Spring 5-5-2023

Investigating the Impact of Prophages on Bacterial Fitness of Streptococcus agalactiae

Caitlin Wiafe-Kwakye
University of Maine, caitlin.tetteh@maine.edu

INVESTIGATING THE IMPACT OF PROPHAGES ON BACTERIAL FITNESS OF *STREPTOCOCCUS AGALACTIAE*

By

Caitlin Selassie Naa Sarku Wiafe-Kwakye

B. Sc. Kwame Nkrumah University of Science and Technology, Ghana, 2011

M.Phil. University of Ghana, Ghana, 2015

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Microbiology)

The Graduate School

The University of Maine

May 2023

Advisory Committee:

Melody N. Neely, Associate Professor of Molecular and Biomedical Sciences, Advisor

Sally D. Molloy, Associate Professor of Genomics, co-advisor

Joshua Kelley, Associate Professor of Biochemistry

Benjamin King, Associate Professor of Bioinformatics

Deborah Bouchard, Associate Professor of Aquatic Animal Health

Keith Hutchison, Professor Emeritus of Biochemistry

INVESTIGATING THE IMPACT OF PROPHAGES ON BACTERIAL FITNESS OF *STREPTOCOCCUS AGALACTIAE*

By Caitlin Selassie Naa Sarku Wiafe-Kwakye

Dissertation advisors: Dr. Melody N. Neely, Dr. Sally D. Molloy

An Abstract of the Dissertation Presented

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

(in Microbiology)

May 2023

Streptococcus agalactiae (Group B Streptococcus or GBS) is a common bacterium found in pregnant women that can cause severe infections in neonates. Although detecting maternal colonization and administering antibiotics during labor can prevent early-onset GBS disease in neonates, antibiotics negatively affect newborns' microbiota, leading to complications like gastrointestinal disorders and immune system dysregulation. Therefore, alternative therapeutic measures are necessary to improve maternal and neonatal outcomes. Understanding GBS disease pathology and developing effective preventive measures and treatments is essential.

GBS evolves from a commensal bacterium to an invasive disease-causing pathogen using various mechanisms, such as adapting to the host immune response, utilizing virulence factors like surface proteins, and regulating gene expression. The GBS genome contains mobile genetic components, including prophages, plasmids, insertion sequences, and transposons, that facilitate mutations and lateral gene transfer. This adaptability allows GBS to develop new virulence factors and antibiotic resistance, enhancing its ability to cause disease and evade host defenses.

Prophages, viral genomes that are integrated into bacterial genomes, may play a critical role in GBS evolution, and understanding their contribution to its virulence could lead to innovative treatments. Bioinformatic analysis of 49 clinical isolates of GBS identified 42 prophages present in their genomes, which can be classified into 5 clusters based on their genomic content, indicating differences in the genetic makeup of the prophages. Further investigation of a hypervirulent GBS strain, found that the only prophage present provides a competitive advantage to the bacterium, possibly by enabling it to better compete for nutrients or resist other bacterial species. The absence of the prophage leads to a metabolic shift, indicating its significant impact on bacterial metabolism and possibly on the pathogen's virulence. Overall, the findings from this dissertation highlight the importance of prophages in GBS pathogenesis and emphasize the need for further research to develop novel therapeutic approaches for the prevention and treatment of GBS infections.

DEDICATION

I dedicate this to my beloved husband, whose unwavering support has been the cornerstone of my research career, and to my two precious children, whose miraculous arrival filled my journey toward earning a Ph.D. with joy and inspiration

ACKNOWLEDGMENTS

I am deeply grateful to the many individuals at the University of Maine who have supported me during my graduate career, particularly in the challenging transition from my home country to this new academic environment. I would like to extend my heartfelt appreciation to the professors and fellow graduate students who provided invaluable guidance and encouragement along the way.

In particular, I would like to express my special thanks to Atefeh Rajaei and Katie Patenaude, former and present graduate members of the Neely lab, for making my graduate experience especially enjoyable. Their camaraderie and support have been a constant source of inspiration.

I would also like to acknowledge the contributions of past and present undergraduate students, especially Amelia, Andrew, Katie, Hannah, and Brandon. Their hard work and dedication have been essential to the research presented in this dissertation.

I am indebted to Dr. Melissa Maginnis, my unofficial mentor, for her invaluable assistance in my professional development and teaching endeavors. Her guidance and support have been instrumental in shaping my career trajectory.

My deepest appreciation goes to my committee members, with a special mention to Keith Hutchison, for their invaluable guidance and expertise in navigating the various aspects of my dissertation.

Finally, I would like to extend my heartfelt gratitude to my mentors, Dr. Melody Neely and Dr. Sally Molloy, for their unwavering support, encouragement, and mentorship. Dr. Melody Neely's calm and steady presence and Dr. Sally Molloy's emphasis on the art of storytelling in science have greatly influenced my development as a researcher, mentor, and teacher. Above all, they have supported me wholeheartedly during the challenging times in my personal life. I cannot thank them enough for their exceptional mentorship and unwavering support.

TABLE OF CONTENTS

DEDICATION……ii
ACKNOWLEDGMENTS……III

INTRODUCTION……1

1.1 Bacterial infections in newborns: Meningitis and Sepsis……1

1.2 Streptococcus agalactiae……2

1.2.1 Microbiological characteristics of Streptococcus agalactiae……2

1.2.2 GBS epidemiological history, colonization, and disease……3

1.2.2.1 History of GBS disease……3

1.2.2.2 GBS colonization and invasive disease……4

1.2.2.3 Disease burden and incidence of GBS disease……5

1.2.3 Treatment and prevention strategies for GBS disease……7

1.2.4 Factors that influence GBS transition from colonization to invasive disease……9

1.2.4.1 Metabolic adaptation of GBS to its environment……10

1.2.4.2 GBS virulence factors……11

1.2.5 Role of evolution in GBS pathogenesis……12

1.3 Bacteriophages……13

1.3.1 Life cycle of bacteriophages……14

1.3.1.1 Lytic cycle……15

1.3.1.2 Lysogenic cycle……16

1.3.1.3 Lytic-Lysogeny decision……17

1.3.2 Phage diversity……17

1.3.3 Prophage impact on bacterial fitness and virulence……19

1.4 GBS and prophages……20

1.5 Summary 23

.. **COMPARATIVE GENOMIC ANALYSIS OF PROPHAGES IN CLINICAL ISOLATES OF STREPTOCOCCUS AGALACTIAE 24**

2.1 Introduction 24

2.2 Materials and Methods 26

2.2.1 Bacterial strains and DNA isolation 26

2.2.2 Genome sequencing, assembly, and prophage isolation 27

2.2.3 GBS prophage database creation and genome clustering 28

2.2.4 Genomic analysis of prophages 28

2.2.5 Data availability 28

2.3 Results 29

2.3.1 Identification of prophages in GBS clinical isolates 29

2.3.2 Prophages in GBS are diverse but show no correlation with serotypes or clonal complexes. 31

2.3.3 Prophages integrate within specific regions of their streptococcal host genome. 33

2.3.4 Prx is encoded on the bacterial chromosome and often in GBS prophages. 42

2.3.5 GBS prophages encode a gene upstream of *prx* with holin-like and transmembrane domains. 45

2.4 Discussion 46

THE PROPHAGE OF THE GBS STRAIN, CNCTC 10/84 PROVIDES A COMPETITIVE ADVANTAGE TO THE LYSOGEN 50

3.1 Introduction 50

3.2 Materials and Methods 52

3.2.1 Bacterial strains, plasmids, and growth conditions 52
3.2.2 Generation of a prophage-cured CNCTC 10/84 strain 52
3.2.3 Phenotypic assays 53
3.2.4 Growth curve and competition assay 54
3.2.5 Virulence assays using a zebrafish infection model. 54
3.2.6 Whole-genome RNA seq transcriptomic analysis 55
3.2.7 Bioinformatic analysis of prophage region 56
3.2.8 Phage induction, propagation, and TEM microscopy 57
3.2.9 Statistical analyses 57
3.3 Results 58
3.3.1 Prophage loss influences the cell wall integrity of CNCTC 10/84. 58
3.3.2 Prophage provides a competitive advantage to the CNCTC 10/84 lysogen *in vitro*. 59
3.3.3 Gene expression and viability of phage particles suggest an alternative mechanism for the competitive advantage of CNCTC 10/84 lysogen. 63
3.4 Discussion 66
GLOBAL TRANSCRIPTOME CHANGES OF GBS STRAIN, CNCTC 10/84 TO PROPHAGE LOSS 72
4.1 Introduction 72
4.2 Materials and Methods 73
4.2.1 Bacterial strains, plasmids, and growth conditions 73
4.2.2 Generation of a prophage-cured CNCTC 10/84 strain 74
4.2.3 RNA isolation 75

4.2.4 Bioinformatic analysis 76
4.2.5 Data availability 77
4.3 Results 77
4.3.1 Transcriptomic changes in CNCTC 10/84 following prophage loss 77
4.3.2 Functional pathway analysis reveals a potential prophage impact on the transcriptome of CNCTC 10/84. 82
4.3.3 ABC transporter genes are differentially expressed in response to prophage loss 85
4.4 Discussion 88
FUTURE DIRECTIONS AND CONCLUSIONS 92
APPENDICES……..91
BIBLIOGRAPHY 125
BIOGRAPHY OF THE AUTHOR 168

LIST OF TABLES

Table 2.1: Characteristics of GBS prophages…… 36

Table 4.1: PCR primers used in this study…….75

Table A.1: Genome assembly data on DMC strains……103

Table A.2: GBS strains and resident prophages……..104

Table A.3: Prophage gene content similarity matrix……98

Table B.1: Significantly differentially expressed GBS genes in the absence of the prophage…105

Table C.1: Filtrates used in phage induction experiments……111

Table C.2: Phage lysates of GBS induction strains on GBS DMC indicator strains……112

Table D.1: Prophage distribution in GBS sequences from Genbank…….115

LIST OF FIGURES

Figure 1.1: GBS invasive disease in neonates…….....5

Figure 1.2: Routes of transmission of GBS in neonates…… ……..9

Figure 1.3: Life cycle of bacteriophages……15

Figure 2.1: Prophage distribution…….. ……30

Figure 2.2: Prophage diversity and genome organization…….32

Figure 2.3: Comparative analysis between GBS prophages……..34

Figure 2.4: Prophage diversity across clusters…….……..35

Figure 2.5: Prophage insertion sites……...……39

Figure 2.6: Identified genes that may contribute to bacterial fitness……41

Figure 2.7: Paratox is highly conserved across clusters……43

Figure 2.8: Paratox is also found on the bacterial host chromosome ……44

Figure 3.1: Prophage loss influences cell wall integrity in CNCTC 10/84……...61

Figure 3.2: Prophages may provide a competitive advantage to CNCTC 10/84……...62

Figure 3.3: Transcription profile of CNCTC 10/84 prophage at different growth phases……66

Figure 3.4 Schematic of possible events supporting prophage-encoded advantage to CNCTC 10/84…….71

Figure 4.1: Quantitative differences in gene expression between lysogen (WT) and non-lysogen (PC) at different growth phases……80

Figure 4.2: Differential expression of bacterial genes in CNCTC 10/84…….81

Figure 4.3: GO and KEGG pathway enrichment analysis……84

Figure 4.4: Simplified representation of genes within KEGG pathway……86

Figure 4.5: Differentially expressed genes involved in ABC transport…….88

Figure 4.6: Model of bacterial gene expression in CNCTC 10/84……91

Figure A.1: Genome comparisons of Cluster A prophages…….100

Figure A.2: Genome comparisons of Cluster B prophages…….101

Figure A.3: Genome comparisons of Cluster C prophages…….102

Figure A.4: Genome comparisons of Cluster D prophages…….103

Figure A.5: Genome comparisons of Cluster E prophages……..104

Figure D.1: mRNA quantity of genes involved T7SS…….120

CHAPTER 1

INTRODUCTION

1.1 Bacterial infections in newborns: Meningitis and Sepsis

Neonatal infections are a significant cause of death worldwide, responsible for approximately 550,000 neonatal deaths annually (1–3). Newborns are at high risk of developing severe infections due to their immature immune systems, which are influenced by genetic, epigenetic, and environmental factors (4–6). This susceptibility is particularly pronounced during the first 28 days of life when their immune systems are still developing (7, 8). Maternal antibodies are transferred to the fetus during pregnancy to protect against infections until the baby's own immune system is strong enough (9). However, the innate immune system, which is the first line of defense against pathogens, is not fully developed at birth, and the adaptive immune system develops during the first years of life (10). Newborns can become exposed to pathogens before birth, during delivery, or after birth, which increases the risk of infection, especially because a newborn's immune system is not yet fully developed (5, 11). Various factors, such as prenatal exposure to environmental pollutants and exposure to vaginal bacteria during delivery, can increase the risk and severity of infections in newborns (12–16).

Neonatal infections can be caused by bacteria, viruses, or fungi, with bacteria being the most common culprit (17–19). Bacterial infections are confirmed through the culture of normally sterile bodily fluids like blood or cerebrospinal fluid. The World Health Organization recommends early identification and treatment of newborns showing signs of serious bacterial infection since it is crucial in reducing mortality and morbidity rates (20). Left untreated,

bacterial infections in neonates can result in long-term issues such as hearing loss, seizures, neurodevelopmental impairment, or death (21, 22). Bacterial infections can be categorized as early onset (within 3 days of birth) or late onset (between 3 and 28 days after birth)(22–24). Early-onset infections typically result from pathogen exposure *in utero* or during delivery, while late-onset infections are more commonly contracted from the environment (25).

Neonatal bacterial infections usually result in pneumonia, sepsis, and meningitis (19). Sepsis is a life-threatening blood infection in newborns less than 28 days old (25) and a major cause of morbidity and mortality in newborns (26). In the United States (US), neonatal bacterial sepsis is the sixth most common cause of infant mortality (27). Bacteremia during early-onset sepsis frequently results in meningitis (22). Bacterial meningitis is a condition where the meninges, the protective membranes around the brain and spinal cord, become acutely inflamed after exposure to bacteria (8, 28). Failure to treat this condition during the neonatal period can lead to severe and lasting neurodevelopmental issues such as hearing loss, cerebral palsy, seizures, developmental delays, and cognitive impairment (8, 29). The mortality rate of bacterial meningitis in neonates can be as high as 40% and the morbidity rate between 20% - 60% (14). Neonatal sepsis and meningitis are usually caused by the same organisms with the two most common pathogens being *Escherichia coli* and *Streptococcus agalactiae* (14, 30, 31).

1.2 Streptococcus agalactiae

1.2.1 Microbiological characteristics of Streptococcus agalactiae

Streptococcus agalactiae or Group B *Streptococcus* (GBS), is the leading cause of neonatal sepsis and meningitis in developed countries (32). GBS is a fast-growing Gram-positive

bacterium that can colonize the gastrointestinal, genitourinary, and upper respiratory tract of approximately 30% of healthy adults (33). GBS gets its name from the Lancefield classification system developed in 1933, where Group B refers to the species-specific carbohydrate 'substance C' found in streptococci (34, 35). GBS is a facultative anaerobe, beta-hemolytic, encapsulated, non-motile, non-spore-forming, catalase-negative cocci that typically occur in chains (36).

GBS is serologically grouped based on type-specific antigens present on the polysaccharide capsule and is currently divided into 10 major serotypes (Ia, Ib, II-IX) (36, 37). In the US and Europe, five serotypes (Ia, Ib, II, III, and V) are associated with human GBS disease (38). Worldwide, serotypes Ia, III, and V are implicated in maternal colonization and disease (39). Serotype III is most commonly associated with neonatal infections, while serotype V is most common in adult infections (39, 40).

Multi-locus sequencing typing (MLST) is another way to classify GBS based on the sequencing of 500-bp fragments of 7 housekeeping genes (41). Strains are assigned to a sequence type (ST) based on their allelic profile, and GBS is grouped into clonal complexes (CC) based on their ST if they share at least five of the seven MLST loci (42, 43). Some CCs are associated with invasive diseases, while others are predominantly colonizers of pregnant women (44, 45).

1.2.2 GBS epidemiological history, colonization, and disease

1.2.2.1 History of GBS disease

GBS was first identified as a cause of mastitis in dairy cows by Nocard and Mollereau in 1887. Because mastitis resulted in reduced milk production (46, 47), the species was named

Streptococcus agalactiae after the Greek words for "no milk" (a- meaning no; and galactos, meaning milk). In 1934, Lancefield and Hare discovered GBS in vaginal swabs from pregnant women, where it was initially associated with low-grade infections (48). In 1968, GBS was linked to neonatal meningitis in Great Britain by Jones and Howells in the first published report of this association (49). By the early 1970s, GBS had become the primary cause of neonatal sepsis and meningitis in the US (50–52), as well as in several other developed countries by the 1980s (53–55).

1.2.2.2 GBS colonization and invasive disease

GBS is primarily found in the outer mucus layer of the colon, although it can also be present in the small intestine on occasion (56, 57). In developed countries, approximately 20-30% of pregnant women are colonized with GBS (58, 59), which can increase the risk of premature delivery and vertical transmission to neonates (60).

Without preventive interventions, 1-2% of neonates born to mothers with GBS colonization may develop early-onset GBS infection (61), and exposure to GBS during the neonatal period may increase the risk of invasive disease (30, 37, 62, 63). While about 1 in 10 newborns may be temporarily colonized by GBS (17, 64–66), only 1% develop invasive disease (40, 53, 62, 67). Invasive disease can occur when GBS is transmitted to a newborn, the bacteria can replicate in the neonate alveoli and adhere to the respiratory epithelium (68). If pulmonary macrophages are unable to clear the bacteria, it can lead to pneumonia. GBS is able to invade pulmonary epithelial and endothelial cells and enter the bloodstream, causing septicemia. Dissemination in the

bloodstream may lead to a breach of the blood-brain barrier and lead to meningitis (68) (Figure 1.1).

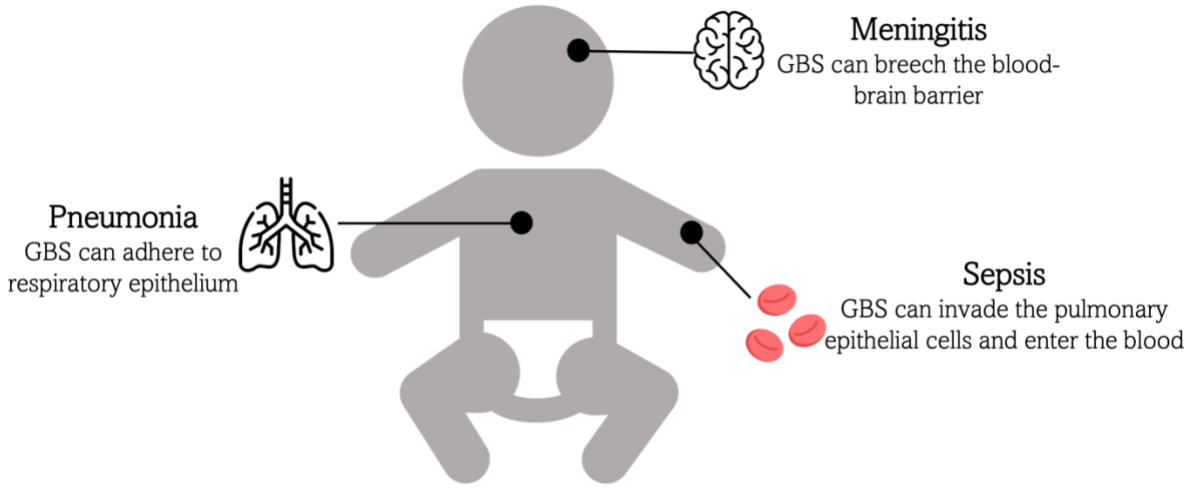


Figure 0.1: GBS invasive disease in neonates.

Neonatal GBS disease is characterized by colonization of the maternal genital tract, invasion of the fetal membranes, and ultimately, infection of the newborn. Once GBS breeches the vaginal epithelium, the bacterium can evade the epithelial barrier of the fetus and adhere to the respiratory epithelium. Evasion of the mucosal immune cells leads to pneumonia (1). GBS can invade the pulmonary epithelial cells and enter the bloodstream resulting in septicemia (2). Breeching the blood-brain barrier leads to meningitis (3)

1.2.2.3 Disease burden and incidence of GBS disease

GBS often causes serious infections in vulnerable populations, such as newborns, pregnant women, and adults with weakened immune systems. In pregnant women, GBS infections can cause chorioamnionitis, sepsis, and endometritis, which can lead to preterm labor, stillbirth, and maternal death (37, 69). In adults with weakened immune systems, GBS infections can cause severe infections such as bloodstream infections, pneumonia, and meningitis (70). While GBS invasive disease typically affects older adults aged 65 and above, especially those who are immunocompromised due to underlying illnesses like diabetes, malignancy, and chronic lung or

kidney diseases (71), neonates bear the highest incidence of GBS disease. The overall prevalence of GBS colonization in pregnant women is estimated to be around 18% globally, with higher rates in some regions such as the Caribbean (34.7%) and Southern Africa (28.9%) (72, 73). In newborns, the incidence of early-onset GBS disease, onset within the first six days of life, is estimated to be around 0.53-1.11 cases per 1,000 live births in the United States and Europe (30) while the incidence of late-onset GBS disease, onset from 7 through 89 days of life, is around 0.32-0.47 cases per 1,000 live births (74).

The incidence of GBS disease in vulnerable populations, particularly newborns and pregnant women, can have significant implications for morbidity and mortality rates (75). These populations are at higher risk for severe infections that can lead to long-term health complications and even death (76). For instance, GBS infections in newborns can cause conditions such as cerebral palsy, hearing and vision loss, and intellectual disabilities, which can persist throughout their lives. In pregnant women, GBS infections can lead to preterm labor, stillbirth, and maternal death. Long-term morbidity associated with GBS infection in babies includes chronic neurological deficits, developmental delays, and behavioral problems that can persist throughout their lives (77–80). Additionally, mortality rates due to GBS disease can reach 5-10%, especially in developing countries with limited access to healthcare (72, 81) . Therefore, prevention and treatment strategies targeting vulnerable populations such as newborns, pregnant women, and immunocompromised adults are crucial to reducing morbidity and mortality rates. Fortunately, various options are available, including screening and prophylaxis for pregnant women and antibiotics for infected individuals.

1.2.3 Treatment and prevention strategies for GBS disease

Universal screening of GBS during pregnancy is a crucial preventive strategy against neonatal infections (82–84). The procedure involves obtaining rectovaginal culture swabs between 35 and 37 weeks of gestation to identify women colonized with GBS, who require intravenous antibiotic prophylaxis during labor or if membranes have ruptured before labor begins (85–87). However, compliance with GBS screening protocols varies across countries and healthcare providers due to various factors such as financial constraints and high rates of Cesarean sections (82, 88–90). Educational interventions have been found effective in promoting compliance with these guidelines. For instance, interactive education programs, electronic reminder systems, and multifaceted intervention strategies have proven useful in improving compliance rates (91).

Intrapartum antibiotic prophylaxis (IAP) is recommended for pregnant women who are colonized with GBS to reduce the risk of transmitting the bacteria to their newborns during labor and delivery (92). This prophylaxis typically consists of intravenous antibiotics administered during labor or if membranes rupture before labor begins. Antibiotic prophylaxis has been shown to significantly reduce the incidence of neonatal GBS infections and associated complications (11). In addition to intravenous antibiotic prophylaxis during labor, antibiotic treatment is recommended for pregnant women with symptomatic GBS infections, such as urinary tract infections, chorioamnionitis, and bacteremia (93). Timely treatment of these infections is essential to prevent severe complications in both the mother and the baby (94). Newborns testing positive for GBS should receive immediate antibiotic treatment to prevent early-onset neonatal sepsis. Close monitoring for symptoms such as fever, respiratory distress, poor feeding, and

lethargy is necessary, as prompt treatment is crucial to prevent life-threatening complications such as meningitis, pneumonia, and septicemia.

Despite efforts to prevent and treat GBS infection, it remains a significant public health concern. GBS disease imposes a considerable economic burden that encompasses medical treatment expenses as well as the long-term consequences of disability and mortality. Hospitalization costs for newborns afflicted with GBS in the US are estimated at over $500 million annually (95), while preventive measures such as screening and prophylaxis in expectant mothers entail significant financial outlays (96). The impact of GBS disease on developing countries is even more severe due to limited healthcare access (72, 97). Additionally, while antibiotics are the most commonly used treatment for GBS, their use raises concerns about antibiotic resistance and their impact on public health in the long term. Overuse of antibiotics during labor can contribute to antibiotic resistance and alter the microbiota of newborns leading to potential long-term effects on their health such as gastrointestinal disorders and immune system dysregulation. Therefore, alternative approaches such as the development of vaccines and other prophylactic measures such as maternal immunization have been explored as potential alternatives to reduce the burden of GBS disease. Yet, we need to better understand the disease pathology and work towards developing effective preventive measures and treatments that can improve maternal and neonatal outcomes.

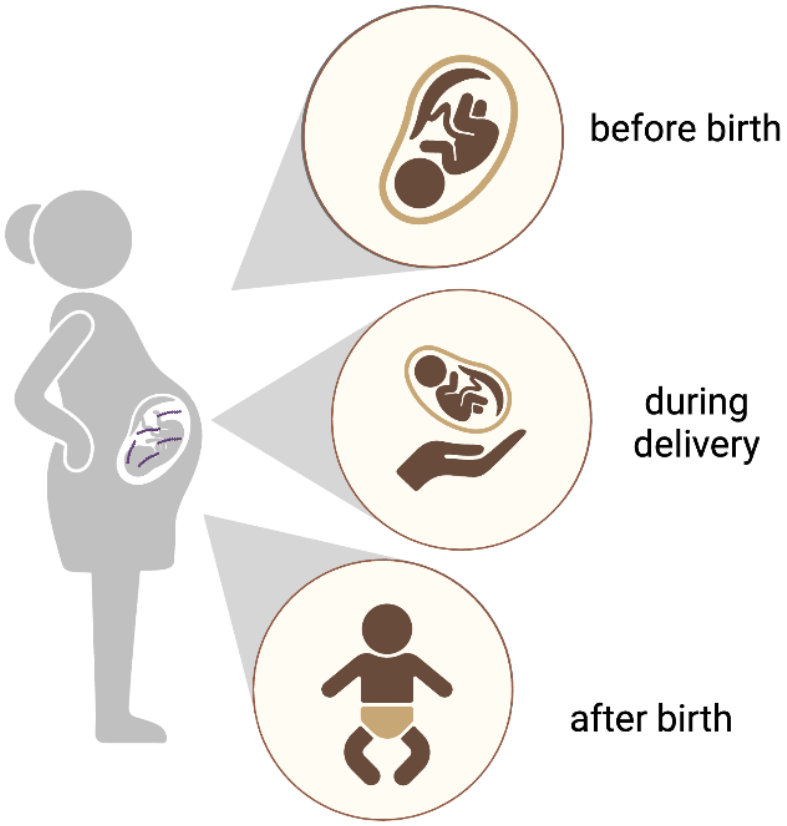


Figure 0.2: Routes of transmission of GBS in neonates

GBS can be transmitted from the mother to the neonate before birth through ascending infection, during delivery as the baby passes through the birth canal. Newborns may also acquire the infection after birth through exposure to the environment.

1.2.4 Factors that influence GBS transition from colonization to invasive disease

Newborns are susceptible to GBS mostly due to immature neonatal phagocytic function, humoral immunity, cell-mediated immunity, or from lack of passively acquired maternal antibodies (36). GBS disease in neonates is divided into early-onset disease (EOD) which occurs during the first week of life and late-onset disease (LOD) manifests between day 7 and 3 months of age (98). Early onset disease is often caused by GBS ascending into the amniotic fluid, multiplying in the baby's respiratory system, and potentially causing pneumonia, sepsis, and meningitis by penetrating the blood-brain barrier (60). The pathogenesis of late-onset GBS disease is not well

understood (75) but it is thought that the bacteria may be transmitted either during childbirth or acquired from contact with environmental sources (99). GBS transforms from a harmless commensal during colonization into an invasive disease-causing pathogen through various mechanisms, such as adapting to the host immune response, utilizing virulence factors like surface proteins, and regulating gene expression during the transition. Understanding these factors is crucial to comprehend the development of GBS-induced diseases.

1.2.4.1 Metabolic adaptation of GBS to its environment

Bacterial replication is essential to colonization and transmission (100), therefore the ability of GBS to process nutrients efficiently greatly influences its survival in the human host. Generally, within the human host, streptococci encounter many microenvironments with unique compositions in terms of the availability of nutrients and trace elements, osmolarity, pH value, oxygen tension, and other variables (101). Therefore, as GBS transitions from a common colonizer to a facultative invasive pathogen, the bacteria must quickly adapt their metabolism to a variety of environments including bacteria-rich and intensely competitive habitats like the respiratory and intestinal mucosa and sterile body sites (i.e., blood, amniotic fluid, or cerebrospinal fluid)(102, 103).

Bacteria are known to frequently use metabolic cues to modulate metabolism and pathogenicity (100). The ability of GBS to adhere to epithelial cells is facilitated by increased acidity within its environment following fermentation (104) and may facilitate colonization by promoting bacterial attachment (105). When GBS is grown in high glucose conditions, the central metabolism of the bacteria adapts to the availability of new nutrients (106). Similarly, in the amniotic fluid, an extremely low-nutrient environment, the majority of changes in GBS transcriptome were

observed to be genes involved in basic bacterial metabolism likely in response to the low amount of nutrients present (107). Examples of the genes expressed by GBS in response to low-nutrient environments include those involved in amino acid biosynthesis, iron acquisition and metabolism, utilization of alternative carbon sources, and genes involved in oxidative stress response (107). These changes in gene expression enable GBS to adapt and survive in low-nutrient environments such as the amniotic fluid, thereby facilitating colonization and contributing to the pathogenesis of invasive GBS infections.

Therefore, the ability of GBS to adapt and survive in low-nutrient environments such as the amniotic fluid is essential for its colonization because it allows the bacteria to modulate its metabolism and pathogenicity in response to changes in nutrient availability, acidity, and oxidative stress. This helps the bacteria to adhere to epithelial cells, colonize new environments and contribute to the pathogenesis of invasive GBS infections.

1.2.4.2 GBS virulence factors

GBS usually colonizes the outer mucus layer of the colon but can also be found in the small intestine (56, 57). To cause an infection, GBS must cross the intestinal epithelium and avoid the immune system (108). This process can be facilitated by several virulence factors such as extracellular surface proteins, capsule polysaccharides, and hemolysin (109). These factors have been shown to increase GBS adhesion, invasion, and evasion of host defenses by inhibiting phagocytic killing and complement-mediated lysis. GBS also has protective factors that help it avoid being attacked by the immune system. All these factors are tightly controlled, and GBS can adapt to different environments using a system called CovS/CovR (110–112). This system allows GBS to sense changes in the environment and regulate its virulence, thereby enabling it to

adapt to different host niches. This ability to adapt and regulate its virulence genes contributes to GBS's success in causing various infections in humans.

1.2.5 Role of evolution in GBS pathogenesis

The genome of GBS is not only composed of chromosomal virulence factors but also includes a wide range of adaptable genetic elements, which enhance its ability to cause disease and survive in different environments. These elements include many mobile genetic components, such as prophages, plasmids, insertion sequences, and transposons, that facilitate mutations and lateral gene transfer (113, 114). As a result, GBS is capable of rapidly evolving and developing new virulence factors, such as capsular polysaccharides that resist complement-mediated killing (115), or antibiotic-resistance genes that provide a selective advantage in antibiotic-rich environments (116). This adaptability of the GBS genome is crucial for its ability to cause disease and evade host defenses.

Lateral gene transfer (LGT) plays a pivotal role in driving GBS genome plasticity by allowing for the exchange of genetic material between organisms through transformation, conjugation, or transduction (117). These mechanisms allow for rapid adaptation to new environments by acquiring virulence factors or antibiotic-resistance genes. Transformation, for example, involves the uptake of DNA from the environment and its incorporation into the bacterial genome. This process has been shown to occur in GBS, as it can acquire antibiotic-resistance genes from other streptococcal species through transformation. Conjugation, on the other hand, involves the transfer of DNA via direct cell-to-cell contact, typically mediated by plasmids or conjugative transposons. In GBS, experimental and *in silico* approaches have shown that large genomic

segments can be exchanged via conjugation between GBS strains (118, 119). For example, conjugation facilitated the acquisition of pilus island 2b (PI-2b), which increased virulence in neonatal invasive disease caused by GBS strains (108). Finally, transduction involves the transfer of genetic material via bacteriophages. Bacteriophages are also major drivers in GBS evolution, and their integration into the GBS genome can confer new traits that enhance virulence or promote survival in diverse environments (120, 121). For example, bacteriophages have been shown to transfer antibiotic-resistance genes between GBS strains (122). It is essential to comprehend how bacteriophages contribute to GBS genome plasticity and pathogenesis to devise effective strategies for controlling and preventing GBS infections.

1.3 Bacteriophages

Bacteriophages, also known as phages, are believed to be the most abundant biological entities on Earth, estimated to be ten times more abundant than bacteria (123). These viruses were first identified in 1915 by Frank W. Twort, who observed 'transparent glassy colonies' that appeared to thrive on micrococci cultures but did not grow alone (124). It was not until 1917 that phages were recognized as viruses that infect bacteria (125), and the term 'bacteriophage', meaning 'bacteria-eater', was coined. Similar to other viruses, phages are small obligate intracellular parasites that cannot grow outside of living cells (126). They can only reproduce, multiply, and spread by infecting a host (127). A remarkable aspect of phages is their diversity, with highly variable virion size, morphology, and genome content (128, 129). Phages have been applied in a variety of fields, encompassing environmental research, veterinary medicine, biotechnology, and agriculture (130). Phages have emerged as a popular remedy for combating infectious diseases caused by antibiotic-resistant bacteria, owing to their notable lytic activity against bacterial

strains (131). In contrast to conventional antibiotics, which can indiscriminately damage other microorganisms or human cells during treatment, phages exhibit a high degree of specificity and selectively target only the host bacteria they infect (132, 133). Thus, phage therapy is being increasingly considered a feasible alternative to traditional antibiotic treatments, particularly in the context of multidrug-resistant bacterial infections such as methicillin-resistant *Staphylococcus aureus* (MRSA), due to its potential efficacy (134). Phages benefit bacteria by promoting genetic transfer, enhancing evolution, increasing diversity, and aiding adaptation in new or changing environments which affects ecosystems (135). To understand their impact on bacterial populations and prevent infections, it is necessary to study their life cycle, extensive diversity, and their role in bacterial evolution and adaptation.

1.3.1 Life cycle of bacteriophages

Phages are studied extensively through research on *Escherichia coli* and lambda phages, which forms the foundation of our knowledge (136). This research has significantly influenced our understanding of the phage life cycle and its relationship with bacterial hosts. The life cycle of phages begins with the attachment of the phage to its host cell through specific receptor recognition(137). The DNA is injected into the cell where it circularizes. Based on the condition of the bacterial cell, a decision is made regarding entry into either lytic or lysogenic phase.

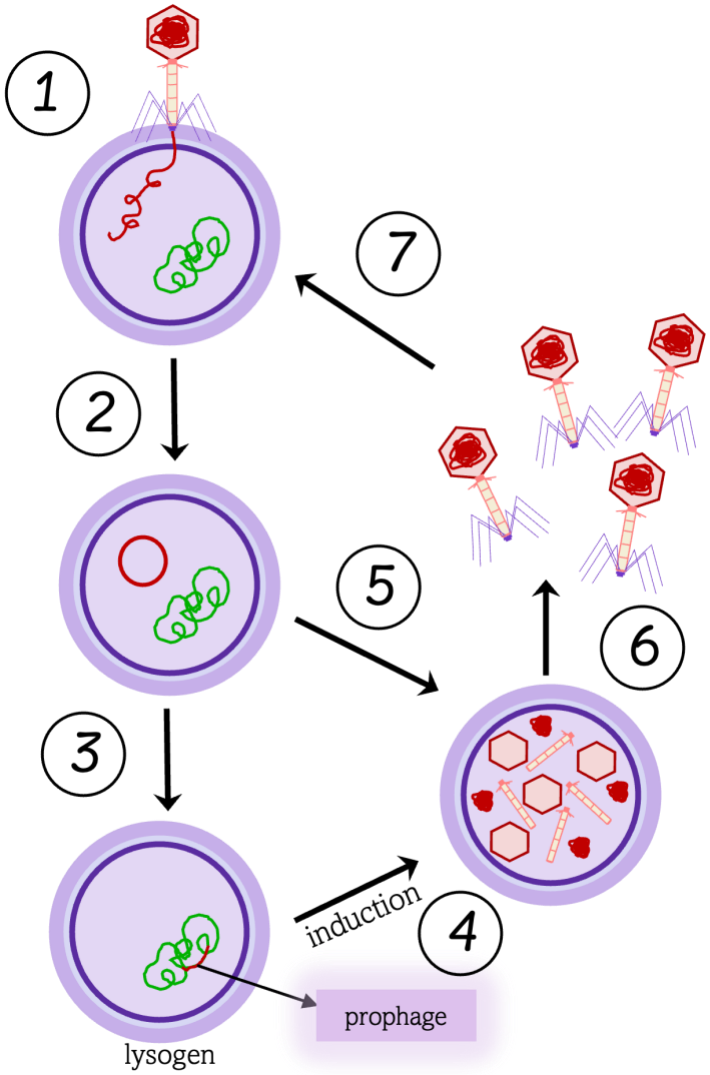


Figure 0.3: Life cycle of bacteriophages

Bacteriophages undergo a complex life cycle consisting of two main stages: the lytic cycle and the lysogenic cycle. The life cycle begins with the phage recognizing and binding to a specific receptor on the bacterial cell. The phage injects its DNA into the bacterial cell (1) and the DNA circularizes (2). Depending on the condition of the bacterial cell, the phage may either enter into the lysogenic phase where it integrates its genome into the bacterial genome. Under stressful conditions, the phage is either induced (4) or enters the lytic cycle (5). The phage then hijacks the bacterial machinery to replicate its DNA and produce new phage particles. Once assembled, the phages are released from the bacterial cell by lysis (6). Newly formed cells can go on to infect new bacterial cells (7).

1.3.1.1 Lytic cycle

In the lytic cycle, phage DNA is replicated, transcribed, and translated using the host cell's machinery, at least initially (138). New phage virions are then assembled and released through lysis of the host cell by proteins encoded by the virus (139). This often results in host cell death. Phage gene expression is carefully regulated to prevent premature lysis and ensure proper progression of this efficient process that maximizes replication while minimizing harm to the bacterial host (140). Lambda virus regulates gene transcription through early, delayed early, and late genes using regulatory proteins like Cro and N (141). This mechanism prevents replication complex pausing and ensures the expression of necessary genes for phage particle formation at different stages of its life cycle to produce mature viruses. During the lytic cycle, phages take advantage of the host cell's resources while avoiding detection by host defense mechanisms (142). This is achieved through a variety of mechanisms, such as the degradation of host cell DNA by phage-encoded nucleases and the modification of phage proteins to evade detection by host cell proteases. Furthermore, the lytic cycle allows phages to propagate in a host population quickly and efficiently.

1.3.1.2 Lysogenic cycle

The lysogenic cycle involves the integration of phage genetic material into the bacterial genome through a site-specific recombination event, resulting in what is known as a prophage. The presence of this prophage turns the infected bacterial cell into a lysogen. During integration, replication of the phage and host genomes occurs concurrently(143). Under stressful conditions, the prophage within a bacterial host can be excised, leading to the initiation of the lytic cycle and eventual lysis of the host cell. This process, known as induction, is triggered by the activation of specific genes within the prophage that lead to its excision from the host DNA(144). The decision to enter into either lytic or lysogenic phase occurs along with the induction process.

Prophage induction occurs either spontaneously, usually at low frequency, or by a wide range of external stressors (145) such as ultraviolet light and antibiotics (mitomycin C). The lysogenic cycle allows phages to persist within a bacterial population by maintaining their genomes as part of the host cell's chromosome. This allows the phage to be passed onto daughter cells during cell division, thereby ensuring its survival in the bacterial population even under non-stressful conditions.

1.3.1.3 Lytic-Lysogeny decision

The lytic versus lysogenic decision is made based on the condition of the bacterial cell. If the bacterial host conditions are favorable for phage replication, then lysis occurs, and the lytic cycle is initiated (146). However, if there are unfavorable conditions such as a lack of nutrients or high levels of stress factors in the bacterial host, then the phage will integrate into the bacterial genome through site-specific recombination, and the lysogenic cycle is initiated (147). This decision is influenced by many factors, including the presence of specific proteins in the phage genome and interactions with host proteins that influence gene expression (148). Moreover, environmental factors, such as nutrient availability and temperature, also play a major role in the activation of the lytic versus lysogenic cycles. Understanding the factors that influence phage decision to become lytic or lysogenic can help in developing strategies for controlling bacterial populations and preventing the spread of infectious diseases.

1.3.2 Phage diversity

Phages exhibit a high degree of genetic diversity, which is dependent on their mode of replication, specific host range, and acquisition of genes through horizontal gene transfer (149). There are two main types of bacteriophages: virulent and temperate phages. Virulent phages

exhibit only the lytic cycle, whereas temperate phages can enter either the lysogenic or lytic phase depending on conditions. Temperate phages can integrate into the host genome and establish a stable and long-term relationship with the host, coexisting for generations without causing any harm to their host cell(150).

Phages have evolved diverse strategies to infect bacteria and replicate their genomes. For example, some bacteriophages use specialized proteins to bind with specific host cell receptors and penetrate the bacterial membrane (151). Once inside the host cell, bacteriophages use various mechanisms to replicate themselves, such as synthesizing their genome using host enzymes and machinery or employing unique transcriptional and translational mechanisms that differ from those of the host cell(152). Furthermore, bacteriophages employ various packaging mechanisms to ensure that their progeny genomes are correctly packaged inside the phage capsid(153, 154).

Phages have very diverse genomes that can range in size from a few kilobases to more than 500 kilobases, and they have both linear and circular genome structures(155, 156). Bacteriophages can exhibit various genomic structures, including double-stranded DNA, single-stranded DNA, double-stranded RNA, and single-stranded RNA genomes. This genomic diversity is thought to be the result of horizontal gene transfer events, such as transduction, transformation, or recombination, that occur between different phages and between phages and their bacterial hosts(153, 157). These events allow phages to acquire new genes, expand their genetic repertoire, and potentially acquire new functions that enhance their fitness in specific environmental conditions. This high degree of diversity ensures that bacteriophages can evolve

and adapt quickly to changing environmental conditions(158). The extensive diversity of phages can have a significant impact on bacterial fitness and virulence, as prophages can introduce new genes into the bacterial genome, altering their phenotype and potentially leading to increased pathogenicity.

1.3.3 Prophage impact on bacterial fitness and virulence

Bacterial-phage interactions are ubiquitous in the biosphere, (159), with temperate phages having a significant impact on bacterial evolution (127). Prophages have developed a mutually beneficial relationship with bacteria, enhancing their colonization, adaptation, and ecological fitness (32, 45, 157, 160). This relationship increases the opportunity for pathogenic bacteria to disseminate and cause infection.

