Microbiome based clinical pathogen detection in bronchoalveolar lavage fluid using Next-Generation Sequencing

Inauguraldissertation
zur Erlangung des Grades eines Doktors der Medizin
des Fachbereichs Medizin
der Justus-Liebig-Universität Gießen

vorgelegt von Bitter, Merle Xenia
aus Neuburg an der Donau

Gießen (2024)

Aus dem Fachbereich Medizin der Justus-Liebig-Universität Gießen

Institut für Medizinische Mikrobiologie

Gutachter: PD Dr. Torsten Hain

Gutachter: Prof. Dr. Andreas Bräuninger

Tag der Disputation: 05.02.2026

Table of contents

1 Introduction

1.1 The human microbiome

Trillions of symbiotic bacteria colonize the human body, such as its skin, the nasal and oral cavities, vagina, intestines, and the respiratory tract (Marsland, Trompette and Gollwitzer, 2015; Natalini, Singh and Segal, 2023). For at least 500 million years, microbes and their hosts have formed cooperative relationships within complex ecosystems (Cho and Blaser, 2012). Those networks are beneficial in many respects, both for the host and the microbes. The microbiome holds great metabolic capacities, which are not mapped in the human genome, such as catabolism of dietary carbohydrates or the synthesis of short chain fatty acids (Mammen and Sethi, 2016; Young, 2017; Yadav, Verma and Chauhan, 2018; Tian *et al.*, 2020; Ney *et al.*, 2023). It can influence different signaling pathways as well as the activity of the immune system in the host organism. Moreover, the local microbiota can thwart the colonization of potentially pathogenic microbes, known as `colonization resistance` (Man, de Steenhuijsen Piters and Bogaert, 2017; Young, 2017; Mendez *et al.*, 2019; Natalini, Singh and Segal, 2023).

Whereas the bacterial composition and load differ between the different compartments, six main colonizers of the human body can be identified at phylum-level: *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Actinobacteria*, *Fusobacteria* and *Cyanobacteria* (Cho and Blaser, 2012; Marsland, Trompette and Gollwitzer, 2015).

The composition in the respective body sites of an individual is relatively stable over the course of time, though varies between individuals (Cho and Blaser, 2012). Factors like dietary habits, season, environment and medical intervention can also change the microbial load, composition, and interaction among the constituents of a microbiome (Cho and Blaser, 2012; Allaband *et al.*, 2019).

Bacteria that are part of the physiological microbiota at one body site can act as pathogens in other parts of the body or under other circumstances. This observation challenges the strict classification into pathogens and commensals within the framework of Koch`s postulates and suggests a holistic view on the microbial community rather than the focus on individual infectious agents (Cho and Blaser, 2012; Wu and Segal, 2018).

1.2 The microbiome of the respiratory tract

1.2.1 The microbiome of the respiratory tract in health

The lung microbiome during health is very dynamic and characterized by a high diversity, combined with a low bacterial density (Fromentin, Ricard and Roux, 2021; Yagi *et al.*, 2021). The microbiome of the lung has not been studied as extensively as e.g., the microbiome of the gut. This can be attributed to the fact that the lung has been thought to be sterile for a long time (Dickson and Huffnagle, 2015) but also to a better access to intestinal samples and the higher biomass in the gut (Huffnagle, Dickson and Lukacs, 2017). Nevertheless, new insights over the last decade regarding the microbiome of the respiratory tract in health and disease indicate its pathophysiological relevance (Dickson, Erb-Downward, *et al.*, 2016).

By now the lung is generally acknowledged to have an own microbiome, which has been sufficiently demonstrated by culture-independent molecular techniques (Dickson, Erb-Downward, *et al.*, 2016; Yagi *et al.*, 2021; Natalini, Singh and Segal, 2023). Notably, the lower respiratory tract is one of the least-populated body sites, with an estimated number of 1 – 10 bacteria per 100 human cells (Marsland, Trompette and Gollwitzer, 2015). In contrast, the ratio of the bacterial cells and all nucleated human cells in the body is about 1:1, according to newer estimations (Allaband *et al.*, 2019).

The microbiome of the respiratory tract is influenced by external factors, like diet, breast-feeding, smoking, pollution and antibiotics (Mendez *et al.*, 2019). Besides, environmental parameters along the upper and lower respiratory tract, such as pH, temperature, partial pressures of O_2 and CO_2, presence of surfactant as well as host-microorganism interactions create unique site-specific growth conditions, so that the precise microbial composition varies at different sites in the airways (Man, de Steenhuijsen Piters and Bogaert, 2017; Natalini, Singh and Segal, 2023). Common bacterial genera found in the respiratory tract during health are listed in Table 1.

Many studies provide evidence for the strong similarity in bacterial composition between the oropharynx and the lower respiratory tract during health whereas the nasal microbiome rather resembles the microbiome of the skin (Bassis *et al.*, 2015; Yagi *et al.*, 2021). These findings led to the understanding that bacteria from the oropharynx are the main source of the lung microbiome. However, the lower respiratory tract exhibits a much

lower biomass than the oropharynx (Huffnagle, Dickson and Lukacs, 2017; Man, de Steenhuijsen Piters and Bogaert, 2017; Wypych, Wickramasinghe and Marsland, 2019). The bacterial density decreases on the way from the carina further distally towards the smaller airways though a bacterial signal is constantly detectable (Martin-Loeches *et al.*, 2020). Variation of the microbial load, composition and diversity within the lung has been observed. However, this spatial variation within the lungs of an individual is small and exceeded by the intersubjective variability at the given anatomic sites (Dickson *et al.*, 2015; Dickson, Erb-Downward, *et al.*, 2016; Man, de Steenhuijsen Piters and Bogaert, 2017).

The bacterial load and composition in the lungs are fundamentally shaped by three factors: immigration, elimination, and relative reproduction rate of the microbial community members. In health, only little reproduction occurs due to the suboptimal growth conditions. During that state the lung microbiome is formed by the balance between immigration and elimination of transient microbes rather than by the reproduction of resident microbes (Dickson, Erb-Downward, *et al.*, 2016). The main source of microbial immigration that shapes the lung microbiome is assumed to be microaspiration of pharyngeal secretion, which occurs during health and disease, and explains the similarity to the oropharyngeal microbiome. Additionally, inhalation of airborne microbes and dispersion along the mucosal layer are regarded to be additional factors (Dickson, Erb-Downward, *et al.*, 2016; Huffnagle, Dickson and Lukacs, 2017; Wu and Segal, 2018; Wypych, Wickramasinghe and Marsland, 2019). Hence, the lower airway microbiome is physiologically more dynamic than the microbiome of the oral cavity or the intestines, frequently changing and influenced by external exposure to bacteria and respiratory clearance mechanisms (Natalini, Singh and Segal, 2023).

Bacteria constantly interact with each other, both in terms of cooperative relationships like cross-feeding, antibiotic degradation, or biofilm production as well as competitive interactions such as iron binding or hydrogen peroxide production (Welp and Bomberger, 2020). An interplay between the respiratory microbiome and the immune response, framed as the lung-lung axis, has been observed and analyzed in multiple studies (Segal *et al.*, 2016; Dickson *et al.*, 2018; Wu *et al.*, 2021).

Locations	Taxa on genus level	Reference
Nasal cavity	- *Corynebacterium* spp. - *Cutibacterium acnes* - *Dolosigranulum* spp. - *Moraxella* spp. - *Neisseria* spp. - *Propionibacterium* spp. - *Staphylococcus* spp. - *Streptococcus* spp.	(Man, de Steenhuijsen Piters and Bogaert, 2017; Welp and Bomberger, 2020; Rhoades *et al.*, 2021; Natalini, Singh and Segal, 2023)
Nasopharynx	- *Corynebacterium* spp. - *Dolosigranulum* spp. - *Haemophilus* spp. - *Moraxella* spp. - *Neisseria* spp. - *Staphylococcus* spp. - *Streptococcus* spp.	(Man, de Steenhuijsen Piters and Bogaert, 2017; Welp and Bomberger, 2020; Di Simone *et al.*, 2023; Natalini, Singh and Segal, 2023)
Oropharynx	- *Bacteroides* spp. - *Fusobacterium* spp. - *Leptotrichia* spp. - *Pevotella* spp. - *Rothia* spp. - *Streptococcus* spp. - *Veillonella* spp.	(Man, de Steenhuijsen Piters and Bogaert, 2017; Welp and Bomberger, 2020; Leitao Filho *et al.*, 2023; Natalini, Singh and Segal, 2023)
Lungs	- *Fusobacterium* spp. - *Porphyromonas* spp. - *Prevotella* spp. - *Rothia* spp. - *Streptococcus* spp. - *Trophyrema whipplei* - *Veillonella* spp.	(de Steenhuijsen Piters *et al.*, 2016; Man, de Steenhuijsen Piters and Bogaert, 2017; Wypych, Wickramasinghe and Marsland, 2019; Welp and Bomberger, 2020; Yagi *et al.*, 2021; Leitao Filho *et al.*, 2023; Natalini, Singh and Segal, 2023)

Table 1: Common bacteria in the different respiratory compartments during health (in alphabetical order).

1.2.2 The microbiome of the respiratory tract during acute infection

During pneumonia, the local microbiome structure in the lung collapses and one or few pathogenic taxa become dominant (Martin-Loeches *et al.*, 2020). However, the traditional pathogenesis concept of one pathogen causing an infectious disease has been increasingly replaced in recent years by a more dynamic model of pathogenesis (Wu and Segal, 2018). Complex interactions between the microbes and the host, rather than solely the emergence of a single infectious agent, determine the development of disease. The local growth

conditions, like pH, temperature, and nutrient supply, improve in favor of bacteria, which are well adapted to conditions shaped by inflammation. Byproducts of the inflammatory host response, such as catecholamines, inflammatory cytokines and free ATP, can function as growth factors for some bacteria, creating a vicious circle. That leads to a community structure which is increasingly influenced by the relative reproduction rates of certain bacteria (Dickson, Erb-Downward, *et al.*, 2016; Huffnagle, Dickson and Lukacs, 2017; Yagi *et al.*, 2021) and characterized by a decreased α-diversity (Kitsios, 2018; Hu *et al.*, 2020). With regard to the clinical picture, an association between microbial dysbiosis and disease severity has been described (Fromentin, Ricard and Roux, 2021).

The strict line between pathogenic and non-pathogenic bacteria becomes increasingly indistinct, as typical pathogens of lower respiratory tract infections like *Streptococcus pneumoniae* are commonly found in the airways of asymptomatic individuals (Mendez *et al.*, 2019). This paradigm shift in the pathogenesis of pneumonia is driven not least by the increasing establishment of next-generation sequencing (NGS) in respiratory research (Wu and Segal, 2018). In their case-control trial, de Steenhuijsen Piters and his colleagues (2016) compared the microbiome profiles of pneumonia patients with those of their healthy controls. They observed that the pneumonia cohort exhibited specific microbiome patterns which were dominated by *Lactobacilli*, *Rothia* and *Streptococcus*. The microbiome profiles of the healthy cohort were more diverse and characterized by taxa such as *Prevotella*, *Veillonella* and *Leptotrichia*.

The detected etiological pathogen spectrum of lower respiratory tract infections is influenced by several factors, like geographical area, patient group or microbiological methods (Woodhead *et al.*, 2011). However, interactions are not only observed between bacteria themselves or bacteria and their hosts but also between bacteria and viruses. Respiratory viral infections can lead to secondary bacterial super-infections. The increased post-viral host susceptibility to bacteria is the endpoint of a multifactorial process including mucosal damage, dysregulated immune responses, and changes in the microbiome. History provides a dramatic example of viral-bacterial interactions when millions of people died from secondary bacterial pneumonia during the Spanish flu pandemic in 1918-1919 (Hanada *et al.*, 2018). But there are also more recent examples (Tarabichi *et al.*, 2015; Mendez *et al.*, 2019; Hoefnagels *et al.*, 2021; Sulaiman *et al.*, 2021; Vitiello, Ferrara and Zovi, 2023). In a longitudinal case-control trial, Kaul and her

colleagues (2020) highlighted the relevance of the inter-kingdom connection when an influenza A infection was associated to specific disturbances in the upper respiratory tract microbiomes of ferrets and humans, which in turn possibly created a niche for opportunistic bacteria like *Pseudomonadales*. Reduced diversity and significantly higher relative abundances of *Pseudomonas* sp. were also observed in Covid-19 patients (Yagi *et al.*, 2021).

The members of the lung microbiome interact not only with each other, but also with bacteria elsewhere in the body, such as in the intestines, known as the gut-lung-axis. Dickson and his colleagues (2016) depicted the link between lung microbiome, gut bacteria, and acute respiratory disease in their study. An enrichment of the lung microbiome with gut-associated bacteria such as *Bacteroides* spp. was found both in a murine model of sepsis and in humans with acute respiratory distress syndrome (ARDS). The bacteria that were found in the bronchoalveolar lavage fluid (BALF) of ARDS patients most likely translocated from the lower gastrointestinal tract, were missed by culture, and correlated with the intensity of the inflammatory host response. Multiple further studies have demonstrated the crucial role of the gut-lung axis in respiratory disease (Hanada *et al.*, 2018; Mendez *et al.*, 2019; Dickson *et al.*, 2020; Martin-Loeches *et al.*, 2020; Stricker *et al.*, 2022).

To conclude, the assumption that pneumonia is mostly caused by one pathogen, is still generally acknowledged. But the theory of a linear pathogenesis that solely one microbe invades a sterile environment was replaced by a new model. The emergence of pneumonia is a dynamic multilayer process that arises from a homeostasis of biodiversity in health and is characterized by a decrease of microbial diversity, an increase of microbial biomass and host inflammation (Dickson, Erb-Downward, *et al.*, 2016).

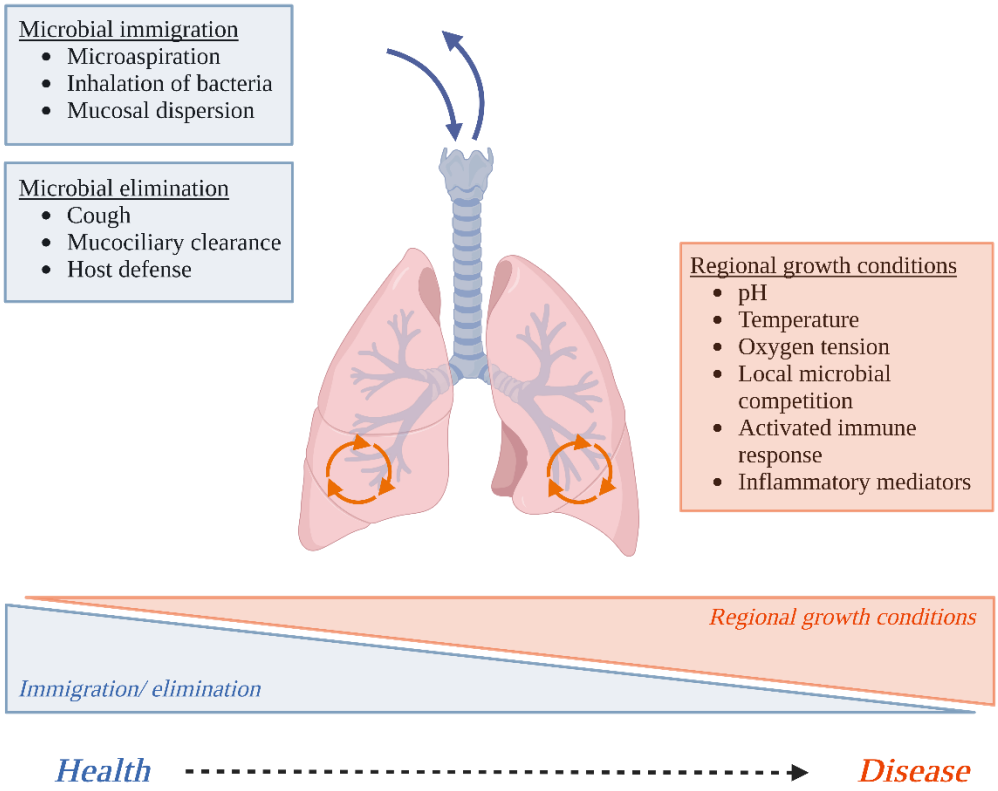


Figure 1: Factors that shape the respiratory microbiome in health and disease.

During health, immigration and elimination mechanisms have the greatest influence on the lung microbiome, which is highly dynamic in this state. During diseases, the taxonomic community structure is increasingly dominated by regional growth conditions. Figure adapted with permission by Dickson, Erb-Downward, et al. *(2016). The figure was created with BioRender ©.*

1.2.3 The microbiome of the respiratory tract during chronic diseases

In chronic respiratory diseases, like COPD or Cystic Fibrosis, changes in the pulmonary microbiome were observed, too (Dickson, Erb-Downward, *et al.*, 2016; Huffnagle, Dickson and Lukacs, 2017; Yagi *et al.*, 2021). The altered lung environment in the course of those diseases leads to changes in the community structure. A shift from *Bacteroidetes* phylum towards *Gammaproteobacteria*, common lung-associated gram-negative pathogens, has been described in numerous chronic diseases (Faner *et al.*, 2017;

Huffnagle, Dickson and Lukacs, 2017), as well as an increased microbial load in the lungs (Wypych, Wickramasinghe and Marsland, 2019).

As the local growth conditions for certain bacteria improve, the bacterial reproduction rates gain influence on the community composition. Contrary to a healthy lung, the α-diversity is decreased in many diseases and during mechanical ventilation, which has been shown in previous studies (Wu and Segal, 2018; Hu *et al.*, 2020; Chen *et al.*, 2021; Fromentin, Ricard and Roux, 2021). The cause of this loss of diversity has not been determined yet. Administration of antibiotics, interspecies competition or immune responses are discussed as possible influencing factors (Faner *et al.*, 2017). Pathobionts, like *Staphylococcus aureus* or *Pseudomonas aeruginosa,* are found in various chronic respiratory diseases, whereas other taxa seem to be specific to certain diseases (Welp and Bomberger, 2020; Fromentin, Ricard and Roux, 2021).

In COPD, an overrepresentation of *Haemophilus*, *Pseudomonas*, *Moraxella*, *Neisseria* and *Streptococcus* was observed (Marsland, Trompette and Gollwitzer, 2015; Mendez *et al.*, 2019). An exacerbation could be linked to the increased presence of *Haemophilus influenzae* in patient sputum (Mendez *et al.*, 2019). Etiopathogenic approaches assume that the compromised immune system due to COPD leads to increased susceptibility to infectious agents, which in turn impairs immune response mechanisms such as the mucociliary clearance, forming a vicious circle (Mendez *et al.*, 2019).

In Cystic Fibrosis (CF), well known bacterial causative agents are *Ps. aeruginosa* [1], *St. aureus*, *Burkholderia cepacia* and *H. influenzae*. However, recent studies suggest that a number of other taxa can be found in the lungs of CF patients, like the species *Stenotrophomonas maltophilia*, MRSA and *Mycobacteroides abscessus* (Mendez *et al.*, 2019; Wypych, Wickramasinghe and Marsland, 2019), and genera such as *Streptococcus*, *Prevotella* and *Veillonella* (Meskini *et al.*, 2021; Reece, Bettio and Renwick, 2021; Widder *et al.*, 2022). Several studies observed a shift in taxonomic composition from childhood to adulthood (Meskini *et al.*, 2021), as well as a connection between the loss of lung function over time with a loss of microbial diversity (Cox *et al.*, 2010; Reece, Bettio and Renwick, 2021).

[1] The exact notation of bacterial taxa is based on the 2022 Revision of the International Code of Nomenclature of Prokaryotes (Oren *et al.*, 2023).

All in all, there seem to be signature alterations in the lung microbiome in certain diseases, often linked to clinical parameters like disease severity. A common factor in all these diseases has been described as a loss of diversity and a dysbiosis dominated by *Proteobacteria* (Faner *et al.*, 2017; Natalini, Singh and Segal, 2023).

1.3 Next-generation sequencing

1.3.1 Types of next-generation sequencing

Over the last decade, massively parallel sequencing techniques combined with bioinformatic data analysis have transformed both research and understanding of the human microbiome fundamentally (Dickson, Erb-Downward, *et al.*, 2016; Man, de Steenhuijsen Piters and Bogaert, 2017; Natalini, Singh and Segal, 2023). In contrast to previous diagnostic methods, next-generation sequencing (NGS) represents an unbiased approach that does not a priori target a specific pathogen but provides insights into the lung microbiome as a complex community (Wu and Segal, 2018).

There are two types of high-throughput DNA sequencing (NGS) protocols: amplicon sequencing and whole genome sequencing. During amplicon sequencing, only one specific region of a gene is sequenced, which serves as a phylogenetic marker. This is most commonly the 16S rRNA gene for bacteria and archaea or the internal transcribed spacer (ITS) for fungi. Amplicon analyses in viruses are not as established since there is no suitable gene found common to all viruses (Allaband *et al.*, 2019).

In contrast, whole genome sequencing targets the entire DNA of a sample, including viral and eukaryotic DNA (Knight *et al.*, 2018; Allaband *et al.*, 2019). While 16S rRNA gene sequencing provides data on the community profile with the relative abundances of its members, whole genome sequencing additionally allows conclusions to be drawn about the functional features of the community (Young, 2017; Knight *et al.*, 2018).

Ultimately, the present RNA transcripts in a microbiome can be mapped and studied, referred to as metatranscriptomics, which enables further perceptions about the functionally active genes at a certain point in time (Young, 2017; Knight *et al.*, 2018).

The type of sequencing is chosen depending on the research hypothesis (Knight *et al.*, 2018; Allaband *et al.*, 2019; Santiago-Rodriguez *et al.*, 2020).

1.3.2 16S rRNA gene sequencing

70S ribosomes, which are composed of a 50S and a 30S subunit, are ubiquitously present in all bacteria. The 16S rRNA belongs to the smaller 30S subunit and is highly conserved due to its essential function in the bacterial organism (Carney *et al.*, 2020). Additionally, the 16S rRNA gene consists of nine variable regions, which can be used to discriminate different bacterial taxa (Allaband *et al.*, 2019).

16S rRNA gene sequencing is an NGS technique which uses the 16S rRNA gene as a marker gene to decipher the structure of a microbial community, including the present microbes and their relative abundances (Young, 2017). The conserved regions in the bacterial 16S rRNA gene are targets to PCR primers and can be amplified across all bacteria. The nine variable regions in the gene, most commonly the V4 region (Figure 2), allow the differentiation between bacterial taxa (Allaband *et al.*, 2019; Carney *et al.*, 2020). By sequencing the products of the 16S rRNA PCR, thousands of short genome sequences can be generated and organized. Sequences, which share 97 % or more similarity are grouped in operational taxonomic units (OTUs) (Dickson, Erb-Downward, *et al.*, 2016). Taxonomic databases can then be used to create a microbial community profile. Despite the variable regions, many bacterial species show strong similarities over the entire length of the 16S rRNA gene. For that reason, the sequencing of variable regions of the 16S rRNA gene is mostly able to discriminate between bacterial taxa at genus level but not at the species level (Knight *et al.*, 2018; Allaband *et al.*, 2019).

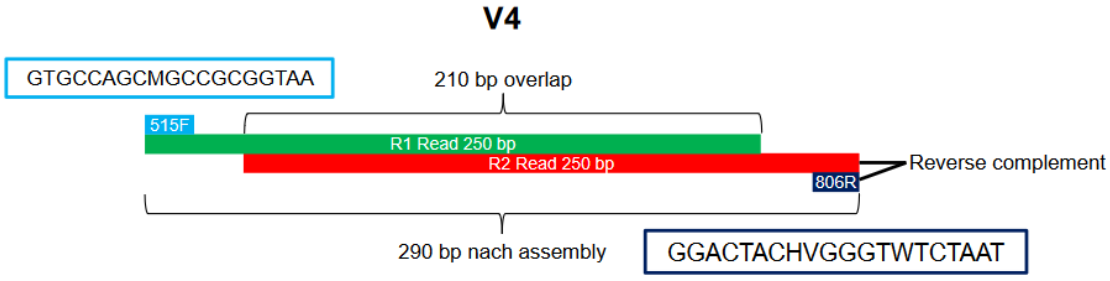


Figure 2: Structure of a V4 16S rRNA sequencing read (source: Markus Weigel).

1.3.3 Alpha and beta diversity

α-diversity provides information about the microbial composition of a given community e.g., within a sample. It quantifies how heterogenous this community is, based on the number of species observed and the proportions they account for. There are different metrics to estimate α-diversity. The observed richness uses the total number of species identified in the respective sample. Chao1 also includes the different abundances of the species present in the sample. Shannon`s diversity index additionally considers the evenness of the species abundances and indicates how reliably the next taxon can be predicted. α-diversity is commonly used for rarefaction curves or comparisons of sample groups (Knight *et al.*, 2018).

β-diversity refers to the diversity between different microbial communities, hence how similar or how different two samples are from each other. As with α-diversity, there are several metrics for estimating β-diversity. The Bray-Curtis metric is quantitative and generates distance matrices of β-diversity differences between all pairs of specimens based on the frequencies of their features. Those distance matrices can then be comprehensibly visualized by principal coordinates analysis (PCoA) (Knight *et al.*, 2018).

1.3.4 Extraction methods

A basic precondition to sequencing the bacterial nucleic acid is to make it accessible through extraction. There are different methods to lyse the cell wall for the extraction process: mechanical (via bead-beating), enzymatic and chemical lysis (Teng *et al.*, 2018; Angebault *et al.*, 2020). The choice of extraction method is important because it can introduce a bias towards specific bacteria, e.g. due to differences in microbial cell walls (Kennedy *et al.*, 2014; Hermans, Buckley and Lear, 2018).

1.4 Lower respiratory tract infections

1.4.1 Epidemiology and etiology of lower respiratory tract infections

Lower respiratory tract infections (LRTI) are the sixth leading cause of death among all ages and the leading cause of death among children younger than 5 years (GBD 2016 Causes of Death Collaborators, 2017). In 2016, LRTI were responsible for nearly 2.38 million deaths worldwide, as the Global Burden of Diseases, Injuries and Risk Factors showed. Particularly at risk are children under five years and those suffering from malnutrition, adults older than 70 years, populations exposed to air pollution as well as people threatened by poverty and with poor access to health care. Among etiologic agents, *Str. pneumoniae* contributed most to mortality, accounting for 1.19 million deaths in 2016 (GBD 2016 Lower Respiratory Infections Collaborators, 2018).

In Germany, pneumonia is the most common infectious disease leading to hospitalization and a grave cause for morbidity and mortality (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021). The most frequent bacterial pathogens of community-acquired pneumonia (CAP) and hospital-acquired pneumonia (HAP) are listed in Table 2. Bacteria that can generally be attributed to the physiological lung flora and therefore do not require therapeutical intervention, are listed in Table 3. This classification is based on the German S3 guidelines for CAP and HAP.

Community-acquired pneumonia	***Hospital-acquired pneumonia***
<u>Common pathogens</u>	*Acinetobacter baumanii*
H. influenzae	Enterobacteria
Str. pneumoniae	- *E. coli*
St. aureus	- *Klebsiella* spp.
	- *Enterobacter* spp.
<u>Rare pathogens</u>	*H. influenzae*
Enterobacteria	*Ps. aeruginosa*
- *Escherichia coli*	*St. aureus*
- *Klebsiella pneumoniae*	*Ste. maltophilia*
- *Proteus mirabilis*	*Str. pneumoniae*
Ps. aeruginosa	

Table 2: Most frequent pathogens of community-acquired pneumonia and hospital-acquired pneumonia according to the German S3 guidelines (Dalhoff* et al.*, 2017; Ewig* et al.*, 2021).

Community-acquired pneumonia	***Hospital-acquired pneumonia***
α-hemolytic *Streptococcus*	α-hemolytic *Streptococcus*
Coagulase-negative *Staphylococcus*	Coagulase-negative *Staphylococcus*
Corynebacterium spp.	*Corynebacterium* spp.
Enterococcus spp.	*Enterococcus* spp.
Haemophilus spp. (except *H. influenzae*)	*Neisseria* spp.
Neisseria spp.	

Table 3: Bacteria with no therapeutical consequence in community-acquired pneumonia and hospital-acquired pneumonia according to the German S3 guidelines (Dalhoff* et al.*, 2017; Ewig* et al.*, 2021).

1.4.2 Pathogen detection in respiratory specimens

The culturing of respiratory samples, like BALF or sputum, is a well-established practice and a firm component in the diagnostic algorithm of respiratory diseases (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021). For CAP, the recommendation for pathogen diagnosis depends on severity and circumstances, like hospitalization (Ewig *et al.*, 2021). For HAP, quantitative or semiquantitative cultures are more liberally recommended in the German guidelines. Molecular genetic analyses for pathogen detection are currently not recommended (Dalhoff *et al.*, 2017).

The timely identification of etiologic microbes in LRTI is critical for the successful treatment of patients. However, pathogens frequently remain undetected by the current culture-based diagnostic procedures (Kalantar and Langelier, 2021). In their systematic review, Shoar and Musher (2020) showed that the studies analyzed did not find an etiologic agent for CAP in more than half of the cases using established microbial tests (culture and PCR), which was in line with several other studies (Muscedere *et al.*, 2012; Jain *et al.*, 2015; Gadsby *et al.*, 2016). Furthermore, some bacteria, like *Treponema pallidum* or *Bartonella* spp. are difficult to cultivate, or even uncultivable (Gu *et al.*, 2021). *Legionella*, *Mycobacterium* or other atypical bacteria are reliant on specialized media and conditions and can take long to grow (Wu and Segal, 2018; Zhu *et al.*, 2023).

In lung research, the implementation of NGS techniques is comparatively young and increasingly carried out for about ten years (Carney *et al.*, 2020). However, there are several studies which compare the bacterial detection yield of culture with the results from NGS techniques in different sample types (Kitsios *et al.*, 2018; Yang *et al.*, 2019; Yoo *et al.*, 2020; Mahnic *et al.*, 2021). A number of studies evaluated the diagnostic yield of metagenomic NGS (mNGS) in BALF samples, mostly in the context of pneumonia. In these studies, the detection rate of mNGS was comparable to that of culture, depending on study design. In addition, it identified several species which were not detected by culture and provided further phylogenetic information useful for clinical practice. Besides, shorter processing times and improved patient outcomes were reported (Charalampous *et al.*, 2019; Wang, Han and Feng, 2019; Fang *et al.*, 2020; Li *et al.*, 2020; Wang *et al.*, 2020; Qian *et al.*, 2021; Xie *et al.*, 2021).

There are not many studies evaluating the performance of 16S rRNA gene sequencing for BALF samples. Dickson et al. (2014) analyzed 46 BALF samples from lung transplant recipients by culture and by 16S rRNA gene sequencing. In NGS, more bacteria could be detected than in culture growth, and in two cases, pathogens were identified that were missed by culture. Besides, a positive culture result was positively associated with culture-independent indices of infection, like DNA burden and low α-diversity. In another study, Zachariah et al. (2018) compared the results of 16S rRNA gene sequencing with conventional culture in 81 BALF samples of pediatric CF patients. NGS detected genera correspondent to cultured species in more than 90 percent and recorded additional genera of potential clinical relevance, like *Staphylococcus*, *Legionella*, and *Pseudomonas*. Bador et al. (2020) reasoned in their study with 132 patients that the overall sensitivity of 16S rRNA gene sequencing was low but that it could be helpful in cases of a negative culture due to a previous antibiotic therapy. 16S rRNA gene sequencing presents a promising approach for analyzing BAL fluid to enhance pathogen detection in a clinical setting and to gain insights into the community structure of the lung microbiome for a more personalized patient management.

1.5 Objectives of this work

The overall purpose is to contribute to the improvement and development of microbiological methodology in clinical pathogen diagnostics. The aim of this work was to descriptively compare bacterial microbiome profiles from V4 16S rRNA gene sequencing of BALF samples with the results from culture-based routine diagnostics. Additionally, as part of a technical analysis, two extraction methods (manual mechanical extraction with the ZymoBIOMICS DNA Miniprep Kit and automated enzymatic extraction with the eMAG platform) were compared in terms of resulting microbiome profiles and contamination. The overall long-term goal is to contribute to the improvement and development of microbiological methodology in clinical pathogen diagnostics.To reach these goals, the following questions were formulated:

1. Do the manual mechanical extraction method and the automated enzymatic extraction platform produce different results?
2. Do the cultured species overlap with the NGS results?
3. Does 16S rRNA gene sequencing with Illumina create an added value to the analysis of the BALF samples?
4. Does long-read sequencing with Nanopore create an added value to the analysis of BALF samples?

2 Material and methods

2.1 Material

2.1.1 Used devices

Device/ model	Producer, head office	REF/ Cat. No
Fragment Analyzer System	Agilent Technologies Inc., Santa Clara, USA	-
2100 Bioanalyzer Instrument	Agilent Technologies Inc., Santa Clara, USA	G2939B
Biometra TAdvanced Basis 230 V (thermal PCR cycler)	Analytik Jena GmbH, Jena, Germany	846-2-070-280
StepOnePlus Real-Time PCR System	Applied Biosystems (Thermo Fisher Scientific), Waltham, USA	15341295
Platefuge Microplate Microcentrifuge with rotor and plate carriers	Benchmark Scientific Inc., Edison, USA	Z742568
Automated extraction platform for nucleic acid purification/ eMAG	bioMérieux SA, Marcy-l`Étoile, France	418591
Centrifuge 5418 R	Eppendorf AG, Hamburg, Germany	5401000010
Hamilton Microlab STAR	HAMILTON Bonaduz AG, Bonaduz, Switzerland	(customized)
LabElite Integrated I.D. Capper (Version 1.0.4)	HAMILTON Bonaduz AG, Bonaduz, Switzerland	(customized)
Invitrogen Qubit 3 Fluorometer	Invitrogen (Thermo Fisher Scientific), Carlsbad, USA	15387293
Multifuge 3L-R	Kendro Laboratory Products, Osterode, Germany	75004370
Milli-Q Direct 8 Water Purification System	Merck Millipore, Billerica, USA	ZR0Q008WW
BioShake iQ	QInstruments GmbH, Jena, Germany	1808-0506
Vortex-Genie 2	Scientific Industries, Bohemia, USA	SI-0236
Horizontal-(24) Microtube (tube holder)	Scientific Industries, Bohemia, USA	SI-H524
Pop-off Cup (single tube holder)	Scientific Industries, Bohemia, USA	146-3011-00
ALPS 50 V Microplate Heat Sealer	Thermo Fisher Scientific Inc., Waltham, USA	AB-1443A
Biological Safety Cabinet Safe 2020 (Clean Bench)	Thermo Fisher Scientific Inc., Waltham, USA	10462614

Thermo Scientific HERAfreeze HFU T Series -86°C Upright Ultra-Low Temperature Freezers	Thermo Fisher Scientific Inc., Waltham, USA	HFU600TV
VWR Galaxy Mini Microcentrifuge	VWR International, Radnor, USA	-
VWR Signature™ Ergonomic High Performance Single-Channel Variable Volume Pipettors: 2 µl, 10 µl, 20 µl, 100 µl, 200 µl and 1000 µl	VWR International, Radnor, USA	613-5258 (2 µl) 613-5259 (10 µl) 613-5260 (20 µl) 613-5262 (100 µl) 613-5263 (200 µl) 613-5265 (1,000 µl)

Table 4: Used devices listed with producer and REF or catalog number.

2.1.2 Expendable material

Expendable material/ model	**Producer, head office**	**REF/ Cat. No**
MicroAmp Fast Optical 96-Well Reaction Plate (0.1 mL)	Applied Biosystems (Thermo Fisher Scientific Inc.), Waltham, USA	4346907
MicroAmp Optical Adhesive Film	Applied Biosystems (Thermo Fisher Scientific Inc.), Waltham, USA	4311971
Microtubes, 5 mL, clear, pre-sterilized	Axygen Scientific, Union City, USA	MCT-500-C-S
Hard-Shell PCR Plates, 96-well, thin-well	Bio-Rad Laboratories, Hercules, USA	HSP9601
SafeSeal-Tips, 200 µl, with filter, low bining	Biozym Scientific GmbH, Hessisch Oldendorf, Germany	770200
DNA LoBind Tubes 1.5 mL	Eppendorf AG, Hamburg, Germany	022431021
Twin.tec PCR Plate 96, semiskirted, red	Eppendorf AG, Hamburg, Germany	951020389
CO-RE Tips with Filters, sterile, 50 µl	HAMILTON Bonaduz AG, Bonaduz, Switzerland	235979
CO-RE Tips with Filters, sterile, 300 µl	HAMILTON Bonaduz AG, Bonaduz, Switzerland	235938
CO-RE Tips with Filters, sterile, 1.000 µl	HAMILTON Bonaduz AG, Bonaduz, Switzerland	235940
Qubit assay tubes	Invitrogen (Thermo Fisher Scientific), Carlsbad, USA	Q32856
Filter Tips, 0,1-10 µl, super slim, 96 pcs/rack, transparent tips, moulded rings, autoclavable	nerbe plus & Co. KG, Winsen/ Luhe, Germany	07-613-8300

Filter tips, 0-20µl, 96 pcs/rack, transparent tip, graduated, autoclavable	nerbe plus & Co. KG, Winsen/ Luhe, Germany	06-622-5300
Filter Tips, 0-100µl, 96 pcs/rack, transparent tip, moulded rings, autoclavable	nerbe plus & Co. KG, Winsen/ Luhe, Germany	07-642-5300
Filter Tips, 100-1.000µl, 96 pcs/rack, transparent tip, moulded rings, autoclavable	nerbe plus & Co. KG, Winsen/ Luhe, Germany	06-693-5300
BlackKnights H with filters, pre-sterilized: 50 µl, 300 µl and 1000 µl	Ritter GmbH, Schwabmünchen, Germany	49010-0105 (50 µl) 49008-0105 (300 µl) 49009-0105 (1,000 µl)
Alu-Sealing Tape, pierceable	SARSTEDT AG & Co. KG, Nümbrecht, Germany	95.1995
Biosphere SafeSeal Tube, 1.5 ml	SARSTEDT AG & Co. KG, Nümbrecht, Germany	72.706.200
Micro tube 1.5 mL DNA LowBind	SARSTEDT AG & Co. KG, Nümbrecht, Germany	72.706.700
PCR-Film, free of DNase/RNase, self-adhesive	SARSTEDT AG & Co. KG, Nümbrecht, Germany	95.1993
Screw cap micro tube, 1.5 ml, sterile	SARSTEDT AG & Co. KG, Nümbrecht, Germany	72.692.005
Abgene 96 Well 0.8mL Polypropylene Deepwell Storage Plate (AB-0765)	Thermo Fisher Scientific Inc., Waltham, USA	12194162
Easy Pierce Heat Sealing Foil	Thermo Fisher Scientific Inc., Waltham, USA	AB-0757

Table 5: Expendable materials listed with producer and REF or catalog number.

