BACTERIAL HETEROGENEITY IN RESPIRATORY INFECTIONS

INVESTIGATIONS OF THE HETEROGENEITY OF HOST-ASSOCIATED BACTERIA DURING RESPIRATORY INFECTIONS

By

DAPHNÉE LAMARCHE, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2020)
Hamilton, Ontario (Biochemistry and Biomedical Sciences)

TITLE: Investigations of the heterogeneity of host-associated bacteria during respiratory infections

AUTHOR: Daphnée Lamarche, B.Sc. (McMaster University)

SUPERVISOR: Dr. Michael G. Surette

NUMBER OF PAGES: xv, 290

Lay Abstract

Respiratory infections are a leading cause of death worldwide, causing more than 2 million deaths in 2016 alone. Despite improving prevention strategies, infections of the lung, leading to pneumonia or more invasive infection, are and will continue to be a major burden on human life and the hospital system. In this thesis, I have studied bacteria found during respiratory infections. I investigated a type of hospital-acquired pneumonia associated with the use of mechanical ventilators (ventilator-associated pneumonia, VAP), where I demonstrated that a shared microbial profile was associated with death in patients hospitalized in the intensive care unit. These patients are part of an ongoing clinical trial to look into probiotics preventing VAP. As such, during my thesis, I evaluated the dose, composition, and stability of probiotics and placebo capsules used in this clinical trial. Finally, I studied the diversity within a group of bacteria called the *Streptococcus* Anginosus/Milleri group, which is an understudied group of bacteria in lung infections but frequent in invasive infections. The bacteria in this group are genetically diverse, which contributes to the difficulty in diagnosis and may contribute to differences in how individuals respond to these infections. This thesis demonstrated the high variability of microbes with the respiratory tract which could contribute from inflammation to death in different clinical populations.

Abstract

Respiratory infections are a leading cause of mortality and morbidity worldwide and will continue to be prevalent in the coming decades. The burden associated with acute respiratory infection is mostly due to the high prevalence of pneumonia, including hospital-acquired pneumonia. Respiratory infections are diverse with respect to clinical features, population targeted, etiology, treatment options, and development of potential complications. Several factors such as the difficulty to identify the etiology of acute respiratory infections and subsequent inappropriate therapeutic plans contribute to the complexity of treating respiratory infections, and to the development of potential complications such as pleural empyema, bacteremia and sepsis. As such, further study of the microbes entering the respiratory tract as potential pathogens or contributing to disease outcome is required.

In this thesis, I have investigated the microbiome of the critically ill as part of the PROSPECT study (NCT02462590) which examines probiotics as a prophylactic treatment to reduce ventilator-associated pneumonia in the intensive care unit (ICU). Using 16S rRNA gene profiling, we identified a novel association between hospital mortality and low microbial α-diversity observed in the lower respiratory tract. Interestingly, this microbial signature was readily identifiable within days of patient admission to the ICU. I also carried out quality control on the probiotic (*Lactobacillus rhamnosus* GG) and investigated the microbiology of adverse events with possible involvement of the probiotic. We demonstrated that the patients enrolled in the PROSPECT study received a constant dose of the probiotic and that 70 % of the placebo used in this study contains no bacteria where the remaining capsules recovered a modest number of non-pathogenic bacteria. Moreover, our results showed that rare adverse events associated with the probiotic consumption could occur, though several events were potentially due to bedside contamination not infection.

Additionally, I performed a comparative genomic study of 151 isolates of the *Streptococcus* Anginosus/Milleri group which represents under-reported respiratory pathogens that may commonly be associated with pneumonia complications such as pleural empyema. This investigation has shown that these isolates are genetically heterogenous, where diversity was observed by the presence and identity of specialized genes and genomic features (e.g. prophages, antimicrobial resistance genes, and biosynthetic gene clusters). Moreover, we further characterized two phenotypically distinct isolates recovered from a single pleural empyema aspirate. The difference in morphotype was the consequence of the presence of a single mutation in a gene involved in exopolysaccharide production.

Taken together, this work characterized the heterogeneity associated with respiratory infections. This heterogeneity, from the microbial community, bacterial group, and isolate level, contributes to the challenge of identifying but also treating respiratory infections such as pneumonia and pleural empyema.

Acknowledgements

This accomplishment was possible due to an extended group of individuals. I could not have selected a better environment for doing a Ph.D. than McMaster University. During my graduate studies, I was fortunate to be surrounded by an exceptional group of wonderful people starting with my supervisor, Dr. Michael Surette. Mike is one of the most brilliant scientists I have encountered, who fosters a collaborative environment within his group and is a generous mentor. Moreover, he is one of the few individuals who can help fix all my issues and calm my anxiety following a 5 minutes talk. I am truly grateful to have you as my supervisor. I want to acknowledge Drs. Jennie Johnstone and Deborah Cook for their tremendous support and outstanding mentoring. I also want to thank the PROSPECT team; I was truly fortunate to be part of this collaborative project. Thanks to Drs. Elena Verdú and Deborah Sloboda for their help and guidance throughout my degree.

I was privileged to work in the Surette lab with an amazing group of smart, exceptionally caring and generous people. I want to profusely thank everyone for their help and support during these times. I want to acknowledge Laura Rossi, our lab's heart and my rock during this process, I will deeply miss our chat in the PCR room. I also want to thank Drs. Fiona Whelan and Dessi Loukov, I admire in both of you your leadership, intelligence and unlimited kindness. I was fortunate to have you as models and to have your continuous support during my Ph.D.

My years spent in grad school would not have been as memorable and enjoyable without my dear 'Sass-ter' (Dr. Louis-Patrick Schenck). Sharing a big breakfast while discussing any topics was truly a pleasure. I have greatly appreciated your support, your help and your friendship during those last years. I admire your commitment to ameliorate my English competency, despite no significant improvement over the years.

Thank you to my family for supporting this bizarre career path, despite you referring to my program as Microbiochemistry. I want to acknowledge my loving uncles (Denis and Chris), who welcomed us in their home every time we needed to evade our daily Hamiltonian routine. We loved every moment spent together and our conversations during dinner. Looking forward to our next trip! Finally thank to my brother Vincent and my sister-in-law Julie (aka 'Lès HE.lfs de la .Nhuit') for hosting game night, those breaks during the writing process were essentials to keep me (and JP) sane.

I would like to thank my wonderful husband Dr. Jean-Philippe Côté, who showed me unconditional love, patience and kindness each day of this process. Thanks for being there to save me before I could drive myself crazy, this thesis exists because of you. Looking forward to us being happy idiots together for the next little while.

Finally, thanks to Charles Darwin, both the scientist and the puppy.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER ONE – Introduction

CHAPTER TWO – Microbial Dysbiosis and Mortality During Mechanical Ventilation: A Prospective Observational study

CHAPTER THREE – Probiotic quality control assessment and microbiology of adverse events in the PROSPECT trial

CHAPTER FOUR – Comparative genomics of the *Streptococcus* Anginosus/Milleri group

CHAPTER FIVE – Characterization of *Streptococcus intermedius* clinical isolates from a pediatric pleural empyema case

APPENDIX A – Appendix to Chapter 2: Supplementary Material

APPENDIX B – PROSPECT Mechanistic Sub-study

APPENDIX C – Quality assessment for the comparative genomic analysis

LIST OF TABLES

LIST OF ABBREVIATIONS

AMR	Antimicrobial resistance
BAL	Bronchoalveolar lavage
BHI	Brain heart infusion
BGC	Biosynthetic gene cluster
CDS	Coding sequence
CRP	C-reactive protein
DNA	Deoxyribonucleic acid
ETA	Endotracheal aspirate
ETT	Endotracheal tube
FDR	False discovery rate
FMT	Fecal microbiota transplantation
GA	Gastric aspirate
GI	Gastrointestinal
HCAI	Health care associated infection
ICU	Intensive care unit
Ig	Immunoglobulin
IL	Interleukin
IL-1RA	Interleukin-1 receptor antagonist
ILY	Intermedilysin
IQR	Interquartile range
JIA	Juvenile idiopathic arthritis
LPS	Lipopolysaccharide
L.r. GG	*Lactobacillus rhamnosus* GG
MOI	Multiplicity of infection
MDR	Multidrug resistant
MRS	De Man, Rogosa and Sharpe
NP	Nasopharynx
NCBI	National Center for Biotechnology Information
OP	Oropharynx
ORF	Open reading frame
OTU	Operational taxonomic unit
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PERMANOVA	Permutational multivariate analysis of variance
PE	Pleural empyema
QIIME	Quantitative Insights Into Microbial Ecology
qPCR	Quantitative polymerase chain reaction
REB	Research ethic board
rRNA	Ribosomal ribonucleic acid
SAG	*Streptococcus* Anginosus/Milleri Group
SCFA	Short chain fatty acid

SD	Standard deviation
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
THY	Todd-Hewitt with 0.5% yeast extract
TLR	Toll-like receptor
VAP	Ventilator-associated pneumonia
VGS	Viridans group streptococci

Chapter 1

Introduction

1.1. The human microbiota

The human body harbours complex microbial communities; this relationship is finely tuned, dynamic and is constantly co-evolving. Collectively, the term microbiota refers to the organisms that comprise microbial communities (fungi, bacteria, archaea, viruses) inhabiting a particular location (Sommer and Bäckhed, 2013). The total number of bacteria in the gastrointestinal (GI) tract is equal to or exceeds the number of host cells. (Sender et al., 2016; Gilbert et al., 2018). These microorganisms are normally commensals or mutualists that confer beneficial advantages to the host through a diverse range of physiological functions such as metabolism, enhancing barrier functions, protection against detrimental organisms and immune homeostasis (Natividad and Verdu, 2013; Sassone-Corsi and Raffatellu, 2015; Sommer and Bäckhed, 2013). In exchange, the host provides an ecological niche and nutrients to these microorganisms (Ubeda and Pamer, 2012). The microbiota also includes potentially 'pathogenic' organisms carried asymptomatically by the host, but that can become virulent in specific conditions and situations (Sommer and Bäckhed, 2013; Bäumler and Sperandio, 2016).

Generally, the microbiota of healthy individuals is stable over time, but modest variability can be observed (Costello et al., 2009; Lozupone et al., 2012; The Human Microbiome Project Consortium, 2012; Faith et al., 2013; David et al., 2014; Voigt et al., 2015; Oh et al., 2016), with the exception of the vaginal microbiome (in certain individuals) and the lower respiratory tract microbiome (Ma et al., 2012; Man et al., 2017). The lower respiratory tract microbiota is considered transient (Man et al., 2017) and consequently could be more prone to temporal fluctuation in its microbial

composition. Within an individual, specific site biogeography is observed across body sites and greatly impacts the microbial composition (Gilbert et al., 2018). Indeed, the site specificity of each microenvironment is driven by changes within diverse variables such as pH, oxygen concentration, temperature, humidity, available space, and density of the microbial populations (Spor et al., 2011; Dickson et al., 2014a; Oh et al., 2016; Man et al., 2017). The loss of separation is observed in diverse conditions such as critical illness (Rogers et al., 2016, 2017; Yeh et al., 2016; Lamarche et al., 2018), elderly (Whelan et al., 2014), and patients taking acid suppressants (Rosen et al., 2015), however the impact of the loss of biogeography on the host outcomes have not been intensively studied.

Several factors influence the composition and the function of the human microbiome. The major compositional drivers between individuals are host genetic, immune-microbiota interactions, and lifestyle (*i.e.* diet, medication, travel, illnesses) (David et al., 2014; Gilbert et al., 2018). Disruption of these microbial ecosystems have been frequently associated with an increased incidence of diseases such as asthma, inflammatory bowel disease, obesity, diabetes and even kidney stones (Morgan et al., 2013; Tyler et al., 2014). These microbial disturbances are commonly referred to as 'dysbiosis', which refers to an altered state in comparison to a control microbiome associated with a precise condition or illness (Hooks and O'Malley, 2017). However, with the exponential increase in popularity of microbiome studies in the last decades, the term has been wrongfully used and associated to exaggerated causality especially in human studies (Walter et al., 2020). The main limitation associated with the term 'dysbiosis' is the absence of a profound understanding and definition of a 'heathy'

microbiota (Shanahan and Hill, 2019), consequently changes or loss of specific taxa cannot be referred to ‘dysbiosis’. Several groups have referred to ‘dysbiosis’ with respect of the functional perturbations following disturbances of the microbial communities rather than focusing exclusively on the microbiota composition (The Human Microbiome Project Consortium, 2012; Lloyd-Price et al., 2016).

Prolonged perturbation of an initially balanced environment to an exacerbated state can induce detrimental health effects for the host associated to functional changes (Lloyd-Price et al., 2016). Perturbation can be stratified by intensity, acute (transient state; modest perturbations) or chronic (degraded stable state)(Lozupone et al., 2012). The microbiome is generally resilient, but if the perturbations are profound its structure, as well as function, could be severely perturbed (Figure 1.1). During these perturbations, environmental selective pressures and microbial networks are altered and consequently abundances of keystone taxa are changed and the delicate balance within host-microbiota is compromised leading to the loss of its beneficial actions for the host.

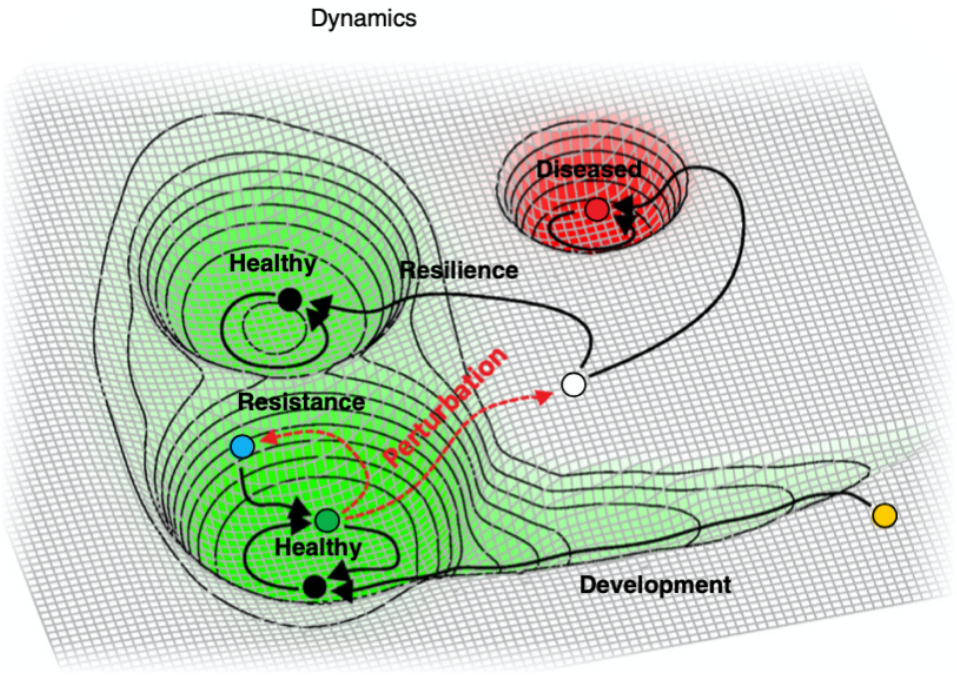


Figure 1.1: Landscape of the human microbiome. Perturbations (black and red arrows) are occurring during adult life impacting the function and structure of host-associated microbial communities. The microbiome could resist these changes (green point), could become perturbed to a transient state of dysbiosis and go to a novel healthy resilient microbiome or become severely altered, where the function is greatly perturbed (red point). The yellow point represents the (unstable) infant microbiome which further matured to a relatively stable microbiome (black point). Figure adapted from Lloyd-Price *et al* (Lloyd-Price et al., 2016). The use of the figure was permitted under the terms of the Creation Commons Attribution 4.0 International License.

1.2. Critical illness

Critical care medicine is a specialized unit dedicated to administrating specialized care to patients facing immediate life-threatening health conditions, where the main objective is preserving organ function. This unit involves a highly trained multidisciplinary team including nutritionists, physiotherapists, pharmacists and infectious disease experts amongst others (Vincent, 2013).

Critical care medicine, as a field, was first introduced in the 1950s at the University of Southern California, where one of the first intensive care units (ICU) was established (Vincent, 2013). Its proposed origin goes back to the 1850s during the Crimean War, where Florence Nightingale demanded that wounded British soldiers were treated in a separate geographical area close to the nursing station (Nightingale, 1863; Weil and Tang, 2011; Vincent, 2013). The concept of creating an environment for high-risk or severely injured patients where lifesaving interventions were administered in a somewhat personalized fashion rapidly gained popularity in developed countries. Over the years, the intensive care units evolved from essentially physical sites administrating personalized care to a modern clinical hub shaped by medical modernization in life-support technology (Vincent, 2013).

Critically ill patients are a complex and heterogeneous population of patients. Patients are admitted to this acute care facility for a myriad of reasons; however, they tend to share similar physiological problems (*i.e.* organ dysfunction) despite the reason for their clinical illness (Marshall, 2010; Dickson, 2016). The main outcomes which are survival and further preservation of quality of life once the illness is controlled are shared in the majority of patients (Marshall, 2010). Moreover, similar treatments are administered to the patients such as mechanical ventilation, nutrition therapy, prophylactic antimicrobial therapy, inotropic agents, vasopressors, and pain suppression therapy. During their stay in the ICU, patients are highly susceptible to hospital-associated complications, such as infections (Cassini et al., 2016; Vincent et al., 2020).

1.2.1. Microbiota of the critically ill patients

Considerable progress has been made for the characterization of the microbiota of critically ill patients in the last decade, following the first report investigating the microbial disturbances associated with critical care published in 1969 (Johanson et al., 1969). In this report, the authors demonstrated that the microbial communities were changing rapidly toward a dominance of gram-negative bacilli, emphasizing the potential association of the gut microbiome in certain ICU pathologies.

This field has rapidly evolved from limited research based on targeted cultured data, mainly in the context of respiratory infection, to extensive mechanistic studies associating patient outcomes to microbial profiles. Critically ill patients represent a highly heterogeneous population of patients in regard to their admission pathology, clinical interventions administered and events occurring during their stay, however, multiple independent studies have now reported similar findings on the dynamic nature of their host-associated bacterial communities. Indeed, studies using both culture-dependent and independent methods have demonstrated that critically ill patients experience pronounced disturbances of their microbial communities which become more severe over time and that microbial diversity in the GI and respiratory tracts of critically ill patients decreases following ICU admission (Zaborin et al., 2014; McDonald et al., 2016; Zakharkina et al., 2017; Berdal et al., 2007; Yeh et al., 2016; Kelly et al., 2016; Shimizu et al., 2010; Lankelma et al., 2017; Lamarche et al., 2018; Hayakawa et al., 2011; Ravi et al., 2019). The loss of diversity was more pronounced in mechanically ventilated patients who developed ventilator-associated pneumonia (Zakharkina et al., 2017). These changes

within the microbial composition of the communities are accompanied by a decrease of 'normal' members of the microbiota, which are usually considered to be health-promoting microorganisms (McDonald et al., 2016; Yeh et al., 2016; Lamarche et al., 2018). Moreover, in contrast to healthy individuals, the bacterial structures are considerably less defined between body sites, emphasizing a loss of separation between anatomical sites and disruption of the normal physiological barrier (Rogers et al., 2016, 2017; Yeh et al., 2016; Lamarche et al., 2018).

1.2.2. Major factors influencing the microbiota of the critically ill patients

Many factors (*e.g.* medical interventions, underlying diseases) can have an effect on the microbial communities within the body. Antibiotic therapy is an intrinsic part of critically ill patient care. A large international study indicated that about 70% of critically ill patients received at least one dose of antimicrobial agent during their stay (prophylactic or therapeutic; (Vincent et al., 2020)). The administration of broad-spectrum antibiotics to critically ill patients increases the risk of colonization and invasive infections (Hurford et al., 2012). It is well established that antimicrobials have a profound impact on the composition of the microbiota. Indeed, in addition to having off-target effects due to broad-spectrum activity, they could also compromise host-microbiota symbiotic relationship between members of complex microbial communities (Willing et al., 2011; Schuijt et al., 2013; Kabat et al., 2014; Estrela et al., 2015), leading to an overall loss of colonization resistance (Pamer, 2016). Moreover, inappropriate antibiotic use could lead to detrimental repercussion such as selecting for multidrug resistant microorganisms

(MDR), depletion of 'health-promoting' bacteria and hospital-associated complication in the neonatal (Rooney et al., 2020) and the adult intensive care unit (Denny et al., 2020).

Moreover, to prevent stress-related mucosal disease such as GI bleeding, prophylactic treatments to reduce gastric acid secretion or to neutralize gastric acid are commonly used in the ICU (Heyland et al., 1996; Ali and Harty, 2009). Common treatments include proton pump inhibitors (PPI), histamine-2 receptor antagonist and sulcrafate, with a preference for the use of PPIs (Daley et al., 2004; Rennke et al., 2013; Freedberg et al., 2014; MacLaren et al., 2014). PPI administration is associated with profound perturbation in the composition of host-associated communities of the GI tract potentially due to a change in pH leading to subsequent bacterial overgrowth (Verdu et al., 1994; Imhann et al., 2016). Despite the absence of pH changes in the lower GI, it has been demonstrated that there is a 3-fold increased risk of developing bacterial overgrowth in the small bowel while on PPI treatment (Freedberg et al., 2014).

Mechanical ventilation is an essential and common component of modern critical care. Although essential, the presence of an endotracheal tube (ETT) impairs first-line defence mechanisms such as the cough reflex, which leads to mucus accumulation requiring tracheobronchial suctioning. Moreover, the ETT disrupts anatomical barrier, and is associated to bacterial translocation through the cuff's leakage and subsequent biofilm formation in the majority of ventilated patients (Craven and Hjalmarson, 2010; Gil-Perotin et al., 2012; Mietto et al., 2013; Joseph et al., 2010). The presence of the ETT alters the normal composition of the microbial communities associated with the LRT (Ramirez et al., 2012; Kelly et al., 2016; Zakharkina et al., 2017). These perturbations

increase the risk of development of nosocomial infections and subsequent detrimental outcomes such as mortality in mechanically ventilated critically ill patients (Ramirez et al., 2012).

Enteral feeding has been reported to decrease the incidence of infection and mortality and ameliorates patient outcomes (Heyland et al., 1995). However, studies have also shown that the presence of a nasogastric tube (NGT) bypasses and compromises normal stimulus related to food consumption such as mastication, saliva production and peristalsis (May et al., 2005; Wu et al., 2014). Inhibition of these stimuli are involved in the decrease of stomach acidity, consequently leading to microbial overgrowth (Giannella et al., 1972; Segal et al., 2006). This bacterial overgrowth is associated with profound perturbation of the composition of the gastric microbiota (Segal et al., 2006; Smith et al., 2011). In addition, the constant opening of the gastro-oesophageal sphincter leads to bacterial translocation to the respiratory tract (Segal et al., 2006; Chapman et al., 2011), and consequently, enteral feeding through NGT increases the risk of aspiration-associated pneumonia (Segal et al., 2006; Smith et al., 2011; Silk and Quinn, 2014).

Opioids (*e.g.* morphine) are frequently prescribed for pain management in the ICU. However, opioids tend to perturb glucose metabolism and are associated with GI dysmotility (Chapple and Deane, 2018). This derangement of the patient's physiology leads to perturbations of the microbial communities, to systemic inflammation favouring bacterial translocation (Banerjee et al., 2016).

Clinical features of critical illness such as pain, catecholamine levels and underlying diseases can also promote changes within the microbial communities

(Dickson, 2016). Indeed, pain and catecholamines (*i.e.* an enzyme produced during stress) negatively impact GI motility. Moreover, catecholamines increase the mucosal inflammation, and alter the pH while promoting the growth of common ICU-associated pathogens (*P. aeruginosa*). Interestingly, a report on trauma patients has demonstrated that the gut microbiota composition was altered as early as ICU admission (Hayakawa et al., 2011).

In addition to changes within the microbial composition of host-associated communities, factors associated with critical illness could potentially promote the virulence of the microorganism. A myriad of factors could promote the virulence of microorganism within the gut microbiota (*e.g.* nutrition, metabolites, alteration within the microbiota composition, bacterial competition, host-stress response mediators) (Cameron and Sperandio, 2015; Dickson, 2016). For instance, using a murine model, morphine (Babrowski et al., 2012) and surgery-associated (Babrowski et al., 2013) injury have been associated with increased virulence phenotypes in *P. aeruginosa* leading to less mucus produced more damage on the intestinal barrier, leading to an enhance mortality rate. Increased mortality was observed in *Caenorhabditis elegans* model, where low complexity microbiota recovered from ICU patients and opioids were administered to the worms (Zaborin et al., 2014).

Essentially, all elements of the GI and LR tracts are perturbed during critical illness, affecting the patient's normal physiological processes and can in turn propagate a detrimental response for the host and disturbed the host-associated microbial communities (Dickson, 2016).

1.2.3. Impact of microbial disturbances for the patient

To the best of our knowledge, only two studies have investigated the consequences of microbial dysbiosis on illness severity and mortality in critically ill patients, including our (described in Chapter 2); however, the predictive aspect of the lung microbiota with illness outcomes have been investigated in different cohorts of patients in various conditions (Rogers et al., 2014; Takahashi et al., 2018; Leitao Filho et al., 2019; O'Dwyer et al., 2019). Dickson and colleagues have investigated outcome predictions using microbial signatures of bronchoalveolar lavage (BAL) samples with ventilator-free days (adjudicated at 28 days post-admission;(Dickson et al., 2020)). They demonstrated that specific microbial signatures (absence of *Lachnospiraceae*) and lung bacterial burden (using 16S rRNA gene copy numbers) were associated with fewer ventilator-free days. Contrary to our findings, they did not observe an association between α-diversity (*i.e.* diversity within sample) and poor outcomes in their cohort of patients. Interestingly, both studies have used specimens collected early during the patient's ICU stay, emphasizing that the airway microbiome could be used to predict disease progression rapidly following ICU admission and perhaps allow for stratification of patients based on risk.

Investigations of the predictive component of GI microbiota are more numerous. Indeed, microbial diversity of the GI tract has been associated with outcomes in patients undergoing allogeneic hematopoietic stem cell transplantation, severe inflammatory response syndrome, and in high-risk patients admitted to the ICU (Shimizu et al., 2010; Taur et al., 2014; Harris et al., 2016; Iapichino et al., 2017). By contrast, in two studies

with ICU patients, no association between survival and microbial diversity of the GI tract was observed (Lankelma et al., 2017; Freedberg et al., 2018). However, the prevalence and abundance of specific members of the GI microbiota (*e.g. Enterococcus*) is associated with poor outcomes (*i.e.* mortality and infections; (Taur et al., 2012; Freedberg et al., 2018))

Severe perturbations ('dysbiosis') impact the structure and composition of host-associated microbial communities in critically ill patients (Figure 1.2). The loss of health-promoting bacteria, and high inflammation due to underlying conditions associated with the acute care setting lead to a loss of colonization resistance. Moreover, the catecholamines and the proinflammatory mediators released by critically ill patients can directly contribute to promoting the growth of prevalent ICU-associated bacteria (*Pseudomonas aeruginosa, Staphylococcus aureus, Streptococcus pneumoniae*) (Huffnagle and Dickson, 2015). In addition, normal host defence mechanisms (gastric pH, mucociliary clearance, physical barrier breached due to the utilization of invasive devices) are altered in critically ill patients, contributing to bacterial overgrowth and the dominance of a single pathogen over time. Indeed, several studies have demonstrated a decrease in abundant organisms or loss of 'normal' host-associated taxa, including the Lachnospiraceae family, *Faecalibacterium*, and *Blautia* genera in the GI tract and *Veillonella, Prevotella*, and *Neisseria* genera in the lower respiratory tract in the ICU cohort (McDonald et al., 2016; Yeh et al., 2016; Lamarche et al., 2018). Several of those depleted microorganisms are known to confer host advantages, such as anti-inflammatory and nutritional benefits via the production of short-chain fatty acid (SCFA;(Sokol et al.,

2008; Arpaia et al., 2013; Levy et al., 2017)). The level of fecal SCFA is drastically decreased upon admission to the ICU (Hayakawa et al., 2011; Yamada et al., 2015). Moreover, perturbation of the indigenous microbiota could lead to harmful repercussions and allow infection by opportunistic secondary potential pathogens such as *C. difficile*, *Candida albicans* or facultative anaerobic gamma proteobacteria such as *Pseudomonas* (Sommer and Bäckhed, 2013).

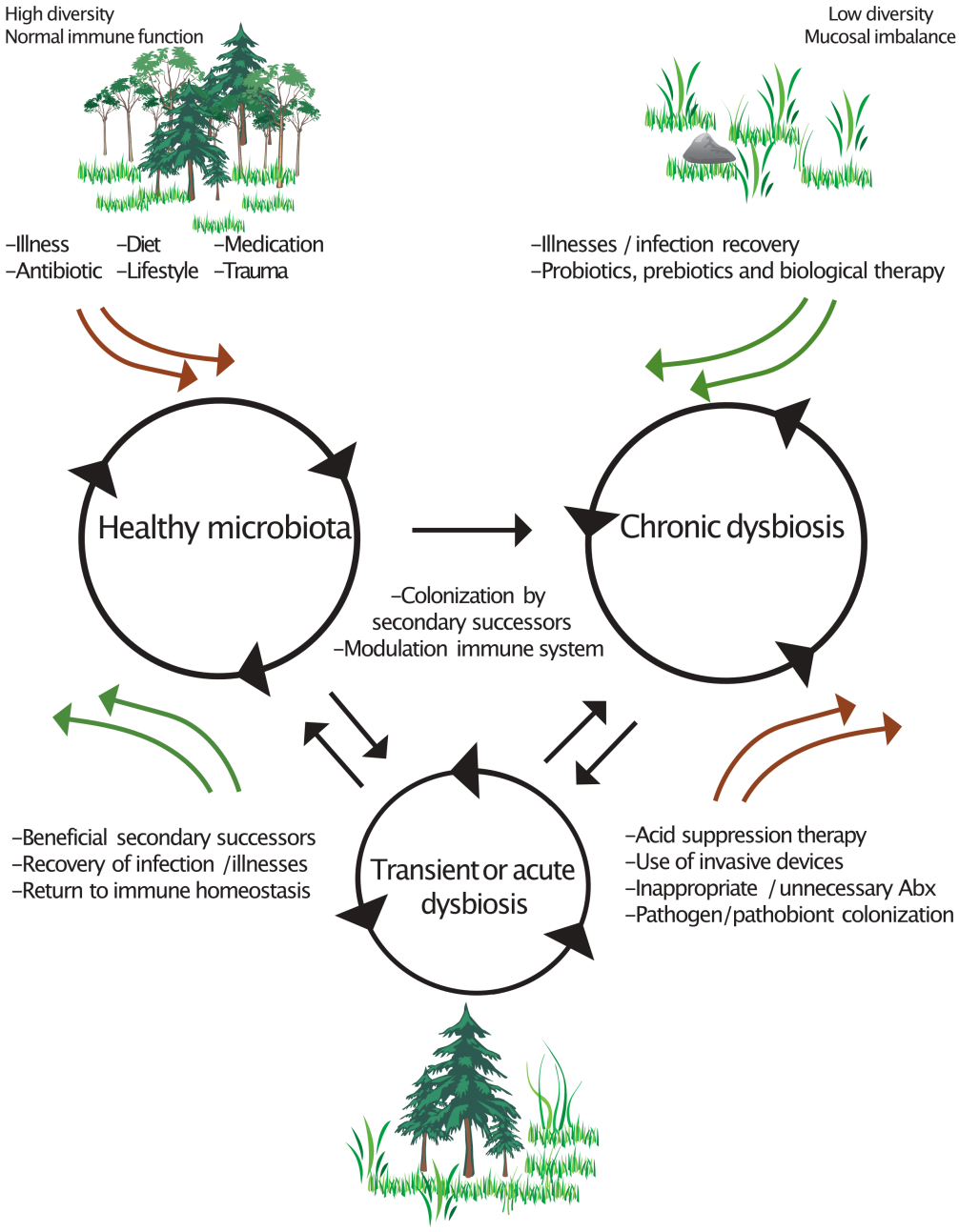


Figure 1.2: Perturbations in the microbiome during health and disease. Despite a relative stability in a healthy individual, numerous factors may trigger alteration of the microbiome such as diet, lifestyle, infection, and in this case critical illness. Depending on the gravity of the perturbations, the composition of the microbiota will be altered. Short administration of antibiotics will lead to an acute dysbiosis while persistent stressors such as critical care will lead to chronic perturbations. Functions dependant on the microbiota such as immune homeostasis or metabolism will be disturbed in either state. However, the recovery time will be longer in the chronic 'dysbiosis' due to the greater perturbation of the host-associated communities and their associated functions.

1.2.4. Therapies targeting the microbiota in critically ill patients

Increasing evidence suggests that therapeutics targeting the microbiota could be a promising avenue to prevent illnesses associated with critical care, despite having established the mechanism of action for some of these treatments (Haak et al., 2017). Therapies targeting the microbiota investigated in ICU patients includes, but are not limited to, probiotics, prebiotics, synbiotics, selective digestive tract decontamination, and fecal microbiota transplantation. Probiotics administration has shown to be a promising strategy to prevent detrimental harm associated to critical illness and is being intensively investigated, therefore probiotics will be further discussed in the next section.

Prebiotics are a dietary component associated with promoting the growth of health-promoting microorganisms. Dietary fiber (*e.g.* non-digestible polysaccharide) is a type of prebiotic which has been investigated in critically ill patients. Their consumption have been associated with a decrease in enteric infections and ventilator-associated pneumonia (VAP) in sepsis cases (Shimizu et al., 2018), systemic inflammatory response syndrome and trauma patients (Shimizu et al., 2013). Moreover, their administration have been associated to an increase in short-chain fatty acid (SCFA) producing microorganism (Fu et al., 2020). SCFAs (*i.e.* butyrate, propionate and acetate) are the most abundant metabolite produced from the fermentation of dietary fiber by certain members of the GI microbiota (Levy et al., 2017). ICU patients have a drastic decrease in SCFAs level rapidly following their ICU admission (Yamada et al., 2015). A combination of prebiotics and probiotics, called synbiotics, have shown a significant reduction in neonatal sepsis

and lower respiratory tract infections in a pediatric cohort in India using a combination of *Lactobacillus plantarum* and fructooligosaccharide (Panigrahi et al., 2017).

Selective digestive tract decontamination (SDD) aims to prevent/eradicate colonization of the oropharynx and the GI tract by potential pathogenic microorganisms, and consequently limit subsequent infections. SDD usually includes an aggressive application of topical, non-absorbable antimicrobials (such as tobramycin, colistin, amphotericin B) in the oropharynx and directly in the stomach and intravenous administration of cephalosporins (Plantinga et al., 2018). Studies have demonstrated a reduction of nosocomial infections and a potential decreased in mortality following SDD (D'Amico et al., 2009; Silvestri et al., 2012; Cavalcanti et al., 2017; Plantinga et al., 2018). Despite the fact that an increase in MDR microorganisms has not been observed in a recent metanalysis (Daneman et al., 2013), a modest study has recently suggested that colonization of the gut by antibiotic-resistant bacteria may occur following ICU discharge (Buelow et al., 2017) and further studies will be needed to comment on the long-term effects of implementing SDD in the ICU, especially in countries with high-MDR rates (Cavalcanti et al., 2017).

Fecal microbiota transplants (FMT) have been recently investigated in the critically ill patient to prevent antibiotic-induced diarrhea and other ICU-associated illnesses such as sepsis. These reports usually investigate a limited number or patients or are presented as case studies, but the results are generally positive (Li et al., 2015; Wei et al., 2016; Dai et al., 2019). However, despite those results, the same safety issue related to FMT such as the potential transfer of pathogens or MDR organisms still exist (Merrick et

al., 2020). The FDA has recently suspended several trials to investigate safety standards following the hospitalization of 4/6 subjects and the death of two patients (DeFilipp et al., 2019; U.S Food and Drug Administration, 2020). Consequently, FMT needs to be further investigated, especially in this vulnerable population of patients before it becomes common use in the ICU.

1.2.5. Probiotics in critical care units

A probiotic is defined as a live microorganism thought to confer a health benefit when administered in adequate amount (Hill et al., 2014). Probiotics have several general strategies to improve the health of the host such as colonization resistance, modulation of the host immune system, protection against physiological stress, modulation of the microbiome and enhancement of the host's epithelial barrier function (Suez et al., 2019). For instance, probiotic strains can compete with pathogenic bacteria for niche colonization, this can happen through changing the conditions of the microenvironment such as decreasing the pH, competing for limiting nutritional resources, and by producing antimicrobial molecules (*e.g.* bacteriocin) or toxic metabolites (*e.g.* hydrogen peroxide), resulting in an increase in host colonization resistance (Plaza-Diaz et al., 2019; Suez et al., 2019). Certain probiotics possess the ability to promote the expression of immune-related genes, and further impact inflammatory marker levels (IL-6, IL-8, TNF-α, Il-1β, C-reactive protein; (Thomas and Versalovic, 2010; Suez et al., 2019)). However, despite their health-promoting potential, defining precise mechanisms of action is challenging (Plaza-Diaz et al., 2019). Probiotics could also improve barrier functions by upregulating

tight-junction proteins, promoting mucus production, and increasing beneficial metabolite levels (Ohland and MacNaughton, 2010). However, mechanism of action greatly varies depending on the strain and the product formulation (Reid, 2016). For instance, *Lactobacillus rhamnosus* GG (LrGG), the most intensely investigated commercially available probiotic, adheres strongly to mucus due to the SpaCBA pili (Kankainen et al., 2009). This pili allows cells to persist longer in contrast to other probiotic strains, and could be potentially immunomodulatory (Segers and Lebeer, 2014).

The efficacy of probiotics has been vigilantly reviewed and demonstrated with various levels of quality evidence for several infections such antibiotic-associated diarrhea (Hempel et al., 2012; Videlock and Cremonini, 2012), necrotizing enterocolitis in preterm infants (Deshpande et al., 2010; Thomas et al., 2017), *Clostridioles difficile*-associated diarrhea (Goldenberg et al., 2017), acute upper respiratory tract (Hao et al., 2015) and VAP (Bo et al., 2014). Various meta-analyses have shown that the use of probiotics has different success rates in different contexts and that their impact may vary depending on the strain, the methodologies used and the population studied (Van den Abbeele et al., 2011; Crooks et al., 2012; Suez et al., 2019).

In the critically ill patient population, a decrease in the incidence of ventilator-associated pneumonia (Petrof et al., 2012; Bo et al., 2014; Weng et al., 2017) and overall ICU-associated infections (Manzanares et al., 2016) have been observed following probiotic administration; however, most probiotic clinical trials are deemed low quality partly due to a limited number of participants or varying/inconsistent methodologies (*e.g.* different strains, dose, duration, or administration route) used during the trial. Moreover,

probiotic administration represents a cost-saving and cost-effective strategy to prevent VAP, *C. difficile*-associated diarrhea and antibiotic-associated antibiotic (Lau et al., 2020).

Despite being generally well tolerated and deemed safe in the general 'healthy' population (Salminen et al., 2002), health concerns have been raised about the potential risks of developing probiotic-associated infections. Indeed, probiotic-related infections are uncommon, however reports of *Lactobacillus*-associated infections do exist (Doron and Snydman, 2015). This highlights the potential risks for vulnerable patients such as the critically ill population (*i.e.* immunocompromised, compromised gastrointestinal tract, multiples underlying illnesses) where safety evidence is more conflicted (Besselink et al., 2008; Hempel et al., 2011; Didari et al., 2014; Yelin et al., 2019). For instance, the infamous clinical trial PROPATRIA demonstrated, in critically ill patients with severe pancreatitis, that mortality was increased in the probiotic arms in comparison to the placebo (Besselink et al., 2008). Trials reporting such negative outcomes associated to probiotics consumption in the ICU are rare, and several meta-analyses have since then suggested they are generally safe and that they should be considered in critically ill patients (Hempel et al., 2011; Manzanares et al., 2016). Moreover, an extensive range of probiotics have been investigated in the ICU and their safety seems comparable, with the exception of *Saccharomyces boulardii* which could potentially be associated with an increased risk for fungemia and consequently should not be used as probiotics in the ICU (Lherm et al., 2002). However, randomized controlled trials (RCTs) investigating this topic need to be aware of the potential complications associated to probiotics, implement

suitable investigation of their safety within the study infrastructure and properly declare adverse event association to their administration.

1.3. Health care-associated infections

Critically ill patients are more susceptible to developing health care-associated infections (HCAIs) during their stay (Cassini et al., 2016), with prevalence of HCAI of 10-51% (Vincent, 2003, 2009; Magill et al., 2018; Vincent et al., 2020). A recent report has demonstrated in an international cohort that 54% of critically ill patients had suspected or proven infections (Vincent et al., 2020) and the mortality rate in patients with a suspected or proven infection was 30% (Vincent et al., 2020) HCAIs are infections acquired following a visit or admission to a facility while receiving treatment. This definition applies to infection occurring 48 hours post-admission to the hospital/health care facility or 30 days after discharged/treatment has been administered (Horan et al., 2008; World Health Organization, 2011). HCAIs represent the most common adverse event associated with hospitalization and their prevalence in acute care hospitals ranged from 3.2 to 12% in high-income countries (Magill et al., 2018; Suetens et al., 2018; Mitchell et al., 2019) with a pooled incidence of 7.6% (World Health Organization, 2011). In Canada, the prevalence of hospitalized patients with an HCAIs was 7.9% in 2017 (Mitchell et al., 2019). HCAIs constitute a major threat for patient safety and are associated with increased length of stay, increased financial cost, and higher risk of long-term detrimental impact (World Health Organization, 2011; Tessier et al., 2019).

HCAIs include but are not limited to bloodstream infections, pneumonia, urinary tract infections, surgical site infections, and *Clostridioides difficile* infections. The causative agents of HCAIs are broad and include organisms that are often carried asymptomatically by individuals (e.g. *Streptococcus pneumoniae, Staphylococcus aureus*) and opportunistic pathogens acquired from the hospital environment (Khan et al., 2017). Common risk-factors for HCAIs include age, emergency admission to the ICU, a length of stay > 7 days, use of invasive devices (catheter, endotracheal tube), and undergoing surgery amongst others (World Health Organization, 2011; Haque et al., 2018).

1.3.1. Ventilator-associated pneumonia

Ventilator-associated pneumonia (VAP) is one of the most common health-care associated infections (World Health Organization, 2011; Edwardson and Cairns, 2019) in critically ill patients and has the highest mortality rate among nosocomial infections in the ICU (Safdar et al., 2005). The number of adult VAP cases are estimated to be around 4,000 per year in Canada (Muscedere et al., 2008) and is associated with longer ICU and hospital stay (Safdar et al., 2005; Kollef et al., 2012). Unsurprisingly, this association is translated to a great economic burden, where studies have estimated the costs associated to VAP range between $10,000 and $40,000 (Muscedere et al., 2008; Kollef et al., 2012; Zimlichman et al., 2013). In Canada, this cost is estimated to be $11,450. (Muscedere et al., 2008).

VAP is defined as a pneumonia that occurs more than 48 hours post endotracheal intubation and mechanical ventilation (Kalanuria et al., 2014). Despite this simple

statement, there is a lack of a clear and established definition in the community, which greatly complicates its diagnosis (Joseph et al., 2010). However, physicians generally agree that VAP should be suspected if new or persistent infiltrates are observed on chest radiography, and if two or more of the following occurred: temperature > 38.3°C, blood leucocytosis or leucopenia, pus in the tracheal aspirate (Kalanuria et al., 2014).

Similarly to HCAIs, infections can also originate from the patient's microbiota and translocate easily to the lower respiratory tract due to the presence of an endotracheal tube (ETT; (Kalanuria et al., 2014; Soussan et al., 2019)). Moreover, exogenous sources of transmission are various and include microbes residing on health care workers, in the hospital environment, and other ICU patients (Huang et al., 2006; Johani et al., 2018; Costa et al., 2019).

Aerobic Gram-negative bacilli, most commonly *Pseudomonas aeruginosa*, are the most common etiological agent recovered from airways samples from infected patients (Chastre and Fagon, 2001; Kalanuria et al., 2014) but the causal agents of VAP are diverse and are not limited to bacteria (Table 1.1). Furthermore, in approximately 30-70% of the cases, VAP is the result of a polymicrobial infection (Carlet, J Torres, 2001). The causative agent seems to depend on the duration of mechanical ventilation. Early-onset VAP is usually caused by *Staphylococcus aureus*, *Streptococcus pneumonia* and *Haemophilus influenza*, while late-onset VAP is frequently caused by Gram-negative bacilli, which are often MDR. Strains can include *Pseudomonas aeruginosa*, *Acinetobacter* species, *Enterobacter* species or methicillin-resistant *Staphylococcus*

aureus and are generally associated with higher mortality rates (Kalanuria et al., 2014; Vincent et al., 2020).

1.3.2. Preventions

Despite the high prevalence of HCAIs infections in hospitalized patients, a significant portion can be prevented (Cassini et al., 2016; Johnstone et al., 2019a). Indeed, the Canadian adverse event study has suggested that 70 000 of the 185 000 hospital admissions associated with an adverse event could have been potentially prevented (Baker, 2004). Prevention methods included the implementation of educational interventions to health care workers, protocols for hand hygiene, environmental decontamination, and the use of personal protective equipment (Safdar and Abad, 2008; Loveday et al., 2014; Yokoe et al., 2014; Saint et al., 2016). Indeed, transmission of microorganisms to the patients by the health-care workers represent a major source of contamination (Pittet et al., 2006). Consequently, a thorough hand wash, cleaning surfaces and refraining from handshakes (Ghareeb et al., 2013; Loveday et al., 2014) have the potential to significantly decreases the incidence of HCAIs.

Table 1.1: Summary list of VAP etiological agents and associated prevalence (Joseph et al., 2010; Kalanuria et al., 2014)

Common causes (prevalence)	Rare/Unsual causes
Gram-Positive cocci	Gram-Positive bacilli
Staphylococcus aureus (20.4%)	*Corynebacterium* species
Coagulase Negative staphylococci (1.4%)	*Listeria monocytogenes*
	Nocardia species
Streptococcus pneumonia And other streptococci (12.1%)	*Serratia* species
Enterococci	Aerobic Gram-Negative bacilli
Enteric Gram-Negative bacilli (14.1%)	*Hafnia alvei*
Escherichia coli	*Stenotrophomonas maltophilia* (1.7%)
Klebsiella species	*Burkholderia cepacia*
Enterobacter species	Gram-Negative cocci
Proteus species	*Neisseria* species (2.6%)
Citrobacter species	*Moraxella* species
Non-Fermentative Gram-Negative bacilli	Anaerobic bacilli
Pseudomonas species (24.4%)	*Bacteroides* species
Acinetobacter species (7.9%)	*Fusabacterium* species
Gram-Negative bacilli	*Prevotella* species
Haemophilus species (9.8%)	*Actinomyces* species
Fungi	Anaerobic cocci
Candida species	*Veillonella* species
	Peptostreptococci
	Viruses
	Influenza and other respiratory viruses
	Herpes simplex viruses
	Cytomegalovirus

The implementation of these preventive measures have successfully decreased the incidence of HCAIs in the last decade (Magill et al., 2018; Johnstone et al., 2019a; Mitchell et al., 2019). Indeed, the Canadian Nosocomial Infection Surveillance program has demonstrated a 30.1% decline in the prevalence of HCAIs between 2009 and 2017

(Mitchell et al., 2019). However, Magill and colleagues emphasized the increased need for more methods in particular against pneumonia (VAP) and *C. difficile* infections where no reduction has been observed in the United States (Magill et al., 2018). Moreover, in Canada, device-associated infection, such as VAP, accounted for 35.6% of HCAIs (Mitchell et al., 2019) and further investigation into prevention strategies to specifically address this issue is needed. Several recommendations have been proposed to prevent VAP such as head bed elevation, administering oral care, limiting the duration of mechanical ventilation, and performing subglottic secretion suctioning (Ramirez et al., 2012). These strategies aim to reduce colonization of the respiratory tract to further prevent translocation of the microorganisms (Joseph et al., 2010).

1.4. Current Limitations for the identification of a causative agent during respiratory infections

Historically, during the pre-antibiotic era, the identification of respiratory infections, such as pneumonia, was achieved in > 90% of the cases, and the etiology was predominantly recognized as *S. pneumoniae* (Bullowa, 1935; Austrian R, 1971). However, following the golden era of antibiotics discovery and the availability of treatment, the interest for identifying the etiology of pneumonia has drastically declined (Bartlett, 2011). A retrospective study from 2009 including 17,000 hospitalized patients in the United States of America have reported that the etiological agent of community-acquired pneumonia was found in below 10% of the cases (Bartlett, 2011).

Numerous studies have demonstrated that an etiological agent for acute respiratory tract infections is only determined in roughly 50% of the cases despite current intensive diagnostic tests (Capelastegui et al., 2012; Psallidas et al., 2014; Jain et al., 2015). In the critical care population, a positive microbiology result is obtained between 50 and 75% depending mostly on the type of pneumonia, where non-ventilated pneumonia and early VAP are associated with a lower rate of identification (~ 50%; (Luyt et al., 2018)). This difference is potentially attributable to the availabilities of relevant sample collection limiting contact with the oropharynx microbiota, which is easier in VAP patients due to the presence of the ETT. Using transtracheal aspiration in non-ventilated patients seems to be promising to identify a potential causative agent (Bartlett, 1977).

Following standard clinical microbiology protocols, many organisms are simply not recovered, and others are dismissed as commensals, greatly impacting the identification of the causative microbial agent. Several causes, in addition to sampling collection protocols, are suspected for this low recovery rate of causative agents for respiratory infections. Indeed, the use of empiric antibiotics prior to sample collection could explain the culture-negative diagnosis (Jain et al., 2015; Gadsby et al., 2016). Moreover, this low recovery suggests the possibility that 'traditional' pathogens do not account for the majority of bacteria associated infections, and that standard clinical microbiology protocols could miss or dismiss 'unconventional' microorganisms with the current regimens and assays. Moreover, with the tremendous progress made in modern medicine such as chemotherapy and clinical interventions (mechanical ventilation,

catheters, etc..), we observed an increased prevalence of vulnerable patients, with impaired immunity (*e.g.* patients with AIDS, cancers, critical illnesses, cystic fibrosis). This population is highly susceptible to respiratory infections and is more likely to have an atypical etiology such as *Pneumocystis jirovecii* pneumonia subsequent to HIV infection or other opportunistic pathogens (Pirofski and Casadevall, 2015).

Another underreported group of microorganisms during respiratory infection is the *Streptococcus* Anginosus/Milleri group (SAG) from the viridans group streptococci. In addition to pneumonia, SAG strains are associated with invasive infections at various body sites such as pleural empyemas, abscesses, and bloodstream infections (Issa et al., 2020).

1.5. The Streptococcus Anginosus/Milleri group

1.5.1. General information

The viridans group streptococci (VGS) is a large heterogeneous group of microorganisms, that includes roughly all the streptococci which are not comprised in the common groups A (*S. pyogenes*; Invasive infections) and B (*S. agalactiae*; a common cause of neonatal sepsis, "strep throat" and pneumonia) streptococci or is not a pneumococci (*S. pneumoniae*) (Doern and Burnham, 2010). Consequently, members of the VGS are genetically heterogeneous and does not share a particular phenotypic trait (Jensen et al., 2013). There are currently 30 different species included in the VGS, divided into 6 different groups*; S. mitis* group*, S. sanguinis* group*, S. bovis* group*, S. salivarius* group*, S. mutans* group, and the *Streptococcus* Anginosus/Milleri group (SAG; (Doern and

Burnham, 2010)). Despite *S. pneumoniae* being a member of the *S. mitis* group, the microorganism is not usually referred to as a member of the VGS since it can be easily distinguished from that group, and rather is classified as pneumococci (Arbique et al., 2004).

The SAG is composed of three species: *S. anginosus, S. intermedius, S. constellatus* (Whiley and Beighton, 1991). SAG species were first identified in 1956 by O. Guthof and referred to as the *Streptococcus* Milleri group to represent this collection of pathogenic VGS strains (Guthof, 1956). The SAG forms a distinct phylogenetic group of streptococci, however, isolates within the SAG are phenotypically diverse, displaying different hemolytic properties, Lancefield group antigen, and production of hydrolytic enzymes (Grinwis et al., 2010a). Despite being a heterogeneous group of microorganisms, members within the SAG share common traits. Indeed, they are usually forming small colonies producing a distinctive butterscotch odour and share the abilities to hydrolyze arginine, produce acetoin from glucose, and the inability to ferment sorbitol (Gossling, 1988a; Doern and Burnham, 2010). Moreover, they are usually slow growers, which tend to prefer anaerobic conditions, and if using aerobic conditions, their growth usually requires CO_2 (Ball and Parker, 1979; Pulliam et al., 1980).

1.5.2. Clinical significance

SAG isolates can asymptomatically colonize various mucosal surfaces (gastrointestinal, urogenital and respiratory tract) and are also recovered during several types of infections across various anatomical sites. Population-based studies investigating the prevalence of

SAG isolates are scarce, especially in the asymptomatic community. Reports have suggested that SAG can be recovered in 15-30% of healthy individuals (Poole and Wilson, 1979; Gossling, 1988a). Moreover, the SAG group is among the most common cause of invasive streptococcal diseases in humans causing more invasive infection than groups A, B, C and G taken together. Indeed, the study performed in the Calgary region in Alberta over a three-year period suggested that the annual incidence for SAG-associated invasive infection was 8.65/100,000 population (Laupland et al., 2006). Another population based-study performed in Israel in a tertiary care hospital over 37 months recorded 346 patient-unique isolates and estimated the prevalence of SAG-associated infections to be 8.8/10,000 of hospital admissions (Siegman-Igra et al., 2012). *S. anginosus* was identified in 82% of the isolates, followed by *S. constellatus* (11%) and *S. intermedius* (6%). Interestingly, they demonstrated that SAG was an underreported causative agent of UTI, where 29% of the isolates were recovered from this type of infection (Siegman-Igra et al., 2012).

A recent Canadian study performed in the general population in British-Columbia, demonstrated that the SAG-associated bacteremia annual incidence is 3.7/100,000 of the population (Laupland et al., 2018). Several other reports have associated SAG to pyogenic infections (characterized by the presence of pus) and have demonstrated the frequent isolation of SAG isolates from abscesses and deep-seated infections at various body sites, pneumonia, bloodstream infections, pleural empyema, and from cystic fibrosis patients during and between pulmonary exacerbations (Gossling, 1988a; Molina et al., 1991; Claridge et al., 2001; Ahmed et al., 2006; Laupland et al., 2006; Parkins et

al., 2008; Sibley et al., 2012; Siegman-Igra et al., 2012; Laupland et al., 2018; Darlow et al., 2020). Moreover, *S. anginosus* and *S. constellatus* were recovered as the numerically dominant organisms in the lower respiratory tract of 15/45 critically ill patients in the FINDPATH study (unpublished). SAG-associated infections are frequently polymicrobial with anaerobic bacteria (*Bacteroides, Prevotella, Peptostreptococcus*), other streptococci, and *Staphylococcus* species amongst others (Gossling, 1988a; Molina et al., 1991; Claridge et al., 2001; Siegman-Igra et al., 2012; Fazili et al., 2017; Furuichi and Horikoshi, 2018a). Interestingly, in a report investigating 60 SAG-associated infections, *S. intermedius* was recovered as the sole microorganism in 89% of the cases, while *S. anginosus* and *S. constellatus* infections were more likely polymicrobial (Claridge et al., 2001).

Despite their pathogenic potential being demonstrated since the first report mentioning the SAG, the entire clinical spectrum has not been widely studied.

1.5.3. Risk factors associated with SAG infections

Several factors contribute to the host's susceptibility to developing infections such as the host's age, sex, immunity, microbiome, genetics, diet and history (past exposure), as well as the infectious inoculum size, and the environment (Casadevall and Pirofski, 2017). For SAG-associated infections, age, sex and prevalence of comorbidities have been reported as risk factors.

The presence of underlying comorbidities seems to be the predominant risk factors for SAG infections, where several reports have shown that >50% of the patients

had a least one underlying conditions (*e.g.* diabetes, cancer, heart-related conditions, heavy alcohol consumption) (Gossling, 1988a; Molina et al., 1991; Siegman-Igra et al., 2012; Suzuki et al., 2016; Laupland et al., 2018; Issa et al., 2020). Moreover, without directly showing association with infection rate, a study has shown that the saliva of alcoholics was enriched for *S. anginosus* (Morita et al., 2005).

The incidence of SAG infections seems to be distributed across all ages; however, it tends to be more prevalent in older individuals. Interestingly, SAG bloodstream infections were neither associated with any cases in infants in one study in Canada nor found to be predominant in the 'very' old group of patients (Laupland et al., 2018). Indeed, the median age reported for several studies was between 35-58 years old (Siegman-Igra et al., 2012; Fazili et al., 2017; Laupland et al., 2018; Issa et al., 2020; Jiang et al., 2020). However, for *S. intermedius* specifically, 17.8% of the cases were in the older age groups (61-70) but only 2% in the > 80 years old group. and the median patient age was 45 ± 24 years old (Issa et al., 2020).

A sex predominance towards male individuals SAG has generally been reported (Molina et al., 1991; Jacobs et al., 1995; Laupland et al., 2006; Suzuki et al., 2016; Fazili et al., 2017; Furuichi and Horikoshi, 2018a; Laupland et al., 2018), however, two population-based studies (Israel and England) did not detect sex preferences (Weightman et al., 2004; Siegman-Igra et al., 2012).

1.5.4. Infectious preferences across species

Evidences indicating that specific species of SAG are enriched in precise illnesses are conflicted and generally SAG do not seem to show any pronounced preferences for infection sites (Siegman-Igra et al., 2012; Junckerstorff et al., 2014). However, reports could suggest that *S. constellatus* tends to be isolated from blood while *S. intermedius* is recovered from infections targeting the central nervous system (Whiley et al., 1992; Jacobs et al., 1995; Belko et al., 2002; Siegman-Igra et al., 2012). *S. anginosus* is less associated with abscesses than *S. constellatus* and *S. intermedius* isolates (Claridge et al., 2001). A recent report has indicated that *S. intermedius* was the most prevalent etiology associated with brain abscess, even surpassing the established (since 1960) most common cause, staphylococci (Darlow et al., 2020).

Interestingly, it has been demonstrated that of the SAG species, *S. intermedius* was associated with worse prognosis, longer hospital stay and higher mortality rates (Junckerstorff et al., 2014). This suggests that although the SAG is forming a distinct phylogenetic group, differences still exist within this group of microorganisms, emphasizing the importance of accurately determining the species while studying this group.

1.5.5 Identification of SAG isolates in a clinical setting

The SAG is part of the viridans group streptococci, which is largely associated with members of the upper respiratory tract microbiota, consequently distinguishing SAG isolates from other streptococci in usually complex and non-sterile environments (*e.g* sputum) is challenging (Ruoff, 1988). Moreover, standard media used for culturing

respiratory specimens (*e.g.* Columbia blood agar) as part of the clinical microbiology protocol does not permit the growth of certain SAG isolates or in some instances are not quantitatively reliable and drastically underestimate its bacterial burden (Figure 1.3; (Sibley et al., 2010a)). In this study, the implementation of a novel type of medium optimized for the recovery of SAG (McKay media) from the standard cystic fibrosis sputum microbiology protocols, demonstrated that SAG isolates were prevalent in the patients' sputum in this specific population (Sibley et al., 2010a).

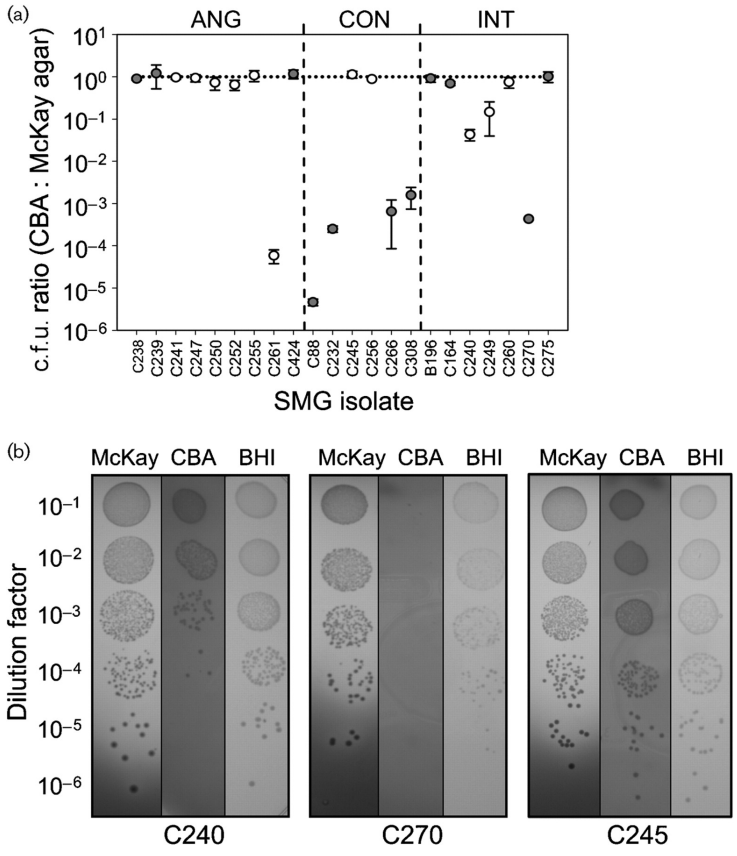


Figure 1.3: Limitations of standard clinical microbiology protocols for the recovery and the quantification of SAG isolates. Mckay media is superior to Columbia blood agar (media commonly used in the clinical laboratory) to recover SAG isolates. Dark and open circles are representing airway and invasive isolates respectively (A). Examples of SAG with variable recovery levels (B) (Sibley et al., 2010a). The use of the figure was permitted under the terms of the Creation Commons Attribution 4.0 International License.

In addition to the recovery limitation due to improper culture conditions, when the SAG isolates are successfully recovered, they are frequently dismissed due to an absence of hemolysis on media supplemented with sheep (or horse) blood. Classically, the clinical microbiology protocols prioritized the investigation of β-hemolytic (*i.e.* producing a complete hemolytic zone of clearance) streptococci (Parkins et al., 2008), however, only about 16% of the strains (38% *S. constellatus*, 12% *S. anginosus* and 7% *S. intermedius*) are β-hemolytic on animal blood media (Whiley et al., 1990). More recently, our group has shown that ~ 40% of the SAG strains (1/61 *S. intermedius*, 46/58 *S. constellatus, S. anginosus* 23/57) were β-hemolytic (Grinwis et al., 2010a), this was in concordance with another report indicating that 34% of the 518 SAG isolates were β-hemolytic (Jacobs et al., 1995). Previous work from our group has demonstrated that a minimum of three tests (Lancefield group, hyaluronidase, and chondoitrin sulfatase production) are required to identify to the species level the SAG isolates (Grinwis et al., 2010a). This suggests that even though biochemical and phenotypical assays are available, those are time-consuming in clinical setting.

Outside the clinic, multilocus sequence analysis based on 7 genes (*map* encoded for a methionine aminopeptidase, *pfl*, a pyruvate formate lyase, *ppaC,* an inorganic pyrophosphatase, *pyk,* a pyruvate kinase, *rpoB,* an RNA polymerase subunit, *sodA,* a superoxide dismutase, and *tuf,* an elongation factor T) has been successful to discriminate between species (Bishop et al., 2009; Jensen et al., 2013)) however its application in the clinic is not realistic. Matrix-associated laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry which is becoming a popular technology in

clinical microbiology to identify bacteria is a promising avenue to pursue for identifying SAG isolates. A recent report has shown that 92% of the 212 SAG isolates included were correctly identify using this technique(Arinto-Garcia et al., 2015). However, this technique still requires growing the bacteria and select the isolates to test on the plate.

Due to all these issues in the clinical microbiology laboratory, the SAG is an unrecognized pathogen during respiratory infections, which is frequently misidentified as a harmless bacterium and as a contaminant for the upper respiratory tract and, consequently, the clinical significance associated to this group is potentially underestimated

1.5.6. Divergent recovery of SAG isolates between respiratory and invasive infections

SAG isolates are under-reported in pneumonia but recovered frequently from pleural empyema, with an incidence between 32 and 52% in Canada (Ahmed et al., 2006; Finley et al., 2008). Pleural empyema (PE) is an infection of the pleural cavity characterized by the presence of pus within the pleura. PE due to bacteria is generally a consequence of unappropriated treated pneumonia. Interestingly, up to 57% of patients with pneumonia are susceptible to develop a pleural effusion that may progress to a pleural empyema (Wrightson and Davies, 2010). Pneumonia caused by SAG isolates tends to exacerbate often (54-73%) to pleural effusion (Okada et al., 2013; Noguchi et al., 2015) more often than other caused including *Streptococcus pneumoniae*, a microorganism frequently associated with lower respiratory tract infections (Okada et al., 2013). Pleural empyema caused by SAG isolates are associated with a good prognosis (83% survival at 12 months

post-infection; (Maskell et al., 2006)). Interestingly, a similar prognosis is observed in culture-negative PE and the authors suggested that this could be potentially SAG-associated PE (Maskell et al., 2006). A small report has demonstrated that the causative agents of PE were identified in 33% of the time from sputum specimens, interestingly these cases were concordant with the results of the pleural fluid culture (Kobashi et al., 2008).

Accordingly, the divergent prevalence of SAG isolates between pleural empyema and lower respiratory infections represents a paradox (Brims et al., 2010) and is likely due to the limitations of the clinical laboratory mentioned above. Indeed, while the identification of an etiological agent in lower respiratory infections is complicated by the presence of a diverse community of microorganisms inhabiting the upper respiratory tract, its identification in usually 'sterile' environments such as the pleural cavity is simple since the microorganism recovered is presumably the culprit. Moreover, anaerobic culture is usually performed for the detection of microorganisms for PE but is not traditionally performed for pneumonia.

1.5.7. Comparative genomics analysis

The first whole bacterial genome sequencing experiment was performed on *Haemophilus influenzae* in 1995, and required a budget of > $1M (Fleischmann et al., 1995). Since then, sequencing technologies have significantly progressed to a place where sequencing whole bacterial genomes can be done within hours and at a negligible cost. A recent publication for our group presents an innovative cost-effective protocol to optimize the

quality of the *de novo* assembly and consequently improve the quality of the downstream analysis (Derakhshani et al., 2020). This protocol combines the high reliability of short reads obtained from Illumina technology with the completeness of long read sequenced via PacBio (Pacific Biosciences) more prone to sequencing errors.

