Analysis of the Role of Host Genetics in Shaping Diversity of the Murine Lung Microbiota

Dissertation
in fulfilment of the requirements for the degree
Doctor rerum naturalium
of the Faculty of Mathematics and Natural Sciences
at the Christian Albrechts University of Kiel

Submitted by Juryung Chung

Department of Evolutionary Genetics,
Max Planck Institute for Evolutionary Biology

Plön, 2021

First referee: Prof. Dr. John F. Baines

Second referee: Prof. Dr. Hinrich Schulenburg

Additional referee: Prof. Dr. Thomas Roeder

Date of oral examination: 28.06.2021

Signature: _____

Table of Contents

Affidavit

Written declaration/Erklärung

Hereby I declare that

i. apart from my supervisor's guidance, the content and design of this dissertation is the product of my own work and only using the sources listed. The co-author's contributions to specific paragraphs are listed in the thesis outline section;
ii. this thesis has not already been submitted either partially or wholly as part of a doctoral degree to another examination body, and no other materials are published or submitted for publication than indicated in the thesis;
iii. the preparation of the thesis has been subjected to the Rules of Good Scientific Practice of the German Research Foundation;
iv. an academic degree has never been withdrawn.

Author contributions

Chapter 1: Prof. Dr. John Baines and Prof. Dr. Saleh Ibrahim designed the study. Dr. Meriem Belheouane dissected the mice. Dr. Sven Künzel performed the MiSeq sequencing. Shauni Doms designed the pipeline for processing the sequences and Dr. Yask Gupta and Dr. Meriem Belheouane designed the pipeline for QTL mapping. I performed all laboratory experiments and data analyses, wrote the chapter with editing from Prof. Dr. John Baines.

Chapter 2: Dr. Danielle Harris provided a *Pelomonas* sample along with valuable discussions on the bacteria. I performed all laboratory experiments and data analyses, and wrote the chapter with editing from Prof. Dr. John Baines.

Chapter 3: Prof. Dr. John Baines and I designed the study. I planned the overall mouse breeding scheme with help from Dr. Sven Künzel. Dr. Marie Vallier and I dissected mice and preserved the tissues. Dr. Sven Künzel performed the MiSeq sequencing. I performed all

laboratory experiments, including genotyping and data analyses, and wrote the chapter with editing from Prof. Dr. John Baines.

Chapter 4: Prof. Dr. John Baines designed the study. Dr. Marie Vallier, Aleksa Cèpic, and I dissected the mice and preserved the tissues. Dr. Sven Künzel performed the MiSeq sequencing. I performed all laboratory experiments and data analyses, and wrote the chapter with editing from Prof. Dr. John Baines.

Plön, April 2021,

Juryung Chung	Prof. Dr. John F. Baines

Abstract

Although lungs were long considered as sterile, the presence and importance of microbes inhabiting the lungs are now widely recognized. Using the 15th generation ("G_{15}") of a mouse advanced intercross line (AIL) population, we performed QTL mapping in an effort to identify host genes that influence lung microbes and may play a role in lung functioning and disease susceptibility. The lung microbiota of the G_{15} mouse population was examined through bacterial 16S rRNA gene amplicon sequencing, whereby *Lactobacillus* was identified as the most abundant genus. High-precision droplet digital PCR (ddPCR) was adapted in order to refine the dataset, remove contaminants, and quantify the absolute load of the candidate genera *Lactobacillus* and *Pelomonas*. The association between host genetic loci and bacterial traits was determined using a QTL linkage mapping approach on the relative abundances of the core microbiota as well as the absolute load data of the two candidate genera. This analysis yielded candidate host genes associated with various lung diseases. For example, two candidate genes, *Il-10* and *Mk2*, were detected from mapping *Lactobacillus* load and were especially interesting as both are involved in regulating inflammatory responses. As such, these genes are known to be involved in lung diseases such as lung cancer, cystic fibrosis, and allergic airway inflammation. To test for potential interaction and feedback between these genes and *Lactobacillus*, their gene expression levels were measured using ddPCR, which revealed a significant negative correlation between *Lactobacillus* load and *Mk2* expression. A follow-up study was performed with *Il-10* knockout (KO) mice, but the initial results revealed no significant differences in lung microbiota composition according to *Il-10* genotype. Finally, a candidate gene analysis was performed in wild type versus *B4galnt2* knock out mice, as this blood group related gene was previously shown to influence the gut microbiota, and is also predicted to be expressed in the lung. This analysis also failed to detect an influence of host genotype. Through this in-depth study of the murine lung microbiota, we generated novel working hypotheses for the impact of candidate bacteria and disease genes on host inflammatory responses in the lung, which will be explored in future work.

Zusammenfassung

Obwohl die Lunge lange Zeit als steril angesehen wurde, sind das Vorhandensein und die Bedeutung von Mikroben, die die Lunge besiedeln, heute allgemein anerkannt. Unter Verwendung der 15. Generation ("G15") einer Mauspopulation einer fortgeschrittenen Kreuzungslinie (AIL) führten wir eine QTL-Kartierung durch, um Wirtsgene zu identifizieren, die die Lungenmikroben beeinflussen und möglicherweise eine Rolle bei der Lungenfunktion und Krankheitsanfälligkeit spielen. Die Lungenmikrobiota der G15-Mauspopulation wurde mittels bakterieller 16S rRNA-Gen-Amplikon-Sequenzierung untersucht, wobei *Lactobacillus* als die an der häufigsten vorkommenden Gattung identifiziert wurde. Hochpräzise digitale Tröpfchen-PCR (ddPCR) wurde angepasst, um den Datensatz zu verfeinern, Verunreinigungen zu entfernen und die absoluten Mengen der Kandidatengattungen *Lactobacillus* und *Pelomonas* zu quantifizieren. Die Assoziation zwischen genetischen Wirtsloci und bakteriellen Merkmalen wurde mithilfe eines QTL-Linkage-Mapping-Ansatzes auf den relativen Häufigkeiten der Kernmikrobiota sowie den absoluten Mengen-Daten der beiden Kandidatengattungen bestimmt. Diese Analyse ergab Kandidaten-Wirtsgene, die mit verschiedenen Lungenkrankheiten assoziiert sind. Zum Beispiel wurden zwei Kandidatengene, *Il-10* und *Mk2*, bei der Kartierung der *Lactobacillus*-Menge entdeckt und waren besonders interessant, da beide an der Regulierung von Entzündungsreaktionen beteiligt sind. Von diesen Genen ist bekannt, dass sie an Lungenerkrankungen wie Lungenkrebs, zystischer Fibrose und allergischer Atemwegsentzündung beteiligt sind. Um eine mögliche Interaktion und Rückkopplung zwischen diesen Genen und *Lactobacillus* zu testen, wurden ihre Genexpressionslevel mittels ddPCR gemessen, was eine signifikante negative Korrelation zwischen der *Lactobacillus*-Menge und der Mk2-Expression ergab. Eine Folgestudie wurde mit *Il-10*-Knockout (KO)-Mäusen durchgeführt, aber die ersten Ergebnisse zeigten keine signifikanten Unterschiede in der Zusammensetzung der Lungenmikrobiota bezüglich des *Il-10*-Genotyps. Schließlich wurde eine Kandidatengen-Analyse in Wildtyp- und *B4galnt2*-KO-Mäusen durchgeführt, da dieses blutgruppenverwandte Gen zuvor gezeigt wurde, dass es die Darm-Mikrobiota beeinflusst und auch in der Lunge exprimiert werden sollte. Auch bei dieser Analyse konnte kein Einfluss des Wirtsgenotyps festgestellt werden. Durch diese eingehende Studie des Lungenmikrobioms von Mäusen haben wir neue Hypothesen für den Einfluss von

Bakterienkandidaten und Krankheitsgenen auf die Entzündungsreaktionen des Wirts in der Lunge generiert, die in zukünftigen Arbeiten untersucht werden sollen.

General Introduction

Microbiota & microbiome

Mammals carry diverse communities of microorganisms on barrier organs such as the skin, gut, oral cavity, vagina, as well as lungs. This group or community of microorganisms is called microbiota, and together with their genetic material, it is known as the microbiome. Microbiota vary among different body sites and also between individuals, but it is important to maintain homeostasis within each community (Bolnick et al., 2014; Larsson et al., 2012; Rakoff-Nahoum et al., 2004; Reikvam et al., 2011). Studies on gut microbiota often highlight the importance of a balance in microbial community structure, since the loss of homeostasis or a bloom of certain taxa is associated with various diseases such as colitis and cancer (Liu et al., 2016; Zackular et al., 2013). As community compositions are distinct, each site can often be characterized by particular microbes and their abundances. For example, human gut microbiota studies are mostly based on fecal samples, which revealed high abundances of Bacteroidetes and Firmicutes at the phylum level, and *Faecalibacterium*, *Bacteroides*, *Prevotella*, *Ruminococcus*, *Bifidobacterium*, and *Roseburia* at the genus level (Arumugam et al., 2011; T. L. A. Nguyen et al., 2015; Vandeputte et al., 2017). On the other hand, the mouse gut is shown to be mainly inhabited by Bacteroidetes and Firmicutes at the phylum level and *Lactobacillus*, *Bacteroides*, *Alistipes*, *Clostridium*, and *Helicobacter* at the genus level (Benson et al., 2010; Linnenbrink et al., 2013). Gut microbiota have been shown to play critical roles for the host such as defense and immunity (Stecher et al., 2010), aging (Ottaviani et al., 2011), drug metabolism (Suwal et al., 2018; Zimmermann et al., 2019), and inflammation/tumorigenesis (Mendes et al., 2019; Zackular et al., 2013).

Microbiota & the host

Variation in microbial communities is affected by many factors such as the environment, diet, and genetics. For gut microbiota, diet (Rodríguez et al., 2015; Schnorr et al. 2014; Smits et al. 2017; Snijders et al. 2016; Wu et al. 2011), intake of antibiotics (Dethlefsen, Huse, Sogin, & Relman, 2008; Maurice, Haiser, & Turnbaugh, 2013), probiotics (Suwal et al. 2018), mode of delivery (Jakobsson et al., 2014; Salminen, Gibson, McCartney, & Isolauri, 2004), and lifestyle (Farré et al. 2018; Obregon-Tito et al. 2015) can affect the community

composition. Among these different factors, the association between microbiota and host genetics (Blekhman et al., 2015) is the primary interest of our study. Benson et al. (2010) used the 4th generation of a mouse intercross population in order to identify the host-microbiota relationship, where the mixed effects model inferred a high fraction of variation that can be explained by genetic factors, while smaller fractions were explained by cohort, family, and litter. Their quantitative trait loci (QTL) analysis revealed several QTL in association with taxa belonging to the core microbiota of the gut, such as *Helicobacter* and *Lactobacillus*. Host genetics also plays a role in microbiota and diet, as the changes in the microbial community in the gut can contribute to obesity and metabolic traits (Org et al., 2015). It is important to note that host genetics is associated with microbiota of other body sites as well, such as the skin (Belheouane et al., 2017; Blekhman et al., 2015).

Microbiota & host co-evolution

As microorganisms play various crucial roles within the host, they have and had several ways to influence host evolution. Several studies have provided evidence of co-evolution between gut microbiota and hosts, as the microbiota diversified dietary niches and shaped host phenotypic plasticity (Alberdi et al., 2016; Groussin et al., 2017; Moran et al., 2019).

A good example of the effect of gut microbiota on the host's niche expansion is the koala, which harbor microbiota including bacteria that are able to digest *Eucalyptus*. Without the bacteria, the consumption and digestion of the plant would be otherwise difficult only with the host's enzymes (Moeller & Sanders, 2020). Humans also have benefited from gut microbiota, particularly with the consumption of milk and digesting or fermenting lactose with lactose-fermenting microorganisms such as *Bifidobacterium*. While lactase, an enzyme to digest lactose, is produced until the weaning phase in mammals, prolonged consumption until- and throughout adulthood in humans has led to an adaptive evolution of persistent lactase production through several mutations in the *LCT* gene (Gerbault et al., 2011). Variation at this gene was found to be strongly associated with *Bifidobacterium* abundance, which may suggest a role of the bacteria in contributing to a shift in prolonged milk consumption as a source of energy for human adults (Goodrich et al., 2016).

While it has been previously shown that host genetics play a role in shaping microbiota (Campbell et al., 2012; Org et al., 2015; Snijders et al., 2016), a shift in microbial composition can provide the host with adaptive plasticity. The gut microbiota of mice exposed to cold

temperature showed a shift in microbial composition, and when mice at room temperature were inoculated with the shifted microbiota, their physiology changed to adapt to colder temperature (Chevalier et al., 2015). Such studies provide evidence suggesting that microbiota and its composition could serve as a means by which the host can evolve adaptive plasticity.

Studying the lung microbiota

While the lungs of diseased individuals were long known to carry microorganisms, healthy lungs were considered sterile until recently (Charlson et al., 2011). With increased interests in the human microbiome and acknowledging its importance to health and fitness, the Human Microbiome Project was launched in 2007, which included the upper respiratory tract, such as mouth and esophagus, as well as the skin, stomach, colon, and vagina (Turnbaugh et al., 2007). The importance and impact of the microbiome was noted in various aspects such as host immune responses, physiology, and metabolism, but as noticeable, lungs or the lower respiratory tract were not included (Turnbaugh et al., 2007).

Previously, the lung microbiome was analyzed with sputum samples or even bronchoalveolar lavage fluids (BALF) using culture-based methods. BALF samples are collected as a bronchoscope enters through either the mouth or nose in order to "flush" the lungs and collect back the fluid. This method has been clinically excellent, but can be challenging for microbiome analysis due to the chance of contamination from the oral or nasal cavity. This method was applied to mice but, there were difficulties and challenges in comparing the microbial community based on BALF and lung tissue simultaneously, as the tissue already has been "flushed" and thus shows lower diversity and richness as well as fewer number of OTUs compared to BALF (Barfod et al., 2013).

Sputum sampling is another method, but also carries a potential for contamination as it travels through the trachea and mouth upon sample collection. Cultivation using sputum or BALF samples also has limitations, as it may provide biased information based on limited or specific bacteria that are able to grow in the given media. Otherwise this requires strain- or species-specific media and is overall very time-consuming. To date, sputum is still collected and used for human lung microbiome studies because it is a non-invasive method for patients and has proven differences in microbial composition between lung cancer patients and controls (Hosgood III et al., 2014). Another study was able to identify potential biomarkers for lung cancer through the metagenomic sequencing of sputum samples (Cameron et al., 2017). Due

to anatomical connection, the microbial species found in the upper airways or oral cavity were detected in the lower respiratory tract (LRT) of humans, especially those with airway diseases such as cystic fibrosis (CF) (Huang & Lynch, 2011). This is most likely due to increased chances of micro-aspiration, which could also occur in healthy individuals, but in less frequency (Bassis et al., 2015). This is also referred to as "island biogeography theory" whereby bacterial species from the "mainland" (mouth) are continuously imported to and affect the "island" habitat (LRT or the lungs), which is most likely to exert selective advantage for specific taxa (Lomolino, 2000; Whiteson et al., 2014). So even if there are transfers of microbes into the LRT from the upper airways, selective pressures in the LRT shape the composition and variability towards specific bacterial species (Dickson et al., 2013), which then allows the formation of distinct microbiota in the LRT. Also due to the nutrient-poor environment in the lungs (compared to e.g. the gut), stronger selection forces may act upon the pool of colonizing microorganisms, selecting those capable of surviving within limited environmental niches (Marsland & Salami, 2015).

With the advancement of technologies and a shift in methodologies from culture- to molecular-based, studies have proven the presence of distinct microbiota in the lungs of healthy humans and various mammals (Bassis et al., 2015; Beck et al., 2012; Charlson et al., 2011; Yun et al., 2014). PCR and sequencing allow an in-depth exploration of microbial diversity with reduced bias. Although the composition varies among different studies and individuals, the major bacterial phyla in the human lungs include Proteobacteria, Firmicutes, and Bacteroidetes. Murine lung microbiota, which also varies among studies and individuals, are mainly composed of Firmicutes, Proteobacteria, Actinobacteria, and Bacteroidetes at phylum level (Barfod et al., 2013; Kostric et al., 2017; Yun et al., 2014). Additionally, it was reported that there is a low biomass in the lungs, indicating a smaller number of resident microbes (Charlson et al., 2011). Although the healthy lung microbiota maintains diversity, such low biomass leads to difficulty in sampling and analyzing due to an increased risk of contamination compared to samples with high biomass such as the gut. Various types of sequencing methods are applied with great resolution and sequencing depth in order to accurately capture the microbial community as well as different methods for effective sequence-based analysis.

We are now informed of the variability in lung microbiota, but our understanding of the function and the role of lung microbiota in health and disease is yet poorly understood.

Similar to gut microbiota, disease and infection are likely to influence and disturb the lung microbial community (Heirali et al., 2017; Iwai et al., 2014; Rogers et al., 2005; Segal et al., 2016). It is unclear, however, whether the dysbiosis or changes in microbial composition is solely a bifactor of the disease or whether such changes induce or exacerbate the disease state by interacting with the host. Although microbiota on such a barrier organ is known to affect tissue homeostasis and defense, we are still on the path to understand the function of the lung microbiota.

The impact of lung microbiota in diseases

As dysbiosis in the gut microbiota has displayed a link to health and disease, similar observations can be found in the lungs. Previous studies have examined microbial communities in the lungs of patients with lung diseases such as cystic fibrosis (CF), asthma, and chronic obstructive pulmonary disease (COPD). First, CF is one of the (well-known) lung diseases, which develops due to a mutation in the cystic fibrosis transmembrane conductance regulator gene, with symptoms including airway infection, mucus burden, impaired liver and pancreatic function, and more. Major pathogens found in CF patients include *Pseudomonas aeruginosa*, *Burkholderia* species, *Staphylococcus aureus*, *Strenotrophomonas maltophilia*, and *Acrhmobacter xylosoxidans* (LiPuma, 2010). However, the major cause of disease is still unknown. Whelan et al. (2017) examined and collected sputum samples of six CF patients over the course of up to one year in the context of three clinical stages, Treatment, Intermediate, and Stable. Although their longitudinal examination did not reveal the cause of the onset of CF exacerbations – acute worsening of symptoms – in these patients, they found significant differences in microbiota among different clinical stages. This reveals microbial compositional changes occurring in disease states. As the host's immune system is modified in these disease states, physiological changes would occur.

Another study by Coburn et al. (2015) examined lung bacterial community in relation to age and disease stages which revealed age-related diversity and structure change. Microbial diversity and lung function regressed with age as the age group of 10 years or younger showed the highest diversity. This study only examines the relationship between lung microbial structure and disease state in different age cohorts, but we are able to see the link between the diversity of microbiota and lung function, thus confirming the impact of lung microbiota on the host.

Another well-known lung disease, asthma, is also characterized by inflamed airway and intermittent wheezing. It has been proposed that the modern industrialized environment, antibiotic treatment, and excessive hygiene are potential causes of the disease (Noverr et al., 2005). Teo et al. (2015) established the association between asthma and modified composition of airway microbiota through the examination of the nasopharyngeal (NP) microbiome in infants. Early colonization of NP by *Streptococcus* at approximately 2 months of age was associated with asthma development at later ages. They suggested the use of antibiotics at early age may affect the airway microbiome, leading to airway infection and asthma. Asthmatic airways were also shown to carry significantly higher load of Proteobacteria, *Prevotella*, and *Haemophilus* compared to healthy controls (Hilty et al., 2010), which is also in accordance with a previous study (Bisgaard et al., 2007).

Lastly, COPD is one of the leading causes of death worldwide and is another type of lung disease characterized by breathing difficulties such as short breath and coughing with increased sputum. The major cause is smoking, although it does appear in people who never smoked, which may be partially due to environmental factors such as dust. Increasing evidence suggests that the lung microbiota plays a role in the progression and severity of COPD (Huang & Lynch, 2011; Huang et al., 2014). Sze et al. (2012) examined the lung tissue microbiome in four different groups – non-smokers, smokers without COPD, patients with severe COPD, and patients with CF – in order to detect and characterize differences in various conditions. They detected a significantly higher bacterial load in CF lungs, but the diversity was significantly lower than other three groups, non-smoker, smoker, and COPD. In the COPD group, a significantly higher abundance of Firmicutes was detected, which was mainly driven by *Lactobacillus*. Distinctive differences in the microbiome between healthy and diseased lungs are thus confirmed through numerous studies. The above findings were only made possible with culture-independent methods, since not all bacteria can be cultivated, which can only lead to biased results.

Aims of the thesis

In this thesis, I aimed to establish a generalized description of the murine lung microbial community and study the role that variation in host genetics plays upon the structure of the microbiota by pinpointing candidate genes using QTL mapping analysis. By pinpointing candidate genes within the identified genomic intervals, further hypotheses can be generated

regarding the evolutionary forces leading to host genetic-based variation for associated microbial traits. I proceed with the study with 4 chapters, each of which examines the lung microbiota in mouse models from different angles. **F**irst, I provide the overall description of the murine lung microbiota with sequence-based methods in an advanced intercross lines (AIL) taken to the 15th generation of intercrossing (i.e. "G_{15}") and examine the variation in the microbial community in relation to host genetics using QTL mapping. Second, I refined the microbial traits by examining them at the absolute abundance level using droplet digital PCR (ddPCR). Taking this data as a trait, QTL mapping was performed, which revealed two strong candidate genes. The identified host candidate genes were prioritized based on their relation to known chronic lung disease genes/pathways. Gene expression was assessed and analyzed using ddPCR in order to further examine the differential expression of candidate host genes as a possible mechanism mediating the association between genetic markers and microbial traits. Third, the initial results of a follow-up study on the interleukin-10 (*Il10*) gene are presented, which was performed based on the mapping results in order to examine the association between host genetics and the candidate genus *Lactobacillus* as well as microbial community structure. Lastly, I present the overview of the lung microbiota in mice according to host genotype at another candidate gene, the blood group-related gene *B4galnt2*.

Chapter I.

Quantitative Trait Locus (QTL) mapping of murine lung-associated bacterial traits

Introduction

Various communities of microbes inhabit barrier organs of mammals, including the skin, gut, vagina, as well as upper and lower airways. These communities of microbes present at given site are called "microbiota", and can be taxonomically identified using the 16S rRNA gene as a phylogenetic marker. Not to be confused with microbiota, "microbiome" refers to these microbes along with the collection of their genome. Studying the microbiome of humans is increasingly important to understanding disease, behavior, physiology, immunity, and more, as exemplified by the launch of a Human Microbiome Project (HMP) (Turnbaugh et al., 2007). By the time HMP was first launched, lungs were still considered sterile or to carry microbes only in the context of disease (Hilty et al., 2010; Rogers et al., 2005). It is now however acknowledged that lungs carry a distinct microbiota even in the healthy state (Beck et al., 2012; Charlson et al., 2012; Charlson et al., 2011; Kostric et al., 2017; Segal et al., 2013). Similar to the gut or other tissue sites, the human lung microbiota also varies intra- and inter-individually, while remaining consistent in the presence of major host-associated taxa such as Firmicutes and Bacteroidetes at the phylum level, and *Pseudomonas*, *Streptococcus*, *Prevotella*, *Fusobacteria*, and *Veillonella* at the genus level (Bassis et al., 2015; Charlson et al., 2012; Charlson et al., 2011; Dickson et al., 2017; Erb-Downward et al., 2011).

The lung microbiota can be affected by various factors that affect its diversity and structure. Such alterations in community composition and/or diversity can lead to susceptibility to disease, infection, and even changes in lung tissue architecture (Boutin et al., 2017; Coburn et al., 2015; Rogers et al., 2005; Roggenbuck et al., 2016; Yun et al., 2014). Environmental factors may include allergens (Noverr et al., 2005), air pollutants (Li et al., 2017), cigarette smoke (Erb-Downward et al., 2011), and lifestyle (Ege et al., 2011; Fujimura et al., 2010; Heirali et al., 2017; Marra et al., 2009; Russell et al., 2013; Yu et al., 2016). Human subjects who smoke or have lung disease, i.e., chronic obstructive pulmonary disease (COPD), were found to have lower microbial diversity compared to healthy individuals (Erb-Downward et al., 2011). As susceptibility to airway diseases is caused by both environmental and genetic factors, it is intriguing to note the crucial role of microbiota within the host (Igartua et al., 2017; M. Y. Lim et al., 2016; Ren et al., 2018). Indeed, lung microbiota can affect the susceptibility to-, or severity of lower respiratory tract infection (Dumas et al., 2018; Gollwitzer et al., 2014; Remot

et al., 2017; Teo et al., 2015), regulate inflammatory status/response (Segal et al., 2016), and may also serve as a biomarker for lung cancer (Hu et al., 2020).

Among different factors, the association between host genetics and the associated microbiota (Blekhman et al., 2015) was the primary question of my study. Benson et al. (2010) pioneered such studies through genetic mapping using a G_4 intercross population of mice, where a mixed effects model inferred a high fraction of variation that can be explained by genetic factors, while smaller fractions were explained by cohort, family, and litter. Their quantitative trait loci (QTL) analysis also revealed several loci in association with core taxa of inhabiting the gut, such as *Helicobacter* and *Lactobacillus* (Benson et al., 2010). Host genetic factors also associate with microbiota together with diet, as changes in the microbial community in the gut can contribute to obesity and metabolic traits (Org et al., 2015). It is important to note that host genetics is associated with microbiota of other body sites such as skin as well, affecting diversity, composition, and intra- and inter-individually (Belheouane et al., 2017; Blekhman et al., 2015).

In this study, we aimed to establish the concept of a murine lung microbiota, by examining the role of host genetic variation in a G_{15} advanced intercross line (AIL), using 16S rRNA gene sequencing and QTL mapping approaches as previously developed using the same resource for the skin (Belheouane et al., 2017; Srinivas et al., 2013). The AIL was produced by intercrossing three laboratory mouse strains, MRL/MpJ, NZM2410/J, and BXD2/TyJ, and one wild-derived strain, CAST/EiJ, for 15 generations with equal sex and strain distributions (Jackson Laboratory (Maine, USA)). These mouse strains are known to be prone to various autoimmune disorders either in later age or spontaneously (early onset) (Andrews et al., 1978; Browman & Crabbe, 2000; Crabbe et al., 1994; Kanno et al., 1992; Rudofsky et al., 1993). Thus, mice with such backgrounds could allow us to examine microbial traits that are associated with loci possibly also linked to (lung) disease susceptibility. Accordingly, Belheouane et al. (2017) typed a total of 270 mice for 53,203 informative SNPs, thus creating a great marker density for use in QTL linkage mapping analysis and making it possible to pinpoint candidate genes that take part in shaping the composition and abundance of members of the lung microbiota.

16S rRNA gene sequencing enables the precise identification of bacteria or entire microbial communities present in a given sample. This has become a target for taxonomy and phylogenetic studies, as the gene contains both highly conserved regions that enable the design of “universal” primers as well as faster evolving “hypervariable” regions that can provide

taxonomic resolution (Doms et al., 2018). However, because PCR with universal primers will in theory amplify any genetic material present in a sample and the lung microbiota, while present, nonetheless clearly displays a low biomass, contamination in such studies can be an issue (Drengenes et al., 2019). In order to control for contaminants, our colleagues in Lübeck used serial dilutions of a mock community and performed 16S rRNA gene sequencing, which led them to observe an increased relative abundance of the putative contaminant genera *Halomonas* and *Shewanella* with increasing dilution factors (Künster et al., unpublished). Based on these results, we hypothesized that we will see a similar pattern of higher fraction of contaminants in samples with lower bacterial counts. Accordingly, we here adapted droplet digital PCR (ddPCR) in order to measure the bacterial load in each sample and examine its correlation with the relative abundance of contaminant taxa. ddPCR is a highly sensitive quantitative PCR method to detect target genes, which works by partitioning a sample into approximately 20,000 droplets, and can be especially suitable for those samples with low biomass. Details for ddPCR can be found in **Chapter 2**.

Overall, we defined the major members of murine lung microbiota in our G_{15} AIL mouse population and identified numerous genomic regions and genes associated with lung microbial abundances. Genes within the genomic regions were often related to various lung diseases, including non-small cell lung cancer (NSCLC). The identification of contaminants proved to be a critical step in analyzing the low bacterial biomass environment of the lung.

Aims

- Define the microbiota in the murine lung through identifying true residents.
- Identify host-microbiota associations using a QTL linkage mapping approach on the relative abundances of core microbial traits.
- Identify and characterize candidate genes within the QTL confidence intervals associated to lung disease and functioning.

Results

Overall lung microbiota composition

For our study, mice from the 15th generation of an AIL population ("G_{15}") were used (Belheouane et al. 2017). DNA and RNA were extracted from a total of 284 samples and subjected to 16S rRNA gene amplicon sequencing targeting V1-V2 region. RNA samples were reverse-transcribed, amplified, and sequenced in order to improve detection of true resident bacteria, as the lungs are known to carry low bacterial load and may be more greatly influenced by contaminant taxa at the level of DNA. From sequencing, we obtained 582,178 unique sequences after the dereplication process, which helps to identify unique or duplicate sequences to be grouped (which then enhance the analysis speed and efficiency). Samples were rarefied to an even sampling depth of 4,000, which resulted in the removal of one sample with 44 reads.

The most abundant phyla include Firmicutes, Proteobacteria, Actinobacteria, and Bacteroidetes (Fig. 1A), in descending order. In comparison, negative extraction controls (NECs), which are blank controls included starting from the tissue extraction step, display a higher abundance of Proteobacteria and Actinobacteria compared to lung samples. The most abundant genera in the lungs are *Lactobacillus*, unclassified *Lachnospiraceae*, *Halomonas*, *Pelomonas*, unclassified *Clostridiales*, and *Shewanella* (Fig. 1B), with *Lactobacillus* being present in 97% of samples.

Halomonas and *Shewanella* are considered to be common lab contaminants (Barfod et al., 2013; Roggenbuck et al., 2016; Weyrich et al., 2019) and are found at appreciable abundance throughout the dataset. Thus, it was necessary to formally assess whether these were contaminants or true lung residents. Based on our hypothesis of an increased fraction of contaminants as the true bacterial load decreases, we assessed the total bacterial load in each sample using droplet digital PCR (ddPCR) (details on the optimization and methods of ddPCR can be found in **Chapter 2**). First, Spearman correlation tests were performed between the total bacterial load and relative abundances of each of the above genera, whereby we found significantly negative correlations for both *Halomonas* and *Shewanella* ($p < 0.05$) (Fig. S1). In contrast, correlations between the total bacterial load and relative abundance of *Lactobacillus*, which is a common microbe in murine lungs (Barfod et al., 2013; Singh et al., 2017; Yun et al., 2014) and the most abundant genus in our samples, was significantly positive ($p < 0.001$) (Fig. S1).

Next, we employed the "decontam" package, which is also based on our working hypothesis that an increased fraction of contaminants should appear as the true bacterial load decreases. Using the frequency method with threshold of 0.1 in R package applied to our ddPCR-derived load estimates, we identified operational taxonomic units (OTUs) belonging to *Halomonas* and *Shewanella* as contaminants. Thus, *Halomonas* and *Shewanella* were accordingly removed from the dataset, following previous studies that identified them as contaminants (Greathouse et al., 2018; Roggenbuck et al., 2016). After filtering, *Lactobacillus*, unclassified *Lachnospiraceae*, *Flavobacterium*, unclassified *Clostridiales*, and *Pelomonas* are the five most abundant taxa at the genus level (Fig. 1B). At the OTU level, *Pelomonas*_OTU4 is the most abundant OTU, followed by *Flavobacterium*_OTU8, *Lactobacillus*_OTU3, and *Lactobacillus*_OTU2, and *Pseudomonas*_OTU6.

Comparison to skin microbiota

As an additional means of verification, the lung microbiota of the AIL G_{15} mice was compared to the skin microbiota of the very same animal. This enables the examination of differences between the two distinct tissue sites, but also serves as a means to confirm and differentiate lung microbiota from the skin, which also has a low bacterial load. For this step the sample size was reduced to 252 in order to obtain the overlap in data between the two tissue sites. In contrast to the lung, the skin microbiota is majorly composed of taxa such as unclassified *Betaproteobacteria*, followed by unclassified *Lachnospiraceae*, unclassified *Clostridiales*, unclassified *Bacteroidetes*, and unclassified *Bacteroidales*, some of which appear at very low abundances in the lungs, (Fig. 2A). A significant compositional difference ($p < 0.001$) is confirmed between the two bacterial populations through beta diversity measures, using Constrained Analysis of Principal Coordinates (CAP) with Bray-Curtis dissimilarity (Fig. 2B). CAP with Bray-Curtis allows quantification of compositional differences between two samples or sites by taking both abundance and presence/absence into account.

