Pedro Miguel Viegas dos Santos

Identification and Characterization of prophages in *Bacteroides* spp. isolates from children with Type 1 Diabetes

Faculdade de Ciências e Tecnologia

2019

Pedro Miguel Viegas dos Santos

Identification and Characterization of prophages in *Bacteroides* spp. isolates from children with Type 1 Diabetes

Mestrado em Biologia Molecular e Microbiana

Trabalho efetuado sob a orientação de:

Maria Leonor Faleiro e Isabel Duarte

Faculdade de Ciências e Tecnologia

2019

Identification and Characterization of prophages in *Bacteroides* spp. isolates from children with Type 1 Diabetes

Authorship Statement

I hereby declare to be the author of this work, which is original and unpublished. Authors and papers consulted are duly cited in the text and are listed in the included references.

“Human subtlety will never devise an invention more beautiful, more simple or more direct than does nature because in her inventions nothing is lacking, and nothing is superfluous.”

- **Leonardo da Vinci**

Acknowledgments

I have to thank two great persons for their tireless support and help provided in all the tasks of the work for this thesis, firstly my advisor **Professora Doutora Maria Leonor Faleiro** of the **Universidade do Algarve**. **Prof. Faleiro** always trusted me to fulfill this important work, without her help and guidance I would never be able to reach the end. Dispensing her precious time to help and guide me through the hardships in the laboratory and in the thesis writing. To my co-advisor **Doutora Isabel Duarte** of the **Universidade do Algarve**, thank you so much for the constant worry and great help in the tasks that I would not fulfill without you. The attention and the work applied by you in this thesis is something that I truly appreciate. I learn a lot with these two great teachers at a personal and professional level.

I would also like to acknowledge **Mestre José Matos**, my partner and friend in the laboratory for all the help to accomplish this work. Without your help and knowledge, accompanied by your experience, I don't think I would have done all of the laboratory work. We worked as a great team, at a professional and human level, sharing great laughs between the work. A great colleague and friend that I would work in my future professional life. To the laboratory technician **Liseta Viegas** who help me always when I needed with a smile and friendly words.

Also for my colleague and great friend **Bruno Várzea** for helping me throughout this journey, sharing knowledge, insight and a great deal of laughs in the lab. We supported each other as great friends do in times of need. A great professional that I would work with in any lab in the world.

Express the profound gratitude to my loving parents that always supported me through these years of study and provide me with the tools to become a man. A special thank you to my loving **Mother** who always battled to see me happy, teaching me that life is short and we due to ourselves to battle the storms in it. Thank you, mom, for the love and trust that you gave me throughout my life. I will live with the great values that you taught me. To my loving **Father** who worked all is life to provide for his family through times of need, thank you for showing me what a great man is, teaching me to never give up no matter what is the incoming problem. To my **Sister, Brother in law, Nephews** and to my **Grandparents** for their support and love that you always provided.

Finally, to my two great colleagues and professionals **Isabel Matos** and **Ana Lopes** for the great times we shared together in the lab. I look forward one day to work also with both. Also, to my lifelong friends that always pushed me to be better.

Thank you

Abstract

Identification and characterization of bacteriophages (phages) in gut microbiome (phageome) has been through the years a difficult task. Bacteriophages display the tools to modulate the bacterial populations in the gut through their abundance and diversity ultimately resulting in dysbiosis. The trigger that starts the metabolic cascade that leads to the onset of type 1 diabetes (T1D) may come via an ample source of biological identities, one of which can be the viral patterns in the bacterial communities. A tool that greatly facilitates the detection of prophages in bacterial genomes is the PHAge Search Tool Enhanced Release (PHASTER) that was used in this study to analyze the prophage patterns of 7 *Bacteroides dorei* strains and 1 *Parabacteroides distasonis* strain that were isolated from children with T1D and healthy Controls. In the current study 9 incomplete prophages (cryptic prophages) were identified and characterized in 7 *B. dorei* strains and in the *P. distasonis* strain. Prophage-like regions in *Bacteroides* isolates from T1D children carrying genes potentially impacting or triggering the autoimmunity attack of β-cells to the onset of T1D were identified. A coding gene for a DEAD/DEAH box helicase in *B. dorei* D8M1 genome and a DNA adenine methylase (*dam*) gene encountered in *B. dorei* D16P1 and D16M14 strains. To establish if stress conditions could trigger the lytic cycle, particularly the exposure of the bacterial strains to mitomycin C, ox-bile and hydrogen peroxide was evaluated. No phage induction was observed under the exposure to the tested inducers. The role of the intestinal virome in the development of T1D is scarce to provide answers for the autoimmune attack of β-cells, but the identification of the *dam* gene in *B. dorei* strains isolated from T1D children evidences that it will be important to dissect its role on the bacterial physiology and its impact on the seroconversion.

Keywords: Type 1 Diabetes *mellitus*; *Bacteroides* spp; Bacteriophages; DNA methylation

Resumo

A identificação e caracterização de bacteriófagos (fagos) no microbioma intestinal foi ao longo dos anos uma tarefa difícil, superada apenas pelos avanços em novas metodologias para o estudo de vírus, como sequenciação de próxima geração, metagenómica, proteómica e bioinformática. A abundância de bacteriófagos no intestino humano pode atingir números elevados, podendo ser considerada a maior concentração de uma entidade biológica no planeta, representando um dos maiores conjuntos genéticos não explorados no nosso planeta. A ferramenta bioinformática PHAge Search Tool Enhanced Release (PHASTER) facilita a deteção de fagos nos genomas bacterianos bem como permite uma rápida identificação e anotação de sequências de fagos nos genomas e plasmídeos bacterianos. Esta ferramenta fornece informações detalhadas sobre a localização genómica dos elementos do fago, a integridade do prófago previsto e suas proteínas componentes, permitindo uma fácil identificação e caracterização do prófago. A doença Diabetes mellitus tipo 1 (DT1), inicialmente designada por diabetes juvenil, é uma doença crónica imunomediada caracterizada pelo ataque autoimune de células β pancreáticas nas ilhotas de Langerhans. A destruição das células β causa hiperglicemia grave, levando a alta morbidade, e se não for diagnosticada em tempo útil, pode resultar numa morte prematura. Geralmente, esta doença é mais comum nos primeiros anos de vida, com metade dos casos em crianças com idade inferior a 15 anos. A prevalência desta doença na população mundial é de 1%, com estudos recentes apontando para um novo início de DT1 em pacientes com mais de 20 anos. A progressão inicial de DT1 pode ser definida em três etapas: aparecimento de autoanticorpos nas células β (soroconversão) (estágio 1); seguido pela perda de células β e hiperglicemia (estágio 2); e finalmente o estágio sintomático (estágio 3). Em crianças finlandesas com alto risco genético para desenvolver DT1, um aumento na população de *B. dorei* foi detetado 8 meses antes da seroconversão para o primeiro autoanticorpo de ilhota, sugerindo que mudanças precoces na microbiota bacteriana intestinal podem ser exploradas para a autoimunidade prevista para DT1. Os bacteriófagos apresentam meios para manipular o ambiente ecológico intestinal através de vários mecanismos que afetam a saúde do hospedeiro pelo ciclo fago-bacteriano-fago. Os fagos ajudam a manter o bacterioma intestinal normal através de sua simbiose estável, usando diferentes ferramentas, como transferência horizontal de genes, lisogenia e controle da competição, mas também podem alterar a homeostase bacteriana levando a uma disbiose.

O iniciador da cascata metabólica para o ataque autoimune às células β pode vir de uma ampla fonte de identidades biológicas, onde as unidades virais nas comunidades bacterianas podem ser um deles. O principal objetivo do presente estudo foi elucidar os padrões de prófagos nos genomas de estirpes de *B. dorei* isoladas de crianças com DT1 e crianças saudáveis da região do Algarve. Outro objetivo foi analisar se a exposição a condições de stress poderiam desencadear o ciclo lítico do prófago, particularmente a exposição ao antibiótico mitomicina C, ox-bile e peróxido de hidrogénio (H_2O_2). Identificamos e caracterizamos 9 prófagos incompletos (profagos crípticos) em 7 estirpes de *B. dorei* e 1 estirpe de *Parabacteroides distasonis* isolados de crianças com DT1 e crianças controlo. A partir dos elementos fágicos identificados, 4 deles nunca foram identificados em bactérias do Filo *Bacteroidetes* de acordo com o Virus-Host Database, fornecendo uma nova visão para estes fagos (*Mannheimia* phage vB_MhM_3927AP2; *Clostridium* phage c-st; *Bacillus* phage phiAGATE; *Synechococcus* phage S-SKS1). Os genomas bacterianos de crianças com DT1 apresentam em média duas regiões fágicas em comparação com os genomas bacterianos das crianças controlo com apenas uma em média, denotando alguma permeabilidade à infeção fágica para aquisição de genes vantajosos ao metabolismo ou fisiologia do hospedeiro. Além disso, a percentagem de similaridade de proteínas fágicas foi maior para os genomas bacterianos de DT1, chegando a 41,17% contra o máximo de 14,28% para genomas bacterianos das crianças controlo com um valor médio de 19,9% e 8,9%, respetivamente. As regiões fágicas identificadas nos genomas bacterianos de DT1 mostraram um tamanho médio de 19,02 kb contra 9,13 kb dos genomas bacterianos de crianças controlo. Algumas das regiões virais identificadas nos genomas bacterianos de crianças com DT1 e controlos evidenciaram similaridade fágica, entre si, sugerindo algumas características comuns. Localizações fágicas em isolados de *Bacteroides* de crianças com DT1 apresentam genes potencialmente impactantes no desencadear do ataque autoimune às células β até ao estabelecimento da doença, em comparação com as regiões de isolados de *B. dorei* de controlos que não apresentaram nenhum gene com esse atributo.

O genoma da estirpe *B. dorei* D8M1 apresentou o maior número de regiões, com um gene que codifica para uma DEAD / DEAH box helicase, com uma possível ação na autoimunidade e possível estímulo na inflamação, mas requer ainda clarificação, nomeadamente os seus reguladores e elementos associados na interação. Outro gene importante que pode influenciar o ataque

autoimune das células β é a DNA adenina metilase (Dam), que tem a capacidade de metilar a posição 6 do azoto na adenina para gerar N-6-metiladenina (m6c) nos seus locais de reconhecimento (repetições palindrómicas de 5 'GATC-3') regulando várias funções no hospedeiro do prófago. O gene *dam* foi identificado no genoma das estirpes *B. dorei* D16P1 e D16M14, evidenciando a mesma similaridade para o mesmo prófago. Um gene órfão de *dam* numa região fágica de uma estirpe de *B. dorei* isolada de uma criança finlandesa com alto risco genético para desenvolver DT1, mostrou um genoma fortemente metilado nos locais de reconhecimento da Dam antes da seroconversão evoluindo para autoimunidade contra células β após 1,5 meses. Em contraste, outra estirpe de *B. dorei* sem o gene *dam*, também de uma criança finlandesa com o mesmo risco genético para DT1, não revelou metilações no seu genoma nem progrediu para autoimunidade contra células β. Até ao momento, não está claro qual é o impacto deste gene no desenvolvimento de DT1. A indução do ciclo lítico pela exposição a mit C, H_2O_2 e ox-bile não foi observada em qualquer uma das estirpes testadas.

A identificação no presente estudo do gene *dam* em regiões idênticas em duas estirpes de *B. dorei* isoladas do mesmo paciente fornece evidências de que os bacteriófagos podem capacitar as estirpes de *B. dorei* com perfis diferenciados, particularmente através da metilação de genes que influenciam a expressão génica através da regulação positiva ou negativa (por exemplo, lisogenia) com a metilação Dam do motivo 5´-GATC-3´, o que poderá explicar, em parte a sua dominância antes da seroconversão em crianças finlandesas e provavelmente em crianças portuguesas. Uma vez que a presença do gene *dam* em estirpes de *B. dorei* isoladas de crianças com DT1 tem sido reportada noutros estudos é necessário analisar o seu papel na fisiologia bacteriana e qual será o seu impacto na seroconversão. As estirpes de *B. dorei* e *P. distasonis* de crianças com DT1 e crianças Controlo do Algarve não apresentaram regiões com prófagos completos nos seus genomas, o que explica a ausência de indução de fagos pelos indutores testados. O papel do viroma intestinal no desenvolvimento de T1D é escasso para fornecer respostas para o ataque autoimune das células β, mas a identificação do gene *dam* em estirpes de *B. dorei* isoladas de crianças com DT1 evidencia a importância da avaliação do impacto deste gene na fisiologia bacteriana e a sua participação na seroconversão.

Palavras chave: Diabetes *mellitus* tipo 1; *Bacteroides* spp; Bacteriofagos; Metilação do ADN

Contents

List of Figures

List of Tables

Glossary

Abi - Abortive infection

BHI - Brain heart infusion medium

BHI+H - Brain heart infusion plus hemin medium

BSH - Bile salt hydrolases

CD - Crohn´s Disease

CRISPR - Clustered regulatory interspaced short palindromic repeats

crRNA - CRISPR RNA

Dam - DNA adenine methyltransferase

DNA - Deoxyribonucleic acid

GTAs - Gene transfer agents

HGT - Horizontal gene transfer

IBD - Intestinal Bowel Disease

LY-BHI - Liquid yeast brain heart infusion medium

Mit C - Mitomycin C

MTase - Methyltransferase

ORF - Open reading frame

PBS - Phosphate Buffered Saline

PHASTER - PHAge Search Tool Enhanced Release

RBP - Receptor-binding protein

REase - Restriction endonuclease

Res - Restriction sites

R-M - Restriction-modification

RNA - Ribonucleic acid

SCFAs - Short-chain fatty acids

Sie - Superinfection exclusion

T1D - Type 1 diabetes *mellitus*

TAE - Tris-Acetate-EDTA buffer

tRNA - transfer RNA

VLPs - Viral-like particles

1 Introduction

1.1 Human gut microbiota

The human gut allocates a large variety of microorganisms, namely bacteria, archaea, fungi (mainly yeasts) and viruses (including bacteriophages). This group of microorganisms is designated as gut microbiota; the collective genes of this microbial community is called the microbiome (Turnbaugh *et al.*, 2007). Several studies have shown that different individuals can differ considerably from one another regarding their microbiomes (Zoetendal, Vaughan and De Vos, 2006; Turnbaugh *et al.*, 2007; Qin *et al.*, 2010; Huttenhower *et al.*, 2012) but over their lifetime this microbial community can be stable (Costello *et al.*, 2009). Both the density and the microbial diversity in the gut increases from the stomach towards the colon (due to the increased availabity of nutrients). The recognized functions of the gut microbiota comprehend the transformation of ingested food and the production of essential vitamins and cofactors (Pereira and Berry, 2017).

The different intestinal anatomic regions, varying in physiology, pH, oxygen tension, digestive flow, substrate availability and host secretions, influence the organization and the distribution of the gut microbiota in biochemical niches in the gastrointestinal system (Flint *et al.*, 2012). Taxonomy and functionality of gut microbiota can vary from one individual to another according to infant gut population development, age, environmental factors, antibiotic usage, transit times, abundance of proteolytic enzymes, bile, diet and disease (Ridlon *et al.*, 2014; Rinninella *et al.*, 2019). These factors affect all microorganisms inhabiting the gut microbiota, including viruses (Minot, Bryson, Chehoud, Gary D Wu, *et al.*, 2013; Lim *et al.*, 2015).

Bacterial composition in the gut of healthy humans may vary from one individual to another, primarily by temporal and spatial differences in combination with other factors that modulate bacterial composition, namely diet, antibiotic, bile acids and disease (Thursby and Juge, 2017).

When a homeostasis state is achieved from stabilization of the selective pressures imposed by the host and microbial competitors, a normal taxonomic constitution can be identified with 5 phyla ruling the bacterial gut microbiota, namely the Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria and Fusobacteria, ordered from highest to lowest abundance (Arumugam *et al.*, 2011). Firmicutes and Bacteroidetes are the two most abundant phyla in human healthy gut microbiota

with constant interactions and variations between them (representing 90% of the bacterial gut microbiota), followed by Actinobacteria, Proteobacteria, Fusobacteria and Verrucomicrobia present in lower numbers (Arumugam *et al.*, 2011).

Firmicutes include more than 200 genera, being the *Clostridium* genus (95%) the most representative. *Bacteroides* and *Prevotella* are the two dominant genera for Bacteroidetes phyla, with the third most abundant phyla Actinobacteria represented mainly by *Bifidobacterium* genus (Arumugam *et al.*, 2011).

The most abundant Archaea genera identified in gut belongs to *Methanobrevibacter* spp. (a methanogenic archaea), while the predominant Fungi phyla Ascomycota is represented by the genera *Saccharomyces*, *Candida* and *Cladosporium* (Hoffmann *et al.*, 2013).

Regarding the gut virome, that constitutes a significant part of the gut microbiome, to date only 1% of viral content is known (Virgin, 2014).

Predominant eukaryotic viruses identified in gut microbiome belong to the Picobirnaviridae (ssRNA), Adenoviridae (dsDNA), Anelloviridae (ssDNA) and Astroviridae (ssRNA) family. Bacteriophage populations are mainly dominated by the *Microviridae* family (ssDNA), followed by *Podoviridae* (dsDNA, order Caudovirales), *Myoviridae* (dsDNA, order Caudovirales), and *Siphoviridae* (dsDNA, order Caudovirales) families in the adult healthy gut (Minot *et al.*, 2013; Mukhopadhya *et al.*, 2019).

The development of bacterial intestinal microbiota starts early in life, varying in three principle stages: from birth to weaning, from weaning to a stabilized diet, and during old age (Flint *et al.*, 2012).

The first acquisition is reported in the neonatal period, were bacterial groups are acquired *in utero* from amniotic fluid and placenta, as seen in bacterial composition of the meconium (earliest stool of a mammalian infant) resembling maternal placenta composition (Dong *et al.*, 2015; Argenio, 2018), with connection to health and immune imprinting still not fully clarified. After birth several factors guide the development of the bacterial content, such as mode of delivery (*e.g.*, Actinobacteria dominates the microbiota of vaginally delivered infants after 1 week, in contrast to infants born by cesarean section for which Firmicutes is the most prevalent bacterial phylum (Bokulich *et al.*, 2016; Hill *et al.*, 2017)), and feeding mode (Sullivan, Farver and Smilowitz, 2015). Other factors, such as gestational age of birth, host genetics, geographical location, use of

antibiotics, maternal diet and weaning duration (transition to solid food) can influence the microbiota development (Zhuang *et al.*, 2019).

The first bacterial groups to colonize the gastrointestinal tract (GI) of infants are facultative anaerobes followed by obligate anaerobes (such as *Bifidobacterium* and *Bacteroides* spp.) after 2 weeks from birth (Eggesbø *et al.*, 2011). During the first year of life, the microbial diversity increases and the gut microbiota starts to resemble adult-like composition with profile variations for each infant (Palmer *et al.*, 2007). At 2-3 years of age the infant microbiota shows similar diversity and functionality to that of an adult (Rodrı *et al.*, 2015).

The patterns of infant intestinal microbiota colonization are very distinct and determined by several factors, including geographical localization (Fallani *et al.*, 2010). Phage predation is another factor than can affect the structure and dynamics of the gut bacteriome in early age. (Milani *et al.*, 2017). It is well established that the bacteriome is established at an early age and changes over time due to diverse factors until its "adult-like" composition. The acquisition of viral content in the gut starts postnatally, this was confirmed through the analysis of the meconium where no viral-like particles (VLPs) were encountered in comparison to 1 week after birth when the VLP numbers reached 10^8/g in faeces (Breitbart *et al.*, 2008). So the virome is gradually acquired after birth, as seen in the study conducted by Lim *et al.* where various eukaryotic viruses were identified from birth to 2 years of age, suggesting acquisition through environmental exposure (Lim *et al.*, 2015).

Bacteriophage load is abundant at early ages being dominated by diverse *Caudovirales* phages (*Siphoviridae*, *Inoviridae*, *Myoviridae* and *Podoviridae* families) and the *Microviridae* family with marked shifts to the domination of the last family after 24 months of age throughout adult life (Minot, Bryson, Chehoud, Gary D Wu, *et al.*, 2013; Lim *et al.*, 2015).

Bacteria and bacteriophage richness are inversely correlated in an age-dependent manner: in the first days after birth a high bacteriophage-low bacteria richness is detected, shifting after 24 months to a low bacteriophage-high bacteria community, demonstrating possible reversed predator-prey cycle (i.e, a predator peak prior to a prey peak) (Cortez and Weitz, 2014).

Substantial differences in the gut virome between adult twins against their co-twins or mother reflects environmental factors driving virome community, structure and possible function (Mukhopadhya *et al.*, 2019). The virome can be strongly influenced by diet, since unrelated individuals through equal diet regime showed gradual virome convergence in community structure (Reyes *et al.*, 2010, 2015).

Concluding that bacterial richness in the gut microbiota, host lifestyle and genetic predisposition will define the virome (Reyes *et al.*, 2015; Mirzaei and Maurice, 2017).

1.1.1 Functionality

A direct impact of the symbiotic relationship between the gut bacterial populations in human health is shown by the essential degradation of substrates, metabolite production and defence against invasion by pathogenic microorganisms (Kosiewicz, Zirnheld and Alard, 2011).

Important functions provided by gut bacteria to the human host include: fermentation of complex carbohydrates for degradation in short-chain fatty acids (SCFAs), providing the energy from undigestible polysaccharides (as plant cell walls and starch) (Koh *et al.*, 2016), vitamin synthesis, and biotransformation of bile acids (Urdaneta and Casadesús, 2017; Rowland *et al.*, 2018).

The SCFAs most abundant in faeces are acetate, propionate and butyrate. From the three, butyrate is the most important as the energy source for human colon epithelial cells (colonocytes). This process starts with the fermentation of carbohydrates (*e.g.* starch) by obligate anaerobes, the majority of which belonging to the Firmicutes taxon (e.g. *Faecalibacterium prausnitzii and Eubacterium rectale* (presently *Agathobacter rectale*), the two most abundant in healthy colon populations) (Walker *et al.*, 2011; Flint *et al.*, 2012; Rowland *et al.*, 2018). Bacteria that provide this process help maintain the colonic epithelium in a C2-skewed metabolic state (this metabolic state is characterized by the high consumption of oxygen) contributing to energy production and limiting oxygen levels in the lumen. Unbalance on the butyrate producing bacterial community (dysbiosis) (e.g. antibiotics) that can lead to more severe dysbiosis, shifting colonic epithelium to a C1-skewed metabolism (characterized by high production of lactate, low oxygen consumption and iNOS enzyme), iNOS generates nitric oxide that in the lumen is converted in nitrate (NO_3^-) and oxygen allowing the expansion of facultative anaerobes leading to possible ulcerative colitis, evidenced by the presence of high lactate levels (Barbetti and Magliocca, 1988; Litvak, Byndloss and Avenue, 2018).

Acetate is the most abundant SCFA in the gut and functions as a major co-factor/metabolite for the growth of different bacteria (*e.g. Faecalibacterium prausnitzii*) contributing to butyrate production (Duncan *et al.*, 2004). Propionate is also an energy source for colonocytes, displaying a role in liver

gluconeogenesis and its production is mainly due to *Bacteroides* species, Negativicutes and some *Clostridium* species (Rowland *et al.*, 2018).

The gut microbiota contributes to dietary intake of vitamins, notably vitamin K and B groups (*e.g.*, biotin (B7), cobalamin (B12), folate (B9), niacin (B3), panthotenic acid (B5), pyridoxine (B6), riboflavin (B2), and thiamine (B1)). Over a quarter of vitamins of bacterial origin participate in dietary intake, mainly four B vitamins (B6, B9, B12, and B3). Bacteria that belong to Bacteroidetes, Fusobacteria and Proteobacteria phyla possess the metabolic pathways for B2 and B7 production, with Firmicutes and Actinobacteria having a smaller quantity of members that participate in the biosynthesis of vitamin B, being the vitamin B2 and B3 the most commonly synthesized. Bacteroidetes are the predicted phyla with the majority of vitamin B biosynthesis (90%), excluding vitamin B12 that is predominantly produced by Fusobacteria phyla (Magnúsdóttir *et al.*, 2015). The lack of capability of the human body to produce these vitamins determine the crucial role of bacteria in nutrition of the human host.

Other functions that are provided by bacterial populations of the gut include the biotransformation of bile salts in the colon. Bile acid synthesis starts in the pericentral hepatocytes from cholesterol to form cholic acid and chenodeoxycholic acid. The last are transformed into bile salts by the conjugation of taurine or glycine and association of Na^+ or K^+ ions, producing a molecule with detergent properties (Russell, 2003; Urdaneta and Casadesús, 2017). Bile salts exist in four types: primary and secondary, conjugated and non-conjugated (Urdaneta and Casadesús, 2017; Rowland *et al.*, 2018). Primary bile salts are produced directly from cholesterol degradation, the most synthesized ones in the human body are cholate and chenodeoxycholate. The follow up of their synthesis is the secretion into the small intestine facilitating lipid digestion and absorption (Russell, 2003). Bile salts that are not recycled back to the liver by absorption on the distal ileum (1-5%) during enterohepatic circulation enter the colon, where bacterial populations belonging to Clostridia convert primary bile salts to secondary through the removal of the hydroxyl group at carbon 7 with their 7α/β-dehydroxylation activity (Wells *et al.*, 2003; Urdaneta and Casadesús, 2017; Rowland *et al.*, 2018). As the result of this activity the predominant secondary bile salts in human faeces are deoxycholate and lithocholate (Ridlon, Kang and Hylemon, 2006).

Deconjugation of bile salts is another important task provided by bacterial populations of the colon, catalyzed by bile salt hydrolases (BSH) that hydrolyse the amid bond between the conjugated glycine or taurine (Urdaneta and Casadesús, 2017; Rowland *et al.*, 2018). BSH genes are

widespread in the bacterial populations of the gut, namely in several genera, such as *Bacteroides*, *Bifidobacterium*, *Clostridium*, *Enteroccocus*, *Lactobacillus* and *Listeria*, and also in the Archaea domain that includes *Methanobrevibacter smithii* and *Methanosphera stadmanae* (Rossocha *et al.*, 2005; Jones *et al.*, 2008).

Both processes of deconjugation and 7α-dehydroxylation facilitates the recovery of bile salts by passive transport in the colon epithelium back to the enterohepatic circulation, regulating the levels of bile salt synthesis of primary and secondary bile salts (Sayin *et al.*, 2013). This attenuates the antimicrobial activity associated with bile salts that exert selective pressures on structural levels in the gut microbiota, such as disruption of bacterial cell membranes by their detergent activity, induced DNA damage, and protein conformational changes, demonstrating the bidirectional relation between gut microbiota and bile salts (Urdaneta and Casadesús, 2017; Rowland *et al.*, 2018).

1.2 The human gut phageome

Identification and characterization of bacteriophages (phages) in the gut microbiome (phageome) has been through the years a difficult task, only surpassed by advances in new methodologies to study viruses, such as next-generation sequencing, metagenomics, proteomics and bioinformatics. Still with the modern advances in virology and viral sequence data bases, the human virome still remains partially unknown (Dalmasso, Hill and Ross, 2014).

Bacteriophage abundance in human gut can reach values as high as 10^{15}, probably the highest concentration of a biological entity on the planet (Dalmasso, Hill and Ross, 2014) representing one of the biggest unexplored genetic pools in the world. Their influence on bacterial population, such as host engineering by phage integration (e.g. gene regulation), alteration of cell physiology (*e.g.* virulence factor, antibiotic resistance, metabolism and phage immunity), phage induction (*e.g.* modification of bacterial populations by lysing competitor bacterial strains, delivering intracellular contents with nutritional value to neighbouring cells turns them into one of the major bacterial gut modulators (Nanda and Thormann, 2015; Howard-varona *et al.*, 2017).

The viral infection of bacterial hosts through the recognition of host-specific surface receptors and phage genome injection into the cell is their main characteristic (Rakhuba *et al.*, 2010). Gut phageome is constituted by tailed phages belonging to the *Caudovirales* order (in *Podoviridae*,

Myoviridae and *Siphoviridae* families) being the most diverse, and *Microviridae* family representing the most abundant (Reyes *et al.*, 2015).

The human gut phageome contains also RNA phages that are transient members of the gut microbiota (Zhang *et al.*, 2006). Studies that focus on RNA phages are scarce when compared with DNA phages, hence RNA phages deserve a full scientific attention due to the high proportion of uncharacterized phages (Siljander, Honkanen and Knip, 2019).

One of the most highly abundant identified phage in the human gut is crAssphage that targets members of the *Bacteroides* genus or Bacteroidetes phylum (Guerin *et al.*, 2018).

1.2.1 The life cycle of Bacteriophages

Bacteriophages can replicate through four life cycles: lytic, lysogenic, pseudolysogenic and chronic regarding their host infection (Weinbauer, 2004; Mirzaei and Maurice, 2017) (**Figure 1.1**). The lytic life cycle is characterized by strict lytic phages (*e.g.* phage T1, phage T4 and phage T7) that through manipulation of bacterial host machinery replicate and propagate by cell lysis (Weinbauer, 2004; Mirzaei and Maurice, 2017). In contrast temperate phages are defined by their lysogenic lifestyle switching to lytic cycle if required (*e.g.*, induction by antibiotics). When the phage is in a lysogenic cycle, the phage genome is integrated in the bacterial chromosome (prophage) (*e.g.*, phage λ and phage P2) or maintained as a plasmid (*e.g.* phage P1) in a latent state replicating with the bacterial host (lysogen) (Weinbauer, 2004; Mirzaei and Maurice, 2017).

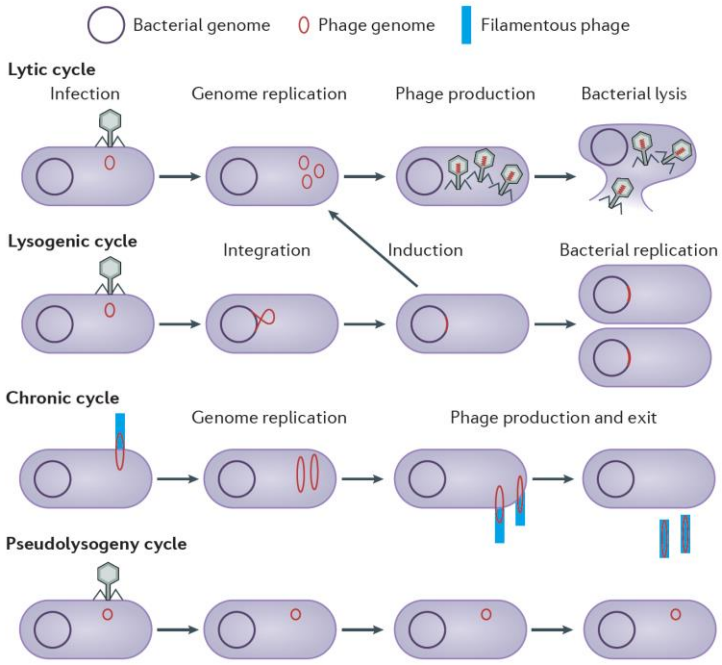

Figure 1.1 Bacteriophage life cycles (Source: Mirzaei and Maurice, 2017)

Chronic life cycle is characterized by the constant replication and release of phages (*e.g*. phage M13), achieved by a budding mechanism. Infected hosts serve as a nest for the viral particles replication without causing cell lysis (Weinbauer, 2004; Mirzaei and Maurice, 2017).

In pseudolysogeny the host contains the phage genetic material as a plasmid-like prophage, that neither integrates (*e.g.*, phage P22) nor enters a lytic cycle (*e.g.*, phage T4). Detection of this type of life cycle is seen in nutrient depleted hosts, where low infection rate success is conferred by loss of enzymatic receptors, or superinfection immunity provided by an already integrated prophage (Cenens *et al.*, 2013; Feiner *et al.*, 2015).

The major steps applied to all bacteriophage life cycles are absorption, separation of the genetic material from the capsid proteins, expression, replication, production of infectious particles, packaging and transmission, with lysogenic and lytic considered to be the dominant life cycles in the gut phageome (Weinbauer, 2004; Mirzaei and Maurice, 2017).

Temperate phages can dominate the gut virome probably by their low **k**ill **t**he **w**inner (KTW) dynamics and large ecological impacts in terms of bacterial fitness (Matos *et al.*, 2013; Mirzaei and Maurice, 2017).

1.2.2 Bacteriophages´ mechanisms

The integration of prophages in the bacterial genome can be fulfilled by site-specific recombination at *att* sites or random site transposition, near non-coding genes (*e.g.*, tRNA genes), low functional genes and intergenic regions. The integration process requires always the presence of *att* sites in the bacterial and phage genome (*att*B and *att*P, respectively), a phage-encoded integrase protein and expression of accessory host factors, such as IHF protein (this protein displays the capability of upregulating the expression of Int topoisomerase and late gene repressors, promoting lysogeny) (**Figure 1.2**) (Ramisetty and Sudhakari, 2019).

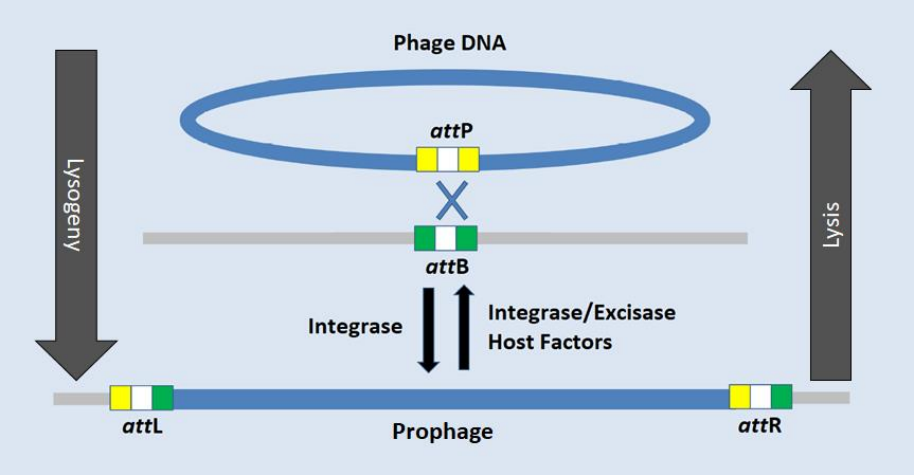

Figure 1.2 Bacteriophage integration and excision positions (Source: Ramisetty and Sudhakari, 2019).

