Aus dem Bereich Innere Medizin V
Theoretische Medizin und Biowissenschaften bzw. Klinische Medizin
der Medizinischen Fakultät
der Universität des Saarlandes, Homburg/Saar

THE AIRWAY MICROBIOTA IN HEALTHY STATE AND CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Dissertation zur Erlangung des Grades eines Doktors der Naturwissenschaften

der Medizinischen Fakultät

der UNIVERSITÄT DES SAARLANDES

2011

vorgelegt von: **Tetyana Zakharkina**
geb. am: 15.02.1984 in Novosibirsk, Russland

To my parents

Angenommen vom Fachbereich Medizin
der Universität des Saarlandes als Dissertation
am

Erstgutachter: Prof. Dr. Dr. Robert Bals
Zweitgutachter:

Prüfungskommission:

Tag der mündlichen Prüfung am

This thesis, containing 26 figures and 13 tables, was typeset with LaTeX using Bitstream Charter and `Latin Modern Typewriter` type faces. The graphical work was prepared using INKSCAPE (vector graphics; publishing of final figures to PDF), PRISM 5 (GraphPad® software, basic biostatistics, curve fitting and scientific graphing), and WEKA 3 (learning algorithms for data mining tasks).

Abstract

Chronic obstructive pulmonary disease (COPD) is the co-occurrence of chronic bronchitis and emphysema. This leads to a limitation of the flow of air to and from the lungs, causing shortness of breath. The role of bacterial infection in COPD remains unclear.

Since there is a growing body of evidence about the presence of core microbial communities in the lower respiratory tract, the main purpose of this work was to investigate their structure more detailed in healthy individuals and their changes linked with COPD. Bronchoalveolar lavages obtained from 9 healthy individuals and 9 stable COPD patients were analyzed using molecular biology-based approaches such as terminal restriction fragment length polymorphism (T-RFLP) and sequencing. Additionally, the *in situ* detection of bacteria was performed on the tracheal tissue sections.

Sequencing data, supported by T-RFLP analysis, showed the presence of core microbiome in the lower respiratory tract. The key genera were *Prevotella*, *Sphingomonas*, *Pseudomonas*, *Acinetobacter*, *Fusobacterium*, *Megasphaera*, *Veillonella*, *Staphylococcus*, and *Streptococcus*. Despite this fact the observed variability in the microbial composition was so high that no pattern could be established both for healthy individual and COPD patient cohorts. On the other hand, the presence of the genera *Mycoplasma* (6 positive samples), *Pseudomonas* (5 positive samples), and *Staphylococcus* (5 positive samples) was highly correlated with disease. Fluorescence *in situ* hybridization with universal eubacterial oligoprobe EUB 338 demonstrated the subepithelial manner of microbial distribution.

In conclusion, non-cultural methods can be used to identify the microbial biota of the airways.

Zusammenfassung

Die chronisch obstruktive Lungenerkrankung (englisch: chronic obstructive pulmonary disease, COPD) ist definiert durch das gleichzeitige Auftreten einer chronischer Bronchitis und eines Lungenemphysems. Dies führt zu einer Einschränkung des Luftflusses sowohl bei der Ein- als auch bei der Ausatmung, was Kurzatmigkeit verursachen kann. Die Rolle von bakteriellen Infektionen bei COPD ist bislang nicht vollständig geklärt.

Da immer mehr Hinweise auf die Anwesenheit von beständiger mikrobieller Besiedelung in der unteren Atemwegen hindeuten, war es der Hauptziel dieser Arbeit, diese Besiedlungen näher zu untersuchen, sowohl im gesunden Zustand als auch bei Patienten mit COPD. Bronchoalveoläre Lavagen von 9 gesunden Personen und 9 stabilen COPD-Patienten wurden mit molekularbiologischen Methoden wie der Sequenzierung und der Analyse terminaler Restriktionsfragmentlängenpolymorphismen (T-RFLPs) untersucht. Zusätzlich wurden in situ Bakterien in Gewebeschnitte der Trachea nachgewiesen.

Sequenzierungsdaten und T-RFLP-Analysen liefern Hinweise auf die Präsenz eines beständigen Mikrobioms in die unteren Atemwegen. Die Hauptgattungen sind *Prevotella*, *Sphingomonas*, *Pseudomonas*, *Acinetobacter*, *Fusobacterium*, *Megasphaera*, *Veillonella*, *Staphylococcus* und *Streptococcus*. Dennoch ist die beobachtete Variabilität der mikrobiellen Zusammensetzung so hoch, dass ein typisches Muster weder für gesunde Individuen noch für COPD-Patienten festgestellt werden konnte. Andererseits korreliert die Anwesenheit von Gattungen wie *Mycoplasma* (6 positive Proben), *Pseudomonas* (5 positive Proben) und *Staphylococcus* (5 positive Proben) stark mit der Krankheit. Die Fluoreszenz-*in-situ*-Hybridisierung mit EUB 338 (englisch: universal eubacterial oligoprobe) zeigt eine klare mikrobielle Verteilung.

Zusammenfassend, konnte in dieser Arberit gezeigt werden, dass nichtkulturelle Verfahren verwendet werden können um das Mikrobiom der Atemwege zu identifizieren.

Contents

1 Introduction

The airways, which are also referred to as the "conducting portion" of the respiratory system, delivers air to the lungs (known as the "respiratory portion"), where gas exchange takes place. The conducting portion consists of the nose, pharynx, larynx, trachea, and bronchi, whereas the respiratory portion consists of the bronchioles, alveolar ducts, alveolar sacs, and alveoli. It is also convenient to divide the tract into the upper respiratory tract (nose and pharynx) and the lower respiratory tract (larynx, trachea, bronchi and lungs). The conducting portion, apart from the bronchi, is colonized by microbes [180].

Chronic obstructive pulmonary disease (COPD) is a respiratory disease associated with chronic inflammation of the lung leading to tissue destruction and emphysema. Frequent bacterial infections of the respiratory tract of COPD patients and chronic bronchitis are known to contribute to the progress and exacerbations of COPD.

Bacterial colonization of the respiratory tract plays a crucial role in the disease progression, as chronic colonization with harmful bacteria contributes to an ongoing inflammation of the lung. Respiratory pathogens are a part of the pulmonary microbial community which have a complex structure and composition. Since harmful microbes compete with a huge variety of standard microbiota representatives in the upper and lower respiratory tract, it is of great interest to monitor the changes in species composition between the healthy state and disease [188, 166, 161, 142, 139]. Moreover, constant exposure of the respiratory tract with the non-sterile air leads to microbe-microbe and microbe-host interactions. However, human airway microbiota, particularly that present in the context of pulmonary disease, remains largely uncharacterized [79], as routinely used classical culture methods do not deliver any information about non-cultivable species. In contrast, molecular biology based approaches that do not depend on cultivating the microbes have high resolution and would allow deep analysis of respiratory samples.

1.1 Chronic obstructive pulmonary disease

COPD has been defined by the Global Initiative for Chronic Obstructive Lung Disease (GOLD): COPD is a preventable and treatable disease with some significant extrapulmonary effects that may contribute to the severity in individual patients. Its pulmonary component is characterized by airflow limitation that is not fully reversible. The airflow limitation is usually progressive and associated with an abnormal inflammatory response of the lung to noxious particles and gases [1]. COPD is not one single disease but an umbrella term used to describe chronic lung diseases that cause limitations in lung airflow. The most important risk factors for COPD are tobacco smoking,

indoor air pollution (e.g. biomass fuel used for cooking and heating), outdoor air pollution, occupational dusts and chemicals (vapors, irritants, and fumes). COPD is not just simply a "smoker's cough", but an under-diagnosed and life threatening lung disease. According to the World Health Organization (WHO), it is estimated that in 2007 about two hundred ten million people had COPD and in 2005 three million people died of COPD. The WHO predicts that COPD will become the third leading cause of death worldwide by 2030 (http://www.who.int/respiratory/copd/en/).

Pathological changes characteristic of COPD are found in the proximal airways, peripheral airways, lung parenchyma and pulmonary vasculature. They include chronic inflammation with increased numbers of neutrophils, macrophages, and cluster of differentitaion 8 positive ($CD8^+$) T lymphocytes [115] in different parts of the lung, and structural changes resulting from repeated injury and repair. The key changes include mucus gland hyperplasia and epithelial damage, whereas subepithelial collagen deposition, angiogenesis, increased smooth muscle and increased proteoglycan deposition are less significant [24]. The inflammation of the respiratory tract of COPD patients appears to be an amplification of the normal inflammatory responses of the respiratory tract to inflammatory stimuli, such as cigarette smoke and bacterial stressors. Some patients develop COPD without smoking. This phenomenon has been linked with α_1-antitrypsin deficiency [107]. α_1-antitrypsin is an acute-phase protein, which is the prototypic member of the serpin super family and a major inhibitor of serine proteases such as neutrophil elastase and proteinase-3. It plays an important role in limiting host tissue injury by proteases at sites of inflammation [172]. There is also a relative lack of information about the inflammatory mechanisms involved in exacerbations of COPD [1]. In mild and moderate exacerbation there is an increase of neutrophil and in some studies also eosinophil number in sputum and in the airway wall [177]. This is associated with increased concentrations of certain mediators, such as TNF-α and interleukin 8 (IL-8) and an increase of biomarkers of oxidative stress.

The pathological characteristics of the respiratory tract of COPD patients and smokers, such as chronic inflammation, result in an impaired lung defense against pathogens which increases susceptibility to pulmonary infections. Two distinct infection cycles in COPD patients are described, that contribute to the progressive loss of lung function:

1. During the acute cycle, the lung becomes more susceptible to repeated acute mucosal infections of the airways from viruses and bacteria, leading to episodes of increased inflammation and worsened symptoms, which are clinically diagnosed as acute exacerbations of (AECOPD) [148].

2. A less well-recognized infection cycle in COPD is the chronic cycle. In this case microbial colonization results in chronic inflammation and lung destruction, con-

ceptualized as the "vicious circle hypothesis" (see Figure 1). This hypothesis posits that once impaired innate lung defense allows microbial pathogens to become established in the lower respiratory tract, the microbial pathogens further impair mucociliary clearance and lung defense due to increased mucus secretion, disrupting ciliary activity and airway epithelial injury [147, 162].

Thus, microbial colonization of the lower airways in patients with COPD can perpetuate itself. Furthermore, the chronic presence of bacteria in the lower airways is not innocuous; rather, it induces inflammation and can contribute to progressive airflow obstruction and lung damage characteristic of this disease [175, 101].

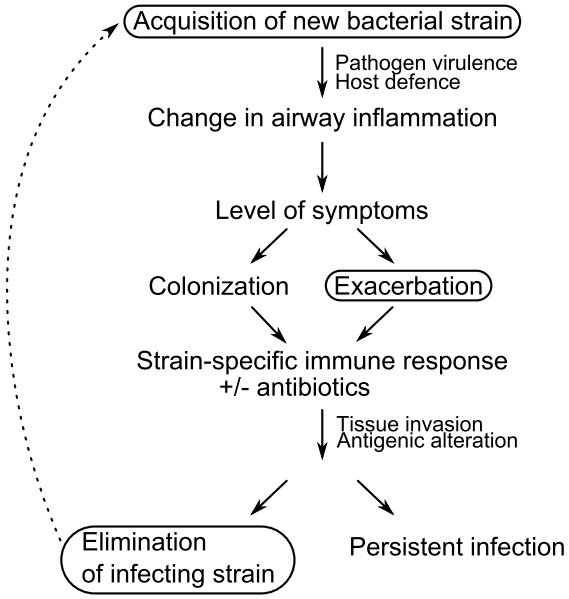

Figure 1: Model of bacterial pathogenesis in COPD (proposed by SETHI) [148]. Once impaired the host immune defence becomes incapable of preventing the colonization by airway pathogens. This leads to exacerbation — worsening of common state and to intensification of symptoms. COPD exacerbation can be defined as "an event in the natural course of the disease characterized by a change in the patient's baseline dyspnea, cough, and/or sputum, that is beyond normal day-to-day variations" [1]. Mucosal damages caused by inflammatory reaction on the pathogen promote the switching of infection to chronicity. In the case of successful elimination of pathogen raises the risk of acquisition of another bacterial strain [159]. RAKHIMOVA et al. in her molecular epidemiological study showed that COPD patients are susceptible to frequent turnover or loss of clones of *P. aeruginosa*, whereas cystic fibrosis patients are chronic carriers of the same strains [126].

HURST and colleagues investigated the relationship of airway inflammation and systemic inflammation for the presence of bacterial pathogens in exacerbated COPD patients [80]. On the cellular level *M. catarrhalis* as well as *H. influenzae* are able not

only to efficiently adhere to but also to invade the airway epithelium [158]. The ability to invade epithelial cells has been discussed as a useful bacterial strategy to colonize the respiratory tract and to avoid extracellular immune recognition by PRRs such as the transmembranous toll-like receptors (TLRs) [5]. SLEVOGT et al. could show that *M. catarrhalis* enters cultured respiratory cells by a trigger-like uptake mechanism. Furthermore, cell surfaces as well as intracellular located PRRs are involved in the pro-inflammatory immune response induced by this bacterium [158]. Their data demonstrated an involvement of both TLR 2 and nucleotide-binding oligomerization domain-containing protein 1 (NOD 1) in the IL-8 response to *M. catarrhalis* [58]. NOD 1 silencing reduced the IL-8 production to approximately 30–35 %, which resembled the deficit in nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) activation in NOD 1 knockout cells.

One of the key points in the COPD management program is prevention and treatment of exacerbations [64]. AECOPD play a crucial role in the course of the disease as they negatively impact on morbidity, mortality, health care costs and health related quality of life [12]. AECOPD are associated with pulmonary and systemic inflammation including the increases of serum C-reactive protein and blood leukocytes level [26, 85]. About 50 % of AECOPD have a bacterial etiology, which is confirmed by several studies [44, 135, 83]. Bacterial microbes most commonly associated with AECOPD are *H. influenzae*, *M. catarrhalis*, *S. pneumonia*, *C. pneumoniae*, *M. pneumoniae*, and *Legionella* species [44, 83, 108]. Also viruses such as rhinovirus, respiratory syncytial virus, influenza virus A, B, and parainfluenza virus 1–3 are often detected in AECOPD.

The identification of the microorganism associated with an exacerbation is important to further understand the relationship between microbes and exacerbations and to develop antimicrobial therapeutic strategies. Current methods to identify associations between microbes and AECOPD are limited. Routine cultural microbiological methods do not allow the identification of viruses and intracellular pathogens. In addition, results can be obtained only after several days. Methods based on nucleic acid testing (NAT) such as the polymerase chain reaction (PCR) have been developed recently [41, 118]. Sputum samples are commonly used as starting material for PCR. This method can be also performed on nasopharyngeal aspirates or samples obtained at bronchoscopy [170].

The collection of exhaled breath condensate (EBC) allows one to noninvasively determine parameters that are associated with physiological or disease processes [76]. In AECOPD, chemical and inflammatory markers were found to be altered, including pH, IL-1β, IL-6, IL-8, IL-10, IL-12 p70, or TNF [60, 92], however, there is only limited data on the detection of microorganisms based on NAT from EBC [57, 84].

The aim of the present study was to evaluate the feasibility and sensitivity of NAT-

based assays to detect exacerbation-associated pathogens from EBC. Patients with AE-COPD were recruited and underwent the collection of EBC and sputum. PCR-based detection assays for various bacterial and pathogens were developed and applied to these materials.

1.1.1 Role of the airway epithelial barrier

The airway epithelium together with alveolar macrophages and dendritic cells play a major role in the initial recognition of bacterial products [16, 61] getting into the lower airways with the air. Since some of these products are potent proinflammatory stimuli it is extremely important for the immune system to distinguish between pathogens and non-pathogenic commensals [174]. This prevents the development of constant inflammation and forms tolerance against harmless microbiota (see Figure 2). This process becomes much more intriguing when taking into account that commensals often share their surface molecules with pathogens [47]. Epithelial cells are equipped with very sensitive recognition tools-toll like receptors (TLRs), NOD like receptors (NLRs) and retinoic acid-inducible gene (RIG)-I-like receptors (RLRs) which recognize a broad variety of microbial structural components [5, 7, 14, 68, 72, 95, 182, 153, 131]. Detailed description, as well as structural and functional analysis of microbial pattern recognition receptors can be found in reviews of AKIRA and MEDZHITOV [5, 106]. After recognition of pathogenic bacteria proinflammatory pathways are activated and cellular components of the adaptive and innate immunity are recruited to the infection site. One key regulator in this process is the NF-κB which translocates from the cytoplasm into the nucleus and activates pro-inflammatory genes in epithelial cells and macrophages [34, 74, 113, 50].