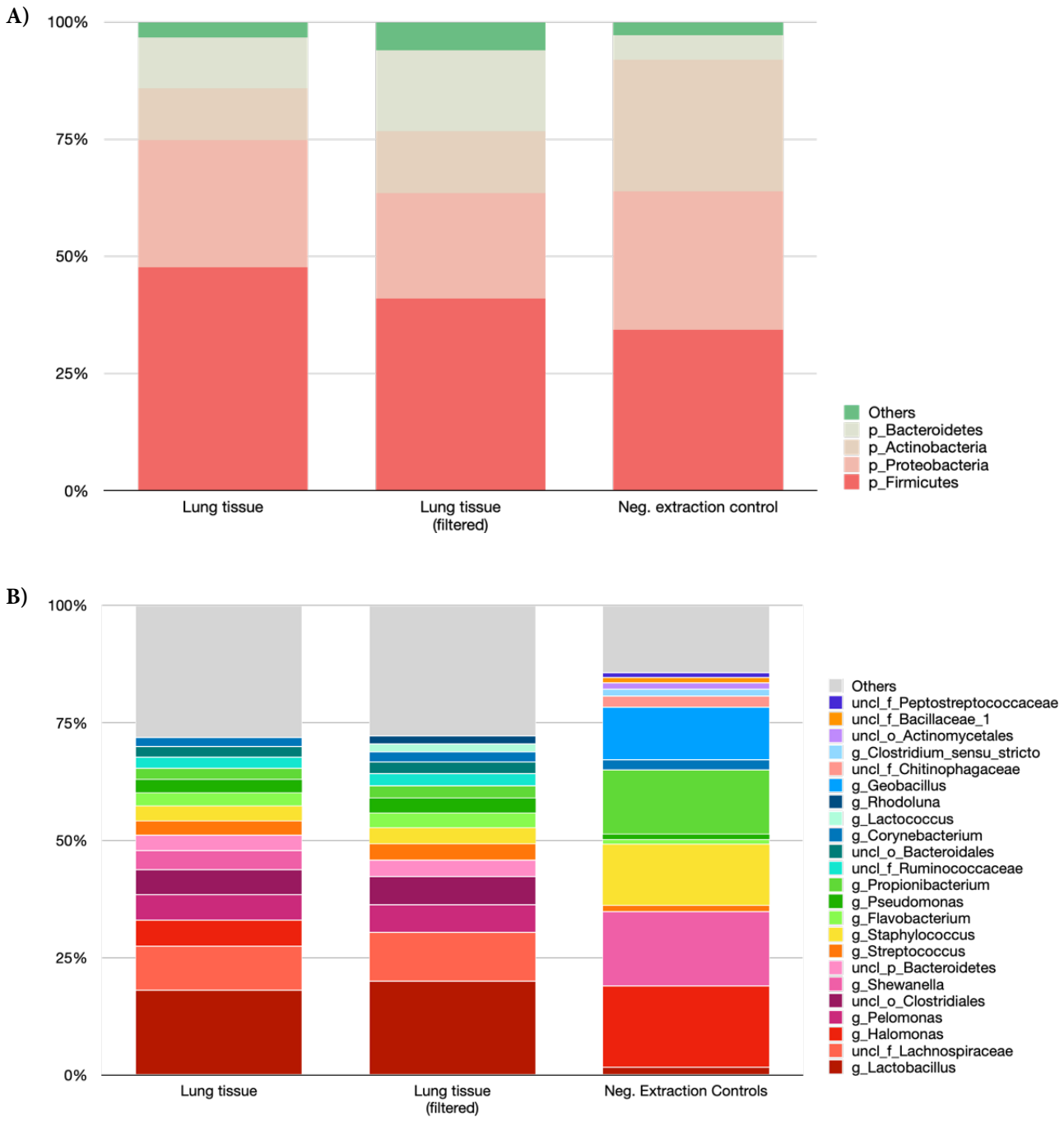

Figure 1. Comparison of microbial composition between lung tissue samples and negative extraction controls. Comparisons were made at the A) phylum and B) genus levels. Compositions of lung tissue samples represent before and after ("filtered") the removal of contaminant taxa. Lung tissue samples (n=252) and negative extraction controls (n=3).

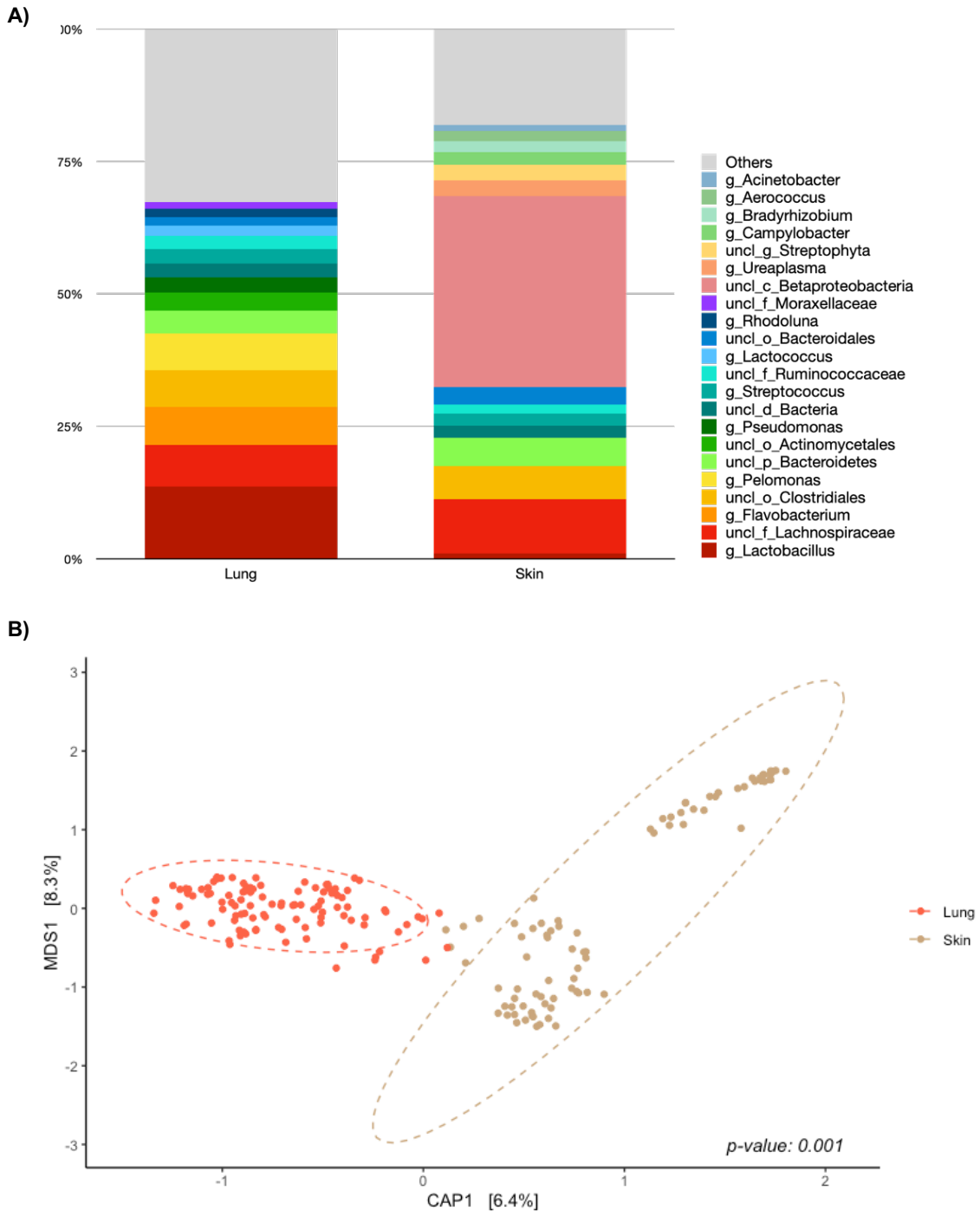

Figure 2. Bacterial community comparison between lung and skin. (A) Composition at genus level. (B) Constrained principle coordinate analysis based on the Bray-Curtis index. Corresponding/matching samples (n=252) were used to compare the two communities.

Core measurable microbiota & summary statistics

After removing contaminants, samples were rarefied to an even sampling depth and a core measurable microbiota (CMM) (Benson et al., 2010) was defined at the genus to phylum and OTU levels. From 16S rRNA gene amplicon sequencing, a large dataset is obtained, whereby many OTUs are present in only a few or single samples. Defining a CMM removes taxa with rare prevalence/abundances. By removing these microbes, analyses can be focused on taxa that are present throughout the dataset, which have a higher probability to be resident bacteria. In this study, the CMM was defined to include taxa that are present in at least 10% of the samples with at least 1% relative abundance. At the genus and OTU levels, the number of CMM taxa represents a small fraction (4.95% and 0.55%, respectively), but their abundances in total represent 91.18% and 49.82% of the entire genus and OTU tables, respectively. On average, genera of the CMM represents 82.4% of each sample's microbial community. The CMM includes a total of 72 taxa from genus to phylum level and 36 OTUs.

Once the CMM was defined, SourceTracker (Knights et al., 2011) was performed as a final step to reduce chances of including contaminants in the dataset, which identifies the level of contamination in each sample using NECs as a "source". After running SourceTracker, nine samples with a fraction of contamination above 20% were removed from the dataset.

Summary statistics were calculated on the relative abundances of each CMM trait (for 243 samples) at the genus to phylum and OTU levels in order to examine variability within each trait (Table S1, 2). *Lactobacillus* is the most abundant genus with an average abundance of 13.20% and ranges from 0.23 to 92.28% throughout the dataset. Others also exhibit considerable variation such as *Pelomonas* ranging from 0 to 87.38%, *Flavobacterium* from 0 to 51.2%. Genera with lower abundance display a lower magnitude of variation such as *Corynebacterium* ranging from 0.03 to 19.35% and *Staphylococcus* from 0 to 4.78%. At the OTU level, *Pelomonas*_OTU4 is the most abundant, ranging from 0 to 87.38% with an overall average of 7.16%. Other OTUs such as *Flavobacterium*_OTU8 ranges from 0 to 51.2%, *Lactobacillus*_OTU3 ranges from 0 to 71.28%, and *Lactobacillus*_OTU2 ranges from 0 to 24.6%. For both the genus and OTU level, traits with higher mean abundance show greater variation (Pearson's corr: $p < 0.001$, r = 0.7975; $p < 0.001$, r = 0.9101, respectively).

In order to perform QTL mapping, the relative abundances of the CMM traits were $\log_{10}$-transformed for input in a mixed effects model in which we incorporate sex, age, and cage effects on the inter-individual variation of the CMM traits. The $\log_{10}$-transformed relative

abundances were used as responses, sex and age as fixed effects, and cage as a random effect. The percentages of total variance explained by these effects were then estimated. At the genus level, these percentages vary from 0 to 15.68%, where, for example, 15.68% of total variance is explained by cage in *Flavobacterium*, whereas 0% is explained for *Pelomonas* and *Streptococcus*. At the OTU level, 14.28% of the total variance is explained by cage in *Flavobacterium*_OTU8 and 0% in *Rhodoluna*_OTU10. Of note, the majority of the variance in our CMM traits remains as residual variation even after we account for sex, age, and cage effects (Table S3, 4 for taxa and OTU levels, respectively).

QTL mapping of the lung microbiota

In order to identify the host genomic regions associated with the variation in lung microbial composition, linkage mapping was performed on the defined CMM traits. In total, 242 samples were included in the mapping. A total of 108 CMM traits, from the genus to phylum and OTU levels, were included for mapping. Overall, four suggestive ($p < 0.1$) QTL (Belheouane et al., 2017) were detected, including three at the genus level, and one from the OTU level (Table 1). The confidence intervals range from 0.14 to 1.9 Mb, and the number of genes included in a given interval ranges from 3 to 29. For example, a region between 62.14 and 64.04 Mb on chromosome 11 is associated with variation in *Alpinimonas,* and a region between 76.66 to 76.80 Mb on chromosome 5 is associated with variation in *Microbacterium*. The number of genes varies within each candidate region with some found adjacent to the peak SNP. For example, genes in proximity to peak SNPs include the *Pmp22* gene on chromosome 11 (*Alpinimonas*) and the *Clock* and *Pdcl2* genes on chromosome 5 (*Microbacterium*). At the OTU level, OTU92_*Alpinimonas* is associated with the same region as *Alpinimonas* at the genus level, but with a different peak SNP.

Candidate gene analysis

With the high resolution stemming from 15 generations of intercrossing, we were able to obtain narrow confidence intervals and identify candidate genes (Table 1). Most genes within the intervals were related to one or more of the following processes: immune response, inflammatory response, cell apoptosis, and DNA repair. For example, the *Pmp22* gene associated with *Alpinimonas* plays a role in lung cancer, where its suppression can help reduce

progression of lung cancer (Qu et al., 2015). A list of candidate genes from significant and suggestive QTLs and their functions are reported in Table 2.

Table 1. QTL statistics of the CMM bacterial traits of G_{15} population.

Trait	Category	Chr	Peak SNP	Position	LOD score	CI (Mb)	Phenotypic Variance	Size	# of Genes
Alpinimonas	Genus	11	JAX00028234	62.99	13.81	62.14 – 64.04	23.11	1.90	29
Kocuria	Genus	5	UNC10015185	119.36	32.79	118.59 – 119.66	9.59	1.07	4
Microbacterium	Genus	5	UNC9497824	76.74	13.85	76.66 – 76.80	7.78	0.14	3
OTU92_*Alpinimonas*	OTU	11	UNC19771585	62.81	14.32	62.14 – 64.04	10.52	1.90	28

(Chr = Chromosome)

Table 2. Genes from QTL intervals and their functions explained.

Trait	Gene	Function (& reference)
Alpinimonas	*Pmp22*	Lung cancer (Qu et al. 2015)
Kocuria	*Med13l*	Neurodevelopmental disorder (NDD) syndrome; heart disease (Ji et al., 2020)
Microbacterium	*Clock*	Pulmonary physiology & pathology (Gebel et al., 2006); inflammation (Kondratov et al., 2006); lung cancer (Papagiannakopoulos et al., 2016)
	Pdcl2	G-protein functioning (Schröder & Lohse, 2000)

Discussion

In this study we performed high-resolution QTL mapping to investigate host-microbiota associations in the murine lung. With high marker density and 15 generations of intercrossing, we were able to identify genomic regions and candidate genes that represent possible novel interactions with bacteria in the lung.

Until recently, lungs were considered sterile in healthy individuals and to carry microbes only in individuals with lung diseases or impaired lung functioning (Cabello et al., 1997; Laurenzi et al., 1961; Segal et al., 2013). As such, the lung was not initially included in the Human Microbiome Project (Turnbaugh et al., 2007). This may be derived from the concept of the structure of the respiratory tract and microaspiration that happens at higher frequency in individuals who smoke or with respiratory tract diseases, but also from the difficulties in obtaining bacteria from sputum with culture-based methods prior to adaptation of 16S rRNA gene sequencing technology (Cabello et al., 1997; Rogers et al., 2009). Nowadays, sampling of human lung microbiota is often done using bronchoalveolar lavage (BAL) fluid instead of sputum, which enhances the profiling of lung microbiota. Using an animal model such as the mouse has advantages over sputum or BALF, since we can excise the tissue directly in order to fully capture the microbiota and avoid contamination from the trachea or mouth. Through recent studies and advancement in technology, the lung microbiota is now recognized as a distinct community, but partially carrying bacteria derived from upper respiratory tract, in part due to microaspiration (Charlson et al., 2011; Dickson et al., 2017; Segal et al., 2013).

In this study, lung microbiota analysis and QTL mapping were performed at the bacterial transcript (RNA) level, in order to capture the active and thus more likely true resident bacteria in the lungs. Although both DNA and RNA can reflect the microbial community within the tissue, they provide different information, especially regarding responses to host. DNA provides an overview of the microbial community, but disregards whether bacteria are dead or alive. On the other hand, RNA-level analysis is capable of providing information of active and live bacteria in the tissue, which are more likely to be interacting with the host and immune responses. As a result, analyzing at RNA level will help reduce environmental noise in such low biomass samples, but may only report those more active at the transcription level at the sampling point (Wagner, 1994; Zakrzewski et al., 2012). As Belheouane et al. (2017) also detected greater efficiency in detecting significant associations between the host and the skin microbial traits, examining at the transcript level was appropriate in this study in order to investigate lung microbiota associated with host genetic variation and biological responses in relation to its bacterial community.

Sequencing results were transformed to relative abundance data in order to visualize and understand the data effectively, and a core measurable microbiota (CMM) was defined that constitutes taxa that are more predominant throughout the samples (Benson et al., 2010). Although sequencing identifies and provides massive data, many are only present in a single or few samples that carry low statistical importance/significance. Defining a CMM thus accounts for taxa that have higher abundance and carry greater potential to functionally interact with the host.

In order to successfully perform analysis and mapping, it was crucial to first define authentic, resident bacterial community of lung microbiota apart from contaminants and environmental noises. Indeed, there are challenges in studying samples of low biomass as the results are more easily hindered by undesirable elements, often derived from the means of sample collection or even lab reagents (Jervis-Bardy et al., 2015; Salter et al., 2014). Negative extraction controls were included in the extraction step in order to examine potential introduction of contaminants from reagents and the environment throughout the sample processing, using R packages such as “SourceTracker” and “decontam”. The two methods are different but are both useful when dealing with contamination. SourceTracker determines the fraction of OTUs or taxa derived from the “source,” such as negative extraction controls, as contaminants that are found in each true sample using a Bayesian approach (Knights et al.,

2011). This can be practical in determining "contaminated" samples, but will not be able to identify specific contaminants. Decontam, on the other hand, adapts simple de novo classification methods in order to identify contaminants based on either their frequency or prevalence. The frequency method is based on the information that contaminants are likely to be negatively correlated to overall bacterial biomass concentration (Jervis-Bardy et al., 2015; Lusk, 2014; Salter et al., 2014; Willner et al., 2012), while the prevalence method uses bacterial abundance data, since the sequences of contaminant taxa are more likely to be prevalent in controls than in true samples (Davis et al., 2018; Lazarevic et al., 2016). Decontam is known to be more useful when dealing with samples with low bacterial load (Lazarevic et al., 2016), and SourceTracker is suggested to be more suitable for samples that are well-defined or with higher-density biomass (Karstens et al., 2019). Independently from decontam, we have hypothesized based on our preliminary tests that the fraction of contaminant taxa would increase with reduction in the bacterial load, which we confirmed by using absolute/total bacterial load based on ddPCR. In addition to the frequency method of decontam, correlation tests between bacterial load and the relative abundance of the putative contaminant taxa *Halomonas* and *Shewanella* revealed a negative correlation, which is consistent with the two taxa being contaminants. *Halomonas* is a halophilic taxon and is thus resistant to salt that is contained in reagents. *Shewanella* is a Gram-negative bacterium that inhabits aquatic environments of warm and temperate regions over the world. As decontam and our own correlation test share the same principle, both tests identified those genera as contaminants. For further analyses, OTUs belonging to these taxa were removed from our dataset, following previous studies on lung microbiota or samples of low biomass (Greathouse et al., 2018; Karstens et al., 2019; Roggenbuck et al., 2016). Indeed, identification and removal of contaminant taxa enables improved observation of the genuine microbial community especially when working with samples of low biomass (Weyrich et al., 2019).

Additionally, lung samples were analyzed together with previously sequenced skin samples in order to compare and confirm distinctiveness between the two low microbial biomass tissue sites (Koskinen et al., 2017). The most dominant genus in each tissue was distinct, with *Lactobacillus* in the lungs and unclassified Betaproteobacteria on the skin. Beta diversity measures with Bray-Curtis dissimilarity confirmed the difference between the two sites. Lastly, it was intriguing to find a genus, *Pelomonas*, which found to be a unique genus in the lungs (Barfod et al., 2013) and was detected throughout the G_{15} mouse samples but not in

NECs or in skin samples. All the steps and procedures provided us with sufficient confirmation that we obtained genuine lung microbiota from lung tissues for further analysis. For future studies, application of decontam with bacterial load would be necessary in order to identify contaminants efficiently, especially for samples with low biomass.

Within the G_{15} population, *Lactobacillus* was the most abundant genus throughout the samples. This is a common microbe found in murine lung microbiota (Barfod et al., 2013; Singh et al., 2017; Yun et al., 2014) and is generally known as a probiotic bacterium. It has been shown to be capable of modulating host immune responses to viral infection in the lungs (Chiba et al., 2013; Villena et al., 2012) and reduce lung injury from *Streptococcus pneumoniae* (Racedo et al., 2006). In a study by Ezendam & Loveren (2008), mice 2-week old or adult mice were administered *Lactobacillus casei* strain Shirota (LcS) and then challenged with ovalbumin in order to induce allergic symptoms in the lungs. This revealed an enhanced immune response from the mice given LcS compared to controls (that were given saline), involving increased eosinophils and lymphocytes. Although it seems that the degree of the effect is dependent on species or strain (Forsythe et al., 2007; Tomosada et al., 2013), these studies demonstrate a general beneficial role of *Lactobacillus* strains in various immune responses.

According to studies in a different field apart from the mammalian lung microbiota, *Pelomonas* contributes to fungal resistance and contraction frequency in hydra (Fraune et al., 2015; Murillo-Rincon et al., 2017). Irrespective of the different host species, it is intriguing to note the mechanism and strategy *Pelomonas* carries. In hydra, small molecules such as amino acids can influence contraction frequency and activity including GABA (gamma-Aminobutyric acid) (Murillo-Rincon et al., 2017). It is intriguing to discover that GABA is one of the amino acids that is also found in pulmonary neuroendorine cells (PNEC), distributed along the alveolar airway epithelium (Yabumoto et al., 2008). Their study shows that GABA can play a role through its receptors in the epithelial cells of murine lungs. Another study shows that GABA is indeed mainly produced in PNEC in primates and can affect mucus production, together with IL-13 (Barrios et al., 2019). Based on these studies, we can speculate a potential role played by *Pelomonas* upon GABA, which could then affect lung functioning.

Streptococcus is another genus with high abundance throughout G_{15} lung samples that was previously found in murine lungs as well as human lungs (Barfod et al., 2013; Beck et al., 2012; C. Cheng et al., 2020; Coburn et al., 2015; Dickson et al., 2013; Scheiermann & Klinman, 2017; Singh et al., 2017). Although it is often detected as one of the dominant genera in the

lungs, *S. pneumoniae* is a well-known cause of lung infection and is a major cause of pneumonia in developing countries (Falade & Ayede, 2011). It is also a leading cause of community-acquired pneumonia, which can be deadly to the elderly (File Jr, 2004), and studies have found an association between *S. pneumoniae* and an increased risk of asthma (Bisgaard et al., 2007; Teo et al., 2015).

Staphylococcus was also found throughout the G_{15} lung samples and was previously found in murine and human lung microbiota (Barfod et al., 2013; Kostric et al., 2017; Remot et al., 2017; Roggenbuck et al., 2016). *Staphylococcus aureus* is another well-known pathogen for respiratory infections and is often called a pathobiont for safely inhabiting the human respiratory tract, but acting pathogenic under certain circumstances (Peres et al., 2015). Again, highlighting the importance of microbial homeostasis, alteration or change in *Staphylococcus* abundance was shown to be associated with asthma in human models (Hilty et al., 2010). In other studies, it was shown to be associated with cystic fibrosis (Dickson et al., 2013; Feigelman et al., 2017; Van Der Gast et al., 2011), along with its association to progression in pulmonary fibrosis depending on its relative abundance (Han et al., 2014). In sum, these taxa play various roles and can modulate the host to enhance or ameliorate inflammation or disease states.

Although no significant QTL were detected, we identified several suggestive QTL and genes within the interval that are related to various lung diseases. These suggestive QTL were detected from *Alpinimonas* on chromosome 11, *Kocuria* on chromosome 5, and *Microbacterium* on chromosome 5 at the genus level, and from OTU92_*Alpinimonas* on chromosome 11 at the OTU level. *Pmp22* was detected adjacent to the peak SNP of QTL of *Alpinimonas*, which was previously found to play a role in lung cancer (Qu et al., 2015). The study found that the suppression of *Pmp22* may affect cell proliferation, invasion, and apoptosis in lung cancer cells, which implies a potential role of the gene in lung cancer therapy. Another gene, *Tekt3*, was also found within the interval of *Alpinimonas* QTL, and is a protein-coding gene belonging to tektin family of proteins. This has been found to be down-regulated in lung squamous cell carcinoma (SCC) metastasis (Tian et al., 2017). It was also found as one of the genes associated to chronic obstructive pulmonary disease (COPD) based on significant and sub-threshold GWAS SNPs (Morrow et al., 2018). These two genes have yet to be studied in depth, but they are interesting findings that have been found to play role in lung diseases and cancer. A similar result was found for OTU92_*Alpinimonas,* which could indicate that this

may be the major species-level OTU of the genus *Alpinimonas* driving the linkage between the bacteria and the host.

From the confidence interval of the *Microbacterium* QTL, there are only three genes detected, which included the *Pdcl2* and *Clock* genes. *Pdcl2* (*phosducin-like 2*) is a member of the phosducin-like protein family, known as a heterotrimeric G protein modulator in humans. Heterotrimeric G protein can mediate its subunit, known as the G protein βγ subunit complex, and different signals such as hormonal and neuronal signaling require a phosducin-like protein for assembly (Lukov et al., 2005). However, its physiological role in lung health or functioning is yet unknown. As one of the core clock genes, *Clock* (*Circadian Locomotor Output Cycles Kaput*) takes part in circadian rhythms, and various diseases result from a disruption of circadian rhythms on different organs such as the skin (Gubareva et al., 2016), heart (Alibhai et al., 2018; Duong et al., 2019; Fletcher et al., 2017), intestines (Le et al., 2017; Voigt et al., 2014), and lung (Gebel et al., 2006; Hadden et al., 2012; Lagishetty et al., 2014; Papagiannakopoulos et al., 2016; Wang et al., 2016), as well as general cancer development (Hashikawa et al., 2017). The relationship between circadian rhythms and host fitness is a widely studied subject, and it is interesting to examine the wide range of tissues and mechanisms that are affected. *Clock* pairs with another gene, *Bmal1*, whereby the two genes regulate the transcription of downstream genes such as *Per2* and *Cry* genes. Not only the physiological disturbance in the rhythm, but also the genetic mutation in clock genes altered host survival and increased the risk of tumorigenesis (Papagiannakopoulos et al., 2016). In the intestines, it has been shown that the microbiota itself maintains circadian rhythmicity and affects host circadian clock function with its metabolites such as short-chain fatty acids (SCFA) (Tahara et al., 2018). Deletion of clock genes can alter the composition and load of the intestinal and fecal microbiota (Liang et al., 2015), which may be applicable to lung microbiota in terms of host health and fitness. We could speculate a potential association between the circadian rhythm of the host and *Microbacterium* in the lungs, in which its metabolites affecting the host's clock-related genes or the disturbance in circadian rhythmicity altering the abundance of *Microbacterium*.

Although lungs had been long known as a sterile tissue, identification of its microbiota has led to broad range of studies including host health, disease, and fitness. The association between the host and lung microbiota is highlighted in this study by performing QTL linkage mapping on a mouse population of great genetic marker density. Narrow confidence intervals included various genes related to cancer and immune responses. We also set out a procedure

for decontaminating the dataset for future studies on samples of low biomass, which helps reduce environmental noise and improve data quality.

Conclusion

Although lungs were considered sterile for a long time, the presence and the importance of microbes inhabiting the lungs are widely recognized as they take part in diseases susceptibility. The overall microbial community of lungs in a mouse mapping population was examined through 16S rRNA gene sequencing, where we identified *Lactobacillus* as the most abundant genus. We further tried to identify associations between host genetics and microbes using QTL linkage mapping on relative abundances data (of the core microbiota). With high genetic marker density, we obtained narrow intervals, but identified only suggestive QTL after permutation-based determination of significance thresholds. Lung disease-related genes were detected in association with microbes such as *Alpinimonas* and *Microbacterium*. Future studies could look into the expression of the genes in association with the associated microbes to further investigate how they affect each other. Another future plan is to examine the gut-lung axis using samples from the same mapping population, in order to identify a 'cross-talk' between microbial communities of the two body sites.

Methods

RNA extraction & library preparation

The G_{15} AIL mouse population was bred by intercrossing 4 strains, MRL/MpJ, NZM2410/J, BXD2/TyJ, and CAST/EiJ, with equal sex and strain distributions for 15 generations (Belheouane et al., 2017; Srinivas et al., 2013). These mice were dissected previously and the entire lung tissues were preserved in RNALater (Invitrogen) at -20°C (by Dr. Marie Vallier and Dr. Meriem Belheouane). Approximately the bottom 1/3 of the left lobe was extracted. Tools were sterilized using 70% ethanol, RNase-Away (Thermo Scientific), and hot sterilizing beads (FST) with temperature at or above 120°C. Tissues were lysed using lysing matrix E (MP Biomedicals) and a Precellys 24 homogenizer (bertin technologies). DNA and RNA were extracted using the AllPrep 96 DNA/RNA kit (QIAGEN) with DNase I treatment (QIAGEN) directly on the columns according to manufacturer's protocol during RNA extraction. At the end, 20µL of RNase-free water was added twice to obtain a total amount of 40µL of clean RNA. For DNA, 30µL of EB buffer was added twice. Concentration was measured on a NanoDrop 1000 (Thermo Scientific) and RNA samples were diluted to 200 ng/µL. Then reverse transcription was performed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) by adding 10 µL RNA. The master mix consisted of 2µL buffer, 0.8µL dNTPs, 2 µL Random Primers, and 4.2µL water for each sample. PCR protocol follows: 25°C for 10 minutes, 37°C for 120 minutes, 85°C for 5 minutes, and incubation at 12°C.

For sequencing, the V1-V2 region of the 16S rRNA gene was amplified using primers 27F and 338R, with unique eight-based MIDs (multiplex identifier), creating 288 different combinations of PCR products. PCR amplifications were conducted in a 25µl template with the following recipe: 5µl of 5x HF buffer (ThermoFisher Scientific), 5µl of 10mM dNTPs, 2µl each of 4µM forward and reverse primers, 0.25µl of Taq polymerase (Phusion High-Fidelity DNA Polymerase, ThermoFisher Scientific), 3µl of cDNA template, and 8.25µl of nuclease-free water. Cycling conditions included: initial denaturation at 98°C for 30 seconds, 30 cycles of 98°C for 9 seconds, 55°C for 30 seconds, and 72°C for 90 seconds, and a final extension at 72°C for 10 minutes. In addition to these samples, 3 negative extraction controls (NEC), one from each extraction plate, and one mock community sample (20ng/µL) (Microbial Community DNA Standard, ZymoBIOMICS) were included. PCR products were run on gels, imaged using a GelDoc XR+ (BioRad) and were quantified with Image Lab (v.6.0.1) (BioRad). Based on the

quantification, these PCR products were mixed into subpools at equal amount of nucleic acid and purified using the MinElute kit (QIAGEN) following the manufacturer's instructions. Purified subpools were quantified on a NanoDrop 1000 (Thermo Scientific), mixed for a final pool with equal concentration of each subpool, and the final pool was sequenced using MiSeq Reagent Kit v2 with 2 x 250bp read length on an Illumina MiSeq.

Sequence processing

Sequences in fastq format were checked for quality using FastQC (v.0.11.7) and trimmed using Trimmomatic (v.0.36). Reads were then merged with VSEARCH (v.2.7.1), chimeras were identified and removed with VSEARCH using the SILVA Gold reference database (Quast et al., 2012). OTU clustering was performed at a 97% similarity threshold with unique sequences detected through dereplication. Afterwards, taxonomy was assigned based on RDP Multi-Classifier (v.16) from the phylum to genus level at a confidence level of 0.8. Sequences for both lung and skin were analyzed together in order to help understand lung community composition, as both carry low bacterial load and are thus prone to contamination. Thirty-two lung samples were removed that were not included in the skin dataset.

Generated OTU and taxonomy tables were transferred to R studio (v.1.2.1335) and were merged with meta data including sample ID, parents, sex, age, and cage information, using "phyloseq" (v.1.28.0) (McMurdie & Holmes, 2013) in order to create a phyloseq object. Dataset was then rarefied to even sampling depth of 4,000 (reads), which removed 1 lung sample that had only 44 reads, leaving us with 252 samples. As two genera commonly considered to be lab contaminants were detected (*Halomonas* and *Shewanella*), we measured total bacterial load using ddPCR to help assess contamination. Details of the ddPCR method can be found in **Chapter 2**. Using this total bacterial load, the frequency method of the R package "decontam" (v.1.4.0) (Davis et al., 2018) was applied with a threshold of 0.1. Spearman correlation tests were also performed between the total bacterial load and relative abundances of *Halomonas* and *Shewanella*. As a result, the two genera were identified as contaminants and thus removed from the dataset before further analysis. We then defined a core measurable microbiota (CMM) with bacterial taxa and OTUs that are present with at least 1% relative abundance and present in at least 10% of the samples.

SourceTracker (v.0.9.5) (Knights et al., 2011) was also adapted in order to determine the fraction of contamination in each sample based on NECs as a "source" (of contamination).

This was run on species-level OTUs of the CMM and NECs. Default parameters were used at a 97% significance threshold. Nine samples with a 20% or more likelihood to be derived from the NEC "source" were removed. The final CMM thus included 243 samples with 36 OTUs and 60 taxa from the genus to phylum level. Summary statistics were performed on the defined final CMM.

For beta diversity measures between lung and skin samples, Constrained Analysis of Principal Coordinates (CAP) (Linnenbrink et al., 2013) was performed using Bray-Curtis dissimilarity. The same analysis was applied for community comparisons among lung samples, NECs and the mock community. Next, the 15 most abundant genera were compared between lung and skin samples as well as between lung samples and NECs. All statistical analyses were performed on R (v.3.6.1).

QTL mapping

Prior to QTL mapping, effects of sex, age, body weight, and cage information on the variation in CMM traits were calculated. Sex, age, and cage showed significance ($p < 0.05$) and were thus included in the mapping. A linear mixed effects analysis was then performed on each CMM trait using "lme4" (v.1.1-10) (Bates et al., 2012). Sex and age were included as fixed effects, and cage information was included as random effects. Variance was estimated using "r.squaredGLMM" in "MuMIn" (v.1.43.6) and "VarCorr" in "lme4" for fixed and random effects, respectively.

Linkage mapping was performed using "DOQTL" (1.6.0) (Gatti et al., 2014) and "QTLRel" (0.2.14) (R. Cheng et al., 2011), whereby we fit an additive model that regresses the log10-transformed relative abundance of CMM traits at taxon and OTU levels on the four founder haplotype contributions. A kinship matrix was defined using "kinship.probs" in "DOQTL" in order to adjust for the kinship between samples (Belheouane et al., 2017). A 3D-array of founder haplotype contributions (# samples x 4 founders x # markers) (Belheouane et al., 2017), kinship matrix, and meta data with sex, age, and cage information were included for linkage mapping.

Permutations were run for each trait by shuffling the genotypic data to define significance thresholds at both the 90th and 95th percentiles of LOD scores. Permutations were run 10,000 times, as it is recommended to run at least 1,000 times (Gatti et al., 2014), and QTL peaks above either threshold for each trait were considered suggestive and significant,

respectively, for mapping associations. QTL confidence intervals were set at a 1.5 LOD score drop on each side of the QTL peak. After QTL mapping, genes within the confidence intervals were closely examined and plotted using “get.mgi.features” and “gene.plot” in “DOQTL” in order to identify potential candidate genes. All QTL mapping analyses were performed in R 3.2.2, except permutations on R 3.6.2 (this was for permutation speed and for running in parallel on our cluster which carries R 3.6.2, and we confirmed the same thresholds are produced from both R versions).