2.1.3 Kits and reagents

Kit	**Method**	**Producer, head office**	**REF/ Cat. No**
Fragment Analyzer dsDNA 935 Reagent Kit (1-1500bp)	Library quantitation and QC	Agilent Technologies Inc., Santa Clara, USA	DNF-935-K1000
High Sensitivity DNA Reagents	Library quantitation and QC	Agilent Technologies Inc., Santa Clara, USA	5067-4626
Nuclease-free water (Ambion)	Wet lab before 16S rRNA V4 amplification	Invitrogen (Thermo Fisher Scientific), Carlsbad, USA	AM9932
AMPure XP	PCR purification	Beckman Coulter, Indianapolis, USA	A63881

500-cycle MiSeq Reagent Nano Kits v2 (Box 1 and 2)	DNA sequencing	Illumina, San Diego, USA	15033625 *and* 15036714
Custom DNA Oligo Primers	DNA barcoding	Integrated DNA Technologies Inc., Coralville, USA	(customized)
Platinum SuperFi PCR Master Mix	16S V4 amplification	Invitrogen (Thermo Fisher Scientific), Carlsbad, USA	12358-050
Quant-iT PicoGreen dsDNA Assay Kit	DNA quantitation	Invitrogen (Thermo Fisher Scientific), Carlsbad, USA	P7589
Qubit dsDNA HS Assay Kit	DNA quantitation	Invitrogen (Thermo Fisher Scientific), Carlsbad, USA	Q33231
NEBNext Library Quant Kit for Illumina	Library quantitation	New England BioLabs Inc., Ipswich, USA	E7630L
NEBNext Ultra II Q5 Master Mix	DNA barcoding	New England BioLabs Inc., Ipswich, USA	M0544S
16S Barcoding Kit 1-24	Nanopore sequencing	Oxford Nanopore Technologies, Oxford, UK	SQK-16S024
Allpex Respiratory Panel 4	Respiratory PCR panel	Seegene, Seoul, Republic of Korea	RP9803X
Ethanol (CAS: 64-17-5)	DNA purification	Sigma-Aldrich, St. Louis, USa	32205-M
ZymoBIOMICS DNA/RNA Shield	Pathogen inactivation and genetic integrity	Zymo Research, Irvine, USA	R1100-250
ZymoBIOMICS DNA Miniprep Kit	DNA extraction	Zymo Research, Irvine, USA	D4300

Table 6: Used kits and reagents listed with methods, producer and REF or catalog number.

2.1.4 Used software

Software program	**Purpose**	**Producer information**	**Version**
Step One Software	Pico Green and qPCR measurement	Applied Biosystems (Thermo Fisher Scientific Inc.), Waltham, USA	2.2.2
ProSize	Analysis of Fragment Analyser measurement	Agilent Technologies Inc., Santa Clara, USA	3.0.1.5
Fragment Analyser	Measurement	Agilent Technologies Inc., Santa Clara, USA	1.2.0.11
GraphPad Prism 9 for Windows 64-bit	Data analysis and visualization	GraphPad Software LLC, San Diego, USA	9.5.0

Hamilton Run Control	Workflow programming with Hamilton pipetting robot	HAMILTON Bonaduz AG, Bonaduz, Switzerland	4.4.0.7740
SILVA ribosomal RNA	Database	Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany	138.1
M3P	Microbiome profile analysis and visualization	Markus Weigel	-
blastn	Analysis of bacterial sequences	National Center for Biotechnology Information, Bethesda, USA	2.12.0+
Epi2me	Processing of Nanopore sequencing data	Oxford Nanopore Technologies (Metrichor), Oxford, UK	5.0.3
mothur	Sequencing data processing	Schloss *et al.* (2009)	1.46.1
vsearch	Sequencing data processing	Rognes *et al.* (2016)	v2.17.1_linux_x86_64

Table 7: Used software and programs listed with purpose, producer and version number.

2.2 Sample acquisition

Bronchoalveolar lavage was performed in patients at the University Hospital Giessen and Marburg, site Giessen (UKGM Giessen, Germany) for sample acquisition from the lower respiratory tract. 150 ml of 0.9 % sodium chloride fluid was instilled and reabsorbed through a bronchoscope. 6 technical controls were included to detect contamination during the lavages. 3 controls consisted of the 0.9 % sodium chloride fluid used for the lavage, and 3 controls were blank rinses of the bronchoscope with the 0.9 % sodium chloride fluid before patient contact. The collection of the technical controls was conducted on three different days. The technical controls were treated the same way as the other samples in the following workflow.

BALF samples then passed through the routine diagnostic procedure and were stored at 4 °C. After picking up the samples at the Institute of Medical Microbiology for further laboratory workflow, the samples were stored at -80 °C.

2.3 Biomolecular methods

2.3.1 Microscopy, bacterial culture, and identification by MALDI-TOF

The BAL fluid samples were cultured according to standard routine protocols of the Institute of Medical Microbiology. All BALF samples were inoculated with a dilution loop onto different agar plates (Columbia sheep blood, Macconkey, chocolate blood and Sabouraud GC agar). The plates were incubated, and results were registered semi quantitively.

From each agar plate, colonies were selected and prepared for species identification by MALDI-TOF.

The procedures described above were performed by the staff of the diagnostic section of the Institute of Medical Microbiology, University Hospital Giessen and Marburg, site Giessen (UKGM Giessen, Germany).

The cultured bacteria were evaluated for potential pathogenicity using the S3 guidelines for community- and hospital-acquired pneumonia and by including microscopic features (granulocytes and epithelial cells) as well as quantitative information from culture growth.

2.3.2 Respiratory PCR panel

The Allpex Respiratory Panel 4 was performed on all specimens with a laboratory request according to the manufacturer`s instructions. It comprised the following bacterial species: *Bordetella parapertussis*, *Bordetella pertussis*, *Chlamydophila pneumoniae*, *H. influenzae*, *Legionella pneumophila*, *Mycoplasma pneumoniae* and *Str. pneumoniae*.

2.3.3 Automated DNA extraction

The DNA extraction with the automated extraction platform for nucleic acid purification eMAG was conducted in close cooperation with the diagnostic division of the Institute of Medical Microbiology, UKGM Giessen. Extraction of nucleic acid from the BAL fluids was carried out according to the standard operation procedure "SAA-PCR-Ger-001-eMAG", which was available in the diagnostic division. 250 µl of each sample was mixed with Proteinase K to lyse the cellular material, virus components, bacteria, and fungi. Nucleases were inactivated by lysis buffer. With magnetic silica and wash buffers, the nucleic acid was isolated and purified. Finally, the nucleic acid was eluted from the magnetic silica particles with an elution buffer at increased temperature. Negative Kit controls were included to identify contamination during the extraction process. Every sample was tested for the presence of SARS-CoV-2 using the Allplex 2019-nCoV Assay. Samples with a positive assay were excluded for safety reasons. The remaining samples were stored at -80 °C for further analysis.

2.3.4 Manual DNA extraction

The manual DNA extraction was carried out with the ZymoBIOMICS DNA Miniprep Kit and according to the producer`s instructions. Initially, 200 µl of each sample was mixed with 800 µl DNA/RNA Shield in a Biosphere Save Seal 1.5 ml tube and incubated for 30 minutes for the inactivation of infectious agents and stabilization of nucleic acids in the specimens. Following, the samples were put into the BioShake iQ with gentle shaking (800 rpm) for 15 minutes at 75 °C. After cooling down for five minutes, the samples were transferred into ZR BashingBead Lysis tubes and processed in the Vortex Genie 2 for 40 minutes at level ten for mechanical lysis. Then, the specimens were filtered and purified with wash buffers in multiple steps according to the protocol and finally eluted in 50 µl DNAse/RNAse free water. Negative kit controls with DNAse/RNAse free water from the extraction kit and Ambion water were included to identify contamination during the extraction process.

2.3.5 DNA quantitation of the extracts

After DNA extraction, the DNA in all extracts was quantified using the Quant-iT PicoGreen dsDNA Assay kit and the Qubit High Sensitivity kit.

The PicoGreen dsDNA Quantitation reagent fluorescently stained double-stranded DNA. It detected dsDNA in a range from 25 pg/mL to 1,000 ng/mL (Molecular Probes - Invitrogen, 2008). As a preparatory step, 45 µl of the Quant-iT PicoGreen dsDNA reagent was diluted with 3955 µl 1x TE buffer. The pipetting steps described below were executed by the Hamilton Microlab STAR, the programming was developed by Jan-Philipp Mengel from the Institute of Medical Microbiology from the Justus-Liebig-University Giessen. 25 µl of the diluted dye was pipetted into each well of a Bio-Rad Hard-Shell 96-Well PCR Plate. 3 µl of the sample was filled up with 22 µl 1x TE buffer to a total volume of 25 µl and added to the wells of the MicroAmp Fast 96-Well plate. A MicroAmp-optical adhesive film was affixed to the plate and the optical measurement was performed in the StepOnePlus Real-Time PCR device, which calculated the mean value of the fluorescence measurement cycles. In parallel, eight λ-DNA standards were treated as the samples, as described above, and measured along as concentration references. A maximum of 88 specimens could be processed in one operation. Additional specimens were measured with the Qubit dsDNA HS assay kit.

During DNA quantitation with the Qubit 1x dsDNA HS assay kit, the dye from the assay kit bound to the double-stranded DNA and emitted a fluorescent signal, which was measured by the Qubit fluorometer. Two standards with a known concentration were used for calibration (Thermo Fisher Scientific, 2021). dsDNA in a range from 10 pg/µL to 100 ng/µL could be detected (Thermo Fisher Scientific, 2020). In Qubit assay tubes, 3 µL of the respective sample was combined with 197 µL of Qubit working solution, which consisted of the Qubit reagent and the Qubit buffer in a ratio of 1:200. For calibration purposes, 10 µL of standard 1 and standard 2 was each combined with 190 µL of Qubit working solution. The prepared assay tubes were vortexed for 2-3 seconds and incubated for 2 minutes at room temperature. For the measurement, the fluorometer was first calibrated with the two standards and then the sample tubes were inserted for DNA quantitation.

2.3.6 16S rRNA V4 PCR and cleanup

In the first PCR, the targeted variable V4 region of the bacterial 16S rRNA gene was to be amplified. Initially, 100 pmol of the Eurofins Genomics primer (forward/ reverse) was diluted with Ambion water to 10 pmol. A 1:1 mixture of the diluted forward and reverse primers was prepared with 1.25 µl of each primer per sample. Following, the primer mix was combined with 2x Platinum SuperFi PCR master mix (12.5 µl per sample) and pipetted on a Bio-Rad Hard-Shell 96-Well PCR plate. 10 µl of the extracts was used as sample input, added to the PCR plate, and mixed by gently pipetting up and down. The PCR plate was sealed with an Easy Pierce Heat sealing foil and the ALPS 50 V Microplate heat sealer preheated to 70 °C. The following PCR cycles were run through in the Biometra TAdvanced PCR cycler:

98°C – 2 min
98°C – 10 sec
55°C – 10 sec } *x 30 cycles*
72°C – 30 sec
72°C – 5 min
4°C – ∞

Negative PCR controls with Ambion water were included to identify contamination during the PCR.

The PCR cleanup was carried out with the AMPure XP beads which were designed for DNA as sample material. The method was based on the Solid Phase Reversible Immobilization (SPRI) technology by Beckman Coulter, Inc. (Indianapolis, USA). 20 µl of AMPure XP beads per sample was combined with the PCR output (25 µl) in a bead/sample ratio of 0.8 and gently mixed by pipetting up and down. The ratio of beads to samples was important as it affected the size selectivity of the beads. A ratio of 0.8 targeted a median DNA size between 300 and 350 base pairs (Beckman Coulter, 2022). During an incubation time of five minutes, the paramagnetic beads selectively bound nucleic acids. Subsequently, an external magnet was applied, causing pellets to form, and immobilizing the nucleic acid along with the beads. The supernatant with unincorporated dNTPs, primers, salts, and other contaminants was removed and the pellets were washed

twice with 80% ethanol. The pellets were then allowed to dry, but not to the point of cracking. Finally, the purified DNA was eluted with 27 µl Milli-Q water by gentle shaking for 2 minutes and transferred into a new Bio-Rad Hard-Shell 96-Well PCR plate. The workflow was conducted with the Hamilton Microlab STAR, the programming was developed by Jan-Philipp Mengel (Institute of Medical Microbiology, University Giessen).

2.3.7 DNA quantitation of the V4 PCR cleanup

The DNA concentration in the samples after the 16S rRNA V4 PCR cleanup was quantified with the Quant-iT PicoGreen dsDNA assay kit and Qubit 1x dsDNA HS assay kit. The quantitation workflow was conducted exactly as described in section 2.3.5. An Excel template for DNA strands with a length of 365 base pairs was used for the analysis of the PicoGreen measurement results.

2.3.8 Index PCR and cleanup

During the index PCR, specific barcodes of a known sequence were attached to the 16S rRNA V4 amplicons to distinguish the different samples in downstream analysis. The same index primers were used in all three batches to ensure comparability (Allaband *et al.*, 2019). At first, NEBNext Ultra II Q5 Master Mix (12,5 µl per sample) was combined with Mili-Q water (2.5 µl per sample) and pipetted on a Bio-Rad Hard-Shell 96-Well PCR plate. 5 µl PCR input and 2.5 µl customed DNA oligo primers (forward/ reverse) were added in each well. The primer tubes were automatically opened and closed by the Hamilton LabElite decapper. The following PCR cycles were run through in the Biometra TAdvanced PCR cycler:

98°C – 30 sec
98°C – 10 sec
65°C – 75 sec] *x 8 cycles*
65°C – 5 min
4°C – ∞

The PCR Cleanup was carried out with the AMPure XP beads as described in section 2.3.6. 25 µl of AMPure XP beads per sample were combined with 22.5 µl of the libraries in a bead/sample ratio of 0.9 which targeted a median DNA size between 300 and 350 base pairs (Beckman Coulter, 2022). The workflow was conducted with the Hamilton Microlab STAR, the programming was developed by Jan-Philipp Mengel (Institute of Medical Microbiology, University Giessen).

2.3.9 DNA quantitation of the index PCR cleanup

The DNA concentration in the libraries was quantified after the index PCR clean up with the Quant-iT PicoGreen dsDNA assay kit and the Qubit 1x dsDNA HS assay kit. The quantitation workflow was conducted as described in section 2.3.5. An Excel template for DNA strands with a length of 438 base pairs was used for the analysis of the PicoGreen measurement results.

2.3.10 Library quantitation and quality control

For quality control the libraries were measured with the Fragment Analyzer and the dsDNA 935 reagent kit (1 bp – 1,500 bp) from Agilent Technologies, which is a method for automated parallel capillary electrophoresis. Firstly, gel and dye as well as 1x conditioning solution were prepared according to producer`s protocol and loaded into the Fragment Analyzer. The plates filled with 1x inlet buffer and with the marker were placed in the designated location. Then, 2 µl of the libraries was mixed with 22 µl of diluent buffer (1x TE) in a Twin.tec PCR plate 96 and finally one drop of mineral oil was added into each well for best preservation. During the measurement, a gel matrix with fluorescent dye was filled into each of the 48 parallel capillaries, a voltage was applied, and the DNA fragments were separated according to size. A light source stimulated the fluorescence emission, which was then measured together with the required running time. The analysis of the measurement results was conducted with the PROSize 3.0 software.

2.3.11 Normalization and pooling

After reviewing the results from the DNA quantitation and quality control, a pool concentration of 4 nM was targeted in all three batches. The pipetting steps were carried out with the Hamilton Microlab STAR, the programming was developed by Jan-Philipp Mengel (Institute of Medical Microbiology, University Giessen). The concentration values from the PicoGreen and the Qubit measurements were used for the pooling. At first, the libraries were diluted to 4 nM with Mili-Q water. Then 5 µl of each library was added to the pool in a Screw Cap micro tube. For samples with a concentration below the pool concentration of 4 nM, 5 µl was pipetted directly to the pool.

2.3.12 Pool quantitation and quality control

Pool quantitation and quality control was conducted with the Agilent Bioanalyzer and additionally by qPCR.

The finished pool was analyzed by automated electrophoresis with the 2100 Bioanalyzer instrument which provided information about library concentration and fragment size. The gel-dye mix was prepared and transferred into the corresponding wells of the DNA microfluidic chip according to producer instructions. DNA marker, ladder and libraries were added, the chip was vortexed for 1 minute and loaded into the 2100 Bioanalyzer instrument for electrophoresis.

Quality control by quantitative PCR (qPCR) was carried out with the NEBNext Library Quant kit for Illumina. Contrary to the electrophoresis, the qPCR as an amplification-based method only measured DNA which contained both adaptor sequences, thus only detected the library molecules that could be sequenced. In turn, qPCR didn`t provide information on the fragment size of the libraries. The qPCR was conducted according to producer`s instructions. 1x Library Quant dilution buffer was prepared and used for 1:1.000 and 1:10.000 dilutions of the pooled libraries. 4 µl of the diluted libraries or DNA standards was combined with 16 µl of Library Quant master mix in a MicroAmp Fast 96-Well plate. For best results, the libraries and the standards ran in triplicates. A MicroAmp-Optical adhesive film was affixed to the plate and the optical measurement happened in the StepOnePlus Real-Time PCR device.

2.3.13 Short-read sequencing with Illumina MiSeq

The sequencing on the Illumina MiSeq system was conducted on a Nano Flow cell with the 500-cycles MiSeq reagent kit and according to Protocol A (Standard Normalization Method) from the Illumina Denature and Dilute Libraries Guide (Document # 15039740v10). At first, 5 µl of the pooled libraries was combined with 5 µl the freshly diluted 0.2 M NaOH in a DNA low bind tube and incubated for 5 minutes at room temperature to denature the DNA strands. Then, 5 µl of Tris HCl (pH 7.0) and 985 µl of HT1 were added as buffer. The final load of 600 µl was prepared consisting of 150 µl of the denatured 20 pM library, 405 µl of HT1 and 45 µl 20 pM PhiX Control. PhiX is a control library which originated from a well-characterized bacteriophage genome with a balanced base composition of approximately 45% GC and 55% AT. This had the advantage that the fluorescent signal was better balanced during the sequencing cycles, thus improving overall run quality, especially for low diversity libraries such as amplicon libraries (Illumina, 2023).

To ensure denaturation, which was essential for the clustering process during sequencing, the finished load was heat-denatured at 95 °C for 3 minutes and instantly placed on ice. Finally, the denatured libraries were loaded in the reagent cartridge, the flow cell was inserted and the sequencing process initiated.

2.3.14 Long-read sequencing with Nanopore

For Nanopore long-read sequencing, library preparation was carried out according to producer`s instruction. Fluid from the same extracts prepared for Illumina sequencing was used. The reagents were kept on ice during the workflow. Three controls with nuclease free water were processed with the 21 samples to identify contamination during the workflow. First the concentrations of the extracts were measured with the Qubit High Sensitivity Kit as described in 2.3.5. 10 ng DNA was prepared in 15 ml Ambion water and then combined with the Taq 2x master mix. 10 µl of a specific 16S barcode was added to every sample and mixed thoroughly. The following PCR cycles were run through in the Biometra TAdvanced PCR cycler:

95°C – 1 min
95°C – 20 sec
55°C – 30 sec } *x 30 cycles*
65°C – 2 min
4°C – ∞

Afterwards, the PCR output was pooled, cleaned up with 30 µl AMPure XP beads (cf. section 2.3.6), eluted with Tris-HCl and again quantified with the Qubit High Sensitivity kit. The pool was combined with flush buffer, sequencing tether, and loading beads and transferred into the Flongle flow cell which was then started in the MinION device.

2.4 Bioinformatic methods

The bioinformatic analysis was conducted by Markus Weigel (Institute of Medical Microbiology, University Giessen). The raw sequencing data sets of the manual Zymo extracts and the automatic eMAG extracts were run through the bioinformatic pipelines both together and separately to check if the OTU assignments differed. Microbiome analysis was performed with Mothur (Kozich *et al.*, 2013). Paired-end reads were joined, primer regions removed and filtered for the expected amplicon length of 253 nt ± 10 nt. Sequences that contained ambiguous nucleotides were excluded. Joined paired-end reads were aligned with the SILVA ribosomal RNA gene database (Quast *et al.*, 2013) and trimmed so that they contained only the hypervariable region V4. The clustering was conducted with a similarity threshold of 97%. After chimera removal using VSEARCH (Rognes *et al.*, 2016), operative taxonomic units (OTUs) were determined and classified with the SILVA ribosomal RNA gene database. For further analysis, all samples were subsampled to 1.000 reads. α-diversity indices and principal coordinate analysis (PCoA) of the Bray–Curtis dissimilarity were created with Mothur. OTUs that could not be classified to the genus level were further analyzed by BLASTn (Altschul *et al.*, 1997) with the 16S ribosomal RNA database from the NCBI RefSeq Targeted Loci Project (NCBI RefSeq Targeted Loci Project, 2019).

2.5 Statistics

An explorative data analysis was conducted. The significance level for the statistical analyses was set at 0.05 and for the prerequisite test (such as testing for Gaussian distribution) at 0.1. The logarithmic transformations were performed with the logarithm base two. Bonferroni correction was applied to correct for multiple related comparisons within a section.

The statistical analysis was carried out and the figures were created with the software GraphPad Prism (Version 9.5.0), apart from some exceptions, which are noted in the subheadings of the respective graphs. The statistical tests were chosen on the basis of the book “Intuitive Biostatistive” by Harvey Motulsky (2017).

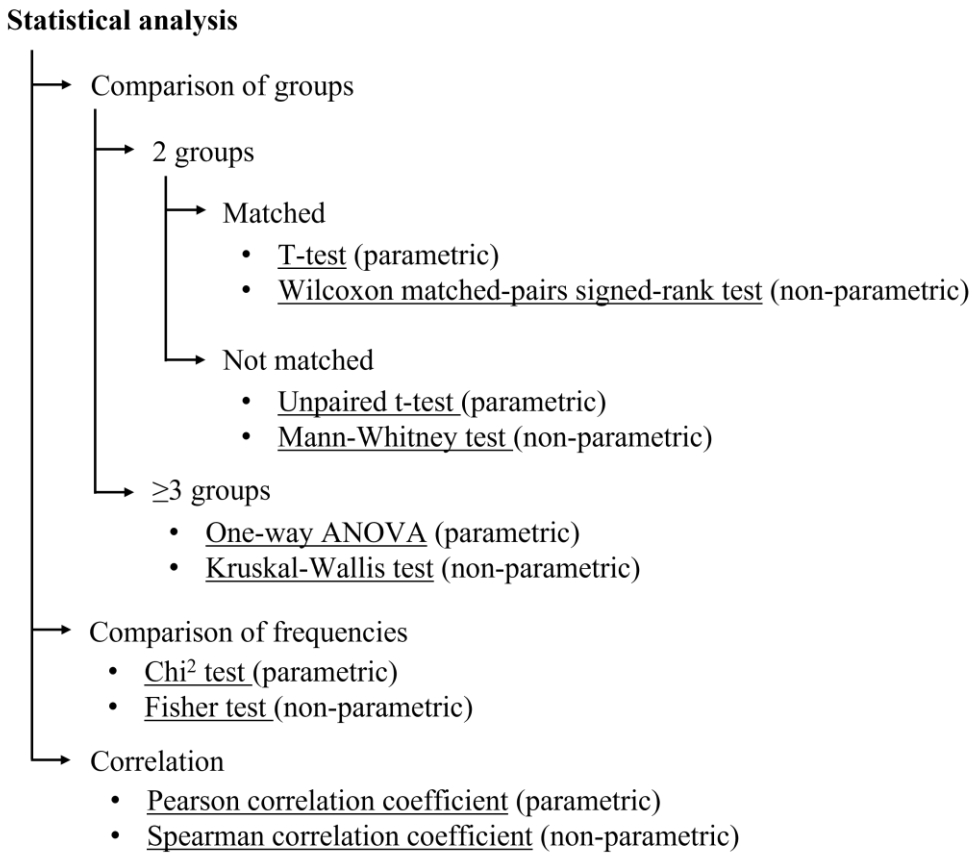


Figure 3: Statistical tests for data analysis.

2.6 Contamination assessment and evaluation of the extraction methods

For the analysis of the Illumina short-read sequencing results, the not subsampled OTU tables were used for the assessment of the read counts. For further analyses, the subsampled data was consulted to allow a better comparison between different microbiome profiles.

The analysis was started with the description of the Illumina short-read sequencing, including sequenced reads and read count distributions of the Zymo and eMAG extracts. Following, the sequencing results from the blank controls were assessed thoroughly. The high abundant genera in the controls were compared with the top five BLAST results to check for concordance. For the analyses of the controls, only the not-subsampled OTU table was considered to have a solid overview of the entire amount of contamination. Based on the analyses of the controls, contamination OTUs were determined for the Zymo and the eMAG extracts separately. Each genus was examined solely and contamination genera were determined based on the following four non-mandatory criteria:

1. The respective genus was abundant in the controls.
2. The respective genus was present mainly either in the Zymo extracts or in the eMAG extracts.
3. The respective genus was evenly distributed in the Zymo or eMAG extracts.
4. The respective genus was known as a common contaminant in microbiome studies (Salter *et al.*, 2014).

OTUs defined as contamination were excluded from further analyses.

Subsequently, the sequencing results and microbiome profiles of the manual Zymo and the automated eMAG extracts were compared. Wet lab data, read counts and the abundances of the taxa were examined. α- and β-diversity were used to assess patterns within and between the microbiomes of different samples. Three metrics for α-diversity were used: observed richness, Chao1 and Shannon diversity index. α-diversity was estimated for samples with over 1,000 read counts. The Bray Curtis index was used to estimate β-diversity.

The microbiome profiles were then assigned to three different categories:

1. Mono-Microbial: One genus was three times as high as any other genus *and* had a relative abundance over 40 percent.
2. Poly-Microbial: Two to three genera were each at least twice as high as any other genus *and* the relative abundance accounted together for more than 50 percent.
3. Multi-Microbial: Any other microbiome profile.

2.7 Comparison of culture results and NGS data

Wet lab parameters and diversity indices were analyzed in relation to the culture results. Subsequently, it was evaluated whether and how the results from culture-based routine diagnostics and Illumina short-read sequencing overlapped. Therefore, two methods were applied: firstly, a table matching the species reported in culture with their corresponding genus, similar to an analysis conducted by Zachariah *et al.* (2018). Secondly, the individual culture results were allocated to their respective microbiome profiles. Afterwards, the Nanopore microbiome profiles of the selected samples were examined at species level. Finally, the results from culture, Illumina and Nanopore sequencing were juxtaposed and analyzed for each sample individually.

3 Results

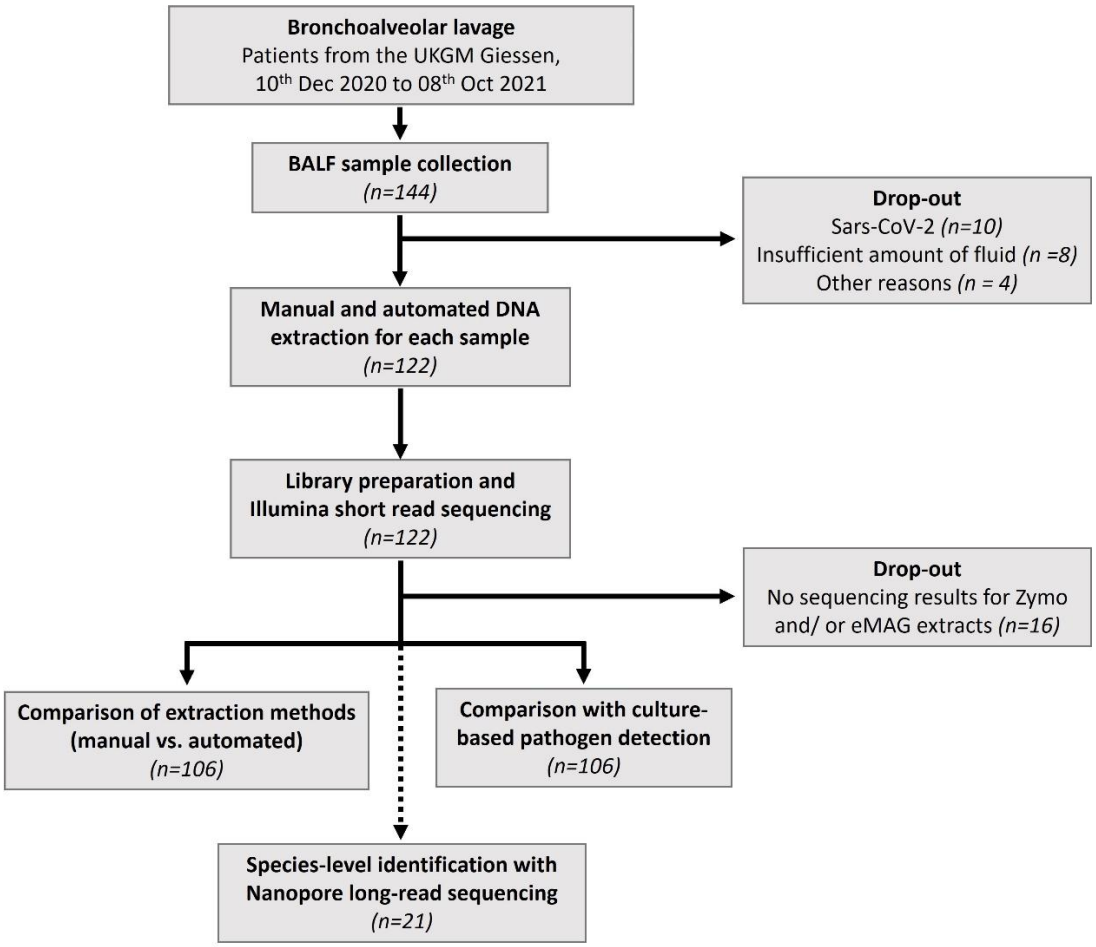


Figure 4: Study flow chart.

In this section, firstly the results of the culture-based routine diagnostics will be presented, and the cultured bacteria will be evaluated according to their potential pathogenicity. Next, the Illumina sequencing data will be depicted, the blank controls analyzed, and the contamination genera identified and excluded from further analyses. Then, the results obtained from the manual Zymo and automated eMAG extraction will be compared. Subsequently, the results from the culture-based routine diagnostics and Illumina short-read sequencing will be opposed and compared. Lastly, selected samples which were sequenced with the Nanopore long-read technique will be analyzed at species level.

3.1 Sample collection and ethical matters

144 BALF samples were collected from December 10th, 2020, until October 8th, 2021, from patients who underwent bronchoalveolar lavage at the UKGM Giessen (Germany). The samples passed through the routine diagnostic procedure and were stored at 4 °C. A median of 1 day (IQR, 1 - 4) after bronchoalveolar lavage, the samples were stored at -80 °C for further laboratory NGS workflow.

The project was reviewed by the ethics committee of the Faculty of Medicine of the Justus-Liebig University Giessen (AZ 158/20). Sample collection was conducted as part of clinical routine diagnostics, with no samples obtained exclusively for research purposes. The surplus sample material, which was not required for clinical diagnostic purposes was appointed to be used for methodological improvement and quality control.

3.2 Culture-based routine diagnostics

3.2.1 Culture-based microbial identification

The BALF samples were cultivated on different media in the diagnostic Institute of the Medical Microbiology. The number of cultivated species in one sample ranged from 0 to 4, with a median of 1 (IQR, 0 – 2; cf. Figure 30 in the appendix). In 39 samples (36.8 %), no bacteria were successfully cultured. In total, 28 species were identified in culture, counting the report “pharyngeal flora” as one species. The three most frequent taxa reported in the culture results were α-hemolytic *Streptococcus*, saprophytic *Neisseria*, and *St. aureus*. A complete overview of the cultured species and the number of samples in which they were detected is shown in Table 8. The results of the respiratory PCR panel are provided in the appendix (Appendix 2: Culture-based microbial identification – further results)

3.2.2 Criteria for pathogenicity in cultured samples

The bacteria which were cultured and listed in the diagnostic reports were evaluated for potential pathogenicity, based on the S3 guidelines for community- and hospital-acquired pneumonia and additional studies when the guidelines did not provide information on the given species (q.v. Table 8). Furthermore, microscopic features (granulocytes and epithelial cells) and quantitative information from the culture report were taken into account. A complete list of the pathogenicity evaluation for every sample is provided in the appendix (Appendix 3: Pathogenicity evaluation for the results of culture-based routine diagnostics).

Species	No. of reports	Pathogenicity	Reference for pathogenicity evaluation
α-hemolytic *Streptococcus*	35	Not pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Saprophytic *Neisseria*	12	Not pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
St. aureus	11	Potentially pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Koagulase-neg. *Staphylococcus*	9	Not pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Prevotella melaninogenica	5	Not pathogenic	(Wypych, Wickramasinghe and Marsland, 2019; Natalini, Singh and Segal, 2023)
Ps. aeruginosa	5	Potentially pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Rothia mucilaginosa	5	Not pathogenic	(Das *et al.*, 2021; Xie *et al.*, 2021)
Enterococcus faecium	4	Not pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Staphylococcus epidermidis	4	Not pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Rothia sp.	3	Not pathogenic	(Das *et al.*, 2021; Xie *et al.*, 2021)
Ste. maltophilia	3	Potentially pathogenic	S3-guideline for HAP (Dalhoff *et al.*, 2017)
Streptococcus mitis	3	Not pathogenic	S3-guideline for CAP and HAP(Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Citrobacter koseri	2	Potentially pathogenic	(Yao *et al.*, 2021)
Enterococcus faecalis	2	Not pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
E. coli	2	Potentially pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Neisseria flava	2	Not pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)

Pharyngeal flora	2	Not pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Actinomyces odontolyticus	1	Not pathogenic	(Könönen and Wade, 2015)
Bordetella bronchiseptica	1	Potentially pathogenic	(Gujju *et al.*, 2021)
H. influenzae	1	Potentially pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Haemophilus parahaemolyticus	1	Not pathogenic	S3-guideline for CAP (Ewig *et al.*, 2021)
Lactobacillus reuteri	1	Not pathogenic	(Zachariah *et al.*, 2018)
Lactobacillus salivarius	1	Not pathogenic	(Zachariah *et al.*, 2018)
Serratia marcescens	1	Potentially pathogenic	S3-guidelines for HAP (Dalhoff *et al.*, 2017)
Streptococcus parasanguinis	1	Not pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Str. pneumoniae	1	Potentially pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Streptococcus Salivarius	1	Not pathogenic	S3-guideline for CAP and HAP (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021)
Veillonella parvula	1	Not pathogenic	(Wypych, Wickramasinghe and Marsland, 2019; Das *et al.*, 2021; Natalini, Singh and Segal, 2023)

Table 8: Species cultivated in culture-based routine diagnostics.

The species are listed with their number of diagnostic reports, the potential pathogenicity and the reference consulted for assessment of potential pathogenicity. "Not pathogenic" specifically meant that these taxa were contextualized as members of the physiological lower respiratory tract microbiota.

9 of 28 cultured species were categorized as potentially pathogenic. The other species were most likely members of the physiological lung flora. As can be seen in Figure 5, in 26 samples at least one potential pathogen was reported. Sample 126 was the only sample in which two potential pathogens were reported (*St. aureus* and *Ser. marcescens*). In 41 samples only bacteria considered part of the physiological flora were cultivated. In 39 samples, the culture report was blank.