The site-specific integration of the phage in the bacterial genome yields the recombinant *attL* and *attR* (*cis* elements) (**Figure 1.2**). The presence of multiple loci in the bacterial genome with *att* site can lead to integration events within genes, disrupting or dysregulating them, affecting bacterial fitness in a tolerable, advantageous, or intolerable way (Ramisetty and Sudhakari, 2019). A mechanism of phage evolution and survival is largely represented by the broad range of hosts that they can infect. An example of this is the utilization of specific markers for integration in different hosts, such as within conserved gene regions that code for functional proteins (*e.g.*, isocitrate dehydrogenase gene) in distantly related species, representing a tolerable recombination (Hill, Gray and Brody, 1989; Ramisetty and Sudhakari, 2019).

Utilization of phage genes by bacterial populations to provide selective advantages, such as antibiotic resistance, proteases, cell division inhibiting factors, auxiliary metabolic genes, and virulence genes demonstrates the co-evolution of bacteria and bacteriophages in the gut microbiota (Virgin, 2014; Wang and Wood, 2016). Not all cases are selective advantages as lysogeny of large prophages may lack ecological benefits, becoming a metabolic burden to their host (Wang and Wood, 2016).

One of the main advantages to a lysogen in the gut bacterial community is superinfection immunity, that prevents multi infections of a lysogen by similar phages. Different mechanisms provide this superinfection immunity or superinfection exclusion (Sie) through conformational changes in phage receptor sites, modifications in DNA transfer system, inactivation of phage lysozymes, DNA phage degradation and abortive exclusion (Ramisetty and Sudhakari, 2019).

Prophage excision depends on the excisase protein to recognize the conserved *att*L and *att*R (cis elements) flanking the prophage genome, which is only achieved efficiently with cooperation of integrase and excisase (**Figure 1.2**) (Ramisetty and Sudhakari, 2019). A successful excision can be confirmed by the formation of *att*B (circularization of the bacterial genome) and *att*P (phage episome formation), ending with the possible host lysis (**Figure 1.2**) (Chen *et al.*, 2019; Ramisetty and Sudhakari, 2019).

Induction from their dormant state can be spontaneous or a forced event through variable biotic or abiotic signals, but many are related to DNA damage (Nanda and Thormann, 2015; Mirzaei and Maurice, 2017). The general mechanism that is linked with prophage induction is the host SOS response. RecA is a key protein to start host SOS response after stress imposed on the cell, its effect on the host starts by the autocatalytic cleavage of the SOS host repressor (LexA) allowing SOS genes to be expressed (*e.g.*, cell inhibition growth and DNA repair) (**Figure 1.3**). At the same time the phage repressor is cleaved triggering the derepression of lytic promoters facilitating excision, replication, genetic material packaging and cell lysis by endolysins (*e.g.*, holin and lysin) (**Figure 1.3**). Well-known factors to induce this SOS-RecA mediated induction are reactive oxygen species (ROS) and antibiotics, such as mitomycin C (Mit C) and fluoroquinolones, through DNA damage (Nanda and Thormann, 2015; Mirzaei and Maurice, 2017). A recent study demonstrated that induction depends on the expression of *recA* gene and can be triggered by short-chain fatty acids and fructose levels in the murine gut in response to the symbiont *Lactobacillus reuteri* (Oh *et al.*, 2019).

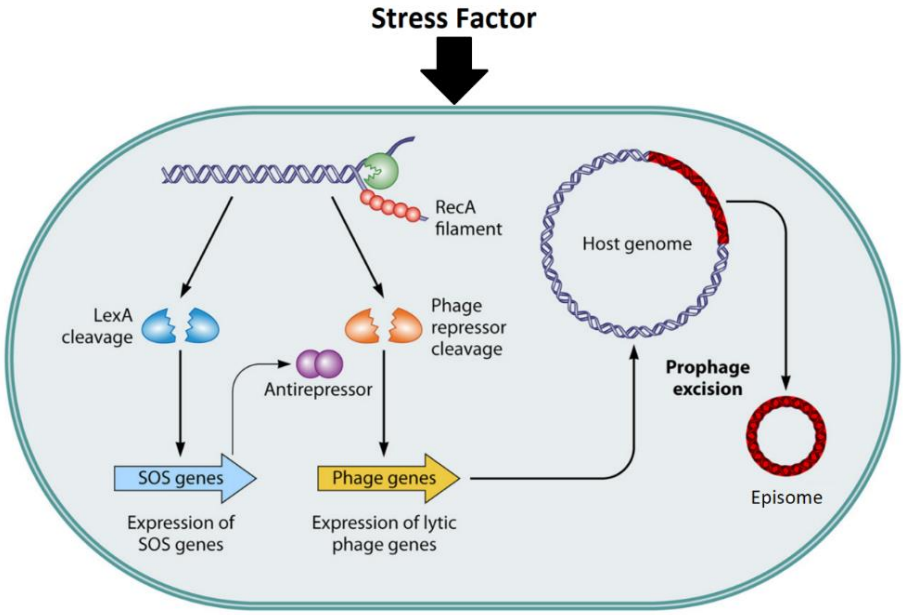

Figure 1.3 Host SOS mechanism (Source: Nanda and Thormann, 2015).

The failure of prophage excision from the bacterial genome due to defective recombinase proteins, homologous recombination with related phages and lack of *cis* elements create cryptic prophages (grounded prophages) in their host's genomes (**Figure 1.4**) (Ramisetty and Sudhakari, 2019).

The prevalence of grounded prophages within various bacterial genomes from human and animal pathogens indicates a co-evolutionary selection of such phage genomes (Casjens, 2003; Bobay, Touchon and Rocha, 2014).

Cryptic prophages are prone to suffer higher rates of deletions, mainly in genes that do not confer fitness advantages to their hosts. Deletion of this genes can be due to the complex process of phage domestication leading to a gene temporal elimination, possibly grounding prophages in the bacterial genome (**Figure 1.4**) (Wang and Wood, 2016).

Phage domestication prevents cell death from phage lytic cycle through organized and rapid inactivation of viable prophages, followed by slow degradation of the integrated phage genes, accomplished with point mutations and deletions (Bobay, Touchon and Rocha, 2014). In the temporal space that genes are eliminated from integrated prophages, they can become a force of horizontal gene transfer (HGT), as cryptic prophages can encode gene transfer agents (GTAs) that

transfer random pieces of the bacterial genome to a recipient cell, but not their own (Lang, Zhaxybayeva and Beatty, 2012). The capability and function of incomplete prophages to impact bacterial communities turn them into one of the greatest movers in bacterial evolutionary mechanics, as seen through horizontal gene transfer with significant impact on gene to gene due to the high abundance of bacteriophages in all ecosystems and niches worldwide (Casjens, 2003; Lang, Zhaxybayeva and Beatty, 2012; Wang and Wood, 2016). These cryptic prophages may provide then selective advantages in bacterial hosts, such as mobile genetic elements, antibiotic resistance, virulence, toxin islands and cell division inhibiting factors (Wang and Wood, 2016).

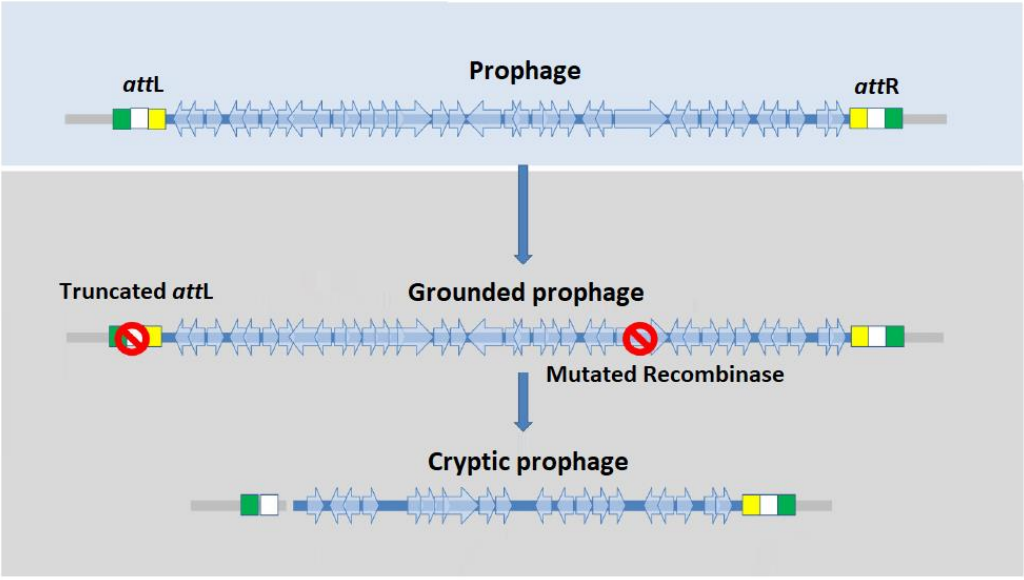

Figure 1.4 Cryptic prophage formation (Source: Ramisetty and Sudhakari, 2019). Integrated prophage in the bacterial chromosome, through site-specific recombination in attP and attB sites, is grounded by mutations in the attL and recombinase gene blocking phage excision. Gradual mutations in the grounded prophage genes, usually deletions, define the cryptic prophage in the bacterial genome.

1.2.3 Bacteria defences and bacteriophage counter-defences

Bacteria and phages present no stopping co-evolution that can lead to diverse bacterial immune mechanisms against bacteriophages and phage counter-defences (Samson *et al.*, 2013; Houte, Buckling and Westra, 2016). The phage immune mechanisms used by bacteria act at different levels on the phage life cycle efficiently avoiding phage infection (Houte, Buckling and Westra, 2016). This mechanisms can consists of surface modifications in phage receptors, DNA and RNA

degradation with restriction-modification (R-M) systems, CRISPR-Cas system immunity and abortive infection (Abi) system (**Figure 1.5**) (Houte, Buckling and Westra, 2016).

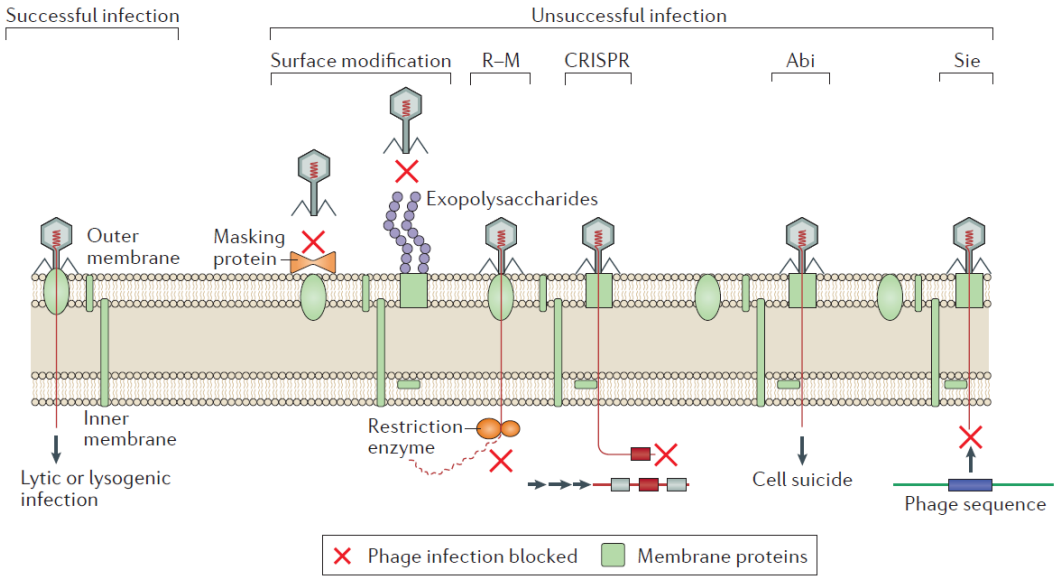

Figure 1.5 Phage host defences. Successful infection by the injection of phage DNA generating a lytic or lysogenic infection. Mechanisms that provide a unsuccessful infection: surface modification by masking phage receptors or block receptors with exopolysaccharides; restriction modification (R-M) applied by restriction endonuclease activity against not methylated phage DNA; CRISPR system where spacer sequences in the CRISPR loci (red block between two grey blocks) match phage region (red block in phage infecting genome) allowing specific cas proteins to cleave specific incoming viral DNA; abortive infection (Abi), infecting phage DNA signals host to apoptosis; superinfection exclusion (Sie), integrate prophage (phage sequence) protects host from reinfection by the same type of phage with different coded mechanism (Source: Mirzaei and Maurice, 2017).

1.2.3.1 Bacterial host surface modification

In the initial stage when the phage receptor-binding protein (RBP) recognizes its host by specific surface receptors, bacteria develop surface modifications mechanisms to avoid recognition (Houte, Buckling and Westra, 2016). The most radical modification is the total loss of phage receptors (*e.g.*, flagellum or the pilus) preventing adsorption but inflicting large fitness deficits to the cell. Less costly bacterial mechanisms presenting the same finality without receptor loss can vary from short temporal downregulation of viral receptors (Høyland-kroghsbo, Mærkedahl and Svenningsen, 2013; Tan and Svenningsen, 2015), receptor mutation (Qimron *et al.*, 2006; Scanlan, Buckling and Hall, 2015), receptor blockage (Nordstrom and Forsgren, 1974; Riede and Eschbach, 1986) or masking (Scanlan and Buckling, 2011).

To counter this bacterial defence, phages evolve to recognize new receptors (Chatterjee and Rothenberg, 2012), ligate to altered receptor structures (Michel *et al.*, 2010) and unblock masked receptors (Scholl, Adhya and Merril, 2005) (Samson *et al.*, 2013)

1.2.3.2 The process of superinfection exclusion

The process of superinfection exclusion (Sie) prevents the initial entry or replication of phage DNA. This bacterial defence mechanism only occurs in lysogens (Houte, Buckling and Westra, 2016). Mediation of Sie can be applied by cell surface alterations, such as modification of phage receptor (Uc-mass *et al.*, 2004), inner membrane alteration (Cumby *et al.*, 2012), immune protein production (Lu and Henning, 1994; Pope *et al.*, 2011; Houte, Buckling and Westra, 2016).

1.2.3.3 The mechanism of Restriction Modification

A widespread and diverse immune mechanism is restriction-modification (R-M) system. R-M system cleaves non-self-unmodified DNA against the protected self-modified DNA (Samson *et al.*, 2013; Houte, Buckling and Westra, 2016). The two components that make part of this system are a methyltransferase (MTase), responsible for DNA methylation, coupled with a restriction endonuclease (REase) that cleaves unmethylated DNA through the recognition of specific restriction sites (Res). This R-M system is divided in four different types (type I to IV) based on subunit composition, specific Res recognition, cofactor requirements and substrate specificity (Roberts *et al.*, 2003; Houte, Buckling and Westra, 2016). The effectiveness of this immune mechanism regarding host protection against phages can reach to 10^8-fold (Houte, Buckling and Westra, 2016).

Phage counter defence to R-M system are anti-restriction strategies, classified as passive and active mechanism. The passive mechanism of phage evasion can be achieved if the infecting phage DNA is protected through rapid modification with MTase, avoiding the recognition from the REase and resisting R-M system targeting. An additional advantage provided by this counter measure is the possibility of propagation to other cells that express similar R-M systems. To counter, these bacteria developed also REases that recognize methylated foreign genetic material. The number of

restriction sites, orientation, distance and incorporation of modified bases in phage genome can elude R-M systems (Samson *et al.*, 2013).

Active mechanism of phage evasion include the masking of Res for REase in phage genome, interference with REase active sites, codification of modification genes (e.g. methyltransferase) and stimulation of host modification enzymes (Samson *et al.*, 2013).

1.2.3.4 The CRISPR-Cas system

Clustered regulatory interspaced short palindromic repeats (CRISPR) and CRISPR- associated Cas proteins form the extremely diverse CRISPR-Cas system that represents the adaptive immune systems of Bacteria and Archaea (Houte, Buckling and Westra, 2016).

The *cas* genes code for the protein machinery that carries the immune response, with the CRISPR loci providing the memory of previous infections. The basic principle of all known CRISPR-cas systems is the acquisition of foreign short stretches of non-repetitive DNA (spacer) and their integration between repeated sequences in the CRISPR loci of the host genome (adaptation), with the spacer sequences inserted matching the respective region in the phage genome (protospacer) (**Figure 1.6**). Transcription of spacers result in a mature CRISPR RNA (crRNA) that forms a complex with one or more Cas proteins. When crRNA-Cas complexes are formed they can bind and cleave complementary nucleic acids of the entering DNA (interference), with additional Cas nucleases, resulting in host immunity (**Figure 1.6**) (Houte, Buckling and Westra, 2016).

The CRISPR-Cas systems can be classified in two types (Class 1 CRISPR-Cas system and Class 2 CRISPR-Cas system) that can provide immunity to phage infection (Oost *et al.*, 2014; Houte, Buckling and Westra, 2016). In the human gut microbiome more than 52.000 spacer sequences were identified in 124 individuals (Mirzaei and Maurice, 2017).

Evasion by mutation is one of the phages mechanisms to evade the CRISPR-Cas system interference, this can be achieve through a single-nucleotide substitution in the protospacer region or in the conserved protospacer-adjacent motif (PAM) (**Figure 1.6**), deletion of protospacer and/or PAM, anti-*CRISPR* genes (Bondy-denomy *et al.*, 2012) and phage-encoded CRISPR-Cas system (Samson *et al.*, 2013).

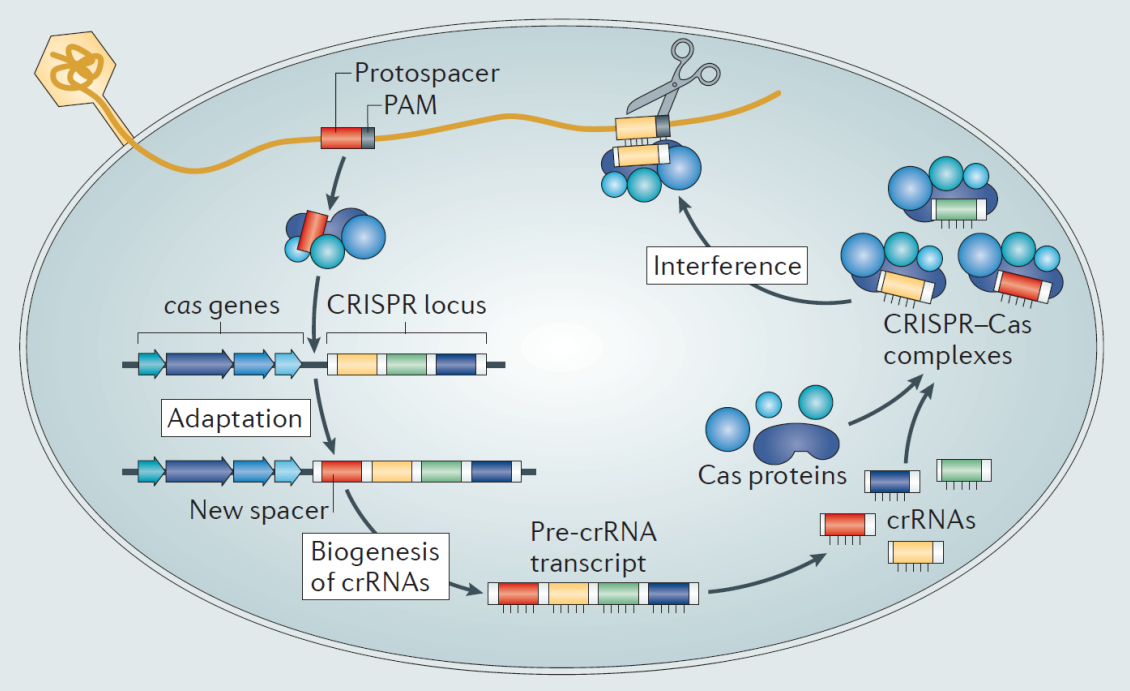

Figure 1.6 CRISPR-Cas system. Acquisition of a non-repetitive sequence from the phage genome (spacer) by Cas proteins, and spacer integration in CRISPR loci in the bacterial chromosome (Adaptation). Transcription of spacer and *cas* genes provided the host with CRISPR-Cas complexes that target spacer homologous sequences in the phage (protospacer) cleaving this regions (interference), resulting in host immunity against the infecting phage. (Source: Samson et al., 2013)

1.2.3.5 The Abortive infection systems

Abortive infection (Abi) systems are the most extreme host defence against phages affecting various steps of the phage replication through host apoptosis, preventing the replication and transmission of phages to neighboring uninfected cells of the phage. This altruistic sacrifice protects the remaining bacterial population from phage infection (Samson *et al.*, 2013; Houte, Buckling and Westra, 2016).

Abi systems range from single protein to a multiple protein complex (Samson *et al.*, 2013; Houte, Buckling and Westra, 2016; Mirzaei and Maurice, 2017) In Gram-negative strains Abi systems, such as the rapid II exclusion (Rex) system and the late inhibitor of phage T4 (Lit) system (encountered in defective prophage e14 of the *E. coli* K-12 strain) protect bacterial populations (Samson *et al.*, 2013).

The Rex systems stops the ATP-dependent processes by suddenly dropping the ATP level with the complex RexA and RexB, the Lit system cleaves the ribosomal elongation factor-Tu when peptide Gol is attached inhibiting protein synthesis, both leading to cell death and the coupled abortion of virus infection (Samson *et al.*, 2013; Houte, Buckling and Westra, 2016).

The toxin-antitoxin (TA) system belongs to a subgroup of Abi systems leading to bacterial death or dormancy after phage infection. This system is composed by a toxin and a neutralizing antitoxin, becoming active with the degradation of the labile neutralizing toxin (Samson *et al.*, 2013).

Resistance against the multiple types of AI systems to promote full phage life cycle are complex, (Samson *et al.*, 2013). Mutations in specific genes, such as the *gol* gene, encoded in a *E. coli* T4 phage lead to a non-activated Lit system allowing phage replication (Champness and Snyder, 1982). Phage containing a mutated *mot*A gene, delays the Rex system activation avoiding depolarization of the membrane, allowing completion of the phage infection cycle (Samson *et al.*, 2013). The phage life cycle can also be completed by a self-coding antitoxin protein, providing defence against the host TA system, as seen in phage T4 that encodes an antitoxin called Dmd against *E.coli* LsoA and RnlA toxins (Otsuka and Yonesaki, 2012).

1.2.3.6 Searching for bacteriophages in bacterial genomes

A tool that greatly facilitates the detection of prophages in the bacterial genomes is PHASTER - PHAge Search Tool Enhanced Release, which shows the capability of rapid identification and annotation of prophage sequences within bacterial genomes and plasmids (Arndt *et al.*, 2016). This tool provides detailed information regarding the genomic location of the phage elements, the completeness of the predicted prophage, and its component proteins allowing an easy access to prophage identification and characterization. This bioinformatics tool has been previously used already in studies focused on the identification and characterization of prophages within bacterial genomes (Lorenz *et al.*, 2016; Crispim *et al.*, 2018).

1.3 Type 1 Diabetes *mellitus* and the gut microbiota

Type 1 diabetes mellitus (T1D), initially called juvenile-onset diabetes, is a chronic immune-mediated disease characterized by the autoimmune attack of pancreatic β-cells in the islet of Langerhans located in the pancreas (**Figure 1.7**) (Katsarou *et al.*, 2017; Siljander, Honkanen and Knip, 2019). The destruction of the β-cells causes severe hyperglycemia leading to high morbidity and if not timely diagnosed may result on premature death. Generally, this disease is more common during the early years of life, with half of the cases in children younger than 15 years of age (Katsarou *et al.*, 2017). T1D prevalence is 1% worldwide with recent studies pointing to a new-onset of T1D in patients older than 20 years (Aathira and Jain, 2014; International Diabetes Federation, 2017). The increasing incidence of T1D in children under 15 years old was observed in Finland, where the number of cases increased in the last five decades from 34 to 65 new cases/100.000/year (Westfall, Ringel and Gardner, 2013). The causes of this increase cannot be explained just by the genetic predisposition to the disease, with the changes introduced by the modern life style will also having a significant contribution (Siljander, Honkanen and Knip, 2019). Autoimmunity is the major factor in the pathophysiology of T1D, with diverse genetic markers identified and associated to the high risk of development of T1D (95%) (Katsarou *et al.*, 2017). Susceptible individuals to T1D harbor a combination of human leukocyte antigen DR3-DQ2 or DR4-DQ8 haplotypes (Kantárová and Buc, 2007). Progression to the onset of T1D can be defined through 3 stages (Katsarou *et al.*, 2017; Siljander, Honkanen and Knip, 2019). This progression starts with the appearance of autoantibodies to β-cells (seroconversion) (Stage 1) followed by β-cell loss and hyperglycemia (Stage 2) until the symptomatic stage (Stage 3) (Katsarou *et al.*, 2017; Siljander, Honkanen and Knip, 2019).

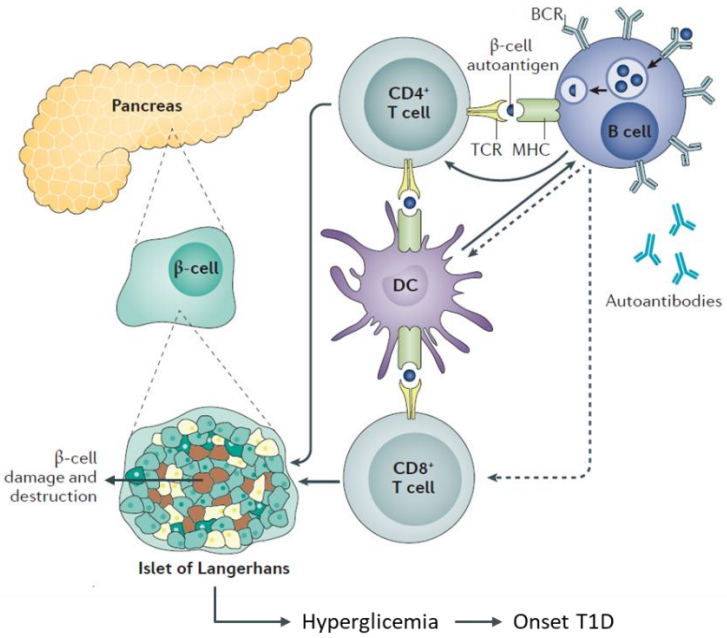

Figure 1.7 β-cell autoimmunity development. Activation of β cells allow the presentation of β-cell autoantigens to CD4+ and CD8+ T cells, together with dendritic cells (DCs). Exposure of β cells to β-cell autoantigens leads to the production of islet-targeting autoantibodies, assisting as asymptomatic biomarkers of T1D. Activation of CD4+ and CD8+ T cells leads to β-cell damage and destruction, impairing the production of insulin allowing hyperglycemia followed by T1D onset. Dashed arrows indicate the potential interactions between β cells and CD8+ T cells and between β cells and DCs. BCR, β cell receptor; TCR, T cell receptor. (Source: Katsarou et al., 2017).

It was reported that children with β-cell autoimmunity show low numbers of *Bifidobacterium* species and increased abundance of *Bacteroides* before the onset of T1D (Goffau *et al.*, 2013). Differences in the microbial metabolic pattern between healthy and seroconverted cases suggested that gut microbiota dysbiosis can provide an initial trigger for the progression of T1D, as low numbers of butyrate-producing, lactate-producing and mucin-degrading bacteria were observed (Brown *et al.*, 2011). These low densities in both bifidobacterial and butyrate-producing bacteria can affect the intestinal epithelial barrier function and cause inflammation, with the proposed trigger for autoimmunity being the increased adhesion and flagella synthesis facilitated by dysbiosis. (Brown *et al.*, 2011; Goffau *et al.*, 2013; Kostic *et al.*, 2015a).

In Finnish children at high genetic risk of developing T1D, an increase in the population of *Bacteroides dorei* was detected 8 months prior to seroconversion to the first islet autoantibody, suggesting that early changes in the bacterial gut microbiota can be explored to predicted T1D autoimmunity (Davis-Richardson *et al.*, 2014). Seroconversion and onset establishment can

comprehend a large time range, namely from weeks to two decades as reported in the study of Ziegler et. al. (2013)

where the progression probability to T1D was greater than 80% after a follow-up study of 15 years with the detection of two islet autoantibodies. Strong evidences support that alterations occur in the gut microbiome before the onset of T1D, but still lacking the knowledge regarding what drives these changes. Dysbiosis may be a secondary stage to seroconversion with bacteriophages playing a role yet unidentified.(Ziegler *et al.*, 2013).

1.3.1 Bacteriophages in health and disease

Bacteriophages represent one of the means to manipulate the gut ecologic environment through various mechanisms impacting host health by a host-bacterial-phage cycle (Mirzaei and Maurice, 2017). Phages help to maintain the normal gut bacteriome through their stable symbiosis using different tools, such as horizontal gene transfer, lysogeny and competition control (Howard-varona *et al.*, 2017).

Gut diseases can be influenced by phage added characteristics, such as virulence factors encoded within phage genomes, such as Shiga toxins (stx1, stx2) (encoded by H-19B phage) in enterohaemorrhagic *E. coli*, effector proteins (Cif, NleH) (encoded by lambdoid phages) in enteropathogenic *E. coli* (Boyd, 2012).

Virome alterations can be correlated with specific diseases as observed in Intestinal Bowel Disease (IBD) (Norman *et al.*, 2015). A study conducted in the UK and USA with patients suffering from IBD highlighted the interaction of the gut virome and bacteria where an enrichment in Caudauvirales phages in Crohns´s disease (CD) and Ulcerative colitis (UC) patients and low bacterial diversity was reported (Norman *et al.*, 2015). This finding lead to the speculation that viral changes may be the primary driver in the inflammatory processes of IDB, with secondary shifts applied to bacterial densities and diversity (Norman *et al.*, 2015). IBD patients detected with a higher occurrence of *Faecalibacterium prausnitzii* temperate phages, Caudovirales from the Norman study, could explain the lower abundance of *F. prausnitzii* in IBD patients gut microbiota (Cao, Shen and Ran, 2014).

The limited number of studies dedicated to RNA bacteriophages in the gut constitute a significant gap that needs to be elucidated, particularly their relationship with the gut immune system and

bacterial populations (Mukhopadhya *et al.*, 2019). These observations suggest that a relation between phages
and their bacterial hosts is established, but the mechanisms involved still escape our grasp, requiring future studies to identify the role of the virome in IBD and other diseases.

1.3.1.1 Type 1 diabetes *mellitus* and bacteriophages

A small number of clinical studies were conducted to access the intestinal virome´s role in the development of T1D (Leonard *et al.*, 2014; Cinek *et al.*, 2016; Zhao *et al.*, 2018). In one study the bacteriome and viruses in the gut at early-onset islet autoimmunity in Finnish children with genetic susceptibility to T1D was analyzed at 3, 6, and 9 months before the seroconversion (Cinek *et al.*, 2016). The children that progressed to T1D showed four bacterial operation taxonomic units (OTUs) significantly less abundant than controls, specifically the reduction in *B. vulgatus* and *Bifidobacterium bifidum* species (Cinek *et al.*, 2016). A relation between CrAssphage abundance and predominant *B. dorei* in seroconverted children, suggested that this phage could contribute to the modification of the intestinal bacteriome, particularly in *Bacteroides* genus towards a defining islet autoimmunity dysbiosis, but a direct correlation of this phage and seroconversion was not observed (Davis-Richardson *et al.*, 2014; Cinek *et al.*, 2016).

Another study in young Finns and Estonians detected low phage diversity and richness in children with autoantibodies (Zhao *et al.*, 2018). This study demonstrated that a predictive succession occurs in the gut microbiome, as viral patterns are consistent with normal intestinal development across individuals sorted by age (Zhao *et al.*, 2018).

To examine further the effect of *B. dorei* dominance in the gut bacteriome prior to seroconversion the DNA methylation patterns were evaluated regarding GATC methylation sites in two *B. dorei* genomes that were collected from two Finnish children with high genetic risk for T1D (Leonard *et al.*, 2014). One of the subjects progressed to autoimmunity after 1.5 months, showing a heavily methylated genome to GATC sites by a single DNA adenine methyltransferase (Dam) in a prophage, in contrast the other sample did not progress to autoimmunity and did not reveal a methylated genome. Regarding this finding, the estimated effects of gut bacterial GATC methylation on autoimmunity outcome are still limited (Leonard *et al.*, 2014). However it is

plausible to speculate that bacteriophages can provide *B. dorei* with selective abilities, particularly through the methylation of genes that can influence gene expression either by up-regulating or down-regulating (*e.g.* lysogeny) within Dam of the 5´-GATC-3´ motif (Løbner-Olesen, Skovgaard and Marinus, 2005; Leonard *et al.*, 2014).

Bacteriophages display the tools to modulate the bacterial populations in the gut through their abundance and diversity affecting the bacterial homeostasis that ultimately can result in dysbiosis. The trigger that starts the metabolic cascade for the autoimmunity attack to β-cells may come by an ample source of biological identities, where viral patterns in the bacterial communities can be one of them.