Supplements

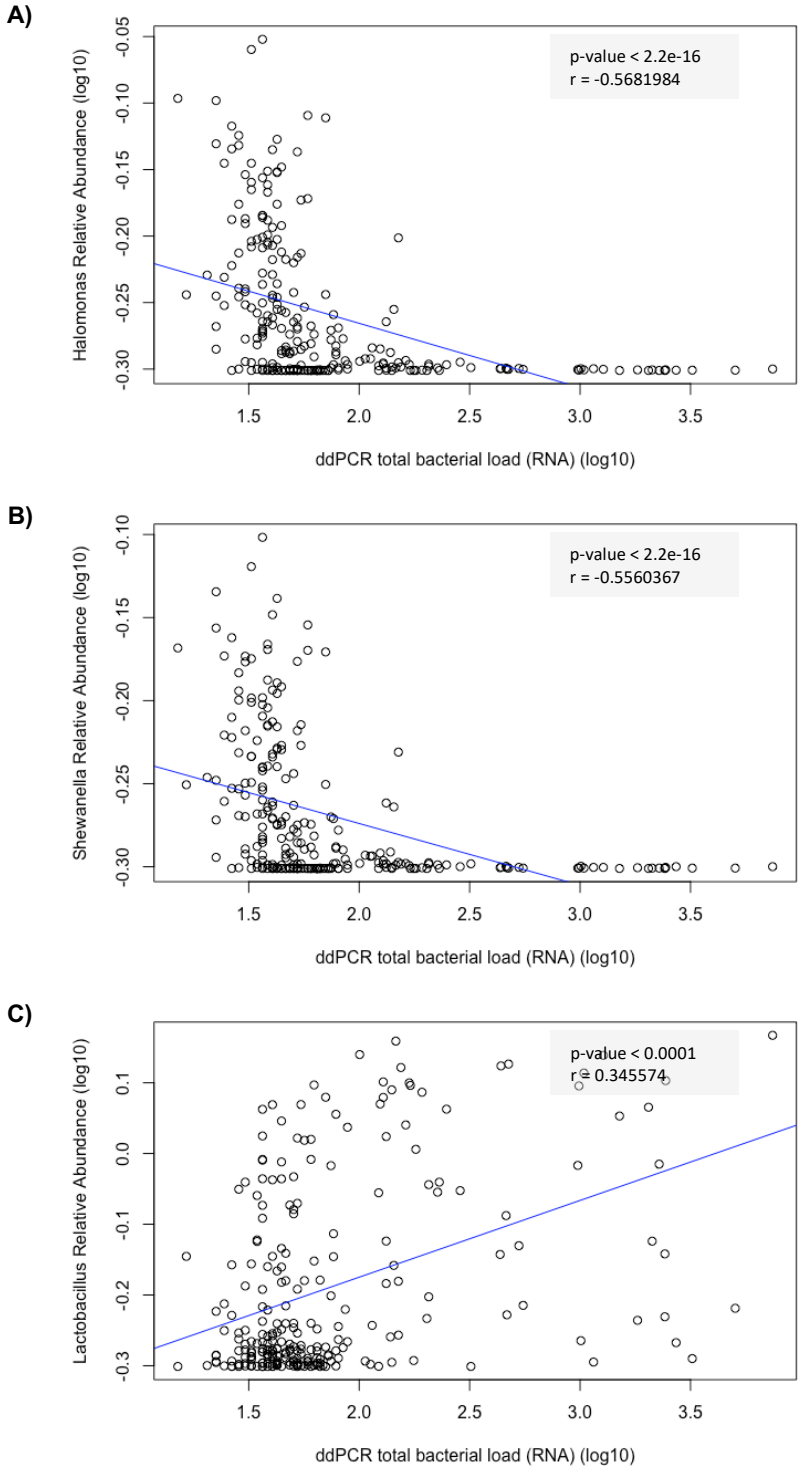

Figure S1. Correlation test with total bacterial load. Correlation between $\log_{10}$-transformed total bacterial load and $\log_{10}$-transformed relative abundance of each of three genera, A) *Halomonas*, B) *Shewanella*, and C) *Lactobacillus* based on 16S rRNA gene sequencing was tested. P-value and Spearman correlation coefficient for each test are shown within each graph (n=252).

Table S1. Summary statistics of CMM traits from genus to phylum levels.

Rank	Trait	min	max	mean	std.dev	coef.var	var
Phylum	Actinobacteria	0.275	42.550	13.243	7.426	0.561	55.153
Class	Actinobacteria	0.275	42.550	13.242	7.427	0.561	55.153
Order	Actinomycetales	0.275	42.425	12.735	7.408	0.582	54.875
Genus	*Corynebacterium*	0	19.363	0.952	1.759	1.848	3.094
Family	*Microbacteriaceae*	0	19.975	2.587	2.804	1.084	7.860
Genus	*Alpinimonas*	0	8.388	0.435	0.992	2.280	0.983
Genus	*Microbacterium*	0	5.488	0.428	0.658	1.538	0.433
Genus	*Rhodoluna*	0	18.613	1.540	2.288	1.486	5.236
Family	*Micrococcaceae*	0	37.125	2.309	4.381	1.897	19.190
Genus	*Kocuria*	0	16.988	0.539	1.451	2.693	2.104
Genus	*Micrococcus*	0	36.838	0.994	3.413	3.433	11.646
Genus	*Rothia*	0	23.488	0.615	2.341	3.809	5.478
Family	*Mycobacteriaceae*	0	30.850	1.279	3.288	2.571	10.812
Genus	*Mycobacterium*	0	30.863	1.292	3.288	2.546	10.812
Genus	Uncl. Actinomycetales	0	36.938	3.513	4.005	1.140	16.041
Phylum	Bacteroidetes	1.275	61.050	17.424	10.560	0.606	111.521
Genus	*Bacteroides*	0	56.563	0.894	4.019	4.494	16.156
Class	Flavobacteriia	0	51.200	8.974	9.938	1.107	98.770
Genus	Uncl. Bacteroidales	0	9.738	1.667	1.835	1.101	3.368
Order	Flavobacteriales	0	51.200	8.974	9.938	1.107	98.770
Family	*Flavobacteriaceae*	0	51.200	7.536	9.577	1.271	91.710
Genus	*Flavobacterium*	0	51.213	7.257	9.614	1.325	92.436
Genus	Uncl. Flavobacteriales	0	42.738	1.328	3.671	2.765	13.474
Genus	Uncl. Bacteroidetes	0	39.963	4.376	4.119	0.941	16.964
Phylum	Firmicutes	4	95.650	40.626	13.179	0.324	173.673
Class	Bacilli	2.475	94.325	20.216	13.352	0.660	178.276
Genus	*Bacillus*	0	8.063	0.584	0.838	1.436	0.702
Genus	*Staphylococcus*	0	4.788	0.472	0.602	1.277	0.363
Order	Lactobacillales	1.25	92.650	18.135	13.529	0.746	183.026
Family	*Lactobacillaceae*	0.225	92.275	13.199	13.742	1.041	188.834
Genus	*Lactobacillus*	0	92.288	13.210	13.742	1.040	188.852
Class	Clostridia	0.25	64.400	19.670	11.233	0.571	126.179
Order	Clostridiales	0.25	64.400	19.660	11.233	0.571	126.185
Genus	*Clostridium sensu stricto*	0	16.363	0.758	1.512	1.996	2.287
Family	*Lachnospiraceae*	0.025	49.450	8.473	6.402	0.756	40.989
Genus	Uncl. *Lachnospiraceae*	0	49.388	7.968	6.271	0.787	39.320
Genus	Uncl. *Ruminococcaceae*	0	54.513	2.565	4.487	1.749	20.135
Family	*Streptococcaceae*	0	37.475	4.597	4.319	0.939	18.650

Genus	*Lactococcus*	0	10.538	1.901	2.090	1.100	4.368
Genus	*Streptococcus*	0	37.088	2.694	4.413	1.638	19.472
Genus	Uncl. Clostridiales	0	34.363	6.941	5.572	0.803	31.046
Phylum	Fusobacteria	0	6.500	0.981	1.363	1.389	1.858
Class	Fusobacteriia	0	6.500	0.981	1.363	1.389	1.858
Order	Fusobacteriales	0	6.500	0.981	1.363	1.389	1.858
Family	*Leptotrichiaceae*	0	6.500	0.973	1.365	1.403	1.864
Genus	*Leptotrichia*	0	6.513	0.984	1.366	1.388	1.865
Phylum	Proteobacteria	0.775	91.225	22.696	14.362	0.633	206.263
Class	Betaproteobacteria	0.2	87.525	12.142	14.933	1.230	222.988
Order	Burkholderiales	0.2	87.525	11.006	14.692	1.335	215.864
Family	*Burkholderiales incertae sedis*	0	40.500	1.150	3.018	2.625	9.107
Genus	*Tepidimonas*	0	40.513	0.979	3.009	3.074	9.056
Genus	Uncl. Burkholderiales	0	14.663	1.154	1.963	1.702	3.855
Family	*Comamonadaceae*	0	87.400	8.227	15.071	1.832	227.122
Genus	*Pelomonas*	0	87.388	7.177	15.258	2.126	232.807
Order	Rhodocyclales	0	39.750	0.787	3.127	3.971	9.777
Family	*Rhodocyclaceae*	0	39.750	0.787	3.127	3.971	9.777
Genus	*Dechloromonas*	0	39.763	0.780	3.128	4.008	9.786
Class	Deltaproteobacteria	0	27.000	1.049	2.163	2.062	4.680
Order	Desulfovibrionales	0	4.725	0.387	0.763	1.975	0.583
Family	*Desulfovibrionaceae*	0	4.725	0.387	0.763	1.975	0.583
Genus	Uncl. *Desulfovibrionaceae*	0	4.513	0.335	0.721	2.154	0.520
Class	Gammaproteobacteria	0.15	38.550	6.676	5.385	0.807	28.994
Order	Enterobacteriales	0	15.725	1.339	2.234	1.668	4.992
Family	*Enterobacteriaceae*	0	15.725	1.339	2.234	1.668	4.992
Genus	*Serratia*	0	14.363	0.980	1.904	1.942	3.624
Order	Pseudomonadales	0.025	37.900	4.378	4.927	1.126	24.277
Family	*Moraxellaceae*	0	35.275	1.577	3.697	2.344	13.667
Genus	Uncl. *Moraxellaceae*	0	35.288	1.348	3.713	2.754	13.783
Family	*Pseudomonadaceae*	0	20.625	2.800	3.797	1.356	14.416
Genus	*Pseudomonas*	0	20.638	2.795	3.786	1.355	14.334
Genus	Uncl. Bacteria	0	12.213	2.626	1.934	0.736	3.739
Genus	Uncl. Proteobacteria	0	9.413	0.743	0.954	1.283	0.910
Genus	Chao1	14	33.000	27.551	3.380	0.123	11.422
Genus	Shannon	0.365	3.036	2.369	0.498	0.210	0.248

(min = minimum, max = maximum, std.dev = standard deviation, coef.var = coefficient variation, var = variation)

Table S2. Summary statistics of CMM traits at OTU level.

	Trait	min	max	mean	std.dev	coef.var	var
Pelomonas	Otu10	0	18.325	1.383	2.243	1.622	5.030
Flavobacterium	Otu14	0	14.175	2.043	2.988	1.463	8.927
Lactobacillus	Otu15	0	8.375	1.517	1.765	1.163	3.114
Lactobacillus	Otu17	0	30.125	1.586	3.405	2.147	11.596
Pseudomonas	Otu19	0	40.500	0.962	3.004	3.123	9.025
Uncl. Actinomycetales	Otu2	0	24.600	3.651	5.406	1.480	29.220
Lactococcus	Otu21	0	5.300	1.090	1.223	1.122	1.497
Lactobacillus	Otu22	0	36.825	0.921	3.409	3.703	11.624
Flavobacterium	Otu23	0	26.675	0.610	2.225	3.649	4.950
Rhodoluna	Otu24	0	39.750	0.768	3.128	4.074	9.786
Uncl. *Moraxellaceae*	Otu25	0	24.400	1.203	2.951	2.452	8.706
Mycobacterium	Otu26	0	35.275	1.319	3.716	2.817	13.809
Uncl. Flavobacteriales	Otu27	0	37.025	0.646	3.027	4.689	9.165
Streptococcus	Otu2761	0	30.850	1.254	3.279	2.615	10.750
Uncl. Bacteroidetes	Otu28	0	17.550	0.854	1.962	2.298	3.849
Serratia	Otu29	0	13.200	0.663	1.566	2.363	2.453
Tepidimonas	Otu3	0	71.275	5.202	10.039	1.930	100.778
Leptotrichia	Otu30	0	14.350	0.968	1.904	1.967	3.624
Micrococcus	Otu31	0	42.725	1.248	3.666	2.938	13.441
Uncl. Burkholderiales	Otu33	0	8.900	0.403	1.065	2.644	1.135
Uncl. Bacteroidetes	Otu36	0	4.500	0.322	0.721	2.238	0.520
Dechloromonas	Otu38	0	26.875	0.761	2.446	3.216	5.984
Uncl. *Lachnospiraceae*	Otu4	0	87.375	7.164	15.258	2.130	232.807
Uncl. *Lachnospiraceae*	Otu40	0	6.875	0.350	0.787	2.247	0.619
Streptococcus	Otu42	0	14.650	0.892	1.815	2.035	3.294
Uncl. Clostridiales	Otu43	0	4.925	0.419	0.929	2.216	0.863
Uncl. Actinomycetales	Otu5499	0	6.500	0.940	1.372	1.460	1.884
Streptococcus	Otu6	0	20.200	2.555	3.747	1.467	14.037
Uncl. Clostridiales	Otu62	0	8.375	0.495	1.011	2.041	1.022
Uncl. Clostridiales	Otu66	0	2.575	0.370	0.469	1.268	0.220
Alpinimonas	Otu7323	0	35.525	0.536	2.612	4.873	6.822
Uncl. Clostridiales	Otu7839	0	4.500	0.493	0.862	1.749	0.743
Uncl. Bacteroidetes	Otu8	0	51.200	5.341	10.098	1.890	101.960
Uncl. *Lachnospiraceae*	Otu8132	0	4.150	0.447	0.756	1.691	0.571
Uncl. Bacteroidetes	Otu9	0	10.525	1.870	2.087	1.116	4.356
Uncl. *Desulfovibrionaceae*	Otu92	0	8.375	0.422	0.992	2.347	0.983
	Chao1	68	625.137	263.711	109.141	0.414	11911.798
	Shannon	0.706	5.340	3.689	0.949	0.257	0.900

(min = minimum, max = maximum, std.dev = standard deviation, coef.var = coefficient variation, var = variation)

Table S3. Portions of variance explained by explanatory variables for genus level.

		Variance component		% Total Variance		Fixed Structure
Rank	**Trait**	**Cage**	**Residual**	**Cage**	**Residual**	**R^2m**
Phylum	Actinobacteria	0.0000	0.0023	0.11	99.89	0.0764
Class	Actinobacteria	0.0000	0.0023	0.12	99.88	0.0764
Order	Actinomycetales	0.0000	0.0023	0.53	99.47	0.0728
Genus	*Corynebacterium*	0.0000	0.0002	1.33	98.67	0.0067
Family	*Microbacteriaceae*	0.0000	0.0004	0.00	100.00	0.1576
Genus	*Alpinimonas*	0.0000	0.0001	0.00	100.00	0.1421
Genus	*Microbacterium*	0.0000	0.0000	0.00	100.00	0.0019
Genus	*Rhodoluna*	0.0000	0.0003	0.00	100.00	0.1079
Family	*Micrococcaceae*	0.0000	0.0010	2.69	97.31	0.0016
Genus	*Kocuria*	0.0000	0.0001	0.35	99.65	0.0178
Genus	*Micrococcus*	0.0000	0.0006	2.95	97.05	0.0009
Genus	*Rothia*	0.0000	0.0003	0.00	100.00	0.0175
Family	*Mycobacteriaceae*	0.0000	0.0005	3.46	96.54	0.0941
Genus	*Mycobacterium*	0.0000	0.0005	3.46	96.54	0.0941
Genus	Uncl. Actinomycetales	0.0001	0.0008	6.49	93.51	0.0198
Phylum	Bacteroidetes	0.0003	0.0036	7.38	92.62	0.0401
Genus	*Bacteroides*	0.0000	0.0006	0.00	100.00	0.0155
Class	Flavobacteriia	0.0004	0.0032	11.90	88.10	0.1070
Genus	Uncl. Bacteroidales	0.0000	0.0002	0.00	100.00	0.0051
Order	Flavobacteriales	0.0004	0.0032	11.90	88.10	0.1070
Family	*Flavobacteriaceae*	0.0005	0.0028	14.59	85.41	0.1387
Genus	*Flavobacterium*	0.0005	0.0028	15.68	84.32	0.1368
Genus	Uncl. Flavobacteriales	0.0000	0.0007	0.00	100.00	0.0113
Genus	Uncl. Bacteroidetes	0.0000	0.0009	0.00	100.00	0.0434
Phylum	Firmicutes	0.0000	0.0041	0.00	100.00	0.0117
Class	Bacilli	0.0000	0.0051	0.67	99.33	0.0482
Genus	*Bacillus*	0.0000	0.0000	0.40	99.60	0.0288
Genus	*Staphylococcus*	0.0000	0.0000	3.54	96.46	0.0055
Order	Lactobacillales	0.0002	0.0052	3.15	96.85	0.0696
Family	*Lactobacillaceae*	0.0003	0.0056	5.72	94.28	0.0519
Genus	*Lactobacillus*	0.0003	0.0056	5.73	94.27	0.0519
Class	Clostridia	0.0000	0.0042	0.00	100.00	0.0929
Order	Clostridiales	0.0000	0.0042	0.00	100.00	0.0929
Genus	*Clostridium sensu stricto*	0.0000	0.0001	1.09	98.91	0.0033
Family	*Lachnospiraceae*	0.0000	0.0018	1.85	98.15	0.0208
Genus	Uncl. *Lachnospiraceae*	0.0000	0.0017	2.07	97.93	0.0219

Genus	Uncl. *Ruminococcaceae*	0.0000	0.0009	0.00	100.00	0.0341
Family	*Streptococcaceae*	0.0000	0.0009	0.00	100.00	0.0424
Genus	*Lactococcus*	0.0000	0.0003	0.88	99.12	0.0632
Genus	*Streptococcus*	0.0000	0.0010	0.00	100.00	0.0529
Genus	Uncl. Clostridiales	0.0000	0.0014	0.00	100.00	0.0931
Phylum	Fusobacteria	0.0000	0.0001	6.38	93.62	0.0178
Class	Fusobacteriia	0.0000	0.0001	6.38	93.62	0.0178
Order	Fusobacteriales	0.0000	0.0001	6.38	93.62	0.0178
Family	*Leptotrichiaceae*	0.0000	0.0001	6.56	93.44	0.0185
Genus	*Leptotrichia*	0.0000	0.0001	6.54	93.46	0.0186
Phylum	Proteobacteria	0.0000	0.0057	0.00	100.00	0.0033
Class	Betaproteobacteria	0.0000	0.0072	0.00	100.00	0.0009
Order	Burkholderiales	0.0000	0.0070	0.00	100.00	0.0008
Family	*Burkholderiales incertae sedis*	0.0000	0.0004	1.04	98.96	0.0583
Genus	*Tepidimonas*	0.0000	0.0004	0.12	99.88	0.0600
Genus	Uncl. Burkholderiales	0.0000	0.0002	5.30	94.70	0.1630
Family	*Comamonadaceae*	0.0000	0.0076	0.00	100.00	0.0191
Genus	*Pelomonas*	0.0000	0.0079	0.00	100.00	0.0213
Order	Rhodocyclales	0.0000	0.0005	0.00	100.00	0.0046
Family	*Rhodocyclaceae*	0.0000	0.0005	0.00	100.00	0.0046
Genus	*Dechloromonas*	0.0000	0.0005	0.13	99.87	0.0053
Class	Deltaproteobacteria	0.0000	0.0003	2.13	97.87	0.0062
Order	Desulfovibrionales	0.0000	0.0000	4.65	95.35	0.0153
Family	*Desulfovibrionaceae*	0.0000	0.0000	4.65	95.35	0.0153
Genus	Uncl. *Desulfovibrionaceae*	0.0000	0.0000	4.86	95.14	0.0115
Class	Gammaproteobacteria	0.0000	0.0015	0.85	99.15	0.0236
Order	Enterobacteriales	0.0000	0.0003	0.00	100.00	0.0061
Family	*Enterobacteriaceae*	0.0000	0.0003	0.00	100.00	0.0061
Genus	*Serratia*	0.0000	0.0002	0.00	100.00	0.0342
Order	Pseudomonadales	0.0000	0.0012	1.30	98.70	0.0557
Family	*Moraxellaceae*	0.0000	0.0007	0.34	99.66	0.0564
Genus	Uncl. *Moraxellaceae*	0.0000	0.0007	0.48	99.52	0.0525
Family	*Pseudomonadaceae*	0.0001	0.0008	8.23	91.77	0.0284
Genus	*Pseudomonas*	0.0001	0.0008	8.46	91.54	0.0285
Genus	Uncl. Bacteria	0.0000	0.0002	2.57	97.43	0.0452
Genus	Uncl. Proteobacteria	0.0000	0.0001	2.94	97.06	0.0522
Genus	Chao1	0.0004	0.0355	0.99	99.01	0.0422
Genus	Shannon	0.0006	0.0131	4.48	95.52	0.0338

Table S4. Portions of variance explained by explanatory variables for OTU level.

		Variance component		% Total Variance		Fixed Structure
	OTU	**Cage**	**Residual**	**Cage**	**Residual**	**R^2m**
Pelomonas	OTU10	0.0000	0.0003	0.00	100.00	0.1007
Flavobacterium	OTU14	0.0001	0.0005	8.94	91.06	0.0302
Lactobacillus	OTU15	0.0000	0.0002	3.28	96.72	0.0306
Lactobacillus	OTU17	0.0000	0.0006	1.27	98.73	0.0653
Pseudomonas	OTU19	0.0000	0.0004	0.10	99.90	0.0593
Uncl. Actinomycetales	OTU2	0.0001	0.0015	8.88	91.12	0.0403
Lactococcus	OTU21	0.0000	0.0001	0.00	100.00	0.0459
Lactobacillus	OTU22	0.0000	0.0006	1.40	98.60	0.0008
Flavobacterium	OTU23	0.0000	0.0003	1.00	99.00	0.0118
Rhodoluna	OTU24	0.0000	0.0005	0.13	99.87	0.0053
Uncl. *Moraxellaceae*	OTU25	0.0000	0.0005	5.52	94.48	0.0350
Mycobacterium	OTU26	0.0000	0.0020	0.36	99.64	0.0610
Uncl. Flavobacteriales	OTU27	0.0000	0.0011	0.00	100.00	0.0594
Streptococcus	OTU2761	0.0000	0.0015	3.09	96.91	0.1088
Uncl. Bacteroidetes	OTU28	0.0000	0.0007	0.00	100.00	0.1574
Serratia	OTU29	0.0000	0.0005	4.55	95.45	0.0169
Tepidimonas	OTU3	0.0000	0.0078	0.00	100.00	0.1322
Leptotrichia	OTU30	0.0000	0.0008	0.00	100.00	0.0366
Micrococcus	OTU31	0.0000	0.0007	0.00	100.00	0.0081
Uncl. Burkholderiales	OTU33	0.0000	0.0001	2.28	97.72	0.0120
Uncl. Bacteroidetes	OTU36	0.0000	0.0000	4.86	95.14	0.0115
Dechloromonas	OTU38	0.0000	0.0003	0.00	100.00	0.0418
Uncl. *Lachnospiraceae*	OTU4	0.0000	0.0079	0.00	100.00	0.0213
Uncl. *Lachnospiraceae*	OTU40	0.0000	0.0000	0.42	99.58	0.0257
Streptococcus	OTU42	0.0000	0.0002	10.84	89.16	0.2075
Uncl. Clostridiales	OTU43	0.0000	0.0001	0.00	100.00	0.0395
Uncl. Actinomycetales	OTU5499	0.0000	0.0001	6.69	93.31	0.0205
Streptococcus	OTU6	0.0001	0.0008	8.51	91.49	0.0357
Uncl. Clostridiales	OTU62	0.0000	0.0001	7.55	92.45	0.0075
Uncl. Clostridiales	OTU66	0.0000	0.0000	0.00	100.00	0.0137
Alpinimonas	OTU7323	0.0000	0.0003	0.00	100.00	0.0075
Uncl. Clostridiales	OTU7839	0.0000	0.0000	3.69	96.31	0.0985
Uncl. Bacteroidetes	OTU8	0.0005	0.0032	14.28	85.72	0.1608
Uncl. *Lachnospiraceae*	OTU8132	0.0000	0.0000	8.40	91.60	0.0170
Uncl. Bacteroidetes	OTU9	0.0000	0.0002	0.99	99.01	0.0633
Uncl. *Desulfovibrionaceae*	OTU92	0.0000	0.0001	0.00	100.00	0.1421

	Chao1	0.0004	0.0355	0.99	99.01	0.0422
	Shannon	0.0006	0.0131	4.48	95.52	0.0338

Chapter II.

Application of droplet digital PCR in the analysis of lung bacterial abundances

Introduction

Droplet digital PCR (ddPCR) is a highly sensitive quantitative PCR method to detect and quantify target nucleic acid sequences. A single PCR mix of 20µL is partitioned into approximately 20,000 droplets using an oil emulsion technique, whereby individual PCR reactions are run within each droplet. The target DNA is detected by fluorescence and is read as either "positive" or "negative" droplet, depending on the presence of the target DNA (Fig. S1). These signals are considered as random events, which are fitted to a Poisson distribution for further quantification (Hindson et al., 2011). Accordingly, unlike quantitative PCR (qPCR), ddPCR does not require a standard curve. ddPCR holds an advantage over qPCR with its higher precision, especially in samples with a low biomass or low abundance of a target gene, thus reducing PCR bias against contaminants or environmental noises (Maheshwari et al., 2017; Sze et al., 2014; Taylor et al., 2017). ddPCR is not only efficient for assessing the total bacterial load (Raveh-Sadka et al., 2015; Sze et al., 2014), but also for assessing the number of target gene copies (Maheshwari et al., 2017) or bacteria with target-specific primers (Baltrušis et al., 2019; Gobert et al., 2018; Srisutham et al., 2017).

In this chapter, genus-specific primers were used to estimate the load of specific bacterial genera of interest, namely *Lactobacillus* and *Pelomonas*, in addition to quantifying the total bacterial load. By targeting these with genus-specific primers, the chances of amplifying non-specific bacteria or environmental noise is reduced.

Lactobacillus was reported as the most abundant bacterial genus throughout the lung tissue samples of the AIL G_{15} mice and is a common microbe in murine lungs (Barfod et al., 2013; Kostric et al., 2017; Remot et al., 2017; Singh et al., 2017; Yun et al., 2014). In previous studies, *Lactobacillus* was shown to be associated with allergic pulmonary inflammation (Tomosada et al., 2013; Villena et al., 2012). Strains of *Lactobacillus rhamnosus* CRL1505 were given nasally to poly(I:C) (an immunostimulant to stimulate viral infections)- and RSV (respiratory syncytial virus)-challenged mice, which increased the level of various cytokines such as IFN-α, IFN-β, IL-6, IL-10, TNF-α, which all play a crucial role in viral infection and inflammation (Tomosada et al., 2013). Those mice treated with lactobacilli carried less viral load and maintained higher weight compared to controls (Tomosada et al., 2013). In another study, oral administration of *Lactobacillus casei* ameliorated an injury caused by *Streptococcus pneumoniae* lung infection, whereby the administration reduced symptoms of infection such as reduced alveolar airspaces, hemorrhage, and increased fibrosis in bronchial walls (Racedo et al., 2006). Oral administration of *L. casei*, in a

form of yogurt, was also shown to help reduce the number of CFU of *Pseudomonas aeruginosa* in the lungs of swiss albino mice (Alvarez et al., 2001). *Lactobacillus* is also known to play beneficial roles within other organs. For example, consumption of food such as yogurt and kimchi, of which the major bacterial component is *Lactobacillus*, helps with digestion and gut health (Kil et al., 2004; Oozeer et al., 2006; Pei et al., 2017; Perdigón et al., 1994). Yogurt with a mix of *Lactobacillus* and *Streptococcus* strains given to BALB/c mice reduced symptoms of inflammatory bowel disease (IBD), with decreased infiltrative immune cells and increased production of IgA cells compared to control mice (Gobbato et al., 2008). The administration *L. casei* was also found to provoke immune responses and delay tumor growth in mice with breast tumors (Aragón et al., 2014). Significant differences in cytokine levels, specifically a reduced level of IL-6, were observed in BALB/c mice given fermented milk, which can play a role in tumor growth and metastasis (Aragón et al., 2014). Overall, *Lactobacillus* strains have been broadly presented as beneficial bacteria to the host in the respiratory tract, intestinal tract, as well as in tumorigenesis in different organs.

Pelomonas was another bacterial genus detected within the core microbiota of the AIL G_{15} mouse lung that caught our attention as a genus previously found uniquely in the lungs (Barfod et al., 2013). It was found to have association with lungs, but its function or role is yet unknown (Borewicz et al., 2013). *Pelomonas* is a genus of Gram-negative bacteria that consists of 3 species, *P. aquatica*, *puraquae*, and *saccharophila*. These are often found in aquatic environments (Gomila et al., 2007) and hence also found on hydra (Murillo-Rincon et al., 2017). With its mechanism in hydra, we hypothesized that *Pelomonas* may be associated with lung functioning and lung disease susceptibility.

In the previous chapter, the total bacterial load estimated using ddPCR was first used to identify putative contaminants such as *Halomonas* and *Shewanella*. In order to further investigate the link between the host and the lung microbial load, in this chapter QTL linkage mapping was performed on the load of total bacteria, *Lactobacillus*, and *Pelomonas*. We herewith demonstrate an improved detection limit in QTL mapping with ddPCR-derived bacterial load information and also a detection of significant QTLs, which included candidate genes that are associated with lung disease susceptibility.

As a result of mapping, we detected two candidate genes, interleukin-10 (*Il-10*) and mitogen-activated protein kinase-activated protein kinase 2 (*Mapkapk2* or *Mk2*), associated with *Lactobacillus* load at the DNA level. We measured the expression level of these candidate

genes using ddPCR in order to examine the relationship between *Lactobacillus* and these candidates at the transcript level.

Aims

- Establish a protocol for ddPCR on lung tissue and estimate the load of total bacteria, *Lactobacillus*, and *Pelomonas*.
- Identify host-microbiota associations using QTL mapping on estimated bacterial loads.
- Identify candidate genes within the confidence intervals of each trait that are potentially associated with lung functioning and/or diseases.

Results

Defining the absolute loads

As mentioned in **Chapter 1**, the total bacterial load in each sample was assessed using ddPCR in order to confirm our hypothesis on determining putative contaminants, *Halomonas* and *Shewanella*. Additionally, loads of *Lactobacillus* and *Pelomonas* were assessed in order to apply them as quantitative traits for linkage mapping analysis. To determine the primer pair for estimating total bacterial load, preliminary tests were performed with three different primer pairs on lung and caecum tissue samples, each with three dilution series, 1:2, 1:5, and 1:100. With this test, a primer pair targeting the V2 hypervariable region of the 16S rRNA gene was selected, (Sze et al., 2014), which produced an appropriate separation between positive and negative droplets and produced accurate changes in the load according to the dilution series. Here, the optimal amount of template nucleic acid was determined to be 10ng in the final 20μL droplet mixture. Primers for *Lactobacillus* were selected similarly, but were additionally tested with DNA from various other bacteria (*Lactobacillus*, *Enterococcus*, *Lachnospiraceae*, *Clostridium*, *Bacteroides*, and *Olsenella*) to verify the target specificity. One primer pair, F-Lacto and R-Lacto (Delroisse et al., 2008) was chosen, which successfully amplified only *Lactobacillus*. A primer pair for *Pelomonas* was tested using *Pelomonas* bacterial DNA. With target specificity of ddPCR, absolute load assessment was performed at both DNA and RNA levels.

Absolute load as a trait for QTL mapping

In the previous chapter, QTL mapping was performed on the relative abundances of a "filtered" dataset without the two contaminant taxa, in order to identify the host genomic regions affecting the variation in the lung microbial traits. Similarly, absolute loads were assessed and mapped as phenotypic traits, including the total bacteria load and two of our candidate taxa, *Lactobacillus* and *Pelomonas*, measured at both the DNA and RNA level for 242 samples. For the two candidates, genus-specific primer pairs were used, which allows target-specific measurement and also helps reduce the environmental noise to which universal primers are prone.

Summary statistics were calculated in order to identify and determine the inter-individual variability of the traits (Table S1), where greater variation was detected for individuals with higher bacterial load ($p < 0.001$, r = 0.9874). Then I tested the effects of variables such as body weight, age, sex, and cage information on the loads using ANOVA. Age, sex, and cage information have significant effects ($p < 0.05$) and were thus included as variables in the QTL mapping. Age and sex were included as fixed effects and cage information as a random effect, and the percentages of total variance explained by these effects were quantified. For example, the fraction of variance explained by cage is 0% for *Lactobacillus* load while 1% is explained for *Pelomonas* load (RNA). Overall, the majority of the variance in our absolute load traits remains as residual variation even after accounting for sex, age, and cage effects (Table S2).

In order to identify host genomic regions associated with variation in lung microbial load, linkage mapping was performed on the ddPCR-derived data. Overall, I identified two significant ($p < 0.05$) QTLs. A region between 131.79 and 133.44 Mb on chromosome 1 is associated with variation in *Lactobacillus* load (DNA level) and a region between 21.98 and 25.57 Mb on chromosome 4 is associated with variation in *Pelomonas* load (RNA level) (Fig. 1, Table 1). The confidence intervals are 1.65 and 3.6 Mb, and the number of genes in a given interval are 22 and nine, respectively.

In order to examine the variation of bacterial load in association with host genetics, the absolute loads as well as the relative abundances were compared among genotype categories at the peak SNPs, UNC1677482 and UNC6891976, associated with *Lactobacillus* and *Pelomonas* loads, respectively (Fig. 2, 3). A Kruskal-Wallis test revealed a significant association between *Lactobacillus* load (DNA) and host genotype, in which the load is significantly lower

in mice with a GG genotype, compared to those with AA or AG genotypes ($p < 0.001$). *Lactobacillus* relative abundances also show significant associations, where mice with an AA genotype carry a significantly higher fraction of *Lactobacillus* compared to the other two genotypes ($p < 0.05$). A Kruskal-Wallis test was also performed for *Pelomonas* load (RNA) on the genotypes of the peak SNP, which in contrast did not reveal any significance. However, the sample distribution is relatively biased towards the GG genotype with 194 samples, compared to only 3 samples for the AA genotype. The total bacterial loads were also examined among different genotypes at both peak SNPs, but no significant associations to the host genotype was detected (Fig. S2, S3).