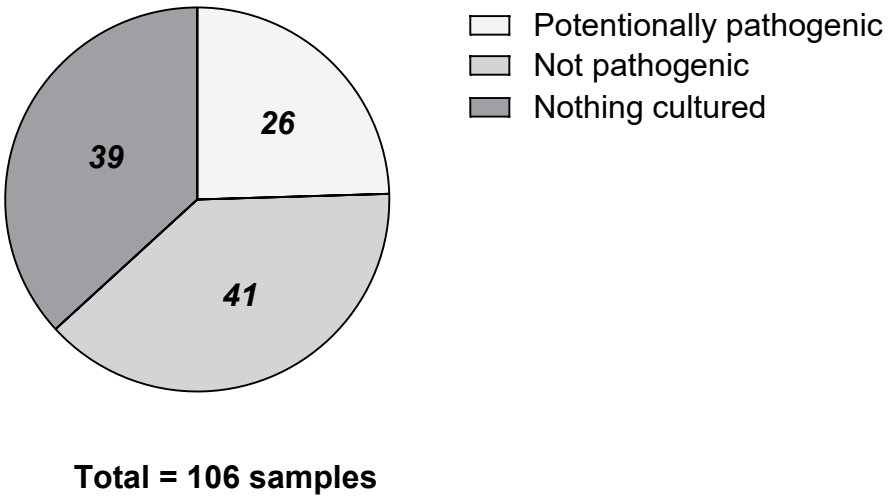


Figure 5: Results of culture-based diagnostics regarding possible pathogenicity.

3.3 Illumina short-read sequencing

In this section, firstly the output of the Illumina short-read sequencing will be presented and analyzed (3.3.1). Subsequently, the data of the blank controls were evaluated (3.3.2), which laid the foundation for the decision on which genera represented potential contaminants (3.3.3). The results of the Zymo and the eMAG methods were compared according to different criteria (3.3.4). Lastly, each microbiome profile was assigned to one of three categories, formed by taxonomic composition of the profiles (3.3.5). This step was conducted to enable the later comparison of the NGS results with the culture results (3.4) and to extend the juxtaposition of the Zymo and the eMAG results (3.3.5).

3.3.1 Wet lab and 16S rRNA sequencing with Illumina

The sequencing run of the manual Zymo extracts generated a total of 694,770 read counts with a median of 5,461 reads per sample (IQR, 3,383 – 8,036). The sequencing run of the automated eMAG extracts produced a total of 685,465 read counts with a median of 5,314 read counts per sample (IQR, 3,323 – 8,604). In Figure 6: Sequenced reads of the manual Zymo and the automated eMAG extracts., the read counts of the sequenced extracts are

plotted for both Zymo and eMAG[2]. The values were distributed lognormally, which has been confirmed by the D`Agostino and Pearson test for lognormal distribution. The results of DNA quantitation at several points in the workflow as well as the Agilent Bioanalyzer electropherograms for pool quantitation and quality control are provided in Appendix 4: Wet lab and 16S rRNA-sequencing with Illumina – further results.

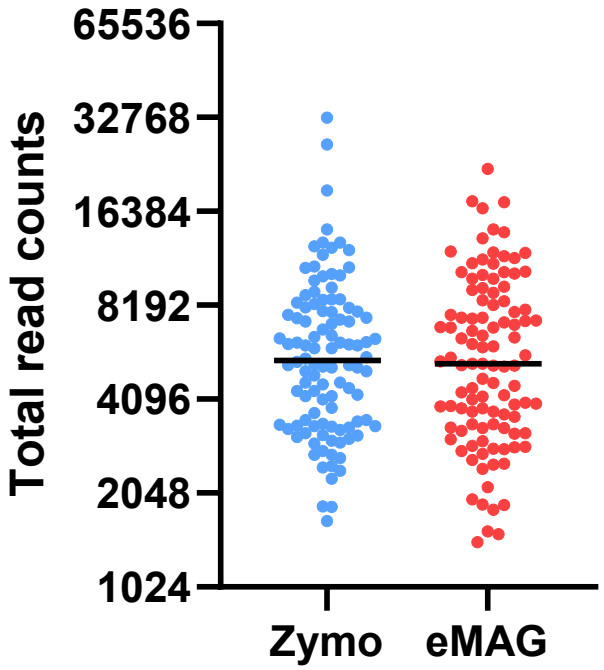


Figure 6: Sequenced reads of the manual Zymo and the automated eMAG extracts.

The values were plotted with their medians on a logarithmic scale (base two) and showed a lognormal distribution (D`Agostino and Pearson test for lognormal distribution).

Figure 7 plots the read counts of all samples in their numerical order. As can be seen, no substantial difference was discernible between batch one (samples one – 51) and batch two (samples 52 – 99). However, batch three (samples 100 – 144) visually appeared differently, with overall lower values both in the Zymo and in the eMAG extracts. The Kruskal-Wallis test was applied for the batch comparison and showed significantly varying medians ($p < 0.001$ for both Zymo and eMAG). A scatter plot, which shows the

[2] The not-subsampled OTU table was used for all analyses of read counts. All other analyses were conducted with the subsampled data, unless otherwise indicated.

read counts of the three batches plotted separately with their medians is provided in Appendix 4: Wet lab and 16S rRNA-sequencing with Illumina – further results.

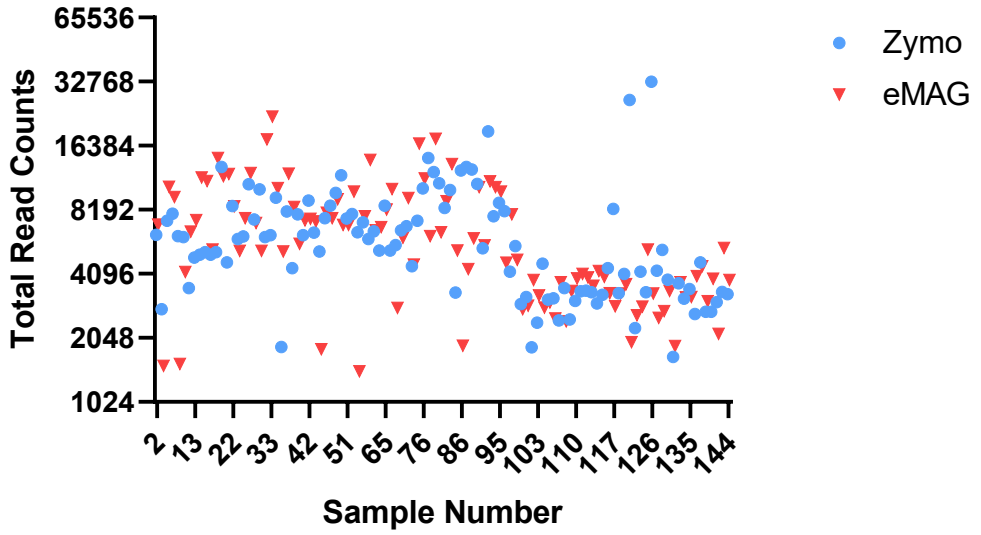


Figure 7: Aligned sequencing reads of the 106 Zymo and eMAG samples.

The blue dots representing the manual Zymo extracts and red triangles for the automated eMAG extracts. The values were plotted on a logarithmic scale (base two).

Question	Data	Performed Test	Result
Were there significant differences between the three batches in the Zymo samples?	Read counts of all Zymo extracts, grouped by the three different batches of laboratory workflow	Kruskal-Wallis test (non-parametric)	***P < 0.0001*** (medians of the three batches varied significantly)
Were there significant differences between the three batches in the eMAG samples?	Read counts of all eMAG extracts, grouped by the three different batches of laboratory workflow	Kruskal-Wallis test (non-parametric)	***P < 0.0001*** (medians of the three batches varied significantly)

Table 9: Statistical tests performed in section 3.3.1.

The significance level was defined as $\alpha = 0.05$ for the family of comparisons. According to the Bonferroni correction, the result was declared statistically significant when its P value was less than 0.025.

3.3.2 Evaluation of the blank controls

Control Type	Number	Median Read counts (IQR)	Most abundant genus	Most abundant family
Zymo Shield Control	9	7,506 (2,850 – 11,673)	*Kocuria*	*Micrococcaceae*
Zymo Kit Control	5	733 (51.5 – 6,617)	*Aquabacterium*	*Comamonadaceae*
eMAG Kit Control	4	4,139 (884 – 7,468)	*Rhodobacteraceae unclassified*	*Rhodobacteraceae*
PCR Control (Zymo)	6	704 (111.3 – 1,096)	*Escherichia-Shigella*	*Enterobacteriaceae*
PCR Control (eMAG)	6	455.5 (50 – 1,425)	*Cutibacterium*	*Propionibacteriaceae*
Bronchoscopy Control	6	1630 (102.5 – 3,382)	*Burkholderia-Caballeronia-Paraburkholderia*	*Burkholderiaceae*

Table 10: Overview of all types of blank controls.

36 blank controls were sequenced and analyzed in detail to make a sound decision about which taxa represented potential contaminants. Table 10 shows a summary of all controls. A total of 118,218 read counts was sequenced in the controls. The number of read counts ranged from 38 to 16,584 per control, with a median of 1,371 (IQR, 165 – 4,993). 21 out of 36 controls exhibited over 1,000 read counts. The control type with the highest median number of read counts was the Zymo Shield control. The dominant taxa varied between the different types of control, which is also depicted in the box plot of Figure 8. A detailed description and analysis of the controls is supplied in Appendix 5: Detailed analysis of the controls.

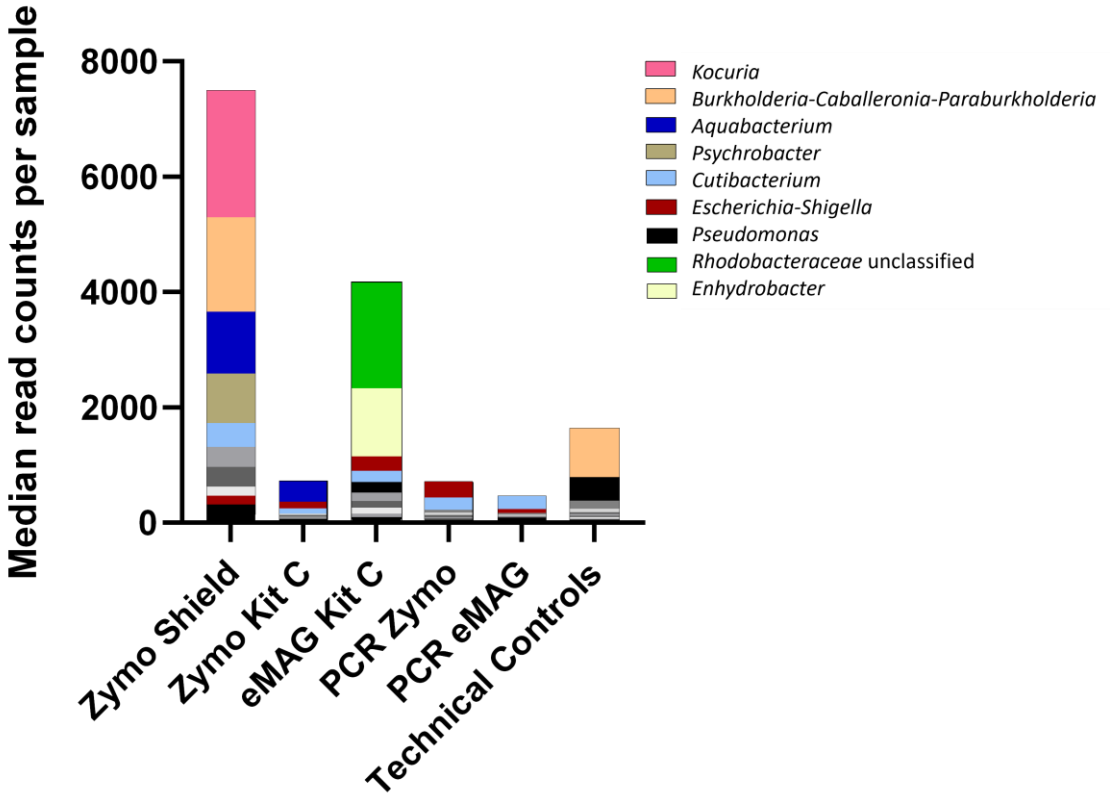


Figure 8: Median sequencing reads per control and dominant taxa in the different control types.

3.3.3 Contamination assessment

The identification of contamination genera was conducted separately for the manual Zymo extracts and the automated eMAG extracts following four non-mandatory criteria, as described in the methods (2.6). Table 11 lists the genera declared as contamination and therefore excluded from further analysis.

Zymo/ eMAG	*Genus*	*Family*	*Dominant OTU*	*Total Read Counts in all controls*	*Probable Source*
Zymo	*Acinetobacter*	*Moraxellaceae*	OTU0038	620	Zymo Shield, Zymo Extraction Kit
Zymo	*Acinetobacter*	*Moraxellaceae*	OTU0087	164	Zymo Shield
Zymo	*Alphaproteobacteria unclassified*	*Alphaproteobacteria unclassified*	OTU0020	2,297	Zymo Shield
Zymo	*Aquabacterium*	*Comamonadaceae*	OTU0005	14,450	Zymo Shield, Zymo Extraction Kit
Zymo	*Burkholderia-Caballeronia-Paraburkholderia*	*Burkholderiaceae*	OTU0003	11,167	Shield
Zymo	*Cupriavidus*	*Burkholderiaceae*	OTU0045	702	Zymo Shield
Zymo + eMAG	*Cutibacterium*	*Propionibacteriaceae*	OTU0009	8,900	All
eMAG	*Enhydrobacter*	*Moraxellaceae*	OTU0019	4,470	eMAG Extraction Kit
eMAG	*Escherichia-Shigella*	*Enterobacteriaceae*	OTU0001	6,595	eMAG Extraction Kit
Zymo	*Halomonas*	*Halomonadaceae*	OTU0060	162	Zymo Shield
Zymo	*Herbaspirillum*	*Oxalobacteraceae*	OTU0043	788	Zymo Shield
Zymo	*Kocuria*	*Micrococcaceae*	OTU0015	15,072	Zymo Shield
Zymo	*Psychrobacter*	*Moraxellaceae*	OTU0021	5,852	Zymo Shield
Zymo	*Ralstonia*	*Burkholderiaceae*	OTU0037	1,581	Zymo Shield, Zymo Extraction Kit
eMAG	*Rhodobacteraceae unclassified*	*Rhodobacteraceae*	OTU0026	7,059	eMAG Extraction Kit
Zymo	*Sphingomonas*	*Sphingomonadaceae*	OTU0061	425	Zymo Shield, Zymo Extraction Kit

Table 11: Genera defined as contamination in the Zymo and eMAG sequencing data.

Only four OTUs were characterized as contamination in the eMAG samples, compared to thirteen OTUs in the Zymo samples. The proportion of the contamination (total contamination read counts / total read counts) was higher in Zymo ($p < 0.0001$), with 30.3 % (210,418 / 694,770) in the manual Zymo extracts and 26.7 % (183,198 / 685,465) in

the automatic eMAG extracts. Noteworthy, in the eMAG extracts, the genus *Escherichia-Shigella* was highly dominant, accounting for 94.0 % of the total eMAG contamination. The contamination in Zymo was more evenly distributed among the different genera, with *Burkholderia-Caballeronia-Paraburkholderia* having the highest abundance (43,15 % of the total Zymo contamination). The Bubble chart (Figure 9) visualizes this feature, showing that *Escherichia-Shigella* was highly prevalent among the eMAG extracts and *Burkholderia-Caballeronia-Paraburkholderia* was most dominant in Zymo. *Cutibacterium* was highly prevalent in the controls, as reflected by a larger bubble, however, did not account for a very large portion of the read counts in the eMAG and Zymo extracts. *Pseudomonas* occurred in the Zymo as well as in the eMAG extracts, though partially also in the controls and was also listed as a constituent contaminant in laboratory contamination (Salter *et al.*, 2014). However, *Pseudomonas* had not been classified as contamination in order to not overlook potential infections. Specimens with *Pseudomonas* had to be carefully and individually examined in further analyses to assess whether actual colonization or contamination was more likely.

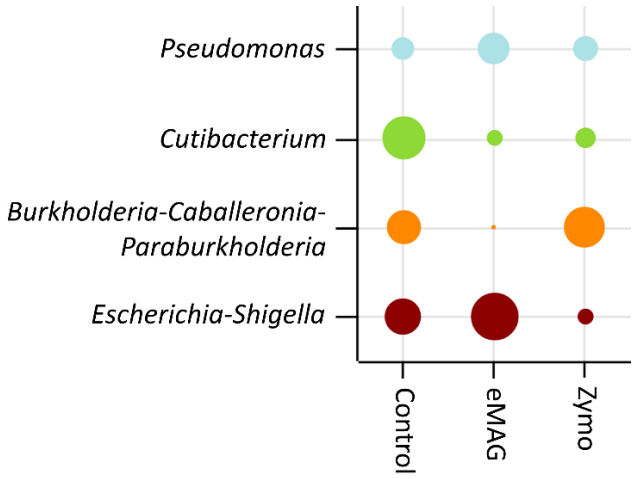


Figure 9: Bubble chart plotting the mean relative abundances of all samples.

Relative abundances were shown for the genera Escherichia-Shigella, Burkholderia-Caballeronia-Paraburkholderia, Cutibacterium and Pseudomonas (visualization: Markus Weigel).

Rhodobacteraceae unclassified and *Enhydrobacter* were the two predominant genera in the eMAG Kit controls (cf. Figure 8), wherefore they were determined as contaminants. However, they were not very abundant in the samples.

The top BLAST results of the contaminant OTUs mostly matched their assigned genus. For OTU0001, the assigned genus was *Escherichia-Shigella* and the top BLAST results were from *E. coli,* a known environmental contaminant. The assigned genus for OTU0003 was *Burkholderia-Caballeronia-Paraburkholderia* and the top BLAST result was *Paraburkholderia fungorum*. For OTU0019, the assigned genus (*Enhydrobacter*) wasn`t concordant with the top BLAST species (*Moraxella osloensis*). For OTU0026, the taxonomic assignment just reached the family level (*Rhodobacteraceae unclassified*), the BLAST result suggested *Paracoccus rhizosphaerae* as sequenced species. *Paracoccus* is also listed as a typical contaminant genus in blank controls (Salter *et al.*, 2014).

Question	**Data**	**Performed Test**	**Result**
Was the proportion of contamination read counts different in Zymo and in eMAG?	Contingency table with contamination read counts	Chi-square with Yates` correction	***P < 0.0001*** (proportion of read counts varied statistically significant)

Table 12: Statistical test performed in section 3.3.3.

The significance level was defined as α = 0.05.

3.3.4 Comparison of the manual Zymo and automated eMAG extraction

The manual Zymo and the automated eMAG extraction methods were compared in terms of read counts, DNA yield, sequenced genera, and diversity indices [3]. The overall number of read counts did not differ substantially between the Zymo and the eMAG extraction method, as shown in Figure 6. This observation could be statistically affirmed by the

[3] As described in the methods, the raw sequencing data of the manual Zymo extracts and the automated eMAG extracts were run through the bioinformatic pipelines both together and separately. The OTU assignments of the two bioinformatic analyses did not differ substantially from each other. For further presentation and analyses, the joint bioinformatic run was used.

Wilcoxon matched pairs signed rank test (p = 0.299). The Pearson correlation analysis of the logarithm-transformed number of read counts from the Zymo and the eMAG extracts indicated a positive correlation of r = 0.42 (95% CI, 0.25 – 0.56; P < 0.001) between the Zymo and the eMAG read counts of the respective samples (graph provided in Appendix 6: Comparison of manual Zymo and automatic eMAG extraction – further results). The DNA yield was assessed in Figure 10 which shows the DNA concentration of the extracted samples. The Wilcoxon matched pairs signed rank test confirmed that there was no significant difference in DNA yield between Zymo and eMAG (p = 0.968).

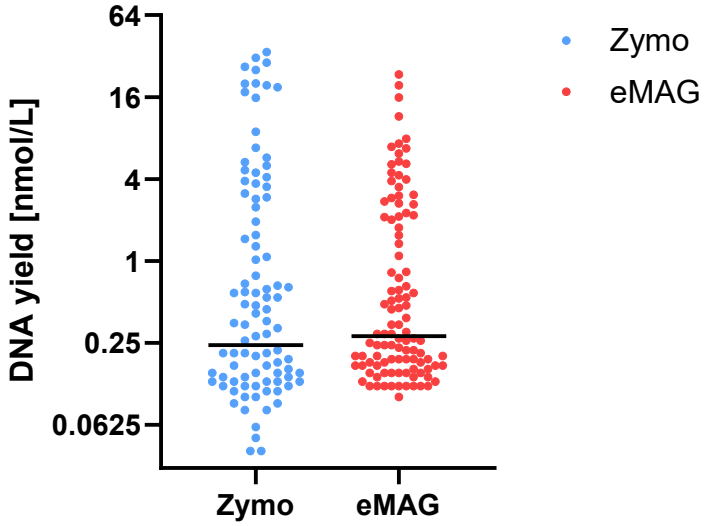


Figure 10: Concentration of the extracted Zymo and eMAG samples.

The values were plotted with their medians on a logarithmic scale (base two).

The Bland-Altman plot was used to further compare the two extraction methods and graphically indicated no systematic difference in read counts in either direction (Figure 11). This observation was confirmed numerically by the relatively small bias of 87.8, calculated from the average difference between the Zymo and eMAG read counts. In addition, the Bland-Altman plot shows that the read counts between the two extraction methods diverged further, the larger their absolute values were, with a few extreme deviations in the range of 15,000 read counts and higher.

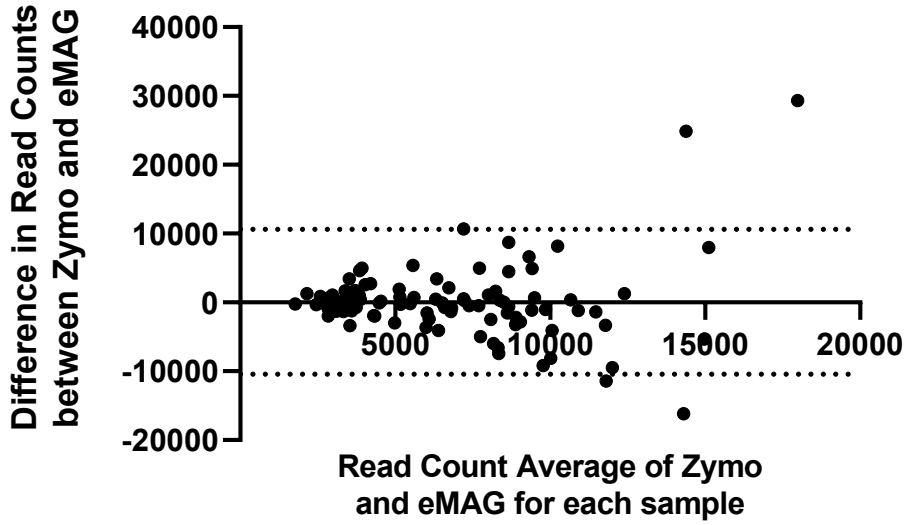


Figure 11: Bland-Altman plot to compare the Zymo and eMAG sequencing reads.

The Y axis represented the differences between the individual Zymo and eMAG values, the x axis the average of those two measurements. The two dotted lines represented the 95% confidence limits of the bias (-10,432 – 10,608). The average bias (average difference between Zymo and eMAG read counts) was 87.8.

Looking at the overall taxonomic composition, Zymo and eMAG differed substantially when contamination was included (Figure 12). In Zymo, the genus *Burkholderia-Caballeronia-Paraburkholderia* was standing out while in the eMAG extracts, *Escherichia-Shigella* was highly abundant. When taking a closer look at the profiles after removing the contamination genera, the two profiles looked more alike. The most abundant genera were *Streptococcus*, *Staphylococcus*, *Pseudomonas*, *Prevotella 7*, *Veillonella* and *Rothia*.

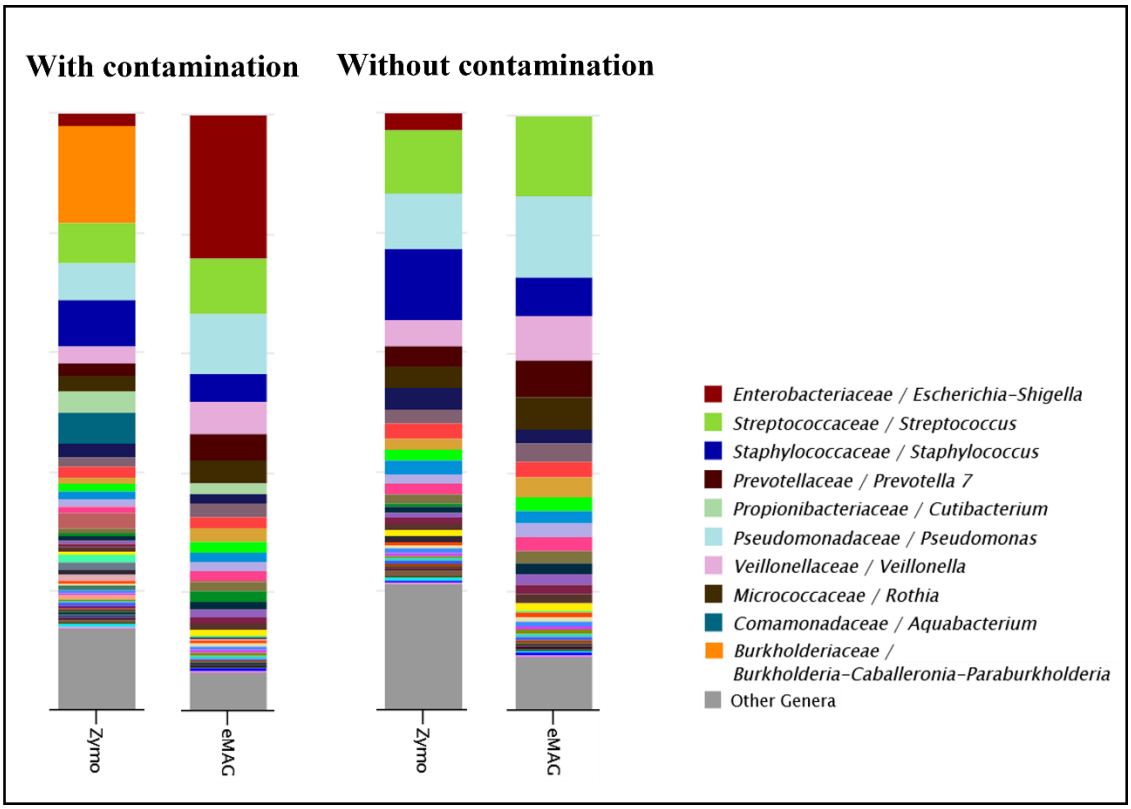


Figure 12: Stacked bar diagram of the joint microbiome profiles.

The microbiome profiles of the Zymo and the eMAG samples are displayed at genus-level both with and without contamination taxa (visualization: Markus Weigel).

A higher number of genera was extracted from the Zymo extracts than from the eMAG extracts (cf. Figure 13). All three α-diversity metrics showed a higher overall diversity in the Zymo extracts (cf. Figure 14), consistent with the fact that α-diversity estimates are based on population richness (the number of taxa). The Wilcoxon matched pairs signed rank test supported statistically that there were significant differences between the Zymo and the eMAG extracts in all three α-diversity indices (all $p < 0.001$).

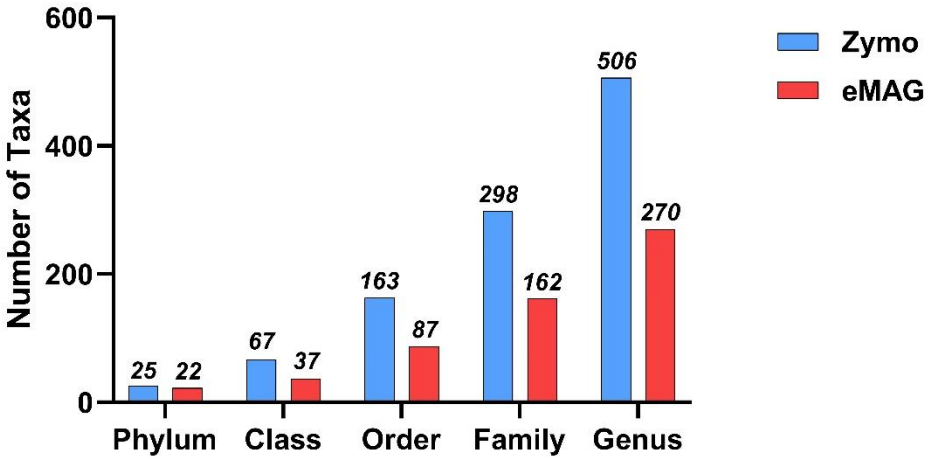


Figure 13: Total number of taxa extracted.

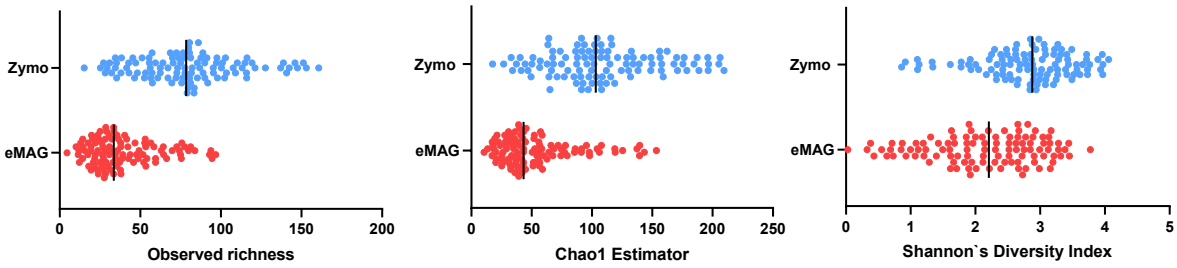


Figure 14: α-diversity estimated by three different metrics.

Observed richness, Chao1 and Shannon diversity indices plotted for the manual Zymo and the automatic eMAG extracts (before exclusion of contaminant genera).

The top six genera which have already been listed, could be retraced in Figure 15. Besides, the Zymo extracts showed lower relative abundances of the given genera overall than the eMAG extracts. However, this observation was not statistically significant in the Wilcoxon matched pairs signed rank test after Bonferroni correction for multiple comparisons ($p = 0.046$).

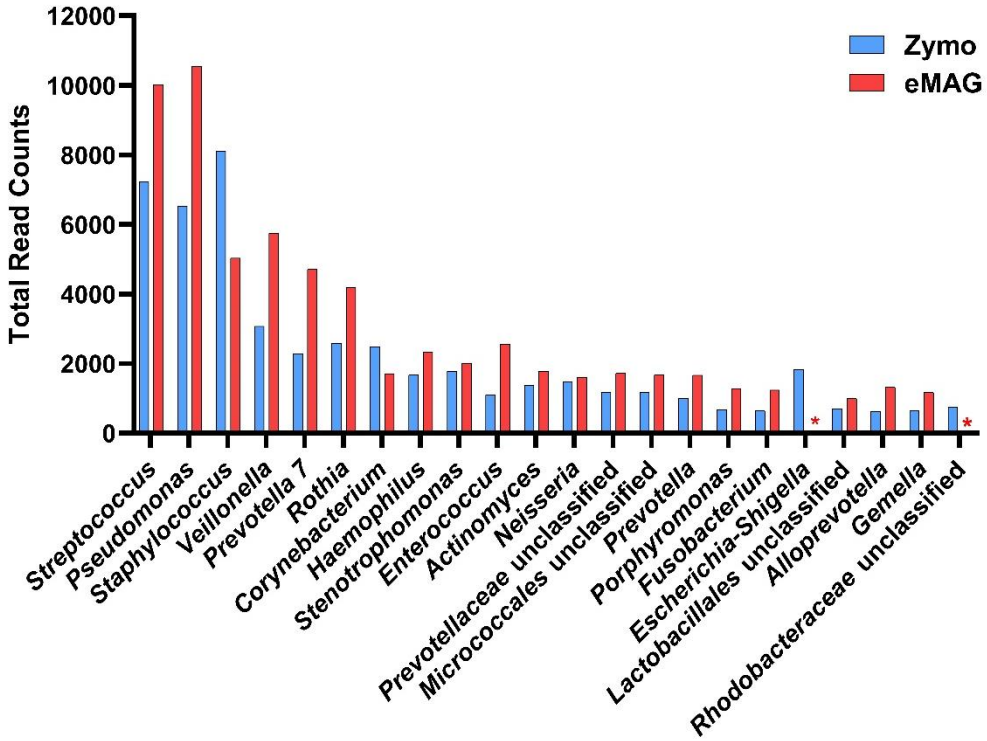


Figure 15: Top 20 sequenced genera for Zymo (blue) and eMAG (red).

Contamination was excluded. Genera with a star (*) were classified as contamination and excluded from analysis.

Figure 16 shows the principal coordinate analysis of the Bray-Curtis coefficient as a metric for β-diversity. The different sample types (Zymo, eMAG and controls) formed overlapping scatter plots. In the area where the Zymo and the eMAG scatter plot touched, the corresponding samples lay very close to each other (e.g., Z24 next to E24, Z70 next to E70, Z102 next to E102, etc.).

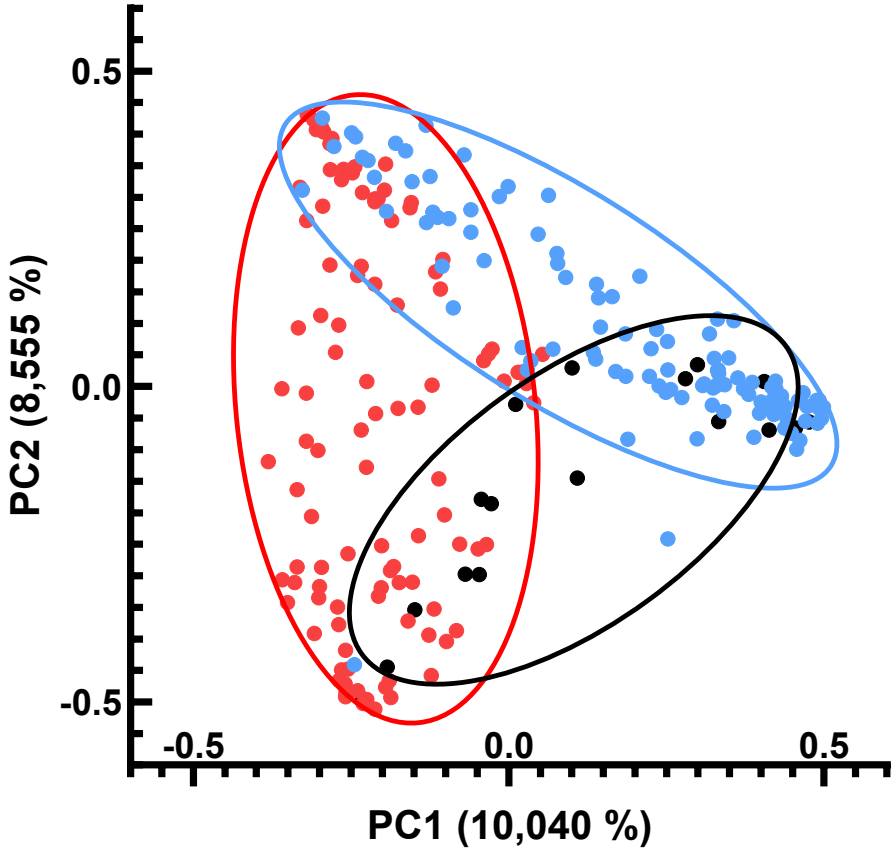


Figure 16: Bray-Curtis dissimilarity coefficient.

The principal coordinate analysis (PCoA) of the Bray-Curtis dissimilarity coefficient shows the β-diversity of the Zymo extracts (blue), the eMAG extracts (red) and the controls (black).

Question	**Data**	**Performed Test**	**Result**
Were there significant differences in the number of read counts between the Zymo and eMAG extracts?	Total absolute read counts of the Zymo and eMAG extracts	Wilcoxon matched pairs signed rank test (two-tailed)	***P = 0.299*** (no significant difference in number of read counts between Zymo and eMAG)

Did the number of read counts correlate between the Zymo and the eMAG extracts?	Logarithm-transformed number of read counts of the Zymo and eMAG extracts	Pearson correlation coefficient	***r = 0.42 (95% CI, 0.25 – 0.56; P < 0.0001)*** (correlation between the numbers of read counts of the Zymo and eMAG extracts)
Were there significant differences in DNA yield between the Zymo and eMAG extracts?	Total absolute read counts of the Zymo and eMAG extracts	Wilcoxon matched pairs signed rank test (two-tailed)	***P = 0.968*** (no significant difference in DNA yield between Zymo and eMAG)
Was there a systematic read count discrepancy between Zymo and eMAG?	Total absolute read counts of the Zymo and eMAG extracts	Bland-Altman plot	***Bias: 87.8 (SD of bias: 5,367; 95% limits of agreement, -10,432 – 10,608)***
Were there significant differences in α-diversity between the Zymo and the eMAG extracts?	Observed richness of the Zymo and the eMAG extracts	Wilcoxon matched pairs signed rank test (two-tailed)	***P < 0.001*** (the observed richness varied significantly between Zymo and eMAG)
Were there significant differences in α-diversity between the Zymo and the eMAG extracts?	Chao1 of the Zymo and the eMAG extracts	Wilcoxon matched pairs signed rank test (two-tailed)	***P < 0.001*** (the Chao1 index varied significantly between the Zymo and eMAG extracts)
Were there significant differences in α-diversity between the Zymo and the eMAG extracts?	Shannon index of the Zymo and the eMAG extracts	Wilcoxon matched pairs signed rank test (two-tailed)	***P < 0.001*** (the Shannon index varied significantly between Zymo and eMAG)
Were there statistically significant differences of the abundances of genera between the Zymo and the eMAG extracts?	Read counts of the individual genera across all samples in the Zymo and the eMAG extracts	Wilcoxon matched pairs signed rank test (two-tailed)	***P = 0.046*** (read counts of the different genera did not vary significantly)

Table 13: Statistical tests performed in section 3.3.4.