2 Objectives

The main aim of the current study was to elucidate the prophage patterns in the genomes of the strains of *B. dorei* that were isolated from children with T1D and healthy controls. Another objective was to establish the stress conditions that could trigger the prophage lytic cycle, particularly the exposure to mitomycin C, ox-bile and hydrogen peroxide.

3 Materials and Methods

3.1 Materials

- Anaerobic Jar, Oxoid (UK).
- AnaeroGen 2.5 L, Atmosphere Generation Systems, Oxoid (UK).
- Analytical Balance AE 200, Mettler (USA).
- Analytical Balance XS-410, Fisher Scientific (Portugal).
- ARE Heating Magnetic Stirrer, VELP Scientifica (USA).
- Autoclave Uniclave 88, AJC (Portugal).
- BH-EN 2000 Class II Series Biosafety Cabinet, Faster (Italy).
- Bio48 Laminar Flow Chamber, Faster (Italy).
- Bullet Blender, Next Advance (USA)
- Electrophoresis Documentation and Analysis System (EDAS) 290, Kodak (USA).
- Electrophoresis Power Supply - EPS 301, Amersham Pharmacia Biotech (UK).
- Gel Electrophoresis Apparatus GNA-100, Pharmacia Biotech (USA).
- GENESYS 10vis Spectrophotometer, Thermo Fisher Scientific (USA).
- Gravity Convection Incubator, Binder (Germany).
- Infinite M200 Microplate Reader, Tecan (Switzerland)
- Mikro 200 Centrifuge, Hettich Zentrifugen (UK).
- Mikro 200R Centrifuge, Hettich Zentrifugen (UK).
- Mini-V/PCR Vertical Laminar Flow Bench, Telstar (Spain).
- Multiplaces Dry Heating Bath, J.P. Selecta (Spain).
- pH-meter GLP 21, Crison (Spain).
- Qubit 2.0 Fluorometer, Thermo Fisher Scientific (USA).
- REAX 2000 Vortex Mixer, Heidolph (Germany).
- SZ Stereo Microscope, Olympus (Japan).
- Ultra-low Temperature Freezer -80ºC U725, Innova New Brunswick Scientific (USA).
- Ultraviolet Transilluminator, UVITEC (UK).
- Universal 320 Centrifuge, Hettich Zentrifugen (UK).

3.1.1 Culture medium

- Brain Heart Broth (Biokar Diagnostics, France) prepared according to the manufacturer's instructions, supplemented with hemin 0.1%, v / v, (Sigma Life Science, USA) and L-cysteine 0.1%, w / v, (Sigma Life Science, USA) (BHI+ H), pH 7.1. Agar at 1.5% (Biokar Diagnostics, France) was added when required to obtain solid medium.

- Brain Heart Broth (Biokar Diagnostics, France) prepared according to the manufacturer's instructions, supplemented with yeast extract 0.5%, w/ v (Biokar Diagnostics, France), hemin 0.1%, v / v, (Sigma Life Science, USA) and L-cysteine 0.1%, w / v, (Sigma Life Science, USA) (LY-BHI), pH 7.1.

3.1.2 Solutions

The solutions used in the present study were as follows:

- Hemin solution 0.1% - 0.1 g hemin (Sigma Life Science, USA), 2 ml 1M NaOH in deionized water to 100 mL – requires protection from light since it is light sensitive
- Hydrogen Peroxide (3%)
- Mitomycin C solution (2 mg/mL) (Sigma Life Science, USA)
- 1 M NaOH solution
- Ox-bile solution 4%
- Phosphate Buffered Saline (PBS) 0.13 M NaCl, 0.002 M KCl, 0.01 M Na_2HPO_4, 0.001M KH_2PO_4
- Tris-Acetate-EDTA buffer (TAE) 50x - 2 M Tris base (Sigma-Aldrich, USA), 1 M Glacial acetic acid (PanReac, USA), 0.5 M EDTA pH 8
- 1 M Tris-HCl buffer pH 8.5 (The pH value was set with 1 M HCl.)

3.1.3 Biological Material

The strains of *Bacteroides dorei* and *Parabacteroides distasonis* used in the present study are listed in Table 3.1.1.

Table 3.1.1 Bacteria used in this study.

Bacteria	Source
Bacteroides dorei DSM 17855	DSMZ
Bacteroides dorei C1P2	Matos, 2018
Bacteroides dorei Sb6	Matos, 2018
Bacteroides dorei Sb8	Matos, 2018
Bacteroides dorei D1P5	Matos, 2018
Bacteroides dorei D8M1	Matos, 2018
Bacteroides dorei D16P1	Matos, 2018
Bacteroides dorei D16M14	Matos, 2018
Parabacteroides distasonis D14MH1	Matos, 2018

3.2 Methods

3.2.1 Recuperation of bacterial samples

Bacteria tested in the present study are listed in Table 3.1.1. The bacterial cells were recovered from -80ºC by transferring 20 µL into BHI + H agar medium and the inoculum was distributed by streaking. The inoculated plates were incubated in an anaerobic jar with an anaerobic sachet inside (Oxoid, UK) at 37ºC for 48 h. An isolated colony was selected from the agar plate, observing the usual cultural characteristics (small to medium round and light blue colony), and then was transferred to 5 mL BHI + H medium, which was sealed with sterile liquid paraffin. The incubation occurred at 37°C for 24 h.

3.2.2 Genomic DNA extraction

Genomic DNA was isolated from 2 mL of the bacterial culture prepared as described in the previous section. DNA was extracted using Wizard Genomic DNA Purification Kit (Promega, USA), according to the manufacturer's protocol for Gram positive and Gram negative bacteria with a modification that included an additional step of lysis with sterile acid-washed glass beads (425-600 µm) (Sigma Life Science, USA) comprising 1/3 of the volume of the bacterial pellet. The mixture was sonicated during 1 min using the bead beater (Next Advance, USA).

3.2.2.1 DNA integrity control and quantification

DNA integrity was analyzed by electrophoresis using a 0.75 % (w/v) agarose gel (Lonza, Denmark). The DNA quantification was performed by using the fluorescence-based Qubit dsDNA BR Assay Kit (Invitrogen, USA).

3.2.3 Genome sequencing

Each bacterial genome was sent to sequence in 10 mM Tris-HCl buffer at the concentrations ranging from 2.7 to 5.0 µg/µL for Standard Whole Genome Service by Illumina next-generation sequencing provided by the company MicrobesNG (UK).

The steps followed to obtain a final draft genome are shown in **Figure 3.1.1**.

The bioinformatics tool Scaffold_builder (http://edwards.sdsu.edu/scaffold_builder/) (Silva et al.,2013) was used to assemble the contigs obtained from the sequencing into scaffolds - termed scaffold building. For this, one published complete genome for each genus was used to guide the assembly of the contigs. The *Bacteroides dorei* genome (NCBI accession number CP007619) was chosen to represent the *Bacteroides* genus, and the *Parabacteroides distasonis* genome (NCBI accession number CP000140) was the representative genome for the *Parabacteroides* genus.

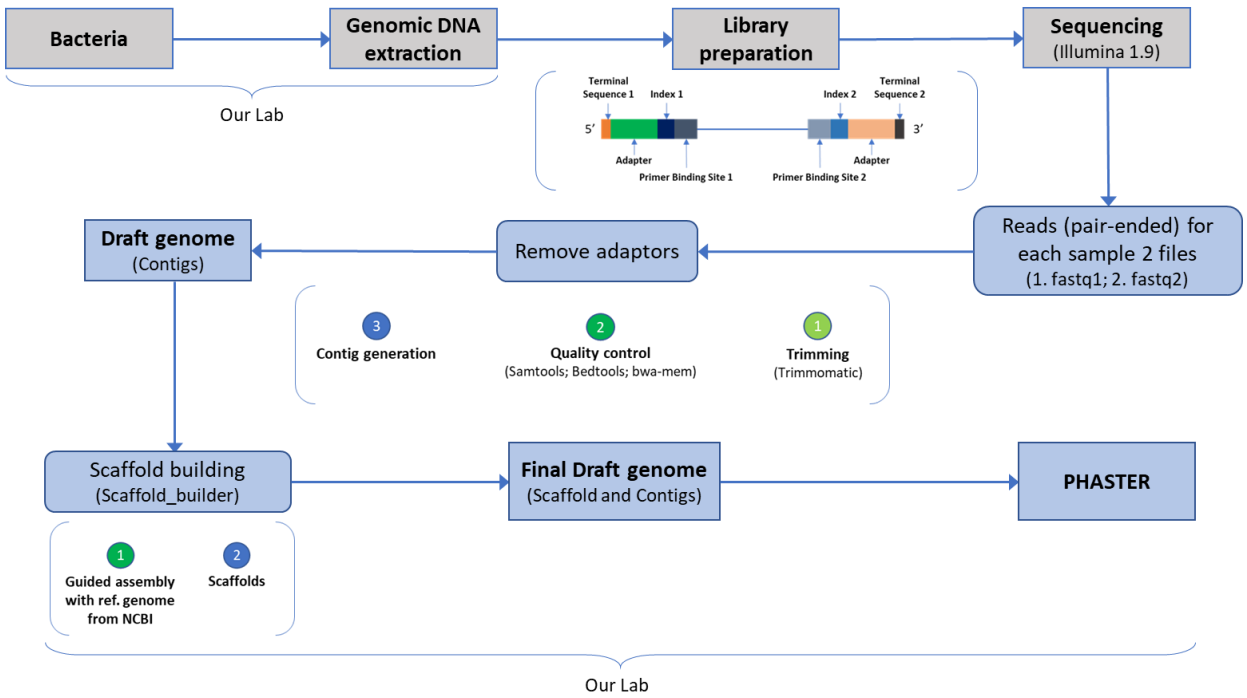

Figure 3.1.1 Workflow of data generation and analysis

3.2.3.1 Quality control of trimmed sequences

Additional quality control for the trimmed reads obtained from the sequencing facility was performed using the bioinformatics tool FastQC (Andrews S., 2010) for each of the sequence files (fastq) provided for each genome used in this study (**Table 3.1.1**).

3.2.4 Prophage induction assays

For induction assays, bacterial cultures were grown in 5 mL LY-BHI medium sealed with sterile liquid paraffin and incubated at 37°C for 24 h. Afterwards, each bacterial culture was diluted 10-fold and 100 µL were transferred to a 96-well flat-bottom microplate (Sarstedt, Germany). When the optical density (OD) at 600 nm reached 0.15 - 0.25, inducer agents were added at final concentrations of 1-7 µg/mL for mitomycin C (Sigma Madrid, Spain), 0.2% and 0.4% for Ox-bile (Sigma-Aldrich, USA), and 100 µM, 50 µM and 10 µM for hydrogen peroxide. The growth was monitored by spectrophotometry (OD_{600nm}) each hour until five hours after the start of the induction process using a microplate reader (Tecan Infinite M200, Tecan, Austria).

3.2.5 PHASTER analysis and data collection

To search for possible prophage-like elements in bacterial genomes and plasmids, the bioinformatics tool PHASTER (PHAge Search Tool Enhanced Release) (http://phaster.ca/) was applied (Arndt *et al*., 2016).

Briefly, PHASTER is a web server (http://phaster.ca/) with the capability of rapid identification and annotation of prophage sequences within bacterial genomes and plasmids (Arndt *et al.*, 2016). The pipeline starts with the submission of sequences to the server, that can either be raw sequences (fasta format) or annotated genome sequences (NCBI's GenBank file format), followed by the detection of sequences similar to known phages from published databases using BLAST (Altschul et al., 1990). After this initial detection a clustering of prophage-like sequences is done regarding the location, proximity and orientation of the sequences within the bacterial genome. The final step in the run is the calculation of the completeness score based on the presence of structural proteins

found and clustered within the genome, classifying potential prophages as: complete, questionable, and incomplete.

Several output results are provided: statistical information regarding prophage prediction, possible structural proteins, genome images showing the locations of the specific prophage-like regions encountered in the genome, and the potentially encoded proteins within each specific region of the predicted prophage.

4 Results

4.1 Identification and characterization of prophages in bacterial samples

Regarding the main goals of this study, the genome of seven *Bacteroides dorei* strains and one *Parabacteroides distasonis* strain that were previously isolated from faecal samples of children with Type 1 Diabetes (T1D) and Control children (Matos, 2018) were sequenced. The assembled scaffold genome of each of these bacterial species was subsequently searched for the presence of potential prophage-like regions using the tool PHASTER (Arndt *et al.*, 2016). The identification and characterization of the prophage elements in each bacterial strain are presented. The prophage elements identified in the tested bacteria are summarized in **Table 4.1.1** and **4.1.2**.

4.1.1 Prophages in *B. dorei* C1P2

In the strain *B. dorei* C1P2, which was isolated from healthy children, one putative prophage region (**Figure 4.1.1**) was detected in region 1 positioned in scaffold 4, sequence coordinates from 139060 bp to 146553 bp, with 7.4 kb total length, and an incompleteness score of 50. Analysis of the sequence revealed similarity to *Mannheimia* phage vB_MhM_3927AP2 (Niu et al., 2015) (14.18% shared orthologous proteins) with 7 open reading frames (ORFs) and no attachment sites. The following coding genes were identified: major capsid protein (911 bp length), prohead protease (986 bp) (both related to head proteins families), terminase (1427 bp), portal protein (with 2534 bp), HNH protein (452 bp) and a N-acetylmuramoyl-L-alanine amidase (464 bp). In the prophage-like region no tRNA was encountered.

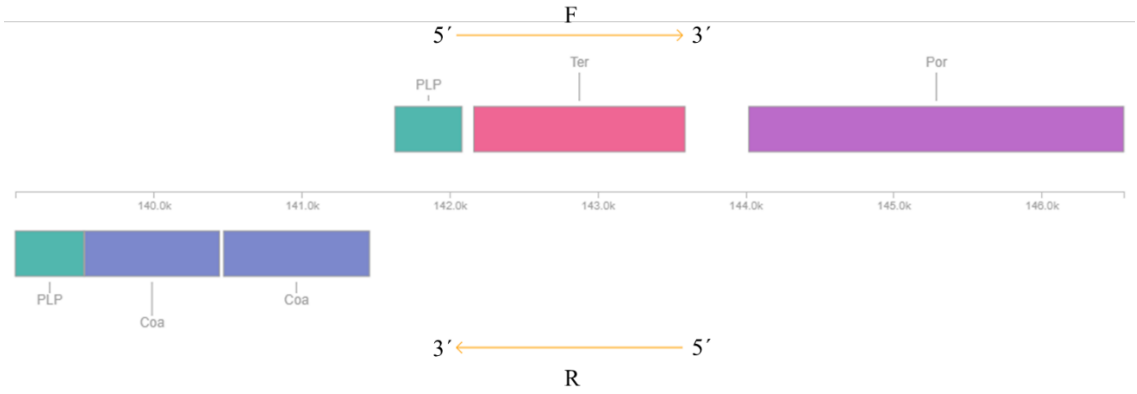

Figure 4.1.1 Predicted organization of the putative Mannheimia phage vB_MhM_3927AP2 on scaffold 4 of B. dorei C1P2 region 1. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): N-acetylmuramoyl-L-alanine amidase (PLP), major capsid protein (Coa), prohead protease (Coa), HNH protein (PLP), terminase (Ter), and portal protein (Por).

4.1.2 Prophages in *B. dorei* Sb6

B. dorei Sb6 isolated from healthy children presented one putative incomplete prophage region (**Figure 4.1.2**). Region 1 presented an incompleteness score of 10, located in scaffold 6 from 188800 bp to 199923 bp, presenting 11.1 kb total length. Sequence analysis showed similarity with *Riemerella* phage RAP44 (Cheng et al., 2012) (4.76% shared orthologous proteins), with 21 ORFs and no attachment sites in the region. Identified coding genes represented a putative RecT protein (842 bp length), a putative exonuclease (599 bp), a methyltransferase (491 bp), an anti-repressor Ant (890 bp), a VRR-NUC domain protein (278 bp), an antirepressor (701 bp), and a lambda repressor-like DNA-binding domain protein (275 bp). In the prophage-like region no tRNA was encountered.

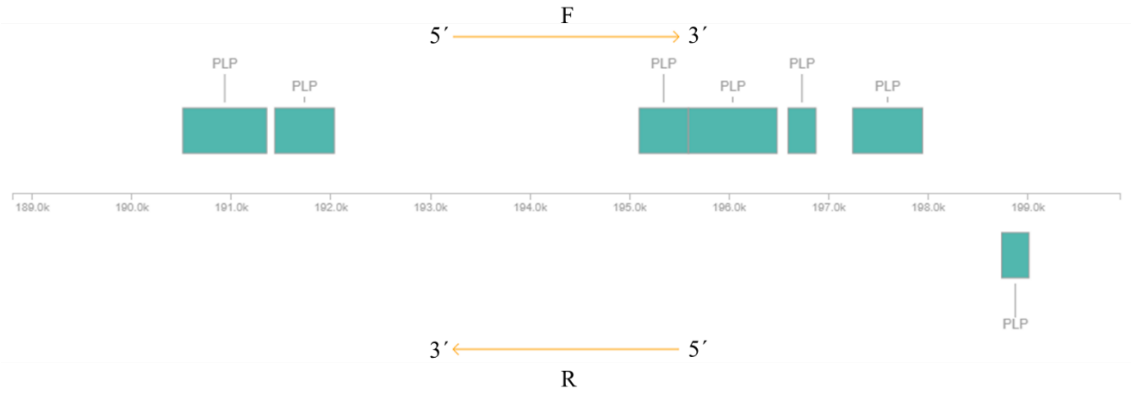

Figure 4.1.2 Predicted organization of the putative Riemerella phage RAP44 on scaffold 6 of B. dorei Sb6 region 1. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): putative RecT protein (PLP), putative exonuclease (PLP), methyltransferase (PLP), anti-repressor Ant (PLP), VRR-NUC domain protein (PLP), antirepressor (PLP), lambda repressor-like DNA-binding domain protein (PLP). PLP is Prophage-like protein.

4.1.3 Prophages in *B. dorei* Sb8

One putative incomplete prophage-like region, region 1 (**Figure 4.1.3**) was identified in the strain *B. dorei* Sb8 isolated from healthy children. This region presents an incompleteness score of 50, is located in scaffold 5 in coordinates 120653 bp to 129619 bp, having 8.9 kb total length. Sequence analysis matched in similarity with *Flavobacterium* phage 1H (Castillo and Middelboe, 2016) (7.69% shared orthologous proteins) with 13 ORFs and two attachment sites. Attachment site attL (12 bp length) in position 120653 bp to 120665 bp, and attR (12 bp length) in location 129619 bp to 129631 bp flank all the remaining ORFs, 5´-end and 3´-end respectively. Genes coding for putative antirepressor protein (521 bp length), putative anti-repressor protein (203 bp), putative endolysin (314 bp), putative endolysin/secretion activator protein (179 bp), integrase (1229 bp), and probable phage-family integrase (1244 bp) were found. In the prophage-like region no tRNA was encountered.

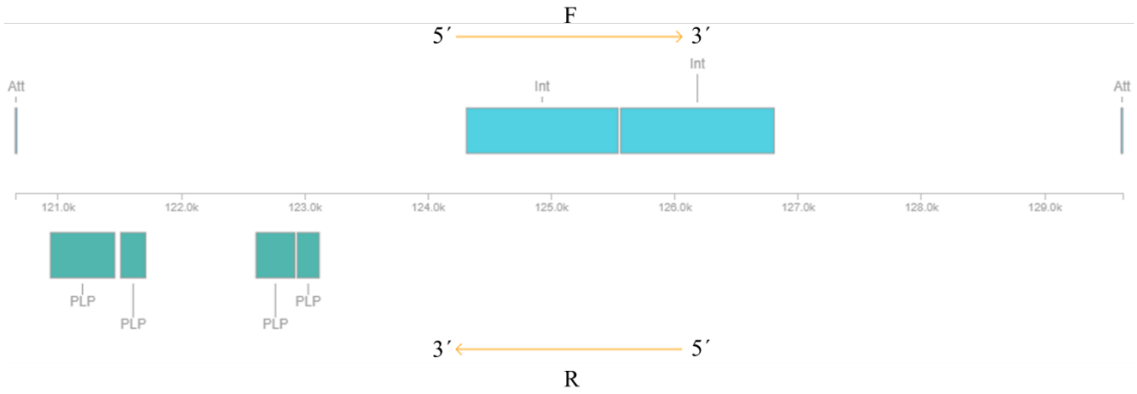

Figure 4.1.3 Predicted organization of the putative Flavobacterium phage 1H on scaffold 5 of B. dorei Sb8 region 1. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): att (Att), putative antirepressor (PLP), putative anti-repressor protein (PLP), putative endolysin (PLP), putative endolysin/secretion activator protein (PLP), integrase (Int), probable phage-family integrase (Int), and attR (Att). PLP is Prophage-like protein.

4.1.4 Prophage in *B. dorei* D1P5

In *B. dorei* D1P5 isolated from a subject with T1D, two putative incomplete phage regions were found, region 1 and 2, with incompleteness score of 10 and 40, respectively.

Region 1 (**Figure 4.1.4**) was encountered in scaffold 6 from 188800 bp to 199923 bp, with 11.1 kb total length. Sequence similarity with *Riemerella* phage RAP44 (Cheng *et al.*, 2012) (4.76% shared orthologous proteins) with 21 ORFs, and no attachment sites. Annotated coding genes represented a putative RecT protein (842 bp length), a putative exonuclease (599 bp), a methyltransferase (491 bp), an anti-repressor Ant (890 bp), an VRR-NUC domain containing protein (278 bp), RNA polymerase binding protein (701), and a lambda repressor-like DNA-binding domain protein (275 bp). In the prophage-like region no tRNA was encountered.

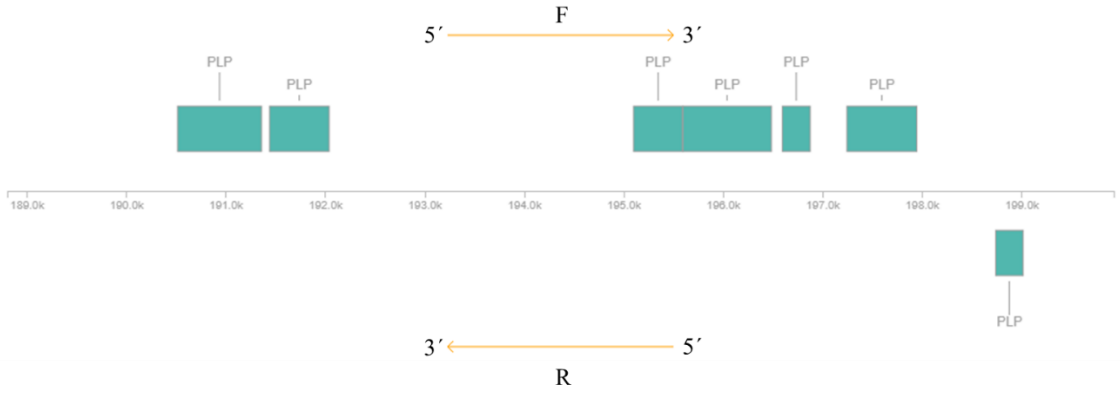

Figure 4.1.4 Predicted organization of the putative Riemerella phage RAP44 on scaffold 6 of B. dorei D1P5 region 1. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): putative RecT protein (PLP), putative exonuclease (PLP), methyltransferase (PLP), anti-repressor Ant (PLP), VRR-NUC domain containing protein (PLP), RNA polymerase binding protein (PLP), lambda repressor-like DNA-binding domain protein (PLP). PLP is Prophage-like protein.

Region 2 (**Figure 4.1.5**) is located in NODE 35 (contig not grouped in a scaffold) from 3612 bp to 12023 bp with 8.4 kb total length. High similarity with the sequence of *Agrobacterium* phage Atu ph07 (20% shared orthologous proteins) (Spollen *et. al.*, 2017), with 10 ORFs and no attachment sites. Coding genes were identified for a putative cobyrinic acid ac-diamide synthase (779 bp length), putative ArdC-like antirestriction protein (1583 bp), major capsid protein (1385 bp), major capsid protein (776 bp), IS21 transposition protein (755 bp) and IS21 transposase (1598 bp). In the prophage-like region no tRNA was encountered.

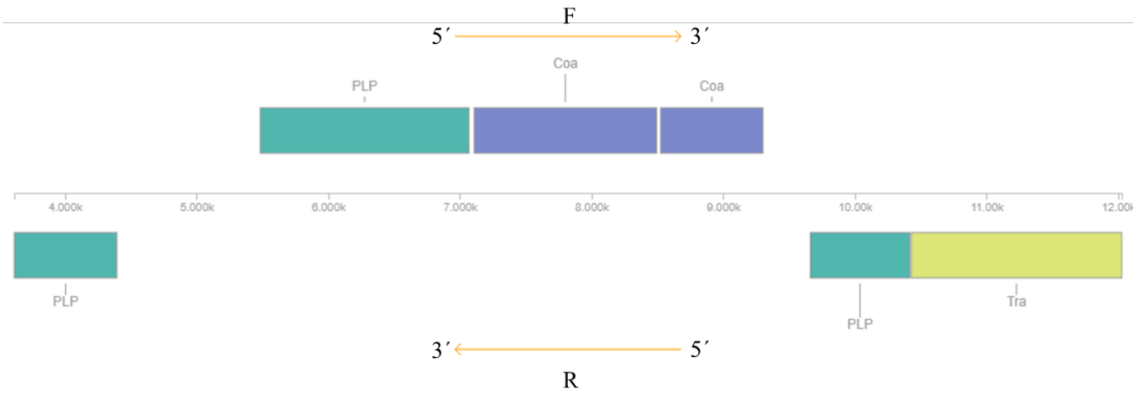

Figure 4.1.5 Predicted organization of the putative Agrobacterium phage Atu ph07 on node 35 of B. dorei D1P5 region 2. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): cobyrinic acid ac-diamide synthase (PLP), putative ArdC-like antirestriction protein (PLP), major capsid protein (Coa), major capsid protein (Coa), IS21 transposition protein (PLP), IS21 transposase (Tra). PLP is Prophage-like protein.

4.1.5 Prophages in *B. dorei* D8M1

B. dorei D8M1 isolated from a subject with T1D exhibit three putative incomplete prophage-like regions, region 1, 2 and 3 with completeness scores of 30, 50 and 10, respectively.

Region 1 (**Figure 4.1.6**) is present in scaffold 4 from 532927 bp to 560790 bp, with 27.8 total kb length. Sequence analysis of region 1 identified similarity with *Clostridium* phage c-st (16.66% shared orthologous proteins) (Sakaguchi *et al.*, 2005), with 12 ORFs and 4 attachment sites in forward position. Two attachment sites were identified: attL (11 bp each length), one in position 532927 bp to 532938 bp and the other in 540207 bp to 540218 bp, oth are separated by 7269 bp with no prophage-like genes detected between them. Both are flanking the genes in forward position of the sequence at the 5´-end. Similarly, two attachment sites attR (11 bp each length) were identified in positions 558850 bp to 558861 bp, and 559910 bp to 559921 bp, separated by 1049 bp with an integrase gene between them, and a gp582 phage-like protein flanking the 3´-end of one attR (559910 bp to 559921 bp). Coding genes for two integrases (806 bp length in the reverse position and a 935 bp length in the forward position, respectively), putative restriction-modification protein (1508 bp), DEAD/DEAH box helicase (2318 bp), and gp582 phage-like protein (680 length). In the prophage-like region no tRNA was encountered.

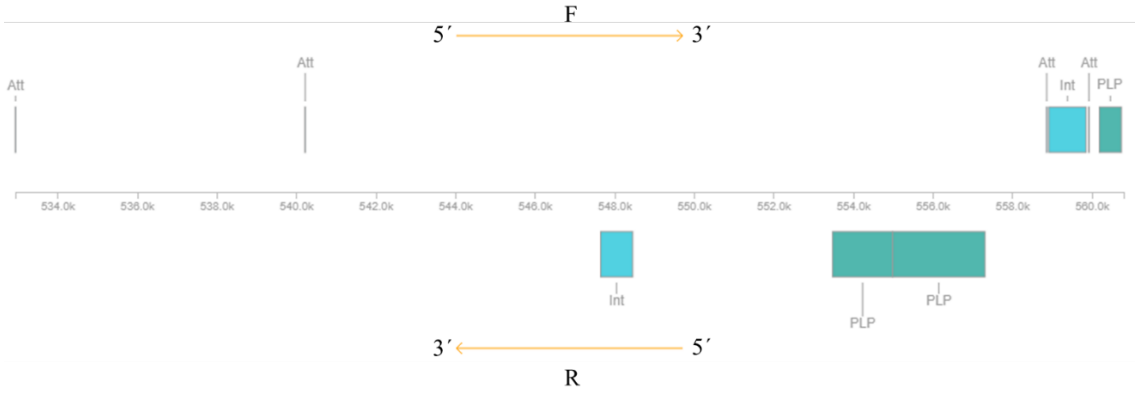

Figure 4.1.6 Predicted organization of the putative Clostridium phage c-st on scaffold 4 of B. dorei D8M1 region 1. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): attL (Att), attL (Att), integrase (Int), putative restriction-modification protein (PLP), DEAD/DEAH box helicase (PLP), attR (Att), integrase (Int), attR (Att), gp582 phage-like protein (PLP). PLP is Prophage-like protein.

Region 2 (**Figure 4.1.7**) is located in scaffold 5 from 80878 bp to 89844 bp with 8.9 kb total length. It presents similarity with the sequence of *Flavobacterium* phage 1H (Castillo and Middelboe, 2016) (7.69% shared orthologous proteins)**,** with 13 ORFs and 2 attachment sites in the forward position. Attachment site attL (12 bp length) from 80878 bp to 80890 bp and attR (12 bp length) from 89844 bp to 89856 bp, both flanking prophage like genes. The identified coding genes were for an integrase (1229 bp length), a probable phage-family integrase (1244 bp), a putative antirepressor (521 bp), a putative anti-repressor protein (203 bp), a putative endolysin (314 bp), and a putative endolysin/secretion activator protein (179 bp). In the prophage-like region no tRNA was encountered.

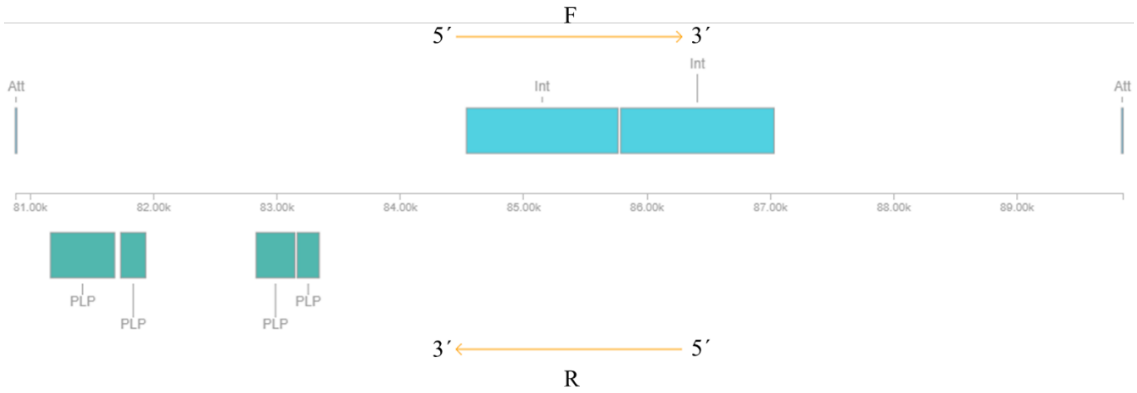

Figure 4.1.7 Predicted organization of the putative Flavobacterium phage 1H on scaffold 5 of B. dorei D8M1 region 2. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): attL (Att), putative antirepressor (PLP), putative anti-repressor protein (PLP), putative endolysin (PLP), putative endolysin/secretion activator protein (PLP), integrase (Int), probable phage-family integrase (Int), attR (Att).

In scaffold 8, PHASTER identified region 3 (**Figure 4.1.8**) ranging from 17386 bp to 28520 bp with 11.1 kb total length. This sequence presented high similarity with *Bacillus* phage phiAGATE (Barylski *et al.*, 2013) (25% shared orthologous proteins), with 8 ORFs and no attachment sites. Genes coding for DnaE-like DNA polymerase III alpha subunit (2954 bp length), DNA polymerase III subunit alpha (1760 bp), DNA primase-helicase (1250 bp), gp6 phage-like protein (500 bp) and putative exonuclease 1 (2222 bp) were all identified in the reverse strand. In the prophage-like region no tRNA was encountered.

Figure 4.1.8 Predicted organization of the putative Bacillus phage phiAGATE on scaffold 8 of B. dorei D8M1 region 3. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): DnaE-like DNA polymerase III alpha subunit (PLP), DNA polymerase III subunit alpha (PLP), DNA primase-helicase (PLP), gp6 phage-like protein (PLP), putative exonuclease 1 (PLP). PLP is Prophage-like protein.

4.1.6 Prophages in *B. dorei* D16P1

In *B. dorei* D16P1 isolated from a subject with T1D, two putative incomplete phage regions were encountered, region 1 and 2 with incompleteness scores of 30 and 50, respectively.

Region 1 (**Figure 4.1.9**) is integrated in scaffold 7 from 131347 bp to 139601 bp with 8.2 kb total length. The region presented high similarity with *Riemerella* phage RAP44 (Cheng et al., 2012) (41.17% shared orthologous proteins), with 17 ORFs and no attachment sites present. The annotated coding genes on this region are for DNA-binding protein (362 bp length), DUF2136 superfamily protein (299 bp), phage portal protein (1487 bp), phage terminase large subunit (1136 bp), ParB-like nuclease (602 bp), and adenine nucleotide alpha hydrolase-like protein (689 bp), all present in the reverse strand. In the prophage-like region no tRNA was encountered.