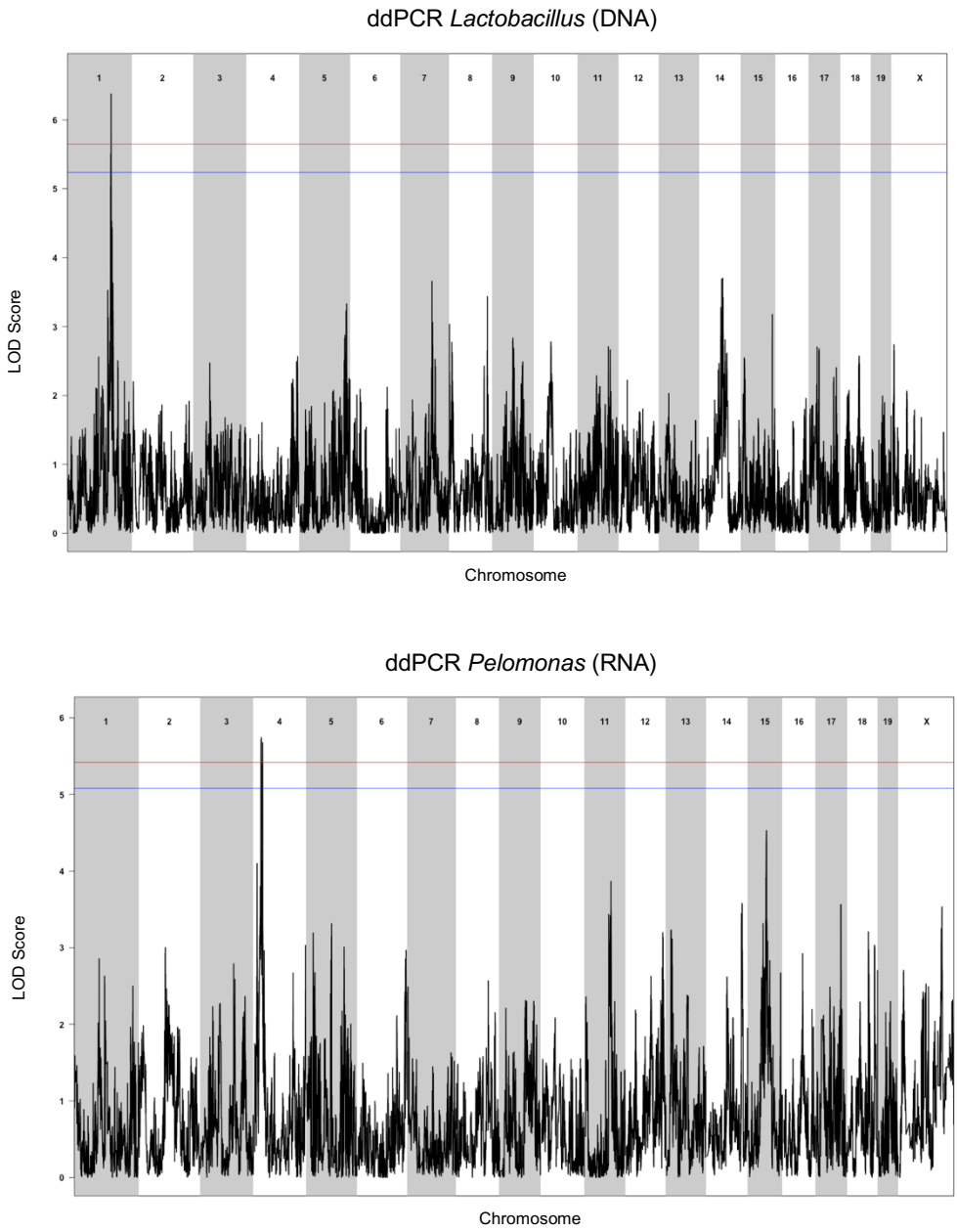

Figure 1. Manhattan plots from QTL mapping of *Lactobacillus* absolute load at the DNA level and *Pelomonas* absolute load at the RNA level. A) Lactobacillus absolute load at the DNA level and B) Pelomonas absolute load at the RNA level. LOD score indicates the -log P value of the association between a locus and a phenotypic trait. Red and blue lines in the plots each indicate significant ($p < 0.05$) and suggestive ($p < 0.1$) thresholds.

Table 1. QTL statistics of the ddPCR-derived bacterial load traits.

	Trait	Chr	Peak SNP	Position	LOD score	CI (Mb)	Phenotypic variance	Size
DNA	*Lactobacillus*	1	UNC1677482	132.96	6.38	131.79 - 133.44	11.43	1.65
RNA	*Pelomonas*	4	UNC6891976	23.43	5.74	21.98 - 25.57	10.35	3.6

(Chr = Chromosome)

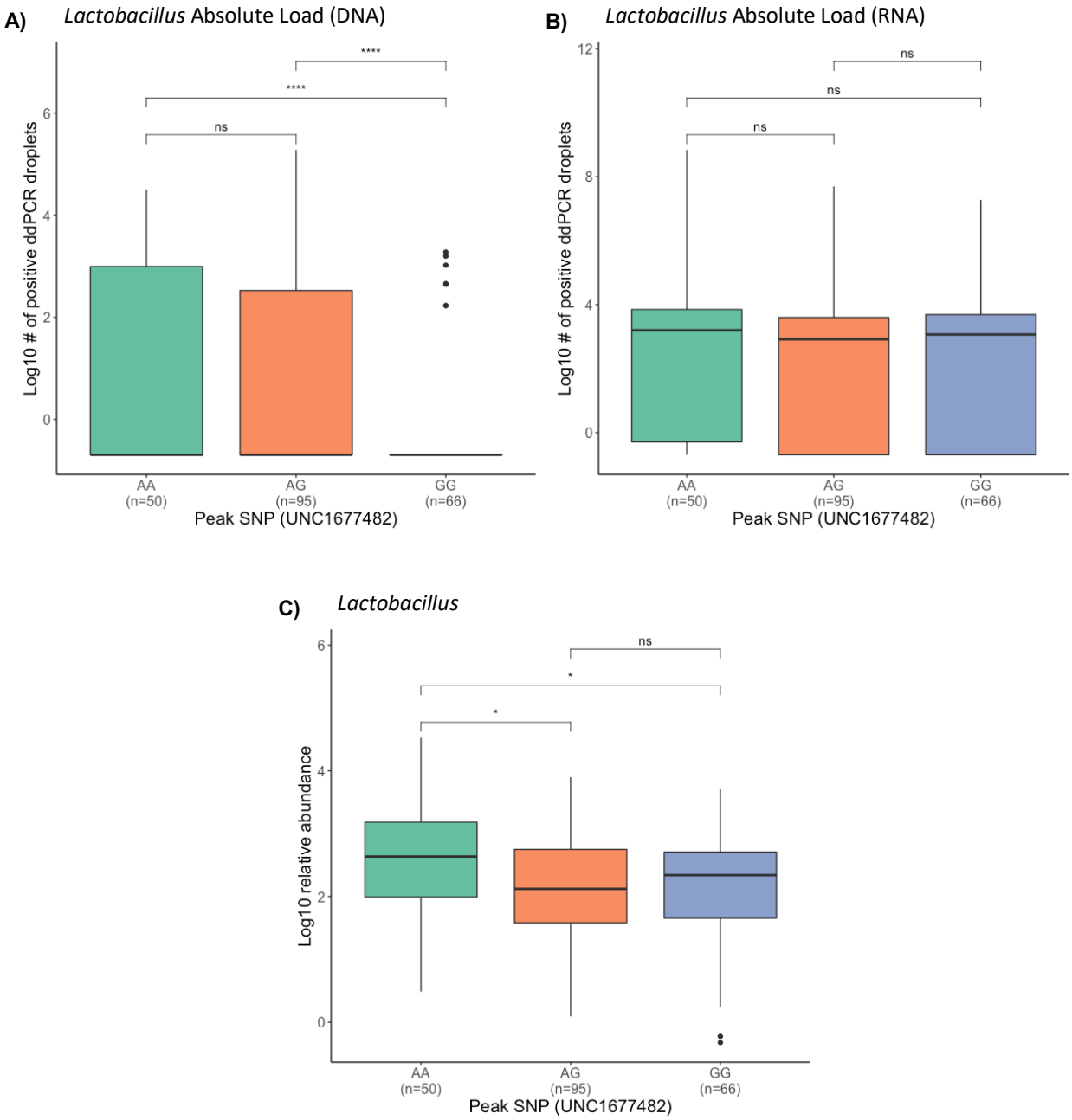

Figure 2. Differences in absolute loads and relative abundances of *Lactobacillus* in each genotype of peak SNP, UNC1677482. Absolute load at A) DNA and B) RNA levels, and C) relative abundance of *Lactobacillus* are compared among genotypes of the peak SNP within the confidence interval from QTL mapping. Loads and relative abundances are $\log_{10}$-transformed and compared using Kruskal-Wallis tests. $p < 0.0001$ ****, $p < 0.001$ ***, $p < 0.01$ **, $p < 0.05$ *, not significant "ns". (n = 211)

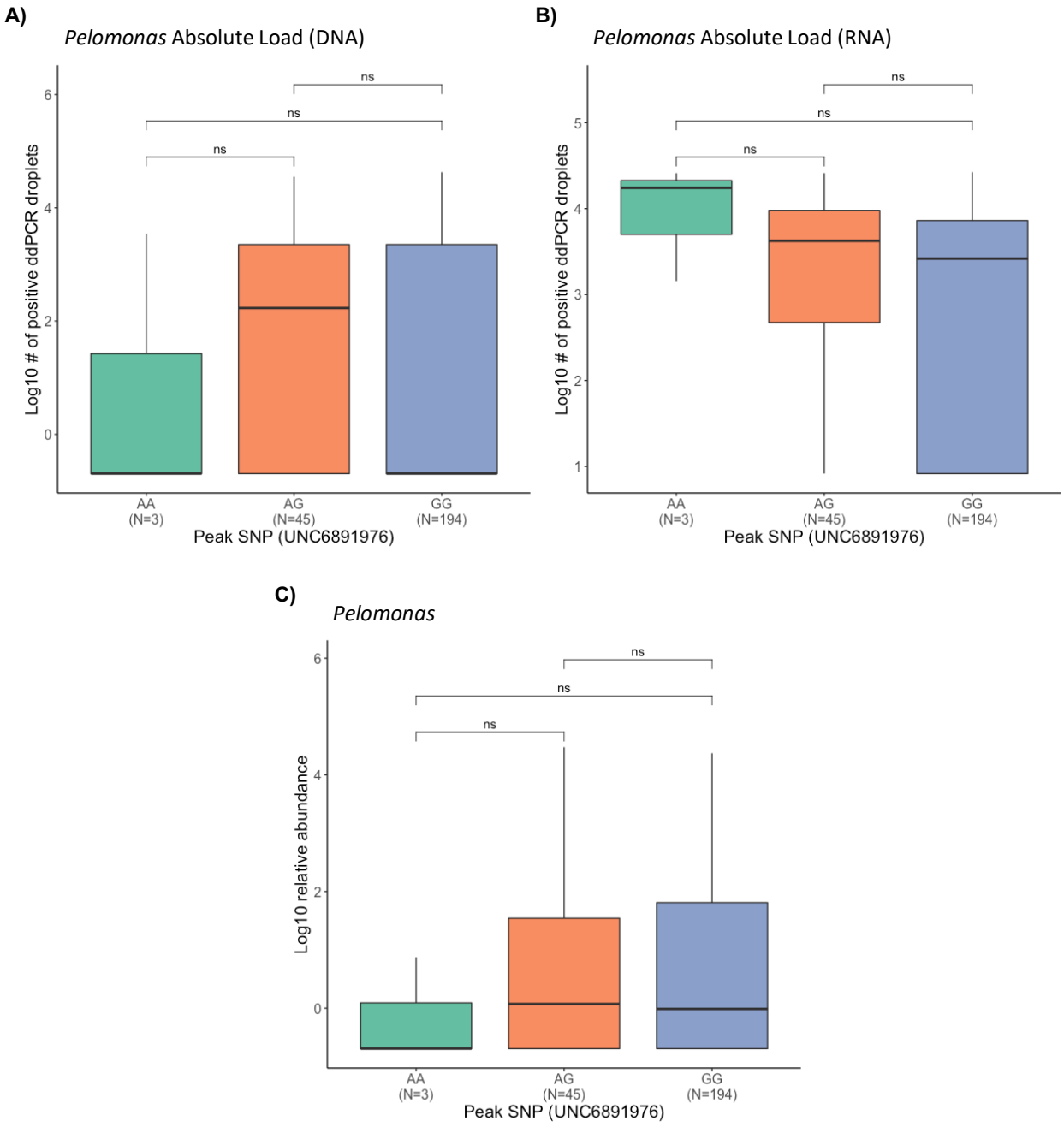

Figure 3. Differences in absolute loads and relative abundances of *Pelomonas* in each genotype of peak SNP, UNC6891976. Absolute load at A) DNA and B) RNA levels, and C) relative abundance of *Pelomonas* are compared among genotypes of the peak SNP within the confidence interval from QTL mapping. Loads and relative abundances are $\log_{10}$-transformed and compared using Kruskal-Wallis tests. $p < 0.0001$ ****, $p < 0.001$ ***, $p < 0.01$ **, $p < 0.05$ *, not significant "ns". (n = 242)

Candidate gene analysis

Within each confidence interval of the *Lactobacillus* and *Pelomonas* QTLs, there are 21 and nine underlying genes, respectively (Fig. 4A, C). Most genes within the intervals are related to one or more of the following processes: immune response, inflammatory response, cell apoptosis, and DNA repair. A number of genes are notable due to their proximity to the peak SNP and their role in lung functioning and disease susceptibility. On chromosome 1, the *Mk2* and the *Il-10* genes are in close proximity to the peak SNP, UNC1677482 (Fig. 4A). *Mk2* is a pro-inflammatory kinase while *Il-10* is an anti-inflammatory cytokine. On chromosome 4 for *Pelomonas* load-associated QTL, several genes are detected within the confidence interval, but the peak SNP appears to be lie in an intergenic region (Fig. 4C). A list of genes especially those related to lung functioning and diseases are shown in Table 2. *Pou3f2* and *Mms22l* are both related to lung cancer, such as non-small-cell lung carcinoma (NSCLC), whereas *Klhl32* is related to COPD. Unlike the rest of the genes, the function of *Klhl32* is poorly known. It is a protein-coding gene that shows a strong association with DNA methylation levels of a specific CpG-site (a cytosine base located adjacent to a guanine base) in subjects with COPD (de Vries et al. 2019).

Next, I estimated the effect of founder haplotype on the significant QTLs. Founder effects are estimated by calculating the association between founder haplotypes at each locus and the phenotypic trait. The chromosome 1 locus is shown to be most affected by the MRL strain while the locus on chromosome 4 is most affected by the CAST strain. For *Lactobacillus* load (DNA level), the MRL strain has the most effect, while *Pelomonas* load (RNA level) is most influenced by the CAST strain (Fig. 4B, D). MRL and NZM are known to be prone to autoimmune diseases, and CAST is a wild-derived line. MRL mice are also known to have an outstanding wound healing ability due to faster dermal cell regeneration along with the presence of healing QTLs such as *heal1* and *heal2* (Blankenhorn et al., 2003; Kench et al., 1999).

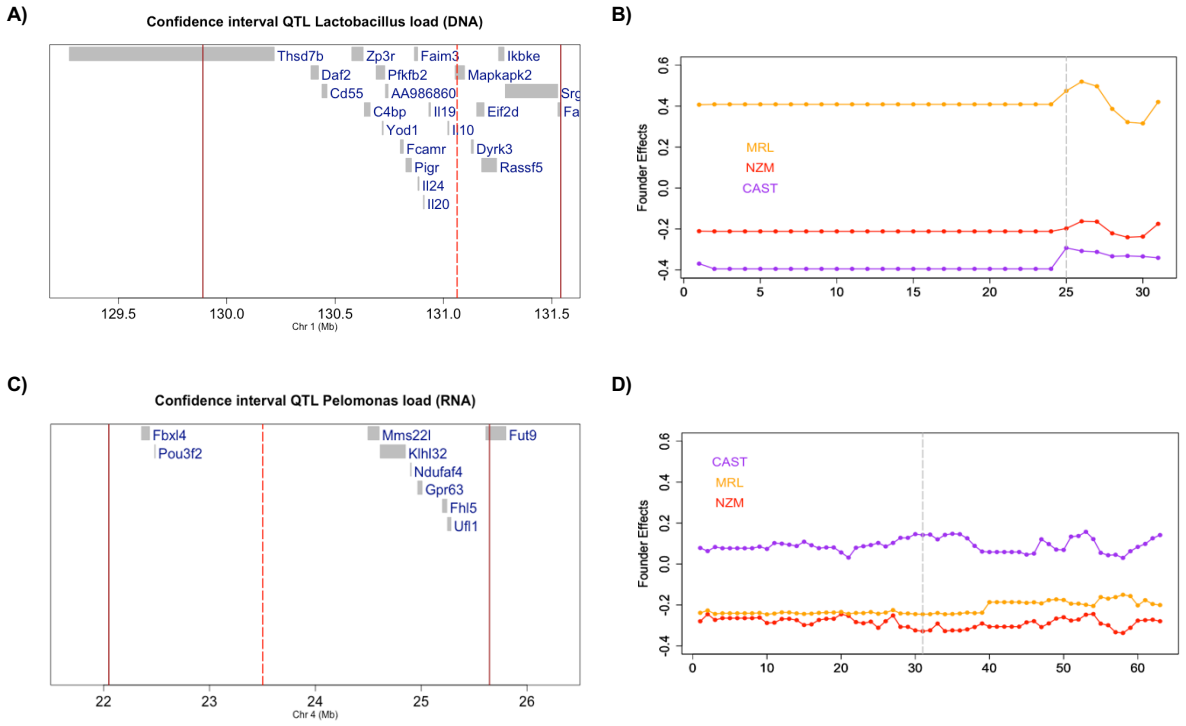

Figure 4. Confidence intervals of gene regions and founder effect plots. A) *Lactobacillus* absolute load (DNA level) on chromosome 1 and B) corresponding founder effects for *Lactobacillus*. C) *Pelomonas* absolute load (RNA level) on chromosome 4 and D) corresponding founder effect plot for *Pelomonas*. The location in Mb on each chromosome is shown on the x-axis. The red dotted line marks the peak SNP and dark red lines on each side indicate lower and upper limit of the interval. Founder effects are shown for each trait with chromosome in Mb on the x-axis with a grey dotted line indicating the peak SNP, while the y-axis represents the level of effect of each founder, CAST, MRL, and NZM.

Table 2. List of candidate genes within confidence intervals and their functions.

Trait	Gene	Functions (& References)
Lactobacillus	*Mk2*	Lung cancer (Liu et al. 2012); inflammatory pulmonary diseases (Qian et al., 2016); inflammation (Scheller, Chalaris, Schmidt-Arras, & Rose-John, 2011)
	Il-10	Tuberculosis (Turner et al. 2002); asthma (Lyon et al. 2004); NSCLC (Vahl et al. 2017; Wang et al. 2011)
	Il-19	Antimicrobial defense in airway epithelial cells (Goodale, Rayack, & Stanton, 2017)
	pIgR	COPD-like phenotype airway inflammation (Richmond et al. 2016)
	Yod1	Chronic hypoxia-induced pulmonary vascular remodeling (Yang et al. 2012)
	Faim3	NSCLC (Gentles et al. 2015)
	Ikbke	NSCLC (Challa et al. 2016)
	Cd55	NSCLC (Dho et al. 2018); hypoxia-associated pulmonary disease (Pandya et al. 2016)
Pelomonas	*Pou3f2*	NSCLC (Ishii et al. 2014)
	Mms22l	Lung carcinogenesis (Nguyen, Ueda, Nakamura, & Daigo, 2012)
	Klhl32	COPD (de Vries et al., 2019)
	Fut9	Bronchopulmonary dysplasia (Chaubey et al., 2020)

Gene expression

We next prioritized two candidate genes due to their close proximity to a peak SNP and their known involvement in the regulation of inflammation in the lung, *Il-10* and *Mk2*. Accordingly, ddPCR assays were performed on these genes in order to explore possible relationships between gene expression and *Lactobacillus* load and/or host genotype. As assessed by Spearman correlation between gene expression levels and *Lactobacillus* load (DNA), *Il-10* expression did not display a significant difference according to genotype. On the other hand, *Mk2* expression showed significantly negative correlation to *Lactobacillus* load ($p < 0.0001$, $r = -0.4837$). In order to evaluate the association between gene expression level and host genotype, a Kruskal-Wallis test was performed. Here, 20 samples were randomly selected for each genotype and a total of 60 samples were used. Primers used are listed in Table 3. This revealed no significant difference according to host genotypes and *Il-10* expression (Fig. 5). *Mk2,* in contrast, is expressed significantly less in individuals with an AA genotype and more in those with GG and heterozygous genotypes.

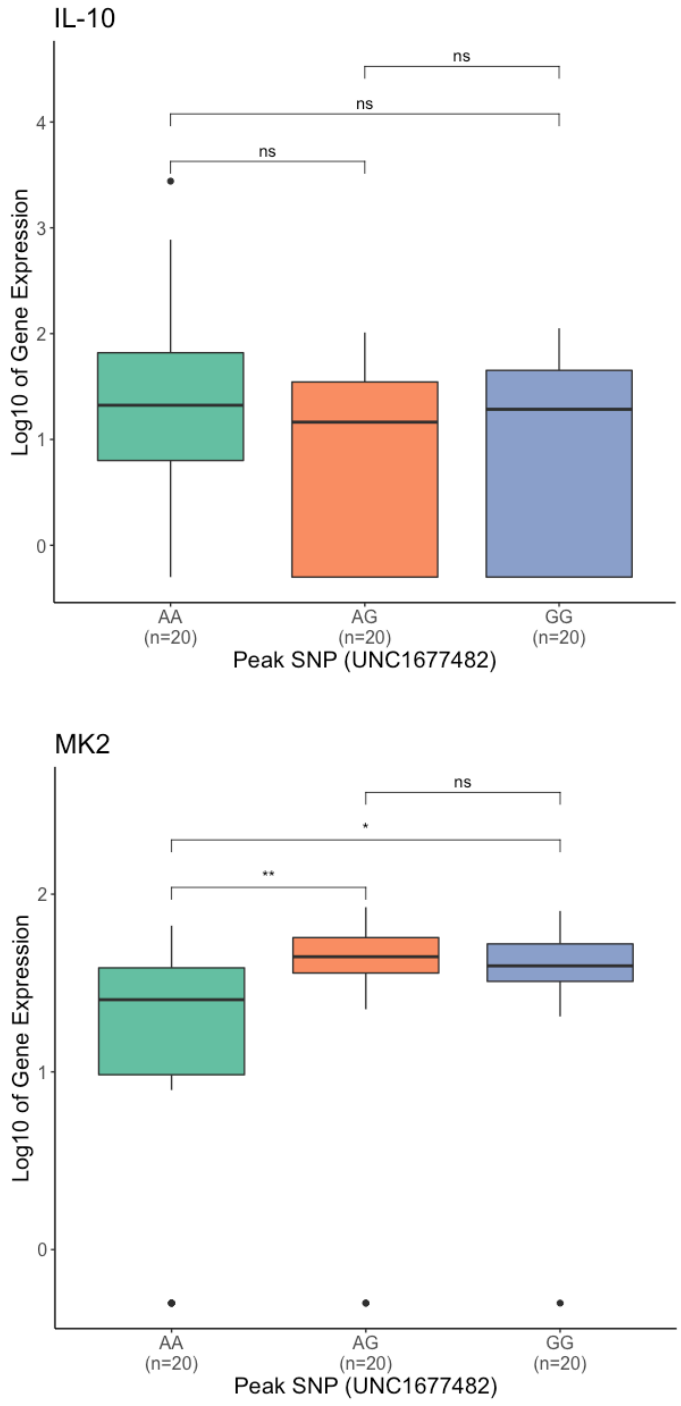

Figure 5. Gene expression levels at the peak SNP. Expression levels of A) *Il-10* and B) *Mk2* are measured for each genotype of the peak SNP, UNC1677482. Kruskal-Wallis test; $p < 0.001$ ***, $p < 0.01$ **, $p < 0.05$ *, not significant "ns". (n=60)

Discussion

Approaching QTL mapping with absolute load data

As lungs display low microbial biomass, and hence a generally low copy number of the bacterial16S rRNA gene (Sze et al., 2014), we are challenged to distinguish between a real signal and environmental noise when using a PCR amplicon-based sequencing approach. In order to capture the true residents, we first analyzed lung microbiota at the bacterial transcript (RNA) level and took a multitiered approach of 16S rRNA gene sequencing and ddPCR.

In this study, we expanded our previous work on the G_{15} AIL population by examining the absolute load of bacteria using ddPCR, a highly sensitive quantitative PCR method for quantifying copies of a target sequence by partitioning samples into approximately 20,000 oil-capsulated droplets. There are advantages in using ddPCR to measure absolute loads, as ddPCR itself carries high precision by creating individual droplets and this can help reduce PCR bias against contaminants or environmental noise, especially by using target-specific primers (Maheshwari et al., 2017; Sze et al., 2014; Taylor et al., 2017). However, as mentioned in a previous study (Sze et al., 2014), the nucleic acid input has to be controlled due to aspects of the partitioning technology of droplets, as well as the calculation based on Poisson modeling. Moreover, one should note the primer compatibility with the samples and DNA concentration when starting ddPCR as primers targeting different hypervariable regions of the 16S rRNA gene have varying levels of sensitivity and specificity. With this PCR method, the chances of amplifying environmental noise can be even further reduced by adapting genus-specific primers. Examining the absolute abundance of specific genera improved our confidence and strength of mapping, and allowed us to identify strong candidate genes as a result. qPCR can be applied to samples with greater bacterial biomass, but ddPCR will be more useful in samples like lungs and skin with low bacterial abundance in order to overcome such challenges. Studies comparing the two methods pointed out the lower detection limit of qPCR as its weakness for studies on low biomass samples, yet showing greater noise to signal ratio shown through negative controls (Sze et al., 2014) and inconsistent results in diluted samples with contamination (Taylor et al., 2017). Taylor et al. (2015) also points out advantages of ddPCR especially for studies over a long period of time due to its high-precision technology and thus being able to produce stable results.

Host-microbiota association

Absolute quantification not only reduced environmental noise but also refined the data for QTL mapping. Two taxa, *Lactobacillus* and *Pelomonas*, were quantified using genus specific primers at both the DNA and RNA levels. While relative abundance based on 16S rRNA gene profiles (**Chapter 1**) of the two did not reveal significant associations with the host genome, their absolute load information, in contrast, did. *Lactobacillus* is a genus known to be found in the murine lungs (Barfod et al., 2013; Singh et al., 2017; Yun et al., 2014) and was the most abundant genus throughout G_{15} murine lung tissue samples. *Pelomonas* was a genus of interest as this was found only in the lungs (but not in the intestines nor vagina) according to a previous study (Barfod et al., 2013). Genes found within the confidence intervals on mouse chromosomes were often related to various lung diseases and host inflammatory responses. *Pou3f2* (class III POU gene 2), one of the candidate genes in the confidence interval of the *Pelomonas* QTL, was found to play a role in proneuronal/neuroendocrine differentiation of lung cancer cells, in which the gene was expressed in small cell lung cancer (SCLC), and its induced expression led to proneuronal/neuroendocrine gene expression in non-SCLC (NSCLC) cells (Ishii et al., 2014). Another gene included within the confidence interval of *Pelomonas* QTL, *Mms22l* (methyl methanesulfonate-sensitivity protein 22-like) was previously found to be over-expressed in clinical and esophageal cancers, playing a role in growth and survival of cancer cells (M. H. Nguyen et al., 2012). This gene was suggested to have an impact on the efficacy of DNA-damaging agents, as the knockout of the gene enhances cancer cell apoptosis (M. H. Nguyen et al., 2012). Based on these observations, we can speculate that *Pelomonas* may play a role in susceptibility to lung cancer and/or other lung diseases in association with these candidate genes. According to studies in a different field apart from the mammalian lung microbiota, *Pelomonas* contributes to fungal resistance and contraction frequency in hydra (Fraune et al., 2015; Murillo-Rincon et al., 2017). Irrespective of the different host species, it is intriguing to note the mechanism and strategy *Pelomonas* carries. In hydra, small molecules such as amino acids can influence contraction frequency and activity including GABA (gamma-Aminobutyric acid) (Murillo-Rincon et al., 2017). It is intriguing to discover that GABA is one of the amino acids that is also found in pulmonary neuroendocrine cells (PNEC), distributed along the alveolar airway epithelium (Yabumoto et al., 2008). The study shows that GABA can play a role through its receptors in the epithelial cells of murine lungs (Yabumoto et al., 2008). Another study shows that GABA is indeed mainly produced in PNEC in primates and can affect mucus production, together with IL-13 (Barrios et al., 2019).

Based on these studies, we speculate a potential role played by *Pelomonas* upon GABA, which could then affect lung functioning. Suggested future studies would be to investigate knockout mice of one of the above candidate genes or examine differences in gene expression according to the presence or absence of *Pelomonas* in the lungs.

Genes that are detected near or at the peak SNP within the interval are more intriguing to examine as this SNP implies a statistically high association between a locus and a phenotypic trait, such as candidate bacterial loads. In order to investigate further into the association between *Lactobacillus* load and the two candidate genes, gene expression analysis was conducted using ddPCR, which revealed a significantly negative correlation between *Lactobacillus* load and the *Mk2* gene. On the other hand, the *Il-10* expression did not show any correlation. Indeed, MK2 as a substrate of $p38^{MAPK}$ pathway plays a role in immune response during bacterial and viral infections, and can affect translation of IL-10 and its transcript stability (Bode et al., 2012; Ehlting et al., 2016). The $p38^{MAPK}$ pathway, which includes MK2 as a downstream substrate, takes part in regulating cell apoptosis, inflammation, and tumorigenesis (Maruyama et al., 2009). Thus, genetic variants in this pathway or in MK2 may contribute to susceptibility to cancer.

Our AIL was produced by intercrossing three laboratory mouse strains, BXD2, MRL, and NZM, and one wild mouse strain, *Mus musculus castaneus,* or CAST. These strains were intercrossed with even sex and strain distribution for 15 generations (Belheouane et al., 2017). As we determined which strain had the most impact on the mapping result, the MRL strain showed the strongest founder effect on the QTL of *Lactobacillus* load on chromosome 1, and CAST on the QTL of *Pelomonas* load on chromosome 4. MRL and NZM are strains developed to be prone to autoimmune diseases, whereby the mice can display symptoms when triggered. With these strains, we are thus able to examine genes associated with host susceptibility to autoimmune diseases.

Candidate genes & gene expression

MK2 is part of the feedback and crosstalk mechanisms that modulate the macrophage response towards inflammatory mediators, and responds to both lipopolysaccharide (LPS)- and virus-induced infection of macrophages, which then induce various other signals such as cytokines (Ehlting et al., 2016; Kotlyarov et al., 1999; Ronkina et al., 2007). Thus, it is involved in transcriptional and post-transcriptional regulation of cytokine expression. In the study by Ehlting et al. (2016), the role of MK2 in viral infection was examined in macrophages as well hepatocytes, where it was found to be required for up-regulation of IL-6, IL-10, and TNF-α. MK2 was also found

to play a role in IL-10 transcript stability by down-regulating TTP, a mRNA-degrading protein (Cao et al., 2003; Marchese et al., 2010; Ronkina et al., 2007; Tudor et al., 2009). IL-10 transcript stability was reduced in *Mk2* knockouts, while it was maintained in TTP knockout macrophages, but was reduced again when those TTP knockout macrophages were given deoxycycline to reconstruct TTP (Ehlting et al., 2016). Here, we recognize the role of MK2 in different pathways, which then affect other pathways and the expression of cytokines. Overall, it is important to note the potential role of MK2 on host inflammation and disease susceptibility, as it controls the expression and production of IL-10 and other cytokines.

The functional relationship between *Lactobacillus* and MK2 and/or IL-10 is yet unclear. Strains of *Lactobacillus* have previously been shown to induce IL-10 and other cytokines and to ameliorate inflammation (Racedo et al., 2006; Villena et al., 2012), but another study showed that the effect was short-termed *in vitro* (Noguchi et al., 2012). Nasal administration of *L. casei* helped lung recovery from *S. pneumoniae* (Racedo et al., 2006) and oral administration of *L. rhamnosus* instigated changes in cytokine levels that resulted in recovery from lung injury induced by poly(I:C), an immunostimulant similar to viral infection, with symptoms including pulmonary inflammation (Villena et al., 2012). It is possible that the presence of *Lactobacillus* and its metabolites may send a positive or negative feedback to $p38^{MAPK}$ and its downstream signals, which then play a role in lung functioning and recovery. Alternatively, active or inactive *Mk2* expression and subsequent IL-10 expression may affect the survival of *Lactobacillus* in the lungs.

As the *Il-10* expression data did not show any correlation to host genotypes, we speculate a potential macrophage specific effect measured in whole lung tissue. In murine lungs, IL-10 is produced majorly via interstitial macrophages (Kawano et al., 2016). However, as we extracted RNA from partial lung tissue, we speculate that it might not be sufficient to capture differences in expression considering where the gene is expressed within the tissue or organ. Although I performed ddPCR in order to estimate the number of macrophages, the primer pair was not specific for interstitial macrophages and accordingly did not produce any significant results (Fig S4). As a follow-up experiment, ddPCR measurement specifically for interstitial macrophages can be conducted. Single cell RNA-sequencing could also be conducted in the future, which is able to capture specific gene expression in each cell.

Conclusion

Although the lungs were traditionally viewed as sterile in healthy individuals, it is becoming increasingly clear that microbes inhabit the lower airway tract and likely contribute to disease susceptibility. Using a QTL mapping approach and high-precision ddPCR to describe bacterial communities without the need to cultivate (i.e. metagenomics), we examined the link between individual lung bacterial groups and markers in the host genome, in an effort to identify host genes that influence lung microbes and may play a role in lung functioning and disease susceptibility. For example, the amount of *Lactobacillus*, a common lung microbe, displays a strong association with a region of mouse chromosome 1 containing interesting candidate genes such as *Il-10* and *Mk2*, which are known to play a role in lung disease. Through our in-depth study on the murine lung microbiota, we generated novel working hypotheses for the impact of candidate bacteria and disease genes on host inflammatory responses in the lung, which will be explored in future work.

Methods

Sample preparation

DNA and RNA extracted from G_{15} AIL mouse population were used for ddPCR. This population is a result of intercrossing 4 mouse strains, MRL/MpJ, NZM2410/J, BXD2/TyJ, and CAST/EiJ, with equal sex and strain distributions for 15 generations (Belheouane et al., 2017; Srinivas et al., 2013). Lung tissues were preserved RNALater (Invitrogen) and at -20C until extraction using AllPrep 96 DNA/RNA kit (QIAGEN). Reverse transcription of RNA samples was performed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) by adding 10µL RNA. Refer to **Chapter 1** for details.