The significance level was defined as α = 0.05 for the family of comparisons. According to the Bonferroni correction, the result was declared statistically significant when its P value was less than 0.007.

3.3.5 Categorization of the microbiome profiles

The microbiome profiles of the 106 samples were categorized into three different groups (mono-, poly- and multi-microbial), as described in the methods (2.6).

Figure 17 shows how many samples were classified in each of the three different categories. In the Zymo extracts, fewer mono-microbial, and more multi-microbial profiles were classified than in the eMAG extracts (c.f. Figure 17). However, the difference between Zymo and eMAG was not statistically significant (P = 0.187). A look at Figure 20 showed that 16 of the 32 mono-microbial eMAG profiles (50.0 %) could be attributed to *Pseudomonas*. In comparison, *Pseudomonas* was the dominant genus in only six of the 23 mono-microbial Zymo extracts (26.1 %). Regarding the poly-microbial profiles, *Pseudomonas* was, besides *Streptococcus*, also very abundant in the in the eMAG extracts. As mentioned in the contamination assessment (3.3.3), the genus *Pseudomonas* had to be evaluated with particular care because some of the *Pseudomonas* occurrences found in the microbiome profiles were likely to be contamination.

A chart showing the top genera found in the poly-microbial profiles is provided in Appendix 7: Categorization of the microbiome profiles - further results.

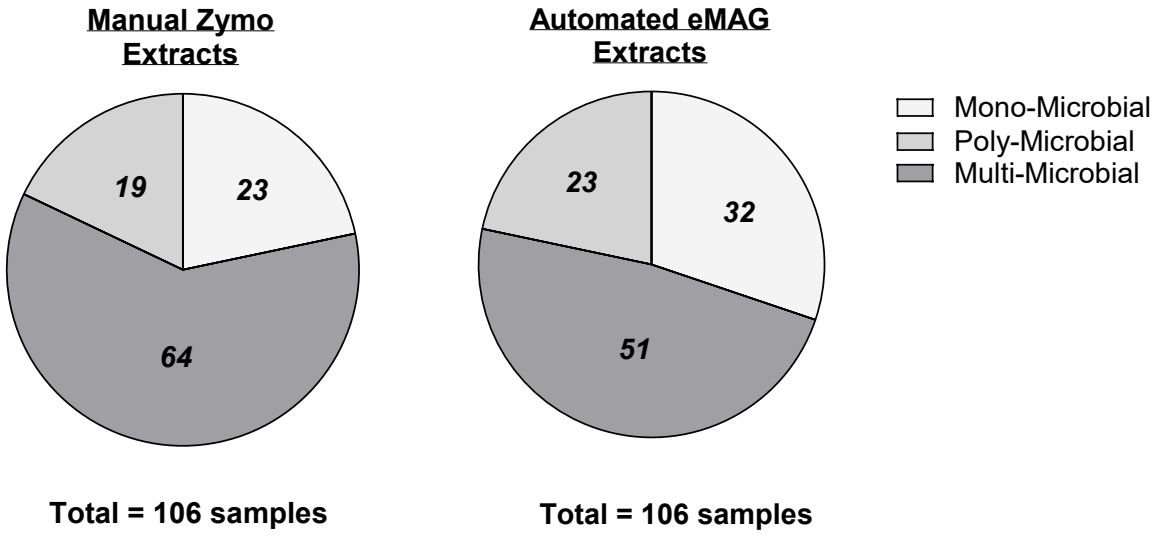


Figure 17: Number of samples with mono-microbial, poly-microbial and multi-microbial microbiome profiles.

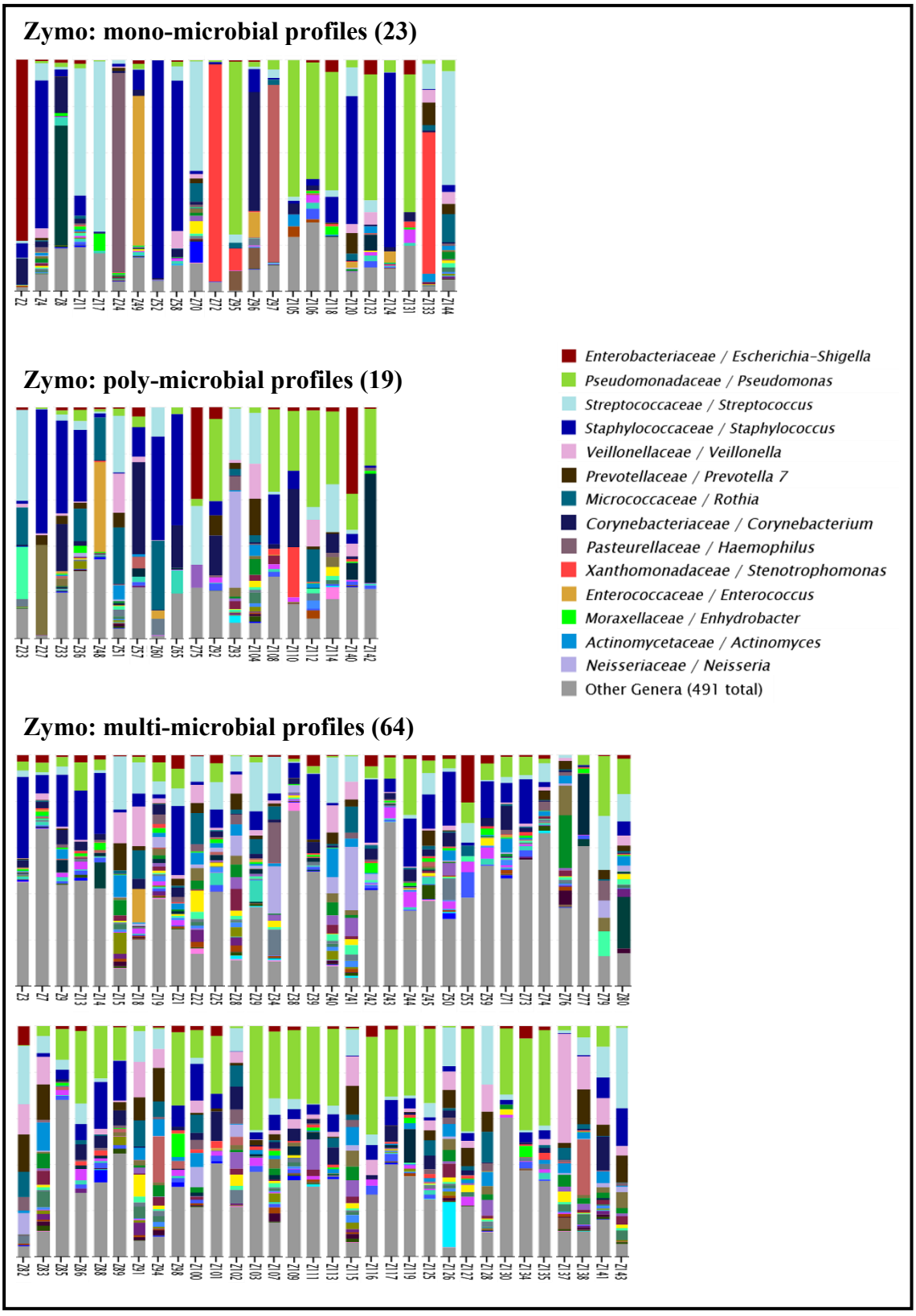


Figure 18: Zymo microbiome profiles arranged by category.

The profiles show the relative abundances of the top 50 genera after the exclusion of contamination genera. All additional genera were summarized as 'Other Genera' (visualization: Markus Weigel).

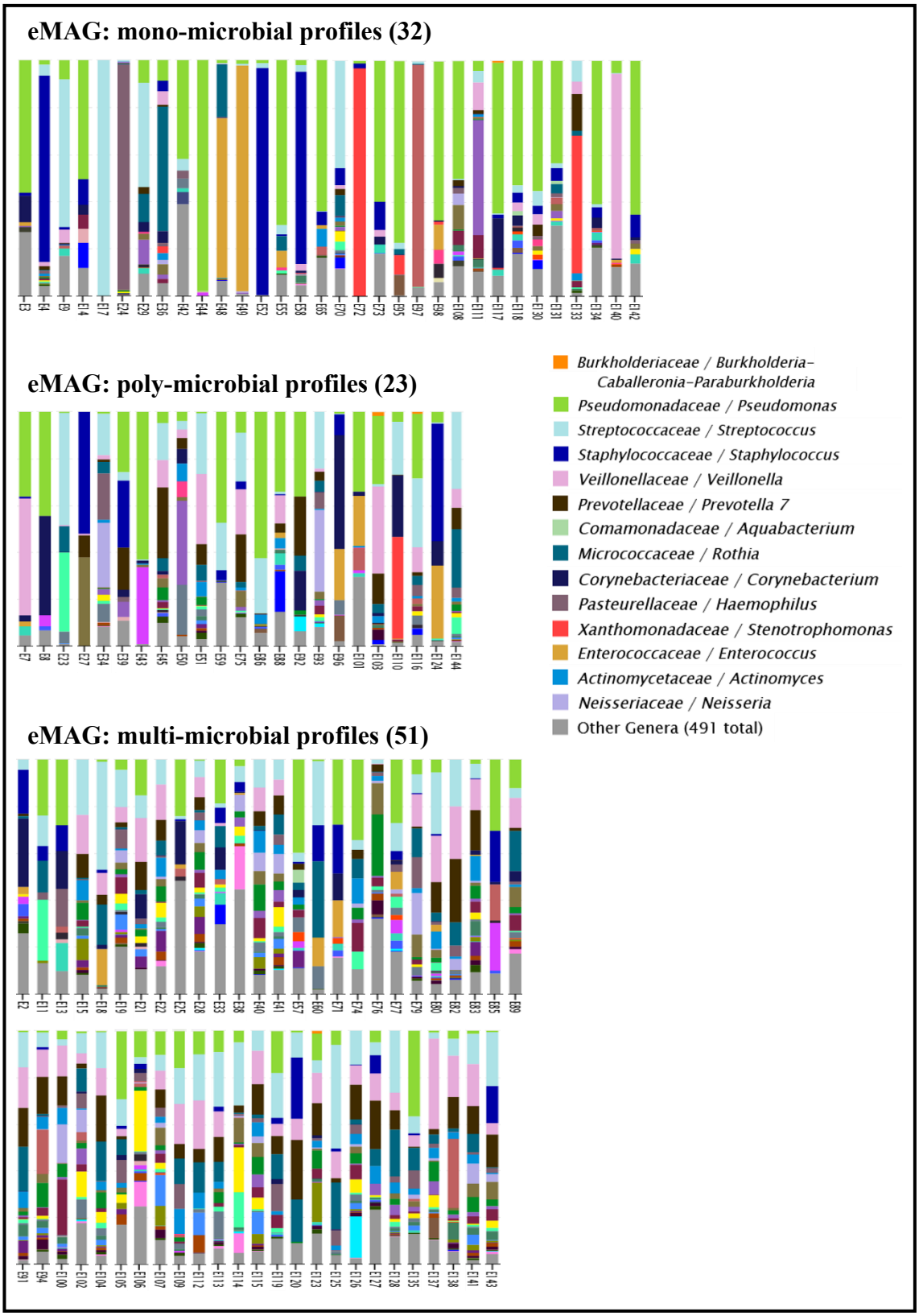


Figure 19: eMAG microbiome profiles arranged category.

The profiles show the relative abundances of the top 50 genera after the exclusion of contamination genera. All additional genera were summarized as 'Other Genera' (visualization: Markus Weigel).

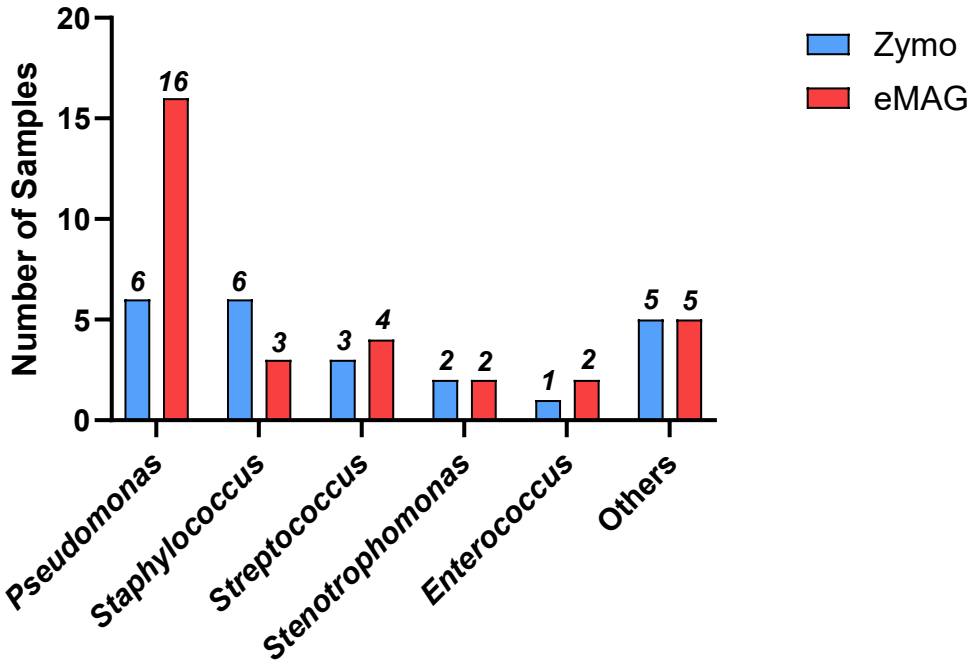


Figure 20: Top genera found in the mono-microbial microbiome profiles.

59 of the 106 samples were classified the same in the Zymo and the eMAG profiles (cf. Table 14). For all 13 overlapping mono-microbial profiles, the dominant genera in the Zymo and in the eMAG extracts matched. Of the 13 samples with a mono-microbial profile in eMAG and a multi-microbial profile in Zymo, in 10 samples *Pseudomonas* was the dominant genus in eMAG. The greatest absolute overlap of the Zymo and the eMAG microbiome profiles was found in the multi-microbial profiles.

	eMAG: Mono-Microbial	**eMAG: Poly-Microbial**	**eMAG: Multi-Microbial**	***Total***
Zymo: Mono-Microbial	13	4	6	***23***
Zymo: Poly-Microbial	6	7	6	***19***
Zymo: Multi-Microbial	13	12	39	***64***
Total	***32***	***23***	***51***	***106***

Table 14: Overlap of the Zymo and the eMAG microbiome profiles.

In the Zymo samples, the extract concentrations of the mono-microbial samples were visually higher than the extract concentrations of the poly- and multi-bacterial samples (Figure 21). However, this could not be affirmed by the Kruskal-Wallis test after Bonferroni correction for multiple comparisons ($p = 0.023$). In the eMAG samples, no significant difference was observed between the three microbiome profiles, which was substantiated by the Kruskal-Wallis test ($p = 0.327$). As mentioned above, some of the mono-microbial eMAG profiles could have been contaminated by *Pseudomonas*. Assuming that some of the *Pseudomonas* occurrences were indeed due to contamination and some mono-microbial profiles with Pseudomonas were actually poly- or multi-microbial, they could have lowered the median extract concentration of the mono-microbial eMAG profiles. When excluding all mono-microbial eMAG profiles with *Pseudomonas* as the dominant genus, the median of the extract concentrations increased from 0.36 (IQR, 0.17 – 3.95) to 2.21 (IQR, 0.30 – 6.10), which is represented by the dotted line in Figure 21. Running the Kruskal-Wallis test again to compare the medians and excluding the mono-microbial eMAG profiles with *Pseudomonas* as the dominant genus, the result showed a statistically significant difference between the medians of the mono-, poly- and multi-microbial samples ($p = 0.011$).

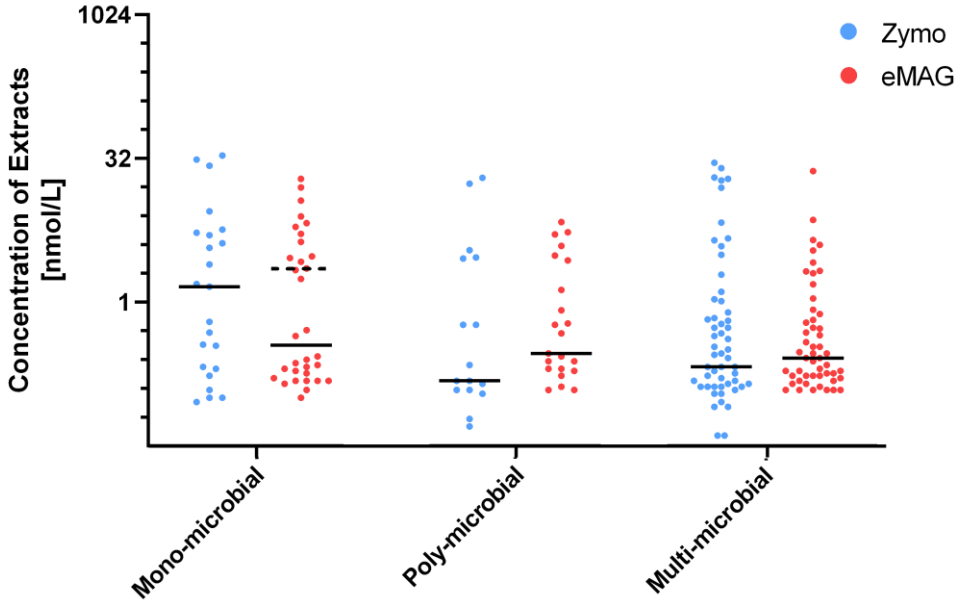


Figure 21: DNA concentration of the Zymo and eMAG extracts, sorted by microbiome profile category.

The dotted line in the scatter plot second from the left represented the median of the extract concentrations of the mono-microbial eMAG profiles after excluding all profiles with Pseudomonas as the dominant genus. The values were plotted with their medians on a logarithmic scale (base two).

Questions	**Data**	**Performed Test**	**Result**
Was the distribution of samples in the respective categories different between Zymo and eMAG?	Amount of samples in the respective categories (Figure 17)	Fisher`s exact test	***P = 0.187*** (distribution was not significantly different)
Were there significant differences between the extract concentrations of the mono-. poly- and multi-bacterial Zymo samples?	Concentration of nucleic acid [nmol/L] in the Zymo extracts, grouped by microbiome profile category	Kruskal-Wallis test (non-parametric)	***P = 0.023*** (medians between the three groups did not vary significantly)
Were there significant differences between the extract concentrations of the mono-. poly- and multi-bacterial eMAG samples?	Concentration of nucleic acid [nmol/L] in the eMAG extracts, grouped by microbiome profile category	Kruskal-Wallis test (non-parametric)	***P = 0.327*** (No differences between the medians of the three groups)
Were there significant differences between the extract concentrations of the mono-. poly- and multi-bacterial samples?	Concentration of nucleic acid [nmol/L] in the eMAG extracts, excluding all mono-microbial samples with *Pseudomonas* as the dominant genus, grouped by microbiome profile category	Kruskal-Wallis test (non-parametric)	***P = 0.011*** (medians between the three groups varied significantly)

Table 15: Statistical tests performed in section 3.3.5.

The significance level was defined as $\alpha = 0.05$ for the family of comparisons. According to the Bonferroni correction, the result was declared statistically significant when its P value was less than 0.017.

3.3.6 Interim summary

In batch three, the median number of read counts was significantly smaller than in batch one and two.

The blank controls were analyzed thoroughly and the decision, which genera were marked as contamination, was made on the basis of four non-mandatory criteria separately for the Zymo and the eMAG samples. The overall proportion of contamination was lower in the eMAG than in Zymo samples. Four OTUs were characterized as contamination in the eMAG samples, which were highly dominated by *Escherichia-Shigella*, and thirteen OTUs were marked as contamination in the Zymo samples.

When comparing the Zymo and the eMAG extracts, the number of sequencing reads did not differ substantially between the Zymo and the eMAG extraction method. The overall taxonomic composition of the Zymo and eMAG samples resembled each other after taking out the identified potential contaminant taxa. The most abundant genera both in Zymo and in eMAG were *Streptococcus*, *Staphylococcus*, *Pseudomonas*, *Prevotella 7*, *Veillonella* and *Rothia*. Overall, a higher number of genera was extracted from the Zymo extracts than from the eMAG extracts and all three α-diversity metrics showed higher numbers in the Zymo extracts.

In the last step of this section, each sample was categorized as mono-, poly- or multi-microbial. It could be observed that in the Zymo extracts, fewer mono-microbial and more multi-microbial profiles were classified than in the eMAG extracts. The genus *Pseudomonas* was highly dominant in eMAG, posing the question if it was colonization or contamination. 59 of the 106 samples were classified the same in the Zymo and the eMAG profiles. The DNA concentration of the extracts tended to be higher in the mono-microbial samples than in the poly- or multi-microbial samples.

3.4 Comparison of culture and short-read sequencing

This section brings together the data from culture-based routine diagnostics and Illumina short-read sequencing. Initially, wet lab data and diversity indices were aligned with the culture results (3.3.1). To assess the overlap of culture and NGS, two methods were

applied (3.4.2). Firstly, a table was generated which matched the species reported in culture with their corresponding genus from NGS. Secondly, the culture results (pathogen, commensal or no growth) were allocated to the microbiome profiles (mono-, poly- or multi-microbial) of their respective samples.

3.4.1 Association of culture results with wet lab and NGS parameters

The DNA yield of the samples with a potential pathogen reported in culture was significantly higher than the samples with a negative culture or a report of a cultivated commensal both in the Zymo and in the eMAG extracts, as shown in Figure 22 ($p = 0.004$ for Zymo and $p = 0.001$ for eMAG).

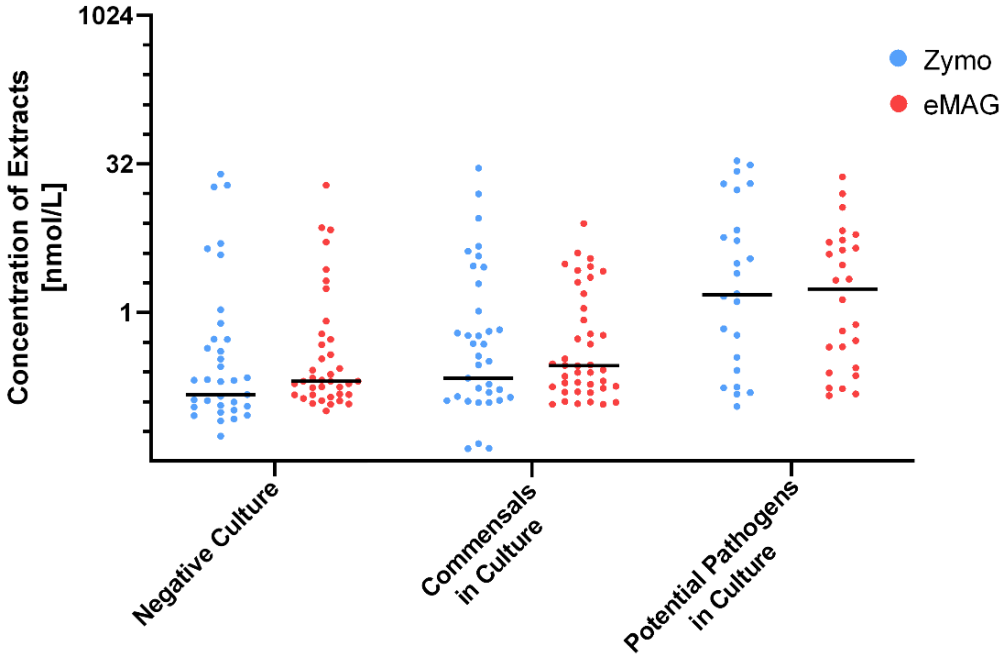


Figure 22: DNA concentration, sorted by culture result.

The values were plotted with their medians on a logarithmic scale (base two).

Subsequently, it was assessed whether α-diversity varied between samples with a different culture result. No visual or statistically significant difference in α-diversity between a negative culture, a cultured commensal or a cultured pathogen was found, independent of diversity index. The only exception was the Shannon diversity index in

the eMAG samples. Here, α-diversity was visually lower in the samples with a potential pathogen in culture (Figure 23). However, after Bonferroni correction for multiple comparisons, this difference was not statistically significant (p = 0.008). The observed richness and the Chao1 index can be found in Appendix 9: Overlap between the culture-based routine diagnostics and Illumina short-read sequencing – further results.

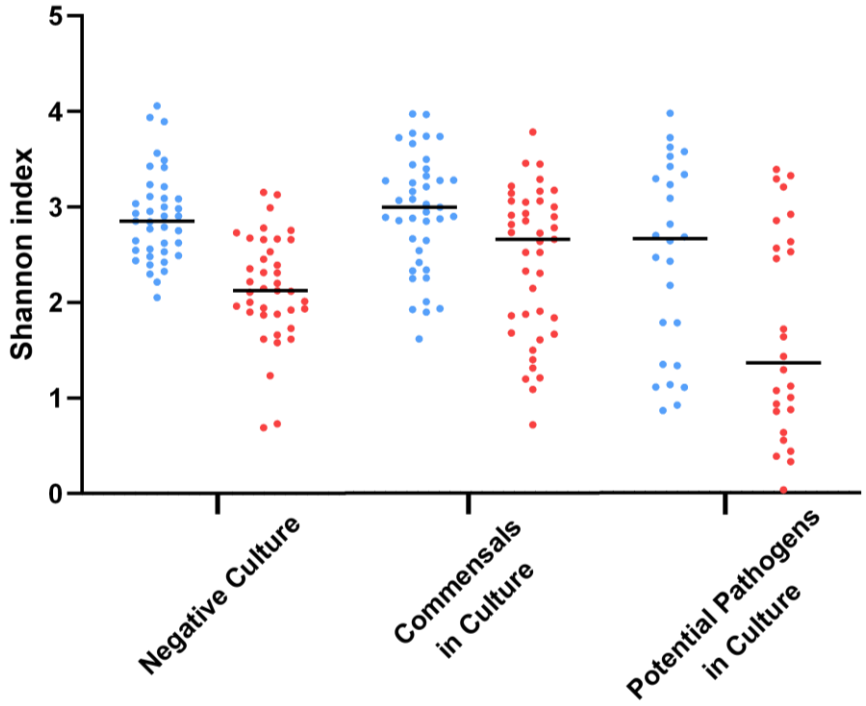


Figure 23: Shannon diversity index, sorted by culture results, shown with the medians.

Null Hypothesis	Data	Performed Test	Result
Were there significant differences in Zymo extract concentrations between samples with potential pathogens, commensals, or a negative culture result?	Concentration of nucleic acid [nmol/L] in the Zymo extracts, grouped by culture result	Kruskal-Wallis test (non-parametric)	***P = 0.004*** (medians between the three culture groups varied significantly)

Were there significant differences in eMAG extract concentrations between samples with potential pathogens, commensals, or a negative culture result?	Concentration of nucleic acid [nmol/L] in the eMAG extracts, grouped by culture result	Kruskal-Wallis test (non-parametric)	***P = 0.001*** (medians between the three culture groups varied significantly)
Were there significant differences in α-diversity between Zymo samples with potential pathogens, commensals, or a negative culture result?	Chao1 index in the Zymo samples, grouped by culture result	Kruskal-Wallis test (non-parametric)	***P = 0.099*** (medians between the three culture groups did not vary significantly)
Were there significant differences in α-diversity between eMAG samples with potential pathogens, commensals, or a negative culture result?	Chao1 index in the eMAG samples, grouped by culture result	Kruskal-Wallis test (non-parametric)	***P = 0.440*** (medians between the three culture groups did not vary significantly)
Were there significant differences in α-diversity between Zymo samples with potential pathogens, commensals, or a negative culture result?	Observed richness in the Zymo samples, grouped by culture result	Kruskal-Wallis test (non-parametric)	***P = 0.379*** (medians between the three culture groups did not vary significantly)
Were there significant differences in α-diversity between eMAG samples with potential pathogens, commensals, or a negative culture result?	Observed richness in the eMAG samples, grouped by culture result	Kruskal-Wallis test (non-parametric)	***P = 0.134*** (medians between the three culture groups did not vary significantly)
Were there significant differences in α-diversity between Zymo samples with potential pathogens, commensals, or a negative culture result?	Shannon diversity index in the Zymo samples, grouped by culture result	Kruskal-Wallis test (non-parametric)	***P = 0.187*** (medians between the three culture groups did not vary significantly)
Were there significant differences in α-diversity between eMAG samples with potential pathogens, commensals, or a negative culture result?	Shannon diversity index in the eMAG samples, grouped by culture result	Kruskal-Wallis test (non-parametric)	***P = 0.008*** (medians between the three culture groups did not vary significantly)

Table 16: Statistical tests performed in section 3.4.1.

The significance level was defined as α = 0.05 for the family of comparisons. According to the Bonferroni correction, the result was declared statistically significant when its P value was less than 0.006.

3.4.2 Overlap between culture results and short-read sequencing data

To assess the overlap between the results from culture based routine diagnostics and the sequencing data, the cultured species were matched with their corresponding genera from the microbiome profiles (Table 17). As the table shows, out of all 118 positive culture reports, a corresponding NGS genus was found in 108 cases (91.5 %) for the Zymo extracts. In 78 cases (66.1 %), these genera were among the top five genera in the respective samples. For the eMAG extracts, in 101 out of the 118 positive culture reports (85.6 %) a corresponding genus was found, and in 77 cases (65.3 %) these genera were among the top five genera in the respective samples.

In all 35 samples with a culture report of a-hemolytic *Streptococcus*, the corresponding genus could be found both in the Zymo and the eMAG extracts, however at varying positions. In comparison, for *St. aureus* and coagulase-negative *Staphylococcus*, there was a relatively high frequency of no corresponding *Staphylococcus* genus in NGS, which occurred in 25.0 % of the Zymo and in 40.0 % of eMAG extracts.

Among the 27 reports of potential pathogens (bold species in Table 17), five didn`t have a corresponding genus in Zymo at all (18.5 %), however, for 16 cultured species, the corresponding genus was found among the top five (59.3 %). Two of the 27 reports of potential pathogens could be attributed to *E. coli*, which could not be evaluated for the eMAG extracts because *Escherichia-Shigella* was categorized as contamination. For the 25 remaining reports, in 8 cases no corresponding genus could be found among the eMAG extracts (32.0 %) and for 16 cultured species (64.0 %), the corresponding genus was found among the top five.

Neither of the two *Citrobacter* reports (sample two and 140) and the one *Serratia* report (sample 126) from culture could be detected in Zymo or eMAG. One reason could have been that the samples did contain *Citrobacter* or *Serratia* (the culture was true positive), but the reads from these taxa were assigned to genera other than *Citrobacter* or *Serratia*. The sequencing data were analyzed with the BLAST algorithm, to check if some OTUs for *Citrobacter* or *Serratia* reads were found in the top five BLAST results of the respective samples. For *Citrobacter*, some fitting BLAST results were found, but none of the corresponding OTU had more than two read counts in the respective samples. When searching for the base sequence of OTU001 (*Escherichia-Shigella*) in BLAST, the first

result from *Cit. koseri* showed a 99.2 % accordance of the nucleotide sequence. For *Serratia*, no fitting BLAST result was found in sample 126.

The overlap of the PCR panel and the Illumina sequencing results is provided in Appendix 8: Overlap between the PCR panel and Illumina short-read sequencing.

Culture-based Diagnostics		Corresponding Genera Detected by Illumina Short Read NGS							
Species	Number of Reports	Most Abundant		Top Five		Any Abundance		Absent	
		Zymo	eMAG	Zymo	eMAG	Zymo	eMAG	Zymo	eMAG
α-hemolytic *Streptococcus*	35	10	11	31	30	35	35	0	0
Saprophytic *Neisseria*	12	1	0	5	6	12	10	0	2
St. aureus	11	5	3	6	6	9	7	2	4
Coagulase-neg. *Staphylococcus*	9	2	2	3	2	6	5	3	4
Pre. melaninogenica	5	0	0	4	3	5	5	0	0
Ps. aeruginosa	5	3	3	4	4	5	4	0	1
Ro. mucilaginosa	5	0	1	3	4	5	5	0	0
Ent. faecium	4	2	2	2	3	4	4	0	0
St. epidermidis	4	2	1	4	3	4	4	0	0
Rothia sp.	3	0	1	2	3	3	3	0	0
Ste. maltophilia	3	2	3	3	3	3	3	0	0
Str. mitis	3	0	0	2	2	3	3	0	0
Cit. koseri	2	0	0	0	0	0	0	2	2
Ent. faecalis	2	0	0	1	1	1	1	1	1
E. coli	2	0	Cont.	0	Cont.	2	Cont.	0	Cont.
N. flava	2	1	2	2	2	2	2	0	0
Act. odontolyticus	1	0	0	0	0	1	1	0	0
Bor. bronchiseptica	1	0	0	1	1	1	1	0	0
H. influenzae	1	1	1	1	1	1	1	0	0
H. parahaemolyticus	1	0	0	0	0	1	1	0	0
Lac. reuteri	1	0	0	1	0	1	1	0	0
Lac. salivarius	1	0	0	0	0	1	1	0	0
Ser. marcescens	1	0	0	0	0	0	0	1	1
Streptococcus parasanguinis	1	0	0	1	0	1	1	0	0
Str. pneumoniae	1	1	1	1	1	1	1	0	0
Str. salivarius	1	0	0	1	1	1	1	0	0
Vei. parvula	1	0	0	0	1	0	1	1	0
Sum	***118***	***30***	***31***	***78***	***77***	***108***	***101***	***10***	***15***

Table 17: Alignment of the species reported in culture with their correspondent genus from Illumina short-read sequencing.

The thick printed species were categorized as potentially pathogenic. The presence of the genus Escherichia was not determined in the eMAG extracts because it was evaluated as contamination. Analysis scheme was modified according to Zachariah et al. (2018).

In addition to the overlap of the detected taxa, it was assessed how the established categories of the culture-based routine diagnostics and the microbiome profiles matched

(Table 18). Among the Zymo extracts, in 10 of 23 monomicrobial profiles (43.5 %), a potential pathogen was detected in culture, in 8 of these 10 cases with corresponding taxa. Among the eMAG extracts, in 11 of 32 monomicrobial profiles (34.4 %), a potential pathogen was cultured, in 8 of these 11 cases with corresponding taxa.

Among the samples with a multi-microbial profile and a potential pathogen cultured, these pathogens were mainly *St. aureus*, *E. coli* and *Ps. aeruginosa* in both the Zymo and the eMAG extract (cf. Appendix 9: Overlap between the culture-based routine diagnostics and Illumina short-read sequencing – further results, Figure 42).

As could be taken from Figure 25, in 13 out of 14 eMAG extracts (92.9 %) with a negative culture and a mono-microbial profile, *Pseudomonas* was the dominant genus in the microbiome profiles. Among the samples with a monomicrobial profile and a potential pathogen or a commensal detected in culture-based routine diagnostic, the dominant taxa in the Zymo and the eMAG extracts were very similar to each other (Figure 25).

	Potential Pathogen in Culture		**Only Commensals in Culture**		**Negative Culture**		***Sum***	
	Zymo	**eMAG**	**Zymo**	**eMAG**	**Zymo**	**eMAG**	**Zymo**	**eMAG**
Mono-Microbial	10	11	7	7	6	14	***23***	***32***
Poly-Microbial	4	5	8	10	7	8	***19***	***23***
Multi-Microbial	12	10	26	24	26	17	***64***	***51***
Sum	***26***	***26***	***41***	***41***	***39***	***39***	***106***	

Table 18: Overlap of the culture categories and the NGS categories.