Accordingly, due to immense technological progress in this filed, comparative genomics using sequencing data is now becoming broadly affordable, rapid and popular (Loman et al., 2012; Besser et al., 2018). Comparative bacterial genome analysis using next-generation sequence data is a powerful tool for several applications such as performing in-depth pathogenicity, evolution and population studies as well as genomic epidemiology amongst others (Guthrie and Gardy, 2017; Besser et al., 2018). Moreover, this type of analysis is more accessible to people with limited bioinformatics skills with the development of an extensive range of software (Edwards and Holt, 2013).

This approach has been used to investigate SAG isolates. While numerous genome announcements presenting a simple overview of the genomes are available (Planet et al., 2013; Rahman et al., 2015, 2016), a few studies have published a more comprehensive analysis using comparative genomics with a single strains of SAG (Srinivasan et al., 2014; Hasegawa et al., 2017a) or multiple isolates (Olson et al., 2013; Issa et al., 2019). In collaboration with our group, Olson *et al.* sequenced and assembled 7 high-quality closed genomes and performed an analysis including 59 publicly available *Streptococcus* genomes (18 SAG isolates). They demonstrated that SAG isolates were rich in mobile elements primarily integrative conjugative elements and bacteriophages and accounted for about 10% of the SAG genomes. They furthermore supported the

previously proposed phylogeny of the SAG and demonstrated that *S. anginosus* was the most diverse species within the SAG (Olson et al., 2013). Issa and colleagues focused their analysis toward the species *S. intermedius* and common virulence factors and demonstrated that the type seven secretion system (T7SS) modules were present on different genomic island, and that 47% of the isolates included in their analysis were carried a resistance gene to nisin, an antimicrobial peptide.

1.5.8. Pathogenicity

SAG species possess a variety of potential virulence factors; adhesins, hemolysin (intermedilysin, streptolysin S-like), hydrolytic enzymes (hyaluronidases, deoxyribonucleases, chondroitin sulfatases, neuraminidases), biofilm, capsule, bacteriocin, and the type 7 secretion system (Nagamune et al., 1996; Kanamori et al., 2004; Grinwis et al., 2010a; Olson et al., 2013, 2013; Asam and Spellerberg, 2014; Asam et al., 2015; Mendonca et al., 2017; Whitney et al., 2017; Issa et al., 2020). The expression of virulence factors tends to be strain and species-specific and consequently fluctuates across SAG isolates (Grinwis et al., 2010a; Olson et al., 2013).

For instance, intermedilysin (ILY), a potent human-specific cytolysin in the family of cholesterol-dependent cytolysins, is produced uniquely by *S. intermedius* (Nagamune et al., 1996; Giddings et al., 2004). The host selectivity of ILY is due to its selectivity for human hCD59. Indeed, the receptor is the glycoprotein hCD59, which is involved in protection against the complement pathway in host cell membranes. CD59 regulates the terminal steps of the alternate complement pathway where its primary role is

to inhibit the formation of the membrane attack complex on host cells by binding the complement proteins (Wickham et al., 2011). ILY is an important virulence factor for *S. intermedius*. Indeed, an *ily* deletion mutant shows less adherence, invasion and cytotoxicity on human hepatic cells (Sukeno et al., 2005) and is highly expressed in the isolates recovered from invasive infections (Nagamune et al., 2000). Interestingly, immunoglobulins recovered from 'healthy' volunteers plasma can neutralize or at least reduced the toxicity associated with ILY, as well as decreased the activities associated with glycosidase (multisubstrate glycolase A and neuraminidases) (Tomoyasu et al., 2017). They further demonstrated that the expression of those two glycosidase could be used to predict the ILY hemolytic activity of *S .intermedius* isolates (Tomoyasu et al., 2018). Previous work from our laboratory has suggested a variable susceptibility to hemolysis between donors (Giraldi, 2015) associated to the presence of antibody against ILY (unpublished).

Study from our group has failed to demonstrate an association between the production of hydrolytic enzymes by an SAG isolate and its capability to cause an infection, suggesting the infection is not induced by the selection of a hypervirulent strain (Grinwis et al., 2010a). This suggests that host factors could promote colonization or infection behaviour observed in SAG isolates. Indeed, our group has demonstrated that the cytokines profiles induced in peripheral blood mononuclear cells could distinguish invasive from isolates from asymptomatic subjects, where invasive isolates were more immunogenic (Kaiser et al., 2014).

1.6. Central paradigm and thesis aims

Central paradigm

Acute respiratory infections are a leading cause of mortality and morbidity worldwide and are predicted to be still prevalent in the next decades. The burden associated with respiratory infection is mostly due to the high prevalence of pneumonia. Several factors such as the difficulty to identify the etiology of acute respiratory infections and subsequent inappropriate therapeutic plans contribute to the complexity of treating respiratory infections, and to the development of potential complications such as pleural empyema, bacteremia and sepsis. Lower respiratory infections are diverse with respect to clinical features, population targeted, etiology, treatment options, and development of potential complications. The etiology of pneumonia and its complications are frequently not identified.

In this study, I have investigated the microbiomes of the critically ill as part of the PROSPECT study which investigates probiotics as a prophylactic treatment to reduce VAP in the ICU. I also carried out additional quality control on the probiotic and investigated the microbiology of adverse events with possible involvement of the probiotic. I finally performed a comparative genomic study of the *Streptococcus* Anginosus/Milleri group which represent potentially under-reported respiratory pathogens and commonly associated with pneumonia complications such as pleural empyema.

I hypothesize that:

1. The baseline bacterial composition (early in the ICU) of the microbiomes of ventilated patients (lower respiratory tract, gastric and stool) will be correlated with clinical outcome.
2. Comparative genomics of the *Streptococcus* Anginosus/Milleri group will define genetic variability associated with phenotypic heterogeneity and pathogenic properties of individual strains.

Research aims

Based on these hypotheses, my research project has been divided into the following aims:

1. Perform a microbial survey of the bacterial communities associated with mechanically ventilated critically ill patients and assess the impact of microbial disturbances on the patient's clinical outcomes within the PROSPECT study (**Chapter 2**).
2. Evaluate the dose consistency and the microbial identity of placebo and probiotic capsules administered during the PROPSECT study as well as investigating the adverse events occurring during this trial and associated with the administration of probiotics (**Chapter 3**).
3. Investigate the genetic diversity observed within clinical isolates within the *Streptococcus* Anginosus/Milleri group (SAG) by comparative genomics (**Chapters 4 and 5**).

Chapter 2

Investigation of the composition and dynamics of the microbiome of critically ill patients

Preface

Research presented as part of this chapter has been previously published as:

Lamarche D, Johnstone J, Zytaruk N, Clarke F, Hand L, Loukov D, Szamosi JC, Rossi L, Schenck LP, Verschoor C, McDonald E, Meade MO, Marshall JC, Bowdish DME, Karachi T, Heels-Ansdell D, Cook D, Surette MG and For the PROSPECT Investigators and the Canadian Critical Care Trials Group. Microbial dysbiosis and mortality during mechanical ventilation: a prospective observational study. *Respiratory Research* **19**, 245 (2018).

Author Contributions: DL is the first author of this published manuscript. Conception and design: DL, JJ, JCM, MOM, DMB, DJC, MGS. Acquisition of data: DL, NZ, FC, LH, DeL, CPV, EM, DHA, TK. Analysis and interpretation of data: DL, MGS, JCS. Drafting or revisiting the article: DL, MGS, JJ, DMB, DJC, JCS, LR, DJC, LPS. Final approval of the manuscript: All authors.

The only changes made on this manuscript were for formatting purposes and for a coherent continuity within the thesis. Appendix A contains the supplementary material published as part of the manuscript.

This work was nested in a larger randomized-control trial PROSPECT. Since the larger study is still ongoing, we are still blinded to the patients' administration of the study product (probiotics versus placebo). Hence, the majority of the data could not be used before this crucial information would be revealed to us. However, addition information and preliminary work that could not be included in the first manuscript has been prepared as part of Appendix B.

Microbial Dysbiosis and Mortality During Mechanical Ventilation: A Prospective Observational Study

Daphnée Lamarche [1,2,3] (lamarcd@mcmaster.ca), Jennie Johnstone [4,5,6] (jennie.Johnstone@oahpp.ca), Nicole Zytaruk [7,8] (zytaruk@mcmaster.ca), France Clarke [7,8] (clarkef@mcmaster.ca), Lori Hand [8,9] (handlori@hhsc.ca) , Dessi Loukov [2,10] (dessiloukov@gmail.com), Jake C. Szamosi [2,3,11] (szamosjc@mcmaster.ca), Laura Rossi [1,2,3] (rossil@mcmaster.ca), Louis P. Schenck [1,2,3] (schenclp@mcmaster.ca), Chris P. Verschoor [8] (cversch@mcmaster.ca), Ellen McDonald [8,11] (emcdonal@mcmaster.ca), Maureen O. Meade [8,9,11] (meadema@hhsc.ca), John C. Marshall [12,13] (MarshallJ@smh.ca), Dawn M.E. Bowdish [2,10] (bowdish@mcmaster.ca), Tim Karachi [9,11] (karachi@mcmaster.ca), Diane Heels-Ansdell [8] (ansdell@mcmaster.ca), Deborah J. Cook [7,8,11] (debcook@mcmaster.ca), Michael G. Surette [1,2,3,11] (Surette@mcmaster.ca) for the PROSPECT Investigators, Canadian Critical Care Trials Group & Canadian Critical Care Translational Biology Group

Affiliations:
[1]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada
[2]Institute for Infectious Disease Research, McMaster University, Hamilton, ON, Canada
[3]Farncombe Family Digestive Health Research Institute, McMaster University, Hamilton, Ontario, Canada
[4]Department of Medicine, University of Toronto, Toronto, ON, Canada
[5]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[6]Public Health Ontario, Toronto, ON, Canada
[7]St. Joseph's Healthcare, Hamilton, ON, Canada
[8]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada
[9]Hamilton Health Sciences, Hamilton, ON, Canada
[10]Department of Pathology and Molecular Medicine, Hamilton, ON, Canada
[11]Department of Medicine, McMaster University, Hamilton, ON, Canada
[12]Department of Surgery, University of Toronto, Toronto, ON, Canada
[13]Interdepartmental Division of Critical Care, University of Toronto, Toronto, ON, Canada

Corresponding author:
Michael G Surette, PhD
Department of Medicine, Health Sciences Bldg, 3N8F, McMaster University
1280 Main Street West
Hamilton, ON, L8S 4K1
Telephone number: +1 905 525 9140 ext. 21964

Abstract

Background: Host-associated microbial communities have important roles in tissue homeostasis and overall health. Severe perturbations can occur within these microbial communities during critical illness due to underlying diseases and clinical interventions, potentially influencing patient outcomes. We sought to profile the microbial composition of critically ill mechanically ventilated patients, and to determine whether microbial diversity is associated with illness severity and mortality.

Methods: We conducted a prospective, observational study of mechanically ventilated critically ill patients with a high incidence of pneumonia in 2 intensive care units (ICUs) in Hamilton, Canada, nested within a randomized trial for the prevention of healthcare-associated infections. The microbial profiles of specimens from 3 anatomical sites (respiratory, and upper and lower gastrointestinal tracts) were characterized using 16S ribosomal RNA gene sequencing.

Results: We collected 65 specimens from 34 ICU patients enrolled in the trial (29 endotracheal aspirates, 26 gastric aspirates and 10 stool specimens). Specimens were collected at a median time of 3 days (lower respiratory tract and gastric aspirates; interquartile range [IQR] 2-4) and 6 days (stool; IQR 4.25-6.75) following ICU admission. We observed a loss of biogeographical distinction between the lower respiratory tract and gastrointestinal tract microbiota during critical illness. Moreover, microbial diversity in the respiratory tract was inversely correlated with APACHE II score ($r = -0.46$, $p = 0.013$) and was associated with hospital mortality (Median Shannon index: Discharged alive; 1.964 vs. Deceased; 1.348, $p = 0.045$*)*.

Conclusions: The composition of the host-associated microbial communities is severely perturbed during critical illness. Reduced microbial diversity reflects high illness severity and is associated with mortality. Microbial diversity may be a biomarker of prognostic value in mechanically ventilated patients.

Trial registration: ClinicalTrials.gov ID NCT01782755. Registered February 4 2013, https://clinicaltrials.gov/ct2/show/NCT01782755

Abstract Word Count: 282

Keywords: microbiome, critical illness, microbial diversity, respiratory tract microbiota, gastrointestinal tract microbiota

Background

The human body harbours trillions of bacterial cells on and within its surfaces and mucous membranes (Lozupone et al., 2012). These microorganisms (microbiota) are largely commensals and mutualists that can confer health advantages to the host. The microbiome is essential for numerous features of host physiology, including metabolism (by degrading otherwise non-digestible molecules), resistance to infection (e.g., via colonization resistance), and immune maturation and homeostasis (Sommer and Bäckhed, 2013; Levy et al., 2017). Normally, the microbiota of healthy individuals is stable over time, although it is sensitive to changes in lifestyle, diet, and illnesses (Lozupone et al., 2012; David et al., 2014). Perturbations of these microbial ecosystems can be associated with several diseases, including inflammatory bowel disease, and *Clostridium difficile* infection, as well as conditions associated with critical illness (i.e., sepsis, acute respiratory distress syndrome, and multiple organ dysfunction syndrome) (Lozupone et al., 2012; Prescott et al., 2015; Yamada et al., 2015; Dickson et al., 2016a; Kitsios et al., 2017; Zakharkina et al., 2017).

The microbiota of patients in the intensive care unit (ICU) fluctuates considerably due to acute disease states associated with critical illness, and common interventions such as mechanical ventilation, antimicrobials, gastric acid suppression, and enteral nutrition (Dickson, 2016; Kitsios et al., 2017). Studies using culture-dependent and culture-independent methods have demonstrated that microbial diversity in the gastrointestinal (GI) and respiratory tracts of critically ill patients decreases following ICU admission, and that critically ill patients experience pronounced disturbances of their microbial communities which become more severe over time (Berdal et al., 2007; Shimizu et al., 2010; Zaborin et

al., 2014; Kelly et al., 2016; McDonald et al., 2016; Yeh et al., 2016; Lankelma et al., 2017; Zakharkina et al., 2017).

The consequences of microbial dysbiosis on illness severity and mortality have been relatively unexplored, particularly in the lower respiratory tract. A better understanding of microbial disturbances in the ICU setting and their impact on clinical outcomes is needed, given the emerging evidence suggesting that therapeutics targeting the microbiota in critical illness may be promising to prevent or treat complications (Silvestri et al., 2012; Li et al., 2015; Manzanares et al., 2016; Wei et al., 2016).

The objectives of this prospective observational study were to investigate the microbial composition at distinct anatomical sites of the respiratory and GI tracts during critical illness, and to evaluate whether the microbial diversity in the first week of the ICU stay of mechanically ventilated patients is associated with illness severity and mortality.

Materials and Methods

Subject recruitment

We recruited critically ill patients 18 years of age or older receiving invasive mechanical ventilation from a medical-surgical and a neuro-trauma ICU in 2 hospitals in Hamilton, Canada. Samples were collected between October 2013 and June 2014 as a translational study nested within a multicenter pilot randomized blinded trial (PROSPECT, NCT02462590) testing the effect of the probiotic *Lactobacillus rhamnosus* GG versus placebo on the risk of ventilator-associated pneumonia (VAP) and other infections (Cook

et al., 2016). Further information concerning patient recruitment is available in the Appendix A.

Healthy donors included in this study were individuals older than 18 years of age without comorbidities who had not received antibiotics in the 6 months before sample collection. Healthy donors were included to investigate the microbial compositional differences between healthy and critically ill individuals; they also participated in other ongoing studies in our laboratory (Moayyedi et al., 2015; Stearns et al., 2015; Potts, 2017).

This study was approved by the Hamilton Integrated Research Ethic Board. All participants or their substitute decision makers provided written informed consent prior to participation.

Sample collection

Endotracheal tube aspirate (ETA), gastric tube aspirate (GA), fecal samples and peripheral venous blood were aseptically collected and transferred by the research coordinator at each ICU. To limit confounders due to probiotic administration, the first available sample following study enrollment from each patient at each body site was used to compare differences in the composition of the microbiota between anatomical sites. All samples included were collected before the 7^{th} day in the ICU. For the healthy cohort, nasopharyngeal and oropharyngeal swabs, fecal samples and bronchoalveolar lavages (BAL) were collected as described elsewhere (Baatjes et al., 2015; Moayyedi et al., 2015; Stearns et al., 2015).

DNA extraction, 16S rRNA gene sequencing, and sequence processing

The genomic DNA extraction and amplification for sequencing on the Illumina MiSeq platform by the McMaster Genomics Facility (Hamilton, Canada) was performed as described previously (Whelan et al., 2014). Paired-end sequences of the v3 region of the 16S rRNA gene were processed through a standardized workflow (Whelan and Surette, 2017). A negative genomic extraction and sequencing controls were conducted to ensure that sequencing contamination was not an issue for low-biomass samples. Our sequencing data and metadata is available at NCBI SRA BioProject PRJNA428805 (SRA accession: SRP128586). Supplementary information is available in the Appendix A.

Data analysis

The α and β diversity estimates were generated in R (R Core Team, 2013) using the 'phyloseq' package (McMurdie and Holmes, 2013). β diversity was calculated on the proportionally normalized operational taxonomic unit (OTU; i.e., a proxy for bacterial 'species') table excluding singletons, and OTUs classified as non-bacterial. The α diversity was calculated using the OTU table excluding only non-bacterial OTUs and rarefied to 2800 reads per sample 100 times, and the mean value of the α diversity measurements was used. Three α diversity metrics were included in this investigation; the Shannon and Simpson diversity indices account for richness (i.e., number of taxa) and evenness (i.e., taxa relative abundances), while the metric Observed Species accounts only for richness (Morgan and Huttenhower, 2012). The within-body site distance-to-

centroid was calculated on a Bray-Curtis distance matrix using the function betadisper from the R package 'Vegan' (Oksanen et al., 2019). The UPGMA consensus tree was generated in QIIME with support established using jackknife (Caporaso et al., 2010). The correlation analysis between patients' metadata and α diversity measurements was performed using the R packages 'Hmisc' and 'corrplot', using the Spearman's rank correlation coefficient (Harrell Jr, 2013; Wei and Simko, 2017). The Kaplan-Meier estimate was generated using the R packages 'survival' and 'survminer' (Therneau and Grambsch, 2000; Kassambara et al., 2017). Survival analysis using Simpson diversity was not included since the same sets of patients were generated in both groups resulting in identical results than using Shannon diversity. The impact of concomitant antimicrobial exposure with sample collection on ETA α diversity was investigated. Additional investigation on the influence of individual antimicrobials was unsuitable given our limited number of samples.

Statistical analysis

Differences across body sites in community composition (biogeography) was investigated using a permutational multivariate analysis of variance (PERMANOVA), performed with the 'Vegan' package in R (Oksanen et al., 2019) on a Bray-Curtis distance matrix, using 100 000 permutations. Differences in within-site inter-sample Bray-Curtis and centroid distances were tested with the 'lmerTest' package in R (Bates and Mächler, 2015), using a linear mixed model with fixed effects of cohort (i.e., Healthy and ICU) and anatomical site, and a random effect of patient on intercept. Compositional differences between the

ICU patients and healthy donors were tested in QIIME using a Kruskal-Wallis test. *P* values were adjusted using Benjamini-Hochberg false discovery rate (FDR) correction. The α diversity, bacterial load, serum cytokines and clinical markers results are presented using the median and interquartile range and were analyzed with a Mann-Whitney test using GraphPad Prism version 6.0 (La Jolla, CA, USA). Correlation analysis was followed with FDR correction to account for multiple-testing (R Core Team, 2013). For the survival analysis, differences in mortality between groups were tested using a log-rank test (R Core Team, 2013). The two population proportions (deceased vs. discharged alive within the low and high diversity groups) were compared using a Fisher's exact test using GraphPad Prism. The significance threshold was set at $p < 0.05$.

Sensitivity analysis

To evaluate whether study product exposure influenced the outcome, *L. rhamnosus* GG sequences (OTU5) were removed from the sensitivity analysis. Results of these analyses were then compared to ensure that the impact of the ingested probiotic administration was minimal and that the study results were not influenced by the presence of *L. rhamnosus* GG. More information is available in the Appendix A.

Results

Demographics of the critically ill cohorts and healthy controls

In total, 34 mechanically ventilated critically ill patients were enrolled; demographic data are presented in Table 2.1. Patients were a mean (standard deviation [SD]) of 66.6 (10.9)

years of age, had an Acute Physiology and Chronic Health Evaluation (APACHE) II score of 25.5 (8.5), and 14/34 (41.2%) were female (Table 2.1). Patients were admitted to the ICU for medical conditions (n=31, 91.2%), surgical conditions (n=1, 2.9%), or trauma (n=2, 5.9%). The median ICU length of stay was 11.5 days (IQR 7-20.5) and 22 (64.7%) patients were alive at hospital discharge.

In total, 65 samples were collected including 29 ETAs, 26 GAs and 10 stool specimens (Appendix A: Table A.1). Data collection was a median of 3 days following ICU admission for ETA and GA samples (IQR 2-4), and 6 days for stool samples (IQR 4.3-6.8).

Samples from 35 healthy adults were included to establish reference points for the biogeography of the healthy microbiome for comparison to the ICU cohort. One specimen type was collected per healthy donor with the exception of the nasopharynx (NP) and oropharynx (OP) swabs, which were both collected from the same individual. In total, 42 samples from 35 healthy adults consisting of 7 BALs, 7 NP swabs and 7 OP swabs, and 21 stool specimens were included (Appendix A: Table A2).

Table 2.1: Demographics and characteristics of ICU patients.

Characteristics	**Patients (n=34)**
Age (years), mean (SD)	66.6 (10.9)
APACHE II, mean (SD)	25.5 (8.5)
Female, n (%)	14 (41.2)
Type of patient, n (%)	
Medical	31 (91.2)
Surgical	1 (2.9)
Trauma	2 (5.9)
Admitting Diagnosis, n (%)	
Pneumonia	13 (38.2)
Sepsis	8 (23.5)
Chronic obstructive pulmonary disease exacerbation	2 (5.9)
Congestive heart failure	2 (5.9)
Respiratory arrest	2 (5.9)
Trauma	2 (5.9)
Alcohol withdrawal	1 (2.9)
Cardiac arrest	1 (2.9)
Cardiogenic shock	1 (2.9)
Laminectomy	1 (2.9)
Renal failure	1 (2.9)
ICU length of stay (days), median (quartile 1, quartile 3)	11.5 (7, 20.5)
Hospital length of stay (days), median (quartile 1, quartile 3)	32 (15, 55)
ICU Mortality, n (%)	5 (14.7)
Hospital Mortality, n (%)	12 (35.3)

Loss of biogeography within the ICU cohort compared to healthy controls

We compared the microbial community structures of patients admitted to the ICU to a healthy cohort to investigate the homogeneity of microbial composition within each body site and to observe any change in biogeography during critical illness (Figure 2.1). The β diversity was measured using the Bray-Curtis dissimilarity metric. Microbial communities segregate substantially and significantly by body site in healthy individuals

(PERMANOVA, $p < 0.001$, $R^2 = 0.529$; Figure 2.1A). However, differences in composition by anatomical site were considerably less pronounced in the ICU cohort than healthy controls (PERMANOVA, $p < 0.001$, $R^2 = 0.082$; Figure 2.1A). Indeed, there is minimal visible clustering in ICU specimen types (Figure 2.1B), and the distance-to-centroid within-body site is significantly higher in the ICU cohort than the healthy controls ($p < 0.001$, Figure 2.1B). Comparison of the pairwise Bray-Curtis dissimilarity among individual body site specimens from the ICU and the healthy cohort demonstrates significantly higher heterogeneity within the ICU cohort for samples collected at those body sites ($p < 0.0001$; Appendix A: Figure A.1). These results show that the distinct biogeographical composition of the microbiome at different body sites is attenuated in ICU patients.

Additionally, the hierarchical clustering of samples shows an overlap between GA and ETA and a subset of stool samples, further indicating a lack of compositional definition between anatomical sites in the ICU samples (Figure 2.1C). Only the microbiota from a subset of ICU stool specimens cluster distinctly from the other samples sites (Figure 2.1A,C). The heterogeneity in the microbial composition of host-associated communities within and between anatomical sites and patients is shown in the taxonomic summaries (Figure 2.2). In summary, these results suggest a lack of separation between distinct anatomical sites, and high heterogeneity within body sites. Comparison with samples from a healthy cohort emphasizes the microbial dysbiosis occurring during critical illness.

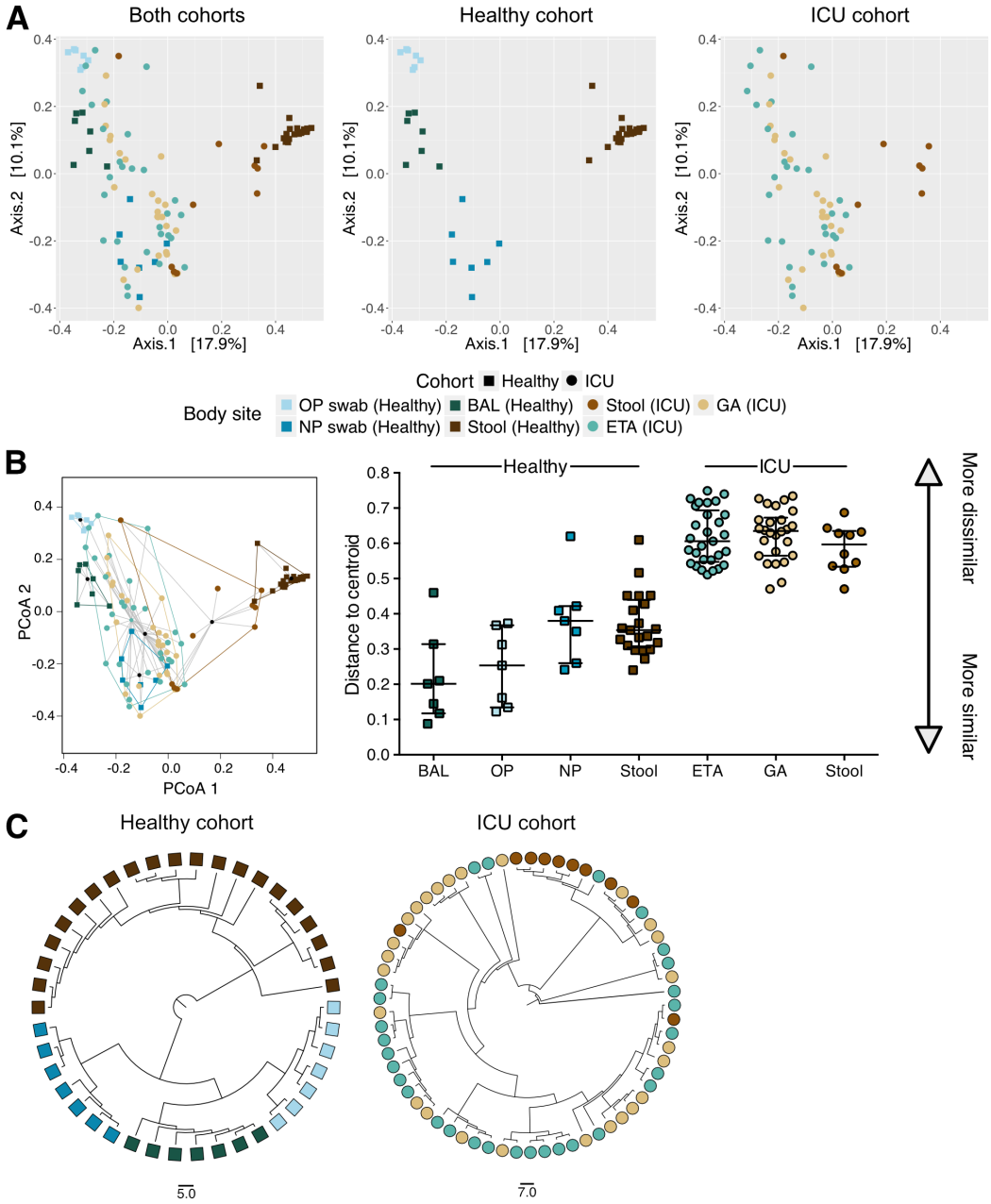


Figure 2.1: Lack of microbial consensus and loss of biogeographical distinction in ICU patients. Principal coordinate analysis (PCoA) ordination using the Bray-Curtis dissimilarity metric between the ICU and healthy cohorts demonstrate that samples collected from the healthy cohort tend to cluster per body sites (PERMANOVA, $p < 0.001$, $R^2 = 0.529$) whereas the samples from different anatomical sites tend to overlap in the ICU cohort (PERMANOVA, $p < 0.001$, $R^2 = 0.082$; A). The ordination plot of group dispersions within body site demonstrates a lack of compositional homogeneity within anatomical sites in the ICU cohort ($p < 0.001$; B). The overlaying lines on the scatter plot show the median distances between the cluster's centroid displayed with a black circle

and each sample within the group and the interquartile range of each site. UPGMA dendogram showing the Bray-Curtis dissimilarity between specimens displays a perfect segregation of samples in the healthy cohort based on collections sites. This is not observed in the ICU cohort (C).

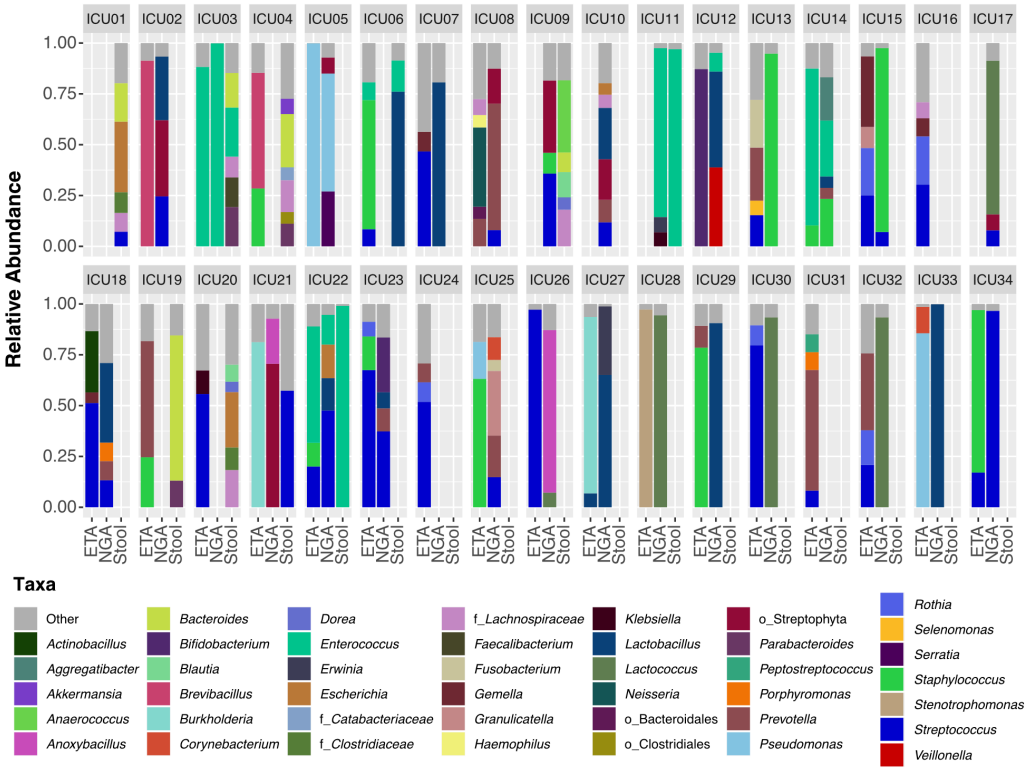


Figure 2.2: Compositional heterogeneity observed within and between anatomical sites in critically ill patients. Taxonomic summaries of the 65 samples included in this study displayed by patients and anatomical sites. Bacterial groups present at less than 5% relative abundance are grouped in the "other" category displayed in gray. Taxonomic groups are labeled according to the highest level resolved if not at the Genus (Order; o_, Family; f_).

Relative abundance of specific bacterial taxa are decreased in the lower respiratory and GI tracts in the ICU cohort when compared to healthy controls

The compositional disparities between ICU and healthy specimens were investigated. We included 29 ETA samples from the ICU cohort, 7 BAL specimens from the healthy

cohort, as well as 10 ICU patient stool samples and 21 healthy subject stool specimens. Analyses showed that the relative abundances of 34 OTUs for respiratory and 29 OTUs for stool samples were significantly different between critically ill vs. healthy individuals following multiple-test correction (Appendix A: Table A3-A4). In respiratory specimens, OTUs with the greatest change in abundance were from the genera *Neisseria, Veillonella, Streptococcus, Staphylococcus,* and *Corynebacterium;* these were significantly decreased in the ICU cohort compared to the healthy cohort (Appendix A: Table A3). OTUs from the Lachnospiraceae family, from the genera *Faecalibacterium, Blautia, Subdoligranulum*, and *Lachnobacterium* were depleted in stool specimens from ICU patients (Appendix A: Table A4). Several of these microorganisms are members of the host-associated communities in healthy individuals (Charlson et al., 2011; Lozupone et al., 2012; The Human Microbiome Project Consortium, 2012; Dickson et al., 2017). Only a few significant OTUs were increased in the ICU patients, indicating that the pathogen expansion tends to be patient-specific. However, the loss of specific bacteria from body sites was more generalized among ICU patients.

Respiratory tract microbial diversity is associated with illness severity in the ICU cohort

To determine whether microbial diversity was associated with clinical parameters, a correlation analysis was performed. Since microbial diversity and environment vary across body sites, correlation was tested separately for each anatomical site. We examined the correlation between α diversity and APACHE II score used to assess illness severity in the ICU (Knaus et al., 1985). Both Shannon and Simpson diversity indices in ETA

specimens were inversely correlated with APACHE II score (r = -0.46, p = 0.013; Figure 2.3A and r = -0.44, p = 0.017; Figure 2.3B, respectively). The association between APACHE II score and Observed Species was not statistically significant, but the magnitude of the correlation was similar (r = -0.31, p = 0.11; Figure 2.3C). No significant correlation between any α diversity metric and APACHE II score was observed with GA samples (Appendix A: Figure A.2).

We also investigated the association between microbial diversity, patient demographic information, and clinical parameters in the correlation analysis (Figure 2.3D). From the 45 correlation analyses included in the correlation matrix of the 29 ETA samples, 6 combinations were significant following multiple-test correction (Appendix A: Table A5).

In summary, our results suggest a possible association between illness severity and ETA microbial diversity in mechanically ventilated patients.

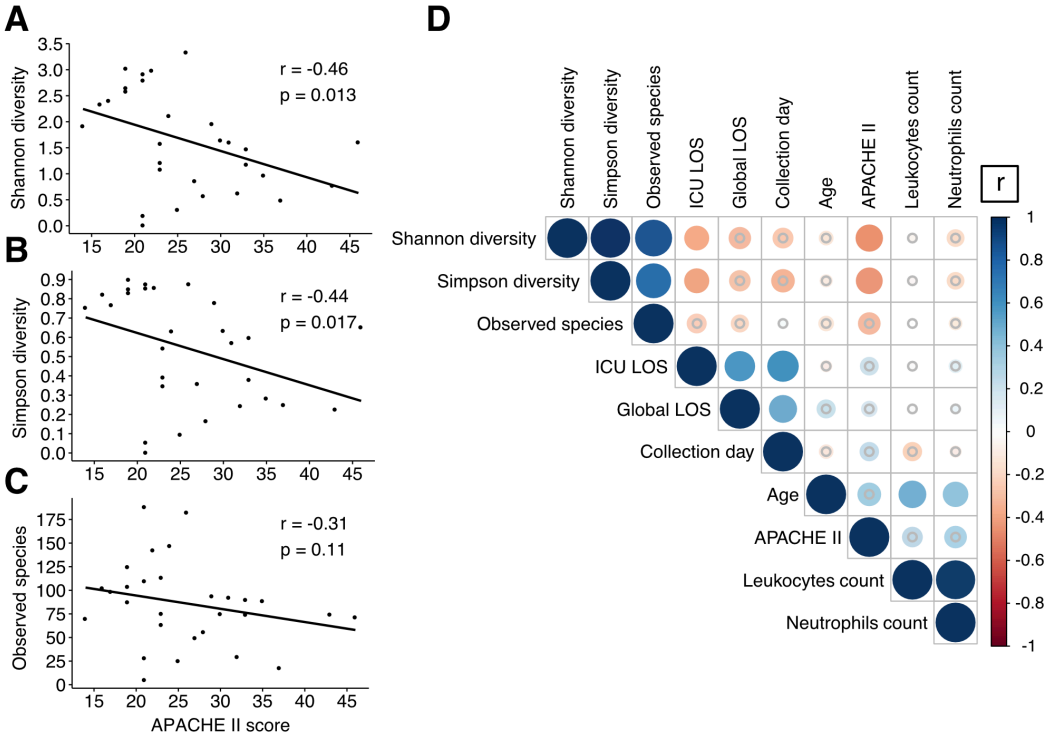


Figure 2.3: Lower respiratory tract microbial diversity is associated with illness severity in critically ill patients. Correlation analysis using Spearman's rank correlation coefficient demonstrated an inverse association between APACHE II score and Shannon diversity (r = -0.46, $p = 0.013$; A), Simpson diversity (r = -0.44, $p = 0.017$; B) and Observed species (r = -0.31, $p = 0.11$). Correlation matrix of clinical parameters and microbial diversity markers showed only a limited number of significant associations (D). The sizes of the circles are dependent on the correlation coefficient value (r). Comparisons that did not achieve significance are represented with a gray circle. LOS represents the length of stay.

Reduced respiratory tract microbial diversity is associated with mortality in critical illness

Samples were stratified based on hospital mortality. Taxonomic summaries are included in the Appendix A (Figure A.3). Within-site compositional differences between deceased and discharged-alive patients were assessed, and no OTUs were found to be significantly different between the groups studied (Appendix A: Table A.6). However, the Shannon

and Simpson diversity indices of the ETAs were significantly lower in patients who died versus those who survived their hospital stay ($p = 0.045$ and $p = 0.0185$; Figure 2.4A-B). These results could not be explained by an increased prevalence of infection in the deceased group since both groups have a similar rate of pneumonia when specimens were collected (9/10 vs. 16/19, $p = 1.00$). Additionally, quantification of 16S rRNA gene in ETA samples demonstrated that bacterial loads were not different between both groups ($p = 0.804$; Appendix A: Figure A.4). No significant difference in ETA α diversity was detected between admission diagnoses (Appendix A: Figure A.5). Furthermore, we failed to detect a correlation between concomitant antimicrobial exposure and Shannon ($r = -0.1$, $p = 0.63$) and Simpson diversity indices ($r = -0.14$, $p = 0.5$) of the ETA microbiota, nor when patients were stratified based on hospital mortality (Appendix A: Figure A.6). Moreover, specimen's collection day was not associated with Shannon (Deceased: $r = 0.099$, $p = 0.79$; Discharged alive: $r = -0.15$, $p = 0.53$) and Simpson diversity indices (Deceased: $r = -0.21$, $p = 0.56$; Discharged alive: $r = -0.16$, $p = 0.51$). GA samples did not exhibit microbial diversity differences between deceased and discharged alive patients (Figure 2.4A-B), or in either anatomical site using observed species (Appendix A: Figure A.7).

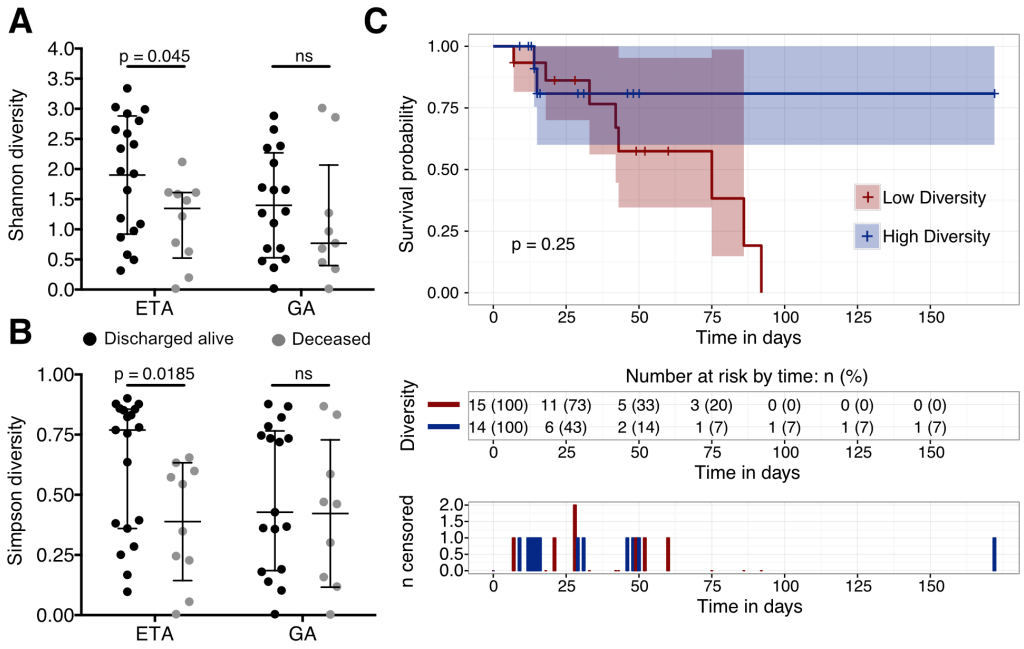


Figure 2.4: Association between microbial diversity and hospital mortality within ICU samples. Shannon (A) and Simpson diversity (B) of ETA and GA specimens shaded by hospital mortality demonstrates a significant reduction in the ETA Shannon diversity in the patients deceased in the hospital versus patients discharged alive. Kaplan-Meier survival curves displayed by high and low microbial diversity groups (C). The censored (i.e., discharged alive) patients are indicated by ticks marks. The threshold for grouping by diversity was the median value of the Shannon diversity measurements for the 29 samples included in this analysis. Confidence intervals are represented by the blue and red shaded areas. Numbers of patients included in the analysis and censored are shown per group under the Kaplan-Meier curve

We compared inflammatory markers and APACHE II score between patients with ETA who were deceased or discharged alive. No significant differences were detected in inflammatory markers (e.g. IL-6, C-reactive protein, leukocyte and neutrophil counts) or in illness severity (Appendix A: Figure A.8).

To further investigate the impact of ETA α diversity on hospital mortality, proportions and time-to-event analyses were performed. Kaplan-Meier curves were generated to compare survival in patients with low versus high Shannon diversity. Although it appeared that patients with lower respiratory tract bacterial diversity (Shannon index < 1.61) were more likely to die in hospital over time than those with more diverse microbiota, this was not statistically significant ($p = 0.25$; Figure 2.4C). From the 29 patients included in the survival analysis, 10 (34.5%) died in hospital; 2/14 (14.3%) in the high microbial diversity group died before discharge versus 8/15 (53.3%) in the low microbial diversity group ($p = 0.05$).

Sensitivity analysis

From the 65 samples included in the study, 31 (from 23 patients) were collected before the administration of *L. rhamnosus* GG or placebo and 34 samples (from 22 patients) were collected following randomization. Our results were not affected by the removal of OTU5, assigned to *Lactobacillus*, that would include the administered probiotic *L. rhamnosus* GG or closely related indigenous strains of *Lactobacillus* from the analysis. Supplementary information concerning the results of the sensitivity analysis is available in the Appendix A.

Discussion

In this prospective observational study, we profiled the composition of microbial communities in a cohort of mechanically ventilated critically ill patients with a high incidence of pneumonia within their first week in the ICU, and how these communities relate to illness severity and clinical outcomes. Our results demonstrate that, in contrast to findings in healthy individuals (The Human Microbiome Project Consortium, 2012), the bacterial community structures are considerably less defined by body site in the upper GI tract and lower respiratory tract. Conditions associated with admission to the ICU and interventions in this setting could compromise normal host barriers and lead to compositional overlap between the airway and stomach, reflecting a loss of microbial separation across anatomical sites in critical illness. This work builds upon early studies of the microbiome in ICU patients (Dickson et al., 2016a; McDonald et al., 2016; Yeh et al., 2016, 2016), by including a control group with lower respiratory tract specimens to compare our findings with those of healthy individuals, and by including additional gastric samples in the ICU cohort to expand the multi-anatomical site analysis. While the proximal gastrointestinal tract is established as an important reservoir of ICU-associated pathogens (Marshall et al., 1988), the role of the stomach as a source of tracheal colonization is controversial, and is not always considered to be a substantial contributor to the pathogenesis of VAP (Chastre and Fagon, 2002). The prominent gastric colonization in ICU patients may be primarily due to use of prophylactic acid suppression to prevent stress-related gastric bleeding (Ali and Harty, 2009). The overlap in microbial compositions of the respiratory tract and stomach has also been demonstrated in non-

critically ill pediatric patients receiving proton-pump inhibitors (Rosen et al., 2015), suggesting that these agents may contribute to the observed loss of biogeographical distinction.

The limited compositional similarities between healthy and ICU cohorts could be explained by the loss of commensal microorganisms during critical illness. We have demonstrated that 'normal' host-associated taxa, including the Lachnospiraceae family, *Faecalibacterium*, and *Blautia* genera in the GI tract and *Veillonella, Prevotella*, and *Neisseria* genera in the lower respiratory tract (Charlson et al., 2011; Lozupone et al., 2012; Dickson et al., 2017) were decreased in relative abundance in the ICU cohort. The decrease of these commensal taxa in the GI tract is consistent with other work (McDonald et al., 2016; Yeh et al., 2016). Several of the depleted OTUs are known to confer host advantages, such as anti-inflammatory and nutritional benefits via the production of short-chain fatty acid (SCFA) (Sokol et al., 2008; Arpaia et al., 2013; Levy et al., 2017). The level of fecal SCFA is drastically decreased upon admission to the ICU (Hayakawa et al., 2011; Yamada et al., 2015). Moreover, perturbation of the indigenous microbiota could lead to harmful repercussions and allow colonization by opportunistic secondary potential pathogens such as *C. difficile*, *Candida albicans* or facultative anaerobic gammaproteobacteria such as *Pseudomonas* (Sommer and Bäckhed, 2013; Scales et al., 2016). Interestingly, the OTUs that were increased in the ICU cohort in comparison to the healthy cohort were common pathogens associated with ICU-acquired infections from the *Enterococcus, Pseudomonas,* and *Staphylococcus* genera (Richards et al., 2000). The limited number of OTUs that were significantly increased in the ICU

cohort versus healthy cohort reflects the heterogeneity of the critically ill population due to various comorbidities and clinical interventions. We also demonstrate a lack of a shared microbial community structure in the ICU cohort toward a microbiota dominated by only few taxa. These results suggest that the emergence of pathogens is patient-specific, while the decrease in relative abundance of commensal taxa is observed more uniformly within the ICU population.

In this study, we demonstrated an inverse association between ETA microbial diversity and both illness severity and hospital mortality. We observed a significant decrease in respiratory microbial diversity in patients who died versus survived their hospital stay. Several studies have established that microbial composition tends to collapse in ICU patients toward the dominance of only a few taxa (Zaborin et al., 2014; Yeh et al., 2016). It has been suggested that the microbial collapse is driven by aggressive antimicrobial administration during critical illness. A large prospective study including 14 414 critically ill patients has demonstrated that 71% of the patients were receiving antibiotics (Vincent, 2009). Nevertheless, we did not see a correlation between concomitant antimicrobial exposure and ETAs microbial diversity in our cohort. This could be due to the fact that the samples included in this study were collected early during critical illness and an effect could be observed with later time points. Zakharkina *et al.* recently demonstrated no association between respiratory tract microbial diversity and antimicrobials in mechanically ventilated patients (Zakharkina et al., 2017). Moreover, it has been demonstrated that lower respiratory tract samples from mechanically ventilated patients with pneumonia tend to have lower α diversity in comparison with patient

without suspected pneumonia (Kelly et al., 2016). The loss of diversity could be due to the predominance of potential pathogens and could explain why this decrease in diversity is observed only with metrics accounting for taxa relative abundances (e.g., Shannon and Simpson diversity). However, survival groups exhibited no detectable difference of bacterial biomass nor rates of pneumonia. These results highlight the potential utilization of α diversity metrics as an index of severity of illness and could, in addition to other clinical markers, improve patient's stratification based on their survival prognosis. To the best of our knowledge, respiratory α diversity has not been identified as a potential predictor of survival outcome in ICU patients. Microbial diversity of the GI tract has been associated with outcome in patients undergoing allogeneic hematopoietic stem cell transplantation, severe inflammatory response syndrome, and in high-risk patients admitted to the ICU (Iapichino et al., 2008; Shimizu et al., 2010; Taur et al., 2014). By contrast, in one study of 34 ICU patients, no association between survival and microbial diversity of the GI tract was observed (Lankelma et al., 2017).

Our analysis was performed on samples collected during the first week of ICU stay (median collection time was 3 days following ICU admission for ETA and GA and 6 days for stool samples). This suggests that early time points could potentially be used to study the association between dysbiosis and an outcome occurring further downstream during the hospital stay (median 28.5 days IQR 14-50). Although there was no significant difference in survival curves between low and high ETA microbial diversity groups, of 29 patients included in the survival analysis, 35% died in hospital whereas 80% of the deceased patients were in the low diversity group. This suggests how ETA microbial

diversity may be a complementary prognostic variable. As an indicator of illness severity akin to organ dysfunction, loss of microbial diversity can be conceptualized as a marker of poor outcome that potentially could be modifiable.

The limited number of samples and patients influences the power of these analyses. Larger studies will be required to increase our confidence in the association between microbial diversity and mortality. The comparison of critically ill patients and healthy individuals was constrained by lack of BAL specimens from the ICU cohort; therefore, more easily obtained ETA samples were used. Moreover, in clinical practice, BAL would not typically be performed early enough or commonly enough in patients without classical immunocompromised states for inclusion in our study. Although the ICU patients received several clinical interventions that could potentially affect the microbial structure of host-associated communities (e.g., antimicrobials, acid suppressants), these are intrinsic to critical care management. Adjustment for confounders such as sex, age, specific medications, and comorbidities was not suitable with our limited sample size, although no significant difference in α diversity measurement was detected between the admission diagnoses. Moreover, despite the great heterogeneity observed in ICU patients, a proinflammatory state is a unifying feature of critical illness, regardless of the reason for admission (i.e., non-infectious conditions such as trauma and pancreatitis, as well as various infectious problems). Changes in colonization resistance and on the host's biologic processes mediated by the microbiota may be similar across critically ill subgroups. Our study was confounded by the collection of samples over the

first 7 days in the ICU. However, there was no correlation of ETAs α diversity with time of sample collection and outcome (deceased or discharged alive).

Strengths of this study include the prospective data collection, protocolized specimen procurement and complete follow-up. We compared specimens from critically ill patients in 2 centers with a control group of healthy individuals. Our study was nested in a randomized trial testing the probiotic *L. rhamnosus* GG versus placebo; to avoid confounding results due to probiotic administration, only samples collected in the first week were included. However, we also performed a sensitivity analysis in which we removed OTU sequences assigned to *L. rhamnosus* GG and our results were not different, increasing confidence in the findings.

Conclusions

In this study, we demonstrated that the composition of the host-associated microbial communities is severely perturbed early in mechanically ventilated critically ill patients. We found that lower respiratory tract microbial diversity is associated with illness severity and may be associated with risk of death. Together, our results suggest that critically ill patients have a potential microbiota signature associated with illness severity and hospital mortality even early in the ICU stay. The prognostic role of these microbial signatures is a promising focus for future research.

Declarations

Acknowledgements

We would like to acknowledge all members of the Surette Laboratory at McMaster University for scientific discussions and especially thank Fiona J. Whelan, Nick Daneman and Wendy Sligl for critical reading of this manuscript. We appreciate the family support for this research, and the assistance of the ICU bedside nurses and research coordinators with sample procurement. This work was conducted in affiliation with the Canadian Critical Care Trials Group and the Canadian Critical Care Translational Biology Group

Funding

The PROSPECT trial was funded by the Canadian Institutes for Health Research (CIHR), the Technology Evaluation for the Elderly, Physicians' Services Incorporated, and the Hamilton Academy of Health Sciences Research Organization. DL is supported by a Doctoral formation award from the Fonds de Recherche du Québec - Santé (FRSQ). DMB, DJC and MGS are supported by the CIHR as Canada Research Chairs. DMB is a Canada Research Chair in Aging and Immunity. DJC is a Canada Research Chair in Knowledge Translation in Critical Care. MGS is a Canada Research Chair in Interdisciplinary Microbiome Research.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

This study was approved by the Hamilton Integrated Research Ethic Board and was performed in accordance with the principles of Good Clinical Practice following the tri-council guidelines. All participants or their substitute decision makers provided written informed consent prior to enrollment.

Consent for publication

Written contentment has been obtained for all participants or their substitute decision makers.

Availability of data and materials

Sequencing data and metadata included in this study is available at NCBI SRA BioProject PRJNA428805 (SRA accession: SRP128586, https://www.ncbi.nlm.nih.gov/sra/SRP128586).

Chapter 3

Probiotic quality control assessment and microbiology of adverse events in the PROSPECT Trial

Preface

Research presented as part of this chapter is being prepared for two publications.

Lamarche D, Johnstone J, Zytaruk N, Rossi L, Szamosi J, Cook D, Surette MG and For the PROSPECT Investigators and the Canadian Critical Care Trials Group. Quality control of probiotic and placebo capsules for the PROSPECT multicentre randomized clinical trial.

Lamarche D, Johnstone J, Zytaruk N, Marshall JC, Lauzier F, Meade M, Cook D, Surette MG and For the PROSPECT Investigators and the Canadian Critical Care Trials Group. Strain level profiling of *Lactobacillus* isolates from adverse infection events in the PROSPECT multicentre randomized clinical trial.

Contributions:

DL, LR conducted the experiments. DL and JS performed the data analysis. MGS, DC, DL, JJ designed the study. NZ oversaw the shipment logistics as well as retrieved and provided clinical information. DL and MGS wrote the manuscript. All authors approved of the final version of the manuscript.

Abstract

Background: The consumption of probiotics to promote health, treat and/or prevent illnesses has increased considerably in recent decades. Their efficacy has been demonstrated in several situations; however, the use of probiotics in vulnerable populations is debated partly due to lack of high-quality evidence (*e.g.* limited numbers of individuals recruited, different doses and identity of the study product) but also a lack of investigation into their safety. Probiotic regimens are generally deemed innocuous in healthy populations; however, concerns have been raised in immunocompromised individuals, patients with an altered (*i.e.* 'leaky') gastrointestinal barrier or critically ill patients. In order to shed light on the efficacy and the safety of the administration of probiotics in these populations, randomized control trials reporting potential adverse events as well as investigating the consistency of the study product administered (placebo and probiotics) are needed.

Methods: As part of a clinical trial investigating the administration of probiotics in mechanically ventilated critically ill patients to prevent healthcare-associated infections (PROSPECT, NCT02462590), we conducted a quality assessment to ensure dose consistency and stability of the administered probiotics (*Lactobacillus rhamnosus* GG, Culturelle™) and to investigate the sterility of the placebo capsules. In addition, microbiological analyses (strain-typing and comparative genomic analysis) were performed on clinical isolates recovered from adverse events during the trial.

Results: 645 capsules were included in the quality control study: 310 probiotic capsules from 6 lots and 335 placebo capsules from 4 lots. All capsules had viable probiotic above

the manufacture specifications (10^{10} CFUs/capsule) and remained stable at 30°C for a 2-week period. Almost 70% of the placebo capsules tested were below the detection limit for viable bacteria (10^2 CFUs/capsule) and the remainder had low level of contaminants (median: 725 [IQR 4800] CFUs/capsule) mainly by other probiotic microorganisms.
To monitor the safety of patients during the trial, all adverse events during the length of the study were reported. 26 *Lactobacillus* isolates were recovered from clinical specimens from the 2653 patients enrolled and 16 were further characterized by strain-typing, from which 13 isolates were identified as the probiotic strain and were considered as an adverse event. Of these, 10/13 patients did not have a *Lactobacillus* isolates recovered from subsequent patients' culture and the initial isolate were considered as potential bedside contaminants of the initial culture.

Conclusions: Taken together, these results demonstrated that the patients included in the PROSPECT clinical trial received a comparable dose of probiotics and that the presence of bacteria in the placebo was negligible compared to the probiotics dose administered during the randomized controlled trial. This study also highlighted the risk of developing probiotic-associated infections in the ICU, even though bedside contamination was a frequent suspected cause, rather than infection. Finally, independent quality control of the probiotic and strain level analysis of potential probiotic strains in adverse events should become the standard for probiotic studies performed in vulnerable patients in order to properly investigate the efficacy as well as the safety of probiotic administration.

Word count: 486 words

Keywords: Probiotics, critically ill, quality assessment, adverse effects, safety, *Lactobacillus*

Introduction

Probiotics are defined as live microorganisms, which when ingested in adequate amounts, confer a health benefit on the host (Hill et al., 2014). According to ClinicalTrials.gov, a common database for clinical trials, 1,486 studies involving the term 'probiotics' have been registered (accessed June 5th, 2020), emphasizing the popularity of probiotic utilization in public health. Several meta-analyses have shown benefits of probiotics for the prevention of antibiotic-associated diarrhea (Hempel et al., 2012; Videlock and Cremonini, 2012), and necrotizing enterocolitis in preterm infants (Deshpande et al., 2010; Thomas et al., 2017). Probiotic administration has been investigated in the critically ill population where the host-associated bacterial communities are highly disturbed due to underlying diseases and diverse clinical interventions (Zaborin et al., 2014; McDonald et al., 2016; Lamarche et al., 2018). In this heterogeneous population, a decrease in the incidence of ventilator-associated pneumonia (Petrof et al., 2012; Bo et al., 2014; Weng et al., 2017) and overall infections (Manzanares et al., 2016) has been observed following probiotic administration; however, most probiotic clinical trials were deemed low quality partly due to limited participants or various methodologies (*e.g.* different strains, dose, duration, or administration route) used during the trial.

Probiotics are considered a natural health product in Canada and the United States, and there is a limited number of quality control studies on dosage and purity using these products. Different standardization procedures for dietary supplements raise questions about dose consistency and identification of the probiotic in the product as efficacy may be due to strain-specific effects. Furthermore, once probiotics leave the

manufacturer, no further analysis is typically performed to examine the integrity of the product. Consequently, several publications have demonstrated a discrepancy with the taxonomy and dose between the advertised product and the ones recovered, emphasizing the importance of quality assessment studies (Fasoli et al., 2003; Marcobal et al., 2008; Angelakis et al., 2011; Goldstein et al., 2015; Lewis et al., 2016). During a concealed-allocated probiotics versus placebo randomized control trial (RCT), the quality assessment of the administered study product and the placebo capsules are not usually monitored and therefore there is little information about the stability of the product during storage and shipping, and the dose consistency.

With the increased use of probiotics for prophylactic use and the treatment of diverse illnesses, health concerns have been raised about the potential risks of developing probiotic-associated infections. Probiotics are deemed harmless in the general 'healthy' population (Salminen et al., 2002); however, evidence is more conflicted for vulnerable patients such as immunocompromised or critically ill patients (Besselink et al., 2008; Hempel et al., 2011; Didari et al., 2014; Yelin et al., 2019). Despite the fact that proven probiotics-related infections are uncommon, reports of *Lactobacillus*-associated infections do exist (Doron and Snydman, 2015), highlighting the potential risks for vulnerable patients such as the critically ill population. In this context, reporting safety outcomes and differentiating infections caused by indigenous strains of *Lactobacillus* versus the administered probiotics is now critical. Unfortunately, strain genotyping rarely occurs for adverse events in probiotic studies. Indeed, the species in the *Lactobacillus* genus, especially the one within the *Lactobacillus casei* group, are not differentiated

easily and its identification requires generally more than one strategy (*e.g.* MALDI-TOF, 16S rRNA, phenotypical tests) (Goldstein et al., 2015; Huang et al., 2018).

The PROSPECT trial (NCT02462590; Cook et al., 2016), investigates the impact of the administration of *Lactobacillus rhamnosus* GG (Culturelle™) or a placebo on the risk of developing ventilator-associated pneumonia and other nosocomial infections in critically ill patients. This trial recruited 2653 patients within 42 centers in Canada and 3 international sites and was performed over a 6-year period. As such, it included multiple manufacturing lots of the study product and placebo. During this trial, there were 26 suspected adverse events involving positive culture potentially associated with the probiotics and identify as *Lactobacillus* or *Lactobacillus rhamnosus*. However, the identification was not performed to the strain level by the respective clinical microbiology laboratories on site.

In this quality assessment study nested in the PROSEPCT trial, additional safety monitoring, quality control, and microbiological analyses were carried out. First, the identity and the dose of the study product (probiotics and placebo) administered for each lot at each study site were determined. In addition, the stability of the probiotics capsules at various temperatures to mimic the potential temperature fluctuations experienced during storage and shipping was examined. Finally, for adverse events associated with positive *Lactobacillus* culture, strain genotyping for the isolates was carried out to determine if the isolate was the study product (*Lactobacillus rhamnosus* GG). Taken together, the observations from this quality assessment study demonstrates that the

inclusion of this additional analysis should be the standard for trials investigating the administration of probiotics in compromised populations such as the critically ill patients.

Materials and Methods

Probiotic procurement

The PROSPECT trial (NCT02462590; Cook et al., 2016) RCT compares *Lactobacillus rhamnosus* GG (Culturelle™ , i-Health, Inc.) or a placebo (microcrystalline cellulose and titanium dioxide preparation; i-Health, Inc.) on the risk of developing ventilator-associated pneumonia and other nosocomial infections in critically ill patients. From each site participating in the PROSPECT Pilot Trial, 1 of every 100 capsules (placebo and probiotic) scheduled for administration to patients were sent via regular post. Shipping and handling procedures followed established protocols ensuring the integrity of the study product during transportation.

The study was approved by the Hamilton Integrated Research Ethic Board (REB) and by each participating centres REB board.

Assessment of the bacterial content of the study product

Bacterial content of the capsules was assessed by opening aseptically the capsule and diluting the content in sterile H_2O and subsequently diluting in culture media for the probiotic capsules. The placebo capsules were not further diluted. Colony-forming units (CFUs) were enumerated on Man, Rogosa and Sharpe (MRS; BD) and Brain Heart Infusion (BHI; BD) agar media with the appropriate dilution depending on the sample

(placebo vs. probiotic). The presence of microorganisms on the surface of the placebo capsules was investigated prior to opening the capsules to test the content by rolling over the capsule on a BHI agar plate. Agar plates were incubated at 37°C for 48 hours in aerobic + 5% CO_2 conditions and an additional incubation at room temperature for >24 hours.

Temperature stability study

We evaluated the stability of *L. rhamnosus* GG capsules (Culturelle™) at 3 temperatures (room temperature [~ 23°C], 30°C and 37°C) for up to 14 days. These probiotic capsules were selected from 2 different lots (Lot L4 and L5). Capsule CFUs were determined on days 0, 7, 10, and 14. CFU assessments at each time point for each of 2 lots tested were performed in 3 randomly selected capsules. The total number of viable bacteria in each of the capsules tested was evaluated by enumerating the diluted content of the probiotic capsule.

Strain typing of Lactobacillus clinical isolates recovered from adverse events

Serious adverse events for the PROSPECT study are described in Johnstone, J., and colleagues (Johnstone et al., 2019b). Briefly, the recovery of an isolate identified to the *Lactobacillus* genus is deemed as an adverse event (AE) when it is recovered from a sterile site or cultured as a prevalent organism from a non-sterile site and is further qualified as a serious adverse event (SAE) when the presence of the microorganism could lead to life-threating situations for the patients. Clinical isolates of *Lactobacillus* spp.

recovered from potential adverse events at each of the sites involved in the PROSPECT study were shipped to McMaster University (Hamilton, Ontario, Canada) for strain typing when available.

Genomic DNA was obtained by adding a colony in 50µL of 5% Chelex and incubating the mixture at 95 °C for 15 minutes. Supernatant was recovered and used as template for PCR amplification. The *Lactobacillus spp*. isolates were tested by PCR in order to determine if they were *Lactobacillus rhamnosus* strain GG by using primers targeting the prophage Lc-Nu as described previously (Brandt and Alatossava, 2003). Colony PCR was performed using Lc-Nu_f (5'- TATCTTGACCAAACTTGACG-3) and Lc-Nu_r (5'- CAATCTGAATGAACAGTTGTC-3'). The PCR conditions included an initial denaturation at 94˚C for 2 min followed by 35 cycles of 94˚C for 30 sec, 54˚C for 30 sec, 72˚C for 45 sec, followed by a final extension at 72˚C for 10 min.
GG-negative isolates were further investigated using Random Amplification of Polymorphic DNA (RAPD) PCR. The PCR reaction was completed in 30µL and included 1µL of template, 20 pmol of RAPD-1254 primer (5′-CCG CAG CCA A-3′), 250 µM of each dNTPs, 3 mM $MgCl_2$ 1 U of Taq polymerase (Invitrogen). The PCR conditions were followed as described previously (Pacheco et al., 1997).

Comparative genomic analysis

Genomic DNA was extracted using a commercially available kit following the manufacture's recommendation (Wizard kit, Promega, A1125). Libraries were constructed using Nextera XT DNA Sample Prep Kit (Illumina) and 2 x 250 paired end

read sequencing was performed on an Illumina MiSeq by the McMaster Genomics Facility (Hamilton, Ontario, Canada). Illumina reads were subsequently assembled using the a5 pipeline (Coil et al., 2015). Genomes were annotated using the RAST tool kit (Brettin et al., 2015) via the PATRIC (Pathosystems Resource Integration Center) web-based interface (Wattam et al., 2014, 2017) and visualized with BRIG (Alikhan et al., 2011). Core genome analysis was performed with Roary (Page et al., 2015) and genomic variants were identified with breseq (Version 0.30.2, (Deatherage and Barrick, 2014)). *L. rhamnosus* reference genomes were downloaded from NCBI (accessed on October 1st, 2019). FastTree (Price et al., 2009) was used to generate a maximum-likelihood phylogenetic tree from the core gene alignment using a generalized time-reversible and the discrete gamma models. FigTree (Rambault, 2009) was used to visualize and modify the tree.

Statistical analysis

We conducted a one-sided t-test with a mu_0 of 10^{10} to ensure that the capsule counts remained above the 10^{10} CFUs/capsule threshold. For the assessment of the study product stability, a one-sided t-test was performed to compare CFU counts between results at room temperature and at 37°C. A p value < 0.05 was deemed significant.