Most pathogenic bacteria carry prophages that encode genes that improve pathogen fitness and virulence through the process of lysogenic conversion (141, 157, 161–166). Lysogenic conversion occurs when the prophage carries genes that provide a selective advantage to the host bacterium, such as genes that produce toxins, antibiotic resistance, or other virulence factors Prophages can regulate virulence genes by producing toxins such as the cholera toxin produced by the CTX phage of *Vibrio cholera* (167), the diphtheria toxin encoded by *Corynebacterium diphtheriae* phages, the Shiga toxin of *Shigella dysenteriae* phages, and the botulinum toxin encoded by *Clostridium botulinum* prophages. These toxins can cause severe illness and are a major public health concern. Prophages can also facilitate the exchange of genetic information among bacteria, enabling them to adapt to challenging environments(160) including the transfer of antibiotic-resistance genes (168). Additionally, prophages can alter the structure of bacterial

genomes through genomic rearrangements or horizontal transfer of genes (157, 169). Prophages may even contribute to bacterial fitness by disrupting genes that lead to an increase in virulence or fitness loss (157, 170, 171)

Phage diversity plays a crucial role in prophage impact on bacterial fitness and virulence. The ability of phages to acquire genes through horizontal gene transfer leads to the emergence of new virulent strains, which can cause epidemics and pandemics(168). Phage genetics and gene content determine the specificity of phages and their interaction with bacteria (172). For instance, *Streptococcus pyogenes* strains are usually poly-lysogenic with prophages constituting about 10% of the bacterial genome(173) and the presence of multiple prophages that encode virulence factors in *S. pyogenes* is thought to contribute to bacterial fitness during colonization, infection, and evasion of the immune system of the human host (174). Experimental evidence indicates that these virulence factors are most often released following lytic induction of *S. pyogenes* prophages (157). Similarly, about 10% of GBS genes that are specific to each strain are associated with phages(114). These phages may carry virulence factors and antibiotic-resistance genes, which may contribute to the pathogenicity of GBS strains.

1.4 GBS and prophages

The first GBS phage was isolated from a bovine source in 1969 (175) giving evidence of the ability of phages to infect GBS. Phages isolated from GBS have been useful in epidemiological investigation and treatment of GBS infections. In the 1980s, bacteriophage typing of epidemiological strains of GBS was used with serotyping to group clinical isolates(176–180). A lysin purified from a GBS phage demonstrated the potential to reduce GBS colonizing the vagina

and oropharynx of mice (181). However, it was not until recently that the impact of lysogeny on GBS disease has been studied. Sequencing of GBS genomes revealed that phage-associated genes account for 10% of strain-specific genes (114, 182). Using randomly amplified polymorphic DNA (RAPD) analysis, van MeeMarquet *et al*. showed that cloned and sequenced prophage DNA fragments from GBS isolates representing major virulence groups were associated with populations of GBS strains able to invade the CNS of neonates (183). In one instance, the sequenced fragments were found in isolates that were 15 years apart suggesting that these prophagic fragments have remained stable over time. Based on their work, it can be hypothesized that GBS may either maintain prophage genes important for virulence and/or fitness or complete prophages have evolved with GBS. In another study, 33 phages induced from GBS lysogenic strains suggest a high level of genetic diversity of prophages (120) and this presents an opportunity to understand their effect on the lifestyle of GBS. The authors demonstrated that lysogenic GBS strains belonging to CC12 and CC19 had lost some catabolic functions, and this may be important to understanding the role of lysogeny in GBS. This is because the loss of catabolic functions in other species has been shown to induce changes in the expression of virulence (157, 171, 184–187). Lysogeny in GBS isolates has been suggested to differ based on prophage group and bacterial clonal lineage (121, 188–190) GBS strains belonging to two clonal lineages, CC1 and CC23, possessed two different profiles of prophagic DNA fragments belonging to two distinct prophage groups. Since both lineages are highly implicated in GBS infections of the skin, bone and joints, these prophage groups are thought to been involved in the evolution of these clones and contributed to their tropism for skin, bone, or joints (190). Similarly, in a large collection of GBS isolates, prophages within a specific group were associated with specific clinical disease. For example, CC17, a lineage commonly

associated with neonatal infections, were found to carry group C prophages while group A prophages seemed to be associated with CC1, a lineage involved in emerging infections in adults (188).

Work to date demonstrates that most genomes of GBS contain one or more prophages, however, their role in bacterial fitness and virulence has not yet been determined experimentally. The majority of the studies on GBS prophages have been focused on the prevalence of prophages in GBS strains in relation to clonal lineages. While clonal lineages have been linked to invasive disease, specific GBS capsular serotypes have also been linked to clinical disease and virulence. As a result, the prevalence of prophages in relation to GBS serotypes needs to be investigated. Furthermore, except for a few strains from Slovakia that included screened pregnant women (189), the majority of the clinical strains employed in these investigations have been focused on GBS invasive illness. The prevalence of prophages in colonization is poorly understood (108).

Although multiple studies have identified regions encoding prophage genes or whole prophages in GBS genomes (114, 121, 188, 191, 192), there is no high-quality annotation of the prophage sequence present in GBS strains. High-quality annotation of prophage sequences in GBS strains will allow for the identification of genes that may contribute to bacterial fitness or virulence during lysogeny. Additionally, comparing isogenic strains that differ only in their prophage regions can lead to a better understanding of the biological contribution of a particular phage to specific bacterial phenotypes, the overall contribution of prophages and the cross-regulation between the prophages in poly-lysogenic strains(193). However, to date, there are no studies to

our knowledge that have investigated the impact of lysogeny on GBS bacterial fitness or virulence by comparing isogenic strains that differ in their prophage regions.

1.5 Summary

Neonates and infants are at risk of life-threatening GBS infections; however, administering antibiotics can adversely affect their microbiota. Therefore, there is a pressing need for alternative therapeutic measures. Bacteriophages play an essential role in the evolution of GBS, and comprehending their contribution to its virulence may lead to innovative treatments. While prophages are frequently present in GBS, their significance to pathogen fitness and virulence remains uncertain. My research utilized bioinformatic techniques to comprehensively annotate GBS prophage genomes and identify genes that may support bacterial fitness. In addition, we compared a GBS strain with its isogenic variant lacking prophages to examine differences in bacterial gene expression, virulence, and fitness using RNA sequencing (RNAseq) and a zebrafish virulence model. Our findings provide insight into the role of prophages during the pathogenesis of GBS infection while also highlighting potential targets for novel therapies.

CHAPTER 2

COMPARATIVE GENOMIC ANALYSIS OF PROPHAGES IN CLINICAL ISOLATES OF STREPTOCOCCUS AGALACTIAE

The work presented in this chapter is in preparation for submission.

2.1 Introduction

Streptococcus agalactiae (group B streptococci or GBS) is a commensal pathogen found on the mucus membranes of the intestinal and vaginal tracts in humans (73, 194). Rectovaginal colonization of pregnant women is a major risk factor for neonatal GBS disease (195). Approximately 50% of babies born through vaginal delivery from women colonized with GBS develop life-threatening infections such as meningitis and sepsis (196). Recommended infection management includes antibiotic treatment of newborns after delivery and Intrapartum antibiotic prophylaxis (IAP) for mothers immediately prior to and during delivery. Although commonly used, IAP does not address the risk of infection in-utero nor when the infection presents in babies over 7 days old (197). Additionally, antibiotic treatment of newborns has long-term negative effects on neonatal microbiota affecting not only metabolism and nutrition but also postnatal development of the immune system (198–201). These long-term effects, in addition to the rise in antibiotic and multi-drug resistance in GBS (202) highlight an increasing need for alternative therapeutic approaches to reduce vaginal colonization and treat neonatal infections. Recent studies have uncovered multiple factors that contribute to GBS colonization and virulence, providing new insights into the development of effective treatments and preventive measures.

GBS adherence to the vaginal epithelium is important for successful bacterial colonization (12). This is enhanced by several GBS determinants including the capsular polysaccharide (CPS) of which 10 different serotypes have been described (Ia, Ib, and II-IX). Six of these capsular serotypes (Ia, Ib, II, III, IV, and V) are most commonly associated with disease in humans (42). GBS is also grouped into clonal complexes (CC) based on their sequence type (ST) if they share at least five of the seven MLST (Multi-locus Sequence Typing) loci (42, 203). Some CCs are associated with invasive disease while others are thought to be predominantly colonizers of pregnant women (44, 204).

Prophages, viral genomes integrated into the bacterial chromosome, can enhance bacterial colonization, environmental adaptation, and ecological fitness, increasing the opportunity for pathogenic bacteria to disseminate and cause infection (160, 204). Phage infection dynamics can drive horizontal gene transfer in bacteria allowing them to adapt to challenging environments (15). Prophages can also change the structure of the bacterial genome by functioning as sites for genomic rearrangements or acting as vehicles for the horizontal transfer of bacterial genes (157, 169). Most genomes of GBS contain one or more prophages (120, 188, 189) yet their role in bacterial fitness and virulence has not yet been described.

Here we report the sequences and diversity of GBS prophages found in 49 GBS clinical strains, including their distribution within serotypes and clonal complexes. Analyses reveal GBS prophage genomic organization, clustering of GBS prophages, regions of recombination in the bacterial chromosome, and identify genes that potentially benefit the bacterial host. One such potential beneficial phage protein is paratox, previously identified in prophage genomes of

Streptococcus pyogenes (205) where it is proposed to prevent the uptake of DNA. In contrast to only being encoded on prophage genomes as in *S. pyogenes*, the gene encoding GBS paratox is found to be present on the bacterial genome in every clinical isolate examined, possibly as a remnant of a mobile genetic element. Moreover, an additional copy of the gene encoding paratox can be found on multiple prophage genomes, suggesting the importance of the conservation of this protein to GBS.

2.2 Materials and Methods

2.2.1 Bacterial strains and DNA isolation

Forty-two GBS clinical isolates collected from the vaginal tracts of pregnant women at Detroit Medical Center (203) were used in this study. Genomic DNA was extracted from GBS samples for whole genome sequencing. Overnight cultures grown at 37°C were pelleted and resuspended in 1 mL TE (10 mM Tris, 1 mM EDTA, pH 8.0) with 25 mg ml^{-1} of lysozyme and 5000 U mL^{-1} mutanolysin and incubated at 37°C for 1 h. Cell pellets were then subjected to a freeze-thaw process (-80°C for 5 min, 37°C for 5 min) and resuspended in 800 µl Nuclei Lysis solution (Wizard Genomic DNA purification kit), followed by incubation at 80°C for 5 min to lyse the cells and then cooled at room temperature. After treatment with RNase solution (3 µl of 10 mg ml^{-1}) for 15 min at 37°C, the sample was cooled to room temperature and protein precipitation solution (Wizard Genomic DNA purification kit) was added to the RNase-treated cell lysate. The DNA was precipitated, rehydrated with 100µl of DNA rehydration solution (Wizard Genomic DNA purification kit) and incubated at 65°C for 1 h. The isolated DNA was sent to the Hubbard Center for Genome Studies (HCG) (Durham, NH) for whole genome sequencing. GBS reference clinical isolates used were 2603 V/R (NC_004116.1), 515 (NZ_CP051004), A909(NC_007432),

CJBIII(NZ_CP063198), CNCTC 10/84 (NZ_CP006910), COHI (NZ_HG939456) and NEM316 (NC_004368.1).

2.2.2 Genome sequencing, assembly, and prophage isolation

Sequenced DMC isolates were assembled and annotated using a sequential list of programs (GitHub tutorial https://github.com/Joseph7e/MDIBL-T3-WGS-Tutorial) organized by Kelley Thomas at the HCG. The sequenced raw reads were analyzed using the following workflow on a Linux server. The quality of the raw reads was examined using FastQC v0.11.5 (206) and exported as HTML figures. Low-quality base reads and adaptors were removed using Trimmomatic v0.36 (207) and exported as paired forward, paired reverse, and unpaired forward and reverse FASTQ files. Genome assembly was performed using SPAdes v3.11.0 (208) to assemble the trimmed read files in a *de novo* fashion. The resulting contiguous sequences were quantified and organized by length. The program QUAST (209) was used to assess overall genome structure and ensured contiguity of the assembled reads. Genomic content was assessed with the program BUSCO (210) which examined the contiguous sequences for common single-copy bacterial orthologs. Each sequence was annotated using PROKKA (211), which individually examined DNA coding sequences, rRNA, tRNA, and ncRNA. Ribosomal RNA sequences were compared against the BLAST nucleotide database (212) to confirm samples as *Streptococcus agalactiae*. Read mapping was performed to calculate the coverage of each contig using BWA-MEM (213) and SAMtools (214). The program Blob_tools (215) BLASTed each contig against a complete nucleotide database to create a taxonomy table. The taxonomy table was filtered according to length (>500 bp), GC content (between 30% and 50% GC), coverage (>4), and, in some cases, species identification (*S. agalactiae*). Heavily contaminated samples were rejected. Filtered contigs were parsed against PHASTER (216) for putative prophage

regions. All prophage genomes were manually examined in Geneious Prime 2021.2 (https://www.geneious.com) for defined genome ends.

2.2.3 GBS prophage database creation and genome clustering

To create the GBS prophage database, GenBank flat files of all prophage genomes were submitted to Dr. Steven Cresawn of James Madison University as input files to be uploaded to the Phamerator website (https://phamerator.org). Multiple techniques were employed, including EMBOSS's polydot function (34) for dot plot analysis, FastANI (217) for ANI analysis, and gene content analysis to identify prophage clusters. Direct comparisons of genomes within and across clusters were performed by visualizing genome maps in Phamerator *Streptococcus* database version 1 (218).

2.2.4 Genomic analysis of prophages

Prophages were annotated automatically using GLIMMER v3.02 and GeneMark v2.5 within DNA Master v5.23.6 (http://cobamide2.bio.pitt.edu) and PECAAN (http://pecaan.kbrinsgd.org) (219, 220). Translational starts were predicted manually based on GeneMark.hmm and conservation across homologs in BLAST and putative gene functions were predicted using BLAST, TMHMM, and HHpred (212, 222, 223). Schematic diagrams of bacterial genes flanking the prophage region and genes surrounding the bacterial paratox were produced in Geneious Prime 2021.2 (https://www.geneious.com). Clustal alignment was performed using clustal omega (224). Graphs were generated with Rstudio (225) and Python 3.9.5 using the packages Matplotlib (221), pandas (226), NumPy (227), and seaborn (228). All figures were edited with Inkscape (https://inkscape.org).

2.2.5 Data availability

Contigs of bacterial genomes in this study can be found at the NCBI BioProject under accession number PRJNA888223 (Table A. 2.)

2.3 Results

2.3.1 Identification of prophages in GBS clinical isolates

The distribution of prophages in GBS clinical isolates was determined by sequencing the genomes of 42 strains collected from the vaginal tracts of pregnant women at the Detroit Medical Center (DMC) in Detroit, Michigan. These 42 GBS strains were previously examined for virulence potential (203). High-coverage draft genome sequences of the 42 vaginal clinical isolates were obtained with an average genome size of 2.03 Mbp (Table A.1). Additionally, prophages were identified in the genome sequences of 7 previously published clinical isolates; 2603 V/R (NC_004116.1), 515 (NZ_CP051004), A909 (NC_007432), CJBIII (NZ_CP063198), CNCTC 10/84 (NZ_CP006910), COHI (NZ_HG939456), NEM316 (NC_004368.1) and used as reference strains (Supplementary Table 2). Analysis of the 42 strains in this study confirmed previous reports that serotypes III and V account for half of the strains (203). In addition, 17 sequence types and ten major clonal complexes were identified, as presented in Table A.2.

A total of 75 prophage regions were identified from the vaginal clinical isolates, of which 36 full-length prophages were extracted. Out of the full-length prophages, 80% (28/36) were extracted from a single contig and 20% (7/36) from two contigs in the sequenced vaginal isolates (Table A.2). Prophages extracted from two contigs were manually inspected to ensure that the genome was complete. The remaining prophage regions could not be determine from the assemblies because they spanned multiple contigs. In addition, seven prophages were isolated

from the reference genomes, out of which four had been previously identified (Javan 5 and Javan 6 in 2603V/R, and Javan 7 and Javan 8 in A909) (192). Altogether, a total of 43 prophages were extracted from the 49 GBS clinical isolates analyzed in this study. The prophage genomes identified in the DMC vaginal isolates were designated with phiDMCxx, where xx is based on the bacterial strain from which they were extracted. A numerical suffix indicates the number of prophages in a single bacterial strain. Out of the GBS genomes examined, most (69.4%, 34/49) had only one prophage, while 12.2% (6/49) had two prophages, and 18.3% (9/49) had no prophages present in their bacterial genome (Table A.2). Prophages could not be extracted from four of the 49 bacterial strains. Among the 43 prophages identified, four were found to be identical, resulting in a total of 39 unique prophages.

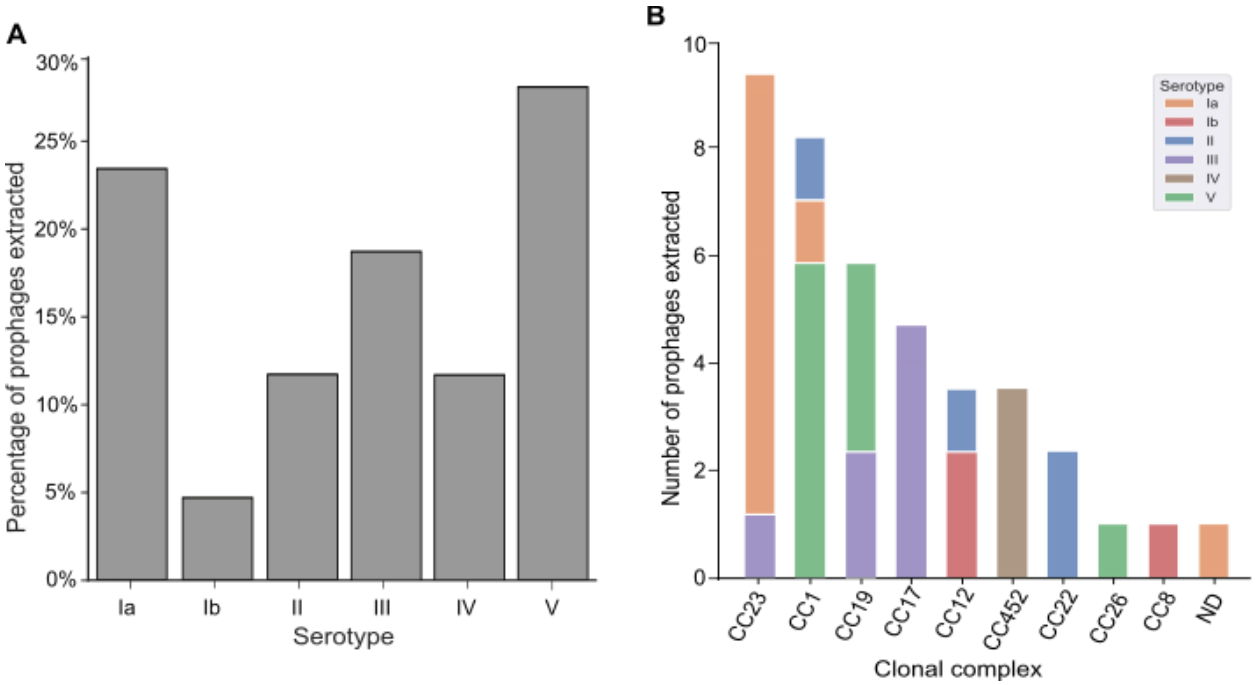


Figure 2.1: Prophage distribution

A. Prophage distribution across GBS serotypes identified in our sample collection. B. Serotype relationship and prophage distribution within individual clonal complexes (CC).

2.3.2 Prophages in GBS are diverse but show no correlation with serotypes or clonal complexes.

GBS strains can produce one of 11 different capsular serotypes (Ia, Ib – X). Specific GBS capsular serotypes have been associated with clinical disease and virulence; therefore, the distribution of prophages across the different serotypes of the GBS vaginal isolates was examined. Serotype V and Ia accounted for 30.2% (13/43) and 23.2% (10/43) of the extracted prophages, respectively (Figure 1A). These serotypes are usually associated with adult infections (70) and account for more than half of the extracted prophages (Figure 1A). Prophages extracted from strains belonging to serotype III, a serotype commonly implicated in neonatal infections (229), constituted 18.6% (8/43) of all prophages with the remaining belonging to serotypes II (11.6%, 5/43), IV (11.6%, 5/43) and Ib (4.6%, 2/43) (Figure 1A).

Clonal complexes (CC) are formed when several sequence types (STs) share six or seven matching alleles, and the number of a particular CC is designated after its ancestor ST or the predominant ST within the clone (230). To determine if there is a relationship between prophage carriage and specific clonal complexes, we evaluated the number of prophages extracted within individual clonal complexes. Over half of the prophages were extracted from CC1 (23.2%, 10/43), CC23 (20.9%, 9/43), and CC19 (13.9%, 6/43) (Figure 1B), clonal complexes common for invasive GBS disease and consistently reported in asymptomatic pregnant women (204).

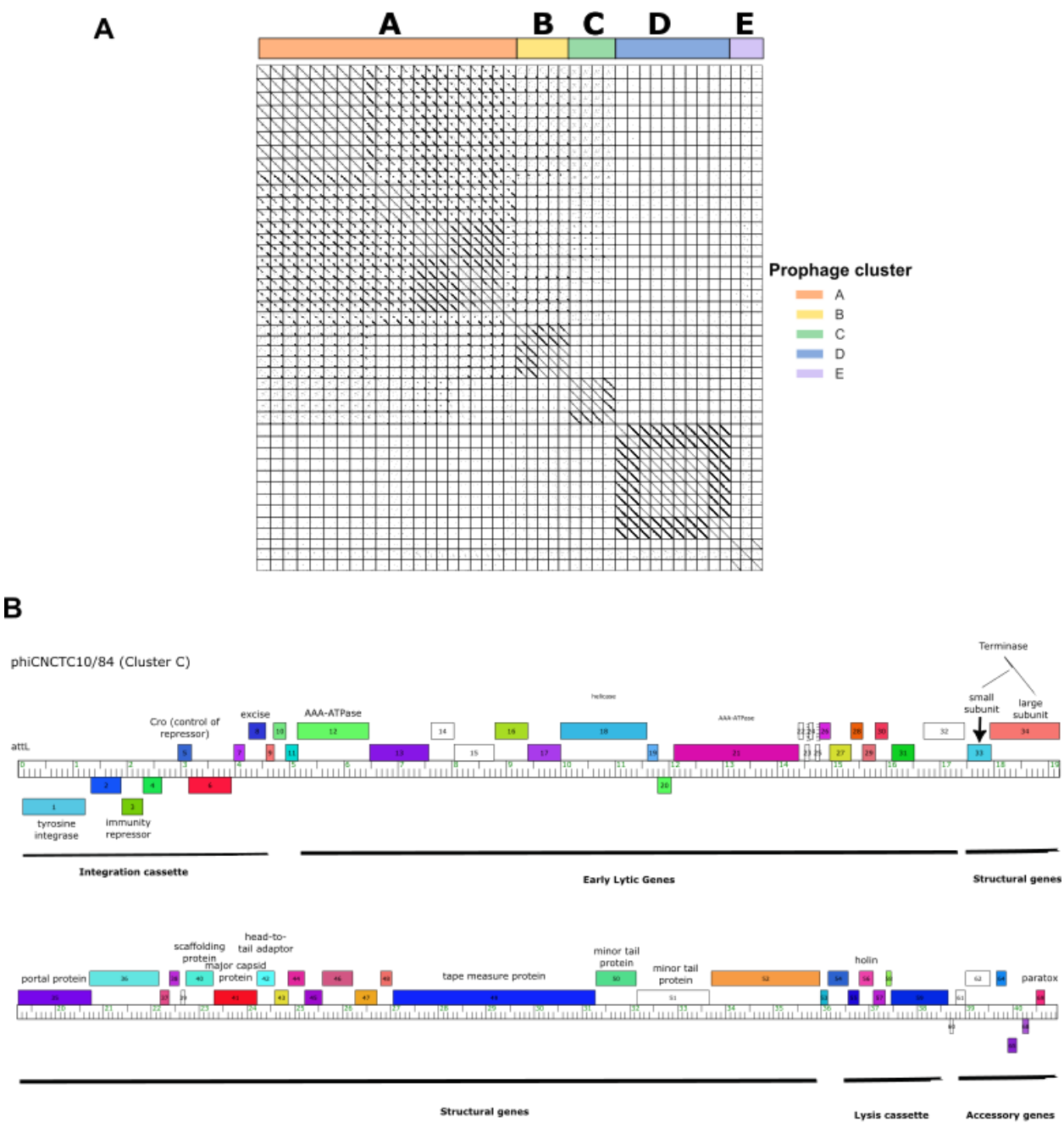


Figure 2.2: Prophage diversity and genome organization

A) Nucleotide sequence comparison of 43 GBS prophages from whole genome sequences concatenated into a single file and compared with itself using polydot (EMBOSS; word size, 15). Dotplot analysis identified five distinct prophage clusters. B) Genome map of Callidus showing genome organization with genes represented as boxes above and below the ruler illustrating genes transcribed in the forward and reverse directions, respectively. The genome coordinates

are represented by the ruler in units of kilobase pairs. Genes are colored according to assigned 'phamilies' with putative gene functions indicated above the genes.

Mosaicism plays a more significant role in the evolution of bacteriophages compared to bacteria (155), therefore further investigation was conducted into the gene arrangement and synteny of GBS prophage genomes. Based on more than 50% nucleotide sequence similarity and shared gene content of over 35%, the GBS prophages can be sorted into five distinct clusters (A–E) (Figure 2A; Table A. 3). The genome sizes of prophages differed across the clusters, with cluster B exhibiting the most consistency at an average size of 36.5kb and a narrow range of variation (600bp) (Figure 3A). Clusters A and C had a similar %GC content to their bacterial host genome (35.6%), whereas clusters B, D, and E had a higher %GC content (Figure 3B). A genome map of selected prophages allowed for visualization of the genetic diversity within and across prophage clusters, revealing significant diversity between clusters (Figure 4). Prophages within each cluster have the same organization, with the left arm encoding the immunity cassette and the early lytic genes and the right arm encoding structural and lysis genes (Figure 2B). Cluster A prophages are the most diverse, with most encoding their own tRNAs. No clear relationship was found between prophage clusters and GBS serotypes or clonal complexes, except for cluster E prophages, which were only found in serotype V strains (Figure 3C; 3D).

2.3.3 Prophages integrate within specific regions of their streptococcal host genome.

Prophages often integrate into tRNA genes and can carry a duplication of the disrupted gene or encode their own tRNA genes. However, integration can sometimes lead to a loss of function and negative effects on the host (231–236). The prophages found in GBS in this study are inserted at nine different locations across the genome, indicating a broad distribution (Figure 5, Table 1). There are two types of integrases - tyrosine and serine - with 25 out of 43 prophages

identified encoding tyrosine integrases (Int-Y) and the remaining 18 encoding serine integrases (Int-S). Most Int-Y prophages use an attB site overlapping host tRNA genes, with one exception (Table 1).

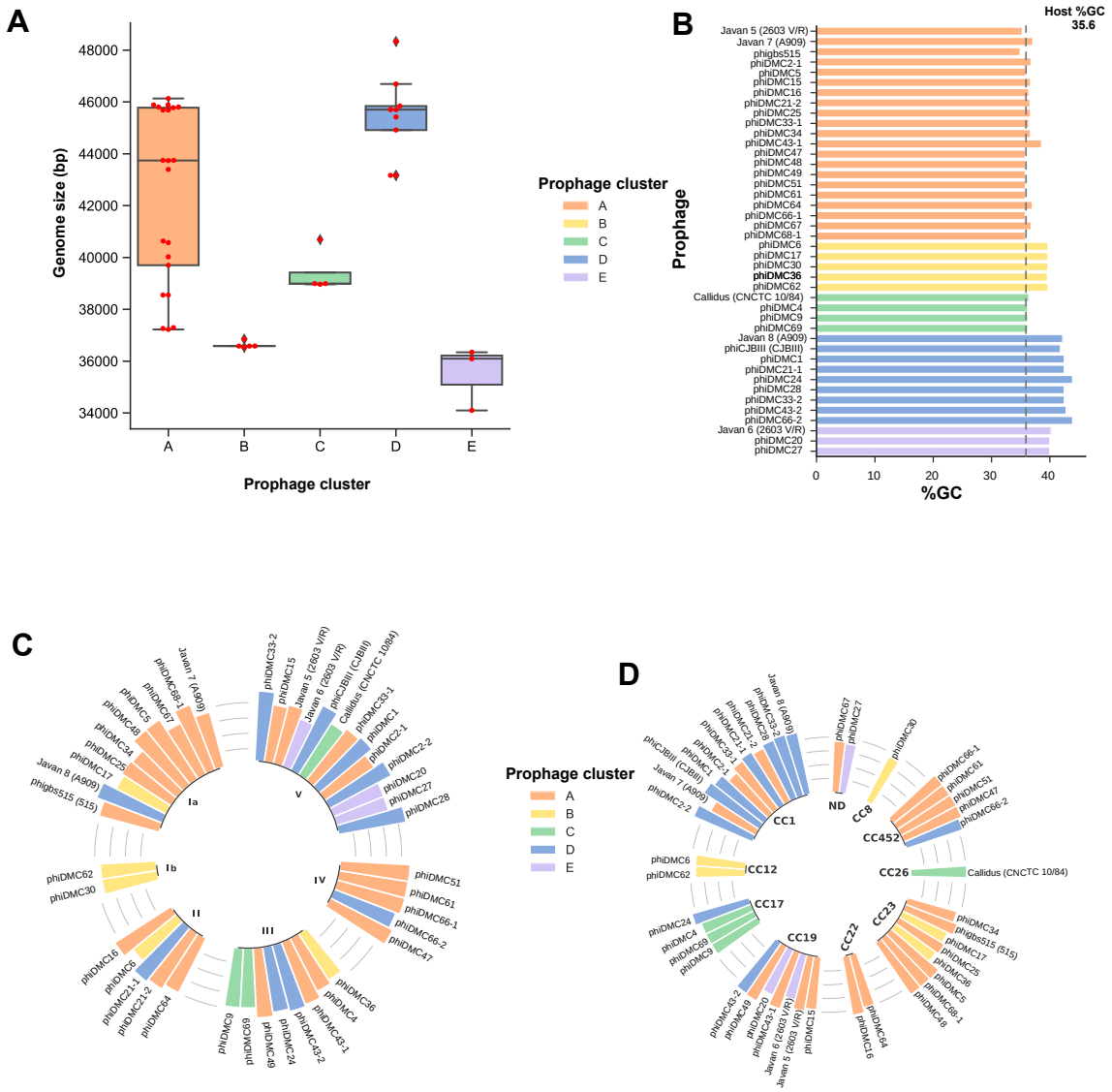


Figure 2.3: Comparative analysis between GBS prophages

A.Genome size distribution of prophages from GBS clinical isolates varies among different clusters. B. The average %GC content of prophages from GBS clinical isolates differs among the different clusters and were mostly higher than the streptococcal host. Dotted line indicates %GC

content of the host. C. Prophage cluster and genome size distribution across the different serotypes. D. Organization of prophage genomes by clonal cluster of the host

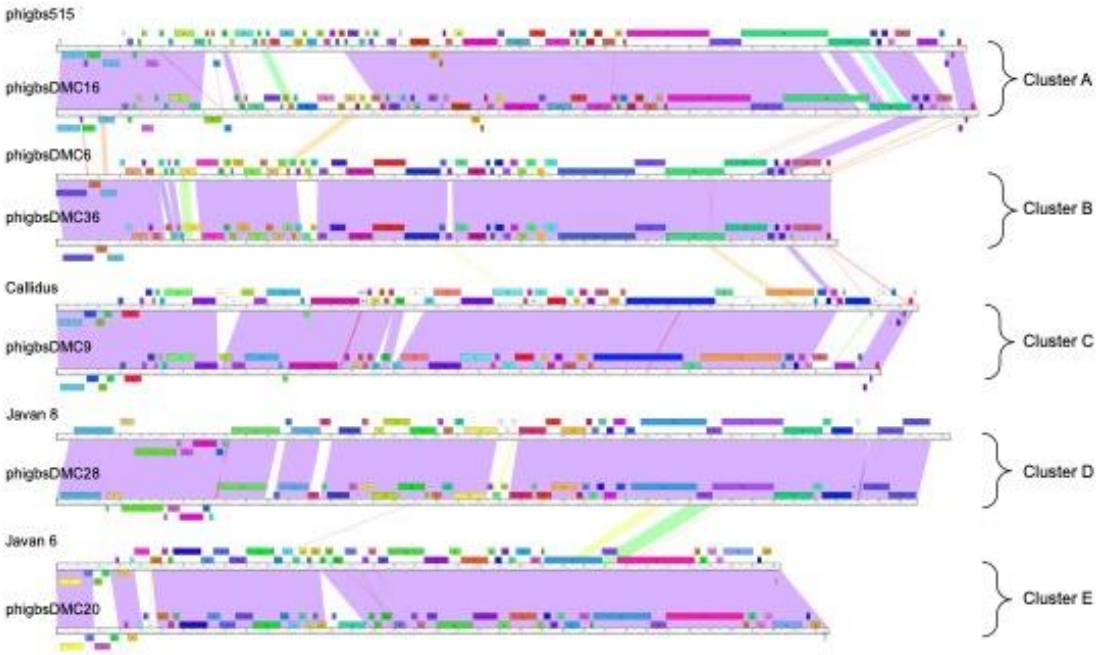


Figure 2.4: Prophage diversity across clusters

Phamerator map of two representative prophage genomes from each cluster demonstrating prophage diversity. Conserved regions between prophage genomes are indicated by regions of violet shading while regions with little or no sequence similarity are white. High similarity is observed within clusters while little similarity is shown between clusters.

Table 2.1: Characteristics of GBS prophages

Prophage	Cluster	Genome size (bp)	% GC	No. of Genes	Type of Integrase	Gene upstream of insertion site (att site)	Inserts into	Gene downstream of insertion site (att site)	attachment site
Javan 5 (2603 V/R)	A	40 574	35.3	78	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
Javan 7 (A909)	A	37 225	37.1	62	Tyrosine	hypothetical protein	N/A	HU family DNA-binding protein	TTATAGTTGGGGCGAATTTGGGGCATAA
phigbs515	A	40 634	34.9	89	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC2-1	A	39 700	36.8	66	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC5	A	46 132	35.9	80	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC15	A	38 551	36.7	68	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC16	A	43 746	36.4	71	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC21-2	A	43 397	36.6	73	Serine	N-acetylmannosamine kinase	acetyl xylan esterase	Sialic acid utilization regulator, RpiR family	GATTTTGATGACTTC
phiDMC25	A	38 551	36.7	68	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC33-1	A	43 746	36.4	72	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC34	A	37 294	36.7	62	Serine	ComGF	N/A	ComGB	TAAATTTTTC
phiDMC43-1	A	43 738	38.6	58	Tyrosine	bacterial ribosome SSU maturation protein RimP	tRNA-Ser-GGA	tRNA (guanine(46)-N(7))-methyltransferase	AATCCCCTCCTCTCCTTT
phiDMC47	A	45 805	35.8	81	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC48	A	45 885	35.8	80	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC49	A	45 685	35.8	84	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC51	A	45 805	35.8	82	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC61	A	45 686	35.8	80	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC64	A	40 022	37	68	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC66-	A	45 884	35.8	80	Tyrosine	integrase	tRNA-	LSU ribosomal protein	ATGTCCCCTGCC

1							Arg	L19p	
phiDMC67	A	37 262	36.8	61	Serine	ComGF	N/A	ComGB	TAAATTTTTC
phiDMC68-1	A	45 777	35.8	80	Tyrosine	integrase	tRNA-Arg	LSU ribosomal protein L19p	ATGTCCCCTGCC
phiDMC6	B	36 849	39.7	57	Serine	ComGD	ComGC	ComGB	TAAATTTTTC
phiDMC17	B	36 582	39.7	54	Serine	ComGD	ComGC	ComGB	TAAATTTTTC
phiDMC30	B	36 585	39.7	56	Serine	ComGD	ComGC	ComGB	TAAATTTTTC
phiDMC36	B	36 581	39.6	54	Serine	ComGD	ComGC	ComGB	TAAATTTTTC
phiDMC62	B	36 534	39.7	54	Serine	ComGD	ComGC	ComGB	TAAATTTTTC
Callidus (CNCTC 10/84)	C	40 696	36.4	67	Tyrosine	hypothetical protein	N/A	HU family DNA-binding protein	TTATAGTTGGGGCGAATTTGGGGCATAA
phiDMC4	C	38 991	36.2	55	Tyrosine	hypothetical protein	n/a	DNA binding protein HbSu	TTATGCCCCAAATTCGCCCCAACTATAA
phiDMC9	C	38 963	36.2	55	Tyrosine	hypothetical protein	N/A	DNA-binding protein HbSu	TTATGCCCCAAATTCGCCCCAACTATAA
phiDMC69	C	38 991	36.2	65	Tyrosine	hypothetical protein	N/A	DNA binding protein HbSu	TTATAGTTGGGGCGAATTTGGGGCATAA
Javan 8 (A909)	D	45 841	42.2	43	Serine	transcriptional regulator AcrR family	N/A	hypothetical protein	ACTTTTGAAAAGGAGA
phiCJBIII (CJBIII)	D	48 336	41.8	46	Serine	transcriptional regulator AcrR family	N/A	hypothetical protein	ACTTTTGAAAAGGAGA
phiDMC1	D	45 705	42.5	45	Serine	transcriptional regulator AcrR family	N/A	hypothetical protein	ACTTTTGAAAAGGAGA
phiDMC2-2	D	46 693	42.5	49	Serine	transcriptional regulator AcrR family	N/A	hypothetical protein	ACTTTTGAAAAGGAGA
phiDMC21-1	D	45 421	42.5	45	Serine	transcriptional regulator AcrR family	N/A	hypothetical protein	ACTTTTGAAAAGGAGA
phiDMC24	D	43 168	43.9	46	Serine	hypothetical protein	N/A	hydrolase (HAD superfamily)	TGGTATAAT
phiDMC28	D	45 702	42.5	45	Serine	transcriptional regulator AcrR family	N/A	hypothetical protein	ACTTTTGAAAAGGAGA
phiDMC33-2	D	46 690	42.5	45	Serine	transcriptional regulator AcrR family	N/A	hypothetical protein	ACTTTTGAAAAGGAGA
phiDMC43-2	D	44 915	42.8	44	Serine	transcriptional regulator AcrR family	N/A	hypothetical protein	ACTTTTGAAAAGGAGA

phiDMC66-2	D	43 168	43.9	47	Serine	hypothetical protein	N/A	hydrolase (HAD superfamily)	TGGTATAAT
Javan 6 (2603 V/R)	E	34 100	40.2	40	Tyrosine	alkyl hydroperoxide reductase protein F	tRNA-Cys	Na+/H+ antiporter	AATCCGTCTACCGCCT
phiDMC20	E	36 343	40	53	Tyrosine	alkyl hydroperoxide reductase protein F	tRNA-Cys	Na+/H+ antiporter	AATCCGTCTACCGCCT
phiDMC27	E	36 093	40	53	Tyrosine	alkyl hydroperoxide reductase protein F	tRNA-Cys	Na+/H+ antiporter	AATCCGTCTACCGCCT

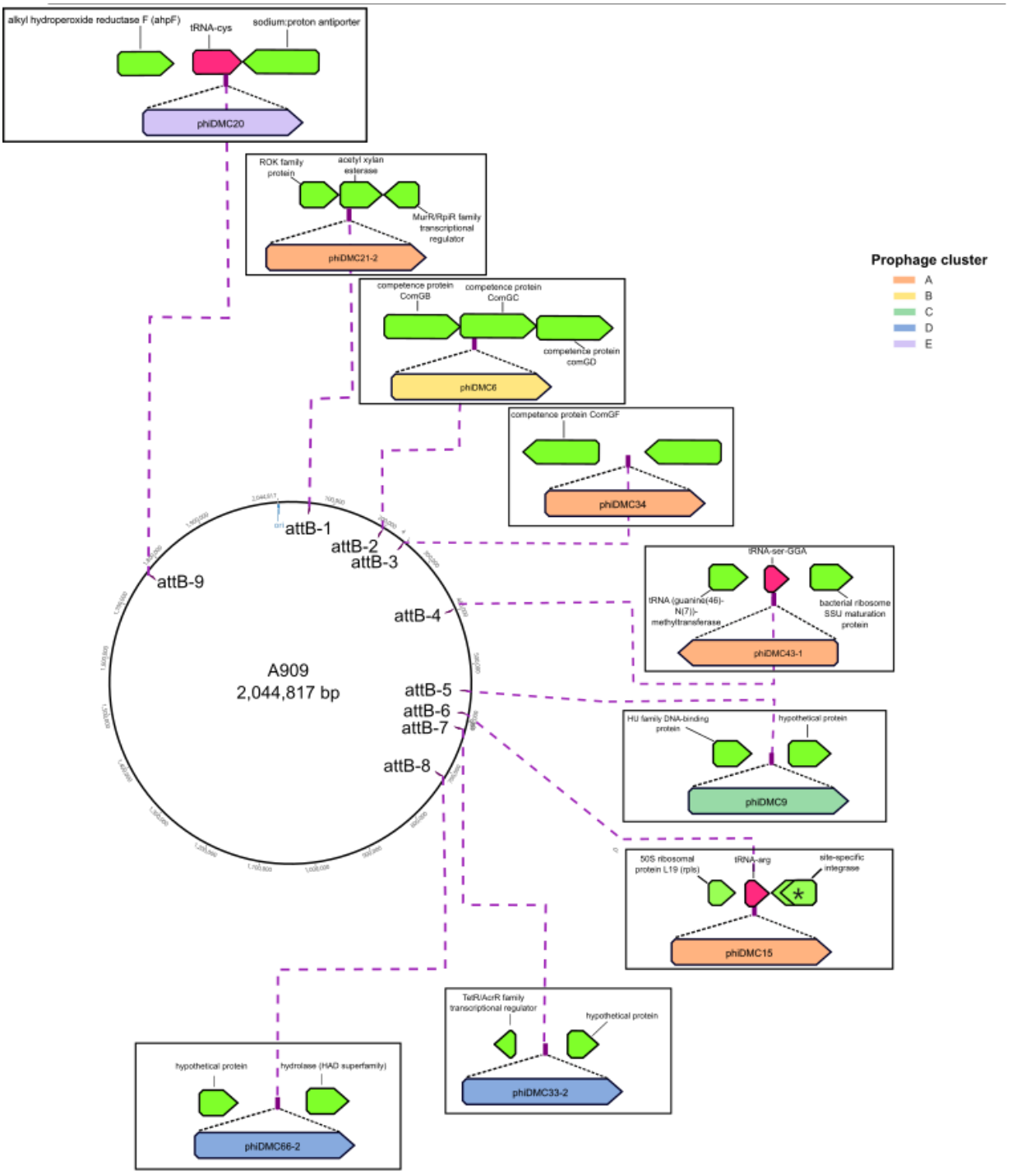


Figure 2.5: Prophage insertion sites

Nine different prophage insertion sites were identified based on the attachment site sequence and is displayed on GBS A909 as a reference genome. Attachment sites are numbered based on their

location from the site of origin. Each box shows site of insertion and flanking chromosomal genes. Genes are colored green and tRNA are colored pink. Representative prophages that use the insertion site are shown and shaded according to cluster.

Most cluster A prophages use an attB site (attB-6) located within a bacterial host arginine tRNA except for phiDMC43-1 which is located within a bacterial host serine-GGA tRNA (attB-4). Cluster E prophages use a different attB site (attB-9) located within a host cysteine tRNA. The common core sequences shared by attB and attP are typically 12-18 bp for cluster A and 16 bp for cluster E, with phage-derived sequences reconstructing the 3' end of the bacterial host tRNA gene. Int-Y prophages integrating at attB-5, a non-tRNA attB site have a longer core sequence of 28 bp, and these are used by Cluster C prophages and a single Cluster A prophage (Javan 7). This attB site is located between a hypothetical protein and an HU family DNA-binding protein. Five attB sites (attB-1, attB-2, attB-3, attB-7, and attB-8) are used by Int-S systems with common core sequences between 10-15 bp. Two of the five attB Int-S sites integrate within open reading frames, which they disrupt. Cluster B prophages integrate at attB-2 within the *com* gene locus as described similarly for the ϕ10403S prophage of *Listeria monocytogenes* (236) and a single Cluster A prophage (phiDMC21-2) integrates at attB-1 within a gene that encodes acetyl xylan esterase, one of the accessory enzymes for xylan degradation. The remaining three Int-S attB sites are located in intergenic regions, specifically between ComGF and ComGB (attB-3), a transcriptional regulator AcrR family protein and a hypothetical protein (attB-7), and a hypothetical protein and HAD family hydrolase (attB-8). Cluster D prophages use attB-7 and attB-8, while attB-3 is used by two cluster A prophages.