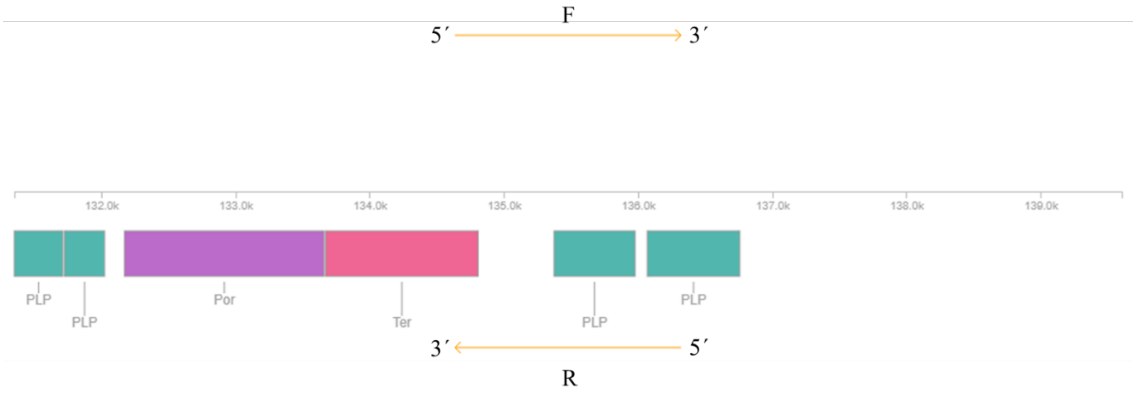

Figure 4.1.9 Predicted organization of the putative Riemerella phage RAP44 on scaffold 7 of B. dorei D16P1 region 1. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): DNA-binding protein (PLP), DUF2136 superfamily protein (PLP), phage portal protein (Por), phage terminase large subunit (Ter), ParB-like nuclease (PLP), adenine nucleotide alpha hydrolase-like protein (PLP). PLP is Prophage-like protein.

Scaffold 12 shows an incomplete prophage-like region, region 2 (**Figure 4.1.10**) from 2.289 bp to 49.798 bp with 47.5 kb total length. Consistent similarity of the sequence with *Flavobacterium* phage 2A (Castillo and Middelboe, 2016) (22.91% shared orthologous proteins), with 48 ORFs and four attachment sites in the forward position. Two attachment sites attL (12 bp length each), one positioned in 6651 bp to 6663 bp and the other in 16527 bp to 16539 bp separated by 9864 bp with a hypothetical protein between them. The other two attachment sites are attR (12 bp length each), the first is located at 34040 bp to 34052 bp, and the second in 49798 bp to 49810 bp, separated by 15746 bp presenting four hypothetical proteins between them. Coding genes identified represent a DNA adenine methylase (839 bp length), and two integrases (1106 bp and 1577 bp length, respectively), all in the reverse strand. In the prophage-like region no tRNA was encountered.

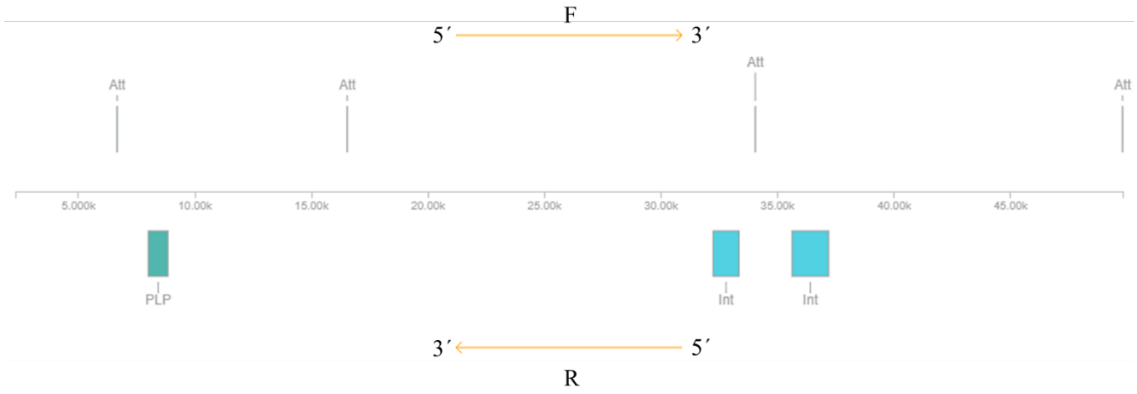

Figure 4.1.10 Predicted organization of the putative Flavobacterium phage 2A on scaffold 12 of B. dorei D16P1 region 2. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): attL (Att), DNA adenine methylase (PLP), attL (Att), integrase (Int), attR (Att), integrase (Int), and attR (Att).

4.1.7 Prophages in *B. dorei* D16M14

One putative incomplete prophage region was found in *B. dorei* D16M14 isolated from a T1D children. Region 1 (**Figure 4.1.11**) is integrated in scaffold 13 with an incompleteness score of 50, ranging from 2.289 bp to 49.798 bp with 47.5 kb total length. Sequence similarity matched *Flavobacterium* phage 2A (Castillo and Middelboe, 2016) (22.91% shared orthologous proteins), 48 ORFs and four attachment sites in the forward position. Two attachment sites attL (with 12 bp length each), one positioned in 6651 bp to 6663 bp, and the other in 16527 bp to 16539 bp, separated by 9864 bp with an hypothetical protein between. The other two attR attachment sites are 12 bp length each, the first being located at 34040 bp to 34052 bp, and the second in 49.798 bp to 49.810 bp, separated by 15746 bp with four hypothetical proteins identified between them. Coding genes were found for DNA adenine methylase (with 839 bp length) and two integrases (1106 bp and 1577 bp length, respectively), all in the reverse position. In the prophage-like region no tRNA was encountered.

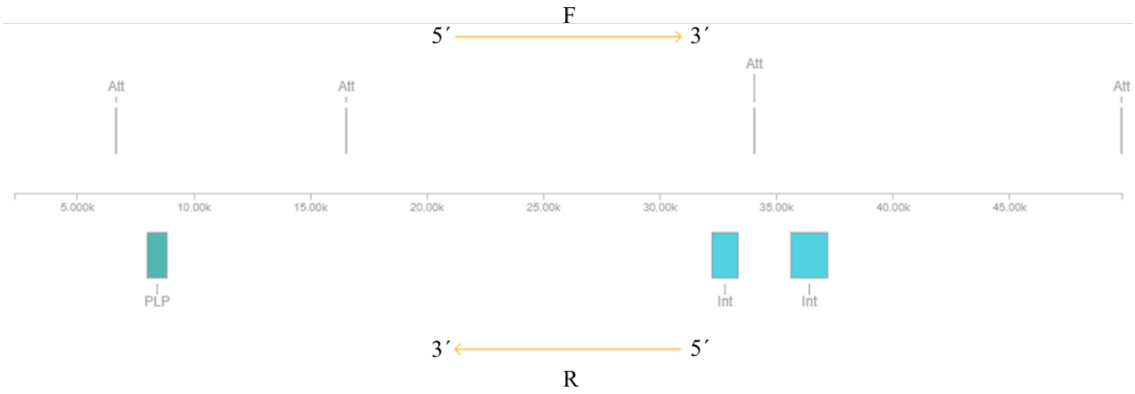

Figure 4.1.11 Predicted organization of the putative Flavobacterium phage 2A on scaffold 13 of B. dorei D16M14 region 1. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): attL (Att), DNA adenine methylase (PLP), attL (Att), integrase (Int), attR (Att), integrase (Int), and attR (Att).

4.1.8 Prophages in *P. distasonis* D14MH1

In *Parabacteroides distasonis* D14MH1 isolated from a T1D children two putative incomplete prophage regions were detected, both in scaffold 3 separated by 382430 bp: region 1 and 2 with incompleteness scores of 10 and 20, respectively.

Region 1 (**Figure 4.1.12**) is located from 125353 bp to 134508 bp with 9.1 kb total length. Significant sequence similarity to *Synechococcus* phage S-SKS1 (Henn *et al.*, 2013) (25% shared orthologous proteins) with 8 ORFs and no attachment sites. Coding genes identified for gp344 phage-like protein (1379 bp length), 6-phosphogluconate dehydrogenase (1460 bp), glucose 6-phosphate dehydrogenase (1466 bp), gp066 phage-like protein (875 bp), and nucleotide-sugar epimerase (1076 bp), all in the forward strand. In the prophage-like region no tRNA was encountered.

Figure 4.1.12 Predicted organization of the putative Synechococcus phage S-SKS1 on scaffold 3 of P. distasonis D16MH1 region 1. Picture was retrieved and adapted from the PHASTER results (Arndt et al.,

2016). Genes with annotated function in viral physiology (from left to right): gp344 phage-like protein (PLP), 6-phosphogluconate dehydrogenase (PLP), glucose 6-phosphate dehydrogenase (PLP), gp066 phage-like protein (PLP), nucleotide-sugar epimerase (PLP). PLP is Prophage-like protein.

Region 2 (**Figure 4.1.13**) ranges from coordinates 516938 bp to 527541 bp, presenting a total length of 10.6 kb. Similarity of the sequence to *Cellulophaga* phage phi39:1 (Holmfeldt *et al.*, 2013) (13.33% shared orthologous proteins) with fifteen ORFs and no attachment sites. Identified genes code for DNA-cytosine methyltransferase (932 bp length), two HNH endonucleases (512 bp and 572 bp length, respectively) separated by a recombinase (with 683 bp) and gp201 phage-like protein (1886 bp), all in the reverse strand. In the prophage-like region no tRNA was encountered.

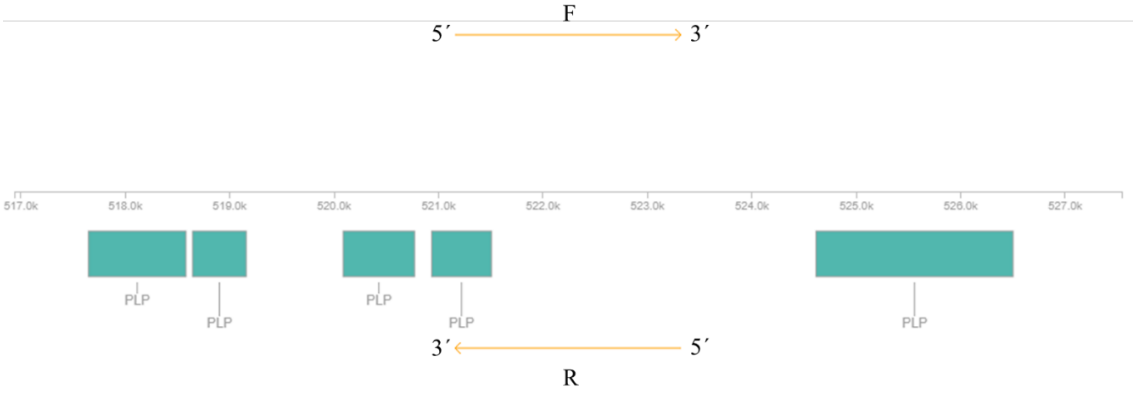

Figure 4.1.13 Predicted organization of the putative Cellulophaga phage phi39:1 on scaffold 3 of P. distasonis D16MH1 region 2. Picture was retrieved and adapted from the PHASTER results (Arndt et al., 2016). Genes with annotated function in viral physiology (from left to right): DNA-cytosine methyltransferase (PLP), HNH endonuclease (PLP), recombinase (PLP), HNH endonuclease (PLP), gp201 phage-like protein (PLP). PLP is Prophage-like protein.

Table 4.1.1 Bacterial species and strains used in this study with prophage-like regions identified. *PHASTER score. prophage similarity identification.

Host	Region	Region Length (kb)	GC%	Left boundary (bp)	Right boundary (bp)	ORFs	Completeness	Score*
Bacteroides dorei C1P2	1	7.4	50.27%	139060	146553	7	Incomplete	50
Bacteroides dorei Sb6	1	11.1	37.27%	188800	199923	21	Incomplete	10
Bacteroides dorei Sb8	1	8.9	44.32%	120653	129619	13	Incomplete	50
Bacteroides dorei D1P5	1	11.1	37.27%	188800	199923	21	Incomplete	10
	2	8.4	47.57%	3612	12023	10	Incomplete	40
Bacteroides dorei D8M1	1	27.8	37.48%	532927	560790	12	Incomplete	30
	2	8.9	44.32%	80878	89844	13	Incomplete	50
	3	11.1	36.44%	17386	28520	8	Incomplete	10
	1	8.2	38.27%	131347	139601	17	Incomplete	30
Bacteroides dorei D16P1	2	47.5	42.05%	2289	49798	48	Incomplete	50
Bacteroides dorei D16M14	1	47.5	42.05%	2289	49798	48	Incomplete	50
	1	9.1	48.53%	125353	134508	8	Incomplete	10
Parabacteroides distasonis D14MH1	2	10.6	44.88%	516938	527541	15	Incomplete	20

Table 4.1.2 Bacterial species and strains used in this study with prophage similarity identification.

Host	Region	Most Similar Phage	Order	Family	% of protein similarity	AttL	AttR
Bacteroides dorei C1P2	1	Mannheimia phage vB_MhM_3927AP2	*Caudovirales*	*Myoviridae*	14.28%		
Bacteroides dorei Sb6	1	Riemerella phage RAP44	*Caudovirales*	*Siphoviridae*	4.76%		
Bacteroides dorei Sb8	1	Flavobacterium phage 1H	*Caudovirales*	*Siphoviridae*	7.69%	1	1
	1	Riemerella phage RAP44	*Caudovirales*	*Siphoviridae*	4.76%		
Bacteroides dorei D1P5	2	Agrobacterium phage Atu ph07	*Caudovirales*	*Myoviridae*	20%		
	1	Clostridium phage c-st	*Caudovirales*	*Siphoviridae*	16.66%	2	2
	2	Flavobacterium phage 1H	*Caudovirales*	*Siphoviridae*	7.69%	1	1
Bacteroides dorei D8M1	3	Bacillus phage phiAGATE	*Caudovirales*	*Herelleviridae*	25%		
	1	Riemerella phage RAP44	*Caudovirales*	*Siphoviridae*	41.17%		
Bacteroides dorei D16P1	2	Flavobacterium phage 2A	*Caudovirales*	*Siphoviridae*	22.91%	2	2
Bacteroides dorei D16M14	1	Flavobacterium phage 2A	*Caudovirales*	*Siphoviridae*	22.91%	2	2
	1	Synechococcus phage S-SKS1	*Caudovirales*	*Myoviridae*	25%		
Parabacteroides distasonis D14MH1	2	Cellulophaga phage phi39:1	*Caudovirales*	*Siphoviridae*	13.33%		

4.2 Prophage Induction

To evaluate if any complete or any prophage elements integrated in each tested bacterial genome could trigger the phage lytic cycle under specific stress conditions, namely mitomycin C (Mit c), hydrogen peroxide and ox-bile. A steep reduction in optical density can be associated to a possible prophage induction. The above bioinformatic results in combination with these induction assays will provide information about the existence of temperate phages in the *Bacteroides* spp. and under which conditions the temperate phage switch to the lytic cycle.

4.2.1 Induction of prophages with Mitomycin C

Mitomycin C is a chemotherapeutic agent that can cause DNA damage (Siegel *et al.*, 1990), being a general signal for prophage induction (Fortier and Moineau, 2007).

4.2.1.1 Mitomycin C in the strain *B. dorei* C1P2

After 4 h of induction of phages in the strain *B. dorei* C1P2 with Mitomycin C (1-7 µg/mL) a growth inhibition was observed but no prophage induction was noticeable in any inducer concentrations (**Figure 4.2.1**).

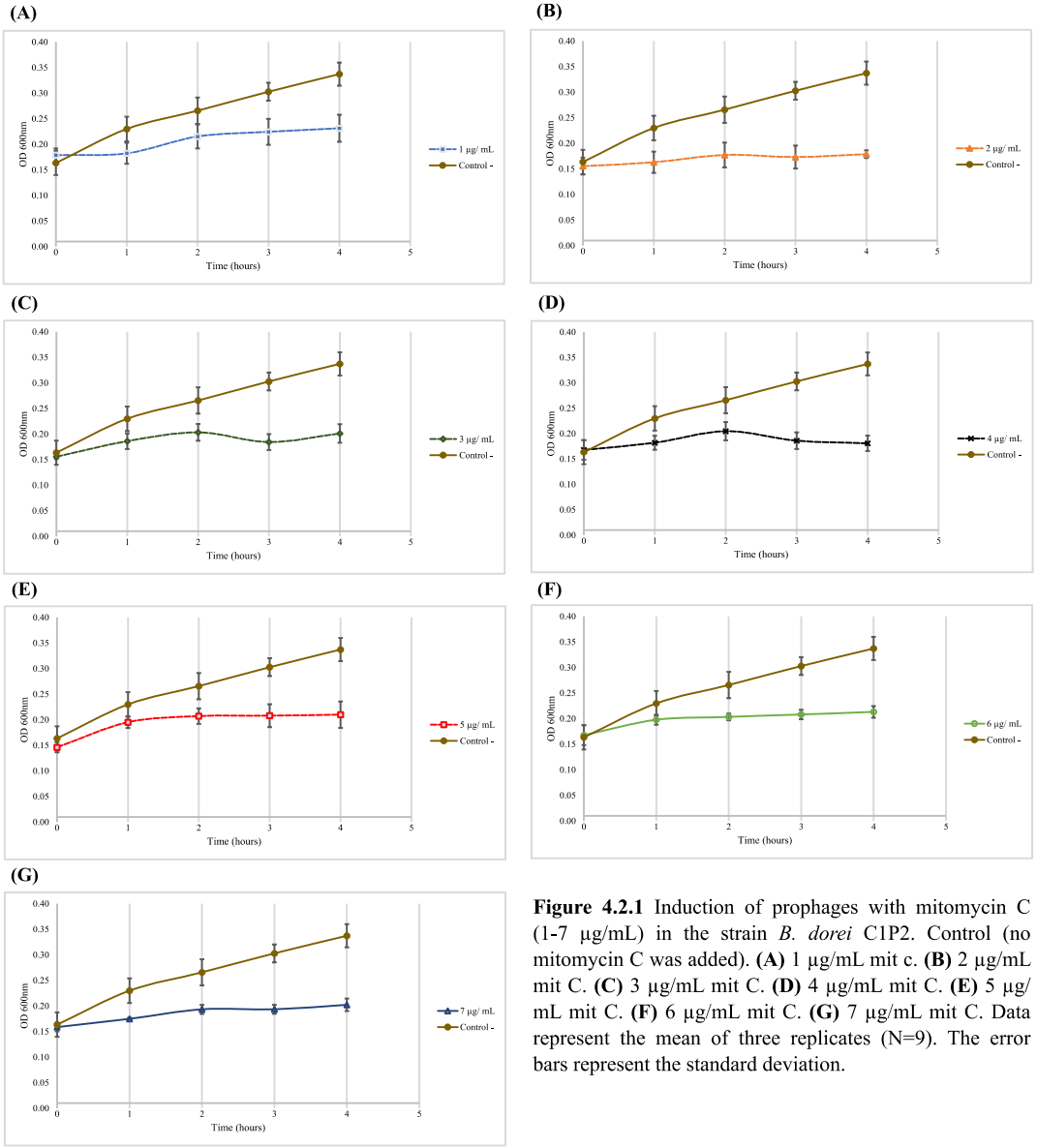

Figure 4.2.1 Induction of prophages with mitomycin C (1-7 µg/mL) in the strain *B. dorei* C1P2. Control (no mitomycin C was added). **(A)** 1 µg/mL mit c. **(B)** 2 µg/mL mit C. **(C)** 3 µg/mL mit C. **(D)** 4 µg/mL mit C. **(E)** 5 µg/mL mit C. **(F)** 6 µg/mL mit C. **(G)** 7 µg/mL mit C. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.1.2 Mitomycin C in the strain *B. dorei* Sb8

The exposure of the strain *B. dorei* Sb8 to mitomycin C resulted in bacterial growth inhibition at all concentrations (**Figure 4.2.2**). No prophage induction was observed after four hours of induction with mitomycin C (**Figure 4.2.2**).

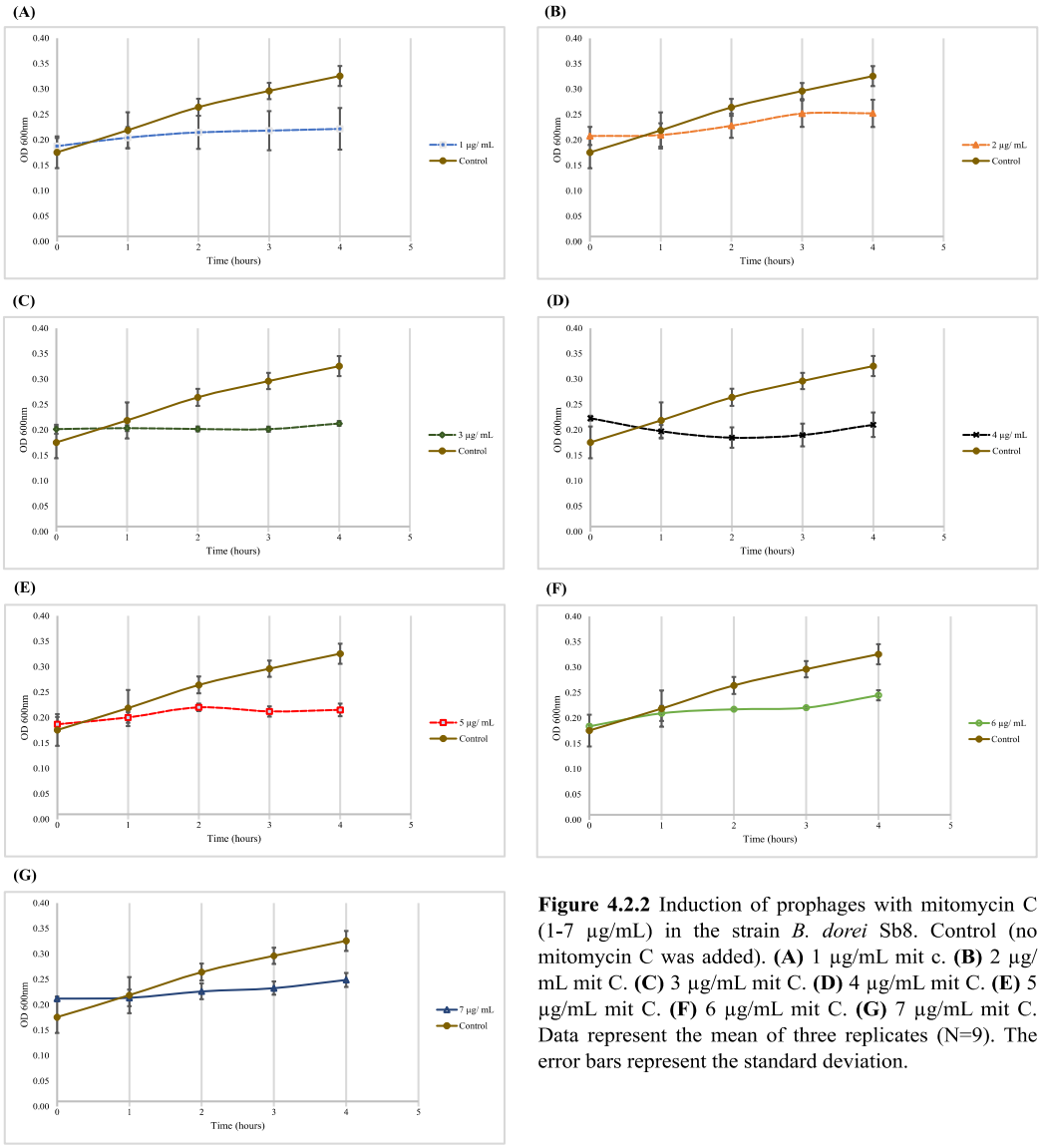

Figure 4.2.2 Induction of prophages with mitomycin C (1-7 µg/mL) in the strain *B. dorei* Sb8. Control (no mitomycin C was added). **(A)** 1 µg/mL mit c. **(B)** 2 µg/ mL mit C. **(C)** 3 µg/mL mit C. **(D)** 4 µg/mL mit C. **(E)** 5 µg/mL mit C. **(F)** 6 µg/mL mit C. **(G)** 7 µg/mL mit C. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.1.3 Mitomycin C in the strain *B. dorei* D1P5

The growth of the strain *B. dorei* D1P5 in the presence of mitomycin C (1-7 µg/mL) was inhibited at all concentrations tested (**Figure 4.2.3**). However, at 6 µg/ mL the decrease in the OD_{600nm} is noticeable which suggest prophage induction at this concentration (**Figure 4.2.3**).

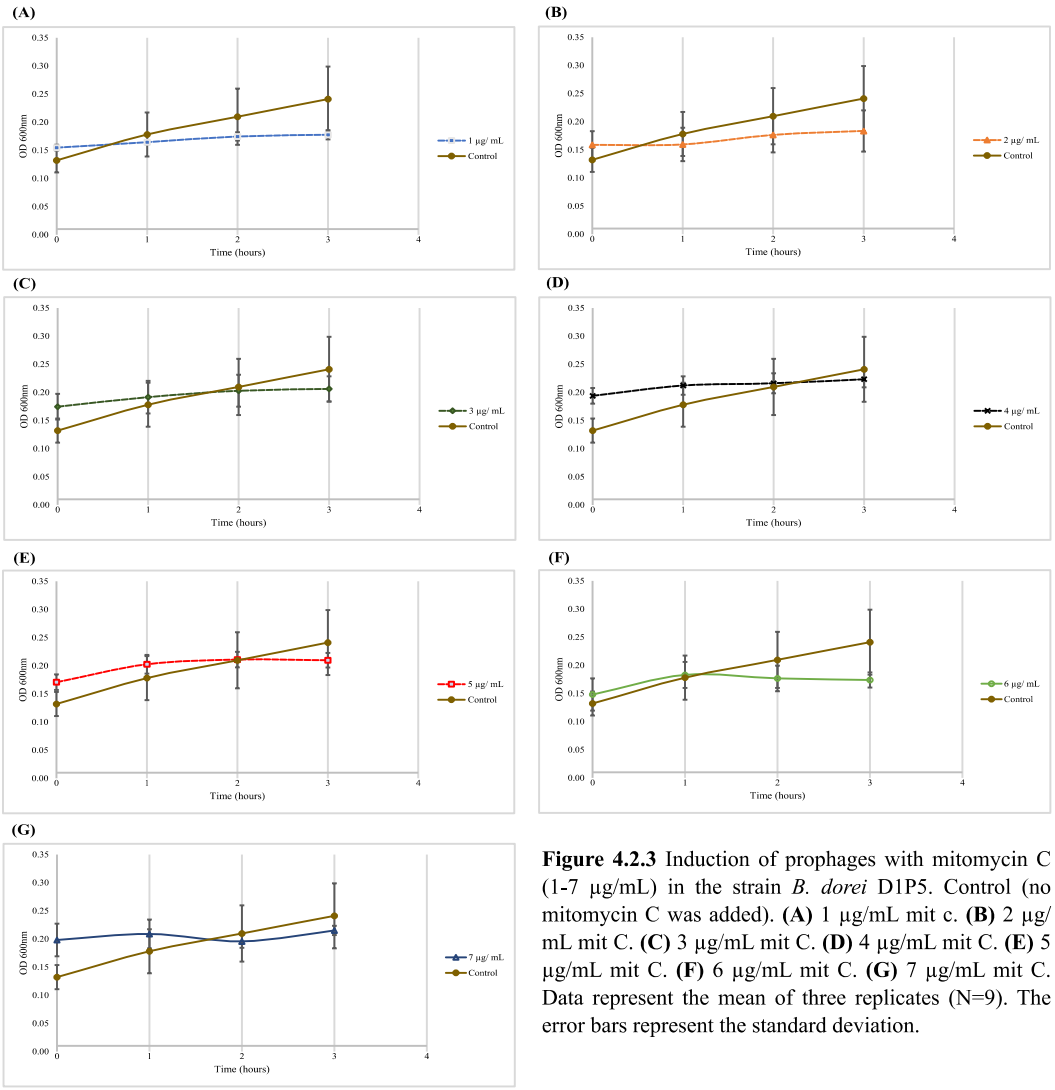

Figure 4.2.3 Induction of prophages with mitomycin C (1-7 µg/mL) in the strain *B. dorei* D1P5. Control (no mitomycin C was added). **(A)** 1 µg/mL mit c. **(B)** 2 µg/ mL mit C. **(C)** 3 µg/mL mit C. **(D)** 4 µg/mL mit C. **(E)** 5 µg/mL mit C. **(F)** 6 µg/mL mit C. **(G)** 7 µg/mL mit C. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.1.4 Mitomycin C in the strain *B. dorei* D8M1

The exposure of the strain *B. dorei* D8M1 to mitomycin C at 1 µg/mL and 2 µg/mL resulted in a marked decrease of the OD_{600nm} suggesting a potential prophage induction at these two concentrations, but particularly at 2 µg/mL (**Figure 4.2.4**). The remaining mitomycin C concentrations just inhibited the bacterial growth but did not cause any noticeable decrease of the OD_{600nm} values after 3 h of induction (**Figure 4.2.4**).

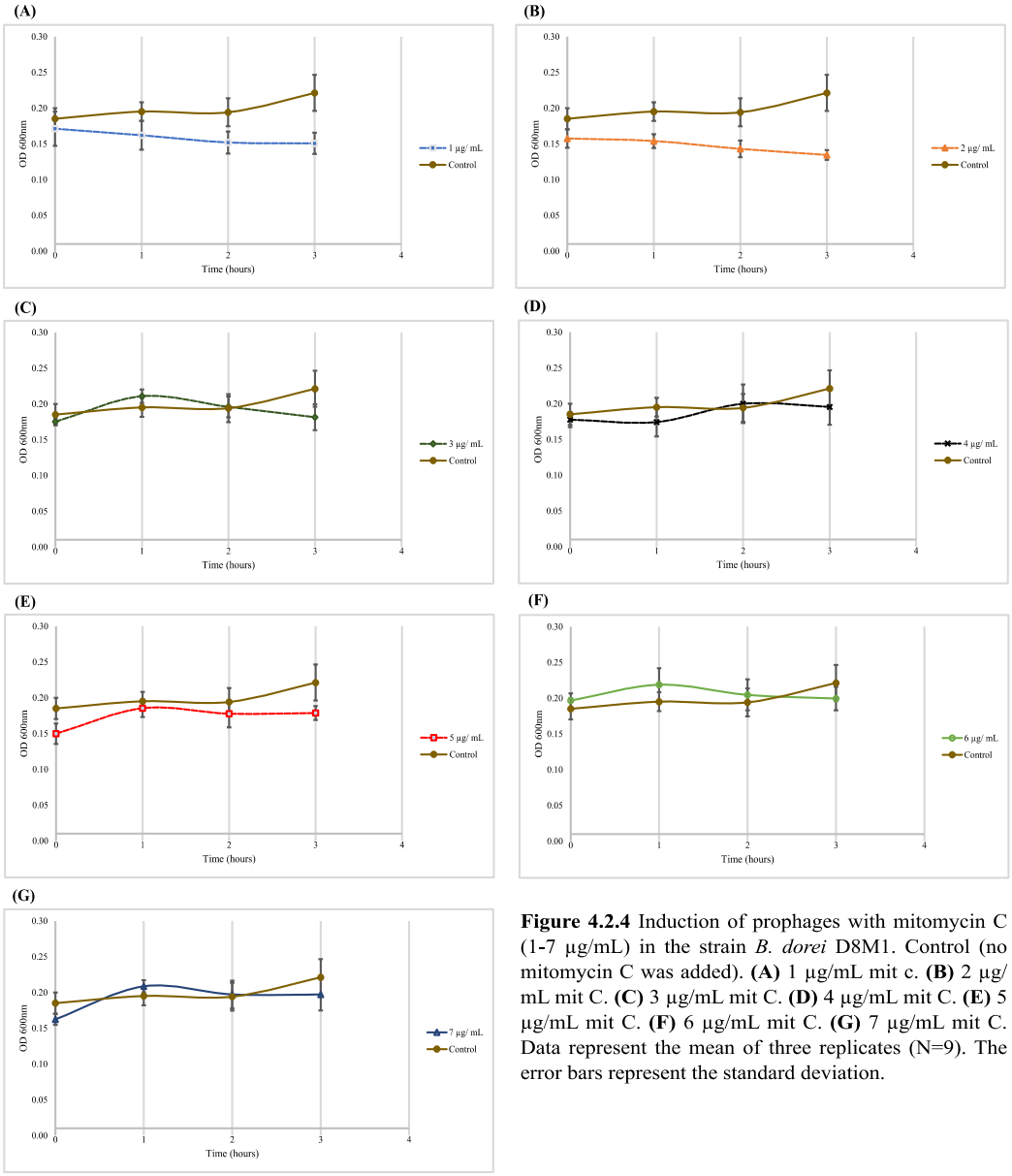

Figure 4.2.4 Induction of prophages with mitomycin C (1-7 µg/mL) in the strain *B. dorei* D8M1. Control (no mitomycin C was added). **(A)** 1 µg/mL mit c. **(B)** 2 µg/ mL mit C. **(C)** 3 µg/mL mit C. **(D)** 4 µg/mL mit C. **(E)** 5 µg/mL mit C. **(F)** 6 µg/mL mit C. **(G)** 7 µg/mL mit C. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.1.5 Mitomycin C in the strain *B. dorei* D16P1

The treatment of the strain *B. dorei* D16P1 with mitomycin C showed that at 1 µg/ mL a marked OD_{600nm} is observed after 5 h of induction. A slight OD_{600nm} decrease is also observed at 2 µg/mL after the same time interval (**Figure 4.2.5**). These results suggest a possible prophage induction, particularly at 1 µg/mL.