Digital droplet PCR (ddPCR) process

i. Determination of primer pair and template load

First, we used DNA and RNA from multiple tissue samples of lung, skin, and caecum, and prepared dilutions (1:2, 1:5, 1:10) in order to confirm the compatibility of our chosen primers to lung tissue samples and determine the concentration of DNA and RNA that allows a good separation between positive and negative droplets. For the purpose of testing, 3 bacterial primer pairs were used (Table 3) and the one highlighted, targeting the V2 hypervariable region of the 16S rRNA gene, was chosen, as it produced a good separation between positive and negative droplets throughout different samples. Based on the dilution series, we concluded to add templates so that 10ng of DNA or cDNA is mixed into the final droplet mixture. DNA samples were measured using NanoDrop 1000 (Thermo Scientific) and were diluted to the lowest concentration (2.5ng/µL) among samples. Concentrations of RNA were also measured (in **Chapter 1**) and normalized to 100ng/µL, and 10uL of each RNA sample was used in reverse transcription for cDNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems).

As the DNA and cDNA template input amount was determined above, only target specificity and annealing temperatures were tested for genus-specific primer pairs. Annealing temperature for primers were determined by gradient PCR, whereby 60°C for the primer pair for total bacteria was chosen. For *Lactobacillus*, primers were tested for target specificity by PCR amplification and gel electrophoresis using different bacterial DNA of *Lactobacillus*, *Enterococcus*, *Lachnospiraceae*, *Clostridium*, *Bacteroides*, and *Olsenella*. The final primer pair,

highlighted in the table, was chosen as it amplified only *Lactobacillus* while other primer pairs also amplified other genera as well. Regarding primer pairs for *Pelomonas,* the sequences of the 16S rRNA gene of 3 species of *Pelomonas* (*P. aquatica, P. puraquae*, and *P. saccharophila*) were obtained through Geneious (v.10.2.3) and used to perform multiple alignment using Clustal W. The alignment was exported in FASTA format and used to design primers on R studio with the "DesignPrimers" function in the R package "DECIPHER" (v.2.12.0) to generate genus-specific primers. Primers were tested on previously described bacterial DNA and *Pelomonas* DNA. *Pelomonas* culture was obtained from AG Prof. Dr. Thomas Bosch, CAU Kiel, and DNA was extracted using DNeasy UltraClean Microbial kit (QIAGEN), following the manufacturer's instructions. The primer pair that amplified only *Pelomonas* was chosen. For assessing the expression level of the two candidate genes, *Il-10* and *Mk2*, only lung samples were used to determine the input concentration. All primers are listed in Table 3 with final pairs highlighted with bold text.

Table 3. List of primers used for ddPCR, including test primers. Final pairs total bacteria and *Lactobacillus* are highlighted.

	Primer	5' – 3'	Reference
Total bacteria	**63F**	**GCAGGCCTAACACATGCAAGTC**	**(Sze et al., 2014)**
	355R	**CTGCTGCCTCCCGTAGGAGT**	
	1048F	GTGSTGCAYGGYYGTCGTCA	(Raveh-Sadka et al. 2015)
	1194R	ACGTCRTCCMCNCCTTCCTC	
	F357	CCTACGGGAGGCAGCAG	(Greathouse et al. 2018)
	R534	ATTACCGCGGCTGCTGG	
Lactobacillus	**F-Lacto**	**GAGGCAGCAGTAGGGAATCTTC**	**(Delroisse et al., 2008)**
	R-Lacto	**GGCCAGTTACTACCTCTATCCTTCTTC**	
	LbLMA1-rev	CTCAAAACTAAACAAAGTTTC	(Dubernet et al. 2002)
	R16-1	CTTGTACACACCGCCCGTCA	
	F_alllact	TGGATGCCTTGGCACTAGGA	(Haarman & Knol, 2006)
	R_alllact	AAATCTCCGGATCAAAGCTTACTTAT	
Pelomonas	357F	CGGGTTGTAAACCGCTTTTGT	
	550R	CGGGGATTTCACCTCTGTCT	
Il-10	IL-10F	TGGCCCAGAAATCAAGGAGC	(Kawane et al., 2010)
	IL-10R	CAGCAGACTCAATACACACT	
Mk2	7F	GGGCACCATG CTGTCGGGCTC	(Sudo et al., 2005)
	424R	CGAGACACT CCATGACAATCAGC	

ii. ddPCR preparation

The 20uL ddPCR mastermix included 1x of 2x QX200™ ddPCR™ EvaGreen SuperMix (BioRad), 1µL each of 2uM forward and reverse primers, H_2O, and 10ng of either DNA or cDNA template. The mastermix was loaded into sample wells of DG8™ Cartridges for the QX200™/QX100™ Droplet Generator (BioRad) with 70µL of QX200™ Droplet Generation Oil for EvaGreen and DG8™ (BioRad). Approximately 35µL of droplets with oil was generated on QX200™ Droplet Generator. Droplets were transferred to 96-well, 250µL, semi-skirted PCR plates (Eppendorf) and sealed using piercable heat seal foil (BioRad). PCR conditions follow: 1 cycle at 95°C for 5 min, 40 cycles at 95°C for 15 sec and 60°C for 1 min, 1 cycle at 4°C for 5 min, 1 cycle at 90°C for 5 min, and 10°C forever. Final products were transferred to QX200™ Droplet Reader and quantified as gene copies per µL using BioRad QuantaSoft (v.1.7.4.0917).

QTL mapping

First, summary statistics on each ddPCR trait were calculated and then Pearson's correlation test was done between the average load and variation in each trait in order to identify and determine inter-variability of the trait throughout the samples. Absolute loads of total bacteria, *Lactobacillus*, and *Pelomonas* were mapped as a phenotypic trait as described in Methods section ("**QTL mapping**") in **Chapter 1**. As two significant ($p < 0.05$) (Belheouane et al., 2017) QTL were detected for two traits, *Lactobacillus* load (DNA) and *Pelomonas* load (RNA), these loads were compared among the genotypes at each peak SNP, UNC1677482 and UNC6891976, respectively, using a Kruskal-Wallis test. Although the genotype data for UNC6891976 was fully available for 242 samples, corresponding data for UNC1677482 were available for only 211 samples.

Supplements

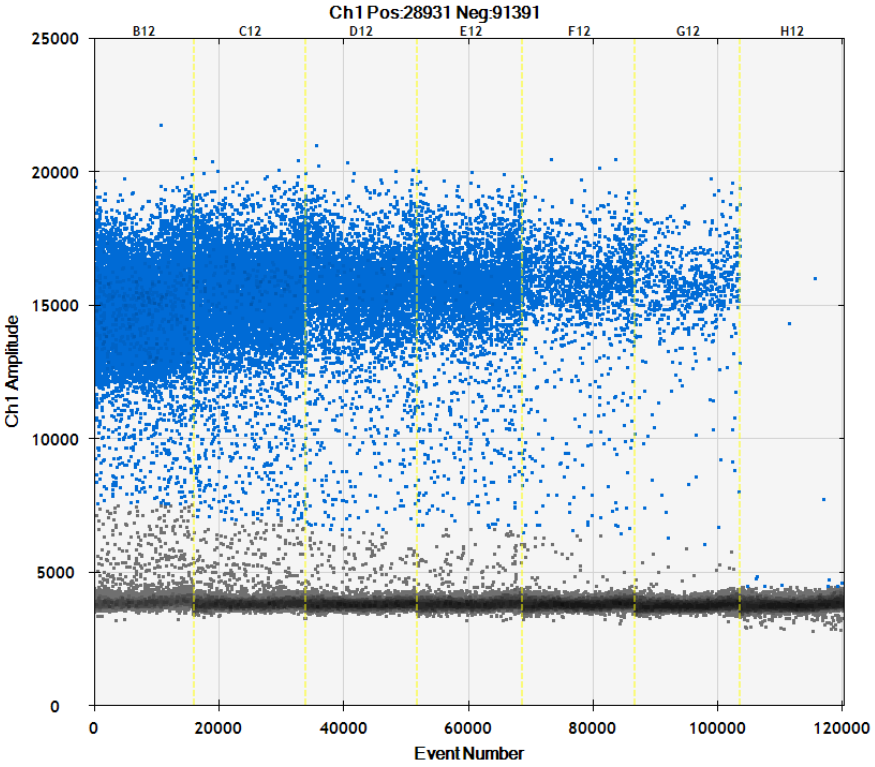

Figure S1. 2D graph of positive and negative droplets from ddPCR. Droplets are generated with a reagent mix including nucleic acid, primers, and ddPCR™ EvaGreen SuperMix, which has a fluorescent dye to help bind to target nucleic acid. This fluorescence is read through QX200 Droplet Reader and the results can be analyzed.

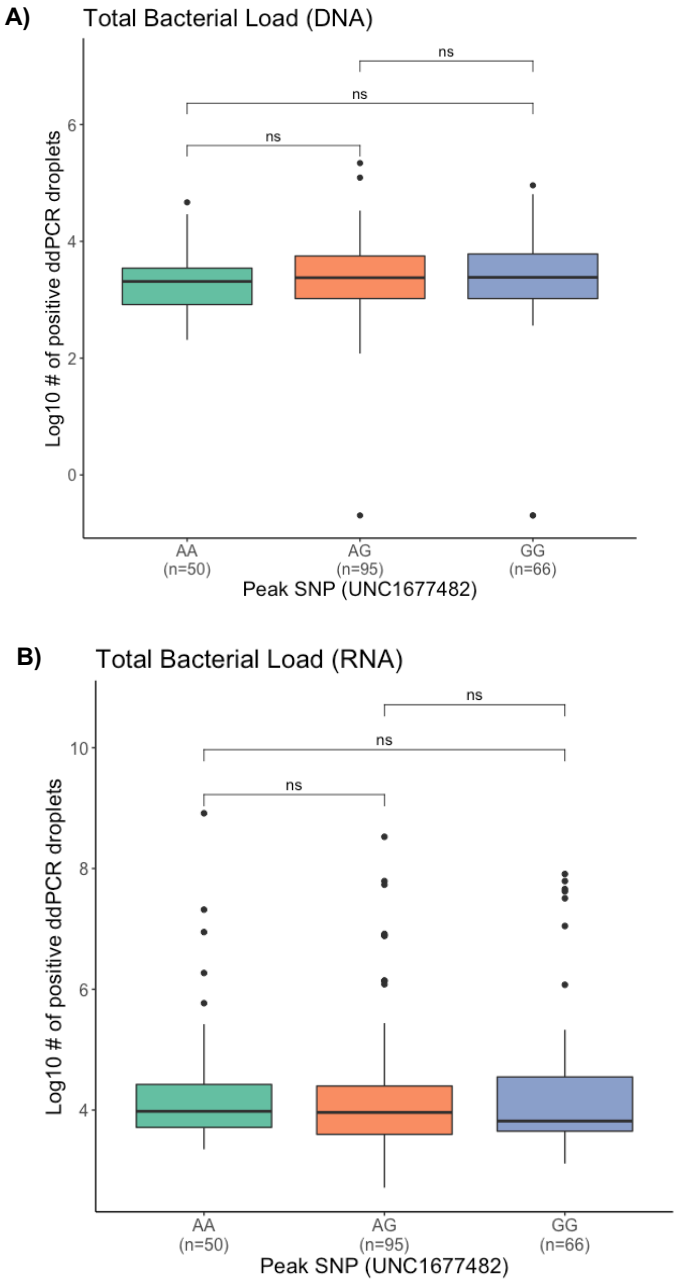

Figure S2. Total bacterial load among genotypes of the peak SNP. Total bacterial loads at A) DNA and B) RNA levels are compared among each genotype of the peak SNP, UNC1677482, of *Lactobacillus* QTL mapping. Kruskal-Wallis test: $p < 0.0001$ ****, $p < 0.001$ ***, $p < 0.01$ **, $p < 0.05$ *, not significant

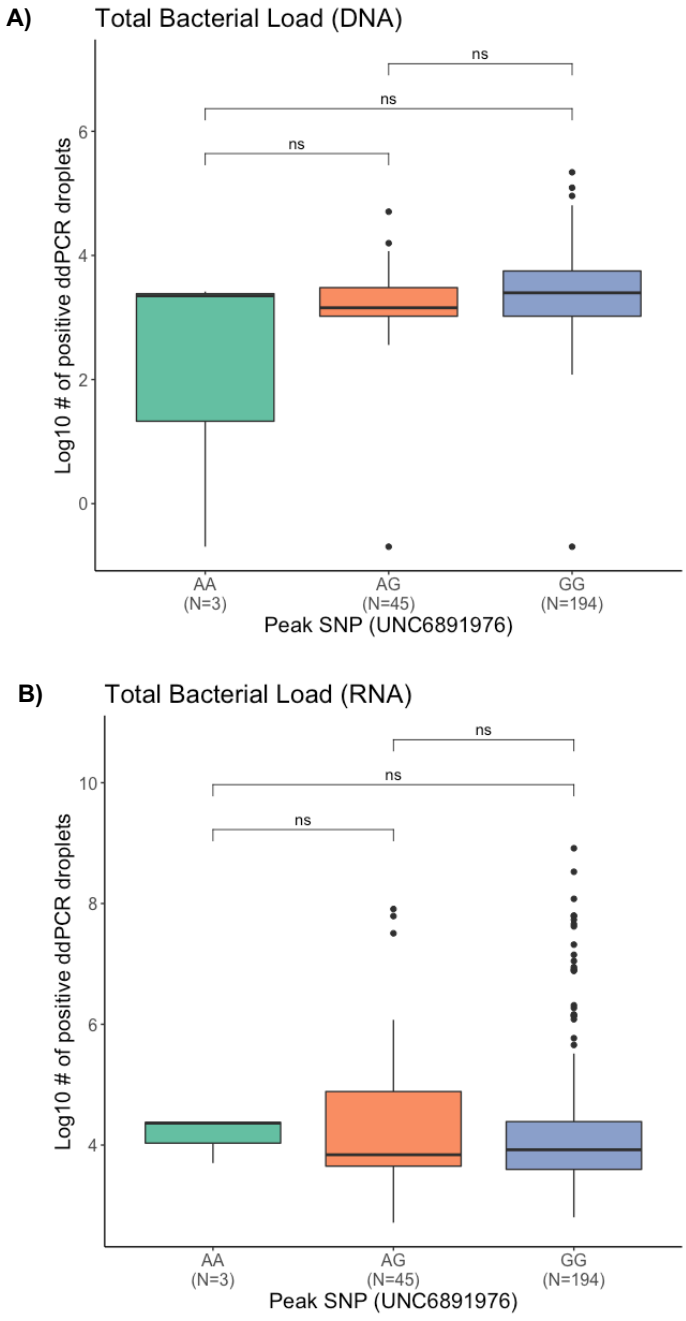

Figure S3. Total bacterial load among genotypes of the peak SNP. Total bacterial loads at A) DNA and B) RNA levels are compared among each genotype of the peak SNP, UNC6891976, of *Pelomonas* QTL mapping. Kruskal-Wallis test: $p < 0.0001$ ****, $p < 0.001$ ***, $p < 0.01$ **, $p < 0.05$ *, not significant

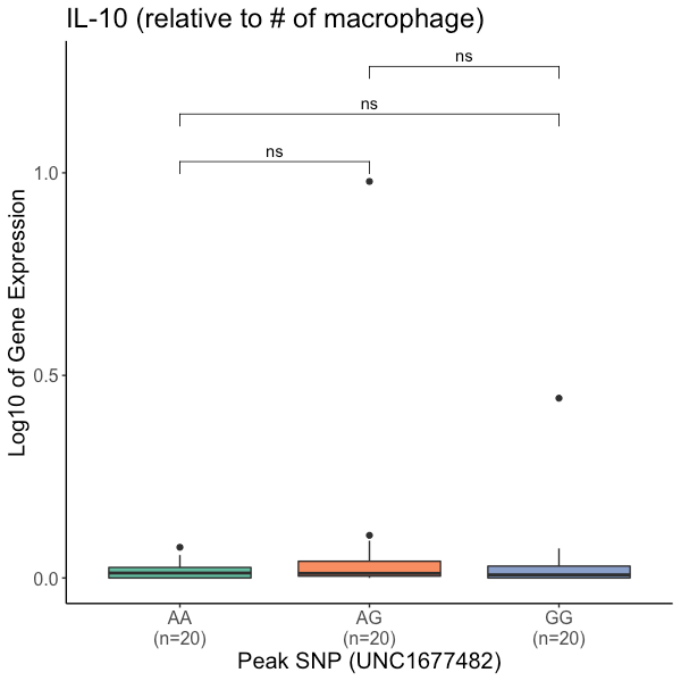

Figure S4. Gene expression level of *Il-10* relative to the number of macrophages. Expression level of *Il-10* relative to the number of macrophages (as estimated by ddPCR targeting *Emr1* of macrophage) in each sample (n=20 for each genotype) was assessed and compared among genotypes of the peak SNP, UNC1677482. Kruskal-Wallis test: $p < 0.0001$ ****, $p < 0.001$ ***, $p < 0.01$ **, $p < 0.05$ *, not significant = ns.

Table S1. Summary statistics of ddPCR CMM traits. (n=242)

	Traits	min	max	mean	std.dev	coef.var	var
DNA	Total Bacteria	0	208	34.511	25.264	0.732	638.276
	Lactobacillus	0	196	7.610	18.316	2.407	335.481
	Pelomonas	0	102	14.810	21.058	1.422	443.419
RNA	Total Bacteria	14.6	7440	243.933	736.298	3.018	542135.125
	Lactobacillus	0	6900	94.047	491.322	5.224	241397.798
	Pelomonas	1	83	29.785	25.155	0.845	632.763

(min = minimum, max = maximum, std.dev = standard deviation, coef.var = coefficient variation, var = variation)

Table S2. Portions of variance explained by explanatory variables.

		Variance component		% Total Variance		Fixed Structure
	Traits	**Cage**	**Residual**	**Cage**	**Residual**	**R^2m**
DNA	Total Bacteria	2.44E-19	9.25E-02	0.00	100.00	0.03645978
	Lactobacillus	6.65E-17	5.59E-01	0.00	100.00	0.01347863
	Pelomonas	2.44E-17	8.01E-01	0.00	100.00	0.000113983
RNA	Total Bacteria	9.38E-03	2.22E-01	4.04	95.96	0.07380843
	Lactobacillus	4.50E-17	9.01E-01	0.00	100.00	0.160099
	Pelomonas	0.003274	0.322789	1.00	99.00	0.06486212

Chapter III.

Analysis of the host candidate gene *Interleukin 10* (*Il-10*)

Introduction

Mammals carry out various mechanisms in order to fight and defend against infections. Many genes are involved in these different mechanisms, including the anti- or pro-inflammatory pathways. Interleukin-10, or IL-10, is a well-known anti-inflammatory cytokine that is produced by numerous immune cells but majorly from macrophages and monocytes in the lungs (Kawano et al., 2016; Sabat et al., 2010), and it modulates the function of several adaptive immunity-related cells. Generally, the release of IL-10 as an immunosuppressive molecule limits and terminates inflammatory responses (Moore et al., 1993), as it takes part in influencing functions of macrophages and/or monocytes. It inhibits antigen presentation by macrophages/monocytes, as well as the release of immune mediator such as TNF-α, IL-1β, IL-6 (Sabat et al., 2010).

In previous studies, mice with the *Il-10* gene deficiency were shown to display colitis unless they are under germ-free condition (Kühn et al., 1993; Madsen et al., 1999). With microorganisms present in the gut, certain groups of bacteria can help reduce the susceptibility to- or severity of inflammation. When probiotics are given, gut inflammation and tumor development can be reduced in *Il-10* knockout (KO) mice (O'mahony et al., 2001). The mechanisms of their probiotic effect are complicated, but the metabolites produced by gut microbes are often suggested (Mendes et al., 2019; O'mahony et al., 2001). Through these *Il-10* KO mouse model studies, we can understand that microorganisms and their byproducts initiate and modulate inflammation, which helps the host regulate mucosal immunity.

While this mouse model is more often applied in studies on gut health, it is also applied on respiratory tract infection studies. In such studies, *Il-10* deficiency in mice led to significantly reduced lung microbial load after *Mycobacterium tuberculosis* (*Mtb*) infection compared to wildtype mice (Redford et al., 2010). In these mice, a prolonged increase in IFN-ɣ and CD4+ T cells was observed throughout the infection, which indicates an enhanced immune response in the absence of *Il-10* in the lungs. Without the anti-inflammatory cytokine, the mice were able to control and defend against *Mtb* infection in association with an improved Th1 response. IL-10 is also associated with other lung diseases such as asthma (S. Lim et al., 1998; Lyon et al., 2004) and chronic obstructive pulmonary disease (COPD) (Seifart et al., 2005). As a crucial modulator in airway inflammation, the frequencies of inflammatory response gene polymorphisms in *Il-10* showed significance between lung cancer patients and controls (Seifart

et al., 2005). Here, genotypes carrying a certain polymorphism of *Il-10* promoter was associated with the presence of small cell lung cancer (SCLC). Lyon et al. (2004) examined six *Il-10* SNPs corresponding to asthma severity, in which a population-based analysis revealed an association between certain loci and asthmatic phenotypes.

In Chapter 2, we detected an association between *Lactobacillus* load at the DNA level and a host genomic region on chromosome 1, which included two candidate genes, one of which was the *Il-10* gene. The exact mechanism behind this association is yet unclear, e.g. whether *Lactobacillus* acts as a probiotic, and/or *Il-10* expression leads to changes in *Lactobacillus* bacterial load. Madsen et al. (1999) detected a decrease in *Lactobacillus* species level in the colon of *Il-10* KO mice and were able to ameliorate symptoms of colitis by restoring them. While there have been studies on the effect of *Lactobacillus* within the gut (through oral administration) and upon lung diseases as probiotics (Chiba et al., 2013; Ezendam & van Loveren, 2008), studies have rarely looked into the bacterial composition or *Lactobacillus* load in the lungs in relation to the *Il-10* gene and associated lung diseases. Tomosada et al. (2013) examined mice with nasally administered *Lactobacillus* species and found modulated immunity, such as an overall increased TNF-α and IL-10 concentrations in bronchoalveolar lavage (BAL) samples. In their study, introducing a viral pathogen caused an increase in pro-inflammatory cytokines such as IFN-α, IFN-β, IFN-ɣ, TNF-α, IL6, and IL10. They also showed that an administration of *Lactobacillus* species increased the level of these cytokines even more, with certain species exerting greater effect than others. Such studies show the immunoregulatory capacity of these species, considering them as "immunobiotics."

In these contexts of previous research and our own QTL mapping results, the following questions arose: 1) do *Lactobacillus* abundances change according to the presence/absence of the *Il-10* gene? 2) how does lung microbial composition vary according to *Il-10* genotype? To answer these questions, I proceeded to perform a follow-up study with *Il-10* KO mice available in our mouse facility (MPI for Evolutionary Biology, Plön, Germany).

Aims

- Identify *Lactobacillus* relative abundance according to *Il-10* genotype.
- Identify changes in lung microbial composition according to *Il-10* genotype.

Results

Lung microbiota overview

In my previous lung microbiota study on the AIL G_{15} population, the *Il-10* gene was detected as one of the candidate genes associated with *Lactobacillus* load through QTL mapping analysis (see Results in **Chapter 2**). Following up on this finding, we started with breeding *Il-10* KO mice that were available in our mouse facility. Selected mice of each genotype were dissected and lung tissue samples were processed following our protocol (see Methods in **Chapter 1**). Using reverse-transcribed RNA samples, 16S rRNA gene sequencing was performed, targeting the V1-V2 region of the gene. After sequencing, 11,992 unique sequences were obtained and the "decontam" package was applied using the prevalence method with threshold of 0.1 in order to remove contaminants.

Overall, *Pseudomonas* is the most abundant genus throughout the samples, followed by *Staphylococcus*, *Propionibacterium*, and *Serratia* (Fig. 1). Although not significant ($p > 0.05$), unclassified *GpI* and *Paracoccus* appear at higher abundance in wildtype mice compared to the other two genotypes, whereas Unclassified *GpI* is not found in KO mice. Although not significant, *Streptococcus* also appears more frequently in homozygous (HO) mice. At the phylum level, Proteobacteria is most abundant, followed by Actinobacteria, Firmicutes, and Cyanobacteria/Chloroplast.

Community variation among genotypes

Variation in the microbial communities among genotypes were examined using alpha and beta diversity measures. No significant difference was found among genotypes for richness, diversity, and evenness indices (Fig. 2).

For beta diversity, Constrained Analysis of Principal Coordinates (CAP) with Bray-Curtis dissimilarity (Linnenbrink et al., 2013) was applied for beta diversity analysis, which also revealed no significant differences among genotypes (Fig. 3A). The effects of variables such as sex, cage, and cross (of F0 – whether the "grandmother" is *Il-10* KO or HO) on beta diversity were evaluated by MANOVA. Here, none of the variable showed a significant effect ($p > 0.05$). Additionally, weighted and unweighted UniFrac distances were measured in order to examine the community structure incorporating bacterial phylogenetic information, but similarly, no significant differences were observed according to *Il-10* genotype (Fig. 3B, C, respectively).

Although the major goal of the study was to examine the *Lactobacillus* according to *Il-10* genotype, its abundance was very low, which may be due to different housing location and/or condition from G_{15} population. Nonetheless, the presence as well as the abundance of *Lactobacillus* was examined and showed a random distribution throughout the samples (Table 1). For the purpose of another study, signs such as colon inflammation and inflamed lymph nodes were noted for male mice. These signs were displayed by mouse ID 1949, which is *Il-10*-deficient.

Table 1. List of samples carrying *Lactobacillus* and their matching metadata. Signs of diarrhea, colon inflammation, and hair loss as well as enlarged lymph nodes are noted with + and – symbols, where + indicates the presence of the symptom(s).

Mouse ID	Cross	Sex	Genotype	Cage	Diarrhea	Colon Inflammation	Lymph Node	Hair Loss
498	BL6f	f	-/-	520	-	NA	NA	-
1932	BL6f	f	+/-	1671	-	NA	NA	+
446	BL6f	f	+/-	1639	-	NA	NA	-
447	BL6f	f	+/-	1639	-	NA	NA	-
1557	BL6f	f	+/-	2859	-	NA	NA	-
1931	BL6f	f	+/+	1671	-	NA	NA	-
1841	BL6f	m	-/-	1204	-	-	+	-
1552	BL6f	m	-/-	2855	-	-	-	-
1083	BL6f	m	-/-	192	-	-	-	-
1084	BL6f	m	+/+	192	-	-	-	-
1544	BL6f	m	+/+	65	-	-	-	-
1559	BL6f	m	+/+	2860	-	-	-	-
9979	IL10f	f	-/-	1629	-	NA	NA	-
1043	IL10f	f	+/-	114	-	NA	NA	-
730	IL10f	f	+/+	1556	-	NA	NA	-
1949	IL10f	m	-/-	23	+	+	+	-
713	IL10f	m	-/-	1514	+	-	-	-
2122	IL10f	m	+/-	3058	-	-	-	-
1950	IL10f	m	+/-	23	-	-	-	-
9974	IL10f	m	+/-	1601	-	-	-	-
1948	IL10f	m	+/-	23	-	-	-	-
724	IL10f	m	+/+	667	-	-	-	-

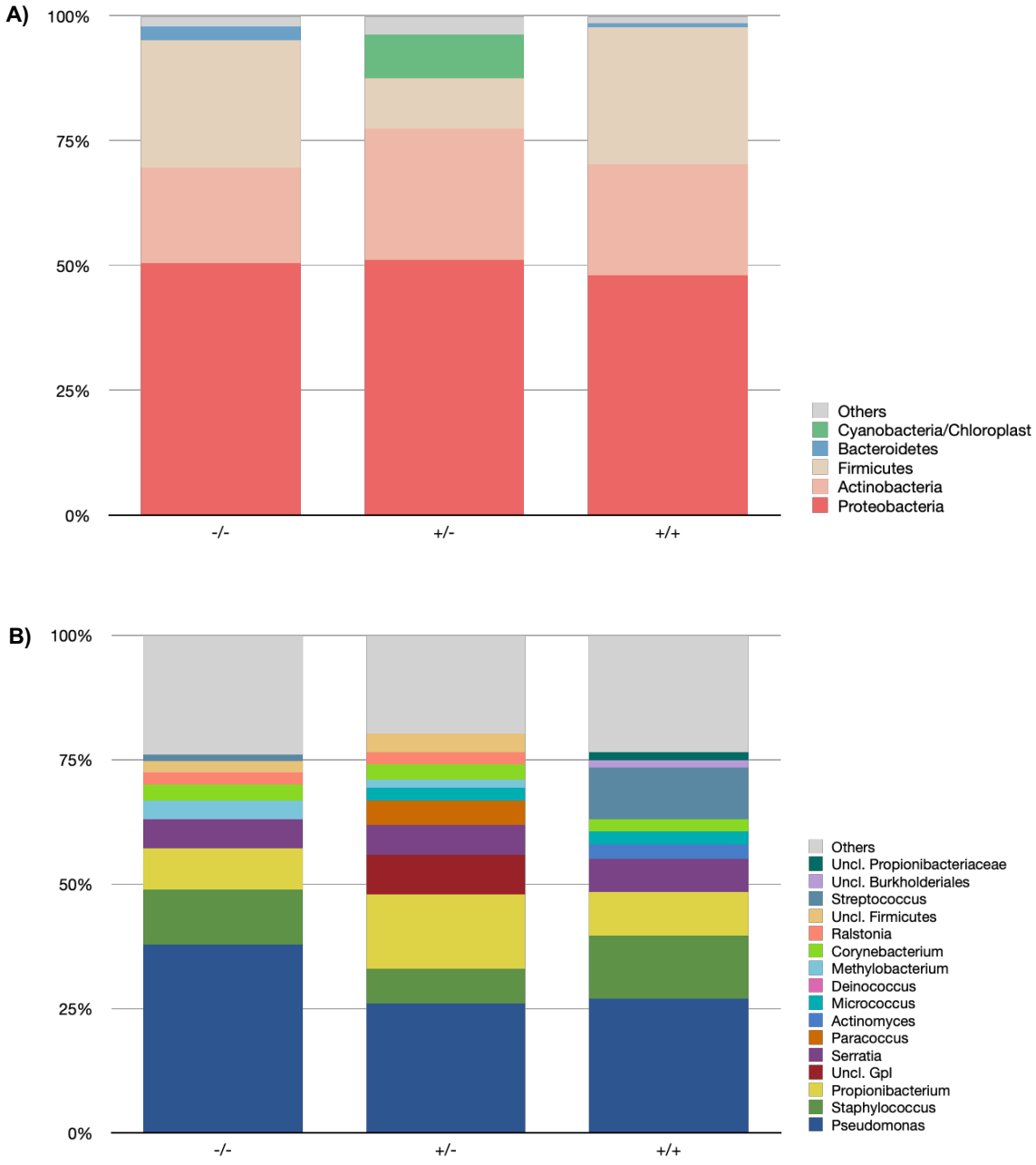

Figure 1. The phylum and genus level comparison among *Il-10* genotypes. A) The four most abundant phyla and B) the ten most abundant genera for each genotype.

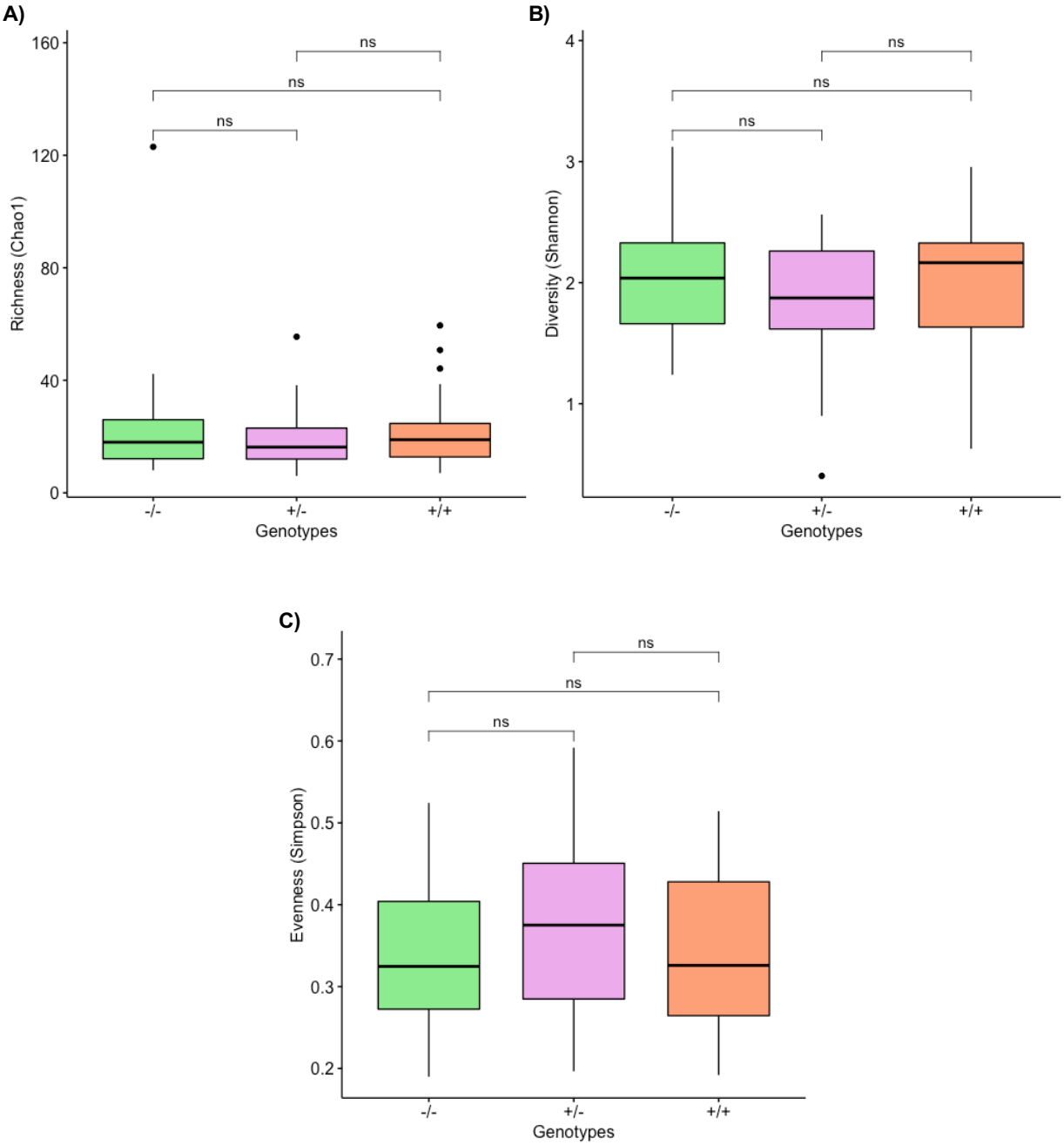

Figure 2. Alpha diversity measures among *Il-10* genotypes. A) Chao1; B) Shannon; C) Simpson.