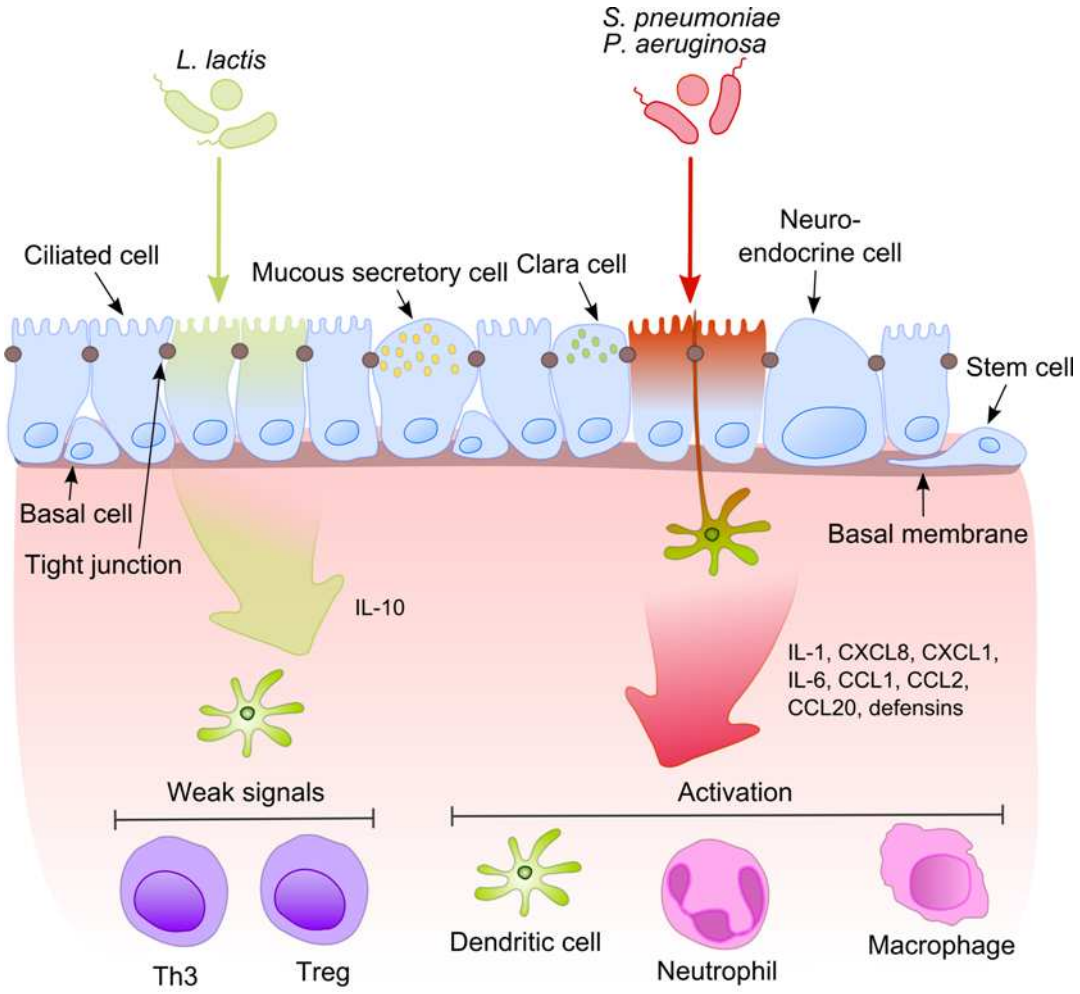

Figure 2: Mechanisms underlying the inflammation in COPD. Airway epithelium has complex structure: consists of at least seven diverse cell types interacting with each other by means of tight junctions [18, 33, 69]. Moreover, epithelial calls can deliver the signals into the underlying tissues taking part in the mechanisms of innate and adaptive immune defence. The key transmitters of the signals are dendritic cells. Once pathogenic bacterium (e.g., *S. pneumoniae*, *P. aeruginosa* [71]) has activated particular pattern recognition receptors on or in epithelial cells, the proinflammatory signaling pathways are activated. This results mainly in IL-1, IL-6 and IL-8 production [65, 90]. These cytokines induce chemotaxis to the site of infection in its target cells (e.g., neutrophils, dendritic cells [48, 129] and macrophages [4]). On the other hand, representatives of standard microbiota cause only weak signaling preventing the inflammation. The mechanism of distinguishing between harmless and harmful bacteria on the molecular as well as on physiological levels is not completely understood.

In contrast, harmless bacteria do not cause the translocation of NF-κB into the nucleus thus preventing the inflammation although they can express the same microbe-associated molecular patterns (MAMPs) [96, 47]. One possible mechanism explaining

this effect was suggested by NEISH showing that non-pathogenic *S. typhimurium* PhoPc and *S. pullorum* are able to prohibit the ubiquitination of NF-κB inhibitor molecule nuclear factor of NF-κB light polypeptide gene enhancer in B-cells inhibitor alpha (IκB-κ) [112]. Another explanation of commensal tolerance of the epithelium refers to the post-translational modification of a protein by the covalent attachment of one or more ubiquitin (Ub) monomers. The inhibition of ubiquitination leads to reduction of inflammation, because only polyubiquitinated (IκB-κ is targeted for degradation by the 26 S proteasome, allowing NF-κB translocation to the nucleus and activation the transcription of effector genes (for example IL-8). Moreover KUMAR et al. showed that probiotic bacteria such as *Lactobacilli* are able to modulate the activity of the Ub-proteasome system via inducing reactive oxygen species (ROS) production in epithelial cells [94]. In mammalian cells, ROS have been shown to serve as critical second messengers in multiple signal transduction pathways in response to proinflammatory cytokines [97]. Bacterially induced ROS causes oxidative inactivation of the catalytic cysteine residue of Ub 12 resulting in incomplete but transient loss of cullin-1 neddylation and consequent effects on NF-κB and β-catenin signaling. Another commensal species, *B. thetaiotaomicron*, attenuates pro-inflammatory cytokine expression by promoting nuclear export of NF-κB subunit RelA, through a peroxisome proliferator activated receptor γ (PPAR-γ)-dependent pathway [88]. PPAR-γ target transcriptionally active Rel A and induce early nuclear clearance limiting the duration of NF-κB action.

The balance between pathogens and commensals is extremely important in the maintenance of homeostasis in the respiratory tract [78].

1.2 Microbiome of the respiratory tract

The airways are continually exposed to a multitude of microorganisms, some of which are able to persist in and even colonize the respiratory tract. This is possible due to the presence of nutrients, oxygen, and optimal growth temperature. There are several host-derived nutrient sources for microbial residents: secretions from airway epithelial cells (especially goblet cells) and submucosal glands, transudate from plasma. Each day about 20 ml to 100 ml of airway surface liquid (ASL) is produced [6]. 90 % of its mass consists of water. ASL contains mucins (MUC5AC and MUC5B), albumin (ranging from 0.48 mg/ml in the trachea to 0.73 mg/ml in the bronchi), prealbumin, immunoglobulins, α_1-antitrypsin, α2-macroglobin, gaptoglobin and other proteins. Lipids are present at an average concentration of 1 % (mostly phospholipids). A range of glycosaminoglycans are present, including heparin sulphate, heparin, chondroitine sulphate, and hyaluronate. The main ions present are Na^+, K^+, and Cl^-. In the alveolar regions of the lungs the alveolar lining fluid (ALF) differs from ASL. ALF contains albumin (9 mg/ml),

phospholipids, vitamin C, vitamin E, reduced glutathione, transferrin (324 µg/ml), and ceruloplasmin (22.2 µg/ml). A major constituent of ALF is a mixture of surface-active compounds known as pulmonary surfactant, which lowers the surface tension of the fluid [180].

Table 1: Key members of the standard microbiota of the respiratory tract

Nasal cavity	**Nasopharynx**	**Trachea**	**Lungs**
Bacteroides sp.	*Bacteroides*	*H. influenzae*	*Fusobacterium* sp.
Diphteroides sp.	*Diphteroids*	*Neisseria* sp.	*P. adiacens*
H. influenzae	*H. influenzae*	S. *pneumoniae*	*V. atypica*
M. pneumoniae	*M. catarrhalis*	*Staphylococcus* sp.	?
Proteus sp.	*N. meningitidis*		
S. epidermidis	*Streptococci*		
S. aureus			

Moreover, the pool of available nutrients is increased by the activities of some members of the microbiota. Macromolecular components of respiratory secretions (proteins, glycoproteins, lipids, nucleic acids) are converted to nutrients (e.g. carbohydrates, amino acids). Thus, the metabolic activity of present bacteria allow for the colonization of new species (see Figure 3). The upper respiratory system is populated by numerous species of bacteria: aerobes, facultative anaerobes, aerotolerant, and anaerobes are all represented [54]. Therefore, the secretions usually contain large numbers of diphteroids, micrococci, staphylococci and smaller numbers of other bacteria that normally inhabit the environment, such as *Bacillus* sp. About 20 % of healthy humans also carry *S. aureus* in the nasal cavity. Further inside the nasal passages, the microbial population resembles that of the nasopharynx. The nasopharynx contains large numbers of microorganisms, mostly α-hemolytic *Streptococci* of the viridans group, nonhemolytic *Streptococci*, *M. catarrhalis* and diphteroids (see Table 1). Anaerobic Gram-negative bacteria, including species of bacteroides, are also normally present in large numbers in the nasopharynx.

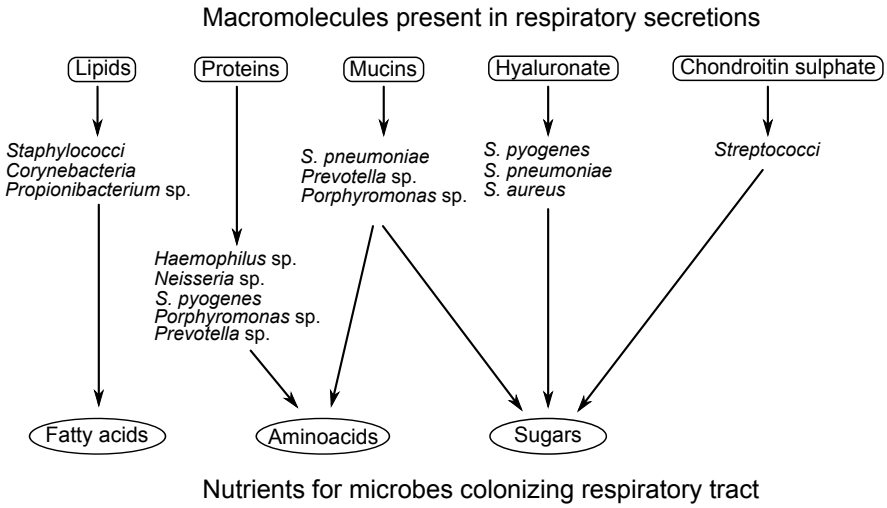

Figure 3: Major metabolic conversions which provide the nutrients for members of indigenous respiratory microbiota [180]. Since different bacteria have various metabolic characteristics the easiest way to provide all of them with nutrients is to convert large complex molecules into smaller universal constituents. In the case of complex microbial communities the most unpretentious members with a wide spectrum of enzymes supply others with fatty acids, amino acids and sugars. Not only new species can be attracted in that way, but total bacterial load of existing community can be increased.

Moreover, commonly pathogenic bacteria such as *Streptococcus*, *H. influenzae* and *N. meningitidis* are often found in this area, especially during the cooler seasons of the year. More virulent species such as *S. pyogenes*, *M. pneumoniae*, *C. diphteriae* and *B. pertussis* are sometimes found, but only in a small percentage of healthy individuals.

Microorganisms in the entire lower tract in the alveolar regions of lungs, except the most common pulmonary pathogens are not detected using conventional culture-based approaches. In diagnostic microbiology, the classification of bacteria present within clinical samples requires the prior culture of bacterial species on selective nutrient media. However, the use of selective media and growth conditions requires assumptions to be made concerning which species will be present. Consequently, sputa or bronchoalveolar lavages (BALs) are assayed routinely for only a limited number of bacterial species considered important, including *H. influenzae*, *S. pneumoniae*, *M. catarrhalis* and *P. aeruginosa*. HILTY et al. have recently published their data concerning the alterations of pulmonary microbiota in asthma and COPD patients. They found that the bronchial tree contains a characteristic microbial biota that differs between health and disease [75]. Pathogenic *Proteobacteria*, particularly *Haemophilus* sp., were much more frequent in bronchi of adult asthmatics or patients with COPD than controls, and *Bacteroidetes* were more frequent in controls.

1.2.1 Bacterial colonization and standard microbiota in COPD

SETHI marked out five different pathways by which bacteria can contribute to the aetiopathogenesis of COPD [163]:

- 10–30 % of acute exacerbations of COPD with consequent morbidity and mortality are caused by bacteria
- Invasion and persistence of bacterial pathogens in respiratory tissues alerts the host response to noxious stimuli such as tobacco smoke inducing the chronic inflammatory response and thus contributing to the pathogenesis of COPD
- Patients with COPD develop hypersensitivity to bacterial antigens which enhances airway hyperreactivity and induces eosinophilic inflammation
- Acute severe childhood respiratory tract bacterial infections damage the immature lung, impairs lung growth thereby predisposing the individual to the development of COPD
- Chronic colonization of the tracheobronchial tree by bacterial pathogens amplifies the chronic airway inflammation present in COPD and accelerates progressive airway obstruction (see Figure 1)

Chronic bacterial colonization is a promising area of research due to its possible implication in the pathogenesis of COPD [148]. ZHANG et al. showed that patients with moderate to severe COPD have bacterial colonization in lower airways even at the stable status [187]. Among the bacterial pathogens, *H. influenzae* was the most frequently isolated pathogen. Both sputum culture and bacterial colony counts could be associated with the severity of colonization. Patients with chronic colonization were associated with a significantly increased frequency of exacerbations ($p = 0.04$) and decline of forced expiratory volume in one second (FEV_1) ($p = 0.035$). An observational study of BANERJEE et al. performed with 67 subjects revealed that those with opportunistic pathogens (*H. influenzae, S. pneumoniae*, *M. catarrhalis*, *H. parainfluenzae, S. aureus, P. aeruginosa* and *K. pneumoniae*) in their sputum exhibited higher neutrophil differential counts, greater sputum supernatant levels of IL-8, leukotriene B, TNF and neutrophil elastase than the non-colonized subjects [20]. The bacteria in the airways are in a constant state of turnover, releasing extracellular products, as well as undergoing lysis with the release of a variety of proteins, lipopoligosaccharides and peptidoglycans. Lipopoligosaccharide (LPS) is a potential proinflammatory stimulus that can lead to development of emphysema. It can therefore be assumed that "colonization" is a low-grade smouldering infection that induces chronic airway inflammation. In the large airways such inflammation would contribute to mucus production and in the small airways to respiratory

bronchitis and progressive airway obstruction [163, 128]. *H. influenzae*, *S. pneumoniae* and *M. catarrhalis* are responsible for 10–30 % of exacerbations.

1.2.2 Bacteria in clinical samples

Within the body of a healthy adult both the beneficial and harmful influences of bacteria are often due to their interactions in complex communities [22]. The major colonization sites are various mucosal surfaces (such as in the gastrointestinal tract, the genitourinary tract, the oral cavity, the nasopharynx, the respiratory tract) and the human skin [121]. These communities, however, remain largely unstudied, leaving their influence upon human development, physiology, immunity, and nutrition almost entirely unknown.

In practical terms, to identify the genus and species of a bacterium the major four types of approaches are used: direct techniques (such as Gram and acid-fast stain), culture techniques (involve cultivating a specimen in nutrient medium), detection of microbial by-products (Voges–Proskauer, gelatinase, hydrogen sulfide production, phenylalanine deaminase tests) and molecular biological techniques [54]. Although the advantages of traditional culture methods are obvious (they are quantitative, isolates can be typed or fingerprinted, antibiotic sensitivity tests can be done, isolates can be archived), they are slow, require skilled and experienced personnel and only cultivable in the used culture medium bacteria can be detected. Culture media are capable of providing the *in vitro* growth of only about 1 % of the total range of bacterial species present. Furthermore, it is questionable whether or not they are actually quantitative [122]. To solve these difficulties, molecular microbiological techniques were developed during the 1960s and 1970s in research laboratories. As a result, over the past one or two decades the identification of the organisms that are infectious to humans have been greatly facilitated by the development and application of specific molecular nucleic acid amplification methods and hybridization (probe) tests [120]. The polymerase chain reaction (PCR) is among the most popular of these methods. PCR is a technique which allows amplification of a single or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence. In general, speed, increased sensitivity and the fact that these methods can be done with inadequately stored specimens are the main reasons to use molecular-based techniques [13, 181, 133]. In addition, databases and phylogenetic trees can be readily generated and automation is possible. High-throughput use of these tests, however, is still somewhat restricted due to the significant costs involved (reagents, apparatus, personnel), the fact that confirmatory tests are often necessary, as bacteria cannot be isolated for further studies, and the methods are not yet widely accepted for medicolegal testing.