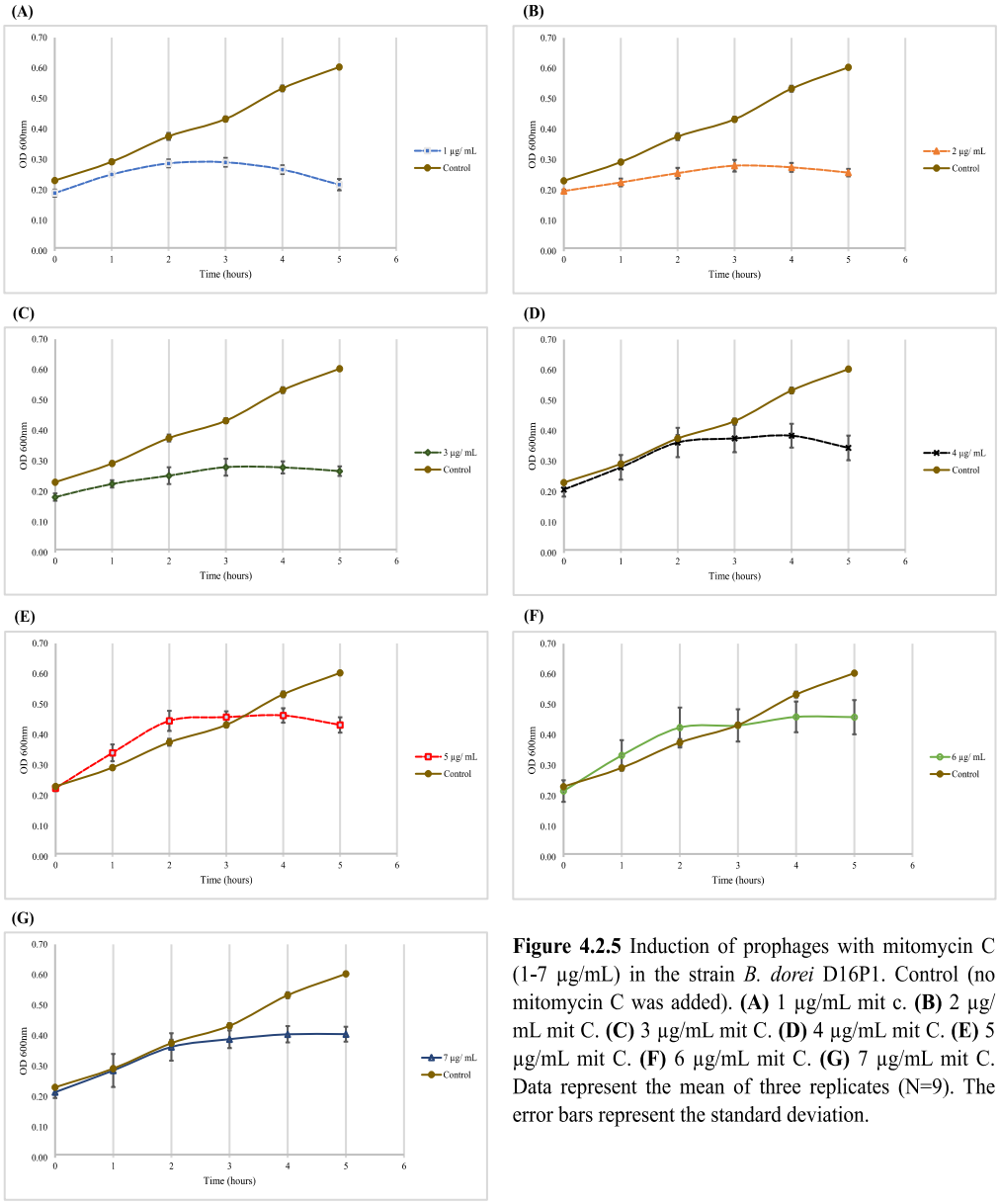

Figure 4.2.5 Induction of prophages with mitomycin C (1-7 µg/mL) in the strain *B. dorei* D16P1. Control (no mitomycin C was added). **(A)** 1 µg/mL mit c. **(B)** 2 µg/ mL mit C. **(C)** 3 µg/mL mit C. **(D)** 4 µg/mL mit C. **(E)** 5 µg/mL mit C. **(F)** 6 µg/mL mit C. **(G)** 7 µg/mL mit C. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.1.6 Mitomycin C in the strain *B. dorei* D16M14

The exposure of the strain *B. dorei* D16M14 to mitomycin C (1-7 µg/mL) inhibited the bacterial growth but the induction of prophages was not observed in any of the concentrations tested (**Figure 4.2.6**).

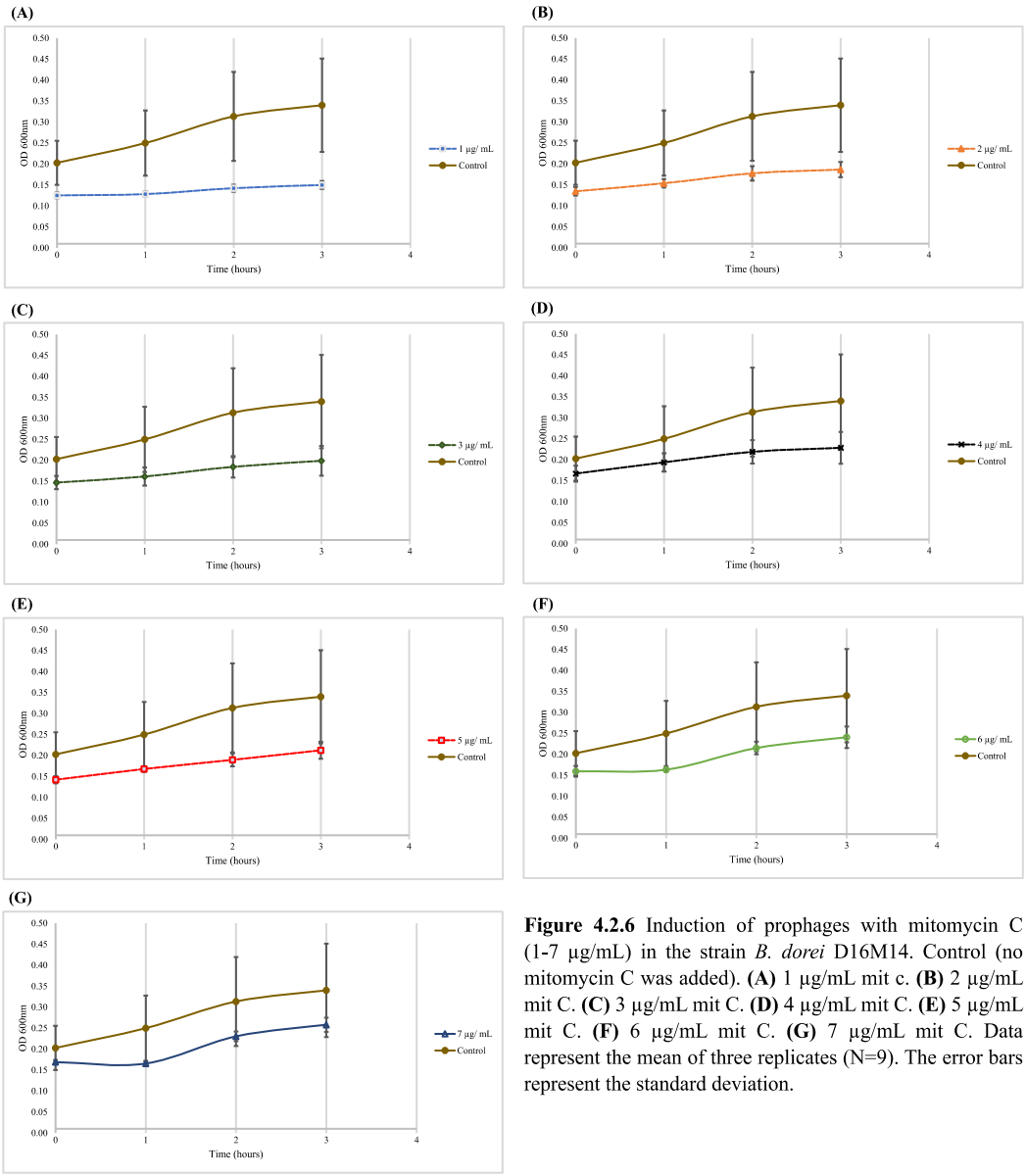

Figure 4.2.6 Induction of prophages with mitomycin C (1-7 µg/mL) in the strain *B. dorei* D16M14. Control (no mitomycin C was added). **(A)** 1 µg/mL mit c. **(B)** 2 µg/mL mit C. **(C)** 3 µg/mL mit C. **(D)** 4 µg/mL mit C. **(E)** 5 µg/mL mit C. **(F)** 6 µg/mL mit C. **(G)** 7 µg/mL mit C. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.1.7 Mitomycin C in the strain *P. distasonis* D14MH1

The growth of *P. distasonis* D14MH1 in the presence of mitomycin C (1-7 µg/mL) was inhibited at all concentrations tested (**Figure 4.2.7**). The bacterial cultures exposed to the different antibiotic concentrations achieved similar OD_{600nm} values after 4h evidencing the absence of prophage induction (**Figure 4.2.7**).

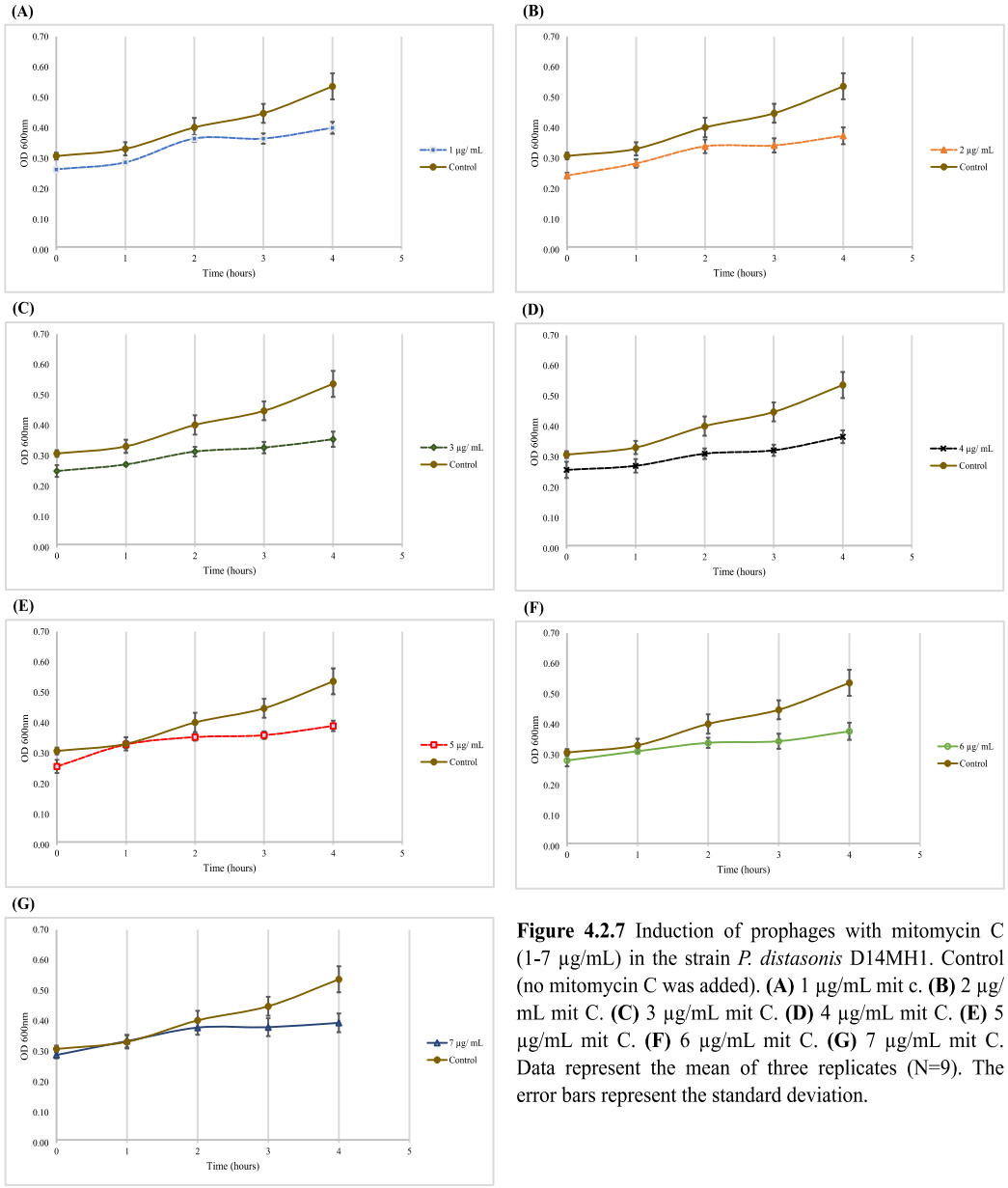

Figure 4.2.7 Induction of prophages with mitomycin C (1-7 µg/mL) in the strain *P. distasonis* D14MH1. Control (no mitomycin C was added). **(A)** 1 µg/mL mit c. **(B)** 2 µg/ mL mit C. **(C)** 3 µg/mL mit C. **(D)** 4 µg/mL mit C. **(E)** 5 µg/mL mit C. **(F)** 6 µg/mL mit C. **(G)** 7 µg/mL mit C. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.2 Induction of prophages with hydrogen peroxide

Hydrogen peroxide is known to cause oxidative stress affecting membrane lipids, proteins, RNA and DNA (Rocha *et al.*, 1996; Tamarit, Cabiscol and Ros, 1998), which may trigger the prophage lytic cycle.

4.2.2.1 Hydrogen peroxide in the strain *B. dorei* C1P2

The challenge of the strain *B. dorei* C1P2 with hydrogen peroxide (H_2O_2) (10 µM, 50 µM, 100 µM) slightly affected the growth of this strain (**Figure 4.2.8**). This result suggests the absence of prophage induction by H_2O_2.

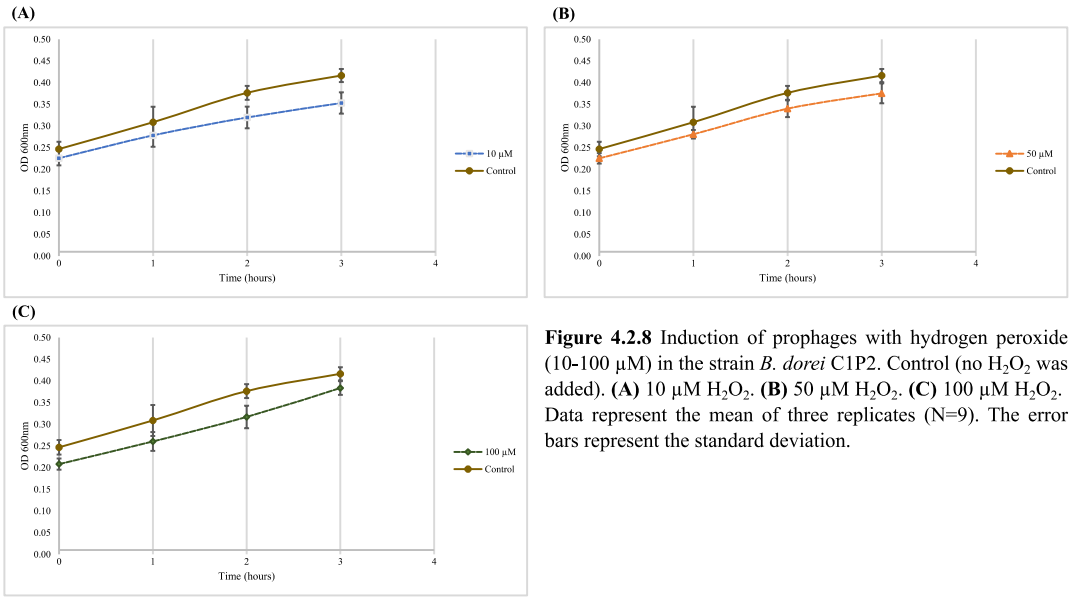

Figure 4.2.8 Induction of prophages with hydrogen peroxide (10-100 µM) in the strain *B. dorei* C1P2. Control (no H_2O_2 was added). **(A)** 10 µM H_2O_2. **(B)** 50 µM H_2O_2. **(C)** 100 µM H_2O_2. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.2.2 Hydrogen peroxide in the strain *B. dorei* Sb6

The exposure of the strain *B. dorei* Sb6 to H_2O_2 caused a slight growth inhibition, particularly at 100 µM concentration (**Figure 4.2.9**). No prophage induction was observed after 3 h induction with H_2O_2 at the concentrations tested (**Figure 4.2.9**).

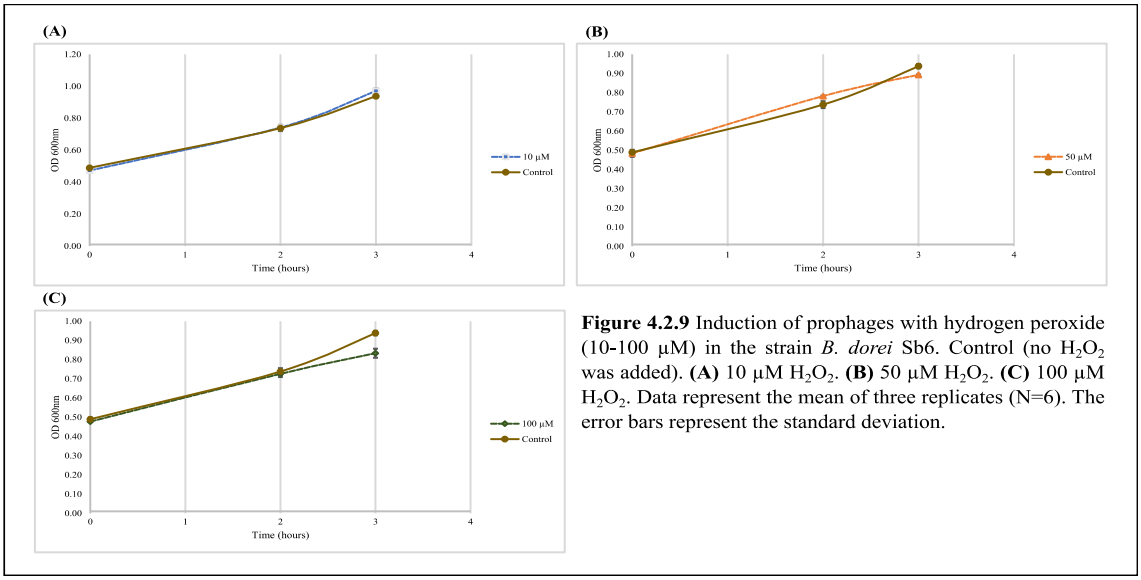

Figure 4.2.9 Induction of prophages with hydrogen peroxide (10-100 µM) in the strain *B. dorei* Sb6. Control (no H_2O_2 was added). **(A)** 10 µM H_2O_2. **(B)** 50 µM H_2O_2. **(C)** 100 µM H_2O_2. Data represent the mean of three replicates (N=6). The error bars represent the standard deviation.

4.2.2.3 Hydrogen peroxide in the strain *B. dorei* Sb8

The treatment of the cultures of *B. dorei* Sb8 with H_2O_2, particularly at 10 µM seems to induce a slight decrease of the OD_{600nm} after 3 h (**Figure 4.2.10**). The two remaining H_2O_2 concentrations (50 µM, 100 µM) inhibited the growth without causing any marked decrease on the OD_{600nm} values. The slight decrease of the OD_{600nm} observed after 3 h suggest a possible prophage induction (**Figure 4.2.10**).

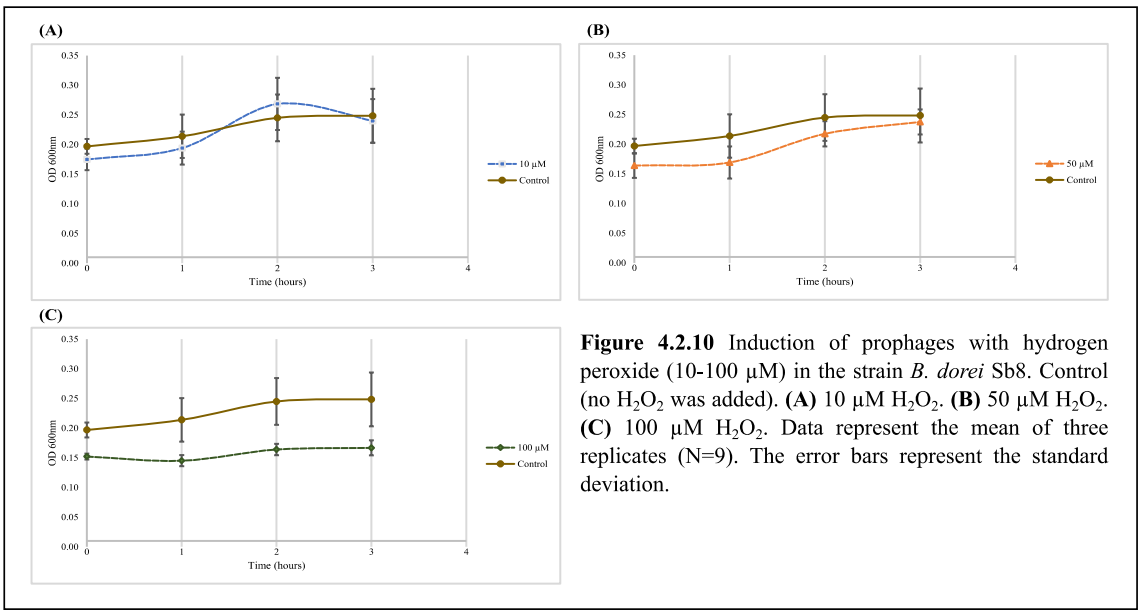

Figure 4.2.10 Induction of prophages with hydrogen peroxide (10-100 µM) in the strain *B. dorei* Sb8. Control (no H_2O_2 was added). **(A)** 10 µM H_2O_2. **(B)** 50 µM H_2O_2. **(C)** 100 µM H_2O_2. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.2.4 Hydrogen peroxide in the strain *B. dorei* D1P5

The strain *B. dorei* D1P5 was able to overcome the treatment with the different concentrations of H_2O_2 (10 µM, 50 µM, 100 µM) (**Figure 4.2.11**). After 4h of exposure to H_2O_2 the treated bacterial cultures achieved similar OD_{600nm} values to the Control.

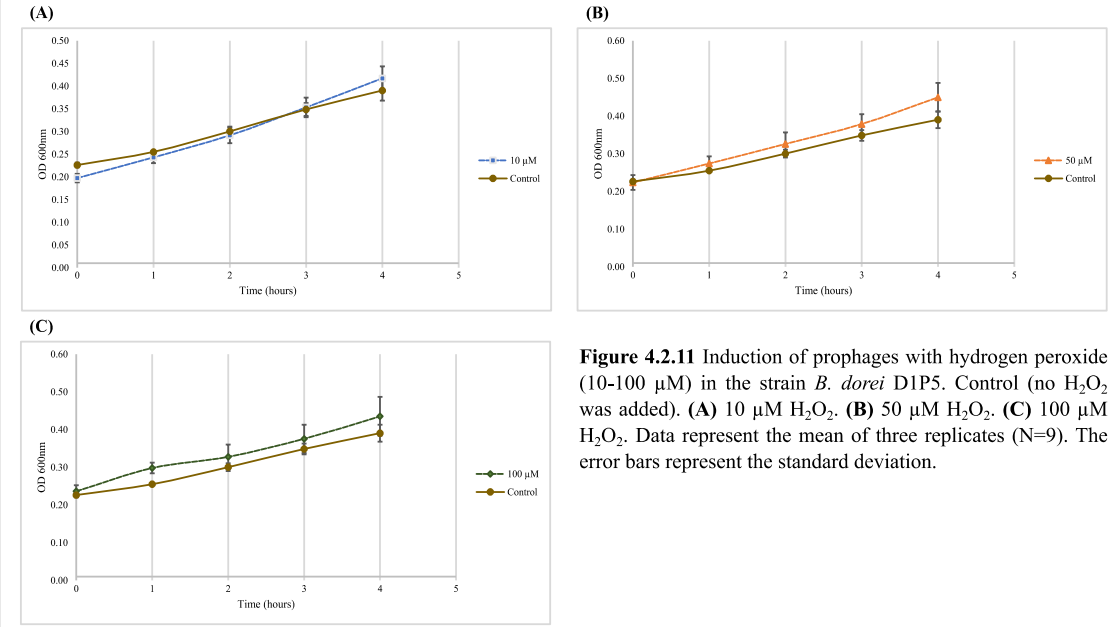

Figure 4.2.11 Induction of prophages with hydrogen peroxide (10-100 µM) in the strain *B. dorei* D1P5. Control (no H_2O_2 was added). **(A)** 10 µM H_2O_2. **(B)** 50 µM H_2O_2. **(C)** 100 µM H_2O_2. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.2.5 Hydrogen peroxide in the strain *B. dorei* D8M1

The exposure of the strain *B. dorei* D8M1 cultures to H_2O_2 at 50 µM and 100 µM resulted in a slight decrease of the OD_{600nm} values after 2 h, but after that time interval the bacterial culture was able to recover from the challenge (**Figure 4.2.12**).

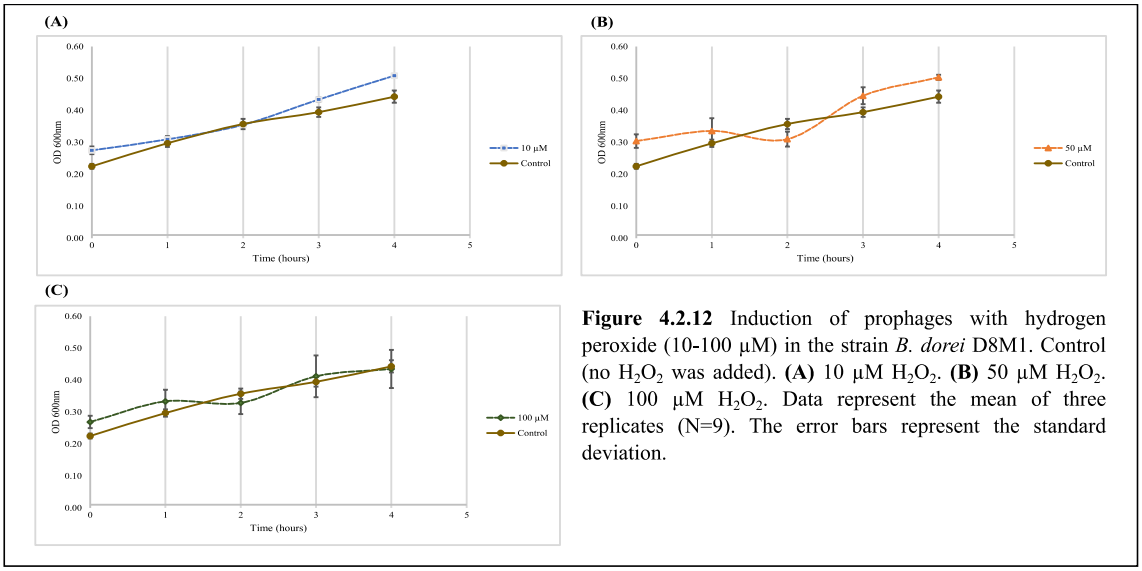

Figure 4.2.12 Induction of prophages with hydrogen peroxide (10-100 µM) in the strain *B. dorei* D8M1. Control (no H_2O_2 was added). **(A)** 10 µM H_2O_2. **(B)** 50 µM H_2O_2. **(C)** 100 µM H_2O_2. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.2.6 Hydrogen peroxide in the strain *B. dorei* D16P1

The growth of the strain *B. dorei* D16P1 exposed to the different concentrations of H_2O_2 (10 µM, 50 µM, 100 µM) was similar to the Control after 5 h evidencing the ability to overcome this stress condition without prophage induction (**Figure 4.2.13**).

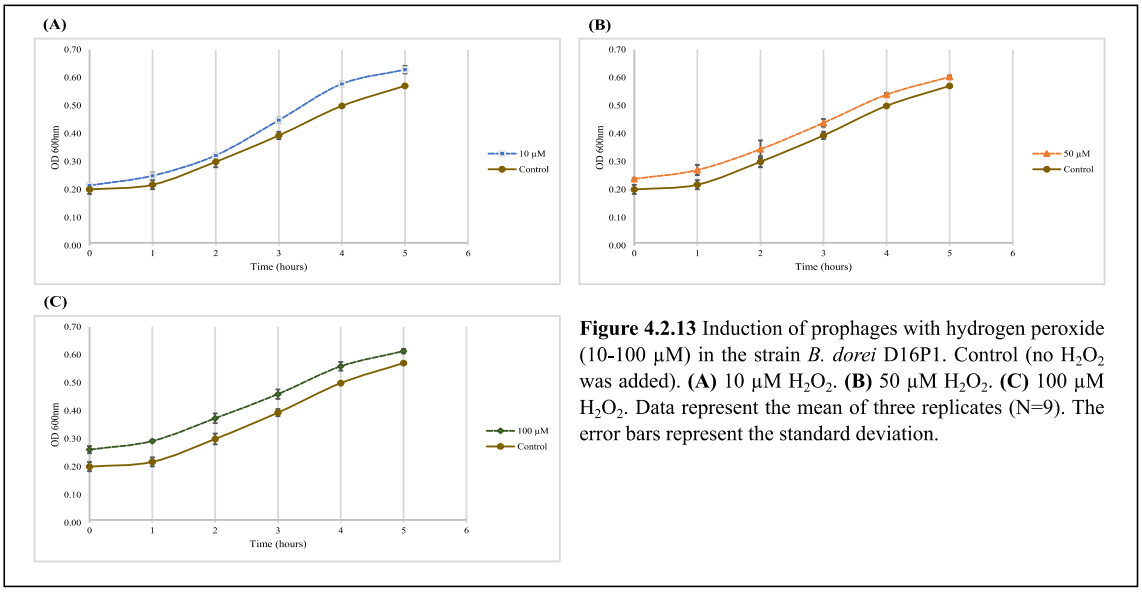

Figure 4.2.13 Induction of prophages with hydrogen peroxide (10-100 µM) in the strain *B. dorei* D16P1. Control (no H_2O_2 was added). **(A)** 10 µM H_2O_2. **(B)** 50 µM H_2O_2. **(C)** 100 µM H_2O_2. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.2.7 Hydrogen peroxide in the strain *B. dorei* D16M14

As observed with the strain D16P1 also for the strain *B. dorei* D16M14 the exposure to the different H_2O_2 concentrations (10 µM, 50 µM, 100 µM) did not affected the bacterial growth and showing no prophage induction (**Figure 4.2.14**).

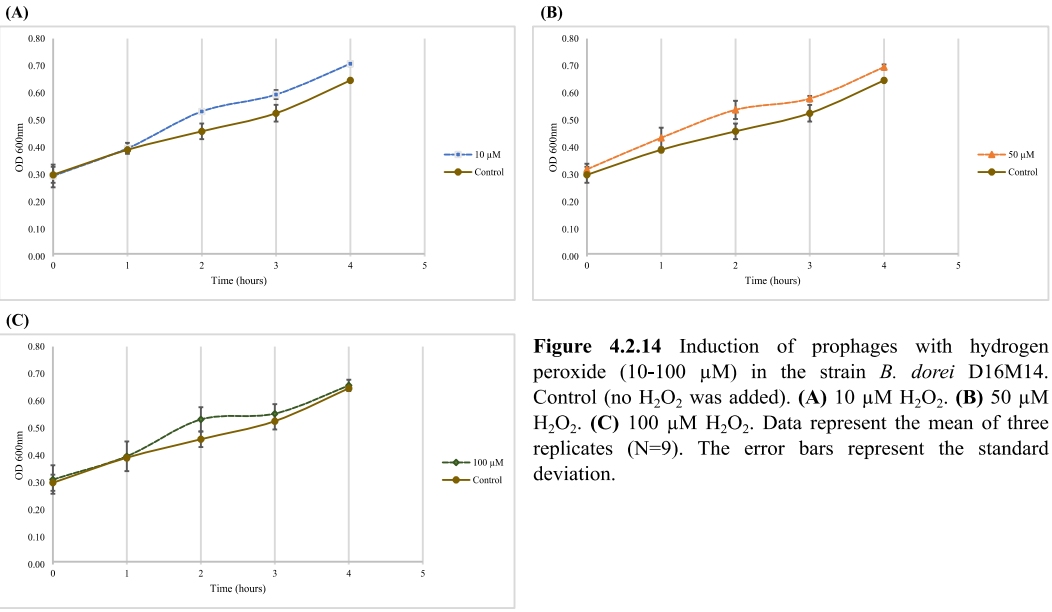

Figure 4.2.14 Induction of prophages with hydrogen peroxide (10-100 µM) in the strain *B. dorei* D16M14. Control (no H_2O_2 was added). **(A)** 10 µM H_2O_2. **(B)** 50 µM H_2O_2. **(C)** 100 µM H_2O_2. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.2.8 Hydrogen peroxide in the strain *P. distasonis* D14MH1

The growth of *P. distasonis* D14MH1 in the presence of the different concentrations of H_2O_2 was similar to the Control evidencing the absence of prophage induction (**Figure 4.2.15**).