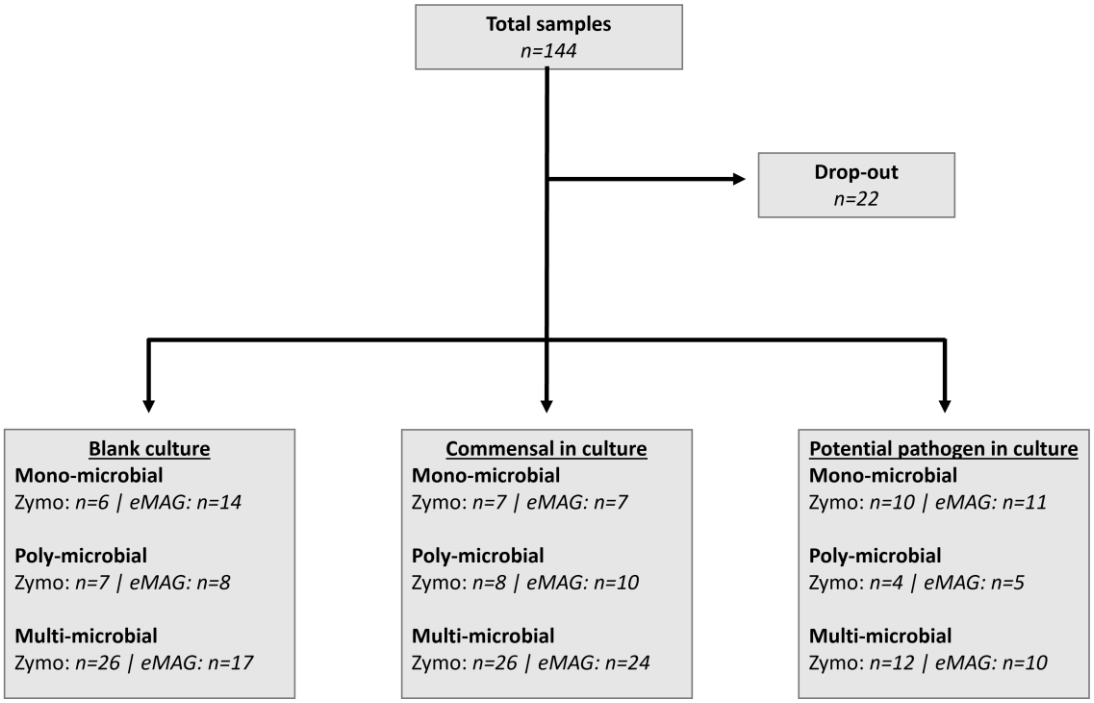


Figure 24: Culture results listed with the respective categories from Illumina short-read sequencing.

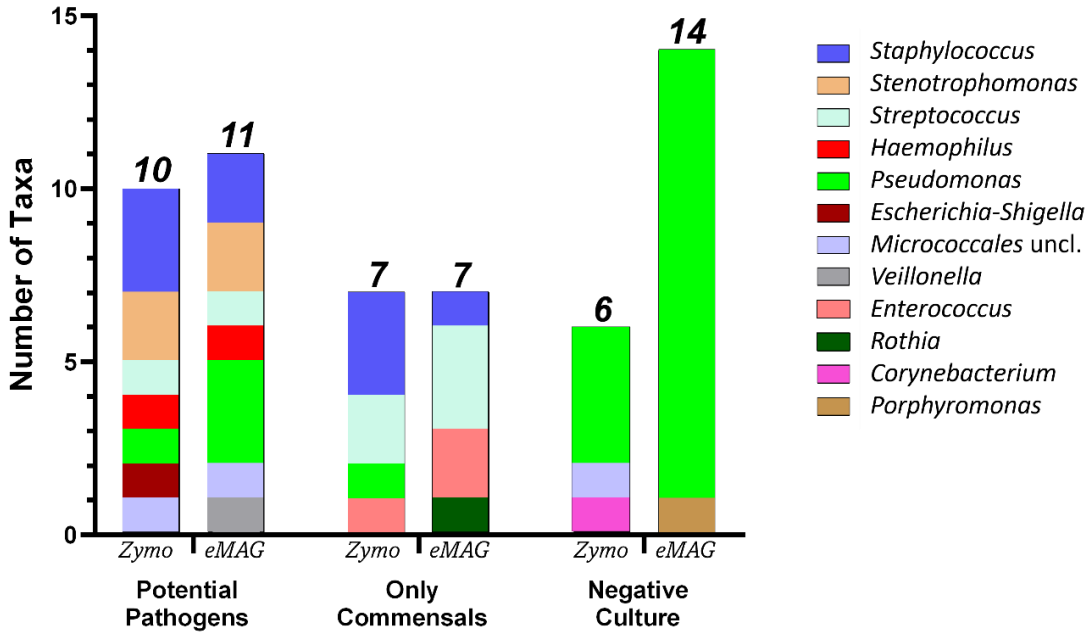


Figure 25: Dominant taxa in the mono-microbial profiles, sorted according to culture results.

Regarding the dominant genera in the mono-microbial profiles, the relative abundances in these microbiome profiles were overall higher in the samples with a potential pathogen detected in culture-based routine diagnostics, than in the samples with commensals

cultured or a negative culture report (Figure 26), which was statistically substantiated by the Kruskal-Wallis test (p = 0.012 for Zymo and p = 0.004 for eMAG).

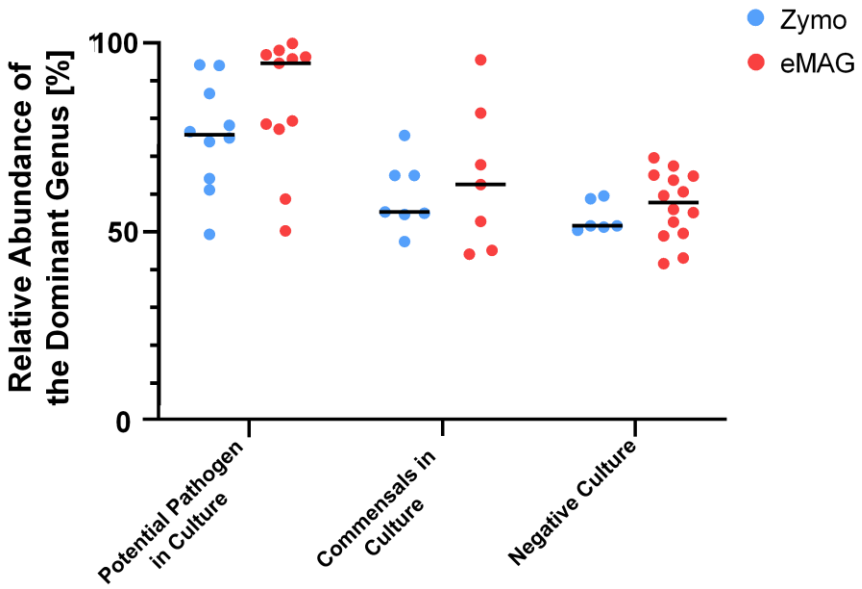


Figure 26: Relative abundances of the dominant genera of the mono-microbial profiles, sorted according to culture results.

Null Hypothesis	Data	Performed Test	Result
Were there differences in relative abundances of the dominant genera between the mono-microbial Zymo samples with potential pathogens, commensals, and a negative culture result?	Relative abundances of the dominant genus [%] in the mono-microbial Zymo samples, grouped by culture result	Kruskal-Wallis test (non-parametric)	***P = 0.012*** (medians between the three groups varied significantly)
Were there differences in relative abundances of the dominant genera between the mono-microbial eMAG samples with potential pathogens, commensals, and a negative culture result?	Relative abundances of the dominant genus [%] in the mono-microbial eMAG samples, grouped by culture result	Kruskal-Wallis test (non-parametric)	***P = 0.004*** (medians between the three groups varied significantly)

Table 19: Statistical tests performed in section 3.4.2.

The significance level was defined as α = 0.05 for the family of comparisons. According to the Bonferroni correction, the result was declared statistically significant when its P value was less than 0.025.

3.4.3 Interim summary

Samples with a potential pathogen in culture exhibited the highest mean DNA concentration of the extracts. α-diversity didn`t vary significantly between samples with different culture results.

The first method to assess, how culture and NGS matched, was a table depicting the alignment of cultured species and sequenced genera (Table 17). In the majority of the culture reports, a corresponding genus was found in NGS (Zymo: 91.5 %; eMAG: 85.6 %). However, the number of corresponding results was smaller when just looking just at the top five sequenced genera (Zymo: 66.1 %; eMAG: 55.3 %). Looking at the reports of cultured potential pathogens, no corresponding genus was present in 18.5 % of the Zymo extracts and 32.0 % of the eMAG extracts. Two cultured potential pathogens could not be detected at all by short-read sequencing (*Citrobacter* and *Serratia*).

The second method was comparing the culture categories with the sequencing categories. In 8 of 23 samples with a monomicrobial profile in Zymo and in 8 of 32 samples in eMAG, a potential pathogen was cultured which matched the NGS genus. In the samples with a multi-microbial profile and a potential pathogen cultured, these cultured species were mainly *St. aureus*, *E. coli* and *Ps. aeruginosa*. In 13 out of 14 eMAG samples with no culture growth and a mono-microbial profile, *Pseudomonas* was the dominant genus in the microbiome profiles. The relative abundances of the dominant genera in the mono-microbial profiles were higher in the samples with a potential pathogen in culture, compared to the samples with commensals or no growth in culture.

3.5 Species identification with Nanopore long-read sequencing

Nanopore long-read sequencing was conducted with 21 selected samples and three blank controls for further microbiome profile evaluation on species level. For this purpose, the manual Zymo extracts were used, due to a more suitable contamination profile and the assumed additional contamination by *Pseudomonas* in the automated eMAG extracts. In section 3.5.1, the Nanopore sequencing output is presented, the microbiome profile categorization from section 2.6 was applied to the data and the results from culture, Illumina short-read and Nanopore long-read sequencing were brought together.

3.5.1 Presentation and analysis of the Nanopore sequencing data

The sequencing run generated a total of 38,470 read counts with a median of 306 reads per sample (IQR, 89 – 1,163). The range of read counts was relatively high, the sample with the highest number generated 18.022 read counts and the sample with the lowest number 38. Appendix 10: Species identification with Nanopore long-read sequencing – further results provides the plotted number of read counts and a table of all samples sequenced with the Nanopore technique, including the two most abundant species with their number of read counts.

The criteria for microbiome profiles from 2.6 were also applied to the Nanopore data, and the samples were divided into mono-, poly- and multi-microbial profiles. The mono-microbial category consisted of six samples, 14 samples were classified as multi-microbial and only one was poly-microbial. Among samples with mono-microbial profiles, a visually higher number of read counts was generated than in the samples with a poly- or a multi-microbial profile, as could be seen in Figure 27. However, this difference was not statistically significant in the Mann-Whitney test ($p = 0.076$), which could be explained by the small sample size.

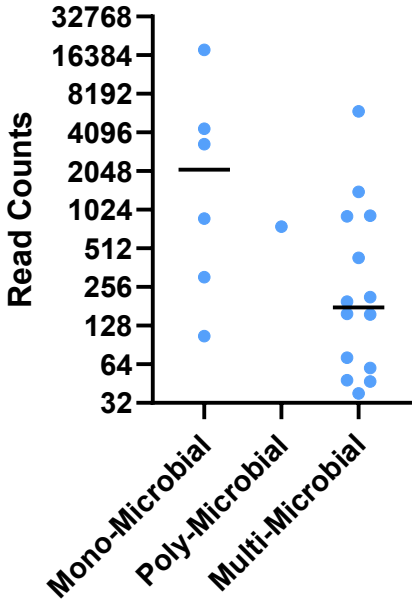


Figure 27: Sequenced reads from Nanopore long-read sequencing, sorted by microbiome profile category.

The values were plotted on a logarithmic scale (base two).

Table 20 summarizes the results from culture and the two sequencing techniques. For three samples no Illumina sequencing results were available due to drop-out. The Venn diagram (Figure 28) visualizes the results from Table 20.

In 13 of 17 positive culture results (76.5 %), at least one cultured species was also found among the top genera from the corresponding Nanopore sequencing results. In 44.4 % (four of nine samples), a cultured species which was potentially pathogenic was also detected by Nanopore sequencing. The species, that were not detected by Nanopore sequencing, although reported in culture-based diagnostics, were *St. aureus*, *Ps. aeruginosa* and *E. coli*. In two samples, Nanopore sequencing detected a potential pathogen, while the culture result was either negative or showed only commensals. In sample 34, *H. influenzae* was recorded by Nanopore sequencing while the culture was negative. The Illumina result was correspondent to the Nanopore result, with *Haemophilus* as second most abundant genus, and in the PCR panel *H. influenzae* was reported as well. In sample eleven, *Str. pneumoniae*, a potential pathogen, was detected as the second most abundant species but was not reported in culture.

In 14 of 18 cases (77.8 %), Illumina and Nanopore sequencing showed overlaps, however, only in eight of 18 cases (44.4 %), the two most abundant taxa coincided fully. Nanopore sequencing matched 22 of the 36 top Illumina taxa (61.1 %).

Illumina short-read sequencing did not register any *Citrobacter* occurrences in the two samples with a positive culture for *Cit. koseri*. One of these cases was sample two. Nanopore sequencing yielded read counts attributed to *Cit. koseri*, thus supported the culture report. In the corresponding Illumina results, *Escherichia-Shigella* was the top genus with a relative abundance of 78.2 %.

Sample Number	Culture-based Routine Diagnostics	Illumina Short- Read Sequencing	Nanopore Long-Read Sequencing
2	*Cit. koseri* *St. epidermidis*	*Escherichia-Shigella* (78.2 %) *Corynebacterium* (11.3 %)	*Cit. koseri* (69.6 %) *St. epidermidis* (4.2 %)
11	α-hemolytic *Streptococcus*	*Streptococcus* (55.1 %) *Staphylococcus* (8.7 %)	*Streptococcus pseudopneumoniae* (37.9 %) *Str. pneumoniae* (13.9 %)
13	*St. aureus*	*Staphylococcus* (21.4 %) *Pseudomonas* (12.2 %)	*Ralstonia pickettii* (18.3 %) *Par. fungorum* (13.3 %)
15	α-hemolytic *Streptococcus* Saprophytic *Neisseria*	*Streptococcus* (23.1 %) *Veillonella* (13.3 %)	*Str. salivarius* (20.0 %) *Veillonella atypica* (12.8 %)
17	*Str. pneumoniae*	*Streptococcus* (73.9 %) *Enhydrobacter* (7.5 %)	*Str. pneumoniae* (87.2 %) *Str. pseudopneumoniae* (2.3 %)
18	α-hemolytic *Streptococcus*	*Veillonella* (17.2 %) *Enterococcus* (14.5 %)	*Veillonella dispar* (23.7 %) *Ent. faecalis* (10.5 %)
19	None	*Pseudomonas* (8.2 %) *Neisseria* (4.7 %)	*Ral. pickettii* (6.9 %) *Neisseria perflava* (5.6 %)
22	α-hemolytic *Streptococcus* Saprophytic *Neisseria* *Ps. aeruginosa*	*Rothia* (13.2 %) *Streptococcus* (11.1 %)	*Str. parasanguinis* (18.1 %) *Campylobacter concisus* (9.6 %)
23	α-hemolytic *Streptococcus* *Ro. mucilaginosa*	*Streptococcus* (39.4%) *Gemella* (22.6 %)	*Streptococcus oralis* (35.1 %) *Gemella haeolysans* (24.9 %)
24	α-hemolytic *Streptococcus* *H. influenzae*	*Haemophilus* (86.6 %) *Streptococcus* (1.3 %)	*H. influenzae* (76.7 %) *H. parahaemolyticus* (1.9 %)
27	*St. aureus*	*Staphylococcus* (54.0 %) *Prevotellaceae uncl.* (39.0 %)	*St. aureus* (74.2 %) *Prevotella oris* (9.7 %)

30	α-hemolytic *Streptococcus* Saprophytic *Neisseria* *E. coli*	-	*H. parahaemolyticus* (12.6 %) *Str. salivarius* (11.1 %)
34	None	*Neisseria* (20.4 %) *Haemophilus* (17.4 %)	*H. influenzae* (8.3 %) *Str. mitis* (8.3 %)
43	*Ps. aeruginosa*	*Pseudmononas* (9.2 %) *Staphylococcus* (5.8 %)	*Ral. pickettii* (8.3 %) *Aquabacterium commune* (6.4 %)
45	α-hemolytic *Streptococcus* *Rothia sp.*	*Staphylococcus* (14.7 %) *Streptococcus* (9.2 %)	*Ral. pickettii* (9.8 %) *Streptococcus australis* (8.4 %)
46	None	-	*Str. salivarius* (10.8 %) *Streptococcus vestibularis* (3.8 %)
48	*Ent. faecium*	*Enterococcus* (39.0 %) *Rothia* (18.4 %)	*Ent. faecium* (55.7 %) *Lactobacillus paracasei* (2.8 %)
49	*Ent. faecium*	*Enterococcus* (65.0 %) *Staphylococcus* (8.7 %)	*Ent. faecium* (59.2 %) *St. epidermidis* (4.9 %)
50	None	*Staphylococcus* (23.3 %) *Lactobacillales uncl.* (9.6 %)	*Ral. pickettii* (9.6 %) *Aq. commune* (8.6 %)
51	α-hemolytic *Streptococcus* *Rothia sp.*	*Streptococcus* (24.5 %) *Rothia* (24.4 %)	*Str. parasanguinis* (31.3 %) *Vei. dispar* (10.3 %)
53	*St. aureus*	-	*Herbaspirillum huttiense* (12.8 %) *Ral. pickettii* (8.5 %)

Table 20: Results from culture, Illumina and Nanopore sequencing.

For Illumina and Nanopore sequencing, only the two most abundant taxa were shown, with the respective relative abundance indicated in parentheses. Potential pathogens were underlined.

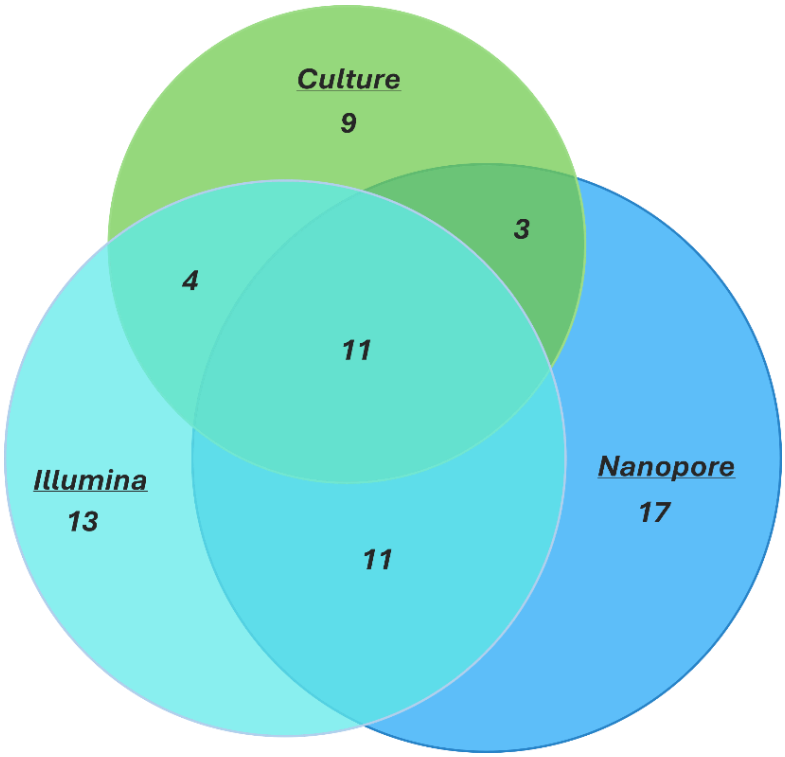


Figure 28: Venn diagram showing the overlapping taxa for culture, Illumina and Nanopore sequencing.

Null Hypothesis	Data	Performed Test	Result
Where there significant differences in read counts between the mono- and multi-bacterial Nanopore samples?	Read counts of the Nanopore samples, grouped by microbiome profile category	Mann-Whitney test (two-tailed)	***P = 0.076*** (No differences between the read counts of the mono- and the multi-bacterial Nanopore samples)

Table 21: Statistical tests performed in section 3.5.1.

The significance level was defined as α = 0.05.

3.5.2 Interim summary

All Nanopore samples were assigned to one of the three categories. The mean number of read counts of the mono-microbial samples were visually higher, however, this could not be confirmed statistically. In 76.5 %, Nanopore sequencing result corresponded with the cultured species. 44.4 % of the cultured potential pathogens were detected by Nanopore sequencing. In two cases, Nanopore sequencing detected a potential pathogen (*H. influenzae* and *Str. pneumoniae*), which was not reported by culture. For the two most abundant taxa, Illumina and Nanopore showed a full or partial overlap in 77.8 % of the cases. Nanopore sequencing matched 22 of the 36 top Illumina taxa (61.1 %).

4 Discussion

4.1 Strengths and limitations

There are both strengths and limitations to microbiome studies in general, and to this study in particular. The study had a large sample size of 144. However, 38 of 144 samples could not be included in the final analysis due to dropout, reducing internal validity.

A common pitfall of 16S rRNA gene sequencing is the risk of contamination during the laboratory workflow. This applies especially to respiratory samples due to the relatively low bacterial biomass in the lower respiratory tract (Salter *et al.*, 2014; Man, de Steenhuijsen Piters and Bogaert, 2017; Carney *et al.*, 2020; Martin-Loeches *et al.*, 2020). For that reason, particular attention was paid to the inclusion and analysis of multiple blank controls.

It is highly questionable to draw strict lines between pathogens and commensals, as illustrated in the introduction. The interactions between bacteria within a microbiome are highly dynamic processes, and taxa which cause infection in one lung may be part of the physiological flora in the other. For this reason, potential pathogenicity was evaluated with great caution (cf. Table 8), considering quantitative culture parameters and the presence of granulocytes (cf. Appendix 3: Pathogenicity evaluation for the results of culture-based routine diagnostics). Each sample was looked at individually to ensure that aggregation of data would not lead to misguided conclusions.

Another limitation was the absence of clinical patient data. The sample group may have had a diverse range of diseases, physiological parameters and personal backgrounds which was not incorporated into the analyses. However, the focus of this work lay on the comparison between NGS and culture, and not on disease or patient-specific microbiome alterations.

Furthermore, the taxonomic resolution of V4 16S rRNA sequencing merely comprises genus-level and does not reach the species-level (Mendez *et al.*, 2019; Martin-Loeches *et al.*, 2020) which limits the applicability for clinical questions. Therefore, some samples were selected for Nanopore long-read sequencing to enable an assessment on species level. Although the microbiome additionally consists of viruses, fungi, and protozoa, this

study, like many others (Allaband *et al.*, 2019), focused only on bacteria, which may also limit the external validity of the study.

An issue all throughout lung microbiome research is that few consistent standard methods have been established to date, due to the relatively young age of the field. Many scientists have exemplified the high variability in results, depending on what laboratory and bioinformatic tools were used (Knight *et al.*, 2018; Greathouse, Sinha and Vogtmann, 2019; Carney *et al.*, 2020; Drengenes *et al.*, 2021). For this reason, it was essential to report laboratory and bioinformatic methodology thoroughly (Greathouse, Sinha and Vogtmann, 2019).

4.2 Evaluation of the extraction method

The choice of the extraction method is a vital affair since the extraction kits were one of the biggest contributors to contamination in this study (cf. 4.4). Many studies showed that the DNA extraction method and kit can substantially influence the sequencing results (Teng *et al.*, 2018; Angebault *et al.*, 2020; Gryp *et al.*, 2020; Neuberger-Castillo *et al.*, 2020). In this section, an answer to question 1 will formulated, which was:

Did the manual extraction method with Zymo and the automated extraction platform eMAG produce different results?

Zymo and eMAG were very similar in terms of DNA yield and read counts. However, deviations could be found in the bacterial community structures (in terms of diversity and extracted genera) and substantial differences in pattern of contamination. The differences between Zymo and eMAG could not only be ascribed to the level of workflow automation since there were additional differences in the protocols and application of these two techniques. For instance, they relied on different molecular methods of extraction: the Zymo kit used a bead-based mechanical method and the eMAG platform extracted nucleic acid enzyme-mediated.

Regarding DNA yield, no difference was observed in the DNA concentration of the extracted samples between Zymo and eMAG. There is no clear trend in the current literature as to which extraction method provides the highest DNA yield. Some studies observed that automated extraction machines produced the highest DNA yield with bead-

beating approach, partially combined with enzymatic lysis (Angebault *et al.*, 2020; Yu *et al.*, 2023). In turn, another study by Lim at al. (2018) did not find any differences in DNA yield between an automated and a manual workflow, however, they found differences between kits. Several studies found that an enzymatic extraction approach provided the highest DNA yield (Vesty *et al.*, 2017; Teng *et al.*, 2018; Rosenbaum *et al.*, 2019). Other studies reported that bead-beating is a better method for total DNA recovery (Lim *et al.*, 2018; Mallott, Malhi and Amato, 2019; Gryp *et al.*, 2020). Another study found no difference between methods with and without bead-beating (Rosenbaum *et al.*, 2019). There are also some studies supporting the assumption that the respective DNA yield does not even have an eminent influence on the final microbiome profile results (Rosenbaum *et al.*, 2019; Angebault *et al.*, 2020; Abundo *et al.*, 2021).

The manual Zymo and the automated eMAG extraction methods were comparable regarding overall read counts (Figure 6) which is in line with a study by Mallott, Malhi and Amato (2019). The Bland-Altman plot (Figure 11), which is generally used to compare two measurement methods, showed that there was no systematic difference in either direction between the Zymo and the eMAG read counts.

The proportion of the contamination among the Zymo extracts (30.3 %) was higher than in the eMAG extracts (26.7 %), as described in 3.3.3 ($P < 0.001$). These proportions were high compared to a study by Willner et al. (2012) who also conducted 16S rRNA microbiome profiling with BALF samples and reported a maximum contamination of eighteen percent. The genera marked as contamination essentially differed between Zymo and eMAG. In Zymo, more contamination genera were determined, but the individual genera were of a lower abundance. In contrast, the eMAG samples were strongly dominated by only one genus, *Escherichia-Shigella*. Although *Pseudomonas* was not found excessively in the eMAG blank controls, it could not ultimately be ruled out as possible contamination. In clinical practice, the predominant *Escherichia-Shigella* contamination combined with the likely *Pseudomonas* contamination might pose a greater problem than the more dispersed pattern of contamination in Zymo, as it may be more difficult to distinguish actual lung pathogens from contamination. The joint microbiome profiles looked similar to each other after excluding the selected contamination genera (Figure 12). These observations are important because they underline the commonly accepted assumption that lung microbiome profiles are

substantially affected by contamination due to their low biomass (Kim *et al.*, 2017; Moffatt and Cookson, 2017; Sui *et al.*, 2020).

The manual Zymo extraction yielded a significantly higher number of genera (Figure 13) respectively with lower relative abundances, which was also reflected in the overall higher α-diversity among the Zymo extracts (Figure 14). There are studies that also found significantly different α-diversity indices between different kits (Lim *et al.*, 2018; Yu *et al.*, 2023). However, a number of other studies concluded that the extraction kit had little influence on α-diversity (Vesty *et al.*, 2017; Mallott, Malhi and Amato, 2019; Mattei *et al.*, 2019; Angebault *et al.*, 2020; Abundo *et al.*, 2021). A study by Lim et al. (2018) with fecal samples and another by Teng et al. (2018) with oral microbiome samples found that including mechanical bead-beating resulted in a higher α-diversity. On the other hand, a study by Rosenbaum et al. (2019) concluded that bead-beating did not have a great influence on the oral microbiome composition. The degree of automation had little influence on the community diversity, as some studies indicated (Lim *et al.*, 2018; Angebault *et al.*, 2020). Mattei et al. (2019) observed that the extraction method producing the highest DNA yield also performed lowest in detecting microbial diversity, which suggests that extraction methods may need to be specifically adapted to research question (e.g. whether DNA yield or microbial diversity is investigated). Regarding the extracted genera after excluding contamination, the top genera in Zymo and eMAG were very similar in this study, which is depicted in the bar chart of Figure 15. This was in line with the findings of a study by Abundo et al. (2021).

The PCoA plot of the Bray-Curtis index (Figure 16) provided information on the taxonomic diversity between all samples and controls. The community structures of samples extracted by one method were overall more similar to each other and therefore formed a cluster. The formations of the Zymo and the eMAG clusters could be partly attributed to the particular contamination patterns of the respective extraction method. This was also observed by Salter et al. (2014) who stated that contamination was the biggest driver of sample arrangement in the PCoA cluster. Other studies found that the extraction method did not have a large effect on the β-diversity of different extraction kits after the contamination was removed (Abundo *et al.*, 2021), or that sample type and host identity were stronger drivers for β-diversity than the extraction method (Shaffer *et al.*, 2022). Lim et al. (2018) found that samples clustered according to whether bead-beating was used or not but irrespective of the degree of automation. The control dots also formed

a cloud due to their own taxonomic composition. The control scatter plot was relatively dispersed, implying that the microbiome profiles of the controls differed more among themselves which led to a greater distance between them.

As shown in Figure 17, more Zymo samples were categorized as multi-microbial (64 samples in Zymo vs. 51 samples in eMAG) and more eMAG samples were categorized as mono-microbial (32 samples in eMAG and 23 samples in Zymo), however, this difference was not statistically significant (P = 0.187). Table 14 showed the overlap of the categorized Zymo and eMAG microbiome profiles. In 47 samples (44.3% of the samples), different categories were assigned for the Zymo and the eMAG extracts, which could be considered very high. The greatest discrepancy occurred in the samples with a mono-microbial profile in eMAG and a multi-microbial profile in Zymo. 10 of these samples had *Pseudomonas* as the dominant genus in the eMAG samples. This raised the question if *Pseudomonas* actually was in all of those BAL fluids or if some could be ascribed to contamination. Another possible reason for the discrepancy in categorization between the eMAG and the Zymo samples is, that the lines between those categories was highly artificial, which is discussed in section 4.3 more deeply.

The studies cited in this section investigated different sample types (such as oral, respiratory, vaginal or fecal samples), some of them were not even human samples (like poultry or vacuumed dust). Study methods as well as results were highly heterogeneous and partly contradictory, and comparability between the studies is therefore limited. Besides, according to current knowledge it is not clear to what extent the extraction method actually contributes to this variation (Mallott, Malhi and Amato, 2019). There is probably not the one best extraction method, but the choice may dependent on the sample type and research hypothesis (Mattei *et al.*, 2019; Sui *et al.*, 2020). More studies are needed to investigate the performance of DNA extraction methods in different settings, which have to be tightly controlled by negative and positive controls.

However, time and personnel resources required for the manual and the automated sequencing technique has not been systematically assessed in this work, it is a substantial aspect to consider when sighting the routine application of NGS. For 50 eMAG samples, about two hours of preparation and post-processing by one person was needed. In contrast, the manual extraction of 18 Zymo samples took one person about one working day, making manual extraction more time and personnel consuming. The longer time

required for manual extraction is in line with experiences from other studies (Angebault *et al.*, 2020; Yu *et al.*, 2023).

In conclusion, it can be stated, that: firstly, both kits extracted bacterial DNA successfully, with comparable results in DNA yield, number of read counts and extracted taxa. Varying results were observed in pattern of contamination and diversity. Secondly, both extraction methods were linked to high levels of contamination, although the contamination profiles of both methods were distinct. The contamination pattern may have been more favorable in the Zymo samples. Thirdly, the overall taxonomic composition looked alike after excluding contamination genera. Lastly, when aiming for routine use of an extraction method, different criteria must be addressed, like the reproducibility and the generation of valid results, the amount of contamination, the turnaround time, as well as the complexity of the workflow.

4.3 Comparison of culture and NGS

In this section, research questions 2-4 will be answered, which were:

2) Did the cultured species overlap with the NGS results?

3) Did V4 16S rRNA gene sequencing with Illumina add value to the analysis of the BALF samples?

4) Did long-read sequencing with Nanopore add value to the analysis of the BALF samples?

Comparing results of culture with microbiome profiles is a challenging undertaking since they rely on different conceptions of the lung and its microbial inhabitants. In many studies in this field, sensitivity and specificity is used to evaluate the detection rate of NGS and compare it with the results of culture-based methods, especially studies that use metagenomic NGS (and not 16S rRNA sequencing) (Wang, Han and Feng, 2019; Chen *et al.*, 2021; Qian *et al.*, 2021; Shi *et al.*, 2023; T. Lin *et al.*, 2023). This approach did not seem reasonable in this work because the microbiome profiles with Illumina short-read sequencing did not reach species level. To find an answer to the posed questions, two methods were applied. Firstly, an analysis similar to that of Zacharia et al. (2018) was performed, who compared 16S rRNA gene sequencing to culture-based bacterial

detection of pediatric BALF specimen. In Table 17, the overlap between the species reported in culture and the genera sequenced by Illumina were assessed. Secondly, culture pathogenicity categories were compared directly with the categories established for the microbiome profiles.

The six most frequent taxa that were sequenced using the Illumina platform were presented in section 3.3.4: *Streptococcus*, *Staphylococcus*, *Pseudomonas*, *Prevotella 7*, *Veillonella* and *Rothia*. Some of them are common bacterial constituents of the lung microbiome in health (cf. Table 1), but some could represent typical pathogens causing pneumonia, e.g. *Str. pneumoniae*, *St. aureus*, *Ps. aeruginosa* (cf. Table 2). Of the 118 species identified in culture, the corresponding genera were detected by short-read sequencing in 91.5 % (Zymo) respectively 85.6 % (eMAG). These numbers were lower than findings from comparable studies by Zacharia et al. (2018) (97.1 %), by Dickson et al. (2014) (95.7%) and by Yoo et a. (2020) (92.0 %). False-negative results in Illumina could be due to e.g. a low microbial load (Yoo *et al.*, 2020). Of the 27 culture reports with potential pathogens, the corresponding genus was found in 81.5 % (Zymo) respectively 68.0 % (eMAG) of the cases. In 59.3 % (Zymo) respectively 64.0 % (eMAG) the potential pathogen belonged to the top five genera of the microbiome profiles. In other words: in about 40 % of the cases, the potentially pathogenic species was either not found at all by Illumina sequencing, or it did not stand out as dominant genus in the microbiome profiles. However, false-positive culture results are also possible. In multiple samples, genera of potential clinical relevance were detected by Illumina sequencing, while the culture result was negative, including genera like *Pseudomonas*, *Haemophilus* and *Staphylococcus*. Due to the resolution only up to genus level, the differentiation between e.g. *St. aureus* and *St. epidermidis* was not possible. Hence, identification at species level is of crucial importance for clinical use.

Two cultured genera were not detected at all by Illumina: *Citrobacter* which was cultured in two samples (sample 2 and 140), and *Serratia*, which was reported in one culture (sample 126). The genera could not be found among the top five BLAST results either. A potential explanation could be that the reads of those taxa were mistakenly assigned to another genus due to a very similar base sequence. In sample 2, the predominant genus in Illumina was *Escherichia-Shigella* with a relative abundance of 78.2 %, while Nanopore aligned with the culture results and identified *Cit. koseri* as the top species (69.6 %). The first result of *Cit. koseri* in BLAST for the sequence of *Escherichia-*

Shigella had a nucleotide sequence similarity of 99.2%. The OTUs were clustered according to a similarity of 97 %. In this sample, it is likely that the OTU was assigned to the wrong genus and that the sequencing depth of Illumina was insufficient.

Nanopore long-read sequencing was able to generate microbiome profiles at species-level in every sequenced sample. In 76.5 %, the cultured species was among the top Nanopore sequencing results, but only in 44.4 % the potential pathogens from culture were detected by Nanopore sequencing. The study design did not allow a conclusive answer to the question of whether the culture was false positive or if Nanopore was false negative. In two samples (9.5 %), potential pathogens were detected by Nanopore, but not by culture (sample 11: *Str. pneumoniae* and sample 34: *H. influenzae*). For the two most abundant taxa, Illumina and Nanopore showed a full or partial overlap in 77.8 % of the cases. Nanopore sequencing matched 22 of the 36 top Illumina taxa (61.1 %). This was in line with a study by Heikema et al. (2020) who found similar bacterial profiles in nasal samples from both platforms. In their review, Chapman and his colleagues (2023) discussed Nanopore sequencing as a promising method for the diagnosis of respiratory tract infections, highlighting the fast preparation time (possible turn-around time of under six hours), which is also supported by several other studies (Mitsuhashi *et al.*, 2017; Charalampous *et al.*, 2019; Wang *et al.*, 2020).