1.3 Molecular techniques to identify microbial species

1.3.1 The rRNA encoding genes

By using PCR to amplify sequences that are specific to a certain species of microorganism and then determining the nucleotide sequence of that specific sequence, one can compare that sequence with the sequences of bacteria that have already been identified. Bacteria having very similar sequences of nucleotides must be closely related. 16 S rRNA gene is such a sequence in the microbial genome, which holds a special place in the characterization of microbial communities.

The bacterial ribosome consists of two subunits: 30 S (small subunit) and 50 S (large subunit). The larger subunit consists of two species of rRNA (5 S and 23 S) and more than 30 associated proteins. The small subunit consists of more than 20 associated proteins and the 16 S rRNA that is approximately 1500 nucleotides long. The genes encoding ribosomal RNAs are ancient. They are functionally constant, universally distributed and comprise of highly conserved sequence domains interspersed with more variable regions (see Figure 4). The 16 S rRNA gene plays in the translational apparatus of the small subunit of the ribosome. This explains its high level of conservation among bacteria. 16 S rRNA gene conservation among the domain *Eubacteria* is more than 60 % even among the most distantly related species [160]. In fact, the 16 S rRNA gene has frequently been called a "molecular chronometer", which reflects the evolutionary distance between two bacteria over time. It has been most commonly employed for identification purposes due to it's optimal size. The sequence which encodes the 5 S subunit is too short and doesn't provide enough information for species identification. The 23 S rRNA gene also has conserved and variable regions and provides possibly better phylogenetic resolution due to its increased length (3000 bp). However, application of 23 S rRNA gene for microbial community analysis is limited by lack of established broad-range PCR amplification and sequencing primers [114]. "Broad-range" PCR primers have been designed to recognize the conserved 16 S rRNA gene sequences and used to amplify intervening variable or diagnostic regions without the need to know any prior sequence or phylogenetic information about the unknown bacterial isolate [89, 167, 146]. The conserved regions are important for classification of higher taxa, while the variable regions can be used for differentiation between closely related species [137].

The first step in 16 S rRNA gene microbial community analysis is the DNA purification. An important point is that this procedure must occur without introducing bias due to e.g. differential lysis or recovery. All bacterial 16 S rRNA genes in the sample are subsequently amplified using primers targeting generally conserved regions.

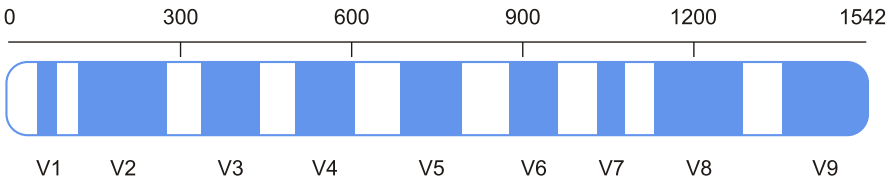

Figure 4: Schematic structure of 16 S rRNA gene

1.3.2 Terminal restriction fragment length polymorphism

For further characterization of amplified bacterial DNA fragments the terminal restriction fragment length polymorphism (T-RFLP) analysis is considered to be the most appropriate tool. T-RFLP analysis generates fragments that differ in length according to the variation in the position of the first specific restriction endonuclease site in ribosomal sequences (see Figure 5). *Msp* I ($^{C\downarrow CGG}_{GGC\uparrow C}$), *Hha* I ($^{G\downarrow CGC}_{C\uparrow GCG}$), and *Rsa* I ($^{GT\downarrow AC}_{CA\uparrow TG}$) are used most often since they were shown to resolve most often unique sequences in model communities. It is important to note that for communities with more than 50 operational taxonomic units (OTUs), none of the restriction enzymes resolved more than 70 % of the total OTUs in the communities with more than 50 operational taxonomic units (OTUs). Thus T-RFLP can most efficiently be used for communities with low or intermediate richness [145]. These fragments are typically fluorescently labeled and so allow their detection on automated DNA sequencing machines. Generated fragments are commonly species - specific in length and can be compared with existing (available on the web) fragment libraries. ROGERS et al. stated using this method could identify in cystic fibrosis patients besides such common bacteria such as *P. aeruginosa* and *H. influenzae* also other species: *P. oris*, *F. gonidoformans*, *B. fragilis*, *Leptotrichi*a-like sp., *C. murliniae* and *S. ventriculi* [132]. KAWANAMI et al. found in (BAL) samples of severe pneumonia patients also *E. faecalis*, *E. casseliflavus*, *V. parvula*, *V. atypica*, *V. dispar*, *L. wadei* and *P. nanceiensis* [87]. Both of these research teams also used sequencing as a more advanced molecular approach.

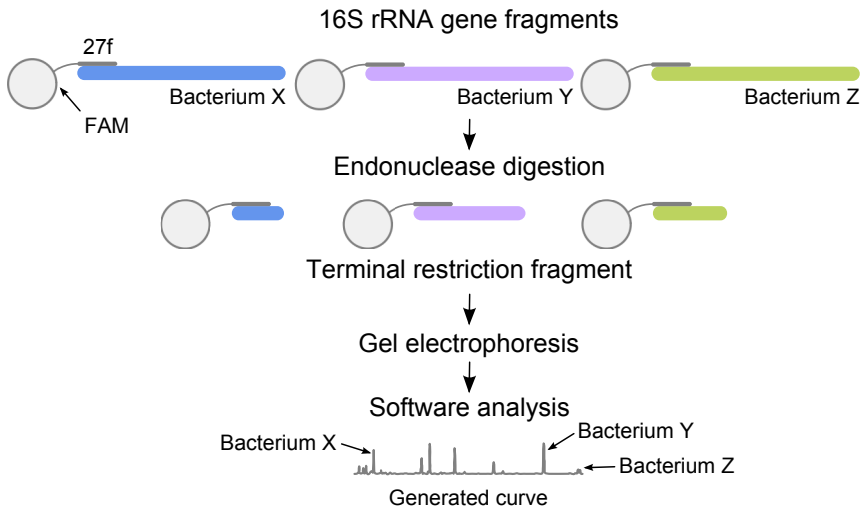

Figure 5: Terminal restriction fragment length polymorphism (T-RFLP) is a method for providing semi-quantitative analysis of the diversity and dinamics in microbial communities. In PCR amplification targeted fragment of ribosomal RNA gene for the domain *Eubacteria* is amplified. The primer set contains the fluorescent labels such as 6-carboxy-fluorescein (FAM) at the 5' end of the forward primer. PCR products are then purified, concentrated and completely digested with one or more endonucleases. Digested products are subjected to size separation with gel or capillary electrophoresis. Each electrophoregram represents a fingerprint of the microbial community. This method is indispensable in routine diagnostic because it allows one to get the overview of present microbial community in short a time.

1.3.3 Sequencing

By means of sequencing it has become possible to determine nucleotide structure of 16 S rRNA genes, compare them with existing available online databases and thus identify new bacterial species associated with respiratory tract [2, 75]. The most popular source in this field is the Basic Local Alignment Search Tool (BLAST): an algorithm for comparing primary biological sequence information, such as the amino acid sequences of different proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Another web tool the ribosomal database project (RDP) provides bacterial and archaeal ribosome related data services to the scientific community, including online data analysis, rRNA derived phylogenetic trees, and aligned and annotated rRNA sequences (http://rdp.cme.msu.edu/). The Greengenes web application also provides access to the current and comprehensive 16 S rRNA gene sequence alignment for browsing, blasting, probing, and downloading

(http://greengenes.lbl.gov/cgi-bin/nph-index.cgi).

Sanger-based sequencing (chain termination method) has dominated the molecular biology landscape over the last three decades, primarily due to the desire of the international community to sequence the entire human genome. As a result, Sanger-based sequencing was quickly adapted to a large-scale, high-throughput automation allowing the parallel sequencing of DNA in up to 384 capillaries at a time. Over the last two decades considerable resources were invested in the development of new alternative sequencing strategies, and as recently as 2005 their utility was demonstrated. These new sequencing strategies called next-generation DNA sequencing techniques provide numerous advantages including high speed and throughput, full automation and expense reduction, and the determination of sequence data from amplified single DNA fragments, negating the need of the *in vitro* cloning of DNA fragments [45]. These alternative strategies for DNA sequencing can be grouped into several categories: microelectrophoretic methods, sequencing by hybridization, real-time observation of single molecules and cyclic-array sequencing. They have recently been implemented in commercial products: 454 pyrosequencing (used in the 454 Genome Sequencers, Roche Applied Science; Basel) [9, 143], Solexa technology (used in the Illumina (San Diego) Genome Analyzer), the SOLiD platform (Applied Biosystems; Foster City, CA, USA), the Polonator (Dover/Harvard) and the HeliScope Single Molecule Sequencer technology (Helicos; Cambridge, MA, USA). Excellent discussion and comparison of all basic principles as well as devices can be found in review papers of SHENDURE and ANSORGE [11, 156, 155]. As soon as such devices appeared on the market, the new specific applications have also become accessible [62]. For example, development of metagenomic sequencing has led to the initiation of an international MetaHIT project, the main purpose of which is to study human intestinal microbial metagenome [125, 25, 29, 49, 67]. This is the branch of the Human Microbiome Initiative with primary postulated goals to determine if there is a core human microbiome, to understand the changes in the human microbiome that can be correlated with human health, and to develop new technological and bioinformatics tools to support these aims [184, 157]. To accelerate new breakthrough in DNA sequencing technologies and to enhance methods of existing analysis the READNA consortium was created. It includes 16 European partners from both academia and industry. In 2008 they introduced the new method to identify bacteria [141]. It is based on the detection of patterns of protein masses by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. The interdisciplinary nature of the consortium will allow the exploration of novel concepts of nucleic acid analysis. Further explosive development of the next-generation DNA sequencing and synthesis techniques will enable improvement of a broad range of applications. CURRAN et al. have evaluated the new diagnostic approach relevant for COPD patients using such a

molecular-based tool as DNA microarray [43]. They described a novel respiratory microarray using PCR product probes for the detection of nine bacterial species in the human respiratory tract. This technique could potentially be a useful method to monitor the populations of bacteria in respiratory samples. Microarrays offer the advantage of both unlimited target detection and simultaneous detection of organisms in clinical specimens with complex microbiota [171]. HUANG et al. used in their recent work the PhyloChip: DNA microarray for rapid profiling of microbial populations to study pulmonary microbial communities of exacerbated COPD patients [79]. This high-density microarray containing 500 000 probes can detect approximately 8500 bacterial taxa. The analysis revealed the presence of over 1200 bacterial taxa representing 140 distinct families. A core community of 75 taxa was detected in all patients. In table 2 I have summarized the most common non-cultural approaches to studying pulmonary microbial communities with some examples of supporting software and tools (see Table 2).

Generally, an optimal procedure for the identification of bacteria should comply with the following criteria: provide a universal bacterial identification capability; identify all bacteria in a sample quantitatively; detect emerging previously uncharacterized bacteria and find the most closely related species; make possible immediate and flexible detection of bacterial samples, as well as high-throughput analysis; require low operational costs; process the transfer of information rapidly; and facilitate sharing of data by establishing easily accessible and regularly updated databases [141, 140].

Table 2: Molecular biology-based approaches.

Method	**Applications**	**Programming tools**
T-RFLP	Whole picture of microbial community, semi-quantitative	GeneMapper — curve plotting
PCR, qPCR	Targeted pathogen identification; quantitative	FastPCR — a free software for design PCR primers; GenEx software — qPCR data analysis
Microarrays	Unlimited target detection, simultaneous identification	Cluster — hierarchical clustering, self-organizing maps; GeneXplorer — web-visualization of microarray datasets; MicroArray Suite — extracting and visualizing DNA microarray data
Sequencing	Identification of new species	RDB and BLAST — sequence databases; DOTUR — estimating species richness; PHRED — sequence quality control; Lusy — sequence quality trimming and vector removal
Next-generation sequencing	Identification of new species and genes, metagenomic studies, comparative genomics, gene regulation studies, targeted resequencing, DNA-protein interactions	CLCbio Genomics Workbench — *de novo* and reference assembly of Sanger sequencing data; RCluster — hierarchical clustering, self-organizing maps; GeneXplorer — web visualization of microarray datasets; MicroArray Suite — extracting and visualizing DNA microarray data; Illumina, Helicos, and SOLiD data Galaxy — interactive and reproducible genomics; Genomatix — integrated solutions for next Ggeneration sequencing data analysis; JMP Genomics — visualization and statistics tool; NextGENe — *de novo* and reference assembly of Illumina, SOLiD and Roche FLX data; SeqMan Genome Analyser — next generation sequence assembly of Illumina, Roche FLX and Sanger data; SHORE — for Short Read, is a mapping and analysis pipeline for short DNA sequences produced on a Illumina Genome Analyzer

2 Aims

The lower airways are constantly exposed to a broad variety of environmental microbes. But there is a lack of information about bacterial diversity in the respiratory tract, because only classical culture techniques are steel used routineously in the diagnostic laboratories. In this thesis it was dealt exclusively with pulmonary microbial communities with a view to determine their compositional changes linked with COPD.

The aims of the study were:

1. To detect the most common respiratory pathogens by the PCR in sputum and exhaled breath condensate samples of AECOPD patients
2. To test the suitability of culture independent molecular biology-based approach — terminal restriction fragment length polymorphism (commonly used in environmental microbiology) for processing of clinical samples
3. To characterise the pulmonary microbial communities presenting in the BAL samples of COPD patients and healthy individuals by sequencing
4. To provide more deeply knowledge about stable core microbiota of human lower respiratory tract

3 Materials and methods

Table 3: Chemicals used in work

Manufacturer	**Reagent**
Carl Roth GmbH, Karlsruhe, Germany	0.5 mm zirconia beads Agarose ultra Ampicillin sodium salt CASO boullion Chloroform:isoamylalcohol 24:1 EDTA Ethidium bromide IPTG LB agar medium LB bouillon Lysozyme $Na_2HPO_4 \cdot 2H_2O$ NaCl Nutrient broth Paraformaldehide Phenol:chloroform:isoamylalcohol PIPES Polyethylene glycol 6000 SDS Sodium EDTA Tris - HCl Tween
Sigma-Aldrich GmbH, Munich, Germany	Boric acid l Fluorescein isothiocyanate (FITC) KCl Methanol RNase Inhibitor Trisma® base Water Molecular Biology Reagent X-gal
QIAGEN GmbH, Hilden, Germany	Proteinase K
Roche Diagnostics, Mannheim, Germany	Triton X-100
MERCK KGaA, Darmstadt, Germany	$CaCl_2$ $NaH_2PO_4 \cdot H_2O$
Finnzymes Oy, Espoo, Finland	DMSO
Mallinckrodt Baker Deutschland, Griesheim, Germany	$MgCl_2 \cdot 6H_2O$
PAA Laboratories GmbH, Pasching, Austria	Dulbecco's PBS

3.1 Participant selection and sampling procedure

The study was approved by the ethics committee of the University of Marburg and Saarland University; all participants gave written informed consent. There were two groups: healthy people, who have never smoked, without immunosuppressive or other life-threatening disorders and haven't any respiratory illnesses for one year prior to the study and COPD patients in a stable state (see Table 4). The diagnosis of COPD is today based on lung function measurement. The qualitative parameter is bronchodilator ratio of forced expiratory volume in 1 s (FEV_1)/vital capacity (VC) or forced vital capacity (FVC). If it is below 0.7 the COPD is diagnosed [98]. The mean value of FEV_1/FVC ratio was 0.576, C reactive protein (CRP) level in serum did not exceed 4.8 mg/l and the mean number of leukocytes was $8.0 \cdot 10^9$ cells/l. The patient was defined as stable if there was no exacerbation for the previous 4 weeks and the patient was not currently having an exacerbation.