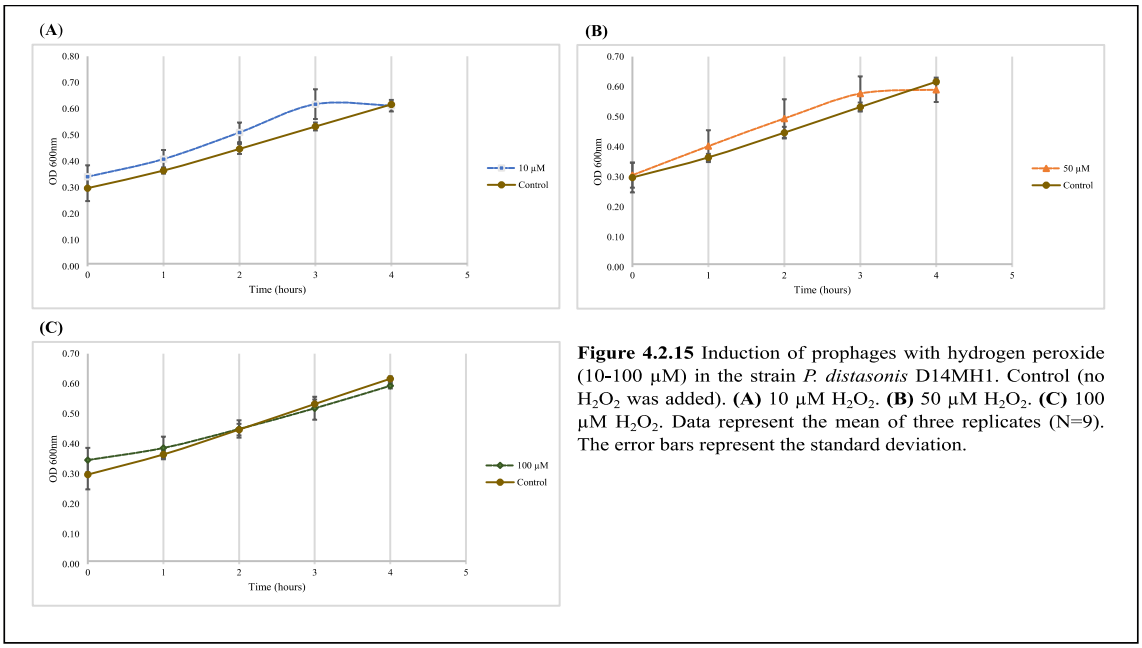

Figure 4.2.15 Induction of prophages with hydrogen peroxide (10-100 µM) in the strain *P. distasonis* D14MH1. Control (no H_2O_2 was added). **(A)** 10 µM H_2O_2. **(B)** 50 µM H_2O_2. **(C)** 100 µM H_2O_2. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.3 Induction of prophages with Ox-bile

As mentioned before bile has antibacterial activity and can induce DNA damage (Van Velkinburgh and Gunn, 1999; Gunn, 2000; Prieto *et al.*, 2004), and so the exposure of *Bacteroides* spp. to ox-bile may trigger prophage induction. This effect was examined, and the results are illustrated in **Figures 4.2.16 - 4.2.23**.

Ox-bile is a bile salt generally used for preparing microbiological culture media, but in certain concentrations can work as antibacterial agent causing DNA damage and provoking an SOS induction in a canonical, RecA-dependent type (Sutton *et al.*, 2000) as seen in *Salmonella enterica* (Prieto *et al.*, 2004). This bile acid can be a potential prophage inducer in this study.

4.2.3.1 Ox-bile in the strain *B. dorei* C1P2

The growth of the strain *B. dorei* C1P2 in the presence of ox-bile (0.2%, 0.4%) was slightly stimulated during 3 h but after 4 h the OD_{600nm} values were similar to the Control (**Figure 4.2.16**). The induction of prophages by ox-bile was not observed.

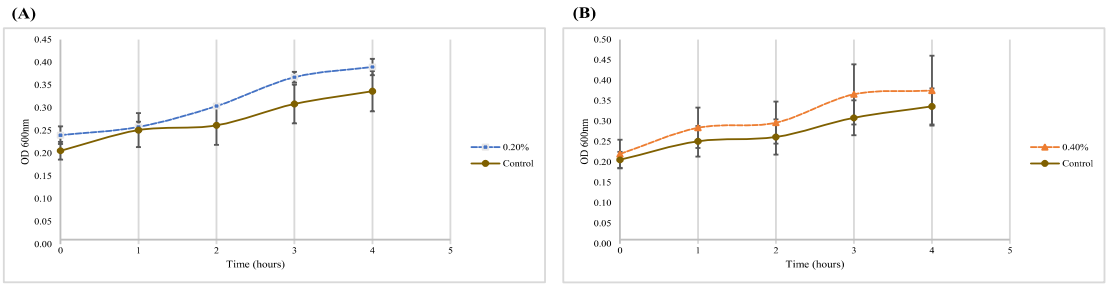

Figure 4.2.16 Induction of prophages with ox-bile (0.2-0.4%) in the strain *B. dorei* C1P2. Control (no ox-bile was added). **(A)** 0.2% ox-bile. **(B)** 0.4% ox-bile. Data represent the mean of three replicates (N=9). The error bars represent the standart.

4.2.3.2 Ox-bile in the strain *B. dorei* Sb6

The growth of the strain *B. dorei* Sb6 exposed to ox-bile (0.2%, 0.4%) was similar to the Control (**Figure 4.2.17**). The induction of prophages by ox-bile was not observed.

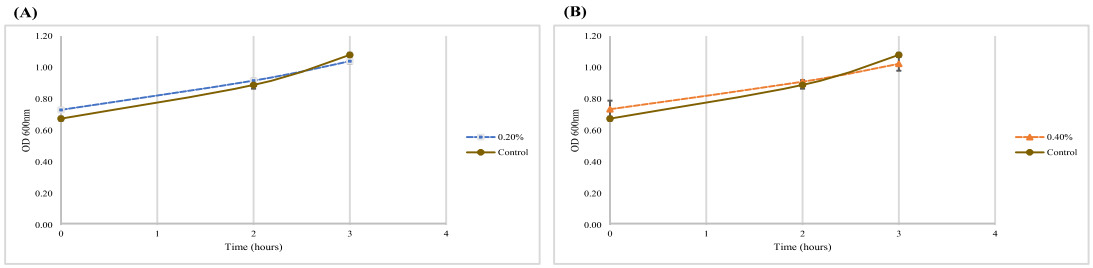

Figure 4.2.17 Induction of prophages with ox-bile (0.2-0.4%) in the strain *B. dorei* Sb6. Control (no ox-bile was added). **(A)** 0.2% ox-bile. **(B)** 0.4% ox-bile. Data represent the mean of three replicates (N=6). The error bars represent the standard.

4.2.3.3 Ox-bile in the strain *B. dorei* Sb8

During the exposure of the strain *B. dorei* Sb8 to ox-bile (0.2%, 0.4%) no prophage induction was detected for both concentrations (**Figure 4.2.18**).

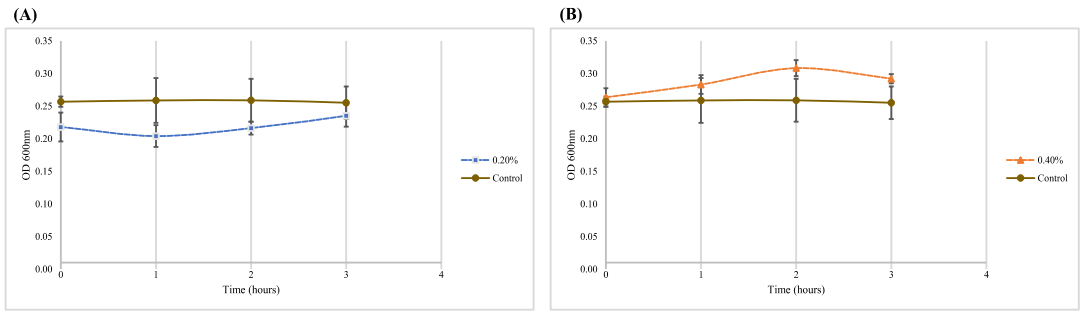

Figure 4.2.18 Induction of prophages with ox-bile (0.2-0.4%) in the strain *B. dorei* Sb8. Control (no ox-bile was added). **(A)** 0.2% ox-bile. **(B)** 0.4% ox-bile. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.3.4 Ox-bile in the strain *B. dorei* D1P5

The growth of the strain *B. dorei* D1P5 in the presence of 0.2 % ox-bile was slightly impaired but after 4 h the OD_{600nm} values were similar to the Control and no prophage induction was observed (**Figure 4.2.19**).

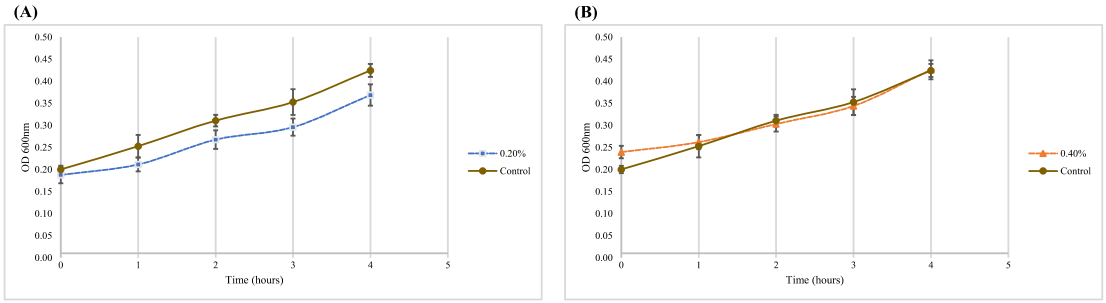

Figure 4.2.19 Induction of prophages with ox-bile (0.2-0.4%) in the strain *B. dorei* D1P5. Control (no ox-bile was added). **(A)** 0.2% ox-bile. **(B)** 0.4% ox-bile. Data represent the mean of three replicates (N=9). The error bars represent the standard.

4.2.3.5 Ox-bile in the strain *B. dorei* D8M1

The exposure of *B. dorei* D8M1 to ox-bile at 0.2% caused a lag phase of about 1 h and then the bacterial culture showed similar OD_{600nm} values to the Control (**Figure 4.2.20**). At the higher ox-bile concentration (0.4%) the strain *B. dorei* D8M1 showed after 2 h a decrease on the OD_{600nm} values in comparison to the Control, but no prophage induction was observed (**Figure 4.2.20**).

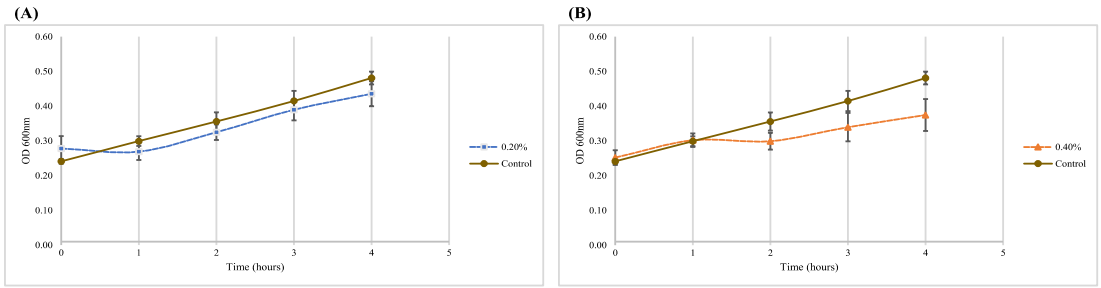

Figure 4.2.20 Induction of prophages with ox-bile (0.2-0.4%) in the strain *B. dorei* D8M1. Control (no ox-bile was added). **(A)** 0.2% ox-bile. **(B)** 0.4% ox-bile. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.3.6 Ox-bile in the strain *B. dorei* D16P1

The growth pattern of the strain *B. dorei* D16P1 treated with ox-bile (0.2%, 0.4%) was similar to the Control showing no prophage induction (**Figure 4.2.21**).

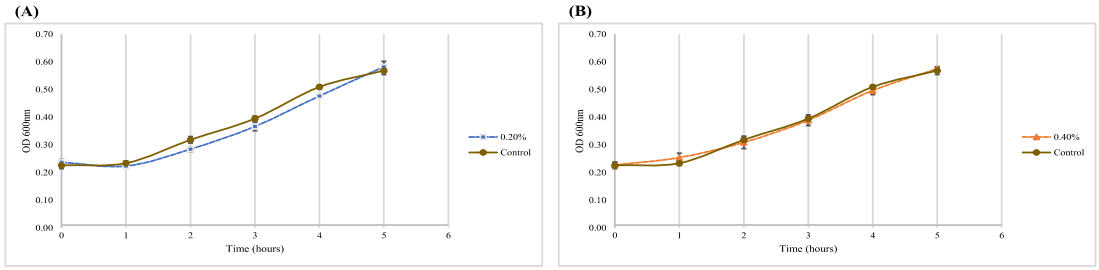

Figure 4.2.21 Induction of prophages with ox-bile (0.2-0.4%) in the strain *B. dorei* D16P1. Control (no ox-bile was added). **(A)** 0.2% ox-bile. **(B)** 0.4% ox-bile. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.3.7 Ox-bile in the strain *B. dorei* D16M14

The growth of the strain *B.* dorei D16M14 was not affected by the exposure to ox-bile (0.2%, 0.4%) evidencing the absence of prophage induction (**Figure 4.2.22**).

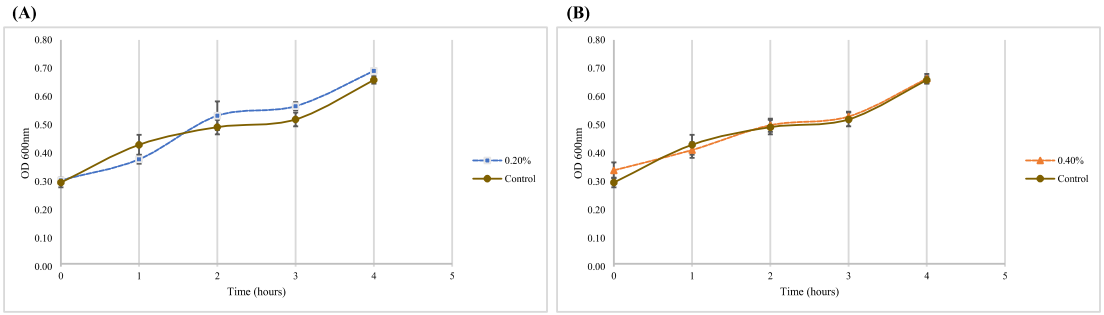

Figure 4.2.22 Induction of prophages with ox-bile (0.2-0.4%) in the strain *B. dorei* D16M14. Control (no ox-bile was added). **(A)** 0.2% ox-bile. **(B)** 0.4% ox-bile. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

4.2.3.8 Ox-bile in the strain *P. distasonis* D14MH1

The growth of *P. distasonis* D14MH1 exposed to ox-bile (0.2%, 0.4%) was affected, particularly after 2h, when the OD_{600nm} values were significantly lower in comparison to the Control (**Figure 4.2.23**). However, the ox-bile challenge did not cause prophage induction.

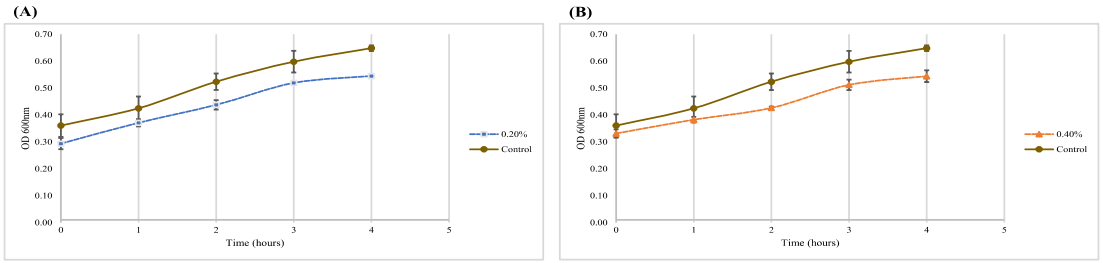

Figure 4.2.23 Induction of prophages with ox-bile (0.2-0.4%) in the strain *P. distasonis* D14MH1. Control (no ox-bile was added). **(A)** 0.2% ox-bile. **(B)** 0.4% ox-bile. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

5 Discussion

The role of the intestinal virome in the development of T1D is scarce to provide answers for the autoimmune attack of β-cells (Siljander, Honkanen and Knip, 2019). Prior to islet autoimmunity a complex dysbiosis is triggered affecting the balance of gut bacteria, fungi, viruses and unicellular eukaryotes that lastly impact the host physiology, development and immunity (Kosiewicz, Zirnheld and Alard, 2011; Cho and Blaser, 2012). These impacts favor proinflammatory signals at early age, particularly sensible for intestinal microbiota and immune system development toward its adult-like composition (Kostic *et al.*, 2015b; Stewart *et al.*, 2018).

Seroconversion to T1D is variable and can range in a time span from weeks to more than two decades (KNIP *et al.*, 2010) as reported by Ziegler *et. al*. (2013) where the probability of progression of T1D was greater than 80% after a follow-up of 15 years with the detection of two islet autoantibodies, reflecting the dynamic process of β-cells autoimmunity (Ziegler *et al.*, 2013). There are strong evidences that pronounced alterations occur in the gut microbiome preceding T1D onset, but what pathways are blocked or engaged are still unknown, and for this aspect viral composition may play an important role in triggering islet autoimmunity (Davis-Richardson *et al.*, 2014; Leonard *et al.*, 2014; Cinek *et al.*, 2016; Zhao *et al.*, 2018).

The bacteriophage CrAssphage is highly abundant in human faeces (Dutilh *et al.*, 2014; Manrique *et al.*, 2016) and it was linked with *B. dorei* but its impact on islet autoimmunity was not established (Cinek et al., 2016)

In the present study, we identified and characterized 9 incomplete prophages (cryptic prophages) in seven *B. dorei* strains and one *P. distasonis* strain from Control and T1D children (Matos, 2018). From the identified prophage-like elements in the scaffolds of each genome, 4 elements were never identified in *Bacteroidetes* according to Virus-Host Database providing new insights regarding these phages (Mannheimia phage vB_MhM_3927AP2; *Clostridium* phage c-st; *Bacillus* phage phiAGATE; *Synechococcus* phage S-SKS1). The performed bioinformatics analyses evidenced that all putative prophages present in the tested *B. dorei* strains and *P. distasonis* genomes are incomplete. The CrAssphage was not identified in any of our strains.

The prevalence of cryptic prophages has been reported within various human and animal bacterial pathogen genomes indicating co-evolutionary selection of such genomes (Casjens, 2003; Bobay, Touchon and Rocha, 2014).

The presence of cryptic prophages in all bacterial genomes analyzed may result from the complex process of phage domestication that can lead to the temporal elimination of their genes, grounding them in bacterial genomes (Ramisetty and Sudhakari, 2019). The follow-up of this, is the accumulation of deletions at higher rates in cryptic prophages, removing not advantageous genes from these regions (Bobay, Touchon and Rocha, 2014). In the temporal space that genes are eliminated from integrated prophages, they can became a force of horizontal gene transfer (HGT), as cryptic prophages can encode gene transfer agents (GTAs) transferring random pieces of bacterial genome to a recipient cell, but not their own (Lang, Zhaxybayeva and Beatty, 2012). The hypothesis that these cryptic prophages can modulate immune system through their impact in development of bacterial gut microbiota prior to the first β-cell autoantibody seems plausible. The presence of incomplete prophages in all genomes may show that *B. dorei* apply an efficient phage domestication, avoiding future lysis by them.

From the eight bacterial genomes analyzed, they were dominated by cryptic phages belonging to order *Caudovirales* (tailed phages) from T1D subjects and respective Controls, particularly *Siphoviridae* family. Supporting factors to this finding in children with ages above 2 years old in T1D patients and healthy subjects are limited by the lack of virome studies in children from the Algarve region. The absence of *att* in some of the genomes abolish completely the possibility of excision from the bacterial genome without affecting the transcription of genes (Ramisetty and Sudhakari, 2019).

The bacterial genomes from T1D children presented in average two prophage-like regions in comparison with the bacterial genomes from Controls that showed only one, denoting some permeability to phage infection for acquisition of advantageous genes to the host metabolism or physiology (Howard-varona *et al.*, 2017). Also, the percentage of phage protein similarity were higher for the bacterial genomes from T1D getting as high as 41.17% against the maximum of 14.28% for Controls with an average value of 19.9% and 8.9%, respectively. The prophage-like regions identified in the bacterial genomes from T1D showed an average length of 19.02 kb against the 9.13 kb of Controls. Some of the prophage-like regions identified in the bacterial genomes from T1D and Controls displayed phage similarity suggesting some common features (Minot, Bryson, Chehoud, Gary D. Wu, *et al.*, 2013; Mirzaei and Maurice, 2017).

Prophage-like regions in *Bacteroides* isolates from T1D children that can carry genes potentially impacting or triggering the autoimmunity attack of β-cells to the onset of T1D

were identified, in comparison with the prophage-like regions in *Bacteroides* isolates from Controls that did not show any gene with this attribute.

The *B. dorei* D8M1 genome showed the highest number of prophage-like regions. Region 1 has similarity to the *Clostridium botulinum* phage, with 5 of 12 ORFs from this region having annotated function. Two integrases were present in this region with a coding gene for DEAD/DEAH box helicase. This enzyme drives the hydrolysis of the nucleoside triphosphate (DNTP) of dsRNA unwinding or RNA-protein complexes remodulation, playing essential roles in RNA metabolism both in eukaryotes and prokaryotes (Jarmoskaite and Russell, 2011; Redder *et al.*, 2015; Perčulija and Ouyang, 2019) and in bacteria these enzymes play a crucial role on adaptation to stress conditions (Redder *et al.*, 2015). It is known that the DHX9 has autoantigen properties being associated with the early stages of systemic lupus erythematosus syndrome (Yamasaki *et al.*, 2007) and now we know that the translocation of the gut pathogen *Enterococcus gallinarum* to the liver and other systemic tissues induces autoimmunity including the production of lupus-related autoantibodies (Vieira *et al.*, 2018). Nevertheless, the role of the DEAD/DEAH box helicases on inflammation triggering autoimmunity still requires clarification, namely their regulators and interaction associates (Perčulija and Ouyang, 2019).

Another important gene that can possibly influence the autoimmune attack of β-cells is DNA adenine methylase (Dam) which has the capability to methylate nitrogen position 6 in adenine to generate N-6-methyladenine (m6c) on recognition sites (palindromic repeats of 5´-GATC-3´) regulating various functions in the phage´s host (Sternberg and Coulby, 1990; Murphy *et al.*, 2008, 2013; Samson *et al.*, 2013). Dam can be used by phages as a R-M escape system (Samson *et al.*, 2013) supporting DNA packaging as observed in Phage P1 DNA, that depends on GATC methylated sequences for viral packaging, when this gene is mutated a severely reduction in viral particles is observed (Sternberg and Coulby,

1990). Other function provided by Dam is to upkeep the phage lysogenic cycle, that can be advantageous to the host as stated in an enterohemorrhagic *E. coli* strain in which a *dam* gene was correlated with the maintenance of phage lysogenic cycle by Dam, allowing the continuous production of Shiga toxins encoded in a phage toxic island (Murphy *et al.*, 2008, 2013). The *dam* gene was identified in the genome of the strains *B. dorei* D16P1 and D16M14, showing the same phage similarity in both. A single-orphan of a *dam* gene in a prophage region of a strain of *B. dorei* isolated from a high genetic

risk Finnish child, that showed a heavily methylated genome in Dam recognition sites prior to seroconversion, and which progressed to β-cells autoimmunity after 1.5 months was reported (Davis-Richardson *et al.*, 2014; Leonard *et al.*, 2014). In contrast another *B. dorei* strain without the *dam* gene, also from a Finnish child with the same genetic risk to T1D, but which did not revealed a methylated genome neither progress to autoimmunity against β-cells was observed (Davis-Richardson *et al.*, 2014; Leonard *et al.*, 2014). So far it is not clear what is the impact of this gene on the onset of T1D (Davis-Richardson *et al.*, 2014; Leonard *et al.*, 2014).

In the phage induction assays carried with *Bacteroides* strains either from T1D or Control children the lytic cycle driven by the exposure to mit C, H_2O_2 and ox-bile was not observed. The observed growth inhibition by the exposure to mitomycin C may involve the bacterial SOS system that can provide protection against the inducer stress on DNA damage or allow the development of antibiotic resistance by generating DNA mutation within the bacterial genome (Casanueva *et al.*, 2008; Fornelos *et al.*, 2018). Resistance to mitomycin C cross-linked DNA damage can be achieved by the presence of a AraC/XylS transcriptional regulator specifically a Reg protein first identified in *B. fragilis*, but not identified in the current study (Casanueva *et al.*, 2008). The drop of the optical density observed in some concentrations of mitomycin C can be due to a bactericide effect upon these strains at specific concentrations.

The exposure of Shiga toxin-producing *E. coli* (STEC) strains to H_2O_2 is one of the factors that can induce the Shiga toxin production. The genes that code for Shiga toxins are settled on lambdoid prophages, so the successful toxin production can only take place after prophage induction and the exposure of these STEC strains to H_2O_2 (3 mM) leads to efficient phage induction (Łoś *et al.*, 2010). For some of the tested *Bacteroides* strains the exposure to H_2O_2 seems to cause stress to the bacterial cells evidenced by the bacterial growth stagnation. *Bacteroides* are not able to grow at concentrations higher than 5 µM of dissolved O_2, and the ability of the tested *Bacteroides* strains to overcome the H_2O_2 stress can probably be associated with an efficient repair system, as the *recT* gene, known to be a mediator of single-stranded conversion in DNA recombination repair (Iyer, Koonin and Aravind, 2002). In a *B. fragilis* strain two genes responsible for the coding of a cytochrome bd oxidase that is responsible for intracellular O_2 levels reduction, helps with the maintenance of the bacterial growth under oxidative stress conditions, providing

an insight for their large presence in gut microbiota, however not identified in the current study (Baughn and Malamy, 2004; Wexler and Goodman, 2017).

Intestinal lysogenic bacteria are exposed to bile salt stress in the GI and then eventually release lytic phages in a competing bacterial community (Broudy, Pancholi and Fischetti, 2001) contributing to the modulation of the microbiota from the GI tract. It was reported that the exposure of lysogenic strains of *Salmonella enterica* serovar Typhimurium ATCC 23555 and 19585 to 0.5 % bile salts induces a high prophage induction (3.34 log PFU/mL) (Kim *et al.*, 2014).

The treatment of the *Bacteroides* strains with 0.2% and 0.4% for Ox-bile did not induce any prophage and did not affect the bacterial growth, except for the *P. distasonis* D14MH1 for which a slight growth decrease was observed. The absence of prophage induction of the Lagaffe and Mushu phages of the intestinal bacterium *Faecalibacterium prausnitzii* by mitomycin C, H_2O_2 and bile salts also was previously reported (Cornuault *et al.*, 2018).

6 Conclusion and Future perspectives

The main goal of this study was the identification and characterization of prophages in *Bacteroides* spp. isolates from children with Type 1 diabetes. The analysis of the genomes with the bioinformatics tool PHASTER revealed the lack of complete prophages within these strains, with a clear difference between the prophage-like regions encountered within T1D isolates and Controls, presenting in average of two prophage-like regions integrated in T1D isolates against only one in strains isolated from Control children.

In the prophage-like regions analyzed, specifically for T1D isolates, two interesting genes were identified that can possibly impact T1D onset. One of these genes codes for a DEAD/DEAH box helicase that participates in human innate immunity possibly providing an insight in the inflammation process triggering autoimmune diseases, but it requires further investigation. Another important gene than can influence the development of T1D is the *dam* gene identified in *B. dorei* D16P1 and D16M14 prophage regions. This gene was also identified as a single-orphan in a prophage region of a *B. dorei* isolated from a high genetic risk Finnish child, with almost all of the Dam recognition sites methylated prior to seroconversion, that progressed to β-cell autoimmunity after 1.5 months. The identification in the current study of a *dam* gene in identical prophage regions from two *B. dorei* strains that were isolated from the same patient provides evidence that bacteriophages may qualify *B. dorei* with selective abilities, particularly through the methylation of genes influencing gene expression either by up-regulating or down-regulating (*e.g.* lysogeny) with the Dam methylation of the 5´-GATC-3´ motif, that can explain its dominance prior to seroconversion. Since us and others are reporting the presence of *dam* gene in *B. dorei* strains isolated from T1D children it is important to dissect its role on the bacterial physiology and its impact on the seroconversion.

The trigger that starts the metabolic cascade to the autoimmunity attack to β-cells may come from an ample source of biological identities, where viral patterns in the bacterial communities can be one of them as perceived by the described genes within prophage regions.

The analyzed *Bacteroides* strains from T1D and Control children from Algarve did not show to carry any complete prophage regions in their genomes, which explains the absence of prophage induction by the tested inducers Mitomycin C, H_2O_2 and Ox-bile.

Finally, this study elucidates the phage patterns in *Bacteroides* spp. isolates from T1D children from the Algarve region, being able to identify genes in prophage regions that can participate on the onset of T1D. The role of the intestinal virome in the development of T1D is scarce to provide answers for the autoimmune attack of β-cells and with the current study we tried to provide answers to complex questions in terms of bacterial and viral patterns within the bacterial communities in the unhealthy gut.

A full analysis of each bacterial genome for virulence islands and toxin would also be an effective way to demonstrate the capacity of *Bacteroides* spp. to disturb the gut environment, providing the maintenance of a dysbiotic state that can allow a dysregulated immune reaction triggering T1D.

7 References

Aathira, R. and Jain, V. (2014) 'Advances in management of type 1 diabetes mellitus', *World Journal of Diabetes*, 5(5), pp. 689–696. doi: 10.4239/wjd.v5.i5.689.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Argenio, V. D. (2018) 'The Prenatal Microbiome : A New Player for Human Health', *High-Throughput*, 7, pp. 1–10. doi: 10.3390/ht7040038.

Arndt, D. *et al.* (2016) 'PHASTER: a better, faster version of the PHAST phage search tool', *Nucleic acids research*, 44(W1), pp. W16–W21. doi: 10.1093/nar/gkw387.

Arumugam, M. *et al.* (2011) 'Enterotypes of the human gut microbiome', *Nature*, 473. doi: 10.1038/nature09944.

Barbetti, F. and Magliocca, F. M. (1988) 'Fecal Lactate and Ulcerative Colitis', *Gastroenterology*. American Gastroenterological Association, 95(6), pp. 1564–1568. doi: 10.1016/S0016-5085(88)80078-7.

Baughn, A. D. and Malamy, M. H. (2004) 'The strict anaerobe Bacteroides fragilis grows in and benefits from nanomolar concentrations of oxygen', *Nature*, 427, pp. 441–444. doi: 10.1038/nature02285.

Bobay, L., Touchon, M. and Rocha, E. P. C. (2014) 'Pervasive domestication of defective prophages by bacteria', *PNAS*, 111(33), pp. 12127–12132. doi: 10.1073/pnas.1405336111.

Bokulich, N. A. *et al.* (2016) 'Antibiotics , birth mode , and diet shape microbiome maturation during early life', *Science Translational Medicine*, 8(343), pp. 1–14.

Bondy-denomy, J. *et al.* (2012) 'Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system', *Nature*. Nature Publishing Group, pp. 1–6. doi: 10.1038/nature11723.

Boyd, E. F. (2012) 'Advances in Virus Research', in *Bacteriophages, Part A*, pp. 91–118. doi: 10.1016/S0074-7696(08)60503-3.

Breitbart, M. *et al.* (2008) 'Viral diversity and dynamics in an infant gut', *Elsevier*, 159. doi: 10.1016/j.resmic.2008.04.006.

Broudy, T. B., Pancholi, V. and Fischetti, V. A. (2001) 'Induction of lysogenic bacteriophage and phage-associated toxin from group A streptococci during coculture with human pharyngeal cells', *Infection and Immunity*, 69(3), pp. 1440–1443. doi: 10.1128/IAI.69.3.1440-1443.2001.

Brown, C. T. *et al.* (2011) 'Gut Microbiome Metagenomics Analysis Suggests a Functional Model for the Development of Autoimmunity for Type 1 Diabetes', *PloS one*, 6(10), pp. 1–9. doi: 10.1371/journal.pone.0025792.

Cao, Y., Shen, J. and Ran, Z. H. (2014) 'Association between Faecalibacterium prausnitzii Reduction and Inflammatory Bowel Disease : A Meta-Analysis and Systematic Review of the Literature', *Gastroenterology Research and Practice*.

Casanueva, A. I. *et al.* (2008) 'An AraC/XylS family transcriptional regulator homologue from Bacteroides fragilis is associated with cell survival following DNA damage', *FEMS Microbiology Letters*, 278(2), pp. 249–256. doi: 10.1111/j.1574-6968.2007.01004.x.

Casjens, S. (2003) 'Prophages and bacterial genomics : what have we learned so far ?', *Molecular Microbiology*, 49(2), pp. 277–300. doi: 10.1046/j.1365-2958.2003.03580.x.

Cenens, W. *et al.* (2013) 'Phage-host interactions during pseudolysogeny: Lessons from the Pid/dgo', *Bacteriophage*. doi: 10.4161/bact.25029.