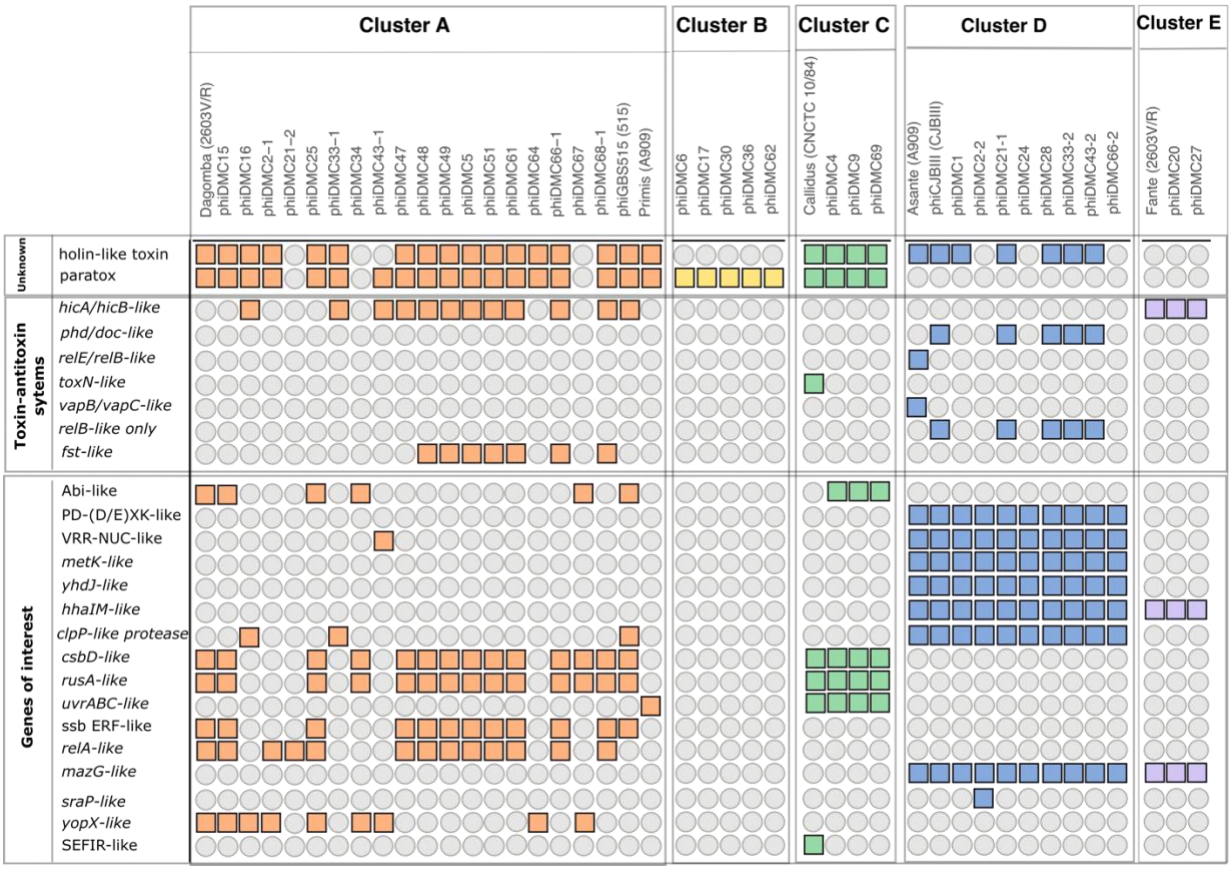


Figure 2.6: Identified genes that may contribute to bacterial fitness

Genes of interest are widely dispersed among clusters. Co-occurrence of the paratox and holtox proteins is diverse and not cluster dependent. Solid colored boxes represent the presence of genes of interest by prophage cluster. The absence of a gene is indicated by a grey circle.

2.1.1 GBS prophages encode multiple toxin-antitoxin systems.

Prophages encode genes that contribute to bacterial survival in several pathogens such as *Vibrio cholerae*, *Escherichia coli*, and *Streptococcus pyogenes* (157). However, their role in GBS is not well understood. GBS prophage genomes are enriched with genes that potentially contribute to virulence including toxin-antitoxin (TA) systems that were present in all clusters except cluster B (Figure 6). TA systems identified were unique to specific clusters. For example, about 30% of cluster A prophages have the *fst*-like TA system, believed to have a role in bacterial adaptation to adverse environmental conditions, promoting survival in harsh or fluctuating environments

(237), while cluster A and cluster E phage genomes have the *hicAB* systems that target cellular RNAs (238). Half of the cluster D prophages have *phd/doc*-like genes, a type II TA system thought to have a role in maintaining the stability of prophages within bacterial genomes (239). Several cluster D prophages also encode RelB, usually associated with its cognate antitoxin, but no *relE* homolog was found in these prophages. The Cluster D prophage, Javan 8, has both *relE/relB*-like and *vapB/vapC*-like pairs, which are type II TA systems that inhibit translation (240). Similarly, only one cluster C prophage, Callidus has a homolog of the *toxN* gene, the toxin of a type III TA system believed to be involved in phage defense (241) (Figure 6). Co-existing prophages appeared to complement each other with TA systems in some instances (Figure 7A).

2.3.4 Prx is encoded on the bacterial chromosome and often in GBS prophages.

A notable feature identified in most GBS prophages is paratox. Paratox, encoded by the *prx* gene, is a conserved protein in streptococcal species including GBS (205). Previous reports in *S. pyogenes* identified the *prx* gene to be located on the distal right arm of prophage genomes, and not on the bacterial genome, suggesting that it was acquired through horizontal gene transfer during lysogeny. Further, this work demonstrated that paratox was involved in the prevention of the uptake of DNA by acting as an inhibitor of the ComR protein in the regulation of competence controlled by the ComRS quorum sensing system (205). Therefore, the GBS prophage genomes were analyzed for a homolog of the paratox encoding gene (*prx*). More than 60% (27/43) of the GBS prophages across clusters A, B, and C contain a *prx* homolog (~98.3% homology to *S. pyogenes* paratox) located at the right terminal end of the prophage, adjacent to the phage attachment site (Figure 8A).

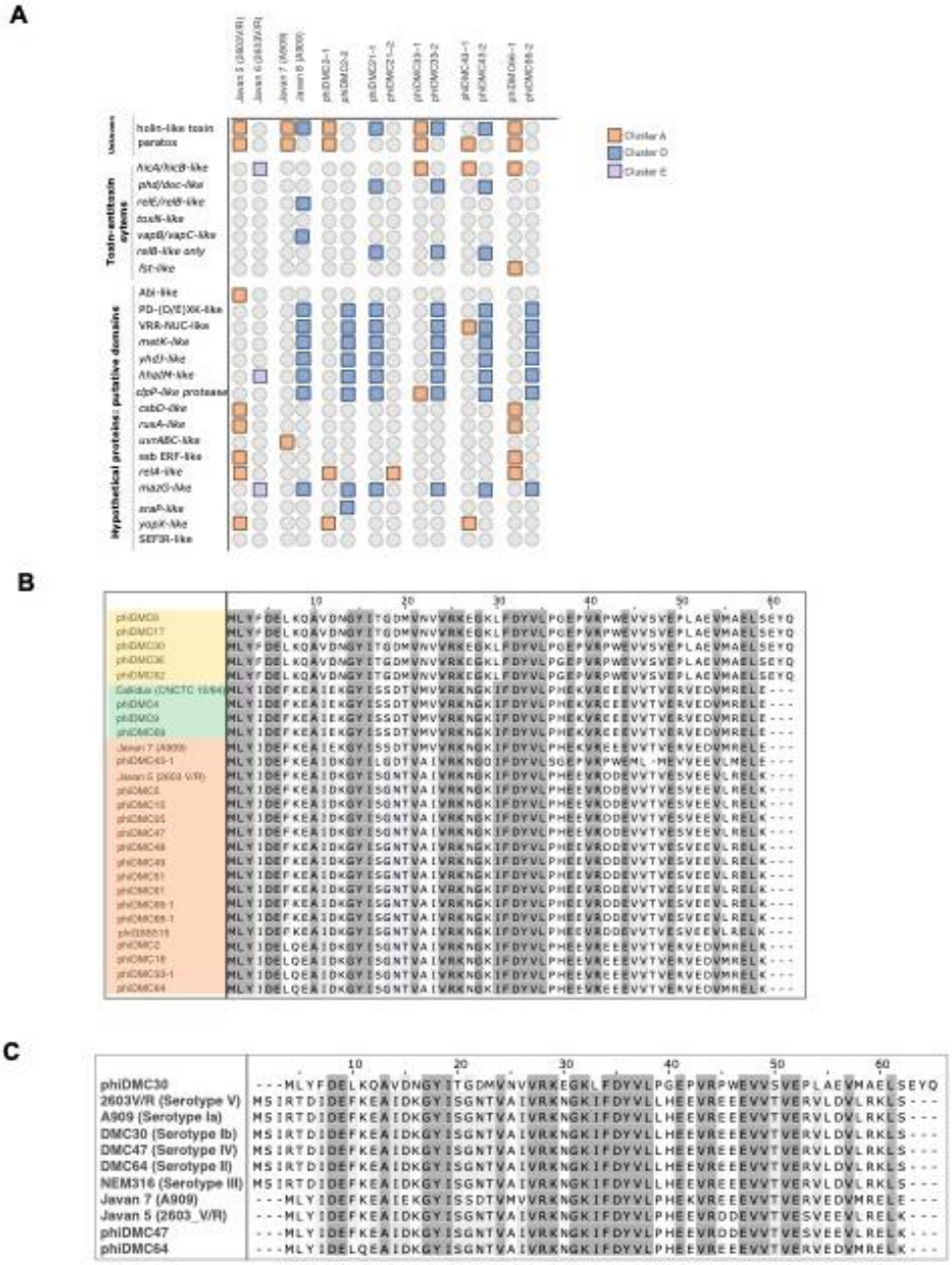


Figure 2.7: Paratox is highly conserved across clusters

A. Co-habiting prophages may complement genes. Solid colored boxes represent the presence of a gene by prophage cluster. Absence of wither gene is indicated by a grey circle. B. Clustal alignment of prophage paratox amino acid sequence by cluster. C. Clustal alignment of amino acid sequences of the host paratox and prophage paratox proteins.

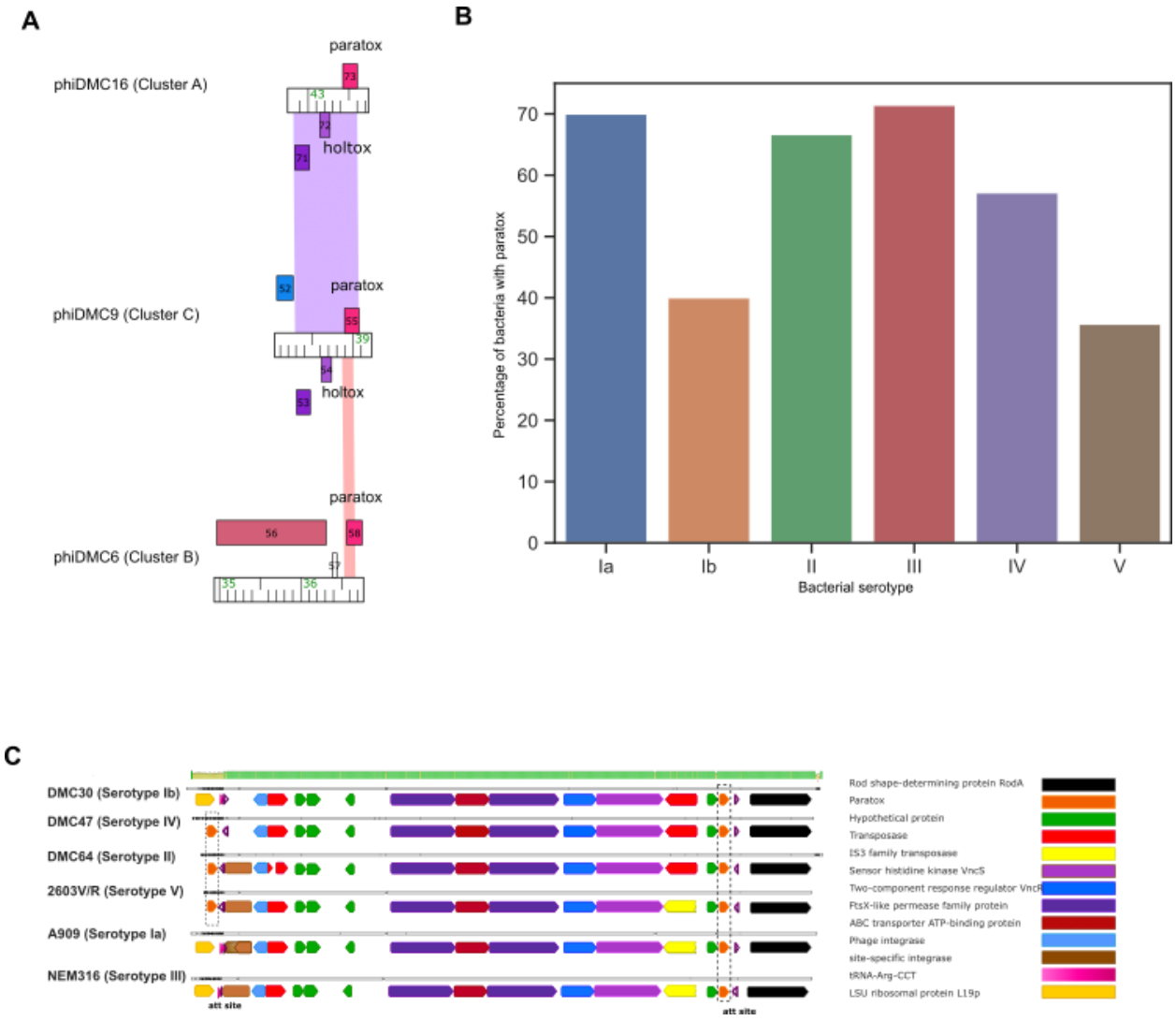


Figure 2.8: Paratox is also found on the bacterial host chromosome

A. Phamerator map alignment of *prx* gene from selected prophages within each cluster (A-C). B. C. Alignment of genome region of *prx* genes found on the host chromosome. The color panel above the selected sequences indicates high-to-low sequence conservation. Grey blocks indicate homologous sequences. Genes are marked by arrows, for which the putative functions are indicated by the colored key. Dotted box indicates *prx* gene.

Despite the observed difference of 3 amino acids at the C-terminus of the paratox proteins in cluster B prophages, a comprehensive examination of the paratox proteins in prophages from clusters A, B, and C indicated a substantial degree of sequence conservation among them (Figure 7B). Considering the notable high conservation of paratox within prophages, its distribution

across various serotypes was evaluated. However, there was no distinct correlation found between the presence of prophages and the number of bacteria per serotype (Figure 8B). This lack of correlation was also observed when examining clonal complexes (data not shown).

An extensive analysis of the GBS host genomes revealed the presence of another *prx* homolog in the chromosome of every GBS clinical isolate in this study. While these host *prx* genes were surrounded by the same sets of genes in every bacterial strain, they were different from the genes surrounding the *prx* gene encoded in the prophage genomes. The region surrounding the *prx* gene on the host genome begins with a phage integrase and ends with the *prx* gene flanked by two attachment sites (Figure 8C). This region also encodes transposase elements and several bacterial genes suggesting that this may be a mobile genetic element. An amino acid alignment of these host-encoded paratox proteins with the prophage paratox proteins showed some conservation of the amino acids (Figure 7C). However, the host-encoded paratox protein contains 3 additional amino acids at the N-terminus than the canonical prophage paratox protein.

2.3.5 GBS prophages encode a gene upstream of *prx* with holin-like and transmembrane domains.

In the *S. pyogenes* prophages that encode paratox, the *prx* gene is always located adjacent to a toxin gene and prophages lacking a toxin gene also lack the *prx* gene (242), therefore the name para-tox (adjacent to a toxin). To investigate whether the *prx* gene is adjacent to a toxin gene in the GBS prophages, the region surrounding *prx* was analyzed. Except for cluster B and a few cluster A prophages, all prophages carrying the *prx* gene have an ORF encoding a putative holin-like gene (previously designated as *holtox* (243) adjacent to *prx* (Figure 8A). The function of the protein encoded by the *holtox* gene is not known, but protein sequence analyses reveal the

presence of a holin domain (PF16935). Alignment of the *prx* genes of selected prophages in Phamerator shows the *prx* genes of clusters A and C are almost always preceded by two genes that encode a hypothetical protein and a putative *holtox* (Figure 8A). However, cluster B prophages have a different gene arrangement with the *prx* gene located a couple of genes downstream of the phage endolysin. Although cluster D and E prophages lack *prx* genes, all but three cluster D prophages contain the *holtox* gene (Figure 6), which suggests that the *prx* and *holtox* genes may not be inherited as a single module. Moreover, the *holtox* gene was not identified in any cluster E prophages.

2.4 Discussion

Prophages play a major role in virulence in many pathogens, including *Streptococcal* pathogens. *S. pyogenes* requires multiple prophage-encoded virulence genes for successful infection (174). While colonization by *S. pyogenes* always has the potential to result in disease, GBS can colonize the human urogenital tract and behave as a commensal, only becoming a pathogen under certain circumstances. In particular, GBS colonization of the vaginal tract of pregnant women is a major risk factor for transmission to the neonate, resulting in life-threatening disease. However, little is known about the role or the presence of prophages in this opportunistic pathogen. Therefore, this study utilized a collection of 42 GBS vaginal isolates from pregnant women, which had the potential to cause neonatal disease, to determine the presence and diversity of prophages. Furthermore, the relationship between prophages and previously described virulence genotypes was investigated. Seven previously analyzed GBS clinical strains were added to the investigation as reference strains.

There is a high level of prophage diversity among the GBS strains, with most strains having at least one prophage in their genome and some strains carrying multiple prophages. This is consistent with other studies (120, 188, 189). Out of the 49 strains investigated, 41 (~84%) had at least one prophage in their genome, with 6 strains carrying 2 complete prophage genomes. Serotypes Ia, III, and V, linked to GBS invasive disease (70, 229), had a higher proportion of prophages. However, this may be due to the fact that most of the bacterial genomes analyzed in this study belonged to these three serotypes explaining why over 70% of the prophages identified were found in these serotypes. Additionally, the absence of prophages in certain strains was not found to be distinctive to a specific serotype.

Cohabiting prophages can work together to regulate bacterial gene expression, prophage induction, and increase antibiotic sensitivity (244–246). In our dataset, 90% of bacterial strains with multiple prophages always carried a cluster A with mostly a cluster D prophage. This may not be a coincidence, as carrying phages from different clusters can increase genetic diversity and reinforce bacterial fitness or virulence. This could occur through gene expression complementation or one prophage stimulating the induction of another prophage (244, 245). However, it is unclear whether prophages interact in poly lysogenic GBS, and further research is needed to test these hypotheses.

GBS prophage integration sites vary among prophage clusters, but prophages within the same cluster tend to integrate into the same site. Prophage integration can disrupt genes, as observed in *Staphylococcus aureus* when lysogenized by phi13 phage, leading to the loss of beta-toxin expression (233). Conversely, prophage excision can impact bacterial phagosomal escape, as seen in *L. monocytogenes*, where excision of the prophage leads to the expression of the *com*

genes and allows for escape from phagosomes (236). Our dataset showed several prophages integrate into the *com* locus, which is in line with other studies on GBS prophages (188, 189, 192), and further research is needed to understand the relevance of this insertion site on the fitness and virulence of GBS strains.

Prophages can spread virulence genes in pathogenic bacteria and increase bacterial fitness during infection, such as enhancing adhesion to epithelial cells, increasing survival in serum, and improving antibiotic resistance (188, 247). Within our dataset, multiple genes of interest were identified including toxin-antitoxin systems. Four of the six TA systems identified in our study have been previously reported (188). Some genes were unique to this study, including a gene with homology to SEFIR/Toll/interleukin-1 receptor domain-containing protein domain and MazG, which is involved in bacterial survival under nutrition stress. Conducting further studies to investigate the specific functions of these genes and their role in GBS pathogenesis may be worthwhile.

A significant finding was that the *prx* gene, which encodes for the paratox protein involved in bacterial competence (205), is present in all 49 clinical isolates of GBS. Paratox was found in both prophage genomes and the bacterial host chromosome, with the latter possibly being a remnant of a previous phage or a mobile genetic element. The region containing the *prx* gene on the bacterial host chromosome has a phage integrase located upstream of an attL site at the 5' end, suggesting that this *prx* gene may be a genomic remnant of a previous phage. The presence of a transposase gene also in this region suggests the *prx* gene may be part of a transposon. Therefore, it is not known whether the *prx* gene on the host chromosome was originally a phage

gene or was inherited from a mobile genetic element. However, the high conservation of the *prx* gene, whether prophage encoded, or host-encoded, suggests having this gene present provides a significant fitness advantage, particularly since all bacterial strains contain the *prx* gene regardless of having a prophage.

Unlike in *S. pyogenes*, the *prx* gene in GBS prophages is not always adjacent to a toxin-encoding gene. In some instances, a putative *holtox* gene is located next to *prx*, but it is unclear whether it functions as a toxin. This putative *holtox* is homologous to a holin-like toxin gene that encodes a protein thought to have antibacterial activity against Gram-positive bacteria and complements a lysis defective bacteriophage (243). Further research into this gene is necessary to provide insight into its function in GBS.

Prophages have been consistently found to contribute to bacterial fitness and virulence. This study investigated the presence and diversity of prophages in 49 clinical isolates of GBS strains. GBS prophages are very diverse with most strains carrying at least one prophage and some carrying multiple prophages. Cohabiting prophages tended to carry specific combinations of prophage clusters, which could impact bacterial fitness or virulence. Prophages were found to integrate into specific sites, including genes within the *com* locus, which could impact bacterial fitness. Notably, all 47 clinical isolates of GBS had the *prx* gene, which is involved in bacterial competence, present on the bacterial host chromosome, and may provide a fitness advantage. Additional research on prophages in GBS is needed, as they may have a significant effect on bacterial fitness and virulence.

CHAPTER 3

THE PROPHAGE OF THE GBS STRAIN, CNCTC 10/84 PROVIDES A COMPETITIVE ADVANTAGE TO THE LYSOGEN

3.1 Introduction

Bacteria in natural environments compete for resources and space to survive by employing various strategies such as antibiotic and toxin synthesis, motility, sporulation, predatory functions, and biofilm formation (248–250). Bacterial competition most often happens within or between bacterial species and can have significant impacts on microbial community dynamics (251, 252). Bacterial competition can promote genetic diversity and adaptation to changing environments. For example, certain bacterial clones tend to grow faster when grown in competition with a different bacteria (253). Similarly, when resources are in abundance, bacteria may prioritize using up the resources as quickly as possible rather than investing energy in competing with other bacterial species for those resources (254). In this way, bacterial competition can lead to an increase in the uptake and utilization of nutrients.

Bacteria exhibit a multifaceted approach to competition, which extends beyond nutrient utilization. In fact, they employ a diverse range of tactics to harm, inhibit growth, and eliminate their competitors. For example, bacteria can use temperate phages against competing bacterial species(254). Temperate phages have two potential life cycles: they can follow a lytic cycle, in which they replicate within bacterial cells and produce infectious virions, followed by lysis of the cell, or alternatively integrate into the bacterial genome and replicate along with the host genome (157). Through lysogenic conversion, a process where phages incorporate their genetic

material into the genome of the host bacteria and alter its phenotype, phages can provide new traits, including virulence factors, and metabolic enzymes to the bacteria they infect (141, 157, 161–166). These beneficial traits provided by prophages enhance bacterial fitness and give lysogenic bacteria an advantage in competing for limited resources (254–256). Additionally, prophages that are induced into the lytic cycle can promote direct competition within bacterial communities when a small subpopulation of lysogenic cells release virions that go on to infect closely related but not identical strains(257, 258) and in that process, the newly lysogenized bacteria can acquire traits from competitors (259).

GBS genomes typically contain one or more prophages, with a high degree of genetic diversity (120, 121, 188–190) as demonstrated in Chapter 2. Lysogenic GBS strains may also possess various fitness factors that promote metabolic activity and growth (183). Additionally, the fact that prophages account for 10% of strain-specific genes in GBS(114, 182), and that they can form infectious particles and either lysogenize or kill other strains (120), suggests that they play an important role in GBS evolution. However, the role of lysogeny on GBS bacterial fitness and virulence has not yet been determined experimentally, specifically by comparing isogenic strains that differ in their prophage regions.

Here, a hypervirulent GBS strain is compared with its corresponding non-lysogenic strain, i.e., cured of its prophage, to understand the contribution of the prophage to bacterial fitness. The findings of this study reveal that the prophage provides a competitive advantage to the lysogen when grown in competition with its isogenic non-lysogenic strain.

3.2 Materials and Methods

3.2.1 Bacterial strains, plasmids, and growth conditions

GBS strain CNCTC 10/84 (serotype V, sequence type 26) was used as a wildtype (WT) strain in this study. Plasmids were maintained in TOP10 (Invitrogen) *Escherichia coli* and used for plasmid construction and vector propagation. GBS strains were grown in Todd Hewitt broth (Accumedia product number) supplemented with 2% yeast extract (THYB) and on THYB agar at 37°C and *E. coli* strains were grown in Luria broth (LB) media (Acumedia) and on LB agar at 37°C. Media was supplemented with antibiotics when needed as follows; chloramphenicol (Cam) 20 µg/ml for *E. coli* and 3 µg/ml for GBS, erythromycin 750 µg/ml for E. coli and 5 µg/ml for GBS, streptomycin 250 µg/ml for GBS. Antibiotics were purchased from Sigma – Aldrich. All restriction enzymes were purchased from New England Biolabs.

3.2.2 Generation of a prophage-cured CNCTC 10/84 strain

To cure CNCTC 10/84 of its prophage (550,775 – 591,450), the prophage excision (*xis*) gene (W903-RS03095) was cloned into the expression vector pLZ12-rofA-pro. The open reading frame of the excise gene was PCR amplified from genomic DNA using primers listed in Table 3.1. The 405-bp PCR product and pLZ12-rofA-pro were digested with *Bam*HI and *Pst*I according to the manufacturer's recommendations (New England Biolabs, Ipswich, MA) and gel purified. After ligation of the excise sequence into the vector using T4 ligase (NEB), 5 µl of ligation mixture was electroporated into TOP10 *E. coli* (Invitrogen) according to the manufacturer's recommendations. The sequence and orientation of the excise insert were confirmed by DNA sequencing. The recombinant plasmid was then transformed into competent GBS as previously described (260) and cam^R colonies were selected for screening. To confirm

the loss of the prophage, cam^R colonies were verified by PCR for the presence of the attB site using primers that span the bacterial attachment site and the loss of a gene specific to the prophage (gp5) (Table 3.1). To ensure that the recombinant plasmid did not influence results in this study, the prophage-cured CNCTC 10/84 strain was cured of its plasmid by growing cultures overnight at 37°C, double diluted, and growth to late-logarithmic phase at 40°C. Serial dilutions were plated on THYB agar plates and plasmid-free colonies were verified by PCR for the presence of the attB site, loss of gp5, and xis gene. Restriction endonuclease digests were carried out according to the manufacturer's recommendations (New England BioLabs, Ispwich, MA)

Table 3.1: Primers used in this study.

Description	Primers	Sequence (5' to 3')	Tm	amplicon size
Primers used to amplify excision (xis) gene	5' Xis-Cphage-BamH1	CGC GGA TCC CTT ATC AAA AAT TCT ACT TAC	57.5	405 bp
	3' Xis-Cphage-Pst1	AAA ACT GCA GCT TCA TTA TGT TAT ACT CCT AAC TG	58.5	
Primers used to detect loss of gp5 in Callidus	5' GBS-Cphage-Gp5-fwd	GAG CAT TTT CAG TGG GTC GC	56.8	260 bp
	3' GBS-Cphage-Gp5-rev	CAC TTT CCA AGA AAC AAC CTC AGG	56.2	
Primers used to detect bacterial attachment site in CNCTC 10/84 after loss of prophage	5' GBS-Callidus-attP-fwd	TGC CAA CCG AAA AAC CTA AC	53.7	150 bp
	3' GBS-Callidus-attP-rev	GTA GGA CAC GCT GAT GCA AA	55.4	
	5' GBS-Callidus-attB-fwd	TGG AGG ATT TGT TAA CAT GG	50.2	455 bp
	3' GBS-Callidus-attB-rev	TTC CTT TGA GCT CTC TAG TCG	53.7	

3.2.3 Phenotypic assays

To determine if there were differences in chain length, overnight cultures were imaged using the Zeiss axiocam and analyzed with ImageJ. Beta-hemolytic activity was tested by spotting 20 µl of

serially diluted overnight cultures on blood agar plates (Thermo Fisher Scientific). Biofilm assays were performed as previously described (261). Antimicrobial susceptibility testing was performed using the Kirby-Bauer disk diffusion method (262). Triton X-100 susceptibility was determined as previously described (263). All experiments were performed in triplicate.

3.2.4 Growth curve and competition assay

Wild-type (WT) and prophage-cured (PC) CNCTC 10/84 strains were grown overnight in THYB at 37°C and diluted 1:50 in THYB and incubated at 37°C without shaking. Absorbance readings (OD600) were recorded every 30 minutes till cultures reached the stationary phase. For competition assay, THYB broth or RPMI was seeded with a spontaneous streptomycin-resistant clone of the WT and an erythromycin-resistant clone of the PC generated by integrative insertion of the pBR-omega plasmid into the chromosome at the attB site using primers, listed in Table 3.1. Wild-type and prophage-cured cultures were individually grown at a final concentration of 2×10^5/mL and for coculture, were combined in a 1:1 ratio with both organisms at a final concentration of 2×10^5/ mL. Single cultures and cocultures of both wild type and prophage-cured were plated for CFU enumeration on selective media every 2 hours for 24 hours. Cell-free supernatants of each sample collected every 2 hours were spotted on GBS strain DMC 22 to test for release of phage particles Competition indices were calculated using colony forming unit (CFU) density per mL, where competition index was calculated as ln $((CFU^{PC}{}_{T24}/CFU^{PC}{}_{T0})/(CFU^{WT}{}_{T24}/CFU^{WT}{}_{T0}))$. Assays were performed with six experimental replicates for THYB and five replicates for RPMI.

3.2.5 Virulence assays using a zebrafish infection model.

All animal experiments were performed under approved International Animal Care and Use Committee protocols at the University of Maine. Anesthetized (26 µg/ml Tricaine for 10 minutes) 2 days post fertilization (dpf) zebrafish larvae (zf5) were injected with 10 cfu per fish of indicated bacterial strain grown to mid-log phase into the yolk sac and monitored for survival over 72 hours at 28˚C in 24-well plates. To confirm bacterial dose, 100ul of serially diluted bacterial suspension was plated on THYB agar plates with the required antibiotics. Experiments were performed in triplicate.

3.2.6 Whole-genome RNA seq transcriptomic analysis

Wild type and prophage-cured CNCTC 10/84 were grown anaerobically overnight in THY at 37°C. Cultures were diluted 1:50 into 50mL THY, incubated anaerobically at 37°C, and grown to early logarithmic (OD600 ~ 0.3) and late logarithmic phase (OD600 ~ 0.7). RNA was isolated using RNeasy mini kit (Qiagen, Germantown, MD) following the manufacturer's protocol with a few modifications. Harvested cells (4 ml) were treated with RNAProtect Bacteria Reagent (Qiagen) and incubated for 5 min at room temperature. Cells were then centrifuged and resuspended in 100 µl of TE (10 mM Tris, 1 mM EDTA, pH 8.0) with 10 µl lysozyme (25mg/ml, Sigma cat #) and mutanolysin (50U/ml-Sigma cat#). Cells were lysed using Lysing Matrix B tubes (MP biomedicals, Irvine, CA) and homogenized for 5 mins in the TissueLyser LT (Qiagen) set for 5 min at 50 Hz. To remove contaminating DNA, RNA was treated with DNAse on the column (Qiagen) followed by a second DNAse treatment using the Turbo DNA-free Kit (Thermo Scientific, Waltham, MA) according to the manufacturer's recommendations. RNA quantity was determined with the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Montchanin, DE, USA). Samples were stored at -80°C until ready to use.

3.2.7 Bioinformatic analysis of prophage region

The isolated RNA was sent to the Hubbard Center for Genome Studies for RNA sequencing on an Illumina HiSeq 2500. Raw sequencing data files were uploaded to the public Galaxy server at usegalaxy.org. For additional RNA sequencing analyses, raw sequencing data files from three different published RNA sequencing data were obtained (NCBI GEO accession number GSE98398, (264); GSE165992,(261); NCBI SRA accession number SRP140532,(265)). Read quality was determined using FastQC(206) and processed using Trimmomatic(207). Reads were aligned to the CNCTC 10/84 reference genome sequence (NCBI RefSeq NZ_CP006910) using Bowtie2 version 2.1.0 (266) with default (– end-to-end) alignment mode and by specifying –sensitive as an additional parameter. Reads that mapped only once to the genome (uniquely mapping reads) were extracted from SAM files by filtering for the "XS:" tag used by bowtie2 for reporting secondary alignments for a given read. Only uniquely mapped reads stored under SAM file format were used for the subsequent operations. For visualization of transcriptional activity across the genome, SAM files were converted to corresponding binary format (BAM files) with SAMtools version 0.1.19 (214)and viewed with Integrative Genome Viewer (IGV)(267). The genome map of the prophage region of CNCTC 10/84 was downloaded from Phamerator (218)and edited with Inkscape (www.inkscape.org). For differential expression analysis and TPM (transcripts per million mapped reads) calculations, uniquely mapped reads were sorted with the kallisto quant command from Kallisto package version 0.5.4 (268)with the following parameters: -m intersection-nonempty -s no -t gene -i locus_tag. This step aimed at producing a matrix composed of raw read counts per gene for each library. The matrix of read counts was then utilized for manual TPM calculations using the following formula: TPM = RPK/ scale

factor, where RPK is the ratio of the number of reads mapped to a gene to the effective length of the gene in base pairs and the scale factor is determined as sum (RPK)/ 1E6. TPM values have been calculated for all libraries, and the arithmetic mean has been used.

3.2.8 Phage induction, propagation, and TEM microscopy

Cell-free supernatants of wild-type bacterial cultures cloned with the *xis* gene were collected, sterile filtered through 0.2 µM pore diameter membrane filter, and concentrated using Amicon tubes, and stored at 4℃ until use. For plaque assays, 10 µl of the phage supernatant and 100 µl of an overnight culture of the indicator strain GBS DMC 22 were mixed in a 15 ml conical tube containing 3 ml top agar (0.45% agar) supplemented with 10 mM $CaCl_2$ and poured onto THYB agar plates. The plates were incubated at room temperature overnight to form plaques on the lawns.

For TEM microscopy, a 1-mL volume of phage lysate (2.0 x 10^9 PFU /mL) was centrifuged at 4°C and 20k x g for 1 hour. Nearly all the supernatant was removed and replaced with 100 µL of phage. The sample was stored overnight at 4°C before preparing grids. Lysate was applied in a 10-µL volume to a carbon-coated copper grid for 2.5 min before washing twice with water. The grid was stained with 10-µL of uranyl acetate. Transmission electron microscopy was conducted at the Leduc Bioimaging Facility at Brown University in Providence, Rhode Island.

3.2.9 Statistical analyses

For zebrafish injections, statistical analyses were conducted using GraphPad Prism 7 software (GraphPad Software, Inc., La Jolla, CA). Kaplan-Meier survival curves were subjected to a log-rank (Mantel-Cox) test, and Bonferroni correction was then used to determine statistical

differences between pairs of treatments. The remaining statistics for figures were calculated using the Microsoft Excel data analysis tool, by first conducting an F-test to determine variance followed by a t-test to determine significance. All significant differences are indicated in the figures, with *, **, ***, and ns indicating P values of <0.05, <0.01, <0.001, and not significant, respectively.

3.3 Results

3.3.1 Prophage loss influences the cell wall integrity of CNCTC 10/84.

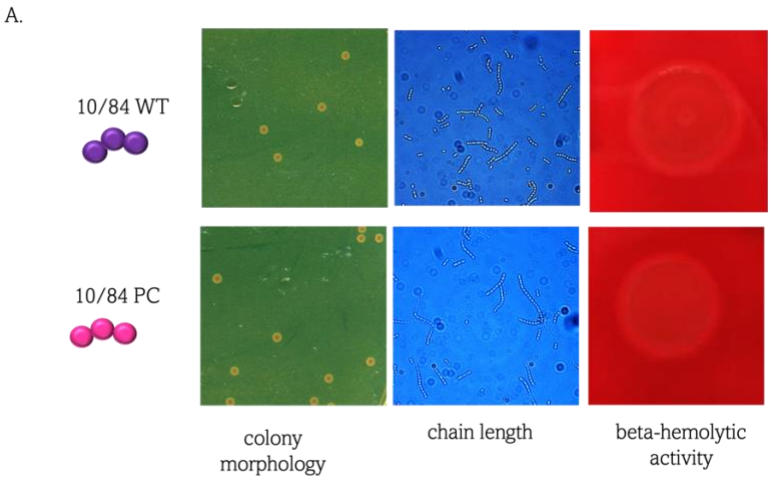
To investigate prophage impact on bacterial fitness, a prophage-cured derivative of the hypervirulent strain, CNCTC 10/84, was isolated by inducing the single prophage within the strain. The prophage-cured strain is annotated as CNCTC 10/84 PC and the wild-type strain as CNCTC 10/84 WT. Both strains were analyzed for phenotypic differences by comparing cell morphology, chain length, and hemolytic activity. No differences were observed between the strains for these assays (Figure 3.1a). Similarly, there was no difference between the two strains in their ability to form biofilms (data not shown). The antibiotic susceptibility profile was compared between the two strains i.e., CNCTC 10/84 WT and CNCTC 10/84 PC. Both strains were sensitive to chloramphenicol, erythromycin, penicillin, and vancomycin but showed intermediate resistance against tetracycline (Figure 3.1b). To investigate whether the presence of prophage affects the autolysis susceptibility of GBS, both strains were subjected to treatment with Triton X-100, a non-ionic detergent. Spot tests were performed on log-phase bacterial cells that were exposed to 0.01% Triton X-100 for 1 hour, 7 hours, and 24 hours. The results showed that the wild-type strain was more susceptible to the treatment compared to the prophage-cured strain. (Figure 3.1c).

3.3.2 Prophage provides a competitive advantage to the CNCTC 10/84 lysogen *in vitro*.

An intact cell wall is fundamental for survival as a compromised cell membrane can result in increased susceptibility to stress, antimicrobial agents, and host immune defenses, which may lead to decreased virulence (269). Since we observed resistance of the prophage-cured strain to autolysis, we wanted to explore the impact of the absence of the prophage on the bacterial stress response. Co-culture of the wild-type and prophage-cured strains in both a nutrient-rich (THYB) and nutrient-poor medium (RPMI), saw a significant difference in the CFU enumeration of the prophage-cured strain compared to the wildtype strain after 24 hours (Figure 3.2b). To better understand the dynamics of growth between the wild type and prophage-cured strain in co-culture, viable counts of each bacterial strain were monitored every 2 hours over 24 hours. After 6 hours of growth in the log phase, it was found that the prophage-cured strain had a cell density 2-fold less than the wild-type strain in co-culture, despite showing similar growth kinetics in monoculture conditions (Figure 3.2c). In contrast, when the prophage-cured strain is co-cultured with a different GBS strain (515) containing a distinct prophage from the CNCTC 10/84 strain, the prophage-cured strain shows normal growth (data not shown). This suggests that the inability of the prophage-cured strain to survive with its wild-type isogenic strain may be attributed to the presence of the prophage.

To determine whether there are differences in virulence between the wild type and prophage-cured strain, zebrafish larvae were challenged with both strains and monitored for survival for 72 hours. By 48 hours, only 50% of zebrafish larvae inoculated with the wildtype strain of GBS had survived while about 60% of zebrafish larvae had survived when inoculated with the prophage-

cured strain. However, by 72 hours, similar survival rates of about 25% were observed in zebrafish inoculated with both strains (Figure 3.2a). Since the two strains showed almost the same level of virulence individually, the question of whether one would have a competitive advantage over the other in vivo was investigated. When zebrafish larvae were co-infected with both strains, a similar rate of over 60% survival was observed as with zebrafish inoculated with the prophage-cured strain alone (Figure 3.2a). This prompted an investigation into the bacterial burden of both strains in co-infected zebrafish. However, there was no difference in colony-forming units (CFU) of bacteria recovered from zebrafish larvae (Figure 3.2b). This suggests that the interactions between the wild type and prophage-cured strains are complex and may not be only due to the presence or absence of the prophage.



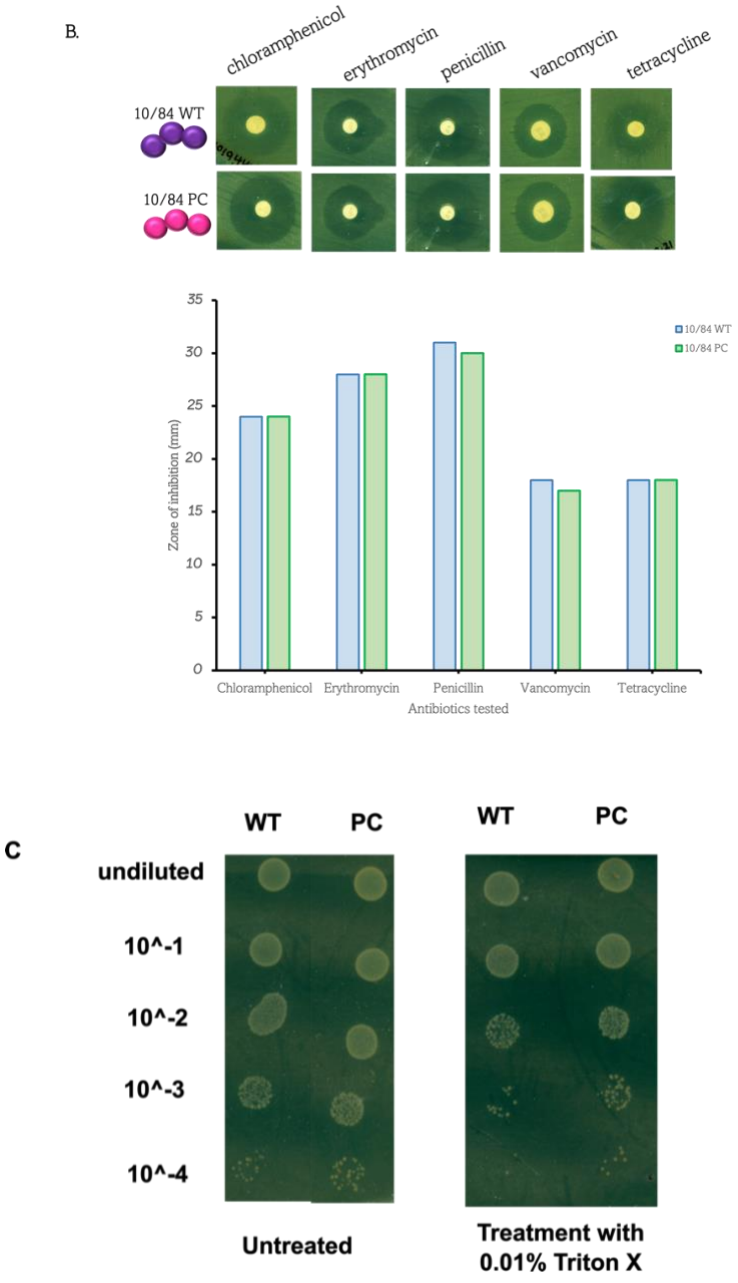


Figure 3.1: Prophage loss influences cell wall integrity in CNCTC 10/84

A) Comparison between wildtype and prophage-cured CNCTC 10/84 strain of phenotypic features; colony morphology, chain length, and beta-hemolytic activity. B) Antibiogram comparison between wild type and phage-cured strains. C) Cell membrane permeability assay comparing wildtype and phage-cured strains treated with 0.01% triton X.