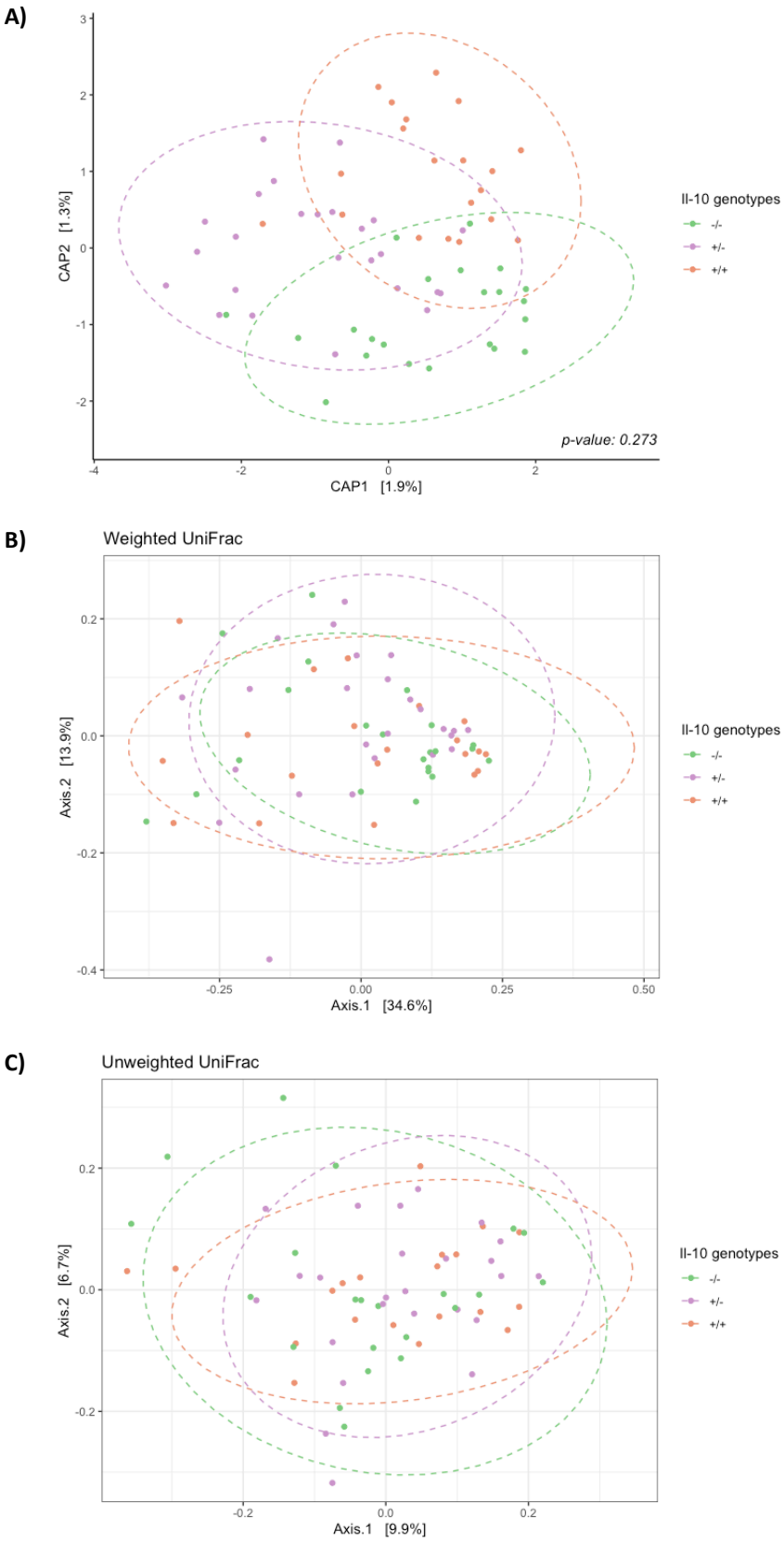

Figure. 3. Bacterial community comparison between *Il-10* genotypes. Beta diversity measures using Constrained Analysis of Principal Coordinates (CAP) with A) Bray-Curtis dissimilarity, B) Weighted UniFrac, and C) Unweighted UniFrac, examining the differences in microbial communities among different *IL-10* genotypes.

Discussion

In the previous chapters, several candidate genes were identified including *Il-10* from QTL mapping of *Lactobacillus* load. In this chapter, a follow-up study was performed using an *Il-10* knockout model. Although the overall results were negative, the *Lactobacillus* abundance is very low, detected only in twenty-two mice. The composition of lung microbiota is also different from that of the G_{15} mouse population, as these mice carried high abundance of *Pseudomonas*, *Propionibacterium*, and *Staphylococcus*, in comparison to a high abundance of *Lactobacillus*, unclassified *Lachnospiraceae*, and *Pelomonas* in G_{15} mice. This may result from different breeding locations and facilities between the two populations, which is indeed one of the many factors that affect microbial composition (Guo et al., 2019). This follow-up project was initiated as *Il-10* KO mice were already available in our facility, which could be used for immediate breeding and further experiments. Such big differences in lung microbial communities between two populations may not have been unexpected, but previous studies have observed such differences according to different housing locations, mouse vendors, and mouse genetic backgrounds (Guo et al., 2019; Mähler & Leiter, 2002). Guo et al. (2019) showed the importance of resident microbiota in cases of lung transplantation using mice from different vendors. As the mice carried varying microbiotas, reactions to transplantation varied, with some exhibiting rejection. In addition to these environmental factors, the genetic backgrounds of the mice, whether BALB/c/129 or C57BL/6, can affect the results (Eisener-Dorman et al., 2009; Seong et al., 2004). Different strains tend to display a wide range of different behaviors, and the BALB/c/129 strain is known to carry genetic contamination, which led to a division of the strain into several lineages (Eisener-Dorman et al., 2009).

The *Il-10* KO mouse model is most commonly used in gut inflammation studies due to its characterization in causing colitis (Büchler et al., 2012; Dann et al., 2014; Kühn et al., 1993; Nagalingam et al., 2013; Sturlan et al., 2001). *Il-10* gene depletion causes colitis in mice under specific-pathogen-free (SPF) condition, but not in a germ-free condition (Sellon et al., 1998). Colitis is affected by different factors such as genetic background and the presence/absence of certain bacteria; thus, the composition of gut microbiota affects the host's susceptibility to diseases. In a study of *Il-10* KO mice with *Helicobacter hepaticus* infection, the severity of disease was dependent on the microbial community structure and the disease was initiated in the presence of indigenous community (Nagalingam et al., 2013). It is also interesting to note

the role of IL-10 in the immune response. When wildtype control and *Il-10* deficient mice were challenged with *Citrobacter rodentium*, those mice deficient in IL-10 performed better in attenuating and controlling infection-associated colitis (Dann et al., 2014). It was an unexpected observation for the authors and is quite intriguing considering the anti-inflammatory role of IL-10. However, this study model with *Il-10* KO mice suggests an enhanced IL-27 production and restricted Th17 response that take over the anti-inflammatory mechanism in absence of IL-10.

The *Il-10* KO mouse model is also used in lung studies, as IL-10 deficiency similarly leads to inflammation in the lungs (Guilbault et al., 2002; Shanley et al., 2000), increases risk of susceptibility to asthma (Kawano et al., 2016; S. Lim et al., 1998) as well as COPD (DeMeo et al., 2008). It is further used to study lung infection with *Pseudomonas aeruginosa* (Guilbault et al., 2002) and *Mycobacterium tuberculosis* (*Mtb*) (Redford et al., 2010) in terms of disease clearance or susceptibility. Guilbault et al. (2002) examined *P. aeruginosa* infection in wildtype and *Il-10* KO mice, revealing higher sensitivity to infection in KO mice, characterized by weight loss, bacterial burden, and cellular inflammatory response. Changes in the lung bacterial load with infection are also relevant. For example, Redford et al. (2010) examined murine lung bacterial load with real-time PCR and detected a reduced bacterial load as well as accelerated response to *Mtb* infection in the absence of IL-10.

Expanding on previous studies, our study aimed to evaluate a potential role of the lung microbiota in *Il-10* deficiency. While the initial results seem negative, several future studies are suggested. During dissection, caecum and colon samples were collected from these mice, which could be used to explore the gut-lung axis. Both the lung and gut carry bacteria that produce the short-chain fatty acids (SCFA) such as Firmicutes, Bacteroidetes, and Proteobacteria. The "crosstalk" between the microbiota of two sites can occur through these SCFAs, either through lymph nodes by antigen-presenting cells or through the bloodstream, as the bacterial products/metabolites are transferred from the intestinal mucosa to the lungs, or vice versa (Invernizzi et al., 2020). Such crosstalk is intriguing, as we would expect the affected microbiota at one site to be followed by changes at another site. Noverr et al. (2005) reported a development of airway allergic disease in mice in the absence of gut microbiota, which was depleted by the use of antibiotics. Another study found an acute lung injury caused by the administration of LPS to the lungs lead to an increase in bacterial load in the caecum and the bloodstream (Sze et al., 2014b). This study demonstrates the potential for crosstalk between

microbiota at different sites, as such acute exposure to LPS disrupts the lung microbiota, which lead to translocation of bacteria into the bloodstream. Similarly, in a study by Schuijt et al. (2016), mice depleted of gut microbiota showed a delay in bacterial infection clearance and reduced pulmonary cytokine concentrations. In the present mouse model, a collective study of both gut and lung microbiota would be more informative and enable us to help better understand the overall mechanism and role played by microbiota in host immunity.

In addition to the further insight that could be gained by examining the gut-lung axis, other drawbacks of the present study could be addressed in the future to assess the potential that the current observations represent a false negative. In the previous chapter, the association between *Lactobacillus* and the host genome was evaluated using ddPCR. ddPCR is able to effectively measure the number of target bacteria especially in samples with low bacterial load. It is likely that lung samples carry low bacterial load, as indicated by difficulties in library preparation and sequencing due to the overwhelming amount of host nucleic acid. Extra steps were required in order to remove host nucleic acid prior to and after sequencing. As the *Il-10* gene was detected from QTL mapping of *Lactobacillus* load, it may be more informative to examine in the same manner in these *Il-10* mice samples. With high-precision ddPCR, effective estimation of *Lactobacillus* load can be performed. Another aspect in the previous study was gene expression results from the *Il-10* and *Mk2* genes. As *Mk2* has been shown to affect the expression *Il-10* (Cao et al., 2003; Marchese et al., 2010; Ronkina et al., 2007; Tudor et al., 2009), examining *Mk2* expression with ddPCR in these mice would provide further information regarding the feedback between the two genes.

Conclusion

Following the detection of the *Il-10* and *Mk2* genes from QTL mapping of *Lactobacillus* load in the previous chapter, I performed an experiment with an *Il-10* KO mouse model. The aims were to examine differences in the lung microbial community and differential abundances of *Lactobacillus* among *Il-10* genotypes. Potentially due to low bacterial load, sequence reads were low compared to the sequencing results from the G_{15} mice, and I was thus not able to obtain optimal results from relative abundance data. However, this chapter reveals preliminary results regarding lung microbiota in *Il-10* KO mouse model, as previous studies mostly focus on bacterial load but not the community composition. As *Il-10* and *Mk2* are involved in immune responses, this *Il-10* KO model would be able to help us better understand the relationship between the two genes and how *Mk2* gene expression levels change in presence or absence of the *Il-10* gene.

Methods

Mice breeding

Il-10 KO and wildtype C57BL/6 mice (Jackson Laboratories, Maine, USA) were first mated at age of 8-10 weeks producing heterogeneous F_1 mice. Here, two crosses were created whereby the F_1 population was generated between *Il-10* KO females and C57BL/6 males, and between C57BL/6 females and *Il-10* KO males, in order to reduce potential "grandmother" (Rausch et al., 2017) or legacy effects. These heterozygous F_1 mice were mated within each cross at age of 10 weeks, which included 11 pairs from the first cross of F_1 mice and 11 pairs from the second cross. F_2 mice were weaned at an age of 3 weeks and were managed so that males or females from the same litter were kept in the same cage. Mice were maintained in individual ventilated cages (IVCs), type II long (Tecniplast®, Greenline) in our specific pathogen-free facility (MPI für Evolutionsbiologie, Plön, Germany) at a 12-h light/dark cycle. Decalcified water and food (1324, fortified, from Altromin) were provided *ad libidum*. Mice were provided with bedding from Rettenmaier, FS14, and nesting material (Baumwollnestlets Plexx®, 14010). All materials were autoclaved prior to introducing into the barrier and to the animals.

Genotyping

Once the F_2 mice reached the age of 3 weeks and weaned, their ears were punched for identification and those ear punches were collected for genotyping. Genotyping was performed using a simple NaOH treatment of the tissue to extract DNA. First, 150uL of 50nM NaOH (Roth) was added to the ear punch tissues and heated at 98°C for 30 minutes. Then the sample tubes were vortexed and returned to heat for 30 more minutes. Next, samples were neutralized with 15uL of 1M Tris at pH 8 and vortexed. Resulting DNA samples were stored at -20°C. Genotyping primers were ordered through Metabion with the following sequences: CCA CAC GCG TCA CCT TAA TA for mutant reverse, CTT GCA CTA CCA AAG CCA CA for common, and GTT ATT GTC TTC CCG GCT GT for wildtype reverse (The Jackson Laboratory). The PCR master mix included 5uL of Multiplex PCR Buffer (QIAGEN), 0.2uL of each primer (10uM), and 2.4uL H2O, and the PCR program followed: 10 minutes at 94°C, 35 cycles of 60 seconds at 94°C, 30 seconds at 67°C, and 60 seconds at 72°C, and extension for 5 minutes at

72°C, and hold at 12°C. PCR products were confirmed on 1% agarose gel that was run at 100V for 45 minutes.

Dissection & DNA/RNA extraction

Ten mice on average (ranging between nine and 15) were selected from each sex of each genotype from each cross for tissue extraction at 4-month or 17-week age. F_2 mice were initially planned for dissection at age of 6-month, which is similar to the age range of G_{15} mice (between 4 and 6 months) (Belheouane et al., 2017) and is when the lung microbiota has developed to a stable stage (Kostric et al., 2017). However, this had to be shortened to 4 months, or approximately 17 weeks, as some of the knockout mice started to suffer with symptoms of gut inflammation such as diarrhea. A total of 84 mice were chosen with even sex and cage distribution as much as possible in order to reduce external effects. For tissue extraction, mice were killed by CO_2 exposure and all tools were cleaned and sterilized with filtered water, 70% ethanol, RNALater, and heating beads at 150°C, in this respective order. Separate sets of tools were used for opening or cutting of skin, rib cage, and lungs in order to avoid contamination as much as possible. The lower ½ of the inferior lobe of the right lung was extracted and kept in RNALater (Invitrogen) at 4°C overnight. The solution was removed the next day and kept at -20°C. Left lungs were extracted and kept in PFA at 4°C for histology in the future. DNA/RNA extraction was performed using 96-well DNA/RNA following the previous protocol (methods section in **Chapter 1**) and only RNA samples were used in this study for analysis. RNA concentrations were measured using a Nanodrop 1000 (ThermoFisher Scientific) with Qubit RNA HS assay kits (Invitrogen). RNA samples were diluted to the same concentration prior to cDNA synthesis for normalization throughout all samples. cDNA synthesis was performed following previous protocol (methods section in **Chapter 1**).

Library preparation & sequencing

Amplicon libraries for sequencing were prepared using cDNA templates according to the methods described in **Chapter 1**. Due to high amounts of nucleic acid from the host, the final pool was run again on 1.5% agarose gel at 50V for 7 hours in order to separate host from bacterial nucleic acid. Gel extraction was performed again and retrieved DNA was used for 16S rRNA gene amplicon sequencing with Illumina MiSeq (2 x 250 bp read length).

Sequence processing

As samples included a high abundance of host sequences, these sequences were removed using KneadData (v.0.6.1), which performs trimming using Trimmomatic (v.0.36). Then reads were merged with VSEARCH (v.2.7.1), chimeras were identified and removed with VSEARCH using the SILVA Gold reference database (Quast et al., 2012). OTU clustering was performed at a 97% similarity threshold with unique sequences detected through dereplication. Afterwards, taxonomy was assigned based on RDP Multi-Classifier (v.16) from phylum to genus level at a confidence level of 0.8. Generated OTU and taxonomy tables were transferred to R studio (v.1.2.1335) and were merged with metadata including sample ID, parents, sex, age, and cage information, using "phyloseq" (v.1.28.0) (McMurdie & Holmes, 2013) in order to create a phyloseq object. The prevalence method of "decontam" (v.1.4.0) (Davis et al., 2018) was performed in order to define contaminants that appear more often in negative extraction controls compared to real samples. OTUs defined as contaminants were removed the dataset and all samples were normalized using the cumulative sum scaling (CSS) normalization method (Joseph et al., 2013) due to low sequence reads. Then alpha diversity measures were calculated with Chao1, Shannon, and Simpson for richness, diversity, and evenness, respectively, between age groups and among genotypes within each age group. Beta diversity was analyzed using Constrained Analysis of Principal Coordinates (CAP) with Bray-Curtis dissimilarity as well as weighted and unweighted UniFrac. Summary statistics were obtained in order to measure variability within each trait. All statistical analyses were performed on R (3.6.2).

Chapter IV.

Analysis of the host blood group-related candidate gene *B4galnt2*

Introduction

The blood group-related gene *B4galnt2* encodes a glycosyltransferase *ß-1,4-N-acetylgalactosaminyltransferase 2* that is involved in the synthesis of the Sd(a)/Cad carbohydrate blood group antigen. *B4galnt2* is commonly expressed in the gastrointestinal tracts of vertebrates, but displays variable tissue-specific expression patterns in wild mouse populations (Johnsen et al., 2009; Linnenbrink et al., 2011). Its expression in the intestinal mucosa has been reported to play a role in determining the composition of microbiota within the gut (Staubach et al., 2012).

Aside from its role in the gut, when the expression of the *B4galnt2* gene is switched from gut to blood vessel by a common allele, termed "*Modifier of von Willebrand Factor-1*" (*Mvwf1*), the coagulation factor von Willebrand factor (VWF) is more quickly cleared from the bloodstream (Mohlke et al., 1999). This results in prolonged bleeding times. This variation appears to have been maintained in the mouse lineage for several million years despite such detrimental effect and fitness cost, possibly due to a protective role in host-pathogen interactions (Linnenbrink et al., 2011). Rausch et al. (2015) determined that the host becomes more resistant to *Salmonella* infection when *B4galnt2* expression is absent in the intestinal epithelium. When infected with *Salmonella*, mice expressing *B4galnt2* in the intestinal epithelium experienced inflammation in the epithelium layer and a reduction in cecum weight. Those that did not express *B4galnt2* in the intestinal epithelium displayed less severe infection and cecal inflammation. These results were found to stem from changes in microbial composition according to the expression of *B4galnt2*-glycans, highlighting the gene's role in host-microbiota interactions.

Both intestines and lungs are characterized by glycosylated mucus secretion. Although not as prominent as its prominent mucin types, such as MUCIN 5AC (MUC5AC) and MUC5B, lung was also found to express MUC2 (Buisine et al., 1999; Dohrman et al., 1994), which is glycosylated by B4GALNT2 (Wei et al., 2012) and is the major mucin in the gut (Velcich et al., 2002). These mucins consist of VWF-like D domains (Sadler, 1998), and they help protect the cell layers in the respiratory and intestinal tracts. The mucin glycans play an important role in protecting mucosal barriers of the respiratory and intestinal tracts (Javitt et al., 2020) as well as host-pathogen interactions (Lamblin et al., 1991; Schulz et al., 2007). Since these mucins are critical for the maintenance of healthy epithelial layers in these tissues, their deficiency or

malfunction leads to inflammation and diseases. Previous studies have shown their overexpression in the lungs to be linked to asthma (McNamara et al., 2004; Morcillo & Cortijo, 2006) and decreased mucins linked to cystic fibrosis (Davies et al., 1999; Henke et al., 2004). B4GALNT2 has also been detected in the lung macrophages as well as lung cancer tissues (Uhlén et al., 2015), which leads to the hypothesis of a potential contribution of this candidate gene in the lung and the lung microbiota.

Here, using a *B4galnt2* mouse model either with- or without the wild type allele [intestinal endothelium expression-, and possibly other tissues such as lung (Uhlén et al., 2015)], we aimed to examine the role of this candidate gene in possibly mediating microbial differences in the lung.

Aims

- Evaluate microbial communities according to *B4galnt2* genotype in the lung.

Results

Examining overall lung microbiota

A total of 18 mice, consisting of eight females and ten males, were selected for this study, for which DNA and RNA were extracted from lung tissues. For each sex, half were wildtype and the other half were KO for the wild type allele of the *B4galnt2* gene. After sequencing, 196,236 unique sequences were obtained (normalized to 10,000 sequences per sample). Overall microbial composition at the genus and OTU levels was examined. At both the DNA and RNA level, *Lactobacillus* was the most abundant genus followed by unclassified Firmicutes and unclassified Bacteroidetes. Also, for both levels, OTU1_*Lactobacillus* was the most abundant OTU, followed by OTU9_Uncl. Firmicutes and OTU26_Uncl. *Enterobacteriaceae* (Fig. 1).

Alpha diversity measures indicated higher diversity and richness at the RNA level compared to DNA (Chao1: $p < 0.0001$, Shannon: $p < 0.001$) (Fig. 2). However, there were no significant differences in diversity and richness between *B4galnt2* genotypes at either level (Chao1: $p = 0.4363$ at DNA level, $p = 0.2581$ at RNA level; Shannon: $p = 0.6078$ at DNA level, $p = 0.9314$ at RNA level) (Fig. 2).

To analyze differences between individuals (beta diversity), Constrained Analysis of Principal Coordinates (CAP) applied to the Bray-Curtis index was used, which indicated significant differences between DNA and RNA level communities ($p < 0.0001$) (Fig. 3). However, similar to alpha diversity, there were no significant differences in beta diversity between *B4galnt2* genotypes ($p = 0.5856$ at DNA level, $p = 0.7340$ at RNA level).

Figure 1. Comparison of lung microbiota composition between RNA and DNA levels. Bacterial communities are compared between RNA and DNA level communities at A) genus and B) OTU levels. The 10 most abundant taxa are shown. Both RNA and DNA communities have high abundance of genus *Lactobacillus* and OTU1_*Lactobacillus*.

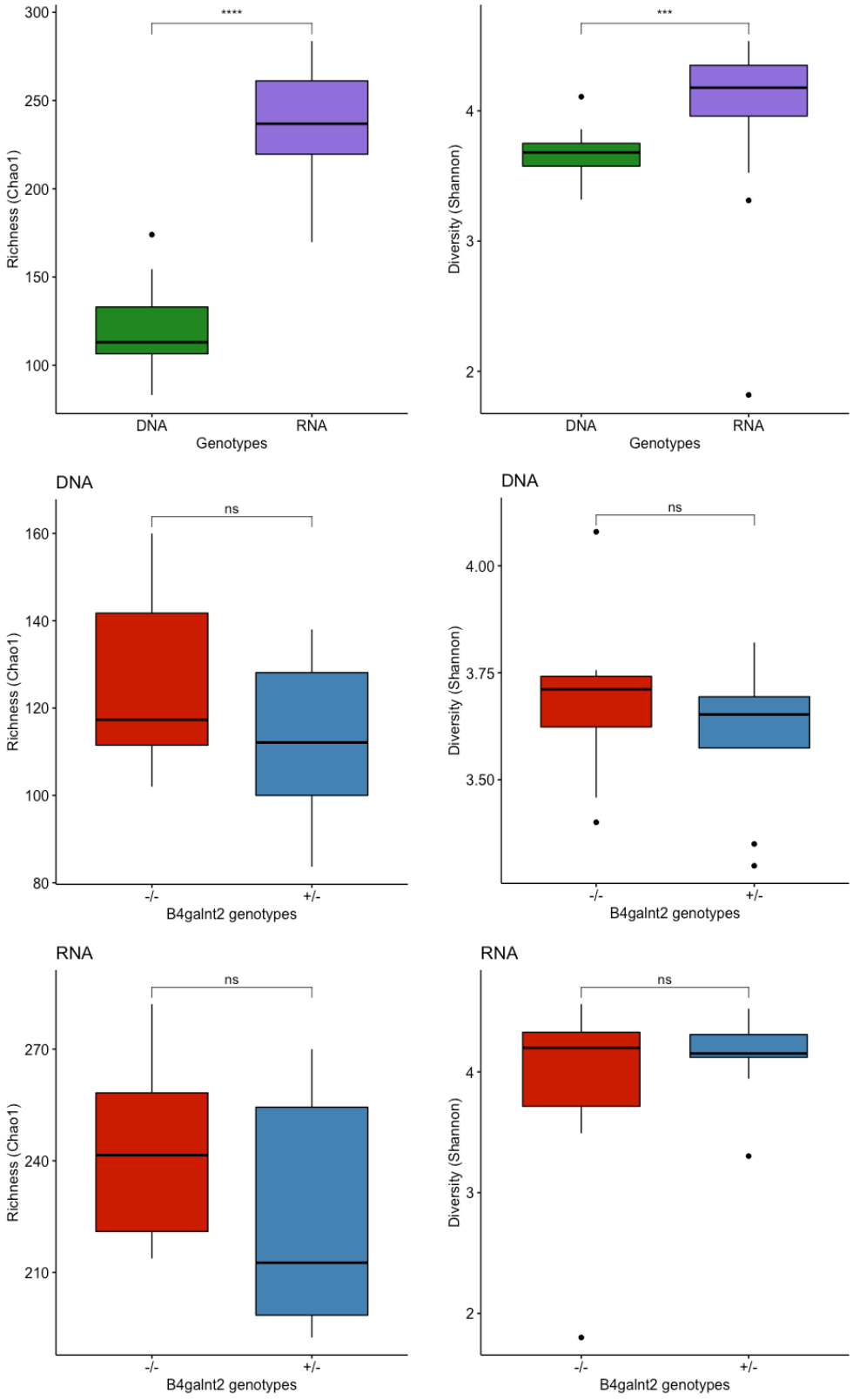

Figure 2. Alpha diversity measures to compare bacterial communities and identify differences. Alpha diversity measures are shown A) between DNA and RNA levels, and between genotypes at B) DNA and C) RNA levels. Richness was observed using Chao1 indices and diversity was observed using Shannon indices.

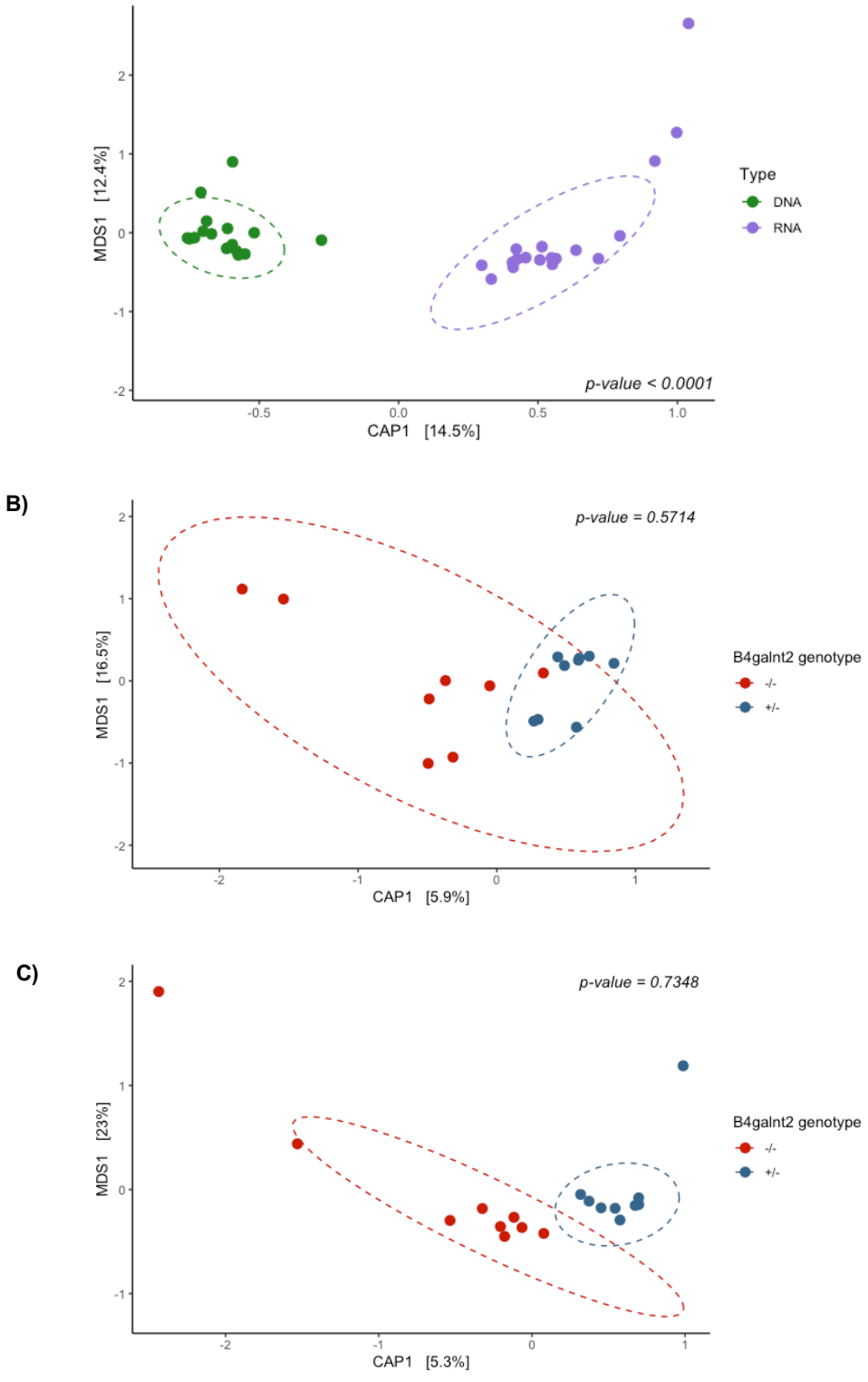

Figure 3. Bacterial community comparison between *B4galnt2* genotypes. Beta diversity measurement using Constrained Analysis of Principal Coordinates (CAP) with Bray-Curtis indices was applied A) between DNA and RNA levels, and between genotypes at B) DNA level and at C) RNA level.

Indicator species

Despite the lack of differences at the overall community level, it remains possible that individual taxa display differences according to *B4galnt2* genotype. To test this hypothesis, indicator species analysis was performed using the R package, "indic.species". Several potential indicator species (Table 1) and genera (Table 2) were detected. Although the adjusted p-values do not indicate significance, we find a nominal level of significance for numerous taxa, which may represent candidates worthy of further examination, e.g. regarding the potential for false negative associations.

Table 1. Indicator species analysis for comparison between *B4galnt2* genotypes. Stat: association statistics; Uncl.: unclassified.

Group	Indicator OTU	Stat	p-value	Adjusted p-value
DNA	OTU1242_*Lactobacillus*	0.626	0.009	0.928
	OTU17_*Lactobacillus*	0.535	0.028	0.928
RNA	OTU165_Uncl. Gammaproteobacteria	0.584	0.029	1.000
	OTU61_Uncl. Firmicutes	0.395	0.030	1.000
	OTU256_*Parcubacteria genera incertae sedis*	0.520	0.032	1.000
	OTU207_*Methylobacterium*	0.469	0.038	1.000
	OTU23_*Lactobacillus*	0.476	0.048	1.000
	OTU209_Uncl. Clostridiales	0.477	0.049	1.000

Table 2. Indicator genera analysis for comparison between *B4galnt2* genotypes. Stat: association statistics; uncl.: unclassified.

Group	Indicator genus	Stat	p-value	Adjusted p-value
DNA	*Brevibacterium*	0.789	0.032	0.625
	Lactobacillus	0.707	0.036	0.625
RNA	*Listeria*	0.828	0.011	1.000
	Uncl. *Streptophyta*	0.834	0.030	1.000
	Psychrobacter	0.790	0.037	1.000

Discussion

Maintenance of differential expression of the *B4galnt2* gene in wild mouse populations is paradoxical given the host's reduced ability of blood coagulation and thus greater fitness cost when expressed in the blood vessels instead of gut epithelium. However, the gene expression may have been shaped by selection on host-pathogen interactions, given *B4galnt2's* demonstrated role in mediating susceptibility to experimental infection by *Salmonella* (Rausch et al., 2015). Here, gut microbial community composition revealed an association to *B4galnt2* expression, which in turn affects the host's susceptibility to infection (Rausch et al., 2015). Although the lung microbiota has not been studied in this mouse model, we aimed to examine the role of this candidate gene in possibly mediating microbial differences in the lung, which is characterized by glycosylated mucus secretions similar to the gut.

Among the first obvious patterns in my data are that distinct differences are present between communities at the DNA and RNA levels. Differences at the DNA and RNA level may suggest a difference in taxa that are active but present at low abundance, or vice versa. DNA samples allow an examination of all bacterial material present, while this may also include dormant or dead bacterial cells. On the other hand, RNA samples provide information on bacterial communities at the active transcript level, which are highly likely to be interacting with the host. However, this could lead to a bias towards active bacteria only at the time of sampling and dissection. As shown in previous studies (Belheouane et al., 2017), such differences should be understood and noted for microbial studies.

Second, *Lactobacillus* was the most abundant genus throughout the samples for both genotypes, which is a known resident bacterial genus in the lungs (Barfod et al., 2013; Singh et al., 2017; Yun et al., 2014) and also abundant in the lung microbiota of G_{15} population (**Chapter 1**). As such, it is interesting to identify a *Lactobacillus* OTU as an indicator species in these mice. Through QTL mapping on the G_{15} mouse population (in Chapter 1, 2), we identified *Lactobacillus* as a candidate genus. As a lung resident and also a candidate genus that may be interacting with the host, *Lactobacillus* would be a promising candidate for further lung microbiota studies. Although indicator species analysis did not find species of significance, several were nominally significant and thus could be further examined, e.g. using ddPCR.

In this study, we conducted a first evaluation of a blood group-related gene in possibly influencing the lung microbiota. Similar to previous chapters, *Lactobacillus* is again observed

as an abundant taxon that could possibly be influenced by differences in glycosylated mucins in the lung. Given that we also collected colon and caecum tissue samples from these mice, future work may also examine both microbiotas in order to evaluate a possible role of *B4galnt2* in the gut-lung axis.