Champness, W. C. and Snyder, L. (1982) 'The gol site: A Cis-acting bacteriophage T4 regulatory region that can affect expression of all the T4 late genes', *Journal of Molecular*

Biology, 155, pp. 395–407. doi: 10.1016/0022-2836(82)90478-8.
Chatterjee, S. and Rothenberg, E. (2012) 'Interaction of Bacteriophage with Its E. coli Receptor, LamB', *Viruses*, 4, pp. 3162–3178. doi: 10.3390/v4113162.
Chen, Y. *et al.* (2019) 'Prophage Excision in Streptococcus pneumoniae Serotype 19A ST320 Promote Colonization : Insight Into Its Evolution From the Ancestral Clone', *Frontiers in Microbiology*, 10(205), pp. 1–10. doi: 10.3389/fmicb.2019.00205.
Cho, I. and Blaser, M. J. (2012) 'The human microbiome : at the interface of health and disease', *Nature Reviews Genetics*, 13(April). doi: 10.1038/nrg3182.
Cinek, O. *et al.* (2016) 'Imbalance of bacteriome profiles within the Finnish Diabetes Prediction and Prevention study: Parallel use of 16S profiling and virome sequencing in stool samples from children with islet autoimmunity and matched controls', *Pediatric Diabetes*, 18(7), pp. 588–598. doi: 10.1111/pedi.12468.
Cornuault, J. K. *et al.* (2018) 'Phages infecting Faecalibacterium prausnitzii belong to novel viral genera that help to decipher intestinal viromes', *Microbiome*. Microbiome, 6(1), p. 65. doi: 10.1186/s40168-018-0452-1.
Cortez, M. H. and Weitz, J. S. (2014) 'Coevolution can reverse predator – prey cycles', *PNAS*, 111(20). doi: 10.1073/pnas.1317693111.
Costello, E. K. *et al.* (2009) 'Bacterial community variation in human body habitats across space and time', *Science*, 326, pp. 1694–1697. doi: 10.1126/science.1177486.
Crispim, J. S. *et al.* (2018) 'Screening and characterization of prophages in Desulfovibrio genomes', *Scientific Reports*, 8(1), pp. 1–10. doi: 10.1038/s41598-018-27423-z.
Cumby, N. *et al.* (2012) 'The Bacteriophage HK97 gp15 Moron Element Encodes a Novel Superinfection Exclusion Protein', *Journal of Bacteriology*, 194(18), pp. 5012–5019. doi: 10.1128/JB.00843-12.
Dalmasso, M., Hill, C. and Ross, R. P. (2014) 'Exploiting gut bacteriophages for human health', *Trends in Microbiology*, pp. 1–7.
Davis-Richardson, A. G. *et al.* (2014) 'Bacteroides dorei dominates gut microbiome prior to autoimmunity in Finnish children at high risk for type 1 diabetes', *Frontiers in Microbiology*, 5(678), pp. 1–11. doi: 10.3389/fmicb.2014.00678.
Dong, X. *et al.* (2015) 'Bacterial communities in neonatal feces are similar to mothers ' placentae', *Canadian Journal of Infectious Diseases & Medical Microbiology*, 26(2), pp. 90–95.
Duncan, S. H. *et al.* (2004) 'Contribution of acetate to butyrate formation by human faecal bacteria', *British Journal of Nutrition*, 91, pp. 915–923. doi: 10.1079/BJN20041150.
Dutilh, B. E. *et al.* (2014) 'A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes', *Nature Communications*. Nature Publishing Group, 5, pp. 1–11. doi: 10.1038/ncomms5498.
Eggesbø, M. *et al.* (2011) 'Development of gut microbiota in infants not exposed to medical interventions', *APMIS*, 119, pp. 17–35. doi: 10.1111/j.1600-0463.2010.02688.x.
Fallani, M. *et al.* (2010) 'Intestinal Microbiota of 6-week-old Infants Across Europe: Geographic Influence Beyond Delivery Mode, Breast-feeding, and Antibiotics', *HEPATOLOGY AND NUTRITION*, 51(1), pp. 77–84. doi: 10.1097/MPG.0b013e3181d1b11e.
Feiner, R. *et al.* (2015) 'A new perspective on lysogeny: prophages as active regulatory switches of bacteria', *Nature Reviews Microbiology*. Nature Publishing Group, 13(10), pp. 641–650. doi: 10.1038/nrmicro3527.
Flint, H. J. *et al.* (2012) 'The role of the gut microbiota in nutrition and health', *Nature Reviews Gastroenterology & Hepatology*. Nature Publishing Group, 9(10), pp. 577–589. doi: 10.1038/nrgastro.2012.156.
Fornelos, N. *et al.* (2018) 'Lytic gene expression in the temperate bacteriophage GIL01

is activated by a phage-encoded LexA homologue', *Nucleic Acids Research*. Oxford University Press, 46(18), pp. 9432–9443. doi: 10.1093/nar/gky646.

Fortier, L. and Moineau, S. (2007) 'Morphological and Genetic Diversity of Temperate Phages in Clostridium difficile', *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 73(22), pp. 7358–7366. doi: 10.1128/AEM.00582-07.

Goffau, M. C. De *et al.* (2013) 'Fecal Microbiota Composition Differs Between Children With b -Cell Autoimmunity and Those Without', *DIABETES*, 62. doi: 10.2337/db12-0526.

Guerin, E. *et al.* (2018) 'Biology and Taxonomy of crAss-like Bacteriophages , the Most Abundant Virus in the Human Gut Biology and Taxonomy the Most Abundant Virus in the Human Gut', *Cell Host and Microbe*. Elsevier Inc., 24, pp. 1–12. doi: 10.1016/j.chom.2018.10.002.

Gunn, J. S. (2000) 'Mechanisms of bacterial resistance and response to bile', *Microbes and Infection*, (2), pp. 907–913.

Hill, C. J. *et al.* (2017) 'Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort', *Microbiome*. Microbiome, pp. 1–18. doi: 10.1186/s40168-016-0213-y.

Hill, C. W., Gray, J. A. and Brody, H. (1989) 'Use of the Isocitrate Dehydrogenase Structural Gene for Attachment of e14 in Escherichia coli K-12', *Journal of Bacteriology*, 171(7), pp. 4083–4085. doi: 10.1128/jb.171.7.4083-4084.1989.

Hoffmann, C. *et al.* (2013) 'Archaea and Fungi of the Human Gut Microbiome : Correlations with Diet and Bacterial Residents', *PLoS ONE*, 8(6). doi: 10.1371/journal.pone.0066019.

Holmfeldt, K. *et al.* (2013) 'Twelve previously unknown phage genera are ubiquitous in global oceans', 110(31), pp. 1–6. doi: 10.1073/pnas.1305956110.

Houte, S. Van, Buckling, A. and Westra, E. R. (2016) 'Evolutionary Ecology of Prokaryotic Immune Mechanisms', *Microbiology and Molecular Biology Reviews*, 80(3), pp. 745–763. doi: 10.1128/MMBR.00011-16.Address.

Howard-varona, C. *et al.* (2017) 'Lysogeny in nature : mechanisms , impact and ecology of temperate phages', *The ISME*. Nature Publishing Group, 11(7), pp. 1511–1520. doi: 10.1038/ismej.2017.16.

Høyland-kroghsbo, N. M., Mærkedahl, R. B. and Svenningsen, S. Lo (2013) 'A Quorum-Sensing-Induced Bacteriophage Defense Mechanism', *mBio*, 4(1), pp. 1–8. doi: 10.1128/mBio.00362-12.Editor.

Huttenhower, C. *et al.* (2012) 'Structure, function and diversity of the healthy human microbiome', *Nature*. Nature Publishing Group, 486(7402), pp. 207–214. doi: 10.1038/nature11234.

International Diabetes Federation (2017) *IDF Diabetes Atlas*.

Iyer, L., Koonin, E. V. and Aravind, L. (2002) 'Classification and evolutionary history of the single-strand annealing proteins, RecT, Redβ, ERF and RAD52', *BMC Genomics*, 3. doi: 10.1186/1471-2164-3-8.

Jarmoskaite, I. and Russell, R. (2011) 'DEAD-box proteins as RNA helicases and chaperones', *Wiley Interdisciplinary Reviews: RNA*, 2(1), pp. 135–152. doi: 10.1002/wrna.50.

Jones, B. V *et al.* (2008) 'Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome', *PNAS*, 105(36), pp. 13580–13585.

Kantárová, D. and Buc, M. (2007) 'Genetic Susceptibility to Type 1 Diabetes Mellitus in Humans', *Physiol. Res.*, 56, pp. 255–266.

Katsarou, A. *et al.* (2017) 'Type 1 diabetes mellitus', *Nature*. Macmillan Publishers

Limited, 3, pp. 1–18. doi: 10.1038/nrdp.2017.16.
Kim, S. *et al.* (2014) 'Survival, prophage induction, and invasive properties of lysogenic Salmonella Typhimurium exposed to simulated gastrointestinal conditions', *Archives of Microbiology*, 196(9), pp. 655–659. doi: 10.1007/s00203-014-1005-z.
KNIP, M. *et al.* (2010) 'Prediction of Type 1 Diabetes in the General Population', *DIABETES CARE*, 33(6), pp. 0–6. doi: 10.2337/dc09-1040.
Koh, A. *et al.* (2016) 'From Dietary Fiber to Host Physiology : Short-Chain Fatty Acids as Key Bacterial Metabolites', *Cell*, 165. doi: 10.1016/j.cell.2016.05.041.
Kosiewicz, M. M., Zirnheld, A. L. and Alard, P. (2011) 'Gut microbiota , immunity , and disease : a complex relationship', 2(September), pp. 1–11. doi: 10.3389/fmicb.2011.00180.
Kostic, A. D. *et al.* (2015a) 'The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Resource The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Diabetes', *Cell Host & Microbe*, 17, pp. 260–273. doi: 10.1016/j.chom.2015.01.001.
Kostic, A. D. *et al.* (2015b) 'The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Resource The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Diabetes', *Cell Host & Microbe*, pp. 260–273. doi: 10.1016/j.chom.2015.01.001.
Lang, A. S., Zhaxybayeva, O. and Beatty, J. T. (2012) 'Gene transfer agents : phage-like elements of genetic exchange', *Nature Reviews Microbiology*. Nature Publishing Group, 10(June). doi: 10.1038/nrmicro2802.
Leonard, M. T. *et al.* (2014) 'The methylome of the gut microbiome: Disparate Dam methylation patterns in intestinal Bacteroides dorei', *Frontiers in Microbiology*, 5(JULY), pp. 1–6. doi: 10.3389/fmicb.2014.00361.
Lim, E. S. *et al.* (2015) 'Early life dynamics of the human gut virome and bacterial microbiome in infants', *Nature Medicine*. Nature Publishing Group, 21(10), pp. 1228–1234. doi: 10.1038/nm.3950.
Litvak, Y., Byndloss, M. X. and Avenue, O. S. (2018) 'Colonocyte metabolism shapes the gut microbiota', *Science*, 362(6418), pp. 1–15. doi: 10.1126/science.aat9076.Colonocyte.
Løbner-Olesen, A., Skovgaard, O. and Marinus, M. G. (2005) 'Dam methylation: Coordinating cellular processes', *Current Opinion in Microbiology*, 8(2), pp. 154–160. doi: 10.1016/j.mib.2005.02.009.
Lorenz, N. *et al.* (2016) 'Identification and initial characterization of prophages in Vibrio campbellii', *PLoS ONE*, 11(5), pp. 1–14. doi: 10.1371/journal.pone.0156010.
Łoś, J. M. *et al.* (2010) 'Hydrogen peroxide-mediated induction of the Shiga toxin-converting lambdoid prophage ST2-8624 in Escherichia coli O157:H7', *FEMS Immunol Med Microbiol*, 58(3), pp. 322–329. doi: 10.1111/j.1574-695X.2009.00644.x.
Lu, M. and Henning, U. (1994) 'Superinfection exclusion by T-even-type coliphages', *Trends in Microbiology*, 2(4), pp. 4–6.
Magnúsdóttir, S. *et al.* (2015) 'Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes', *Frontiers in Genetics*, 6(148). doi: 10.3389/fgene.2015.00148.
Manrique, P. *et al.* (2016) 'Healthy human gut phageome', *PNAS*, 113(37), pp. 10400–10405. doi: 10.1073/pnas.1601060113.
Matos, R. C. *et al.* (2013) 'Enterococcus faecalis Prophage Dynamics and Contributions to Pathogenic Traits', *PLOS Genetics*, 9(6). doi: 10.1371/journal.pgen.1003539.
Michel, A. *et al.* (2010) 'Bacteriophage PhiX174 's Ecological Niche and the Flexibility of Its Escherichia coli Lipopolysaccharide Receptor', *Applied and Environmental*

Microbiology, 76(21), pp. 7310–7313. doi: 10.1128/AEM.02721-09.
Milani, C. *et al.* (2017) 'The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota', *Microbiology and Molecular Biology Reviews*, 81(4), pp. 1–67.

Minot, S., Bryson, A., Chehoud, C., Wu, G. D., *et al.* (2013) 'Rapid evolution of the human gut virome', *PNAS*, 110(30). doi: 10.1073/pnas.1300833110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1300833110.
Minot, S., Bryson, A., Chehoud, C., Wu, G. D., *et al.* (2013) 'Rapid evolution of the human gut virome', *PNAS*, 110(30), pp. 12450–12455. doi: 10.1073/pnas.1300833110.
Mirzaei, M. K. and Maurice, C. F. (2017) 'Ménage à trois in the human gut : interactions between host , bacteria and phages', *Nature Publishing Group*. Nature Publishing Group, 15(7), pp. 397–408. doi: 10.1038/nrmicro.2017.30.
Mukhopadhya, I. *et al.* (2019a) 'The gut virome : the " missing link " between gut bacteria and host immunity ?', *Therapeutic Advances in Gastroenterology Review*, 12, pp. 1–17. doi: 10.1177/1756284819836620.
Mukhopadhya, I. *et al.* (2019b) 'The gut virome : the " missing link " between gut bacteria and host immunity?', *Therapeutic Advances in Gastroenterology*, 12, pp. 1–17. doi: 10.1177/1756284819836620.
Murphy, J. *et al.* (2013) 'Bacteriophage orphan DNA methyltransferases: Insights from their bacterial origin, function, and occurrence', *Applied and Environmental Microbiology*, 79(24), pp. 7547–7555. doi: 10.1128/AEM.02229-13.
Murphy, K. C. *et al.* (2008) 'Dam methyltransferase is required for stable lysogeny of the shiga toxin (Stx2)-encoding bacteriophage 933W of enterohemorrhagic Escherichia coli O157:H7', *Journal of Bacteriology*, 190, pp. 438–441. doi: 10.1128/JB.01373-07.
Nanda, A. M. and Thormann, K. (2015) 'Impact of Spontaneous Prophage Induction on the Fitness of Bacterial Populations and Host-Microbe Interactions', *Journal of Bacteriology*, 197(3), pp. 410–419. doi: 10.1128/JB.02230-14.
Nordstrom, K. and Forsgren, A. (1974) 'Effect of Protein A on Adsorption of Bacteriophages to Staphylococcus aureus', *Journal of Virology*, 14(2), pp. 198–202.
Norman, J. M. *et al.* (2015) 'Disease-Specific Alterations in the Enteric Virome in Article Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease', *Cell*. Elsevier Inc., 160(3), pp. 447–460. doi: 10.1016/j.cell.2015.01.002.
Oh, J. *et al.* (2019) 'Dietary Fructose and Microbiota-Derived Short- Chain Fatty Acids Promote Bacteriophage Production in the Gut Symbiont Lactobacillus reuteri Article Dietary Fructose and Microbiota-Derived Short-Chain Fatty Acids Promote Bacteriophage Production in the Gut', *Cell Host & Microbe*. Elsevier Inc., 25, pp. 1–12. doi: 10.1016/j.chom.2018.11.016.

Oost, J. Van Der *et al.* (2014) 'Unravelling the structural and mechanistic basis of CRISPR – Cas systems', *Nature Reviews Microbiology*, 12. doi: 10.1038/nrmicro3279.
Otsuka, Y. and Yonesaki, T. (2012) 'Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins', *Molecular Microbiology*, 83(4), pp. 669–681. doi: 10.1111/j.1365-2958.2012.07975.x.
Palmer, C. *et al.* (2007) 'Development of the Human Infant Intestinal Microbiota', *PLoS BIOLOGY*, 5(7). doi: 10.1371/journal.pbio.0050177.
Perčulija, V. and Ouyang, S. (2019) *Diverse Roles of DEAD/DEAH-Box Helicases in Innate Immunity and Diseases, Helicases from All Domains of Life*. doi: 10.1016/b978-0-12-814685-9.00009-9.
Pereira, C. and Berry, D. (2017) 'Microbial nutrient niches in the gut', *Environmental*

Microbiology, 19, pp. 1366–1378. doi: 10.1111/1462-2920.13659.
Pope, W. H. *et al.* (2011) 'Expanding the Diversity of Mycobacteriophages : Insights into Genome Architecture and Evolution', *PLoS ONE*, 6(1). doi: 10.1371/journal.pone.0016329.
Prieto, A. I. *et al.* (2004) 'Bile-Induced DNA Damage in Salmonella enterica', 1794(December), pp. 1787–1794. doi: 10.1534/genetics.104.031062.
Qimron, U. *et al.* (2006) 'Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage', *PNAS*, 103(50), pp. 19039–19044.
Qin, J. *et al.* (2010) 'A human gut microbial gene catalogue established by metagenomic sequencing', *Nature*, 464. doi: 10.1038/nature08821.
Rakhuba, D. V. *et al.* (2010) 'Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell', *Polish Journal of Microbiology*, 59(3), pp. 145–155.
Ramisetty, B. C. M. and Sudhakari, P. A. (2019) 'Bacterial " Grounded " Prophages : Hotspots for Genetic Renovation and Innovation', *Frontiers in Genetics*, 10(65), pp. 1–17. doi: 10.3389/fgene.2019.00065.
Redder, P. *et al.* (2015) 'Bacterial versatility requires DEAD-box RNA helicases', *FEMS Microbiology Reviews*, 39(3), pp. 392–412. doi: 10.1093/femsre/fuv011.
Reyes, A. *et al.* (2010) 'Viruses in the fecal microbiota of monozygotic twins and their mothers', *Nature*, 466(7304), pp. 334–338. doi: 10.1038/nature09199.Viruses.
Reyes, A. *et al.* (2015) 'Gut DNA viromes of Malawian twins discordant for severe acute malnutrition', *PNAS*, 112(38), pp. 11941–11946. doi: 10.1073/pnas.1514285112.
Ridlon, J. M. *et al.* (2014) 'Bile Acids and the Gut Microbiome', *Current Opinion in Microbiology*, 30(3), pp. 332–338. doi: 10.1097/MOG.0000000000000057.Bile.
Ridlon, J. M., Kang, D. and Hylemon, P. B. (2006) 'Bile salt biotransformations by human intestinal bacteria', *Journal of Lipid Research*, 47, pp. 241–259. doi: 10.1194/jlr.R500013-JLR200.
Riede, I. and Eschbach, M. (1986) 'Evidence that TraT interacts with OmpA of Escherichia coli', *Federation of European Biochemical Societies*, 205(2).
Rinninella, E. *et al.* (2019) 'What is the Healthy Gut Microbiota Composition ? A Changing Ecosystem across Age , Environment , Diet , and Diseases', *Microorganisms*, (7). doi: 10.3390/microorganisms7010014.
Roberts, R. J. *et al.* (2003) 'A nomenclature for restriction enzymes , DNA methyltransferases , homing endonucleases and their genes', *Nucleic Acids Research*, 31(7), pp. 1805–1812. doi: 10.1093/nar/gkg274.
Rocha, E. R. *et al.* (1996) 'Oxidative Stress Response in an Anaerobe , Bacteroides fragilis : a Role for Catalase in Protection against Hydrogen Peroxide', 178(23), pp. 6895–6903.
Rodrı, J. M. *et al.* (2015) 'The composition of the gut microbiota throughout life, with an emphasis on early life', *Microbial Ecology in Health & Disease*, 1, pp. 1–17.
Rossocha, M. *et al.* (2005) 'Conjugated Bile Acid Hydrolase Is a Tetrameric N-Terminal Thiol Hydrolase with Specific Recognition of Its Cholyl but Not of Its Tauryl Product', *Biochemistry*, 2(222), pp. 5739–5748.
Rowland, I. *et al.* (2018) 'Gut microbiota functions : metabolism of nutrients and other food components', *European Journal of Nutrition*. Springer Berlin Heidelberg, 57(1), pp. 1–24. doi: 10.1007/s00394-017-1445-8.
Russell, D. W. (2003) 'The Enzymes, Regulation, and Genetics of Bile Acid Synthesis', *Annual Review of Biochemistry*, 72, pp. 137–174. doi: 10.1146/annurev.biochem.72.121801.161712.
Sakaguchi, Y. *et al.* (2005) 'The genome sequence of Clostridium botulinum type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny',

102(48).
Samson, J. E. *et al.* (2013) 'Revenge of the phages : defeating bacterial defences', *Nature Reviews Microbiology*. Nature Publishing Group, 11(10), pp. 675–687. doi: 10.1038/nrmicro3096.
Sayin, S. I. *et al.* (2013) 'Gut Microbiota Regulates Bile Acid Metabolism by Reducing the Levels of Tauro-beta-muricholic Acid , a Naturally Occurring FXR Antagonist', *Cell Metabolism*, 17, pp. 225–235. doi: 10.1016/j.cmet.2013.01.003.
Scanlan, P. D. and Buckling, A. (2011) 'Co-evolution with lytic phage selects for the mucoid phenotype of Pseudomonas fluorescens SBW25', *The ISME Journal*. Nature Publishing Group, 6, pp. 1148–1158. doi: 10.1038/ismej.2011.174.
Scanlan, P. D., Buckling, A. and Hall, A. R. (2015) 'Coevolution with bacteriophages drives genome-wide host evolution and constrains the acquisition of abiotic-beneficial mutations', *Bacteriophage*, 5(June), pp. 7–9.
Scholl, D., Adhya, S. and Merril, C. (2005) 'Escherichia coli K1 's Capsule Is a Barrier to Bacteriophage T7', *Applied and Environmental Microbiology*, 71(8), pp. 4872–4874. doi: 10.1128/AEM.71.8.4872.
Siegel, D. *et al.* (1990) 'Metabolism of Mitomycin C by DT-Diaphorase : Role in Mitomycin C-induced DNA Damage and Cytotoxicity in Human Colon Carcinoma Cells1', 50, pp. 7483–7489.
Siljander, H., Honkanen, J. and Knip, M. (2019) 'Microbiome and type 1 diabetes', *EBioMedicine*, 46, pp. 512–521.
Sternberg, N. and Coulby, J. (1990) 'Cleavage of the bacteriophage P1 packaging site (pac) is regulated by adenine methylation', *Proceedings of the National Academy of Sciences of the United States of America*, 87, pp. 8070–8074. doi: 10.1073/pnas.87.20.8070.
Stewart, C. J. *et al.* (2018) 'Temporal development of the gut microbiome in early childhood from the TEDDY study', *Nature*. Springer US. doi: 10.1038/s41586-018-0617-x.
Sullivan, A. O., Farver, M. and Smilowitz, J. T. (2015) 'The Influence of Early Infant-Feeding Practices on the Intestinal Microbiome and Body Composition in Infants', *Nutrition and Metabolic Insights*, 8, pp. 1–9. doi: 10.4137/NMI.S29530.TYPE.
Sutton, M. D. *et al.* (2000) 'The SOS Response : Recent Insights into umuDC -Dependent Mutagenesis and DNA Damage tolerance', 34.
Tamarit, J., Cabiscol, E. and Ros, J. (1998) 'Identification of the Major Oxidatively Damaged Proteins in Escherichia coli Cells Exposed to Oxidative Stress', 273(5), pp. 3027–3032.
Tan, D. and Svenningsen, L. (2015) 'Quorum Sensing Determines the Choice of Antiphage Defense Strategy in Vibrio anguillarum', *mBio*, 6(3), pp. 1–10. doi: 10.1128/mBio.00627-15.Editor.
Thursby, E. and Juge, N. (2017) 'Introduction to the human gut microbiota', *Biochemical Journal (2017)*, 474, pp. 1823–1836. doi: 10.1042/BCJ20160510.
Turnbaugh, P. J. *et al.* (2007) 'The Human Microbiome Project', *Nature*, 449(7164), pp. 804–810. doi: 10.1038/nature06244.
Uc-mass, A. *et al.* (2004) 'An orthologue of the cor gene is involved in the exclusion of temperate lambdoid phages. Evidence that Cor inactivates FhuA receptor functions', *Virology*, 329, pp. 425–433. doi: 10.1016/j.virol.2004.09.005.
Urdaneta, V. and Casadesús, J. (2017) 'Interactions between Bacteria and Bile Salts in the Gastrointestinal and Hepatobiliary Tracts', *Frontiers in Medicine*, 4(October), pp. 1–13. doi: 10.3389/fmed.2017.00163.
Van Velkinburgh, J. C. and Gunn, J. S. (1999) 'PhoP-PhoQ-Regulated Loci Are Required

for Enhanced Bile Resistance in Salmonella spp .', *INFECTION AND IMMUNITY*, 67(4), pp. 1614–1622.
Vieira, S. M. *et al.* (2018) 'Translocation of a gut pathobiont drives autoimmunity in mice and humans', *Science*, 359(6380), pp. 1156–1161. doi: 10.1126/science.aar7201.
Virgin, H. W. (2014) 'The Virome in Mammalian Physiology and Disease', *Cell*. Elsevier Inc., 157(1), pp. 142–150. doi: 10.1016/j.cell.2014.02.032.
Walker, A. W. *et al.* (2011) 'Dominant and diet-responsive groups of bacteria within the human colonic microbiota', *The ISME Journal*. Nature Publishing Group, pp. 220–230. doi: 10.1038/ismej.2010.118.
Wang, X. and Wood, T. K. (2016) 'Cryptic prophages as targets for drug development', *Drug Resistance Updates*. Elsevier Ltd, 27, pp. 30–38. doi: 10.1016/j.drup.2016.06.001.
Weinbauer, M. G. (2004) 'Ecology of prokaryotic viruses', *FEMS Microbiology Reviews*, 28, pp. 127–181. doi: 10.1016/j.femsre.2003.08.001.
Wells, J. E. *et al.* (2003) 'Development and application of a polymerase chain reaction assay for the detection and enumeration of bile acid 7 a -dehydroxylating bacteria in human feces', *Elsevier Science*, 331, pp. 127–134. doi: 10.1016/S0009-8981(03)00115-3.
Westfall, J. M., Ringel, M. and Gardner, J. (2013) 'Incidence of Type 1 Diabetes in Finland', *JAMA*, 310(4), pp. 429–430. doi: 10.1001/jama.2013.8739.
Wexler, A. G. and Goodman, A. L. (2017) 'An insider's perspective: Bacteroides as a window into the microbiome', *Nature Microbiology*. doi: 10.1038/nmicrobiol.2017.26.
Yamasaki, Y. *et al.* (2007) 'Autoantibodies to RNA helicase A: A new serologic marker of early lupus', *Arthritis and Rheumatism*, 56(2), pp. 596–604. doi: 10.1002/art.22329.
Zhang, T. *et al.* (2006) 'RNA Viral Community in Human Feces : Prevalence of Plant Pathogenic Viruses', *PLoS Biology*, 4(1). doi: 10.1371/journal.pbio.0040003.
Zhao, G. *et al.* (2018) 'Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children', *PNAS*, 115(48). doi: 10.1073/pnas.1817913115.
Zhuang, L. *et al.* (2019) 'Intestinal Microbiota in Early Life and Its Implications on Childhood Health', *Genomics, Proteomics & Bioinformatics*. Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China, 17(1), pp. 13–25. doi: 10.1016/j.gpb.2018.10.002.
Ziegler, A. G. *et al.* (2013) 'Seroconversion to Multiple Islet Autoantibodies and Risk of Progression to Diabetes in Children', *JAMA*, 309(23).
Zoetendal, E. G., Vaughan, E. E. and De Vos, W. M. (2006) 'A microbial world within us', *Molecular Microbiology*, 59(6), pp. 1639–1650. doi: 10.1111/j.1365-2958.2006.05056.x.

8 Attachments

8.1 Quality control for the trimmed reads of each genome

8.1.1 Quality control for the trimmed reads of *B. dorei* C1P2

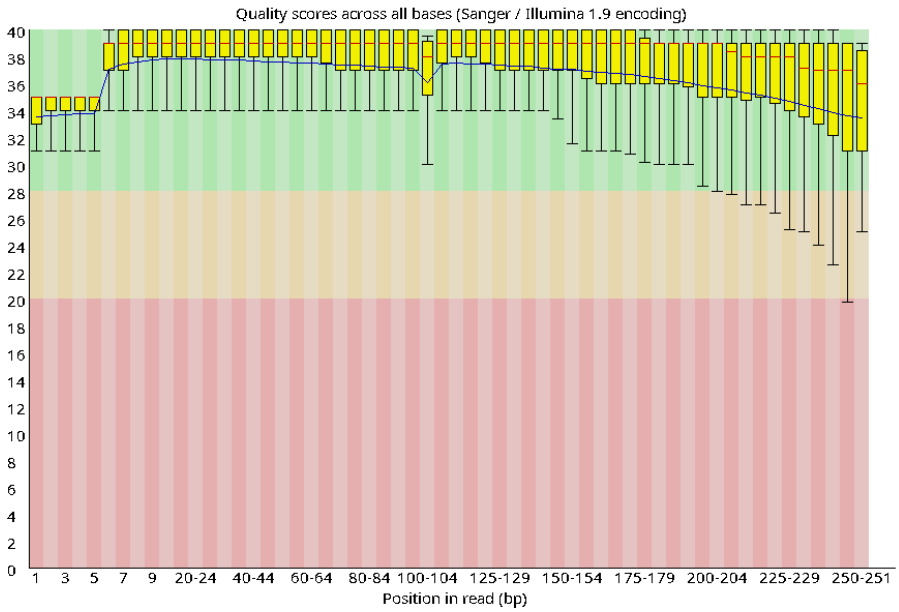

Figure 8.1 Quality control for the trimmed reads obtained from Illumina for *B. dorei* C1P2 file fastq1. Obtained with FastQC (Andrews S., 2010).

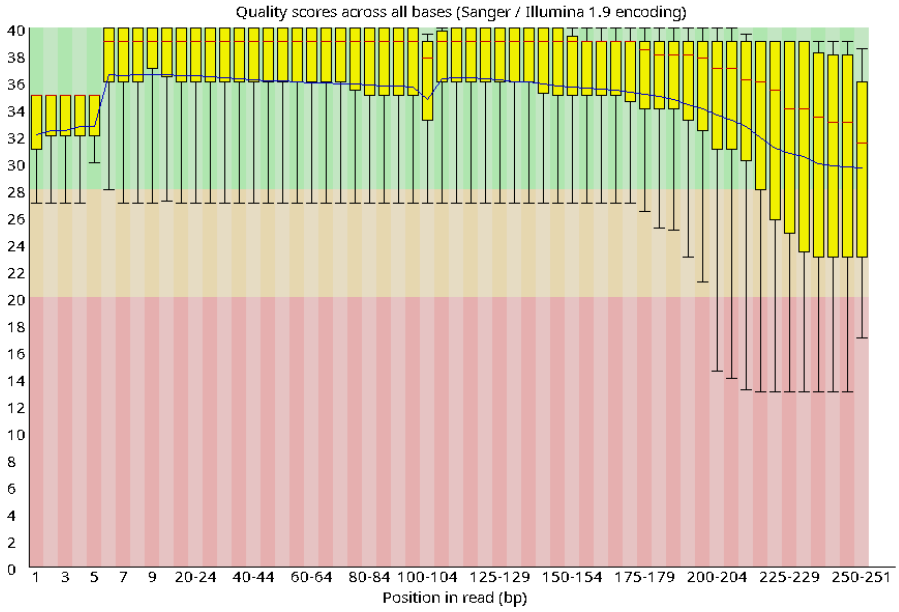

Figure 8.2 Quality control for the trimmed reads obtained from Illumina for *B. dorei* C1P2 file fastq2. Obtained with FastQC (Andrews S., 2010).

8.1.2 Quality control for the trimmed reads of *B. dorei* Sb6

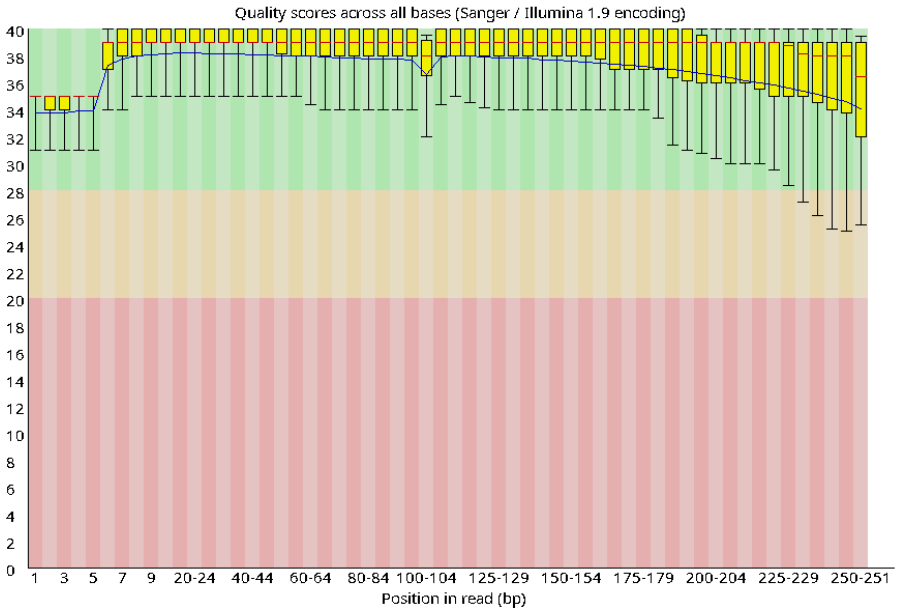

Figure 8.3 Quality control for the trimmed reads obtained from Illumina for *B. dorei* Sb6 file fastq1. Obtained with FastQC (Andrews S., 2010).