The microbiome profiles provided additional qualitative and quantitative information on the bacterial composition such as diversity indices in every successfully sequenced sample. As stated in the introduction, the microbiome of an infected lung is characterized by a high microbial mass and the dominance of one or few pathogenic genera which goes along with a low bacterial community diversity. Dickson et al. (2014) referred to these parameters as culture-independent indices of infection. In their study with BAL fluid, they showed a statistically significant relationship between these indices and bacterial isolation in culture. In this work, the DNA yield (i.e. concentration of double-stranded DNA measured in the extracts of BAL fluid) was significantly higher in the samples with a potential pathogen in culture (Figure 22). However, when looking at the individual values this metric could not be used to accurately distinguish these samples from those with a negative culture or a cultured commensal. Consequently, it could not be used as a single discriminator to determine whether a potential infection is present but rather as an additional puzzle piece to evaluate the condition of the lung microbiome. Kitsios et al. (2018) did not find significant differences in DNA yield between culture positive and

culture negative samples, but they identified the bacterial abundances in the sequencing results as the best predictor of culture positivity. This was consistent with the findings of this work, as the mono-microbial samples with cultured pathogens showed significantly higher relative abundances of the dominant genus than samples with a commensal or no growth (Figure 26). However, as with the DNA yield, this could not be used as a sharp predictor for culture positivity. Regarding the diversity indices, no significant differences could be observed between the different culture results (Table 16). This result contradicts a number of studies which associated lower diversity indices with isolated pathogens in culture (Kitsios *et al.*, 2018; Langelier *et al.*, 2018; Chen *et al.*, 2021). Culture-independent indices of infection have also been linked to clinical parameters. Dickson et al. (2020) observed that the outcome of critical patients could be predicted by key features of the lung microbiome such as microbial burden. And the ratio of procalcitonin / Simpson`s index (for α-diversity) was shown to be a valuable prognostic marker for patients with severe bacterial pneumonia (Sun *et al.*, 2023). Community instability of the human gut microbiome was shown to be a strong predictor for clinical disease activity (Knight *et al.*, 2018). If this applies also to the lung microbiome is to be investigated in longitudinal studies. Culture-independent indices of infection could provide valuable additional information for respiratory diagnostics; however, they must always be seen as part of a big picture.

The establishment of microbiome categories into mono-, poly- and multi-microbial helped to contextualize the profiles and to compare them with the culture results. However, it is an artificial classification, it simplifies the complex composition of the lung microbiome and the actual benefit of such categories in the clinical practice is questionable. The assumption on which the classification was based was that “mono-microbial” indicated a microbiome profile dominated by one taxon, which could be a sign of disturbed microbial homeostasis, as explained in the introduction. Additional to a dysbiosis, a diseased lung microbiome is characterized by a high microbial biomass (Fromentin, Ricard and Roux, 2021; Yagi *et al.*, 2021). In Figure 21, it was tested whether the mono-microbial samples had a higher DNA concentration than the poly- or multi-microbial sample. In Zymo, the visual findings showed a higher DNA concentration in the mono-microbial samples, but this could not be confirmed statistically after correction for multiple comparisons. In eMAG, the mono-microbial samples showed a higher DNA concentration only after excluding samples allegedly contaminated with *Pseudomonas*,

so that some of these may have been incorrectly assigned to the "mono-microbial" category.

All in all, the results of this work were consistent with existing studies, that 16S rRNA gene sequencing can (currently) not replace culture in the detection of bacteria in the lower respiratory tract. There is an overlap between culture results and the sequencing results, but they differ in a substantial number of cases and culture remains the gold standard. However, NGS could represent a valuable extension to increase the microbial detection rate, as many studies previously found (Yoo *et al.*, 2020; Qian *et al.*, 2021; Shi *et al.*, 2023). This applies especially for techniques which reach the sequencing depths of species-level, like Nanopore sequencing. The strengths of NGS are that it can detect a broad spectrum of bacteria and captures additional features of the complex ecosystem in the lungs (Wang, Han and Feng, 2019; Chen *et al.*, 2021). It can detect searched and non-searched organisms, eliminating the bias towards fast-growing or easily culturable species (Carney *et al.*, 2020; Tozzo, Delicati and Caenazzo, 2022). Besides, it can be applied in patients pretreated with antibiotics (Kitsios *et al.*, 2018; Yoo *et al.*, 2020) and therefore offers advantages, especially in critically ill patients with multiple risk factors, e.g. in intensive care units.

4.4 Controls, contamination assessment and quality evaluation

Contamination during sample preparation can have a profound impact on the sequencing results (Dickson, Erb-Downward, *et al.*, 2016). Distinguishing between sequencing data of the actual lung microbiome and contamination is one of the biggest challenges in respiratory microbiome studies due to the low bacterial biomass (Man, de Steenhuijsen Piters and Bogaert, 2017; Greathouse, Sinha and Vogtmann, 2019; Carney *et al.*, 2020), whereas in stool samples the sheer mass of gut microbiome usually exceeds potential contamination originated during sampling or laboratory workflow (Dickson, Erb-Downward, *et al.*, 2016).

Therefore, great focus was attached to performing and analyzing blank controls. Contamination during the passage of the bronchoscope through the upper respiratory tract is believed to be rather small, when conducted by a skilled bronchoscopist (Bassis *et al.*, 2015; Dickson *et al.*, 2017; Yagi *et al.*, 2021). This was in line with the findings of this

work. We analyzed six technical bronchoscope controls, three of them solely with the sodium chloride fluid used for the lavage (without flushing), and three that contained the sodium chloride fluid after flushing it through the bronchoscope. The six controls had a median number of read counts of 1,630 which was one of the lower numbers among the different control types (cf. Figure 8). The three controls with solely sodium chloride exhibited more read counts than the fluid after flushing through the bronchoscope (7,443 vs. 4,133), which suggested that the main polluter during bronchoscopy was the sodium chloride fluid used for the rinse and not the bronchoscope itself.

Additional sources of contamination include molecular biology grade water, PCR reagents and DNA extraction kits (Salter *et al.*, 2014), but when it comes to the specific polluter, findings differ. A study by Drengenes et al. (2019) found the extraction kit to be the major contributor of the contamination, which partly corresponded to the findings of this work. As shown in Figure 8, the eMAG kit controls were second highest in read counts. However, the read counts of the Zymo kit controls were at a very low level (except of one outlier). Major contributor in this work seemed to be the Zymo Shield control. Another study identified the PCR master mix as the main source of contamination (Stinson, Keelan and Payne, 2019), which could not be confirmed in this study.

In this work, the most prevalent genera present in the controls were *Aquabacterium, Burkholderia*, *Cutibacterium, Kocuria*, *Enhydrobacter*, *Escherichia, Pseudomonas, Psychrobacter* and *Rhodobacter*. The taxonomic profiles varied between control type, however these findings were overall conform with the common contaminant genera described in literature (Salter *et al.*, 2014; Drengenes *et al.*, 2019; Stinson, Keelan and Payne, 2019).

A difficult matter was how to deal with the genus *Pseudomonas*. As shown in Figure 25, *Pseudomonas* was often the dominant bacterium in the samples with both a mono-microbial profile and a negative culture, most notably among the eMAG extracts. *Pseudomonas* is known as a common contaminant (Salter *et al.*, 2014; Zinter *et al.*, 2019). On the other hand, *Pseudomonas* is also a dreaded pathogen in pneumonia, especially in hospital-acquired pneumonia and patients with risk factors for multi-resistant pathogens (Dalhoff *et al.*, 2017; Ewig *et al.*, 2021). The same fundamental issue applies to a number of other genera, such as *Staphylococcus* (Zinter *et al.*, 2019). Examining and evaluating

each case individually and including clinical patient data is crucial in such cases to distinguish between colonization and contamination.

A case-by-case approach to identify the contamination taxa might not be optimal, as it also relies on subjective judgement and therefore diminishes objectivity. Nevertheless, determining criteria and examining each sample closely appeared to be the best way to keep the balance between excluding contamination while leaving the original microbiome constituents. Further solutions to the problem of contamination identification could involve bioinformatic tools using statistical models, as described in some studies (Davis *et al.*, 2018; Drengenes *et al.*, 2019, 2021; Zinter *et al.*, 2019).

In Figure 7, a substantial difference between batches one and two and batch three was apparent and also statistically confirmed. This so-called batch effect (Faner *et al.*, 2017) most likely arose from differences in the laboratory workflow which underscores the susceptibility of the results to variations in the workflow, and consequently the importance of a highly standardized methodology.

4.5 NGS in respiratory research

This work followed an observational, cross-sectional approach, which has been one of the predominant study types in respiratory microbiome research (de Steenhuijsen Piters *et al.*, 2016; Young, 2017; Carney *et al.*, 2020; Yagi *et al.*, 2021). Observational studies are important, especially for a young research field, to explore microbial community structures in health and disease and to generate new research hypotheses. However, they don`t answer questions concerning causation or underlying (patho-)mechanisms since they neither track microbial changes over time nor directly manipulate variables. Longitudinal and interventional studies are needed, which are still rare in the field (Young, 2017; Zachariah *et al.*, 2018; Carney *et al.*, 2020; Yagi *et al.*, 2021; Pérez-Cobas *et al.*, 2023).

In this study, 16S rRNA gene sequencing and Nanopore sequencing were used to generate microbiome profiles. One important aspect to keep in mind when analyzing and comparing microbiome data is that different methods often produce different results. Along the way from DNA extraction to microbiome profiles there are many different choices to make, concerning the type of sequencing (e.g., 16S rRNA gene, metagenome

and metatranscriptome sequencing), the sequencing platform, the variable region (V1–3 or V4), the reference data bases for OTU assignment and bioinformatic processing workflows. Most decisions for or against a method will influence the sequencing result, as studies have shown (Knight *et al.*, 2018; Allaband *et al.*, 2019; Carney *et al.*, 2020; Hu *et al.*, 2020; Drengenes *et al.*, 2021; Meskini *et al.*, 2021).

The different types of sequencing have their own strengths and weaknesses. Marker gene sequencing, such as 16S rRNA, gives a broad overview on present taxa and their abundances within the microbiome. It is cost-effective and applicable to many different research questions and study designs. However, it produces profiles with a low resolution (genus-level), does not include viruses or fungi and can be biased towards certain bacteria, which arises from primer affinity and overamplification during PCR (Faner *et al.*, 2017; Knight *et al.*, 2018; Di Simone *et al.*, 2023). The V4 region, which was targeted in this work, is widely used and is supposed to generate an accurate map of the lung microbiome (Mendez *et al.*, 2019; Drengenes *et al.*, 2021). Whether and how 16S rRNA gene sequencing will find entry into the clinical practice remains to be evaluated in studies with a clinically-oriented approach (Bador *et al.*, 2020). While marker gene sequencing provides only phylogenetic information, metagenomic NGS also displays the functional activity (such as virulence genes and antimicrobial resistances) of a bacterial community (Young, 2017; Carney *et al.*, 2020). Combined host/ microbe mNGS is able to capture the dynamic interactions of pathogens, commensals and host responses, giving further insights into the mechanisms in the lungs during diseases such as pneumonia (Kalantar and Langelier, 2021).

The advancements of NGS techniques have resulted in reduced cost for microbiome research. The processing and analysis of the vast amount of high-dimensional data sets generated by NGS remains a challenge and is increasingly assisted by machine-learning algorithms (Man, de Steenhuijsen Piters and Bogaert, 2017; Knight *et al.*, 2018; Martin-Loeches *et al.*, 2020; Kalantar and Langelier, 2021). Additionally, the lack of standardized protocols and guidelines for NGS-based microbiome profiling poses difficulties in comparing studies and conducting meta-analyses (Faner *et al.*, 2017; Man, de Steenhuijsen Piters and Bogaert, 2017; Knight *et al.*, 2018; Abundo *et al.*, 2021; Yagi *et al.*, 2021).

All in all, much multidisciplinary groundwork is needed to establish standardized procedures for generating and processing microbiome data. Randomized controlled trials with combined omics-approaches are necessary to reveal underlying mechanisms and causations of the respiratory microbiome in health and disease (Faner *et al.*, 2017; Man, de Steenhuijsen Piters and Bogaert, 2017), also including the non-bacterial constituents of the microbiome (Young, 2017). In addition, geographical, environmental and ethnic differences need to be further investigated, as current knowledge is based on a biased research population mainly from North America and Europe (Gaulke and Sharpton, 2018; Allaband *et al.*, 2019; Skallevold, Vallenari and Sapkota, 2021; Abdill, Adamowicz and Blekhman, 2022; L. Lin *et al.*, 2023).

4.6 Transition into clinical practice

In the last section (4.5), it became clear, that the application of microbiome-based diagnostics and therapies is currently more the subject of research rather than practice. However, increasing knowledge, further development and decreasing costs will bring the establishment of NGS in a clinical settings within reach (Allaband *et al.*, 2019; Charalampous *et al.*, 2019; Chapman *et al.*, 2023). This section gives some examples for possible future microbiome-related applications in the clinical practice.

In diagnostics of infectious diseases, two scopes of application seem realistic. Firstly, microbiome profiles could be used for the detection of pathogens in combination with culture (Byrd and Segre, 2016; Flynn and Dooley, 2021; Meskini *et al.*, 2021; Liu *et al.*, 2022) or as rapid point-of-care approach (Wu and Segal, 2018). Secondly, features in the microbiome could be used as diagnostic or prognostic biomarkers (Faner *et al.*, 2017; de Koff *et al.*, 2021; Yagi *et al.*, 2021; Strouts, McAllister and Tsalik, 2023). For example, Dickson et al. showed in their prospective, observational cohort study, that some features of the lung microbiome could predict the outcome of ventilated ICU patients and that the presence of gut-associated bacteria (e.g., *E. coli*, *Enterobacter* spp., *Kl. pneumoniae*) in the lung was associated with the occurrence of acute respiratory distress syndrome (Dickson *et al.*, 2020). Microbiome characteristics might also be helpful for the planning of pharmacotherapies. The composition of the airway microbiomes is likely to influence the antibiotic susceptibility of *Ps. aeruginosa* in patients with cystic fibrosis (Reece,

Bettio and Renwick, 2021). And the metabolism of the gut microbiome could possibly affect the bioavailability of digoxin, a drug used for cardiac arrhythmias (Young, 2017).

Alterations in the lung microbiome have also been observed in patients with lung cancer. A study which compared BAL fluid of patients with lung cancer and patients with benign lung mass found differences in the microbial composition between these two groups (Lee *et al.*, 2016). Another study linked the development of metastases with an overrepresentation of *Legionella* (Yu *et al.*, 2016). There are more findings connecting specific changes in the microbiome of the lung or the saliva to different types of lung cancer and other clinical parameters (Skallevold, Vallenari and Sapkota, 2021; Wong-Rolle *et al.*, 2021; Yagi *et al.*, 2021; Bou Zerdan *et al.*, 2022; Marshall *et al.*, 2022), which raises the possibility that features of the lung microbiome may also serve as potential biomarker in oncology.

When considering the therapeutic potential of the microbiome, there are multiple possible strategies to target and manipulate the microbiome. Since changes in the microbiome were shown to play part in multiple diseases, the basic concept is to (re-)construct a beneficial microbiome structure (Young, 2017). Fecal microbiota transplantation (FMT) is a prime example for microbiome-based therapy, depicting an established treatment option for recurrent and severe antibiotics-associated *Clostridioides difficile* infections (CDI). Multiple studies have demonstrated the effectiveness of the administration of donor-feces to restore the homeostasis of the gut microbiome (Frisbee and Petri Jr, 2020). Since 2014, FMT is recommended in the guidelines of the European Society of Clinical Microbiology and Infectious Diseases for recurrent CDI (Debast *et al.*, 2014).

There is no clinically established therapeutical option regarding the lung microbiome, yet. Most approaches in their research stages target the lung microbiome through the gut-lung axis, by dietary manipulation, probiotics or microbial metabolites, such as short-chain fatty acids (Carney *et al.*, 2020; Martin-Loeches *et al.*, 2020). Fundamental and transitional research is needed, to develop diagnostic methods and personalized microbiome therapies that stand up to the demands of routine clinical application (Taroncher-Oldenburg *et al.*, 2018; Carney *et al.*, 2020; Chotirmall *et al.*, 2022).

5 Summary

Lower respiratory tract infections (LRTI) are a major cause for morbidity and mortality worldwide. A precise and time-effective pathogen identification is crucial for the successful treatment of patients with LRTI. At present, culture is the gold standard for routine diagnostics. However, in around half of the cases, pathogens remain undetected. Next-generation sequencing (NGS) represents a promising approach for an unbiased microbial detection and could provide additional information for clinical practice. In this descriptive work, bacterial microbiome profiles of 144 bronchoalveolar lavage fluid samples were generated with Illumina V4 16S rRNA gene sequencing and compared to the corresponding culture results. 21 samples were selected for Nanopore long-read sequencing, to reach a higher taxonomic resolution. Additionally, manual mechanical (Zymo) and automated enzymatic (eMAG) extraction methods were compared.

The results of the two extraction methods showed similarities in some respects (number of sequenced reads and the DNA yield), but variations in others (contamination and α-diversity). After the removal of the contamination genera, Zymo and eMAG presented with a similar community composition. The most abundant genera were *Streptococcus*, *Staphylococcus*, *Pseudomonas*, *Prevotella 7*, *Veillonella* and *Rothia*. The cultured bacteria could be retraced by Illumina sequencing (Zymo: 91.5%; eMAG: 85.6%), however, not always appearing among the top sequenced genera of the microbiome profiles. In 76.5 %, Nanopore sequencing result corresponded with the cultured species. In two cases, Nanopore sequencing detected a potential pathogen (*H. influenzae* and *Str. pneumoniae*), which was not reported in culture. A higher DNA yield could be observed in samples with a potential pathogen in culture. α-diversity did not vary significantly between the different culture results.

In conclusion, NGS results showed a partial overlap with culture as the current diagnostic gold standard in LRTI. NGS could represent a valuable extension to increase the microbial detection rate since it captures a broad spectrum of bacteria. Furthermore, NGS reveals additional microbial features of the complex ecosystem in the lungs.

6 Zusammenfassung

Infektionen der unteren Atemwege sind eine führende Ursache von Morbidität und Mortalität weltweit. Eine korrekte und rasche Identifikation der Krankheitserreger ist entscheidend für die erfolgreiche Behandlung von Patientinnen und Patienten. Derzeit ist die Kultur der diagnostische Goldstandard. In etwa der Hälfte der Fälle gelingt es jedoch nicht, die Krankheitserreger zu identifizieren. Next-Generation Sequencing (NGS) ist eine vielversprechende Methode für die mikrobielle Erregerbestimmung und könnte zusätzliche Informationen für die klinische Praxis liefern. In dieser deskriptiven Arbeit wurden bakterielle Mikrobiom-Profile von 144 bronchoalveolären Lavage-Proben mittels Illumina V4 16S rRNA Sequenzierung erstellt und mit den entsprechenden Kulturergebnissen verglichen. 21 Proben wurden zudem mit Nanopore sequenziert, um eine höhere taxonomische Auflösung zu erzielen. Zusätzlich wurden die manuelle, mechanische (Zymo) und die automatisierte, enzymatische (eMAG) Extraktion gegenübergestellt.

Beide Extraktionsmethoden generierten ähnliche Ergebnisse bei der Anzahl der Read Counts und der DNA-Konzentration. Unterschiede konnten bei Kontamination und α-Diversität beobachtet werden. Nach Entfernung der Kontamination wiesen Zymo und eMAG eine ähnliche taxonomische Zusammensetzung auf. Die häufigsten Gattungen waren *Streptococcus*, *Staphylococcus*, *Pseudomonas*, *Prevotella 7*, *Veillonella* und *Rothia*. Viele der kultivierten Taxa konnten durch Illumina-Sequenzierung detektiert werden (Zymo: 91,5 %; eMAG: 85,6 %), jedoch nicht immer als herausstechende Gattung innerhalb der Mikrobiom-Profile. In 76,5 % stimmte das Ergebnis der Nanopore-Sequenzierung mit der Kultur überein. In zwei Fällen detektierte Nanopore potenzielle Pathogene (*H. influenzae* und *Str. pneumoniae*), die nicht in der Kultur nachgewiesen werden konnten. In Proben mit einem potenziellen Pathogen in der Kultur konnte eine höhere extrahierte DNA-Konzentration verzeichnet werden, bei der α-Diversität gab es keine signifikanten Unterschiede zwischen den verschiedenen Kulturergebnissen.

Zusammenfassend wiesen die NGS-Ergebnisse teilweise Überschneidungen mit der Kultur auf, welche derzeit als diagnostischer Goldstandard gilt. NGS könnte sich als wertvolle Ergänzung erweisen, um die mikrobielle Nachweisrate zu steigern, da es ein breites Spektrum von Bakterien erfasst, und zusätzliche mikrobielle Charakteristika des komplexen Ökosystems in der Lunge beleuchtet.

7 Abbreviations

ARDS	acute respiratory distress syndrome
BAL(F)	bronchoalveolar lavage (fluid)
bp	base pair
CAP	community-acquired pneumonia
CDI	*Clostridioides difficile* infection
CF	cystic fibrosis
CO_2	carbon dioxide
COPD	chronic obstructive pulmonary disease
(ds)DNA	double-stranded desoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
FMT	fecal microbiota transplantation
HAP	hospital-acquired pneumonia
IQR	interquartile range
LRTI	lower respiratory tract infection
(m)NGS	(metagenomic) next-generation sequencing
MRSA	methicillin-resistant *Staphylococcus aureus*
nt	nucleotides
OTU	operational taxonomic unit
PCoA	principal coordinates analysis
PCR	polymerase chain reaction
rpm	revolutions per minute
(r)RNA	(ribosomal) ribonucleic acid
SARS-CoV-2	severe acute respiratory syndrome coronavirus type 2
UKGM	Universitätsklinikum Gießen und Marburg (University Hospital Giessen and Marburg)
uncl	unclassified

8 List of figures

9 List of tables

10 References

Abdill, R.J., Adamowicz, E.M. and Blekhman, R. (2022) 'Public human microbiome data are dominated by highly developed countries', *PLoS biology*, 20(2), p. e3001536. Available at: https://doi.org/10.1371/journal.pbio.3001536.

Abundo, M.E.C. *et al.* (2021) 'Assessment of two DNA extraction kits for profiling poultry respiratory microbiota from multiple sample types', *PloS One*, 16(1), p. e0241732. Available at: https://doi.org/10.1371/journal.pone.0241732.

Allaband, C. *et al.* (2019) 'Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians', *Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association*, 17(2), pp. 218–230. Available at: https://doi.org/10.1016/j.cgh.2018.09.017.

Altschul, S.F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Research*, 25(17), pp. 3389–3402. Available at: https://doi.org/10.1093/nar/25.17.3389.

Angebault, C. *et al.* (2020) 'Combined bacterial and fungal targeted amplicon sequencing of respiratory samples: Does the DNA extraction method matter?', *PloS One*, 15(4), p. e0232215. Available at: https://doi.org/10.1371/journal.pone.0232215.

Bador, J. *et al.* (2020) '16S rRNA PCR on clinical specimens: Impact on diagnosis and therapeutic management', *Médecine Et Maladies Infectieuses*, 50(1), pp. 63–73. Available at: https://doi.org/10.1016/j.medmal.2019.09.014.

Bassis, C.M. *et al.* (2015) 'Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals', *mBio*, 6(2), pp. e00037-15. Available at: https://doi.org/10.1128/mBio.00037-15.

Beckman Coulter (2022) *Möglichkeiten mit AMPure XP*. Available at: https://www.beckman.de/reagents/genomic/cleanup-and-size-selection/pcr/ampure-xp-capabilities (Accessed: 2 July 2023).

Bou Zerdan, Maroun *et al.* (2022) 'The Lung Microbiota and Lung Cancer: A Growing Relationship', *Cancers*, 14(19), p. 4813. Available at: https://doi.org/10.3390/cancers14194813.

Byrd, A.L. and Segre, J.A. (2016) 'Adapting Koch's postulates', *Science*, 351(6270), pp. 224–226. Available at: https://doi.org/10.1126/science.aad6753.

Carney, S.M. *et al.* (2020) 'Methods in Lung Microbiome Research', *American Journal of Respiratory Cell and Molecular Biology*, 62(3), pp. 283–299. Available at: https://doi.org/10.1165/rcmb.2019-0273TR.

Chapman, R. *et al.* (2023) 'Nanopore-Based Metagenomic Sequencing in Respiratory Tract Infection: A Developing Diagnostic Platform', *Lung*, 201(2), pp. 171–179. Available at: https://doi.org/10.1007/s00408-023-00612-y.

Charalampous, T. *et al.* (2019) 'Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection', *Nature Biotechnology*, 37(7), pp. 783–792. Available at: https://doi.org/10.1038/s41587-019-0156-5.

Chen, Y. *et al.* (2021) 'Application of Metagenomic Next-Generation Sequencing in the Diagnosis of Pulmonary Infectious Pathogens From Bronchoalveolar Lavage Samples', *Frontiers in Cellular and Infection Microbiology*, 11, p. 541092. Available at: https://doi.org/10.3389/fcimb.2021.541092.

Cho, I. and Blaser, M.J. (2012) 'The human microbiome: at the interface of health and disease', *Nature Reviews. Genetics*, 13(4), pp. 260–270. Available at: https://doi.org/10.1038/nrg3182.

Chotirmall, S.H. *et al.* (2022) 'Therapeutic Targeting of the Respiratory Microbiome', *American Journal of Respiratory and Critical Care Medicine*, 206(5), pp. 535–544. Available at: https://doi.org/10.1164/rccm.202112-2704PP.

Cox, M.J. *et al.* (2010) 'Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients', *PloS One*, 5(6), p. e11044. Available at: https://doi.org/10.1371/journal.pone.0011044.

Dalhoff, K. *et al.* (2017) *S3-Leitlinie zur Epidemiologie, Diagnostik und Therapie erwachsener Patienten mit nosokomialer Pneumonie – Update 2017*, *AWMF online*. Available at: https://register.awmf.org/assets/guidelines/020-013l_S3_Nosokomiale_Pneumonie_Erwachsener_2017-11.pdf (Accessed: 27 January 2024).

Das, S. *et al.* (2021) 'A prevalent and culturable microbiota links ecological balance to clinical stability of the human lung after transplantation', *Nature Communications*, 12(1), p. 2126. Available at: https://doi.org/10.1038/s41467-021-22344-4.

Davis, N.M. *et al.* (2018) 'Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data', *Microbiome*, 6(1), p. 226. Available at: https://doi.org/10.1186/s40168-018-0605-2.

Debast, S.B. *et al.* (2014) 'European Society of Clinical Microbiology and Infectious Diseases: Update of the Treatment Guidance Document for Clostridium difficile Infection', *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 20(Suppl 2), pp. 1–26. Available at: https://doi.org/10.1111/1469-0691.12418.

Di Simone, S.K. *et al.* (2023) 'Understanding respiratory microbiome-immune system interactions in health and disease', *Science Translational Medicine*, 15(678), p. eabq5126. Available at: https://doi.org/10.1126/scitranslmed.abq5126.

Dickson, R.P. *et al.* (2014) 'Analysis of culture-dependent versus culture-independent techniques for identification of bacteria in clinically obtained bronchoalveolar lavage fluid', *Journal of Clinical Microbiology*, 52(10), pp. 3605–3613. Available at: https://doi.org/10.1128/JCM.01028-14.

Dickson, R.P. *et al.* (2015) 'Spatial Variation in the Healthy Human Lung Microbiome and the Adapted Island Model of Lung Biogeography', *Annals of the American Thoracic*

Society, 12(6), pp. 821–830. Available at: https://doi.org/10.1513/AnnalsATS.201501-029OC.

Dickson, R.P., Singer, B.H., *et al.* (2016) 'Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome', *Nature Microbiology*, 1(10), p. 16113. Available at: https://doi.org/10.1038/nmicrobiol.2016.113.

Dickson, R.P., Erb-Downward, J.R., *et al.* (2016) 'The Microbiome and the Respiratory Tract', *Annual Review of Physiology*, 78, pp. 481–504. Available at: https://doi.org/10.1146/annurev-physiol-021115-105238.

Dickson, R.P. *et al.* (2017) 'Bacterial Topography of the Healthy Human Lower Respiratory Tract', *mBio*, 8(1), pp. e02287-16. Available at: https://doi.org/10.1128/mBio.02287-16.

Dickson, R.P. *et al.* (2018) 'The Lung Microbiota of Healthy Mice Are Highly Variable, Cluster by Environment, and Reflect Variation in Baseline Lung Innate Immunity', *American Journal of Respiratory and Critical Care Medicine*, 198(4), pp. 497–508. Available at: https://doi.org/10.1164/rccm.201711-2180OC.

Dickson, R.P. *et al.* (2020) 'Lung Microbiota Predict Clinical Outcomes in Critically Ill Patients', *American Journal of Respiratory and Critical Care Medicine*, 201(5), pp. 555–563. Available at: https://doi.org/10.1164/rccm.201907-1487OC.

Dickson, R.P. and Huffnagle, G.B. (2015) 'The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease', *PLoS Pathogens*, 11(7), p. e1004923. Available at: https://doi.org/10.1371/journal.ppat.1004923.

Drengenes, C. *et al.* (2019) 'Laboratory contamination in airway microbiome studies', *BMC Microbiology*, 19(1), p. 187. Available at: https://doi.org/10.1186/s12866-019-1560-1.

Drengenes, C. *et al.* (2021) 'Exploring protocol bias in airway microbiome studies: one versus two PCR steps and 16S rRNA gene region V3 V4 versus V4', *BMC genomics*, 22(1), p. 3. Available at: https://doi.org/10.1186/s12864-020-07252-z.

Ewig, S. *et al.* (2021) *S3-Leitlinie zur Behandlung von erwachsenen Patienten mit ambulant erworbener Pneumonie – Update 2021*, *AWMF online*. Available at: https://register.awmf.org/assets/guidelines/020-020l_S3_Behandlung-von-erwachsenen-Patienten-mit-ambulant-erworbener-Pneumonie__2021-05.pdf (Accessed: 27 January 2024).

Faner, R. *et al.* (2017) 'The microbiome in respiratory medicine: current challenges and future perspectives', *The European Respiratory Journal*, 49(4), p. 1602086. Available at: https://doi.org/10.1183/13993003.02086-2016.

Fang, X. *et al.* (2020) 'Diagnostic Value of Metagenomic Next-Generation Sequencing for the Detection of Pathogens in Bronchoalveolar Lavage Fluid in Ventilator-Associated Pneumonia Patients', *Frontiers in Microbiology*, 11, p. 599756. Available at: https://doi.org/10.3389/fmicb.2020.599756.

Flynn, M. and Dooley, J. (2021) 'The microbiome of the nasopharynx', *Journal of Medical Microbiology*, 70(6), p. 001368. Available at: https://doi.org/10.1099/jmm.0.001368.

Frisbee, A.L. and Petri Jr, W.A. (2020) 'Considering the Immune System during Fecal Microbiota Transplantation for Clostridioides difficile Infection', *Trends in Molecular Medicine*, 26(5), pp. 496–507. Available at: https://doi.org/10.1016/j.molmed.2020.01.009.

Fromentin, M., Ricard, J.-D. and Roux, D. (2021) 'Respiratory microbiome in mechanically ventilated patients: a narrative review', *Intensive Care Medicine*, 47(3), pp. 292–306. Available at: https://doi.org/10.1007/s00134-020-06338-2.

Gadsby, N.J. *et al.* (2016) 'Comprehensive Molecular Testing for Respiratory Pathogens in Community-Acquired Pneumonia', *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 62(7), pp. 817–823. Available at: https://doi.org/10.1093/cid/civ1214.

Gaulke, C.A. and Sharpton, T.J. (2018) 'The influence of ethnicity and geography on human gut microbiome composition', *Nature Medicine*, 24(10), pp. 1495–1496. Available at: https://doi.org/10.1038/s41591-018-0210-8.

GBD 2016 Causes of Death Collaborators (2017) 'Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016', *Lancet (London, England)*, 390(10100). Available at: https://doi.org/10.1016/S0140-6736(17)32152-9.

GBD 2016 Lower Respiratory Infections Collaborators (2018) 'Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016', *The Lancet. Infectious Diseases*, 18(11), pp. 1191–1210. Available at: https://doi.org/10.1016/S1473-3099(18)30310-4.

Greathouse, K.L., Sinha, R. and Vogtmann, E. (2019) 'DNA extraction for human microbiome studies: the issue of standardization', *Genome Biology*, 20(1), p. 212. Available at: https://doi.org/10.1186/s13059-019-1843-8.

Gryp, T. *et al.* (2020) 'Comparison of five assays for DNA extraction from bacterial cells in human faecal samples', *Journal of Applied Microbiology*, 129(2), pp. 378–388. Available at: https://doi.org/10.1111/jam.14608.

Gu, W. *et al.* (2021) 'Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids', *Nature Medicine*, 27(1), pp. 115–124. Available at: https://doi.org/10.1038/s41591-020-1105-z.

Gujju, V.R. *et al.* (2021) 'Bordetella bronchiseptica infections in patients with HIV/AIDS: A case report and review of the literature', *Medicine*, 100(51), p. e28244. Available at: https://doi.org/10.1097/MD.0000000000028244.

Hanada, S. *et al.* (2018) 'Respiratory Viral Infection-Induced Microbiome Alterations and Secondary Bacterial Pneumonia', *Frontiers in Immunology*, 9, p. 2640. Available at: https://doi.org/10.3389/fimmu.2018.02640.

Heikema, A.P. *et al.* (2020) 'Comparison of Illumina versus Nanopore 16S rRNA Gene Sequencing of the Human Nasal Microbiota', *Genes*, 11(9), p. 1105. Available at: https://doi.org/10.3390/genes11091105.

Hermans, S.M., Buckley, H.L. and Lear, G. (2018) 'Optimal extraction methods for the simultaneous analysis of DNA from diverse organisms and sample types', *Molecular Ecology Resources*, 18(3), pp. 557–569. Available at: https://doi.org/10.1111/1755-0998.12762.

Hoefnagels, I. *et al.* (2021) 'The Role of the Respiratory Microbiome and Viral Presence in Lower Respiratory Tract Infection Severity in the First Five Years of Life', *Microorganisms*, 9(7), p. 1446. Available at: https://doi.org/10.3390/microorganisms9071446.

Hu, Y. *et al.* (2020) 'Metagenomic analysis of the lung microbiome in pulmonary tuberculosis - a pilot study', *Emerging Microbes & Infections*, 9(1), pp. 1444–1452. Available at: https://doi.org/10.1080/22221751.2020.1783188.

Huffnagle, G.B., Dickson, R.P. and Lukacs, N.W. (2017) 'The respiratory tract microbiome and lung inflammation: a two-way street', *Mucosal Immunology*, 10(2), pp. 299–306. Available at: https://doi.org/10.1038/mi.2016.108.

Illumina (2023) *What is the PhiX Control v3 Library and what is its function in Illumina Next Generation Sequencing*. Available at: https://knowledge.illumina.com/library-preparation/general/library-preparation-general-reference_material-list/000001545 (Accessed: 2 July 2023).

Jain, S. *et al.* (2015) 'Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults', *The New England Journal of Medicine*, 373(5), pp. 415–427. Available at: https://doi.org/10.1056/NEJMoa1500245.

Kalantar, K.L. and Langelier, C.R. (2021) 'Host-Microbe Metagenomics: a Lens To Refocus Our Perspective on Infectious and Inflammatory Diseases', *mSystems*, 6(4), p. e0040421. Available at: https://doi.org/10.1128/mSystems.00404-21.

Kaul, D. *et al.* (2020) 'Microbiome disturbance and resilience dynamics of the upper respiratory tract during influenza A virus infection', *Nature Communications*, 11(1), p. 2537. Available at: https://doi.org/10.1038/s41467-020-16429-9.

Kennedy, N.A. *et al.* (2014) 'The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing', *PloS One*, 9(2), p. e88982. Available at: https://doi.org/10.1371/journal.pone.0088982.

Kim, D. *et al.* (2017) 'Optimizing methods and dodging pitfalls in microbiome research', *Microbiome*, 5(1), p. 52. Available at: https://doi.org/10.1186/s40168-017-0267-5.

Kitsios, G.D. *et al.* (2018) 'Respiratory Microbiome Profiling for Etiologic Diagnosis of Pneumonia in Mechanically Ventilated Patients', *Frontiers in Microbiology*, 9, p. 1413. Available at: https://doi.org/10.3389/fmicb.2018.01413.

Kitsios, G.D. (2018) 'Translating Lung Microbiome Profiles into the Next-Generation Diagnostic Gold Standard for Pneumonia: a Clinical Investigator's Perspective', *mSystems*, 3(2), pp. e00153-17. Available at: https://doi.org/10.1128/mSystems.00153-17.

Knight, R. *et al.* (2018) 'Best practices for analysing microbiomes', *Nature Reviews. Microbiology*, 16(7), pp. 410–422. Available at: https://doi.org/10.1038/s41579-018-0029-9.

de Koff, E.M. *et al.* (2021) 'Microbial and clinical factors are related to recurrence of symptoms after childhood lower respiratory tract infection', *ERJ Open Research*, 7(2), pp. 00939–02020. Available at: https://doi.org/10.1183/23120541.00939-2020.

Könönen, E. and Wade, W.G. (2015) 'Actinomyces and related organisms in human infections', *Clinical Microbiology Reviews*, 28(2), pp. 419–442. Available at: https://doi.org/10.1128/CMR.00100-14.

Kozich, J.J. *et al.* (2013) 'Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform', *Applied and Environmental Microbiology*, 79(17), pp. 5112–5120. Available at: https://doi.org/10.1128/AEM.01043-13.

Langelier, C. *et al.* (2018) 'Metagenomic Sequencing Detects Respiratory Pathogens in Hematopoietic Cellular Transplant Patients', *American Journal of Respiratory and Critical Care Medicine*, 197(4), pp. 524–528. Available at: https://doi.org/10.1164/rccm.201706-1097LE.