Results

Capsules included in the quality assessment studies

In total, 771 capsules were sent and processed between September 2013 and June 2019 (370 probiotic and 401 placebo capsules; Figure 3.1). From those 771 capsules, 742 were sent from centers enrolled in the PROSPECT trial and 29 capsules were sent from the method center (PROSPECT leading site). 38 participating centers provided capsules for the quality assessment sub-study. 126 capsules were excluded from the analysis due to various reasons, but largely due to damage during transportation and loss of the capsules during storage (Figure 3.2). 645 capsules were included in the final quality assessment analysis, including probiotic capsules from 6 different lots (28 from lot 1, 2 lot 2, 52 lot 3, 178 lot 4, 3 lot 5, 47 lot 6) and placebo capsules from 5 lots (25 from lot 1, 34 lot 2, 50 lot 3, 226 lot 4).

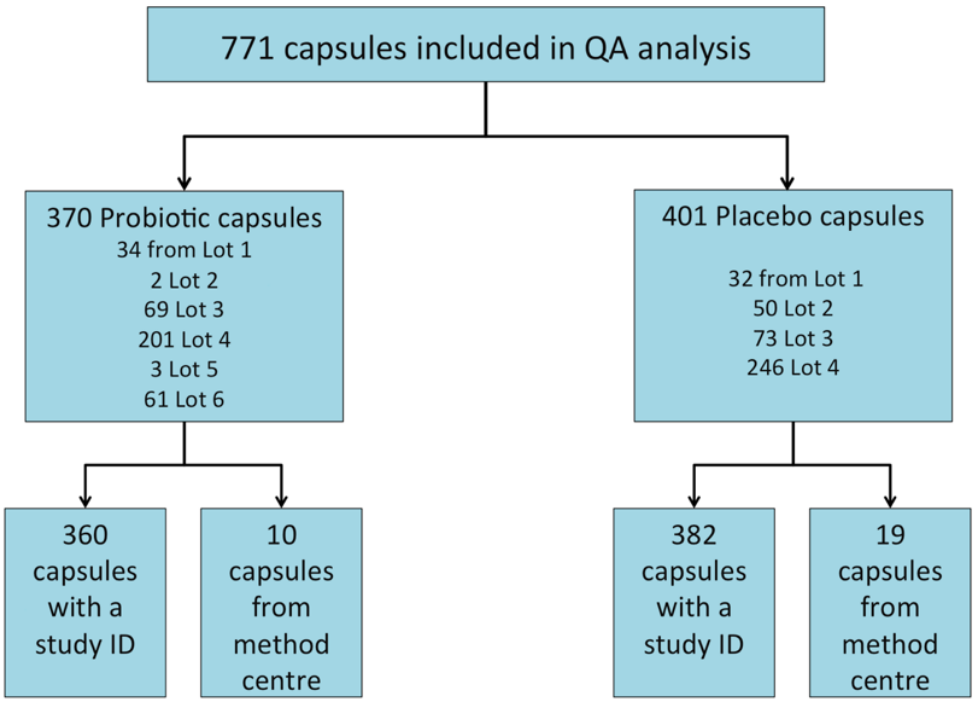
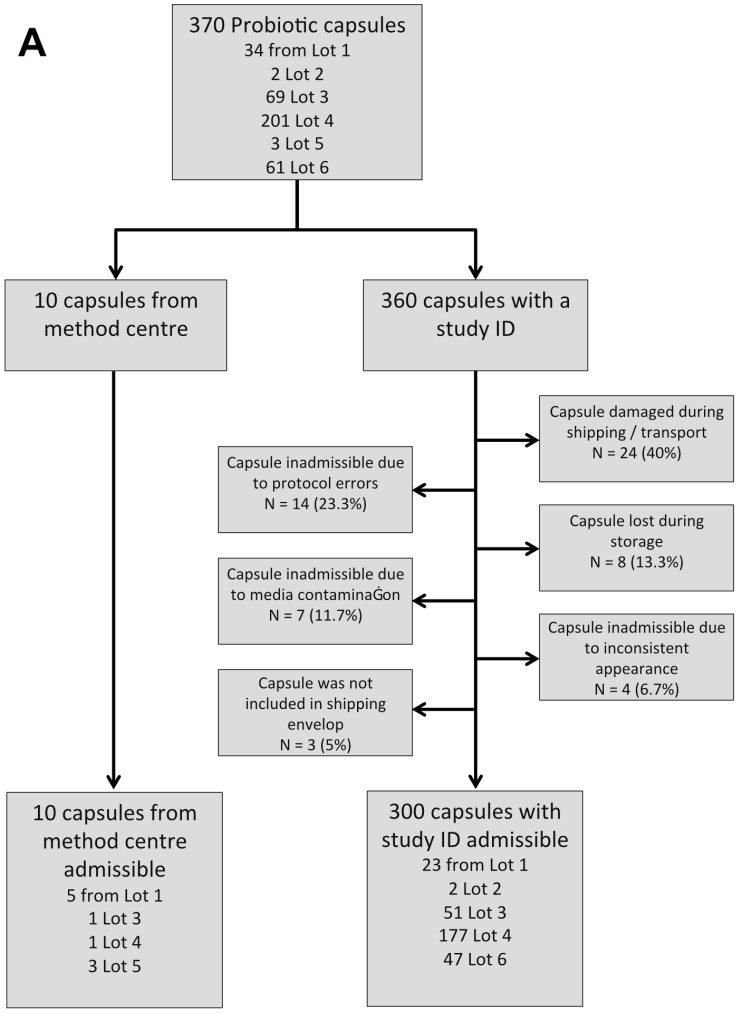


Figure 3.1: Capsules received as part of the quality assessment study.


A
370 Probiotic capsules
34 from Lot 1
2 Lot 2
69 Lot 3
201 Lot 4
3 Lot 5
61 Lot 6
10 capsules from method centre
360 capsules with a study ID
Capsule damaged during shipping / transport
N = 24 (40%)
Capsule inadmissible due to protocol errors
N = 14 (23.3%)
Capsule lost during storage
N = 8 (13.3%)
Capsule inadmissible due to media contamination
N = 7 (11.7%)
Capsule inadmissible due to inconsistent appearance
N = 4 (6.7%)
Capsule was not included in shipping envelop
N = 3 (5%)
10 capsules from method centre admissible
5 from Lot 1
1 Lot 3
1 Lot 4
3 Lot 5
300 capsules with study ID admissible
23 from Lot 1
2 Lot 2
51 Lot 3
177 Lot 4
47 Lot 6

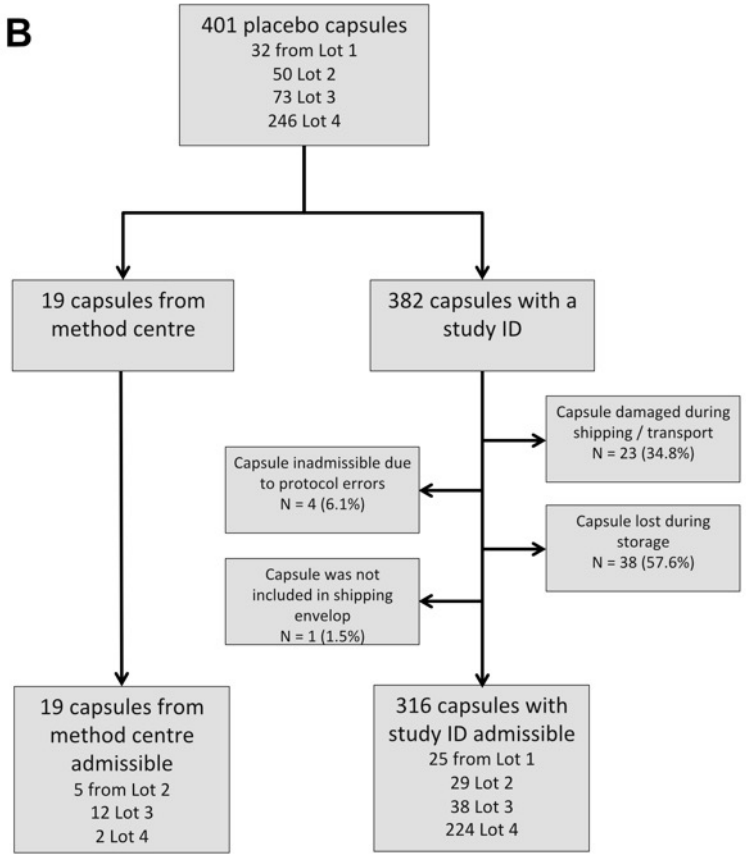


Figure 3.2: Organization schematic of the capsules included in the final quality assessment analysis. Flow-chart of the probiotic (A) and placebo (B) capsules included and stratification of exclusion criteria. A total of 66 capsules sent from the study sites were excluded due to various reasons. The percentage on the figure represents the number of capsules excluded from the 66 capsules for each individual category.

Probiotic capsules contain the expected dose of Lactobacillus rhamnosus GG

The dose consistency investigation demonstrated that every probiotic capsule (n=310) contained at least 10^{10} CFUs of *L. rhamnosus* GG (Figure 3.3A). The median CFUs recovered from all tested samples was $3.4x10^{10}$ (IQR $1.86x10^{10}$) CFUs/capsule (Figure 3.3A). All but one lot tested remained within the 3 standard deviations recommended

threshold (Figure 3.3B). Indeed, the capsules from lot 1 displayed higher bacterial content generally and a greater variability compared to the other lots tested.

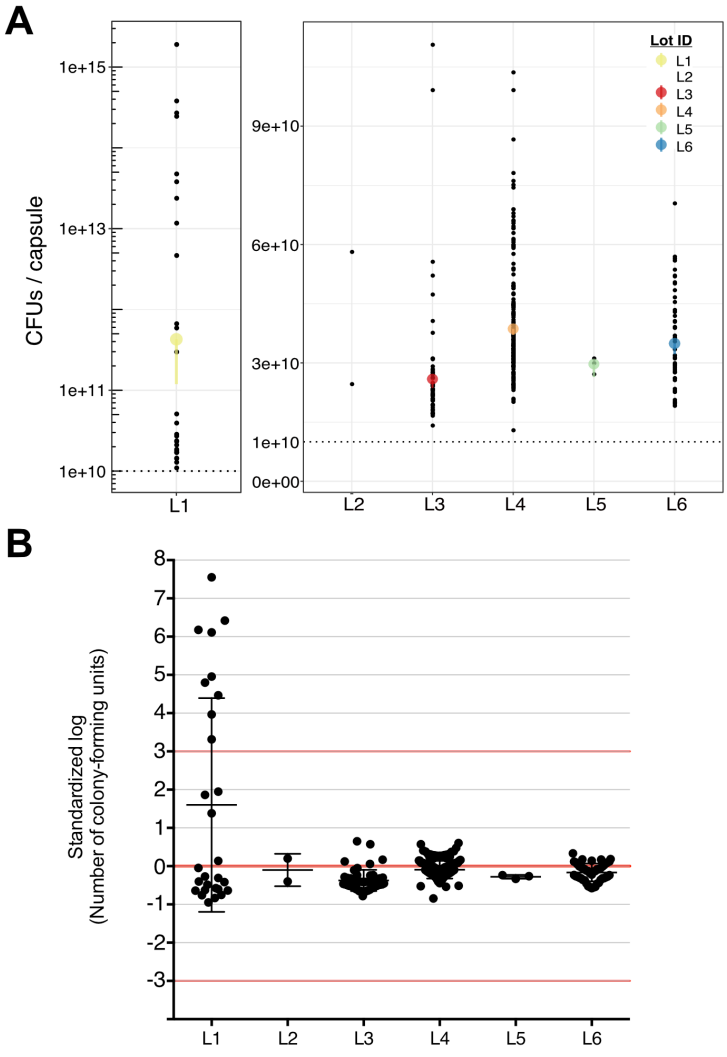


Figure 3.3: Bacterial CFUs for each probiotic capsule tested remained above the expected 10^{10} colony-forming units threshold. Bacterial load was measured by serial dilution. CFUs/capsule (black points) and estimated means (colored points) are displayed by lot (A: Due to higher variability, lot 1 was plotted on a separate scale). Bacterial CFUs visualized on a logarithmic scale using the mean and the overlaying lines represent the standard deviation (B).

Modest and lot-dependent presence of bacteria in the placebo capsules

From the 335 placebo capsules, no microorganisms were detected for 225 capsules (67.2%; limit of detection = 10^2 CFUs/capsule), and the remaining 110 recovered modest bacterial growth (Median 725 [IQR 4800] CFUs/capsule; Figure 3.4). Manufacturing quality control specifications have a maximum threshold of $1.5x10^3$ lactic acid bacterial CFUs in placebo capsules. 47 placebo capsules exceeded the threshold, with 34/47 capsules from a single lot (Figure 3.4A). The presence of bacteria seems to be variable between lots as well as within lots (Figure 3.4B), where lot 4 had both extremes, capsules with low vs. elevated loads. Microorganisms were recovered from the outside of the placebo capsule about 85% of the time (51/315 capsules had no microorganisms recovered from the outside, 20 capsules were not tested).

16S rRNA gene sequencing was performed on the diverse morphotypes recovered to investigate the identity of the bacteria recovered (Table 3.1). Interestingly, *L. rhamnosus* GG and other species used commercially as probiotics strains such as *Lactobacillus acidophilus*, *Lactobacillus rhamnosus* (different than the GG strain) and *Bacillus coagulans,* were the most frequent organisms recovered.

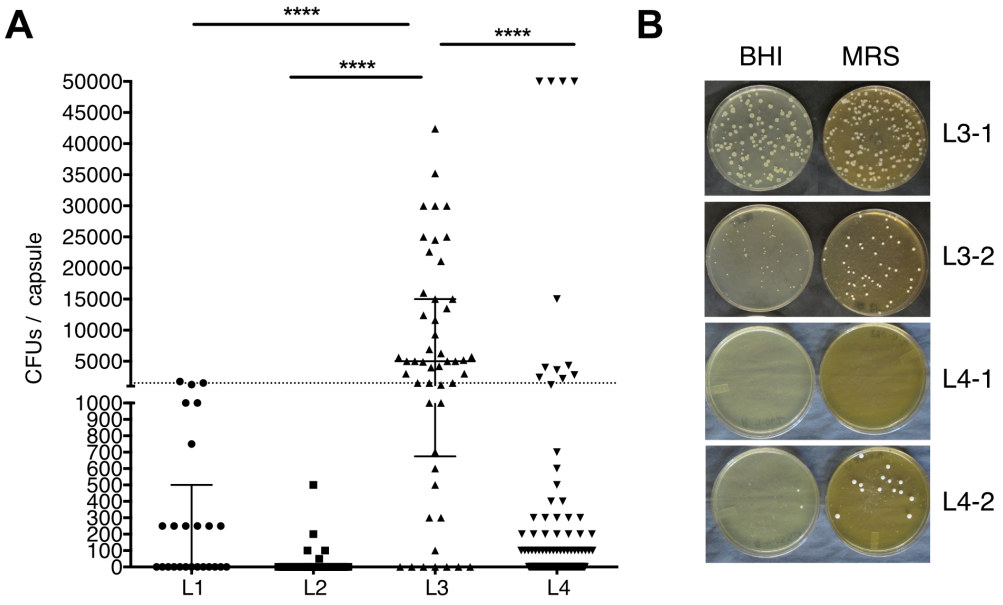


Figure 3.4: Bacterial contamination of the placebo capsules was not detected in the majority of samples and low levels observed in the remaining samples. The placebo bacterial counts were obtained by serial dilution and enumeration on agar plates. CFUs are organized by lots and the mean and standard deviation are displayed (A; Kruskal-Wallis and Dunn's multiple comparison tests, **** indicates a p value ≤ 0.001). The dotted line represents the manufacturer's threshold (1.5×10^3 CFUs/capsule). Variability within the bacterial count and colonies morphology is observed between and within lot (B)

Table 3.1: Taxonomic identification of the most common morphotypes recovered from placebo capsules

Organism identified from placebo capsules	Frequency
Lactobacillus rhamnosus	Common
Lactobacillus acidophilus	Common
Bacillus coagulans	Common
Lactobacillus gasseri	Rare
Lactobacillus casei	Rare
Lactobacillus plantarum	Rare
Bacillus subtilis	Rare
Micrococcus luteus	Rare
Staphylococcus capitis (CoNS)	Rare
Staphylococcus caprae (CoNS)	Rare
Staphylococcus pasteuri (CoNS)	Rare
Staphylococcus haemolyticus (CoNS)	Rare
Staphylococcus epidermidis (CoNS)	Rare
Staphylococcus hominis (CoNS)	Rare
Staphylococcus petrasii (CoNS)	Rare
Staphylococcus aureus	Rare
Corynebacterium mucifaciens	Rare
Fungi	Rare

Common is defined by being present at > 10% of the capsules tested and rare is < 10% of the capsules tested. CoNS: Coagulase-negative staphylococci

The probiotics are sensitive to temperature changes

During the PROSPECT trial, one lot of probiotic study product was kept at 35˚C for 24 hours during shipping to an international site. This prompted us to investigate the stability of the bacterial content of the capsule at higher temperatures, to mimic the conditions potentially observed during storage and shipping. A study was performed on two lots (L4 and L5). In total, 3 capsules from both lots were tested at baseline and 3 capsules for each of 3 different temperatures (Room temperature, 30 and 37°C) were tested at day 7, 10 and 14 for each lot (n=60). Of the 60 capsules tested, 4 (6.7%) were excluded from the

analysis due to contamination of the culture medium. Following the 14 days, all capsules tested (at all temperatures) maintained more than the minimal threshold of 10^{10} CFUs/capsule of *L. rhamnosus* GG. Storage at room temperature and 30°C maintained similar stability, whereas capsules stored at 37°C had a slight, but statistically significant, decrease in CFUs compared to room temperature storage at days 10 and 14 (Figure 3.5) while remaining above the minimal threshold of 10^{10} CFUs/capsule. A modest decrease in CFUs between the room temperature controls and 37°C was observed at day 10 and 14 for both lots (Lot 4 day 10 $p = 0.006$, day 14 $p = 0.0003$; lot 5 day 10 $p = 0.027$, day 14 $p = 0.0009$). Both lots were significantly above the 10^{10} threshold at room temperature and 30°C at all timepoints. Lot 4 estimated means was not above the threshold at 37°C on day 14, and lot 5 did not have sufficient data (n=2) at 37°C on day 14 to conduct a statistical test.

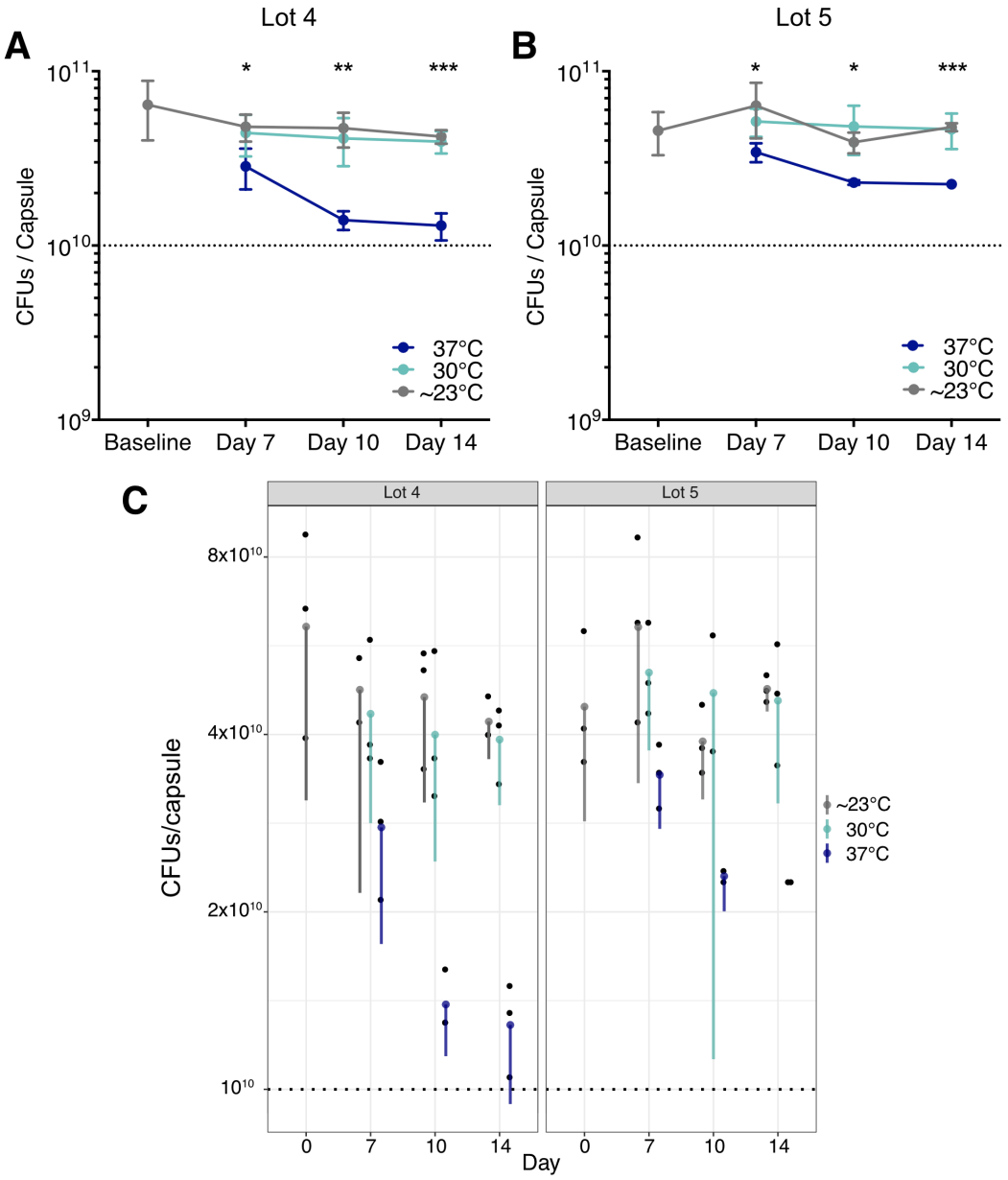


Figure 3.5: Capsules remained above the 10^{10} threshold when stored at room temperature or at 30°C and decreased towards the threshold at 37°C. Bacterial counts for each condition for lot 4 and lot 5 was assessed. This experiment was performed in triplicate. CFUs results are presented using the mean and the overlaying lines represent the standard deviation for lot 4 (A) and lot 5 (B). Statistical analysis was performed using an unpaired one-sided t test at day 7, 10 and 14 between capsules stored at ~ 23°C and 37°C (***p value ≤ 0.001, ** a p value ≤ 0.01, * a p value ≤ 0.05). Observed CFUs/capsule (black points) and estimated means (coloured points) in lots 4 and 5 over a two-week period are displayed (C). The lower end of the 95% confidence interval of the mean is shown with a coloured line.

Microbiological analysis of Lactobacillus isolates associated with AE

During the PROSPECT study, 26 potential adverse events (AE) associated with a *Lactobacillus* positive culture were recorded. From these 26 potential AEs, 16 clinical isolates were preserved and sent to the Surette laboratory for further strain level analysis (Table 3.2). The *L. rhamnosus* GG (*L. r* GG)-specific PCR identified 13/16 isolates as the study product. To confirm the negative isolates, the three isolates were genotyped via Random Amplification of Polymorphic DNA (RAPD) PCR. The three profiles (AE2, AE8, AE11) were distinct from the *L. r* GG strain (Supplementary Figure 3.1). Whole genome sequencing was performed for AE2 and AE8 and confirmed the result of the *L.r* GG-specific PCR and RAPD profiles that the isolates were not *L. rhamnosus* GG (Figure 3.6) and that they are more closely related to *L. rhamnosus* ATCC 8530 according to whole genome sequencing analysis. Both genomes are similar in number of coding sequences, genome sizes, GC content, specialized genes, sequences and assembly's overall quality (Table 3.3).

Table 3.2: Clinical isolates recovered from adverse events

ID	*L. r* GG PCR	Source	Additional information	Subsequent blood cultures	Final Results
AE1	+	Blood; Arterial line		Negative	AE
AE2	-	Blood		NA (discharged)	no AE *
AE3	+	Blood		Negative	AE
AE4	+	Blood; Arterial line	1/4 cultures positive; multiples organisms recovered	Negative	AE
AE5	+	Blood; Central line	1/4 cultures positive	NA (deceased)	AE
AE6	+	Venous catheter; Tip central jugular	tip was positive for *Candida* and blood cultures were negative	NA	AE
AE7	+	Blood; Arterial line	Central line negative	Negative	AE
AE8	-	Gallbladder fluid		NA	No AE *
AE9	+	Blood; Central line	Mixed with other organisms; Arterial and PICC lines cultures were negatives	Negative	AE
AE10	+	Blood		Negative	AE
AE11	-	ETA	Low abundance; Mixed with *Candida tropicalis*	Negative	no AE †
AE12	+	Pelvic abscess fluid	2/3 cultures were positives and mixed with *E. coli* and gram-negative bacilli	Negative	AE
AE13	+	Urine; Foley catheter	Mixed with *Escherichia coli*	Negative	AE
AE14	+	Blood, Catheter PICC line	1/3 cultures positive, 1/3 recovered *Enterococcus faecalis*	Negative	SAE
AE15	+	Blood; Central line	1/2 culture positive	Negative	AE
AE16	+	Blood: Multiple lines	Multiples lines positive	Negative	SAE

Samples were collected outside the surveillance period* or was not the dominant microorganism in a non-sterile site† and did not qualify as an adverse event.

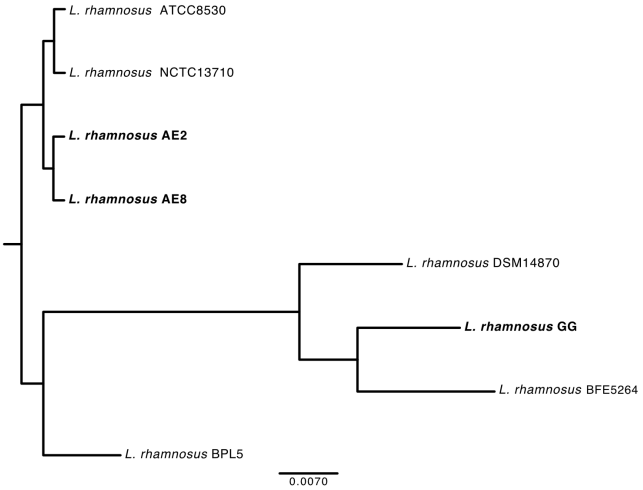


Figure 3.6: Maximum likelihood phylogeny generated from the alignment of 2103 core genes within 8 *Lactobacillus rhamnosus* strains. The SNPs core gene tree demonstrates the great similarity between AE2 and AE8 and highlights an increased divergence between the *L. r* GG reference and AE2 and AE8. Figtree was used to edit the tree.

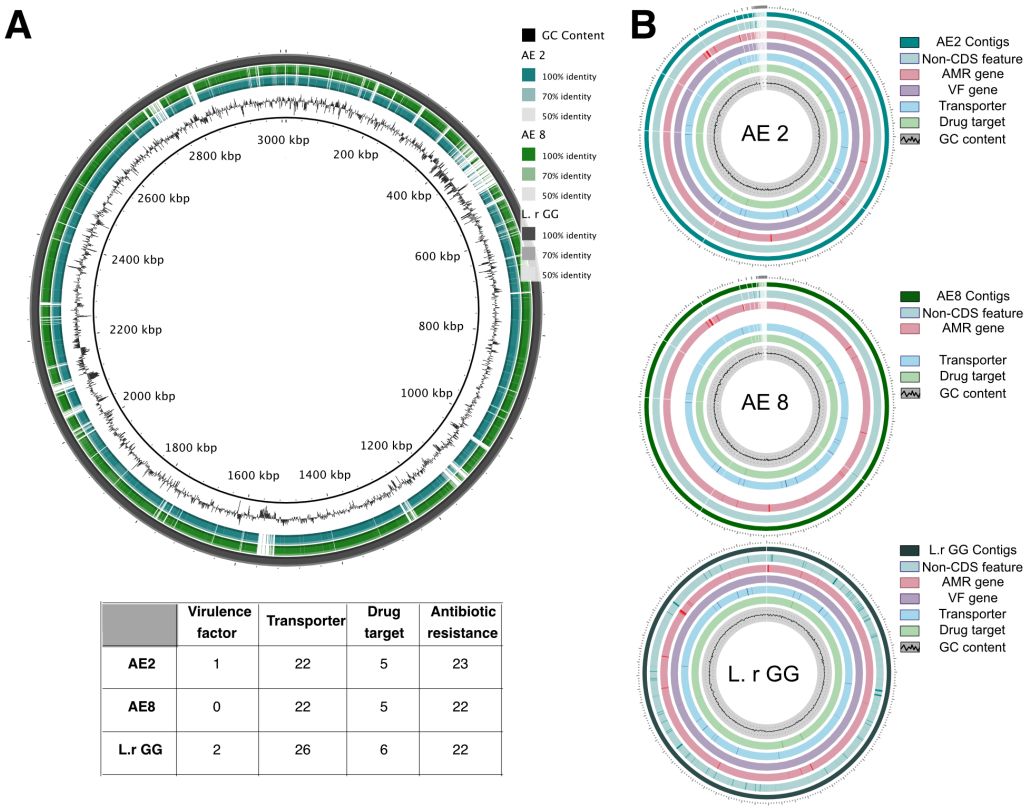
Table 3.3: Assembly and annotation statistics

Assembly	**AE2**	**AE8**
# contigs	25	20
Largest contig (bp)	1049485	1049484
Total length (bp)	2935032	2932897
GC (%)	46.7	46.71
N50 (bp)	708839	710149
L50	2	2
# CDS	2926	2924
# tRNA	58	63

CDS: coding sequence

Pangenome analysis reveals that *L. r* GG and the isolates from AE2 and AE8 share 2269 coding sequences (CDS) in common and 1287 genes were shared between at least 2 isolates (Figure 3.7). When investigating the presence of small sequence variants

(SNVs; including single nucleotide polymorphisms and indels) between AE2, AE8 and a close reference genome (ATCC 8530), only 37 divergent SNPs were identified between the reference genome and the clinical *Lactobacillus* isolates (Table 3.4; where 18 and 19 SNPs were, respectively, unique to AE2 and AE8). SNVs identified at low coverage position were excluded from this final count. Interestingly, although very similar, the clinical isolates AE2 and AE8 were recovered from different patients in different provinces, more than a year apart.



	Virulence factor	Transporter	Drug target	Antibiotic resistance
AE2	1	22	5	23
AE8	0	22	5	22
L.r GG	2	26	6	22

Figure 3.7: Comparison of *L. rhamnosus* recovered from adverse event and *L. rhamnosus* GG. A) Genomic comparison between the strains AE2 (teal) and AE8 (green) with the closed reference *L.r* GG (black) showed a high similarity between the patient's isolates. B) Visualization of specialized genes obtained via PATRIC and RAST for AE2 (teal), AE8 (green), and *L.r* GG (black). A summary table of the presence of specialized genes identified by RAST is included.

Table 3.4: SNPs position and identity between AE2 and AE8

Strain	Mutation	Annotation	Gene description
AE2	T→A	intergenic (-64/+9)	acetyltransferase, GNAT family/N-acetyl-L,L-diaminopimelate deacetylase
AE2	C→T	intergenic (+84/+227)	hypothetical protein/hypothetical protein
AE2	G→A	A119V (GCC→GTC)	NAD-dependent protein deacetylase of SIR2 family
AE2	C→T	S281N (AGT→AAT)	Phenylalanyl-tRNA synthetase beta chain
AE2	+ GC	coding (7-8/807 nt)	hypothetical protein
AE2	T→A	intergenic (-5/-142)	hypothetical protein/hypothetical protein
AE2	T→C	intergenic (-190/-150)	3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form /hypothetical protein
AE2	T→C	intergenic (-204/-136)	
AE2	C→G	intergenic (-205/+239)	Transcriptional regulator, MarR family/hypothetical protein
AE2	+ CT	intergenic (-228/+215)	
AE2	C→T	intergenic (-248/+196)	
AE2	+ CGGG	intergenic (-82/-98)	Deoxyuridine 5'-triphosphate nucleotidohydrolase /hypothetical protein
AE2	+ TT	intergenic (-99/-83)	
AE2	A→G	intergenic (-102/-81)	
AE2	C→T	intergenic (-132/-51)	
AE2	+ AC	intergenic (-145/-37)	
AE2	C→T	intergenic (+2/-335)	hypothetical protein/hypothetical protein
AE2	C→T	intergenic (+8/-329)	
AE2	A→G	intergenic (+10/-327)	
AE8	C→T	H342Y (CAT→TAT)	Putative transcriptional antiterminator, BglG family / PTS system
AE8	G→T	T124T (ACC→ACA)	Phosphoenolpyruvate-dihydroxyacetone phosphotransferase
AE8	C→G	E156Q (GAA→CAA)	Cation-transporting ATPase, E1-E2 family
AE8	C→T	G87R (GGG→AGG)	hypothetical protein

Strain	Mutation	Annotation	Gene description
AE8	G→A	A21T (GCT→ACT)	Transcriptional regulator, AcrR family
AE8	C→T	S30L (TCG→TTG)	Cold shock protein of CSP family
AE8	C→A	D163Y (GAT→TAT)	hypothetical protein
AE8	A→T	R126R (CGT→CGA)	Efflux ABC transporter, permease/ATP-binding protein MdlB
AE8	C→T	S249N (AGT→AAT)	Phosphopantothenoylcysteine decarboxylase/ Phosphopantothenoylcysteine synthetase
AE8	Δ1 bp	coding (246/978 nt)	3'->5' exoribonuclease Bsu YhaM
AE8	C→T	W19* (TGG→TAG)	Regulatory protein Spx
AE8	Δ21 bp	intergenic (-133/+33)	Pyruvate oxidase/Glycine betaine ABC transport system, glycine betaine-binding protein OpuAC
AE8	G→C	P100R (CCC→CGC)	D-lactate dehydrogenase
AE8	C→A	D89Y (GAT→TAT)	LSU ribosomal protein L10p (P0)
AE8	Δ1 bp	coding (1598/2055 nt)	Kup system potassium uptake protein
AE8	G→A	A1074V (GCC→GTC)	DNA-directed RNA polymerase beta subunit
AE8	C→T	S85S (TCG→TCA)	Beta-propeller domains of methanol dehydrogenase type
AE8	G→A	intergenic (-210/+17)	Glycerol-3-phosphate ABC transporter, ATP-binding protein UgpC

green is synonymous, blue is nonsynonymous and red is nonsense, underlined bases indicate the position of a SNP

Of the potential 16 AEs, three were not considered as AE because two were outside the surveillance period (before randomization and 2.5 months post-study) and one did not meet the AE's definition (not the predominant organism in a site considered as non-sterile, in this case an endotracheal aspirate). Interestingly, these three isolates were the only ones that did not correspond to *L. rhamnosus* GG. For 10/13 of the AEs, bedside contamination was suspected since only one of the daily blood cultures was positive, or subsequent blood cultures were negative for *Lactobacillus*. Follow-up cultures were not obtainable for three patients (Table 3.2).

Discussion

As natural health products, probiotics are less stringently regulated than traditional therapeutics. The requirements for quality control are minimal although many manufacturers exceed the required testing. While the safety of probiotics in a healthy population is not disputed, the use of probiotics in patient populations such as the critically ill merits further scrutiny with respect to the product and adverse events that may be directly associated with the study product (*e.g.* positive culture from sterile sites or predominant organism in a non-sterile sites). For the PROSPECT trial, we have carried out extensive quality control of probiotic and placebo capsules and performed additional microbial analysis on *Lactobacillus* isolates from suspected AEs. To the best of our knowledge, this constitutes the first quality assessment study nested in a larger randomized control trial (RCT) looking at the consistency and identity of the

administered study product as well as investigating other potential limitations such as stability of the product at various temperatures and monitoring the safety of the product.

Discrepancies between the expected and recovered viable dose in commercially available probiotics have been previously reported (Fasoli et al., 2003; Marcobal et al., 2008; Angelakis et al., 2011; Goldstein et al., 2015; Lewis et al., 2016; Morovic et al., 2016); however, Culturelle™ products have indicated similar viable counts between the advertised and the recovered amount (Toscano et al., 2013; Goldstein et al., 2014). In our quality assessment, all probiotics capsules tested (n=310, from 6 lots) were above the manufacturer's specifications ($>10^{10}$ CFUs/capsule). Moreover, no decrease in viable organisms to below the manufacturing threshold was observed, even when approaching the expiration date. We did, however, observe an effect of temperature on the stability of the probiotic. While no significant difference in CFUs was detected between the capsules incubated at room temperature and those incubated at 30°C after 2 weeks, we observed a negative impact on the bacteria recovered from the probiotic capsules at higher temperatures (37°C) after 7 days. Importantly, the product still maintained the minimal dose of viable organisms. This suggests that the capsules remained above the manufacturer's specification during storage and shipping during the PROSPECT trial. Hence, we are confident that the PROSPECT Pilot Trial capsules tested during this quality control study reflect the doses prescribed by the protocol and that, consequently, patients in the probiotics arm received a consistent dose during the trial.

For the placebo capsules, the majority of the capsules were sterile (225/335). However, we detected low level contamination (median = 750 [IQR 4800], average 6.46±

$12.1x10^3$ CFUs/capsule) in the remaining ones. The presence of bacteria seemed to be lot-dependent, where more bacterial contaminants were recovered in one lot compared to the 4 other lots (median Lot1-2-4: 0, Lot3: $5x10^3$ CFUs/capsule), emphasizing the need to test every lot used in the study. Lot 3 was discontinued from the larger trial once we noticed the contamination, limiting the potential discrepancy between lots during the final analysis of the clinical trial outcomes. The vast majority of the identified contaminants were probiotic organisms at doses that would not be expected to have biological activity. Numerous host factors such as genetics, microbiota composition and immunity, modify the intestinal environment which influences the efficiency of colonization by novel bacterial strains. Consequently, colonization and engraftment from probiotics are highly variable between individuals (Derrien and van Hylckama Vlieg, 2015; Maldonado-Gómez et al., 2016). Only a limited number of studies have investigated the relationship between dose and colonization especially at lower doses, but it has been suggested that a high inoculum (>10^{10} CFUs) is necessary for *L.r* GG transient colonization and that subjects receiving fewer bacteria (from 10^6 to 10^9 CFUs/capsule) failed to have detectable amounts in feces of healthy individuals (Saxelin et al., 1991). Several aspects of the host physiology in a disease state, as well as the impact of common clinical interventions (*e.g.* antibiotics, proton-pump inhibitors), could impact colonization by the probiotic strains. Consequently, more studies are necessary to comment on the colonization and the potential level of engraftment of probiotics in critically ill patients and its impact on long-term clinical outcomes. However, at a dose of at least 1 million times less potent than the

probiotics capsules, we can hypothesize that the biological activity of the placebo would be negligible in comparison to the administration of the regular probiotics.

Probiotics are 'generally recognized as safe' in the general healthy population. However, clinical complications due to probiotics administration is a genuine concern for the use of live biotherapeutics in patients. *Lactobacillus*-related bacteremia is a possibility in patients with comorbidities and/or immunocompromised (Cannon et al., 2005; Doron and Snydman, 2015; Kothari et al., 2019). Currently, AEs are inconsistently reported and in order to evaluate the safety of the probiotic outside of the healthy population scope, RCTs need to properly report them. As part of the PROSPECT trial safety monitoring, 26 potential adverse events associated with *Lactobacillus* were recorded; however, only 16 had retained the *Lactobacillus* isolate. From those 16 events, 13 were confirmed as a true AE. A recent study has reported an increased risk for ICU patients to develop probiotic-associated bacteremia, with 6 cases reported in 522 patients (Yelin et al., 2019). This highlights the need for more safety monitoring studies. In our study, from the 16 available isolates, 13 were identified as the study product, *L. rhamnosus* GG, and three identified as another strain of *L. rhamnosus*. This suggests that probiotics, as well as indigenous strains, have the potential to generate an infection, as previously demonstrated (Doron and Snydman, 2015). Interestingly, our comparative analysis of two out of the three recovered strains of *Lactobacillus rhamnosus* (non-GG strain) demonstrated a high similarity between the two isolates, suggesting a close ancestor, though they were from two separate patients, in different provinces collected at different times. This could

potentially suggest another commercially available strain used in probiotic preparation or in the food industry (dairy product).

Bedside contamination was suspected for 10/13 confirmed AE isolates, based on the fact that only one of multiple blood cultures were positive on one day or if all subsequent cultures were negative despite no treatment regimens in place to target *Lactobacillus*. This issue highlights the importance of properly establishing a protocol for the administration of probiotics in the intensive care unit, in order to prevent/limit these accidental contaminations (Skljarevski et al., 2016).

Conclusions

With disputed evidence demonstrating the efficacy and safety of probiotics in susceptible populations such as critically ill patients, stronger and better-designed RCT are required. Indeed, several aspects of safety monitoring and quality assessment are not reported or simply omitted from RCTs. To limit research biases, this type of study should become the standard for probiotics studies in vulnerable patients (*e.g*. immunocompromised, compromised GI tract).

Acknowledgements

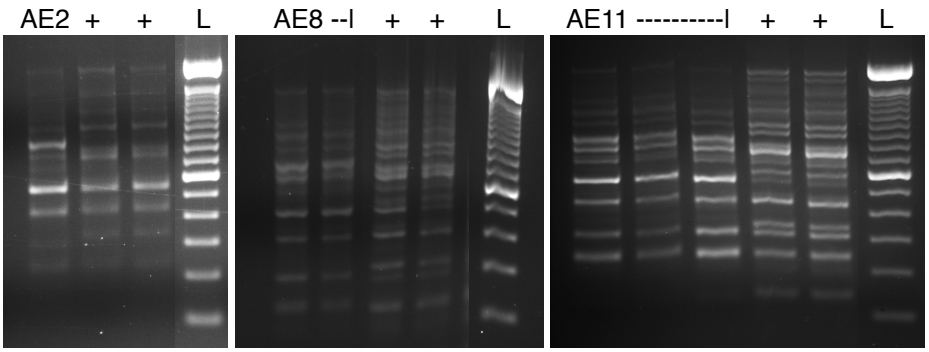
We would like to thank the study coordinators and pharmacists at each PROSPECT site for coordinating the shipment of the study product as well as the clinical microbiology lab for isolating and sending the clinical isolates to Dr. Surette's laboratory. Special thanks to Erika Szymkiewicz for organizing the capsules database and to the investigator team at

the King Abdullah International Medical Research Center (Riyadh, Saudi Arabia) lead by Dr. Yaseen Arabi for reviewing the protocol for the stability study. This work was conducted in affiliation with the Canadian Critical Care Trials Group and the Canadian Critical Care Translational Biology Group.

Ethics declarations

This study was approved by the Hamilton Integrated Research Ethic Board (REB #13–170 and #13–238) and was performed in accordance with the principles of Good Clinical Practice following the Tri-Council guidelines.

Supplementary Material



Supplementary Figure 3.1: RAPD profiles confirmed that AE2, AE8 and AE11 are not *L. rhamnosus* GG. RAPD PCRs and gels were run independently for each isolate. Several controls from different colonies were used to confirm the results. Gels were run two hours at 100V and stained with ethidium bromide. + : *L. rhamnosus* GG.

Chapter 4

Comparative Genomics of the *Streptococcus* Anginosus/Milleri Group

Preface

Research presented as part of this chapter has been prepared for publication as:

Lamarche D, Szymkiewicz R, Surette MG. Comparative Genomics of the *Streptococcus* Milleri/Anginosus Group. 2020. *written for publication*

Contributions:

DL conducted the experiments. DL and RS performed the data analysis. MGS and DL designed the study. DL and MGS wrote the manuscript. All authors approved the final revision of the manuscript.

Abstract

The *Streptococcus* Anginosus/Milleri group (SAG) forms a distinct phylogenetic group of three closely related species: *S. anginosus, S. intermedius,* and *S. constellatus.* The SAG isolates asymptomatically colonize various mucosal surfaces; however, these species have the potential to cause serious infection. Despite the fact that the SAG cause more invasive infections than Group A or Group B *Streptococcus*, less research has focused on this group than other pathogenic streptococci. To investigate the genomic heterogeneity and virulence potential within this group, we performed a comparative genomic analysis of 151 SAG genomes (95 originating from our group and 56 retrieved from NCBI) representing 70, 56 and 25 isolates of *S. anginosus*, *S. intermedius*, and *S. constellatus*, respectively. Whole genome-based phylogenetic analysis confirmed that the three streptococcal species clustered separately from each other, and that the *S. intermedius* clade is more divergent from the other two species. *S. anginosus* isolates exhibited the greatest within-species diversity. *In silico* predictions did not detect the presence of antibiotic resistance genes in the majority of strains (90/151) but did identify 697 biosynthetic clusters with unknown functions and predicted 16 putative novel phages.

Word count: 184 words

Keywords: *Streptococcus* Anginosus Group, *Streptococcus* Milleri Group, comparative genomics, *S. anginosus*, *S. intermedius*, *S. constellatus*.

Introduction

Respiratory infections are a leading cause of illness and mortality worldwide (World Health Organization, 2018a). The World Health Organization has predicted that lower respiratory tract infections will still be among the leading cause of death in 2030 (World Health Organization, 2018b). Despite the clinical and economic burden associated with respiratory infections, the identification of a defined etiological agent is achieved in less than half of cases (Capelastegui et al., 2012; Psallidas et al., 2014; Jain et al., 2015). This low identification rate could be explained by clinical aspects (*e.g.* administration of antimicrobials complicates the recovery/growth of microorganisms or unavailability of lower respiratory tract specimens) but also by the failure of standard clinical microbiology protocols. Indeed, current diagnostic methods have limitations. For example, the routine protocol for the culture of clinical specimens is only optimized to recover certain bacteria (Sibley et al., 2010a; Jain et al., 2015). Furthermore, the rise of previously unsuspected and atypical microorganisms, due to a steadily increasing population of immunocompromised patients, and the fact that many microorganisms are dismissed as harmless but have an underappreciated pathogenic potential complicate the identification of the culprit microorganism.

An unappreciated group of microorganisms causing respiratory infections is the *Streptococcus* Anginosus/Milleri group (SAG) The SAG form a distinct phylogenetic group of streptococci composed of three closely related species: *S. anginosus, S. intermedius,* and *S. constellatus* (Whiley and Hardie, 1989; Whiley and Beighton, 1991). The SAG asymptomatically colonizes various mucosal surfaces (gastrointestinal,

urogenital and upper respiratory tract) and can be recovered in 15-30% of healthy individuals at various mucosal sites (Poole and Wilson, 1979; Gossling, 1988b). However, the SAG is one of the most common causes of invasive streptococcal diseases in humans (Laupland et al., 2006; Reißmann et al., 2010; Darlow et al., 2020). The SAG can be isolated from abscesses, pneumonia, pleural empyema, bloodstream infection and from the lower respiratory tract of cystic fibrosis patients (Shinzato and Saito, 1995; Maskell et al., 2006; Finley et al., 2008; Sibley et al., 2008, 2010b; Asam and Spellerberg, 2014; Laupland et al., 2018; Yamazaki et al., 2019; Parkins et al., 2008). While individual SAG species do not generally present distinct preference in the clinic for a type of infection, *S. intermedius* and *S. constellatus* are more frequently recovered from abscesses, while *S. anginosus* is predominantly identified in blood and urinary tract infections (Whiley et al., 1992; Claridge et al., 2001; Furuichi and Horikoshi, 2018b; Thomas-White et al., 2018; Darlow et al., 2020). Interestingly, studies have demonstrated that infections associated with *S. intermedius* were generally linked to worse prognosis, longer hospital length of stay, and higher mortality rates (Junckerstorff and Murray, 2016) and were associated with more pyogenic infections than the other SAG species (Kobo et al., 2017).

The SAG possesses a variety of potential virulence factors (*e.g* adhesins, hydrolytic enzymes, cytotoxins, a type 7 secretion system, and capsules) (Nagamune et al., 1996; Grinwis et al., 2010a; Hasegawa et al., 2017a; Whitney et al., 2017; Issa et al., 2019). No association between a virulence phenotype displayed by an SAG isolate (*i.e.* production of hydrolytic enzymes) and its capability to cause infection has been

demonstrated, suggesting that infections by the SAG are not caused by hypervirulent strains (Grinwis et al., 2010a). However, the cytokine response of human peripheral blood mononuclear cells to heat killed SAG strains did show differences between isolates from infections versus colonization sites, with a stronger response found in infection isolates (Kaiser et al., 2014).

With the SAG being usually dismissed as harmless or atypical pathogenic bacteria, only a limited number of studies have been performed to investigate genetic heterogeneity within this group of microorganisms. In this report, we undertake a comparative genomics study of 151 isolates of SAG including 95 from our own group (7 already available publicly; (Olson et al., 2013)) and 56 from publicly available databases. We observed that, as expected, each of the species form distinct lineages with the *S. anginosus* displaying the greatest intra-species heterogeneity. The accessory genome is characterized by numerous phages and a large repertoire of uncharacterized biosynthetic gene clusters (primarily bacteriocin-like). This genetic heterogeneity is consistent with the known phenotypic heterogeneity that has made the SAG challenging to diagnose clinically. This expanded database of SAG genomes should facilitate further research and improved diagnostics of this group of underappreciated pathogens.

Materials and Methods

Bacterial strains

The bacterial strains used for this study were isolated from clinical samples recovered from healthy subjects or patients with acute or chronic illness episodes as part of other

ongoing studies in our laboratory (Supplementary Table S4.1; both at the University of Calgary, Calgary, Canada, and McMaster University, Hamilton, Canada) and clinical isolates from Calgary Laboratory Services (Calgary, Canada) as reported previously (Grinwis et al., 2010a, 2010b). Reference strains were obtained from the American Type Culture Collection (ATCC).

Isolates were cultured on McKay agar (semi-selective media for the SAG*;* (Sibley et al., 2010a)) at 37°C in the presence of 5% CO_2. The full 16S rRNA gene was sequenced to confirm the species identification as described previously (Sibley et al., 2008). Liquid cultures were grown in Todd-Hewitt broth supplemented with 0.5% yeast extract (THY, BD) at 37°C in an anaerobic chamber (5% CO_2, 5% H_2, 90% N_2; Shel Labs) overnight. The genomic DNA was extracted using the Wizard Genomic DNA Purification kit (Promega, A1125) according to the manufacturer's recommendations and was stored at -20°C.

Publicly available genomes were added to the analysis. High-quality SAG genomes available from NCBI (accessed: March 2018) were included for subsequent analysis (Supplementary Table S4.1).

Sequencing, assembly, and annotation

The 88 SAG isolates isolated from healthy and acutely/chronically ill individuals were sequenced on the Illumina MiSeq platform by the McMaster Sequencing facility (Ontario, Canada) using Nextera DNA library preparation and standard protocols to generate 2x250 bp paired-end reads. Read quality was assessed via FastQC (Andrews Simon, 2010).

Paired-end Illumina reads were *de novo* assembled using the a5 pipeline (v20160825, Coil et al., 2015). Contigs smaller than 1000 bp with an average coverage below 10-fold were excluded from the analysis. Contigs were ordered with Mauve Multiple Genome Alignment (build date Feb 13 2015, Rissman et al., 2009) to the appropriate species reference: *S. intermedius* B196, *S. constellatus* C1050, or *S. anginosus* C1051. Bacterial gene prediction and annotation was performed using Prokka (v1.12-beta, Seemann, 2014) using high-quality species-specific reference genomes (B196, C1050, C1051, Olson et al., 2013). Bacteriophages gene prediction and annotation was performed using RAST tool kit (Brettin et al., 2015) via the PATRIC (Pathosystems Resource Integration Center) web-based interface (Wattam et al., 2014, 2017). New genome assemblies will be deposited in Genbank upon submission of this work as a peer-reviewed manuscript.

Quality assessment

The coverage information per nucleotide position was obtained by first processing the raw reads using Trimmomatic using default parameters (Bolger et al., 2014), followed by mapping the trimmed reads back to the draft genome using Bowtie2 (v2.2.6; Langmead and Salzberg, 2012). Mapping the trimmed reads to the draft genome consisted of converting the output SAM files from Bowtie2 to BAM files, sorting the BAM files via SAMtools (Li et al., 2009) and calculating the coverage from the BAM files using BEDtools (v2.17.0, Quinlan and Hall, 2010). Depth of coverage of each assembly was determined by taking the mean of the individual coverage values for each position on the assembly. A taxonomic assignment was based on core gene phylogeny and by using

Kraken (Wood and Salzberg, 2014) and by extracting the predicted 16S rRNA gene from the assemblies using RNAmmer (v2.3.3, Lagesen et al., 2007) and by confirming its closest match using the extended human oral microbiome database (Escapa et al., 2018). 16S rRNA sequences extracted with RNAmmer were used to generate maximum-likelihood (ML) phylogenetic trees using a generalized time-reversible (GTR) model under the discrete gamma model with 20 rate categories with FastTree (v2.1.7; Price et al., 2009). Contigs identified as phiX174 (which is used in Illumina sequencing as control DNA and can appear as minor contaminant in the sequence output files) via Blast (Altschul et al., 1990) were manually culled from the draft assemblies. Quality of the draft assembly, the size of the draft genome, the number of contigs, N50, and L50 were assessed with Quast (v4.5, Gurevich et al., 2013).

Pangenome analysis

Pangenome was inferred with Roary (v3.8.2, options "-p 10 -r -t 11 - e -n -o clustered_proteins"; Page et al., 2015) using the gene annotations obtained from Prokka (Seemann, 2014). Association between the pangenome and phenotype was assessed using Scoary (v1.6.16; Brynildsrud et al., 2016). The heatmap showing the presence and absence of genes within the strains pangenome was made using the roary_plots.py (github.com/sanger-pathogens/Roary/blob/master/contrib/roary_plots/roary_plots.py).

In silico predictions

Assembled genomes were processed through PHASTER (Arndt et al., 2016) to identify

prophages, the Comprehensive Antibiotic Resistance Database (CARD, v3.0.7) Resistance Gene Identifier (v.5.0.0; McArthur et al., 2013; Jia et al., 2017) for the identification of antibiotic resistance genes, and antiSMASH (v.5.1.1, options "--knownclusterblast --subcluterblast --asf --inclusive --smcogs"; Medema et al., 2011; Blin et al., 2019) for the identification of potential biosynthetic gene clusters. Predicted intact prophage sequences were extracted from the draft assemblies, and aligned with MAFFT (v7.215, using the model "FFT-NS-2"; Katoh and Standley, 2013). Maximum-likelihood (ML) phylogenetic trees using a generalized time-reversible (GTR) model under the discrete gamma model with 20 rate categories with FastTree (v2.1.7; Price et al., 2009) The presence of a nisin-like biosynthetic gene cluster was further examined by investigating NisB homologues using BLAST (v2.2.28+, using the tblastx default parameters; Altschul, 1997). The search was performed using the amino acid sequences of NisB (accessed on uniprot: Name: NISB_LACLL, Accession: P20103, Organism: *Lactococcus lactis* subsp. *lactis*) against all contigs from the 151 strains. The NisB homologues were aligned with MUSCLE (v3.8.31, default parameters; Edgar, 2004). A maximum-likelihood tree inferred from alignment used the Jones-Taylor-Thornton (JTT) model and CAT approximation (Lartillot and Philippe, 2004) as recommended as the default parameters by FastTree. FigTree (v1.4.2; Rambault, 2009) and iTOL (v5.6; Letunic and Bork, 2019) was used to visualize the phylogenies. Sequence alignments and prophage gene annotations were visualized with Geneious (v9.0.5; https://www.geneious.com). Figures were made using the ggplot2 package in R (R Core Team, 2013; Wickham Hadley, 2016).

Variant calling analysis

Sequence variants were identified with Snippy (v4.0-pre_20180729; https://github.com/tseemann/snippy) and the annotation were obtained via SnpEff (Cingolani et al., 2012). SNPs were visualized with Geneious.

Statistical analysis

Differences between the numbers of biosynthetic gene clusters (BGCs) with specific function across species was investigated using a Kruskal-Wallis test with Benjamini-Hochberg procedure for multiple testing correction. A p value < 0.05 was deemed significant.

Results

Description of the strains selected for this study.

To perform a comprehensive investigation and to avoid potential bias within the analysis, we included strains from diverse locations (anatomical and geographical sites) collected from on-going studies in our laboratory as well as reference genomes obtained from NCBI. Following quality assessment, 151 SAG strains (70 *S. anginosus*, 25 *S. constellatus,* and 56 *S. intermedius*; Supplementary Table S4.1) were included in the analysis, 95 were collected and sequenced by our laboratory (or collaborators) and 56 were recovered from NCBI. The isolates represent various anatomical sites and were collected from healthy subjects and infectious episodes in human subjects with the

exception of a single isolate from the NCBI database that was acquired from a chimpanzee. The final quality assessment summarized by species and of each genome assembly are listed in Table 1 and Table C.1(Appendix C), respectively. All draft genomes had an N50 (*i.e.* the contig length where 50% of the bases of the assembly are covered) greater than 65 kbp (196028 ± 65344 bp) and the mean coverage of the in-house isolates was 461-fold (SD 253). Additional information about the quality assessment is available in Appendix C.

Table 4.1: Summary of assemblies' statistics per species for the 88 draft genomes in-house assembled

Species	# of isolates	# of Contigs	GC content (%)	Assembly length	N50	L50
S. anginosus	31	40 (24)	38.73 (0.18)	1958197 (87912)	171384 (87176)	5 (2)
S. constellatus	14	31 (11)	37.98 (0.10)	1917273 (61053)	251607 (193657)	3.5 (2)
S. intermedius	43	16 (4)	37.57 (0.68)	1958398 (38385)	353928 (219279)	2 (1)

Mean values and (SD) are displayed

The species for each strain was confirmed with the core gene phylogeny, Kraken, predicted full-length 16S rRNA gene sequences and further verified by the identification of the intermedilysin gene, specific to *S. intermedius*. The species identification based on the 16S rRNA gene sequences predicted by RNAmmer was congruent with the core gene phylogeny for 144/151 strains. The identification obtained with Kraken for the in-house strains matched the whole genome core gene phylogeny. The extracted 16S genes from two *S. anginosus* genomes retrieved from NCBI were identified as *Micrococcus luteus*

and *Corynebacterium tuberculostearium* (Sang_CCUG39159, and BV1, respectively) (Supplementary Table S4.2). In the case of BV1, a second 16S gene correctly matched *S. anginosus*. The extracted 16S genes from these strains likely come from contamination during the genome assembly, but this could not be further investigated as no raw read sequences were available. The remaining 5 strains with incongruities between core gene/kraken taxonomy and 16S gene were four *S. constellatus* identified as *S intermedius* by the 16S rRNA gene, and one *S. anginosus* genome with an *S. constellatus* 16S rRNA gene. These were unexpected findings and merit further investigation.

High genetic diversity was observed within and between the SAG species

The phylogenetic relationship between the 151 isolates was investigated by aligning the 440 core genes shared between the three species. The phylogenetic tree is highly congruent with isolate species identification and suggests that *S. constellatus* and *S. anginosus* isolates are more closely related to each other than to the *S. intermedius* clade (Figure 4.1). Moreover, the tree topology suggests higher genetic diversification within the *S. anginosus* clade than the other two SAG species. We have identified predicted subspecies clades with thin the core gene phylogeny based on previous work by other investigators (Jensen et al., 2013; Babbar et al., 2017; Bauer et al., 2020).

To further investigate the genomic diversity observed within the SAG, we performed a pangenome analysis using Roary. Core genes were defined as present in at least in 150 (> 99%) of the isolates, soft genes as present in 144 to 149 (95-99%) of the strains, shell genes as present in 23 to 143 (15-95%) of genomes, and cloud genes as

present in less than 22 (<15%) of the strains included. The number of genes found in each category are listed in Table 4.2. Pangenome analysis demonstrated a high heterogeneity in terms of the total gene diversity within the SAG where *S. anginosus* was determined to be the most diverse species with a pangenome of 9,835 genes followed by *S. constellatus* (5,412 genes) and *S. intermedius* (5,140 genes) (Table 4.2).

Table 4.2: Core and accessory genes organized by species and prevalence

Pangenome	***S. anginosus***	***S. constellatus***	***S. intermedius***	**SAG**
Core genes (present in ≥ 150 isolates)	777	978	1223	440
Soft genes (present in 144 to 149 isolates)	167	156	98	46
Shell genes (Present in 23 to 143 isolates)	1597	1417	937	2851
Cloud genes (Present in ≤ 22 isolates)	7294	2861	2882	11869
Total genes	9835	5412	5140	15206

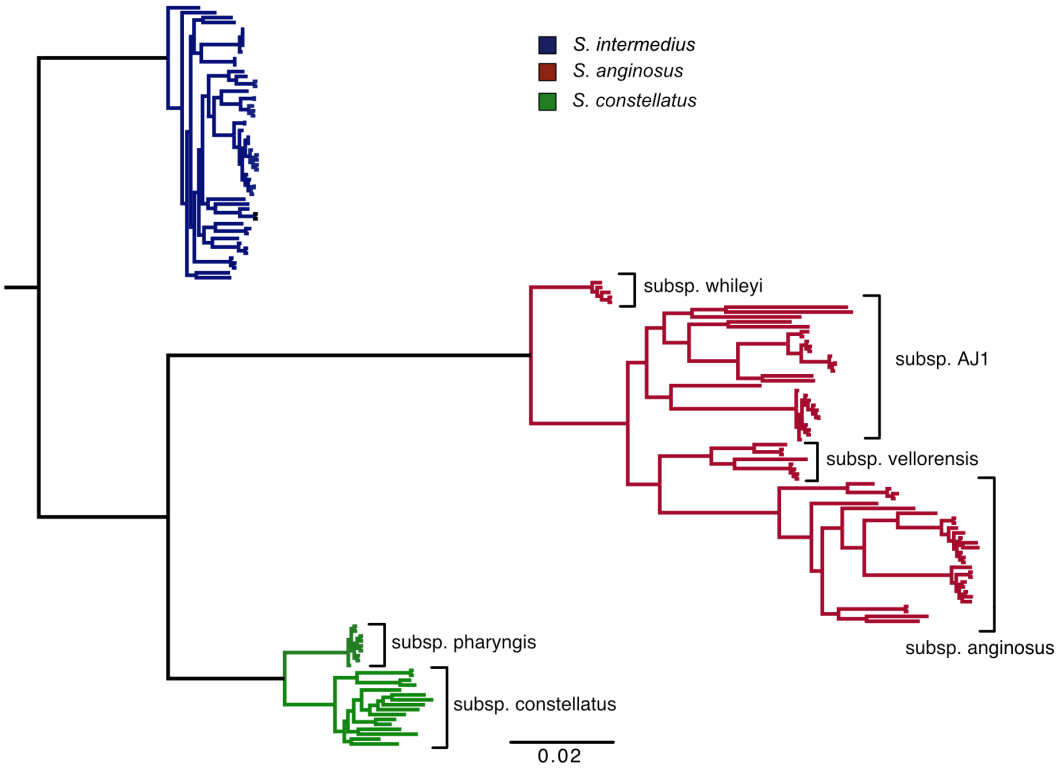


Figure 4.1: Maximum likelihood phylogeny generated from the alignment of 440 core genes within 151 SAG strains. The SNP core genes phylogenetic tree supports the three species-specific clades within the SAG and highlights an increased genetic diversity within the *S. anginosus* species. Figtree was used to visualize the tree. Branch color emphasizes species clades. Branch lengths represent the number of substitutions per site. (subsp.: subspecies)

The pangenome and phylogenetic analyses support the presence of two subspecies within the *S. constellatus* (*pharyngis* and *constellatus*) clade as previously proposed (Whiley et al., 1999). Previous work in our research group has shown that isolates of the subsp. *pharyngis* were not naturally competent in contrast to the other SAG species and subspecies, despite having a functional *comCDE* pathway (competence-related genes, (Lacroix, 2014)).The two *S. constellatus* subspecies demonstrate divergence within their respective pangenomes, where *S. constellatus* subsp. *pharyngis* displays a more

conservative pangenome (higher number of core genes and restricted accessory genome; Supplementary Figure S4.1) when compared to *S. constellatus* subsp. *constellatus*. To ensure this difference was not a bias introduced due to a divergent number of isolates within each subspecies (subsp. *constellatus:* 16 vs. 9 for subsp. *pharyngis),* two simulations were performed on the subsp. *constellatus* to include 9 isolates which further supported these conclusions (Supplementary Figure S4.1BC).

Low prevalence of antibiotic resistance genes within the SAG

The 151 SAG strains were processed through the Comprehensive Antibiotic Resistance Database (CARD) Resistance Gene Identifier (RGI) in order to identify antimicrobial resistance genes. 109 ORFs in 61 strains were identified as potential antimicrobial resistance genes (Figure 4.2). The CARD RGI database did not identify potential resistance genes for the remaining 90 strains. At least one antimicrobial resistance gene was identified in 50% (28/56) of *S. intermedius* strains, 44% (11/25) of *S. constellatus* and 31.4% (22/70) of *S. anginosus* strains. Resistance to five drug classes was predicted, where tetracycline (n=66) was the most frequently identified, followed by macrolide (n=29), glycopeptide (n=8) and aminoglycoside (n=6). In addition, one strain harbored vancomycin resistance genes, which is unusual in streptococci, as reported previously (Srinivasan et al., 2014). Our group had previously demonstrated that ribosomal mutation (substitution of adenine residue) is a prevalent mechanism conferring macrolide resistance in SAG isolates (Grinwis, 2012). Surprisingly, macrolides resistance due to

point mutation in the 23S rRNA gene (CARD identifier: 3004125) was not detected by CARD.

The SAG isolates are known to be resistant to sulfafoxamides and genome analysis reveals that the target of this class of drugs (dihydropteroate synthase) is absent in these organisms accounting for their resistance (data not shown). Dihydropteroate synthase is involved in the tetrahydrofolic biosynthesis pathway which is essential for folate synthesis and is a precursor of several nucleotides and amino acids (McCullough and Maren, 1973). The absence of the gene for dihydropteroate synthetase implies some alternative dihydrofolate biosynthetic pathway or a requirement of an intermediate after this enzyme. However, this resistance was not identified in the *in silico* prediction since the RGI predicts resistance alleles of dihydropteroate synthetase, and the absence of the enzyme is not used to predict resistance.