Table 4: Participant characteristics

Parameter	**Lavage study**		**AECOPD study**
	Healthy	**COPD**	
Age (yrs)	27 ± 2.7	63 ± 6.3	70.7 ± 8.05
Sex, male (%)	55.6	55.6	18 (62.1)
Smokers/never-smokers, n	0 / 9	9 / 0	4 / 23 / 2
Pack years	—	53.9 ± 25.2	42.24 ± 32.7
FEV 1, % predicted	—	50.2 ± 14.4	39.35 ± 11.35
FVC , % predicted	—	57.6 ± 15.74	54.54 ± 15.38
Cells/ml, $\cdot 10^5$	0.6	6.7	—
Leukocytes, $\cdot 10^9$ /L	—	8 ± 1.89	11.93 ± 4.85
Total DNA concentration in BAL, ng / µL	195.6	201.3	—

Bronchoscopy was performed by Prof. Dr. Dr. Bals and Dr. Rentz using a flexible fiberoptic bronchoscope following mild sedation and local anesthesia with xylocaine 2 %. Sterile isotonic saline was instilled into the bronchi in 10 ml aliquots and collected bronchoalveolar lavage fluid samples were immediately sent to the research laboratory.

Sputum and EBC samples from 29 patients (mean age 70 years, range from 52 to 83) with AECOPD were collected between December 2008 and April 2009 (see Table 4). AECOPD were defined in patients with known COPD presented with at least two of the following symptoms (increased dyspnoea, increased sputum purulence, increased sputum production) [64]. The study was approved by the ethics committee of the University of Marburg and informed consent was obtained from all patients.

EBC was collected according to the published ERS guidelines [76]. In short, EBC

samples were collected during 10 min of quiet breathing through a single-use disposable RTube™ collector (Respiratory Research, Inc., Charlottesville, USA), while subjects were wearing a nose clip. The aluminum sleeve of the device had been cooled to an initial temperature of −20 °C prior to collection. After collection, the plunger was used to pool the condensed material within the tube into a single sample (about 1.0 ml). Samples were stored in reaction tubes at a temperature of 4 °C. Immediately after EBC collection patients proceeded to expectorated sputum sampling procedure. Sputa were collected into sterile tubes applying a standard operation protocol. Each sputum and EBC sample was divided into two equal parts used for the DNA and RNA isolation (approximately 500 µl for each sample type). To prevent RNA degradation, EBCs were mixed with 10 volumes of RNAlaterTM RNA stabilization reagent (QIAGEN GmbH, Hilden, Germany).

3.2 Nucleic acid amplification tests

3.2.1 DNA purification

BAL samples were stored on ice prior to the DNA extraction. First lavages were centrifuged by 2500 rpm at 4 °C for 10 min. Then the supernatant was discarded and the cell pellet was used as a starting material for the DNA isolation.

Total DNA was purified by using a combination of a bead-beating method with the standard phenol:chloroform extraction. Each cell pellet was resuspended in 750 µl of sodium phosphate buffer and 250 µl of TNS solution. Then 0.7 g of sterile zirconium beads were added. After bead beating for 45 s at 6.5 m/s (using the homogenizer Precellys® 24, PEQLAB Biotechnologie GmbH, Erlangen, Germany), the samples were put on ice. To sediment the remaining cells homogenates were centrifuged for 10 min at 4 °C. The supernatants were placed in 2 ml vials on ice and the proteins were denaturated, with one volume of phenol:chloroform:isoamylalcohol solution. Following the spinning down for 10 min at 4 °C 800 µl of supernatant were added to the equal volume of chloroform:isoamylalcohol. After centrifugation under the same conditions 650 µl of supernatant were mixed thoroughly with two volumes of PEG 6000 and inverted 10 times. The extracted DNA was precipitated by spinning for 80 min at 4 °C. The liquid phase was removed by pipetting and 500 µl of ice-cold 70 % ethanol was added to wash the the nucleic acid pellet. After the spinning down and removing the ethanol the DNA was dried briefly at room temperature for 5 min. Finally it was eluted in 50 µl of EB buffer and additionally incubated at 37 °C for two hours to enhance the DNA recovery yield. Following each DNA isolation its concentration was determined using NanoDrop™ 8000 (Thermo Fisher Scientific Inc., Wilmington, USA). As an additional step, we performed an overnight proteinase K digestion prior to the extraction in the case of sputum to lyse the mucous components.

Table 5: Buffers and solutions used for the DNA extraction

Solution	Composition	Features
Sodium phosphate buffer	112.87 mM Na_2HPO_4 7.12 mM NaH_2PO_4	pH = 8, sterilize by filtering or autoclaving
TNS solution	500 mM Tris-HCl, 100 mM NaCl, 10 % SDS (weight/volume)	pH = 8, sterilize by filtering or autoclaving
Phenol:chloroform: isoamylalcohol	25:24:1	pH = 7.5–8.0
Chloroform: isoamylalcohol	24:1	
PEG 6 000 solution	30 % (weight/volume) polyethylene glycol 6 000 in 1.6 M NaCl	prepare with RNAse free water, sterilize by autoclaving
EB buffer	10 mM Tris-HCl	pH = 8, prepare with RNAse free water, sterilize by filtering or autoclaving

As an additional step, an overnight proteinase K treatment was performed prior to the extraction to lyse the mucous components of sputum. DNA was extracted from positive controls (*M. catarrhalis* MCCM 01214, *L. pneumophila* DSM 7513, *M. pneumoniae* ATCC 29342 and *C. pneumonia* VR-1360; and clinical isolates *S. aureus*, *H. influenzae*, *S. pneumoniae*) using the DNeasy© Blood & Tissue Kit (QIAGEN GmbH, Hilden, Germany) according to the user's manual. RNA was extracted using the NucleoSpin© RNA L kit (Macherey-Nagel, GmbH, Düren, Germany) and its protocol for the total RNA purification from cultured cells and tissue. Reverse transcription was performed using the Fermentas cDNA synthesis kit (Fermentas, St. Leon-Rot. Germany). As positive controls the RNA isolated from influenza viruses AH 1, AH 3 and respiratory syncytial virus was used (kindly provided by Dr. Markus Eickmann, Institute for Virology, Philipps University of Marburg). Nucleic acids isolated both from sputum and EBC specimens were dissolved in a final volume of 50 µl of elution buffer (10 mM Tris-HCl, pH = 8.5 for DNA) or sterile RNase-free water (for RNA) and stored at −20 °C.

3.2.2 Polymerase chain reaction

Pure total DNA (100 ng) was used as a starting material for the PCR-based amplification of the 16 S rRNA gene. Used primer set generates a 900 bp long PCR product from *E. coli* DH5α DNA. This enhances the accuracy rate in the sequence database search. The following oligonucleotides were obtained from Metabion GmbH (Martinsried, Germany): forward primer Ba27f, 5'-AGA GTT TGA TCC TGG CTC AG-3'; reverse primer Ba907r, 5'-

CCG TCA ATT CCT TTR AGT TT-3'. PCR mixtures were composed of 1 x PCR buffer, 2 % of DMSO, 0.2 mM concentrations of each deoxinucleoside triphosphate, 0.5 μM concentrations of each primer, and 1 U of Phusion Hot Start DNA polymerase (Finnzymes Oy, Espoo, Finland) in a final volume of 50 μl. This polymerase has proofreading activity and needs the heating procedure of 98 °C for 30 s to activate the enzyme. The samples were subjected to the initial denaturation step of 94 °C for 3 min and to 40 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 45 s, and extension at 72 °C for 1 min and 30 s, followed by a final extension step at 72 °C for 7 min. Amplification was carried out by using a MyCycler Thermal Cycler (Bio-Rad, Munich, Germany). In each amplification run the total *E. coli* DNA (100 ng) was used as a positive control and 5 μl of Water Molecular Biology Reagent as a negative control. PCR products were stored at −20 °C.

To detect typical pulmonary pathogens in sputum and EBC samples conventional PCR approach with specific primer sets was used. Identified pathogens with corresponding primers are listed in Table 6 [28, 73, 108, 173, 77, 154, 32]. All reactions were performed in final volume of 25 μl using Phusion DNA Polymerase (Finnzymes, Inc., Espo, Finnland) in a Bio-Rad iCycler (Bio-Rad, Munich, Germany). The PCR conditions are indicated in the references. A positive and a negative (water) control were included with each PCR assay. All control PCR reactions were performed in separate 0.2 ml PCR tubes with flat caps (Fisher Scientific GmbH, Schwerte, Germany) to avoid any possible contamination of clinical samples with control DNA.

Table 6: Primer used for pathogen detection assays

Pathogen	Forward primer (s) (5'-3')	Reverse primer (s) (5'-3')
S. aureus	CGC ATT GAT GGT GAT ACG GTT	AGC CAA GCC TTG ACG AAC TAA TGC
H. influenzae	CTA CGC ATT TCA CCG CTA CAC	CGT ATT ATC GGA AGA TGA AAG TGC
M. catarrhalis	CTA CGC ATT TCA CCG CTA CAC	CCC ATA AGC CCT GAC GTT AC
S. pneumoniae	CTA CGC ATT TCA CCG CTA CAC	AAG GTG CAC TTG CAT CAC TAC C
L. pneumophila[1]	AAG AAT AGC CTG CGT CCG AT	GTC AAC TTA TCG CGT TTG CT GAG GGT TGA TAG GTT AAG AGC
M. pneumoniae	AGG CTC AGG TCA ATC TGG CGT GGA	GGA TCA AAC AGA TCG GTG ACT GGG T
Influenza virus AH 1[2]	CAG ATG CAG ACA CAA TAT GT ATA GGC TAC CAT GCG AAC AA	AAA CCG GCA ATG GCT CCA AA CTT AGT CCT GTA ACC ATC CT
Influenza virus AH 3[3]	CAG ATT GAA GTG ACT AAT GC AGC AAA GCT TTC AGC AAC TG	GTT TCT CTG GTA CAT TCC GC GCT TCC ATT TGG AGT GAT GC
Respiratory syncytial virus	GTC TTA CAG CCG TGA TTA GG GAT GTT ACG GTG GGG AGT CT	GGG CTT TCT TTG GTT ACT TC GTA CAC TGT AGT TAA TCA CA
C. pneumoniae	GTT GTT CAT GAA GGC CTA CT	GTG TCA TTC GCC AAG GTT AA

3.2.3 Agarose gel electrophoresis and purification of PCR products

The 1.5 % (w/v) agarose gels prepared on tris-borate buffer (0.045 M Tris-borate, 0.001 M EDTA, pH = 8.0) were used to analyze PCR products. Addition of 0.01 % of as intercalating agent such as ethidium bromide (v/v) allowed visualizing the DNA fragments under the ultra violet light (Gel Doc™ 2000, Bio-Rad, Munich, Germany). Voltage (E143, Consort nv, Turnhout, Belgium) of 5 V/cm (measured as the distence between the electrodes) was applied to the gel tank (HU 13, Bio-Rad, Munich, Germany). Marker DNAs (O'GeneRuler ™, Fermentas GmbH, St. Leon-Rot, Deutschland) were loaded into slots on both the right and left sides of the gel and 6 x Orange DNA Loading Dye (Fermentas GmbH, St. Leon-Rot, Germany) was mixed with each DNA sample prior to the pipetting.

[1] Seminested PCR
[2] Nested PCR
[3] Nested PCR

3.2.4 Terminal restriction fragment length polymorphism analysis

The terminal restriction fragment length polymorphism (T-RFLP) analysis was introduced at the end of the 1990s [104]. The method involves restriction enzyme digestion of PCR amplified 16 S rRNA gene. The PCR amplification entails fluorescent-labeled forward primer, which allows detection of the digests by a DNA sequencer. Since the locations of the restriction sites are not random, but have a phylogenetic background, T-RFLP provides knowledge of the taxonomic position of the test bacteria [10, 31, 103] (see Figure 6).

The T-RFLP analysis was performed using FAM (6-carboxyfluorescein)-marked PCR products. For the PCR step the forward primer Ba27f (5'-AGA GTT TGA TCC TGG CTC AG-3') was labeled with FAM (Metabion GmbH, Martinsried, Germany). The labeled PCR products were cut from the agarose gel and purified with a commercial kit (NucleoSpin®Extract II, MACHEREY-NAGEL GmbH & Co. KG, Dueren, Germany) and then subjected to restriction enzyme digestion. The reaction mixture composed of 120 ng of PCR product, 2.5 U of restriction enzyme *Msp*I (Promega GmbH, Mannheim, Germany) 1 µl of 10 x reaction buffer and 1 µg of acetylated bovine serum albumine (BSA) supplied by the same manufacturer. The final volume was adjusted to 10 µl with deionized sterile water. Following incubation for 3 h at 37 °C in the dark, samples were proceeded to the separation and detection of the digested products via capillary electrophoresis on ABI 310 Genetic Analyzer (Applied Biosystems Deutschland GmbH, Darmstadt, Germany) using the The GeneScan™ 500 ROX™ XL size standard (Applied Biosystems Deutschland GmbH, Darmstadt, Germany). It is a ROX™ dye-labeled size standard containing 16 ROX™ dye-labeled, single-stranded DNA fragments suitable for the reproducible sizing of fragment analysis data. Generated data was analysed with GeneMapper® Software (Applied Biosystems Deutschland GmbH, Darmstadt, Germany). For each sample the analysis was performed in duplicate.

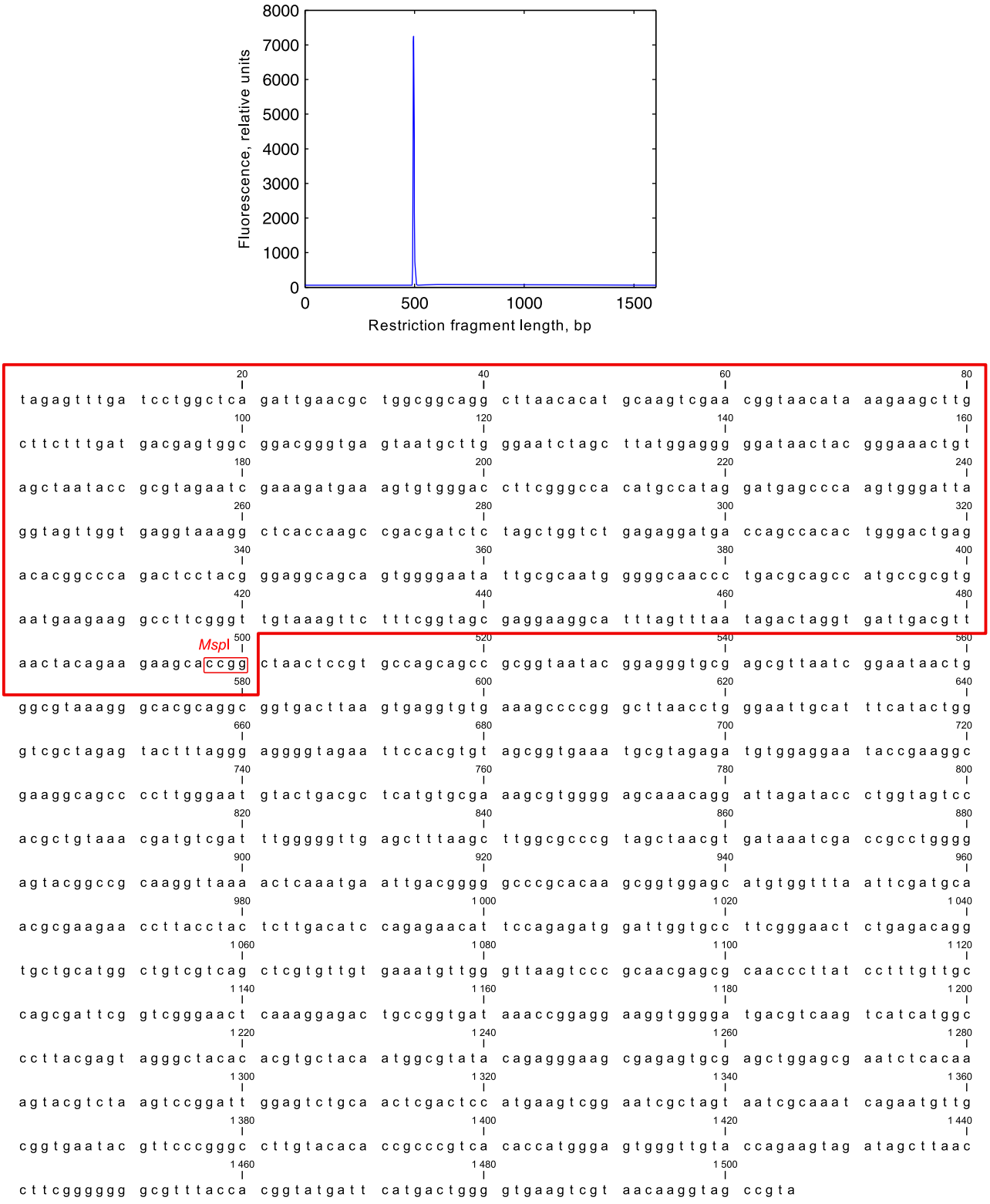

Figure 6: A control T-RFLP profile, generated by enzymatic digestion with *Msp*I of amplified 16 S rRNA gene of *H. parainfluenzae* strain CIP 102513. In the reference sequence obtained from the GenBank of the National Center for Biotechnology Information the restriction site for *Msp*I is marked with a red box.