Lee, S.H. *et al.* (2016) 'Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions', *Lung Cancer (Amsterdam, Netherlands)*, 102, pp. 89–95. Available at: https://doi.org/10.1016/j.lungcan.2016.10.016.

Leitao Filho, F.S. *et al.* (2023) 'Characterization of the Lower Airways and Oral Microbiota in Healthy Young Persons in the Community', *Biomedicines*, 11(3), p. 841. Available at: https://doi.org/10.3390/biomedicines11030841.

Li, Y. *et al.* (2020) 'Application of metagenomic next-generation sequencing for bronchoalveolar lavage diagnostics in critically ill patients', *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology*, 39(2), pp. 369–374. Available at: https://doi.org/10.1007/s10096-019-03734-5.

Lim, M.Y. *et al.* (2018) 'Comparison of DNA extraction methods for human gut microbial community profiling', *Systematic and Applied Microbiology*, 41(2), pp. 151–157. Available at: https://doi.org/10.1016/j.syapm.2017.11.008.

Lin, L. *et al.* (2023) 'The airway microbiome mediates the interaction between environmental exposure and respiratory health in humans', *Nature Medicine*, 29(7), pp. 1750–1759. Available at: https://doi.org/10.1038/s41591-023-02424-2.

Lin, T. *et al.* (2023) 'Microbiological diagnostic performance of metagenomic next-generation sequencing compared with conventional culture for patients with community-

acquired pneumonia', *Frontiers in Cellular and Infection Microbiology*, 13, p. 1136588. Available at: https://doi.org/10.3389/fcimb.2023.1136588.

Liu, H. *et al.* (2022) 'Diagnostic Significance of Metagenomic Next-Generation Sequencing for Community-Acquired Pneumonia in Southern China', *Frontiers in Medicine*, 9, p. 807174. Available at: https://doi.org/10.3389/fmed.2022.807174.

Mahnic, A. *et al.* (2021) 'Comparison Between Cultivation and Sequencing Based Approaches for Microbiota Analysis in Swabs and Biopsies of Chronic Wounds', *Frontiers in Medicine*, 8, p. 607255. Available at: https://doi.org/10.3389/fmed.2021.607255.

Mallott, E.K., Malhi, R.S. and Amato, K.R. (2019) 'Assessing the comparability of different DNA extraction and amplification methods in gut microbial community profiling', *Access Microbiology*, 1(7), p. e000060. Available at: https://doi.org/10.1099/acmi.0.000060.

Mammen, M.J. and Sethi, S. (2016) 'COPD and the microbiome', *Respirology (Carlton, Vic.)*, 21(4), pp. 590–599. Available at: https://doi.org/10.1111/resp.12732.

Man, W.H., de Steenhuijsen Piters, W.A.A. and Bogaert, D. (2017) 'The microbiota of the respiratory tract: gatekeeper to respiratory health', *Nature Reviews. Microbiology*, 15(5), pp. 259–270. Available at: https://doi.org/10.1038/nrmicro.2017.14.

Marshall, E.A. *et al.* (2022) 'Distinct bronchial microbiome precedes clinical diagnosis of lung cancer', *Molecular Cancer*, 21(1), p. 68. Available at: https://doi.org/10.1186/s12943-022-01544-6.

Marsland, B.J., Trompette, A. and Gollwitzer, E.S. (2015) 'The Gut–Lung Axis in Respiratory Disease', *Annals of the American Thoracic Society*, 12(Suppl 2), pp. S150–S156. Available at: https://doi.org/10.1513/AnnalsATS.201503-133AW.

Martin-Loeches, I. *et al.* (2020) 'The importance of airway and lung microbiome in the critically ill', *Critical Care (London, England)*, 24(1), p. 537. Available at: https://doi.org/10.1186/s13054-020-03219-4.

Mattei, V. *et al.* (2019) 'Evaluation of Methods for the Extraction of Microbial DNA From Vaginal Swabs Used for Microbiome Studies', *Frontiers in Cellular and Infection Microbiology*, 9, p. 197. Available at: https://doi.org/10.3389/fcimb.2019.00197.

Mendez, R. *et al.* (2019) 'Lung inflammation and disease: A perspective on microbial homeostasis and metabolism', *IUBMB Life*, 71(2), pp. 152–165. Available at: https://doi.org/10.1002/iub.1969.

Meskini, M. *et al.* (2021) 'An Overview on the Upper and Lower Airway Microbiome in Cystic Fibrosis Patients', *Tanaffos*, 20(2), pp. 86–98.

Mitsuhashi, S. *et al.* (2017) 'A portable system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer', *Scientific Reports*, 7(1), p. 5657. Available at: https://doi.org/10.1038/s41598-017-05772-5.

Moffatt, M.F. and Cookson, W.O. (2017) 'The lung microbiome in health and disease', *Clinical Medicine (London, England)*, 17(6), pp. 525–529. Available at: https://doi.org/10.7861/clinmedicine.17-6-525.

Molecular Probes - Invitrogen (2008) *Quant-iT^{TM} PicoGreen ® dsDNA Reagent and Kits*. Available at: https://tools.thermofisher.com/content/sfs/manuals/mp07581.pdf (Accessed: 2 July 2023).

Motulsky, H. (2017) *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking*. 4th edn. New York: Oxford University Press.

Muscedere, J.G. *et al.* (2012) 'The adequacy of timely empiric antibiotic therapy for ventilator-associated pneumonia: an important determinant of outcome', *Journal of Critical Care*, 27(3), p. 322.e7–14. Available at: https://doi.org/10.1016/j.jcrc.2011.09.004.

Natalini, J.G., Singh, S. and Segal, L.N. (2023) 'The dynamic lung microbiome in health and disease', *Nature Reviews. Microbiology*, 21(4), pp. 222–235. Available at: https://doi.org/10.1038/s41579-022-00821-x.

NCBI RefSeq Targeted Loci Project (2019) *NCBI RefSeq Targeted Loci Project*. Available at: https://www.ncbi.nlm.nih.gov/refseq/targetedloci/ (Accessed: 29 February 2024).

Neuberger-Castillo, L. *et al.* (2020) 'Method Validation for Extraction of DNA from Human Stool Samples for Downstream Microbiome Analysis', *Biopreservation and Biobanking*, 18(2), pp. 102–116. Available at: https://doi.org/10.1089/bio.2019.0112.

Ney, L.-M. *et al.* (2023) 'Short chain fatty acids: key regulators of the local and systemic immune response in inflammatory diseases and infections', *Open Biology*, 13(3), p. 230014. Available at: https://doi.org/10.1098/rsob.230014.

Oren, A. *et al.* (2023) 'International Code of Nomenclature of Prokaryotes. Prokaryotic Code (2022 Revision)', *International Journal of Systematic and Evolutionary Microbiology*, 73(5a). Available at: https://doi.org/10.1099/ijsem.0.005585.

Pérez-Cobas, A.E. *et al.* (2023) 'The respiratory tract microbiome, the pathogen load, and clinical interventions define severity of bacterial pneumonia', *Cell Reports. Medicine*, 4(9), p. 101167. Available at: https://doi.org/10.1016/j.xcrm.2023.101167.

Qian, Y.-Y. *et al.* (2021) 'Improving Pulmonary Infection Diagnosis with Metagenomic Next Generation Sequencing', *Frontiers in Cellular and Infection Microbiology*, 10, p. 567615. Available at: https://doi.org/10.3389/fcimb.2020.567615.

Quast, C. *et al.* (2013) 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic Acids Research*, 41(Database issue), pp. D590-596. Available at: https://doi.org/10.1093/nar/gks1219.

Reece, E., Bettio, P.H. de A. and Renwick, J. (2021) 'Polymicrobial Interactions in the Cystic Fibrosis Airway Microbiome Impact the Antimicrobial Susceptibility of Pseudomonas aeruginosa', *Antibiotics (Basel, Switzerland)*, 10(7), p. 827. Available at: https://doi.org/10.3390/antibiotics10070827.

Rhoades, N.S. *et al.* (2021) 'Acute SARS-CoV-2 infection is associated with an increased abundance of bacterial pathogens, including Pseudomonas aeruginosa in the nose', *Cell Reports*, 36(9), p. 109637. Available at: https://doi.org/10.1016/j.celrep.2021.109637.

Rognes, T. *et al.* (2016) 'VSEARCH: a versatile open source tool for metagenomics', *PeerJ*, 4, p. e2584. Available at: https://doi.org/10.7717/peerj.2584.

Rosenbaum, J. *et al.* (2019) 'Evaluation of Oral Cavity DNA Extraction Methods on Bacterial and Fungal Microbiota', *Scientific Reports*, 9(1), p. 1531. Available at: https://doi.org/10.1038/s41598-018-38049-6.

Salter, S.J. *et al.* (2014) 'Reagent and laboratory contamination can critically impact sequence-based microbiome analyses', *BMC Biology*, 12, p. 87. Available at: https://doi.org/10.1186/s12915-014-0087-z.

Santiago-Rodriguez, T.M. *et al.* (2020) 'Metagenomic Information Recovery from Human Stool Samples Is Influenced by Sequencing Depth and Profiling Method', *Genes*, 11(11), p. 1380. Available at: https://doi.org/10.3390/genes11111380.

Schloss, P.D. *et al.* (2009) 'Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities', *Applied and Environmental Microbiology*, 75(23), pp. 7537–7541. Available at: https://doi.org/10.1128/AEM.01541-09.

Segal, L.N. *et al.* (2016) 'Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype', *Nature Microbiology*, 1, p. 16031. Available at: https://doi.org/10.1038/nmicrobiol.2016.31.

Shaffer, J.P. *et al.* (2022) 'A comparison of six DNA extraction protocols for 16S, ITS and shotgun metagenomic sequencing of microbial communities', *BioTechniques*, 73(1), pp. 34–46. Available at: https://doi.org/10.2144/btn-2022-0032.

Shi, X.-Q. *et al.* (2023) 'Metagenomic next-generation sequencing vs. conventional detection methods for detecting the pulmonary infections', *European Review for Medical and Pharmacological Sciences*, 27(10), pp. 4752–4763. Available at: https://doi.org/10.26355/eurrev_202305_32486.

Shoar, S. and Musher, D.M. (2020) 'Etiology of community-acquired pneumonia in adults: a systematic review', *Pneumonia (Nathan Qld.)*, 12, p. 11. Available at: https://doi.org/10.1186/s41479-020-00074-3.

Skallevold, H.E., Vallenari, E.M. and Sapkota, D. (2021) 'Salivary Biomarkers in Lung Cancer', *Mediators of Inflammation*, 2021, p. 6019791. Available at: https://doi.org/10.1155/2021/6019791.

de Steenhuijsen Piters, W.A.A. *et al.* (2016) 'Dysbiosis of upper respiratory tract microbiota in elderly pneumonia patients', *The ISME journal*, 10(1), pp. 97–108. Available at: https://doi.org/10.1038/ismej.2015.99.

Stinson, L.F., Keelan, J.A. and Payne, M.S. (2019) 'Identification and removal of contaminating microbial DNA from PCR reagents: impact on low-biomass microbiome

analyses', *Letters in Applied Microbiology*, 68(1), pp. 2–8. Available at: https://doi.org/10.1111/lam.13091.

Stricker, S. *et al.* (2022) 'Respiratory and Intestinal Microbiota in Pediatric Lung Diseases—Current Evidence of the Gut–Lung Axis', *International Journal of Molecular Sciences*, 23(12), p. 6791. Available at: https://doi.org/10.3390/ijms23126791.

Strouts, F.R., McAllister, L.B. and Tsalik, E.L. (2023) 'Viewing both sides of the coin for infectious disease diagnosis', *The Journal of Clinical Investigation*, 133(8), p. e169242. Available at: https://doi.org/10.1172/JCI169242.

Sui, H.-Y. *et al.* (2020) 'Impact of DNA Extraction Method on Variation in Human and Built Environment Microbial Community and Functional Profiles Assessed by Shotgun Metagenomics Sequencing', *Frontiers in Microbiology*, 11, p. 953. Available at: https://doi.org/10.3389/fmicb.2020.00953.

Sulaiman, I. *et al.* (2021) 'Microbial signatures in the lower airways of mechanically ventilated COVID-19 patients associated with poor clinical outcome', *Nature Microbiology*, 6(10), pp. 1245–1258. Available at: https://doi.org/10.1038/s41564-021-00961-5.

Sun, G. *et al.* (2023) 'Ratio of procalcitonin/Simpson's dominance index predicted the short-term prognosis of patients with severe bacterial pneumonia', *Frontiers in Cellular and Infection Microbiology*, 13, p. 1175747. Available at: https://doi.org/10.3389/fcimb.2023.1175747.

Tarabichi, Y. *et al.* (2015) 'The administration of intranasal live attenuated influenza vaccine induces changes in the nasal microbiota and nasal epithelium gene expression profiles', *Microbiome*, 3, p. 74. Available at: https://doi.org/10.1186/s40168-015-0133-2.

Taroncher-Oldenburg, G. *et al.* (2018) 'Translating microbiome futures', *Nature Biotechnology*, 36(11), pp. 1037–1042. Available at: https://doi.org/10.1038/nbt.4287.

Teng, F. *et al.* (2018) 'Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling', *Scientific Reports*, 8(1), p. 16321. Available at: https://doi.org/10.1038/s41598-018-34294-x.

Thermo Fisher Scientific (2020) *Qubit™ 1X dsDNA HS Assay Kits*. Available at: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017455_Qubit_1X_dsDNA_HS_Assay_Kit_UG.pdf (Accessed: 2 July 2023).

Thermo Fisher Scientific (2021) *Qubit fluorometers and assays Accurate , specific , and sensitive quantification*. Available at: https://static.fishersci.eu/content/dam/fishersci/en_EU/lifescience/16419_NEW_Minisite/Molecular_Biology/Qubit_Fluorometers_and_Assays.pdf (Accessed: 2 July 2023).

Tian, L. *et al.* (2020) 'Deciphering functional redundancy in the human microbiome', *Nature Communications*, 11(1), p. 6217. Available at: https://doi.org/10.1038/s41467-020-19940-1.

Tozzo, P., Delicati, A. and Caenazzo, L. (2022) 'Human microbiome and microbiota identification for preventing and controlling healthcare-associated infections: A systematic review', *Frontiers in Public Health*, 10, p. 989496. Available at: https://doi.org/10.3389/fpubh.2022.989496.

Vesty, A. *et al.* (2017) 'Evaluating the Impact of DNA Extraction Method on the Representation of Human Oral Bacterial and Fungal Communities', *PloS One*, 12(1), p. e0169877. Available at: https://doi.org/10.1371/journal.pone.0169877.

Vitiello, A., Ferrara, F. and Zovi, A. (2023) 'The direct correlation between microbiota and SARS-CoV-2 infectious disease', *Inflammopharmacology*, 31(2), pp. 603–610. Available at: https://doi.org/10.1007/s10787-023-01145-9.

Wang, J., Han, Y. and Feng, J. (2019) 'Metagenomic next-generation sequencing for mixed pulmonary infection diagnosis', *BMC Pulmonary Medicine*, 19(1), p. 252. Available at: https://doi.org/10.1186/s12890-019-1022-4.

Wang, K. *et al.* (2020) 'Metagenomic Diagnosis for a Culture-Negative Sample From a Patient With Severe Pneumonia by Nanopore and Next-Generation Sequencing', *Frontiers in Cellular and Infection Microbiology*, 10, p. 182. Available at: https://doi.org/10.3389/fcimb.2020.00182.

Welp, A.L. and Bomberger, J.M. (2020) 'Bacterial Community Interactions During Chronic Respiratory Disease', *Frontiers in Cellular and Infection Microbiology*, 10, p. 213. Available at: https://doi.org/10.3389/fcimb.2020.00213.

Widder, S. *et al.* (2022) 'Association of bacterial community types, functional microbial processes and lung disease in cystic fibrosis airways', *The ISME journal*, 16(4), pp. 905–914. Available at: https://doi.org/10.1038/s41396-021-01129-z.

Willner, D. *et al.* (2012) 'Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples', *PloS One*, 7(4), p. e34605. Available at: https://doi.org/10.1371/journal.pone.0034605.

Wong-Rolle, A. *et al.* (2021) 'Unexpected guests in the tumor microenvironment: microbiome in cancer', *Protein & Cell*, 12(5), pp. 426–435. Available at: https://doi.org/10.1007/s13238-020-00813-8.

Woodhead, M. *et al.* (2011) 'Guidelines for the management of adult lower respiratory tract infections - summary', *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 17(Suppl 6), pp. 1–24. Available at: https://doi.org/10.1111/j.1469-0691.2011.03602.x.

Wu, B.G. *et al.* (2021) 'Episodic Aspiration with Oral Commensals Induces a MyD88-dependent, Pulmonary T-Helper Cell Type 17 Response that Mitigates Susceptibility to Streptococcus pneumoniae', *American Journal of Respiratory and Critical Care Medicine*, 203(9), pp. 1099–1111. Available at: https://doi.org/10.1164/rccm.202005-1596OC.

Wu, B.G. and Segal, L.N. (2018) 'The Lung Microbiome and Its Role in Pneumonia', *Clinics in Chest Medicine*, 39(4), pp. 677–689. Available at: https://doi.org/10.1016/j.ccm.2018.07.003.

Wypych, T.P., Wickramasinghe, L.C. and Marsland, B.J. (2019) 'The influence of the microbiome on respiratory health', *Nature Immunology*, 20(10), pp. 1279–1290. Available at: https://doi.org/10.1038/s41590-019-0451-9.

Xie, F. *et al.* (2021) 'Clinical metagenomics assessments improve diagnosis and outcomes in community-acquired pneumonia', *BMC Infectious Diseases*, 21(1), p. 352. Available at: https://doi.org/10.1186/s12879-021-06039-1.

Yadav, M., Verma, M.K. and Chauhan, N.S. (2018) 'A review of metabolic potential of human gut microbiome in human nutrition', *Archives of Microbiology*, 200(2), pp. 203–217. Available at: https://doi.org/10.1007/s00203-017-1459-x.

Yagi, K. *et al.* (2021) 'The Lung Microbiome during Health and Disease', *International Journal of Molecular Sciences*, 22(19), p. 10872. Available at: https://doi.org/10.3390/ijms221910872.

Yang, L. *et al.* (2019) 'Metagenomic identification of severe pneumonia pathogens in mechanically-ventilated patients: a feasibility and clinical validity study', *Respiratory Research*, 20(1), p. 265. Available at: https://doi.org/10.1186/s12931-019-1218-4.

Yao, Y. *et al.* (2021) 'Carbapenem-Resistant Citrobacter spp. as an Emerging Concern in the Hospital-Setting: Results From a Genome-Based Regional Surveillance Study', *Frontiers in Cellular and Infection Microbiology*, 11, p. 744431. Available at: https://doi.org/10.3389/fcimb.2021.744431.

Yoo, I.Y. *et al.* (2020) 'Comparison of 16S Ribosomal RNA Targeted Sequencing and Culture for Bacterial Identification in Normally Sterile Body Fluid Samples: Report of a 10-Year Clinical Laboratory Review', *Annals of Laboratory Medicine*, 40(1), pp. 63–67. Available at: https://doi.org/10.3343/alm.2020.40.1.63.

Young, V.B. (2017) 'The role of the microbiome in human health and disease: an introduction for clinicians', *BMJ (Clinical research ed.)*, 356, p. j831. Available at: https://doi.org/10.1136/bmj.j831.

Yu, G. *et al.* (2016) 'Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features', *Genome Biology*, 17(1), p. 163. Available at: https://doi.org/10.1186/s13059-016-1021-1.

Yu, K.-M. *et al.* (2023) 'Optimization of DNA extraction and sampling methods for successful forensic microbiome analyses of the skin and saliva', *International Journal of Legal Medicine*, 137(1), pp. 63–77. Available at: https://doi.org/10.1007/s00414-022-02919-6.

Zachariah, P. *et al.* (2018) 'Culture-Independent Analysis of Pediatric Bronchoalveolar Lavage Specimens', *Annals of the American Thoracic Society*, 15(9), pp. 1047–1056. Available at: https://doi.org/10.1513/AnnalsATS.201802-146OC.

Zhu, N. *et al.* (2023) 'Identification and comparison of Chlamydia psittaci, Legionella and Mycoplasma pneumonia infection', *The Clinical Respiratory Journal*, 17(5), pp. 384–393. Available at: https://doi.org/10.1111/crj.13603.

Zinter, M.S. *et al.* (2019) ‘Towards precision quantification of contamination in metagenomic sequencing experiments’, *Microbiome*, 7(1), p. 62. Available at: https://doi.org/10.1186/s40168-019-0678-6.

Appendix

Appendix 1: Permission granting document

Permission Granting Document

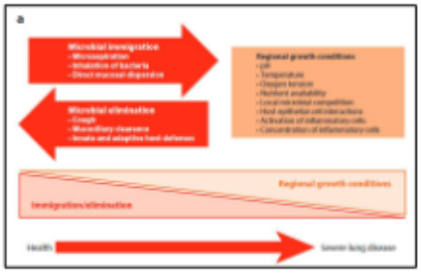
I, Dr. Robert Dickson, hereby grant permission to Merle Bitter to utilize the figure described below in her dissertation titled "Microbiome-based clinical pathogen detection in bronchoalveolar lavage fluid using Next-Generation-Sequencing."

Figure Information:

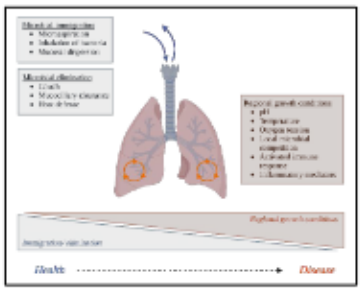
- **Figure Title or Description:** Ecological modeling of the respiratory microbiome.
- **Figure Number (if applicable):** Figure 2a.
- **Source/Publication:** Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The Microbiome and the Respiratory Tract. Annu Rev Physiol. 2016;78:481-504. doi:10.1146/annurev-physiol-021115-105238
- **Context of Use in the dissertation:** Explicating the role of the lung microbiome in acute lower respiratory tract infections as part of the introduction. The figure was slightly altered for the use in the dissertation.

Conditions and Acknowledgments:

1. The use of the figure is restricted to the specified dissertation.
2. Proper attribution will be given in the following format: "Figure adapted with permission by Dickson et al., 2016 [Reference of the paper]."
3. Modifications to the figure were made with the express approval of the original copyright holder.



Original Figure



Adapted Figure

I acknowledge that I have read and understood the terms and conditions outlined above and grant permission for the specified use of the figure in the dissertation.

[redacted]

Copyright Holder's Full Name

March 6, 2024

Signature and Date

Figure 29: Permission granting document.

Appendix 2: Culture-based microbial identification – further results

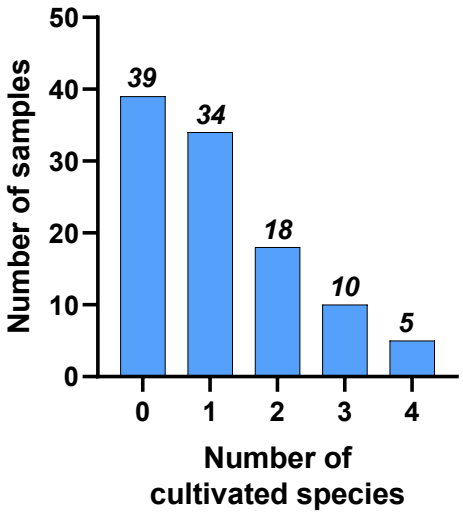


Figure 30: Number of cultivated species in all 106 samples.

Respiratory PCR panel

In 81 samples a respiratory PCR Panel was performed as part of routine diagnostics. No bacterial DNA was detected in 73 of these 81 samples. In four samples, only the DNA of *H. influenzae* was found. In three samples, only the DNA of *Str. pneumoniae* was reported and in one sample, records of nucleic acid from both *H. influenzae* and *Str. pneumoniae* were listed.

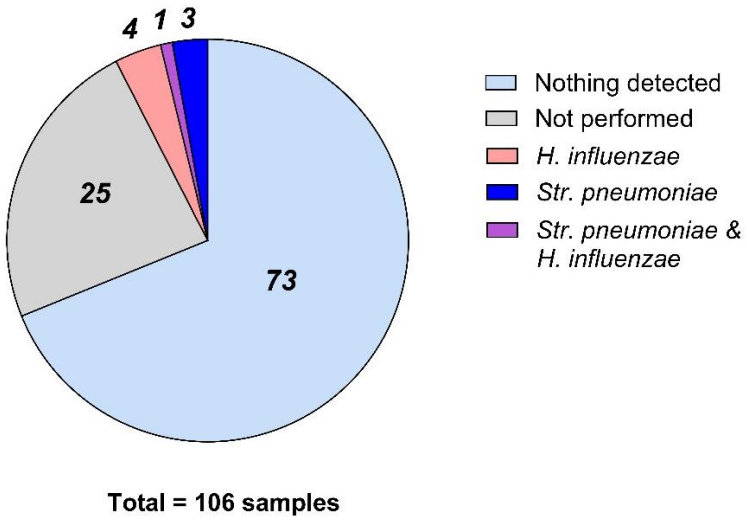


Figure 31: Results of the respiratory PCR panel.

Appendix 3: Pathogenicity evaluation for the results of culture-based routine diagnostics

Sample ID	***Possible Pathogen***	***Amount***	***Microscopy***	***PCR Panel***
2	*Cit. koseri*	plenty	Granulocytes 10-25 No epithelial cells	Not performed
3	-	-	Granulocytes >25 No epithelial cells	Nothing detected
4	*St. aureus*	plenty	Granulocytes >25 Epithelial cells <10	Nothing detected
7	-	-	Granulocytes <10 No epithelial cells	Nothing detected
8	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
9	-	-	No Granulocytes No epithelial cells	Nothing detected
11	-	-	Granulocytes >25 No epithelial cells	Not performed
13	*St. aureus*	mass	No information	Not performed
14	*St. aureus*	mass	No information	Not performed
15	-	-	Granulocytes >25 Epithelial cells <10	*H. influenzae*
17	*Str. pneumoniae*	mass	Granulocytes >25 No epithelial cells	*Str. pneumoniae*
18	-	-	Granulocytes >25 No epithelial cells	Nothing detected
19	-	-	Granulocytes 10-25 No epithelial cells	Nothing detected
21	-	-	No Granulocytes No epithelial cells	Not performed
22	*Ps. aeruginosa*	sparse	No Granulocytes No epithelial cells	Not performed
23	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
24	*H. influenzae*	mass	Granulocytes >25 No epithelial cells	*H. influenzae*
25	-	-	Granulocytes >25 No epithelial cells	Nothing detected
27	*St. aureus*	mass	Granulocytes >25 No epithelial cells	Not performed
28	-	-	Granulocytes >25 No epithelial cells	Nothing detected
29	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
33	-	-	Granulocytes <10 No epithelial cells	Nothing detected

34	-	-	Granulocytes >25 No epithelial cells	*H. influenzae*
36	-	-	Granulocytes >25 Epithelial cells 10-25	Nothing detected
38	-	-	Granulocytes <10 No epithelial cells	Nothing detected
39	-	-	No Granulocytes No epithelial cells	Nothing detected
40	-	-	No Granulocytes No epithelial cells	Nothing detected
41	-	-	Granulocytes <10 Epithelial cells <10	Nothing detected
42	-	-	Granulocytes 10-25 No epithelial cells	Nothing detected
43	*Ps. aeruginosa*	few	Granulocytes >25 No epithelial cells	Nothing detected
44	*Ps. aeruginosa*	plenty	Granulocytes >25 No epithelial cells	Nothing detected
45	-	-	Granulocytes >25 No epithelial cells	Nothing detected
48	-	-	Granulocytes >25 No epithelial cells	Nothing detected
49	-	-	Granulocytes >25 No epithelial cells	Nothing detected
50	-	-	Granulocytes >25 No epithelial cells	Nothing detected
51	-	-	Granulocytes 10-25 Epithelial cells <10	Nothing detected
52	*St. aureus*	few	Granulocytes 10-25 Epithelial cells <10	Not performed
55	-	-	Granulocytes >25 No epithelial cells	Nothing detected
57	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
58	-	-	Granulocytes >25 No epithelial cells	Nothing detected
59	-	-	Granulocytes <10 Epithelial cells <10	Nothing detected
60	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
65	-	-	Granulocytes 10-25 No epithelial cells	Nothing detected
70	-	-	Granulocytes >25 Epithelial cells <10	Not performed
71	-	-	Granulocytes >25 No epithelial cells	Nothing detected

72	*Ste. maltophilia*	plenty	Granulocytes >25 No epithelial cells	Nothing detected
73	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
74	-	-	No Granulocytes No epithelial cells	Nothing detected
75	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
76	-	-	Granulocytes >25 No epithelial cells	Not performed
77	-	-	Granulocytes >25 No epithelial cells	Nothing detected
79	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
80	-	-	Granulocytes >25 Epithelial cells <10	Not performed
82	*E. coli*	plenty	Granulocytes 10-25 Epithelial cells <10	Not performed
83	-	-	Granulocytes >25 Epithelial cells <10	Not performed
85	-	-	Granulocytes <10 No epithelial cells	Nothing detected
86	-	-	Granulocytes >25 No epithelial cells	Nothing detected
88	-	-	Granulocytes >25 No epithelial cells	Nothing detected
89	-	-	Granulocytes 10-25 No epithelial cells	Nothing detected
91	*St. aureus*	sparse	No Granulocytes No epithelial cells	Nothing detected
92	-	-	No Granulocytes No epithelial cells	Nothing detected
93	*St. aureus*	sparse	No information	*Str. pneumoniae;* *H. influenzae*
94	-	-	No Granulocytes No epithelial cells	Not performed
95	*Ps. aeruginosa*	mass	Granulocytes >25 Epithelial cells <10	Not performed
96	-	-	Granulocytes <10 Epithelial cells <10	Nothing detected
97	*Bor. bronchiseptica*	few	Granulocytes 10-25 No epithelial cells	Nothing detected
98	-	-	Granulocytes >25 No epithelial cells	Nothing detected
100	-	-	No Granulocytes No epithelial cells	Not performed

101	-	-	Granulocytes 10-25 Epithelial cells <10	Nothing detected
102	*St. aureus*	few	No information	Nothing detected
103	-	-	Granulocytes 10-25 Epithelial cells <10	Nothing detected
104	-	-	Granulocytes >25 No epithelial cells	Not performed
105	-	-	Granulocytes >25 No epithelial cells	Not performed
106	-	-	No information	Nothing detected
107	*Ps. aeruginosa*	plenty	Granulocytes >25 Epithelial cells <10	Nothing detected
108	-	-	Granulocytes <10 Epithelial cells <10	Not performed
109	-	-	Granulocytes 10-25 No epithelial cells	Nothing detected
110	*Ste. maltophilia*	sparse	No Granulocytes No epithelial cells	Not performed
111	-	-	Granulocytes >25 No epithelial cells	Nothing detected
112	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
113	-	-	Granulocytes >25 Epithelial cells <10	Not performed
114	-	-	Granulocytes >25 Epithelial cells <10	Not performed
115	*E. coli*	plenty	Granulocytes >25 Epithelial cells <10	*Str. pneumoniae*
116	-	-	Granulocytes >25 No epithelial cells	Nothing detected
117	-	-	Granulocytes 10-25 Epithelial cells <10	Nothing detected
118	-	-	Granulocytes <10 No epithelial cells	Nothing detected
119	-	-	Granulocytes 10-25 Epithelial cells <10	Nothing detected
120	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
123	-	-	Granulocytes >25 No epithelial cells	Nothing detected
124	-	-	Granulocytes >25 No epithelial cells	Nothing detected
125	-	-	No information	Nothing detected
126	*St. aureus;* *Ser. marcescens*	mass; plenty	Granulocytes >25 Epithelial cells <10	Nothing detected
127	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected

128	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
130	-	-	Granulocytes >25 No epithelial cells	Nothing detected
131	-	-	No Granulocytes Epithelial cells <10	Not performed
133	*Ste. maltophilia*	mass	Granulocytes 10-25 No epithelial cells	Nothing detected
134	-	-	Granulocytes >25 No epithelial cells	Nothing detected
135	-	-	Granulocytes >25 Epithelial cells <10	Nothing detected
137	-	-	Granulocytes >25 Epithelial cells <10	*H. influenzae*
138	-	-	Granulocytes <10 Epithelial cells <10	Nothing detected
140	*Cit. koseri*	plenty	No Granulocytes No epithelial cells	Not performed
141	-	-	Granulocytes >25 Epithelial cells <10	*Str. pneumoniae*
142	-	-	Granulocytes 10-25 Epithelial cells <10	Nothing detected
143	*St. aureus*	mass	Granulocytes >25 Epithelial cells <10	Not performed
144	*St. aureus*	few	Granulocytes <10 No epithelial cells	Nothing detected

Table 22: Pathogenicity evaluation for the culture results for every sample and result of the PCR panel.

Appendix 4: Wet lab and 16S rRNA-sequencing with Illumina – further results

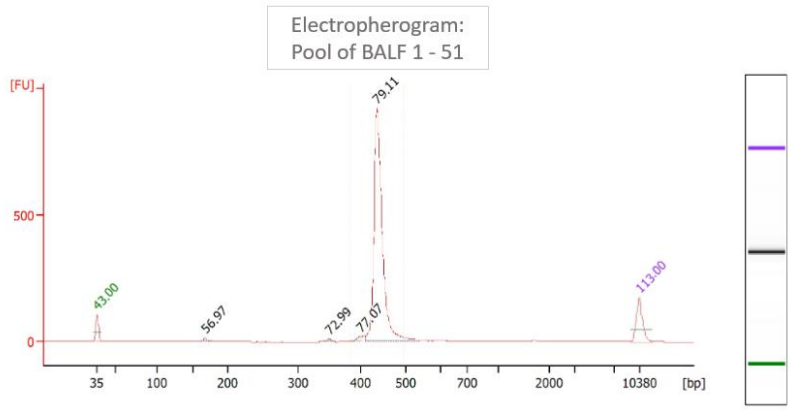
		Concentration of Extracts [nmol/L]		***Concentration of PCR-1 Clean-up [nmol/L]***		***Concentration of PCR-2 Clean-up [nmol/L]***	
Sample ID	**Days until freeze**	**Zymo**	**eMAG**	**Zymo**	**eMAG**	**Zymo**	**eMAG**
2	20	34,17	23,48	19,5	21,7	91,4	9,6
3	1	0,34	0,22	4,8	-0,1	77,6	1,6
4	13	1,55	0,44	21,4	0,9	99,4	15,4
7	4	0,47	0,24	4,2	0,1	65,8	3,1
8	15	0,2	0,24	7,8	0,2	91	1,1
9	33	-0,05	0,19	1,2	0,2	15,5	2,2
11	6	0,62	0,18	4,4	0,2	79,6	3,4
13	4	0,11	0,17	3,8	0,6	56,3	12,3
14	4	-0,14	0,15	2,1	0,2	40,1	5,2
15	33	0,04	0,17	38	15,8	102,8	58,3
17	1	2,48	2,16	56,5	80,6	87	84,9
18	1	4,64	0,48	10,2	1,8	71,3	28,5
19	5	-0,02	0,15	2,5	0,3	34,3	4,4
21	1	-0,08	0,19	0,8	0,2	19,2	5,6
22	2	-0,1	0,14	26,1	9,5	85,8	65,9
23	6	-0,08	0,17	14,1	6,6	92	70,9
24	6	1,45	0,51	34,5	54,2	99,4	56,9
25	4	0,44	0,14	2,2	0,5	32,6	5,9
27	4	20,02	5,39	43,3	67,8	107,5	92
28	4	0,32	0,16	21,1	10,3	91,5	70,8
29	20	0,64	0,25	2,4	0,7	44,2	14,6
33	1	-0,12	0,12	1,9	0,3	51	6,7
34	7	19,54	6,9	20,7	15,7	69,2	40,6
36	2	0,05	0,16	2,1	0,5	20,4	8,1
38	3	-0,08	0,12	3	0,7	55,5	10,6
39	2	-0,16	0,12	1,5	0,3	50,2	3,3
40	0	-0,18	0,12	4	3	56,5	47,2
41	0	-0,12	0,12	32,6	27,2	102,4	83,3
42	0	-0,1	0,14	2,5	0,3	41,4	7,4
43	1	3,13	1,34	4,5	1,1	78,3	7,7
44	0	20,15	6,15	4,5	3,8	35,8	4,5
45	0	0,04	0,12	2,5	1,3	33,9	21,1
48	11	2,86	*7,9*	0,1	*3,49*	59	*51,19*
49	15	0,48	*2,9*	0,7	*2,31*	31,2	*38,74*
50	1	0,08	*0*	0,1	*0,33*	40,1	*5,64*
51	0	-0,01	*0,2*	21,7	*2,07*	94,3	*44,62*
52	0	26,75	11,55	21,6	9,1	94,3	18,3

55	2	0,14	0,15	1,1	0,8	7,7	6,9
57	0	0,22	0,2	1,3	0,8	22,1	3,3
58	1	4,13	2,64	8,8	2,8	99,2	32,6
59	0	0,78	0,2	1,8	0,7	32,4	2,6
60	2	2,93	1,54	37	24,8	98,7	94
65	0	0,12	0,12	1,6	0,6	12,7	2,7
70	1	8,91	2,25	24	8	102,2	67,9
71	0	18,83	7,27	7,1	4,1	42,9	2,2
72	1	31,04	6,71	16,9	6,4	55,2	68,9
73	2	25,35	5,18	3,1	2,4	15,8	4,1
74	0	0,15	0,12	2,9	0,8	71,6	6,9
75	1	0,14	0,58	1,1	1,3	4,6	23,3
76	1	1,94	2,6	29,5	27,8	93,7	75,6
77	3	3,86	2,1	10,1	1,7	89	10,5
79	4	0,17	0,12	0,7	1	5	7,7
80	1	0,19	0,22	1,8	1,3	18,4	19,3
82	1	0,68	0,45	36,3	51,6	107,2	99
83	0	0,29	0,34	3,4	4,2	60,3	40,6
85	3	0,09	0,13	1,5	0,8	27,4	2,2
86	2	0,54	0,47	2,1	0,9	33	2,2
88	1	28,67	3,07	8,4	2,6	28	3,8
89	1	0,14	0,19	1	1	9,9	11,4
91	1	0,18	0,17	4,2	18	61,7	74,5
92	1	0,11	0,13	1,3	0,7	26	2
93	1	17,37	5,12	33,8	40,9	94,3	92,1
94	2	0,13	0,15	5,5	6,6	73,9	63
95	0	5,75	3,01	60,5	53	105,9	55,9
96	0	5,02	2,73	9,9	9,6	87,4	62,4
97	0	5,32	*15,9*	15,5	42,6	77,4	68,4
98	1	4,45	*19,5*	3,2	2,1	73	6,1
100	1	0,21	0,29	2,3	4,8	30,1	45,5
101	1	0,16	0,29	2,3	2	31,3	26,8
102	1	0,59	0,65	33,3	51,6	110,3	110,9
103	2	1,07	0,82	3,4	2,9	45,3	23,2
104	1	0,58	1,09	8,7	34,7	103,7	81
105	1	0,09	0,17	1	2	9,8	20,7
106	6	0,21	0,54	2,5	2,8	31,7	23
107	7	0,15	0,24	2,4	2,7	28,7	23,1
108	1	0,15	0,2	2	1,8	15,2	21
109	1	0,13	0,19	2	2,2	24,2	27,7
110	1	0,15	0,27	2,7	6,7	32,3	36,9
111	1	0,54	1,75	2,8	5,3	32,2	33,4
112	0	0,58	0,83	8,9	51,2	41,4	79,4
113	1	0,41	0,61	3,5	1,4	34	4,4
114	1	0,12	0,18	2,1	5,5	13	31,3
115	1	1,28	2,12	14,7	36,3	93,8	74,5

116	0	0,12	0,19	1	1,6	8	13,8
117	4	0,21	0,27	1,2	1,9	21	12,9
118	3	0,1	0,18	0,8	1,6	5,8	10,1
119	4	0,13	0,26	1,2	2,2	5,9	13,8
120	3	3,7	2,01	13,3	7,2	92,2	25
123	0	0,12	0,3	1	1,5	9,9	16
124	0	0,36	0,6	3,7	4,4	42,2	30,3
125	0	0,13	0,26	1,2	4,2	20,9	49,9
126	0	0,26	0,75	26,5	69,2	121,5	59,1
127	2	0,66	3,48	1,7	4,2	16,8	22,6
128	1	15,81	3,97	13	37,9	68,9	63,3
130	6	0,08	0,15	1,1	0,8	7,1	5,2
131	4	0,1	0,15	0,8	0,7	7,6	6
133	6	0,17	0,23	17,6	26,7	75,5	61,6
134	1	0,11	0,21	1,2	1,2	22,7	13,3
135	1	0,21	0,34	1,3	1,3	23,7	6,4
137	3	0,28	0,38	7,4	33,6	59,7	98,3
138	2	1,02	0,53	4,3	3,8	72,5	79,6
140	1	3,5	4,26	3	4,9	30,4	*38,4*
141	1	0,13	0,29	1,3	12	20,7	*91,66*
142	2	0,06	0,1	1,7	1,5	9,8	*7,92*
143	2	6,79	4,46	16,8	28,5	105,9	*35,02*
144	3	0,35	3,87	2,6	32,3	42,8	*122,1*

Table 23: Results of concentration measurement with the PicoGreen assay for dsDNA for each sample at different time points of the workflow.