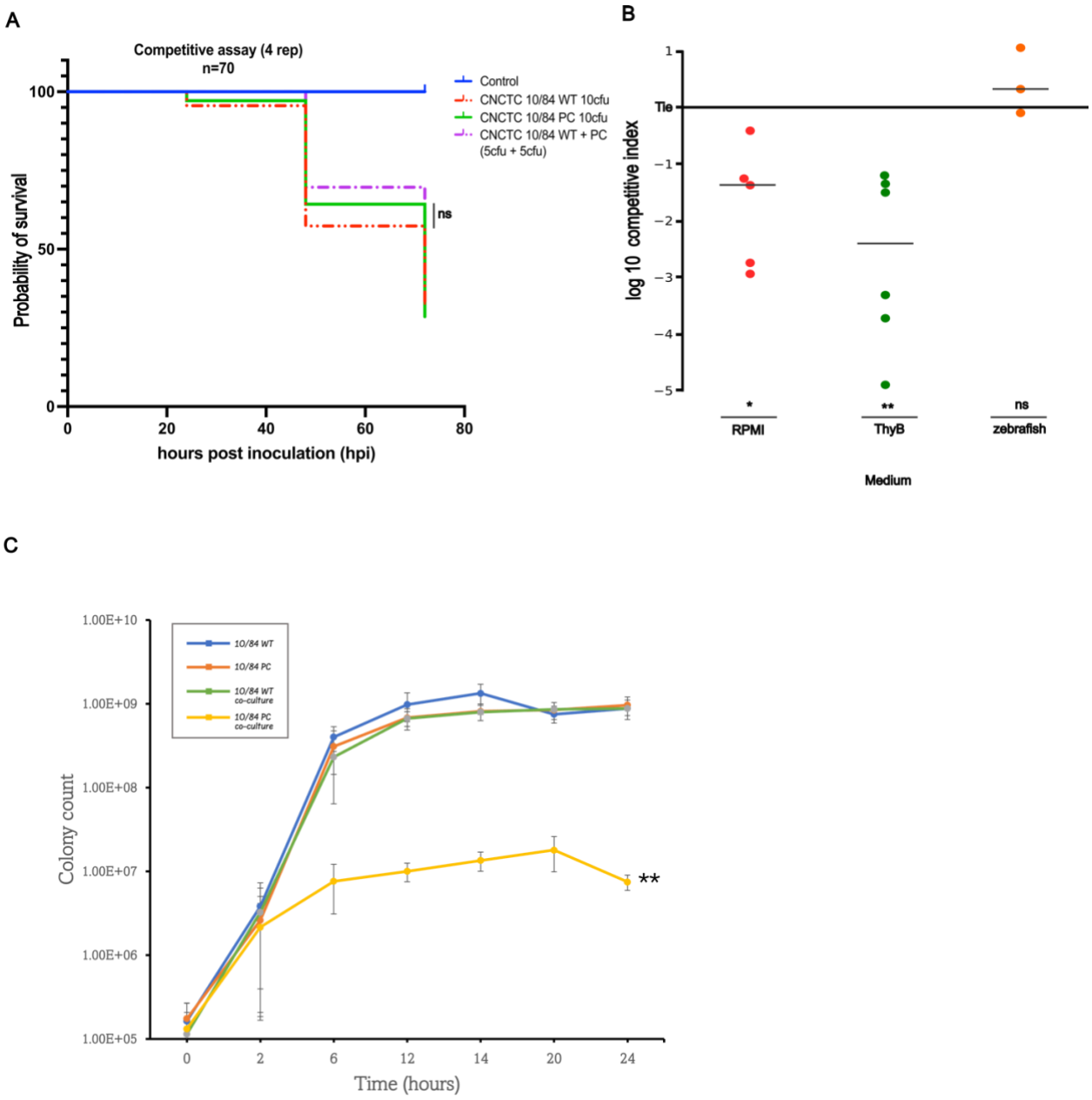


Figure 3.2: Prophages may provide a competitive advantage to CNCTC 10/84

A) Two-day-post-fertilization (dpf) zebrafish were challenged with 10cfu of CNCTC 10/84 lysogen (wildtype: WT) and non-lysogen (phage-cured (PC) in solo infection and 5cfu of each strain in a co-infection assay. Zebrafish larvae were monitored for survival over 72 hours. The probability of survival was plotted against hours post inoculation and represents 4 experimental replicates with a total number of 70 fish. B) Log competitive index of CNCTC 10/84 lysogen (wildtype: WT) and non-lysogen (phage-cured (PC) after growth for 24hrs in rich medium (THYB) and minimal medium (RPMI) and zebrafish co-infection. Each colored point represents the competitive index of 5, 6, and 3 experimental replicates of RPMI, THYB, and zebrafish larvae respectively. Horizontal bars within the points indicate the mean competitive index. The

student t-test paired two samples of means was used to compare the difference between the initial CFU and final CFU of WT and PC. *, $p < 0.05$; **, $p < 0.01$. C) Growth curve over 24h of WT and PC in monoculture and co-culture. Each colored line represents the growth curve of WT, PC, WT in coculture (co-WT), and PC in coculture (co-PC). Error bars indicate standard error from four experimental replicates. The student t-test paired two samples of means was used to compare the difference between WT and PC in co-culture. **, $p < 0.01$.

3.3.3 Gene expression and viability of phage particles suggest an alternative mechanism for the competitive advantage of CNCTC 10/84 lysogen.

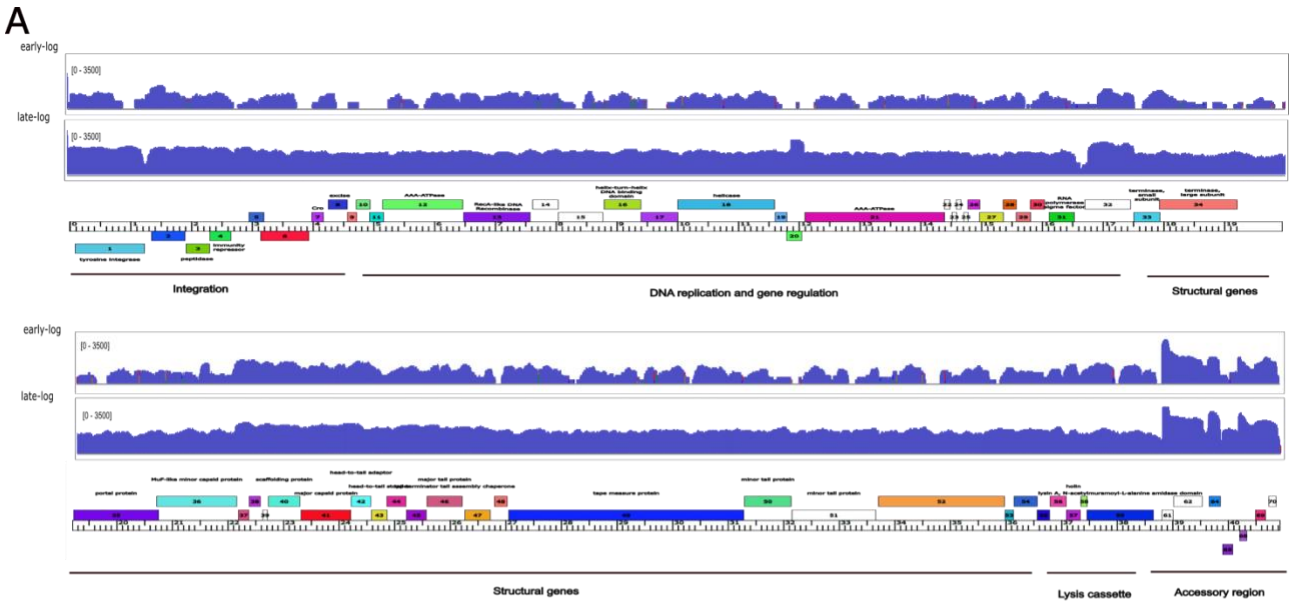
Spontaneous prophage induction in a small subset of a strain can promote phage spread and increase lysogen survival when grown in mixed populations (270). Prophage-carrying bacteria are typically immune to infection by the same virus, but the excision of prophage from the genome and its subsequent release can potentially infect and eliminate competing strains, making temperate phages a valuable biological weapon for the host strains (254). We have shown that the wild type grew twice as much as the prophage-cured strain in co-culture, but the latter recovered when co-cultured with a different GBS strain. This suggests that the CNCTC 10/84 strain prophage may be spontaneously induced, which could lead to the lysis of bacterial cells in the phage-cured strain, ultimately inhibiting its growth. To determine if the prophage is spontaneously induced, supernatants of bacterial culture of the wild type at each growth phase were tested for the presence of phage particles using spot titers. However, plaques were not observed in monoculture or in co-culture. This could either mean that the prophage does not form viable phage particles or the induced phage particles were not present in sufficient quantities to cause the formation of visible plaques.

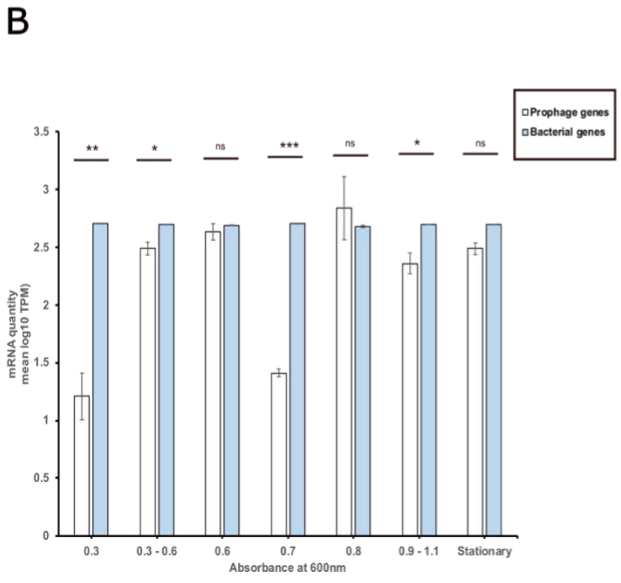
Since no plaques were detected, RNAseq was utilized to gain a more comprehensive understanding of the prophage transcriptome and investigate the possibility of phage release at

undetectable levels. A visual inspection of the transcriptional activity of the prophage was conducted at the early log and late log phases grown in monoculture. Gene expression was observed across the whole prophage region within the wild type (Figure 3.3a), suggesting that the lysogen is not stable and phage particles may be released into the growth medium. Consequently, we aimed to determine the quantity of phage particles released into the growth medium, as a significant release may indicate the potential bacterial lysis of the prophage-cured strain during co-culture. We hypothesized the mean TPM (transcripts per million mapped reads) values for prophage genes that would be higher compared to bacterial genes since the induction of prophage into the lytic cycle leads to an increase in gene expression, We also aimed to determine whether the quantity of phage particles released remains constant during different growth stages, as there was a noticeable difference in growth rate between the two strains throughout (Figure 3.2c). The relative gene expression within each library of the prophage region compared to the bacterial region was evaluated using TPM values as reported in Materials and Methods. The mean $\log_{10}$ TPM for prophage genes was low compared to the bacterial genes at four different time points, and about the same for the remaining three growth phases (Fig 3.3b). The data suggests a certain degree of spontaneous prophage induction may be happening within a subset of the bacterial population of the CNCTC 10/84 wild type (lysogen) but may be occurring at a low frequency.

To determine whether the prophage was inducible and could produce viable phage particles, we over-expressed the excise gene in the wild type and concentrated the supernatant. This resulted in the successful capture of phage particles. The isolated and purified phage, Callidus was confirmed with PCR targeting the attP site and selected phage genes. Next-generation genome

sequencing also confirmed the sequence identity of Callidus. TEM experiments confirmed a phage with *Siphoviridae* morphology with a tail length of 180 nm and a head diameter of 60 nm (Fig 4.3c)





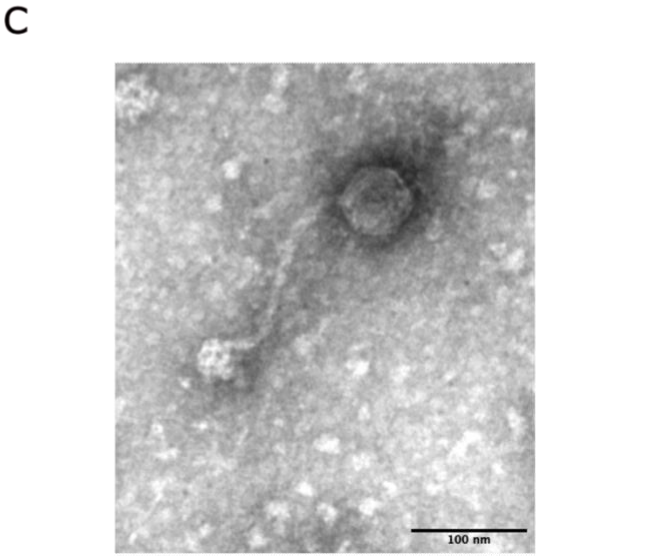


Figure 3.3: Transcription profile of CNCTC 10/84 prophage at different growth phases

A) Transcription profile of Callidus during early log and late log phases. The Callidus genome map, which displays gene boxes colored by their putative function and grouped by function, spans two lines. The density of sequence reads is plotted above each map, and the vertical scales indicate the absolute number of reads. B) Mean relative gene expression from RNA-seq analyses within each library showing using transcripts per million mapped reads (TPM) values for each growth phase. Data were compared with data analysis tool in Excel (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.005$; ns, not significant). RNA-seq analyses were performed on RNA-seq data from the present data set,(261, 264, 265). C) Representative TEM microscopy images of propagated and isolated phage Callidus.

3.4 Discussion

Prophages can contribute to bacterial competition by encoding genes that provide a fitness advantage to the lysogen, allowing it to outcompete other bacteria in the environment (254–256). For example, some prophages encode bacterial toxins that kill other bacteria or provide genes that confer antibiotic resistance allowing survival in the presence of antibiotics while other strains without the prophage would be eliminated. The induction of prophages in response to environmental stimuli, such as exposure to antibiotics, can provide a selective advantage to the lysogen and allow it to dominate the bacterial population (258). Therefore, understanding the dynamics of prophage induction and the consequences of prophage-mediated bacterial competition is important for understanding bacterial pathogenesis and evolution. In this study, a hypervirulent GBS strain was compared with its isogenic strain differing only in the prophage region to determine phenotypic and virulence differences. Additionally, the competitiveness of the wild type, containing the prophage, was evaluated in comparison to the strain that had been cured of the prophage.

No significant phenotypic differences were detected between the wild-type and prophage-cured strains with the assays used in this study. However, exposing both strains to a detergent showed

that the wild type was more vulnerable to autolysis of the cell membrane (Figure 3.1c). This result suggests that the phage-cured strain may have a more robust cell membrane than the wild-type strain or that the wild-type may be more susceptible to membrane-disrupting agents, such as Triton X-100. Alternatively, it is possible that the wild-type strain is more susceptible to lysis because the phage produces lysin, which weakens the cell wall to some degree. The lysogen (i.e., the bacterial cell that contains the prophage) may be more susceptible to membrane-disrupting agents because the prophage itself can produce toxins or other virulence factors that can damage the host cell membrane (271). These virulence factors may be produced in response to certain environmental conditions such as treatment with a detergent or as part of the prophage life cycle. In addition, the lysogen may be less able to repair or replace damaged membrane components due to the diversion of cellular resources toward maintaining the prophage (272). In contrast, the prophage-cured strain lacks the prophage and thus would not produce any virulence factors that could damage the host cell membrane. This strain may also have more cellular resources available to repair or replace damaged membrane components, making it less susceptible to membrane-disrupting agents. This finding stipulates that the prophage may play a role in cell wall integrity therefore it will be interesting to determine if there are differences in the expression of bacterial genes involved in nutrient uptake or stress between the wild-type and prophage-cured strains. If differential gene expression is observed in genes associated with nutrient uptake or stress response, additional experiments will be necessary to investigate these factors. This could help determine whether the wild-type is less efficient at taking up certain nutrients or more sensitive to stressors and whether this difference contributes to its susceptibility to membrane-disrupting agents.

In a zebrafish larvae yolk sac model of infection, the absence of the prophage did not seem to attenuate virulence in this hypervirulent strain (Fig 3.2a). This could be due to several reasons. One reason would be that the prophage may produce virulence factors that can contribute to bacterial pathogenicity, however, they may not play a significant role in this particular infection model. Additionally, it is possible that other factors, such as the ability of the bacteria to evade the host immune system, may be more important determinants of virulence in this context. In zebrafish research, the ability to visualize the innate immune response in the transparent embryo and the ability to study the adult with both adaptive and innate immune functions present limitless opportunities for research in a wide range of areas(273). Yolk sac injections are a common method used to assess general aspects of infection, including dissemination and interactions with the host immune system (274). On the other hand, otic vesicle injections can be used to visualize and study interactions between the bacterial pathogen and specific immune cells of the host (274). Therefore, immune responses between the wild-type and prophage-cured strains can be compared using the otic vesicle model of infection in zebrafish larvae. Additionally, infection of adult zebrafish with streptococcal species is well established ((273, 275) and it is possible that differences in virulence, as well as the adaptive immune response to both strains, may be observed between the two strains in the adult zebrafish.

While it was not surprising that the bacterial load of both the wild-type and prophage-cured strain was similar in the co-infection (Fig 3.2b), the minimal delay in zebrafish response to the presence of both strains suggests that the interaction between the wild-type and prophage-cured strains is complex and may involve factors other than just the presence or absence of the prophage. Therefore, it was not surprising that in a controlled environment, i.e., *in vitro*, it was

determined that the wild-type has a competitive advantage over the prophage-cured strain when co-cultured together (Fig 3.2b). Bacterial competition can happen within or between bacterial species and is often observed as growth limitation or inhibition (252). While it is possible that the competitiveness by the wildtype may be due to growth limitation in the context of metabolism or availability of nutrients, the fact that the prophage-cured strain seems to enter the stationary phase as early as 6 hours after sub-culture (Fig 3.2c), suggests this may not be the case. The growth dynamics of both strains in co-culture indicate that the prophage-cured strain is somehow being inhibited by the wild-type strain. One possibility is that the prophage in the wild-type strain is being induced under non-inducing conditions, a phenomenon known as spontaneous prophage induction (270). Spontaneous induction is usually linked to the activity of an SOS response system (276) which is because of the bacteria encountering stressful conditions. It is therefore not surprising that while it was observed that there were reads that mapped to almost all the prophage genes from the RNA sequencing data (Fig 3.3a), phage particles were not detected from the supernatants of the wild type at different growth phases. Although it is likely that phage particles are present but may not be infective or may not be able to form plaques due to defects in their genome, the evidence that the prophage was induced and could be propagated as shown in Fig 3.3c dispels this notion. A likely possibility is that the phage particles are present in very low numbers or are bound to or sequestered by bacterial cells or other particles in the culture medium, making them inaccessible to detection methods such as spot titers. Additionally, even if there are phage particles present in appreciable numbers we expect to see a decrease in cell viability through infection and amplification of phage particles, however, we only see a steady state of growth in the prophage-cured strain.

Bacteria can produce a variety of secondary metabolites and antimicrobial compounds that inhibit the growth of other bacteria in their environment further contributing to competition (250, 277). Thus, entering the stationary phase early may be the response of the prophage-cured strain to a compound or secondary metabolite produced by the wild-type strain. Bacteria can also coordinate a diverse range of social behaviors through communication with each other (278). Bacteria cooperate with their kin to increase the fitness of the population (279). When the bacterial population reaches a critical cell density, the QS (quorum sensing) system is activated, allowing the coordination of numerous bacterial processes (280). Therefore, it is possible that the prophage-cured strain may be required to reach a specific cell density to survive whatever stressful condition may be happening in the co-culture.

Taken together, this study demonstrates that in GBS, the presence of prophages can confer a competitive advantage that provides fitness benefits to the lysogen. As summarized in Figure 3.4, the competitive advantage may be due to multiple factors. However, it is not entirely clear the factor contributing to this advantage, and additional work is needed to determine specific factors underlying the benefits provided by the prophage.

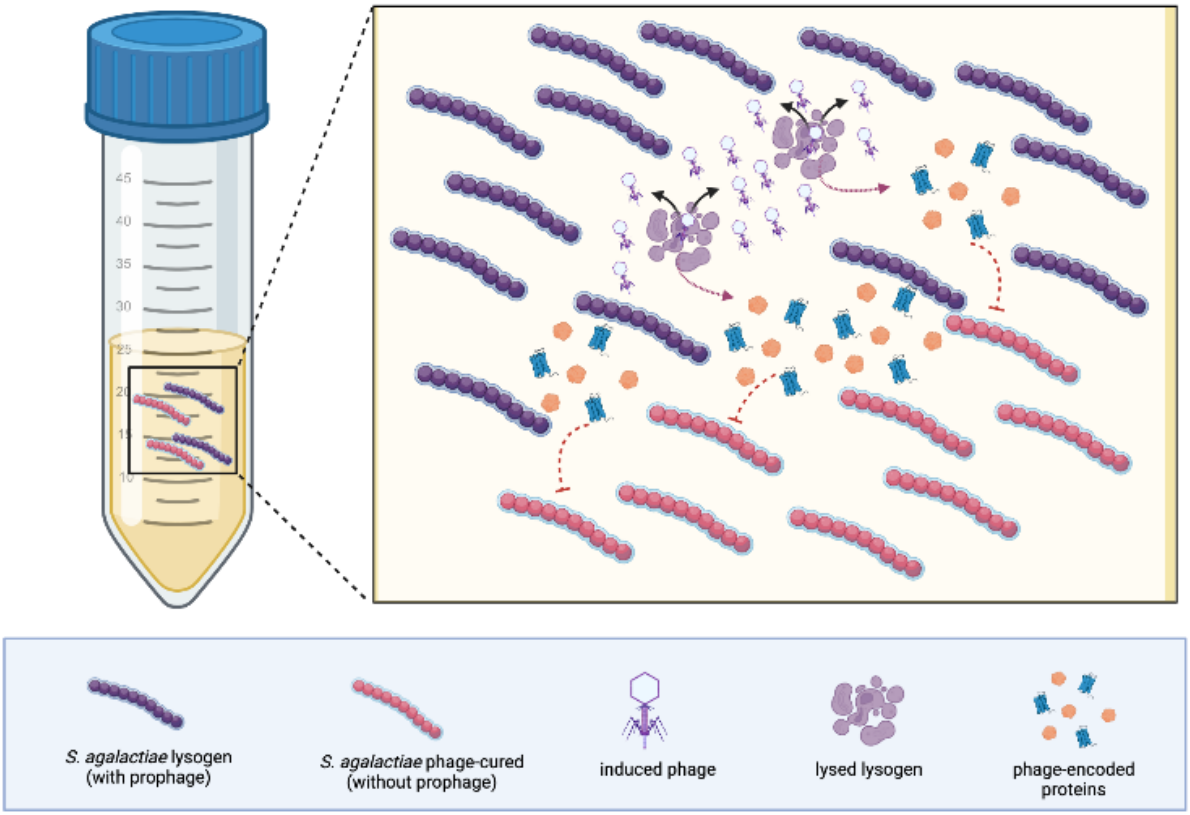


Figure 3.4 Schematic of possible events supporting prophage-encoded advantage to CNCTC 10/84

Co-culture of the GBS lysogen with its isogenic strain, lacking the prophage demonstrates a prophage-encoded competitive advantage. Multiple factors may be contributing to this competitive advantage. The prophage may be induced during co-culture leading to lysis of bacterial cells and release of the prophage. The phage may then lyse the prophage-cured strain leading to growth inhibition. Alternatively, phage and bacteria encoded proteins may be released by the lysogen that may inhibit growth of the prophage-cured strain.

CHAPTER 4

GLOBAL TRANSCRIPTOME CHANGES OF GBS STRAIN, CNCTC 10/84 TO PROPHAGE LOSS

4.1 Introduction

Bacterial-bacteriophage interactions are common in the biosphere (159) and prophages, viral genomes integrated into bacterial genomes, can enhance bacterial colonization, adaptation, and fitness, potentially increasing the opportunity for pathogenic bacteria to cause infection (157, 160, 204, 281). Through lysogenic conversion, a process where phages incorporate their genetic material into the genome of the host bacteria and alter its phenotype, phages can provide new traits, including virulence factors, and metabolic enzymes to the bacteria they infect (141, 157, 161–166). These virulence factors can be transferred to non-pathogenic bacteria or increase virulence by producing toxins. Prophages can also interfere with bacterial gene expression or encode genes that contribute to colonization and fitness. Examples include the STX prophage of the Shiga toxin *Escherichia coli* (282)which transfers virulence factors, and the CTX phage of *Vibrio cholera*, which produces toxins (157). The *Staphylococcus aureus* beta-toxin-encoding gene is disrupted upon integration of bacteriophage φ13 (233) and *Streptococcus pyogenes* secretes the DNase Spd1, encoded by prophage genes, to escape neutrophil clearance during initial colonization and establish infection (28).

GBS prophages exhibit high levels of genetic diversity (120, 121, 188–190) and contain genes that may contribute to bacterial fitness and virulence as shown in Chapter 2. Additionally, lysogenic GBS strains have exhibited a loss of certain catabolic functions, such as the inability to utilize substrates like arbutin, D-cellobiose, D-galactose, and D-ribose (185). However, the role

of prophages in bacterial fitness and virulence has not yet been experimentally determined, particularly in GBS. Comparing isogenic bacterial strains that differ only in their prophage region can provide a better understanding of the biological role of a specific phage and its contribution to bacterial virulence(193, 283). For example, prophage-cured strains have been used to study prophages as a molecular regulatory switch (236, 284). Moreover, studies have investigated the expression of bacterial genes by introducing phages into bacterial strains lacking prophages and comparing them to the wildtype strain. RNAseq analysis demonstrates differential expression of bacterial genes is mostly associated with sugar transport, nucleotide biosynthesis, and metabolism (285, 286). Although this approach provides insight into how phages affect bacterial gene expression, it does not consider the impact of naturally occurring phages that have co-evolved with their bacterial hosts. To our knowledge, no research has investigated differential gene expression in GBS strains by comparing isogenic strains that differ in their prophage regions, particularly naturally occurring prophages. Therefore, such studies could provide valuable insights into the specific genes and pathways affected by prophages in GBS.

This study presents the comprehensive transcriptional changes in GBS upon prophage loss. Through high throughput RNA sequencing, we compared the transcriptome of a wildtype lysogenic GBS strain with that of a non-lysogenic (phage-cured) strain. Our results demonstrate differential expression of multiple bacterial genes upon prophage loss, particularly those involved in metabolic pathways such as ABC transporter pathways.

4.2 Materials and Methods

4.2.1 Bacterial strains, plasmids, and growth conditions

This study used GBS strain CNCTC 10/84 (serotype V, sequence type 26) as a wildtype (WT) strain. Plasmids were maintained in TOP10 (Invitrogen) Escherichia coli and used for plasmid construction and vector propagation. GBS strains were grown in Todd Hewitt broth (Acumedia) supplemented with 2% yeast extract (THY) and on THY agar at 37°C and *E. coli* strains were grown in Luria broth (LB) media (Acumedia) and on LB agar at 37°C. Media was supplemented with antibiotics when needed as follows; chloramphenicol (Cam) 20µg/ml for *E. coli* and 3µg /ml for GBS. Antibiotics were purchased from Sigma-Aldrich. All restriction enzymes were purchased from New England Biolabs.

4.2.2 Generation of a prophage-cured CNCTC 10/84 strain

To cure CNCTC 10/84 of its prophage (550,775 – 591,450), the prophage excision (*xis*) gene (W903-RS03095) was cloned into the expression vector pLZ12-rofA-pro. The open reading frame of the excise gene was PCR amplified from genomic DNA using primers listed in Table 4.1. The 405-bp PCR product and pLZ12-rofA-pro were digested with *Bam*HI and *Pst*I according to the manufacturer's recommendations (New England Biolabs, Ipswich, MA) and gel purified. After ligation of the excise sequence into the vector using T4 ligase (NEB), 5 µl of ligation mixture was electroporated into TOP10 *E. coli* (Invitrogen) according to the manufacturer's recommendations. The sequence and orientation of the excise insert were confirmed by DNA sequencing. The recombinant plasmid was then transformed into competent GBS as previously described _ and cam^R colonies were selected for screening. To confirm the loss of the prophage, cam^R colonies were verified by PCR for the presence of the attB site using primers that span the bacterial attachment site and the loss of a gene specific to the prophage (gp5) (Table 4.1). To ensure that the recombinant plasmid did not influence results in this study,

the prophage-cured CNCTC 10/84 strain was cured of its plasmid by growing cultures overnight at 37°C, double diluted, and growth to late-logarithmic phase at 40°C. Serial dilutions were plated on THY agar plates and plasmid-free colonies were verified by PCR for the presence of the attB site, loss of gp5, and xis gene. Restriction endonuclease digests were carried out according to the manufacturer's recommendations (New England BioLabs, Ispwich, MA)

Table 4.1: PCR primers used in this study.

Description	Primers	Sequence (5' to 3')	Tm	amplicon size
Primers used to amplify excision (xis) gene	5' Xis-Cphage-BamH1	CGC GGA TCC CTT ATC AAA AAT TCT ACT TAC	57.5	405 bp
	3' Xis-Cphage-Pst1	AAA ACT GCA GCT TCA TTA TGT TAT ACT CCT AAC TG	58.5	
Primers used to detect loss of gp5 in Callidus	5' GBS-Cphage-Gp5-fwd	GAG CAT TTT CAG TGG GTC GC	56.8	260 bp
	3' GBS-Cphage-Gp5-rev	CAC TTT CCA AGA AAC AAC CTC AGG	56.2	
Primers used to detect bacterial attachment site in CNCTC 10/84 after loss of prophage	5' GBS-Callidus-attP-fwd	TGC CAA CCG AAA AAC CTA AC	53.7	150 bp
	3' GBS-Callidus-attP-rev	GTA GGA CAC GCT GAT GCA AA	55.4	
	5' GBS-Callidus-attB-fwd	TGG AGG ATT TGT TAA CAT GG	50.2	455 bp
	3' GBS-Callidus-attB-rev	TTC CTT TGA GCT CTC TAG TCG	53.7	

4.2.3 RNA isolation

Wild type and prophage-cured CNCTC 10/84 were grown anaerobically overnight in THY at 37°C. Three replicate cultures of each strain were diluted 1:50 into 50mL THY, incubated anaerobically at 37°C, and grown to early logarithmic (OD600 ~ 0.3) and late logarithmic phase (OD600 ~ 0.7). RNA was isolated using RNeasy mini kit (Qiagen, Germantown, MD) following the manufacturer's protocol with a few modifications. Harvested cells (4 ml) were treated with RNAProtect Bacteria Reagent (Qiagen) and incubated for 5 min at room temperature. Cells were

then centrifuged and resuspended in 100 µl of TE (10 mM Tris, 1 mM EDTA, pH 8.0) with 10 µl lysozyme (25mg/ml, Sigma cat #) and mutanolysin (50U/ml-Sigma cat#). Cells were lysed using Lysing Matrix B tubes (MP biomedicals, Irvine, CA) and homogenized for 5 mins in the TissueLyser LT (Qiagen) set for 5 min at 50 Hz. To remove contaminating DNA, RNA was treated with DNAse on the column (Qiagen) followed by a second DNAse treatment using the Turbo DNA-free Kit (Thermo Scientific, Waltham, MA) according to the manufacturer's recommendations. RNA quantity was determined with the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Montchanin, DE, USA). Samples were stored at -80°C until ready to use.

4.2.4 Bioinformatic analysis

RNA was sent to the Hubbard Center for Genomes studies for quality control analysis, library preparation, and paired-end sequencing on an Illumina Hiseq2500. Raw sequencing data files were uploaded to the public Galaxy server at usegalaxy.org (287). Read quality was determined using FastQC v0.11.5 (206)and processed using Trimmomatic with the FastQC output as a guide(207). Reads were aligned to the CNCTC 10/84 reference genome sequence (NCBI RefSeq NZ_CP006910) using Bowtie2 version 2.1.0 (266)with default (– end-to-end) alignment mode and by specifying –sensitive as an additional parameter. Reads that mapped only once to the genome (uniquely mapping reads) were extracted from SAM files by filtering for the "XS:" tag used by bowtie2 for reporting secondary alignments for a given read. Only uniquely mapped reads stored under SAM file format were used for the subsequent operations. For visualization of transcriptional activity across the genome, SAM files were converted to corresponding binary format (BAM files) with SAMtools version 0.1.19 (214)and viewed with Integrative Genome

Viewer (IGV). Processed reads were quantitated using Kallisto (268) by aligning in a strand-specific orientation to the CNCTC 10/84 transcriptome using a FastA coding transcript (retrieved from GenBank). The R statistical package, DeSeq2 (288) was used for pair-wise comparisons of gene expression from the Kallisto quantification output. Genes with low expression levels i.e. reads < 10 were removed. Genes were considered significantly regulated if Log2 fold change (Log2FC) was greater than 1.0 and the p-adjusted (padj) was less than 0.05. Differentially expressed genes were analyzed with a publicly available gene set enrichment search tool, Genome2D for functional patterns (289). Network analysis of gene sets was performed using the string app in Cytoscape (290). Figures of differentially expressed genes were generated in RStudio(225). Figures of network analysis were generated in Cytoscape (290). All figures were edited with Inkscape (www.inkscape.org).

4.2.5 Data availability

The RNA-Seq data will be deposited in Gene Expression Omnibus (GEO) (currently in progress)

4.3 Results

4.3.1 Transcriptomic changes in CNCTC 10/84 following prophage loss

RNA-Seq analyses were performed on the lysogenic (WT) and non-lysogenic (phage-cured; PC) strain of CNCTC 10/84 grown in ThyB medium at early-log and late-log growth phases to assess the impact of prophage loss on their transcriptome (Figure 4.1a). These time points were selected because of the high metabolic activity and cell division of bacteria during the logarithmic phase (291). A summary of the RNA-seq data is presented in Table 4.2, and the complete transcriptomic data for the WT and PC strains are reported in Table C.1.

Table 4.2: Summary of RNAseq data.

Strain	Timepoint	Replicate	Total reads
CNCTC 10/84 wildtype	early-log	1	5398640
CNCTC 10/84 wildtype	early-log	2	9461694
CNCTC 10/84 wildtype	early-log	3	11625014
CNCTC 10/84 wildtype	late-log	1	9641009
CNCTC 10/84 wildtype	late-log	2	11851535
CNCTC 10/84 wildtype	late-log	3	12805969
CNCTC 10/84 prophage-cured	early-log	1	11300558
CNCTC 10/84 prophage-cured	early-log	2	12083697
CNCTC 10/84 prophage-cured	early-log	3	9689393
CNCTC 10/84 prophage-cured	late-log	1	13889286
CNCTC 10/84 prophage-cured	late-log	2	12661432
CNCTC 10/84 prophage-cured	late-log	3	9274564

Principal component analysis (PCA) was used to assess the data variability and quality (Figure 4.1b), which confirmed that the data is reproducible and shows a correlation between biological replicate samples. Additionally, the PCA analysis revealed unique patterns of gene expression in the bacteria when the prophage is present or absent, accounting for 82% of the variance in the data.

Figure 4.2a presents a plot comparing the relative expression of bacterial genes in the PC strain compared to that of the WT CNCTC 10/84 strain across the whole genome. The results showed

that approximately 19% (374/1967) of all transcripts exhibited differential expression during each experimental growth phase (Figure 4.1c; 4.2a), indicating transcriptome plasticity in response to prophage loss during the log phase. The most differences in expression between the WT and PC strains were observed during the late logarithmic phase (LL), where 12.5% of all bacterial transcripts were differentially expressed (Fig 4.1c; 4.2c.).

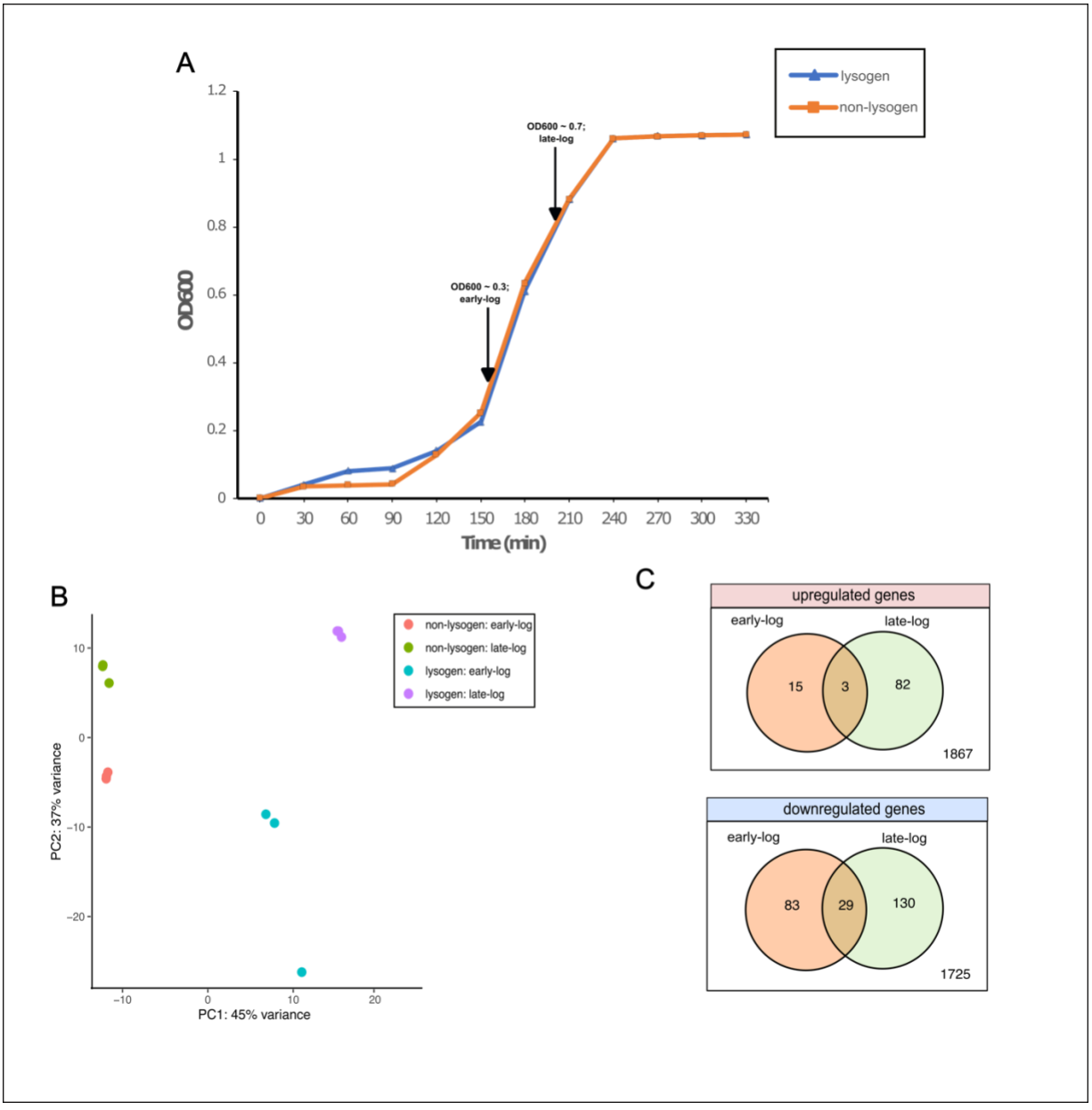


Figure 4.1: Quantitative differences in gene expression between lysogen (WT) and non-lysogen (PC) at different growth phases

A) Growth of GBS CNCTC 10/84 in THY. Arrows mark time points of sample collection for RNA isolation. B) PCA plot analysis of RNAseq data. Each circle represents a single biological replicate. Each experimental condition is designated with a separate color. Distinct clusters of three replicates denote highly reproducible reads. C) Venn diagram of differentially expressed genes at early logarithmic (early-log) and late logarithmic (late-log) growth phases. Venn diagram with red header indicates upregulated genes and blue header indicates downregulated genes. Orange shade represents bacterial genes at early log and green shade represents bacterial genes at late log. Differentially expressed genes represent genes with a $\log_2$ fold change of 1.