3.2.5 Analysis of T-RFLP data

The electropherogram supplied with Excel table was obtained as output data after capillary electrophoresis during T-RFLP analysis (see Figure 7). Electropherogram represents the composition of microbial community. The *x* axis expresses the length in bp of cut terminal fluorescently marked fragments and the *y* axis — the relative fluorescence intensity of each fragment. The relative fluorescence intensity is the indirect degree of particular fragment frequency in a given sample.

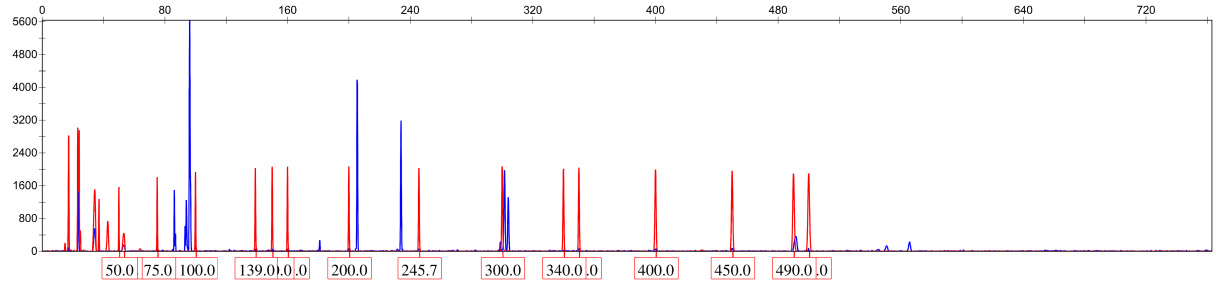

Figure 7: Electropherogram

Terminal restriction fragments between 50 bp and 700 bp were included in the analysis. Although the longer fragments can be measured as well by means of an extrapolation option on the software, but it is connected with the loss of accuracy. Additionally the lowest signal threshold of 50 relative units of fluorescence intensity was taken into account. This prevented the recording of false peaks and background noise [55]. Tables were extracted for each sample with peak size *vs.* fluorescence intensity and terminal restriction fragments (TRFs) that differed by ±1 bp in different profiles were considered as identical in order to compare T-RFLP profiles between different samples. Final T-RFLP analysis comprehends the comparison of terminal restriction fragments of different samples, in the first place the identification of unique and common peaks.

Microbial diversity was estimated with the Shannon-Wiener diversity index (H') and species richness value using the Excel software (Microsoft, USA). This index takes into account both the number of species present and the proportion of the total accounted:

$$H_S = -\sum_{i=1}^{S} p_i \log p_i,$$

where S — the number of species and p_i — the relative abundance of each species, calculated as the proportion of individuals of a given species to the total number of individuals in the community. It takes into account the number of phylotypes as well as the evenness of band intensity on a gel. In contrast, the phylotype richness ignores the

relative abundance of each phylotype and focuses only on the total number of unique phylotypes in a particular community.

The relative abundance of each TRF within a given T-RFLP pattern was calculated as the peak height of the respective TRF divided by the total peak height of all TRFs detected within a fragment length range between 50 bp and 600 bp.

3.2.6 A-tailing procedure

Since the Phusion Hot Start DNA polymerase generates blunt-ended fragments PCR products were modified using the A-tailing procedure. Volume of 5 µl of purified PCR fragment was mixed with 5 U of SupraTherm™ *Taq* DNA polymerase (Ares Bioscience GmbH, Cologne, Germany), 0.2 mM concentrations of each deoxinucleoside triphosphate and 1 µl of 10 x PCR reaction buffer containing 15 mM $MgCl_2$ (Ares Bioscience GmbH, Colog ne, Germany). Deionized sterile water was added to a final volume of 10 µl. The reaction mixture was than incubated at 70 °C for 30 min. Using this method, only one insert could be ligated into the vector (as opposed to multiple insertions that can occur with blunt-ended cloning). In addition, with T-vector cloning there is no need to dephosphorylate the vector, and there is a low background of religated vector.

3.2.7 Ligation

PCR amplification of DNA samples with universal eubacterial primer set resulted in generating a mixture of 16 S rRNAgene amplicons belonging to different bacteria. This enabled the determination of microbial composition of the sample using sequencing. To separate the amplicons we performed the ligation reaction following the transformation of competent *E. coli* DH5α with the obtained plasmid vector [91].

For cloning the pGEM®-T Easy Vector System obtained from Promega GmbH (Mannheim, Germany) was used (see Figure 8). Each poly (A)-tailed PCR product in a volume of 2 µl was mixed with 5 µl 2 x rapid ligation buffer for T4 DNA Ligase, 1 µl (corresponds to 50 ng) of pGEM®-T Easy Vector and 1 µl of T4 DNA Ligase (3 Weiss U/µl). The total volume was adjusted to 10 µl with deionized nuclease-free sterile water. Positive control (provided in the kit insert DNA) as well as background control (without DNA insert) and transformation control (0.1 ng of uncut plasmid DNA) were included. After mixing by pippeting, the reactions were incubated overnight at 4 °C.

The appropriate amount of the PCR products to include in the ligation reaction was calculated using the following equation:

$$\frac{ng\,of\,vector \cdot kb\,size\,of\,insert}{kb\,size\,of\,vector} \cdot \frac{vector\,molar\,ratio}{kb\,size\,of\,vector} = ng\,of\,insert.$$

Since the insert:vector molar ratio 3:1 provides good results it was used in the calculations.

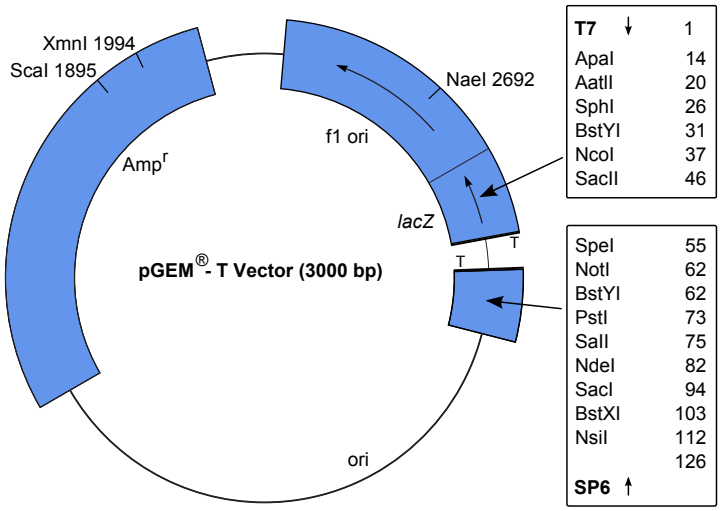

Figure 8: pGEM ® -T Vector Map and Sequence Reference Points

$$\frac{50\,ng\,vector \cdot 0.9\,kb\,insert}{3.0\,kb\,vector} \cdot \frac{3}{1} = 45\,ng\,of\,insert.$$

The pGEM®-T Easy Vector has been linearized at base 60 with EcoRV and a T added to both 3'-ends. The EcoRV site will not be recovered upon ligation of the vector and insert.

The pUC / M13 forward pUC / M13 reverse primers were used for sequencing of the inserts. Using the universal eubacterial primer set would lead to the mismatches caused by contamination of isolated plasmid DNA with chromosomal DNA of the host strain *E. coli* DH5α.

3.2.8 Preparation of chemically ultra-competent *E. coli* cells

The given protocol is thoroughly described in the paper of INOUE et al. [82]. It is based on slow growth, $MnCl_2$-mediated competence and shock-freezing.

First one isolated colony of *E. coli* DH5α was picked from an over night LB agar growth plate (37 °C) and incubated in 3 ml of LB bouillon. The next day cells were grown at 37 °C and the optical density[4] was measured at 578 nm (Ultrospec™ 2100 *pro* UV/Visible Spectrophotometer, GE Healthcare Europe GmbH, Munich, Germany) to determine the cell concentration:

$$OD_{578\,nm} \cdot 10^9 = cells/ml.$$

Three individual flasks containing 400 ml of autoclaved SOB medium (see Table 7) were inoculated with $1{\cdot}10^9$, $2{\cdot}10^9$, and $4{\cdot}10^9$ *E. coli* DH5α cells.

Table 7: Materials used for cloning

Name	Composition	Protocol
SOC medium	20 g bacto-tryptone 5 g bacto-yeast extract 0.5 g NaCl 20 mM glucose 1 L H_2O	pH = 7.0, adjust with 5 N NaOH; sterilize by filtration throgh a 0.22-micron filter
SOB medium	20 g bacto-tryptone 5 g yeast extract 8.55 mM NaCl 2.5 mM KCl 10 mM $MgCl_2$ 1 L H_2O	pH = 7.0, sterilize by autoclaving
Transformation buffer	10 mM PIPES 15 mM $CaCl_2$ 250 mM KCl 55 mM $MnCl_2$[5]	pH = 6.7, sterilize by filtering
LB / ampicillin / IPTG / X-Gal plates	100 mg / ml ampicillin stock solution[6] 35 g of LB agar medium 1 L H_2O 100 mg / ml X-gal stock solution[7] 100 mM IPTG stock solution[8]	Cool down the autoclaved LB agar medium to 55 °C, add ampicillin stock solution to the end concentration 100 µg / ml, pour 25 ml in a 10 cm dish.

After 48 hours of growth at 17 °C the OD_{578} = 0.4–0.6 was reached and the culture was chilled for 10 min on ice. The cells were pelleted at 2500 g (Heraeus Megafuge 1.0R, DJB Labcare Ltd, Buckinghamshire, England) at 4 °C. After this, the cells were re-suspended in 100 ml of ice-cold transformation buffer (see Table 7), incubated for 10 min on ice and pelleted as above. Then 20 ml of ice-cold transformation buffer and 7 % (v/v) of DMSO were added and the suspension was incubated for additional 10 min on ice. It was extremely important not to exceed this time period, otherwise the survival rate of cells would rapidly decline. Finally the cells were aliquoted into sterile 1.5 ml tubes and shock-frozen immediately in liquid nitrogen. Each aliquot (50 µl) was stored at −80 °C.

[5]The buffer was adjusted to near final volume and the pH was titrated prior addition of MgCl, since a precipitate will be formed

[6]dissolved in sterile deionized water and stored at –20 °C

[7]dissolved in DMF and stored in the dark at –20 °C

[8]dissolved in deionized nuclease-free sterile water and stored at –20 °C

3.2.9 Transformation using pGEM®-T Easy Vector

Tubes containing the ligation reactions were centrifuged shortly to sediment the contents on the bottom of the tubes. Then 2 µl of each reaction was transferred into a 1.5 ml microcentrifuge tube (Biozym Biotech Trading GmbH, Hess. Oldendorf, Germany) and put on ice. Highly competent *E. coli* DH5α aliquots were removed from the −80 °C freezer and thawed on ice for 5 min. Cells were mixed by gently flicking the tubes and 50 µl of each aliquot was carefully transferred into each tube with the ligation reaction mixture. After gentle flicking, the tubes were placed on ice for 20 min, heat-shocked for 45 s in a water bath at exactly 42 °C without shaking and then returned on ice for 2 min. Next 200 µl of room-temperature SOC medium was added to the tubes containing cells transformed with ligation reactions and they were incubated for 1.5 hrs at 37 °C with shaking (150 rpm). 25 µl and 50 µl of cell suspensions were plated on LB agar plates containing ampicillin, IPTG and X-Gal, and incubated overnight (16 h to 24 h) at 37 °C. The next day transformation efficiency was evaluated using the following initial data: 50 µl of competent cells were mixed with 0.1 ng of uncut plasmid DNA and dissolved in 200 µl of SOC medium. Thus the concentration of plasmid vector DNA was:

$$0.1\,ng \cdot 1\,ml/0.25\,ml = 0.025\,ng/ml.$$

Transformation efficiency:

$$\frac{178\,CFU}{0.000625\,ng} = 2.8 \cdot 10^5\,CFU/ng = 2.8 \cdot 10^8\,CFU/\mu g\,DNA.$$

Recombinant clones were identified by blue-white screening because successful cloning of an insert into the pGEM®-T Easy Vector interrupts the coding sequence of β-galactosidase.

3.3 Phylogenetic tree construction

Obtained sequences were assembled into a consensus sequence using the DNA sequence analysis software Ugene (Unipro Novosibirsk, Russia). For sequence alignment analysis, approximately 900 bps of the sequence (5'–3', with primer ends) were compared with those available in the GenBank of the National Center for Biotechnology Information by BLAST search. Identification of the species level was defined as at least 98 % similarity of the 16 S rRNA gene sequence to the sequence of its closest bacterial relative in the GenBank database using online software BlastN (National Center for Biotechnology Information (NCBI), USA). The multiple alignment analysis was performed using Ugene. Multiple sequence alignment algorithms insert gaps in order to align the sequences to maximize similarity according to the evolutionary model summarized in the substitution matrix. Gaps correspond to an insertion or deletion of a substring (sometimes a

single residue). Gaps can occur because of single mutations, DNA slippage. The "gap open cost" is the error of introducing gaps in an alignment and the "gap extension cost" is the error of every extension past the initial gap. Varying the "gap open" and "gap extension" costs not only produces very different alignments but produces different distributions of phylogeny scores [138]. We used such gap settings: gap open cost — 10, gap extension cost — 1, end gap cost — as any other. Gaps and regions of alignments for which homology of residues could not be reasonably assumed were excluded from the phylogenetic analysis. The construction of phylogenetic trees is closely related to that of multiple alignments. Some progressive alignment techniques even use a phylogenetic tree to determine the order of the progressive adding of sequences [17]. The neighbor-joining tree was constructed using Ugene and uploaded to the Interactive Tree of Life project (http://itol.embl.de/) for annotation [102]. Bootstrapping was carried out 1 000 times to evaluate the tree statistically.

3.3.1 Sequence analysis

The Sequin program (http://www.ncbi.nlm.nih.gov/projects/Sequin/) was used to submit the sequences to the GenBank® database. Chimeras (DNA sequences composed of DNA from two or more parents: artifacts made during the PCR process [53]) and vector sequence fragments were removed using the Black Box Chimera Check (B2C2) software (http://www.researchandtesting.com/B2C2) and the Vector Screening tool (RunVecScreen, Sequin) respectively [66]. The method of chimeric sequences exclusion extracts 100 bp long contiguous start and end regions of each FASTA format input sequence. These regions are then aligned against a custom BLAST database containing 7 199 high quality sequences consisting of a representative, verified set of bacteria with full taxonomic information (at all 7 major levels). Each species of bacteria was only included once to minimize search space. When sequences are aligned against this database, taxonomic information of each of the regions is evaluated and the percentage of matching taxonomic levels is calculated from species back to the kingdom level. The higher the taxonomic identity between both regions, the less likely it is that the sequence is chimeric [66]. The parameters of search were as follows: region of analysis — ends, number of base pairs to analyze — 100, minimal sequence length to be considered — 100, lower threshold — 4, upper threshold — 6. The short sequences and "Definite Chimeras" were removed.

3.3.2 OTUs definition

The phylogenetic tree was constructed using distance neighbor-joining method (Phylip package). of phylogenetic analysis. A basic concept of neighbor-joining method is the

concept of "neighbors". The distances are assumed to be additive, meaning that the pairwise distance between two organisms can be achieved by adding the distances of each branch in the tree. A pair of operational taxonomic units (or aligned sequences) are neighbors if they are connected by a node. Operational taxonomic unit (OTU) — is just that: a defined level which taxonomists use to discuss or compare organisms, the terminal level at which that taxonomy classifies the sequences. While it might be all the way down to the specific strain for one taxonomy it might only be to sub-order for another (http://greengenes.lbl.gov/cgi-bin/nph-index.cgi). Although the OTU cut-off depends on the diversity of a particular microbial community, it is considered that OTUs that are defined by distances of less than 0.03 correspond to a strain-level delineation, of 0.03 correspond to a species, of 0.05 correspond to a genus, of 0.15 correspond to a class, and of 0.20 to 0.30 correspond to a phylum [144, 119]. To determine the distances between aligned sequences the distance matrix view was used (Unipro Ugene, Novosibirsk, Russia).