Values in italic were measured with the Qubit High Sensitivity Kit. In addition, days until freeze are noted.



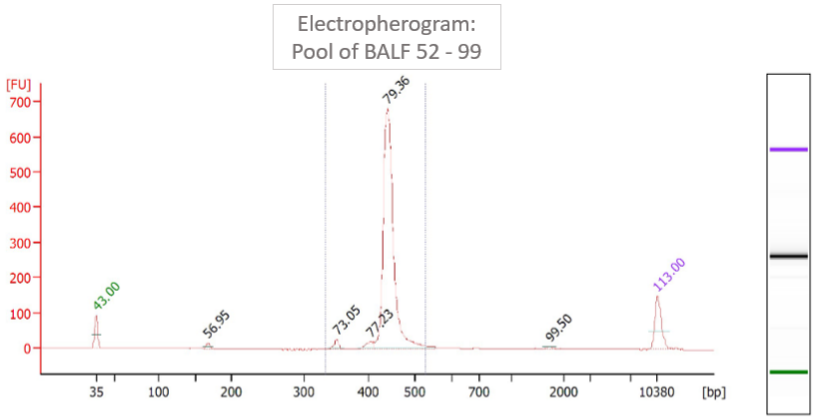
Overall Results for sample 1 : **BALF1-51 gereinigt**

Number of peaks found:	4	Corr. Area 1:	1.519,8
Noise:	0,3		

Peak table for sample 1 : **BALF1-51 gereinigt**

Peak	Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
1	35	125,00	5.411,3	Lower Marker
2	168	14,25	128,6	
3	351	6,15	26,6	
4	401	14,47	54,6	
5	436	1.078,28	3.743,5	
6	10.380	75,00	10,9	Upper Marker

Figure 32: Electropherogram from Agilent Bioanalyzer of the pooled DNA from samples one to 51 for DNA quantitation and quality control.



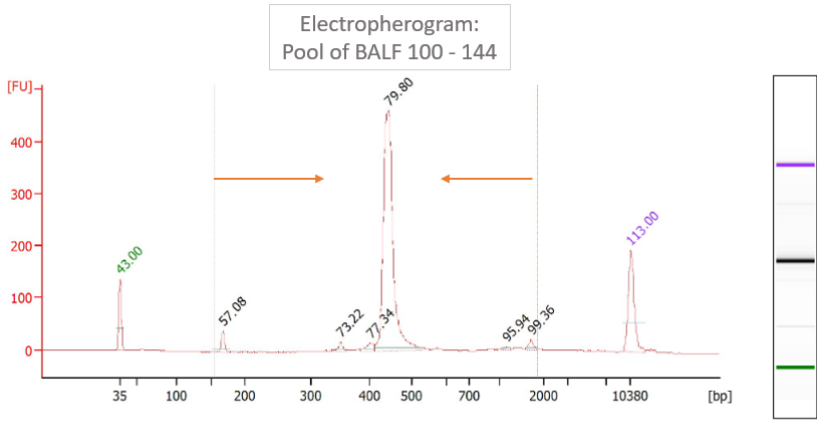
Overall Results for sample 2 : **BALF 52-99 gereinigt**

Number of peaks found:	5	Corr. Area 1:	1.364,2
Noise:	0,5		

Peak table for sample 2 : **BALF 52-99 gereinigt**

Peak	Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
1	35	125,00	5.411,3	Lower Marker
2	168	18,28	165,2	
3	351	23,69	102,2	
4	404	25,79	96,7	
5	441	1.095,49	3.765,9	
6	1.716	3,75	3,3	
7	10.380	75,00	10,9	Upper Marker

Figure 33: Electropherogram from Agilent Bioanalyzer of the pooled DNA from samples 52 to 99 for DNA quantitation and quality control.



Overall Results for sample 1 : **MB BAL 120-144 Pool**

Number of peaks found:	6	Corr. Area 1:	1.188,1
Noise:	0,3		

Peak table for sample 1 : **MB BAL 120-144 Pool**

Peak	Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
1	35	125,00	5.411,3	Lower Marker
2	168	31,78	286,4	
3	351	11,75	50,7	
4	403	15,14	56,9	
5	446	576,61	1.958,1	
6	1.160	3,51	4,6	
7	1.711	8,10	7,2	
8	10.380	75,00	10,9	Upper Marker

Figure 34: Electropherogram from Agilent Bioanalyzer of the pooled DNA from samples 99 to 144 for DNA quantitation and quality control.

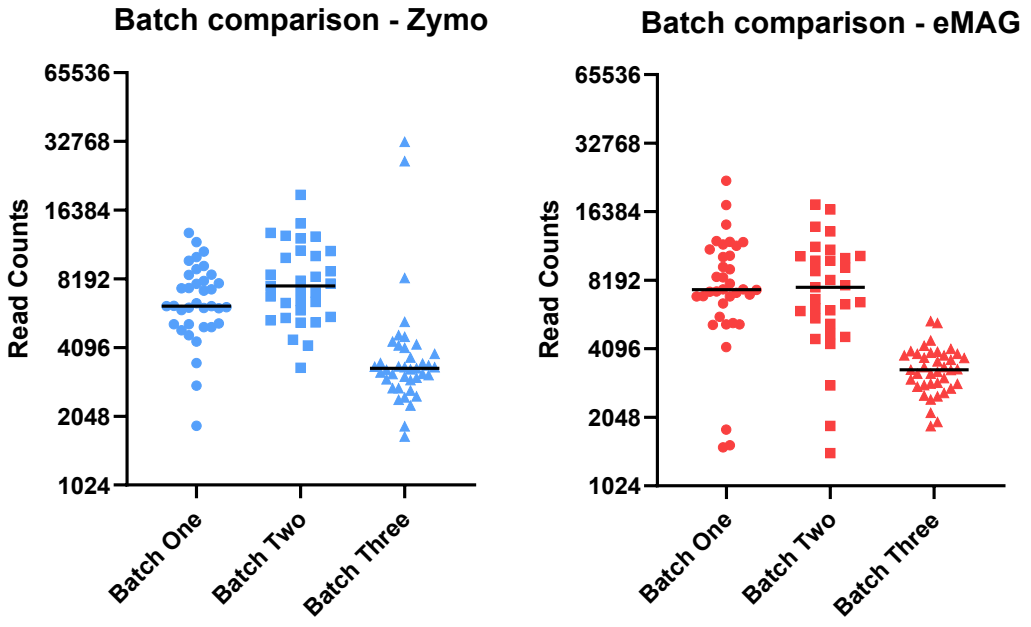


Figure 35: Scatter plots of the read counts for Zymo and for eMAG, separated by batch, with their medians.

Appendix 5: Detailed analysis of the controls

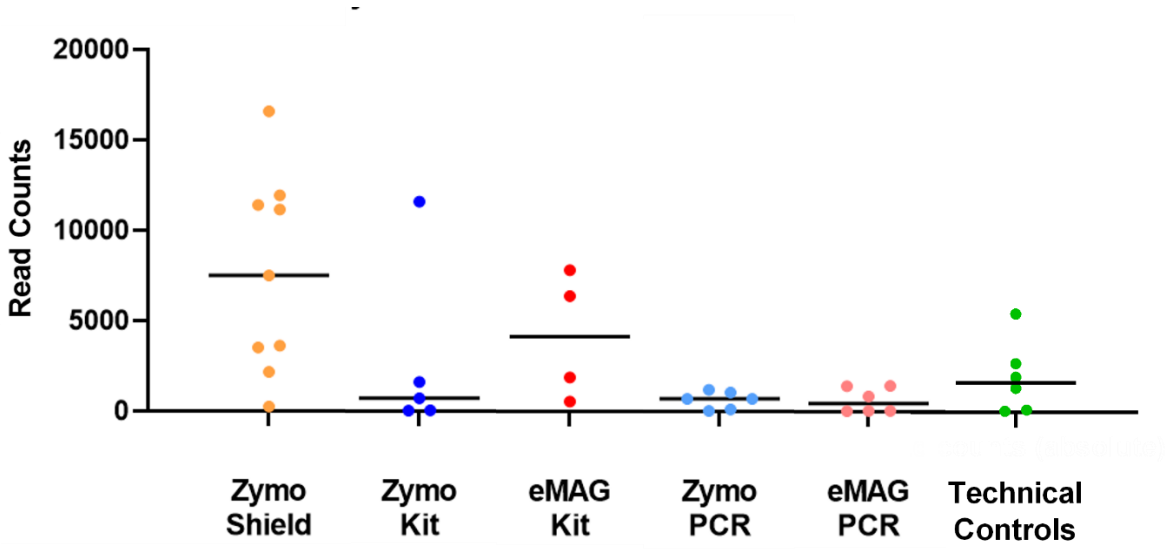


Figure 36: Number of read counts of each control, arranged and color-coded by control type, as well as the median for every control type.

In the Zymo Shield controls, the number of read counts varied strongly. The lowest number of read counts had control C33 (263), and by far the highest number, not just of the Zymo Shield controls but of all controls, had control C22 with 16,584 read counts. The read counts from the third batch were overall lower than from the first and second batch. Four genera were most abundant in the Zymo Shield controls: *Kocuria*, *Burkholderia-Caballeronia-Paraburkholderia*, *Aquabacterium* and *Psychrobacter*. The controls showed homogeneity in the genera present, albeit partial differences in the relative abundances of those genera. In the first batch, *Aquabacterium* and *Psychrobacter* were represented stronger, while the last batch was more dominated by *Burkholderia-Caballeronia-Paraburkholderia* and, to some extent, *Pseudomonas*. Control C22 from batch two was protruding with the highest number of absolute read counts (16,584) and *Kocuria* as the major genus with a relative abundance of 88.8%. 97.4% of all *Kocuria* read counts in all controls (not just in the Zymo Shield controls) were in control C22. Only five other controls had any read counts which were assigned to the *Kocuria* genus, none of which were higher than 255 read counts and four of them also being Zymo Shield controls. When excluding *Kocuria* from the microbiome profiles, C22 resembled the other two kit controls from its batch. Those findings could indicate an individual contamination in the process of DNA extraction or library preparation.

The five Zymo Kit controls didn`t have many read counts overall, with a median of 733 read counts. One outlier, however, was control ZC4 from batch one with 11,607 read counts. Thus, a large portion of the dominant taxa shown in Figure 8 could be attributed to this control. *Aquabacterium*, *Escherichia-Shigella* and *Cutibacterium* were, in descending order, the most frequently sequenced taxa in control ZC4. *Escherichia-Shigella* and *Cutibacterium* were most abundant in the controls ZC9 and C25. The other two controls possessed neglectable low numbers of read counts (47 and 56).

The four eMAG Kit controls featured a variable number of read counts with the two highest numbers of read counts in the first and the lower numbers in the two last batches. The microbiome profiles showed a heterogenous picture with *Enhydrobacter* as the dominant genus in control EKit1 (relative abundance of 67.7 %) and *Rhodobacteraceae unclassified* as the dominant genus in control EKit2 (86.7 %), both from batch one. 6,736 of the total 7,059 read counts in all controls (95,4 %) of *Rhodobacteraceae unclassified* were attributable to control EKit2. 4,159 of the total 4,470 read counts in all controls (93,0 %) of *Enhydrobacter* were attributable to control EKit1. Those observation underlined the occurrence of these taxa specifically in the eMAG Kit controls. Figure 8 also emphasized those two taxa as the dominant ones among the eMAG Kit controls. The other two eMAG Kit controls, which featured lower numbers of read counts, presented more heterogenous microbiome profiles, with no single taxon dominating.

The PCR controls of both the Zymo and the eMAG samples showed overall low numbers of read counts. They also resembled each other in taxonomic composition, both being dominated by the two genera of *Cutibacterium* and *Escherichia-Shigella*. The Zymo PCR controls had higher abundances of *Escherichia-Shigella* (40.9 % of all read counts in all Zymo PCR controls), followed by *Cutibacterium* (30,4 %). In the eMAG PCR controls, it was reverse, with *Cutibacterium* in the first place (60.0 % of all read counts in all eMAG PCR controls) and *Escherichia-Shigella* following (16.5 %). Across the batches, the number of read counts exhibited discrete differences, again batch three having slightly lower numbers than batch one and two. The three batches of both the Zymo and the eMAG PCR controls were homogeneous in terms of the genera present as well as the relative abundances.

The bronchoscopy controls consisted of three blank bronchoscope flushes with 0.9 % sodium chloride fluid (MB4-6) and three controls of solely the 0.9 % sodium chloride

fluid used for the routine BALF procedure (MB1-3), collected on three different days. The median number of read counts of the six bronchoscopy controls was 1,630 (IQR, 102.5 – 3,382), with the lowest number of read counts being 59 (control MB2) and 5,447 as the highest number of read counts (control MB1). *Burkholderia-Caballeronia-Paraburkholderia* was the most prevalent genus, accounting for 44.6 % of the total read counts in all 6 technical controls. *Pseudomonas* was the second most prevalent genus with 10.0 %. Overall, the taxonomic composition of the six controls was very similar. Fewer read counts were obtained from the three bronchoscopy flushes (4,133) than from the three sodium chloride controls (7,443).

As can be seen in Figure 37, the observed richness, Chao 1 and the Shannon index as parameters for α-diversity, were highest in the Zymo Shield controls. That could be attributed in part to the higher total number of read counts. Control C22 was grouped along the other Zymo Shield Controls in the higher diversity ranges when consulting the observed richness and Chao1. However, the Shannon index for C22 was one of the lowest (0.67) as it also assessed evenness and the microbiome profile of C22 was highly dominated by *Kocuria*, and therefore very uneven.

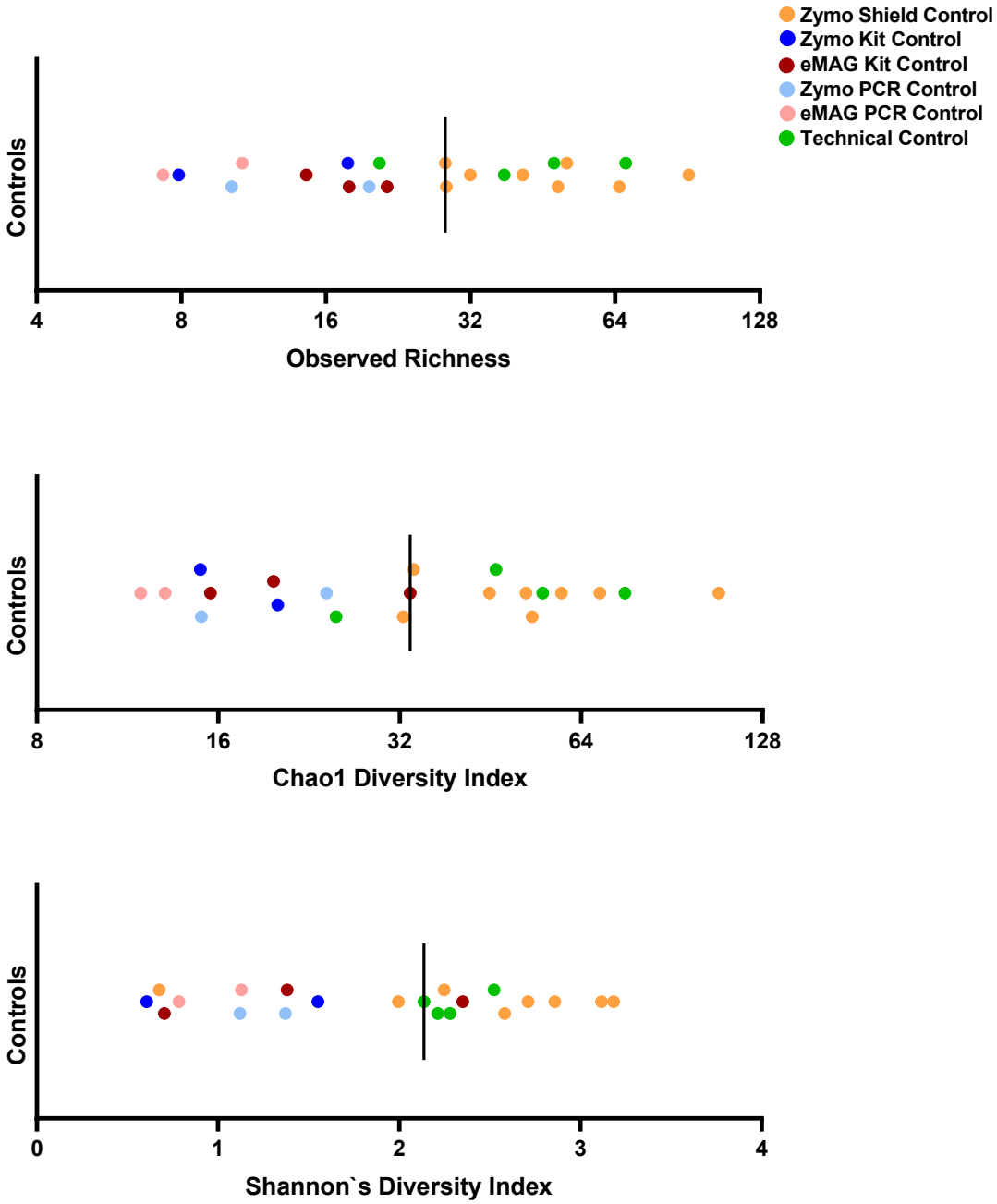


Figure 37: Three α-diversity metrics (observed richness, Chao1 and Shannon) of all 17 controls with over 1,000 read counts with color marking for control type.

Appendix 6: Comparison of manual Zymo and automatic eMAG extraction – further results

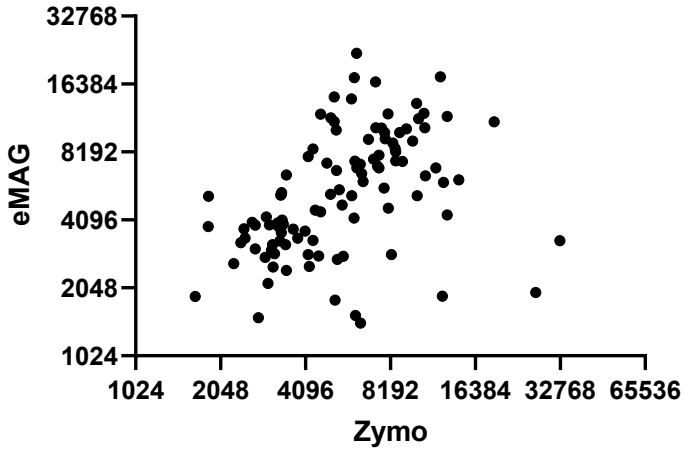


Figure 38: Scatter plot showing the number of read counts of the Zymo and the eMAG samples.

Appendix 7: Categorization of the microbiome profiles - further results

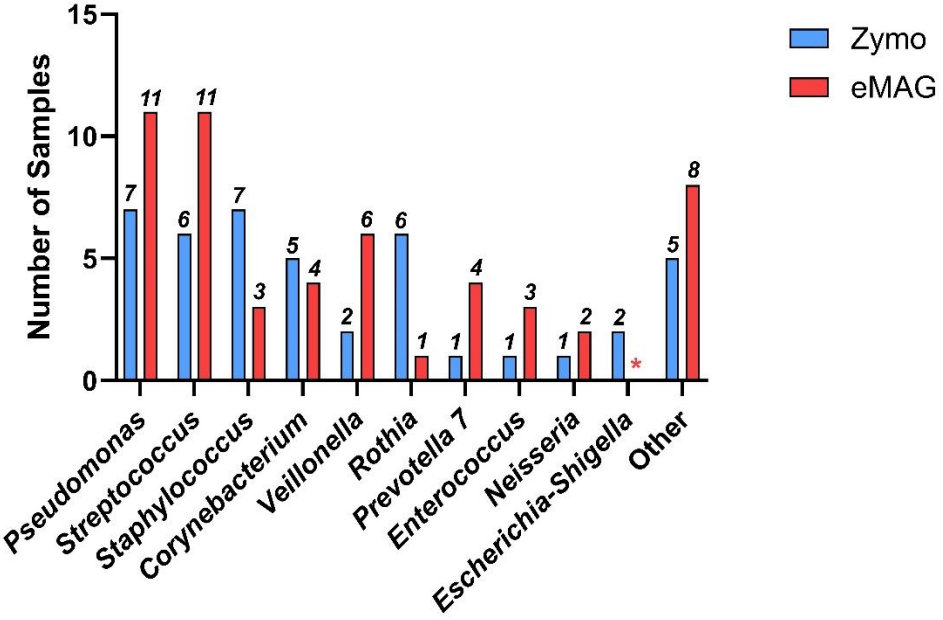


Figure 39: Top genera found in the poly-microbial microbiome profiles.

Genera with a star (*) were classified as contamination and excluded from analysis.

Appendix 8: Overlap between the PCR panel and Illumina short-read sequencing

All the nine PCR panel reports could be retraced by Illumina short-read sequencing. The one sample in which *Haemophilus* was most abundant in both Zymo and eMAG was also the sample with the one positive culture result for *H. influenzae*.

PCR Panel		Corresponding Genera Detected by Illumina Short Read NGS							
Species	Number of Reports	Most Abundant		Top Five		Any Abundance		Absent	
		Zymo	eMAG	Zymo	eMAG	Zymo	eMAG	Zymo	eMAG
H. influenzae	5	1	1	3	3	5	5	0	0
Str. pneumoniae	4	1	1	4	4	4	4	0	0
Sum	***9***	***2***	***2***	***7***	***7***	***9***	***9***	***0***	***0***

Table 24: Alignment of the species reported in the PCR panel with their correspondent genus from Illumina short-read sequencing.

Analysis scheme modified according to Zacharia et al. (2018).

Appendix 9: Overlap between the culture-based routine diagnostics and Illumina short-read sequencing – further results

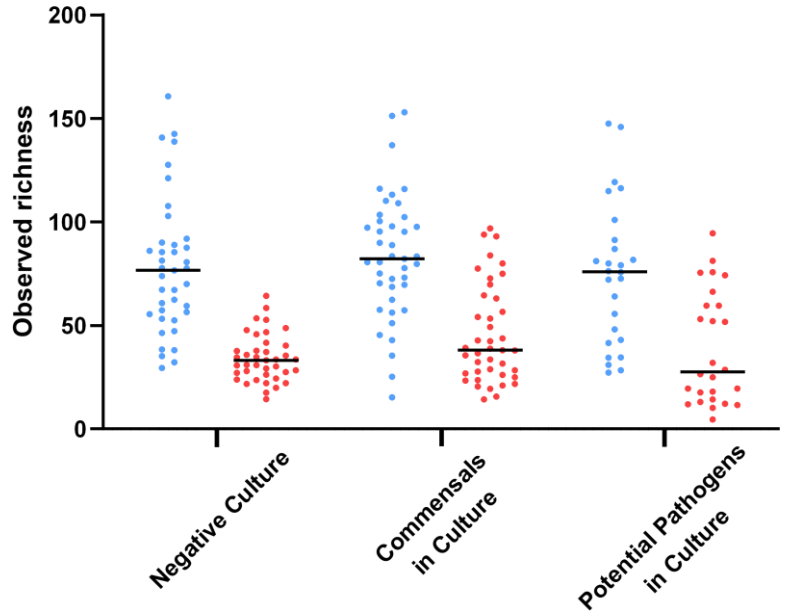


Figure 40: Observed richness representing α-diversity, sorted by culture result with the medians.

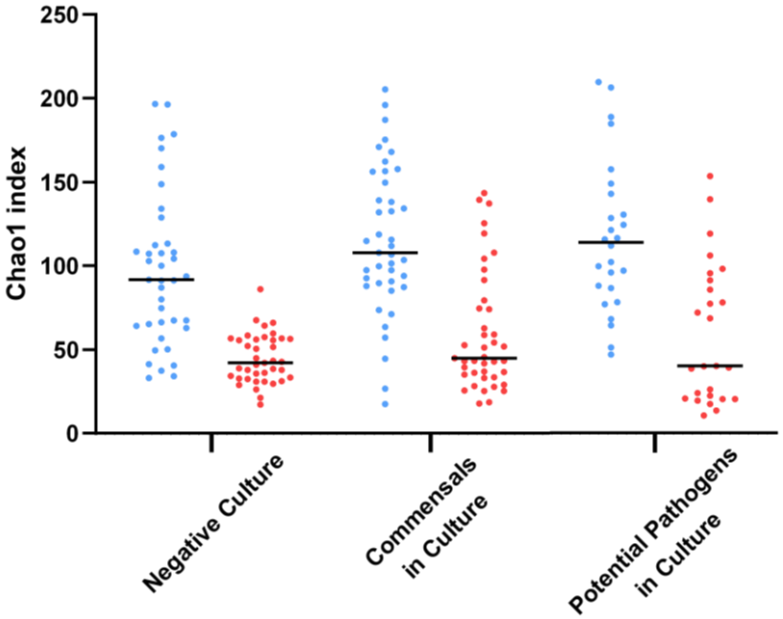


Figure 41: Chao1 index representing α-diversity, sorted by culture result with the medians.

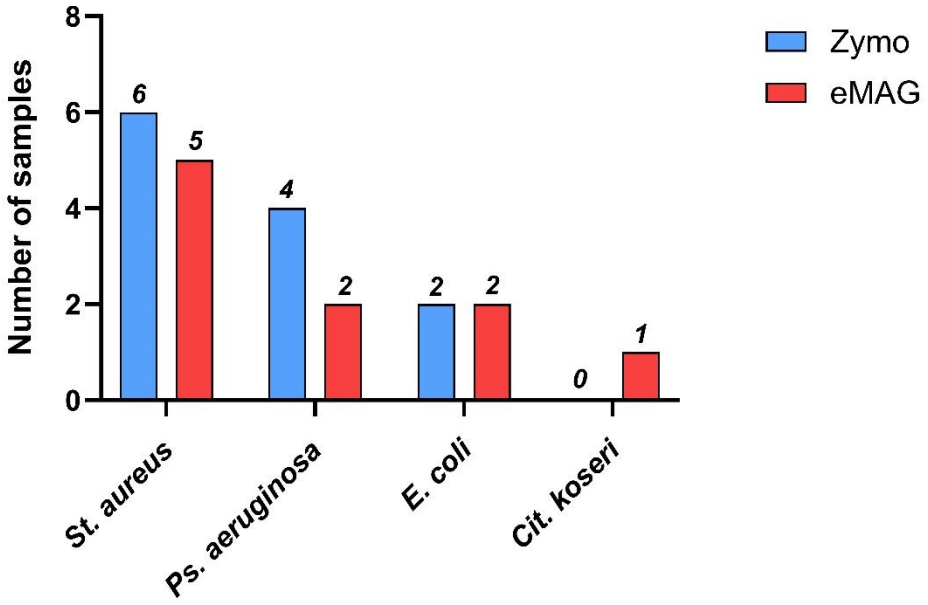


Figure 42: Potentially pathogenic species recorded in culture for samples with a multi-microbial profile in Illumina short-read sequencing.

Appendix 10: Species identification with Nanopore long-read sequencing – further results

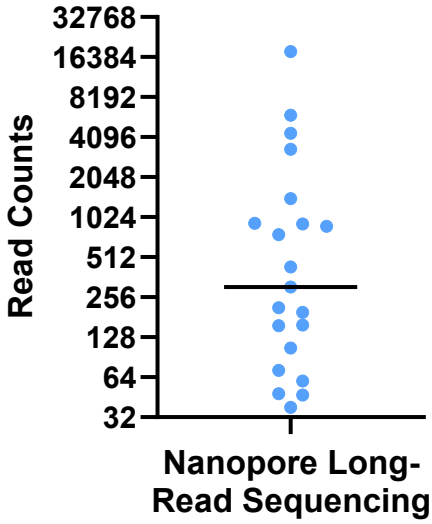


Figure 43: Sequencing reads from Nanopore long-read sequencing.

The values were plotted on a logarithmic scale (base two) and the D`Agostino and Pearson test indicated a lognormal distribution.

Sample Number	Number of Read Counts	Most Abundant Species (No. 1)	Read Counts of Species No. 1	Second Abundant Species (No. 2)	Read Counts of Species No. 2
24	18022	*H. influenzae*	13815	*H. parahaemolyticus*	337
15	5981	*Str. salivarius*	1198	*Vei. atypica*	764
2	4369	*Cit. koseri*	3041	*St. epidermidis*	185
27	3321	*St. aureus*	2465	*Pre. oris*	321
51	1407	*Str. parasanguinis*	440	*Vei. dispar*	145
22	919	*Str. parasanguinis*	166	*Cam. concisus*	88
30	912	*H. parahaemolyticus*	115	*Str. salivarius*	101
17	872	*Str. pneumoniae*	760	*Str. pseudopneumoniae*	20
23	754	*Str. oralis*	265	*Gem. haeolysans*	188
11	433	*Str. pseudopneumoniae*	164	*Str. pneumoniae*	60
49	306	*Ent. faecium*	181	*St. epidermidis*	15
45	214	*Ral. pickettii*	21	*Str. australis*	18
50	197	*Ral. pickettii*	19	*Aq. commune*	17
46	158	*Str. salivarius*	17	*Str. vestibularis*	6
43	156	*Ral. pickettii*	13	*Aq. commune*	10
48	106	*Ent. faecium*	59	*Lac. paracasei*	3
C24	75	*Paraburkholderia bryophila*	17	*Par. fungorum*	10
19	72	*Ral. pickettii*	5	*N. perflava*	4
13	60	*Ral. pickettii*	11	*Par. fungorum*	8
34	48	*H. influenzae*	4	*Str. mitis*	4
53	47	*Her. huttiense*	6	*Ral. pickettii*	4
18	38	*Vei. dispar*	9	*Ent. faecalis*	4
C4	2	*Aq. commune*	1	*H. influenzae*	1
PCR-C	1	*Cit. koseri*	1	None	-

Table 25: Results of Nanopore long-read sequencing for all samples including the two most abundant species with their number of read counts.

Publikationsverzeichnis

Der damalige Stand der Forschungsarbeit wurde im Rahmen des Science Days 2021 (Fachbereiches Medizin, Justus-Liebig-Universität Gießen) am 12.11.2021 unter dem Namen „Mikrobiom-basierter klinischer Erregernachweis durch NGS bei Infektionen der unteren Atemwege“ als Poster-Vortrag vorgestellt. Der Vortrag gewann die Auszeichnungen „Beste Methodik“ und „Bestes Poster“ im Bereich der klinischen/ patientennahen Forschung. Co-Autorinnen und -Autoren waren: Markus Weigel, Benjamin Ott, Jan Philipp Mengel, Alexandra Amend, Johanna Fitzner, Judith Schmiedel, Heiko Slanina, Can Imirzalioglu und Torsten Hain.

Ehrenwörtliche Erklärung

“Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, und dass die vorgelegte Arbeit weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.”

_____ _____

Ort/ Datum *Unterschrift*

Danksagung

Der größte Dank gilt meinem Doktorvater PD Dr. Torsten Hain, der bei Problemen oder Fragen immer ein offenes Ohr hatte und mir mit seinem Vertrauen und seiner Anleitung die Chance gegeben hat, unglaublich viel zu lernen. Ich hätte mir keine bessere Betreuung vorstellen können! Ich danke meinem Betreuer PD Dr. Can Imirzalioglu, der mir durch seine langjährige klinische Expertise eine hervorragende Unterstützung war und wertvolles Feedback gegeben hat. Großer Dank und Wertschätzung geht an meine Arbeitsgruppe um Alexandra Amend, Benjamin Ott, Jan-Philipp Mengel und Markus Weigel, die mir auf unschätzbare Weise im Labor und bei bioinformatischen Fragen zur Seite gestanden haben, und ohne die ich mehr als einmal verzweifelt wäre.

Ich danke dem Team der Medizinischen Mikrobiologie des Universitätsklinikums Gießen, im Besonderen Dr. Judith Schmiedel, Johanna Fitzner und Katharina Matis, die bei praktischen Fragen und Anliegen immer hilfsbereit und unterstützend waren.

Zudem geht Dank an Dr. Anita Windhorst, die mich bei meiner statistischen Auswertung großartig beraten hat, sowie an Prof. Dr. Khodr Tello, der bei Anliegen rund um die Bronchoskopie eine große Hilfe war.

Mit großer Freude habe ich am JLU TRAINEE-Programm teilgenommen, welches ideell sowie finanziell eine wichtige Unterstützung für mich war.

Meine WG-MitbewohnerInnen Konstantin, Leo, Rebecca und Sophia haben mir das Home-Office zum Zuhause gemacht. Konstantin bin ich außerdem sehr dankbar für seinen wertvollen fachlichen Input.

Und schließlich möchte ich meiner Familie danken. Meinen Eltern, Claudia und Andreas, für die vielseitige, liebevolle Unterstützung und die zahlreichen Chancen, die sie mir ermöglicht haben. Danke an Lea, mit der ich Freude und Frust teilen konnte und auf deren Rat und Tat ich mich immer verlassen konnte. Und an Niklas und David für eure Unterstützung.