Methods

Mice breeding & dissection

B4galnt2 wildtype and KO mice were maintained in our specific pathogen-free facility (MPI für Evolutionsbiologie, Plön, Germany) at 12-h light/dark cycle. Decalcified water and food (1324, fortified, from Altromin) were provided *ad libidum*. Mice were provided with bedding from Rettenmaier, FS14, and nesting material (Baumwollnestlets Plexx®, 14010). All materials were autoclaved prior to introducing into the barrier and to the animals.

A total of 18 mice were selected, consisting of four mice of each genotype for females and five for males for dissection and further analysis. Dissection was performed at age between 21 to 37 weeks, with an average of 32 weeks. Several tissues were collected including small intestine, colon tissue, caecum tissue and content, and left and right lungs. Samples were preserved overnight in either RNALater (Invitrogen) for DNA/RNA extraction or 4% PFA for histology. Afterwards, RNALater was removed and those samples were preserved at -20°C. Samples in PFA were transferred to PBS and stored at 4°C.

Processing samples for sequencing

DNA and RNA were extracted from 18 left lung tissue samples using AllPrep DNA/RNA Mini Kit (QIAGEN). DNase I treatment was performed on RNA extraction columns using DNase I solution (STEMCELL solution), where 28U (14µL of 10x solution and 66µL H_2O adding up to 80µL) was added to each sample. DNA and RNA were eluted with 50µL EB twice to obtain final volume of 100µL each. Blank well was included in the extraction step as a negative extraction control. Concentration of RNA was measured on NanoDrop 1000 (ThermoFisher Scientific) and samples were diluted to 200ng/µL. Then reverse transcription was performed on these samples using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) by adding 10µL of the template. A library was prepared for both DNA and cDNA samples targeting the V1-V2 region of the 16S rRNA gene with a primer pair of 27F-338R, with unique eight-based MIDs (multiplex identifier). PCR amplifications were conducted in a 25µL volume with the following recipe: 5µL of 5x HF buffer (ThermoFisher Scientific), 5µL of 10mM dNTPs, 4µM of forward and reverse primers (2µL each), 0.25µl of Fusion Taq II (Finnzym), 3µL of template, and 8.25µL of nuclease-free water. Cycling conditions include: initial denaturation at 98°C for 30 seconds, 30 cycles of 98°C for 9 seconds,

55°C for 30 seconds, and 72°C for 90 seconds, and final extension at 72°C for 10 minutes. Negative extraction controls and mock communities were included in the library. Sequencing was performed using the Illumina MiSeq Reagent Kit v2 with 2 x 250bp read length.

Statistical analysis

Sequences in fastq format were checked for quality using fastQC (v.0.11.7) and trimmed using Trimmomatic (v.0.36). Reads were then merged with VSEARCH (v.2.7.1) and chimeras were identified and removed with VSEARCH using SILVA Gold reference database (Quast et al., 2012). OTU clustering was performed at a 97% similarity threshold with unique sequences detected through dereplication. Afterwards, taxonomy was assigned based on RDP Multi-Classifier (v.16) from phylum to genus level at a confidence level of 0.8.

Generated OTU and taxonomy tables were transferred to R studio (v.1.2.1335) and merged with meta data including sample ID, age, sex, cage information, and genotype, using "phyloseq" (1.28.0) (McMurdie & Holmes, 2013) in order to create a phyloseq object. Dataset was then rarefied to an even sampling depth of 10,000 (reads).

Alpha and beta diversity measures were compared between DNA and RNA levels and between genotypes at the DNA and RNA levels using the "microbiome" package (1.6.0). Then candidate species were identified using the function MULTIPATT with "indicspecies" (1.7.6) (Cáceres & Legendre, 2009) on R. The statistical significance was tested with 10,000 permutations and the p-values were adjusted using the "BH" correction method (Benjamini & Hochberg, 1995). All statistical analyses were performed on R 3.6.1.

General conclusion and perspectives

Although the lungs of healthy individuals have historically been considered sterile, recent studies indicate the presence of a resident microbial community (Beck et al., 2012; Charlson et al., 2012; Charlson et al., 2011; Kostric et al., 2017; Segal et al., 2013). Similar to other body sites, a healthy lung microbiota generally helps maintain homeostasis, while disease or infection disturbs this homeostasis. The disruption leads to alterations including changes in microbial load and shifts in microbial community composition through a blooming of certain bacterial taxa or through an immigration of non-resident bacteria. The mechanisms are not yet clear, but studies have suggested different factors affecting the lung microbiota. Environmental factors such as smoke, dust, and cigarette smoking are a few examples, in addition to host genetic variation. Host genetic factors are involved in various diseases known to also involve barrier organ-associated microbes, such as inflammatory bowel disease (IBD) and chronic obstructive pulmonary diseases (COPD). The role of host genetics in influencing microbes is complex, but previous studies have shown associations between regions of the host genome and individual microbial abundances (Benson et al., 2010; Campbell et al., 2012; Liu et al., 2016). Snijders et al. (2016) examined the influence of host genetics on the mouse gut microbiome using QTL mapping approach where they detected Lactobacillales abundance associated to arthritis, rheumatic disease, and diabetes. Similarly, Belheouane et al. (2017) showed host-microbiota association in the skin microbiota.

Through my thesis projects, I aimed to explore the impact of host genetics on the lung microbiota using mouse models and a combination of methods including ddPCR and QTL mapping. By examining the results, candidate genes that influence microbes and are also known to be associated with disease can be pinpointed to serve as potential targets of novel intervention, e.g. involving microbes or their products.

Lung microbiota of AIL population

Using the 15th generation of an AIL mouse population, narrow intervals (ranging from 0.14 to 3.6 Mb) were obtained, whereby candidate genes can be pinpointed. First, the lung microbiota of G_{15} mice population was examined and protocols were established for lung tissue samples for any onward studies. In addition to the relative abundances data based on sequencing, absolute loads were also assessed through ddPCR, which helped us identify

contaminants, and QTL mapping based on these absolute load data revealed significant QTL hits that were not detected at the level of relative abundance. ddPCR was especially advantageous for lung tissue samples that carry low bacterial load. Using this technique, effective analysis was performed using genus-specific primers, as we detected significant associations, for example, between *Lactobacillus* load at the DNA level and a SNP on mouse chromosome 1. This microbial trait displayed a significant association with two genes, *Il-10* and *Mk2*, which are both involved in immune responses. IL-10 is a well-known anti-inflammatory cytokine and MK2 is a pro-inflammatory kinase which is a downstream substrate of the $p38^{MAPK}$ pathway and is crucial in controlling the expression and stability of immune cytokine mRNAs. Although the exact mechanism of this association is unknown, I speculate that changes in the gene expression or regulation of one or both of these genes affects the survival of the bacteria through mucus or epithelium layer immunity. Alternatively, metabolites produced by the bacteria affect the immune system and thus the gene expression levels.

In an effort to test the above hypotheses, an *Il-10* KO mouse model was adapted, presented in the third chapter. Although significant differences were not observed at the level of relative abundance, I plan to approach the lung samples with ddPCR in order to estimate the absolute load of *Lactobacillus*. The expression of *Mk2* will also be assessed and compared among different genotypes, which could unravel the feedback (or association) between the two genes, *Il-10* and *Mk2*, within the lungs. We also plan to examine the gut-lung axis using corresponding caecum or colon samples of these mice as well as the G_{15} mouse population in order to identify a potential "cross-talk" between microbial communities of the two tissue sites. This cross-talk may take place via an exchange of short-chain fatty acids (SCFA) between the tissues or even by translocating bacteria through oropharynx reflux or also through blood and lymph (Mathieu et al., 2018). Although the function of bacteria and metabolites in the lungs is only partially known, a link between the gut and lung modulated by bacteria is evident. For example, oral administration of probiotic species improved lung function, as the symptoms from infection were ameliorated (Chiba et al., 2013; Racedo et al., 2006; Villena et al., 2012). Studies also showed that disturbances in the gut microbiota from an intake of antibiotics affects lung diseases and inflammation (Metsälä et al., 2015; Russell et al., 2015; Russell et al., 2013). Examination of the gut microbiota may in part help us understand the influence of *Lactobacillus* on host health and the candidate genes. We also plan to carry on to further inspect

the role of our candidate genus and genes, for example, by examining the *Mk2* gene in KO mice and examining the lung microbiome with- and without nasal administration of *Lactobacillus*.

Acknowledgement

First of all, I would like to thank my supervisor, Prof. Dr. John F. Baines, for the support and guidance throughout my PhD stages. Thank you for believing in me and providing me with the greatest opportunities and experiences in the field of microbiome and evolutionary medicine. I had spent the most valuable years studying under your supervision and thus have grown one step further. I would also like to thank Dr. Meriem Belheouane for your help and support. You pulled me through and helped me expand my perspectives. As a team, I am very grateful to be a part of Leibniz ScienceCampus "Evolutionary Medicine of the Lung (EvoLung)" with my PhD project. It was a very eye-opening experience with amazing researchers and their ideas as we shared constructive feedback and discussions at our meetings.

I would like to thank all members of our research group. Thank you, Dr. Marie Vallier, for great ideas and discussions whenever I had questions regarding my data, experimental design, etc. Thanks to Hanna Fokt and Shauni Doms as we worked as a great team in endless sessions of mouse dissection. And thank you, Aleksa Cèpić, for sharing thoughtful ideas and helping me whenever I had questions especially regarding sequencing and programming. And thank you, Dr. Nadia Andrea Andreani, for your support and insightful comments, while I was writing up this thesis. I would like to give special thanks to Dr. Sven Künzel for his help on sequencing and mouse breeding, and to the mouse team at Max-Planck-Institute for Evolutionary Biology in Plön, especially Anastasia Vock and Anika Jonas.

I am also grateful to be a member of International Max Planck Research Schools (IMPRS) for my PhD. This well-organized program put many students from all over the world together and allowed us to be more active scientists and share ideas. Special thanks to my colleagues, especially Dr. Anuradha Mukerjee, Dr. Devika Bhave, Johana Fajardo, and Gillian Dureux, for supports, discussions, and friendly memories we shared throughout our journey.

Thank you, all my relatives who supported me throughout my study and who always cheer me whenever I visit them in Korea. I have learned that I have an extraordinarily amazing family beside me. Lastly, I would like express my deepest, earnest love and appreciation to my parents who gave me unconditional love and support throughout my life. I cannot thank you enough for providing me the opportunity to study and pursue my dreams. I am the luckiest

person to have you as parents. My dad, who himself was very devoted to science, would have been very proud to see this.

Curriculum Vitae

PERSONAL INFORMATION

Name and surname / Title	Cecilia Juryung Chung, M.S.
Date of birth	November 15, 1988
Nationality	Dual citizenship in USA and South Korea
Current residence	Plön, Germany.
Phone number	+49 (0)176 30133651 (mobile), +49 (0)4522 763619 (office)
E-mail	cjchung@evolbio.mpg.de

CURRENT POSITION

May 2017 to Present	PhD researcher/student at Max-Planck-Institute for Evolutionary Biology, Plön, Germany.
Main research topics	The role of host genetics in shaping diversity of the murine lung microbiota of various mouse models.

EDUCATION

Jun 2015	Completed the Master's degree in Biology at Seoul National University (Seoul, South Korea) with a thesis titled: "Effects of Sodium Fluoroacetate and *Bd* on Survival, Metamorphosis, and Behavior of *Bufo gargarizans* tadpoles". Supervisor: Dr. Bruce Waldman
Jun 2011	Completed the Bachelor's degree in Animal Science at University of California, Davis (Davis, USA).

RESEARCH EXPERIENCE

Jan 2017 to Present	PhD Research at Max-Planck-Institute for Evolutionary Biology (Plön, Germany) *Main research topics:* The role of host genetics on the community variation in the murine lung microbiota of Advanced Intercross Line population, candidate gene analysis in association with candidate genus in terms of chronic lung diseases, lung microbiome in various mouse models.
Dec 2015 to Sep 2016	Researcher at Molecular Ecology group at Yeungnam University (Daegu, South Korea). *Main research topics:* Conservation genetics of *Cottus hangiongensis*; construction of DNA barcoding for Decapods in Korea.
Jan 2013 to Jun 2015	Master's degree at Behavioral Ecology group at Seoul National University

(Seoul, South Korea).

Main research topic: Effect of sodium fluoroacetate and *Bd* on survival and behavior of tadpoles of *Bufo gargarizans*. Supervisor: Dr. Bruce Waldman

RESEARCH IMPACT

JOURNAL PAPERS

1. Belheouane, M., Vallier, M., Čepić, A., **Chung, C. J.**, Ibrahim, S., & Baines, J. F. (2020). Assessing similarities and disparities in the skin microbiota between wild and laboratory populations of house mice. *The ISME Journal*. doi: 10.1038/s41396-020-0690-7

2. Ham, S. J., Lee, S. Y., Song, S., **Chung, J.R.**, Choi, S., Chung, J. (2015). Interaction between RING1 (R1) and Ubiquitin-Like (UBL) Domain Is Critical for the Regulation of Parkin Activity. *The Journal of Biological Chemistry*, Advance Online Publication. doi: 10.1074/jbc.M115.687319

3. Kim, H., Yang, J., Kim, M.J., Choi, S., **Chung, J.R.**, Kim, J.M., Yoo, Y.H., Chung, J., Koh, H. (2015). Tumor necrosis factor receptor-associated protein 1 (TRAP1) mutation and TRAP1 inhibitor gamitrinib-triphenylphosphonium (G-TPP) induce a forkhead box O (FOXO)-dependent cell protective signal from mitochondria. *The Journal of Biological Chemistry*. Advance Online Publication. doi: 10.1074/jbc.M115.656934

ABSTRACTS

1. Joni Lund, Sabine Bartel, **Juryung Chung**, Zane Orinska, Skadi Kull, Uta Jappe, John Baines, Susanne Krauss-Etschmann (2019) House dust mite (HDM)-induced "asthma" phenotypes differ in 4 mouse strains – link to the microbiome?, 17th ERS Lung Science Conference.

2. **Juryung Chung**, John Baines (2017) "Unravelling the role of host genetics in shaping variation in the lung microbiota", KEC symposium on Pathogen Evolution 2017, Kiel, Germany.

CONFERENCES & MEETINGS

Poster: NDI3 conference, September 2019, Borstel, Germany.
Talk: IMPRS 9th retreat, September 2019, Översee, Germany
Talk & Poster: EvoLung Meeting, June 2019, Borstel, Germany.
Poster: Aquavit, 2019, Plön, Germany.
Poster: IMPRS 8th retreat, 2018, Breklum, Germany.
Talk & Poster: EvoLung Meeting, September 2018, Kettendorf, Germany.
Poster: IMPRS 7th retreat, 2017, Leck, Germany.
Oral: Aquavit, 2017, Plön, Germany.

WORKSHOPS

Scientific Writing Workshop by Inflammation at Interfaces, Cluster of Excellence in Inflammation Research, Timmendorfer Strand, Germany.

16S Metabarcoding Workshop, Berlin, Germany

ORGANIZING EXPERIENCE

Organizing IMPRS 8th retreat, 2018 (Plön & Breklum, Germany)

Organizing Aquavit 2017 (Plön, Germany)

Bibliography

Alberdi, A., Aizpurua, O., Bohmann, K., Zepeda-Mendoza, M. L., & Gilbert, M. T. P. (2016). Do vertebrate gut metagenomes confer rapid ecological adaptation? *Trends in ecology & evolution, 31*(9), 689-699.

Alibhai, F. J., Reitz, C. J., Peppler, W. T., Basu, P., Sheppard, P., Choleris, E., . . . Martino, T. A. (2018). Female ClockΔ19/Δ19 mice are protected from the development of age-dependent cardiomyopathy. *Cardiovascular research, 114*(2), 259-271.

Alvarez, S., Herrero, C., Bru, E., & Perdigon, G. (2001). Effect of Lactobacillus casei and yogurt administration on prevention of Pseudomonas aeruginosa infection in young mice. *Journal of food protection, 64*(11), 1768-1774.

Andrews, B. S., Eisenberg, R. A., Theofilopoulos, A., Izui, S., Wilson, C. B., McConahey, P., . . . Dixon, F. (1978). Spontaneous murine lupus-like syndromes. Clinical and immunopathological manifestations in several strains. *The Journal of experimental medicine, 148*(5), 1198-1215.

Aragón, F., Carino, S., Perdigón, G., & de LeBlanc, A. d. M. (2014). The administration of milk fermented by the probiotic Lactobacillus casei CRL 431 exerts an immunomodulatory effect against a breast tumour in a mouse model. *Immunobiology, 219*(6), 457-464.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., . . . Batto, J.-M. (2011). Enterotypes of the human gut microbiome. *Nature, 473*(7346), 174-180.

Baltrušis, P., Halvarsson, P., & Höglund, J. (2019). Molecular detection of two major gastrointestinal parasite genera in cattle using a novel droplet digital PCR approach. *Parasitology research, 118*(10), 2901-2907.

Barfod, K. K., Roggenbuck, M., Hansen, L. H., Schjørring, S., Larsen, S. T., Sørensen, S. J., & Krogfelt, K. A. (2013). The murine lung microbiome in relation to the intestinal and vaginal bacterial communities. *BMC microbiology, 13*(1), 303.

Barrios, J., Kho, A. T., Aven, L., Mitchel, J. A., Park, J.-A., Randell, S. H., . . . Ai, X. (2019). Pulmonary neuroendocrine cells secrete γ-aminobutyric acid to induce goblet cell hyperplasia in primate models. *American journal of respiratory cell and molecular biology, 60*(6), 687-694.

Bassis, C. M., Erb-Downward, J. R., Dickson, R. P., Freeman, C. M., Schmidt, T. M., Young, V. B., . . . Huffnagle, G. B. (2015). Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *MBio, 6*(2), e00037-00015.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., . . . Scheipl, F. (2012). Package 'lme4'. *CRAN. R Foundation for Statistical Computing, Vienna, Austria*.

Beck, J. M., Young, V. B., & Huffnagle, G. B. (2012). The microbiome of the lung. *Translational Research, 160*(4), 258-266.

Belheouane, M., Gupta, Y., Künzel, S., Ibrahim, S., & Baines, J. F. (2017). Improved detection of gene-microbe interactions in the mouse skin microbiota using high-resolution QTL mapping of 16S rRNA transcripts. *Microbiome, 5*(1), 59.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., . . . Hua, K. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences, 107*(44), 18933-18938.

Bisgaard, H., Hermansen, M. N., Buchvald, F., Loland, L., Halkjaer, L. B., Bønnelykke, K., . . . Thorsen, S. V. (2007). Childhood asthma after bacterial colonization of the airway in neonates. *New England Journal of Medicine, 357*(15), 1487-1495.

Blankenhorn, E. P., Troutman, S., Clark, L. D., Zhang, X.-M., Chen, P., & Heber-Katz, E. (2003). Sexually dimorphic genes regulate healing and regeneration in MRL mice. *Mammalian genome, 14*(4), 250-260.

Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q., Bukowski, R., Bell, J. T., . . . Gevers, D. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome biology, 16*(1), 191.

Bode, J. G., Ehlting, C., & Häussinger, D. (2012). The macrophage response towards LPS and its control through the p38MAPK–STAT3 axis. *Cellular signalling, 24*(6), 1185-1194.

Bolnick, D. I., Snowberg, L. K., Hirsch, P. E., Lauber, C. L., Org, E., Parks, B., . . . Svanbäck, R. (2014). Individual diet has sex-dependent effects on vertebrate gut microbiota. *Nature communications, 5*, 4500.

Borewicz, K., Pragman, A. A., Kim, H. B., Hertz, M., Wendt, C., & Isaacson, R. E. (2013). Longitudinal analysis of the lung microbiome in lung transplantation. *FEMS microbiology letters, 339*(1), 57-65.

Boutin, S., Graeber, S. Y., Stahl, M., Dittrich, A. S., Mall, M. A., & Dalpke, A. H. (2017). Chronic but not intermittent infection with Pseudomonas aeruginosa is associated with global changes of the lung microbiome in cystic fibrosis. *European Respiratory Journal, 50*(4).

Browman, K. E., & Crabbe, J. C. (2000). Quantitative trait loci affecting ethanol sensitivity in BXD recombinant inbred mice. *Alcoholism: Clinical and Experimental Research, 24*(1), 17-23.

Büchler, G., Wos-Oxley, M. L., Smoczek, A., Zschemisch, N.-H., Neumann, D., Pieper, D. H., . . . Bleich, A. (2012). Strain-specific colitis susceptibility in IL10-deficient mice depends on complex gut microbiota–host interactions. *Inflammatory bowel diseases, 18*(5), 943-954.

Buisine, M.-P., Devisme, L., Copin, M.-C., Durand-Réville, M., Gosselin, B., Aubert, J.-P., & Porchet, N. (1999). Developmental mucin gene expression in the human respiratory tract. *American journal of respiratory cell and molecular biology, 20*(2), 209-218.

Cabello, H., Torres, A., Celis, R., El-Ebiary, M., De La Bellacasa, J. P., Xaubet, A., . . . Soler, N. (1997). Bacterial colonization of distal airways in healthy subjects and chronic lung disease: a bronchoscopic study. *European Respiratory Journal, 10*(5), 1137-1144.

Cáceres, M. D., & Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology, 90*(12), 3566-3574.

Cameron, S. J., Lewis, K. E., Huws, S. A., Hegarty, M. J., Lewis, P. D., Pachebat, J. A., & Mur, L. A. (2017). A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. *PloS one, 12*(5), e0177062.

Campbell, J. H., Foster, C. M., Vishnivetskaya, T., Campbell, A. G., Yang, Z. K., Wymore, A., . . . Podar, M. (2012). Host genetic and environmental effects on mouse intestinal microbiota. *The ISME journal, 6*(11), 2033-2044.

Cao, H., Dzineku, F., & Blackshear, P. J. (2003). Expression and purification of recombinant tristetraprolin that can bind to tumor necrosis factor-α mRNA and serve as a substrate for mitogen-activated protein kinases. *Archives of biochemistry and biophysics, 412*(1), 106-120.

Charlson, E. S., Bittinger, K., Chen, J., Diamond, J. M., Li, H., Collman, R. G., & Bushman, F. D. (2012). Assessing bacterial populations in the lung by replicate analysis of samples from the upper and lower respiratory tracts. *PloS one, 7*(9), e42786.

Charlson, E. S., Bittinger, K., Haas, A. R., Fitzgerald, A. S., Frank, I., Yadav, A., . . . Collman, R. G. (2011). Topographical continuity of bacterial populations in the healthy human respiratory tract. *American journal of respiratory and critical care medicine, 184*(8), 957-963.

Cheng, C., Wang, Z., Wang, J., Ding, C., Sun, C., Liu, P., . . . Gu, B. (2020). Characterization of the lung microbiome and exploration of potential bacterial biomarkers for lung cancer. *Translational Lung Cancer Research, 9*(3), 693.

Cheng, R., Abney, M., Palmer, A. A., & Skol, A. D. (2011). QTLRel: an R package for genome-wide association studies in which relatedness is a concern. *BMC genetics, 12*(1), 1-3.

Chevalier, C., Stojanović, O., Colin, D. J., Suarez-Zamorano, N., Tarallo, V., Veyrat-Durebex, C., . . . Hagemann, S. (2015). Gut microbiota orchestrates energy homeostasis during cold. *Cell, 163*(6), 1360-1374.

Chiba, E., Tomosada, Y., Vizoso-Pinto, M. G., Salva, S., Takahashi, T., Tsukida, K., . . . Villena, J. (2013). Immunobiotic Lactobacillus rhamnosus improves resistance of infant mice against respiratory syncytial virus infection. *International immunopharmacology, 17*(2), 373-382.

Coburn, B., Wang, P. W., Caballero, J. D., Clark, S. T., Brahma, V., Donaldson, S., . . . Tullis, D. E. (2015). Lung microbiota across age and disease stage in cystic fibrosis. *Scientific reports, 5*, 10241.

Crabbe, J. C., Belknap, J. K., Mitchell, S. R., & Crawshaw, L. I. (1994). Quantitative trait loci mapping of genes that influence the sensitivity and tolerance to ethanol-induced hypothermia in BXD recombinant inbred mice. *Journal of Pharmacology and Experimental Therapeutics, 269*(1), 184-192.

Dann, S. M., Le, C., Choudhury, B. K., Liu, H., Saldarriaga, O., Hanson, E. M., . . . Eckmann, L. (2014). Attenuation of intestinal inflammation in interleukin-10-deficient mice infected with Citrobacter rodentium. *Infection and immunity, 82*(5), 1949-1958.

Davies, J. R., Svitacheva, N., Lannefors, L., Kornfält, R., & Carlstedt, I. (1999). Identification of MUC5B, MUC5AC and small amounts of MUC2 mucins in cystic fibrosis airway secretions. *Biochemical Journal, 344*(2), 321-330.

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome, 6*(1), 226.

Delroisse, J.-M., Boulvin, A.-L., Parmentier, I., Dauphin, R. D., Vandenbol, M., & Portetelle, D. (2008). Quantification of Bifidobacterium spp. and Lactobacillus spp. in rat fecal samples by real-time PCR. *Microbiological research, 163*(6), 663-670.

DeMeo, D. L., Campbell, E. J., Barker, A. F., Brantly, M. L., Eden, E., McElvaney, N. G., . . . Stoller, J. K. (2008). IL10 polymorphisms are associated with airflow obstruction in severe α1-antitrypsin deficiency. *American journal of respiratory cell and molecular biology, 38*(1), 114-120.

Dickson, R. P., Erb-Downward, J. R., Freeman, C. M., McCloskey, L., Falkowski, N. R., Huffnagle, G. B., & Curtis, J. L. (2017). Bacterial topography of the healthy human lower respiratory tract. *MBio, 8*(1).

Dickson, R. P., Erb-Downward, J. R., & Huffnagle, G. B. (2013). The role of the bacterial microbiome in lung disease. *Expert review of respiratory medicine, 7*(3), 245-257.

Dohrman, A., Tsuda, T., Escudier, E., Cardone, M., Jany, B., Gum, J., . . . Basbaum, C. (1994). Distribution of lysozyme and mucin (MUC2 and MUC3) mRNA in human bronchus. *Experimental lung research, 20*(4), 367-380.

Doms, S., Hermes, B.-M., & Baines, J. F. (2018). Evolutionary Perspectives on the Human Gut Microbiome. In *The Gut Microbiome in Health and Disease* (pp. 67-78): Springer.

Drengenes, C., Wiker, H. G., Kalananthan, T., Nordeide, E., Eagan, T. M., & Nielsen, R. (2019). Laboratory contamination in airway microbiome studies. *BMC microbiology, 19*(1), 1-13.

Dumas, A., Corral, D., Colom, A., Levillain, F., Peixoto, A., Hudrisier, D., . . . Neyrolles, O. (2018). The host microbiota contributes to early protection against lung colonization by Mycobacterium tuberculosis. *Frontiers in immunology, 9*, 2656.

Duong, A. T., Reitz, C. J., Louth, E. L., Creighton, S. D., Rasouli, M., Zwaiman, A., . . . Bailey, C. D. (2019). The Clock Mechanism Influences Neurobiology and Adaptations to Heart Failure in ClockΔ 19/Δ 19 Mice With Implications for Circadian Medicine. *Scientific reports, 9*(1), 1-18.

Ege, M. J., Mayer, M., Normand, A.-C., Genuneit, J., Cookson, W. O., Braun-Fahrländer, C., . . . von Mutius, E. (2011). Exposure to environmental microorganisms and childhood asthma. *New England Journal of Medicine, 364*(8), 701-709.

Ehlting, C., Trilling, M., Tiedje, C., Le-Trilling, V. T. K., Albrecht, U., Kluge, S., . . . Hengel, H. (2016). MAPKAP kinase 2 regulates IL-10 expression and prevents formation of intrahepatic myeloid cell aggregates during cytomegalovirus infections. *Journal of hepatology, 64*(2), 380-389.

Eisener-Dorman, A. F., Lawrence, D. A., & Bolivar, V. J. (2009). Cautionary insights on knockout mouse studies: the gene or not the gene? *Brain, behavior, and immunity, 23*(3), 318-324.

Erb-Downward, J. R., Thompson, D. L., Han, M. K., Freeman, C. M., McCloskey, L., Schmidt, L. A., . . . Sundaram, B. (2011). Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PloS one, 6*(2), e16384.

Ezendam, J., & van Loveren, H. (2008). Lactobacillus casei Shirota administered during lactation increases the duration of autoimmunity in rats and enhances lung inflammation in mice. *British Journal of Nutrition, 99*(1), 83-90.

Falade, A., & Ayede, A. (2011). Epidemiology, aetiology and management of childhood acute community-acquired pneumonia in developing countries--a review. *African journal of medicine and medical sciences, 40*(4), 293.

Feigelman, R., Kahlert, C. R., Baty, F., Rassouli, F., Kleiner, R. L., Kohler, P., . . . von Mering, C. (2017). Sputum DNA sequencing in cystic fibrosis: non-invasive access to the lung microbiome and to pathogen details. *Microbiome, 5*(1), 20.

File Jr, T. M. (2004). Streptococcus pneumoniae and community-acquired pneumonia: a cause for concern. *The American Journal of Medicine Supplements, 117*(3), 39-50.

Fletcher, E. K., Morgan, J., Kennaway, D. R., Bienvenu, L. A., Rickard, A. J., Delbridge, L. M., . . . Young, M. J. (2017). Deoxycorticosterone/salt-mediated cardiac inflammation and fibrosis are dependent on functional CLOCK signaling in male mice. *Endocrinology, 158*(9), 2906-2917.

Forsythe, P., Inman, M. D., & Bienenstock, J. (2007). Oral treatment with live Lactobacillus reuteri inhibits the allergic airway response in mice. *American journal of respiratory and critical care medicine, 175*(6), 561-569.

Fraune, S., Anton-Erxleben, F., Augustin, R., Franzenburg, S., Knop, M., Schröder, K., . . . Bosch, T. C. (2015). Bacteria–bacteria interactions within the microbiota of the ancestral metazoan Hydra contribute to fungal resistance. *The ISME journal, 9*(7), 1543-1556.

Fujimura, K. E., Johnson, C. C., Ownby, D. R., Cox, M. J., Brodie, E. L., Havstad, S. L., . . . Wegienka, G. (2010). Man's best friend? The effect of pet ownership on house dust microbial communities. *The Journal of allergy and clinical immunology, 126*(2), 410.

Gatti, D. M., Svenson, K. L., Shabalin, A., Wu, L.-Y., Valdar, W., Simecek, P., . . . Palmer, A. (2014). Quantitative trait locus mapping methods for diversity outbred mice. *G3: Genes, Genomes, Genetics, 4*(9), 1623-1633.

Gebel, S., Gerstmayer, B., Kuhl, P., Borlak, J. r., Meurrens, K., & Müller, T. (2006). The kinetics of transcriptomic changes induced by cigarette smoke in rat lungs reveals a specific program of defense, inflammation, and circadian clock gene expression. *Toxicological Sciences, 93*(2), 422-431.

Gerbault, P., Liebert, A., Itan, Y., Powell, A., Currat, M., Burger, J., . . . Thomas, M. G. (2011). Evolution of lactase persistence: an example of human niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1566), 863-877.

Gobbato, N., Rachid, M., & Perdigón, G. (2008). Anti-inflammatory effect of yoghurt in an experimental inflammatory bowel disease in mouse. *The Journal of dairy research, 75*(4), 497.

Gobert, G., Cotillard, A., Fourmestraux, C., Pruvost, L., Miguet, J., & Boyer, M. (2018). Droplet digital PCR improves absolute quantification of viable lactic acid bacteria in faecal samples. *Journal of microbiological methods, 148*, 64-73.

Gollwitzer, E. S., Saglani, S., Trompette, A., Yadava, K., Sherburn, R., McCoy, K. D., . . . Marsland, B. J. (2014). Lung microbiota promotes tolerance to allergens in neonates via PD-L1. *Nature medicine, 20*(6), 642-647.

Gomila, M., Bowien, B., Falsen, E., Moore, E. R., & Lalucat, J. (2007). Description of Pelomonas aquatica sp. nov. and Pelomonas puraquae sp. nov., isolated from

industrial and haemodialysis water. *International journal of systematic and evolutionary microbiology, 57*(11), 2629-2635.

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., . . . Ley, R. E. (2016). Genetic determinants of the gut microbiome in UK twins. *Cell host & microbe, 19*(5), 731-743.

Greathouse, K. L., White, J. R., Vargas, A. J., Bliskovsky, V. V., Beck, J. A., von Muhlinen, N., . . . Robles, A. I. (2018). Interaction between the microbiome and TP53 in human lung cancer. *Genome biology, 19*(1), 123.

Groussin, M., Mazel, F., Sanders, J. G., Smillie, C. S., Lavergne, S., Thuiller, W., & Alm, E. J. (2017). Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nature communications, 8*(1), 1-12.

Gubareva, E., Maidin, M., Tyndyk, M., Vinogradova, L., & Panchenko, A. (2016). Role of clock proteins in skin carcinogenesis in 14-month-old SHR mice maintained under disrupted light regimen. *Voprosy Onkologii, 62*(5), 666-670.

Guilbault, C., Stotland, P., Lachance, C., Tam, M., Keller, A., Thompson-Snipes, L., . . . Stevenson, M. M. (2002). Influence of gender and interleukin-10 deficiency on the inflammatory response during lung infection with Pseudomonas aeruginosa in mice. *Immunology, 107*(3), 297-305.

Guo, Y., Wang, Q., Li, D., Onyema, O. O., Mei, Z., Manafi, A., . . . Barker, T. H. (2019). Vendor-specific microbiome controls both acute and chronic murine lung allograft rejection by altering CD 4+ Foxp3+ regulatory T cell levels. *American Journal of Transplantation, 19*(10), 2705-2718.

Hadden, H., Soldin, S. J., & Massaro, D. (2012). Circadian disruption alters mouse lung clock gene expression and lung mechanics. *Journal of applied physiology, 113*(3), 385-392.

Han, M. K., Zhou, Y., Murray, S., Tayob, N., Noth, I., Lama, V. N., . . . Huffnagle, G. B. (2014). Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *The Lancet Respiratory Medicine, 2*(7), 548-556.

Hashikawa, K.-i., Katamune, C., Kusunose, N., Matsunaga, N., Koyanagi, S., & Ohdo, S. (2017). Dysfunction of the circadian transcriptional factor CLOCK in mice resists chemical carcinogen-induced tumorigenesis. *Scientific reports, 7*(1), 1-11.

Heirali, A. A., Workentine, M. L., Acosta, N., Poonja, A., Storey, D. G., Somayaji, R., . . . Parkins, M. D. (2017). The effects of inhaled aztreonam on the cystic fibrosis lung microbiome. *Microbiome, 5*(1), 51.