3.4 Fluorescence *in situ* hybridization

The *in situ* detection of nucleic acid sequences (genes on chromosomes, mRNA in tissues) provides a direct visualization of the spatial location of specific sequences. The *in situ* hybridization takes advantage of the specific annealing of complementary nucleic acid molecules (DNA or RNA) through hydrogen bonds formed between bases attached to the sugar phosphate backbone. This base pairing underlies the formation of a double stranded-complex, in which one strand has opposite orientation to the other with respect to the sugar phosphate backbone. The sequence is read from 5' to 3' (this refers to the position of the sugar at which the phosphate residues are attached). Any nucleic acid sequence can therefore be specifically detected by use of a probe that is the "antisense" reverse complementary sequence [178, 30]. At the end of the 1990s AMANN et al. had optimized this method for the identification of bacteria in polymicrobial communities [8].

FISH was performed on human and mice lung tissue sections. In order to preserve the cell and tissue structure, specimens were fixed with formaldehyde for 12 h at 4 °C. Freshly prepared 4 % (w/vol) paraformaldehyd solution in 0.1 M phosphate buffered saline (pH = 7.2) was used as a fixative agent. SM 2000R sliding microtome (Leica Mikrosysteme Vertrieb GmbH, Wetzlar, Germany) was used to prepare the microsections. 1 µm to 2 µm paraffin sections were collected on bio-adhesive coated SuperFrost PLUS® glass slides (VWR International GmbH, Darmstadt, Germany). Prior to FISH, slides were dewaxed: xylene — 20 min, xylene — 10 min, 100 % ethanol — 5 min, 100 % ethanol — 5 min, 96 % ethanol — 5 min and 70 % ethanol — 5 min.

In order to detect the general bacterial colonization in lower airways, the universal eubacterial oligonucleotide probe EUB 338 (5'-GCT GCC TCC CGT AGT AGG AGT-3') was used. This probe is complementary to a region of the 16 S rRNA gene. The 5' end of the probe EUB 338 was labeled with the fluorophore Alexa 647 (absorption 650 nm, emission 668 nm; Invitrogen GmbH, Darmstadt, Germany). The labeled probe was commercially purchased from IBA GmbH, Goettingen, Germany. The NONEUB 338 probe with the same 5' end labeling (5'-CGA CGG AGG GCA TCC TCA-3'), which is complementary to the EUB 338, was used as a negative control. Pure culture of *S. aureus*, resuspended in 20 µl of sterile water and fixed on a glass slide, was used as a positive control.

To facilitate the penetration of the probe for the gram-positive bacteria, the slides were incubated in 25 µl of freshly prepared, pre-warmed lysozyme work solution (20 mg / ml) at 37 °C for 10 min in a humid chamber. Lysozyme work solution contained 2 mM of sodium EDTA, 1.2 % (vol/vol) of Triton X-100 and 20 mM of Tris-HCl (pH = 8.0). After the incubation slides were rinsed in distilled H_2O and air dried.

The glass slides were than exposed to 20 µl of hybridization buffer, containing 100 ng of the oligonucleotide probe and incubated at 46 °C for 36 hs in a humid atmosphere in the dark. Then the slides were covered with cover glasses and put into 50 ml polyethylene tubes (SARSTEDT AG & Co., Nuembrecht, Germany). The humidity conditions were reached by putting pieces of blotting paper into a polyethylene tubes and soaking them with the remaining hybridization buffer.

Table 8: Hybridization buffer.

Stock reagent	**Volume, µl**	**Final concentration**
5M NaCl	360	900 mM
1M Tris-HCl	40	20 mM
10% SDS (added last to avoid precipitation)	2	0.01 %
Formamide	400	20 %
RNase inhibitor	333 of 1:5 concentrate	—
dH_2O	add to 2 ml (870)	—

The stability of hybrids depends on the temperature, the monovalent cation concentration, the percentage of C - G bases and the hybrid size. The melting temperature (T_m) is the temperature at which 50 % of the two DNA strands are denatured [46]. For a hybrid size of less than 200 base pairs, the T_m follows the next equation:

$$T_m = 85.1\,°\text{C} + 16.6 \cdot \log[Na^+] + 0.41(\%\ G - \ C) - 820/L - 0.65 \cdot (\%\ F),$$

where Na^+ is the ionic concentration (in moles per liter) of sodium salt, L — the size

of the probe in base pairs and *F* — the percentage of formamide in the hybiridization buffer. Thus:

$$T_{mEUB338} = 81.5 + 16.6 \cdot \log 5 + (0.41 \cdot 62) - 820\,/\,21 - 0.65 \cdot 20 = 79\,^{\circ}\text{C}.$$

The hybridization experiments are usually performed at (T_m−20 °C), but the adding of formamide allowed us to lower the hybridization temperature to 46 °C.

Table 9: Washing buffer for FISH

Stock reagent	**Volume, µl**	**Final concentration**
5M NaCl	225	225 mM
1M Tris-HCl	1000	20 mM
10 % SDS (added last to avoid precipitation)	50	0.01 %
0.5M EDTA	500	5 mM
dH_2O	add to 50 ml	

After hybridization the slides were quickly transferred into a preheated washing buffer and incubated for 15 min at 48 °C in a water bath (Köttermann GmbH & Co KG, Uetze/ Haenigsen, Germany), rinsed with dH_2O for several seconds, air dried and embedded in Antifade/DAPI (absorption 358 nm, emission 461 nm) mounting medium (AppliChem GmbH, Darmstadt, Germany). The glass slides with the stained samples were examined on a fluorescence research microscope Olympus BX 53® (Olympus America Inc., Melville, New York, USA) using cellSens® software control.

3.5 Statistical analysis

Although the analysis of T-RFLP data has developed considerably over the last decade, there remains a lack of consensus about which statistical analyses offer the best means for finding trends in this data. There are several common statistical approaches described in the literature: principal component analysis (PCA), correspondence analysis, detrended correspondence analysis, and the additive main effects and multiplicative interaction model [42]. T-RFLP results obtained in this study were analyzed using PCA. PCA is a projection method that helps visualize the main information contained in large data tables. PCA involves a mathematical procedure that transforms a number of possibly correlated variables (genera detected in microbial communities of different origins) into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The result is a new set of variables that represent linear combinations of the

original variables that are uncorrelated and reflect the most important structure of the data. Values for each sample projected onto these “loadings” are then calculated and called “scores”. The PCA analysis was performed by J. Pauling (Max-Planck-Institute for Informatics, Saarbruecken).

All cell culture experiments were performed in triplicate. Nonparametric unpaired two-tailed T-test with confidence interval 95 % was carried out to evaluate statistically the obtained data.

4 Results

A total of 29 patients with AECOPD were recruited to study. Ten most common respiratory pathogens were detected by PCR. Both EBSs and sputa were tested on the presence of any of them. To study the bacterial communities more deeply and to determine if there are compositional changes in pulmonary microbiota linked with COPD 18 BAL of adult subjects were studied. This consisted of two groups: 9 healthy individuals and 9 COPD patients.

4.1 Bacterial DNA is present in BAL

As mentioned in Table 4, the average total DNA concentration isolated from BAL samples was 200 ng/µl. The major part of it is composed of nucleic and metahondrial eukaryotic DNA. Therefore, not knowing the exact amount of target bacterial DNA I first optimized the PCR conditions. After 40 cycles of amplification all the samples showed positive results viewed bands in the agarose gel (see Figure 9).

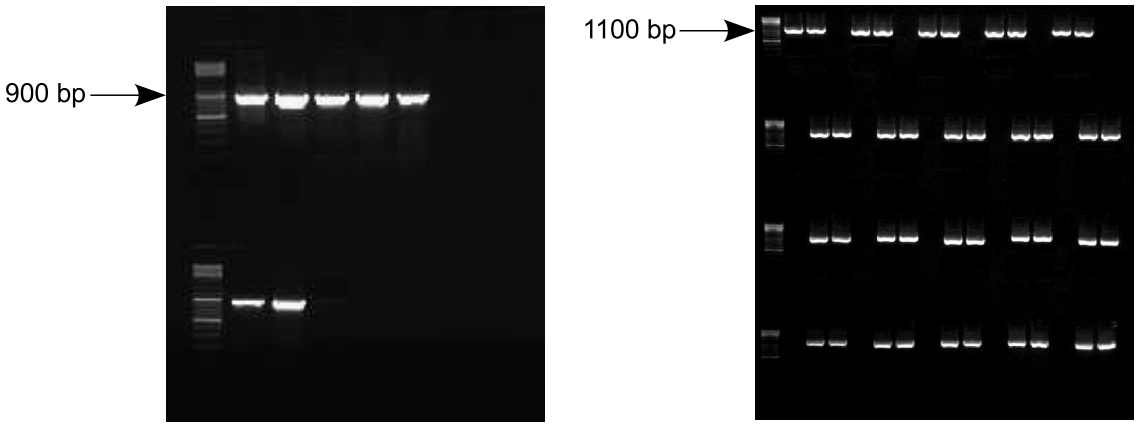

Figure 9: Agarose gel electrophoresis of amplified 16 S rRNA gene fragments. Total DNA extracted from the cell pellets of BAL samples was used as a starting material for the PCR (left). In each ampligication run the total *E. coli* DNA (100 ng) was used as a positive control and 5 µl of Water Molecular Biology Reagent as a negative control. After 40 cycles of amplification distinguishable bands were visible under the ultra-violet light. The amplificates were cut out of the gel, cleaned up and cloned into *E. coli* host cells. To check if the cloning procedure was successful isolated plasmid DNA was amplified with M 13 universal primer set (right). The possible contamination of the plasmid DNA with the genomic DNA of *E. coli* has made the use of primers specific for the 16 S rRNA gene undesirable.

4.2 Pathogens in sputum and exhaled breath condensate

A total of 29 patients with AECOPD were recruited during a four month study period (December 2008 – April 2009). Both exhaled breath condensates and sputum were analysed for each of the 10 pathogens. Patients with ID numbers 1, 2, 4, 7, 8, 13, 16, 19, 20, and 28 received antibiotic treatment (see Figure 12). Antibiotics included quinolones, third-generation cephalosporins, macrolides and β-lactams.

Firstly, I analyzed whether NA can be extracted from EBC. DNA and RNA concentrations were determined in both sample types directly after the isolation procedure. DNA could be detected in all sputa and about 90 % of EBCs (25 probes). RNA could be extracted from 84 % of sputa and from 72 % of EBCs (24 and 21 probes respectively) (Figure 10). Other samples revealed no detectable RNA or DNA. While the amount of RNA was equivalent for sputum and EBC, substantially more DNA could be isolated from sputum.

The next step was to determine whether species-specific PCR can be sucsessfully performed on EBC samples. Different bacteria and viruses were identified in both sample types. About 80 % of the sputa were positive, whereas 50 % of EBCs revealed a positive result. Figure 11 summarizes the results for all detected microorganisms.

One critical issue is whether EBC and sputum samples reveal similar results. Figure 12 displays the distribution of the positive results in individual patients. Surprisingly, the data obtained from both materials did not correlate well. Only one patient had *S. aureus* simultaneously in both samples. *C. pneumoniae* was not found in the studied patient group. Six individuals had negative results in EBC and sputum. Remarkably, influenza AH3 virus presence was always associated with bacterial infection (in most cases *S. aureus*). Detection of *S. pneumonia* was associated with the presence of *S. aureus* in almost all cases. Viruses were identified in sputum samples of patients with antibiotic administration. Some condensates showed primarily the presence of *S. aureus* DNA in these patients. All sputa (except patient with ID 19) of such patients were bacteria-negative.

Co-infection is a common phenomenon in AECOPD. The presence of multiple patho gens was detected in 18 of the 29 patients (62 %) when taking the results of both EBC and sputum analysis into account. Figure 12 shows the distribution of the number of detected pathogens. *M. catarrhalis*, *H. influenza*, and *M. pneumoniae* could be identified only in the presence of other pathogens. Eight subjects revealed both bacteria and viruses in their samples. No correlation was found between the number of individual pathogens and clinical severity, CRP, leukocytes, or FEV_1 (data not shown).

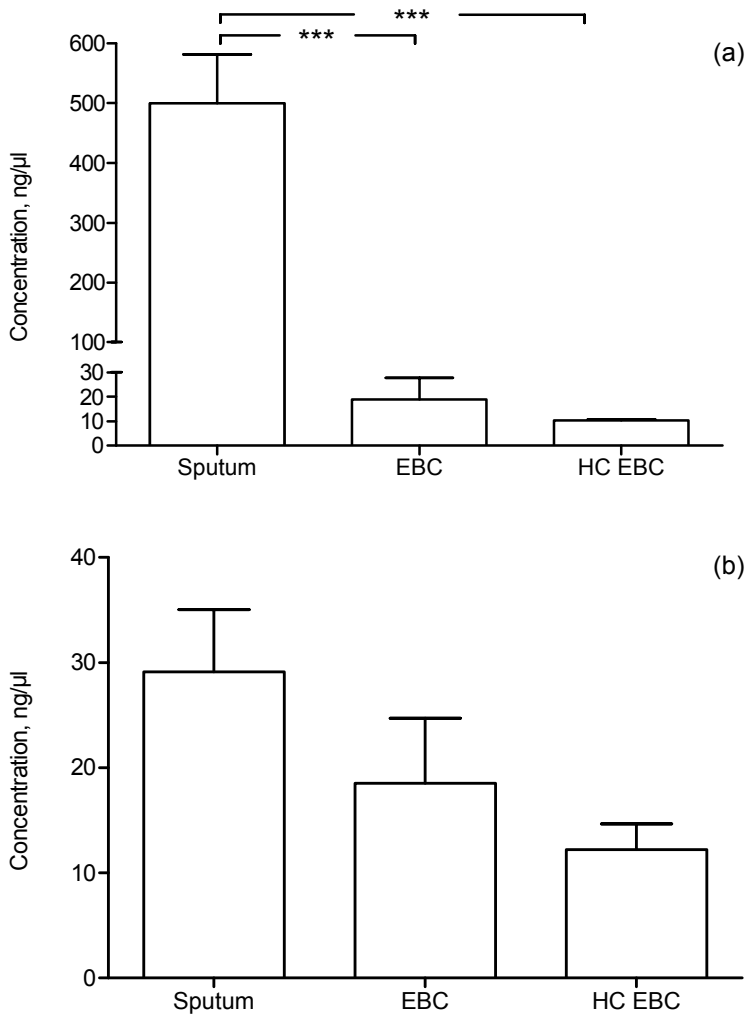

Figure 10: DNA (a) and RNA (b) concentrations in exhaled breath condensate (EBC) and sputum samples. No significant difference was detected between RNA concentrations in the two studied sample types (A). Significantly more DNA was isolated from sputum as compared with EBC (B). Ten control condensates (HC EBC) collected from healthy individuals were included. *** indicates $p < 0.0001$.

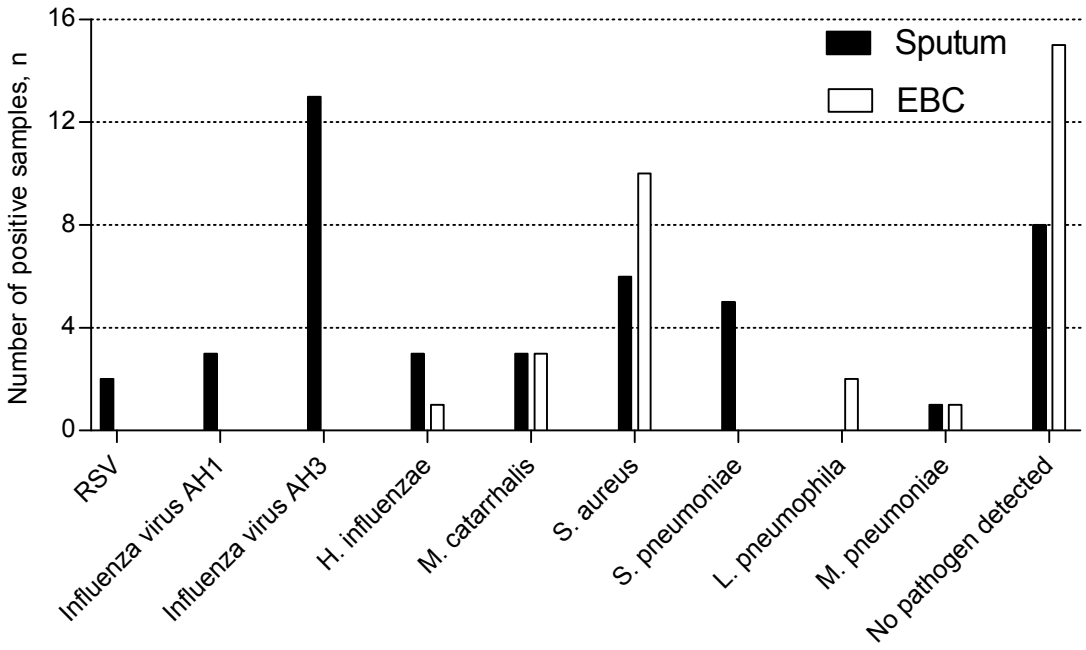

Figure 11: Bacterial and viral nucleic acids can be detected by PCR. PCR was performed on the nucleic acids isolated from sputum (black) and exhaled breath condensate (EBC) (white). Differences between both materials were analysed by χ^2-squared test with Yate's correction: P (RSV) = 0.4718, influenza virus AH1 = 0.2357, influenza virus AH3 = 0.0002, *Haemophilus influenzae* = 0.6043, *Moraxella catarrhalis* = 0.6664, *Staphylococcus aureus* = 0.3781, *Streptococcus pneumoniae* = 0.0613, *Legionella pneumophila* = 0.4718, *Mycoplasma pneumoniae* = 0.4718, and for the samples without pathogen detection = 0.1073.