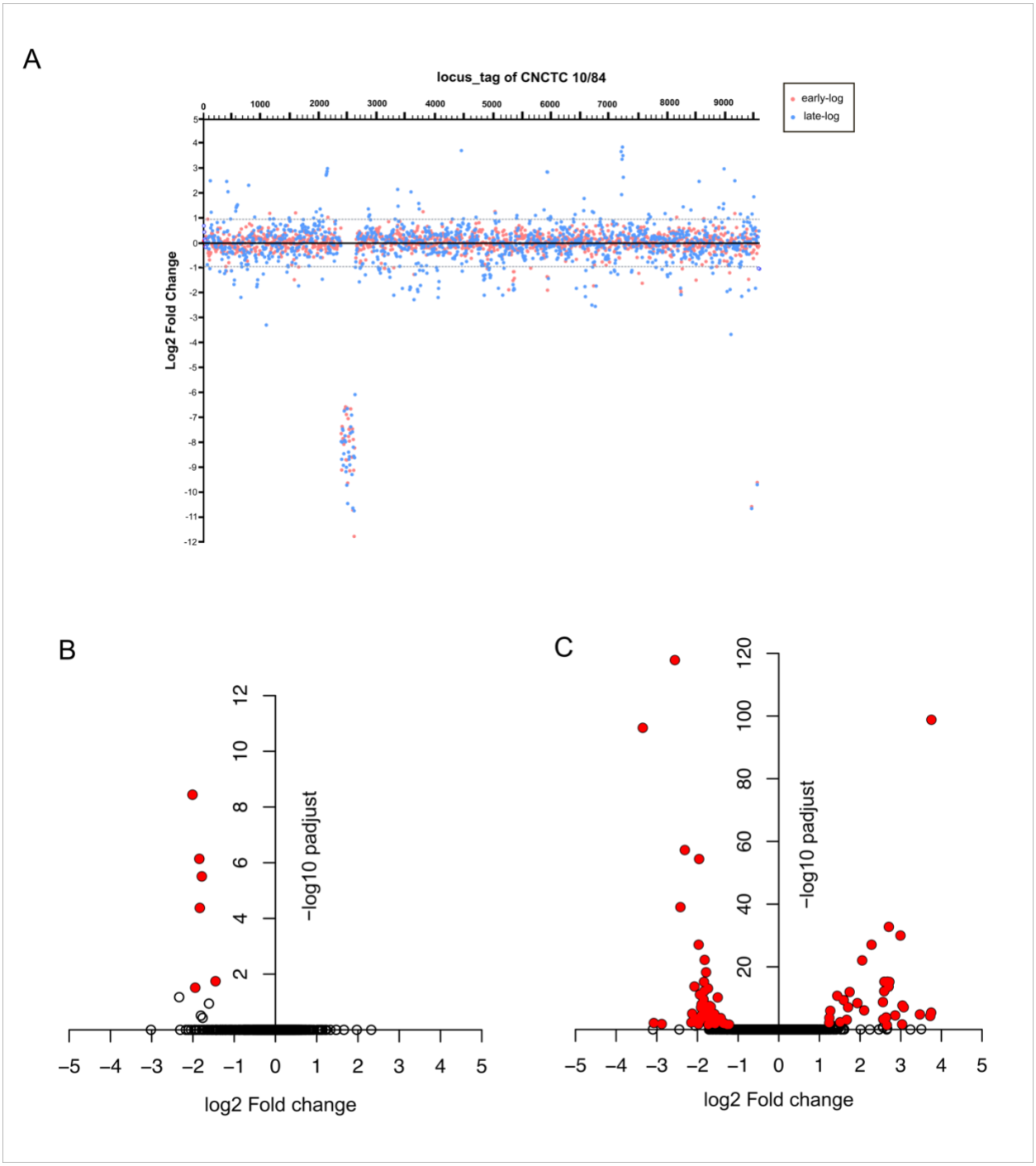


Figure 4.2: Differential expression of bacterial genes in CNCTC 10/84

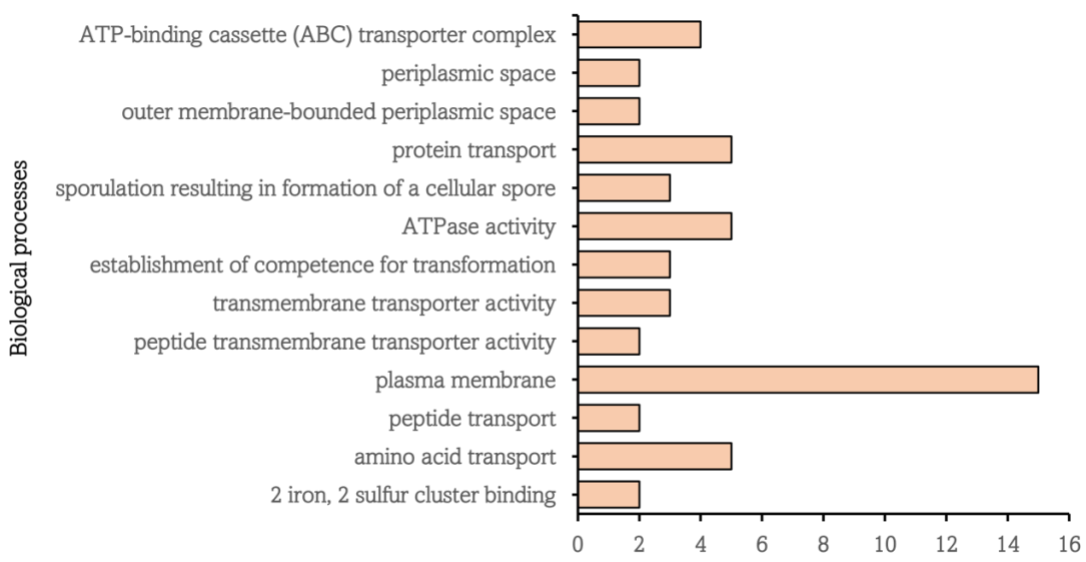
A) Dot plot of differential gene expression across the whole genome of CNCTC 10/84. Each dot corresponds to a single gene. Genes are shown along the horizontal axis in reference to their locus tag, which is indicated on the top horizontal scale. The log2 fold change in expression of the non-lysogen relative to that in the lysogen is presented on the y-axis. The horizontal dotted lines indicate the 1-fold change cutoff range that was used in the analysis. Genes expressed during early log are colored in red, and those that are expressed during late log are colored in

blue. Genes with a log2 fold change less than -5 are prophage genes. Volcano plot showing differentially expressed bacterial genes of non-lysogen compared to lysogen at early log and late log phases are shown in B and C. Points in red represent significantly expressed genes with Padj < 0.05. The total number of analyzed genes was 1967 and they are represented by the average log2 fold change values.

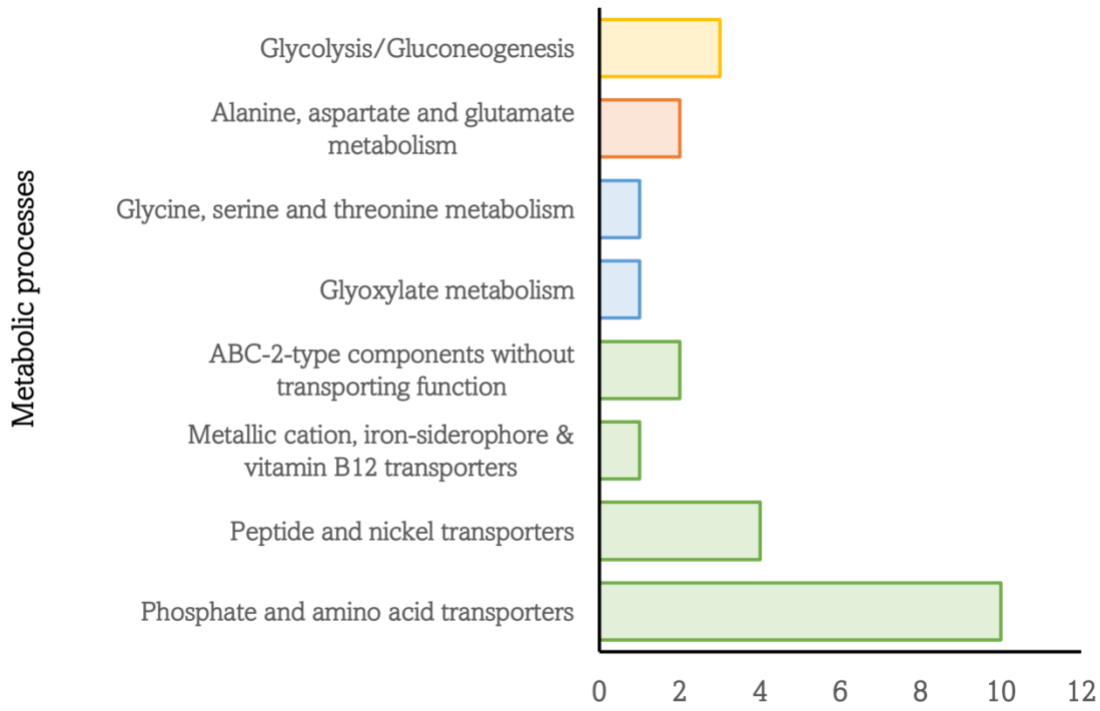
4.3.2 Functional pathway analysis reveals a potential prophage impact on the transcriptome of CNCTC 10/84.

Considering the observed differential gene expression of nearly one-fifth of bacterial genes, we investigated the functional pathways affected by the loss of the prophage to gain insights into how the bacteria was impacted at a molecular level. Examining genes that displayed significant changes in expression (log2 Fold change 1; adjusted p-value < 0.05) at both time points, resulted in a list of 95 genes. Gene ontology (GO) enrichment analyses were performed on the list of differentially expressed genes, of which 13 GO terms were associated with cellular processes like protein and molecule transport, energy metabolism, competence for transformation, sporulation, and nutrient uptake by bacteria (Fig 3.3a). The most enriched GO term was “plasma membrane”, indicating that a large number of differentially expressed genes are involved or related to this structure.

A


Biological processes
ATP-binding cassette (ABC) transporter complex
periplasmic space
outer membrane-bounded periplasmic space
protein transport
sporulation resulting in formation of a cellular spore
ATPase activity
establishment of competence for transformation
transmembrane transporter activity
peptide transmembrane transporter activity
plasma membrane
peptide transport
amino acid transport
2 iron, 2 sulfur cluster binding
0
2
4
6
8
10
12
14
16

B


Metabolic processes
Glycolysis/Gluconeogenesis
Alanine, aspartate and glutamate metabolism
Glycine, serine and threonine metabolism
Glyoxylate metabolism
ABC-2-type components without transporting function
Metallic cation, iron-siderophore & vitamin B12 transporters
Peptide and nickel transporters
Phosphate and amino acid transporters
0
2
4
6
8
10
12

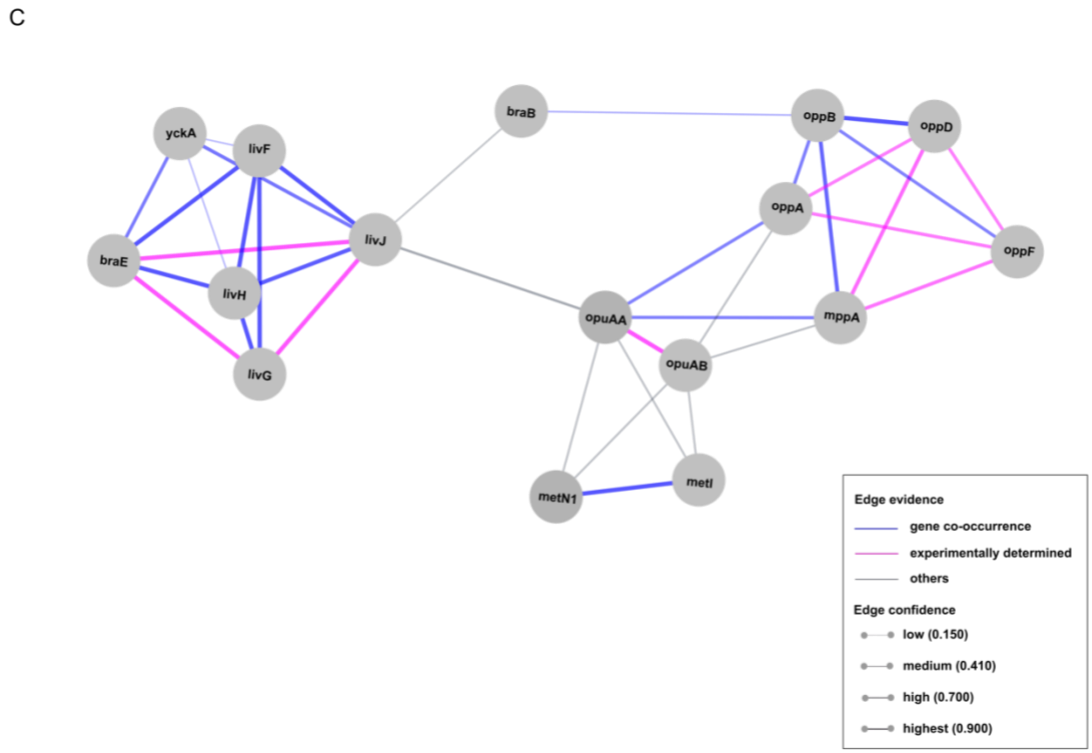


Figure 4.3: GO and KEGG pathway enrichment analysis

A) Gene ontology pathway enrichment analysis of bacterial genes with the most change in gene expression from early log to late log with the loss of the prophage. The DE genes were involved in 13 gene ontology groups defined by Genome2D. B) KEGG pathway enrichment analysis on over-expressed DE genes at late-log in the prophage-cured strain. C) Protein-protein interactions between genes involved in the ABC transporter pathway. Nodes represent each gene from the ABC transporter pathway overexpressed in the prophage-cured strain and edges indicate protein-protein associations. Edge color represents evidence of protein interactions, blue for gene co-occurrence, purple for proteins whereby interactions have been experimentally determined, and grey for all other interactions including from curated databases, gene fusions, gene neighborhood, text mining, co-expression, or protein homology. The density of the edges represents the extent to which a network is fully connected, and the edge confidence is based on a confidence score cut-off of 0.41.

To further investigate the functional pathways affected by the prophage loss in CNCTC 10/84, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. The most enriched pathways were related to metabolic processes, including pyruvate metabolism, glycolysis/gluconeogenesis, and amino acid metabolism, as well as the ABC transporters pathway, which plays a crucial role in nutrient uptake and multidrug resistance in

bacteria (Figure 3.3B). Notably, the differential expression of the genes involved in the metabolic pathways was observed only during the late log phase. Specifically, genes involved in alanine, aspartate, and glutamate metabolism (argG and argH) were upregulated, while genes involved in pyruvate metabolism (pdhB, pdhC), glycolysis/gluconeogenesis (pflB, gloA), and fructose bisphosphatase (fbp) were downregulated (Figure 4.4).

Among the 95 significantly differentially expressed genes, 17 genes were involved in ABC transporter activities, such as phosphate and amino acid transport (livG, livH, livG, metI, metN1), peptide and nickel transport (oppA, oppB, oppD, oppF), metallic cation, iron-siderophore, and vitamin B12 transport, and ABC-2-type components without transporting function (brae, braD).

4.3.3 ABC transporter genes are differentially expressed in response to prophage loss

As most of the genes that showed differential expression were linked to ABC transport, we conducted a protein-protein interaction analysis to explore the functional associations between these genes. The analysis generated a network of 16 nodes and 36 edges with an average clustering coefficient of 0.683 and a network diameter of 4.5. The PPI enrichment value was < 1.0e-16, indicating that the interactions between the genes were higher than expected. This suggests that the proteins are likely biologically interconnected. Notably, we identified several highly connected nodes, including oppA, oppB, oppD, oppF, and mppA, which are known to interact experimentally. We also analyzed the normalized gene expression using TPM and found that 12 out of the 17 genes were upregulated in the prophage-cured strain during the late log phase. Among these, livG, livH, livF, and mppA showed the highest fold change of almost 4.

Only two genes, opuAA and opuAB, which are associated with glycine betaine transport, showed changes in gene expression in both early and late log phases.

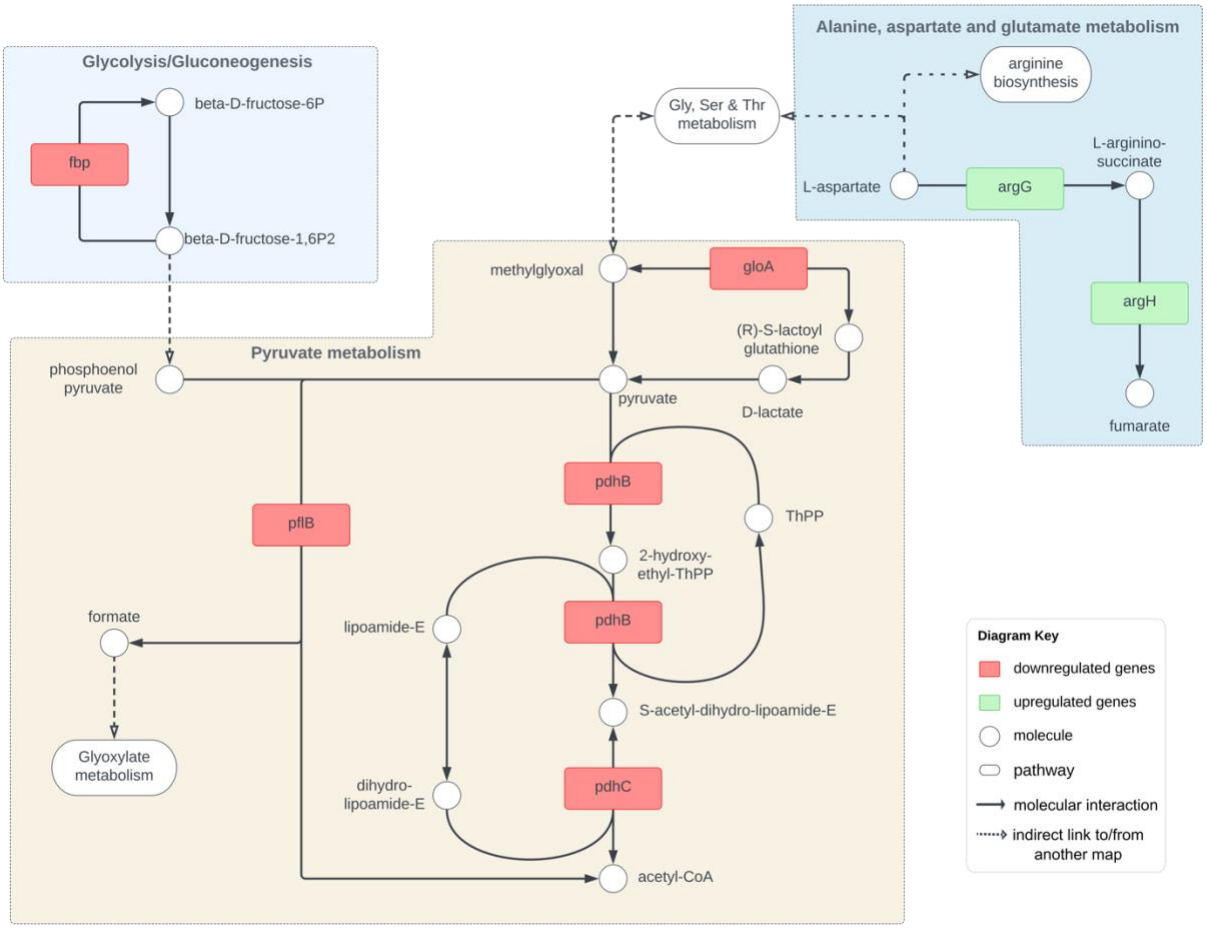


Figure 4.4: Simplified representation of genes within KEGG pathway

Simplified representation of metabolic pathways and differentially regulated genes (log2 fold change ± 1; adjusted p-value < 0.05). Each pathway is represented by a unique color. Red shaded rectangles represent downregulated genes and green represents upregulated genes. Round shapes indicate molecules involved in the pathway and pathways are shown by oval shaped objects. Solid lines show the molecular interactions and dotted lines show indirect link to/from another pathway map. Arrows indicates source to target molecule.

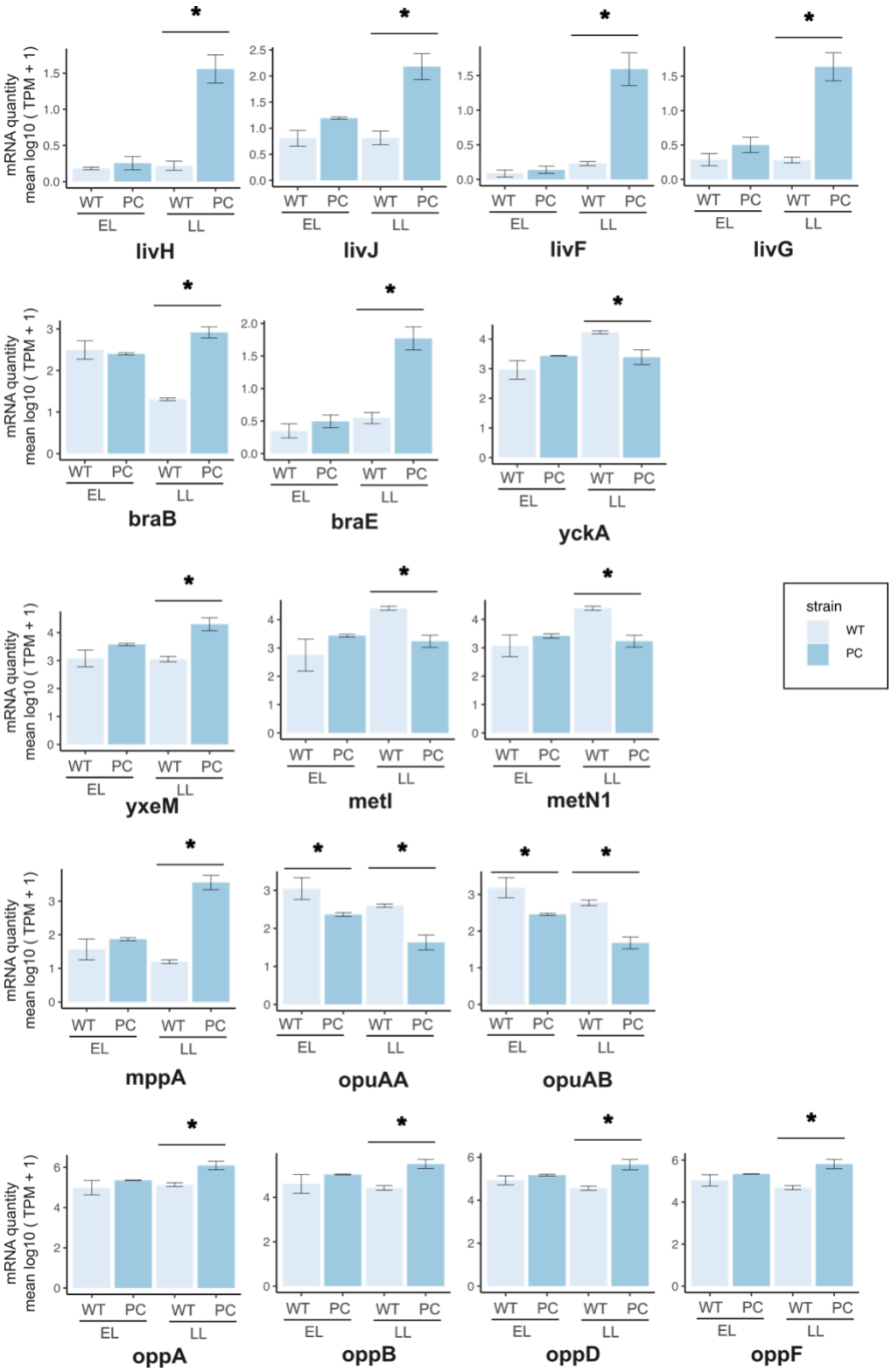

mRNA quantity
mean log10 (TPM + 1)
WT
PC
EL
LL
livH
livJ
livF
livG
braB
braE
yckA
yxeM
metI
metN1
mppA
opuAA
opuAB
oppA
oppB
oppD
oppF
strain
WT
PC

Figure 4.5: Differentially expressed genes involved in ABC transport

Expression of genes involved with ABC transporters from RNA-seq analyses showing mean transcripts per million mapped reads (TPM) values for each strain lysogen; (WT) vs non-lysogen (PC) and timepoint; early-log (EL) vs late-log (LL). Data were compared using DESeq2 (*, P-adj , 0.05; ±2-fold).

4.4 Discussion

Temperate phages unquestionably have a significant impact on bacterial evolution (127), and through lysogeny, phages have developed a mutually beneficial relationship with bacteria. Several studies have described the transcriptome of prophage genomes during lysogeny in various bacterial hosts such as *E. coli*, *L. lactis*, and *S. thermophilus* (232, 285, 292, 293). Previous studies by (285, 286) have provided limited information on the effect of prophage presence on bacterial gene expression, leaving a knowledge gap regarding the impact of naturally occurring prophages on bacterial gene expression. This study aims to address this gap by analyzing the transcriptome changes in the GBS clinical isolate CNCTC 10/84 in response to the loss of a naturally occurring prophage, thereby providing insight into the impact of the prophage on bacterial gene expression in this specific GBS strain.

Around 19% of GBS genes exhibited differential expression globally, with approximately 12.5% of this differential expression happening in the late-log growth phase. In general, bacteria enter the early-log phase after the lag phase, where they begin to multiply, while the late-log phase is characterized by actively growing cells that may be preparing to enter the stationary phase. The observation of variation in the genes differentially expressed between these two phases, coupled with the absence of the prophage, implies that the prophage may play a role in regulating genes during different phases of growth.

During the early-log phase, the downregulation of differentially expressed genes was associated with arginine deaminase, glycine betaine transport, and the integral component of the membrane. However, in the late-log phase, most differentially expressed genes were related to ABC transporters and metabolism. The study found that the loss of the prophage led to differences in the expression of genes related to different growth phases, suggesting that the prophage may be involved in modulating genes at different stages of growth. Specifically, 17 different ABC transporters were identified that may be regulated or indirectly controlled by the prophage.

ABC transporters are responsible for exporting a variety of glycans in cell-surface glycoconjugates in both Gram-positive and Gram-negative bacteria (294) and have been linked to prophages. For example, the substrate binding protein DppA1 of the ABC transporter DppBCDF, which is responsible for the uptake of dipeptides and tripeptides regulates Pf5 prophage of *Pseudomonas aeruginosa* to reduce biofilm formation (295). Deleting dppA1 genes induced bacteriophage Pf5 genes during biofilm formation and increased lytic phage particles by a million-fold, indicating the importance of the phage in stress response or gene transfer within biofilms (295).

Notably, the results of the current study indicate that ABC transporters involved in oligopeptide transport were upregulated in the phage-cured strain during the late log phase, suggesting that the presence of the prophage leads to a downregulation of these genes. This modulation of oligopeptide transport in response to the environment may be influenced by the prophage, but further research is needed to investigate the specific role of prophages in regulating these genes.

To summarize, this study aimed to explore the overall transcriptomic response of GBS CNCTC 10/84 to the loss of prophage. Although only a few genes were differentially expressed during early-log phase, a significant proportion of bacterial genes (15%) showed differential expression during late-log phase. These findings suggest that the absence of the prophage triggers cellular changes in CNCTC 10/84, which indicates that GBS prophages might modulate bacterial genes to provide a fitness advantage (Fig 4.6). The diversity of prophages found in various GBS strains (chapter 2) implies that a specific prophage could offer unique advantages to one strain, while a different prophage could offer different advantages. Therefore, this study provides only a snapshot of how prophage might be modulating bacterial gene expression in a single GBS strain. Further research is necessary to investigate prophage impact on bacterial gene expression in other GBS strains.

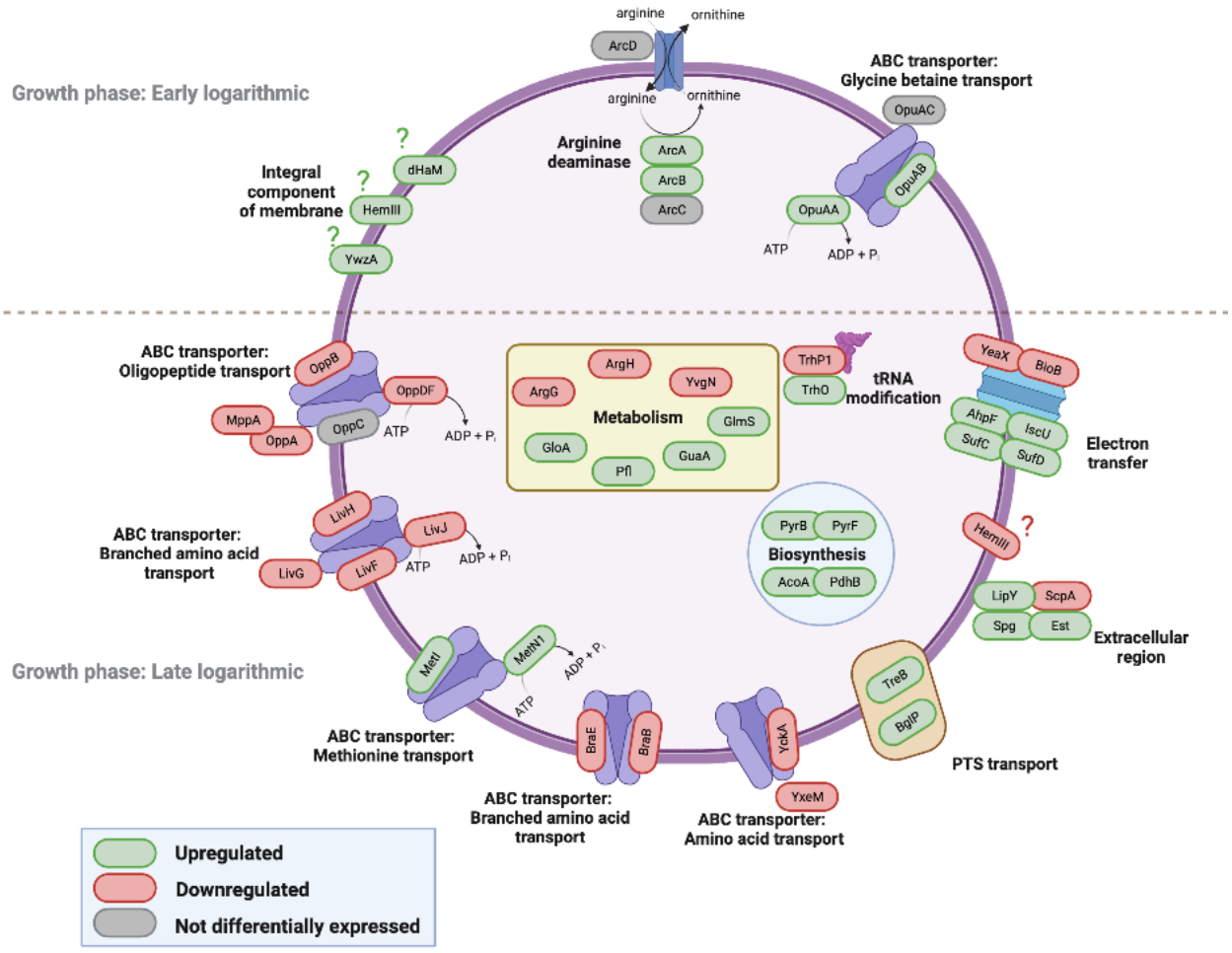


Figure 4.6: Model of bacterial gene expression in CNCTC 10/84

This model depicts the significantly expressed bacterial genes in CNCTC10/84 when the prophage is present, grouped according to their respective gene ontology terms. Genes positively impacted by the prophage are shown in green, the negatively impacted genes are in red, and the genes that are not significantly differentially expressed are in grey. The dashed lines represent the separation between differential gene expression in the early log and late log phases. Created with BioRender

CHAPTER 5

FUTURE DIRECTIONS AND CONCLUSIONS

The work presented in my dissertation explores the role of prophages in *Streptococcus agalactiae*, a bacterium commonly found in pregnant women that can cause severe infections in neonates, including sepsis and meningitis. Accurate detection of maternal colonization and proper antibiotic prophylaxis during labor are crucial to preventing early-onset GBS disease in neonates.

Several studies have identified specific GBS serotypes and clonal complexes associated with high rates of neonatal disease. For example, serotypes Ia, Ib, II, III, and V are commonly associated with neonatal infections, with III being the most virulent and common worldwide. Similarly, clonal complexes such as CC17 and CC19 have been found to be associated with neonatal disease in various geographical locations.

Molecular epidemiology studies can provide important information on the distribution, virulence, and evolution of GBS. For instance, identification of high-risk serotypes or clonal complexes can guide the development of effective vaccine strategies against GBS colonization and infection, reducing neonatal morbidity and mortality caused by this pathogen.

GBS acquires new genetic traits as it evolves, affecting its virulence, resistance to antibiotics, and other factors that impact human health and disease treatment. These new genetic traits can be acquired through horizontal gene transfer with the help of phages, plasmids, or transposons.

Prophages, viral genomes that integrate into bacterial genomes, are known to influence bacterial evolution by providing a mechanism for the exchange of genetic material among different strains. About 10% of strain-specific genes in GBS are thought to be encoded by prophage elements (114), making it essential to understand the role of prophages in GBS pathogenesis to better understand how this pathogen evolves and adapts to changing environmental conditions.

While several studies have identified prophages in GBS, our comparative genomics approach is the first to provide a comprehensive analysis of prophages in clinical isolates of GBS. We identified 42 diverse prophages that can be divided into clusters based on their genomic content. These prophages integrate at different sites in the bacterial host, and prophages within the same clusters tend to have the same attachment sites. Although there is no clear association between prophage clusters and serotypes/clonal complexes, it is possible that a larger dataset is needed to establish this relationship. In our research, we discovered that GBS has several beneficial prophage genes, including paratox. This protein, which is known to inhibit the uptake of DNA in *S. pyogenes*, was found in both the chromosome and prophage genomes of GBS. which is a remarkable finding that has not been previously reported. By studying the function of paratox in GBS and investigating its relationship with the prophage genome, we may gain valuable insights into its significance. Additionally, we identified a gene called holtox located upstream of paratox whose function is not yet known and requires further exploration.

A comprehensive analysis of the impact of prophages on bacterial fitness and virulence in GBS is still lacking, despite the identification of several GBS prophages in previous studies. As part of our efforts to address this, we generated an isogenic strain of GBS lacking a prophage and compared its phenotype to that of the wild-type (lysogenic) strain. Our results showed that the

prophage provided a competitive advantage to the lysogen of CNCTC 10/84, possibly by enabling it to better compete for nutrients or resist other bacterial species. Bacterial competition for nutrients in a polymicrobial environment may play a role in the ability of CNCTC 10/84 to survive. Determining the specific mechanisms by which the prophage provides a competitive advantage to the lysogen would provide new knowledge on how prophages increase bacterial fitness. Similarly, it is crucial to investigate whether the growth inhibition observed in the phage-cured strain is due to a low level induction of the prophage in the wild-type strain and release of phage particles that result in subsequent infection and lysis of the phage-cured strain. Overexpression of the prophage repressor protein in the wild type strain could help to determine if the growth inhibition observed in the phage cured strain is actually due to a low level of bacterial killing from phage infection. Further work, such as competitive assays with other GBS strains, would be necessary to gain a better understanding of the mechanism by which the prophage is providing a competitive advantage to the lysogen. We also observed a delay in overall killing with the prophage-cured strain compared to the wild type *in vivo*, but the mechanisms underlying this effect and its relevance to other routes of infection and host immune response remain unclear. Therefore, further investigations using different routes of infection and/or animal models and immune assays are needed to fully understand the impact of this prophage on GBS pathogenesis and host interaction.

Although there have been numerous studies on phage gene expression during lysogeny, few have investigated prophage-regulated bacterial gene expression. The finding that the absence of the prophage in CNCTC 10/84 results in a shift in expression of numerous metabolic genes suggests that the prophage has a significant impact on bacterial metabolism. This could be a promising

area of investigation for exploring the bacterial competition observed in the absence of the prophage. However, the study only examined early and late log growth phases, indicating that more research is necessary to understand the impact of the prophage on bacterial gene expression in all growth phases, including lag and stationary phases. Validating the findings with qPCR and generating knockout strains of differentially expressed genes could provide insight into why the metabolic shift occurred.

While this dissertation provides insight into the impact of prophages on bacterial fitness in GBS, it only provides a snapshot of what may be happening. To gain a better understanding of prophage impact, more phage-cured strains need to be generated. Since different serotypes possess diverse prophages, it is plausible that various prophages are affecting different bacterial strains. This emphasizes the need to comprehend the impact of these prophages in diverse GBS strains.

A. APPENDIX A

Table A.1 Genome assembly data on DMC strains. Contigs of bacterial genomes in this study can be found at the NCBI BioProject under accession number PRJNA888223. Sequencing data is presented here.

BioSample	Accession	Isolate ID	Forward Reads	Reverse Reads	Initial Contigs #	Final Contigs #	Total Length of Assembly	Final Coverage
SAMN31274409	JAPCZP000000000	DMC1	3802746	3802746	125	22	2081623	793.178
SAMN31274410	JAPCZQ000000000	DMC2	4377434	4377434	2303	81	2181137	825.015
SAMN31274411	JAPCZR000000000	DMC3	1568594	1612078	48	30	2085749	266.967
SAMN31274412	JAPCZS000000000	DMC4	1091004	1116982	180	37	1993949	227.955
SAMN31274413	JAPCZT000000000	DMC5	2056877	2112458	35	21	2049500	348.178
SAMN31274414	JAPCZU000000000	DMC6	2118761	2165216	53	36	2174718	325.475
SAMN31274415	JAPCZV000000000	DMC9						
SAMN31274416	JAPCZW000000000	DMC10	1398847	1437015	39	29	2086193	248.683
SAMN31274417	JAPCZX000000000	DMC13	2950143	3007619	897	33	1992663	488.422
SAMN31274418	JAPCZY000000000	DMC14	1845749	1903402	97	25	2081059	322.287
SAMN31274419	JAPCZZ000000000	DMC15	1650328	1695264	52	41	2059121	302.371
SAMN31274420	JAPDAA000000000	DMC16	1751664	1783490	119	97	2142229	305.162
SAMN31274421	JAPDAB000000000	DMC17	1812490	1852412	1132	57	2095964	185.529
SAMN31274422	JAPDAC000000000	DMC20	2291597	2357602	705	53	2010660	373.607
SAMN31274423	JAPDAD000000000	DMC21	2399927	2467203	52	28	2117475	385.188
SAMN31274425	JAPDAE000000000	DMC24	1614289	1653096	704	38	2012645	262.725
SAMN31274426	JAPDAF000000000	DMC25	1918765	1974745	529	29	2092268	292.528
SAMN31274427		DMC27	1681195	1681195	598	331	1869127	27.3649
SAMN31274428	JAPDAG000000000	DMC28	1340165	1340165	155	24	2067178	250.38
SAMN31274429	JAPDAH000000000	DMC29	1680386	1680386	837	79	2191516	291.718
SAMN31274430	JAPDAI000000000	DMC30	1622879	1666193	41	25	2159481	279.745
SAMN31274431	JAPDAJ000000000	DMC33	1794251	1829965	42	21	2126729	302.502
SAMN31274432	JAPDAK000000000	DMC34	1843750	1883878	25	25	2033496	306.582

SAMN31274433	JAPDAL000000000	DMC36	1827016	1882035	37	22	2033755	334.477
SAMN31274434		DMC38	1993102	2039692	1204	932	1778117	10.4898
SAMN31274435	JAPDAM000000000	DMC39	1709522	1709522	46	20	2105602	293.148
SAMN31274436	JAPDAN000000000	DMC43	3798110	3798110	443	49	2185156	428.62
SAMN31274437	JAPDAO000000000	DMC47	2373096	2440622	222	38	2022693	398.553
SAMN31274438	JAPDAP000000000	DMC48	4978046	4978046	2741	66	2052193	820.247
SAMN31274439	JAPDAQ000000000	DMC49	2141229	2141229	543	41	2088593	341.366
SAMN31274440	JAPDAR000000000	DMC51	1916532	1947120	165	32	2016641	348.242
SAMN31274441	JAPDAS000000000	DMC52	1453104	1484830	410	90	2149962	228.268
SAMN31274442	JAPDAT000000000	DMC56	1885521	1938687	717	134	2127783	314.763
SAMN31274443	JAPDAU000000000	DMC59	2188495	2248028	65	52	2176593	337.277
SAMN31274444	JAPDAV000000000	DMC61	2099053	2143079	76	30	2015563	362.584
SAMN31274445	JAPDAW000000000	DMC62	717346	731726	150	46	2135495	115.622
SAMN31274446	JAPDAX000000000	DMC64						
SAMN31274447	JAPDAY000000000	DMC66	1535116	1562154	317	35	2093544	243.903
SAMN31274448	JAPDAZ000000000	DMC67	1261600	1291564	34	22	2036351	233.979
SAMN31274449	JAPDBA000000000	DMC68	1769663	1816239	64	17	2018609	319.539
SAMN31274450	JAPDBB000000000	DMC69	1554890	1594488	566	59	2086157	202.212
SAMN31274451	JAPDBC000000000	DMC70	2219361	2286869	196	66	2197855	21.901

Table A. 2: GBS strains and resident prophages. GBS strains with their respective clinical attributes and the resident prophages present in the strains. GBS strains that did not have an intact prophage are indicated under Prophage as 'none'. GBS strains where an intact prophage could not be bioinformatically extracted are indicated as 'unextractable'. n/a indicates not applicable.