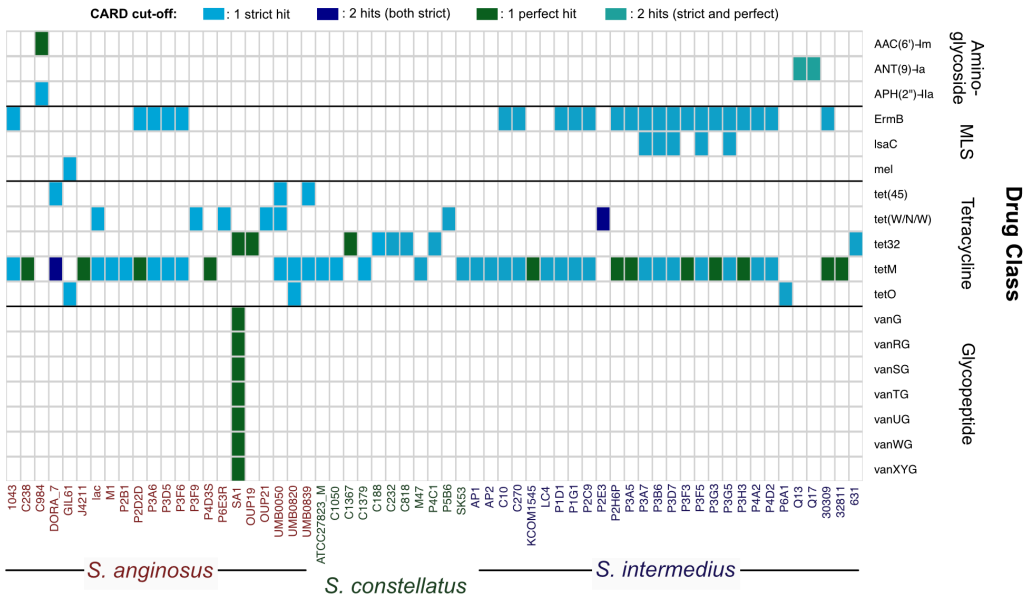


Figure 4.2: Predictions of antibiotic resistance genes of SAG isolates.

Resistance to tetracycline and erythromycin are the most frequently observed resistance. Predicted resistance profiles were obtained via the CARD RGI. Colors within the heatmap display the number of hits and their quality threshold. Strains names are colored by species. The 90 isolates without antibiotic resistance gene identification were omitted from the figure. MLS: Macrolide-lincosamide-streptogramin.

The majority of secondary metabolite biosynthetic gene clusters identified are of unknown function

1442 potential biosynthetic gene clusters (BGCs) were identified using the antibiotics & Secondary Metabolite Analysis Shell (antiSMASH) comprehensive pipeline. Each SAG isolate contained 5-14 biosynthetic gene clusters (median 10 [IQR 1] clusters; Figure 4.3). The majority of the identified clusters had an unknown function and the second most abundant function was associated with bacteriocin production (Table 4.3). The majority of the identified BGCs are small ribosomal encoded peptides, however 27 BGCs associated with nonribosomal peptide synthetases were predicted in the species *S. anginosus* and *S. intermedius* and one type II polyketide synthase was identified in the *S. intermedius* strain TYG1620. Each species had a similar average number of BGCs per strain: *S. anginosus*, 9.11 (638/70); *S. constellatus*, 10.08 (252/25); *S. intermedius* 9.86 (552/56) clusters/strain. The number of BGCs with unknown function is significantly decreased in the *S. intermedius* species while no differences was detected between *S. anginosus* and *S. constellatus* (Figure 4.4). Moreover, *S. intermedius* is enriched with specific classes, namely lantipeptide ($p < 0.001$) and lantipeptide-saccharide ($p = 0.002$) BGCs. These classes of BGCs were virtually absent from the other two species.

Table 4.3: Biosynthetic gene cluster types identified *in silico*

Cluster type	**Frequency**
Bacteriocin	221 (15.33)
Bacteriocin- Non-ribosomal peptide synthetase	1 (0.07)
Butyrolactone	1 (0.07)
Lantipeptide	35 (2.43)
Lantipeptide-saccharide	10 (0.69)
Non-ribosomal peptide synthetase	26 (1.8)
Possible fatty acid	151 (10.47)
Possible saccharide	291 (20.18)
Sactipeptide	8 (0.55)
Type II PKS (Polyketide synthase) - Fatty acid	1 (0.07)
Unknown function	697 (48.34)

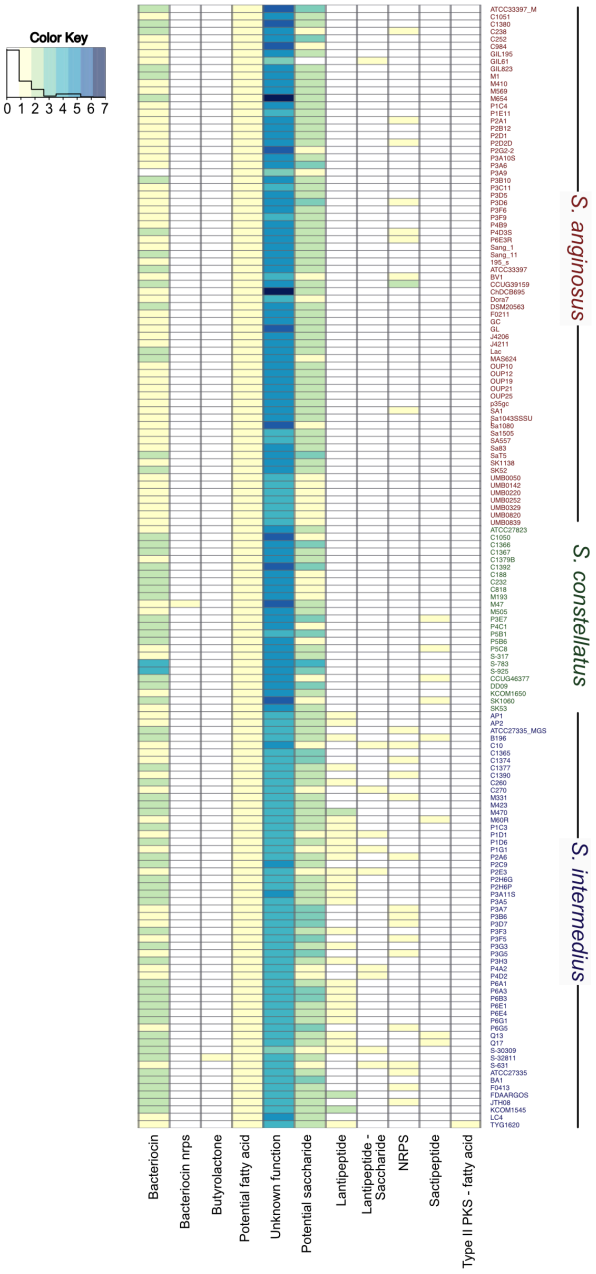


Figure 4.3: Heterogeneity within the presence and function of secondary metabolites biosynthetic gene clusters within the SAG. Gene clusters were identified with antiSMASH. Color displays the occurrence of the class of BGCs.

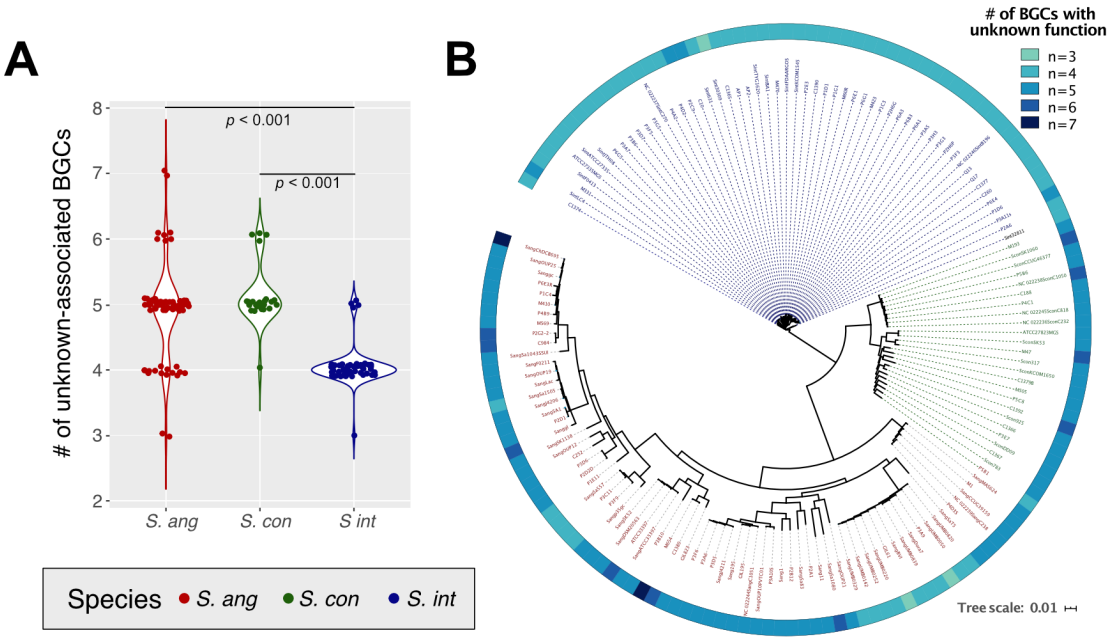


Figure 4.4: Differences within the numbers of BGCs associated to unknown function across SAG species. The species *S. intermedius* has significantly less predicted BCGs associated to unknown function than the two other SAG species (A). Visualization of the number of BCGs with unknown function by species and subspecies with the SAG core gene phylogeny (B). Significance was determined using a Kruskal-Wallis test.

S. intermedius was enriched for secondary metabolite BGCs sharing similarities with genes associated with nisin biosynthesis. Nisin is a bacteriocin of the lantibiotic family produced by *Lactococcus lactis* subsp. *lactis* with uncommon broad-spectrum antimicrobial activity. An intact nisin Q/A BGC was present in 9 *S. intermedius* strains (C270 as the prototypical strain; Figure 4.5). This cluster was associated with a larger conserved cluster predicted to be a polysaccharide biosynthetic gene cluster found broadly across the *S. intermedius*. The nisin Q/A BGC was absent from a small number of strains (*e.g.* AP1, n = 6). The remaining strains appear to have lost part of this region (*e.g.* B196, n = 26) and several key genes were lost from the nisin Q/A cluster. Whether

this partial cluster still produced a molecule with biological activity needs further investigation.

One gene lost in these strains (represented by B196) was *nisB*. However, a blast search for *nisB* genes identified two other related genes. One *nisB* homologue was associated with a separate biosynthetic gene cluster which had no other nisin-like genes identified and we suggest this *nisB* is a paralogue. A more distantly related paralogue was also present in the *S. intermedius* (Supplementary Figure S4.2). These examples highlight the diversity and plasticity of these small BGCs in the SAG.

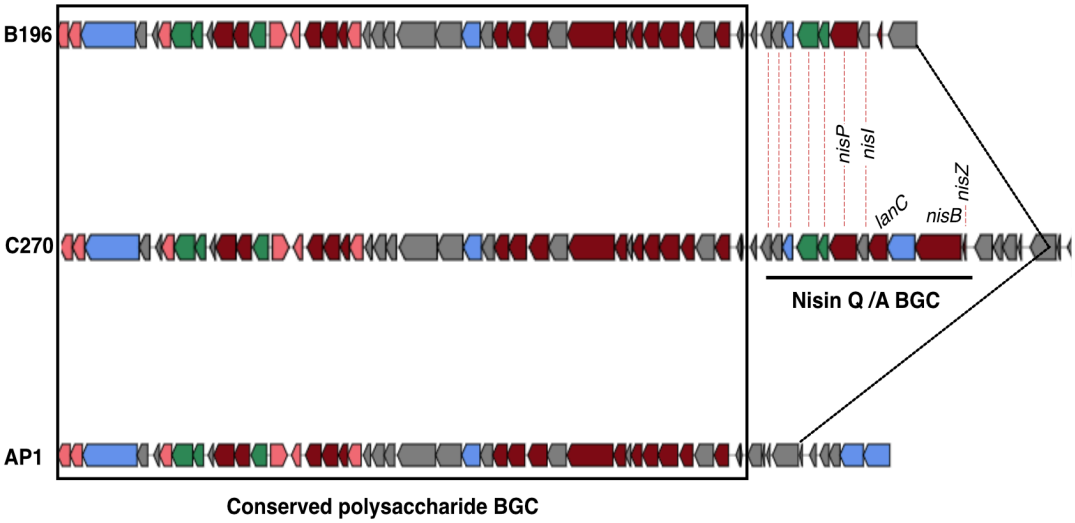


Figure 4.5. Conserved polysaccharide BGC associated with nisin biosynthesis within SAG isolates. The BGC region containing a Nisin Q/A like cluster (represented by C270) showing the genes of this cluster in red. The cluster is not intact in most other strains (e.g. B196) with several key genes in the nisin BGC absent. A small number of strains do not have any part of the nisin BGC (e.g. AP1). The region is flanked by a large conserved polysaccharide BGC and a recombinase gene (indicated by the dashed blue line). Biosynthetic gene clusters were predicted and the graphic extracted from antiSMASH.

Phage predictions in the SAG genomes

Prediction of lysogenic phage in these bacterial genomes was performed with PHASTER (Arndt et al., 2016). It has three classifications based on a completeness score (150 being the highest score) assigned to each identified prophages defined as follow: i) Intact prophage regions are represented with a score of > 90; ii) Questionable prophages are defined by a score between 60 and 90 and ; iii) Incomplete prophages have a score < 60. Based on this classification, PHASTER identified 16 intact, 20 questionable and 115 incomplete prophages within the 151 isolates (Figure 4.6A). Only one intact prophage was predicted in *S. intermedius* and was present in two isolates of the same strain (collected from the same sample). PHASTER estimates similarity by using the number of genes shared between the predicted prophage and the reference prophage as percent of total genes in the predicted prophage. Using this metric, the observed prophage-related open-reading frame (ORF) similarity between the predicted intact prophage within SAG strains and each of their closest reference phages used for identification was below 37% similar (Figure 4.6B), suggesting novel prophages within SAG strains. The sequences alignments between the most common prophages and the potential novel prophages identified in SAG strains support this lack of similarity (Supplementary Figure S4.3).

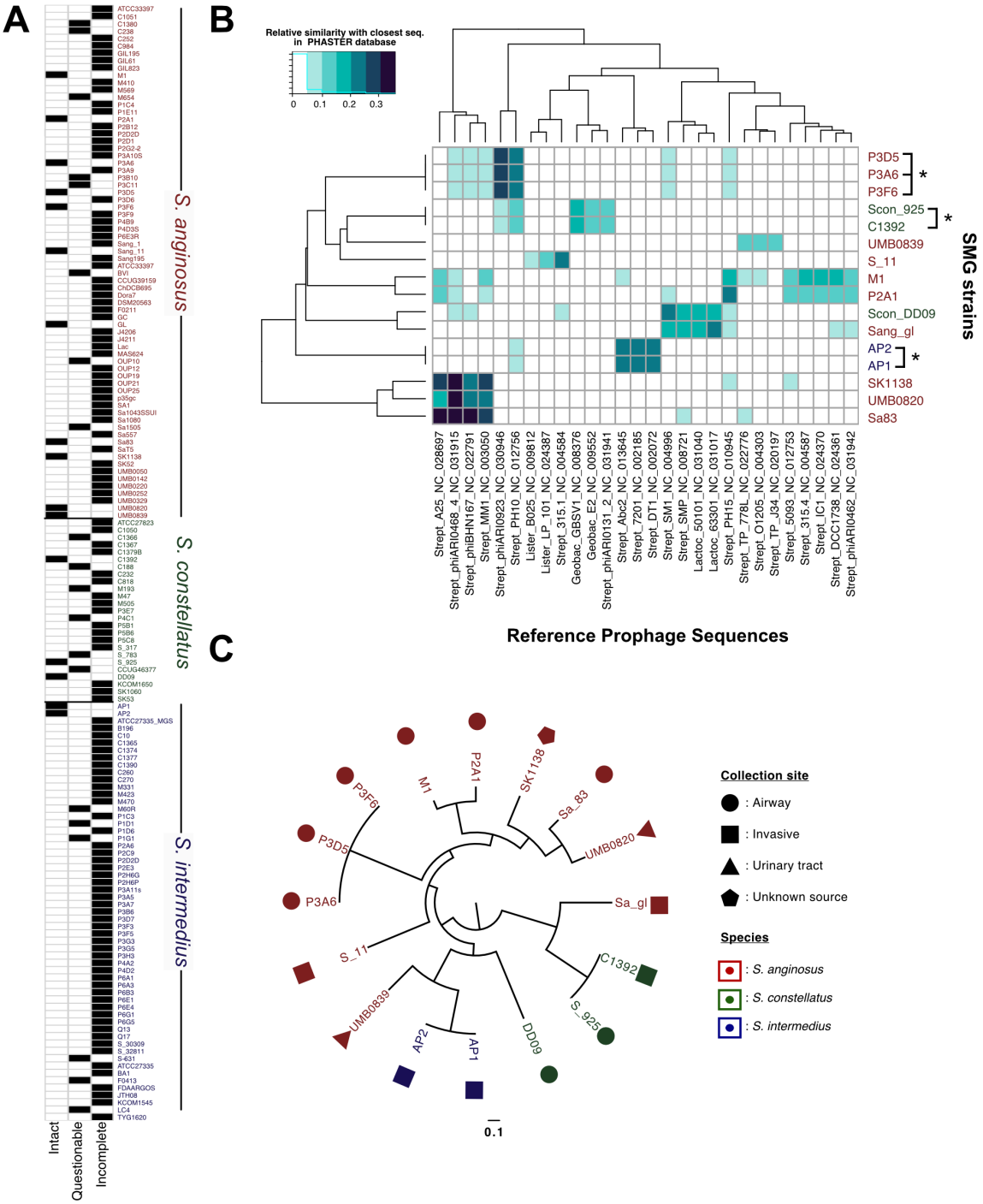


Figure 4.6: Identification of putative prophages within SAG isolates emphasizes the lack of similarity to known phages. PHASTER was used to assess the presence (in black) or the absence (in white) of intact, questionable and incomplete prophages (A). ORF similarity between each intact prophage (n = 16), to closely related available phage genome was low (<37%). Similarity values for the comparisons < 5% similarity are in

white. (B). Similarity was established via PHASTER and was defined by the number of ORFs shared between the reference and the predicted prophage region on the number of total ORFs in the prophage region. The predicted prophage region was estimated with PHASTER. Asterisks indicate SAG strains with identical prophage patterns. Phylogenetic tree of the intact prophages, midpoint rooted (C). Each strain name is coloured by species and each shape represents the isolate's collection site.

Evidence for shared strains between patients

A few isolates share the same core gene phage sequences (strains AP1, AP2; P3D5, P3F6, P3A6; C1392, 925_Scon; Figure 4.6C). AP1 and AP2 are very closely related strains isolated from a single sample from one individual, however, the other strains with similar phages were collected from different patients. The sequences of the predicted prophages in strains P3D5, P3F6 and P3A6 are identical. The genomes of these three strains are similar with only 7 genes different. Variant analysis has identified 40 SNPs divergent between the three strains (Supplementary Table S4.3). The strains were recovered from endotracheal aspirates from patients admitted in the same intensive care unit and from sputum samples collected in an individual with asthma from the same geographical region, raising the possibility of shared strains of SAG within the same geographical region between different individuals.

The predicted prophages from strain C1392 and Scon_925 are highly similar (19 SNVs; Figure 4.7A), and this prompted a closer look at the relatedness of these two genomes, Surprisingly, the two bacterial genomes have an average nucleotide identity of 99.03% with 435 genes different between them (Figure 4.7B). Interestingly, the bacteria were recovered from two patients at different times and different countries. One was recovered in a bronchoalveolar lavage from an ICU patient in the USA between August

2012 and 2013 (Roach et al., 2015), and C1392 was isolated from a brain abscess prior to 2008 in Alberta, Canada.

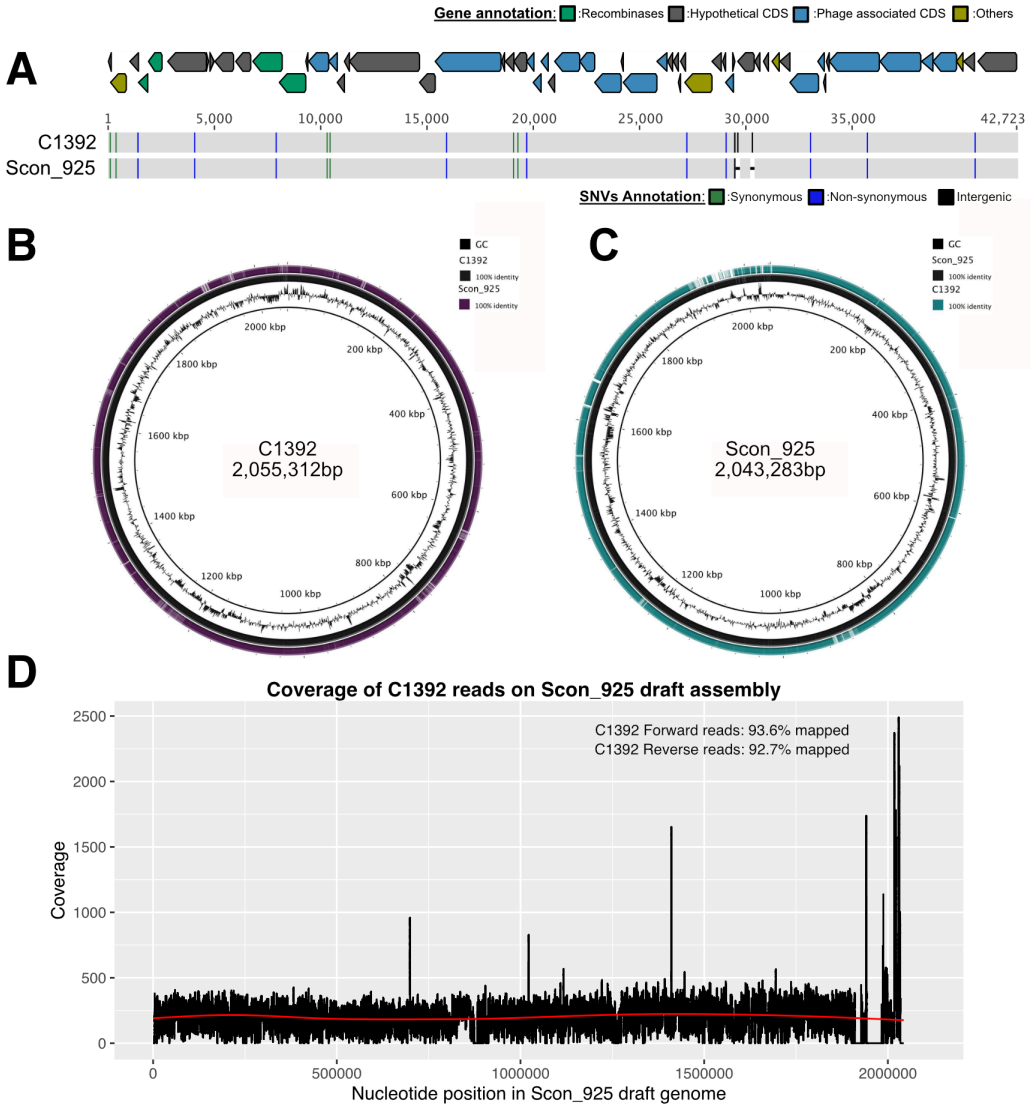


Figure 4.7: Predicted prophages identified between two unrelated strains suggest strain similarity. Nucleotide alignments demonstrated the predicted prophages of strain C1392 and Scon_925 differ by only 19 SNVs, when using C1392 reads on Scon_925 draft assembly (A). These strains are 99.03 % similar to each other (based on the whole genome); B and C show where the small percentage of differences are in the context of C1392 as the reference (B) and Scon_925 as the reference(C). Coverage information per nucleotide of C1392 reads on Scon_925 genome highlights the high genetic similarity between the two isolates (D). The regression line of the coverage information is displayed in red.

Discussion

The *Streptococcus* Anginosus/Milleri Group (SAG) is composed of three distinct species (*anginosus*, *constellatus* and *intermedius*), which are able to colonize mucosal surfaces asymptomatically in humans as well as infect different body sites. The SAG is a clinically underreported group of microorganisms, often not identified or captured by standard culture protocols in the clinical laboratory, and which is frequently dismissed as a commensal during respiratory infection when successfully cultured. Consequently, in-depth epidemiologic, microbiologic and genetic investigations are scarce, whereas multiple case reports are published describing various conditions/illnesses arising due to 'rare' SAG microorganisms. During this investigation, we added 88 new genomic assemblies and carried out comparative genomics with an additional 63 publicly available SAG strains, where 7 of which were previously published by our group (Olson et al., 2013).

The phylogenetic relationship within the three SAG species was investigated by performing a core genome analysis. Our phylogenetic analysis supports the established three species. The core gene phylogeny has demonstrated high heterogeneity between the three species (*intermedius, anginosus* and *constellatus*). Our results confirm that *S. intermedius* is the most genetically divergent of the three species (Jensen et al., 2013; Arinto-Garcia et al., 2015).

The core gene tree distinguished two clades of *S. constellatus* that agree with the assignment of two subspecies (*S. constellatus* subsp. *constellatus* and *pharyngis*) previously suggested (Whiley et al., 1999). However, we fail to detect a distinct third

cluster which might represent the subspecies *Viborgensis* proposed previously (Jensen et al., 2013). Our analysis was not the first to fail to distinguish the subspecies *Viborgensis* (Arinto-Garcia et al., 2015). However, the strains from the subtype *S. constellatus* subsp. *Viborgensis* used in previous studies were not included in our analysis since their whole genome were not sequenced and originate from a defined geographical location (Viborg, Denmark), hence due to these reasons we cannot conclude that this subspecies, if it exists, would not be resolved by core gene SNP analysis. Our analysis suggested that more subgroups within the *S. anginosus* species were present then the two generally accepted in the field (subspecies *whileyi* and *anginosus).* Our analysis supports the groups proposed by Jensen *et al.*, which included the subspecies AJ1 and by Babbar *et al*. whom recently introduced the subspecies *vellorensis,* bringing the number of subspecies to four (Jensen et al., 2013; Babbar et al., 2017). Other studies have suggested the presence of more than three subspecies within *S. anginosus* (Jacobs et al., 2000; Babbar et al., 2017). As demonstrated by Jensen and colleagues (Jensen et al., 2013), the subspecies *whileyi* was more divergent from the other three and *S. anginosus* subspecies *anginosus* was the most heterogeneous subspecies as previously published (Arinto-Garcia et al., 2015).

The SAG is capable of horizontal gene transfer via natural transformation (Grinwis et al., 2010a; Grinwis, 2012; Lacroix, 2014; Bauer et al., 2020). Indeed, the SAG can uptake and integrate extracellular DNA, where competence is induced by the competence signaling peptide via a positive feedback loop (Håvarstein et al., 1996). Our results demonstrate that *S. constellatus* subsp. *pharyngis*, which is not naturally competent in comparison to the other SAG (Lacroix, 2014), have a smaller pangenome.

The reduced number of accessory genes observed in the *S. constellatus* subsp. *pharyngis* isolates relative to the other *S. constellatus* strains suggest that horizontal gene transfer through natural competence is a significant contributor to the pangenome within this species.

Our results further highlight a greater heterogeneity within the resistome profiles, the presence of prophages and secondary metabolite biosynthetic gene clusters. The majority of the strains (90/151) included in this analysis did not possess any of the antibiotic resistance genes included in the Comprehensive Antibiotic Resistance Database (McArthur et al., 2013; Jia et al., 2017). This result is corroborated with *in vitro* results demonstrating a similar prevalence of antibiotic resistance genes among SAG strains (Grinwis et al., 2010b). However, these results are based on *in silico* predictions and despite good general concordances (Su et al., 2018), discrepancies between *in vitro* predictions can exist. Indeed, a recent report has indicated that only 64-71% of the *in silico* predicted resistance profiles can be confirmed *in vitro* using ResFinder and ARG-ANNOT (software similar in function to CARD) (Sabino et al., 2019). Consequently, the predictions of this investigation should be further tested *in vitro*. Moreover, resistance to macrolides due to point mutation within the 23S rRNA gene were not detected with the CARD RGI, despite being included in their database. Interestingly, our group has shown that this ribosomal point mutation was common across SAG isolates (Grinwis, 2012), accordingly, AMR genes have been potentially underestimated. Moreover, SAG strains are intrinsically resistance to sulfonamide antibiotics since they don't possess its target,

dihydropteroate synthase. Due to the absence of the target, this type of resistance to sulfanoamide was not detected via CARD RGI.

The presence of biosynthetic gene clusters was evenly distributed among the three SAG species, except for the BGCs with unknown function and associated to lantipeptide, which are, respectively, decreased and increased in *S. intermedius* in comparison to other species. Draft assemblies generated from *S. intermedius* isolates were in general of better quality compared to the two other species and could explain the decline in the number of predicted BGCs with unknown function in *S. intermedius* isolates. The large number of unknown BGCs could be novel genes but may also be a result of a poorer genome assembly where one BGC get split between contigs inflating the number of predicted BCGs.

We identified a predicted lantipeptide cluster in 39/51 *S. intermedius* isolates, and absent from the other two species, homologous to nisin Q This lantipeptide was confirmed previously in two *S. intermedius* strains (B196 and C270), where it was shown that C270 has a region homologous to the protein involved in the transport of nisin (NisF). No similar hits were found for the novel region in B196 and it was annotated as LanB (Alkhalili and Canbäck, 2018). Moreover, a recent report identified that around 50% of *S. intermedius* isolates carried the nisin immunity protein (Issa et al., 2019). Our results support this 50% prevalence within the *S. intermedius* strains.

The presence of prophages has been predicted in SAG isolates (Olson et al., 2013; Rahman et al., 2015, 2016; Miller-Ensminger et al., 2018; Issa et al., 2019; Brassil et al., 2020) and functional prophages have been confirmed via induction (Brassil et al., 2020).

Moreover, despite the fact that this was not investigated in the current analysis, phage can contribute to virulence (Brüssow et al., 2004; Rezaei Javan et al., 2019; Sweere et al., 2019) and bacterial self-recognition (Song et al., 2019), which will be greatly beneficial for a strain's survival and persistence within the host. A recent report in bioRxiv suggested that a *S. angionosus* phage-like element enhances virulence in the *Galleria* mellonella (wax moth) infection model, however no further evidence was presented in that report (McCullor et al., 2019).

The identification of potential prophages in what we thought were unrelated strains is unexpected and prompted us to investigate the strain relatedness. The high similarity of the whole genome between strains C1392 and Scon_925 suggests a common source. It is, however, surprising since these strains were recovered in different patients and countries. The identification of three very closely related recovered from endotracheal aspirates from two patients admitted in the same intensive care unit and the third from a sputum sample of a severe asthmatic using the same health care facility suggests potential transmission. These results are surprising considering we previously found no evidence for patient to patient spread within adult cystic fibrosis patients attending the same clinic (Sibley et al., 2010b). These findings merit further investigation.

Reports suggest different pathogenic potential between SAG species despite conflicting reports (Whiley et al., 1992; Jacobs et al., 1995; Belko et al., 2002; Siegman-Igra et al., 2012). *S. intermedius* was associated with worse prognosis, with longer hospital length of stay, and higher mortality rates in comparison to the other members of the SAG (Junckerstorff and Murray, 2016). The expression of virulence factors could

explain the species association with a poorer clinical outcome. Indeed, *S. intermedius* expresses a unique human-specific cytolysins from the family of cholesterol-dependent cytolysin, intermedilysin, and numerous hydrolytic enzymes such as sialidases, hyaluronidases (with a chondroitin sulfatase activity) (Facklam, 2002; Grinwis et al., 2010a). Intermedilysin deletion mutants show less adherence, invasion, and cytotoxicity on human hepatic cells (Sukeno et al., 2005) and was highly expressed in isolates recovered from invasive infection (Nagamune et al., 2000). Moreover, there is subspecies specific virulence profiles within the SAG species (Grinwis et al., 2010a; Jensen et al., 2013), for instance the *S. anginosus* subspecies *whileyi* and *S. constellatus* subspecies *viborgensis* included in the study were all β-hemolytic (Jensen et al., 2013). This suggests that even though the SAG forms a distinct phylogenetic group, differences still exist within this group of microorganisms, emphasizing the importance of accurately determining the species while studying this group. However, distinguishing and identifying the SAG isolates to the species level has been challenging in the clinical microbiology laboratory (Parkins et al., 2008; Sibley et al., 2010a). Indeed, commercially available biochemical kits are unreliable (Limia et al., 2000; Arinto-Garcia et al., 2015) due to a recognized intraspecies variability. SAG identification requires biochemical testing involving a minimum of three tests (Grinwis et al., 2010a), or require complex identification methods such as multilocus sequence analysis or MALDI-TOF mass spectrometry (although the last method frequently misidentifies isolates of the *S. intermedius* species). Even though whole-genome sequencing techniques are becoming

more affordable and rapid, their implementation for routine clinical practices for taxonomic identification purposes is not in our reach yet.

This comparative genomics analysis was performed on 151 strains from the SAG, where 88 have not been published previously. To the best of our knowledge, this represent the largest analysis of whole genome data performed on the SAG. Our analysis supports the species classification proposed in the field previously using multilocus sequence analysis and further highlights the strain level heterogeneity within each species. Furthermore, this study has demonstrated the enriched presence of BGCs associated to lantipeptides in the strains of *S. intermedius*. Interestingly, highly similar strains were identified following our prophage investigation. This genetic similarity was not expected since strains were recovered from different patients in diverse studies.

Acknowledgments

We would like to acknowledge Dr. Fiona J. Whelan and Mr. Shahrokh Shekarriz for the helpful and invaluable insights on comparative genomics analysis and for critical review of this manuscript as well as Ms. Laura Rossi, Ms. Michelle Shah, Ms. Christine Chow, Dr. Julie Kaiser, and Dr. Jennifer Lau for obtaining certain SAG genomes. Thank you to the entire Surette lab for the scientific discussions.

Ethics declarations

NA

Supplementary Material

Supplementary Table S4.1: Strains used in the comparative genomic analysis of 151 SAG isolates

Species	Strain ID	Category	Supplementary information	Patient or population	Assembly Level	Accession #	Source	Reference
S. ang	1505	Human-associated			Draft	GCA_000474115.1	NCBI	
S. ang	1_2_62CV	Human-associated			Draft	GCA_000186545.1	NCBI	
S. ang	1043_SSUI	Airway	BAL	ICU patient	Draft	GCA_001069165.1	NCBI	(Roach et al., 2015)
S. ang	1080_SANG	NA	Fluid	ICU patient	Draft	GCA_001070115.1	NCBI	
S. ang	557_SANG	Airway	BAL	ICU patient	Draft	GCA_001076575.1	NCBI	
S. ang	83_SANG	Airway	BAL	ICU patient	Draft	GCA_001074525.1	NCBI	
S. ang	ATCC33397	Human-associated	Unknown	Unknown	Draft	GCA_002088025.1	NCBI	(Whiley and Beighton, 1991)
S. ang	BVI	Urogenital tract	Vaginal fluid (swab)	Female symptomatic	Draft	GCA_900096985.1	NCBI	(Zuñiga-Bahamon et al., 2016)
S. ang	CCUG 39159	Airway		Healthy	Draft	GCA_000257765.1	NCBI	
S. ang	ChDC B695	Airway	Oral	Unknown	Draft	GCA_001510585.1	NCBI	
S. ang	Dora7	GI	Infant stool	Premature male infant	Draft	GCA_000508545.1	NCBI	
S. ang	DSM 20563	Human-associated	Unknown	Unknown	Draft	GCA_000373605.1	NCBI	
S. ang	F0211	Airway	Unknown	Healthy	Draft	GCA_000184365.2	NCBI	
S. ang	gc	Invasive	Dental abscess		Draft	GCA_002999485.1	NCBI	(Fisher and Russell, 1993)
S. ang	gl	Invasive	Dental abscess		Draft	GCA_002999525.1	NCBI	

Species	Strain ID	Category	Supplementary information	Patient or population	Assembly Level	Accession #	Source	Reference
S. ang	J4206	Invasive	Blood	Hospitalized	Complete	GCA_001697145.1	NCBI	(Palacio et al., 2011; Rahman et al., 2016)
S. ang	J4211	Unknown	Unknown	Hospitalized	Complete	GCA_001412635.1	NCBI	(Rahman et al., 2015)
S. ang	LAC	Invasive	Dental abscess		Draft	GCA_002999615.1	NCBI	(Fisher and Russell, 1993)
S. ang	MAS624	Human-associated			Complete	GCA_000478925.1	NCBI	
S. ang	OUP10	Invasive	Dental abscess		Draft	GCA_002999595.1	NCBI	(Fisher and Russell, 1993)
S. ang	OUP12	Invasive	Dental abscess		Draft	GCA_002999655.1	NCBI	
S. ang	OUP19	Invasive	Dental abscess		Draft	GCA_002999635.1	NCBI	
S. ang	OUP21	Invasive	Dental abscess		Draft	GCA_002999555.1	NCBI	
S. ang	OUP25	Invasive	Dental abscess		Draft	GCA_002999575.1	NCBI	
S. ang	p35gc	Invasive	Dental abscess		Draft	GCA_002999515.1	NCBI	
S. ang	S_11	Invasive	Dental abscess		Draft	GCA_002999475.1	NCBI	
S. ang	SA1	Urinary tract	Urine	Quadriplegic woman	Complete	GCA_000831165.1	NCBI	(Srinivasan et al., 2014)
S. ang	SK1138	Human-associated	Unknown	Healthy	Draft	GCA_000287595.1	NCBI	
S. ang	SK52	Human-associated	Unknown	Healthy	Draft	GCA_000214555.2	NCBI	(Whiley and Beighton, 1991)
S. ang	T5	Human-associated	Unknown	Unknown	Draft	GCA_000474155.1	NCBI	
S. ang	UMB0050	Urinary tract	Urine	Asymptomatic female	Draft	GCA_002848125.1	NCBI	(Thomas-White et al., 2018)
S. ang	UMB0142	Urinary tract	Urine	Asymptomatic female	Draft	GCA_002848085.1	NCBI	
S. ang	UMB0220	Urinary tract	Urine	Asymptomatic female	Draft	GCA_002848065.1	NCBI	

Species	Strain ID	Category	Supplementary information	Patient or population	Assembly Level	Accession #	Source	Reference
S. ang	UMB0252	Urinary tract	Urine	Asymptomatic female	Draft	GCA_002860965.1	NCBI	(Thomas-White et al., 2018)
S. ang	UMB0329	Urinary tract	Urine	Asymptomatic female	Draft	GCA_002847945.1	NCBI	
S. ang	UMB0820	Urinary tract	Urine	Asymptomatic female	Draft	GCA_002860925.1	NCBI	
S. ang	UMB0839	Urinary tract	Urine	Asymptomatic female	Draft	GCA_002860945.1	NCBI	
S. con	317_SINT	Airway	BAL	ICU patient	Draft	GCA_001072275.1	NCBI	(Roach et al., 2015)
S. con	783_SANG	Airway	BAL	ICU patient	Draft	GCA_001074375.1	NCBI	
S. con	925_SCON	Airway	BAL	ICU patient	Draft	GCA_001075725.1	NCBI	
S. con	CCUG 46377	NA			Draft	GCA_000474135.1	NCBI	
S. con	DD09	Airway/ Animal	Oral	Chimpanzee	Draft	GCA_001579115.1	NCBI	(Denapaite et al., 2016)
S. con	KCOM1650	Invasive	Oral Cyst fluid		Draft	GCA_000814045.1	NCBI	
S. con	SK1060	Airway	Throat		Draft	GCA_000223295.2	NCBI	
S. con	SK53	Airway			Draft	GCA_000257785.1	NCBI	
S. int	30309	Invasive	Pyogenic fluid from abscess	Hospitalized	Draft	GCA_002879585.1	NCBI	
S. int	32811	Invasive	Pyogenic fluid from abscess		Draft	GCA_002879575.1	NCBI	
S. int	631_SCON	Airway	BAL	ICU patient	Draft	GCA_001073635.1	NCBI	(Roach et al., 2015)
S. int	ATCC27335				Draft	GCA_000413475.1	NCBI	
S. int	BA1	Invasive	Epidural abscess		Draft	GCA_000313655.1	NCBI	(Planet et al., 2013)
S. int	F0413				Draft	GCA_000234035.1	NCBI	
S. int	FDAARGOS_233	Invasive	Abscess	Pediatric	Complete	GCA_002073355.2	NCBI	
S. int	JTH08				Complete	GCA_000306805.1	NCBI	
S. int	KCOM1545	Oral	Endodontic infection		Complete	GCA_001296205.1	NCBI	

Species	Strain ID	Category	Supplementary information	Patient or population	Assembly Level	Accession #	Source	Reference
S. int	LC4	Invasive	Pyogenic fluid from abscess		Draft	GCA_002879755.1	NCBI	
S. int	TYG1620	Invasive	Brain abscess	Pediatric	Complete	GCA_002356055.1	NCBI	(Hasegawa et al., 2017b)
S. ang	195	NA			Draft		Lab	
S. ang	ATCC33397	Airway	Throat	Unknown	Draft		Lab	
S. ang	C1051	Invasive	Blood	Hospitalized	Complete	CP003860	Lab	(Olson et al., 2013)
S. ang	C1380	Invasive	Empyema	CF	Draft		Lab	(Mendonca et al., 2017)
S. ang	C238	Airway	Sputum	CF	Complete	CP003861	Lab	(Olson et al., 2013)
S. ang	C252	Invasive	Unknown	Hospitalized	Draft		Lab	(Mendonca et al., 2017)
S. ang	C984	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. ang	GIL195	GI	Stool	Healthy	Draft		Lab	
S. ang	GIL61	GI	Stool	Healthy	Draft		Lab	
S. ang	GIL823	GI	Stool	Healthy	Draft		Lab	
S. ang	M1	Airway	Sputum	Cystic fibrosis	Draft		Lab	(Mendonca et al., 2017)
S. ang	M410	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. ang	M569	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. ang	M654	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. ang	P1C4	Airway	Sputum	Asthma	Draft		Lab	
S. ang	P1E11	Airway	ETA	ICU	Draft		Lab	
S. ang	P2A1	Airway	ETA	ICU	Draft		Lab	
S. ang	P2B12	Airway	ETA	ICU	Draft		Lab	
S. ang	P2D1	Airway	ETA	ICU	Draft		Lab	
S. ang	P2D2D	Airway	Sputum	Asthma	Draft		Lab	
S. ang	P2G2_2	Airway	ETA	ICU	Draft		Lab	

Species	Strain ID	Category	Supplementary information	Patient or population	Assembly Level	Accession #	Source	Reference
S. ang	P3A10S	Oral	Oral	Healthy	Draft		Lab	
S. ang	P3A6	Airway	Sputum	Asthma	Draft		Lab	
S. ang	P3A9	Urinary tract	Urine	Patient with UTI	Draft		Lab	
S. ang	P3B10	Oral	Oral	Healthy	Draft		Lab	
S. ang	P3C11	Airway	ETA	ICU	Draft		Lab	
S. ang	P3D5	Airway	ETA	ICU	Draft		Lab	
S. ang	P3D6	Airway	Sputum	Asthma	Draft		Lab	
S. ang	P3F6	Airway	ETA	ICU	Draft		Lab	
S. ang	P3F9	Airway	Oropharynx	Pediatric	Draft		Lab	
S. ang	P4B9	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. ang	P4D3S	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. ang	P6E3R	Invasive	Empyema	Hospitalized	Draft		Lab	
S. con	ATCC27823	Invasive	Empyema	Hospitalized	Draft		Lab	(Mendonca et al., 2017)
S. con	C1050	Invasive	Blood	Hospitalized	Complete	CP003859	Lab	(Olson et al., 2013)
S. con	C1366	Invasive	Blood	Hospitalized	Draft		Lab	(Mendonca et al., 2017)
S. con	C1367	Invasive	Blood	Hospitalized	Draft		Lab	
S. con	C1379B	Invasive	Empyema	Hospitalized	Draft		Lab	
S. con	C1392	Invasive	Brain	Hospitalized	Draft		Lab	
S. con	C188	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. con	C232	Airway	Sputum	CF	Complete	CP003800	Lab	(Olson et al., 2013)
S. con	C818	Airway	Sputum	CF	Complete	CP003840	Lab	
S. con	M193	Airway	Sputum	Cystic fibrosis	Draft		Lab	(Mendonca et al., 2017)
S. con	M47	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. con	M505	Airway	Sputum	Cystic fibrosis	Draft		Lab	

Species	Strain ID	Category	Supplementary information	Patient or population	Assembly Level	Accession #	Source	Reference
S. con	P3E7	Airway	ETA	ICU	Draft		Lab	
S. con	P4C1	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. con	P5B1	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. con	P5B6	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. con	P5C8	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. int	AP1	Invasive	Empyema	Pediatric	Draft		Lab	
S. int	AP2	Invasive	Empyema	Pediatric	Draft		Lab	
S. int	ATCC27335				Draft		Lab	(Mendonca et al., 2017)
S. int	B196	Invasive	Hip abscess	CF	Complete	CP003857	Lab	(Olson et al., 2013)
S. int	C10	Airway	Sputum	Cystic fibrosis	Draft		Lab	
S. int	C1365	Invasive	Blood	Hospitalized	Draft		Lab	(Mendonca et al., 2017)
S. int	C1374	Invasive	Blood	Hospitalized	Draft		Lab	
S. int	C1377	Invasive	Blood	Hospitalized	Draft		Lab	
S. int	C1390	Invasive	Brain	Hospitalized	Draft		Lab	
S. int	C260	Invasive	Unknown	Hospitalized	Draft		Lab	
S. int	C270	Airway	Sputum	CF	Complete	CP003858	Lab	(Olson et al., 2013)
S. int	M331	Airway	Sputum	CF	Draft		Lab	(Mendonca et al., 2017)
S. int	M423	Airway	Sputum	CF	Draft		Lab	
S. int	M470	Airway	Sputum	CF	Draft		Lab	
S. int	M60R	Airway	Sputum	CF	Draft		Lab	
S. int	P1C3	Airway	BAL	Healthy	Draft		Lab	
S. int	P1D1	Airway	Sputum	Asthma	Draft		Lab	
S. int	P1D6	Airway	Sputum	Asthma	Draft		Lab	
S. int	P1G1	Airway	Sputum	Asthma	Draft		Lab	

Species	Strain ID	Category	Supplementary information	Patient or population	Assembly Level	Accession #	Source	Reference
S. int	P2A6	Airway	ETA	ICU	Draft		Lab	
S. int	P2C9	Airway	ETA	ICU	Draft		Lab	
S. int	P2E3	Airway	Sputum	Asthma	Draft		Lab	
S. int	P2H6G	Airway	ETA	ICU	Draft		Lab	
S. int	P2H6P	Airway	ETA	ICU	Draft		Lab	
S. int	P3A11s	Airway	BAL	Pediatric	Draft		Lab	
S. int	P3A5	Airway	BAL	Asthma	Draft		Lab	
S. int	P3A7	Airway	Sputum	Asthma	Draft		Lab	
S. int	P3B6	Airway	Sputum	Asthma	Draft		Lab	
S. int	P3D7	Airway	ETA	ICU	Draft		Lab	
S. int	P3F3	Airway	ETA	ICU	Draft		Lab	
S. int	P3F5	Airway	Sputum	Asthma	Draft		Lab	
S. int	P3G3	Airway	ETA	ICU	Draft		Lab	
S. int	P3G5	Airway	Sputum	Asthma	Draft		Lab	
S. int	P3H3	Airway	BAL	Asthma	Draft		Lab	
S. int	P4A2	Airway	Sputum	CF	Draft		Lab	
S. int	P4D2	Airway	Sputum	CF	Draft		Lab	
S. int	P6A1	Invasive	Blood	Hospitalized	Draft		Lab	
S. int	P6A3	Invasive	Empyema	Hospitalized	Draft		Lab	
S. int	P6B3	Invasive	Empyema	Hospitalized	Draft		Lab	
S. int	P6E1	Invasive	Blood	Hospitalized	Draft		Lab	
S. int	P6E4	Invasive	Liver	Hospitalized	Draft		Lab	
S. int	P6G1	Invasive	Blood	Hospitalized	Draft		Lab	
S. int	P6G5	Invasive	Brain	Hospitalized	Draft		Lab	

Species	Strain ID	Category	Supplementary information	Patient or population	Assembly Level	Accession #	Source	Reference
S. int	Q13	Tn library B196			Draft		Lab	
S. int	Q17	Tn library B196			Draft		Lab	

S.int: *S. intermedius*; *S.ang*: *S. anginosus*; *S.con*: *S. constellatus*
ETA: Endotracheal aspirate
BAL: Bronchoalveolar lavage
Tn: Transposon

Supplementary Table S4.2: Taxonomic misidentification using the 16S rRNA predictions

Strains ID	Core gene phylogeny ID	Kraken ID	*ily*	16S rRNA gene length (nt)	Taxonomic identification based on 16S rRNA gene prediction	Identities (%)	# of 16S rRNA gene seq. predicted	Source
BVI	*S. ang*	NA	-	1134	*Corynebacterium tuberculostearicum*	99.1	2	NCBI
C1392	*S. con*	*S. con*	-	1544	*S. intermedius*	99.3	1	In-house
M505	*S. con*	*S. con*	-	1544	*S. intermedius*	99.2	1	In-house
P3D6	*S. ang*	*S. ang*	-	1258	*S. constellatus*	99	1	In-house
P3E7	*S. con*	*S. con*	-	1544	*S. intermedius*	99.3	1	In-house
Sang_CCUG 39159	*S. ang*	NA	-	1200	*Micrococcus luteus*	99.5	1	NCBI
Scon_925	*S. con*	NA	-	1544	*S. intermedius*	99.3	1	NCBI

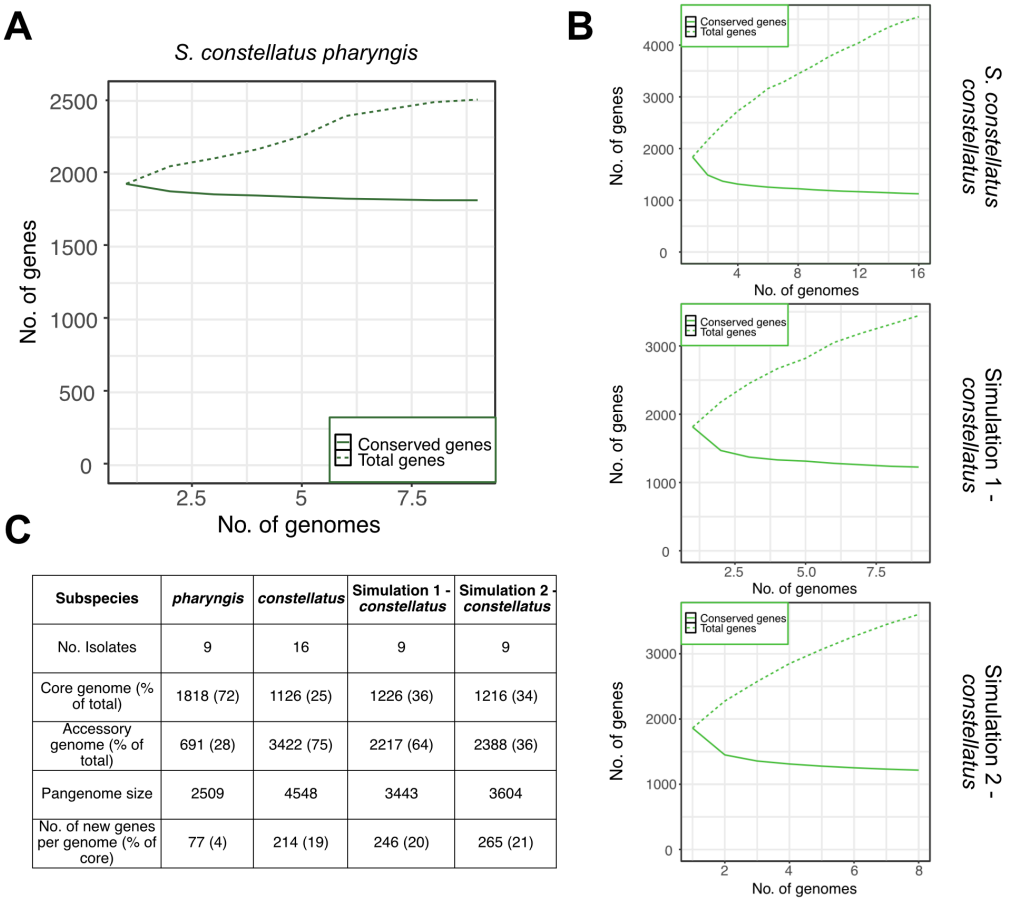
S.int: *S. intermedius*; *S.ang*: *S. anginosus*; *S.con*: *S. constellatus*

Supplementary Table S4.3: SNPs identification and annotation between *S. intermedius* P3A6, P3D5, P3F6

ID	Position in C1051 (bp)	C1051 (ref)	P3A6	P3D5	P3F6	Annotation	Gene name
SNP1	68791	C	C	T	C	upstream	*adhE*
SNP2	68795	T	T	G	T	upstream	*adhE*
SNP3	443966	G	G	G	C	Asn65Lys	*pilD*
SNP4	443968	T	T	T	G	Asn65His	*pilD*
SNP5	443969	G	G	G	A	Cys64Cys	*pilD*
SNP6	443975	A	A	A	C	Arg62Arg	*pilD*
SNP7	443978	G	G	G	A	Cys61Cys	*pilD*
SNP8	443981	A	A	A	G	Arg60Arg	*pilD*
SNP9	637228	T	G	T	T	upstream	SAIN_0617
SNP10	666678	G	G	T	G	Asp312Tyr	*atpD*
SNP11	721111	C	Δ1 bp	C	C	Ala135Asp	*ftsY*
SNP12	721112	T	C	NC	NC	Ala135Ala	*ftsY*
SNP13	815666	T	T	C	C	His40His	*hemN*
SNP14	815670	A	A	C	C	Ile42Leu	*hemN*
SNP15	815675	G	G	A	A	Gln43Gln	*hemN*
SNP16	815678	G	G	A	A	Glu44Glu	*hemN*
SNP17	878486	A	G	A	G	upstream	*pyrF*
SNP18	1002596	C	C	C	A	Val630Val	*topA*
SNP19	1002598	C	C	C	T	Val630Met	*topA*
SNP20	1074552	T	T	T	C	Lys919Lys	SAIN_1030
SNP21	1076064	T	T	C	C	Lys415Lys	SAIN_1030
SNP22	1077404	T	C	T	C	upstream	SAIN_1029
SNP23	1553538	G	G	T	T	upstream	SAIN_1486
SNP24	1553540	T	T	C	C	upstream	SAIN_1486
SNP25	1553542	G	G	T	T	upstream	SAIN_1486
SNP26	1561830	T	C	T	C	upstream	*ctpC*
SNP27	1576721	A	A	A	C	upstream	SAIN_1506
SNP28	1576752	A	A	C	C	upstream	SAIN_1506
SNP29	1576754	A	A	T	T	upstream	SAIN_1506
SNP30	1576760	T	T	C	C	upstream	SAIN_1506

ID	Position in C1051 (bp)	C1051 (ref)	P3A6	P3D5	P3F6	Annotation	Gene name
SNP31	1576761	G	G	C	C	upstream	SAIN_1506
SNP32	1576764	C	C	G	G	upstream	SAIN_1506
SNP33	1576765	A	A	G	G	upstream	SAIN_1506
SNP34	1577446	A	A	G	G	upstream	SAIN_1506
SNP35	1577964	G	G	G	A	upstream	SAIN_1506
SNP36	1577970	T	C	T	C	upstream	SAIN_1506
SNP37	1577973	C	T	C	T	upstream	SAIN_1506
SNP38	1868327	C	C	A	C	upstream	*dltB*
SNP39	1868330	T	T	A	T	upstream	*dltB*
SNP40	1868335	A	G	G	A	upstream	*dltB*

green is synonymous, and blue is nonsynonymous



Subspecies	*pharyngis*	*constellatus*	Simulation 1 - *constellatus*	Simulation 2 - *constellatus*
No. Isolates	9	16	9	9
Core genome (% of total)	1818 (72)	1126 (25)	1226 (36)	1216 (34)
Accessory genome (% of total)	691 (28)	3422 (75)	2217 (64)	2388 (36)
Pangenome size	2509	4548	3443	3604
No. of new genes per genome (% of core)	77 (4)	214 (19)	246 (20)	265 (21)

Figure S4.1: The pangenome of *S. constellatus* subspecies *pharyngis* is more conserved/closed than *S. constellatus* subspecies *constellatus*. Pangenome analysis displaying the number of genes by genome added in the analysis for *S. constellatus* subsp. *pharyngis* (A) and *S. constellatus* subsp. *constellatus* (B). Two simulations with 9 random strains of *S. constellatus* subsp. *constellatus* have been performed to avoid bias. Summary and statistics of the pangenome analysis between *S. constellatus* subsp. *constellatus* and subsp. *pharyngis* (C).

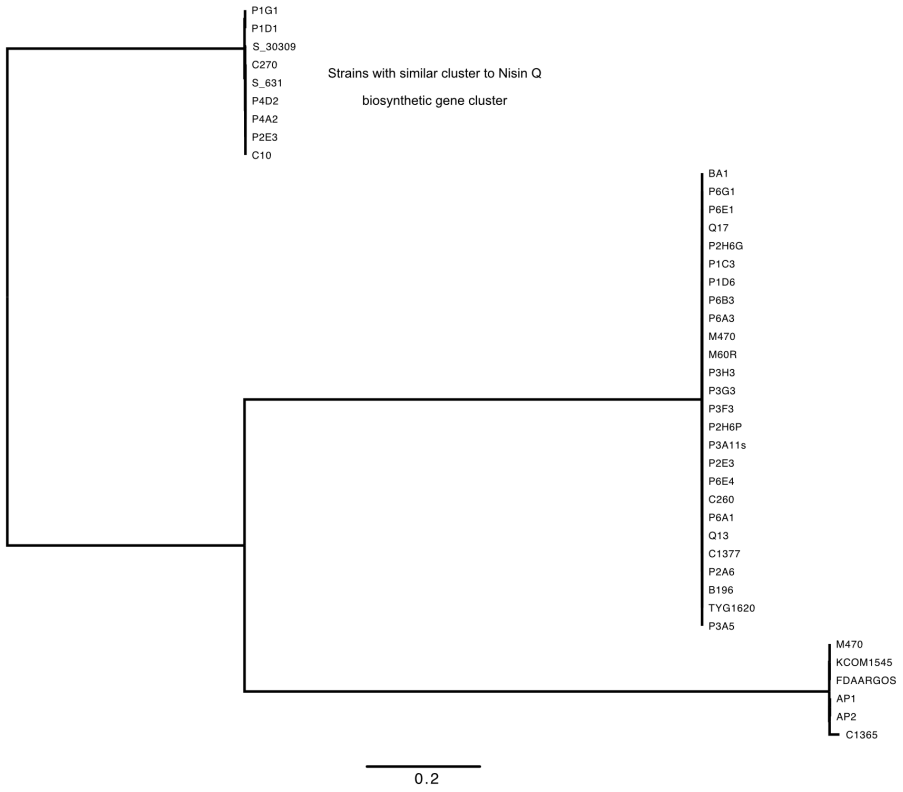


Figure S4.2: Phylogenetic tree inferred using the alignment of NisB-like regions highlights three distinct clades in within the NisB-positive *S. intermedius* strains. Sequences similar to NisB were identified with BLAST and were aligned with MUSCLE and the tree was obtained using Fasttree.

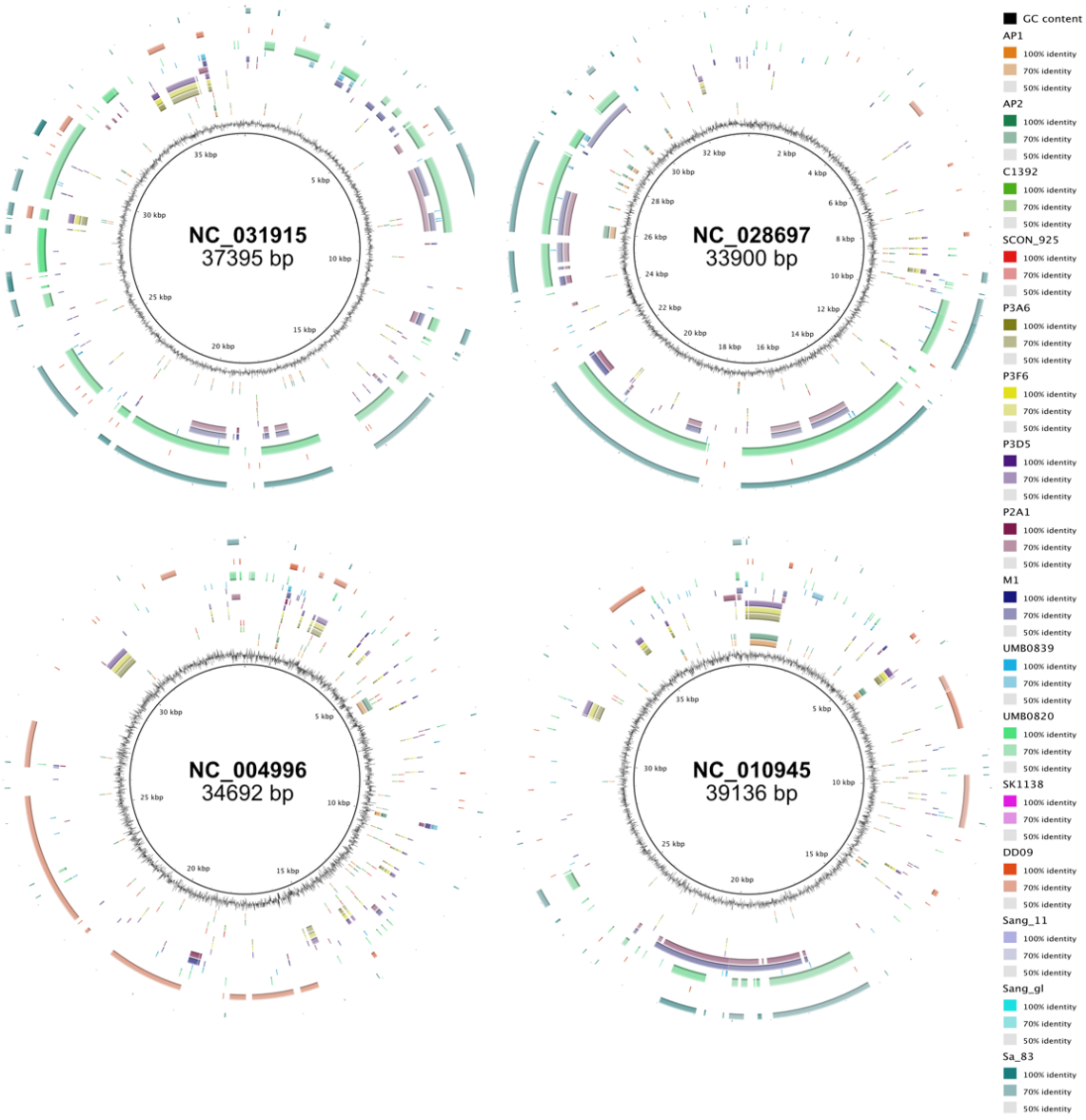


Figure S4.3: Limited genetic similarity between the complete prophages sequences extracted from the SAG strains and their most similar phage sequences available on NCBI. Sequences identified with PHASTER were extracted and aligned with the 4 most common references prophages using BRIG. Each color represents one of the potential 16 intact prophages found within the 151 SAG strains included.

Chapter 5

Characterization of *Streptococcus intermedius* clinical isolates from a pediatric pleural empyema case

Preface

Research presented as part of this chapter has been prepared for the thesis publication.

Lamarche D, Loukov D, Castro K, Howell L, Johnson C, Bowdish D, Surette MG. Characterization of Streptococcus intermedius clinical isolates from a pediatric pleural empyema case (2020). *written for publication*

Contributions:

DL, DLo and KC performed the experiments. DL performed the data analysis. MGS, DB, and DL designed the study. DL and MGS wrote the manuscript. All authors approved the final revision of the manuscript.

Abstract

Acute respiratory infections and their potential complications are a leading cause of morbidity and mortality worldwide. Pleural empyema is a pyogenic infection of the pleural cavity and is a common complication of pneumonia. The incidence of empyema has increased worldwide in the last two decades. In Canada, the most common microorganisms recovered in pleural empyema are the *Streptococcus* Anginosus/Milleri group (SAG) with an incidence between 32 and 52%. The SAG isolates are often dismissed as commensals when recovered from respiratory tract specimens but are common invasive pathogens. This group of microorganisms can transition from commensalism to infection in particular cases, though little is known about their host interactions. It has been proposed that both strain and host variability contribute to the virulence potential of this group of bacteria.

In this study, we have investigated the genetics and host interactions of two phenotypically distinct *S. intermedius* isolates recovered from a pediatric pleural empyema patient. Comparative genomics identified a mutation in exopolysaccharide genes that resulted in the distinct phenotypes. To investigate the immunogenicity of these isolates and whether host response might have contributed to the selection of the mutant strain, we carried out a preliminary cytokine response assay of healthy donors' peripheral blood mononuclear cells (PBMCs). Additional SAG strains were included in the comparison. Variable induction of interleukin-8, in response to strains was observed. However, despite a phenotypical difference between the isolates recovered in the pediatric pleural empyema patient, the cytokine responses were similar and did not show

detectable induction of IL-8 secretion from PBMCs.

Word count: 253

Keywords: *Streptococcus* Milleri*, Streptococcus anginosus, Streptococcus intermedius,* pleural empyema.

Introduction

Respiratory infections are a leading cause of illness and mortality worldwide (World Health Organization, 2018a). Moreover, respiratory infections will continue to be the leading cause of death (World Health Organization, 2018b) due to, among others, a sustained increase in the incidence of pneumonia and pleural empyema (PE). PE is a pyogenic infection of the pleural cavity associated with high morbidity and mortality worldwide, and is a common complication of pneumonia (Wrightson and Davies, 2010). The incidence of empyema has increased in the last two decades in Canada (Finley et al., 2008) as well as worldwide despite prevention strategies (Ahmed et al., 2006; Brims et al., 2010; Grijalva et al., 2011). Common microorganisms recovered from PE are from the *Streptococcus* Anginosus/Milleri group (SAG) with an incidence between 32 and 52% in Canada (Ahmed et al., 2006), around 37% in Japan (Yamazaki et al., 2019), and 25% in Europe (Maskell et al., 2006; Meyer et al., 2011). SAG-induced PEs require surgery more often than other pathogens (Porta et al., 1998; Noguchi et al., 2015; Kobashi et al., 2008), but seem to have better prognoses (83% survival rate at 12 months) in comparison to PE of other bacteriological causes (Maskell et al., 2006). Staphylococcal, enterobacterial and mixed aerobic infections have a 55% survival rate at 12 months (Maskell et al., 2006).