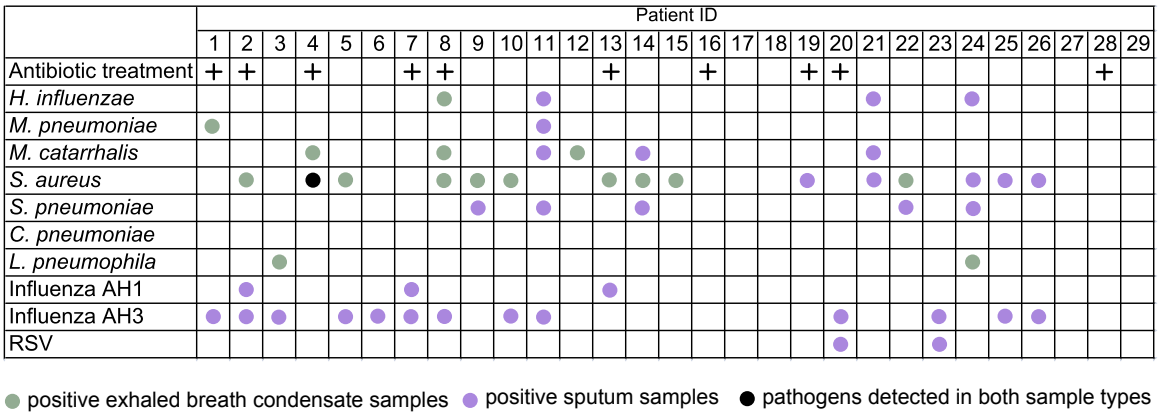

Figure 12: Distribution of positive sputum or EBC results in individual patients. Only *S. aureus* was detected in both sputum and EBC in one patient.

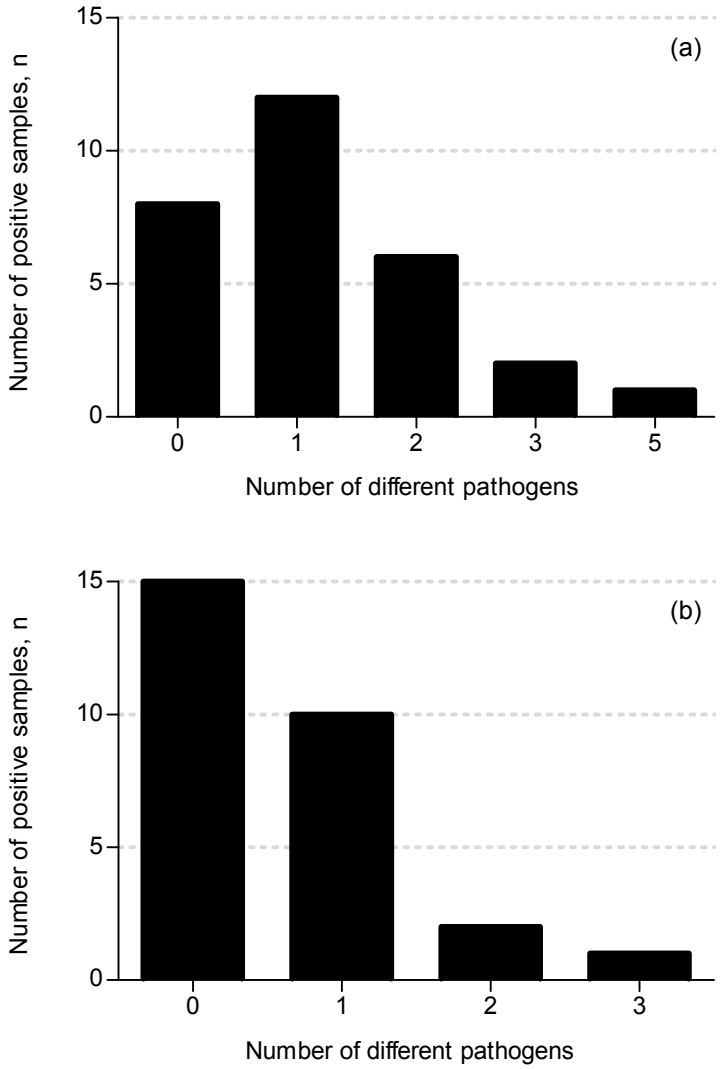

Figure 13: Microbes identified in sputum samples of exacerbated COPD patients using a conventional PCR set with specific primers (a). Exacerbation status is linked to acute infection (viral or bacterial, and in rare cases fungal). Common pulmonary pathogens were detected only in about 72 % of examined samples. Influenza virus AH1, *S. aureus* and *S. pneumoniae* appeared to be the most common exacerbative agents. 31 % of patients had mixed infections with 2, 3, or even 5 pathogens detected simultaneously (b).

COPD patients who have potentially pathogenic microorganisms in their sputum show an exaggerated airway inflammatory response and poorer health status [20].

4.3 T-RFLP is a suitable tool to study pulmonary microbiota

T-RFLP analysis of obtained 18 BAL samples from healthy individuals and COPD patients included the comparison of terminal restriction fragments with RDP database ((R10. U26) 703 222 Good Quality (> 1200) Bacterial). This comparison resulted in

phylogenetic assignment of fragments and was performed using a web-based tool Microbial Community analysis III (MiCA 3) with the following parameters: T-RFLP analysis (PAT+), digest sensitivity — 2 mismatches within 2 bases from 5' of forward primer, window size — forward match ± 1 bp.

Some fragments could not be matched and assigned because of some gaps in the database. If no appropriate match was found in the database the fragment was marked as "\-". Some fragments were identified as "uncultured bacteria". They could also not be phylogenetically assigned.

Fluorescence intensity reflects the relative abundance of particular bacterium in the sample. But it is important to note that abundance is calculated on the basis of 16 S rRNA gene copies. This parameter has high diversity amongst the species and does not correspond necessarily to a number of bacterial cells. All terminal fragments were classified as "of the most frequent occurrence ribotypes" if the relative fluorescence of peaks was > 3 %.

4.3.1 Bacteria in healthy individual BAL samples

As example, Figure 14 illustrates the bacterial community detected in the healthy person with ID number 5. All terminal restriction fragment (TRF) profiles represent the ribotypes of frequent occurrence. *Prevotella, Bacteroides* and *S. umbrinus* NRRL B - 2572T (96 bp, 205 bp and 234 bp respectively) were found more frequently in ID number 5 sample (see Figure 14). Potentially pathogenic genera as *Pseudomonas* and *Mycoplasma* were also present. Two distinct representatives of the genus *Lactobacillus* composed 2.6 % of total microbial community.

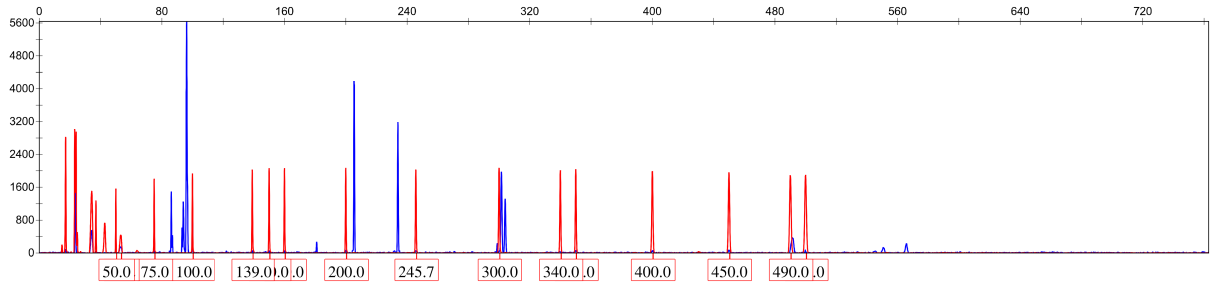

Figure 14: T-RFLP profile obtained from healthy individual with ID number 5. Peaks with size 17 bp, 24 bp and 34 bp were excluded from the analysis. Species richness - 15. \- (53 bp) - affiliation of the fragment could not be deduced.

In ID the number 6 sample (see Figure 15) the genera *Prevotella* and *Veillonella* dominated. *Prevotella* was one of the most common bacteria also in ID number 7 and 9 samples. Interestingly in this sample uncultured *Legionella* sp. (498 bp) was also found.

It composed almost 2 % of the community. Clones of *Actinomyces* and *Megasphaera* appended the species richness. *S. pneumoniae* (552 bp) with 9 % of relative abundance was the third major ribotype.

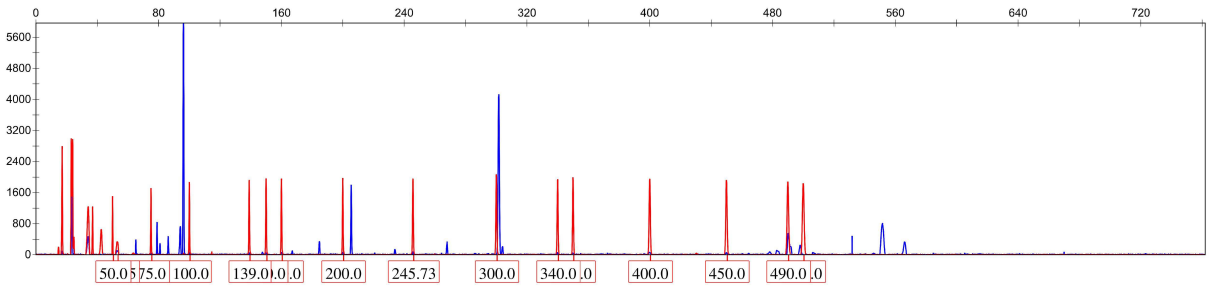

Figure 15: T-RFLP profile obtained from healthy individual with ID number 6. Peaks with size 17 bp, 24 bp and 34 bp were excluded from the analysis. Species richness - 19. \- (53 bp) - affiliation of the fragment could not be deduced.

In Table 10 I summarized T-RFLP analysis results obtained in the healthy individual group. They illustrate bacterial genera distribution between different individuals. No common pattern of species distribution could be defined. All studied samples showed highly heterogeneous patterns. Some bacterial genera, such as *Bacteroidetes*, *Lactobacillus*, *Megasphaera*, *Prevotella* and *Streptomyces* were most common and appeared in two thirds of studied samples. In contrast, *Actinomyces*, *Bacillus*, *Corynebacterium*, *Lactococcus*, *Micrococcus*, *Peptostreptoocccus* and some other genera could be found only in certain individuals. Interestingly, the presence of genus *Lactobacillus* completely correlated with distribution of *Megasphaera* between the individuals and partly with occurrence of *Mezorhizobium*, *Pseudomonas* and *Streptomyces*. But representatives of *Lachnospiraceae* in samples with ID numbers 1 and 2 (see Table 10) were found simultaneously with genus *Xanthomonas*.

Table 10: Terminal restriction profiles in healthy individual samples

Potential representative	**ID1**	**ID2**	**ID3**	**ID4**	**ID5**	**ID6**	**ID7**	**ID8**	**ID9**
\- (no match was found)		+		+	+	+	+		
Acinetobacter				+		+			+
Actinomyces						+			
Bacillus						+			
Bacteroidetes	+	+		+	+	+	+	+	+
Clostridiales	+	+		+	+		+		
Corynebacterium								+	
Firmicutes oral clone	+		+					+	
Flavobacterium			+	+					
Fusobacterium		+		+		+			
Lachnospiraceae	+	+							
Lactobacillus			+	+	+	+	+		+
Lactococcus			+						
Legionella				+		+			
Megasphaera			+	+	+	+	+		+
Mesorhizobium			+	+	+		+		+
Microbacteriaceae								+	
Micrococcus			+						
Mycoplasma			+	+	+			+	
Neisseria				+	+		+		
Peptostreptococcus			+						
Porphyromonadaceae			+		+				
Prevotella				+	+	+	+	+	+
Pseudomonas			+		+	+	+		+
Ralstonia						+	+		+
Rhizobium			+						
Selenomonas				+					
Sphingomonadales	+	+		+		+		+	+
Staphylococcus				+					
Stenotrophomonas		+							
Streptococcus	+	+	+			+		+	
Streptomyces	+	+		+	+	+	+		+
uncultured bacteria	+	+	+	+				+	+
Veillonellaceae					+	+	+		+
Xanthomonas	+	+							

4.3.2 Molecular fingerprint analysis of COPD patient BAL samples

Figure 16 shows relative abundance of ribotypes detected in COPD patient with ID number 1. *X. campestris* EGS09 (411 bp) and *Sphingomonas* sp. SaS3 (436 bp) dominated in this patient making up 17 % and 19 % of total community respectively. Oral clones of *Actinomyces* and some *Firmicutes* bacterium were also presented in total relative abun-

dance of about 14 %. Potential pathogen *C. glutamicum* ATCC 27021 made up 3.5 % of microbial community.

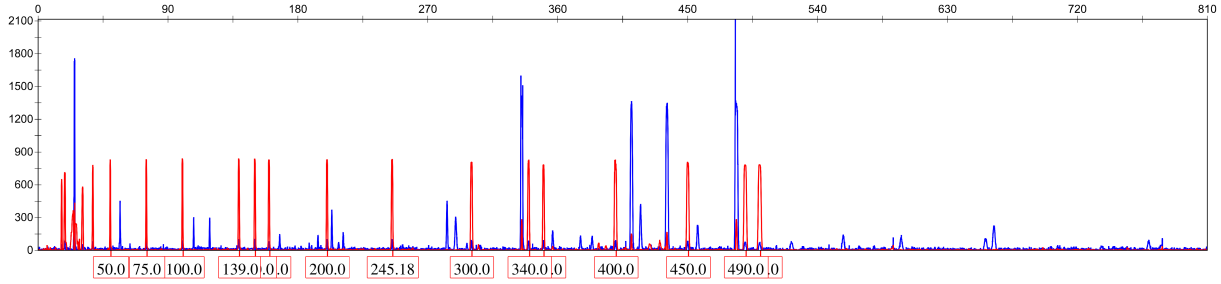

Figure 16: T-RFLP profile obtained from COPD patient with ID number 1. Species richness - 17. \- (357 bp) - affiliation of the fragment could not be deduced.

Low biodiversity was characteristic for the sample of COPD patient with ID number 4 (see Figure 17). The most dominant ribotype appeared to be *Sphingobacterium* sp. Bmc112 (550 bp) with relative abundance of 68 %. That was also significant for that sample as only a single pathogenic genus *Pseudomonas* (482 bp) was detected. *Prevotella* (96 bp), *Lactobacillus* (179 bp) and *Clostridiales* (204 bp) ribotypes were rare (< 1 %).

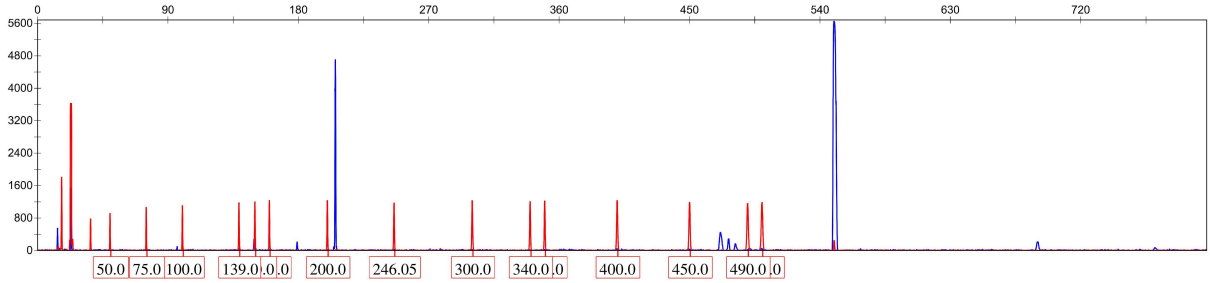

Figure 17: T-RFLP profile obtained from COPD patient with ID number 4. Peaks with size 14 bp, 23 bp and 690 bp were excluded from the analysis. Species richness - 9.