Bacterial strain	Serotype	Sequence type (ST)	Clonal complex (CC)	Bacterial accession number/contig number	Prophage	Coordinates
2603 V/R	V	110	19	NC_004116.1	Javan 5	558 773 - 599 346
					Javan 6	1 833 089 - 1 867 188
515	Ia	23	23	NZ_CP051004	phiGBS515	558 712 - 599 345
A909	Ia	7	1	NC_007432	Javan 7	548 935 - 586 159
					Javan 8	654 882 - 700 722
CJBIII	V	1	1	NZ_CP063198	phiCJBIII	610 250 - 658 585
CNCTC 10/84	V	26	26	NZ_CP006910	Callidus	550 755 - 591 450
COHI	III	17	17	NZ_HG939456	none	n/a
NEM316	III	23	23	NC_004368.1	none	n/a
DMC 1	V	1	1	NODE_1_length_599266_cov_280.701108 - NODE_4_length_161306_cov_318.978626	phiDMC1*	1 - 41877/ 157 477 - 161 306
DMC 2	V	827	1	NODE_3_length_166000_cov_356.650576	phiDMC2-1	19 476 - 59 175
				NODE_14_length_42782_cov_314.025741 - NODE_1_length_600106_cov_289.469613	phiDMC2-2*	38 596 - 42 783/ 1- 42 863
DMC 3	Ib	12	12	n/a	none	n/a
DMC 4	III	17	17	NODE_6_length_126496_cov_105.233340	phiDMC4**	69 292 - 108 282
DMC 5	Ia	23	23	NODE_5_length_122198_cov_172.878489/ NODE_1_length_553758_cov_158.382558	phiDMC5	107 152 - 122 203/ 1 - 31 081
DMC 6	II	12	12	NODE_1_length_322061_cov_150.271571	phiDMC6	20 785 - 57 633
DMC 9	III	17	17	NODE_6_length_126496_cov_152.635282	phiDMC9	
DMC 10	Ib	268	12	n/a	none	n/a
DMC 13	IV	468	452	n/a	none	n/a
DMC 14	Ib	12	12	n/a	none	n/a
DMC 15	V	1233	19	NODE_12_length_68891_cov_159.687424	phiDMC15	810 - 39 360
DMC 16	II	22	22	NODE_8_length_60059_cov_158.772108	phiDMC16	10 346 - 54 091

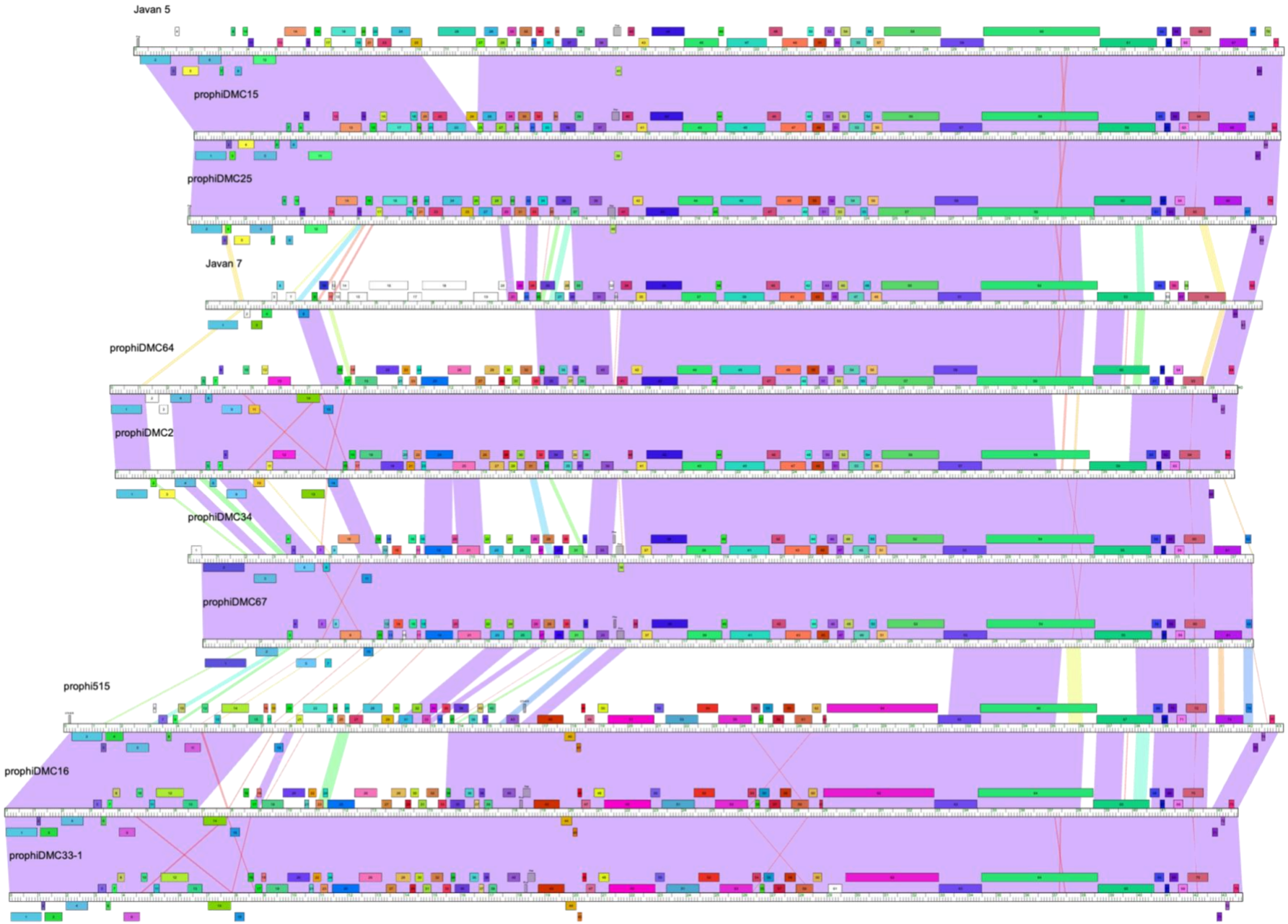
DMC 17	Ia	23	23	NODE_9_length_96718_cov_105.006781	phiDMC17	19 150 - 55 731
DMC 20	V	1233	19	NODE_6_length_120704_cov_196.757259 / NODE_10_length_76177_cov_227.733373	phiDMC20	1 - 4553/ 44 389 - 76 179
DMC 21	II	1	1	NODE_16_length_40263_cov_177.997160/ NODE_13_length_48478_cov_193.092097	phiDMC21-1	36 437 - 40 264/ 1 - 41 593
				NODE_6_length_106360_cov_244.738518	phiDMC21-2	28 465 - 71 862
DMC 24	III	17	17	NODE_35_length_49021_cov_122.545384	phiDMC24	2 385 - 45 552
DMC 25	Ia	23	23	NODE_1_length_482712_cov_137.953509	phiDMC25	337 011 - 375 561
DMC 27	V			NODE_14_length_62871_cov_36.570429	phiDMC27	17 805 - 53 897
DMC 28	V	1	1	NODE_1_length_426519_cov_109.332004/ NODE_5_length_161318_cov_122.882475	phiDMC28	1874 - 44 712/ 1 - 48 538
DMC 29	III	27	19	n/a	none	n/a
DMC 30	Ib	8	8	NODE_8_length_115828_cov_177.017061	phiDMC30	60 089 - 96 673
DMC 33	V	1	1	NODE_3_length_170046_cov_150.976059	phiDMC33-1	19 476 - 63 221
				NODE_14_length_42784_cov_134.518391/ NODE_1_length_600713_cov_128.183997	phiDMC33-2	38 958 - 42 785/ 1 - 42 862
DMC 34	Ia	23	23	NODE_6_length_123204_cov_146.566808	phiDMC34	66 758 - 104 051
DMC 36	III	23	23	NODE_6_length_96719_cov_160.926205	phiDMC36	40 988 - 77 569
DMC 38	IV			n/a	none	n/a
DMC 39	V	1	1	n/a	none	n/a
DMC 43	III	19	19	NODE_10_length_107641_cov_208.298556	phiDMC43-1	
				NODE_14_length_86051_cov_199.828977/ NODE_66_length_13680_cov_174.668634	phiDMC43-2	44 964 - 86 051/ 1 - 3827
DMC 47	IV	452	452	NODE_1_length_286846_cov_195.021334	phiDMC47	107 777 - 153 581
DMC 48	Ia	23	23	NODE_3_length_199465_cov_344.276219	phiDMC48	46 676 - 92 560
DMC 49	III	19	19	NODE_1_length_50795_cov_123.840070	phiDMC49	105 097 - 150 781
DMC 51	IV	452	452	NODE_2_length_286846_cov_168.591931	phiDMC51***	107 777 - 153 581
DMC 52	III	27	19	n/a	unextractable	n/a
DMC 56	II	22	22	n/a	unextractable	n/a
DMC 59	III	19	19	n/a	unextractable	n/a
DMC 61	IV	452	452	NODE_1_length_287056_cov_172.238812	phiDMC61	133 453 - 179 138
DMC 62	Ib	12	12	NODE_11_length_56240_cov_63.336624	phiDMC62	431 - 36 964
DMC 64	II	22	22	NODE_18_length_71407_cov_21.310901	phiDMC64	11 829 - 51 850
DMC 66	IV	468	452	NODE_1_length_286917_cov_106.947613	phiDMC66-1	133 258 - 179 141
				NODE_19_length_66792_cov_160.058636	phiDMC66-	20 947 - 64 154

					2****	
DMC 67	Ia	1733	ND	NODE_6_length_96865_cov_139.495390	phiDMC67	40 454 - 77 715
DMC 68	Ia	23	23	NODE_1_length_651246_cov_138.757236	phiDMC68-1	498 319 - 544 095
DMC 69	III	17	17	NODE_1_length_260544_cov_83.279271	phiDMC69	69 291 - 108 281
DMC 70	III	19	19	n/a	unextractable	n/a

* phiDMC1 same as phiDMC28
** phiDMC4 same as phiDMC9
*** phiDMC51 same as phiDMC47
**** phiDMC66-2 same as phiDMC24

Table A. 3. Prophage gene content similarity matrix. Table shows the gene content similarity matrix of selected prophages with compared to each other based on 'phamilies'. Prophages are considered to be in the same cluster if they have a gene content similarity value > 0.35.

	Javan 8	Callidus	Javan 5	Javan 6	Javan 7	phigbs515	phigbsCJBIII	phigbsDMC15	phigbsDMC16	phigbsDMC17	phigbsDMC2	phigbsDMC20	phigbsDMC21-1	phigbsDMC24	phigbsDMC25	phigbsDMC27	phigbsDMC28	phigbsDMC30	phigbsDMC33-1	phigbsDMC33-2	phigbsDMC34	phigbsDMC36	phigbsDMC43-1	phigbsDMC43-2	phigbsDMC47	phigbsDMC48	phigbsDMC49	phigbsDMC5	phigbsDMC6	phigbsDMC61	phigbsDMC62	phigbsDMC64	phigbsDMC66-1	phigbsDMC67	phigbsDMC68	phigbsDMC69	phigbsDMC9
Javan 8	1.0000	0.0182	0.0183	0.0698	0.0196	0.0354	0.6939	0.0185	0.0367	0.0215	0.0189	0.0674	0.7292	0.6538	0.0185	0.0674	0.6792	0.0213	0.0367	0.7660	0.0202	0.0217	0.0870	0.7872	0.0165	0.0167	0.0167	0.0167	0.0206	0.0167	0.0217	0.0185	0.0167	0.0202	0.0167	0.0104	0.0104
Callidus	0.0182	1.0000	0.0373	0.0256	0.1000	0.0432	0.0090	0.0455	0.0444	0.0333	0.0382	0.0165	0.0180	0.0175	0.0455	0.0165	0.0171	0.0163	0.0444	0.0180	0.0156	0.0336	0.0667	0.0179	0.0559	0.0638	0.0563	0.0638	0.0323	0.0563	0.0336	0.0455	0.0638	0.0156	0.0563	0.6892	0.6892
Javan 5	0.0183	0.0373	1.0000	0.0171	0.3100	0.3091	0.0091	0.9296	0.1765	0.1182	0.4362	0.0252	0.0182	0.0177	0.9296	0.0252	0.0172	0.1376	0.1765	0.0182	0.4176	0.1193	0.0948	0.0180	0.2500	0.2627	0.2627	0.2627	0.1339	0.2627	0.1193	0.4124	0.2627	0.4176	0.2521	0.0783	0.0783
Javan 6	0.0698	0.0256	0.0171	1.0000	0.0182	0.0331	0.0575	0.0172	0.0431	0.0196	0.0357	0.8727	0.0690	0.0667	0.0172	0.8727	0.0645	0.0396	0.0431	0.0690	0.0092	0.0198	0.0485	0.0682	0.0397	0.0400	0.0400	0.0400	0.0189	0.0400	0.0198	0.0351	0.0400	0.0092	0.0400	0.0396	0.0396
Javan 7	0.0196	0.1000	0.3100	0.0182	1.0000	0.1709	0.0097	0.3265	0.1667	0.0943	0.3913	0.0088	0.0194	0.0189	0.3265	0.0088	0.0183	0.0636	0.1667	0.0194	0.2842	0.0952	0.1321	0.0192	0.1626	0.1736	0.1639	0.1736	0.0909	0.1639	0.0952	0.3978	0.1736	0.2979	0.1639	0.1038	0.1038
phigbs515	0.0354	0.0432	0.3091	0.0331	0.1709	1.0000	0.0263	0.3241	0.6222	0.0932	0.2051	0.0323	0.0351	0.0342	0.3241	0.0323	0.0333	0.1017	0.6044	0.0351	0.1739	0.0940	0.1368	0.0348	0.3684	0.3964	0.3717	0.3964	0.0992	0.3717	0.0940	0.2222	0.3964	0.1739	0.3717	0.0744	0.0744
phigbsCJBIII	0.6939	0.0090	0.0091	0.0575	0.0097	0.0263	1.0000	0.0092	0.0273	0.0213	0.0093	0.0556	0.8261	0.7400	0.0092	0.0556	0.8000	0.0211	0.0273	0.8667	0.0200	0.0215	0.0753	0.8085	0.0082	0.0083	0.0083	0.0083	0.0204	0.0083	0.0215	0.0092	0.0083	0.0200	0.0083	0.0000	0.0000
phigbsDMC15	0.0185	0.0455	0.9296	0.0172	0.3265	0.3241	0.0092	1.0000	0.1880	0.1193	0.4565	0.0168	0.0183	0.0179	1.0000	0.0168	0.0174	0.1389	0.1880	0.0183	0.4222	0.1204	0.0957	0.0182	0.2521	0.2759	0.2542	0.2759	0.1351	0.2542	0.1204	0.4316	0.2759	0.4222	0.2542	0.0885	0.0885
phigbsDMC16	0.0367	0.0444	0.1765	0.0431	0.1667	0.6222	0.0273	0.1880	1.0000	0.1062	0.3980	0.0420	0.0364	0.0354	0.1880	0.0420	0.0345	0.1053	0.9722	0.0364	0.1491	0.1071	0.1416	0.0360	0.2773	0.3017	0.2797	0.3017	0.1026	0.2797	0.1071	0.4184	0.3017	0.1491	0.2797	0.0678	0.0678
phigbsDMC17	0.0215	0.0333	0.1182	0.0196	0.0943	0.0932	0.0213	0.1193	0.1062	1.0000	0.1215	0.0190	0.0213	0.0206	0.1193	0.0190	0.0200	0.7869	0.1062	0.0213	0.1515	0.9815	0.1089	0.0211	0.0976	0.0984	0.0984	0.0984	0.8667	0.0984	0.9815	0.1193	0.0984	0.1515	0.0984	0.0283	0.0283
phigbsDMC2	0.0189	0.0382	0.4362	0.0357	0.3913	0.2051	0.0093	0.4565	0.3980	0.1215	1.0000	0.0259	0.0187	0.0182	0.4565	0.0259	0.0177	0.1204	0.3980	0.0187	0.4000	0.1226	0.1071	0.0185	0.2672	0.3036	0.2696	0.3036	0.1171	0.2696	0.1226	0.9143	0.3036	0.4157	0.2696	0.0614	0.0614
phigbsDMC20	0.0674	0.0165	0.0252	0.8727	0.0088	0.0323	0.0556	0.0168	0.0420	0.0190	0.0259	1.0000	0.0667	0.0645	0.0168	0.9630	0.0625	0.0385	0.0420	0.0667	0.0089	0.0192	0.0374	0.0659	0.0388	0.0391	0.0391	0.0391	0.0183	0.0391	0.0192	0.0254	0.0391	0.0089	0.0391	0.0286	0.0286
phigbsDMC21-1	0.7292	0.0180	0.0182	0.0690	0.0194	0.0351	0.8261	0.0183	0.0364	0.0213	0.0187	0.0667	1.0000	0.6415	0.0183	0.0667	0.8367	0.0211	0.0364	0.9535	0.0200	0.0215	0.0860	0.8085	0.0164	0.0165	0.0165	0.0165	0.0204	0.0165	0.0215	0.0183	0.0165	0.0200	0.0165	0.0103	0.0103
phigbsDMC24	0.6538	0.0175	0.0177	0.0667	0.0189	0.0342	0.7400	0.0179	0.0354	0.0206	0.0182	0.0645	0.6415	1.0000	0.0179	0.0645	0.6034	0.0204	0.0354	0.6731	0.0194	0.0208	0.0833	0.6604	0.0160	0.0161	0.0161	0.0161	0.0198	0.0161	0.0208	0.0179	0.0161	0.0194	0.0161	0.0100	0.0100
phigbsDMC25	0.0185	0.0455	0.9296	0.0172	0.3265	0.3241	0.0092	1.0000	0.1880	0.1193	0.4565	0.0168	0.0183	0.0179	1.0000	0.0168	0.0174	0.1389	0.1880	0.0183	0.4222	0.1204	0.0957	0.0182	0.2521	0.2759	0.2542	0.2759	0.1351	0.2542	0.1204	0.4316	0.2759	0.4222	0.2542	0.0885	0.0885
phigbsDMC27	0.0674	0.0165	0.0252	0.8727	0.0088	0.0323	0.0556	0.0168	0.0420	0.0190	0.0259	0.9630	0.0667	0.0645	0.0168	1.0000	0.0625	0.0385	0.0420	0.0667	0.0089	0.0192	0.0374	0.0659	0.0388	0.0391	0.0391	0.0391	0.0183	0.0391	0.0192	0.0254	0.0391	0.0089	0.0391	0.0286	0.0286
phigbsDMC28	0.6792	0.0171	0.0172	0.0645	0.0183	0.0333	0.8000	0.0174	0.0345	0.0200	0.0177	0.0625	0.8367	0.6034	0.0174	0.0625	1.0000	0.0198	0.0345	0.8750	0.0189	0.0202	0.0808	0.7500	0.0156	0.0157	0.0157	0.0157	0.0192	0.0157	0.0202	0.0174	0.0157	0.0189	0.0157	0.0097	0.0097
phigbsDMC30	0.0213	0.0163	0.1376	0.0396	0.0636	0.1017	0.0211	0.1389	0.1053	0.7869	0.1204	0.0385	0.0211	0.0204	0.1389	0.0385	0.0198	1.0000	0.1053	0.0211	0.1500	0.8000	0.0762	0.0208	0.0968	0.0976	0.0976	0.0976	0.8833	0.0976	0.8000	0.1182	0.0976	0.1500	0.0976	0.0377	0.0377
phigbsDMC33-1	0.0367	0.0444	0.1765	0.0431	0.1667	0.6044	0.0273	0.1880	0.9722	0.1062	0.3980	0.0420	0.0364	0.0354	0.1880	0.0420	0.0345	0.1053	1.0000	0.0364	0.1491	0.1071	0.1416	0.0360	0.2773	0.3017	0.2797	0.3017	0.1026	0.2797	0.1071	0.4184	0.3017	0.1491	0.2797	0.0678	0.0678
phigbsDMC33-2	0.7660	0.0180	0.0182	0.0690	0.0194	0.0351	0.8667	0.0183	0.0364	0.0213	0.0187	0.0667	0.9535	0.6731	0.0183	0.0667	0.8750	0.0211	0.0364	1.0000	0.0200	0.0215	0.0860	0.8478	0.0164	0.0165	0.0165	0.0165	0.0204	0.0165	0.0215	0.0183	0.0165	0.0200	0.0165	0.0103	0.0103
phigbsDMC34	0.0202	0.0156	0.4176	0.0092	0.2842	0.1739	0.0200	0.4222	0.1491	0.1515	0.4000	0.0089	0.0200	0.0194	0.4222	0.0089	0.0189	0.1500	0.1491	0.0200	1.0000	0.1531	0.0727	0.0198	0.1750	0.1765	0.1864	0.1765	0.1456	0.1864	0.1531	0.3913	0.1765	0.9355	0.1864	0.0455	0.0455
phigbsDMC36	0.0217	0.0336	0.1193	0.0198	0.0952	0.0940	0.0215	0.1204	0.1071	0.9815	0.1226	0.0192	0.0215	0.0208	0.1204	0.0192	0.0202	0.8000	0.1071	0.0215	0.1531	1.0000	0.1100	0.0213	0.0984	0.0992	0.0992	0.0992	0.8500	0.0992	1.0000	0.1204	0.0992	0.1531	0.0992	0.0286	0.0286
phigbsDMC43-1	0.0870	0.0667	0.0948	0.0485	0.1321	0.1368	0.0753	0.0957	0.1416	0.1089	0.1071	0.0374	0.0860	0.0833	0.0957	0.0374	0.0808	0.0762	0.1416	0.0860	0.0727	0.1100	1.0000	0.0851	0.1301	0.1311	0.1311	0.1311	0.1048	0.1311	0.1100	0.1053	0.1311	0.0727	0.1311	0.0660	0.0660
phigbsDMC43-2	0.7872	0.0179	0.0180	0.0682	0.0192	0.0348	0.8085	0.0182	0.0360	0.0211	0.0185	0.0659	0.8085	0.6604	0.0182	0.0659	0.7500	0.0208	0.0360	0.8478	0.0198	0.0213	0.0851	1.0000	0.0163	0.0164	0.0164	0.0164	0.0202	0.0164	0.0213	0.0182	0.0164	0.0198	0.0164	0.0102	0.0102
phigbsDMC47	0.0165	0.0559	0.2500	0.0397	0.1626	0.3684	0.0082	0.2521	0.2773	0.0976	0.2672	0.0388	0.0164	0.0160	0.2521	0.0388	0.0156	0.0968	0.2773	0.0164	0.1750	0.0984	0.1301	0.0163	1.0000	0.8721	0.8721	0.8721	0.0945	0.8721	0.0984	0.2627	0.8721	0.1750	0.9398	0.0794	0.0794
phigbsDMC48	0.0167	0.0638	0.2627	0.0400	0.1736	0.3964	0.0083	0.2759	0.3017	0.0984	0.3036	0.0391	0.0165	0.0161	0.2759	0.0391	0.0157	0.0976	0.3017	0.0165	0.1765	0.0992	0.1311	0.0164	0.8721	1.0000	0.8824	1.0000	0.0952	0.8824	0.0992	0.3097	1.0000	0.1765	0.8824	0.0887	0.0887
phigbsDMC49	0.0167	0.0563	0.2627	0.0400	0.1639	0.3717	0.0083	0.2542	0.2797	0.0984	0.2696	0.0391	0.0165	0.0161	0.2542	0.0391	0.0157	0.0976	0.2797	0.0165	0.1864	0.0992	0.1311	0.0164	0.8721	0.8824	1.0000	0.8824	0.0952	1.0000	0.0992	0.2650	0.8824	0.1765	0.9277	0.0800	0.0800
phigbsDMC5	0.0167	0.0638	0.2627	0.0400	0.1736	0.3964	0.0083	0.2759	0.3017	0.0984	0.3036	0.0391	0.0165	0.0161	0.2759	0.0391	0.0157	0.0976	0.3017	0.0165	0.1765	0.0992	0.1311	0.0164	0.8721	1.0000	0.8824	1.0000	0.0952	0.8824	0.0992	0.3097	1.0000	0.1765	0.8824	0.0887	0.0887
phigbsDMC6	0.0206	0.0323	0.1339	0.0189	0.0909	0.0992	0.0204	0.1351	0.1026	0.8667	0.1171	0.0183	0.0204	0.0198	0.1351	0.0183	0.0192	0.8833	0.1026	0.0204	0.1456	0.8500	0.1048	0.0202	0.0945	0.0952	0.0952	0.0952	1.0000	0.0952	0.8500	0.1150	0.0952	0.1456	0.0952	0.0367	0.0367
phigbsDMC61	0.0167	0.0563	0.2627	0.0400	0.1639	0.3717	0.0083	0.2542	0.2797	0.0984	0.2696	0.0391	0.0165	0.0161	0.2542	0.0391	0.0157	0.0976	0.2797	0.0165	0.1864	0.0992	0.1311	0.0164	0.8721	0.8824	1.0000	0.8824	0.0952	1.0000	0.0992	0.2650	0.8824	0.1765	0.9277	0.0800	0.0800
phigbsDMC62	0.0217	0.0336	0.1193	0.0198	0.0952	0.0940	0.0215	0.1204	0.1071	0.9815	0.1226	0.0192	0.0215	0.0208	0.1204	0.0192	0.0202	0.8000	0.1071	0.0215	0.1531	1.0000	0.1100	0.0213	0.0984	0.0992	0.0992	0.0992	0.8500	0.0992	1.0000	0.1204	0.0992	0.1531	0.0992	0.0286	0.0286
phigbsDMC64	0.0185	0.0455	0.4124	0.0351	0.3978	0.2222	0.0092	0.4316	0.4184	0.1193	0.9143	0.0254	0.0183	0.0179	0.4316	0.0254	0.0174	0.1182	0.4184	0.0183	0.3913	0.1204	0.1053	0.0182	0.2627	0.3097	0.2650	0.3097	0.1150	0.2650	0.1204	1.0000	0.3097	0.4066	0.2650	0.0696	0.0696
phigbsDMC66-1	0.0167	0.0638	0.2627	0.0400	0.1736	0.3964	0.0083	0.2759	0.3017	0.0984	0.3036	0.0391	0.0165	0.0161	0.2759	0.0391	0.0157	0.0976	0.3017	0.0165	0.1765	0.0992	0.1311	0.0164	0.8721	1.0000	0.8824	1.0000	0.0952	0.8824	0.0992	0.3097	1.0000	0.1765	0.8824	0.0887	0.0887
phigbsDMC67	0.0202	0.0156	0.4176	0.0092	0.2979	0.1739	0.0200	0.4222	0.1491	0.1515	0.4157	0.0089	0.0200	0.0194	0.4222	0.0089	0.0189	0.1500	0.1491	0.0200	0.9355	0.1531	0.0727	0.0198	0.1750	0.1765	0.1765	0.1765	0.1456	0.1765	0.1531	0.4066	0.1765	1.0000	0.1765	0.0455	0.0455
phigbsDMC68	0.0167	0.0563	0.2521	0.0400	0.1639	0.3717	0.0083	0.2542	0.2797	0.0984	0.2696	0.0391	0.0165	0.0161	0.2542	0.0391	0.0157	0.0976	0.2797	0.0165	0.1864	0.0992	0.1311	0.0164	0.9398	0.8824	0.9277	0.8824	0.0952	0.9277	0.0992	0.2650	0.8824	0.1765	1.0000	0.0800	0.0800
phigbsDMC69	0.0104	0.6892	0.0783	0.0396	0.1038	0.0744	0.0000	0.0885	0.0678	0.0283	0.0614	0.0286	0.0103	0.0100	0.0885	0.0286	0.0097	0.0377	0.0678	0.0103	0.0455	0.0286	0.0660	0.0102	0.0794	0.0887	0.0800	0.0887	0.0367	0.0800	0.0286	0.0696	0.0887	0.0455	0.0800	1.0000	1.0000
phigbsDMC9	0.0104	0.6892	0.0783	0.0396	0.1038	0.0744	0.0000	0.0885	0.0678	0.0283	0.0614	0.0286	0.0103	0.0100	0.0885	0.0286	0.0097	0.0377	0.0678	0.0103	0.0455	0.0286	0.0660	0.0102	0.0794	0.0887	0.0800	0.0887	0.0367	0.0800	0.0286	0.0696	0.0887	0.0455	0.0800	1.0000	1.0000


Javan 5
prophiDMC15
prophiDMC25
Javan 7
prophiDMC64
prophiDMC2
prophiDMC34
prophiDMC67
prophi515
prophiDMC16
prophiDMC33-1

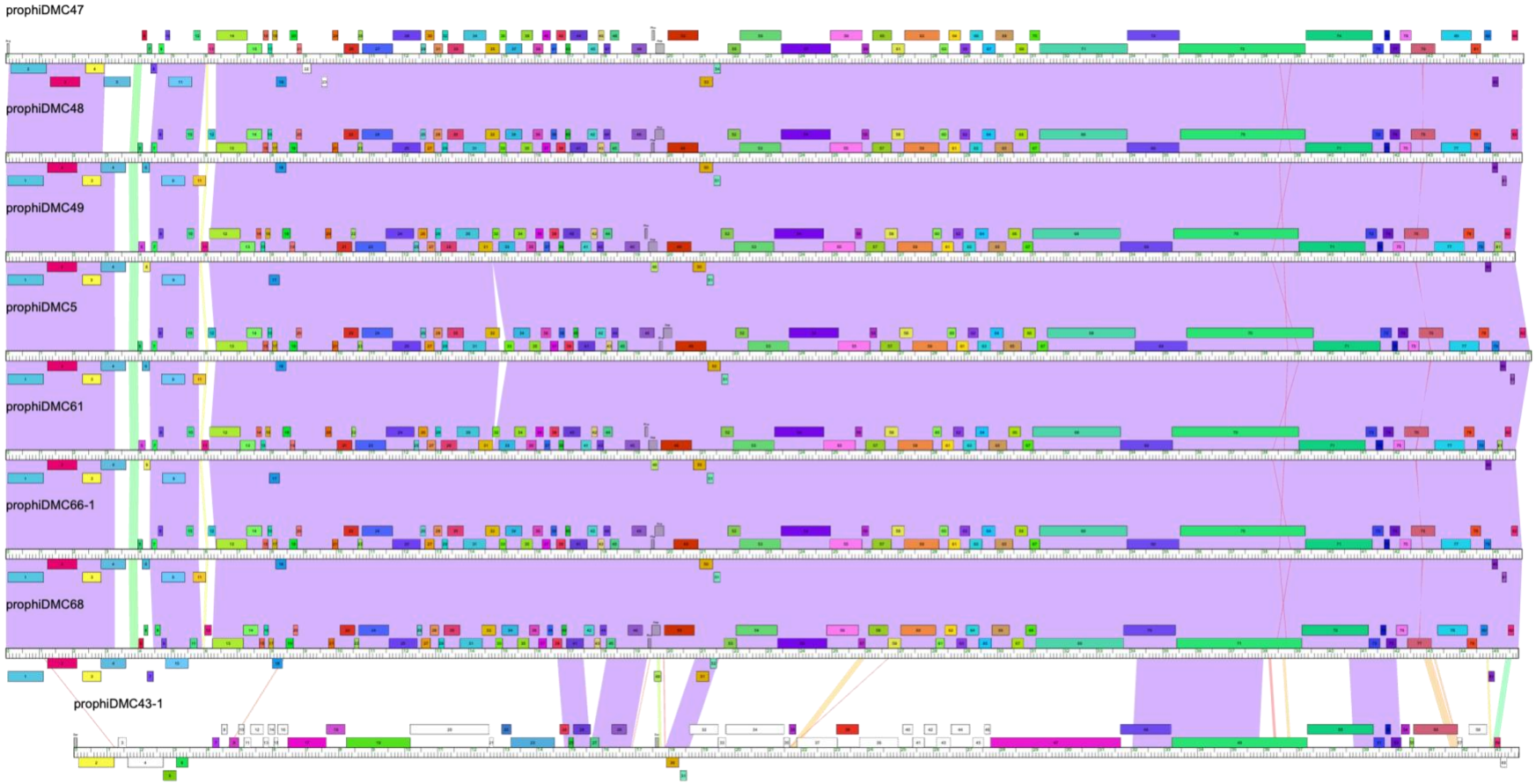


Figure A.1: Genome comparisons of Cluster A prophages. Genome similarity in nucleotide sequence is illustrated using a color spectrum. The colors range from violet to red, where violet represents the highest similarity and red the lowest. Genes are represented as colored boxes, positioned above or below each genome, which indicate their direction of transcription. The color of the boxes corresponds to their gene phamilies. The maps were generated using phamerator software and the *Streptococcus agalactiae* (version 1) database.

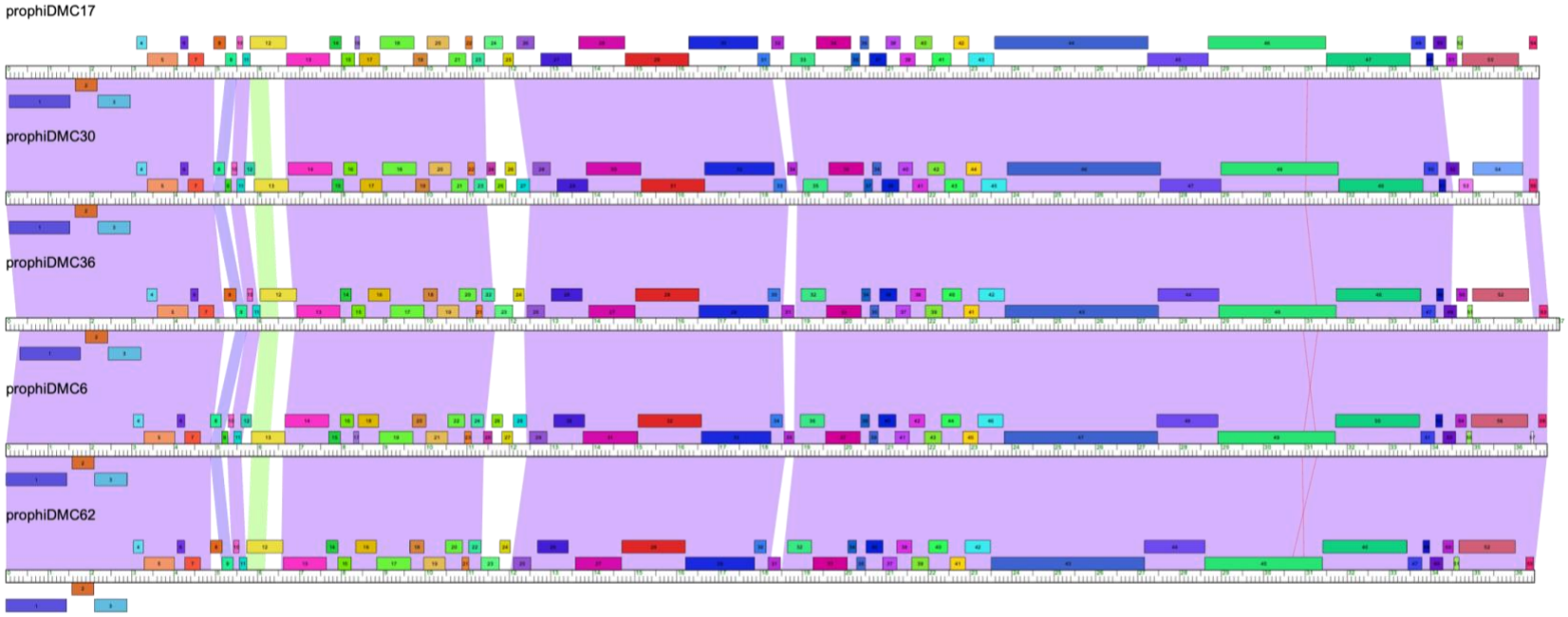


Figure A.2: Genome comparisons of Cluster B prophages. Genome similarity in nucleotide sequence is illustrated using a color spectrum. The colors range from violet to red, where violet represents the highest similarity and red the lowest. Genes are represented as colored boxes, positioned above or below each genome, which indicate their direction of transcription. The color of the boxes corresponds to their gene phamilies. The maps were generated using phamerator software and the *Streptococcus agalactiae* (version 1) database.

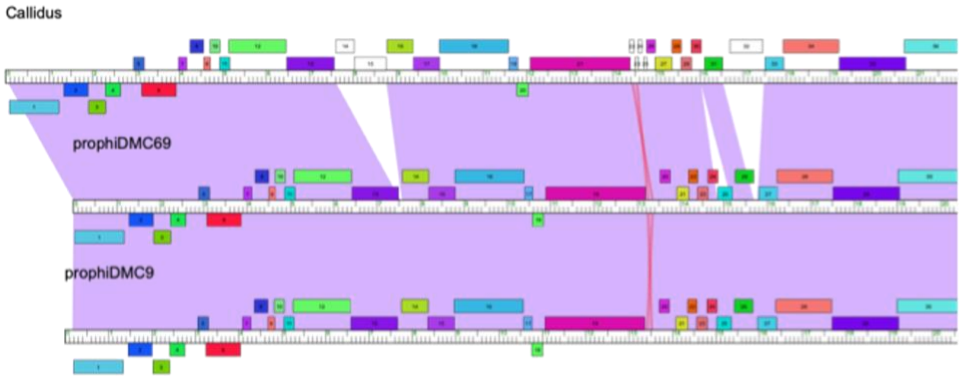


Figure A.3: Genome comparisons of Cluster C prophages. Genome similarity in nucleotide sequence is illustrated using a color spectrum. The colors range from violet to red, where violet represents the highest similarity and red the lowest. Genes are represented as colored boxes, positioned above or below each genome, which indicate their direction of transcription. The color of the boxes corresponds to their gene phamilies. The maps were generated using phamerator software and the *Streptococcus agalactiae* (version 1) database.

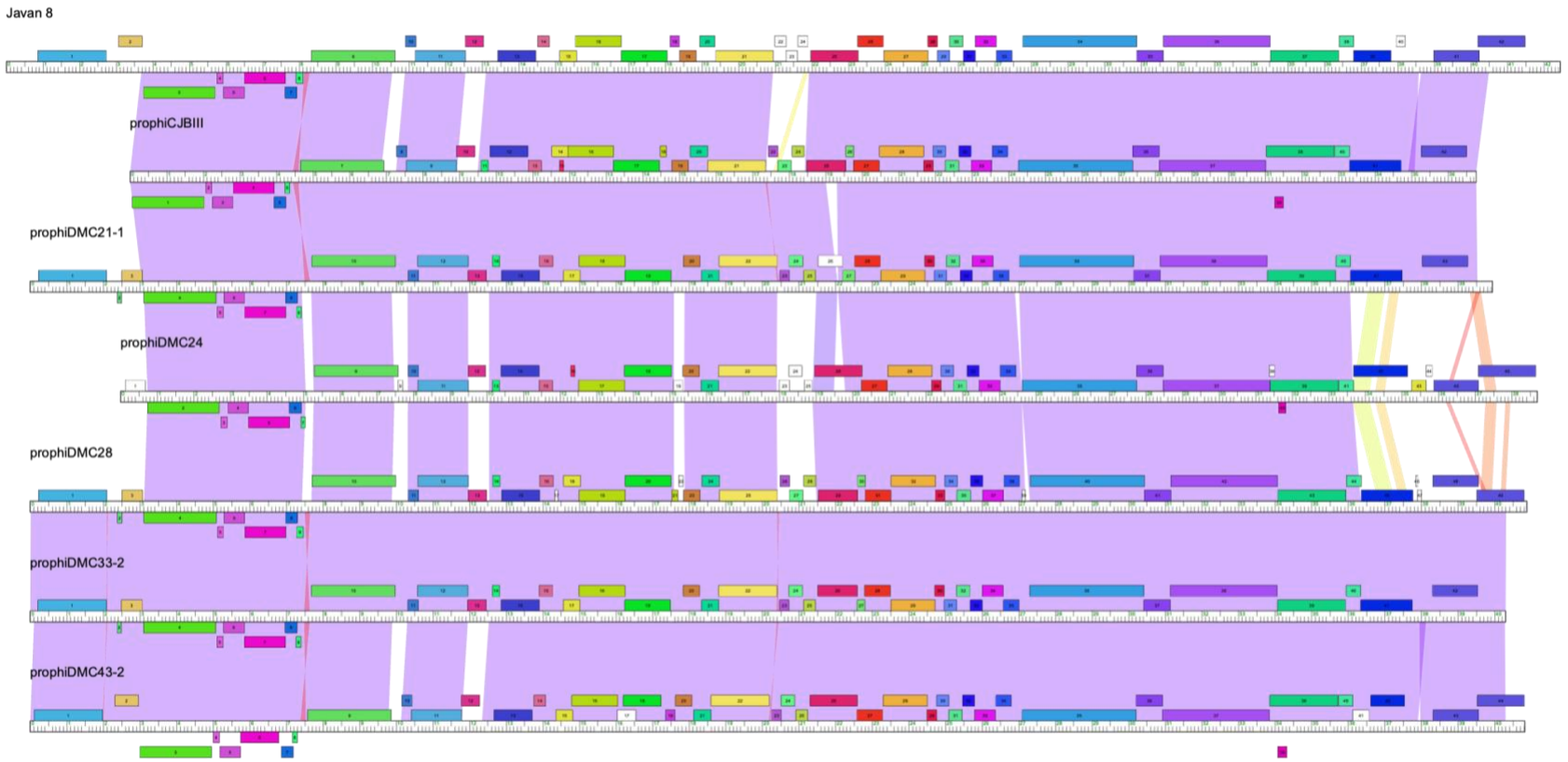


Figure A.4: Genome comparisons of Cluster D prophages. Genome similarity in nucleotide sequence is illustrated using a color spectrum. The colors range from violet to red, where violet represents the highest similarity and red the lowest. Genes are represented as colored boxes, positioned above or below each genome, which indicate their direction of transcription. The color of the boxes corresponds to their gene phamilies. The maps were generated using phamerator software and the *Streptococcus agalactiae* (version 1) database.

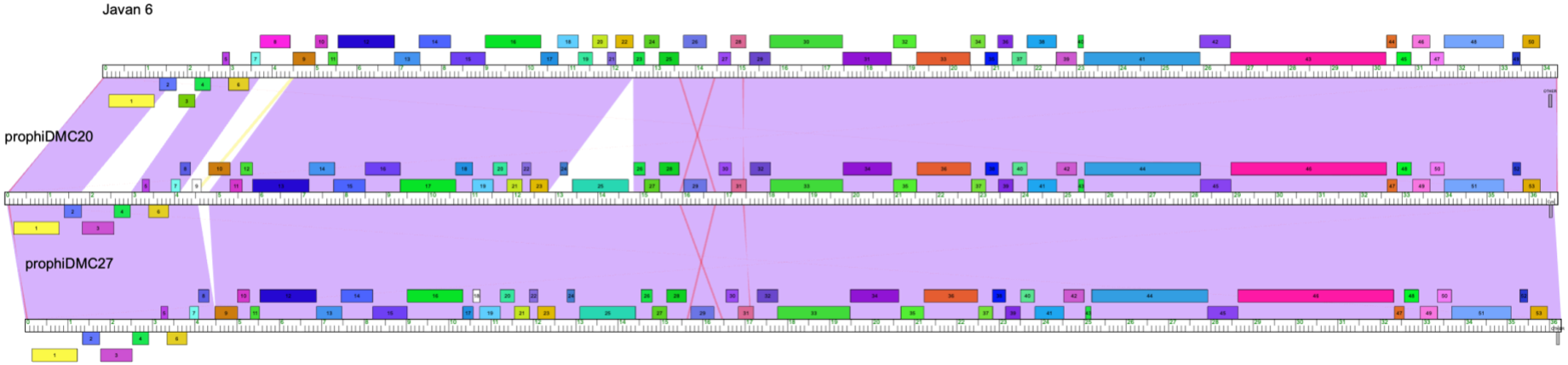


Figure A.5: Genome comparisons of Cluster E prophages. Genome similarity in nucleotide sequence is illustrated using a color spectrum. The colors range from violet to red, where violet represents the highest similarity and red the lowest. Genes are represented as colored boxes, positioned above or below each genome, which indicate their direction of transcription. The color of the boxes corresponds to their gene phamilies. The maps were generated using phamerator software and the *Streptococcus agalactiae* (version 1) database.