Henke, M. O., Renner, A., Huber, R. M., Seeds, M. C., & Rubin, B. K. (2004). MUC5AC and MUC5B mucins are decreased in cystic fibrosis airway secretions. *American journal of respiratory cell and molecular biology, 31*(1), 86-91.

Hilty, M., Burke, C., Pedro, H., Cardenas, P., Bush, A., Bossley, C., . . . Pachter, L. (2010). Disordered microbial communities in asthmatic airways. *PloS one, 5*(1), e8578.

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., . . . Legler, T. C. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical chemistry, 83*(22), 8604-8610.

Hosgood III, H. D., Sapkota, A. R., Rothman, N., Rohan, T., Hu, W., Xu, J., . . . Wu, G. (2014). The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. *Environmental and molecular mutagenesis, 55*(8), 643-651.

Hu, Y., Cheng, M., Liu, B., Dong, J., Sun, L., Yang, J., . . . Jin, Q. (2020). Metagenomic analysis of the lung microbiome in pulmonary tuberculosis-a pilot study. *Emerging microbes & infections, 9*(1), 1444-1452.

Huang, Y. J., & Lynch, S. V. (2011). The emerging relationship between the airway microbiota and chronic respiratory disease: clinical implications. *Expert review of respiratory medicine, 5*(6), 809-821.

Huang, Y. J., Sethi, S., Murphy, T., Nariya, S., Boushey, H. A., & Lynch, S. V. (2014). Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. *Journal of clinical microbiology, 52*(8), 2813-2823.

Igartua, C., Davenport, E. R., Gilad, Y., Nicolae, D. L., Pinto, J., & Ober, C. (2017). Host genetic variation in mucosal immunity pathways influences the upper airway microbiome. *Microbiome, 5*(1), 16.

Invernizzi, R., Lloyd, C. M., & Molyneaux, P. L. (2020). Respiratory microbiome and epithelial interactions shape immunity in the lungs. *Immunology, 160*(2), 171-182.

Ishii, J., Sato, H., Yazawa, T., Shishido-Hara, Y., Hiramatsu, C., Nakatani, Y., & Kamma, H. (2014). Class III/IV POU transcription factors expressed in small cell lung cancer cells are involved in proneural/neuroendocrine differentiation. *Pathology international, 64*(9), 415-422.

Iwai, S., Huang, D., Fong, S., Jarlsberg, L. G., Worodria, W., Yoo, S., . . . Segal, M. (2014). The lung microbiome of Ugandan HIV-infected pneumonia patients is compositionally and functionally distinct from that of San Franciscan patients. *PloS one, 9*(4), e95726.

Javitt, G., Khmelnitsky, L., Albert, L., Bigman, L. S., Elad, N., Morgenstern, D., . . . Fass, D. (2020). Assembly Mechanism of Mucin and von Willebrand Factor Polymers. *Cell, 183*(3), 717-729. e716.

Jervis-Bardy, J., Leong, L. E., Marri, S., Smith, R. J., Choo, J. M., Smith-Vaughan, H. C., . . . Rogers, G. B. (2015). Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. *Microbiome, 3*(1), 19.

Ji, W., Ferdman, D., Copel, J., Scheinost, D., Shabanova, V., Brueckner, M., . . . Ment, L. R. (2020). De novo damaging variants associated with congenital heart diseases contribute to the connectome. *Scientific reports, 10*(1), 1-11.

Johnsen, J. M., Teschke, M., Pavlidis, P., McGee, B. M., Tautz, D., Ginsburg, D., & Baines, J. F. (2009). Selection on cis-regulatory variation at B4galnt2 and its influence on von Willebrand factor in house mice. *Molecular biology and evolution, 26*(3), 567-578.

Joseph, N., Paulson, C., Corrada Bravo, H., & Pop, M. (2013). Robust methods for differential abundance analysis in marker gene surveys. *Nat. Methods, 10*, 1200-1202.

Kanno, H., Nose, M., Itoh, J., Taniguchi, Y., & Kyogoku, M. (1992). Spontaneous development of pancreatitis in the MRL/Mp strain of mice in autoimmune mechanism. *Clinical & Experimental Immunology, 89*(1), 68-73.

Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W. T., Wolfe, A. J., . . . McWeeney, S. (2019). Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *MSystems, 4*(4).

Kawane, K., Tanaka, H., Kitahara, Y., Shimaoka, S., & Nagata, S. (2010). Cytokine-dependent but acquired immunity-independent arthritis caused by DNA escaped from degradation. *Proceedings of the National Academy of Sciences, 107*(45), 19432-19437.

Kawano, H., Kayama, H., Nakama, T., Hashimoto, T., Umemoto, E., & Takeda, K. (2016). IL-10-producing lung interstitial macrophages prevent neutrophilic asthma. *International immunology, 28*(10), 489-501.

Kench, J. A., Russell, D. M., Fadok, V. A., Young, S. K., Worthen, G. S., Jones-Carson, J., . . . Nemazee, D. (1999). Aberrant wound healing and TGF-β production in the autoimmune-prone MRL/+ mouse. *Clinical immunology, 92*(3), 300-310.

Kil, J.-H., Jung, K.-O., Lee, H.-S., Hwang, I.-K., Kim, Y.-J., & Park, K.-Y. (2004). Effects of kimchi on stomach and colon health of Helicobacter pylori-infected volunteers. *JOURNAL OF FOOD SCIENCE AND NUTRITION-NEW SERIES-, 9*(2), 161-166.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., . . . Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature methods, 8*(9), 761.

Kondratov, R. V., Kondratova, A. A., Gorbacheva, V. Y., Vykhovanets, O. V., & Antoch, M. P. (2006). Early aging and age-related pathologies in mice deficient in BMAL1, the core componentof the circadian clock. *Genes & development, 20*(14), 1868-1873.

Koskinen, K., Pausan, M. R., Perras, A. K., Beck, M., Bang, C., Mora, M., . . . Moissl-Eichinger, C. (2017). First insights into the diverse human archaeome: specific detection of archaea in the gastrointestinal tract, lung, and nose and on skin. *MBio, 8*(6).

Kostric, M., Milger, K., Krauss-Etschmann, S., Engel, M., Vestergaard, G., Schloter, M., & Schöler, A. (2017). Development of a Stable Lung Microbiome in Healthy Neonatal Mice. *Microbial Ecology*, 1-14.

Kotlyarov, A., Neininger, A., Schubert, C., Eckert, R., Birchmeier, C., Volk, H.-D., & Gaestel, M. (1999). MAPKAP kinase 2 is essential for LPS-induced TNF-α biosynthesis. *Nature cell biology, 1*(2), 94-97.

Kühn, R., Löhler, J., Rennick, D., Rajewsky, K., & Müller, W. (1993). Interleukin-10-deficient mice develop chronic enterocolitis. *Cell, 75*(2), 263-274.

Lagishetty, V., Parthasarathy, P. T., Phillips, O., Fukumoto, J., Cho, Y., Fukumoto, I., . . . Lockey, R. F. (2014). Dysregulation of CLOCK gene expression in hyperoxia-induced lung injury. *American Journal of Physiology-Cell Physiology, 306*(11), C999-C1007.

Lamblin, G., Lhermitte, M., Klein, A., Houdret, N., Scharfman, A., Ramphal, R., & Roussel, P. (1991). The carbohydrate diversity of human respiratory mucins: a protection of the underlying mucosa. *Am Rev Respir Dis, 144*(3 Pt 2), S19-S24.

Larsson, E., Tremaroli, V., Lee, Y. S., Koren, O., Nookaew, I., Fricker, A., . . . Bäckhed, F. (2012). Analysis of gut microbial regulation of host gene expression along the length

of the gut and regulation of gut microbial ecology through MyD88. *Gut, 61*(8), 1124-1131.

Laurenzi, G. A., Potter, R. T., & Kass, E. H. (1961). Bacteriologic flora of the lower respiratory tract. *New England Journal of Medicine, 265*(26), 1273-1278.

Lazarevic, V., Gaïa, N., Girard, M., & Schrenzel, J. (2016). Decontamination of 16S rRNA gene amplicon sequence datasets based on bacterial load assessment by qPCR. *BMC microbiology, 16*(1), 73.

Le, H. P. T., Nakamura, Y., Oh-Oka, K., Ishimaru, K., Nakajima, S., & Nakao, A. (2017). The frequency of Th17 cells in the small intestine exhibits a day–night variation dependent on circadian clock activity. *Biochemical and biophysical research communications, 490*(2), 290-295.

Li, N., He, F., Liao, B., Zhou, Y., Li, B., & Ran, P. (2017). Exposure to ambient particulate matter alters the microbial composition and induces immune changes in rat lung. *Respiratory research, 18*(1), 1-10.

Liang, X., Bushman, F. D., & FitzGerald, G. A. (2015). Rhythmicity of the intestinal microbiota is regulated by gender and the host circadian clock. *Proceedings of the National Academy of Sciences, 112*(33), 10479-10484.

Lim, M. Y., Yoon, H. S., Rho, M., Sung, J., Song, Y.-M., Lee, K., & Ko, G. (2016). Analysis of the association between host genetics, smoking, and sputum microbiota in healthy humans. *Scientific reports, 6*(1), 1-9.

Lim, S., Crawley, E., Woo, P., & Barnes, P. J. (1998). Haplotype associated with low interleukin-10 production in patients with severe asthma. *The Lancet, 352*(9122), 113.

Linnenbrink, M., Johnsen, J. M., Montero, I., Brzezinski, C. R., Harr, B., & Baines, J. F. (2011). Long-term balancing selection at the blood group-related gene B4galnt2 in the genus Mus (Rodentia; Muridae). *Molecular biology and evolution, 28*(11), 2999-3003.

Linnenbrink, M., Wang, J., Hardouin, E. A., Künzel, S., Metzler, D., & Baines, J. F. (2013). The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Molecular ecology, 22*(7), 1904-1916.

LiPuma, J. J. (2010). The changing microbial epidemiology in cystic fibrosis. *Clinical microbiology reviews, 23*(2), 299-323.

Liu, S., da Cunha, A. P., Rezende, R. M., Cialic, R., Wei, Z., Bry, L., . . . Weiner, H. L. (2016). The host shapes the gut microbiota via fecal microRNA. *Cell host & microbe, 19*(1), 32-43.

Lomolino, M. V. (2000). A call for a new paradigm of island biogeography. *Global Ecology and Biogeography, 9*(1), 1-6.

Lukov, G. L., Hu, T., McLaughlin, J. N., Hamm, H. E., & Willardson, B. M. (2005). Phosducin-like protein acts as a molecular chaperone for G protein βγ dimer assembly. *The EMBO journal, 24*(11), 1965-1975.

Lusk, R. W. (2014). Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PloS one, 9*(10), e110808.

Lyon, H., Lange, C., Lake, S., Silverman, E. K., Randolph, A. G., Kwiatkowski, D., . . . Laird, N. (2004). IL10 gene polymorphisms are associated with asthma phenotypes in children. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society, 26*(2), 155-165.

Madsen, K. L., Doyle, J. S., Jewell, L. D., Tavernini, M. M., & Fedorak, R. N. (1999). Lactobacillus species prevents colitis in interleukin 10 gene–deficient mice. *Gastroenterology, 116*(5), 1107-1114.

Maheshwari, Y., Selvaraj, V., Hajeri, S., & Yokomi, R. (2017). Application of droplet digital PCR for quantitative detection of Spiroplasma citri in comparison with real time PCR. *PloS one, 12*(9).

Mähler, M., & Leiter, E. H. (2002). Genetic and environmental context determines the course of colitis developing in IL–10-deficient mice. *Inflammatory bowel diseases, 8*(5), 347-355.

Marchese, F. P., Aubareda, A., Tudor, C., Saklatvala, J., Clark, A. R., & Dean, J. L. (2010). MAPKAP kinase 2 blocks tristetraprolin-directed mRNA decay by inhibiting CAF1 deadenylase recruitment. *Journal of Biological Chemistry, 285*(36), 27590-27600.

Marra, F., Marra, C. A., Richardson, K., Lynd, L. D., Kozyrskyj, A., Patrick, D. M., . . . FitzGerald, J. M. (2009). Antibiotic use in children is associated with increased risk of asthma. *Pediatrics, 123*(3), 1003-1010.

Marsland, B. J., & Salami, O. (2015). Microbiome influences on allergy in mice and humans. *Current opinion in immunology, 36*, 94-100.

Maruyama, J., Naguro, I., Takeda, K., & Ichijo, H. (2009). Stress-activated MAP kinase cascades in cellular senescence. *Current medicinal chemistry, 16*(10), 1229-1235.

Mathieu, E., Escribano-Vazquez, U., Descamps, D., Cherbuy, C., Langella, P., Riffault, S., . . . Thomas, M. (2018). Paradigms of lung microbiota functions in health and disease, particularly, in asthma. *Frontiers in physiology, 9*, 1168.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one, 8*(4), e61217.

McNamara, N., Gallup, M., Khong, A., Sucher, A., Maltseva, I., Fahy, J. V., & Basbaum, C. (2004). Adenosine up-regulation of the mucin gene, MU2, in asthma. *The FASEB journal, 18*(14), 1770-1772.

Mendes, V., Galvao, I., & Vieira, A. T. (2019). Mechanisms by which the gut microbiota influences cytokine production and modulates host inflammatory responses. *Journal of interferon & cytokine research, 39*(7), 393-409.

Metsälä, J., Lundqvist, A., Virta, L., Kaila, M., Gissler, M., & Virtanen, S. (2015). Prenatal and post-natal exposure to antibiotics and risk of asthma in childhood. *Clinical & Experimental Allergy, 45*(1), 137-145.

Moeller, A. H., & Sanders, J. G. (2020). Roles of the gut microbiota in the adaptive evolution of mammalian species. *Philosophical Transactions of the Royal Society B, 375*(1808), 20190597.

Mohlke, K. L., Purkayastha, A. A., Westrick, R. J., Smith, P. L., Petryniak, B., Lowe, J. B., & Ginsburg, D. (1999). Mvwf, a dominant modifier of murine von Willebrand factor, results from altered lineage-specific expression of a glycosyltransferase. *Cell, 96*(1), 111-120.

Moore, K. W., O'garra, A., Malefyt, R. d., Vieira, P., & Mosmann, T. R. (1993). Interleukin-10. *Annual review of immunology, 11*(1), 165-190.

Moran, N. A., Ochman, H., & Hammer, T. J. (2019). Evolutionary and ecological consequences of gut microbial communities. *Annual Review of Ecology, Evolution, and Systematics, 50*, 451-475.

Morcillo, E. J., & Cortijo, J. (2006). Mucus and MUC in asthma. *Current opinion in pulmonary medicine, 12*(1), 1-6.

Morrow, J. D., Cho, M. H., Platig, J., Zhou, X., DeMeo, D. L., Qiu, W., . . . Bueno, R. (2018). Ensemble genomic analysis in human lung tissue identifies novel genes for chronic obstructive pulmonary disease. *Human genomics, 12*(1), 1.

Murillo-Rincon, A. P., Klimovich, A., Pemöller, E., Taubenheim, J., Mortzfeld, B., Augustin, R., & Bosch, T. C. (2017). Spontaneous body contractions are modulated by the microbiome of Hydra. *Scientific reports, 7*(1), 1-9.

Nagalingam, N. A., Robinson, C. J., Bergin, I. L., Eaton, K. A., Huffnagle, G. B., & Young, V. B. (2013). The effects of intestinal microbial community structure on disease manifestation in IL-10-/-mice infected with Helicobacter hepaticus. *Microbiome, 1*(1), 15.

Nguyen, M. H., Ueda, K., Nakamura, Y., & Daigo, Y. (2012). Identification of a novel oncogene, MMS22L, involved in lung and esophageal carcinogenesis. *International journal of oncology, 41*(4), 1285-1296.

Nguyen, T. L. A., Vieira-Silva, S., Liston, A., & Raes, J. (2015). How informative is the mouse for human gut microbiota research? *Disease models & mechanisms, 8*(1), 1-16.

Noguchi, S., Hattori, M., Sugiyama, H., Hanaoka, A., Okada, S., & Yoshida, T. (2012). Lactobacillus plantarum NRIC1832 enhances IL-10 production from CD4+ T cells in vitro. *Bioscience, biotechnology, and biochemistry*, 120404.

Noverr, M. C., Falkowski, N. R., McDonald, R. A., McKenzie, A. N., & Huffnagle, G. B. (2005). Development of allergic airway disease in mice following antibiotic therapy and fungal microbiota increase: role of host genetics, antigen, and interleukin-13. *Infection and immunity, 73*(1), 30-38.

O'mahony, L., Feeney, M., O'halloran, S., Murphy, L., Kiely, B., Fitzgibbon, J., . . . Collins, J. (2001). Probiotic impact on microbial flora, inflammation and tumour development in IL-10 knockout mice. *Alimentary pharmacology & therapeutics, 15*(8), 1219-1225.

Oozeer, R., Leplingard, A., Mater, D. D., Mogenet, A., Michelin, R., Seksek, I., . . . Corthier, G. (2006). Survival of Lactobacillus casei in the human digestive tract after consumption of fermented milk. *Applied and environmental microbiology, 72*(8), 5615-5617.

Org, E., Parks, B. W., Joo, J. W. J., Emert, B., Schwartzman, W., Kang, E. Y., . . . Gunsalus, R. (2015). Genetic and environmental control of host-gut microbiota interactions. *Genome research, 25*(10), 1558-1569.

Ottaviani, E., Ventura, N., Mandrioli, M., Candela, M., Franchini, A., & Franceschi, C. (2011). Gut microbiota as a candidate for lifespan extension: an ecological/evolutionary perspective targeted on living organisms as metaorganisms. *Biogerontology, 12*(6), 599-609.

Papagiannakopoulos, T., Bauer, M. R., Davidson, S. M., Heimann, M., Subbaraj, L., Bhutkar, A., . . . Jacks, T. (2016). Circadian rhythm disruption promotes lung tumorigenesis. *Cell metabolism, 24*(2), 324-331.

Pei, R., Martin, D. A., DiMarco, D. M., & Bolling, B. W. (2017). Evidence for the effects of yogurt on gut health and obesity. *Critical reviews in food science and nutrition, 57*(8), 1569-1583.

Perdigón, G., Rachid, M., De Budeguer, M. V., & Valdez, J. C. (1994). Effect of yogurt feeding on the small and large intestine associated lymphoid cells in mice. *Journal of Dairy Research, 61*(4), 553-562.

Peres, A. G., Stegen, C., Li, J., Xu, A. Q., Levast, B., Surette, M. G., . . . Madrenas, J. (2015). Uncoupling of pro-and anti-inflammatory properties of Staphylococcus aureus. *Infection and immunity, 83*(4), 1587-1597.

Qian, F., Deng, J., Wang, G., D Ye, R., & W Christman, J. (2016). Pivotal role of mitogen-activated protein kinase-activated protein kinase 2 in inflammatory pulmonary diseases. *Current Protein and Peptide Science, 17*(4), 332-342.

Qu, H., Zhu, M., Tao, Y., & Zhao, Y. (2015). Suppression of peripheral myelin protein 22 (PMP22) expression by miR29 inhibits the progression of lung cancer. *Neoplasma, 62*(6), 881-886.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research, 41*(D1), D590-D596.

Racedo, S., Villena, J., Medina, M., Agüero, G., Rodríguez, V., & Alvarez, S. (2006). Lactobacillus casei administration reduces lung injuries in a Streptococcus pneumoniae infection in mice. *Microbes and infection, 8*(9-10), 2359-2366.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., & Medzhitov, R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell, 118*(2), 229-241.

Rausch, P., Künzel, S., Suwandi, A., Grassl, G. A., Rosenstiel, P., & Baines, J. F. (2017). Multigenerational influences of the Fut2 gene on the dynamics of the gut microbiota in mice. *Frontiers in microbiology, 8*, 991.

Rausch, P., Steck, N., Suwandi, A., Seidel, J. A., Künzel, S., Bhullar, K., . . . Vallance, B. A. (2015). Expression of the blood-group-related gene B4galnt2 alters susceptibility to Salmonella infection. *PLoS pathogens, 11*(7), e1005008.

Raveh-Sadka, T., Thomas, B. C., Singh, A., Firek, B., Brooks, B., Castelle, C. J., . . . Morowitz, M. J. (2015). Gut bacteria are rarely shared by co-hospitalized premature infants, regardless of necrotizing enterocolitis development. *Elife, 4*, e05477.

Redford, P. S., Boonstra, A., Read, S., Pitt, J., Graham, C., Stavropoulos, E., . . . O'Garra, A. (2010). Enhanced protection to Mycobacterium tuberculosis infection in IL-10-deficient mice is accompanied by early and enhanced Th1 responses in the lung. *European journal of immunology, 40*(8), 2200-2210.

Reikvam, D. H., Erofeev, A., Sandvik, A., Grcic, V., Jahnsen, F. L., Gaustad, P., . . . Johansen, F.-E. (2011). Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression. *PloS one, 6*(3), e17996.

Remot, A., Descamps, D., Noordine, M.-L., Boukadiri, A., Mathieu, E., Robert, V., . . . Hammad, H. (2017). Bacteria isolated from lung modulate asthma susceptibility in mice. *The ISME journal, 11*(5), 1061-1074.

Ren, L., Zhang, R., Rao, J., Xiao, Y., Zhang, Z., Yang, B., . . . Shang, Y. (2018). Transcriptionally active lung microbiome and its association with bacterial biomass and host inflammatory status. *MSystems, 3*(5).

Rogers, G. B., Carroll, M. P., Serisier, D. J., Hockey, P. M., Kehagia, V., Jones, G. R., & Bruce, K. D. (2005). Bacterial activity in cystic fibrosis lung infections. *Respiratory research, 6*(1), 49.

Rogers, G. B., Daniels, T. W., Tuck, A., Carroll, M. P., Connett, G. J., David, G. J., & Bruce, K. D. (2009). Studying bacteria in respiratory specimens by using conventional and molecular microbiological approaches. *BMC pulmonary medicine, 9*(1), 14.

Roggenbuck, M., Anderson, D., Barfod, K. K., Feelisch, M., Geldenhuys, S., Sørensen, S. J., . . . Gorman, S. (2016). Vitamin D and allergic airway disease shape the murine lung microbiome in a sex-specific manner. *Respiratory research, 17*(1), 1-18.

Ronkina, N., Kotlyarov, A., Dittrich-Breiholz, O., Kracht, M., Hitti, E., Milarski, K., . . . Gaestel, M. (2007). The mitogen-activated protein kinase (MAPK)-activated protein kinases MK2 and MK3 cooperate in stimulation of tumor necrosis factor biosynthesis and stabilization of p38 MAPK. *Molecular and cellular biology, 27*(1), 170-181.

Rudofsky, U., Evans, B., Balaban, S., Mottironi, V., & Gabrielsen, A. (1993). Differences in expression of lupus nephritis in New Zealand mixed H-2z homozygous inbred strains of mice derived from New Zealand black and New Zealand white mice. Origins and initial characterization. *Laboratory investigation; a journal of technical methods and pathology, 68*(4), 419.

Russell, S. L., Gold, M. J., Reynolds, L. A., Willing, B. P., Dimitriu, P., Thorson, L., . . . Mohn, W. W. (2015). Perinatal antibiotic-induced shifts in gut microbiota have differential effects on inflammatory lung diseases. *Journal of Allergy and Clinical Immunology, 135*(1), 100-109. e105.

Russell, S. L., Gold, M. J., Willing, B. P., Thorson, L., McNagny, K. M., & Finlay, B. B. (2013). Perinatal antibiotic treatment affects murine microbiota, immune responses and allergic asthma. *Gut microbes, 4*(2), 158-164.

Sabat, R., Grütz, G., Warszawska, K., Kirsch, S., Witte, E., Wolk, K., & Geginat, J. (2010). Biology of interleukin-10. *Cytokine & growth factor reviews, 21*(5), 331-344.

Sadler, J. E. (1998). Biochemistry and genetics of von Willebrand factor. *Annual review of biochemistry, 67*(1), 395-424.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., . . . Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology, 12*(1), 87.

Scheiermann, J., & Klinman, D. M. (2017). Three distinct pneumotypes characterize the microbiome of the lung in BALB/cJ mice. *PloS one, 12*(7), e0180561.

Schröder, S., & Lohse, M. J. (2000). Quantification of the tissue levels and function of the G-protein regulator phosducin-like protein (PhlP). *Naunyn-Schmiedeberg's archives of pharmacology, 362*(4-5), 435-439.

Schuijt, T. J., Lankelma, J. M., Scicluna, B. P., e Melo, F. d. S., Roelofs, J. J., de Boer, J. D., . . . Belzer, C. (2016). The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. *Gut, 65*(4), 575-583.

Schulz, B. L., Sloane, A. J., Robinson, L. J., Prasad, S. S., Lindner, R. A., Robinson, M., . . . Packer, N. H. (2007). Glycosylation of sputum mucins is altered in cystic fibrosis patients. *Glycobiology, 17*(7), 698-712.

Segal, L. N., Alekseyenko, A. V., Clemente, J. C., Kulkarni, R., Wu, B., Chen, H., . . . Blaser, M. J. (2013). Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome, 1*(1), 19.

Segal, L. N., Clemente, J. C., Tsay, J.-C. J., Koralov, S. B., Keller, B. C., Wu, B. G., . . . Morris, A. (2016). Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nature microbiology, 1*, 16031.

Seifart, C., Plagens, A., Dempfle, A., Clostermann, U., Vogelmeier, C., von Wichert, P., & Seifart, U. (2005). TNF-α, TNF-β, IL-6, and IL-10 polymorphisms in patients with lung cancer. *Disease markers, 21*(3), 157-165.

Sellon, R. K., Tonkonogy, S., Schultz, M., Dieleman, L. A., Grenther, W., Balish, E., . . . Sartor, R. B. (1998). Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infection and immunity, 66*(11), 5224-5231.

Seong, E., Saunders, T. L., Stewart, C. L., & Burmeister, M. (2004). To knockout in 129 or in C57BL/6: that is the question. *Trends in genetics, 20*(2), 59-62.

Shanley, T. P., Vasi, N., & Denenberg, A. (2000). Regulation of chemokine expression by IL-10 in lung inflammation. *Cytokine, 12*(7), 1054-1064.

Singh, N., Vats, A., Sharma, A., Arora, A., & Kumar, A. (2017). The development of lower respiratory tract microbiome in mice. *Microbiome, 5*(1), 61.

Snijders, A. M., Langley, S. A., Kim, Y.-M., Brislawn, C. J., Noecker, C., Zink, E. M., . . . Huang, Y. (2016). Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome. *Nature microbiology, 2*, 16221.

Srinivas, G., Möller, S., Wang, J., Künzel, S., Zillikens, D., Baines, J. F., & Ibrahim, S. M. (2013). Genome-wide mapping of gene–microbiota interactions in susceptibility to autoimmune skin blistering. *Nature communications, 4*.

Srisutham, S., Saralamba, N., Malleret, B., Rénia, L., Dondorp, A. M., & Imwong, M. (2017). Four human Plasmodium species quantification using droplet digital PCR. *PloS one, 12*(4), e0175771.

Staubach, F., Künzel, S., Baines, A. C., Yee, A., McGee, B. M., Bäckhed, F., . . . Johnsen, J. M. (2012). Expression of the blood-group-related glycosyltransferase B4galnt2 influences the intestinal microbiota in mice. *The ISME journal, 6*(7), 1345-1355.

Stecher, B., Chaffron, S., Käppeli, R., Hapfelmeier, S., Freedrich, S., Weber, T. C., . . . von Mering, C. (2010). Like will to like: abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. *PLoS Pathog, 6*(1), e1000711.

Sturlan, S., Oberhuber, G., Beinhauer, B. G., Tichy, B., Kappel, S., Wang, J., & Rogy, M. A. (2001). Interleukin-10-deficient mice and inflammatory bowel disease associated cancer development. *Carcinogenesis, 22*(4), 665-671.

Sudo, T., Kawai, K., Matsuzaki, H., & Osada, H. (2005). p38 mitogen-activated protein kinase plays a key role in regulating MAPKAPK2 expression. *Biochemical and biophysical research communications, 337*(2), 415-421.

Suwal, S., Wu, Q., Liu, W., Liu, Q., Sun, H., Liang, M., . . . Liu, Z. (2018). The probiotic effectiveness in preventing experimental colitis is correlated with host gut microbiota. *Frontiers in microbiology, 9*, 2675.

Sze, M. A., Abbasi, M., Hogg, J. C., & Sin, D. D. (2014). A comparison between droplet digital and quantitative PCR in the analysis of bacterial 16S load in lung tissue samples from control and COPD GOLD 2. *PloS one, 9*(10).

Sze, M. A., Dimitriu, P. A., Hayashi, S., Elliott, W. M., McDonough, J. E., Gosselink, J. V., . . . Hogg, J. C. (2012). The lung tissue microbiome in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine, 185*(10), 1073-1080.

Tahara, Y., Yamazaki, M., Sukigara, H., Motohashi, H., Sasaki, H., Miyakawa, H., . . . Shibata, S. (2018). Gut microbiota-derived short chain fatty acids induce circadian clock entrainment in mouse peripheral tissue. *Scientific reports, 8*(1), 1-12.

Taylor, S. C., Laperriere, G., & Germain, H. (2017). Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Scientific reports, 7*(1), 1-8.

Teo, S. M., Mok, D., Pham, K., Kusel, M., Serralha, M., Troy, N., . . . Hollams, E. (2015). The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell host & microbe, 17*(5), 704-715.

Tian, F., Zhao, J., Fan, X., & Kang, Z. (2017). Weighted gene co-expression network analysis in identification of metastasis-related genes of lung squamous cell carcinoma based on the Cancer Genome Atlas database. *Journal of thoracic disease, 9*(1), 42.

Tomosada, Y., Chiba, E., Zelaya, H., Takahashi, T., Tsukida, K., Kitazawa, H., . . . Villena, J. (2013). Nasally administered Lactobacillus rhamnosus strains differentially modulate respiratory antiviral immune responses and induce protection against respiratory syncytial virus infection. *BMC immunology, 14*(1), 40.

Tudor, C., Marchese, F. P., Hitti, E., Aubareda, A., Rawlinson, L., Gaestel, M., . . . Dean, J. L. (2009). The p38 MAPK pathway inhibits tristetraprolin-directed decay of interleukin-10 and pro-inflammatory mediator mRNAs in murine macrophages. *FEBS letters, 583*(12), 1933-1938.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature, 449*(7164), 804-810.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., . . . Asplund, A. (2015). Tissue-based map of the human proteome. *Science, 347*(6220).

Van Der Gast, C. J., Walker, A. W., Stressmann, F. A., Rogers, G. B., Scott, P., Daniels, T. W., . . . Bruce, K. D. (2011). Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *The ISME journal, 5*(5), 780-791.

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., . . . Darzi, Y. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature, 551*(7681), 507-511.

Velcich, A., Yang, W., Heyer, J., Fragale, A., Nicholas, C., Viani, S., . . . Augenlicht, L. (2002). Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science, 295*(5560), 1726-1729.

Villena, J., Chiba, E., Tomosada, Y., Salva, S., Marranzino, G., Kitazawa, H., & Alvarez, S. (2012). Orally administered Lactobacillus rhamnosus modulates the respiratory immune response triggered by the viral pathogen-associated molecular pattern poly (I: C). *BMC immunology, 13*(1), 1-15.

Voigt, R. M., Forsyth, C. B., Green, S. J., Mutlu, E., Engen, P., Vitaterna, M. H., . . . Keshavarzian, A. (2014). Circadian disorganization alters intestinal microbiota. *PloS one, 9*(5), e97500.

Wagner, R. (1994). The regulation of ribosomal RNA synthesis and bacterial cell growth. *Archives of Microbiology, 161*(2), 100-109.

Wang, C.-Y., Hsieh, M.-J., Hsieh, I.-C., Shie, S.-S., Ho, M.-Y., Yeh, J.-K., . . . Wang, C.-C. (2016). CLOCK modulates survival and acute lung injury in mice with polymicrobial sepsis. *Biochemical and biophysical research communications, 478*(2), 935-941.

Wei, X., Yang, Z., Rey, F. E., Ridaura, V. K., Davidson, N. O., Gordon, J. I., & Semenkovich, C. F. (2012). Fatty acid synthase modulates intestinal barrier function through palmitoylation of mucin 2. *Cell host & microbe, 11*(2), 140-152.

Weyrich, L. S., Farrer, A. G., Eisenhofer, R., Arriola, L. A., Young, J., Selway, C. A., . . . Cooper, A. (2019). Laboratory contamination over time during low-biomass sample analysis. *Molecular ecology resources, 19*(4), 982-996.

Whelan, F. J., Heirali, A. A., Rossi, L., Rabin, H. R., Parkins, M. D., & Surette, M. G. (2017). Longitudinal sampling of the lung microbiota in individuals with cystic fibrosis. *PloS one, 12*(3), e0172811.

Whiteson, K. L., Bailey, B., Bergkessel, M., Conrad, D., Delhaes, L., Felts, B., . . . Maughan, H. (2014). The upper respiratory tract as a microbial source for pulmonary infections in cystic fibrosis. Parallels from island biogeography. *American journal of respiratory and critical care medicine, 189*(11), 1309-1315.

Willner, D., Daly, J., Whiley, D., Grimwood, K., Wainwright, C. E., & Hugenholtz, P. (2012). Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PloS one, 7*(4), e34605.

Yabumoto, Y., Watanabe, M., Ito, Y., Maemura, K., Otsuki, Y., Nakamura, Y., . . . Watanabe, K. (2008). Expression of GABAergic system in pulmonary neuroendocrine cells and airway epithelial cells in GAD67-GFP knock-in mice. *Medical molecular morphology, 41*(1), 20-27.

Yu, G., Gail, M. H., Consonni, D., Carugno, M., Humphrys, M., Pesatori, A. C., . . . Landi, M. T. (2016). Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. *Genome biology, 17*(1), 1-12.

Yun, Y., Srinivas, G., Kuenzel, S., Linnenbrink, M., Alnahas, S., Bruce, K. D., . . . Schaible, U. E. (2014). Environmentally determined differences in the murine lung microbiota and their relation to alveolar architecture. *PloS one, 9*(12), e113466.

Zackular, J. P., Baxter, N. T., Iverson, K. D., Sadler, W. D., Petrosino, J. F., Chen, G. Y., & Schloss, P. D. (2013). The gut microbiome modulates colon tumorigenesis. *MBio, 4*(6).

Zakrzewski, M., Goesmann, A., Jaenicke, S., Jünemann, S., Eikmeyer, F., Szczepanowski, R., . . . Schlüter, A. (2012). Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscriptome sequencing. *Journal of biotechnology, 158*(4), 248-258.

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., & Goodman, A. L. (2019). Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature, 570*(7762), 462-467.