The SAG is composed of three species (*S. anginosus, S. constellatus* and *S. intermedius*) and represents a distinct phylogenetic group within the *Streptococcus* genus; however, the SAG species are phenotypically and genetically heterogeneous (Grinwis et al., 2010a; Jensen et al., 2013; Sibley et al., 2010b). SAG species asymptomatically

colonize various mucosal surfaces (gastrointestinal, urogenital and upper respiratory tract) in about 15-30% of healthy individuals (Poole and Wilson, 1979; Gossling, 1988b). Nevertheless, the SAG is associated with several illnesses such as systemic and deep-seated infections, pneumonia, pleural empyema and pulmonary exacerbation in cystic fibrosis patients (Shinzato and Saito, 1995; Laupland et al., 2006, 2018; Maskell et al., 2006; Parkins et al., 2008; Asam and Spellerberg, 2014; Yamazaki et al., 2019). Members of the SAG are under-reported in respiratory infections and hence their high prevalence in PE represents a paradox (Brims et al., 2010).

The complex interaction between SAG organisms and the host leading to commensalism or infection are likely driven by strain-specific and host-specific responses. When comparing isolates of SAG from respiratory tract to invasive isolates, no differences in phenotypes distinguished the two groups from each other (Grinwis et al., 2010a). However, we demonstrated that invasive SAG isolates are more pro-inflammatory compared to respiratory isolates (Kaiser et al., 2014). Moreover, the PBMC response to specific strains was donor dependent suggesting both strain and host variability contribute to the virulence potential of this group of bacteria (Kaiser et al., 2014). A heterogeneity within the host-response across strains was also demonstrated in *Staphylococcus aureus* and *Streptococcus pyogenes* (Peres et al., 2015; Sela et al., 2018).

Herein, we carried out a case study of a pediatric patient with juvenile idiopathic arthritis (JIA) who developed a pneumonia that progressed to pleural empyema. Two distinct colony morphologies of *S. intermedius* were recovered from the pleural fluid. The genetics and host interactions of the *S. intermedius* isolates recovered from this

pediatric pleural empyema patient were investigated.

Materials and Methods

Subject

A 2-year old female patient with systemic juvenile idiopathic arthritis (JIA) was admitted at her local hospital (Mesa, Arizona, USA) for a bone marrow aspirate. Patients with JIA, an illness characterized by elevated circulating pro-inflammatory cytokines (*e.g.* IL-6, Il-1β, TNF-α), are generally treated using disease-modifying antirheumatic drugs (DMARDs). This patient was receiving the DMARD anakinra (Kineret™), an IL-1β receptor antagonist (ILRA), to manage the JIA.

In addition, two healthy adult female donors (23 and 26 years old) were included to investigate the heterogeneity in the immune response to SAG isolates. We could not match the age of the patient with healthy donors.

All participants or their legal guardians provided written informed consent prior to participation. Healthy donors included in this study were individuals older than 18 years of age without comorbidities who had not experienced fever or infection for 4 weeks before sample collection. This study was approved by the Hamilton Integrated Research Ethics Board.

Microbial communities profiling

The pleural empyema isolates recovered from the pleural fluid recovered from the local clinical microbiology laboratory and provided by Dr. Mochon (Clinical Director,

Infectious Diseases Division Laboratory Sciences of Arizona/Sonora Quest Laboratories, Clinical Assistant Professor, Department of Pathology University of Arizona, College of Medicine). The two distinct SAG isolates were recovered from the original Tryptic soy agar slant when further cultured on agar media (Brain Heart infusing agar, Todd-Hewitt agar supplemented with 0.5% yeast extract, BD). The 16S rRNA gene was sequenced (variable regions V1-V5) as described in (Lau et al., 2016) to confirm their identification as *S. intermedius.*

The bacterial communities from the patient's oropharynx swab post-antibiotics was assessed by 16S rRNA amplicon sequencing as described in (Lamarche et al., 2018). Targeted culturing was performed by plating the specimen on McKay agar media, a semi-selective media for the isolation of the SAG isolates (Sibley et al., 2010a) in anaerobic (5% CO_2, 5% H_2, 90% N_2) and aerobic + 5% CO_2 conditions at 37°C for 24 hours. SAG-specific PCR was performed on each of the colony types (Olson et al., 2010). 16S rRNA gene was sequenced (V1-V5) for all the colony types recovered for the specimen for species identification.

Bacterial strains

Additional *S. intermedius* strains used in the study were isolated from clinical samples from past projects (Grinwis et al., 2010a; Sibley et al., 2010a) (Table 5.1). Bacteria strains were recovered from a glycerol frozen stock on Todd-Hewitt agar supplemented with 0.5% yeast extract (THY, BD) at 37 °C for 48 hours in aerobic conditions + 5% CO_2. Colonies were inoculated in THY broth and incubated for 24 hours in aerobic conditions

+ 5% CO_2. Liquid cultures were homogenized at 1000 rpm for 2 minutes (PowerLyzer™ 24, MoBio), were further centrifuged, and cell pellets were washed and centrifuged three subsequent times and re-suspended in PBS (pH 7). Cells suspensions were heat-killed for 10 min at 75°C. CFUs were enumerated before the heat-killed treatment to standardize the inoculum and after to assess the viability of the bacterial preparation.

Table 5.1: Strains used in this study

Strains ID	Site	Invasive	Ref
B196	Hip	+	(Grinwis et al., 2010a)
C1390	Brain	+	
C1365	Blood	+	
C1377	Blood	+	
M60R	Airway	-	
C270	Airway	-	
M331	Airway	-	
AP1	Empyema	+	This study
AP2	Empyema	+	This study

Antimicrobial disc diffusion assay

Isolates were inoculated from an agar plate into 4 mL of Todd-Hewitt broth supplemented with 0.5% yeast extract (THY, BD) and incubated at 37 °C overnight in aerobic conditions + 5% CO_2. Sensitivity to the antibiotics was assessed as described in (Grinwis et al., 2010b) with minor distinctions. Briefly, liquid cultures were standardized to an OD of 0.1 and plated on THY agar. The following antibiotics disks (Oxoid) were manually laid on the inoculated agar: Ceftriaxone (30µg), cefepime (30µg), tetracycline (30 µg),

azithromycin (15 µg), clindamycin (2 µg), erythromycin (15µg). The zones of inhibition were measured 24 hours after and the measurements were used to assess resistance or sensitivity following the standards of the Clinical laboratory Standards Institute. The experiment was repeated twice to confirm the antibiotic susceptibility profiles.

Comparative genomics

SAG isolates were sequenced on the Illumina platform by the McMaster Genomics Facility (Hamilton, Canada) using standard protocols to generate paired-end reads. Illumina reads were subsequently assembled using the a5 pipeline (Coil et al., 2015). Contigs were ordered via Mauve Contig Mover (Rissman et al., 2009) against the *S. intermedius* B196 reference genome (Accession number: NC_022246.1) and the annotation was performed using RAST tool kit (Brettin et al., 2015) via the PATRIC (Pathosystems Resource Integration Center) web-based interface (Wattam et al., 2014, 2017). PHASTER web interface was used to identify prophages within the draft assembly (Arndt et al., 2016). BRIG (Blast Ring Image Generator, Alikhan et al., 2011) was used to visualize the genomes. Pangenome differences between the two isolates were assessed with Roary (v3.8.2, Page et al., 2015) and variants were identified with Snippy (v4.0-pre_20180729, https://github.com/tseemann/snippy) and were visualized with Geneious (v8.1.9, https://www.geneious.com). Pangenome wide association was assessed with Scoary (v1.6.16; Brynildsrud et al., 2016). Two additional high-quality complete *S. intermedius* genomes were included in the comparative genomics analysis (B196: NC_22246.1 and C270: NC_02237; Olson et al., 2013).

PBMC isolation and activation

Peripheral blood mononuclear cells (PBMCs) were recovered from heparinized blood via density gradient centrifugation using Ficoll-Plaque separation media (GE Healthcare) and Leucosep tubes (Greiner Bio-One) and stored at -120°C in DMSO and heat-inactivated human AB serum (Corning). PBMCs were quickly thawed, washed and seeded to a density of $1x10^5$ cells/wells in 200 µL of RPMI-1640 supplemented with 10% fetal bovine serum (FBS; Gibco), 100 U/ml penicillin G, 100 µg/ml streptomycin and 2mM L-glutamine using round-bottom 96 well plates (Corning). Following an overnight rest period at 37°C in atmospheric conditions supplemented with 5% CO_2, PBMCs were stimulated with the heat-killed bacteria at a multiplicity of infection (MOI) of 1:1, 10 ng/mL of LPS (*Escherichia coli* O55:B5, Sigma-Aldrich) for 24 hours at 37°C in the presence of 5% CO_2. Cells were subsequently centrifuged at 1,500 x g for 5 minutes and supernatants were recovered and stored at -20°C until further use. PBMCs viability following the stimulation was assessed by quantifying lactate dehydrogenase (LDH) released using the CytoTox96 non-radioactive cytotoxicity assay kit (Promega) following the manufacturer's recommendations.

Cytokine quantification

Triplicate reactions were pooled from individual PBMC stimulation experiments and the quantification of cytokines was performed in duplicates. The release IL-8 was quantified in the supernatant using the Deluxe and the Human Il-8 ELISA MAX™ Standard

(BioLegend) following the manufacturer's instructions. Cytokine production was normalized with the LPS-PBMC control to correct for donor and experimental heterogeneity.

Statistical analysis

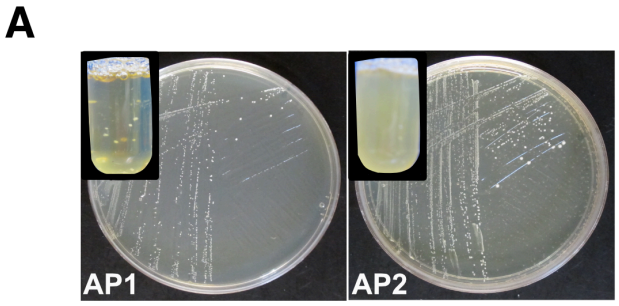
Differences within the cytokine profiles obtained from various bacterial stimuli and donors were assessed with non-parametric statistical analyses. The Benjamini-Hochberg procedure was used for multiple test correction. The significance threshold was set at $p < 0.05$.

Results

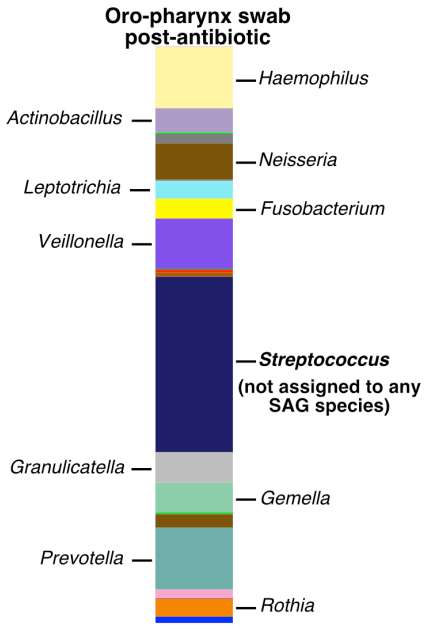
Two phenotypically distinct S. intermedius isolates were recovered from the pleural cavity during the infection

Following clinical standard care procedure at the hospital, the pediatric patient subsequently developed bacterial pneumonia, which then spread to the pleural cavity. *Streptococcus intermedius* was recognized as the causative agent and the patient was treated with antibiotics. While a single species was identified from the pleural fluid of the patient, two different morphotypes were recovered (AP1, AP2). The two isolates are visually distinct from each other on agar and in liquid media (Figure 5.1A). AP1 colonies are small and rough, and the cells aggregate in liquid. In contrast, AP2 colonies are larger and smoother and cells grow uniformly in liquid. The antibiotic sensitivity profiles are identical between the two strains.

Pleural fluid was not available for analysis; however, we obtained an oropharyngeal (OP) specimen (post-antibiotic treatment) and carried out 16S rRNA gene amplicon sequencing and targeted culturing to profile the bacterial community present. *S. intermedius* was not identified from the OP sample taken following antibiotic treatment (Figure 5.1B). None of the *Streptococcus* related operational taxonomic units (OTUs) were assigned to SAG or matched to the specific isolated AP1 and AP2 (Figure 5.1C). Semi-selective culturing on McKay agar, recovered many isolates from the genus *Streptococcus*, but none of the isolates were assigned to any SAG species (Supplementary Table S5.1).



Antibiotic	AP1	AP2
	Results	Results
Azithromycin	Susceptible	Susceptible
Erythromycin	Susceptible	Susceptible
Clindamycin	Susceptible	Susceptible
Ceftriaxone	Susceptible	Susceptible
Cefepime	Susceptible	Susceptible
Tetracycline	**Resistant**	**Resistant**



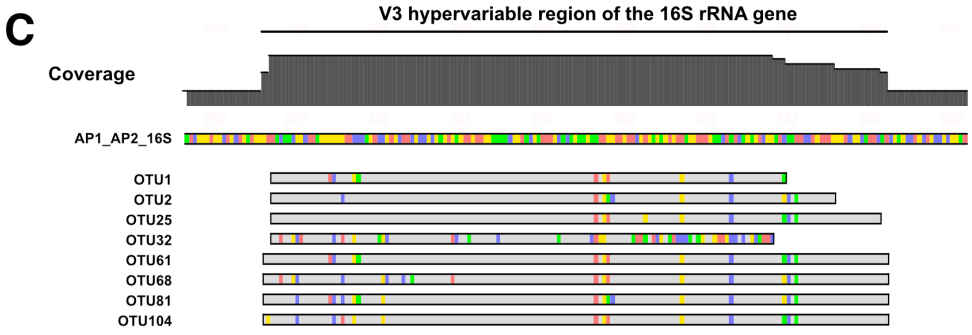


Figure 5.1: Microbiological information about the clinical isolates recovered from the pleural fluid during a pleural empyema. Two isolates identified as *S. intermedius* were recovered from the pleural fluid (A) and shared identical antibiotic sensitivity profiles (B). The microbial profile of an oropharyngeal swab identified specific *Streptococcus* related operational taxonomic units (OTUs) (B), however none of these sequences match *S. intermedius*. The alignment of the OTUs assigned to the *Streptococcus* genus to the 16S rRNA gene of AP1 and AP2 (used as reference; identical) highlights the sequences differences between the streptococcal OTUs and the *S. intermedius* AP1-AP2 16S rRNA gene v3 variable region (C).

High genetic similarity between AP1 and AP2

In order to investigate the difference between both morphotypes, a comparative genomic analysis was performed. The analysis revealed that the genomes of the two morphotypes (AP1 and AP2) recovered from the patient are almost identical according to genome size, GC content, number of coding sequences and number and identity of specialized genes (Figure 5.2, Table 5.2).

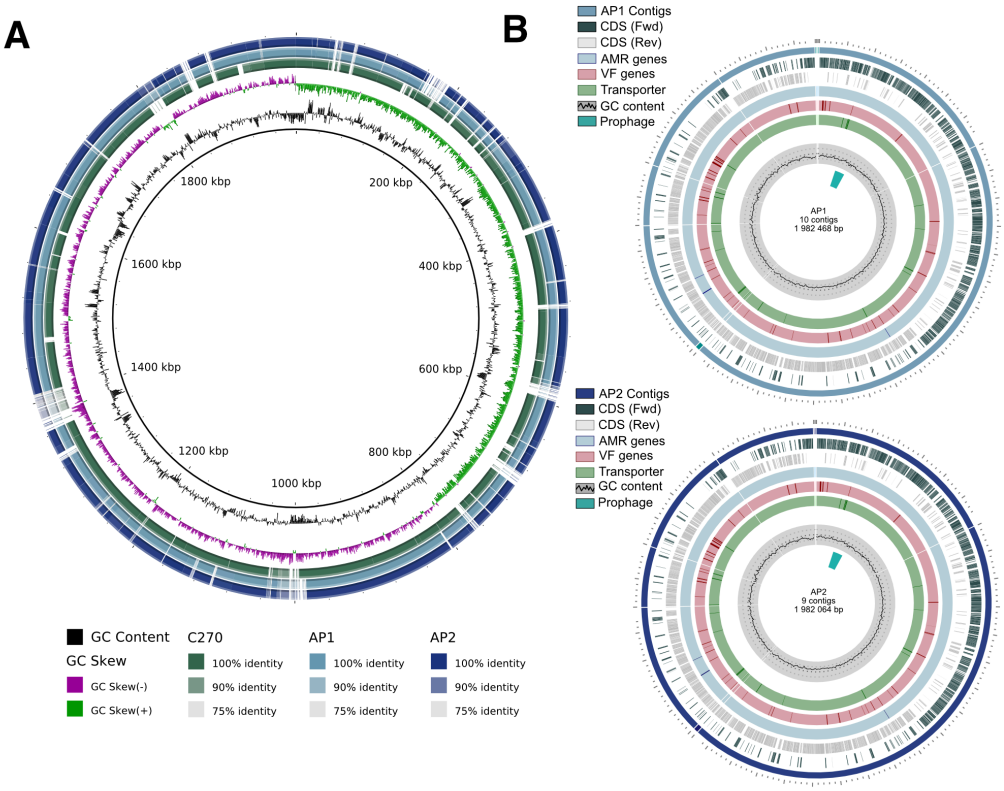


Figure 5.2: Limited genetic diversity observed between strains AP1 and AP2. Genomic comparison between the strains AP1 (teal) and AP2 (marine) with the reference C270 (dark green) versus B196 showed a high similarity between the patient's isolates. Visualization of specialized genes obtained via PATRIC and RAST for AP1 (teal) and AP2 (marine).

Table 5.2: Genomic characteristics of AP1 and AP2 isolates

Characteristics	AP1	AP2
Genome size (bp)	1 982 468	1 982 064
GC content (%)	37.63	37.63
CDS	2018	2015
tRNA	59	55
Hypothetical proteins	490	486
Virulence Factor	2	2
Transporter	33	33
Antibiotic resistance	1	1

CDS: Coding sequences

The genes associated with virulence factors code for a component of a manganese ABC transporter (VFG001359) and for a UTP--glucose-1-phosphate uridylyltransferase (Hyaluronic acid capsule; VFG000964). The isolates harbour a resistance gene for tetracycline, matching their phenotypic data, and both contain an intact prophage. The closest reference for the prophage is the *Streptococcus* phage YMC_2011 (NC_018285) sharing 16/59 (27%) of the prophages related-genes followed by two other *Streptococcus* prophages (*Streptococcus thermophilus* bacteriophage DT1, NC_002072 and *Streptococcus mutans* bacteriophage M102 NC_002185; (Tremblay and Moineau, 1999; van der Ploeg, 2007)) with 15 shared genes. Therefore, there is limited genetic similarities between the prophage identified from the clinical strain (AP1 and AP2) and the three closest references (Figure 5.3) suggesting a novel prophage.

The pan-genome analysis demonstrated a total of 1984 functional genes within both isolates where 1979 of these genes were conserved in both genome assemblies. The two draft genomes differ by only 5 genes that code mostly for hypothetical proteins,

except one gene that was identified to be a nucleoside-diphosphate-sugar epimerase and GAF domain. The last gene is present only in AP2, while the other unique hypothetical four are only present in AP1. The four genes present in AP1 were small and their cumulative size covered roughly the length of the region obtained from the gene unique to AP2.

Phenotypic difference between AP1 and AP2 is due to one SNP in pel operon

Due to the great genetic similarity between the phenotypically different isolates, a variant calling analysis was performed and demonstrated that the core genome of the two isolates differs by only 3 single-nucleotide polymorphisms (SNP): a non-synonymous SNP, an intergenic SNP, and finally a SNP causing the introduction of a premature stop codon (Table 5.3).

Table 5.3: SNPs position between isolates AP1 and AP2

Location	AP1	AP2	Effect
Intergenic (35bp upstream of SIR_0749; homologous to CutC)	C	G	-
SIR_1175 (homologous to *glmU)*	C	T	Non-syn (G304S)
SIR_1599 (renamed *pelH*;homologous to *pelD*)	C	T	+ Stop codon (W369*)

Syn: Synonymous. B196 gene identifiers were used to indicate the location of the SNPs

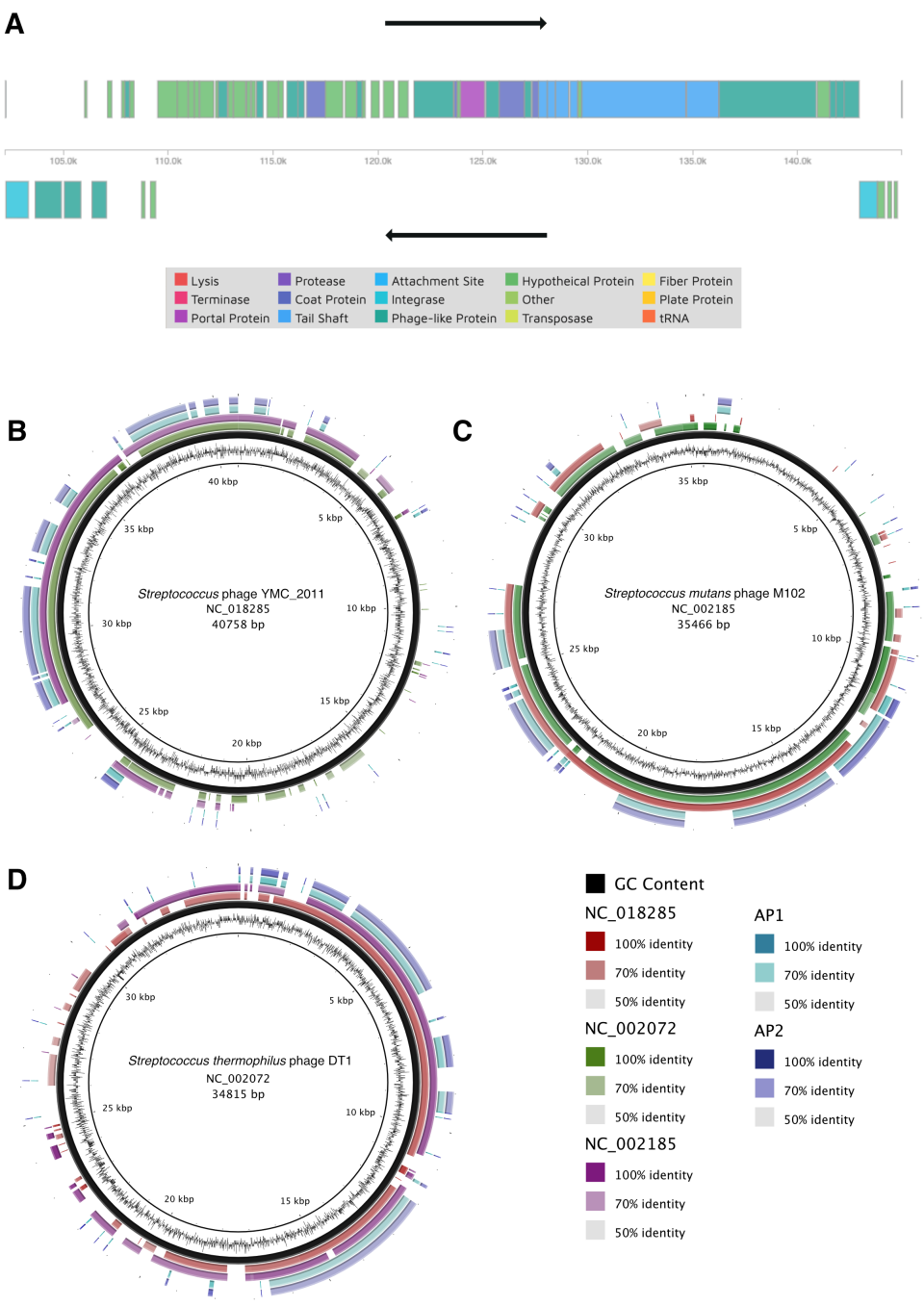


Figure 5.3: Lack of similarities between the closest references prophages sequences and the novel prophage identified in AP1 and AP2. Linear map of the prophage predicted from AP1 and AP2 assemblies (A). Genomic comparison between the prophages identified by PHASTER in AP1 (teal) and AP2 (marine) with the three closest NCBI reference available in PHASTER (B, C, D) suggested the identification of a novel prophage.

The non-synonymous SNP causes a change in the amino acid sequence at position 304 in the region coding for a predicted GlmU, where AP1 has a glycine versus a serine for AP2. GlmU is involved in the biosynthesis of uridine diphosphate N-acetylglucosamine (UDP-GlcNac), a precursor of peptidoglycan and a major component of the cell wall. The other SNP was found 35bp upstream of a gene coding for the copper homeostasis protein (CutC) involved in copper ion binding. Finally, the last SNP leads to the introduction of a premature stop codon in the gene SIR_1599 of AP2 (Figure 5.4AB), a gene homologous to *pelD* in *Pseudomonas aeruginosa*. The *pel* operon is involved in the production of an exopolysaccharide in *P. aeruginosa* (Whitney et al., 2012). Due to its similarity with *pelD,* the gene SIR_1599 was renamed *pelH* (personal communication, Dr. Lynne Howell). Moreover, preliminary analyses have indicated that the SNP located in *pelH* could explain the different phenotype between the two isolates (Castro, Turner, and Howell, unpublished data; Figure 4C). Indeed, deletion of the *pelH* gene prevented the aggregation phenotype observed for AP1.

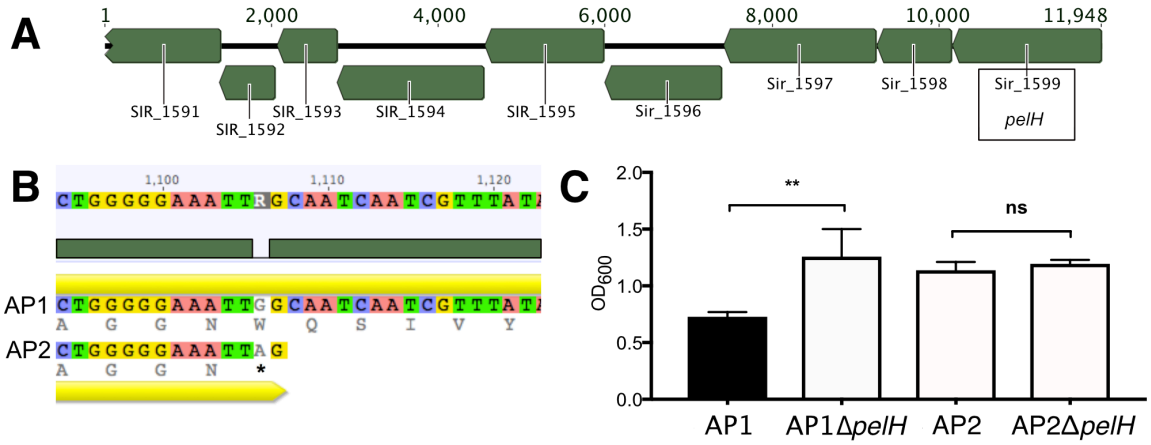


Figure 5.4: Change in morphotype between AP1 and AP2 is due to a SNP introducing a premature stop codon in the *pel* operon. Schematic representation of the *pel* operon in *S. intermedius* (A). SNP in AP2 is introducing a stop codon in *pelH* (B) and *pelH* deletion is leading to the loss of aggregation phenotype in AP1 (C). Significance was determined using a Kruskal-Wallis test.

Profiling the immunogenicity of S. intermedius strains for IL-8 production

In order to assess the immunogenicity of AP1 and AP2, cytokine responses of PBMCs were assessed and compared to an additional panel of 7 *S. intermedius* reference strains (Table 5.1). Results were normalized to show the changes in IL-8 production relative to the LPS positive control rather than the unstimulated condition in order to control for the variability of IL-8 production across donors. Our results demonstrated that there is a significant difference in IL-8 production depending on the *S. intermedius* strain ($p < 0.001$; pairwise comparisons are included in Supplementary Table S5.2.); however, there is no significant difference detected between AP1 and AP2 IL-8 induction ($p = 0.32$) and between AP2 and the unstimulated control ($p = 0.22$) showing no induction of IL-8. IL-8 induction was similar between donors (Supplementary Figure 5.1A; $p = 0.82$). There was

no detectable IL-8 induction by either of these strains. Additionally, the observed variability of IL-8 production across SAG isolates was not explained by cytotoxicity of specific strains, which was homogenous across the strains (Supplementary Figure 5.1). This result was expected since heat-killed bacteria were used for the stimulation, consequently inactivating the potent cytotoxin, intermedilysin.

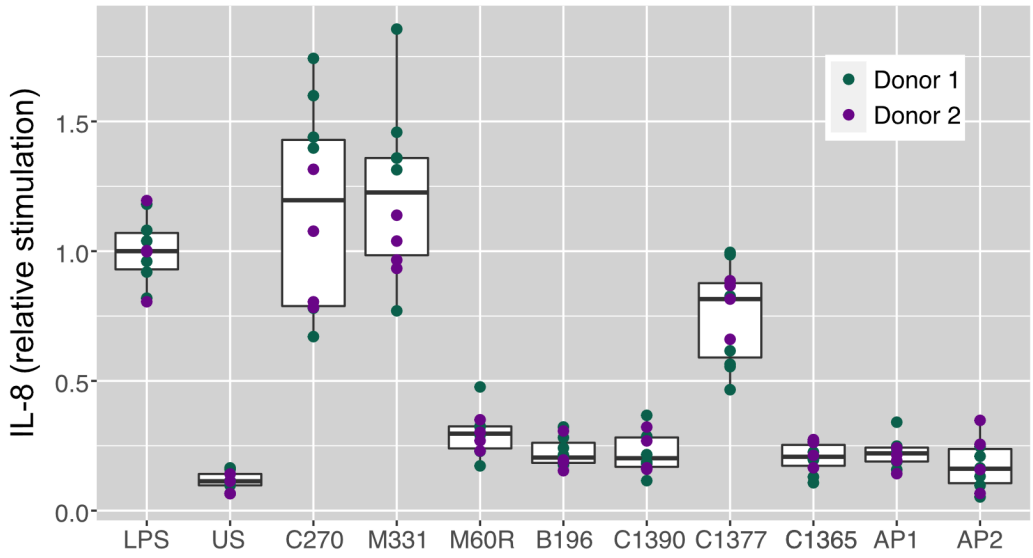


Figure 5.5: Heterogenity of IL-8 secretion between *S.intermedius* isolates. PBMC were stimulated with 9 heat-killed strains of *S. intermedius* at a MOI of 1. Levels of IL-8 were measured in the cell culture supernatant after 24 hours by ELISA. Significant differences were observed for the IL-8 production between strains ($p < 0.001$, Kruskall-Wallis test), but no difference was detected between AP1 and AP2. Results for donor 1 were pooled from 3 independent experiments and results for donor 2 were pooled from 2 independent experiments. Donors are sex-matched and no significant difference between donors was detected ($p = 0.82$, Mann-Whitney test). The median and interquartile range are displayed.

To determine what may be driving the strain-dependant heterogeneity in IL-8 induction, a comparative genomic assessment was performed. The pangenome analysis demonstrated that isolates do not cluster based on their ability to induce IL-8 production

or on the collection sites (Figure 5.6). The computational tool Scoary highlighted 13 genes potentially involved in the strains' capacity to induce IL-8 production, where 11/13 genes were coding for hypothetical proteins (Table 5.4). From the predicted 13 genes, four genes pairs were identified. One strain in the IL-8 – groups seems to be a outlier (M60R), interestingly, this strain has demonstrated variable immunogenicity depending on the host in a previous study from our group (Kaiser et al., 2014). The genes identified lost statistical significance with multiple test corrections; however, this was expected due to the modest number of strains included in the analysis.

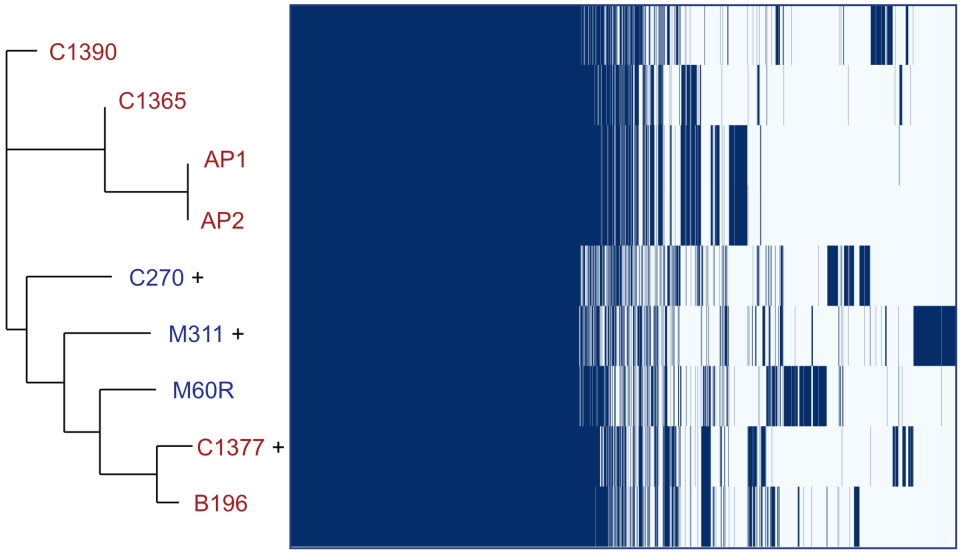


Figure 5.6: *S. intermedius* strains are not clustering by their ability to induce IL-8 or by their collection site. Accessory binary genes tree and heatmap displaying the 9 *S. intermedius* strains used for the PBMC-SAG stimulation demonstrated that isolates did not cluster based on the IL-8 production level or their collection site (*i.e* invasive versus airway). Invasive strains are labelled in red, and strains recovered from the airways are labelled in blue. The (+) represents a strain associated with high IL-8 production.

Table 5.4: Genes associated with IL-8 production predicted identified with Scoary

Gene ID	Annotation	Number (#) of strains with the gene associated to IL-8 production	
		# IL-8 + strains (n = 3)	# IL-8 – strains (n = 6)
group_1074	hypothetical protein	0	5 (absent in M60R)
group_604	hypothetical protein	0	5 (absent in M60R)
group_1077	hypothetical protein	0	5 (absent in M60R)
group_325	hypothetical protein	3	1 (Present in B196)
lppC	Putative lipoprotein LppC	0	5 (absent in M60R)
group_555	hypothetical protein	0	5 (absent in M60R)
group_603	hypothetical protein	0	5 (absent in M60R)
group_602	hypothetical protein	0	5 (absent in M60R)
group_1249	hypothetical protein	3	1 (B196)
group_1248	hypothetical protein	3	1 (B196)
yafQ	mRNA interferase YafQ	0	5 (absent in M60R)
group_308	hypothetical protein	0	5 (absent in M60R)
group_309	hypothetical protein	0	5 (absent in M60R)

IL-8 +: C270, C1377, M331 IL-8 -: AP1, AP2, C1365, C1390 M60R, B196

Discussion

The *Streptococcus* Anginosus/Milleri Group (SAG) is a genetically and phenotypically heterogenous group of microorganisms asymptomatically inhabiting various mucosal surfaces. However, the SAG has the pathogenic potential to cause a diverse range of infections in select but yet undefined conditions. Despite its prevalence in the clinic, the SAG is still an underreported group of microorganisms in respiratory and invasive infections. The goal of this study was to characterize two clinical isolates recovered from

a pleural cavity infection in a pediatric patient, to explore the immunogenicity of SAG clinical isolates in healthy volunteers.

Here we characterized *S. intermedius* recovered from a pediatric patient (2 years old of age, female). The patient developed a pulmonary infection, which exacerbated to a pleural empyema. Interestingly, two distinct morphotypes, both identified as *Streptococcus intermedius*, were recovered from the pleural fluid. One morphotype displayed smaller and rougher colonies on agar and was aggregating in liquid culture (AP1), while the second (AP2) displayed a more uniform growth in liquid culture and was bigger and smoother on solid media. Our results suggest that the phenotypic difference within the morphotype was due to a single nucleotide polymorphism (SNP) in the *pelH* gene, leading to the introduction of a premature stop codon in the sequence of the strain AP2 and the loss of the aggregating and biofilm phenotype (Turner, Castro, Howell *et al,* Unpublished*)*. The *pelH* gene in *S. intermedius* shares sequence homology with *pelD* from the *pelABCDEFG* operon in *Pseudomonas aeruginosa*, where it is involved in the regulation of PEL polysaccharide production, an essential component for biofilm production in certain strains of *P. aeruginosa* (Whitney et al., 2012).
Studies have reported different virulence potential between rough and smooth morphotypes (Jordan, 1926; Jiménez de Bagüés et al., 2004; Howard et al., 2006; Shea et al., 2011; Mancilla, 2016) in multiple bacterial species. It has been demonstrated that aggregation is generally beneficial for bacteria by providing increased resistance to external stress such as desiccation, antibiotics, and predation (to both other microorganisms and the immune system) (Trunk et al., 2018; Grinberg et al., 2019).

Moreover, despite being more easily detected by components of the immune system and potentially mounting a greater immune response, aggregating bacterial cells are more challenging to kill (Galdiero et al., 1988; Süßmuth et al., 2000). Unpublished work from our group has demonstrated that *S. intermedius* B196 induces a host-dependent robust activation of NF-κB, but is resistant to killing by macrophages (Stearns and Pelka, unpublished). Moreover, beneficial impacts of bacterial aggregation are more pronounced for larger aggregates; where larger aggregates are more resistant to phagocytosis (Galdiero et al., 1988). It is likely that aggregation also leads to more tolerance to environmental stressors (Monier and Lindow, 2003).

While the aggregating behaviour may be beneficial for the bacteria, the presence of the non-aggregating AP2 morphotype within the pleural fluid is interesting. It is established that changes within the environment are a major determinant of genetic diversity. Consequently, the initial transition of the bacterial strain from the respiratory tract (initial site of infection) to the pleural cavity (where the isolates were recovered) or the dynamic microenvironment-induced empyema could have selected a specific bacterial phenotype potentially more adapted to its new niche. Indeed, phenotypic diversification could facilitate the establishment of the microorganism in a novel environment (Sheppard et al., 2018). The microenvironment of a pleural cavity infection is continuously evolving from the first signs of parapneumonic effusion to a pleural empyema. Briefly, during the pleural effusion exacerbation, pulmonary interstitial fluid accumulates in the pleural space and this exacerbation is associated with microbiological and biochemical changes (i.*e.* the microenvironment progressively become more acidic (until pH < 7.2) and

microorganisms tend to be recovered from the pleural cavity) (Light, 2006). It has been demonstrated that the SAG isolates were thriving in typical conditions associated with pleural empyemas, such as low pH and partial pressure of oxygen (Maskell et al., 2006); however rapid biochemical changes and a highly inflamed microenvironment could have led to these genotypic and phenotypic changes. The inflamed environment is filled with protein-rich edema, consequently bacteria cells are potentially more metabolically active and the requirement of producing an exopolysaccharide matrix is hypothetically not necessary.

While this study primarily assessed the role of PBMCs, other aspects of the immune response to SAG could control the host's susceptibility to infection. Preliminary results obtained in collaboration with Dr. Dawn Bowdish (McMaster University, Canada) have suggested that AP2 seems to induce greater IgG and IgA binding than AP1 and could suggest that exopolysaccharide production could help to evade the immune system initially, further supporting the translocation of the strain to the pleural cavity.

Variable production of cytokines from different isolates of SAG has previously been observed (Kaiser et al., 2014), however, IL-8 has not been investigated previously. The comparative genomic analysis failed to significantly detect genes involved in the production of IL-8. Induction of the immune response by the bacterium is dependent on a myriad of factors. Predominantly, it is initiated when pattern recognition receptors (PRRs) activation sense conserved microbial motifs (teichoic acid, peptidoglycan, lipopolysaccharides in gram-negative bacteria, flagellin, etc.). The difference in cytokine production between SAG strains could be due to differential expression of components of

the cell wall or the presence of a capsule in particular strains which could lead to differential induction of PRRs and TLRs. Indeed, certain cell wall components could protect the bacteria from the immune system (Peres et al., 2015; Gerlach et al., 2018). While the exact mechanism for this response across SAG isolates is unknown, *Staphylococcus aureus* strains also demonstrate variability in host immune responses as altered expression of pro- and anti-inflammatory cell wall components (Peres et al., 2015). Furthermore, *S. aureus* and *S. pyogenes* also induce strain-specific responses within the adaptive immune system (Sela et al., 2018). Nonetheless, hosts may adapt to the presence of certain strains, as a recent study identified a heightened immune response to exogenous *S. aureus* strains as opposed to the strain they were colonized with (Reiss-Mandel et al., 2018). This study suggests that the host itself may influence the outcome of host-SAG interactions (elimination, commensalism or disease). In this study, we were not able to comment on the host-strains specificity since the patient's PBMC were not assessed. However, we compared IL-8 induction of both isolates (API and AP2) with cells collected in healthy volunteers and we demonstrated that both strains elicit similar IL-8 response, and potentially similar neutrophil recruitment. In only investigating PBMC responses, we cannot comment of the neutrophil-mediated defence. It is possible that neutrophils have a harder time dealing with bacterial aggregates during an infection (Rakita et al., 1999; Trunk et al., 2018), however this warrants further investigation. Pleural empyema is often a complication of pneumonia where the etiological agent may not be identified. Standard clinical microbiology for lower respiratory tract infections will not routinely recover the SAG even if it is present (Sibley et al., 2010a). Various studies

have demonstrated that despite current diagnostic tests, an etiological agent is only determined in about 50% of the cases for respiratory infections (Capelastegui et al., 2012; Psallidas et al., 2014; Jain et al., 2015). Even though this remains controversial, we suspect that SAG isolates are underestimated as the causative agents identified during respiratory infections such as pneumonia. Indeed, work in our laboratory using sputum from cystic fibrosis patients has demonstrated that the current protocol in the clinical laboratory in the hospital setting is not optimal for the recovery as well as the quantification of SAG isolates (Sibley et al., 2010). This has been observed (but not yet published) in smaller cohorts of non-cystic fibrosis patients with acute respiratory infections. In a cohort of patients with suspected ventilator-associated pneumonia, we found that SAG species were recovered as the dominant organism in 15/45 patients suggesting that SAG isolates can be cultured under the proper conditions from lower respiratory tract samples during respiratory infections (unpublished data). A study performed in Japan has demonstrated that SAG isolates were causing more pneumonia than *S. pneumoniae* (Shinzato and Saito, 1995) and However, the clinical absence of SAG in pneumonia could also potentially suggest a preference for closed-space infections such as abscesses and infections of the pleural cavity.

The lack of causative agent could be explained by a myriad of reasons starting with the administration of an effective antibiotic treatment prior to sample collection. SAG isolates are not particularly resistant to antibiotics and could have been killed. Moreover, the absence of culture under anaerobic conditions, which is implemented for

invasive infections such as pleural empyema, could also explain the absence of causative agents especially the one driving in these conditions.

The determination of an etiology is important for patient well-being. Appropriate treatment will decrease potential complications such as the exacerbations of pneumonia to infection of the pleural cavity as well as decreasing the use of ineffective antimicrobials which could lead to an increase prevalence of AMR genes. A study has already demonstrated that about 30% of the antibiotic administered were not necessary (Fleming-Dutra et al., 2016). During this AMR crisis, we need to decrease the use of unnecessary antimicrobial therapy.

This study contains several limitations. The lack of investigation using the patient's PBMC was the main limitation. This experiment would have allowed us to comment on the immunogenicity of the strains in their host and assess how immune modifying drugs influenced cytokine responses to SAG isolates to further comment on the patient's susceptibility to SAG infections. In addition, we did not recruit age-matched pediatric healthy volunteers. Several studies have demonstrated that age is a factor influencing the cytokine profiles (ter Horst et al., 2016; Decker et al., 2017; Rea et al., 2018), hence, the use of our cohort of healthy adult donors is not suitable for comparison to the JIA subject. Moreover, the pleural fluid as well as a pre-antibiotic OP swab were not available to examine the microbial communities. Consequently, we were not able to determine if the *S. intermedius* isolates were present in the upper respiratory tract of the patient.

The results obtained from this case study suggest that a unique SNP in an exopolysaccharide gene (*pelH*) was responsible for the phenotypic differences observed in the aggregation behavior. This changes in phenotype between the two isolates recovered in the pleural empyema from the same patient did not drive different cytokine response, however this phenotype could impact other aspects of the immune response which could control the host's susceptibility to infection.

Acknowledgments

Thank to Dr. Kyle Novakowski, Dr. Brian Mochon, and Ms. Sue Colvin for organizing the shipment of the samples to McMaster University and to Dr. L. Patrick Schenck for the tremendous scientific contributions for the data analysis and for the critical reading of the manuscript. Finally, thanks to all members of the Surette laboratory for the scientific discussions.

Ethics declarations

This study was approved by the Hamilton Integrated Research Ethic Board and was performed in accordance with the principles of Good Clinical Practice following the tri-council guidelines. All participants or their legal guardian provided written informed consent prior to enrollment.

Supplementary Material

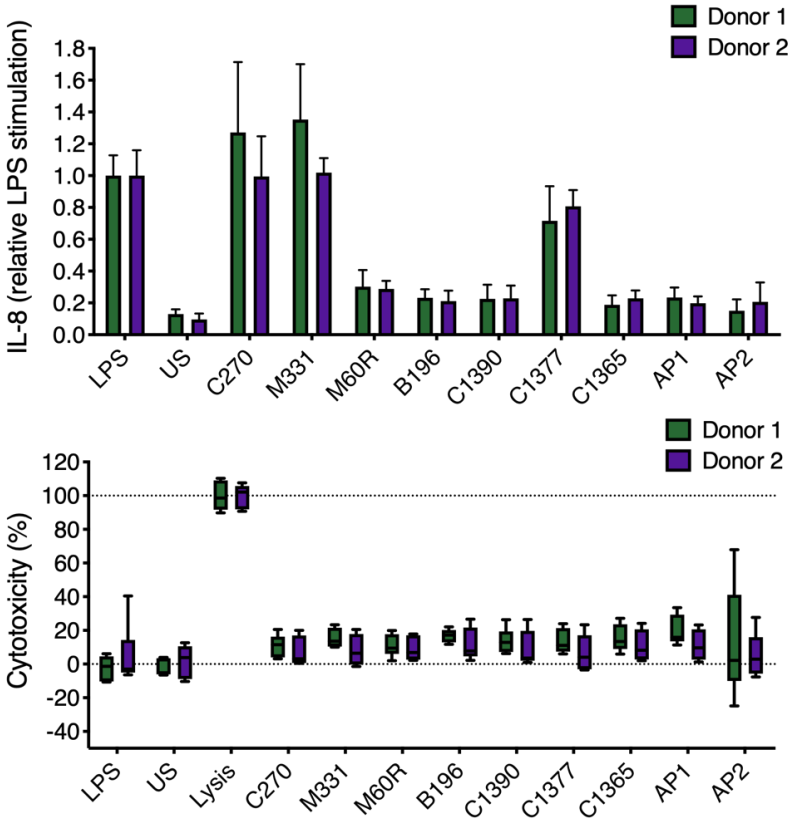
Supplementary table S5.1: Identity and abundances of the bacterial isolates recovered from the patient oropharyngeal swab on McKay agar.

Isolate	Culture conditions (Environment, dilution)	Abundance (CFUs/mL)	Taxonomic identification	Identity (%)
CJ1-A1	AE, 10^{-1}	6.00E+03	*Streptococcus vestibularis*	99.2
CJ1-A2	AE, 10^{-1}	5.00E+02	*Granulicatella adiacens*	99.2
CJ1-A3	AE, 10^{-1}	2.50E+04	*Gemella haemolysans*	96.5
CJ1-A4	AE, 10^{-1}	100	*Gemella haemolysans*	99.5
CJ1-B1	AE, 10^{-3}	2.00E+05	*Streptococcus mitis/oralis*	99.8
CJ1-B2	AE, 10^{-3}	3.00E+04	*Streptococcus vestibularis*	99.4
CJ1-B3	AE, 10^{-3}	4.00E+04	*Streptococcus vestibularis*	99.2
CJ1-B4	AE, 10^{-3}	1.00E+03	*Gemella haemolysans*	99.3
CJ1-C1	ANA, 10^{-4}	4.00E+06	*Gemella haemolysans*	99.4
CJ1-C2	ANA, 10^{-4}	2.00E+05	*Streptococcus dentisani*	99.3
CJ1-C3	ANA, 10^{-3}	3.00E+04	*Streptococcus vestibularis*	99.6
CJ1-C4	ANA, 10^{-3}	5.00E+04	*Gemella haemolysans*	99.5
CJ1-C6	ANA, 10^{-3}	8.00E+04	*Streptococcus parasanguinis*	99.6
CJ1-C7	ANA, 10^{-3}	3.00E+04	*Gemella haemolysans*	99.7
CJ1-C8	ANA, 10^{-1}	2.50E+04	*Streptococcus vestibularis*	99.6
CJ1-C9	ANA, Swab	NA	*Gemella haemolysans*	99.3

CFU: Colony-forming unit

Supplementary Table S5.2: Comparison of IL-8 production differences between each PBMC stimulation conditions (adjusted p value displayed following the Benjamini-Hochberg procedure).

	LPS	US	C270	M331	M60R	B196	C1390	C1377	C1365	AP1
US	4.10E-05	-	-	-	-	-	-	-	-	-
C270	0.72248	4.10E-05	-	-	-	-	-	-	-	-
M331	0.16108	4.10E-05	0.89584	-	-	-	-	-	-	-
M60R	2.60E-05	4.10E-05	2.60E-05	2.60E-05	-	-	-	-	-	-
B196	2.60E-05	8.70E-05	2.60E-05	2.60E-05	0.05908	-	-	-	-	-
C1390	2.60E-05	0.00045	2.60E-05	2.60E-05	0.08453	0.93468	-	-	-	-
C1377	0.00574	2.60E-05	0.04157	0.00131	2.60E-05	2.60E-05	2.60E-05	-	-	-
C1365	2.60E-05	0.0034	2.60E-05	2.60E-05	0.00794	0.73397	0.75263	2.60E-05	-	-
AP1	2.60E-05	0.00045	2.60E-05	2.60E-05	0.04157	0.89584	1	2.60E-05	0.89584	-
AP2	2.60E-05	0.22283	2.60E-05	2.60E-05	0.01017	0.21669	0.18309	2.60E-05	0.3275	0.3275



Supplementary Figure S5.1: The observed variability in IL-8 production within the *S. intermedius* species is independent of cytotoxicity. PBMC were stimulated with 9 heat-killed strains of *S. intermedius* at a MOI of 1. Levels of IL-8 were measured in the cell culture supernatant after 24 hours by ELISA. Results are coming from 3 independent experiments using donor 1 and 2 independent experiments using donor 2. The median and interquartile range are displayed. No observed variability within *S.intermedius* strains cytoxicity (B). PBMC viability was determined via LDH detection following 24 hours stimulation and normalized to a 100% lysis control

Chapter 6

Conclusions

In this thesis, I present the work I have performed while investigating the heterogeneity associated with respiratory infections and related frequent complications (pleural empyema).

As part of the PROSPECT randomized controlled trial investigating the effect of probiotic administration on the prevalence of hospital-acquired infections, I have surveyed the composition of the host-associated communities of mechanically ventilated critically ill patients at various anatomical sites (upper and lower GI and lower respiratory tracts) using 16S rRNA gene sequencing (**Chapter 2**). We reported a novel association between the microbial diversity of the lower respiratory tract and hospital mortality in this population of patients. In addition to this, I have performed a quality assessment to investigate the study product (*Lactobacillus rhamnosus* GG and placebo composed of microcrystalline cellulose) administered during the PROSPECT trial to ensure the patients received a dose and a product consistent with the manufacturer's stated content. I have performed a microbiology analysis such as strain-typing to further characterize potential adverse events associated with the administration of probiotics. To the best of our knowledge, this is the first quality assessment of this magnitude nested directly in the larger study infrastructure of a clinical trial (**Chapter 3**). This study confirms that the probiotic dose was stable in time but also at a higher temperature (2-weeks up to 30°C), and that placebo contained none or modest level of microorganism. I also investigated, the genetic diversity of the *Streptococcus* Anginosus/Milleri group (SAG), a potential underrecognized group of microorganisms during pneumonia but prevalent in pleural cavity infection (common complication related to pneumonia) and other pyogenic

infections. We reported that the SAG isolates formed a genetically heterogeneous group with respect to the presence of overall and specialized genes (AMR, BGC) and the presence of prophages (**Chapter 4**). I have also characterized two phenotypically divergent isolates of *Streptococcus intermedius* recovered from the pleural cavity during a pleural empyema. We demonstrated that the phenotype difference was due to the presence of a single SNP introducing a stop codon into a gene involved in exopolysaccharide production (**Chapter 5**).

Taken together, the results gathered in my doctoral thesis emphasize that the heterogeneity associated with respiratory infections contributes to the challenge of identifying but also treating respiratory infections such as pneumonia. Two hypotheses were proposed for each arm of my project and were tested.

First hypothesis: Investigations into the PROSPECT study

The first hypothesis was that the baseline bacterial composition (early after ICU admission) of the microbiomes of mechanically ventilated patients (lower respiratory tract, gastric and stool) is correlated with clinical outcomes. Together with the co-authors of that study, we presented a novel association between microbial diversity of the lower-respiratory tract microbiota and patient outcomes in the critical care population, where lower α-diversity (*i.e.* diversity observed within specimen) is associated with hospital mortality. However, this association was not observed using gastric aspirates specimens and was not tested with stool samples due to the limited number of specimens recovered. Indeed, stool specimens are sporadically available in the ICU since the GI physiology is

perturbed during critical care. To prevent this limitation, reports have suggested the use of a more easily accessible type of specimen, a rectal swab. Even though compositional differences were observed between stool and rectal swab (Bansal et al., 2018; Fair et al., 2019), the latter was able to detect microbial perturbations similarly to stool specimens (Bansal et al., 2018). Using rectal swabs in our study could have prevented this shortage of specimens and increased greatly our analysis power.

Moreover, we have demonstrated that the biogeography between anatomical sites was lost during critical illness. However, we did not have the required number of specimens to further test this compositional homogenization within and across patients. Our number of specimens was limited since we focus only on samples collected early in the ICU to prevent or limit the potential analysis bias associated with probiotic administration. However, having samples from multiples sites collected in the same patient will be a strength for the larger mechanistic sub-study, once we are unblinded to the probiotic or placebo administration. Reports using culture-independent methods (16S rRNA gene sequencing) have indicated that gut-associated bacteria were enriched in the LRT during critical illness and that their presence was further associated with worst prognosis for the patients (Dickson et al., 2016, 2020). These reports suggest potential causal link between colonization of the lower airway with intestinal microbes and clinical outcomes; however, this requires further scrutiny.

We encountered major challenges when analyzing the microbial composition of critically ill patients using culture-independent approaches (*e.g.* 16S rRNA gene sequencing, qPCR) including low biomass, presence of PCR inhibitors, and host cells and

DNA as well as the prevalence of fungi in the specimen. Accordingly, negative controls (extraction, PCR, and sequencing,) are required to distinguish a signal associated with contamination from low biomass specimens. Indeed, common taxa associated with extraction kit contamination can be human-associated microorganisms (*e.g.* *Lachnospiraceae*, *Lactobacillus* spp., *Propionibacterium* spp.) (Glassing et al., 2016). The majority of ICU patients are receiving antibiotics during their stay which could impact the bacterial biomass at various sites as well as introducing amplification bias due to the detection of DNA from dead bacterial cells in the microbial profiles. For instance, 16S microbial profiles enriched with GI-associated microorganisms could be obtained for certain clinical specimens even in absence of any growth using cultured-enriched method, and those were excluded from subsequent analysis (data not shown). Further, to limit the amplification of DNA from dead bacterial cells following antibiotic treatment, the use of a relatively novel method based on DNA sequencing and the use of propidium monoazide (PMA) has been proposed. PMA can covalently bind free DNA following photoactivation. Once DNA is linked with PMA, PCR amplification will be inhibited and hence DNA from dead bacteria will not be present in the microbial profiles of the clinical samples (Exterkate et al., 2015). This technique could be promising but could be prone to overestimate the proportion of living cells within the samples and required the dilution of samples, which could be problematic with low biomass samples and could lead to the loss of very rare bacterial taxa (Papanicolas et al., 2019).

High levels of inflammatory mediators are observed during critical illness leading to the infiltration of immune cells into the LRT. The presence of host cells and their DNA can be detrimental for the PCR amplification, sequencing and downstream analysis (Rossi L, personal communication, (Lundberg et al., 2013)). A protocol using acellular supernatant obtained by centrifugation attempts to reduce host cells background in bronchoalveolar lavage samples (Dickson et al., 2014b). However, adding this extra processing step during the extraction protocol significantly reduced bacterial signal (52% of the samples were positive for a bacterial signal using the acellular supernatant versus 96% using the whole BAL) and induced changes within the composition and diversity of the microbiota (Dickson et al., 2014b).

Finally, PCR inhibitors present in the sample such as organic matters (Schrader et al., 2012) can interfere with the PCR amplification and affect downstream analysis, such as quantification of bacterial DNA. For instance, qPCR and quantitative culturing are generally in agreement to estimate bacterial load (Stearns et al., 2015). In our dataset for respiratory specimens, amplification with qPCR was complicated by the presence of inhibitors and was not consistent across samples. To limit the impact of PCR inhibitors or bias due to host DNA, a relatively novel technology called droplet digital PCR can be used. This technique divides the samples in thousands of fractions and the amplifications are carried within this smaller pool (Taylor et al., 2017). Despite the fact that Droplet digital PCR is a promising technology, its implementation is not common due to its cost.

Even though 16S rRNA gene sequencing was instrumental to the microbiome field expansion in the last decades, this technique is associated with a few limitations. The

main one is its low taxonomic resolution which rarely goes beyond the genus level. Some hypervariable regions provide sufficient taxonomic information to differentiate phylogenetically close organisms, but often 16S rRNA gene sequencing is not able to differentiate genera of specific groups such as the *Lachnospiraceae* family. Moreover, other molecular approaches such as metagenomics are now available. However, despite the fact that metagenomics is a powerful method, it is highly constrained by the abundant presence of host DNA in samples collected from the LRT of critically ill patients and potentially not suitable for this cohort of patients. In addition, conclusions obtained from a microbiome analysis are highly susceptible to the methodology and data processing pipeline used in absence of a gold standard methodology. Indeed, biases associated with the extraction method used, PCR amplification, library preparation are potentially introduced with this approach. Accordingly, to ensure complete transparency, reproducibility and generalizability within the field, scientists need to provide sufficient information on their approaches and need to make their analysis pipeline clear and understandable as well as providing their code for the data analysis. This could facilitate the mutual sharing of knowledge as well as supporting standardized methodologies within this diverse field. Finally, molecular approaches are not able to discriminate between live and dead material, potentially overestimating the bacterial load in the specimens as well as introducing a potential analysis bias.

Implementation of culturing for investigating the composition of the microbiota of ICU patients could help overcome these limitations and could increase our confidence in the data obtained with molecular approaches. Our group has shown the potential of a

cultured-enriched approach in the GI and respiratory tracts (Sibley et al., 2011; Lau et al., 2016; Whelan et al., 2020). They demonstrated that having multiples conditions (different media and oxygen requirements) increased the recovery of bacterial taxa in the clinical specimens and demonstrated that the majority of the lower respiratory tract and gut microbiota are in fact culturable (Sibley et al., 2011; Lau et al., 2016; Whelan et al., 2020). This technique allows to achieve a more genuine representation of the microbial communities inhabiting the studied environment and allows the recovery and further preservation of the bacterial strains. Moreover, having a strain library could allow us to perform additional mechanistic investigations to build on the data obtained from molecular approaches. For instance, isolating the bacterial strains from the critically ill patients' samples open a whole new type of analysis where we could investigate the virulence potential or the functionality of the strains using phenotypic assay and simple host models, or investigate the prevalence and evolution of common AMR genes in this population. Moreover, a database including clinical information about patient's history, medications and infections is available for the patients enrolled in the PROSPECT study. An analysis comparing direct sequencing 16S profiles to cultured-enriched 16S profiles (Sibley et al., 2011; Lau et al., 2016; Whelan et al., 2020) could be also interesting to pursue. In the course of the mechanistic study, 96 samples were cultured using this approach (Appendix B) and a preliminary strain library of 380 isolates has been built and are available for future studies.

Second hypothesis: Investigations into the Streptococcus Anginosus/Milleri group

The second hypothesis for my thesis was that comparative genomics of the SAG may define genetic variability associated with phenotypic heterogeneity and pathogenic properties of individual strains. However, I was not able to fully investigate this hypothesis and my analysis was limited to the characterization of the genetic variability observed across SAG isolates. I observed that within the three SAG species, *S. anginosus* displays the greatest intra-species heterogeneity. Moreover, I have demonstrated that the accessory genome of the SAG is characterized by the presence of numerous phages and a large repertoire of uncharacterized biosynthetic gene clusters (primarily bacteriocin-like). This genetic heterogeneity is consistent with the known phenotypic heterogeneity previously reported.

This expanded database of SAG genomes should facilitate future research and improve diagnostic of this group of underappreciated respiratory pathogens. Indeed, SAG isolates are still considered harmless members of the upper respiratory, GI and urogenital tract microbiota, despite several reports demonstrating the SAG pathogenic potential in multiple type of infections including life-threatening infections (Issa et al., 2020). This is made even worse by the limitations of current clinical microbiology laboratory, which are not optimal for SAG recovery (Sibley et al., 2010a; Junckerstorff and Murray, 2016). Accordingly, the clinical presence and significance of the SAG are underestimated due to its phenotypic variability, complicating its identification, and its prevalence in healthy individuals. Recent data about the asymptomatic carriage of SAG isolates are missing in the field. Indeed, their prevalence in the general 'healthy' population is based on older

studies (Poole and Wilson, 1979; Gossling, 1988a) with less advanced identification methods, with limited number of participants, and no information about the time period for which strains are carried.

The complex interaction between SAG organisms and the host leading to asymptomatic carriage or infection is not defined yet but is likely driven by strain-specific and host-specific responses. Currently, our comparative genomics analysis (**Chapter 4**) has omitted an in-depth investigation into SAG pathogenicity. Such an analysis should be further explored in the future using comprehensive curated databases such as the virulence factors database (Chen et al., 2016). However, our group has failed to distinguish between strains recovered from the airways and strains recovered from an invasive source, using phenotypic assays including virulence-associated phenotypes and biochemical profiling (Grinwis et al., 2010a). More recently, our group showed that cytokine profiles can be used to differentiate respiratory tract from invasive strains since the latter are inducing a more pro-inflammatory response (Kaiser et al., 2014). Furthermore, the immune response provoked by a particular strain could vary between individuals (Kaiser et al., 2014). Surprisingly, I have demonstrated that the two morphotypes recovered from a pleural empyema did not induce the production of IL-8, a potent pro-inflammatory cytokine, when using peripheral blood mononuclear cells (**Chapter 5**), but we did not investigate if the morphotypes had different virulence potential.

Individual hosts may adapt to the presence of certain strains, shown by a recent study that identified a heightened immune response to exogenous *S. aureus* strains as

opposed to the strain they were colonized with (Reiss-Mandel et al., 2018). This suggests that the host itself may influence the transition between the commensal and pathogenic lifestyles displayed by SAG in the general population. Consequently, a mutual relationship could exist between the host and microbe in which specific strains may promote a tolerogenic immune response that facilitates their colonization while others are perceived as pathogens. An investigation of the immune response induced by a host endogenous strains using matched host cells should be performed to further explore this current hypothesis. In addition to this potential tolerogenic immune response specific to host and strains, individual cytokine profiles are dependent of host genetics, age and gender, but also environmental factors, seasons cycling and composition of host-associated microbial communities (Li et al., 2016a, 2016b; Schirmer et al., 2016; ter Horst et al., 2016). All these factors could partly explain individual differential susceptibility to illnesses and immune-related conditions.

Conclusions

The investigations within this thesis contributed to our understanding of the bacterial heterogeneity observed during respiratory infections. Several of the conclusions made within my thesis are based on molecular approaches, a powerful method for the detection of genomic features and to investigate the association between clinical outcomes and the presence of bacterial taxa. This work was limited by several factors explained above.

Despite this, I believed that the work performed in this report has set the foundation for a better understanding of the microbial dynamics of mechanically

ventilated critically ill patients and the potential factors contributing to poor outcomes in the ICU patients. The observed decreased in microbial diversity of the lower respiratory tract sample is not the cause of the exacerbated illness/outcomes but rather a proxy of the general health of the patients, such as classification methods commonly used in the ICU based on a pathophysiologic scoring system (*e.g.* APACHE II) attributed early during their stay in the ICU. A recent report has demonstrated that the microbial diversity observed in the lower respiratory tract was a better prognosis tool to predict mortality in sepsis patients than APACHE II, SOFA, and SAPS II scores (Sepsis Lung Microbiome Study Group, 2020). The results taken together could suggest that microbial diversity, estimated using microbiome analysis, could be used as a prognosis tool to further stratify patients based on outcomes and this early in the ICU.

The loss of biogeography in ICU patients needs to be further investigated in future studies. We hypothesize that the degree of this microbial homogenization could be potentially associated with clinical outcomes and used as a proxy for illness severity in this population of patients. This lack of microbial consensus and the further homogenization between body sites could be due to underlying illnesses and conditions as well as clinical interventions. In the last decades, our understanding of the beneficial impacts associated with the microbiota has considerably improved. This novel knowledge of the importance of the microbiota in conjunction with the observation that few key taxa are depleted during critical illness and that there is a loss of separation between anatomical sites in critically ill patients suggest that certain clinical interventions provided in the ICU need to be re-evaluated. For instance, the use of acid suppressant

therapy, which has a profound impact on microbial disturbances in ICU patients, is now under investigation (Alhazzani et al., 2016). The epidemiology of critically ill patients has changed since the prophylactic use of PPI has been proposed. Indeed, GI bleeding is now rare and the necessity and benefit of PPI administration to mitigate bleeding is potentially lower than the risk of subsequent infections such as *C. difficile* infection and pneumonia. Moreover, nutritional goals are rarely meet during critical illness and more research is needed in this field more specifically on the impact of diet on the microbiome and its association with SCFA production and clinical outcomes. Furthermore, in order to limit the administration of antibiotic due to the patient's health and public health concerns, studies are ongoing to investigate the optimal timing for the administration of antibiotics but also to prevent the overuse of treatment (Denny et al., 2020; Wunderink et al., 2020). This could prevent inappropriate treatment and consequently decrease the chances of selection of AMR genes in this cohort of patients, where ~70% received treating and prophylactic antibiotics currently.