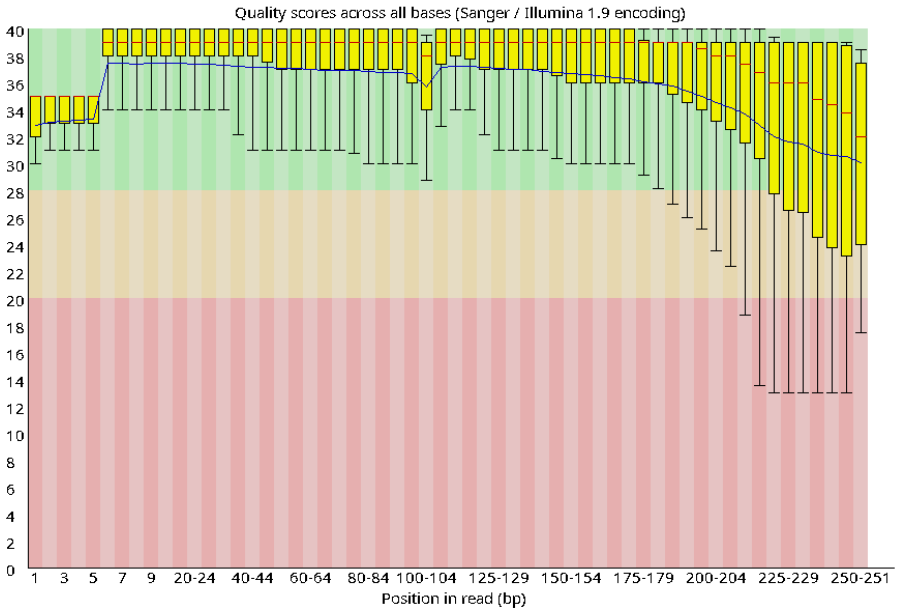

Figure 8.4 Quality control for the trimmed reads obtained from Illumina for *B. dorei* Sb6 file fastq2. Obtained with FastQC (Andrews S., 2010).

8.1.3 Quality control for the trimmed reads of *B. dorei* Sb8

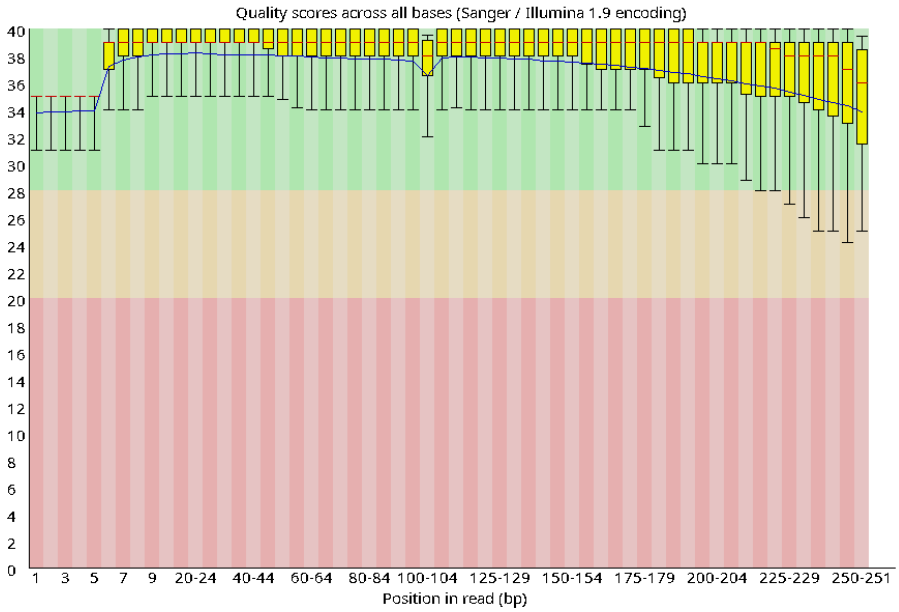

Figure 8.5 Quality control for the trimmed reads obtained from Illumina for *B. dorei* Sb8 file fastq1. Obtained with FastQC (Andrews S., 2010).

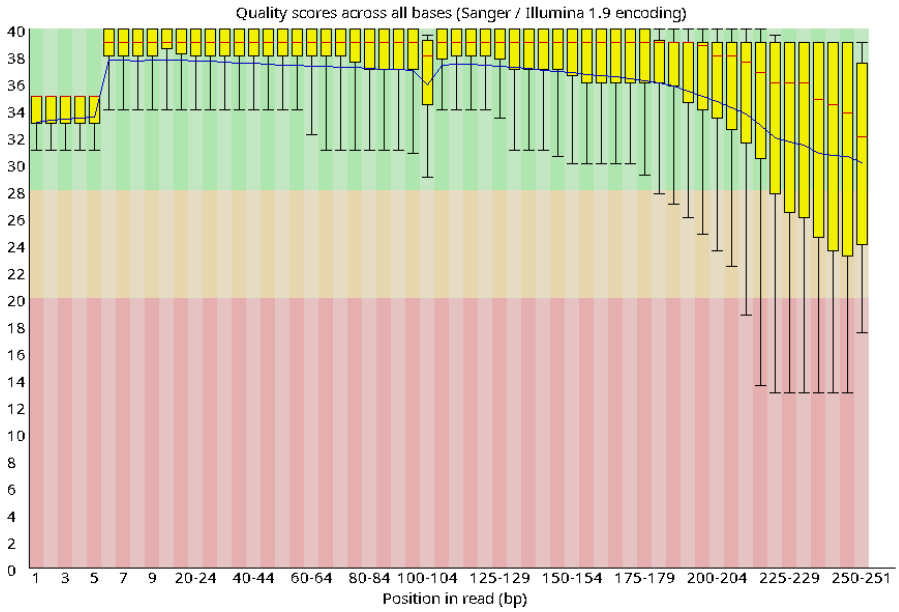

Figure 8.6 Quality control for the trimmed reads obtained from Illumina for *B. dorei* Sb8 file fastq2. Obtained with FastQC (Andrews S., 2010).

8.1.4 Quality control for the trimmed reads of *B. dorei* D1P5

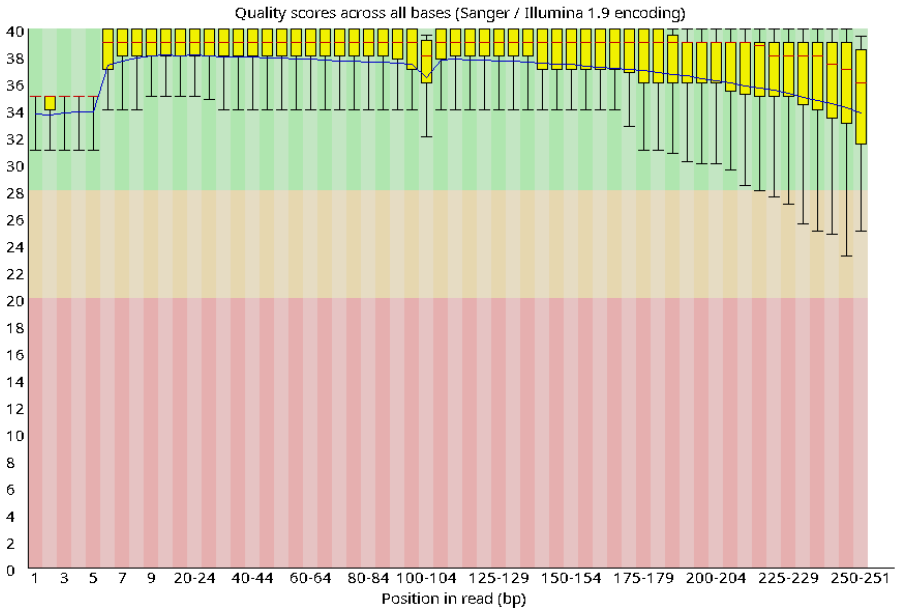

Figure 8.7 Quality control for the trimmed reads obtained from Illumina for *B. dorei* D1P5 file fastq1. Obtained with FastQC (Andrews S., 2010).

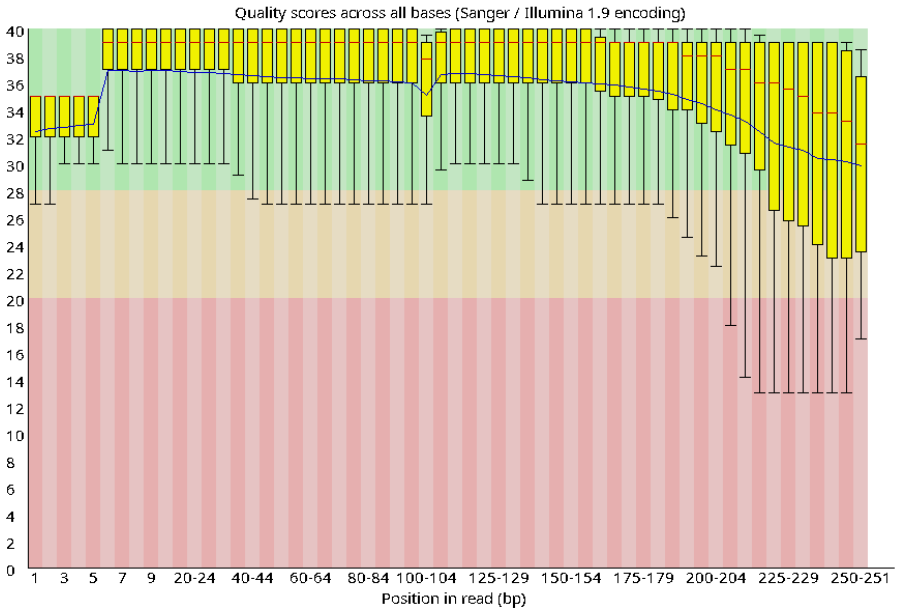

Figure 8.8 Quality control for the trimmed reads obtained from Illumina for *B. dorei* D1P5 file fastq2. Obtained with FastQC (Andrews S., 2010).

8.1.5 Quality control for the trimmed reads of *B. dorei* D8M1

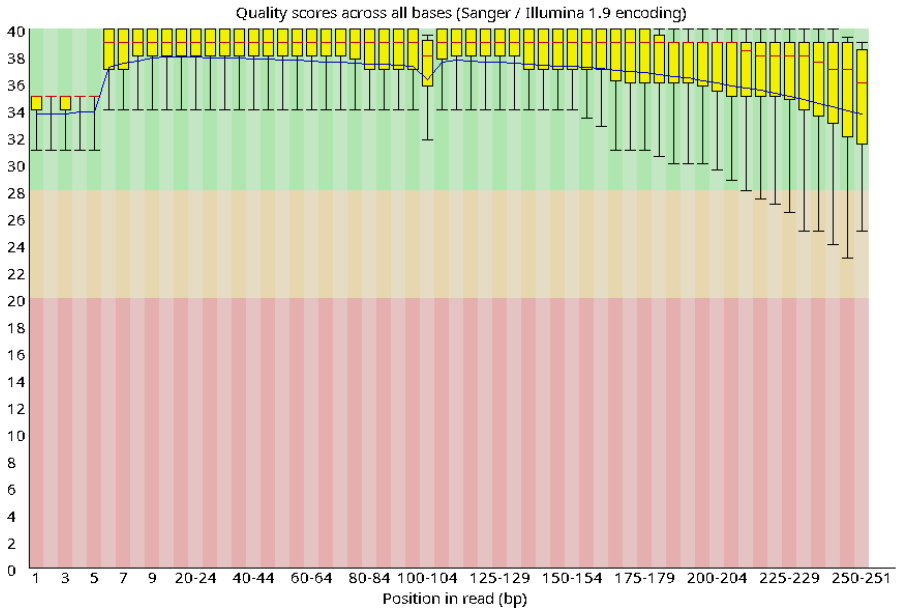

Figure 8.9 Quality control for the trimmed reads obtained from Illumina for *B. dorei* D8M1 file fastq1. Obtained with FastQC (Andrews S., 2010).

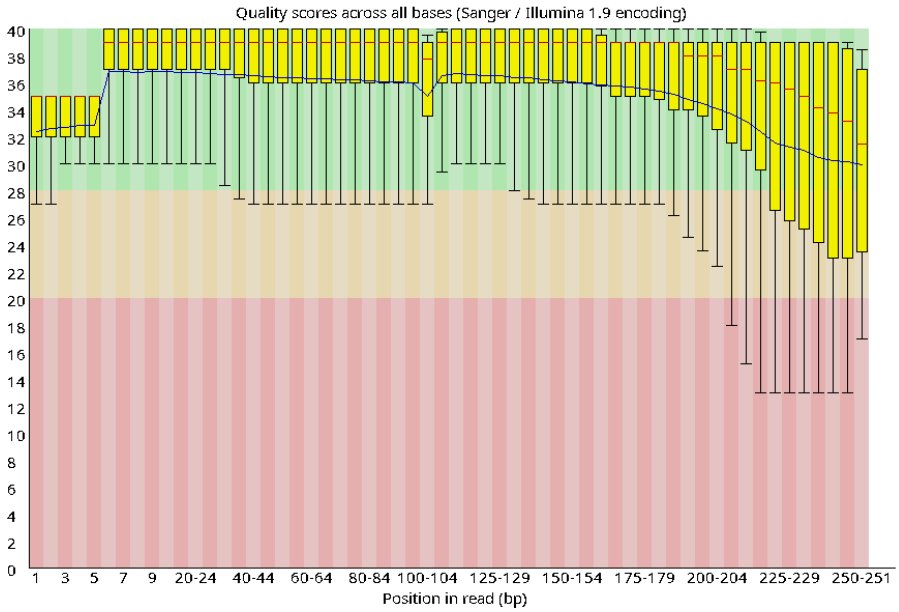

Figure 8.10 Quality control for the trimmed reads obtained from Illumina for *B. dorei* D8M1 file fastq2. Obtained with FastQC (Andrews S., 2010).

8.1.6 Quality control for the trimmed reads of *B. dorei* D16P1

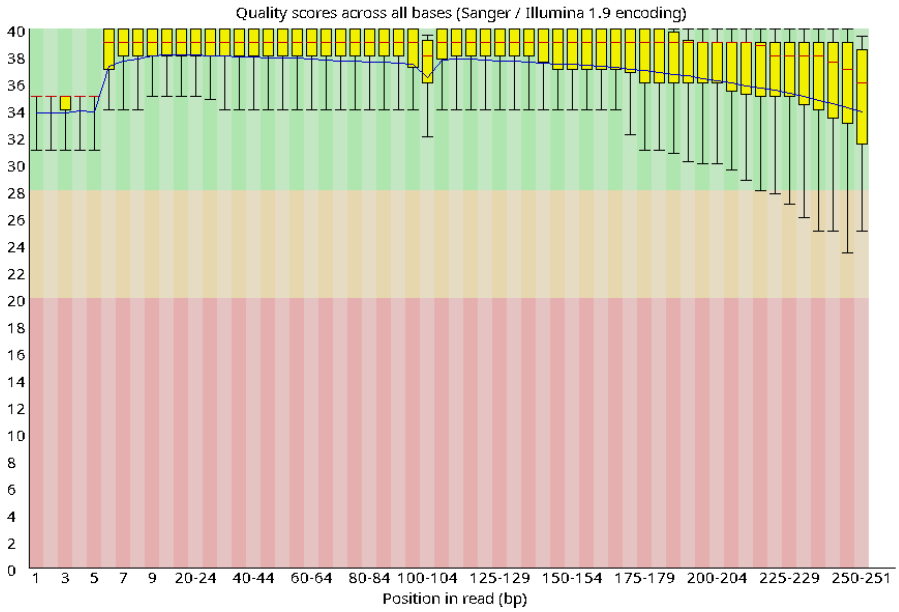

Figure 8.11 Quality control for the trimmed reads obtained from Illumina for *B. dorei* D16P1 file fastq1. Obtained with FastQC (Andrews S., 2010).

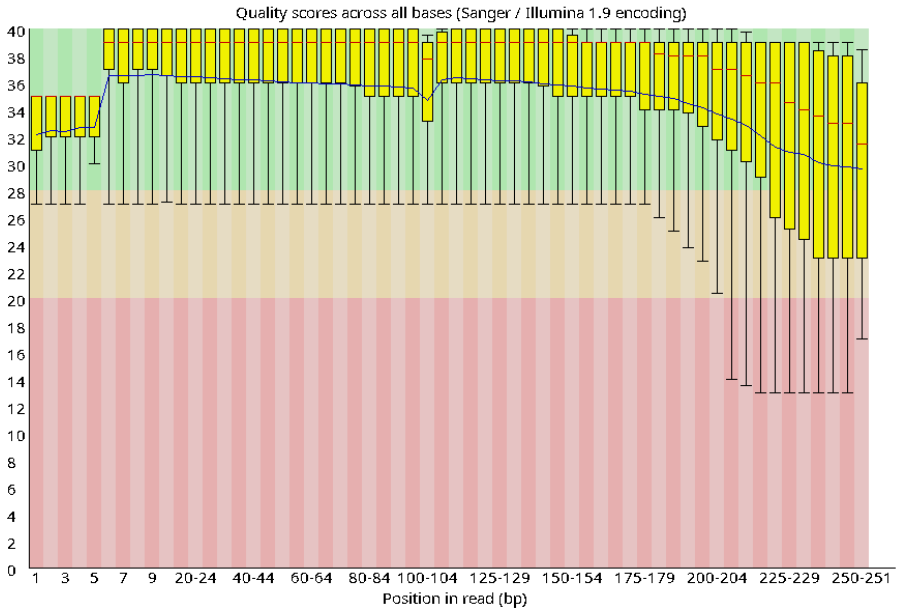

Figure 8.12 Quality control for the trimmed reads obtained from Illumina for *B. dorei* D16P1 file fastq2. Obtained with FastQC (Andrews S., 2010).

8.1.7 Quality control for the trimmed reads of *B. dorei* D16M14

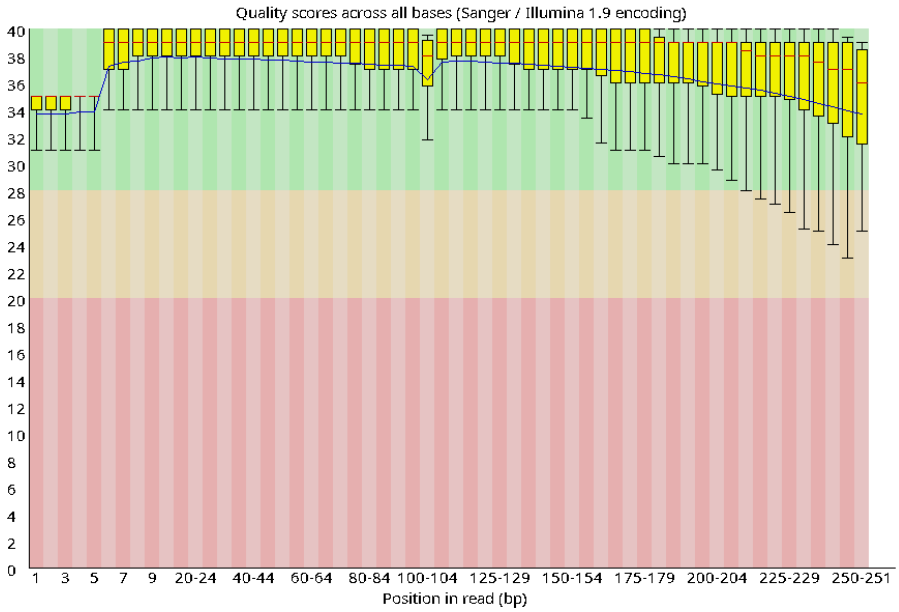

Figure 8.13 Quality control for the trimmed reads obtained from Illumina for *B. dorei* D16M14 file fastq1. Obtained with FastQC (Andrews S., 2010).

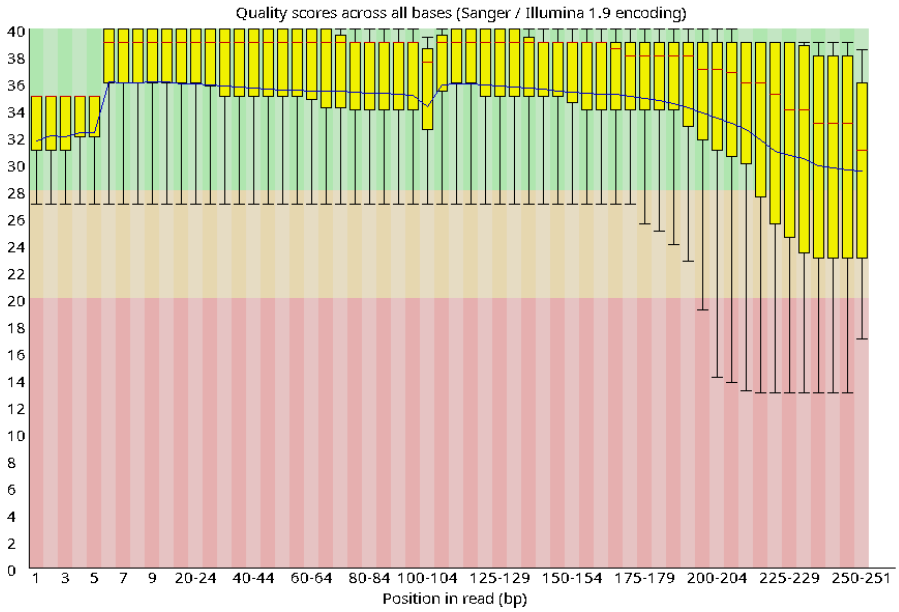

Figure 8.14 Quality control for the trimmed reads obtained from Illumina for *B. dorei* D16M14 file fastq2. Obtained with FastQC (Andrews S., 2010).

8.1.8 Quality control for the trimmed reads of *P. distasonis* D14MH1

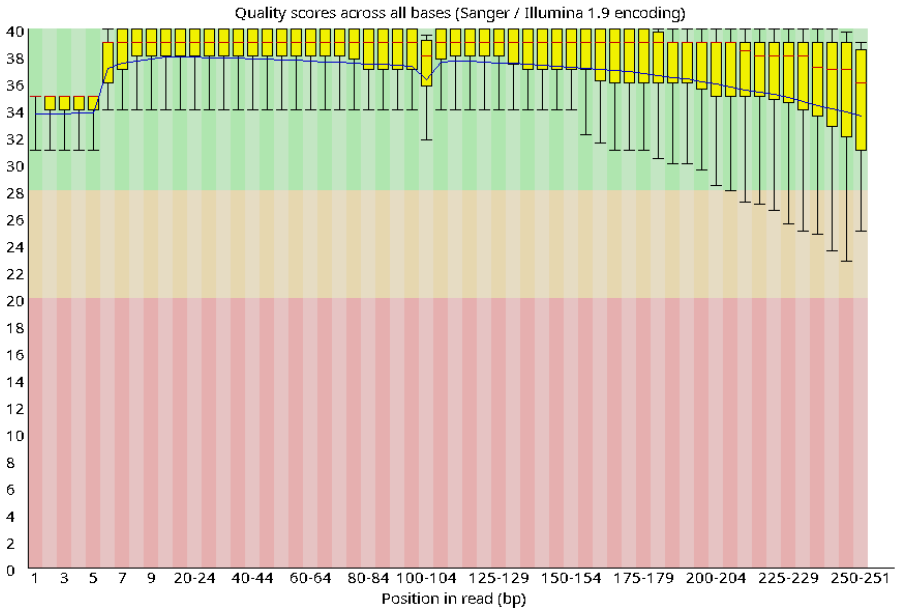

Figure 8.15 Quality control for the trimmed reads obtained from Illumina for *P. distasonis* D14MH1 file fastq1. Obtained with FastQC (Andrews S., 2010).

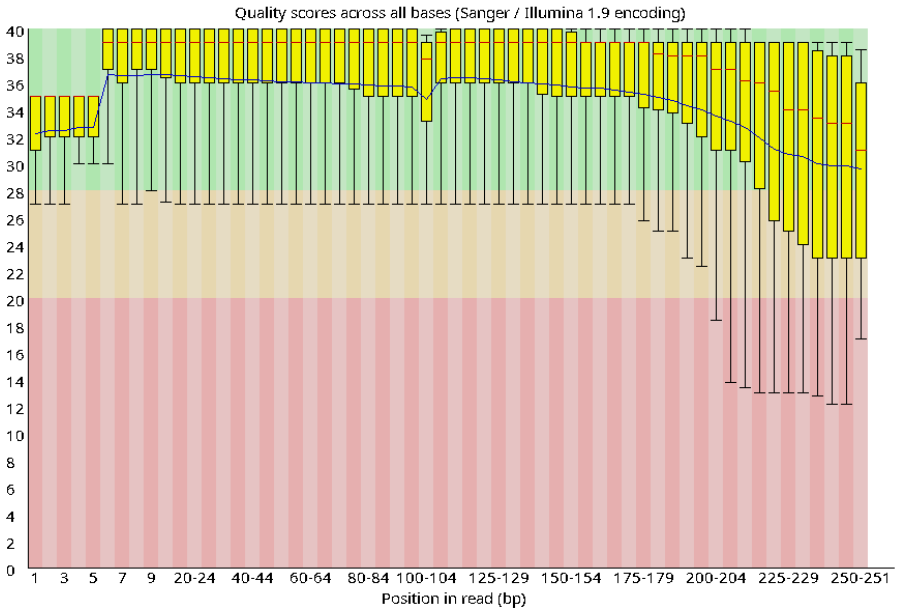

Figure 8.16 Quality control for the trimmed reads obtained from Illumina for *P. distasonis* D14MH1 file fastq2. Obtained with FastQC (Andrews S., 2010).

8.2 Induction essays control

8.2.1 Mitomycin C essays control

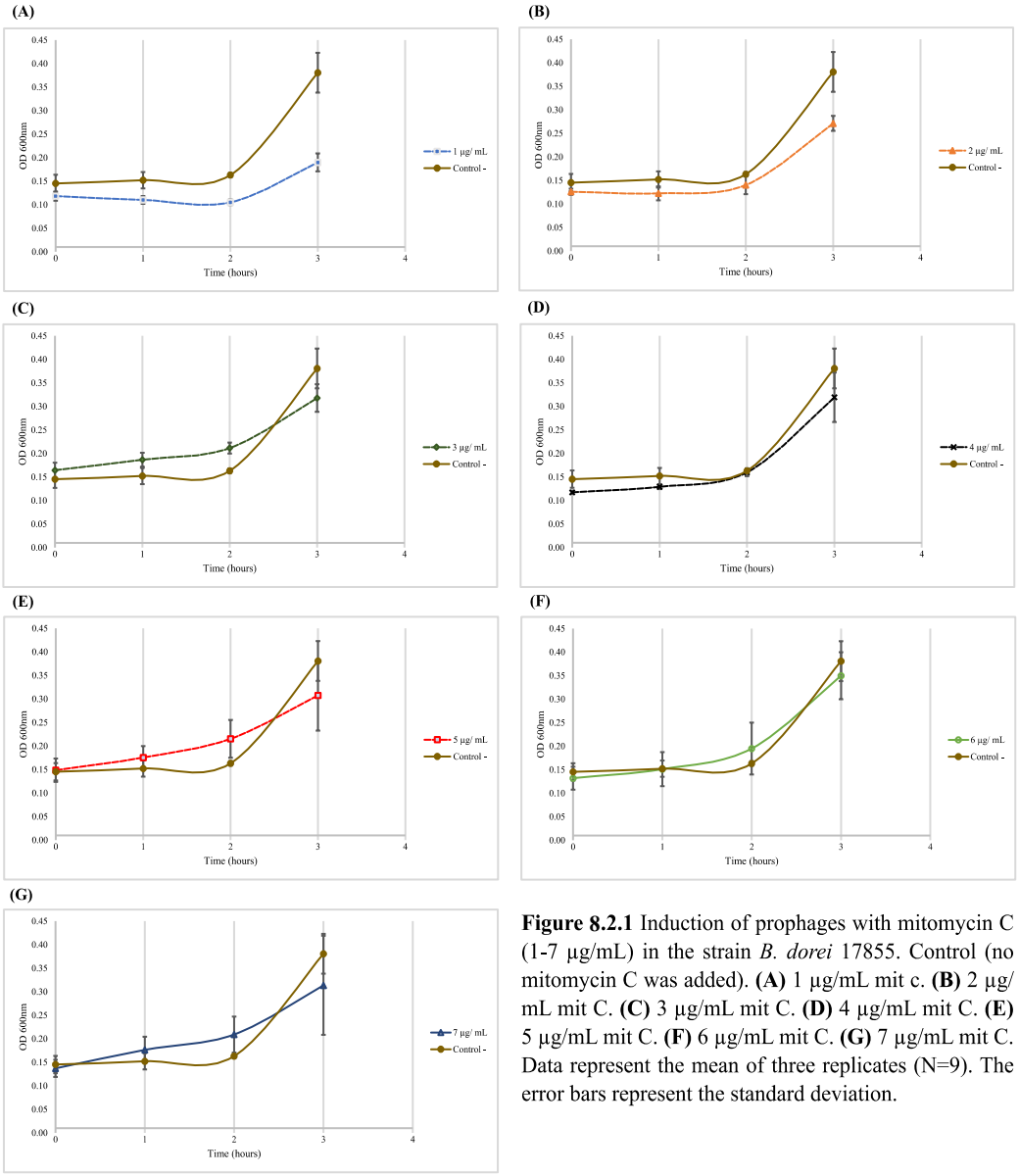

Figure 8.2.1 Induction of prophages with mitomycin C (1-7 µg/mL) in the strain *B. dorei* 17855. Control (no mitomycin C was added). **(A)** 1 µg/mL mit c. **(B)** 2 µg/mL mit C. **(C)** 3 µg/mL mit C. **(D)** 4 µg/mL mit C. **(E)** 5 µg/mL mit C. **(F)** 6 µg/mL mit C. **(G)** 7 µg/mL mit C. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

8.2.2 Hydrogen peroxide essay control

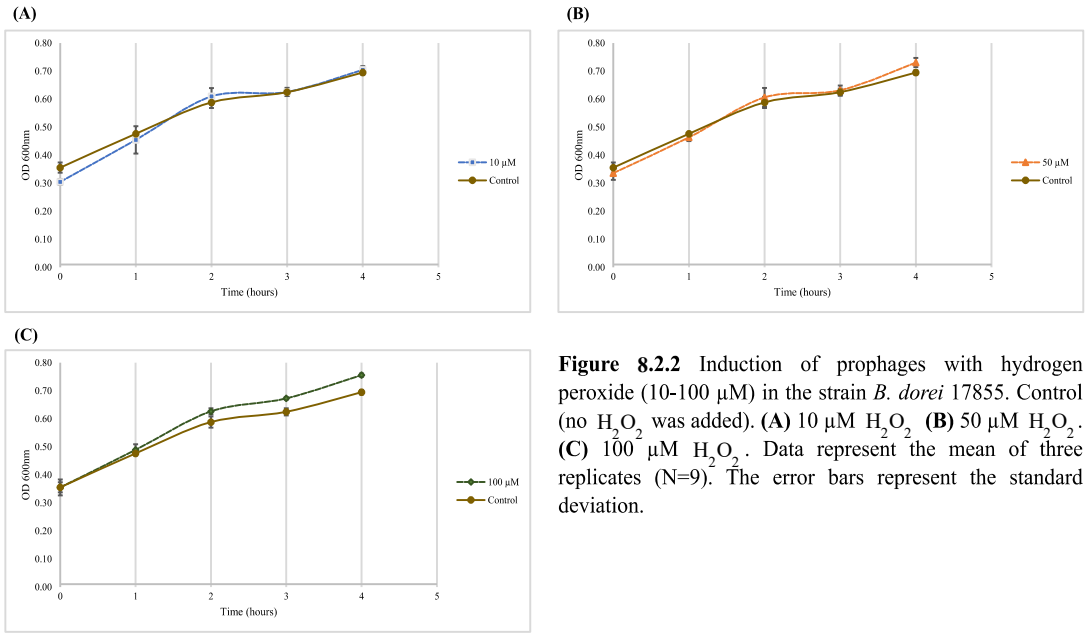

Figure 8.2.2 Induction of prophages with hydrogen peroxide (10-100 µM) in the strain *B. dorei* 17855. Control (no H_2O_2 was added). **(A)** 10 µM H_2O_2 **(B)** 50 µM H_2O_2. **(C)** 100 µM H_2O_2. Data represent the mean of three replicates (N=9). The error bars represent the standard deviation.

8.2.3 Ox-bile essay control

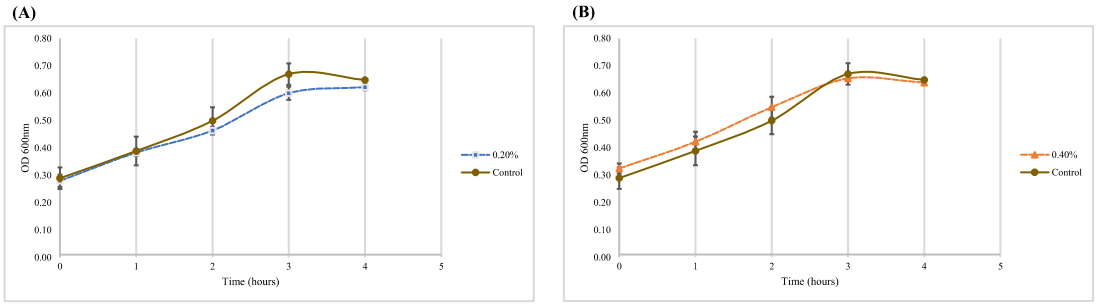

Figure 8.2.3 Induction of prophages with ox-bile (0.2%-0.4%) in the strain *B. dorei* 17855. Control (no ox-bile was added). (A) 0.2% ox-bile. (B) 0.4% ox-bile.