APPENDIX B

Table B.1: Significantly differentially expressed GBS genes in the absence of the prophage. Differentially expressed genes filtered out by absolute log2fold change > 1 and adjusted p-value of 0.05. Early log shaded orange and late log shaded green.

	locus_tag	CDS name	early log log2foldchange	early log adjusted p-value	late log log2foldchange	late log adjusted p-value
1	W903_RS00325	peptidoglycan DD-metalloendopeptidase family protein CDS	-0.367264416	1	2.596990895	5.58E-16
2	W903_RS00875	LacI family transcriptional regulator CDS	1.130175252	1	2.66019471	0.04563294
3	W903_RS00900	argininosuccinate synthase CDS	-0.303358828	1	1.933990458	3.89E-09
4	W903_RS00905	argH CDS	-0.269227749	1	1.704381012	7.89E-08
5	W903_RS00980	sufC CDS	-0.122641515	1	-1.970568691	9.05E-28
6	W903_RS00985	sufD CDS	-0.420295099	1	-1.9021155	3.16E-05
7	W903_RS00990	cysteine desulfurase CDS	-0.204566396	1	-1.563997088	0.01525854
8	W903_RS00995	SUF system NifU family Fe-S cluster assembly protein CDS	-0.145218022	1	-1.452747786	0.00165048
9	W903_RS01015	peptide ABC transporter substrate-binding protein CDS	-0.210064379	1	1.267458559	1.07E-06
10	W903_RS01020	ABC transporter permease CDS	-0.147502844	1	1.437980635	1.77E-11
11	W903_RS01035	ATP-binding cassette domain-containing protein CDS	-0.330232374	1	1.510357537	0.0046238
12	W903_RS01050	23S rRNA	-1.451472059	1	-2.877946386	0.01740962
13	W903_RS01110	DUF1033 family protein CDS	-0.160281525	1	-2.310538638	5.55E-58
14	W903_RS01260	treP CDS	-1.105660153	1	-1.565270897	0.00879872
15	W903_RS01295	ferric reductase-like transmembrane domain-containing protein CDS	0.055313144	1	2.281970131	9.05E-28
16	W903_RS01515	GNAT family N-acetyltransferase CDS	0.239804314	1	-2.073272775	1.90E-14
17	W903_RS01520	GNAT family N-acetyltransferase CDS	0.10770164	1	-1.689216409	0.00174911
18	W903_RS01655	PTS transporter subunit EIIC CDS	-0.087724663	1	-3.344741869	5.83E-97

19	W903_RS01730	pepC CDS	-0.451603336	1	1.672089281	0.00080539
20	W903_RS02330	alpha/beta hydrolase CDS	-1.20098187	1	-1.973796332	0.0196337
21	W903_RS02850	biotin transporter BioY CDS	-0.62121205	1	2.640991382	0.00014379
22	W903_RS02855	bioB CDS	-0.401112766	1	2.720690061	6.30E-16
23	W903_RS02860	hypothetical protein CDS	0.026949089	1	2.702102962	2.16E-14
24	W903_RS02865	thiolase family protein CDS	0.102230176	1	2.993144908	1.06E-30
25	W903_RS02870	AMP-binding protein CDS	0.304901461	1	2.707150195	1.80E-33
26	W903_RS03380	fructose-bisphosphatase class III CDS	0.253833976	1	-1.583605444	1.09E-05
27	W903_RS04015	U32 family peptidase CDS	0.733154867	1	2.101412611	7.88E-07
28	W903_RS04200	ATP-binding cassette domain-containing protein CDS	0.164336906	1	-1.844654644	5.46E-06
29	W903_RS04205	ABC transporter permease CDS	-0.219327613	1	-1.834843953	5.99E-13
30	W903_RS04210	methionine ABC transporter substrate-binding protein CDS	-0.055275048	1	-1.853776482	4.84E-09
31	W903_RS04215	DEAD/DEAH box helicase CDS	-0.133330279	1	2.051397171	9.09E-23
32	W903_RS04255	alpha/beta hydrolase CDS	-1.159031152	1	-2.149691154	0.00577172
33	W903_RS04270	superoxide dismutase SodA CDS	-0.150540459	1	-1.222004583	0.02621227
34	W903_RS04295	glycerate kinase CDS	0.580772514	1	-1.894029353	4.05E-05
35	W903_RS04300	GntP family permease CDS	0.651910919	1	-1.939258009	7.10E-12
36	W903_RS04320	VIT family protein CDS	-0.025132266	1	1.597328397	4.16E-10
37	W903_RS04720	thiamine pyrophosphate-dependent dehydrogenase E1 component subunit alpha CDS	-0.636123426	1	-2.076502015	0.00019866
38	W903_RS04725	alpha-ketoacid dehydrogenase subunit beta CDS	-0.515148356	1	-1.833137114	0.0001049
39	W903_RS04730	dihydrolipoamide acetyltransferase CDS	-0.366050901	1	-1.614684919	7.91E-05
40	W903_RS04910	glmS CDS	0.696648863	1	-1.501214957	6.14E-11
41	W903_RS05105	peptide ABC	-0.174176966	1	3.754263349	1.54E-99

		transporter substrate-binding protein CDS				
42	W903_RS05510	aspartate carbamoyltransferase catalytic subunit CDS	-0.357000498	1	-2.043243896	0.00486126
43	W903_RS05525	pyrF CDS	-0.117668931	1	-1.597570756	0.01016067
44	W903_RS05590	NADH-dependent flavin oxidoreductase CDS	-0.159964888	1	-1.340280853	0.014623
45	W903_RS05715	hypothetical protein CDS	-0.050017244	1	-1.898367603	9.62E-09
46	W903_RS05770	redox-sensing transcriptional repressor Rex CDS	0.394732893	1	-1.958869922	4.12E-55
47	W903_RS05935	Asp23/Gls24 family envelope stress response protein CDS	-1.61157028	0.11420893	-1.640021749	0.00983341
48	W903_RS05945	Asp23/Gls24 family envelope stress response protein CDS	-1.169575583	1	-1.734091895	0.02726271
49	W903_RS05950	DUF2273 domain-containing protein CDS	-1.199979537	1	-1.701560255	0.01253659
50	W903_RS05960	GlsB/YeaQ/YmgE family stress response membrane protein CDS	-1.452513776	0.01794173	-1.704670694	2.87E-08
51	W903_RS06495	DUF1836 domain-containing protein CDS	-1.516638978	1	2.863480796	3.22E-05
52	W903_RS06500	hemolysin III family protein CDS	-2.0105898	3.61E-09	2.672431557	5.31E-16
53	W903_RS06510	glutathione S-transferase family protein CDS	0.299125746	1	-1.42661697	0.00023462
54	W903_RS07070	transporter substrate-binding domain-containing protein CDS	0.049423965	1	1.742973031	1.08E-12
55	W903_RS07085	rhodanese-related sulfurtransferase CDS	-0.383089657	1	-1.716626949	3.72E-07
56	W903_RS07090	hypothetical protein CDS	0.172396628	1	-2.129172715	8.82E-06
57	W903_RS07220	helix-turn-helix domain-containing protein CDS	-0.531318648	1	-2.419073629	9.29E-40

58	W903_RS07230	CPBP family intramembrane metalloprotease CDS	-0.108896024	1	1.240926627	0.00772808
59	W903_RS07275	LPXTG cell wall anchor domain-containing protein CDS	-0.087167967	1	-2.554933484	1.41E-118
60	W903_RS07285	aldo/keto reductase CDS	-0.24925818	1	-1.621812462	2.90E-05
61	W903_RS07295	gloA CDS	-0.152604474	1	-1.740136956	7.68E-14
62	W903_RS07710	CBS domain-containing protein CDS	-0.128467827	1	3.073317907	8.27E-08
63	W903_RS07715	ABC transporter ATP-binding protein CDS	0.054864995	1	2.569949599	0.00061141
64	W903_RS07720	ABC transporter ATP-binding protein CDS	0.410772166	1	3.471722102	1.55E-05
65	W903_RS07725	branched-chain amino acid ABC transporter permease CDS	0.128583973	1	3.726309514	4.90E-05
66	W903_RS07730	branched-chain amino acid ABC transporter permease CDS	0.061275049	1	3.749232106	4.12E-06
67	W903_RS07735	ABC transporter substrate-binding protein CDS	0.271015078	1	2.562702165	1.78E-09
68	W903_RS07865	amino acid ABC transporter permease CDS	0.003683232	1	-1.352227547	0.00825695
69	W903_RS07990	brnQ CDS	-0.878504502	1	2.593150814	6.15E-13
70	W903_RS08395	pflB CDS	-0.017412839	1	-1.366154251	0.01176628
71	W903_RS08475	NAD(P)/FAD-dependent oxidoreductase CDS	-0.110234753	1	-1.26646802	0.02726271
72	W903_RS08730	ABC transporter permease/substrate binding protein CDS	-1.785192107	3.09E-06	-1.736129213	1.55E-05
73	W903_RS08735	glycine betaine/L-proline ABC transporter ATP-binding protein CDS	-1.844819225	7.25E-07	-1.911279745	9.64E-08
74	W903_RS08910	ahpC CDS	-0.386547457	1	-1.671949576	6.17E-08
75	W903_RS08915	ahpF CDS	-0.358976757	1	-1.850896037	2.38E-10
76	W903_RS08950	MFS transporter CDS	-0.098768246	1	1.244547394	0.0001959
77	W903_RS09130	lacA CDS	0.721519754	1	3.034307276	0.02317844

78	W903_RS09555	S8 family serine peptidase CDS	-0.042481744	1	3.04256236	2.23E-08
79	W903_RS09600	YSIRK signal domain/LPXTG anchor domain surface protein CDS	-0.956414475	1	-1.931761862	0.00113046
80	W903_RS09665	ATP-binding cassette domain-containing protein CDS	-0.071706248	1	-1.82283986	6.35E-23
81	W903_RS09685	VOC family protein CDS	-0.751772771	1	-3.07354568	0.00849539
82	W903_RS05960	GlsB/YeaQ/YmgE family stress response membrane protein CDS	-1.452513776	0.01794173	-1.704670694	2.87E-08
83	W903_RS06500	hemolysin III family protein CDS	-2.0105898	3.61E-09	2.672431557	5.31E-16
84	W903_RS08730	ABC transporter permease/substrate binding protein CDS	-1.785192107	3.09E-06	-1.736129213	1.55E-05
85	W903_RS08735	glycine betaine/L-proline ABC transporter ATP-binding protein CDS	-1.844819225	7.25E-07	-1.911279745	9.64E-08
86	W903_RS10095	arcA CDS	-1.944944531	0.03068612	-0.972762875	1
87	W903_RS10105	argF CDS	-1.833356775	4.20E-05	-1.271598188	1

APPENDIX C

Induction of phages from GBS strains

Phage induction is when an integrated prophage excises from the bacterial genome under stressful conditions. Prophage induction occurs either spontaneously, usually at low frequency, or by a wide range of external stressors (145, 296) such as ultraviolet light and antibiotics (e.g., mitomycin C). Mitomycin-induced phages from GBS characterized using a multilocus typing system showed a high level of genetic diversity and molecular groups of the phages identified demonstrated their ability to infect or integrate into bacterial cells intraspecies (120). To better understand the contribution of our prophages to GBS lifestyle, we attempted to isolate phages from two GBS strains used in this study (characterized in Chapter 2). GBS strains, A909 and DMC9, were treated with either mitomycin C (mitC) or mitomycin and chloroform (mitC + $ChCl_3$) for 3 hours to induce their respective phages. Supernatants were added to two GBS strains, NEM316 and DMC64 grown to mid-log, and incubated overnight as an enrichment step to allow the phage to propagate. Spot tests of filtrates from A909 and DMC9 were then spotted on a range of DMC strains (as indicator strains) and observed after 24 hours for clearings. While DMC9 showed the most promise, with bacterial lysis observed in more than 50% of indicator strains, we were unable to capture the phage with subsequent experiments. However, we were successful in capturing one of the two phages present in A909 (Javan 7). Spontaneously induced phages were also captured from DMC27, DMC28, DMC39, and DMC49. Further work is needed to characterize these phages, determine host range, and investigate their impact on bacterial fitness when introduced into other GBS strains.

Table C. 1: Filtrates used in phage induction experiments.

Filtrate ID	GBS strain to be induced	treatment	recipient GBS strain for enrichment step
A1	A909	none	NEM316
A2		mitC	
A3		mitC + $ChCl_3$	
A4		none	DMC 64
A5		mitC	
A6		mitC + $ChCl_3$	
D1	DMC 9	none	NEM316
D2		mitC	
D3		mitC + $ChCl_3$	
D4		none	DMC 64
D5		mitC	
D6		mitC + $ChCl_3$	

Table C. 2:

	Phage lysates												Comments
Strains	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	
DMC 12	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 13	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 14	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 15	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 16	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 17	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 18	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 19	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 21	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 22	-	+	+	+	-	-	-	-	-	+	-	-	
DMC 23	-	-	-	+	-	-	-	-	-	+	-	-	
DMC 24	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 25	-	-	-	-	-	-	-	-	-	+	-	-	
DMC 26	-	-	-	-	-	-	-	-	-	+	-	-	
DMC 27	-	-	-	-	-	-	-	-	-	-	-	-	**spontaneous induction**
DMC 28	-	-	-	-	-	-	-	-	-	-	-	-	**spontaneous induction**
DMC 29	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 30	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 31	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 32	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 33	-	-	-	-	-	-	-	-	-	-	-	-	

DMC 34	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 35	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 36	-	-	-	-	-	-	-	-	-	+	-	-	
DMC 37	-	-	-	-	-	-	-	-	-	-	-	-	
DMC 38	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 39	-	-	-	+	-	-	-	-	-	+	+	+	**spontaneous induction**
DMC 40	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 41	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 42	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 43	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 44	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 45	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 46	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 47	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 48	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 49	-	-	-	+	-	-	-	-	-	+	+	+	**spontaneous induction**
DMC 50	-	-	-	+	-	-	-	-	-	+	+	+	
DMC 51	-	-	-	+	-	-	-	-	-	+	+	+	

Appendix D
Prophage association with T7SS

The data presented here was analyzed as part of the work submitted for publication; Spencer *et al.,* 2023, Heterogeneity of the group B streptococcal type VII secretion system and influence on colonization of the female genital tract. bioRxiv 2023.01.25.525443; doi: https://doi.org/10.1101/2023.01.25.525443 Author contribution was only in data analysis. The abstract of the paper is presented here:

Abstract

Type VIIb secretion systems (T7SSb) in Gram-positive bacteria facilitate physiology, interbacterial competition, and/or virulence via EssC ATPase-driven secretion of small α-helical proteins and toxins. Recently, we characterized T7SSb in group B Streptococcus (GBS), a leading cause of infection in newborns and immunocompromised adults. GBS T7SS comprises four subtypes based on variation in the C-terminus of EssC and the repertoire of downstream effectors; however, the intra-species diversity of GBS T7SS and impact on GBS-host interactions remains unknown. Bioinformatic analysis indicates that GBS T7SS loci encode subtype-specific putative effectors, which have low inter-species and inter-subtype homology but contain similar domains/motifs and therefore may serve similar functions. We further identify orphaned GBS WXG100 proteins. Functionally, we show that GBS T7SS subtype I and III strains secrete EsxA in vitro and that in subtype I strain CJB111, esxA1 appears to be differentially transcribed from the T7SS operon. Further, we observe subtype-specific effects of GBS T7SS on host colonization, as subtype I but not subtype III T7SS promotes GBS vaginal persistence. Finally, we observe that T7SS subtypes I and II are the predominant subtypes in

clinical GBS isolates. This study highlights the potential impact of T7SS heterogeneity on host-GBS interactions.

Table D.1: Prophage distribution in GBS sequences from Genbank. Prophage regions were searched in additional sequences from Genbank for analysis of prophage association with T7SS.

Sequence name	Prophage cluster	paratox on chromosome	phage paratox	holtox
NC_004368	none	TRUE	FALSE	FALSE
NC_004116	A & E	TRUE	TRUE	FALSE
NC_007432	A & D	TRUE	TRUE	TRUE
NZ_CP063198	D	TRUE	TRUE	TRUE
NZ_CP006910	C	TRUE	FALSE	FALSE
NZ_HG939456	none	TRUE	FALSE	FALSE
NZ_CP051004	A	TRUE	TRUE	FALSE
NZ_CP007570	none	TRUE	FALSE	FALSE
NZ_CP007571	A	TRUE	TRUE	FALSE
NZ_CP007572	A	TRUE	TRUE	FALSE
NZ_CP007631	A & D	TRUE	TRUE	TRUE
NZ_CP007632	A	TRUE	TRUE	TRUE
NZ_CP008813	none	FALSE	FALSE	FALSE
NZ_CP010319	none	FALSE	FALSE	FALSE
NZ_CP010867	D	TRUE	FALSE	TRUE
NZ_CP010874	none	TRUE	FALSE	FALSE
NZ_CP010875	none	TRUE	FALSE	FALSE
NZ_LT714196	not clustered	TRUE	FALSE	TRUE
NZ_CP011325	D	TRUE	FALSE	TRUE
NZ_CP011326	D	TRUE	FALSE	TRUE
NZ_CP011327	D	TRUE	FALSE	TRUE
NZ_CP012419	none	TRUE	FALSE	FALSE
NZ_CP012480	none	TRUE	FALSE	FALSE
NZ_CP012503	C & D	TRUE	FALSE	TRUE
NZ_CP013202	D	TRUE	FALSE	TRUE
NZ_CP013908	D	TRUE	FALSE	FALSE

NZ_CP016391	none	TRUE	FALSE	FALSE
NZ_CP019978	D	TRUE	FALSE	FALSE
NZ_CP020387	A	TRUE	TRUE	FALSE
NC_021486	none	FALSE	FALSE	FALSE
NZ_AP018935	D	TRUE	FALSE	TRUE
NZ_CP030845	none	TRUE	FALSE	FALSE
NZ_CP044091	A	TRUE	TRUE	TRUE
NZ_CP049938	D	TRUE	FALSE	TRUE
NZ_CP053027	D	TRUE	FALSE	TRUE
NZ_LR134512	D	TRUE	FALSE	TRUE
NZ_LR134520	D	TRUE	FALSE	TRUE
NZ_LS483387	none	TRUE	FALSE	FALSE
NZ_LT545678	A & C	TRUE	TRUE	FALSE
CP051841	none	FALSE	FALSE	FALSE
CP051843	none	FALSE	FALSE	FALSE
CP051844	none	TRUE	FALSE	FALSE
CP051845	none	TRUE	FALSE	FALSE
CP051846	none	TRUE	FALSE	FALSE
CP051847	none	FALSE	FALSE	FALSE
NC_018646	D	FALSE	FALSE	TRUE
NC_021485	D	TRUE	FALSE	TRUE
NZ_CP020449	A	TRUE	TRUE	TRUE
NZ_CP021769	none	TRUE	FALSE	FALSE
NZ_CP021770	not clustered	TRUE	FALSE	TRUE
NZ_CP021771	none	TRUE	FALSE	FALSE
NZ_CP021772	D	TRUE	FALSE	FALSE
NZ_CP021773	none	TRUE	FALSE	FALSE
NZ_CP021862	E	TRUE	FALSE	FALSE
NZ_CP021863	none	TRUE	FALSE	FALSE
NZ_CP021864	none	TRUE	FALSE	FALSE
NZ_CP021865	none	TRUE	FALSE	FALSE
NZ_CP021866	none	TRUE	FALSE	FALSE

NZ_CP021868	D	TRUE	FALSE	FALSE
NZ_CP021869	C	TRUE	TRUE	TRUE
NZ_CP021870	none	TRUE	FALSE	FALSE
NZ_CP022537	none	TRUE	FALSE	FALSE
NZ_CP025026	none	TRUE	FALSE	FALSE
NZ_CP025027	none	TRUE	FALSE	FALSE
NZ_CP025028	none	TRUE	FALSE	FALSE
NZ_CP025029	none	TRUE	FALSE	FALSE
NZ_CP026082	none	TRUE	FALSE	FALSE
NZ_CP026084	none	TRUE	FALSE	FALSE
NZ_CP029561	none	TRUE	FALSE	FALSE
NZ_CP029749	D	TRUE	FALSE	TRUE
NZ_CP033822	A	TRUE	TRUE	TRUE
NZ_CP034315	D	TRUE	FALSE	TRUE
NZ_CP036376	E	TRUE	FALSE	FALSE
NZ_CP041998	D	TRUE	FALSE	TRUE
NZ_CP041999	D	TRUE	FALSE	TRUE
NZ_CP042000	D	TRUE	FALSE	TRUE
NZ_CP042001	A & D	TRUE	TRUE	FALSE
NZ_CP042002	D	TRUE	FALSE	TRUE
NZ_CP044090	A & E	TRUE	TRUE	TRUE
NZ_LS483342	none	TRUE	TRUE	FALSE
FO393392	none	FALSE	FALSE	FALSE
CP007482	none	FALSE	FALSE	FALSE
HF952106	none	FALSE	FALSE	FALSE
CP007565	none	FALSE	FALSE	FALSE
CP051842	none	TRUE	FALSE	FALSE
CP051848	none	FALSE	FALSE	FALSE
NZ_CP059383	D	TRUE	FALSE	TRUE
CP011328	none	FALSE	FALSE	FALSE
NZ_CP011329	A & D	TRUE	TRUE	TRUE
NZ_LR134265	C	TRUE	TRUE	TRUE

CP029632	none	FALSE	FALSE	FALSE
CP018623	none	FALSE	FALSE	FALSE
CP015976	none	FALSE	FALSE	FALSE
NZ_CP058666	D	TRUE	FALSE	TRUE
CP019804	none	FALSE	FALSE	FALSE
CP019805	none	FALSE	FALSE	FALSE
CP019806	none	FALSE	FALSE	FALSE
CP019807	none	FALSE	FALSE	FALSE
CP003919	none	FALSE	FALSE	FALSE
CP019800	none	FALSE	FALSE	FALSE
CP019801	none	FALSE	FALSE	FALSE
CP019802	none	FALSE	FALSE	FALSE
CP019803	none	FALSE	FALSE	FALSE
CP019808	none	FALSE	FALSE	FALSE
CP019809	none	FALSE	FALSE	FALSE
CP019810	none	FALSE	FALSE	FALSE
CP019811	none	FALSE	FALSE	FALSE
CP019812	none	FALSE	FALSE	FALSE
CP019813	none	FALSE	FALSE	FALSE
CP019814	none	FALSE	FALSE	FALSE
CP019815	none	FALSE	FALSE	FALSE
CP019816	none	FALSE	FALSE	FALSE
CP019817	none	FALSE	FALSE	FALSE
CP019818	none	FALSE	FALSE	FALSE
CP019819	none	FALSE	FALSE	FALSE
CP019820	none	FALSE	FALSE	FALSE
CP019821	none	FALSE	FALSE	FALSE
CP019822	none	FALSE	FALSE	FALSE
CP019823	none	FALSE	FALSE	FALSE
CP019824	none	FALSE	FALSE	FALSE
CP019825	none	FALSE	FALSE	FALSE
CP019826	none	FALSE	FALSE	FALSE

CP019827	none	FALSE	FALSE	FALSE
CP019828	none	FALSE	FALSE	FALSE
CP019829	none	FALSE	FALSE	FALSE
CP019830	none	FALSE	FALSE	FALSE
CP019831	none	FALSE	FALSE	FALSE
CP019832	none	FALSE	FALSE	FALSE
CP019833	none	FALSE	FALSE	FALSE
CP019834	none	FALSE	FALSE	FALSE
CP019835	none	FALSE	FALSE	FALSE
CP019836	none	FALSE	FALSE	FALSE
CP019837	none	FALSE	FALSE	FALSE
NZ_CP031556	A&D	TRUE	TRUE	TRUE
NZ_CP019979	A&D	TRUE	TRUE	TRUE
NZ_CP021867	D	TRUE	FALSE	TRUE
CP038809	none	FALSE	FALSE	FALSE
NZ_CP016501	D	TRUE	FALSE	TRUE
		82	21	38

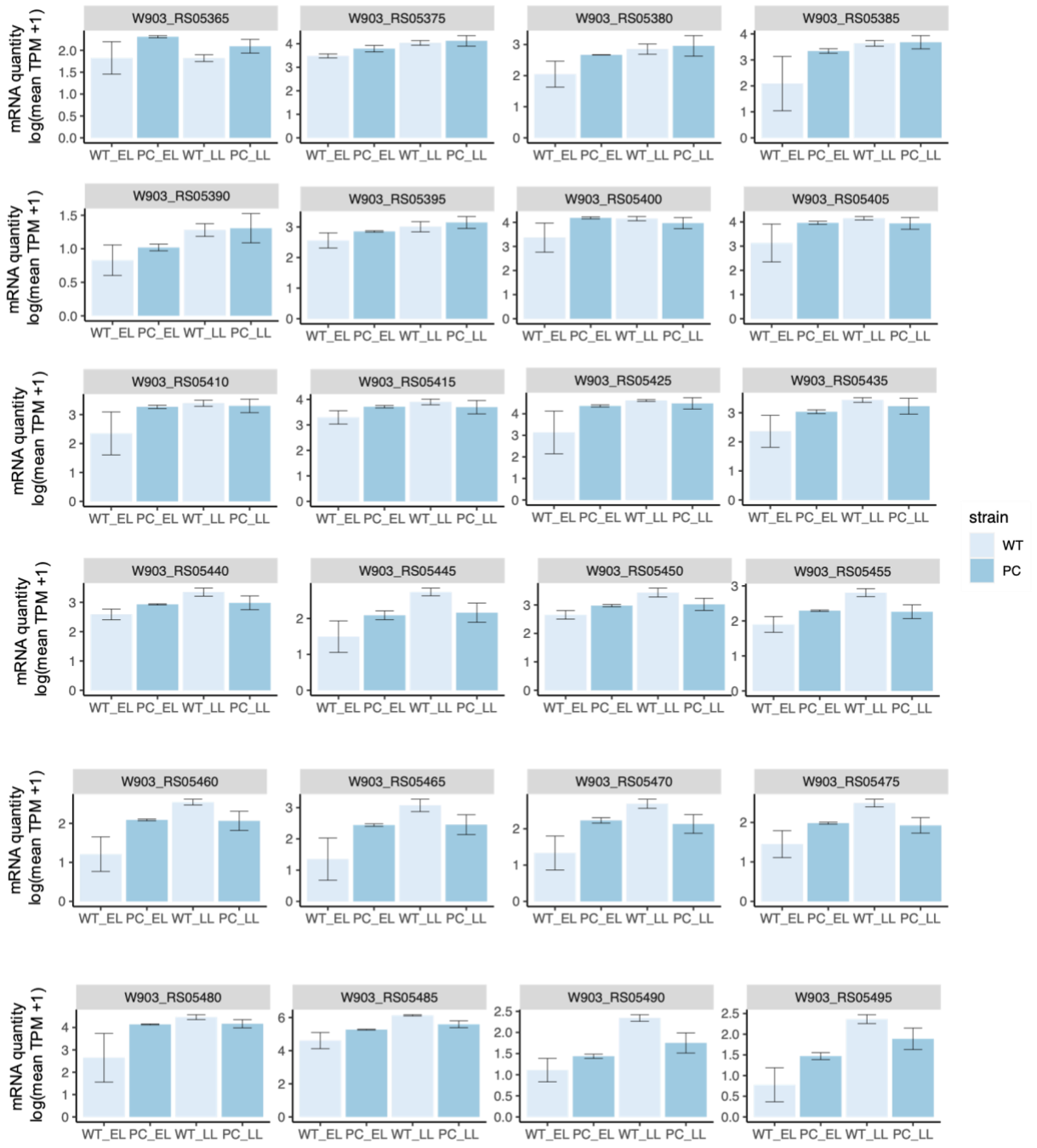


Figure D. 1: mRNA quantity of genes involved T7SS. Genes in the T7SS were not differentially expressed in the absence of the prophage.

BIBLIOGRAPHY

1. Vergnano S, Buttery J, Cailes B, Chandrasekaran R, Chiappini E, Clark E, Cutland C, de Andrade SD, Esteves-Jaramillo A, Guinazu JR, Jones C, Kampmann B, King J, Kochhar S, Macdonald N, Mangili A, de Menezes Martins R, Velasco Muñoz C, Padula M, Muñoz FM, Oleske J, Sanicas M, Schlaudecker E, Spiegel H, Subelj M, Sukumaran L, Tagbo BN, Top KA, Tran D, Heath PT. 2016. Neonatal infections: Case definition and guidelines for data collection, analysis, and presentation of immunisation safety data. Vaccine 34:6038–6046.

2. Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, Rudan I, Campbell H, Cibulskis R, Li M, Mathers C, Black RE. 2012. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. The Lancet 379:2151–2161.

3. Neonatal mortality. UNICEF DATA. https://data.unicef.org/topic/child-survival/neonatal-mortality/. Retrieved 20 March 2023.

4. Cuenca AG, Wynn JL, Moldawer LL, Levy O. 2013. Role of Innate Immunity in Neonatal Infection. Am J Perinatol 30:105–112.

5. Simonsen KA, Anderson-Berry AL, Delair SF, Davies HD. 2014. Early-Onset Neonatal Sepsis. Clin Microbiol Rev 27:21–47.

6. Basha S, Surendran N, Pichichero M. 2014. Immune Responses in Neonates. Expert Rev Clin Immunol 10:1171–1184.

7. Kollmann TR, Kampmann B, Mazmanian SK, Marchant A, Levy O. 2017. Protecting the Newborn and Young Infant from Infectious Diseases: Lessons from Immune Ontogeny. Immunity 46:350–363.

8. Tavares T, Pinho L, Bonifácio Andrade E. 2022. Group B Streptococcal Neonatal Meningitis. Clin Microbiol Rev 35:e00079-21.

9. Palmeira P, Quinello C, Silveira-Lessa AL, Zago CA, Carneiro-Sampaio M. 2012. IgG Placental Transfer in Healthy and Pathological Pregnancies. Clin Dev Immunol 2012:985646.

10. Simon AK, Hollander GA, McMichael A. 2015. Evolution of the immune system in humans from infancy to old age. Proc Biol Sci 282:20143085.

11. Prevention of Group B Streptococcal Early-Onset Disease in Newborns | ACOG. https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2020/02/prevention-of-group-b-streptococcal-early-onset-disease-in-newborns. Retrieved 20 March 2023.

12. Patras KA, Nizet V. 2018. Group B Streptococcal Maternal Colonization and Neonatal Disease: Molecular Mechanisms and Preventative Approaches. Front Pediatr 6:27.

13. Heath PT, Balfour G, Weisner AM, Efstratiou A, Lamagni TL, Tighe H, O'Connell LAF, Cafferkey M, Verlander NQ, Nicoll A, McCartney AC, PHLS Group B Streptococcus Working Group. 2004. Group B streptococcal disease in UK and Irish infants younger than 90 days. Lancet 363:292–294.

14. Bundy LM, Noor A. 2022. Neonatal MeningitisStatPearls. StatPearls Publishing, Treasure Island (FL).

15. KHALESSI N, AFSHARKHAS L. 2014. Neonatal Meningitis: Risk Factors, Causes, and Neurologic Complications. Iran J Child Neurol 8:46–50.

16. Tingskov Pedersen C-E, Eliasen AU, Ketzel M, Brandt J, Loft S, Frohn LM, Khan J, Brix S, Rasmussen MA, Stokholm J, Chawes B, Morin A, Ober C, Bisgaard H, Pedersen M, Bønnelykke K. 2023. Prenatal exposure to ambient air pollution is associated with early life immune perturbations. Journal of Allergy and Clinical Immunology 151:212–221.

17. Stoll BJ, Hansen NI, Sánchez PJ, Faix RG, Poindexter BB, Van Meurs KP, Bizzarro MJ, Goldberg RN, Frantz ID, Hale EC, Shankaran S, Kennedy K, Carlo WA, Watterberg KL, Bell EF, Walsh MC, Schibler K, Laptook AR, Shane AL, Schrag SJ, Das A, Higgins RD. 2011. Early Onset Neonatal Sepsis: The Burden of Group B Streptococcal and E. coli Disease Continues. Pediatrics 127:817–826.

18. Hornik CP, Fort P, Clark RH, Watt K, Benjamin DK, Smith PB, Cohen-Wolkowiez M. 2012. Early and Late Onset Sepsis in Very-Low-Birth-Weight Infants from a Large Group of Neonatal Intensive Care Units. Early Hum Dev 88:S69–S74.

19. Newborn infections. https://www.who.int/teams/maternal-newborn-child-adolescent-health-and-ageing/newborn-health/newborn-infections. Retrieved 24 December 2022.

20. WHO-MCA-17.01-eng.pdf.

21. Weston EJ, Pondo T, Lewis MM, Martell-Cleary P, Morin C, Jewell B, Daily P, Apostol M, Petit S, Farley M, Lynfield R, Reingold A, Hansen NI, Stoll BJ, Shane AL, Zell E, Schrag SJ. 2011. The Burden of Invasive Early-Onset Neonatal Sepsis in the United States, 2005–2008. Pediatr Infect Dis J 30:937–941.

22. Sewell E, Roberts J, Mukhopadhyay S. 2021. ASSOCIATION OF INFECTION IN THE NEONATE AND LONG-TERM NEURODEVELOPMENTAL OUTCOME. Clin Perinatol 48:251–261.

23. Polin RA, the COMMITTEE ON FETUS AND NEWBORN, Papile L-A, Baley JE, Bhutani VK, Carlo WA, Cummings J, Kumar P, Tan RC, Wang KS, Watterberg KL. 2012. Management of Neonates With Suspected or Proven Early-Onset Bacterial Sepsis. Pediatrics 129:1006–1015.

24. Stoll BJ, Hansen N, Fanaroff AA, Wright LL, Carlo WA, Ehrenkranz RA, Lemons JA, Donovan EF, Stark AR, Tyson JE, Oh. W, Bauer CR, Korones SB, Shankaran S, Laptook AR, Stevenson DK, Papile L-A, Poole WK. 2002. Late-Onset Sepsis in Very Low Birth Weight Neonates: The Experience of the NICHD Neonatal Research Network. Pediatrics 110:285–291.

25. Singh M, Alsaleem M, Gray CP. 2022. Neonatal SepsisStatPearls. StatPearls Publishing, Treasure Island (FL).

26. Oumer M, Abebaw D, Tazebew A. 2022. Time to recovery of neonatal sepsis and determinant factors among neonates admitted in Public Hospitals of Central Gondar Zone, Northwest Ethiopia, 2021. PLOS ONE 17:e0271997.

27. 2022. Infant Mortality | Maternal and Infant Health | Reproductive Health | CDC. https://www.cdc.gov/reproductivehealth/maternalinfanthealth/infantmortality.htm. Retrieved 21 March 2023.

28. Brouwer MC, Tunkel AR, Beek D van de. 2010. Epidemiology, Diagnosis, and Antimicrobial Treatment of Acute Bacterial Meningitis. Clinical Microbiology Reviews 23:467.

29. Kim KS. 2010. Acute bacterial meningitis in infants and children. The Lancet Infectious Diseases 10:32–42.

30. Puopolo KM, Mukhopadhay S, Frymoyer A, Benitz WE. 2021. The Term Newborn. Clinics in Perinatology 48:471–484.

31. Le Guennec L, Coureuil M, Nassif X, Bourdoulous S. 2020. Strategies used by bacterial pathogens to cross the blood–brain barrier. Cellular Microbiology 22.

32. Davies HG, Carreras-Abad C, Le Doare K, Heath PT. 2019. Group B Streptococcus: Trials and Tribulations. The Pediatric Infectious Disease Journal 38:S72.

33. Murray PR, Rosenthal KS, Pfaller MA. 2020. Medical Microbiology E-Book. Elsevier Health Sciences.

34. Lancefield RC. 1933. A SEROLOGICAL DIFFERENTIATION OF HUMAN AND OTHER GROUPS OF HEMOLYTIC STREPTOCOCCI. Journal of Experimental Medicine 57:571–595.

35. Lancefield RC. THE ANTIGENIC COMPLEX OF STREPTOCOCCUS HA~MOLYTICUS.

36. Patterson MJ. 1996. Streptococcus, p. . *In* Baron, S (ed.), Medical Microbiology, 4th ed. University of Texas Medical Branch at Galveston, Galveston (TX).

37. Edwards: CHAPTER 12-Group B Streptococcal Infections - Google Scholar. https://scholar.google.com/scholar_lookup?author=M.+S.+Edwards&author=V.+Nizet&author=C.+J.+Baker&publication_year=2011&title=CHAPTER+12%E2%80%93group+B+streptococcal+infections&journal=Infectious+diseases+of+the+fetus+and+newborn.+7th+edn&pages=419-469. Retrieved 4 January 2023.

38. Lin FY, Clemens JD, Azimi PH, Regan JA, Weisman LE, Philips JB, Rhoads GG, Clark P, Brenner RA, Ferrieri P. 1998. Capsular polysaccharide types of group B streptococcal isolates from neonates with early-onset systemic infection. J Infect Dis 177:790–792.

39. Bianchi-Jassir F, Paul P, To K-N, Carreras-Abad C, Seale AC, Jauneikaite E, Madhi SA, Russell NJ, Hall J, Madrid L, Bassat Q, Kwatra G, Le Doare K, Lawn JE. 2020. Systematic review of Group B Streptococcal capsular types, sequence types and surface proteins as potential vaccine candidates. Vaccine 38:6682–6694.

40. Trijbels-Smeulders MAJM, Kimpen JLL, Kollée LAA, Bakkers J, Melchers W, Spanjaard L, Wannet WJB, Hoogkamp-Korstanje MAA. 2006. SEROTYPES, GENOTYPES, AND ANTIBIOTIC SUSCEPTIBILITY PROFILES OF GROUP B STREPTOCOCCI CAUSING NEONATAL SEPSIS AND MENINGITIS BEFORE AND AFTER

INTRODUCTION OF ANTIBIOTIC PROPHYLAXIS. The Pediatric Infectious Disease Journal 25:945.

41. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan M-S, Kunst F, Glaser P, Rusniok C, Crook DWM, Harding RM, Bisharat N, Spratt BG. 2003. Multilocus Sequence Typing System for Group B Streptococcus. J Clin Microbiol 41:2530–2536.

42. McGee L, Chochua S, Li Z, Mathis S, Rivers J, Metcalf B, Ryan A, Alden N, Farley MM, Harrison LH, Snippes Vagnone P, Lynfield R, Smelser C, Muse A, Thomas AR, Schrag S, Beall BW. 2021. Multistate, Population-Based Distributions of Candidate Vaccine Targets, Clonal Complexes, and Resistance Features of Invasive Group B Streptococci Within the United States, 2015–2017. Clinical Infectious Diseases 72:1004–1013.

43. Burcham LR, Spencer BL, Keeler LR, Runft DL, Patras KA, Neely MN, Doran KS. 2019. Determinants of Group B streptococcal virulence potential amongst vaginal clinical isolates from pregnant women. PLoS ONE 14:e0226699.

44. Gori A, Harrison OB, Mlia E, Nishihara Y, Chan JM, Msefula J, Mallewa M, Dube Q, Swarthout TD, Nobbs AH, Maiden MCJ, French N, Heyderman RS. 2020. Pan-GWAS of Streptococcus agalactiae Highlights Lineage-Specific Genes Associated with Virulence and Niche Adaptation. mBio 11:e00728-20.

45. Shabayek S, Spellerberg B. 2018. Group B Streptococcal Colonization, Molecular Characteristics, and Epidemiology. Front Microbiol 9:437.

46. Keefe GP. 1997. Streptococcus agalactiae mastitis: a review. Can Vet J 38:429–437.

47. Ruegg PL. 2017. A 100-Year Review: Mastitis detection, management, and prevention. J Dairy Sci 100:10381–10397.

48. Lancefield RC, Hare R. 1935. THE SEROLOGICAL DIFFERENTIATION OF PATHOGENIC AND NON-PATHOGENIC STRAINS OF HEMOLYTIC STREPTOCOCCI FROM PARTURIENT WOMEN. Journal of Experimental Medicine 61:335–349.

49. Jones HE, Howells CH. 1968. Neonatal meningitis due to Streptococcus agalactiae. Postgrad Med J 44:549–551.

50. Anthony BF, Okada DM. 1977. The Emergence of Group B Streptococci in Infections of the Newborn Infant. Annual Review of Medicine 28:355–369.

51. McCracken GH. 1973. Group B streptococci: The new challenge in neonatal infections. The Journal of Pediatrics 82:703–706.

52. Boyer KM, Gadzala CA, Burd LI, Fisher DE, Paton JB, Gotoff SP. 1983. Selective Intrapartum Chemoprophylaxis of Neonatal Group B Streptococcal Early-Onset Disease. I. Epidemiologic Rationale. The Journal of Infectious Diseases 148:795–801.

53. Fluegge K, Siedler A, Heinrich B, Schulte-Moenting J, Moennig M-J, Bartels DB, Dammann O, von Kries R, Berner R, German Pediatric Surveillance Unit Study Group. 2006. Incidence and clinical presentation of invasive neonatal group B streptococcal infections in Germany. Pediatrics 117:e1139-1145.

54. Kalliola S, Vuopio-Varkila J, Takala AK, Eskola J. 1999. Neonatal group B streptococcal disease in Finland: a ten-year nationwide study. Pediatr Infect Dis J 18:806–810.

55. Neto MT. 2008. Group B streptococcal disease in Portuguese infants younger than 90 days. Arch Dis Child Fetal Neonatal Ed 93:F90-93.

56. Anthony BF, Carter JA, Eisenstadt R, Rimer DG. 1983. Isolation of Group B Streptococci from the Proximal Small Intestine of Adults. The Journal of Infectious Diseases 147:776.

57. Kneafsey PD, Kelly JK, Church DL, Rapp EF, Lafreniere R. 1987. Phlegmonous duodenitis complicating multiple myeloma: a successfully treated case. Am J Gastroenterol 82:1322–1325.

58. Jones N, Oliver K, Jones Y, Haines A, Crook D. 2006. Carriage of group B streptococcus in pregnant women from Oxford, UK. J Clin Pathol 59:363–366.

59. Bergeron MG, Ke D, Ménard C, François FJ, Gagnon M, Bernier M, Ouellette M, Roy PH, Marcoux S, Fraser WD. 2000. Rapid Detection of Group B Streptococci in Pregnant Women at Delivery. N Engl J Med 343:175–179.

60. Johri AK, Paoletti LC, Glaser P, Dua M, Sharma PK, Grandi G, Rappuoli R. 2006. Group B Streptococcus: global incidence and vaccine development. 12. Nat Rev Microbiol 4:932–942.

61. Verani JR, McGee L, Schrag SJ, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention

(CDC). 2010. Prevention of perinatal group B streptococcal disease--revised guidelines from CDC, 2010. MMWR Recomm Rep 59:1–36.

62. Kunze M, Ziegler A, Fluegge K, Hentschel R, Proempeler H, Berner R. 2011. Colonization, serotypes and transmission rates of group B streptococci in pregnant women and their infants born at a single University Center in Germany 39:417–422.

63. Kwatra G, Cunnington MC, Merrall E, Adrian PV, Ip M, Klugman KP, Tam WH, Madhi SA. 2016. Prevalence of maternal colonisation with group B streptococcus: a systematic review and meta-analysis. The Lancet Infectious Diseases 16:1076–1084.

64. Matsubara K, Yamamoto G. 2009. Invasive group B streptococcal infections in a tertiary care hospital between 1998 and 2007 in Japan. International Journal of Infectious Diseases 13:679–684.

65. Stoll BJ, Hansen NI, Bell EF, Shankaran S, Laptook AR, Walsh MC, Hale EC, Newman NS, Schibler K, Carlo WA, Kennedy KA, Poindexter BB, Finer NN, Ehrenkranz RA, Duara S, Sánchez PJ, O'Shea TM, Goldberg RN, Van Meurs KP, Faix RG, Phelps DL, Frantz ID, Watterberg KL, Saha S, Das A, Higgins RD. 2010. Neonatal Outcomes of Extremely Preterm Infants From the NICHD Neonatal Research Network. Pediatrics 126:443–456.

66. Skoff TH, Farley MM, Petit S, Craig AS, Schaffner W, Gershman K, Harrison LH, Lynfield R, Mohle-Boetani J, Zansky S, Albanese BA, Stefonek K, Zell ER, Jackson D, Thompson T, Schrag SJ. 2009. Increasing Burden of Invasive Group B Streptococcal Disease in Nonpregnant Adults, 1990–2007. Clinical Infectious Diseases 49:85–92.

67. Trijbels-Smeulders MAJM, Kollée LAA, Adriaanse AH, Kimpen JLL, Gerards LJ. 2004. Neonatal group B streptococcal infection: incidence and strategies for prevention in Europe. The Pediatric Infectious Disease Journal 23:172.

68. Janowski AB, Newland JG. 2017. From the microbiome to the central nervous system, an update on the epidemiology and pathogenesis of bacterial meningitis in childhood. F1000Res 6:F1000 Faculty Rev-86.

69. MacDorman MF, Kirmeyer S. 2009. Fetal and perinatal mortality, United States, 2005. Natl Vital Stat Rep 57:1–19.

70. Farley MM, Strasbaugh LJ. 2001. Group B Streptococcal Disease in Nonpregnant Adults. Clinical Infectious Diseases 33:556–561.

71. Schwartz B, Schuchat A, Oxtoby MJ, Cochi SL, Hightower A, Broome CV. 1991. Invasive group B streptococcal disease in adults. A population-based study in metropolitan Atlanta. JAMA 266:1112–1114.

72. Gonçalves BP, Procter SR, Paul P, Chandna J, Lewin A, Seedat F, Koukounari A, Dangor Z, Leahy S, Santhanam S, John HB, Bramugy J, Bardají A, Abubakar A, Nasambu C, Libster R, Yanotti CS, Horváth-Puhó E, Sørensen HT, Beek D van de, Bijlsma MW, Gardner WM, Kassebaum N, Trotter C, Bassat Q, Madhi SA, Lambach P, Jit M, Lawn JE, Søgaard KK, Kassel MN van, Snoek L, Gier B de, Ende A van der, Hahné SJM, Harden LM, Ghoor A, Mbatha S, Lowick S, Laughton B, Jaye T, Lala SG, Sithole P, Msayi J, Kumalo N, Msibi TN, Arumugam A, Murugesan N, Rajendraprasad N, Priya M, Mabrouk A, Katana PV, Mwangome E, Newton CR, Mucasse H, Aerts C, Massora S, Medina V,

Rojas A, Amado D, Llapur CJ, Hossain AKMT, Rahman QS, Ip M, Seale A, Heath PT, Doare KL, Khalil A, Schrag SJ, Meulen AS, Mason E, Blau DM, Arifeen SE, Assefa N, Onyango D, Sow SO, Mandomando I, Ogbuanu I, Kotloff KL, Scott JAG, Gurley ES, Barr BAT, Mahtab S. 2022. Group B streptococcus infection during pregnancy and infancy: estimates of regional and global burden. The Lancet Global Health 10:e807–e819.

73. Russell NJ, Seale AC, O'Driscoll M, O'Sullivan C, Bianchi-Jassir F, Gonzalez-Guarin J, Lawn JE, Baker CJ, Bartlett L, Cutland C, Gravett MG, Heath PT, Le Doare K, Madhi SA, Rubens CE, Schrag S, Sobanjo-Ter Meulen A, Vekemans J, Saha SK, Ip M, GBS Maternal Colonization Investigator Group. 2017. Maternal Colonization With Group B Streptococcus and Serotype Distribution Worldwide: Systematic Review and Meta-analyses. Clin Infect Dis 65:S100–S111.

74. Mukhopadhyay S, Wade KC, Puopolo KM. 2019. Drugs for the Prevention and Treatment of Sepsis in the Newborn. Clin Perinatol 46:327–347.