Several therapeutic options targeting the microbiota to prevent complications associated with critical care (*i.e.* nosocomial infections), such as the PROSPECT study, are currently investigated. These therapies could be cost-effective supplements to the current treatment administered during critical illness to prevent detrimental complications and mortality. Even though the investigation into the impact of probiotics administrations on the microbial communities in critically ill (via the PROSPECT mechanistic sub-study) has not been possible in this thesis since the study is still blinded, we did, however, improve our understanding of the microbial perturbances in this highly heterogeneous

population of patients. This knowledge will be required to identify potential mechanisms of action and comment on the effectiveness of those strategies. Although probiotics are considered safe in the general population, there is potential for risk in vulnerable populations such as the critically ill. I investigated *Lactobacillus* associated adverse events in the PROSPECT study (*e.g.* positive blood culture). I demonstrated that three instances of positive *Lactobacillus* blood culture were not due to the study product.

The *Streptococcus* Anginosus/Milleri group represents under-appreciated pathogens of the respiratory tract. I undertook a comparative genomics study that highlights the species and strain heterogeneity in this group that has contributed to the challenges in routine identification. This work could provide a foundation for improved diagnostics. My analysis of 151 strains from the SAG demonstrated the heterogeneity observed within the accessory genomes and has identified relevant specialized genes. The biological relevance of those genomic features (such as prophages, antimicrobial resistance genes, and biosynthetic genes clusters) still has yet to be determined. However, we build on previous reports investigating this underreported etiology in respiratory infections and emphasized that the SAG should be further investigated due to its pathogenic potential.

Routine application of microbiome and comparative genomics concepts to the field of critical illness and respiratory infections especially its integration in the health care-associated practices is becoming more feasible. These approaches could improve the detection of potential etiological agents of infections potentially missed by current clinical microbiology laboratory protocols and could help identify high-risk population

within the intensive care units, and all of this taken together could be beneficial for the patient health. Currently, its implementation is not practical partly due to technological limitations and due to the delay associated with sample processing (nucleic acid extraction, library preparation, sequencing, downstream analysis). Moreover, even though, a microbial signature associated with poor outcomes is identified, we still ignore how to infer this novel data in the patient's therapeutic plan to limit detrimental repercussions. Accordingly, more fundamental studies into the mechanistic impact of microbial perturbations and on improving microorganism's detection in a timely fashion are required to meaningfully integrate next-generations sequencing to the health-care system.

Bibliography

Ahmed, R. A., Marrie, T. J., and Huang, J. Q. (2006). Thoracic Empyema in Patients with Community-Acquired Pneumonia. *Am. J. Med.* 119, 877–883.

Alhazzani, W., Guyatt, G., Marshall, J. C., Hall, R., Muscedere, J., Lauzier, F., et al. (2016). Re-evaluating the Inhibition of Stress Erosions (REVISE): a protocol for pilot randomized controlled trial. *Ann. Saudi Med.* 36, 427–433.

Ali, T., and Harty, R. F. (2009). Stress-Induced Ulcer Bleeding in Critically Ill Patients. *Gastroenterol. Clin. North Am.* 38, 245–265.

Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12, 402. doi:10.1186/1471-2164-12-402.

Alkhalili, R. N., and Canbäck, B. (2018). Identification of Putative Novel Class-I Lanthipeptides in Firmicutes: A Combinatorial *In Silico* Analysis Approach Performed on Genome Sequenced Bacteria and a Close Inspection of Z-Geobacillin Lanthipeptide Biosynthesis Gene Cluster of the Thermophilic Geobacillus sp. Strain ZGt-1. *Int. J. Mol. Sci.* 19, 2650.

Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

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

Angelakis, E., Million, M., Henry, M., and Raoult, D. (2011). Rapid and Accurate Bacterial Identification in Probiotics and Yoghurts by MALDI-TOF Mass Spectrometry. *J. Food Sci.* 76, M568–M572.

Arbique, J. C., Poyart, C., Trieu-Cuot, P., Quesne, G., Carvalho, M. d. G. S., Steigerwalt, A. G., et al. (2004). Accuracy of Phenotypic and Genotypic Testing for Identification of *Streptococcus pneumoniae* and Description of *Streptococcus pseudopneumoniae* sp. nov. *J. Clin. Microbiol.* 42, 4686–4696.

Arinto-Garcia, R., Pinho, M. D., Carriço, J. A., Melo-Cristino, J., and Ramirez, M. (2015). Comparing Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry and Phenotypic and Molecular Methods for Identification of Species within the *Streptococcus* anginosus Group. *J. Clin. Microbiol.* 53, 3580–3588.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–21.

Arpaia, N., Campbell, C., Fan, X., Dikiy, S., van der Veeken, J., deRoos, P., et al. (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* 504, 451–455. doi:10.1038/nature12726.

Asam, D., Mauerer, S., and Spellerberg, B. (2015). Streptolysin S of *Streptococcus anginosus* exhibits broad-range hemolytic activity. *Med. Microbiol. Immunol.* 204, 227–237.

Asam, D., and Spellerberg, B. (2014). Molecular pathogenicity of *Streptococcus anginosus*. *Mol. Oral Microbiol.* 29, 145–155. doi:10.1111/omi.12056.

Austrian R (1971). "Of Man and pneumococcus, an historical paradox.," in *Of microbes and life* (New York and London), 165–73.

Baatjes, A. J., Smith, S. G., Watson, R., Howie, K., Murphy, D., Larché, M., et al. (2015). T regulatory cell phenotypes in peripheral blood and bronchoalveolar lavage from non-asthmatic and asthmatic subjects. *Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol.* 45, 1654–1662.

Babbar, A., Kumar, V. N., Bergmann, R., Barrantes, I., Pieper, D. H., Itzek, A., et al. (2017). Members of a new subgroup of *Streptococcus anginosus* harbor virulence related genes previously observed in *Streptococcus pyogenes*. *Int. J. Med. Microbiol. IJMM* 307, 174–181.

Babrowski, T., Holbrook, C., Moss, J., Gottlieb, L., Valuckaite, V., Zaborin, A., et al. (2012). *Pseudomonas aeruginosa* virulence expression is directly activated by morphine and is capable of causing lethal gut derived sepsis in mice during chronic morphine administration. *Ann. Surg.* 255, 386–393.

Babrowski, T., Romanowski, K., Fink, D., Kim, M., Gopalakrishnan, V., Zaborina, O., et al. (2013). The intestinal environment of surgical injury transforms *Pseudomonas aeruginosa* into a discrete hypervirulent morphotype capable of causing lethal peritonitis. *Surgery* 153, 36–43.

Baker, G. R. (2004). The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *Can. Med. Assoc. J.* 170, 1678–1686.

Ball, L. C., and Parker, M. T. (1979). The cultural and biochemical characters of *Streptococcus* milleri strains isolated from human sources. *J. Hyg. (Lond.)* 82, 63–78.

Banerjee, S., Sindberg, G., Wang, F., Meng, J., Sharma, U., Zhang, L., et al. (2016). Opioid-induced gut microbial disruption and bile dysregulation leads to gut barrier compromise and sustained systemic inflammation. *Mucosal Immunol.* 9, 1418–1428.

Bansal, S., Nguyen, J. P., Leligdowicz, A., Zhang, Y., Kain, K. C., Ricciuto, D. R., et al. (2018). Rectal and Naris Swabs: Practical and Informative Samples for Analyzing the Microbiota of Critically Ill Patients. *mSphere* 3, e00219-18.

Bartlett, J. G. (1977). Diagnostic accuracy of transtracheal aspiration bacteriologic studies. *Am. Rev. Respir. Dis.* 115, 777–782. doi:10.1164/arrd.1977.115.5.777.

Bartlett, J. G. (2011). Diagnostic Tests for Agents of Community-Acquired Pneumonia. *Clin. Infect. Dis.* 52, S296–S304.

Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G., and Neufeld, J. D. (2011). Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads. *Appl. Environ. Microbiol.* 77, 3846–3852.

Bates, D., and Mächler, M. (2015). Ben Bolker, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67.

Bauer, R., Neffgen, N., Grempels, A., Furitsch, M., Mauerer, S., Barbaqadze, S., et al. (2020). Heterogeneity of *Streptococcus* anginosus ß-hemolysis in relation to CRISPR/Cas. *Mol. Oral Microbiol.* 35, 56–65. doi:10.1111/omi.12278.

Bäumler, A. J., and Sperandio, V. (2016). Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* 535, 85–93. doi:10.1038/nature18849.

Belko, J., Goldmann, D. A., Macone, A., and Zaidi, A. K. M. (2002). Clinically significant infections with organisms of the *Streptococcus* milleri group: *Pediatr. Infect. Dis. J.* 21, 715–723. doi:10.1097/00006454-200208000-00002.

Berdal, J. E., Bjørnholt, J., Blomfeldt, A., Smith-Erichsen, N., and Bukholm, G. (2007). Patterns and dynamics of airway colonisation in mechanically-ventilated patients. *Clin. Microbiol. Infect.* 13, 476–480.

Besselink, M. G., van Santvoort, H. C., Buskens, E., Boermeester, M. A., van Goor, H., Timmerman, H. M., et al. (2008). Probiotic prophylaxis in predicted severe acute pancreatitis: a randomised, double-blind, placebo-controlled trial. *The Lancet* 371, 651–659.

Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L., and Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin. Microbiol. Infect.* 24, 335–341.

Bishop, C. J., Aanensen, D. M., Jordan, G. E., Kilian, M., Hanage, W. P., and Spratt, B. G. (2009). Assigning strains to bacterial species via the internet. *BMC Biol.* 7, 3.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81–W87.

Bo, L., Li, J., Tao, T., Bai, Y., Ye, X., Hotchkiss, R. S., et al. (2014). Probiotics for preventing ventilator-associated pneumonia. *Cochrane Database Syst. Rev.*

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* 30, 2114–2120.

Brandt, K., and Alatossava, T. (2003). Specific identification of certain probiotic *Lactobacillus rhamnosus* strains with PCR primers based on phage-related sequences. *Int. J. Food Microbiol.* 84, 189–196.

Brassil, B., Mores, C. R., Wolfe, A. J., and Putonti, C. (2020). Characterization and spontaneous induction of urinary tract *Streptococcus anginosus* prophages. *J. Gen. Virol.* 101, 685–691. doi:10.1099/jgv.0.001407.

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., and Gerdes, S. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci \ldots. *Sci. Rep.*

Brims, F. J. H., Lansley, S. M., Waterer, G. W., and Lee, Y. C. G. (2010). Empyema thoracis: new insights into an old disease. *Eur. Respir. Rev.* 19, 220–228.

Brüssow, H., Canchaya, C., and Hardt, W.-D. (2004). Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiol. Mol. Biol. Rev.* 68, 560–602. doi:10.1128/MMBR.68.3.560-602.2004.

Brynildsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 17, 238.

Buelow, E., Bello González, T. d. j., Fuentes, S., de Steenhuijsen Piters, W. A. A., Lahti, L., Bayjanov, J. R., et al. (2017). Comparative gut microbiota and resistome profiling of intensive care patients receiving selective digestive tract decontamination and healthy subjects. *Microbiome* 5, 88.

Bullowa, J. G. M. (1935). THE RELIABILITY OF SPUTUM TYPING AND ITS RELATION TO SERUM THERAPY. *J. Am. Med. Assoc.* 105, 1512..

Cameron, E. A., and Sperandio, V. (2015). Frenemies: Signaling and Nutritional Integration in Pathogen-Microbiota-Host Interactions. *Cell Host Microbe* 18, 275–284.

Cannon, J. P., Lee, T. A., Bolanos, J. T., and Danziger, L. H. (2005). Pathogenic relevance of Lactobacillus: a retrospective review of over 200 cases. *Eur. J. Clin. Microbiol. Infect. Dis.* 24, 31–40.

Capelastegui, A., España, P. P., Bilbao, A., Gamazo, J., Medel, F., Salgado, J., et al. (2012). Etiology of community-acquired pneumonia in a population-based study: Link between etiology and patients characteristics, process-of-care, clinical evolution and outcomes. *BMC Infect. Dis.* 12, 134.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336.

Carlet, J Torres, A. (2001). Ventilator-associated pneumonia. *Eur. Respir. J.* 4, 1034–1045.

Casadevall, A., and Pirofski, L. (2017). What Is a Host? Attributes of Individual Susceptibility. *Infect. Immun.* 86, e00636-17.

Cassini, A., Plachouras, D., Eckmanns, T., Abu Sin, M., Blank, H.-P., Ducomble, T., et al. (2016). Burden of Six Healthcare-Associated Infections on European Population Health: Estimating Incidence-Based Disability-Adjusted Life Years through a Population Prevalence-Based Modelling Study. *PLOS Med.* 13, e1002150.

Cavalcanti, A. B., Lisboa, T., and Gales, A. C. (2017). Is Selective Digestive Decontamination Useful for Critically Ill Patients?: *SHOCK* 47, 52–57.

Chapman, M. J., Nguyen, N. Q., and Deane, A. M. (2011). Gastrointestinal Dysmotility: Clinical Consequences and Management of the Critically Ill Patient. *Gastroenterol. Clin. North Am.* 40, 736–739.

Chapple, L., and Deane, A. (2018). From dysmotility to virulent pathogens: implications of opioid use in the ICU. *Curr. Opin. Crit. Care* 24, 118–123.

Charlson, E. S., Bittinger, K., Haas, A. R., Fitzgerald, A. S., Frank, I., Yadav, A., et al. (2011). Topographical Continuity of Bacterial Populations in the Healthy Human Respiratory Tract. *Am. J. Respir. Crit. Care Med.* 184, 957–963.

Chastre, J., and Fagon, J. (2001). State of the Art Ventilator-associated Pneumonia. 1997. doi:10.1164/rccm.2105078.

Chastre, J., and Fagon, J.-Y. (2002). Ventilator-associated Pneumonia. *Am. J. Respir. Crit. Care Med.* 165, 867–903.

Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* 44, D694–D697.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w ^{1118} ; iso-2; iso-3. *Fly (Austin)* 6, 80–92.

Claridge, J. E., Attorri, S., Musher, D. M., Hebert, J., and Dunbar, S. (2001). *Streptococcus intermedius*, *Streptococcus constellatus*, and *Streptococcus anginosus* ("Streptococcus milleri group") are of different clinical importance and are not equally associated with abscess. *Clin. Infect. Dis.* 32, 1511–1515.

Coil, D., Jospin, G., and Darling, A. E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinforma. Oxf. Engl.* 31, 587–589.

Cook, D. J., Johnstone, J., Marshall, J. C., Lauzier, F., Thabane, L., Mehta, S., et al. (2016). Probiotics: Prevention of Severe Pneumonia and Endotracheal Colonization Trial—PROSPECT: a pilot trial. *Trials* 17, 377.

Costa, D. M., Johani, K., Melo, D. S., Lopes, L. K. O., Lopes Lima, L. K. O., Tipple, A. F. V., et al. (2019). Biofilm contamination of high-touched surfaces in intensive care units: epidemiology and potential impacts. *Lett. Appl. Microbiol.* 68, 269–276.

Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science* 326, 1694–1697.

Craven, D. E., and Hjalmarson, K. I. (2010). Ventilator-associated tracheobronchitis and pneumonia: thinking outside the box. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 51 Suppl 1, S59-66.

Crooks, N. H., Snaith, C., Webster, D., Gao, F., and Hawkey, P. (2012). Clinical review: Probiotics in critical care. *Crit. Care Lond. Engl.* 16, 237.

Dai, M., Liu, Y., Chen, W., Buch, H., Shan, Y., Chang, L., et al. (2019). Rescue fecal microbiota transplantation for antibiotic-associated diarrhea in critically ill patients. *Crit. Care* 23, 324.

Daley, R. J., Rebuck, J. a, Welage, L. S., and Rogers, F. B. (2004). Prevention of stress ulceration: current trends in critical care. *Crit. Care Med.* 32, 2008–2013.

D'Amico, R., Pifferi, S., Torri, V., Brazzi, L., Parmelli, E., and Liberati, A. (2009). Antibiotic prophylaxis to reduce respiratory tract infections and mortality in adults receiving intensive care. *Cochrane Database Syst. Rev.* doi:10.1002/14651858.CD000022.pub3.

Daneman, N., Sarwar, S., Fowler, R. A., and Cuthbertson, B. H. (2013). Effect of selective decontamination on antimicrobial resistance in intensive care units: a systematic review and meta-analysis. *Lancet Infect. Dis.* 13, 328–341.

Darlow, C. A., McGlashan, N., Kerr, R., Oakley, S., Pretorius, P., Jones, N., et al. (2020). Microbial aetiology of brain abscess in a UK cohort: prominent role of *Streptococcus intermedius*. *J. Infect.*, S0163445320301237.

David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta, A., et al. (2014). Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* 15, R89.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., et al. (2014c). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563.

De Gregoris, T. B., Aldred, N., Clare, A. S., and Burgess, J. G. (2011). Improvement of phylum- and class-specific primers for real-time PCR quantification of bacterial taxa. *J. Microbiol. Methods* 86, 351–356.

Deatherage, D. E., and Barrick, J. E. (2014). "Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq," in *Engineering and Analyzing Multicellular Systems* Methods in Molecular Biology., eds. L. Sun and W. Shou (New York, NY: Springer New York), 165–188.

Decker, M.-L., Gotta, V., Wellmann, S., and Ritz, N. (2017). Cytokine profiling in healthy children shows association of age with cytokine concentrations. *Sci. Rep.* 7, 17842.

DeFilipp, Z., Bloom, P. P., Torres Soto, M., Mansour, M. K., Sater, M. R. A., Huntley, M. H., et al. (2019). Drug-Resistant *E. coli* Bacteremia Transmitted by Fecal Microbiota Transplant. *N. Engl. J. Med.* 381, 2043–2050.

Denapaite, D., Rieger, M., Köndgen, S., Brückner, R., Ochigava, I., Kappeler, P., et al. (2016). Highly Variable Streptococcus oralis Strains Are Common among Viridans Streptococci Isolated from Primates. *mSphere* 1, mSphere.00041-15,

Denny, K. J., De Wale, J., Laupland, K. B., Harris, P. N. A., and Lipman, J. (2020). When not to start antibiotics: avoiding antibiotic overuse in the intensive care unit. *Clin. Microbiol. Infect.* 26, 35–40.

Derakhshani, H., Bernier, S. P., Marko, V. A., and Surette, M. G. (2020). Completion of draft bacterial genomes by long-read sequencing of synthetic genomic pools. *BMC Genomics* 21, 519.

Derrien, M., and van Hylckama Vlieg, J. E. T. (2015). Fate, activity, and impact of ingested bacteria within the human gut microbiota. *Trends Microbiol.* 23, 354–366.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072.

Deshpande, G., Rao, S., Patole, S., and Bulsara, M. (2010). Updated Meta-analysis of Probiotics for Preventing Necrotizing Enterocolitis in Preterm Neonates. *PEDIATRICS* 125, 921–930.

Dickson, R. P. (2016). The microbiome and critical illness. *Lancet Respir. Med.* 4, 59–72.

Dickson, R. P., Erb-Downward, J. R., Freeman, C. M., McCloskey, L., Falkowski, N. R., Huffnagle, G. B., et al. (2017). Bacterial Topography of the Healthy Human Lower Respiratory Tract. *mBio* 8, e02287–16.

Dickson, R. P., Erb-Downward, J. R., and Huffnagle, G. B. (2014a). Towards an ecology of the lung: new conceptual models of pulmonary microbiology and pneumonia pathogenesis. *Lancet Respir. Med.* 2, 238–246.

Dickson, R. P., Erb-Downward, J. R., Prescott, H. C., Martinez, F. J., Curtis, J. L., Lama, V. N., et al. (2014b). Cell-associated bacteria in the human lung microbiome. *Microbiome* 2, 28.

Dickson, R. P., Schultz, M. J., van der Poll, T., Schouten, L. R., Falkowski, N. R., Luth, J. E., et al. (2020). Lung Microbiota Predict Clinical Outcomes in Critically Ill Patients. *Am. J. Respir. Crit. Care Med.* 201, 555–563.

Dickson, R. P., Singer, B. H., Newstead, M. W., Falkowski, N. R., Erb-Downward, J. R., Standiford, T. J., et al. (2016a). Enrichment of the lung microbiome with gut

bacteria in sepsis and the acute respiratory distress syndrome. *Nat. Microbiol.* 1, 16113.

Dickson, R. P., Singer, B. H., Newstead, M. W., Falkowski, N. R., Erb-Downward, J. R., and Standiford, T. J. (2016b). Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. *Nat. Microbiol.* 1.

Didari, T., Solki, S., Mozaffari, S., Nikfar, S., and Abdollahi, M. (2014). A systematic review of the safety of probiotics. *Expert Opin. Drug Saf.* 13, 227–239.

Doern, C. D., and Burnham, C.-A. D. (2010). It's Not Easy Being Green: the Viridans Group Streptococci, with a Focus on Pediatric Clinical Manifestations. *J. Clin. Microbiol.* 48, 3829–3835.

Doron, S., and Snydman, D. R. (2015). Risk and Safety of Probiotics. *Clin. Infect. Dis.* 60, S129–S134.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

Edwards, D. J., and Holt, K. E. (2013). Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microb. Inform. Exp.* 3, 2.

Edwardson, S., and Cairns, C. (2019). Nosocomial infections in the ICU. *Anaesth. Intensive Care Med.* 20, 14–18.

Escapa, I. F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F. E., and Lemon, K. P. (2018). New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. *mSystems* 3, e00187-18.

Estrela, S., Whiteley, M., and Brown, S. P. (2015). The demographic determinants of human microbiome health. *Trends Microbiol.* 23, 134–141.

Exterkate, R. A. M., Zaura, E., Brandt, B. W., Buijs, M. J., Koopman, J. E., Crielaard, W., et al. (2015). The effect of propidium monoazide treatment on the measured bacterial composition of clinical samples after the use of a mouthwash. *Clin. Oral Investig.* 19, 813–822.

Facklam, R. (2002). What Happened to the Streptococci: Overview of Taxonomic and Nomenclature Changes. *Clin. Microbiol. Rev.* 15, 613–630.

Fair, K., Dunlap, D. G., Fitch, A., Bogdanovich, T., Methé, B., Morris, A., et al. (2019). Rectal Swabs from Critically Ill Patients Provide Discordant Representations of the Gut Microbiome Compared to Stool Samples. *mSphere* 4, e00358-19.

Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., et al. (2013). The long-term stability of the human gut microbiota. *Science* 341, 1237439–1237439.

Fasoli, S., Marzotto, M., Rizzotti, L., Rossi, F., Dellaglio, F., and Torriani, S. (2003). Bacterial composition of commercial probiotic products as evaluated by PCR-DGGE analysis. *Int. J. Food Microbiol.* 82, 59–70.

Fazili, T., Riddell, S., Kiska, D., Endy, T., Giurgea, L., Sharngoe, C., et al. (2017). *Streptococcus* anginosus Group Bacterial Infections. *Am. J. Med. Sci.* 354, 257–261.

Finley, C., Clifton, J., FitzGerald, J. M., and Yee, J. (2008). Empyema: An Increasing Concern in Canada. *Can. Respir. J.* 15, 85–89.

Fisher, L. E., and Russell, R. R. B. (1993). The Isolation and Characterization of Milleri Group Streptococci from Dental Periapical Abscesses. *J. Dent. Res.* 72, 1191–1193.

Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E., Kerlavage, A., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512.

Fleming-Dutra, K. E., Hersh, A. L., Shapiro, D. J., Bartoces, M., Enns, E. A., File, T. M., et al. (2016). Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011. *JAMA* 315, 1864.

Freedberg, D. E., Lebwohl, B., and Abrams, J. a. (2014). The Impact of Proton Pump Inhibitors on the Human Gastrointestinal Microbiome. *Clin. Lab. Med.* 34, 771–785.

Freedberg, D. E., Zhou, M. J., Cohen, M. E., Annavajhala, M. K., Khan, S., Moscoso, D. I., et al. (2018). Pathogen colonization of the gastrointestinal microbiome at intensive care unit admission and risk for subsequent death or infection. *Intensive Care Med.* 44, 1203–1211.

Fu, Y., Moscoso, D. I., Porter, J., Krishnareddy, S., Abrams, J. A., Seres, D., et al. (2020). Relationship Between Dietary Fiber Intake and Short-Chain Fatty Acid–Producing Bacteria During Critical Illness: A Prospective Cohort Study. *J. Parenter. Enter. Nutr.* 44, 463–471.

Furuichi, M., and Horikoshi, Y. (2018a). Sites of infection associated with *Streptococcus* anginosus group among children. *J. Infect. Chemother. Off. J. Jpn. Soc. Chemother.* 24, 99–102.

Furuichi, M., and Horikoshi, Y. (2018b). Sites of infection associated with *Streptococcu*s anginosus group among children. *J. Infect. Chemother.* 24, 99–102.

Gadsby, N. J., Russell, C. D., McHugh, M. P., Mark, H., Conway Morris, A., Laurenson, I. F., et al. (2016). Comprehensive Molecular Testing for Respiratory Pathogens in Community-Acquired Pneumonia. *Clin. Infect. Dis.* 62, 817–823.

Galdiero, F., Carratelli, C. R., Nuzzo, I., Bentivoglio, C., and Galdiero, M. (1988). Phagocytosis of bacterial aggregates by granulocytes. *Eur. J. Epidemiol.* 4, 456–460.

Gerlach, D., Guo, Y., De Castro, C., Kim, S.-H., Schlatterer, K., Xu, F.-F., et al. (2018). Methicillin-resistant *Staphylococcus aureus* alters cell wall glycosylation to evade immunity. *Nature* 563, 705–709.

Ghareeb, P. A., Bourlai, T., Dutton, W., and McClellan, W. T. (2013). Reducing pathogen transmission in a hospital setting. Handshake verses fist bump: a pilot study. *J. Hosp. Infect.* 85, 321–323.

Giannella, R. a, Broitman, S. a, and Zamcheck, N. (1972). Gastric acid barrier to ingested microorganisms in man: studies in vivo and in vitro. *Gut* 13, 251–256.

Giddings, K. S., Zhao, J., Sims, P. J., and Tweten, R. K. (2004). Human CD59 is a receptor for the cholesterol-dependent cytolysin intermedilysin. *Nat. Struct. Mol. Biol.* 11, 1173–1178.

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., and Knight, R. (2018). Current understanding of the human microbiome. *Nat. Med.* 24, 392–400.

Gil-Perotin, S., Ramirez, P., Marti, V., Sahuquillo, J. M., Gonzalez, E., Calleja, I., et al. (2012). Implications of endotracheal tube biofilm in ventilator-associated pneumonia response: a state of concept. *Crit. Care* 16, R93.

Giraldi, K. (2015). Mechanisms of host recognition and immune evasion of members of the *Streptococcus* anginosus/milleri group. Available at: http://hdl.handle.net/11375/18306.

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., and Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog.* 8, 24.

Goldenberg, J. Z., Yap, C., Lytvyn, L., Lo, C. K.-F., Beardsley, J., Mertz, D., et al. (2017). Probiotics for the prevention of *Clostridium difficile*-associated diarrhea in adults and children. *Cochrane Database Syst. Rev.*

Goldstein, E. J. C., Citron, D. M., Claros, M. C., and Tyrrell, K. L. (2014). Bacterial counts from five over-the-counter probiotics: Are you getting what you paid for? *Anaerobe* 25, 1–4.

Goldstein, E. J. C., Tyrrell, K. L., and Citron, D. M. (2015). *Lactobacillus* Species: Taxonomic Complexity and Controversial Susceptibilities. *Clin. Infect. Dis.* 60, S98–S107.

Gossling, J. (1988a). Occurrence and Pathogenicity of the *Streptococcus* milleri Group. *Clin. Infect. Dis.* 10, 257–285.

Gossling, J. (1988b). Occurrence and Pathogenicity of the *Streptococcus* milleri Group. *Clin. Infect. Dis.* 10, 257–285.

Grijalva, C. G., Zhu, Y., Nuorti, J. P., and Griffin, M. R. (2011). Emergence of parapneumonic empyema in the USA. *Thorax* 66, 663–668.

Grinberg, M., Orevi, T., and Kashtan, N. (2019). Bacterial surface colonization, preferential attachment and fitness under periodic stress. *PLOS Comput. Biol.* 15, e1006815.

Grinwis, M. E. (2012). Phenotypic and genotypic characteristics of respiratory isolates of the streptococcus milleri group from patients with cystic fibrosis. [Ph.D. Thesis]

Grinwis, M. E., Sibley, C. D., Parkins, M. D., Eshaghurshan, C. S., Rabin, H. R., and Surette, M. G. (2010a). Characterization of *Streptococcus* milleri Group Isolates from Expectorated Sputum of Adult Patients with Cystic Fibrosis. *J. Clin. Microbiol.* 48, 395–401.

Grinwis, M. E., Sibley, C. D., Parkins, M. D., Eshaghurshan, C. S., Rabin, H. R., and Surette, M. G. (2010b). Macrolide and Clindamycin Resistance in *Streptococcus* milleri Group Isolates from the Airways of Cystic Fibrosis Patients. *Antimicrob. Agents Chemother.* 54, 2823–2829.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinforma. Oxf. Engl.* 29, 1072–1075.

Guthof, O. (1956). [Pathogenic strains of Streptococcus viridans; streptocci found in dental abscesses and infiltrates in the region of the oral cavity]. *Zentralblatt Bakteriol. Parasitenkd. Infekt. Hyg. 1 Abt Med.-Hyg. Bakteriol. Virusforsch. Parasitol. Orig.* 166, 553–564.

Guthrie, J. L., and Gardy, J. L. (2017). A brief primer on genomic epidemiology: lessons learned from *Mycobacterium tuberculosis*: A primer on *M. tuberculosis* genomic epidemiology. *Ann. N. Y. Acad. Sci.* 1388, 59–77.

Haak, B. W., Levi, M., and Wiersinga, W. J. (2017). Microbiota-targeted therapies on the intensive care unit: *Curr. Opin. Crit. Care* 23, 167–174.

Hao, Q., Dong, B. R., and Wu, T. (2015). Probiotics for preventing acute upper respiratory tract infections. *Cochrane Database Syst. Rev.* 5, CD006895.

Haque, M., Sartelli, M., McKimm, J., and Abu Bakar, M. (2018). Health care-associated infections – an overview. *Infect. Drug Resist.* 11, 2321–2333.

Harrell Jr, F. E. (2013). *Hmisc: Harrell miscellaneous. R package version 3.12-2*. Computer software]. Available from http://cran. r-project.

Harris, B., Morjaria, S. M., Littmann, E. R., Geyer, A. I., Stover, D. E., Barker, J. N., et al. (2016). Gut Microbiota Predict Pulmonary Infiltrates after Allogeneic Hematopoietic Cell Transplantation. *Am. J. Respir. Crit. Care Med.* 194, 450–463.

Hasegawa, N., Sekizuka, T., and, Y. S. I., and 2017 (2017a). Characterization of the pathogenicity of Streptococcus intermedius TYG1620 isolated from a human brain abscess based on the complete genome sequence with \ldots. *Infect. Immun.* 85.

Hasegawa, N., Sekizuka, T., Sugi, Y., Kawakami, N., Ogasawara, Y., Kato, K., et al. (2017b). Characterization of the Pathogenicity of Streptococcus intermedius TYG1620 Isolated from a Human Brain Abscess Based on the Complete Genome Sequence with Transcriptome Analysis and Transposon Mutagenesis in a Murine Subcutaneous Abscess Model. *Infect. Immun.* 85, e00886–16.

Håvarstein, L. S., Gaustad, P., Nes, I. F., and Morrison, D. A. (1996). Identification of the streptococcal competence-pheromone receptor. *Mol. Microbiol.* 21, 863–869.

Hayakawa, M., Asahara, T., Henzan, N., Murakami, H., Yamamoto, H., Mukai, N., et al. (2011). Dramatic changes of the gut flora immediately after severe and sudden insults. *Dig. Dis. Sci.* 56, 2361–2365.

Hempel, S., Newberry, S. J., Maher, A. R., Wang, Z., Miles, J. N. V., Shanman, R., et al. (2012). Probiotics for the Prevention and Treatment of Antibiotic-Associated Diarrhea: A Systematic Review and Meta-analysis. *JAMA* 307, 1959.

Hempel, S., Newberry, S., Ruelaz, A., Wang, Z., Miles, J. N. V., Suttorp, M. J., et al. (2011). Safety of probiotics used to reduce risk and prevent or treat disease. *Evid. ReportTechnology Assess.*, 1–645.

Heyland, D., Cook, D. J., Winder, B., Brylowski, L., Van deMark, H., and Guyatt, G. (1995). Enteral nutrition in the critically ill patient: A prospective survey. *Crit. Care Med.* 23, 1055–1060.

Heyland, D. K., Tougas, G., King, D., and Cook, D. J. (1996). Impaired gastric emptying in mechanically ventilated, critically ill patients. *Intensive Care Med.* 22, 1339–1344.

Hill, C., Guarner, F., Reid, G., Gibson, G. R., Merenstein, D. J., Pot, B., et al. (2014). The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat. Rev. Gastroenterol. Hepatol.* 11, 506–514.

Hooks, K. B., and O'Malley, M. A. (2017). Dysbiosis and Its Discontents. *mBio* 8, mBio.01492-17, e01492-17.

Horan, T. C., Andrus, M., and Dudeck, M. A. (2008). CDC/NHSN surveillance definition of health care–associated infection and criteria for specific types of infections in the acute care setting. *Am. J. Infect. Control* 36, 309–332.

Howard, S. T., Rhoades, E., Recht, J., Pang, X., Alsup, A., Kolter, R., et al. (2006). Spontaneous reversion of *Mycobacterium abscessus* from a smooth to a rough morphotype is associated with reduced expression of glycopeptidolipid and reacquisition of an invasive phenotype. *Microbiology* 152, 1581–1590.

Huang, C.-H., Li, S.-W., Huang, L., and Watanabe, K. (2018). Identification and Classification for the *Lactobacillus casei* Group. *Front. Microbiol.* 9.

Huang, S. S., Datta, R., and Platt, R. (2006). Risk of Acquiring Antibiotic-Resistant Bacteria From Prior Room Occupants. *Arch. Intern Med.* Oct 9; 166(18): 1945-51.

Huffnagle, G. B., and Dickson, R. P. (2015). The bacterial microbiota in inflammatory lung diseases. *Clin. Immunol.* 159, 177–182.

Hurford, A., Morris, A. M., Fisman, D. N., and Wu, J. (2012). Linking antimicrobial prescribing to antimicrobial resistance in the ICU: Before and after an antimicrobial stewardship program. *Epidemics* 4, 203–210.

Iapichino, G., Callegari, M. L., Marzorati, S., Cigada, M., Corbella, D., Ferrari, S., et al. (2008). Impact of antibiotics on the gut microbiota of critically ill patients. *J. Med. Microbiol.* 57, 1007–1014.

Iapichino, G., Lankelma, J. M., and Joost, W. W. (2017). Gut microbiota disruption in critically ill patients: Discussion on" Critically ill patients demonstrate large interpersonal variation of intestinal microbiota. Intensive Care Med. May; 43(5): 718-719.

Imhann, F., Bonder, M. J., Vich Vila, A., Fu, J., Mujagic, Z., Vork, L., et al. (2016). Proton pump inhibitors affect the gut microbiome. *Gut* 65, 740–748.

Issa, E., Salloum, T., Panossian, B., Ayoub, D., Abboud, E., and Tokajian, S. (2019). Genome Mining and Comparative Analysis of Streptococcus intermedius Causing Brain Abscess in a Child. *Pathogens* 8, 22.

Issa, E., Salloum, T., and Tokajian, S. (2020). From Normal Flora to Brain Abscesses: A Review of *Streptococcus intermedius*. *Front. Microbiol.* 11, 826.

Jacobs, J. A., Pietersen, H. G., Stobberingh, E. E., and Soeters, P. B. (1995). *Streptococcus anginosus*, *Streptococcus constellatus* and *Streptococcus intermedius* Clinical Relevance, Hemolytic and Serologic Characteristics. *Am. J. Clin. Pathol.* 104, 547–553.

Jacobs, J. A., Schot, C. S., and Schouls, L. M. (2000). The *Streptococcus* anginosus species comprises five 16S rRNA ribogroups with different phenotypic characteristics and clinical relevance. *Int. J. Syst. Evol. Microbiol.* 50, 1073–1079.

Jain, S., Self, W. H., Wunderink, R. G., Fakhran, S., Balk, R., Bramley, A. M., et al. (2015). Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *N. Engl. J. Med.* 373, 415–427.

Jensen, A., Hoshino, T., and Kilian, M. (2013a). Taxonomy of the Anginosus group of the genus Streptococcus and description of *Streptococcus anginosus* subsp. whileyi subsp. nov. and *Streptococcus constellatus* subsp. viborgensis subsp. nov. *Int. J. Syst. Evol. Microbiol.* 63, 2506–2519.

Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573.

Jiang, S., Li, M., Fu, T., Shan, F., Jiang, L., and Shao, Z. (2020). Clinical Characteristics of Infections Caused by *Streptococcus* Anginosus Group. *Sci. Rep.* 10, 9032.

Jiménez de Bagüés, M. P., Terraza, A., Gross, A., and Dornand, J. (2004). Different Responses of Macrophages to Smooth and Rough *Brucella* spp.: Relationship to Virulence. *Infect. Immun.* 72, 2429–2433.

Johani, K., Abualsaud, D., Costa, D. M., Hu, H., Whiteley, G., Deva, A., et al. (2018). Characterization of microbial community composition, antimicrobial resistance and biofilm on intensive care surfaces. *J. Infect. Public Health* 11, 418–424.

Johanson, W. G., Pierce, A. K., and Sanford, J. P. (1969). Changing Pharyngeal Bacterial Flora of Hospitalized Patients: Emergence of Gram-Negative Bacilli. *N. Engl. J. Med.* 281, 1137–1140.

Johnstone, J., Garber, G., and Muller, M. (2019a). Health care-associated infections in Canadian hospitals: still a major problem. *Can. Med. Assoc. J.* 191, E977–E978.

Johnstone, J., Heels-Ansdell, D., Thabane, L., Meade, M., Marshall, J., Lauzier, F., et al. (2019b). Evaluating probiotics for the prevention of ventilator-associated pneumonia: a randomised placebo-controlled multicentre trial protocol and statistical analysis plan for PROSPECT. *BMJ Open* 9, e025228.

Jordan, E. O. (1926). THE INTERCONVERTIBILITY OF "ROUGH" AND "SMOOTH" BACTERIAL TYPES. *JAMA J. Am. Med. Assoc.* 86, 177.

Joseph, N. M., Sistla, S., Dutta, T. K., Badhe, A. S., and Parija, S. C. (2010). Ventilator-associated pneumonia: A review. *Eur. J. Intern. Med.* 21, 360–368.

Junckerstorff, R. K., and Murray, R. J. (2016b). Streptococcus anginosus Group Bacteria: No Longer a Case of Mistaken Identity. *Clin. Microbiol. Newsl.* 38, 35–39.

Junckerstorff, R. K., Robinson, J. O., and Murray, R. J. (2014). Invasive *Streptococcus* anginosus group infection—does the species predict the outcome? *Int. J. Infect. Dis.* 18, 38–40.

Kabat, A. M., Srinivasan, N., and Maloy, K. J. (2014). Modulation of immune development and function by intestinal microbiota. *Trends Immunol.*, 1–11.

Kaiser, J. C., Verschoor, C. P., Surette, M. G., and Bowdish, D. M. (2014). Host cytokine responses distinguish invasive from airway isolates of the *Streptococcus* milleri/anginosis group. 14, 1–11.

Kalanuria, A. A., Zai, W., and Mirski, M. (2014). Ventilator-associated pneumonia in the ICU. *Crit. Care*. Mar 18; 18(2): 208.

Kanamori, S., Shinzato, T., Saito, A., and Kusano, N. (2004). The role of the capsule of the *Streptococcus* milleri group in its pathogenicity. *J. Infect. Chemother.* 10, 105–109.

Kankainen, M., Paulin, L., Tynkkynen, S., von Ossowski, I., Reunanen, J., Partanen, P., et al. (2009). Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein. *Proc. Natl. Acad. Sci.* 106, 17193–17198.

Kassambara, A., Kosinski, M., and Biecek, P. (2017). survminer: Drawing Survival Curves using "ggplot2".

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinforma. Oxf. Engl.* 28, 1647–1649.

Kelly, B. J., Imai, I., Bittinger, K., Laughlin, A., Fuchs, B. D., Bushman, F. D., et al. (2016). Composition and dynamics of the respiratory tract microbiome in intubated patients. *Microbiome* 4, 7.

Khan, H. A., Baig, F. K., and Mehboob, R. (2017). Nosocomial infections: Epidemiology, prevention, control and surveillance. *Asian Pac. J. Trop. Biomed.* 7, 478–482.

Kitsios, G. D., Morowitz, M. J., Dickson, R. P., Huffnagle, G. B., McVerry, B. J., and Morris, A. (2017). Dysbiosis in the intensive care unit: Microbiome science coming to the bedside. *J. Crit. Care* 38, 84–91.

Knaus, W. A., Draper, E. A., Wagner, D. P., and Zimmerman, J. E. (1985). APACHE II: a severity of disease classification system. *Crit. Care Med.* 13.

Kobashi, Y., Mouri, K., Yagi, S., Obase, Y., and Oka, M. (2008). Clinical analysis of cases of empyema due to Streptococcus milleri group. *Jpn. J. Infect. Dis.* 61, 484–486.

Kobo, O., Nikola, S., Geffen, Y., and Paul, M. (2017). The pyogenic potential of the different *Streptococcus* anginosus group bacterial species: retrospective cohort study. *Epidemiol. Infect.* 145, 3065–3069.

Kollef, M. H., Hamilton, C. W., and Ernst, F. R. (2012). Economic Impact of Ventilator-Associated Pneumonia in a Large Matched Cohort. *Infect. Control Hosp. Epidemiol.* 33, 250–256.

Kothari, D., Patel, S., and Kim, S.-K. (2019). Probiotic supplements might not be universally-effective and safe: A review. *Biomed. Pharmacother.* 111, 537–547.

Lacroix, A.-M. (2014). INVESTIGATION OF COMPETENCE HETEROGENEITY IN STREPTOCOCCUS MILLERI GROUP CLINICAL ISOLATES. [M.Sc. Thesis]

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108.

Lamarche, D., Johnstone, J., Zytaruk, N., Clarke, F., Hand, L., Loukov, D., et al. (2018). Microbial dysbiosis and mortality during mechanical ventilation: a prospective observational study. *Respir. Res.* 19, 245.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

Lankelma, J. M., van Vught, L. A., Belzer, C., Schultz, M. J., van der Poll, T., de Vos, W. M., et al. (2017). Critically ill patients demonstrate large interpersonal variation in intestinal microbiota dysregulation: a pilot study. *Intensive Care Med.* 43, 59–68.

Lartillot, N., and Philippe, H. (2004). A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. *Mol. Biol. Evol.* 21, 1095–1109.

Lau, J. T., Whelan, F. J., Herath, I., Lee, C. H., Collins, S. M., Bercik, P., et al. (2016). Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Med.* 8, 1.

Lau, V. I., Rochwerg, B., Xie, F., Johnstone, J., Basmaji, J., Balakumaran, J., et al. (2020). Probiotics in hospitalized adult patients: a systematic review of economic evaluations. *Can. J. Anesth. Can. Anesth.* 67, 247–261.

Laupland, K. B., Pasquill, K., Parfitt, E. C., Dagasso, G., and Steele, L. (2018). *Streptococcus anginosus* group bloodstream infections in the western interior of British Columbia, Canada. *Infect. Dis.* 50, 423–428.

Laupland, Ross, T., Church, D. L., and Gregson, D. B. (2006). Population-based surveillance of invasive pyogenic streptococcal infection in a large Canadian region. *Clin. Microbiol. Infect.* 12, 224–230.

Leitao Filho, F. S., Alotaibi, N. M., Ngan, D., Tam, S., Yang, J., Hollander, Z., et al. (2019). Sputum Microbiome Is Associated with 1-Year Mortality after Chronic Obstructive Pulmonary Disease Hospitalizations. *Am. J. Respir. Crit. Care Med.* 199, 1205–1213.

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259.

Levy, M., Blacher, E., and Elinav, E. (2017). Microbiome, metabolites and host immunity. *Curr. Opin. Microbiol.* 35, 8–15.

Lewis, Z. T., Shani, G., Masarweh, C. F., Popovic, M., Frese, S. A., Sela, D. A., et al. (2016). Validating bifidobacterial species and subspecies identity in commercial probiotic products. *Pediatr. Res.* 79, 445–452.

Lherm, T., Monet, C., Nougière, B., Soulier, M., Larbi, D., Le Gall, C., et al. (2002). Seven cases of fungemia with *Saccharomyces boulardii* in critically ill patients. *Intensive Care Med.* 28, 797–801.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* 25, 2078–2079.

Li, Q., Wang, C., Tang, C., He, Q., Zhao, X., Li, N., et al. (2015). Successful treatment of severe sepsis and diarrhea after vagotomy utilizing fecal microbiota transplantation: a case report. *Crit. Care* 19, 37.

Li, Y., Oosting, M., Deelen, P., Ricaño-Ponce, I., Smeekens, S., Jaeger, M., et al. (2016a). Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. *Nat. Med.* 22, 952–960.

Li, Y., Oosting, M., Smeekens, S. P., Jaeger, M., Aguirre-Gamboa, R., Le, K. T. T., et al. (2016b). A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans. *Cell* 167, 1099-1110.

Light, R. W. (2006). Parapneumonic Effusions and Empyema. *Proc. Am. Thorac. Soc.* 3, 75–80.

Limia, A., Alarcón, T., Jiménez, M. L., and López-Brea, M. (2000). Comparison of Three Methods for Identification of *Streptococcus* milleri Group Isolates to Species Level. *Eur. J. Clin. Microbiol. Infect. Dis.* 19, 128–131.

Lloyd-Price, J., Abu-Ali, G., and Huttenhower, C. (2016). The healthy human microbiome. *Genome Med.* 8, 51.

Loman, N. J., Constantinidou, C., Chan, J. Z. M., Halachev, M., Sergeant, M., Penn, C. W., et al. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.* 10, 599–

Loveday, H. P., Wilson, J. A., Pratt, R. J., Golsorkhi, M., Tingle, A., Bak, A., et al. (2014). epic3: National Evidence-Based Guidelines for Preventing Healthcare-Associated Infections in NHS Hospitals in England. *J. Hosp. Infect.* 86, S1–S70. d

Lozupone, C. a., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature* 489, 220–230.

Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., and Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nat. Methods* 10, 999–1002.

Luyt, C.-E., Hékimian, G., Koulenti, D., and Chastre, J. (2018). Microbial cause of ICU-acquired pneumonia: hospital-acquired pneumonia versus ventilator-associated pneumonia. *Curr. Opin. Crit. Care* 24, 332–338.

Ma, B., Forney, L. J., and Ravel, J. (2012). The vaginal microbiome: rethinking health and diseases. *Annu. Rev. Microbiol.* 66, 371–389.

MacLaren, R., Reynolds, P. M., and Allen, R. R. (2014). Histamine-2 receptor antagonists vs proton pump inhibitors on gastrointestinal tract hemorrhage and infectious complications in the intensive care unit. *JAMA Intern. Med.* 174, 564–74..

Magill, S. S., O'Leary, E., Janelle, S. J., Thompson, D. L., Dumyati, G., Nadle, J., et al. (2018). Changes in Prevalence of Health Care–Associated Infections in U.S. Hospitals. *N. Engl. J. Med.* 379, 1732–1744.

Maldonado-Gómez, M. X., Martínez, I., Bottacini, F., O'Callaghan, A., Ventura, M., van Sinderen, D., et al. (2016). Stable Engraftment of *Bifidobacterium longum* AH1206 in the Human Gut Depends on Individualized Features of the Resident Microbiome. *Cell Host Microbe* 20, 515–526.

Man, W. H., de Steenhuijsen Piters, W. A. A., and Bogaert, D. (2017). The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat. Rev. Microbiol.* 15, 259–270.

Mancilla, M. (2016). Smooth to Rough Dissociation in Brucella: The Missing Link to Virulence. *Front. Cell. Infect. Microbiol.* Jan 5; 5:98.

Manzanares, W., Lemieux, M., Langlois, P. L., and Wischmeyer, P. E. (2016). Probiotic and synbiotic therapy in critical illness: a systematic review and meta-analysis. *Crit. Care* 19, 262.

Marcobal, A., Underwood, M. A., and Mills, D. A. (2008). Rapid Determination of the Bacterial Composition of Commercial Probiotic Products by Terminal Restriction Fragment Length Polymorphism Analysis: *J. Pediatr. Gastroenterol. Nutr.* 46, 608–611. d

Marshall, J. C. (2010). Critical illness is an iatrogenic disorder: *Crit. Care Med.* 38, S582–S589.

Marshall, J. C., Christou, N. V., Horn, R., and Meakins, J. L. (1988). The microbiology of multiple organ failure. The proximal gastrointestinal tract as an occult reservoir of pathogens. *Arch. Surg. Chic. Ill 1960* 123, 309–315.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10.

Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., and Neufeld, J. D. (2012). PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 13, 31.

Maskell, N. A., Batt, S., Hedley, E. L., Davies, C. W. H., Gillespie, S. H., and Davies, R. J. O. (2006). The Bacteriology of Pleural Infection by Genetic and Standard Methods and Its Mortality Significance. *Am. J. Respir. Crit. Care Med.* 174, 817–823.

May, G. a O., Reynolds, N., Smith, A. R., Macfarlane, G. T., and Kennedy, A. (2005). Effect of pH and Antibiotics on Microbial Overgrowth in the Stomachs and Duodena of Patients Undergoing Percutaneous Endoscopic Gastrostomy Feeding. *J. Clin. Microbiol.* 43, 3059–3065.

McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., et al. (2013). The Comprehensive Antibiotic Resistance Database. *Antimicrob. Agents Chemother.* 57, 3348–3357.

McCullor, K., Rahman, M., King, C., and McShan, W. M. (2019). Characterization of *Streptococcus pneumoniae* phage-like element SpnCI reveals an enhanced virulent phenotype in the acute invertebrate infection model *Galleria mellonella*. *bioRxiv*, 670141.

McCullough, J. L., and Maren, T. H. (1973). Inhibition of dihydropteroate synthetase from *Escherichia coli* by sulfones and sulfonamides. *Antimicrob. Agents Chemother.* 3, 665–669.

McDonald, D., Ackermann, G., Khailova, L., Baird, C., Heyland, D., Kozar, R., et al. (2016). Extreme Dysbiosis of the Microbiome in Critical Illness. *mSphere* 1, e00199-16.

McMurdie, P. J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* 8, e61217–11.

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39, W339–W346.

Mendonca, M. L., Szamosi, J. C., Lacroix, A.-M., Fontes, M. E., Bowdish, D. M., and Surette, M. G. (2017). The sil Locus in *Streptococcus* Anginosus Group: Interspecies Competition and a Hotspot of Genetic Diversity. *Front. Microbiol.* 7.

Merrick, B., Allen, L., Masirah M Zain, N., Forbes, B., Shawcross, D. L., and Goldenberg, S. D. (2020). Regulation, risk and safety of Faecal Microbiota Transplant. *Infect. Prev. Pract.* 2, 100069.

Meyer, C. N., Rosenlund, S., Nielsen, J., and Friis-Møller, A. (2011). Bacteriological aetiology and antimicrobial treatment of pleural empyema. *Scand. J. Infect. Dis.* 43, 165–169.

Mietto, C., Pinciroli, R., Patel, N., and Berra, L. (2013). Ventilator associated pneumonia: evolving definitions and preventive strategies. *Respir. Care* 58, 990–1007.

Miller-Ensminger, T., Garretto, A., Brenner, J., Thomas-White, K., Zambom, A., Wolfe, A. J., et al. (2018). Bacteriophages of the Urinary Microbiome. *J. Bacteriol.* Mar 12; 200(7);e00738-17.

Mitchell, R., Taylor, G., Rudnick, W., Alexandre, S., Bush, K., Forrester, L., et al. (2019). Trends in health care–associated infections in acute care hospitals in Canada: an analysis of repeated point-prevalence surveys. *Can. Med. Assoc. J.* 191, E981–E988.

Moayyedi, P., Surette, M. G., Kim, P. T., Libertucci, J., Wolfe, M., Onischi, C., et al. (2015). Fecal Microbiota Transplantation Induces Remission in Patients With Active Ulcerative Colitis in a Randomized Controlled Trial. *Gastroenterology* 149, 102–109.e6.

Molina, J.-M., Leport, C., Bure, A., Wolff, M., Michon, C., and Vilde, J.-L. (1991). Clinical and Bacterial Features of Infections Caused by *Streptococcus milleri*. *Scand. J. Infect. Dis.* 23, 659–666.

Monier, J.-M., and Lindow, S. E. (2003). Differential survival of solitary and aggregated bacterial cells promotes aggregate formation on leaf surfaces. *Proc. Natl. Acad. Sci.* 100, 15977–15982.

Morgan, X. C., and Huttenhower, C. (2012). Chapter 12: Human microbiome analysis. *PLoS Comput. Biol.* 8, e1002808.

Morgan, X. C., Segata, N., and Huttenhower, C. (2013). Biodiversity and functional genomics in the human microbiome. *Trends Genet. TIG* 29, 51–8.

Morita, E., Narikiyo, M., Yokoyama, A., Yano, A., Kamoi, K., Yoshikawa, E., et al. (2005). Predominant presence of Streptococcus anginosus in the saliva of alcoholics. *Oral Microbiol. Immunol.* 20, 362–365.

Morovic, W., Hibberd, A. A., Zabel, B., Barrangou, R., and Stahl, B. (2016). Genotyping by PCR and High-Throughput Sequencing of Commercial Probiotic Products Reveals Composition Biases. *Front. Microbiol.* 7, 1747.

Muscedere, J. G., Martin, C. M., and Heyland, D. K. (2008). The impact of ventilator-associated pneumonia on the Canadian health care system. *J. Crit. Care* 23, 5–10.

Nagamune, H., Ohnishi, C., Katsuura, A., Fushitani, K., Whiley, R. A., Tsuji, A., et al. (1996). Intermedilysin, a Novel Cytotoxin Specific for Human Cells, Secreted by *Streptococcus intermedius* UNS46 Isolated from a Human Liver Abscess. *Infect. Immun.* Aug; 64(8):3093-100.

Nagamune, H., Whiley, R. A., Goto, T., Inai, Y., Maeda, T., Hardie, J. M., et al. (2000). Distribution of the Intermedilysin Gene among the Anginosus Group Streptococci and Correlation between Intermedilysin Production and Deep-Seated Infection with *Streptococcus intermedius*. *J. Clin. Microbiol.* Jan; 38(1): 220-26.

Natividad, J. M. M., and Verdu, E. F. (2013). Modulation of intestinal barrier by intestinal microbiota: Pathological and therapeutic implications. *Pharmacol. Res.* 69, 42–51.

Nightingale, F. (1863). *Notes on hospitals*. 3rd Edition. London: Longman, Green, Longman , Roberts, and Green.

Noguchi, S., Yatera, K., Kawanami, T., Yamasaki, K., Naito, K., Akata, K., et al. (2015). The clinical features of respiratory infections caused by the *Streptococcus* anginosus group. *BMC Pulm. Med.* 15, 133.

O'Dwyer, D. N., Ashley, S. L., Gurczynski, S. J., Xia, M., Wilke, C., Falkowski, N. R., et al. (2019). Lung Microbiota Contribute to Pulmonary Inflammation and Disease Progression in Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* 199, 1127–1138.

Oh, J., Byrd, A. L., Park, M., NISC Comparative Sequencing Program, Kong, H. H., and Segre, J. A. (2016). Temporal Stability of the Human Skin Microbiome. *Cell* 165, 854–866.

Ohland, C. L., and MacNaughton, W. K. (2010). Probiotic bacteria and intestinal epithelial barrier function. *Am. J. Physiol.-Gastrointest. Liver Physiol.* 298, G807–G819. doi:10.1152/ajpgi.00243.2009.

Okada, F., Ono, A., Ando, Y., Nakayama, T., Ishii, H., Hiramatsu, K., et al. (2013). High-resolution CT findings in *Streptococcus milleri* pulmonary infection. *Clin. Radiol.* 68, e331–e337.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2019). *vegan: Community Ecology Package*. Available at: https://CRAN.R-project.org/package=vegan.

Olson, A. B., Kent, H., Sibley, C. D., Grinwis, M. E., Mabon, P., Ouellette, C., et al. (2013). Phylogenetic relationship and virulence inference of *Streptococcus* Anginosus Group: curated annotation and whole-genome comparative analysis support distinct species designation. *BMC Genomics*. Dec 17; 14: 895

Olson, A. B., Sibley, C. D., Schmidt, L., Wilcox, M. A., Surette, M. G., and Corbett, C. R. (2010). Development of Real-Time PCR Assays for Detection of the *Streptococcus milleri* Group from Cystic Fibrosis Clinical Specimens by Targeting the cpn60 and 16S rRNA Genes. *J. Clin. Microbiol.* 48, 1150–1160.

Pacheco, A. B. F., Guth, B. E. C., Soares, K. C. C., Nishimura, L., Almeida, D. F. D., and Ferreira, L. C. S. (1997). Random Amplification of Polymorphic DNA Reveals Serotype- Specific Clonal Clusters among Enterotoxigenic *Escherichia coli* Strains Isolated from Humans. *J. Clin. Microbiol*. Jun; 35(6): 1521-5.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinforma. Oxf. Engl.* 31, 3691–3693.

Palacio, F., Lewis, J. S., Sadkowski, L., Echevarria, K., and Jorgensen, J. H. (2011). Breakthrough Bacteremia and Septic Shock Due to *Streptococcus anginosus* Resistant to Daptomycin in a Patient Receiving Daptomycin Therapy. *Antimicrob. Agents Chemother.* 55, 3639–3640.

Pamer, E. G. (2016). Resurrecting the intestinal microbiota to combat antibiotic-resistant pathogens. *Science* 352, 535–538.

Panigrahi, P., Parida, S., Nanda, N. C., Satpathy, R., Pradhan, L., Chandel, D. S., et al. (2017). A randomized synbiotic trial to prevent sepsis among infants in rural India. *Nature* 548, 407–412.

Papanicolas, L. E., Choo, J. M., Wang, Y., Leong, L. E. X., Costello, S. P., Gordon, D. L., et al. (2019). Bacterial viability in faecal transplants: Which bacteria survive? *EBioMedicine* 41, 509–516.

Parkins, M. D., Sibley, C. D., Surette, M. G., and Rabin, H. R. (2008). The *Streptococcus* milleri group—An unrecognized cause of disease in cystic fibrosis: A case series and literature review. *Pediatr. Pulmonol.* 43, 490–497.

Peres, A. G., Stegen, C., Li, J., Xu, A. Q., Levast, B., Surette, M. G., et al. (2015). Uncoupling of Pro- and Anti-Inflammatory Properties of *Staphylococcus aureus*. *Infect. Immun.* 83, 1587–1597.

Petrof, E. O., Dhaliwal, R., Manzanares, W., Johnstone, J., Cook, D., and Heyland, D. K. (2012). Probiotics in the critically ill. *Crit. Care Med.* 40, 3290–3302.

Pirofski, L., and Casadevall, A. (2015). What is infectiveness and how is it involved in infection and immunity? *BMC Immunol.* 16.

Pittet, D., Allegranzi, B., Sax, H., Dharan, S., Pessoa-Silva, C. L., Donaldson, L., et al. (2006). Evidence-based model for hand transmission during patient care and the role of improved practices. *Lancet Infect. Dis.* 6, 641–652.

Planet, P. J., Rampersaud, R., Hymes, S. R., Whittier, S., Della-Latta, P. A., Narechania, A., et al. (2013). Genome Sequence of the Human Abscess Isolate *Streptococcus intermedius* BA1. *Genome Announc.* 1.

Plantinga, N. L., de Smet, A. M. G. A., Oostdijk, E. A. N., de Jonge, E., Camus, C., Krueger, W. A., et al. (2018). Selective digestive and oropharyngeal decontamination in medical and surgical ICU patients: individual patient data meta-analysis. *Clin. Microbiol. Infect.* 24, 505–513.

Plaza-Diaz, J., Ruiz-Ojeda, F. J., Gil-Campos, M., and Gil, A. (2019). Mechanisms of Action of Probiotics. *Adv. Nutr.* 10, S49–S66.

Poole, P. M., and Wilson, G. (1979). Occurrence and cultural features of *Streptococcus milleri* in various body sites. *J. Clin. Pathol.* 32, 764–768.

Porta, G., Rodríguez-Carballeira, M., Gómez, L., Salavert, M., Freixas, N., Xercavins, M., et al. (1998). Thoracic infection caused by *Streptococcus milleri*. *Eur. Respir. J.* 12, 357–362.

Potts, R. H. G. (2017). Investigating the gut microbiome of Generalized Anxiety Disorder, Major Depressive Disorder and Bipolar patients. [M.Sc. Thesis]

Prescott, H. C., Dickson, R. P., Rogers, M. A. M., Langa, K. M., and Iwashyna, T. J. (2015). Hospitalization Type and Subsequent Severe Sepsis. *Am. J. Respir. Crit. Care Med.* 192, 581–588.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650.

Psallidas, I., Corcoran, J., and Rahman, N. (2014). Management of Parapneumonic Effusions and Empyema. *Semin. Respir. Crit. Care Med.* 35, 715–722.

Pulliam, L., Porschen, R. K., and Hadley, W. K. (1980). Biochemical properties of CO2-dependent streptococci. *J. Clin. Microbiol.* 12, 27–31.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* 26, 841–842.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing Available at: http://www.R-project.org/.

Rahman, M., Nguyen, S. V., McCullor, K. A., King, C. J., Jorgensen, J. H., and McShan, W. M. (2015). Complete Genome Sequence of *Streptococcus anginosus* J4211, a Clinical Isolate. *Genome Announc.* 3.

Rahman, M., Nguyen, S. V., McCullor, K. A., King, C. J., Jorgensen, J. H., and McShan, W. M. (2016). Comparative Genome Analysis of the Daptomycin-Resistant *Streptococcus anginosus* Strain J4206 Associated with Breakthrough Bacteremia. *Genome Biol. Evol.* 8, 3446–3459.

Rakita, R. M., Vanek, N. N., Jacques-Palaz, K., Mee, M., Mariscalco, M. M., Dunny, G. M., et al. (1999). Enterococcus faecalis Bearing Aggregation Substance Is Resistant to Killing by Human Neutrophils despite Phagocytosis and Neutrophil Activation. *Infect. Immun.* 67, 6067–6075.

Rambault, A. (2009). *FigTree*. Available at: Computer program distributed by the author, website: http://tree.bio.ed.ac.uk/software/figtree/.

Ramirez, P., Bassi, G. L., and Torres, A. (2012). Measures to prevent nosocomial infections during mechanical ventilation. *Curr. Opin. Crit. Care* 18, 86–92.

Ravi, A., Halstead, F. D., Bamford, A., Casey, A., Thomson, N. M., van Schaik, W., et al. (2019). Loss of microbial diversity and pathogen domination of the gut microbiota in critically ill patients. *Microb. Genomics*. Sept; 5(9); e000293.

Rea, I. M., Gibson, D. S., McGilligan, V., McNerlan, S. E., Alexander, H. D., and Ross, O. A. (2018). Age and Age-Related Diseases: Role of Inflammation Triggers and Cytokines. *Front. Immunol.* 9, 1111e.

Reid, G. (2016). Probiotics: definition, scope and mechanisms of action. *Best Pract. Res. Clin. Gastroenterol.* 30, 17–25.

Reiss-Mandel, A., Rubin, C., Zayoud, M., Rahav, G., and Regev-Yochay, G. (2018). Staphylococcus aureus Colonization Induces Strain-Specific Suppression of Interleukin-17. *Infect. Immun.* 86, S344.

Reißmann, S., Friedrichs, C., Rajkumari, R., Itzek, A., Fulde, M., Rodloff, A. C., et al. (2010). Contribution of *Streptococcus anginosus* to Infections Caused by Groups C and G Streptococci, Southern India. *Emerg. Infect. Dis.* 16, 656–663.

Rennke, S., Wachter, R. M., and Sumant, R. (2013). "Chapter 27: Strategies To Prevent Stress-Related Gastrointestinal Bleeding (Stress Ulcer Prophylaxis): Brief Update Review," in *Making Health Care Safer II: An Updated Critical Analysis of the Evidence for Patient Safety Practices.*

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

Richards, M. J., Edwards, J. R., Culver, D. H., and Gaynes, R. P. (2000). Nosocomial infections in combined medical-surgical intensive care units in the United States. *Infect. Control Hosp. Epidemiol.* 21.

Rissman, A. I., Mau, B., Biehl, B. S., Darling, A. E., Glasner, J. D., and Perna, N. T. (2009). Reordering contigs of draft genomes using the Mauve aligner. *Bioinforma. Oxf. Engl.* 25, 2071–2073.

Roach, D. J., Burton, J. N., Lee, C., Stackhouse, B., Butler-Wu, S. M., Cookson, B. T., et al. (2015). A Year of Infection in the Intensive Care Unit: Prospective Whole Genome Sequencing of Bacterial Clinical Isolates Reveals Cryptic Transmissions and Novel Microbiota. *PLOS Genet.* 11, e1005413.

Rogers, G. B., Zain, N. M. M., Bruce, K. D., Burr, L. D., Chen, A. C., Rivett, D. W., et al. (2014). A Novel Microbiota Stratification System Predicts Future Exacerbations in Bronchiectasis. *Ann. Am. Thorac. Soc.* 11, 496–503.

Rogers, M. B., Aveson, V., Firek, B., Yeh, A., Brooks, B., Brower-Sinning, R., et al. (2017). Disturbances of the Perioperative Microbiome Across Multiple Body Sites in Patients Undergoing Pancreaticoduodenectomy: *Pancreas* 46, 260–267.

Rogers, M. B., Firek, B., Shi, M., Yeh, A., Brower-Sinning, R., Aveson, V., et al. (2016). Disruption of the microbiota across multiple body sites in critically ill children. *Microbiome* 4, 66.

Rooney, A. M., Timberlake, K., Brown, K. A., Bansal, S., Tomlinson, C., Lee, K.-S., et al. (2020). Each Additional Day of Antibiotics Is Associated With Lower Gut Anaerobes in Neonatal Intensive Care Unit Patients. *Clin. Infect. Dis.* 70, 2553–2560.

Rosen, R., Hu, L., Amirault, J., Khatwa, U., Ward, D. V., and Onderdonk, A. (2015). 16S Community Profiling Identifies Proton Pump Inhibitor Related Differences in Gastric, Lung, and Oropharyngeal Microflora. *J. Pediatr.* 166, 917–923.

Ruoff, K. L. (1988). *Streptococcus anginosus* ("*Streptococcus milleri*"): the unrecognized pathogen. *Clin. Microbiol. Rev.* 1, 102–108.

Sabino, Y. N. V., Santana, M. F., Oyama, L. B., Santos, F. G., Moreira, A. J. S., Huws, S. A., et al. (2019). Characterization of antibiotic resistance genes in the species of the rumen microbiota. *Nat. Commun.* 10, 5252.

Safdar, N., and Abad, C. (2008). Educational interventions for prevention of healthcare-associated infection: A systematic review: *Crit. Care Med.* 36, 933–940.

Safdar, N., Dezfulian, C., Collard, H. R., and Saint, S. (2005). Clinical and economic consequences of ventilator-associated pneumonia: A systematic review. *Crit. Care Med.* 33, 2184–2193.

Saint, S., Greene, M. T., Krein, S. L., Rogers, M. A. M., Ratz, D., Fowler, K. E., et al. (2016). A Program to Prevent Catheter-Associated Urinary Tract Infection in Acute Care. *N. Engl. J. Med.* 374, 2111–2119.

Salminen, M. K., Tynkkynen, S., Rautelin, H., Saxelin, M., Vaara, M., Ruutu, P., et al. (2002). *Lactobacillus* Bacteremia during a Rapid Increase in Probiotic Use of *Lactobacillus rhamnosus* GG in Finland. *Clin. Infect. Dis.* 35, 1155–1160.

Sassone-Corsi, M., and Raffatellu, M. (2015). No Vacancy: How Beneficial Microbes Cooperate with Immunity To Provide Colonization Resistance to Pathogens. *J. Immunol.* 194, 4081–4087.

Saxelin, M., Elo, S., Salminen, S., and Vapaatalo, H. (1991). Dose Response Colonisation of Faeces after Oral Administration of *Lactobacillus casei* Strain GG. *Microb. Ecol. Health Dis.* 4, 209–214.

Scales, B. S., Dickson, R. P., and Huffnagle, G. B. (2016). A tale of two sites: how inflammation can reshape the microbiomes of the gut and lungs. *J. Leukoc. Biol.* 100, 943–950.

Schirmer, M., Smeekens, S. P., Vlamakis, H., Jaeger, M., Oosting, M., Franzosa, E. A., et al. (2016). Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell*. Dec 15; 167(7): 1897.

Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012). PCR inhibitors - occurrence, properties and removal. *J. Appl. Microbiol.* 113, 1014–1026.

Schuijt, T. J., van der Poll, T., de Vos, W. M., and Wiersinga, W. J. (2013). The intestinal microbiota and host immune interactions in the critically ill. *Trends Microbiol.* 21, 221–9.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* 30, 2068–2069.

Segal, R., Pogoreliuk, I., Dan, M., Baumoehl, Y., and Leibovitz, a. (2006). Gastric microbiota in elderly patients fed via nasogastric tubes for prolonged periods. *J. Hosp. Infect.* 63, 79–83.

Segers, M. E., and Lebeer, S. (2014). Towards a better understanding of *Lactobacillus rhamnosus* GG - host interactions. *Microb. Cell Factories* 13, S7.

Sela, U., Euler, C. W., Correa da Rosa, J., and Fischetti, V. A. (2018). Strains of bacterial species induce a greatly varied acute adaptive immune response: The contribution of the accessory genome. *PLoS Pathog.* Jan 11;14(1): e1006726.

Sender, R., Fuchs, S., and Milo, R. (2016). Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell* 164, 337–340.

Sepsis Lung Microbiome Study Group (2020). Could lung bacterial dysbiosis predict ICU mortality in patients with extra-pulmonary sepsis? A proof-of-concept study. *Intensive Care Med.* Aug 7.

Shanahan, F., and Hill, C. (2019). Language, numeracy and logic in microbiome science. *Nat. Rev. Gastroenterol. Hepatol.* 16, 387–388.

Shea, P. R., Beres, S. B., Flores, A. R., Ewbank, A. L., Gonzalez-Lugo, J. H., Martagon-Rosado, A. J., et al. (2011). Distinct signatures of diversifying selection revealed by genome analysis of respiratory tract and invasive bacterial populations. *Proc. Natl. Acad. Sci.* 108, 5039–5044.

Sheppard, S. K., Guttman, D. S., and Fitzgerald, J. R. (2018). Population genomics of bacterial host adaptation. *Nat. Rev. Genet.* 19, 549–565.

Shimizu, K., Ogura, H., Asahara, T., Nomoto, K., Morotomi, M., Tasaki, O., et al. (2013). Probiotic/synbiotic therapy for treating critically ill patients from a gut microbiota perspective. *Dig. Dis. Sci.* 58, 23–32.

Shimizu, K., Ogura, H., Tomono, K., Tasaki, O., Asahara, T., and Nomoto, K. (2010). Patterns of gram-stained fecal Flora as a quick diagnostic marker in patients with severe SIRS. *Dig Sci Springer US* 56.

Shimizu, K., Yamada, T., Ogura, H., Mohri, T., Kiguchi, T., Fujimi, S., et al. (2018). Synbiotics modulate gut microbiota and reduce enteritis and ventilator-associated

pneumonia in patients with sepsis: a randomized controlled trial. *Crit. Care* 22, 239.

Shinzato, T., and Saito, A. (1995). The Streptococcus milleri Group as a Cause of Pulmonary Infections. *Clin. Infect. Dis.* 21, S238–S243.

Sibley, C. D., Church, D. L., Surette, M. G., Dowd, S. E., and Parkins, M. D. (2012). Pyrosequencing reveals the complex polymicrobial nature of invasive pyogenic infections: microbial constituents of empyema, liver abscess, and intracerebral abscess. *Eur. J. Clin. Microbiol. Infect. Dis.* 31, 2679–2691.

Sibley, C. D., Grinwis, M. E., Field, T. R., Eshaghurshan, C. S., Faria, M. M., Dowd, S. E., et al. (2011). Culture Enriched Molecular Profiling of the Cystic Fibrosis Airway Microbiome. *PLoS ONE* 6, e22702.

Sibley, C. D., Grinwis, M. E., Field, T. R., Parkins, M. D., Norgaard, J. C., Gregson, D. B., et al. (2010a). McKay agar enables routine quantification of the "Streptococcus milleri" group in cystic fibrosis patients. *J. Med. Microbiol.* 59, 534–540.

Sibley, C. D., Parkins, M. D., Rabin, H. R., Duan, K., Norgaard, J. C., and Surette, M. G. (2008). A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. *Proc. Natl. Acad. Sci.* 105, 15070–15075. d

Sibley, C. D., Sibley, K. A., Leong, T. A., Grinwis, M. E., Parkins, M. D., Rabin, H. R., et al. (2010b). The *Streptococcus milleri* Population of a Cystic Fibrosis Clinic Reveals Patient Specificity and Intraspecies Diversity. *J. Clin. Microbiol.* 48, 2592–2594.

Siegman-Igra, Y., Azmon, Y., and Schwartz, D. (2012). Milleri group streptococcus–a stepchild in the viridans family. *Eur. J. Clin. Microbiol. Infect. Dis.* 31, 2453–2459.

Silk, D. B. a., and Quinn, D. G. (2014). Dual-Purpose Gastric Decompression and Enteral Feeding Tubes Rationale and Design of Novel Nasogastric and Nasogastrojejunal Tubes. *J. Parenter. Enter. Nutr.* Jul; 39(5): 531-43.

Silvestri, L., de la Cal, M. A., and van Saene, H. K. F. (2012). Selective decontamination of the digestive tract: the mechanism of action is control of gut overgrowth. *Intensive Care Med.* 38, 1738–1750.

Skljarevski, S., Barner, A., and Bruno-Murtha, L. A. (2016). Preventing avoidable central line–associated bloodstream infections: Implications for probiotic administration and surveillance. *Am. J. Infect. Control* 44, 1427–1428.

Smith, A. R., Macfarlane, S., Furrie, E., Ahmed, S., Bahrami, B., Reynolds, N., et al. (2011). Microbiological and immunological effects of enteral feeding on the upper gastrointestinal tract. *J. Med. Microbiol.* 60, 359–365.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L. G., Gratadoux, J.-J., et al. (2008). *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U. S. A.* 105, 16731–16736.

Sommer, F., and Bäckhed, F. (2013). The gut microbiota–masters of host development and physiology. *Nat. Rev. Microbiol.* 11, 227–238.

Song, S., Guo, Y., Kim, J.-S., Wang, X., and Wood, T. K. (2019). Phages Mediate Bacterial Self-Recognition. *Cell Rep.* 27, 737-749.e4.

Soussan, R., Schimpf, C., Pilmis, B., Degroote, T., Tran, M., Bruel, C., et al. (2019). Ventilator-associated pneumonia: The central role of transcolonization. *J. Crit. Care* 50, 155–161.

Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat. Rev. Microbiol.* 9, 279–290.

Srinivasan, V., Metcalf, B. J., Knipe, K. M., Ouattara, M., McGee, L., Shewmaker, P. L., et al. (2014). *vanG* Element Insertions within a Conserved Chromosomal Site Conferring Vancomycin Resistance to *Streptococcus agalactiae* and *Streptococcus anginosus*. *mBio* 5, e01386-14.

Stearns, J. C., Davidson, C. J., McKeon, S., Whelan, F. J., Fontes, M. E., Schryvers, A. B., et al. (2015). Culture and molecular-based profiles show shifts in bacterial communities of the upper respiratory tract that occur with age. *ISME J.* 9, 1246–1259.

Su, M., Satola, S. W., and Read, T. D. (2018). Genome-Based Prediction of Bacterial Antibiotic Resistance. *J. Clin. Microbiol.* 57, e01405-18.

Suetens, C., Latour, K., Kärki, T., Ricchizzi, E., Kinross, P., Moro, M. L., et al. (2018). Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017. *Eurosurveillance* 23.

Suez, J., Zmora, N., Segal, E., and Elinav, E. (2019). The pros, cons, and many unknowns of probiotics. *Nat. Med.* 25, 716–729.

Sukeno, A., Nagamune, H., Whiley, R. A., Jafar, S. I., Aduse-Opoku, J., Ohkura, K., et al. (2005). Intermedilysin is essential for the invasion of hepatoma HepG2 cells by *Streptococcus intermedius*. *Microbiol. Immunol.* 49, 681–694.

Süßmuth, S. D., Muscholl-Silberhorn, A., Wirth, R., Susa, M., Marre, R., and Rozdzinski, E. (2000). Aggregation Substance Promotes Adherence, Phagocytosis, and Intracellular Survival of Enterococcus faecalis within Human Macrophages and Suppresses Respiratory Burst. *Infect. Immun.* 68, 4900–4906.

Suzuki, H., Hase, R., Otsuka, Y., and Hosokawa, N. (2016). Bloodstream infections caused by *Streptococcus* anginosus group bacteria: A retrospective analysis of 78 cases at a Japanese tertiary hospital. *J. Infect. Chemother.* 22, 456–460.

Sweere, J. M., Van Belleghem, J. D., Ishak, H., Bach, M. S., Popescu, M., Sunkari, V., et al. (2019). Bacteriophage trigger antiviral immunity and prevent clearance of bacterial infection. *Science*. Mar 29; 363(6434): eaat9691.

Takahashi, Y., Saito, A., Chiba, H., Kuronuma, K., Ikeda, K., Kobayashi, T., et al. (2018). Impaired diversity of the lung microbiome predicts progression of idiopathic pulmonary fibrosis. *Respir. Res.* 19, 34.

Taur, Y., Jenq, R. R., Perales, M.-A., Littmann, E. R., Morjaria, S., Ling, L., et al. (2014). The effects of intestinal tract bacterial diversity on mortality following allogeneic hematopoietic stem cell transplantation. *Blood* 124, 1174–1182.

Taur, Y., Xavier, J. B., Lipuma, L., Ubeda, C., Goldberg, J., Gobourne, A., et al. (2012). Intestinal Domination and the Risk of Bacteremia in Patients Undergoing Allogeneic Hematopoietic Stem Cell Transplantation. *Clin. Infect. Dis.* 55, 905–914.

Taylor, S. C., Laperriere, G., and Germain, H. (2017). Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci. Rep.* 7, 2409.

ter Horst, R., Jaeger, M., Smeekens, S. P., Oosting, M., Swertz, M. A., Li, Y., et al. (2016). Host and Environmental Factors Influencing Individual Human Cytokine Responses. *Cell* 167, 1111-1124.e13.

Tessier, L., Guilcher, S. J. T., Bai, Y. Q., Ng, R., and Wodchis, W. P. (2019). The impact of hospital harm on length of stay, costs of care and length of person-centred episodes of care: a retrospective cohort study. *Can. Med. Assoc. J.* 191, E879–E885.

The Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi:10.1038/nature11234.

Therneau, T. M., and Grambsch, P. M. (2000). *Modeling survival data: extending the Cox model*. New York: Springer.

Thomas, C. M., and Versalovic, J. (2010). Probiotics-host communication: Modulation of signaling pathways in the intestine. *Gut Microbes* 1, 148–163.

Thomas, J. P., Raine, T., Reddy, S., and Belteki, G. (2017). Probiotics for the prevention of necrotising enterocolitis in very low-birth-weight infants: a meta-analysis and systematic review. *Acta Paediatr.* 106, 1729–1741.

Thomas-White, K., Forster, S. C., Kumar, N., Van Kuiken, M., Putonti, C., Stares, M. D., et al. (2018). Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nat. Commun.* 9, 1557.

Tomoyasu, T., Matoba, M., Takao, A., Tabata, A., Whiley, R. A., Maeda, N., et al. (2018). Rapid screening method for detecting highly pathogenic *Streptococcus intermedius* strains carrying a mutation in the *lacR* gene. *FEMS Microbiol. Lett.* 365, 385.

Tomoyasu, T., Yamasaki, T., Chiba, S., Kusaka, S., Tabata, A., Whiley, R. A., et al. (2017). Positive- and Negative-Control Pathways by Blood Components for Intermedilysin Production in *Streptococcus intermedius*. *Infect. Immun.* 85, e00379–17.

Toscano, M., Vecchi, E. de, Rodighiero, V., and Drago, L. (2013). Microbiological and genetic identification of some probiotics proposed for medical use in 2011. *J. Chemother.* 25, 156–161.

Tremblay, D. M., and Moineau, S. (1999). Complete Genomic Sequence of the Lytic Bacteriophage DT1 of *Streptococcus thermophilus*. *Virology* 255, 63–76.

Trunk, T., Khalil, H. S., and Leo, J. C. (2018). Bacterial autoaggregation. *AIMS Microbiol.* 4, 140–164.

Tyler, A. D., Smith, M. I., and Silverberg, M. S. (2014). Analyzing the Human Microbiome: A “How To” guide for Physicians. *Am. J. Gastroenterol.*, 1–11.

Ubeda, C., and Pamer, E. G. (2012). Antibiotics, microbiota, and immune defense. *Trends Immunol.* 33, 459–466.

U.S Food and Drug Administration (2020). Safety Alert Regarding Use of Fecal Microbiota for Transplantation and Risk of Serious Adverse Events Likely Due to Transmission of Pathogenic Organisms. Available at: https://www.fda.gov/vaccines-blood-biologics/safety-availability-biologics/safety-

alert-regarding-use-fecal-microbiota-transplantation-and-risk-serious-adverse-events-likely.

Van den Abbeele, P., Roos, S., Eeckhaut, V., MacKenzie, D. A., Derde, M., Verstraete, W., et al. (2011). Incorporating a mucosal environment in a dynamic gut model results in a more representative colonization by lactobacilli. *Microb. Biotechnol.* 5, 106–115.

van der Ploeg, J. R. (2007). Genome sequence of *Streptococcus mutans* bacteriophage M102. *FEMS Microbiol. Lett.* 275, 130–138.

Verdu, E., Viani, F., Armstrong, D., Fraser, R., Siegrist, H. H., Pignatelli, B., et al. (1994). Effect of omeprazole on intragastric bacterial counts, nitrates, nitrites, and N-nitroso compounds. *Gut* 35, 455–460.

Videlock, E. J., and Cremonini, F. (2012). Meta-analysis: probiotics in antibiotic-associated diarrhoea. *Aliment. Pharmacol. Ther.* 35, 1355–1369.

Vincent, J.-L. (2003). Nosocomial infections in adult intensive-care units. *THE LANCET* 361, 10.

Vincent, J.-L. (2009). International Study of the Prevalence and Outcomes of Infection in Intensive Care Units. *JAMA* 302, 2323.

Vincent, J.-L. (2013). Critical care - where have we been and where are we going? *Crit. Care* 17, S2.

Vincent, J.-L., Sakr, Y., Singer, M., Martin-Loeches, I., Machado, F. R., Marshall, J. C., et al. (2020). Prevalence and Outcomes of Infection Among Patients in Intensive Care Units in 2017. *JAMA* 323, 1478.

Voigt, A. Y., Costea, P. I., Kultima, J. R., Li, S. S., Zeller, G., Sunagawa, S., et al. (2015). Temporal and technical variability of human gut metagenomes. *Genome Biol.* 16, 73.

Walter, J., Armet, A. M., Finlay, B. B., and Shanahan, F. (2020). Establishing or Exaggerating Causality for the Gut Microbiome: Lessons from Human Microbiota-Associated Rodents. *Cell* 180, 221–232.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., et al. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* 42, D581–91.

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., et al. (2017). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* 45, D535–D542.

Wei, T., and Simko, V. (2017). *R package "corrplot": Visualization of a Correlation Matrix*. Available at: https://github.com/taiyun/corrplot.

Wei, Y., Yang, J., Wang, J., Yang, Y., Huang, J., Gong, H., et al. (2016). Successful treatment with fecal microbiota transplantation in patients with multiple organ dysfunction syndrome and diarrhea following severe sepsis. *Crit. Care* 20, 332.

Weightman, N. C., Barnham, M. R. D., and Dove, M. (2004). Streptococcus milleri group bacteraemia in North Yorkshire, England (1989-2000). *Indian J. Med. Res.* 119 Suppl, 164–167.

Weil, M. H., and Tang, W. (2011). From Intensive Care to Critical Care Medicine: A Historical Perspective. *Am. J. Respir. Crit. Care Med.* 183, 1451–1453.

Weng, H., Li, J.-G., Mao, Z., Feng, Y., Wang, C.-Y., Ren, X.-Q., et al. (2017). Probiotics for Preventing Ventilator-Associated Pneumonia in Mechanically Ventilated Patients: A Meta-Analysis with Trial Sequential Analysis. *Front. Pharmacol.* 8, 717.

Whelan, F. J., and Surette, M. G. (2017). A comprehensive evaluation of the sl1p pipeline for 16S rRNA gene sequencing analysis. *Microbiome BioMed Cent.* 5.

Whelan, F. J., Verschoor, C. P., Stearns, J. C., Rossi, L., Luinstra, K., Loeb, M., et al. (2014). The Loss of Topography in the Microbial Communities of the Upper Respiratory Tract in the Elderly. *Ann. Am. Thorac. Soc.* 11, 513–521.

Whelan, F. J., Waddell, B., Syed, S. A., Shekarriz, S., Rabin, H. R., Parkins, M. D., et al. (2020). Culture-enriched metagenomic sequencing enables in-depth profiling of the cystic fibrosis lung microbiota. *Nat. Microbiol.* Feb; 5(2): 379-390.

Whiley, R. A., and Beighton, D. (1991). Emended Descriptions and Recognition of *Streptococcus constellatus*, *Streptococcus intermedius*, and *Streptococcus anginosus* as Distinct Species. *Int. J. Syst. Bacteriol.* 41, 1–5.

Whiley, R. A., Beighton, D., Winstanley, T. G., Fraser, H. Y., and Hardie, J. M. (1992). *Streptococcus intermedius*, *Streptococcus constellatus*, and *Streptococcus anginosus* (the Streptococcus milleri group): association with different body sites and clinical infections. *J. Clin. Microbiol.* 30, 243–244.

Whiley, R. A., Fraser, H., Hardie, J. M., and Beighton, D. (1990). Phenotypic differentiation of *Streptococcus intermedius*, *Streptococcus constellatus*, and

Streptococcus anginosus strains within the "Streptococcus milleri group". *J. Clin. Microbiol.* 28, 1497–1501.

Whiley, R. A., Hall, L. M. C., Hardie, J. M., and Beighton, D. (1999). A study of small-colony, β-haemolytic, Lancefield group C streptococci within the anginosus group: description of *Streptococcus constellatus* subsp. pharyngis subsp. nov., associated with the human throat and pharyngitis. *Int. J. Syst. Evol. Microbiol.* 49, 1443–1449.

Whiley, R. A., and Hardie, J. M. (1989). DNA-DNA Hybridization Studies and Phenotypic Characteristics of Strains within the 'Streptococcus milleri Group.' *J. Gen. Microbiol.* 135, 2623–2633.

Whitney, J. C., Colvin, K. M., Marmont, L. S., Robinson, H., Parsek, M. R., and Howell, P. L. (2012). Structure of the Cytoplasmic Region of PelD, a Degenerate Diguanylate Cyclase Receptor That Regulates Exopolysaccharide Production in *Pseudomonas aeruginosa*. *J. Biol. Chem.* 287, 23582–23593.

Whitney, J. C., Peterson, S. B., Kim, J., Pazos, M., Verster, A. J., Radey, M. C., et al. (2017). A broadly distributed toxin family mediates contact-dependent antagonism between gram-positive bacteria. *eLife* 6, e26938.

Wickham Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York Available at: https://ggplot2.tidyverse.org.

Wickham, S. E., Hotze, E. M., Farrand, A. J., Polekhina, G., Nero, T. L., Tomlinson, S., et al. (2011). Mapping the Intermedilysin-Human CD59 Receptor Interface Reveals a Deep Correspondence with the Binding Site on CD59 for Complement Binding Proteins C8 and C9. *J. Biol. Chem.* 286, 20952–20962.

Willing, B. P., Russell, S. L., and Finlay, B. B. (2011). Shifting the balance: antibiotic effects on host-microbiota mutualism. *Nat. Rev. Microbiol.* 9, 233–243.

Wood, D. E., and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15, R46.

World Health Organization (2011). Report on the burden of endemic health care-associated infection worldwide. Geneva Available at: http://apps.who.int/iris/bitstream/10665/80135/1/9789241501507_ eng.pdf.

World Health Organization (2018a). Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva.

World Health Organization (2018b). Global Health Estimates 2016-2060 Summary Tables: Projection of Deaths by Cause, Age, Sex, by World Bank Income Group. Geneva, Switzerland.

Wrightson, J., and Davies, R. (2010). The Approach to the Patient with a Parapneumonic Effusion. *Semin. Respir. Crit. Care Med.* 31, 706–715.

Wu, W. M., Yang, Y. S., and Peng, L. H. (2014). Microbiota in the stomach: new insights. *J. Dig. Dis.* 15, 54–61.

Wunderink, R. G., Srinivasan, A., Barie, P. S., Chastre, J., Dela Cruz, C. S., Douglas, I. S., et al. (2020). Antibiotic Stewardship in the Intensive Care Unit. An Official American Thoracic Society Workshop Report in Collaboration with the AACN, CHEST, CDC, and SCCM. *Ann. Am. Thorac. Soc.* 17, 531–540.

Yamada, T., Shimizu, K., Ogura, H., Asahara, T., Nomoto, K., Yamakawa, K., et al. (2015). Rapid and Sustained Long-Term Decrease of Fecal Short-Chain Fatty Acids in Critically Ill Patients With Systemic Inflammatory Response Syndrome. *J. Parenter. Enter. Nutr.* 39, 569–577.

Yamazaki, A., Ito, A., Ishida, T., and Washio, Y. (2019). Polymicrobial etiology as a prognostic factor for empyema in addition to the renal, age, purulence, infection source, and dietary factors score. *Respir. Investig.* 57, 574–581.

Ye, Y. (2010). Identification and quantification of abundant species from pyrosequences of 16S rRNA by consensus alignment. in *2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Hong Kong, China: IEEE), 153–157.

Yeh, A., Rogers, M. B., Firek, B., Neal, M. D., Zuckerbraun, B. S., and Morowitz, M. J. (2016). Dysbiosis Across Multiple Body Sites in Critically Ill Adult Surgical Patients: *SHOCK* 46, 649–654.

Yelin, I., Flett, K. B., Merakou, C., Mehrotra, P., Stam, J., Snesrud, E., et al. (2019). Genomic and epidemiological evidence of bacterial transmission from probiotic capsule to blood in ICU patients. *Nat. Med.* Nov; 25(11): 1728-1732.

Yokoe, D. S., Anderson, D. J., Berenholtz, S. M., Calfee, D. P., Dubberke, E. R., Ellingson, K. D., et al. (2014). A Compendium of Strategies to Prevent Healthcare-Associated Infections in Acute Care Hospitals: 2014 Updates. *Infect. Control Hosp. Epidemiol.* 35, 967–977.

Zaborin, A., Smith, D., Garfield, K., Quensen, J., Shakhsheer, B., Kade, M., et al. (2014). Membership and Behavior of Ultra-Low-Diversity Pathogen Communities Present in the Gut of Humans during Prolonged Critical Illness. *mBio* 5, e01361-14.

Zakharkina, T., Martin-Loeches, I., Matamoros, S., Povoa, P., Torres, A., Kastelijn, J. B., et al. (2017). The dynamics of the pulmonary microbiome during mechanical ventilation in the intensive care unit and the association with occurrence of pneumonia. *Thorax* 72, 803–810.

Zimlichman, E., Henderson, D., Tamir, O., Franz, C., Song, P., Yamin, C. K., et al. (2013). Health Care–Associated Infections: A Meta-analysis of Costs and Financial Impact on the US Health Care System. *JAMA Intern. Med.* 173, 2039.

Zuñiga-Bahamon, A., Tobar-Tosse, F., Guillermo-Ortega, J., Wibberg, D., and Tauch, A. (2016). Draft Genome Sequence of *Streptococcus anginosus* BVI, a New Vaginal Pathogen Candidate. *Genome Announc.* 4, e01417-16, /ga/4/6/e01417-16.atom.

Appendix A

Appendix to Chapter 2: Supplementary Material

Supplementary Methods

Subject recruitment

Patients included in this study were ≥18 years of age and were expected to need mechanical ventilation for at least 72 hours. Exclusion criteria were immunocompromised patients (HIV < 200 CD4 T cells/µl, chronic immunosuppressive medications, prior organ or haematological transplant, absolute neutrophil count <500 cells/µl), patients with an increased risk of endovascular infections, patients with gastroesophageal or intestinal injury or recent surgery of the oesophagus, stomach, bowel, hepatobiliary tree, spleen or pancreas in the prior 72 hours, patients with suspected or documented ischemia, or severe acute pancreatitis, patients unable to receive enteral medications, pregnant patients, those undergoing life-support withdrawal, or patients who had been mechanically ventilated for more than 72 hours. Patients enrolled in PROSPECT were administered the study product (probiotic or placebo) twice per day via the nasogastric tube and clinical parameters and demographic data were collected.

DNA extraction and 16S rRNA gene sequencing

The genomic DNA extraction was performed as described in Stearns J.C., *et al.,* 2015 (Stearns et al., 2015). Briefly, 300 µl of ETA, GA or 0.1 g of fecal sample were added to a lysis buffer (800 µl of 200 mM of NaH_2PO_4 pH 8 and 100 µl of guanidine thyocyanate-EDTA-*N*-Lauroylsarcosine) and mechanically homogenized with 0.2 g of 0.1 mm glass beads for 3 min at 3000 rpm for ETA and GA and with 0.2 g of 2.8 mm and 0.2 g of 0.1 mm glass beads (MoBio, Carlsband, CA, USA) for 2 cycles of 3 min at 3000 rpm for

fecal samples. Samples were then incubated for 1 h at 37 ˚C with 50 µL lysozyme (100 mg/ml), 50 µL mutanolysin (10 U/µl) and 10 µl RNase A (10 mg/ml) followed by 1.5 h incubation at 65 ˚C with the addition of 25 µL proteinase K (20 mg/ml), 25 µL sodium dodecyl sulfate (25%) and 75 µL NaCl (5 M). The supernatant was collected, and the DNA was subsequently extracted via a phenol/chloroform extraction and purified using a Zymo Research DNA Clean and Concentrator™-25 columns (Cedarlane, Burlington, ON, Canada). The DNA was stored at -20 ˚C until further use.

Variable region 3 of the 16S rRNA gene was amplified as previously described with modest modifications (Bartram et al., 2011). Briefly, PCRs were prepared in a 50 µl reaction mixture including 200 µM of each dNTP, 1.5 mM of $MgCl_2$, 5 pmol of barcoded primers and 2.5 U of Taq polymerase. Subsequently, the mixture was divided into three reactions and the PCR amplification was carried out. The PCR conditions included an initial denaturation at 94˚C for 2 min followed by 30 cycles of 94˚C for 30 sec, 50˚C for 30 sec, 72˚C for 30 sec and finished by a final extension at 72˚C for 10 min. PCR amplicons were sequenced via Illumina MiSeq Personal Sequencer system by the McMaster Genomics Facility (Hamilton, Ontario, Canada).

Sequence processing

The sequencing data were processed through a standardized pipeline (Whelan and Surette, 2017). Briefly, the sequences were trimmed of the forward and reverse primers using Cutadapt (Martin, 2011) and the paired-end reads were subsequently aligned using PANDAseq (Masella et al., 2012). The alignment primers were then removed with

Cutadapt (Martin, 2011) and low-quality reads were removed using sickle with a quality threshold of 30 (https://github.com/najoshi/sickle). Next, reads were clustered in operational taxonomic units (OTUs) using the program AbundantOTU+ (Ye, 2010) based on a 97% similarity threshold. Finally, taxonomy was assigned using the Ribosomal Database Project (RDP) classifier (Wang et al., 2007) against the 2011 Greengenes reference database to the genus level (DeSantis et al., 2006) using Quantitative Insights Into Microbial Ecology (QIIME; (Caporaso et al., 2010)).

Measurement of serum cytokines

Venous blood was collected and centrifuged at 1.5 g for 10 min at 25°C and serum was stored at -140°C until processed. Serum C-reactive protein was measured by sandwich ELISA using monoclonal capture and detection antibodies (clone 5 - #ab8279, clone 6 - #ab24462; Abcam INC., Ontario, Canada) and purified human CRP as a standard (Aviva Systems Biology Corp., California, USA). Remaining serum cytokines were measured by Milliplex MAP Human High Sensitivity T Cell kit (Millipore, Ontario, Canada) as per manufacturer's protocol. Measurements were performed in duplicate and the average value was reported.

Quantification of bacterial load

Quantification of bacterial 16S rRNA encoding genes was performed by quantitative polymerase chain reaction (qPCR) on a Bio-Rad CFX96 thermocycler (Bio-Rad, Ontario, Canada) as previously described (De Gregoris et al., 2011). Briefly, reactions were

carried in a 20 µL mixture including 40 ng of template, 10 pmol of each primers, 10 µg of bovine serum albumin and SsoFast EvaGreen supermix (Bio-Rad, Ontario, Canada). The qPCR program included an initial denaturation at 98˚C for 2 min followed by 40 cycles of 98˚C for 5 sec and 60˚C for 30 sec followed by the generation of the melt curve with an 0.5°C increments for 5 seconds from 65°C up to 95°C. Reactions were performed in triplicates.

Sensitivity analysis

OTU sequences assigned to the *Lactobacillus* genus were aligned to the *L. rhamnosus* GG 16S rRNA gene sequence using the software Geneious version 7.1.5 (Kearse et al., 2012). This was used to identify the OTU associated with the probiotic administered to patients in the probiotic group of the PROSPECT trial. To reduce downstream bias of this OTU it was removed *in silico* and the microbiota analysis was repeated.

Supplementary Results

Sensitivity analysis

Once OTU5 was removed for the sensitivity analysis, two GA samples had less than 2800 reads and were excluded from the subsequent analysis. Our main conclusions (the loss of biogeographical distinction between anatomical sites (Figure A.9), an inverse correlation between ETA α diversity and APACHE II score (Figure A.10), decreased ETA α diversity in patients who died in hospital (Figure A.11), and the tendency of patients to have worse prognosis when they had lower respiratory microbial diversity (Figure A.12)

were all unaffected by the removal of OTU5. The primary difference between the two analyses lay in the significant compositional differences between stool samples from healthy controls and ICU patients, where the abundance of OTU5 was significantly increased within the ICU population (Table A.4). However, we recognize that we did not eliminate or correct for the indirect impact that the probiotics administration could have systemically.

Supplementary Tables

Table A.1: ICU samples collected and additional information concerning the patients included in the study

Patient ID	Samples Collection day			Status at discharge	ICU LOS‡	Global LOS§	APACHE II
	ETA*	GA †	Stool				
11001			3	Deceased	16	332	39
11002	6	4		Alive	30	52	28
11003	5	3	5	Deceased	18	18	43
11004	2		3	Deceased	6	75	33
11005	3	4		Deceased	10	42	21
11008	6		7	Deceased	11	14	24
11009	1	2		Alive	3	14	22
11010	3	3		Alive	7	12	26
11011		3	7	Alive	54	63	19
11012		5		Alive	5	12	19
11014		4	4	Deceased	4	4	26
11015	3	2		Alive	18	49	35
11016	1	1		Alive	7	15	19
11017	2	2		Alive	4	28	23
11018	4	2		Alive	21	46	29
11019	1			Alive	4	13	21
11020		2		Alive	5	5	6
11022	3	6		Deceased	15	15	46
11023	4		6	Alive	12	48	16
11024	3		6	Alive	12	172	19
11025	7	4	7	Deceased	34	43	23
11026	3	3	6	Alive	7	50	30
11027	2	2		Alive	22	29	14
21001	2			Alive	6	9	17
21002	3	3		Deceased	86	86	31
21003	2	2		Alive	7	7	25
21004	4	4		Deceased	9	33	32
21005	5	5		Deceased	73	92	21
21006	7	4		Alive	60	60	33
21008	2	2		Deceased	7	7	23
21009	3			Alive	31	31	19
21010	3	3		Alive	7	16	21

Patient ID	Samples Collection day			Status at discharge	ICU LOS‡	Global LOS§	APACHE II
	ETA*	GA †	Stool				
21011	1	4		Alive	19	28	37
21012	2	4		Alive	15	21	27

*ETA: endotracheal tube aspirate †GA: gastric tube aspirate ‡LOS: Length of stay §Global LOS: represents the days between ICU admission and hospital discharge

Table A.2: Samples collected from healthy donors

Specimen's type	Donors (n)	Sample's ID	Age	Sex	Reference
BAL	7	BAL1	NA	NA	(Baatjes et al., 2015)
		BAL2	NA	NA	
		BAL3	NA	NA	
		BAL4	NA	NA	
		BAL5	NA	NA	
		BAL6	NA	NA	
		BAL7	NA	NA	
NP swab	7	NP1	NA	Female	(Stearns et al., 2015)
		NP2	NA	Female	
		NP3	NA	NA	
		NP4	NA	NA	
		NP5	NA	Male	
		NP6	NA	Male	
		NP7	NA	Male	
OP swab	7	OP1	NA	Female	
		OP2	NA	Female	
		OP3	NA	NA	
		OP4	NA	NA	
		OP5	NA	Male	
		OP6	NA	Male	
		OP7	NA	Male	
Stool	21	S1	NA	NA	(Moayyedi et al., 2015)
		S2	NA	NA	
		S3	NA	NA	
		S4	NA	NA	
		S5	NA	NA	
		S6	NA	NA	
		S7	40	Male	(Potts, 2017)
		S8	25	Male	
		S9	39	Female	
		S10	55	Female	
		S11	40	Male	
		S12	29	Male	
		S13	20	Female	
		S14	23	Male	

Specimen's type	Donors (n)	Sample's ID	Age	Sex	Reference
Stool	21	S15	60	Female	(Potts, 2017)
		S16	54	Female	
		S17	29	Female	
		S18	39	Male	
		S19	64	Male	
		S20	39	Female	
		S21	24	Female	

Table A.3: OTUs significantly different in the lower respiratory tract between healthy donors and ICU patients

Taxonomy			Relative abundance			*p* value (after FDR)
Order	Family	Genus	Healthy	ICU	Δ	
Neisseriales	Neisseriaceae	*Neisseria*	36.26	1.30	34.95	0.001
Clostridiales	Veillonellaceae	*Veillonella*	3.34	0.18	3.16	0.001
Lactobacillales	Streptococcaceae	*Streptococcus*	2.81	0.11	2.70	< 0.001
Bacillales	Staphylococcaceae	*Staphylococcus*	2.40	0.08	2.32	< 0.001
Actinomycetales	Corynebacteriaceae	*Corynebacterium*	2.23	0.03	2.21	< 0.001
Clostridiales	Veillonellaceae	*Megasphaera*	1.44	0.11	1.33	< 0.001
Pasteurellales	Pasteurellaceae	*Haemophilus*	0.88	0.18	0.70	0.003
Caulobacterales	Caulobacteraceae	*Caulobacter*	0.67	0.01	0.67	< 0.001
Clostridiales	Veillonellaceae	*Veillonella*	0.61	0.04	0.56	< 0.001
Clostridiales	Veillonellaceae	*Selenomonas*	0.36	0.06	0.29	0.001
Fusobacteriales	Fusobacteriaceae	*Fusobacterium*	0.22	0.00	0.22	0.003
Bacteroidales	Prevotellaceae	*Prevotella*	0.19	0.00	0.19	< 0.001
Enterobacteriales	Enterobacteriaceae	*Proteus*	0.19	0.00	0.19	< 0.001
Bacteroidales	Prevotellaceae	*Prevotella*	0.18	0.01	0.17	0.008
Actinomycetales	Actinomycetaceae	*Actinomyces*	0.20	0.06	0.14	0.031
Sphingomonadale	Sphingomonadaceae		0.14	0.02	0.12	0.003
Burkholderiales	Comamonadaceae	*Caldimonas*	0.12	0.00	0.12	< 0.001
Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	0.10	0.00	0.10	< 0.001
Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	0.09	0.00	0.09	0.003
Clostridiales	ClostridialesFamilyX	*Eubacterium*	0.10	0.02	0.08	0.012
Actinomycetales			0.14	0.07	0.08	0.006
Clostridiales	Veillonellaceae	*Veillonella*	0.08	0.00	0.08	0.002
Clostridiales	Veillonellaceae		0.07	0.00	0.07	0.028
Fusobacteriales	Fusobacteriaceae	*Fusobacterium*	0.07	0.01	0.06	0.045
Sphingomonadale	Sphingomonadaceae	*Novosphingobium*	0.06	0.00	0.06	< 0.001
Clostridiales	Veillonellaceae	*Selenomonas*	0.08	0.03	0.05	0.035
Erysipelotrichales	Erysipelotrichaceae	*Bulleidia*	0.11	0.07	0.03	0.040
Neisseriales	Neisseriaceae	*Neisseria*	0.08	0.05	0.03	0.040
Fusobacteriales	Fusobacteriaceae	*Fusobacterium*	1.09	1.07	0.02	0.013
Lactobacillales	Streptococcaceae	*Streptococcus*	1.00	2.05	-1.05	0.048
Mycoplasmatales	Mycoplasmataceae	*Mycoplasma*	0.00	0.26	-0.26	0.012
Clostridiales	Peptostreptococcacea	*Peptostreptococcus*	0.42	0.56	-0.14	0.029
Clostridiales	Veillonellaceae	*Selenomonas*	0.03	0.05	-0.02	0.005
Neisseriales	Neisseriaceae	*Neisseria*	0.16	0.17	-0.02	0.044

Table A.4: OTUs that are significantly different in stool between healthy donors and ICU patients

Taxonomy			Relative abundance			*p* value (after FDR)
Order	Family	Genus	Healthy	ICU	Δ	
Clostridiales	Lachnospiraceae		28.71	6.83	21.88	0.003
Clostridiales	Ruminococcaceae	*Faecalibacterium*	8.41	1.53	6.88	0.005
Clostridiales	Lachnospiraceae	*Blautia*	7.95	2.60	5.35	0.030
Clostridiales	Lachnospiraceae		2.81	0.13	2.68	0.003
Clostridiales	Ruminococcaceae	*Subdoligranulum*	1.19	0.07	1.12	0.004
Clostridiales	Lachnospiraceae	*Lachnobacterium*	1.02	0.01	1.01	0.002
Clostridiales	Ruminococcaceae	*Oscillospira*	0.99	0.03	0.96	0.004
Coriobacteriales	Coriobacteriaceae	*Collinsella*	0.37	0.02	0.35	0.013
Clostridiales			0.40	0.06	0.34	0.045
Clostridiales	Lachnospiraceae	*Shuttleworthia*	0.33	0.01	0.32	0.042
Clostridiales	Ruminococcaceae	*Clostridium*	0.28	0.00	0.28	0.013
Clostridiales	Ruminococcaceae	*Oscillospira*	0.30	0.04	0.26	0.042
Erysipelotrichales	Erysipelotrichaceae	*Clostridium*	0.26	0.00	0.26	0.013
Clostridiales			0.25	0.00	0.25	0.003
Clostridiales	Lachnospiraceae		0.31	0.07	0.24	0.005
Clostridiales			0.24	0.03	0.21	0.004
Clostridiales			0.19	0.00	0.19	0.030
Clostridiales	Lachnospiraceae	*Anaerostipes*	0.21	0.04	0.17	0.029
Clostridiales			0.16	0.00	0.16	0.023
Clostridiales	Ruminococcaceae		0.15	0.03	0.13	0.021
Clostridiales	Lachnospiraceae	*Ruminococcus*	0.11	0.01	0.10	0.020
Lactobacillales	Enterococcaceae	*Enterococcus*	0.00	23.39	-23.39	0.002
Lactobacillales	Lactobacillaceae	*Lactobacillus*	0.00	7.87	-7.87	0.000
Erysipelotrichales	Erysipelotrichaceae		0.01	0.54	-0.53	0.025
Bacillales	Staphylococcaceae	*Staphylococcus*	0.00	0.53	-0.53	0.004
Erysipelotrichales	Erysipelotrichaceae	*Clostridium*	0.01	0.39	-0.38	0.020
Lactobacillales	Enterococcaceae	*Enterococcus*	0.00	0.11	-0.11	0.030
Bacteroidales	Porphyromonadaceae	*Parabacteroides*	0.00	0.05	-0.05	0.030
Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	0.00	0.04	-0.03	0.032

Table A.5: Correlation matrix results using Spearman rank coefficient correlation between metadata and α diversity metrics of ETAs

Variable 1	Variable 2	β coefficient	*p* value	
			No FDR	With FDR
Age	Collection time	-0.12	0.535	0.730
Age	Leukocytes count	0.46	0.015	0.068
Age	Neutrophils count	0.36	0.062	0.216
APACHE II score	Age	0.34	0.072	0.229
APACHE II score	Collection time	0.22	0.248	0.486
APACHE II score	Leukocytes count	0.22	0.253	0.455
APACHE II score	Neutrophils count	0.26	0.185	0.396
Collection time	Leukocytes count	-0.24	0.210	0.430
Collection time	Neutrophils count	-0.12	0.563	0.745
Global Length of stay	Age	0.21	0.264	0.496
Global Length of stay	APACHE II score	0.15	0.427	0.686
Global Length of stay	Collection time	0.48	0.008	0.053
Global Length of stay	ICU Length of stay	0.57	0.001	0.009
Global Length of stay	Leukocytes count	-0.03	0.697	0.829
Global Length of stay	Neutrophils count	0.00	0.998	0.998
Global Length of stay	Observed Species	-0.19	0.319	0.553
Global Length of stay	Shannon diversity	-0.30	0.113	0.283
Global Length of stay	Simpson diversity	-0.28	0.148	0.351
ICU Length of stay	Age	-0.08	0.674	0.843
ICU Length of stay	APACHE II score	0.21	0.284	0.492
ICU Length of stay	Collection time	0.61	< 0.001	0.004
ICU Length of stay	Leukocytes count	-0.04	0.829	0.949
ICU Length of stay	Neutrophils count	0.08	0.710	0.860
Neutrophils count	Leukocytes count	0.96	< 0.001	< 0.001
Observed Species	Age	-0.14	0.471	0.687
Observed Species	APACHE II score	-0.31	0.107	0.282
Observed Species	Collection time	0.04	0.843	0.949
Observed Species	ICU Length of stay	-0.23	0.220	0.430
Observed Species	Leukocytes count	0.02	0.928	0.989
Observed Species	Neutrophils count	-0.07	0.726	0.860
Observed Species	Shannon diversity	0.85	< 0.001	< 0.001
Observed Species	Simpson diversity	0.77	< 0.001	< 0.001
Shannon diversity	Age	-0.12	0.523	0.730
Shannon diversity	APACHE II score	-0.46	0.013	0.082

Table continues next page

Variable 1	Variable 2	β coefficient	*p* value	
			No FDR	With FDR
Shannon diversity	Collection time	-0.26	0.174	0.412
Shannon diversity	ICU Length of stay	-0.38	0.045	0.167
Shannon diversity	Leukocytes count	-0.03	0.897	0.982
Shannon diversity	Neutrophils count	-0.14	0.471	0.687
Simpson diversity	Age	-0.10	0.624	0.802
Simpson diversity	APACHE II score	-0.44	0.017	0.078
Simpson diversity	Collection time	-0.33	0.076	0.229
Simpson diversity	ICU Length of stay	-0.38	0.041	0.167
Simpson diversity	Leukocytes count	-0.01	0.953	0.982
Simpson diversity	Neutrophils count	-0.14	0.473	0.687
Simpson diversity	Shannon diversity	0.98	< 0.001	< 0.001

Global length of stay is the number of days between ICU admission and hospital discharge
FDR: False discovery rate

Table A.6: Compositional differences between respiratory specimens of patients deceased versus discharged alive from the hospital.

Taxonomy				Deceased	Discharged alive	Delta	p value	p value after FDR
Class	Order	Family	Genus					
Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	12.20%	4.59%	7.61%	0.016	0.667
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.00%	0.63%	-0.63%	0.032	0.667
Bacilli	Gemellales	Gemellaceae	*Gemella*	0.00%	0.22%	-0.22%	0.037	0.667
Actinobacteria	Coriobacteriales	Coriobacteriaceae	*Atopobium*	0.04%	0.58%	-0.53%	0.040	0.667
Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae		0.03%	0.01%	0.02%	0.047	0.667
Betaproteobacteria	Burkholderiales	Burkholderiaceae	*Burkholderia*	16.75%	0.00%	16.75%	0.047	0.667
Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Scardovia*	0.04%	0.00%	0.04%	0.047	0.667
Clostridia	Clostridiales	Veillonellaceae	*Veillonella*	0.09%	0.68%	-0.59%	0.050	0.667
Bacteroidia	Bacteroidales	Porphyromonadaceae	*Tannerella*	0.00%	0.02%	-0.02%	0.051	0.667
				0.00%	0.02%	-0.02%	0.051	0.667
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.00%	0.08%	-0.08%	0.052	0.667
Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	*Bulleidia*	0.00%	0.11%	-0.11%	0.057	0.667
TM7-3	EW055			0.00%	0.02%	-0.02%	0.081	0.667
				0.00%	0.02%	-0.02%	0.081	0.667
Actinobacteria	Actinomycetales	Actinomycetaceae	*Actinomyces*	0.00%	0.02%	-0.02%	0.081	0.667
Clostridia	Clostridiales	Veillonellaceae	*Selenomonas*	0.00%	0.13%	-0.13%	0.081	0.667
Bacteroidia	Bacteroidales			0.01%	0.01%	0.00%	0.092	0.667
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.02%	0.28%	-0.26%	0.107	0.667
Clostridia	Clostridiales	Ruminococcaceae		0.00%	0.04%	-0.03%	0.108	0.667
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.02%	0.61%	-0.59%	0.115	0.667
Actinobacteria	Actinomycetales	Actinomycetaceae	*Actinomyces*	0.03%	0.42%	-0.39%	0.119	0.667

Spirochaetes	Spirochaetales	Spirochaetaceae	*Spirochaeta*	0.00%	0.06%	-0.06%	0.126	0.667
Actinobacteria	Coriobacteriales	Coriobacteriaceae		0.00%	0.05%	-0.05%	0.126	0.667
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	0.00%	0.26%	-0.26%	0.126	0.667
Clostridia	Clostridiales	Veillonellaceae	*Anaeroglobus*	0.00%	0.03%	-0.03%	0.126	0.667
Flavobacteria	Flavobacteriales	Flavobacteriaceae	*Capnocytophaga*	0.00%	0.04%	-0.04%	0.126	0.667
Flavobacteria	Flavobacteriales	Flavobacteriaceae	*Capnocytophaga*	0.00%	0.02%	-0.02%	0.126	0.667
Clostridia	Clostridiales			0.00%	0.12%	-0.12%	0.126	0.667
Bacilli	Bacillales	Staphylococcaceae	*Staphylococcus*	0.03%	0.01%	0.02%	0.133	0.667
Clostridia	Clostridiales	Lachnospiraceae		0.02%	0.15%	-0.13%	0.140	0.667
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.08%	2.69%	-2.62%	0.143	0.667
Chloroplast	Streptophyta			0.14%	0.03%	0.11%	0.143	0.667
Gammaproteobacteria	Pasteurellales	Pasteurellaceae	*Haemophilus*	0.01%	0.17%	-0.15%	0.145	0.667
Bacilli	Lactobacillales	Carnobacteriaceae	*Granulicatella*	0.13%	1.36%	-1.23%	0.149	0.667
Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	*Campylobacter*	0.04%	0.22%	-0.18%	0.150	0.667
Actinobacteria	Actinomycetales			0.00%	0.10%	-0.10%	0.151	0.667
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	6.06%	11.56%	-5.50%	0.154	0.667
Actinobacteria	Actinomycetales	Corynebacteriaceae	*Corynebacterium*	0.08%	0.00%	0.08%	0.168	0.667
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	0.03%	0.00%	0.03%	0.168	0.667
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	0.19%	0.00%	0.19%	0.168	0.667
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	7.30%	9.66%	-2.35%	0.169	0.667
Clostridia	Clostridiales	Veillonellaceae	*Dialister*	0.02%	0.14%	-0.12%	0.171	0.667
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Leclercia*	0.00%	0.02%	-0.02%	0.178	0.667
Flavobacteria	Flavobacteriales	Flavobacteriaceae	*Capnocytophaga*	0.00%	0.03%	-0.02%	0.178	0.667
Clostridia	Clostridiales	ClostridialesFamilyXIII.IncertaeSedis	*Eubacterium*	0.00%	0.03%	-0.02%	0.178	0.667
TM7-3	CW040	F16		0.00%	0.03%	-0.02%	0.178	0.667

Actinobacteria	Actinomycetales	Actinomycetaceae	*Actinomyces*	0.06%	0.17%	-0.12%	0.180	0.667
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Klebsiella*	0.00%	0.03%	-0.03%	0.187	0.667
Actinobacteria	Actinomycetales	Micrococcaceae	*Rothia*	1.20%	4.49%	-3.29%	0.189	0.667
Bacteroidia	Bacteroidales	Bacteroidaceae	*Bacteroides*	0.29%	0.19%	0.09%	0.190	0.667
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.00%	0.02%	-0.02%	0.193	0.667
Gammaproteobacteria	Alteromonadales	Pseudoalteromonadaceae	*Pseudoalteromonas*	0.00%	0.03%	-0.03%	0.193	0.667
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.00%	0.19%	-0.19%	0.193	0.667
Bacilli	Bacillales	Bacillaceae	*Pontibacillus*	0.06%	0.00%	0.06%	0.193	0.667
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.18%	0.00%	0.18%	0.193	0.667
Actinobacteria	Actinomycetales	Actinomycetaceae	*Actinomyces*	0.00%	0.03%	-0.03%	0.193	0.667
Actinobacteria	Actinomycetales	Actinomycetaceae	*Actinomyces*	0.00%	0.05%	-0.05%	0.193	0.667
Clostridia	Clostridiales	ClostridialesFamilyXI.IncertaeSedis		0.06%	0.41%	-0.35%	0.196	0.667
Clostridia	Clostridiales	Lachnospiraceae	*Shuttleworthia*	0.00%	0.20%	-0.20%	0.196	0.667
Bacteroidia	Bacteroidales	Bacteroidaceae	*Bacteroides*	0.02%	0.01%	0.01%	0.198	0.667
Actinobacteria	Actinomycetales	Actinomycetaceae	*Actinomyces*	0.02%	0.08%	-0.07%	0.199	0.667
Actinobacteria	Actinomycetales	Corynebacteriaceae	*Corynebacterium*	0.02%	0.06%	-0.05%	0.203	0.667
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Serratia*	0.07%	0.15%	-0.08%	0.205	0.667
Bacilli	Turicibacterales	Turicibacteraceae		0.01%	0.07%	-0.05%	0.222	0.697
Clostridia	Clostridiales	Lachnospiraceae		0.58%	1.23%	-0.65%	0.223	0.697
Clostridia	Clostridiales	ClostridialesFamilyXIII.IncertaeSedis	*Mogibacterium*	0.01%	0.22%	-0.20%	0.226	0.697
Clostridia	Clostridiales	Veillonellaceae	*Veillonella*	0.01%	0.06%	-0.05%	0.228	0.697
Betaproteobacteria	Neisseriales	Neisseriaceae	*Neisseria*	0.36%	1.80%	-1.44%	0.235	0.710
Bacteroidia	Bacteroidales	Bacteroidaceae	*Bacteroides*	0.00%	0.02%	-0.02%	0.256	0.710
Clostridia	Clostridiales	Veillonellaceae	*Selenomonas*	0.00%	0.09%	-0.09%	0.257	0.710
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.00%	0.05%	-0.04%	0.257	0.710

Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.03%	0.10%	-0.07%	0.258	0.710
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.23%	1.12%	-0.89%	0.258	0.710
Actinobacteria	Actinomycetales	Corynebacteriaceae		0.01%	0.02%	-0.01%	0.259	0.710
Fusobacteria	Fusobacteriales	Fusobacteriaceae	*Fusobacterium*	0.02%	0.04%	-0.02%	0.259	0.710
Clostridia	Clostridiales	Ruminococcaceae	*Oscillospira*	0.00%	0.02%	-0.01%	0.270	0.718
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Erwinia*	0.00%	0.02%	-0.01%	0.284	0.718
Bacteroidia	Bacteroidales			0.00%	0.02%	-0.01%	0.284	0.718
Clostridia	Clostridiales	ClostridialesFamilyXI .IncertaeSedis	*Finegoldia*	0.00%	0.10%	-0.10%	0.296	0.718
Gammaproteobacteria	Vibrionales	Vibrionaceae	*Vibrio*	0.00%	0.02%	-0.02%	0.296	0.718
Bacteroidia	Bacteroidales	Porphyromonadaceae	*Tannerella*	0.00%	0.02%	-0.02%	0.296	0.718
Gammaproteobacteria	Pasteurellales	Pasteurellaceae	*Aggregatibacter*	0.00%	0.03%	-0.03%	0.296	0.718
Clostridia	Clostridiales	Lachnospiraceae		0.00%	0.02%	-0.02%	0.296	0.718
Betaproteobacteria	Neisseriales	Neisseriaceae	*Neisseria*	0.00%	0.04%	-0.04%	0.296	0.718
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Klebsiella*	0.03%	0.63%	-0.60%	0.297	0.718
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	0.11%	0.07%	0.03%	0.304	0.727
Clostridia	Clostridiales	Lachnospiraceae	*Clostridium*	0.02%	0.01%	0.01%	0.315	0.736
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.01%	0.06%	-0.05%	0.315	0.736
Actinobacteria	Actinomycetales	Micrococcaceae	*Rothia*	0.07%	0.20%	-0.14%	0.328	0.756
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Serratia*	0.01%	0.02%	-0.01%	0.331	0.756
Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*	0.04%	0.15%	-0.12%	0.340	0.761
Betaproteobacteria	Neisseriales	Neisseriaceae	*Neisseria*	0.04%	0.24%	-0.21%	0.343	0.761
Betaproteobacteria	Neisseriales	Neisseriaceae	*Simonsiella*	0.01%	0.04%	-0.03%	0.351	0.761
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	0.01%	0.01%	-0.01%	0.363	0.761
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.02%	0.04%	-0.02%	0.364	0.761
Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*	0.01%	0.13%	-0.12%	0.364	0.761

Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.03%	0.01%	0.02%	0.364	0.761
Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Alloscardovia*	0.02%	0.13%	-0.11%	0.366	0.761
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.18%	0.63%	-0.45%	0.370	0.761
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	0.15%	0.09%	0.06%	0.386	0.761
Clostridia	Clostridiales	Lachnospiraceae	*Dorea*	0.36%	0.16%	0.19%	0.386	0.761
Bacilli	Gemellales	Gemellaceae	*Gemella*	0.59%	3.08%	-2.49%	0.387	0.761
Clostridia	Clostridiales	Ruminococcaceae	*Faecalibacterium*	0.06%	0.16%	-0.10%	0.393	0.761
Actinobacteria	Actinomycetales			0.03%	0.01%	0.02%	0.399	0.761
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Escherichia*	0.41%	0.59%	-0.18%	0.404	0.761
Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	0.01%	0.05%	-0.03%	0.412	0.761
Clostridia	Clostridiales	Ruminococcaceae	*Oscillospira*	0.00%	0.02%	-0.02%	0.422	0.761
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.03%	0.28%	-0.25%	0.423	0.761
Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	0.06%	0.14%	-0.08%	0.432	0.761
Bacilli	Bacillales	Staphylococcaceae	*Staphylococcus*	15.60%	11.94%	3.67%	0.449	0.761
Mollicutes	Mycoplasmatales	Mycoplasmataceae	*Mycoplasma*	0.01%	0.05%	-0.03%	0.456	0.761
Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*	0.09%	4.82%	-4.73%	0.460	0.761
Bacilli	Lactobacillales	Enterococcaceae	*Enterococcus*	9.80%	0.14%	9.66%	0.464	0.761
Bacteroidia	Bacteroidales	Porphyromonadaceae	*Porphyromonas*	0.01%	0.02%	-0.01%	0.464	0.761
Clostridia	Clostridiales	Lachnospiraceae	*Lachnobacterium*	0.01%	0.03%	-0.02%	0.464	0.761
Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*	0.01%	0.02%	-0.02%	0.464	0.761
Actinobacteria	Actinomycetales	Corynebacteriaceae	*Corynebacterium*	0.08%	0.91%	-0.83%	0.464	0.761
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.07%	0.27%	-0.19%	0.467	0.761
Bacilli	Bacillales	Bacillaceae	*Oceanobacillus*	0.00%	0.06%	-0.06%	0.468	0.761
Bacteroidia	Bacteroidales			0.00%	0.04%	-0.04%	0.468	0.761
Bacteroidia	Bacteroidales			0.00%	0.09%	-0.09%	0.468	0.761
Bacteroidia	Bacteroidales			0.00%	0.08%	-0.08%	0.468	0.761

Bacteroidia	Bacteroidales			0.00%	0.02%	-0.02%	0.468	0.761
Clostridia	Clostridiales	Veillonellaceae	*Selenomonas*	0.00%	0.07%	-0.07%	0.468	0.761
Clostridia	Clostridiales	Ruminococcaceae	*Eubacterium*	0.00%	0.02%	-0.02%	0.468	0.761
Gammaproteobacteria	Pasteurellales	Pasteurellaceae	*Haemophilus*	0.00%	0.02%	-0.02%	0.468	0.761
Bacteroidia	Bacteroidales			0.00%	0.06%	-0.06%	0.468	0.761
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.01%	0.05%	-0.03%	0.484	0.781
Clostridia	Clostridiales	Peptostreptococcaceae	*Peptostreptococcus*	0.03%	0.84%	-0.81%	0.519	0.831
Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	0.16%	0.04%	0.12%	0.533	0.844
Clostridia	Clostridiales	Ruminococcaceae	*Oscillospira*	0.02%	0.05%	-0.03%	0.536	0.844
Clostridia	Clostridiales	Veillonellaceae	*Veillonella*	0.01%	0.05%	-0.04%	0.540	0.844
Clostridia	Clostridiales	Lachnospiraceae	*Blautia*	0.24%	0.23%	0.01%	0.560	0.847
Actinobacteria	Coriobacteriales	Coriobacteriaceae	*Collinsella*	0.01%	0.04%	-0.03%	0.582	0.847
Flavobacteria	Flavobacteriales	Flavobacteriaceae	*Capnocytophaga*	0.02%	0.06%	-0.04%	0.582	0.847
Gammaproteobacteria	Pasteurellales	Pasteurellaceae	*Haemophilus*	0.06%	0.23%	-0.17%	0.588	0.847
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.01%	0.02%	-0.01%	0.592	0.847
Betaproteobacteria	Neisseriales	Neisseriaceae	*Neisseria*	0.00%	0.02%	-0.02%	0.592	0.847
Actinobacteria	Actinomycetales	Actinomycetaceae	*Actinomyces*	0.01%	0.05%	-0.04%	0.592	0.847
Bacteroidia	Bacteroidales	Bacteroidaceae	*Bacteroides*	0.03%	0.02%	0.01%	0.594	0.847
Bacteroidia	Bacteroidales	Bacteroidaceae	*Bacteroides*	0.05%	0.05%	-0.01%	0.594	0.847
Bacilli	Bacillales	Staphylococcaceae	*Staphylococcus*	0.24%	0.00%	0.23%	0.602	0.847
Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae		0.04%	0.01%	0.03%	0.602	0.847
Bacilli	Bacillales	Bacillaceae	*Anoxybacillus*	0.01%	0.02%	-0.01%	0.603	0.847
Mollicutes	Mycoplasmatales	Mycoplasmataceae	*Mycoplasma*	0.28%	0.25%	0.03%	0.607	0.847
Clostridia	Clostridiales	Veillonellaceae	*Veillonella*	0.15%	0.20%	-0.05%	0.613	0.847
Bacteroidia	Bacteroidales	Porphyromonadaceae	*Parabacteroides*	0.00%	0.02%	-0.01%	0.619	0.847

Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	0.83%	2.70%	-1.87%	0.620	0.847
Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	*Akkermansia*	0.01%	0.02%	-0.01%	0.623	0.847
Clostridia	Clostridiales	Clostridiaceae		0.00%	0.04%	-0.03%	0.646	0.847
Clostridia	Clostridiales	Clostridiaceae	*Clostridium*	0.00%	0.02%	-0.02%	0.646	0.847
Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Parascardovia*	0.00%	0.02%	-0.01%	0.646	0.847
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.11%	4.48%	-4.37%	0.655	0.847
Fusobacteria	Fusobacteriales	Fusobacteriaceae	*Leptotrichia*	0.07%	0.21%	-0.14%	0.662	0.847
Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	*Campylobacter*	0.01%	0.02%	-0.02%	0.666	0.847
Fusobacteria	Fusobacteriales	Fusobacteriaceae	*Fusobacterium*	0.01%	0.01%	0.00%	0.666	0.847
Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	0.03%	0.06%	-0.04%	0.669	0.847
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.02%	0.08%	-0.06%	0.669	0.847
Fusobacteria	Fusobacteriales	Fusobacteriaceae	*Fusobacterium*	0.29%	1.48%	-1.19%	0.670	0.847
Bacteroidia	Bacteroidales	Bacteroidaceae	*Bacteroides*	0.06%	0.03%	0.04%	0.673	0.847
Bacteroidia	Bacteroidales	Porphyromonadaceae	*Porphyromonas*	0.01%	0.02%	-0.01%	0.674	0.847
Clostridia	Clostridiales	Clostridiaceae		0.07%	0.02%	0.05%	0.674	0.847
Bacilli	Bacillales	Paenibacillaceae	*Brevibacillus*	5.64%	4.80%	0.84%	0.676	0.847
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.00%	0.04%	-0.04%	0.676	0.847
Clostridia	Clostridiales	Veillonellaceae	*Selenomonas*	0.00%	0.07%	-0.07%	0.676	0.847
Betaproteobacteria	Neisseriales	Neisseriaceae	*Neisseria*	0.02%	0.06%	-0.04%	0.691	0.861
Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Moraxella*	0.02%	0.02%	-0.01%	0.714	0.880
Clostridia	Clostridiales	Ruminococcaceae		0.04%	0.02%	0.02%	0.715	0.880
Gammaproteobacteria	Pasteurellales	Pasteurellaceae	*Aggregatibacter*	0.11%	0.04%	0.07%	0.732	0.895
				0.03%	0.11%	-0.08%	0.737	0.897
Clostridia	Clostridiales	Lachnospiraceae		0.02%	0.07%	-0.04%	0.744	0.900
Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	0.05%	0.01%	0.05%	0.759	0.908
Gammaproteobacteria	Pasteurellales	Pasteurellaceae	*Actinobacillus*	3.00%	0.19%	2.81%	0.760	0.908

Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	*Stenotrophomonas*	9.75%	0.02%	9.73%	0.770	0.916
Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	*Bifidobacterium*	0.01%	0.01%	0.00%	0.780	0.920
Clostridia	Clostridiales	Lachnospiraceae		0.01%	0.08%	-0.06%	0.791	0.920
Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	0.01%	0.01%	0.00%	0.792	0.920
Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	0.02%	0.02%	0.00%	0.793	0.920
Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	0.05%	0.07%	-0.02%	0.800	0.920
Clostridia	Clostridiales	Clostridiaceae		0.01%	0.05%	-0.04%	0.800	0.920
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	0.20%	0.07%	0.13%	0.807	0.922
Clostridia	Clostridiales	Lachnospiraceae		0.06%	0.07%	-0.01%	0.820	0.930
Clostridia	Clostridiales	Veillonellaceae	*Megasphaera*	0.03%	0.14%	-0.11%	0.823	0.930
Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	*Catenibacterium*	0.01%	0.02%	-0.01%	0.846	0.950
Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	0.90%	0.36%	0.54%	0.852	0.950
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	0.02%	0.02%	0.00%	0.854	0.950
Actinobacteria	Actinomycetales	Propionibacteriaceae	*Propionibacterium*	0.04%	0.03%	0.01%	0.866	0.959
Bacilli	Lactobacillales	Enterococcaceae	*Enterococcus*	0.18%	7.14%	-6.95%	0.880	0.968
Bacteroidia	Bacteroidales	Porphyromonadaceae		0.00%	0.02%	-0.02%	0.896	0.976
Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	*Bulleidia*	0.00%	0.02%	-0.02%	0.896	0.976
Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	0.00%	0.03%	-0.03%	0.931	0.980
Clostridia	Clostridiales	Eubacteriaceae	*Eubacterium*	0.00%	0.02%	-0.01%	0.931	0.980
Clostridia	Clostridiales	Veillonellaceae	*Selenomonas*	0.00%	0.04%	-0.04%	0.931	0.980
Clostridia	Clostridiales	ClostridialesFamilyXIII.IncertaeSedis	*Eubacterium*	0.01%	0.03%	-0.02%	0.931	0.980
Gammaproteobacteria	Pasteurellales	Pasteurellaceae	*Haemophilus*	0.09%	0.09%	0.00%	0.933	0.980
Bacilli	Bacillales	Staphylococcaceae	*Staphylococcus*	0.15%	0.07%	0.08%	0.935	0.980
Bacteroidia	Bacteroidales	Porphyromonadaceae	*Porphyromonas*	0.14%	0.60%	-0.45%	0.940	0.980
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.01%	0.04%	-0.03%	0.944	0.980

Betaproteobacteria	Burkholderiales	Oxalobacteraceae	*Janthinobacterium*	0.03%	0.00%	0.02%	0.965	0.980
Bacilli	Lactobacillales	Lactobacillaceae	*Lactobacillus*	0.09%	0.02%	0.07%	0.965	0.980
Bacteroidia	Bacteroidales	Prevotellaceae	*Prevotella*	0.02%	0.01%	0.00%	0.965	0.980
Clostridia	Clostridiales	Ruminococcaceae	*Oscillospira*	0.01%	0.01%	0.00%	0.974	0.980
Bacteroidia	Bacteroidales			0.01%	0.02%	0.00%	0.974	0.980
Bacilli	Lactobacillales	Carnobacteriaceae	*Granulicatella*	0.27%	0.07%	0.20%	0.974	0.980
Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae		0.01%	0.02%	-0.01%	0.977	0.980
Bacteroidia	Bacteroidales	Porphyromonadaceae	*Parabacteroides*	0.04%	0.03%	0.01%	0.979	0.980
Bacilli	Lactobacillales	Streptococcaceae	*Lactococcus*	0.29%	0.08%	0.21%	0.980	0.980

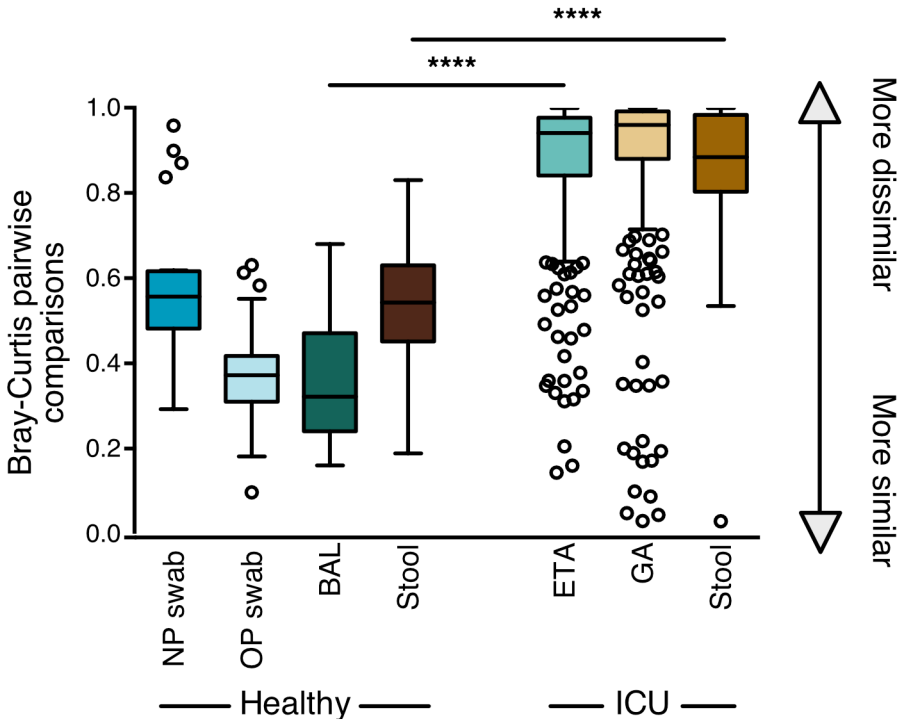


Figure A.1: Greater heterogeneity within anatomical site in the ICU cohort in comparison to a healthy cohort. Tukey' box plots of the pairwise comparisons of Bray-Curtis dissimilarity values within each types of specimens demonstrate that the microbial composition of ICU patients tends to be more variable between individuals compared to healthy individuals with respiratory and gastric specimens. The overlaying lines show the median and the interquartile range of each site. (**** indicates a p value < 0.001).