75. Schuchat A. 1998. Epidemiology of Group B Streptococcal Disease in the United States: Shifting Paradigms. Clin Microbiol Rev 11:497–513.

76. Steer PJ, Russell AB, Kochhar S, Cox P, Plumb J, Gopal Rao G. 2020. Group B streptococcal disease in the mother and newborn—A review. Eur J Obstet Gynecol Reprod Biol 252:526–533.

77. Baker CJ, Barrett FF. 1973. Transmission of group B streptococci among parturient women and their neonates. The Journal of Pediatrics 83:919–925.

78. Chin KC, Fitzhardinge PM. 1985. Sequelae of early-onset group B hemolytic streptococcal neonatal meningitis. The Journal of Pediatrics 106:819–822.

79. Edwards MS, Rench MA, Haffar AAM, Murphy MA, Desmond MM, Baker CJ. 1985. Long-term sequelae of group B streptococcal meningitis in infants. The Journal of Pediatrics 106:717–722.

80. Wald ER, Bergman I, Taylor HG, Chiponis D, Porter C, Kubek K. 1986. Long-term outcome of group B streptococcal meningitis. Pediatrics 77:217–221.

81. Lawn JE, Bianchi-Jassir F, Russell NJ, Kohli-Lynch M, Tann CJ, Hall J, Madrid L, Baker CJ, Bartlett L, Cutland C, Gravett MG, Heath PT, Ip M, Le Doare K, Madhi SA, Rubens CE, Saha SK, Schrag S, Sobanjo-ter Meulen A, Vekemans J, Seale AC. 2017. Group B Streptococcal Disease Worldwide for Pregnant Women, Stillbirths, and Children: Why, What, and How to Undertake Estimates? Clinical Infectious Diseases 65:S89–S99.

82. Pangerl S, Sundin D, Geraghty S. 2021. Group B Streptococcus Screening Guidelines in Pregnancy: A Critical Review of Compliance. Matern Child Health J 25:257–267.

83. Schrag SJ, Zell ER, Lynfield R, Roome A, Arnold KE, Craig AS, Harrison LH, Reingold A, Stefonek K, Smith G, Gamble M, Schuchat A. 2002. A Population-Based Comparison of Strategies to Prevent Early-Onset Group B Streptococcal Disease in Neonates. New England Journal of Medicine 347:233–239.

84. Money DM, Bc V, Dobson S, Bc V, Embree J. 2004. THE PREVENTION OF EARLY-ONSET NEONATAL GROUP B STREPTOCOCCAL DISEASE.

85. Joachim A, Matee MI, Massawe FA, Lyamuya EF. 2009. Maternal and neonatal colonisation of group B streptococcus at Muhimbili National Hospital in Dar es Salaam, Tanzania: prevalence, risk factors and antimicrobial resistance. BMC Public Health 9:437.

86. Berg AWV den, Houtman-Roelofsen RL, Oostvogel PM, Dekker FW, Dörr PJ, Sprij AJ. 2010. Timing of Group B Streptococcus Screening in Pregnancy: A Systematic Review. GOI 69:174–183.

87. Yow MD, Mason EO, Leeds LJ, Thompson PK, Clark DJ, Gardner SE. 1979. Ampicillin Prevents Intrapartum Transmission of Group B Streptococcus. JAMA 241:1245–1247.

88. Albright CM, MacGregor C, Sutton D, Theva M, Hughes BL, Werner EF. 2017. Group B Streptococci Screening Before Repeat Cesarean Delivery: A Cost-Effectiveness Analysis. Obstetrics & Gynecology 129:111.

89. Brzychczy-Wloch M, Wojkowska-Mach J, Helwich E, Heczko PB. 2013. Incidence of maternal GBS colonization and neonatal GBS disease among Very Low Birth Weight Polish neonates. Med Sci Monit 19:34–39.

90. Ledger WJ, Blaser MJ. 2013. Are we using too many antibiotics during pregnancy? A Commentary. BJOG 120:1450–1452.

91. Schuchat A, Roome A, Zell ER, Linardos H, Zywicki S, O'Brien KL. 2002. Integrated Monitoring of a New Group B Streptococcal Disease Prevention Program and Other Perinatal Infections. Matern Child Health J 6:107–114.

92. Ohlsson A, Shah VS. 2014. Intrapartum antibiotics for known maternal Group B streptococcal colonization. Cochrane Database of Systematic Reviews https://doi.org/10.1002/14651858.CD007467.pub4.

93. Norwitz ER, Greenberg JA. 2009. Antibiotics in Pregnancy: Are They Safe? Rev Obstet Gynecol 2:135–136.

94. Szymański S, Szczerba K, Konstanty- Kurkiewicz V, Daszkiewicz A, Sipak- Szmigiel O. 2018. Infections of GBS's etiology and use of perinatal antibiotic prophylaxis. J Pre Clin Clin Res 12:54–58.

95. Berger M, Xu X, Williams J, Van de Ven C, Mozurkewich E. 2012. Early hospital discharge of infants born to group B streptococci-positive mothers: a decision analysis: Early hospital discharge of infants from GBS-positive women. BJOG: An International Journal of Obstetrics & Gynaecology 119:439–448.

96. Turrentine MA, Ramirez MM, Mastrobattista JM. 2009. Cost-Effectiveness of Universal Prophylaxis in Pregnancy with Prior Group B Streptococci Colonization. Infect Dis Obstet Gynecol 2009:934698.

97. Seale AC, Bianchi-Jassir F, Russell NJ, Kohli-Lynch M, Tann CJ, Hall J, Madrid L, Blencowe H, Cousens S, Baker CJ, Bartlett L, Cutland C, Gravett MG, Heath PT, Ip M, Le Doare K, Madhi SA, Rubens CE, Saha SK, Schrag SJ, Sobanjo-ter Meulen A, Vekemans J, Lawn JE. 2017. Estimates of the Burden of Group B Streptococcal Disease Worldwide for Pregnant Women, Stillbirths, and Children. Clin Infect Dis 65:S200–S219.

98. Baker CJ, Edwards MS. 1988. Group B Streptococcal Infections. Annals of the New York Academy of Sciences 549:193–202.

99. Le Doare K, Heath PT. 2013. An overview of global GBS epidemiology. Vaccine 31:D7–D12.

100. Rohmer L, Hocquet D, Miller SI. 2011. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. Trends in Microbiology 19:341–348.

101. Willenborg J, Goethe R. 2016. Metabolic traits of pathogenic streptococci. FEBS Letters 590:3905–3919.

102. Goldenberg RL, Hauth JC, Andrews WW. 2000. Intrauterine infection and preterm delivery. N Engl J Med 342:1500–1507.

103. Nitsche-Schmitz PD, Rohde M, Chhatwal GS. 2007. Invasion mechanisms of Gram-positive pathogenic cocci. Thromb Haemost 98:488–496.

104. Tamura GS, Kuypers JM, Smith S, Raff H, Rubens CE. 1994. Adherence of group B streptococci to cultured epithelial cells: roles of environmental factors and bacterial surface components. Infect Immun 62:2450–2458.

105. Yamamoto Y, Poyart C, Trieu-Cuot P, Lamberet G, Gruss A, Gaudu P. 2005. Respiration metabolism of Group B Streptococcus is activated by environmental haem and quinone and contributes to virulence. Molecular Microbiology 56:525–534.

106. Di Palo B, Rippa V, Santi I, Brettoni C, Muzzi A, Metruccio MME, Grifantini R, Telford JL, Paccani SR, Soriani M. 2013. Adaptive Response of Group B Streptococcus to High

Glucose Conditions: New Insights on the CovRS Regulation Network. PLoS One 8:e61294.

107. Sitkiewicz I, Green NM, Guo N, Bongiovanni AM, Witkin SS, Musser JM. 2009. Transcriptome Adaptation of Group B Streptococcus to Growth in Human Amniotic Fluid. PLoS One 4:e6114.

108. Landwehr-Kenzel S, Henneke P. 2014. Interaction of Streptococcus agalactiae and Cellular Innate Immunity in Colonization and Disease. Front Immunol 5:519.

109. Flaherty RA, Borges EC, Sutton JA, Aronoff DM, Gaddy JA, Petroff MG, Manning SD. 2019. Genetically distinct Group B Streptococcus strains induce varying macrophage cytokine responses. PLoS One 14:e0222910.

110. Jiang S, Wessels MR. 2014. BsaB, a Novel Adherence Factor of Group B Streptococcus. Infection and Immunity 82:1007–1016.

111. Lembo A, Gurney MA, Burnside K, Banerjee A, Reyes M de los, Connelly JE, Lin W-J, Jewell KA, Vo A, Renken CW, Doran KS, Rajagopal L. 2010. Regulation of CovR expression in Group B streptococcus impacts blood-brain barrier penetration. Mol Microbiol 77:431–443.

112. Jiang S-M, Cieslewicz MJ, Kasper DL, Wessels MR. 2005. Regulation of Virulence by a Two-Component System in Group B Streptococcus. Journal of Bacteriology 187:1105–1113.

113. Teatero S, Ferrieri P, Martin I, Demczuk W, McGeer A, Fittipaldi N. 2017. Serotype Distribution, Population Structure, and Antimicrobial Resistance of Group B Streptococcus Strains Recovered from Colonized Pregnant Women. J Clin Microbiol 55:412–422.

114. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." Proc Natl Acad Sci U S A 102:13950–13955.

115. Cieslewicz MJ, Chaffin D, Glusman G, Kasper D, Madan A, Rodrigues S, Fahey J, Wessels MR, Rubens CE. 2005. Structural and Genetic Diversity of Group B Streptococcus Capsular Polysaccharides. Infect Immun 73:3096–3103.

116. Alves-Barroco C, Rivas-García L, Fernandes AR, Baptista PV. 2020. Tackling Multidrug Resistance in Streptococci – From Novel Biotherapeutic Strategies to Nanomedicines. Frontiers in Microbiology 11.

117. Richards VP, Lang P, Pavinski Bitar PD, Lefébure T, Schukken YH, Zadoks RN, Stanhope MJ. 2011. Comparative genomics and the role of lateral gene transfer in the

evolution of bovine adapted Streptococcus agalactiae. Infection, Genetics and Evolution 11:1263–1275.

118. Brochet M, Couvé E, Glaser P, Guédon G, Payot S. 2008. Integrative conjugative elements and related elements are major contributors to the genome diversity of Streptococcus agalactiae. J Bacteriol 190:6913–6917.

119. Brochet M, Rusniok C, Couvé E, Dramsi S, Poyart C, Trieu-Cuot P, Kunst F, Glaser P. 2008. Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of Streptococcus agalactiae. Proc Natl Acad Sci U S A 105:15961–15966.

120. Domelier A-S, van der Mee-Marquet N, Sizaret P-Y, Héry-Arnaud G, Lartigue M-F, Mereghetti L, Quentin R. 2009. Molecular Characterization and Lytic Activities of Streptococcus agalactiae Bacteriophages and Determination of Lysogenic-Strain Features. Journal of Bacteriology 191:4776–4785.

121. Salloum M, van der Mee-Marquet N, Valentin-Domelier A-S, Quentin R. 2011. Diversity of Prophage DNA Regions of Streptococcus agalactiae Clonal Lineages from Adults and Neonates with Invasive Infectious Disease. PLoS ONE 6:e20256.

122. Giovanetti E, Brenciani A, Morroni G, Tiberi E, Pasquaroli S, Mingoia M, Varaldo PE. 2015. Transduction of the Streptococcus pyogenes bacteriophage Φm46.1, carrying resistance genes mef(A) and tet(O), to other Streptococcus species. Frontiers in Microbiology 5.

123. Weitz JS, Poisot T, Meyer JR, Flores CO, Valverde S, Sullivan MB, Hochberg ME. 2013. Phage-bacteria infection networks. Trends Microbiol 21:82–91.

124. Twort FW. 1915. AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. The Lancet 186:1241–1243.

125. d'Herelle F. 1917. An invisible microbe that is antagonistic to the dysentery bacillus. CR Acad Sci 165:373–375.

126. Salmond GPC, Fineran PC. 2015. A century of the phage: past, present and future. Nat Rev Microbiol 13:777–786.

127. Garriss G, Henriques-Normark B. 2020. Lysogeny in Streptococcus pneumoniae. 10. Microorganisms 8:1546.

128. Keen EC. 2015. A century of phage research: Bacteriophages and the shaping of modern biology. Bioessays 37:6–9.

129. Kasman LM, Porter LD. 2022. BacteriophagesStatPearls. StatPearls Publishing, Treasure Island (FL).

130. Abril AG, Carrera M, Notario V, Sánchez-Pérez Á, Villa TG. 2022. The Use of Bacteriophages in Biotechnology and Recent Insights into Proteomics. Antibiotics (Basel) 11:653.

131. Lin DM, Koskella B, Lin HC. 2017. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. World J Gastrointest Pharmacol Ther 8:162–173.

132. Skurnik M, Pajunen M, Kiljunen S. 2007. Biotechnological challenges of phage therapy. Biotechnol Lett 29:995–1003.

133. Carlton RM. 1999. Phage therapy: past history and future prospects. Arch Immunol Ther Exp (Warsz) 47:267–274.

134. Gupta R, Prasad Y. 2011. Efficacy of Polyvalent Bacteriophage P-27/HP to Control Multidrug Resistant Staphylococcus aureus Associated with Human Infections. Curr Microbiol 62:255–260.

135. Batinovic S, Wassef F, Knowler SA, Rice DTF, Stanton CR, Rose J, Tucci J, Nittami T, Vinh A, Drummond GR, Sobey CG, Chan HT, Seviour RJ, Petrovski S, Franks AE. 2019. Bacteriophages in Natural and Artificial Environments. Pathogens 8:100.

136. Goodfellow PN. 1987. A Genetic Switch: Gene Control and Phage λ. J Med Genet 24:789–790.

137. King A, Adams M, Carstens E, Lefkowitz E. 2012. Order–Caudovirales. Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses, eds AMQ King, MJ Adams, EB Carstens, and EJ Lefkowitz (Amstadam: Elsevier) 39–45.

138. De Smet J, Hendrix H, Blasdel BG, Danis-Wlodarczyk K, Lavigne R. 2017. Pseudomonas predators: understanding and exploiting phage–host interactions. Nature Reviews Microbiology 15:517–530.

139. Hobbs Z, Abedon ST. 2016. Diversity of phage infection types and associated terminology: the problem with 'Lytic or lysogenic.' FEMS microbiology letters 363.

140. Grabowski Ł, Łepek K, Stasiłojć M, Kosznik-Kwaśnicka K, Zdrojewska K, Maciąg-Dorszyńska M, Węgrzyn G, Węgrzyn A. 2021. Bacteriophage-encoded enzymes

destroying bacterial cell membranes and walls, and their potential use as antimicrobial agents. Microbiological Research 248:126746.

141. Casjens SR, Hendrix RW. 2015. Bacteriophage lambda: Early pioneer and still relevant. Virology 479–480:310–330.

142. Stone E, Campbell K, Grant I, McAuliffe O. 2019. Understanding and Exploiting Phage–Host Interactions. Viruses 11:567.

143. Broussard GW, Oldfield LM, Villanueva VM, Lunt BL, Shine EE, Hatfull GF. 2013. Integration-Dependent Bacteriophage Immunity Provides Insights into the Evolution of Genetic Switches. Molecular Cell 49:237–248.

144. Henrot C, Petit M-A. 2022. Signals triggering prophage induction in the gut microbiota. Molecular Microbiology 118:494–502.

145. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. 2017. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. 7. ISME J 11:1511–1520.

146. Golding I. 2011. DECISION MAKING IN LIVING CELLS: LESSONSFROM A SIMPLE SYSTEM. Annu Rev Biophys 40:63–80.

147. Schneider CL. 2017. Bacteriophage-Mediated Horizontal Gene Transfer: Transduction, p. 1–42. *In* Harper, D, Abedon, S, Burrowes, B, McConville, M (eds.), Bacteriophages: Biology, Technology, Therapy. Springer International Publishing, Cham.

148. Roucourt B, Lavigne R. 2009. The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. Environmental Microbiology 11:2789–2805.

149. Mavrich TN, Hatfull GF. 2017. Bacteriophage evolution differs by host, lifestyle and genome. Nat Microbiol 2:17112.

150. Clokie MR, Millard AD, Letarov AV, Heaphy S. 2011. Phages in nature. Bacteriophage 1:31–45.

151. Dowah ASA, Clokie MRJ. 2018. Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. Biophys Rev 10:535–542.

152. Rampersad S, Tennant P. 2018. Replication and Expression Strategies of Viruses. Viruses 55–82.

153. Borodovich T, Shkoporov AN, Ross RP, Hill C. 2022. Phage-mediated horizontal gene transfer and its implications for the human gut microbiome. Gastroenterol Rep (Oxf) 10:goac012.

154. Purohit PK, Inamdar MM, Grayson PD, Squires TM, Kondev J, Phillips R. 2005. Forces during Bacteriophage DNA Packaging and Ejection. Biophysical Journal 88:851–866.

155. Hatfull GF, Hendrix RW. 2011. Bacteriophages and their Genomes. Curr Opin Virol 1:298–303.

156. M. Iyer L, Anantharaman V, Krishnan A, Burroughs AM, Aravind L. 2021. Jumbo Phages: A Comparative Genomic Overview of Core Functions and Adaptions for Biological Conflicts. Viruses 13:63.

157. Brüssow H, Canchaya C, Hardt W-D. 2004. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. Microbiol Mol Biol Rev 68:560–602.

158. Federici S, Nobs SP, Elinav E. 2021. Phages and their potential to modulate the microbiome and immunity. Cell Mol Immunol 18:889–904.

159. Shkoporov AN, Turkington CJ, Hill C. 2022. Mutualistic interplay between bacteriophages and bacteria in the human gut. 12. Nat Rev Microbiol 20:737–749.

160. Penadés JR, Chen J, Quiles-Puchalt N, Carpena N, Novick RP. 2015. Bacteriophage-mediated spread of bacterial virulence genes. Current Opinion in Microbiology 23:171–178.

161. Feiner R, Argov T, Rabinovich L, Sigal N, Borovok I, Herskovits AA. 2015. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. 10. Nat Rev Microbiol 13:641–650.

162. Little JW. 2005. Lysogeny, Prophage Induction, and Lysogenic Conversion, p. 37–54. *In* Phages. John Wiley & Sons, Ltd.

163. Los M, Kuzio J, Mcconnell M, Kropinski A, Wegrzyn G, Christie G. 2010. Lysogenic Conversion in Bacteria of Importance to the Food Industry. Bacteriophage in the Detection and Control of Foodborne Pathogens 157–198.

164. Bondy-Denomy J, Davidson AR. 2014. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. J Microbiol 52:235–242.

165. Obeng N, Pratama AA, Elsas JD van. 2016. The Significance of Mutualistic Phages for Bacterial Ecology and Evolution. Trends in Microbiology 24:440–449.

166. Harrison E, Brockhurst MA. 2017. Ecological and Evolutionary Benefits of Temperate Phage: What Does or Doesn't Kill You Makes You Stronger. BioEssays 39:1700112.

167. Das B, Bischerour J, Barre F-X. 2011. Molecular mechanism of acquisition of the cholera toxin genes. Indian J Med Res 133:195–200.

168. Colavecchio A, Cadieux B, Lo A, Goodridge LD. 2017. Bacteriophages Contribute to the Spread of Antibiotic Resistance Genes among Foodborne Pathogens of the Enterobacteriaceae Family – A Review. Front Microbiol 8:1108.

169. Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. 2003. Prophage Genomics. Microbiol Mol Biol Rev 67:238–276.

170. Al Mamun AA, Tominaga A, Enomoto M. 1997. Cloning and characterization of the region III flagellar operons of the four Shigella subgroups: genetic defects that cause loss of flagella of Shigella boydii and Shigella sonnei. J Bacteriol 179:4493–4500.

171. Maurelli AT, Fernández RE, Bloch CA, Rode CK, Fasano A. 1998. “Black holes” and bacterial pathogenicity: a large genomic deletion that enhances the virulence of Shigella spp. and enteroinvasive Escherichia coli. Proc Natl Acad Sci U S A 95:3943–3948.

172. Koskella B, Meaden S. 2013. Understanding Bacteriophage Specificity in Natural Microbial Communities. Viruses 5:806–823.

173. Vojtek I, Pirzada ZA, Henriques-Normark B, Mastny M, Janapatla RP, Charpentier E. 2008. Lysogenic Transfer of Group A Streptococcus Superantigen Gene among Streptococci. J Infect Dis 197:225–234.

174. Banks DJ, Beres SB, Musser JM. 2002. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. Trends Microbiol 10:515–521.

175. Russell H, Norcross NL, Kahn DE. 1969. Isolation and Characterization of Streptococcus agalactiae Bacteriophage. Journal of General Virology 5:315–317.

176. Stringer J 1980. THE DEVELOPMENT OF A PHAGE-TYPING SYSTEM FOR GROUP-B STREPTOCOCCI. Journal of Medical Microbiology 13:133–144.

177. Anthony BF, Okada DM, Hobel CJ. 1979. Epidemiology of the group B streptococcus: Maternal and nosocomial sources for infant acquisitions. The Journal of Pediatrics 95:431–436.

178. Band JD, Clegg HW, Hayes PS, Facklam RR, Stringer J, Dixon RE, Maxted WR, Fraser DW. 1981. Transmission of Group B Streptococci: Traced by Use of Multiple Epidemiologic Markers. American Journal of Diseases of Children 135:355–358.

179. Boyer KM, Vogel LC, Gotoff SP, Gadzala CA, Stringer J, Maxted WR. 1980. Nosocomial Transmission of Bacteriophage Type 7/11/12 Group B Streptococci in a Special Care Nursery. American Journal of Diseases of Children 134:964–966.

180. Noya FJD, Rench MA, Metzger TG, Colman G, Naidoo J, Baker CJ. 1987. Unusual Occurrence of an Epidemic of Type Ib/c Group B Streptococcal Sepsis in a Neonatal Intensive Care Unit. The Journal of Infectious Diseases 155:1135–1144.

181. Cheng Q, Nelson D, Zhu S, Fischetti VA. 2005. Removal of Group B Streptococci Colonizing the Vagina and Oropharynx of Mice with a Bacteriophage Lytic Enzyme. Antimicrobial Agents and Chemotherapy 49:111–117.

182. Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, Peterson S, Wessels MR, Paulsen IT, Nelson KE, Margarit I, Read TD, Madoff LC, Wolf AM, Beanan MJ, Brinkac LM, Daugherty SC, DeBoy RT, Durkin AS, Kolonay JF, Madupu R, Lewis MR, Radune D, Fedorova NB, Scanlan D, Khouri H, Mulligan S, Carty HA, Cline RT, Van Aken SE, Gill J, Scarselli M, Mora M, Iacobini ET, Brettoni C, Galli G, Mariani M, Vegni F, Maione D, Rinaudo D, Rappuoli R, Telford JL, Kasper DL, Grandi G, Fraser CM. 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. Proc Natl Acad Sci U S A 99:12391–12396.

183. van der Mee-Marquet N, Domelier A-S, Mereghetti L, Lanotte P, Rosenau A, van Leeuwen W, Quentin R. 2006. Prophagic DNA Fragments in Streptococcus agalactiae Strains and Association with Neonatal Meningitis. Journal of Clinical Microbiology 44:1049–1058.

184. Day WA, Fernández RE, Maurelli AT. 2001. Pathoadaptive mutations that enhance virulence: genetic organization of the cadA regions of Shigella spp. Infect Immun 69:7471–7480.

185. Domelier A-S, van der Mee-Marquet N, Grandet A, Mereghetti L, Rosenau A, Quentin R. 2006. Loss of catabolic function in Streptococcus agalactiae strains and its association with neonatal meningitis. J Clin Microbiol 44:3245–3250.

186. Pallen MJ, Wren BW. 2007. Bacterial pathogenomics. Nature 449:835–842.

187. Shelburne SA, Keith D, Horstmann N, Sumby P, Davenport MT, Graviss EA, Brennan RG, Musser JM. 2008. A direct link between carbohydrate utilization and virulence in the major human pathogen group A Streptococcus. Proc Natl Acad Sci U S A 105:1698–1703.

188. van der Mee-Marquet N, Diene SM, Barbera L, Courtier-Martinez L, Lafont L, Ouachée A, Valentin A-S, Santos SD, Quentin R, François P. 2018. Analysis of the prophages carried by human infecting isolates provides new insight into the evolution of Group B Streptococcus species. Clin Microbiol Infect 24:514–521.

189. Lichvariková A, Soltys K, Szemes T, Slobodnikova L, Bukovska G, Turna J, Drahovska H. 2020. Characterization of Clinical and Carrier Streptococcus agalactiae and Prophage Contribution to the Strain Variability. Viruses 12:1323.

190. Salloum M, van der Mee-Marquet N, Domelier A-S, Arnault L, Quentin R. 2010. Molecular Characterization and Prophage DNA Contents of Streptococcus agalactiae Strains Isolated from Adult Skin and Osteoarticular Infections. J Clin Microbiol 48:1261–1269.

191. Domelier A-S, van der Mee-Marquet N, Sizaret P-Y, Héry-Arnaud G, Lartigue M-F, Mereghetti L, Quentin R. 2009. Molecular Characterization and Lytic Activities of Streptococcus agalactiae Bacteriophages and Determination of Lysogenic-Strain Features. Journal of Bacteriology 191:4776–4785.

192. Rezaei Javan R, Ramos-Sevillano E, Akter A, Brown J, Brueggemann AB. 2019. Prophages and satellite prophages are widespread in Streptococcus and may play a role in pneumococcal pathogenesis. 1. Nat Commun 10:4852.

193. Shmidov E, Zander I, Lebenthal-Loinger I, Karako-Lampert S, Shoshani S, Banin E. 2021. An Efficient, Counter-Selection-Based Method for Prophage Curing in Pseudomonas aeruginosa Strains. Viruses 13:336.

194. Gabrielsen C, Mæland JA, Lyng RV, Radtke A, Afset JE. 2017. Molecular characteristics of Streptococcus agalactiae strains deficient in alpha-like protein encoding genes. J Med Microbiol 66:26–33.

195. Kwatra G, Adrian PV, Shiri T, Buchmann EJ, Cutland CL, Madhi SA. 2015. Natural acquired humoral immunity against serotype-specific group B Streptococcus rectovaginal colonization acquisition in pregnant women. Clinical Microbiology and Infection 21:568.e13-568.e21.

196. Effectiveness of intrapartum antibiotic prophylaxis for early-onset group B Streptococcal infection: An integrative review | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S1871519217301336?token=0255F68E13436531D872A4B2A17EE2BC79EE3C4A2D1567262FACF385B3568ED5F12BA824F0342FAA

81B001639672742E&originRegion=us-east-1&originCreation=20221205181842. Retrieved 5 December 2022.

197. Vornhagen J, Quach P, Boldenow E, Merillat S, Whidbey C, Ngo LY, Adams Waldorf KM, Rajagopal L. 2016. Bacterial Hyaluronidase Promotes Ascending GBS Infection and Preterm Birth. mBio 7:e00781-16.

198. Edmond KM, Kortsalioudaki C, Scott S, Schrag SJ, Zaidi AKM, Cousens S, Heath PT. 2012. Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis. Lancet 379:547–556.

199. Penders J, Thijs C, van den Brandt PA, Kummeling I, Snijders B, Stelma F, Adams H, van Ree R, Stobberingh EE. 2007. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. Gut 56:661–667.

200. Björkstén B. 2004. Effects of intestinal microflora and the environment on the development of asthma and allergy. Springer Semin Immunopathol 25:257–270.

201. Lundin A, Bok CM, Aronsson L, Björkholm B, Gustafsson J-A, Pott S, Arulampalam V, Hibberd M, Rafter J, Pettersson S. 2008. Gut flora, Toll-like receptors and nuclear receptors: a tripartite communication that tunes innate immunity in large intestine. Cell Microbiol 10:1093–1103.

202. Hayes K, O'Halloran F, Cotter L. 2020. A review of antibiotic resistance in Group B Streptococcus: the story so far. Crit Rev Microbiol 46:253–269.

203. Burcham LR, Spencer BL, Keeler LR, Runft DL, Patras KA, Neely MN, Doran KS. 2019. Determinants of Group B streptococcal virulence potential amongst vaginal clinical isolates from pregnant women. PLoS ONE 14:e0226699.

204. Shabayek S, Spellerberg B. 2018. Group B Streptococcal Colonization, Molecular Characteristics, and Epidemiology. Front Microbiol 9:437.

205. Mashburn-Warren L, Goodman SD, Federle MJ, Prehna G. 2018. The conserved mosaic prophage protein paratox inhibits the natural competence regulator ComR in Streptococcus. 1. Sci Rep 8:16535.

206. Andrews S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

207. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

208. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19:455–477.

209. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.

210. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

211. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

212. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403–410.

213. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM https://doi.org/10.6084/M9.FIGSHARE.963153.V1.

214. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

215. Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. 6:1287. F1000Research https://doi.org/10.12688/f1000research.12232.1.

216. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16-21.

217. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. 1. Nat Commun 9:5114.

218. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics 12:395.

219. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679.

220. Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451-454.

221. Hunter JD. 2007. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering 9:90–95.

222. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580.

223. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248.

224. Sievers F, Higgins DG. 2014. Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences, p. 105–116. *In* Russell, DJ (ed.), Multiple Sequence Alignment Methods. Humana Press, Totowa, NJ.

225. RStudio Team. 2020. RStudio:Integrated Development Environment for R. RStudio, PBC, Boston, Massachusetts.

226. McKinney W. 2010. Data Structures for Statistical Computing in Python, p. 56–61. *In* . Austin, Texas.

227. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. 2020. Array programming with NumPy. 7825. Nature 585:357–362.

228. Waskom M, Botvinnik O, O'Kane D, Hobson P, Lukauskas S, Gemperline DC, Augspurger T, Halchenko Y, Cole JB, Warmenhoven J, Ruiter JD, Pye C, Hoyer S, Vanderplas J, Villalba S, Kunter G, Quintero E, Bachant P, Martin M, Meyer K, Miles A, Ram Y, Yarkoni T, Williams ML, Evans C, Fitzgerald C, Brian, Fonnesbeck C, Lee A, Qalieh A. 2017. Mwaskom/Seaborn: V0.8.1 (September 2017). Zenodo.

229. Heath PT, Jardine LA. 2014. Neonatal infections: group B streptococcus. BMJ Clin Evid 2014:0323.

230. Luan S-L, Granlund M, Sellin M, Lagergård T, Spratt BG, Norgren M. 2005. Multilocus Sequence Typing of Swedish Invasive Group B Streptococcus Isolates Indicates a Neonatally Associated Genetic Lineage and Capsule Switching. Journal of Clinical Microbiology 43:3727–3733.

231. Campbell AM. 1992. Chromosomal insertion sites for phages and plasmids. J Bacteriol 174:7495–7499.

232. Ventura M, Canchaya C, Pridmore D, Berger B, Brüssow H. 2003. Integration and Distribution of Lactobacillus johnsonii Prophages. Journal of Bacteriology 185:4603–4608.

233. Coleman D, Knights J, Russell R, Shanley D, Birkbeck TH, Dougan G, Charles I. 1991. Insertional inactivation of the Staphylococcus aureusβ-toxin by bacteriophage φ13 occurs by site-and orientation-specific integration of the φ 13 genome. Molecular Microbiology 5:933–939.

234. Lee CY, Iandolo JJ. 1986. Lysogenic conversion of staphylococcal lipase is caused by insertion of the bacteriophage L54a genome into the lipase structural gene. Journal of Bacteriology 166:385–391.

235. Carey JN, Mettert EL, Fishman-Engel DR, Roggiani M, Kiley PJ, Goulian M. 2019. Phage integration alters the respiratory strategy of its host. eLife 8:e49081.

236. Rabinovich L, Sigal N, Borovok I, Nir-Paz R, Herskovits AA. 2012. Prophage Excision Activates Listeria Competence Genes that Promote Phagosomal Escape and Virulence. Cell 150:792–802.

237. Koyanagi S, Lévesque CM. 2013. Characterization of a Streptococcus mutans Intergenic Region Containing a Small Toxic Peptide and Its cis-Encoded Antisense Small RNA Antitoxin. PLoS One 8:e54291.

238. Makarova KS, Grishin NV, Koonin EV. 2006. The HicAB cassette, a putative novel, RNA-targeting toxin-antitoxin system in archaea and bacteria. Bioinformatics 22:2581–2584.

239. Lehnherr H, Maguin E, Jafri S, Yarmolinsky MB. 1993. Plasmid Addiction Genes of Bacteriophage P1: doc, which Causes Cell Death on Curing of Prophage, and phd, which Prevents Host Death when Prophage is Retained. Journal of Molecular Biology 233:414–428.

240. Yang QE, Walsh TR. 2017. Toxin–antitoxin systems and their role in disseminating and maintaining antimicrobial resistance. FEMS Microbiol Rev 41:343–353.

241. LeRoux M, Laub MT. 2022. Toxin-Antitoxin Systems as Phage Defense Elements. Annu Rev Microbiol 76:annurev-micro-020722-013730.

242. Aziz RK, Edwards RA, Taylor WW, Low DE, McGeer A, Kotb M. 2005. Mosaic Prophages with Horizontally Acquired Genes Account for the Emergence and Diversification of the Globally Disseminated M1T1 Clone of Streptococcus pyogenes. Journal of Bacteriology 187:3311–3318.

243. Rajesh T, Anthony T, Saranya S, Pushpam PL, Gunasekaran P. 2011. Functional characterization of a new holin-like antibacterial protein coding gene tmp1 from goat skin surface metagenome. Appl Microbiol Biotechnol 89:1061–1073.

244. Argov T, Sapir SR, Pasechnek A, Azulay G, Stadnyuk O, Rabinovich L, Sigal N, Borovok I, Herskovits AA. 2019. Coordination of cohabiting phage elements supports bacteria–phage cooperation. Nat Commun 10:5288.

245. Lemire S, Figueroa-Bossi N, Bossi L. 2011. Bacteriophage Crosstalk: Coordination of Prophage Induction by Trans-Acting Antirepressors. PLoS Genet 7:e1002149.

246. Cushman J, Freeman E, McCallister S, Schumann A, Hutchison KW, Molloy SD. 2021. Increased whiB7 expression and antibiotic resistance in Mycobacterium chelonae carrying two prophages. BMC Microbiol 21:176.

247. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL. 2016. Prophages mediate defense against phage infection through diverse mechanisms. ISME J 10:2854–2866.

248. Stubbendieck RM, Straight PD. 2016. Multifaceted Interfaces of Bacterial Competition. J Bacteriol 198:2145–2155.

249. Flemming H-C, Wingender J. 2010. The biofilm matrix. 9. Nat Rev Microbiol 8:623–633.

250. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. 2010. Bacterial competition: surviving and thriving in the microbial jungle. 1. Nat Rev Microbiol 8:15–25.

251. The Ecology and Evolution of Microbial Competition | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S0966842X16300749?token=F7E933DE6718E3C9E802654570721A577DFC1B8F21B05E97DF1ABA1A0E5230387BCFA729929F3DEBDD0C2684410F73AE&originRegion=us-east-1&originCreation=20230312172843. Retrieved 12 March 2023.

252. Cornforth DM, Foster KR. 2013. Competition sensing: the social side of bacterial stress responses. 4. Nat Rev Microbiol 11:285–293.

253. Pekkonen M, Ketola T, Laakso JT. 2013. Resource Availability and Competition Shape the Evolution of Survival and Growth Ability in a Bacterial Community. PLOS ONE 8:e76471.

254. Granato ET, Meiller-Legrand TA, Foster KR. 2019. The Evolution and Ecology of Bacterial Warfare. Current Biology 29:R521–R537.

255. Seed KD, Yen M, Shapiro BJ, Hilaire IJ, Charles RC, Teng JE, Ivers LC, Boncy J, Harris JB, Camilli A. 2014. Evolutionary consequences of intra-patient phage predation on microbial populations. eLife 3:e03497.

256. Bobay L-M, Touchon M, Rocha EPC. 2014. Pervasive domestication of defective prophages by bacteria. Proceedings of the National Academy of Sciences 111:12127–12132.

257. Li G, Cortez MH, Dushoff J, Weitz JS. 2020. When to be temperate: on the fitness benefits of lysis vs. lysogeny. Virus Evolution 6:veaa042.

258. Joo J, Gunny M, Cases M, Hudson P, Albert R, Harvill E. 2006. Bacteriophage-mediated competition in Bordetella bacteria. Proc Biol Sci 273:1843–1848.

259. Haaber J, Leisner JJ, Cohn MT, Catalan-Moreno A, Nielsen JB, Westh H, Penadés JR, Ingmer H. 2016. Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. 1. Nat Commun 7:13333.

260. Hanson BR, Runft DL, Streeter C, Kumar A, Carion TW, Neely MN. 2012. Functional Analysis of the CpsA Protein of Streptococcus agalactiae. J Bacteriol 194:1668–1678.

261. Dammann AN, Chamby AB, Catomeris AJ, Davidson KM, Tettelin H, Pijkeren J-P van, Gopalakrishna KP, Keith MF, Elder JL, Ratner AJ, Hooven TA. 2021. Genome-Wide fitness analysis of group B Streptococcus in human amniotic fluid reveals a transcription factor that controls multiple virulence traits. PLOS Pathogens 17:e1009116.

262. Kirby-Bauer-Disk-Diffusion-Susceptibility-Test-Protocol-pdf.pdf.

263. Rajaei A, Rowe HM, Neely MN. 2022. The LCP Family Protein, Psr, Is Required for Cell Wall Integrity and Virulence in Streptococcus agalactiae. Microorganisms 10:217.

264. Hooven TA, Catomeris AJ, Bonakdar M, Tallon LJ, Santana-Cruz I, Ott S, Daugherty SC, Tettelin H, Ratner AJ. 2017. The Streptococcus agalactiae Stringent Response Enhances Virulence and Persistence in Human Blood. Infect Immun 86:e00612-17.

265. Patras KA, Derieux J, Al-Bassam MM, Adiletta N, Vrbanac A, Lapek JD, Zengler K, Gonzalez DJ, Nizet V. 2018. Group B Streptococcus Biofilm Regulatory Protein A Contributes to Bacterial Physiology and Innate Immune Resistance. The Journal of Infectious Diseases 218:1641–1652.

266. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

267. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative Genomics Viewer. Nat Biotechnol 29:24–26.

268. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. 5. Nat Biotechnol 34:525–527.

269. Yadav AK, Espaillat A, Cava F. 2018. Bacterial Strategies to Preserve Cell Wall Integrity Against Environmental Threats. Front Microbiol 9:2064.

270. Nanda AM, Thormann K, Frunzke J. 2015. Impact of Spontaneous Prophage Induction on the Fitness of Bacterial Populations and Host-Microbe Interactions. J Bacteriol 197:410–419.

271. Casjens S. 2003. Prophages and bacterial genomics: what have we learned so far? Molecular Microbiology 49:277–300.

272. Wagner PL, Waldor MK. 2002. Bacteriophage Control of Bacterial Virulence. Infect Immun 70:3985–3993.

273. Phelps HA, Neely MN. 2005. Evolution of the Zebrafish Model: From Development to Immunity and Infectious Disease. Zebrafish 2:87–103.

274. Neely MN. 2017. The zebrafish as a model for human bacterial infections. Bacterial Pathogenesis: Methods and Protocols 245–266.

275. Neely MN, Pfeifer JD, Caparon M. 2002. Streptococcus-Zebrafish Model of Bacterial Pathogenesis. Infection and Immunity 70:3904–3914.

276. Alexeeva S, Guerra Martínez JA, Spus M, Smid EJ. 2018. Spontaneously induced prophages are abundant in a naturally evolved bacterial starter culture and deliver competitive advantage to the host. BMC Microbiology 18:120.

277. Srinivasan R, Kannappan A, Shi C, Lin X. 2021. Marine Bacterial Secondary Metabolites: A Treasure House for Structurally Unique and Effective Antimicrobial Compounds. Mar Drugs 19:530.

278. León-Félix J, Villicaña C. 2021. The Impact of Quorum Sensing on the Modulation of Phage-Host Interactions. J Bacteriol 203:e00687-20.

279. Wall D. 2016. Kin recognition in bacteria. Annu Rev Microbiol 70:143–160.

280. Abisado RG, Benomar S, Klaus JR, Dandekar AA, Chandler JR. 2018. Bacterial Quorum Sensing and Microbial Community Interactions. mBio 9:e02331-17.

281. Davies EV, Winstanley C, Fothergill JL, James CE. 2016. The role of temperate bacteriophages in bacterial infection. FEMS Microbiology Letters 363:fnw015.

282. Wagner PL, Neely MN, Zhang X, Acheson DWK, Waldor MK, Friedman DI. 2001. Role for a Phage Promoter in Shiga Toxin 2 Expression from a Pathogenic Escherichia coli Strain. J Bacteriol 183:2081–2085.

283. Euler CW, Juncosa B, Ryan PA, Deutsch DR, McShan WM, Fischetti VA. 2016. Targeted Curing of All Lysogenic Bacteriophage from Streptococcus pyogenes Using a Novel Counter-selection Technique. PLoS One 11:e0146408.

284. Tran PM, Feiss M, Kinney KJ, Salgado-Pabón W. 2019. ϕSa3mw Prophage as a Molecular Regulatory Switch of Staphylococcus aureus β-Toxin Production. Journal of Bacteriology 201:e00766-18.

285. Ainsworth S, Zomer A, de Jager V, Bottacini F, van Hijum SAFT, Mahony J, van Sinderen D. 2013. Complete Genome of Lactococcus lactis subsp. cremoris UC509.9, Host for a Model Lactococcal P335 Bacteriophage. Genome Announc 1:e00119-12.

286. Sekulovic O, Fortier L-C. 2015. Global Transcriptional Response of Clostridium difficile Carrying the ϕCD38-2 Prophage. Appl Environ Microbiol 81:1364–1374.

287. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544.

288. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:550.

289. Baerends RJ, Smits WK, de Jong A, Hamoen LW, Kok J, Kuipers OP. 2004. Genome2D: a visualization tool for the rapid analysis of bacterial transcriptome data. Genome Biology 5:R37.

290. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research 13:2498.

291. Lempp M, Lubrano P, Bange G, Link H. 2020. Metabolism of non-growing bacteria. Biological Chemistry 401:1479–1485.

292. Seegers JFML, Mc Grath S, O'Connell-Motherway M, Arendt EK, van de Guchte M, Creaven M, Fitzgerald GF, van Sinderen D. 2004. Molecular and transcriptional analysis of the temperate lactococcal bacteriophage Tuc2009. Virology 329:40–52.

293. Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. 2005. Switches in Bacteriophage Lambda Development. Annual Review of Genetics 39:409–429.

294. Cuthbertson L, Kimber MS, Whitfield C. 2007. Substrate binding by a bacterial ABC transporter involved in polysaccharide export. Proceedings of the National Academy of Sciences 104:19529–19534.

295. Lee Y, Song S, Sheng L, Zhu L, Kim J-S, Wood TK. 2018. Substrate Binding Protein DppA1 of ABC Transporter DppBCDF Increases Biofilm Formation in Pseudomonas aeruginosa by Inhibiting Pf5 Prophage Lysis. Front Microbiol 9:30.

296. Phage Therapy: Beyond Antibacterial Action - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5974148/. Retrieved 29 March 2023.

BIOGRAPHY OF THE AUTHOR

Caitlin Wiafe-Kwakye was born and raised in Accra, Ghana on May 12, 1988. She graduated from Wesley Girls' High School, Cape Coast, Ghana in 2006. She received a Bachelor of Science degree in Biochemistry from the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana in 2011. After her degree, she worked in the Ablordey lab at Noguchi Memorial Institute for Medical Research. She received a Master of Philosophy in Medical Microbiology from the University of Ghana in 2016. Caitlin is a candidate for the Doctor of Philosophy degree in Microbiology from the University of Maine in May 2023.