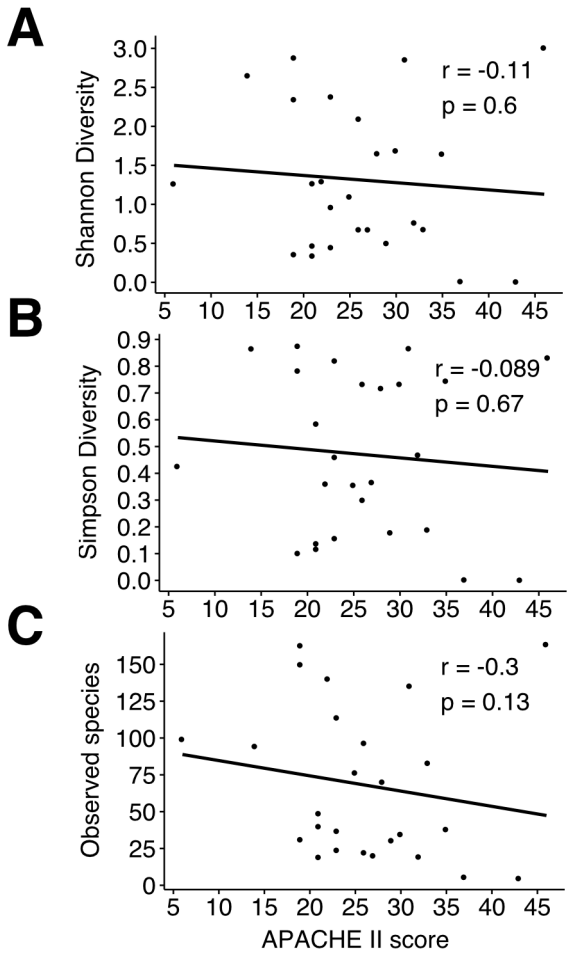


Figure A.2: Gastric microbial diversity is not associated with illness severity. Correlation analysis using Spearman's rank correlation coefficient indicating that APACHE II score is not significantly correlated with Shannon (r = -0.11, $p = 0.6$; A), Simpson diversity (r = -0.089, $p = 0.67$; B) and Observed Species (r = -0.3, $p = 0.13$; C).

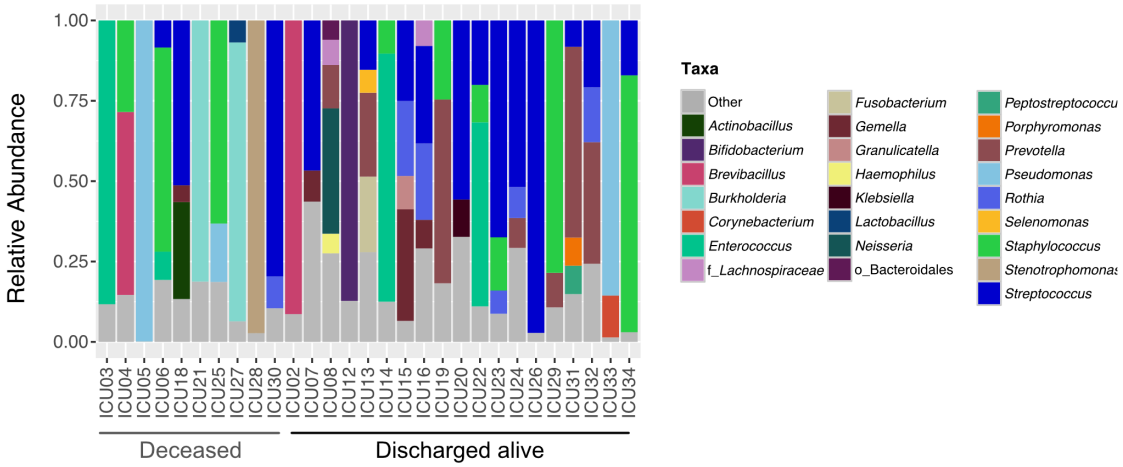


Figure A.3: Microbial profiles of the ETA specimens collected. Taxonomic summaries of the ETAs samples included in this study displayed by patients and hospital mortality (discharged alive or deceased). Bacterial groups present at least than 5% are grouped in the "other" category displayed in gray. Taxonomic summaries are labeled according to the highest level resolved (order; o_, family; f_).

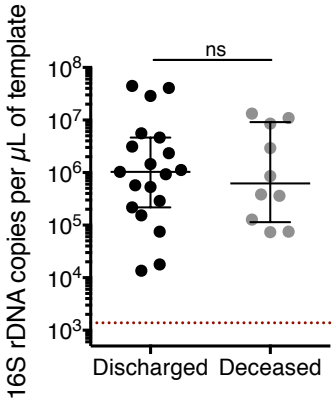


Figure A.4: Lack of association between hospital mortality and bacterial load in lower respiratory tract samples. Bacterial biomass estimates via quantification of the 16S rRNA gene by quantitative polymerase chain reaction failed to demonstrate a difference between patient's deceased and discharged alive from the hospital. Measurements were performed in triplicates and the average was reported. Overlaying lines show the median and interquartile range of each group. The red dotted line represents the limit of detection. (ns refers to a p value > 0.05).

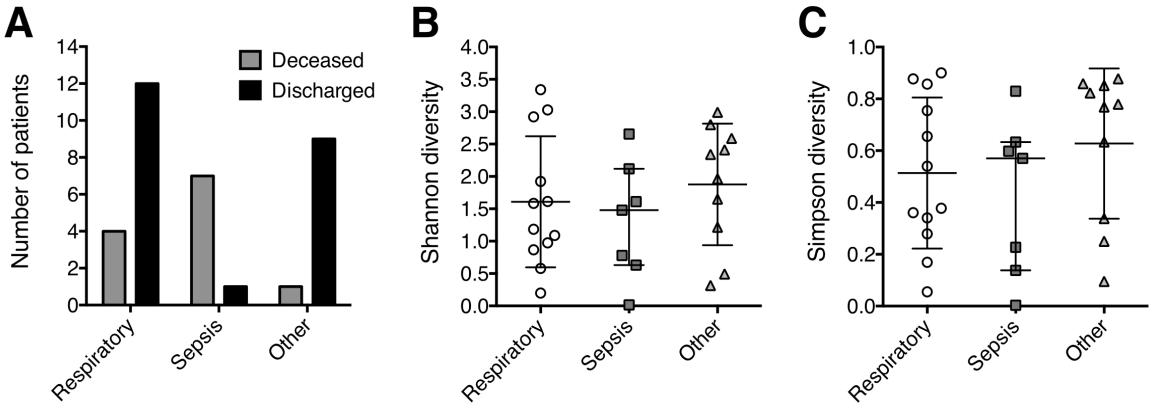


Figure A.5: Absence of detectable difference within microbial diversity between categories of admission. Population proportion of deceased and discharged alive patients by category of admission (A) demonstrates that the mortality burden arises from patient with a sepsis admission diagnosis. Shannon diversity (B) and Simpson diversity (C) of ETA specimens shaded by category of admission demonstrate no detectable differences in the α diversity measurements between patients admitted due to a respiratory, sepsis or other diagnoses.

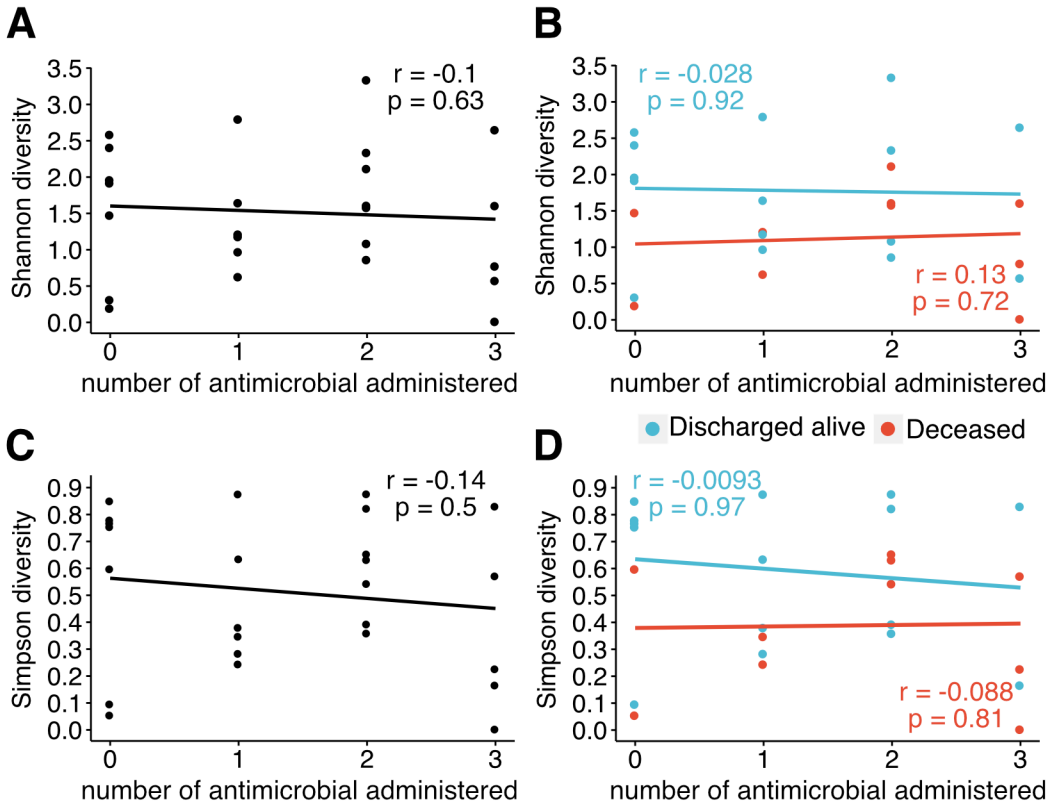


Figure A.6: Antimicrobials exposure is not associated with ETA microbial diversity. ETA α diversity using Shannon (A-B) and Simpson diversity indexes (B-C) was not significantly correlated with concomitant antimicrobial exposure at sample collection day. Correlation analysis was performed using Spearman's rank correlation coefficient.

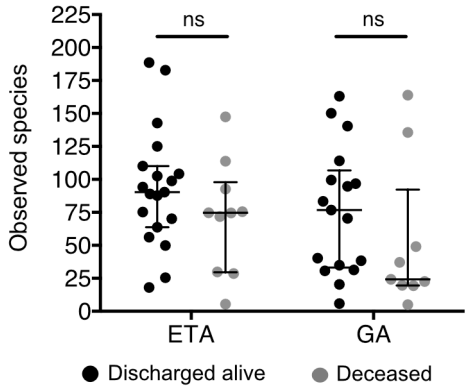


Figure A.7: No association between ICU samples microbial diversity and hospital mortality using the Observed Species index. Observed Species of ETA and GA samples demonstrated no significant reduction of the microbial diversity in the patients deceased in the hospital versus the patient discharged alive. Overlaying lines show the median and interquartile range of each group.

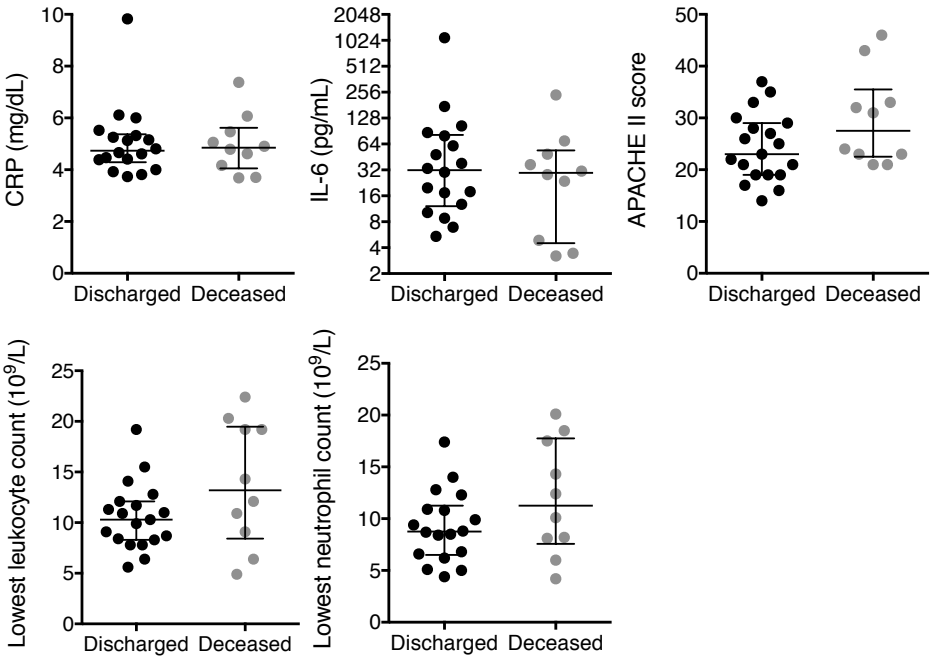


Figure A.8: Inflammatory markers and APACHE II score are not statistically different between deceased and discharged alive patients. Inflammatory markers and illness severity scores were acquired from 29 patients (10 deceased and 19 discharged alive from the hospital). None of the comparisons was statistically different between the deceased and discharged alive from the hospital groups. Overlaying lines show the median and interquartile range of each group.

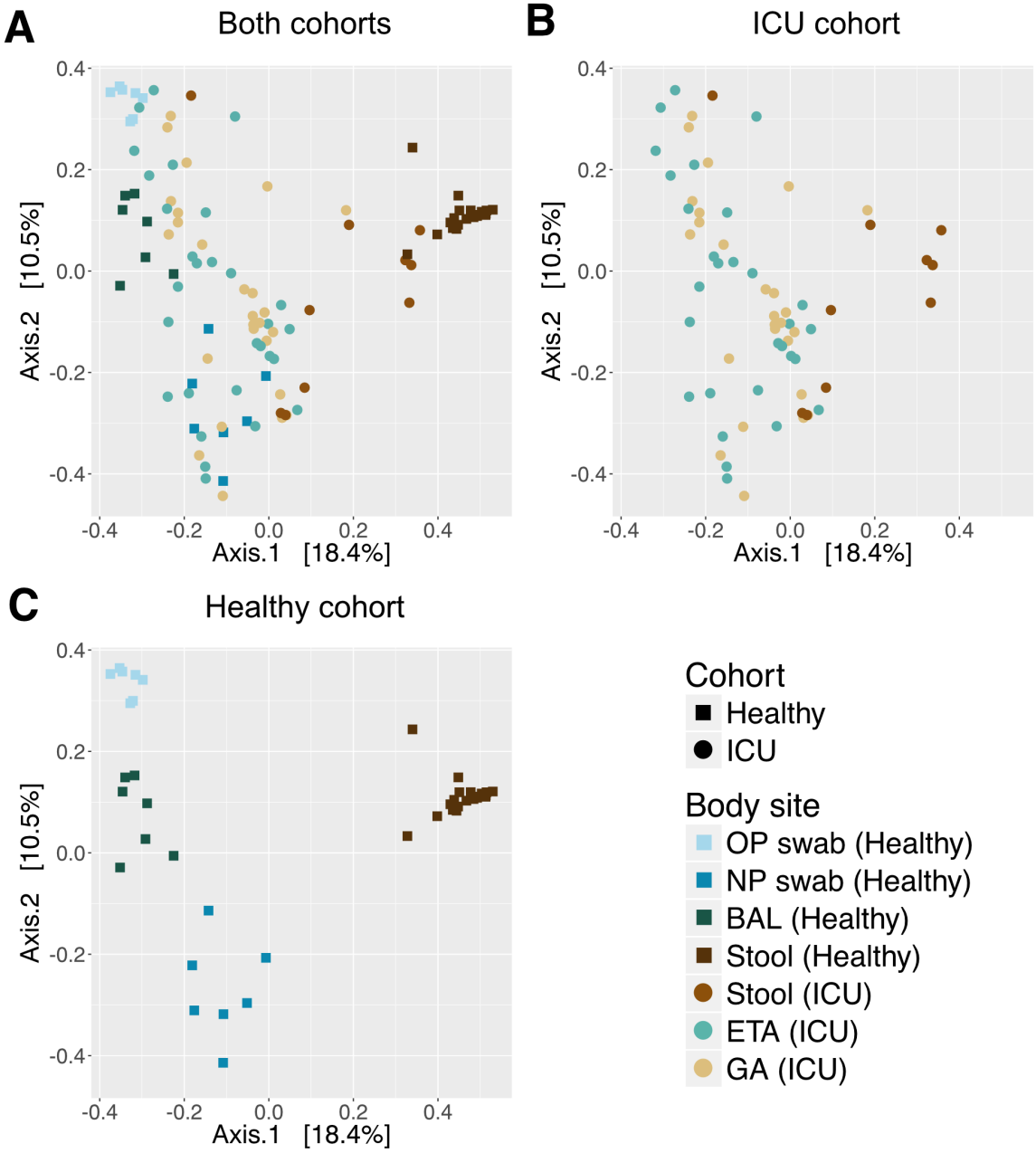


Figure A.9: OTU5 does not influence the loss of biogeographical distinction in ICU patients. Principal coordinates analysis (PCoA) using Bray-Curtis dissimilarity metric between the ICU and healthy cohorts (A) are showing that samples collected from healthy cohort tend to cluster per sample sites (B) whereas the samples from different anatomical sites tend to overlap in the ICU cohort (C).

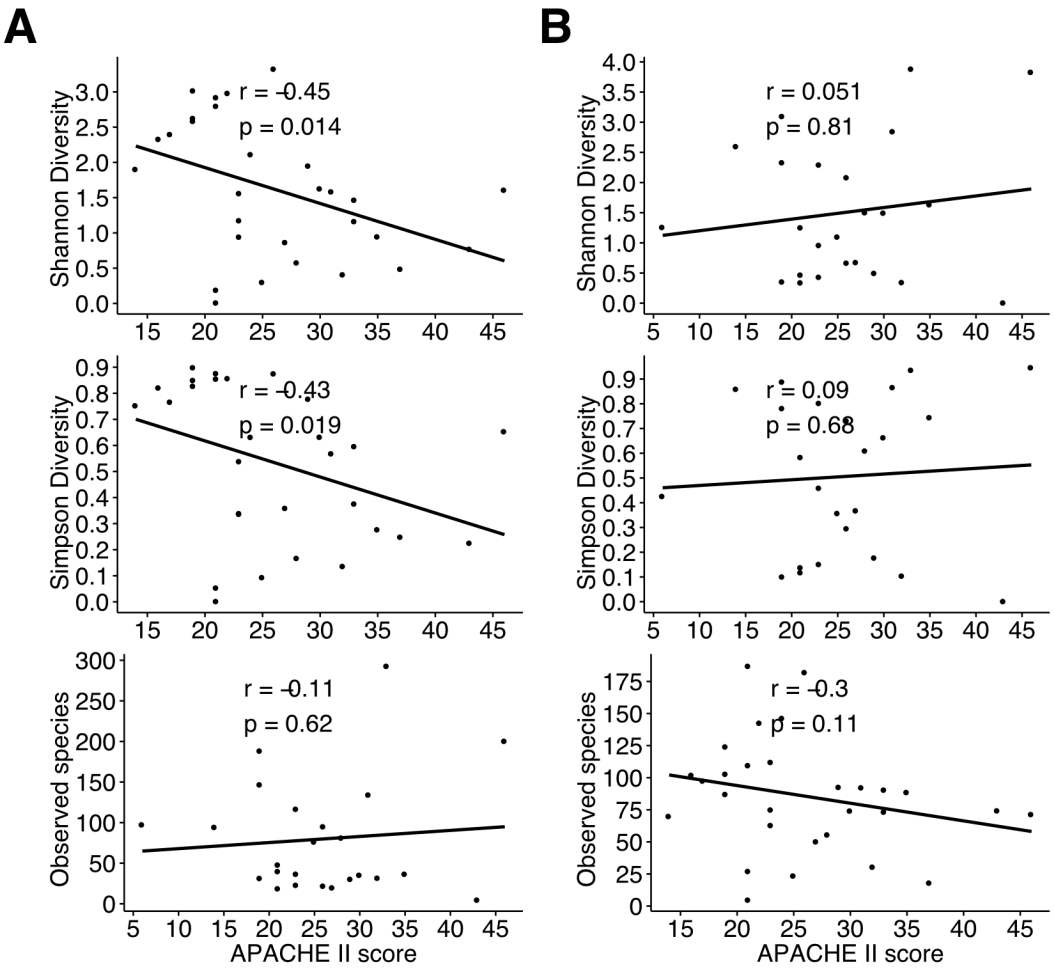


Figure A.10: APACHE II score association with α diversity remains when OTU5 is removed from the analysis. ETA α diversity is inversely correlated with APACHE II score using Shannon and Simpson diversity but not with Observed Species (A). This association is not observed with GA samples (B). Correlation analysis was performed using Spearman's rank correlation coefficient.

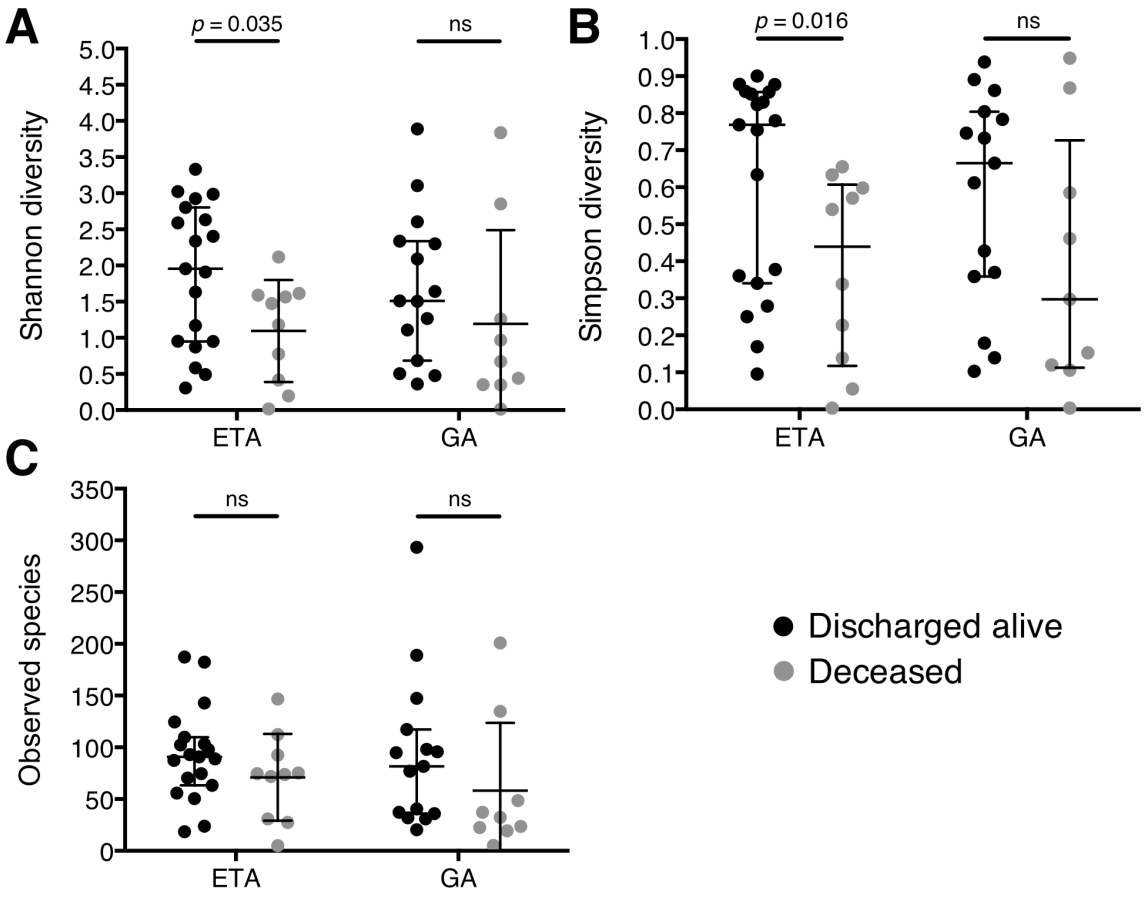


Figure A.11: OTU5 removal does not impact the decrease within microbial diversity observed in the deceased group for ETAs. Shannon (A), Simpson diversity (B) and Observed Species (C) of ETA and GA specimens displayed by hospital mortality demonstrated a significant reduction in the ETA microbial diversity using Shannon and Simpson diversity in the patients deceased in the hospital versus patients discharged alive.

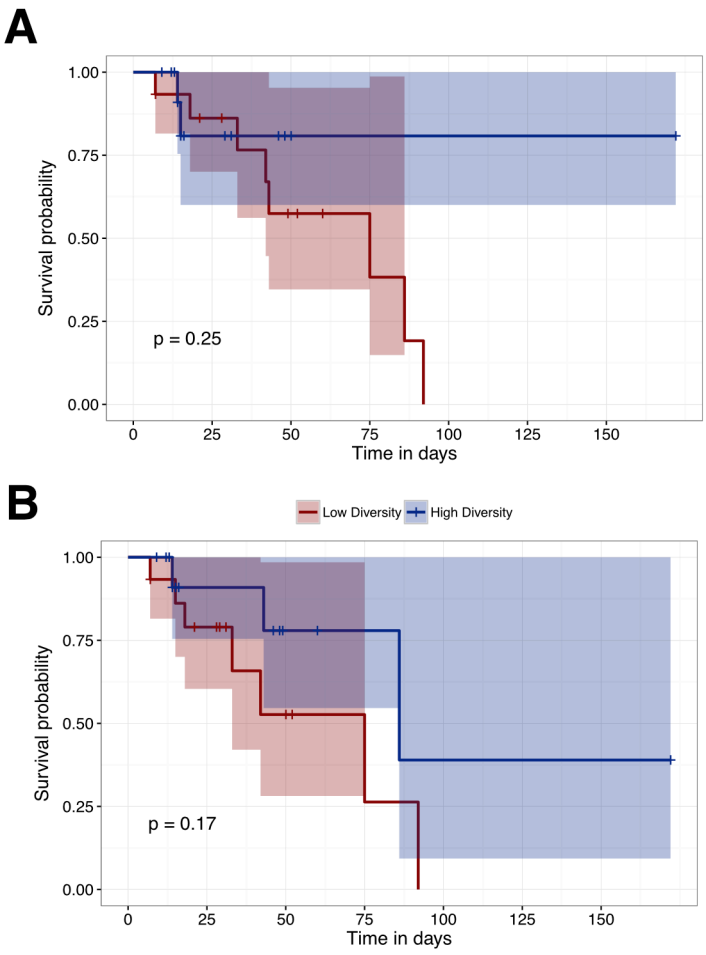


Figure A.12: Association between ICU samples microbial diversity and patient's outcomes when OTU5 is removed. Kaplan-Meier survival curves displayed by microbial diversity groups showing patients censored (i.e. discharged alive) and deceased within time for the ETA samples. The threshold for the diversity group was the median value of the Shannon diversity and Simpson diversity (A) and Observed Species (B) measurements for the 29 samples included in this analysis

Appendix B

Appendix to Chapter 2: Prospect Mechanistic Sub-study

Mechanistic study within the PROSPECT trial

The PROSPECT study stands for Probiotics to prevent Severe Pneumonia and Endotracheal Colonization Trial. The goal of this randomized controlled trial (RCT) is to assess the clinical impact of the administration of *Lactobacillus rhamnosus* GG on the prevalence of ventilator-associated pneumonia (VAP) and other nosocomial infections. During the study, a capsule containing $>1x10^{10}$ Colony Forming Units (CFU) of *Lactobacillus rhamnosus* GG (*L. r* GG; Culturelle®) or placebo (microcrystalline cellulose) has been given twice daily through a nasogastric tube to critically ill mechanically ventilated patients (Cook et al., 2016; Johnstone et al., 2019).

The PROSPECT trial includes a mechanistic sub-study investigating the biological impact of administrating *L.r* GG on the host-associated communities of critically ill patients. In order to do this, we performed a microbiome analysis and measured circulatory inflammatory mediators and related bacterial products from the patient's plasma. The 50 patients in the sub-study were initially part of a feasibility study but were later incorporated into the larger trial and as noted below we remain blinded to the control versus treatment group. Investigations into the host's immunology of the PROSPECT patients have been performed by Dr. Bowdish laboratory (McMaster University) and their manuscript entitled *Evidence for immune dysregulation in mechanically ventilated critically ill patients* has been submitted to the Journal of Intensive Critical Care.

The focus of the microbiome analysis nested within the PROSPECT mechanistic study was to determine how microbiota composition and dynamics interplay with

probiotics and VAP susceptibility. We believed that additional insights as to the bacterial composition of host-associated communities in terms of their perturbations, function, and virulence will be vital to prevent ICU-induced adverse events, such as respiratory infections. Moreover, we hypothesized that probiotics could have a beneficial effect on the stability of the microbial composition and could decrease the carriage of pathogens in the lower respiratory tract and hence could reduce VAP prevalence in the ICU.

The PROSPECT includes 2653 patients recruited from 42 ICUs in Canada, and 3 international sites and has finish enrolling patients, but the study has not been closed yet. Accordingly, information about the administration of the probiotic and placebo is still blinded/unknown to researchers and any subsequent analysis could be biased and should be avoided. However, while waiting for this information to be available to us, we performed a small prospective study focussing on the samples collected early in the ICU (65 samples; collected before or rapidly following the administration of the probiotics and placebo treatment), this work is presented in **Chapter 2**. This Appendix will present the methodology and a brief overview of the preliminary work performed using the entire dataset collected during the PROSPECT mechanistic sub-study. The full analysis of this data awaits unblinding of the trial data which includes those patients from our mechanistic sub-study.

Method

As part of the mechanistic study, three types of specimens were collected: endotracheal tube aspirate (ETA), nasogastric tube aspirate (NGA) and stool samples (Table A1) and

various methods have been used to assess the microbial composition of these different populations such as direct sequencing (as described in Chapter 2) and culturing (Figure A.1). Samples were collected prior to the first administration of the probiotic or the placebo (baseline), three times a week until the patient is extubated, enrolled for more than 60 days or has passed away. Samples processing and analysis for the culture-independent profiling was performed as described previously (Lamarche et al., 2018).

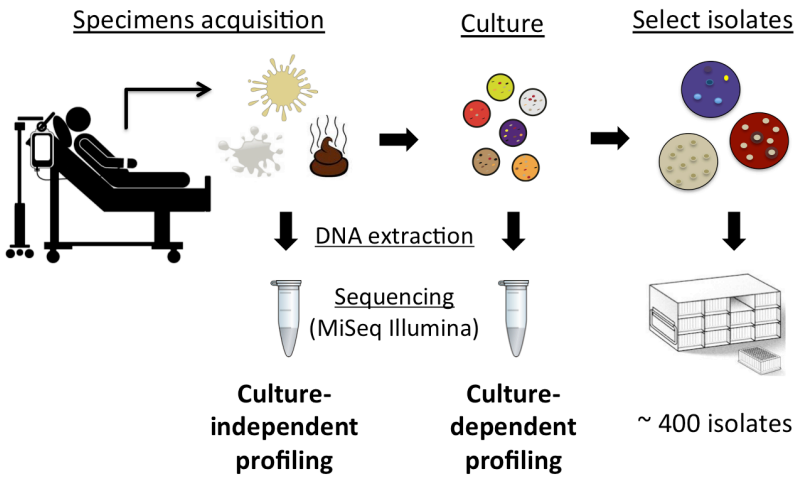


Figure B.1: Experimental approach to perform molecular profiling of the clinical samples collected during the PROSPECT study via culture-dependent and culture-independent methods.

For the culture-dependent profiling, every sample obtained at baseline and at day 7 have generally been cultured on several selective and non-selective media in both aerobic and anaerobic conditions (12 types of media). Following the appropriate incubation time at 37°C (3 days in aerobic condition and 5 days in anaerobic condition), colonies with a specific morphotype have been isolated in order to build an isolate

collection. The colonies picked to build the strain collection show either one of the following phenotype (β-hemolytic, growth on McKay agar (semi-selective media for the isolation of *Streptococcus* Anginosus/Milleri group) or growth on Man, Rogosa and Sharpe (MRS) plate (selective media for the isolation of *Lactobacillus*, *Streptococcus* and *Enterococcus*)). Subsequently, the microorganisms growing on the plate were recovered within one unique pool per specimen and were stored at -80°C until further use.

Results

As part of the mechanistic study, 444 samples were collected in 42 patients from three local ICUs (Table B.1) from October 2013 to July 2014. From these samples, 96 specimens (59 ETA, 35 NGA, 2 BAL) were cultured from 30 patients. The strain library is composed of 380 isolates with specific morphotypes. Indeed, 200 colonies were isolated from McKay agar, 90 were β-hemolytic under aerobic conditions, 40 were β-hemolytic under anaerobic conditions and 50 colonies were isolated from MRS agar (Table B.2).

Table B.1: Specimens received as part of the PROSPECT mechanistic sub-study

Specimens	Received	# Of Patients
Endotracheal Aspirate	228	42
Nasogastric Aspirate	158	36
Stool	49	20
Bronchoscopy sample	9	5
Total	444	42

Table B.2: Taxonomic identification of the isolates based on V1-V5 16S rRNA gene

Isolate ID	Prevalence
Staphylococcus epidermidis	23.67%
Enterococcus durans	13.78%
Staphylococcus aureus	10.60%
Lactobacillus rhamnosus	10.25%
Enterococcus faecalis	9.89%
Lactobacillus gasseri	3.89%
Lactobacillus casei	2.47%
Streptococcus salivarius	2.47%
Lactobacillus paracasei	2.12%
Streptococcus parasanguinis	2.12%
Streptococcus anginosus	1.77%
Pseudomonas aeruginosa	1.41%
Streptococcus infantis	1.41%
Dermabacter jinjuensis	1.06%
Corynebacterium aurimucosum	1.06%
Streptococcus mitis	1.06%
Bacteroides vulgatus	0.71%
Klebsiella pneumoniae	0.71%
Lactobacillus fermentum	0.71%
Lactobacillus salivarius	0.71%
Streptococcus australis	0.71%
Streptococcus mutans	0.71%
Actinomyces odontolyticus	0.35%
Clostridium perfringens	0.35%
Gemella sanguinis	0.35%
Arthrobacter cumminsii	0.35%
Lactobacillus brevis	0.35%
Lactobacillus pentosus	0.35%
Paeniclostridium sordellii	0.35%
Pseudomonas fluorescens	0.35%
Streptococcus constellatus	0.35%
Streptococcus gordonii	0.35%
Streptococcus pneumoniae	0.35%
Streptococcus sp.	0.35%

Since the larger PROSPECT study is currently still active, information regarding the probiotic and placebo administration are still unavailable. Several limitations such as low bacterial biomass, high host DNA, and high fungi in certain specimens collected from ICU patients resulted in the obtention of a poor microbial profile. Microbial profiles were assessed and were manually curated from the analysis when deemed low-quality (low reads, enriched in GI taxa contamination, no growth on any culture media). Consequently, the collection of specimen post curation includes 306 of the 444 (68.9%) samples recovered. Interestingly, the percentage of reads assigned to bacteria following the reads processing pipeline was significantly decreased in the low-quality samples (Figure B.2-A-B). Moreover, low-quality samples had significantly fewer reads (Figure B.2-B).

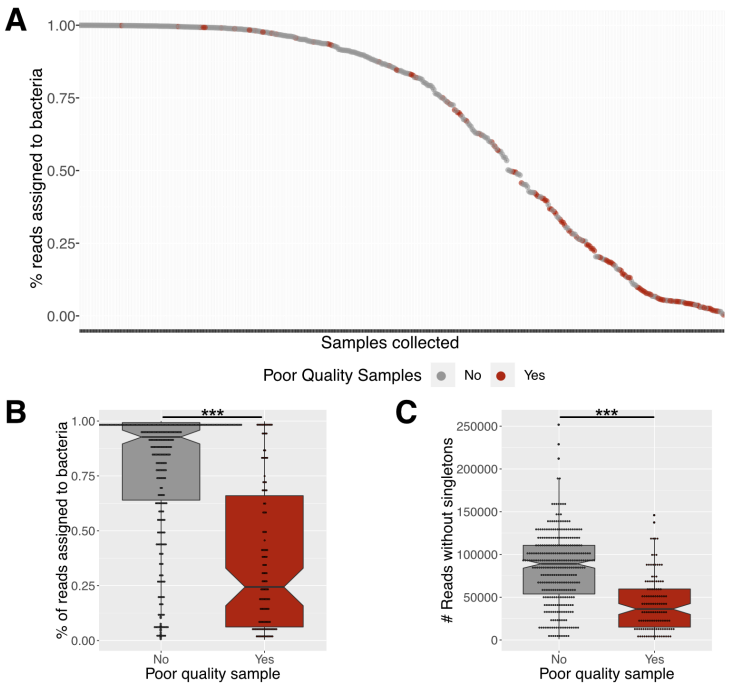


Figure B.2: Read profiles of each sample collected for the PROSPECT study. Low-quality samples have significantly fewer reads assigned to bacteria (A,B) and overall reads (C) than the other samples included in the study. Significance was assessed using a Mann-Whitney test.

Once low-quality samples filter out from the dataset, we demonstrated a loss of a clear distinction between anatomical sites in several patients, suggesting a homogenization of the different microbial communities (Figure B.3).

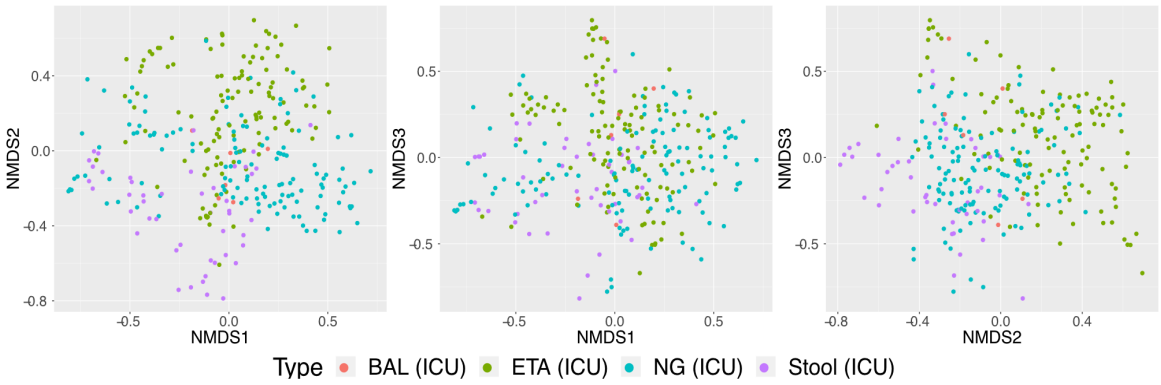


Figure B.3: Lack of microbial consensus and loss of biogeography in ICU patients. Non-Metric multidimensional scaling (NMDS) ordination using Bray-Curtis dissimilarity metric between the ICU patients. Samples collected from critically ill patients tend to overlap. The stress computed for this ordination was 0.21.

Preliminary conclusions

Microbial communities of mechanically ventilated critically ill patients differ significantly from each other and lack stability over time potentially reflecting changing clinical status and therapeutic interventions. We hypothesized that the absence of a core host-associated bacterial community could be explained by the bacterial translocation observed between the different sites due to the introduction of a new subset of microorganisms. The larger analysis for the mechanistic study has not been initiated yet. We expect to have access to the full dataset of information shortly.

Appendix C

Appendix to Chapter 4: Quality assessment for the Comparative Genomics analysis

Supplementary Results

Quality assessment of the draft genome assemblies

To ensure our analysis included only high-quality assemblies, a quality assessment of the in-house draft genomes based on contig coverage and length, was performed on 2513 contigs from 88 strains (31 *S. anginosus*, 14 *S. constellatus*, and 43 *S. intermedius*; Figure C.1). As is the case in other biological fields, no standardized quality threshold has been agreed upon for the analysis of bacterial genomes. We deemed that contigs smaller than 1000 bp with an average coverage (i.e. average number of times a base is covered with trimmed reads) below 10-fold were unreliable and were excluded from the analysis. A cluster of contigs was found below this threshold (Figure C.2A) and was removed from the analysis. Following this guideline, 129 contigs did not meet the set threshold of coverage (67 for *S. anginosus*, 9 for *S. constellatus* and 53 for *S. intermedius*) and were excluded from the analysis (Figure C.2B). The median length of the excluded contigs was 578 bp (IQR 112) and median coverage was 2-fold (IQR 2).

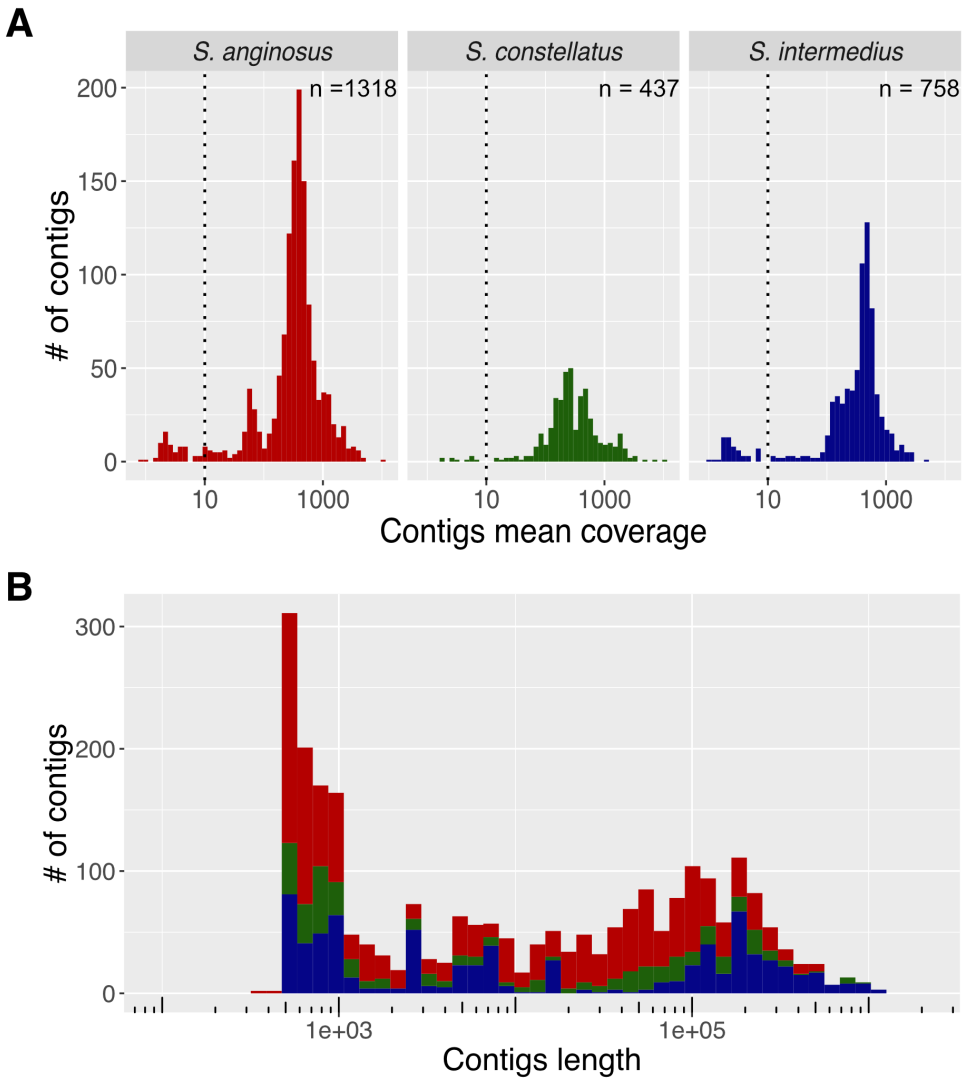


Figure C.1: Quality assessment of the contigs used for the assembly of the Surette laboratory strains. Contigs mean coverage (A) and length (B) for the 2513 contigs originating from 88 strains (Species are displayed by color, where red is *S. anginosus,* green is *S. constellatus* and blue is *S. intermedius*)

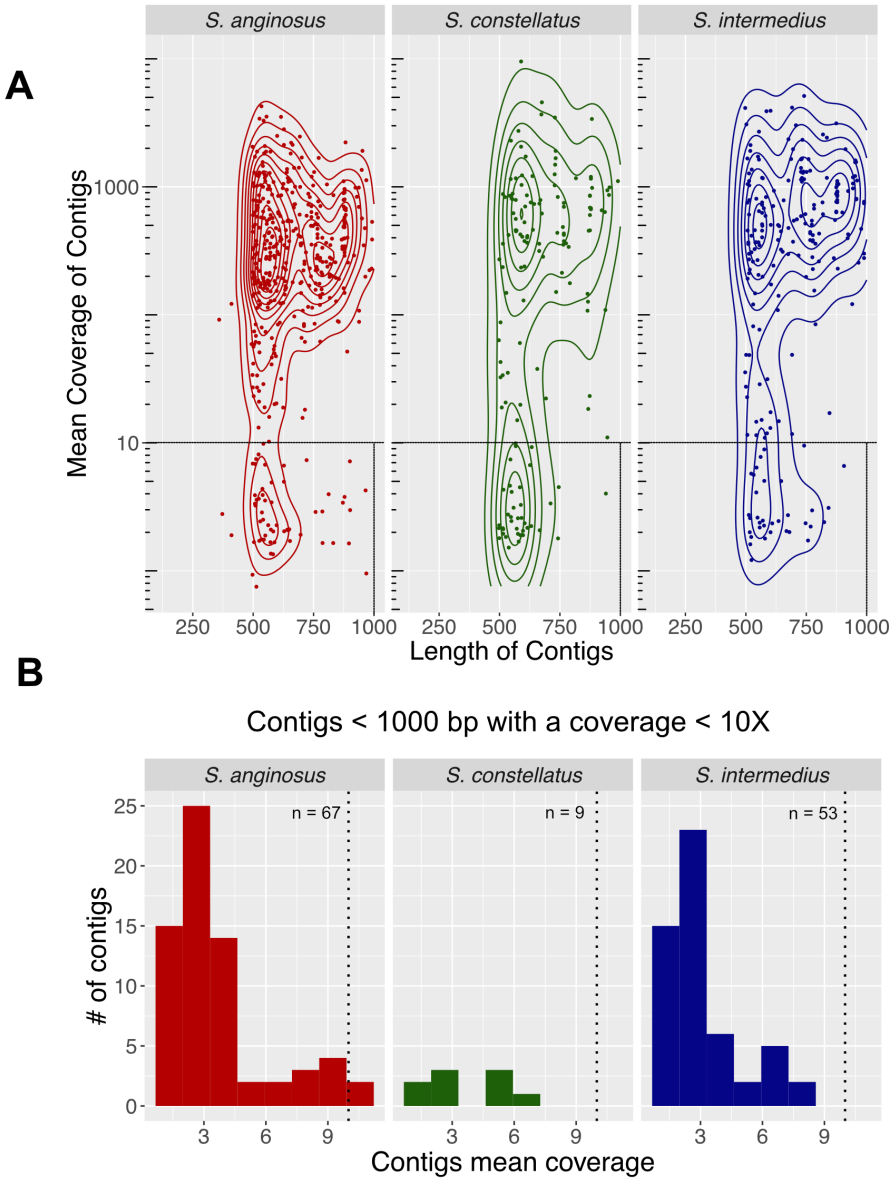


Figure C.2: Establishment of the quality threshold for the downstream-standardized analysis. Visualization of the coverage information of all contigs below 1 kbp (A; each point represents a contig) and quality assessment of the contigs below the quality assessment threshold of 10-fold coverage with a minimum length of 1000 bp (B).

All draft genomes had an N50 greater than 65 kbp (196028 ± 65344 bp) and the mean coverage of the in-house isolates was 461-fold (SD 253). The quality of the genomes assembled in-house exceeds the predetermined thresholds stated above (Figure C.3). The final quality assessment summarized by each genome assembly are listed in Table C.1.

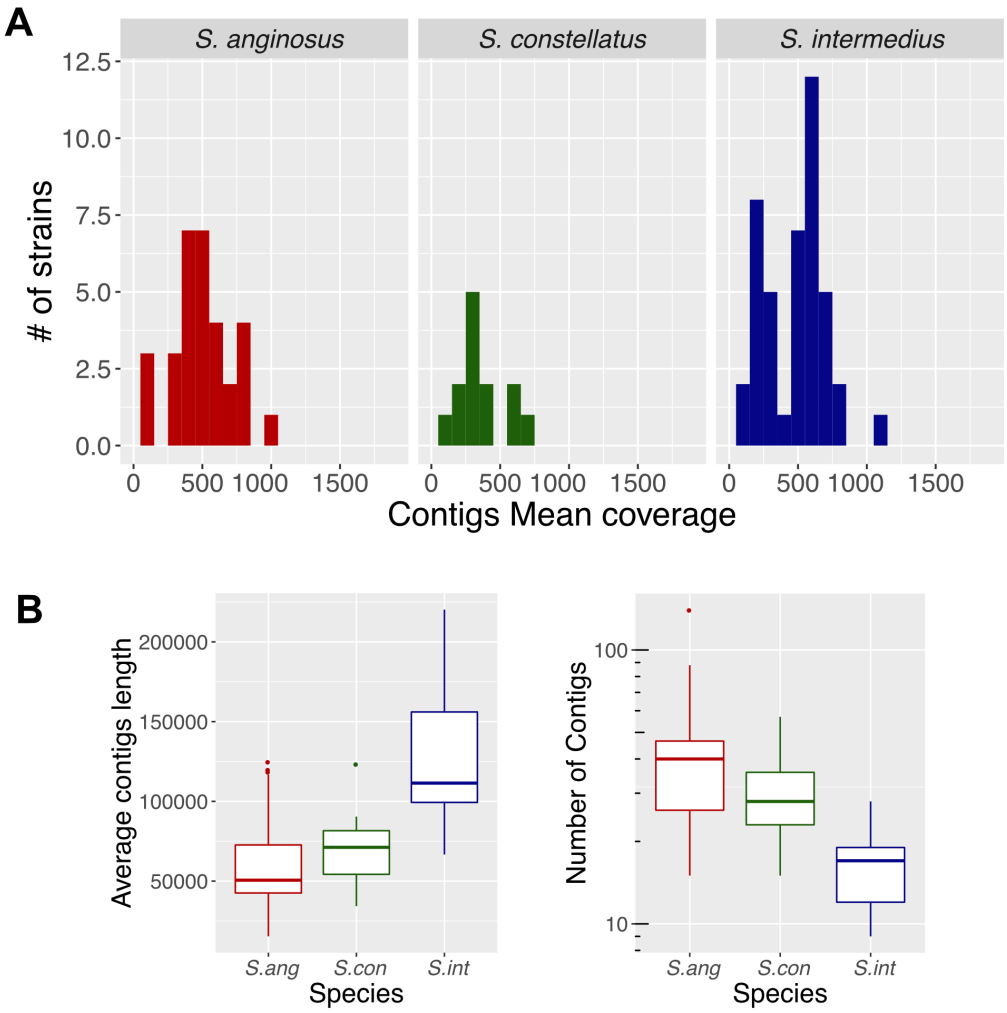


Figure C.3: Final quality assessment of the draft assemblies performed in the Surette laboratory. Coverage, contigs length and number of contigs summarized by strains (A) and species (B) suggest that the assemblies are of good quality.

Supplementary Table C.1: Assemblies' statistics for the 151 SAG isolates

Species	Strain	# of contigs	GC content (%)	N50	L50	Assembly length (bp)	Source
S.ang	ATCC33397	29	38.63	141142	5	1904329	In-house
S.ang	C1380	32	38.81	178074	4	1967055	In-house
S.ang	C252	43	38.73	158567	5	1923944	In-house
S.ang	C984	42	38.61	130958	6	1929725	In-house
S.ang	GIL195	15	38.75	231154	4	1798712	In-house
S.ang	GIL61	88	38.9	76889	9	2103536	In-house
S.ang	GIL823	16	38.68	481736	2	1896147	In-house
S.ang	M1	140	38.48	65344	11	2151590	In-house
S.ang	M410	45	38.7	155646	5	2027025	In-house
S.ang	M569	46	38.88	188252	5	1957182	In-house
S.ang	M654	46	38.85	96805	8	1957270	In-house
S.ang	P1C4	41	39.09	147980	5	1836406	In-house
S.ang	P1E11	35	38.67	156408	4	1903646	In-house
S.ang	P2A1	32	38.72	155081	5	2026735	In-house
S.ang	P2B12	15	38.8	227229	3	1872240	In-house
S.ang	P2D1	37	38.59	194485	4	2008702	In-house
S.ang	P2D2D	40	38.59	138936	6	2023157	In-house
S.ang	P2G2-2	37	38.73	155246	5	1920427	In-house
S.ang	P3A10S	25	38.67	197726	4	1899748	In-house
S.ang	P3A6	18	38.68	297491	3	1998399	In-house
S.ang	P3A9	54	38.73	85015	8	1918532	In-house
S.ang	P3B10	26	38.79	245251	3	1932018	In-house
S.ang	P3C11	47	38.94	98299	5	1937096	In-house
S.ang	P3D5	18	38.68	297491	3	1998337	In-house
S.ang	P3D6	42	38.71	114822	6	1967747	In-house
S.ang	P3F6	19	38.68	297491	3	1999505	In-house
S.ang	P3F9	51	38.76	98310	4	1948383	In-house
S.ang	P4B9	47	38.87	93743	7	1906604	In-house
S.ang	P4D3S	50	38.12	95453	7	2197688	In-house
S.ang	P6E3R	51	38.96	103026	8	1989055	In-house
S.ang	Sang195	26	38.76	208864	4	1803153	In-house
S.con	ATCC27823	44	37.94	89245	8	1865687	In-house
S.con	C1366	35	38.02	175051	4	1925815	In-house

Species	Strain	# of contigs	GC content (%)	N50	L50	Assembly length (bp)	Source
S.con	C1367	23	38.04	245001	4	1903894	In-house
S.con	C1379B	20	38.05	904642	1	1809127	In-house
S.con	C1392	27	37.79	243937	4	2055302	In-house
S.con	C188	23	38.09	208571	2	1918810	In-house
S.con	M193	23	38.02	208212	2	1894475	In-house
S.con	M47	36	37.78	219608	4	1944465	In-house
S.con	M505	15	37.9	266316	2	1850718	In-house
S.con	P3E7	40	37.98	222989	4	1994695	In-house
S.con	P4C1	29	38.09	179507	2	1920852	In-house
S.con	P5B1	57	38.04	129055	6	1954537	In-house
S.con	P5B6	25	38.04	208262	2	1926495	In-house
S.con	P5C8	30	37.93	222096	4	1876952	In-house
S.int	AP1	10	37.63	353442	2	1982468	In-house
S.int	AP2	9	37.63	353817	2	1982064	In-house
S.int	ATCC27335	11	37.61	316075	2	1920136	In-house
S.int	C10	12	37.51	297456	3	2014626	In-house
S.int	C1365	28	37.66	149527	4	1867408	In-house
S.int	C1374	15	37.67	379096	2	1976475	In-house
S.int	C1377	12	37.55	1119414	1	1946658	In-house
S.int	C1390	11	37.62	306252	2	1919181	In-house
S.int	C260	11	37.59	464900	2	1947842	In-house
S.int	M331	18	37.51	468885	2	2006142	In-house
S.int	M423	18	37.57	212077	3	1959842	In-house
S.int	M470	22	37.56	226398	3	1909028	In-house
S.int	M60R	17	37.56	303792	3	2005307	In-house
S.int	P1C3	21	37.71	216677	3	1913473	In-house
S.int	P1D1	21	37.57	216989	3	1981718	In-house
S.int	P1D6	18	37.55	356353	2	1968524	In-house
S.int	P1G1	18	37.57	216989	3	1980924	In-house
S.int	P2A6	16	37.61	1128145	1	1966028	In-house
S.int	P2C9	15	37.66	339456	2	1920032	In-house
S.int	P2E3	17	37.57	292797	3	1987311	In-house
S.int	P2H6G	21	37.71	211882	3	1913253	In-house
S.int	P2H6P	18	37.52	280820	3	1984444	In-house

Species	Strain	# of contigs	GC content (%)	N50	L50	Assembly length (bp)	Source
S.int	P3A11S	18	37.47	430152	2	1957562	In-house
S.int	P3A5	23	37.52	219336	3	1986869	In-house
S.int	P3A7	12	37.57	353206	3	1986920	In-house
S.int	P3B6	9	37.57	367964	2	1986734	In-house
S.int	P3D7	11	37.57	363414	2	1987086	In-house
S.int	P3F3	19	37.52	281212	3	1984981	In-house
S.int	P3F5	11	37.57	298921	2	1986557	In-house
S.int	P3G3	20	37.52	224113	3	1985037	In-house
S.int	P3G5	15	37.57	298522	3	1987824	In-house
S.int	P3H3	22	37.52	281219	3	1986102	In-house
S.int	P4A2	18	37.55	220495	2	1943074	In-house
S.int	P4D2	19	37.55	225768	2	1943439	In-house
S.int	P6A1	17	37.39	219302	2	2008449	In-house
S.int	P6A3	15	37.57	219074	3	1923073	In-house
S.int	P6B3	15	37.57	219074	3	1922886	In-house
S.int	P6E1	15	37.61	339493	3	1883921	In-house
S.int	P6E4	12	37.58	469750	2	1947910	In-house
S.int	P6G1	17	37.62	339477	3	1884730	In-house
S.int	P6G5	15	37.73	298498	2	1887372	In-house
S.int	Q13	22	37.47	1038282	1	1988889	In-house
S.int	Q17	20	37.47	300380	2	1988828	In-house
S.ang	C1051	1	38.97	1911706	1	1911706	NML*
S.ang	C238	1	38.23	2233640	1	2233640	NML*
S.con	C1050	1	38.13	1991156	1	1991156	NML*
S.con	C232	1	38.18	1935414	1	1935414	NML*
S.con	C818	1	38.18	1935662	1	1935662	NML*
S.int	B196	1	37.56	1996214	1	1996214	NML*
S.int	C270	1	37.65	1960728	1	1960728	NML*
S.ang	1505	52	38.56	102421	6	2086609	NCBI
S.ang	1_2_62CV	7	38.81	1137758	1	1821055	NCBI
S.ang	1043_SSUI	60	38.83	84987	8	1971454	NCBI
S.ang	1080_SANG	32	38.81	109331	5	1810732	NCBI
S.ang	557_SANG	64	38.87	75683	9	1954993	NCBI
S.ang	83_SANG	50	38.81	91024	7	1838141	NCBI

Species	Strain	# of contigs	GC content (%)	N50	L50	Assembly length (bp)	Source
S.ang	ATCC33397	9	38.63	1099967	1	1904351	NCBI
S.ang	BVI	26	38.87	122398	6	2014025	NCBI
S.ang	CCUG 39159	83	38.5	63813	12	2294507	NCBI
S.ang	ChDC B695	2	38.88	1293982	1	1853716	NCBI
S.ang	Dora7	138	38.68	120006	5	2080850	NCBI
S.ang	DSM 20563	14	38.61	369100	2	1938196	NCBI
S.ang	F0211	63	38.44	134311	6	1990786	NCBI
S.ang	gc	35	38.89	97362	6	1806763	NCBI
S.ang	gl	27	38.61	201121	4	1936371	NCBI
S.ang	J4206	1	38.62	2001352	1	2001352	NCBI
S.ang	J4211	1	38.99	1924513	1	1924513	NCBI
S.ang	LAC	61	38.41	110506	6	2097799	NCBI
S.ang	MAS624	1	38.34	2122284	1	2122284	NCBI
S.ang	OUP10	12	38.68	313242	3	1863386	NCBI
S.ang	OUP12	35	38.8	93156	6	1831256	NCBI
S.ang	OUP19	22	38.66	264398	3	2036387	NCBI
S.ang	OUP21	26	38.34	115953	5	1997328	NCBI
S.ang	OUP25	86	38.69	92228	7	1962788	NCBI
S.ang	p35gc	52	38.92	70748	8	1788808	NCBI
S.ang	S_11	24	38.85	244072	3	1850270	NCBI
S.ang	SA1	1	38.86	2036353	1	2036353	NCBI
S.ang	SK1138	13	38.61	241357	3	1958191	NCBI
S.ang	SK52	109	38.65	33673	20	1892386	NCBI
S.ang	T5	42	39.05	99521	6	1787417	NCBI
S.ang	UMB0050	43	38.55	229639	3	1974824	NCBI
S.ang	UMB0142	59	38.39	118737	3	1957958	NCBI
S.ang	UMB0220	40	38.43	118741	3	1939067	NCBI
S.ang	UMB0252	49	38.39	156410	3	1964978	NCBI
S.ang	UMB0329	53	38.38	118742	3	1957055	NCBI
S.ang	UMB0820	28	38.89	228873	3	1908765	NCBI
S.ang	UMB0839	23	38.76	502951	2	2115397	NCBI
S.con	317_SINT	48	38.02	85815	8	1852029	NCBI
S.con	783_SANG	69	37.88	77371	8	1910933	NCBI
S.con	925_SCON	124	37.91	46324	15	2033343	NCBI

Species	Strain	# of contigs	GC content (%)	N50	L50	Assembly length (bp)	Source
S.con	CCUG 46377	66	37.94	143311	6	1944826	NCBI
S.con	DD09	26	37.98	1171761	1	1850771	NCBI
S.con	KCOM1650	5	38.04	1225689	1	1965746	NCBI
S.con	SK1060	10	37.99	533492	2	1963180	NCBI
S.con	SK53	54	37.92	70189	9	1840061	NCBI
S.int	30309	12	37.51	453711	2	1956646	NCBI
S.int	32811	15	37.65	319038	2	1971034	NCBI
S.int	631_SCON	19	37.77	296563	2	1968557	NCBI
S.int	ATCC27335	10	37.66	266223	3	1951449	NCBI
S.int	BA1	48	37.73	209077	4	1964874	NCBI
S.int	F0413	8	37.63	1090009	1	1922374	NCBI
S.int	FDAARGOS_233	1	37.66	1913894	1	1913894	NCBI
S.int	JTH08	1	37.71	1933610	1	1933610	NCBI
S.int	KCOM1545	1	37.57	1908201	1	1908201	NCBI
S.int	LC4	15	37.76	284306	2	1914382	NCBI
S.int	TYG1620	1	37.55	2006877	1	2006877	NCBI

S.int: *S. intermedius*; *S.ang*: *S. anginosus*; *S.con*: *S. constellatus*