Some ribotypes (*Bacteroidales*, *Clostridiaceae*, *Lactobacillus*, and *Prevotella*) were present in almost all studied samples, while *Actinomyces*, *Alkaliphilus*, *Gemella*, *Legionella*, *Neisseria* sp., *Peptostreptococcus*, *Selenomonas*, *Streptobacillus,* and *Treponema* detection characterized only particular patients (Table 11). No pattern, which could be linked to the COPD state, was found. But presence of *Lactobacillus* correlated mostly with detection of *Bacteroidales*. Genera *Mycoplasma* (6 positive samples), *Pseudomonas* (5 positive samples) and *Staphylococcus* (5 positive samples) as potential pathogens seem to play the major role in the persistent chronic infection in COPD patients.

In order to compare the community structure, the species richness and Shannon diversity index were calculated for two sample groups. No difference between COPD patients and healthy individuals could be seen (Figure 18).

Thus, T-RFLP analysis showed highly diverse composition of pulmonary microbial communities in healthy individuals as well as in COPD patients.

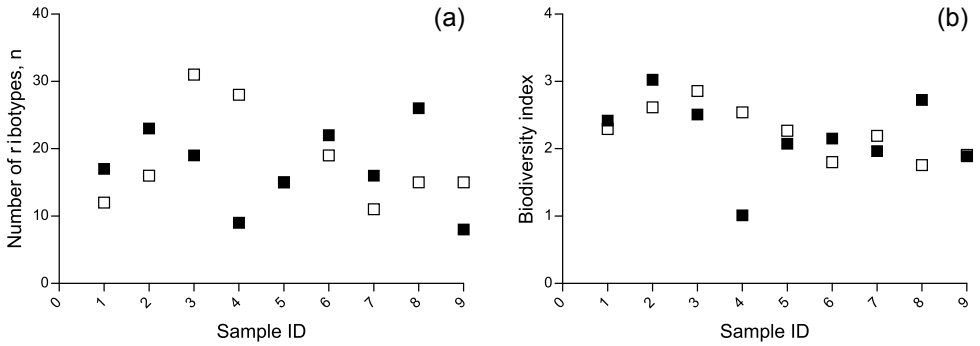

Figure 18: Species richness (a) and diversity index (b). Terminal restriction fragments shorter than 50 bp were excluded from the analysis. Empty squares correspond to healthy individuals, filled squares — to COPD patients.

Table 11: Terminal restriction profiles in COPD patient samples

Potential representative	**ID1**	**ID2**	**ID3**	**ID4**	**ID5**	**ID6**	**ID7**	**ID8**	**ID9**
\- (no match was found)		+	+					+	
Acinetobacter							+	+	
Actinomyces	+								
Alkaliphilus		+							
Bacillus	+		+		+	+	+	+	
Bacteroidales		+	+	+	+	+	+	+	+
Clostridiaceae	+	+	+	+	+	+		+	
Corynebacterium	+		+		+	+			
Firmicutes oral clone	+	+		+			+		+
Flavobacterium						+	+		
Fusobacterium	+	+	+			+		+	
Gemella								+	
Lachnospiraceae	+	+							
Lactobacillus		+	+	+	+	+	+	+	
Legionella							+		
Lysinibacillus	+	+							
Megasphaera						+	+	+	
Mycobacterium								+	
Mycoplasma			+		+	+	+	+	+
Neisseria sp.								+	
Peptostreptococcus								+	
Porphyromonadaceae			+				+		
Prevotella			+	+	+	+	+	+	+
Pseudomonas			+	+		+		+	+
Ralstonia						+		+	
Selenomonas								+	
Sphingomonas	+	+		+		+			
Staphylococcus			+			+	+	+	+
Streptobacillus								+	
Streptococcus	+	+					+	+	
Streptomyces		+				+		+	
Treponema						+			
uncultured bacteria	+	+	+	+	+	+	+	+	+
Veillonella			+		+	+		+	
Xanthomonas	+	+							

4.4 Sequencing data

Altogether about 800 sequences were included in the analyses. Gaps and regions of alignments for which homology of residues could not be reasonably assumed were excluded. An average of at least 880 nucleotides were included in the phylogenetic

analysis of eubacterial clones.

Most of the respiratory bacterial sequences had a relatively high level of similarity with their closest counterparts found in public databases (the mean similarity of 96.3 % with the cultured bacteria). The phylogenetic relationships of bacterial sequences were analyzed with 151 phylotypes and closely related reference sequences obtained from the GenBank database.

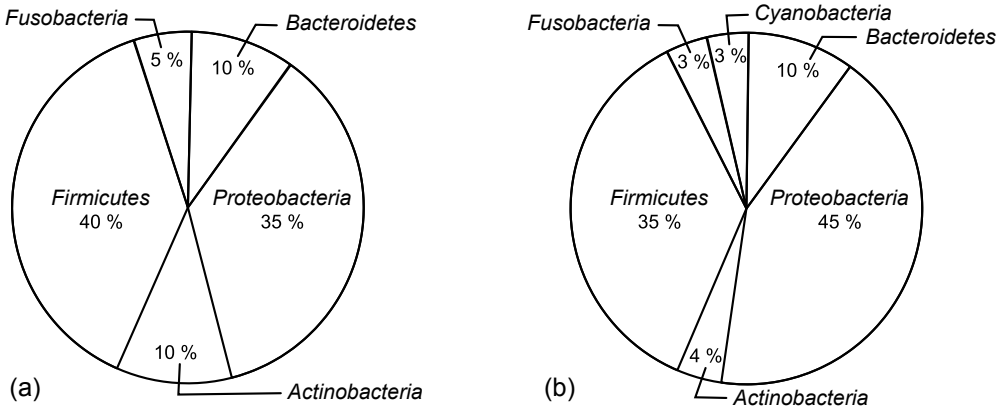

Figure 19: Relative distribution of identified phyla in healthy individuals (A) and COPD patients (B).

Most of the sequences matched in the database with phyla *Firmicutes* and *Proteobacteria* (Figure 19), which composed 40 and 35 % in healthy individuals, and 35 and 45 % — in COPD patients respectively. Other bacteria belonged to *Fusobacteria*, *Bacteroidetes*, and *Actinobacteria*. The phylum *Cyanobacteria* was characteristic only for the COPD patient group.

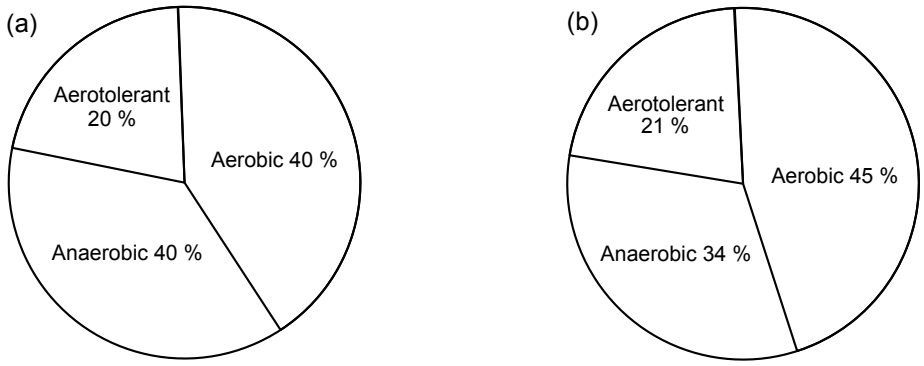

Figure 20: Relation to the oxygen. (a) — healthy individuals, (b) — COPD patients.

Interestingly, anaerobic representative of standard microbiota made up 40 % of the total community in the healthy individuals, and about 34 % in the COPD patients (see Figure 20). Almost 20 % of identified bacteria could tolerate the oxygen in both sample types.

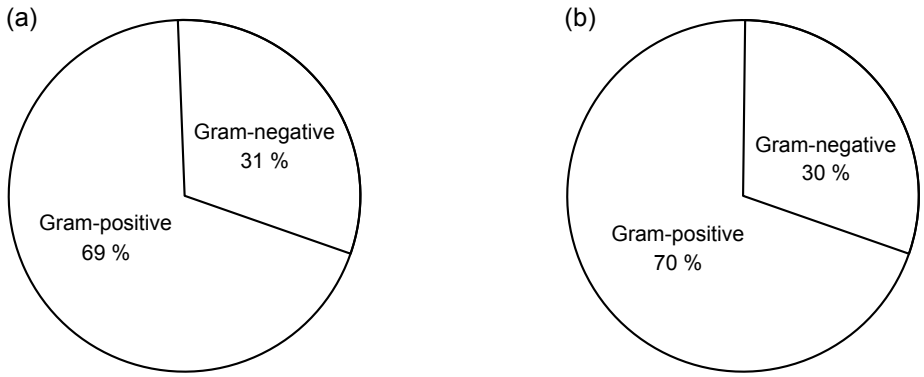

Figure 21: Cell wall composition. (a) — healthy individuals, (b) — COPD patients.

Gram-negative bacteria cause infections in the respiratory tract most often. Our results showed that one third of bacterial representatives had such a type of cell wall (Figure 21).

4.4.1 Phylogenetic tree evaluation

The phylogenetic trees were constructed using the clustering neighbor-joining method. After construction, the reliability of trees was estimated by nonparametric bootstrapping. Bootstrapping is a statistical technique that tests the sampling errors by repeatedly sampling trees through slightly perturbed datasets [183]. By doing so, the robustness of the original tree could be assessed. All the bootstrapped trees are summarized into a consensus tree based on a majority rule. It is considered, that a bootstrap value of 70 % approximately corresponds to 95 % statistical confidence.

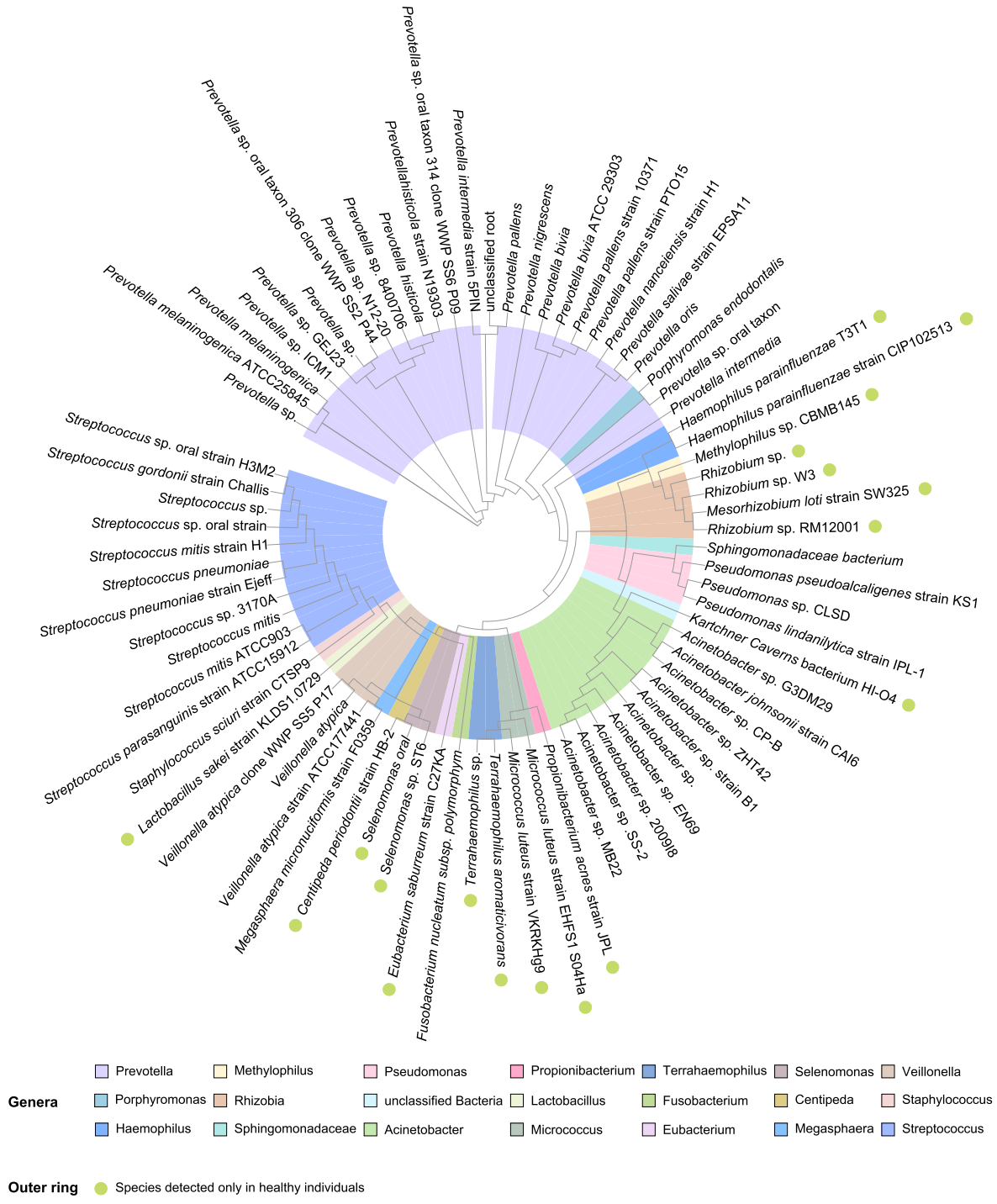

Figure 22: Phylogenetic tree ilustrating the standard microbiota present in healthy lungs. Different genera are denoted with different colors with one exception: the family *Sphingomonadaceae* represents the taxon of higher rank. Bacteria marked with green circles were detected only in healthy individuals, but not in COPD patients.

This dendrogram (see Figure 22) represents the phylogenetic relationships of analyzed bacterial sequences and therefore shows that genera *Prevotella*, *Streptococcus* and *Acinetobacter* were more diverse, representing multiple species and strains. Despite the

fact that *Rhizobium radiobacter* infection has been associated with long-term indwelling catheters, and several cases have also been reported in immunocompromised individuals, three representatives of this genus were detected in healthy individuals only. This genus was present also in respiratory secretions of CF patients [39]. There are no known interactions of the nearest relative of *Rhizobium* sp. *Mesorhizobium loti* with humans or animals. But in our study it was found that nitrogen-fixing *M. loti* was characteristic for the healthy state. *H. parainfluenzae*, *Kartchner Caverns* bacterium, *P. acnes*, *M. luteus*, *Terrahaemophilus* sp., *E. saburreum*, *Selenomonas* sp., *S. periodontii*, and *L. sakei* made up 24 % of the total microbial community and were also present only in healthy individual group samples. Nonsporeforming, Gram-positive, anaerobic bacterium *P. acnes* is found all over the body and generally produce lactic acid, propionic acid, and acetic acid from glucose. Similarly, *M. luteus* has been previously isolated from human skin. The defining metabolical characteristics of *Micrococcus* are the abilities to aerobically produce acid from glucose glycerol, aesculin hydrolysis, arginine dihydrolase, and conversion of nitrate to nitrite.

One difference between T-RFLP analysis and sequencing results cannot be ignored: lactic acid bacteria of genus *Lactobacillus* was found in both sample types using the T-RFLP approach, whereas its detection was successful only in healthy volunteers samples when the clone sequencing was applied.

A second distance-based neighbor-joining tree was constructed with 76 phylotypes (Figure 23). 29 distinct phylotypes were detected in the studied cohort of patients. Genera *Streptococcus*, *Veillonella* and *Prevotella* were most common and had the highest diversity. The diagram depicts that there were more than 20 unique ribotypes (composing 36.8 % of total community) characteristic of the COPD state. For example, *Gemella* sp. — Gram-positive, facultatively anaerobic bacteria, which have been previously found to be involved in pulmonary exacerbations of cystic fibrosis patients.

A common microbial community of 9 bacterial genera was identified in all samples analyzed. This core group included members of the *Prevotella*, *Sphingomonadaceae*, *Pseudomonas*, *Acinetobacter*, *Fusobacterium*, *Megasphaera*, *Veillonella*, *Staphylococcus*, and *Streptococcus*. The sequencing data was also examined for the presence of the atypical bacteria, *M. pneumoniae*, *L. pneumophila*, and *C. pneumoniae*, which are associated with COPD exacerbations. None were detected by this method, although *Mycoplasma* and *Legionella* related species were identified in several samples using the T-RFLP approach.

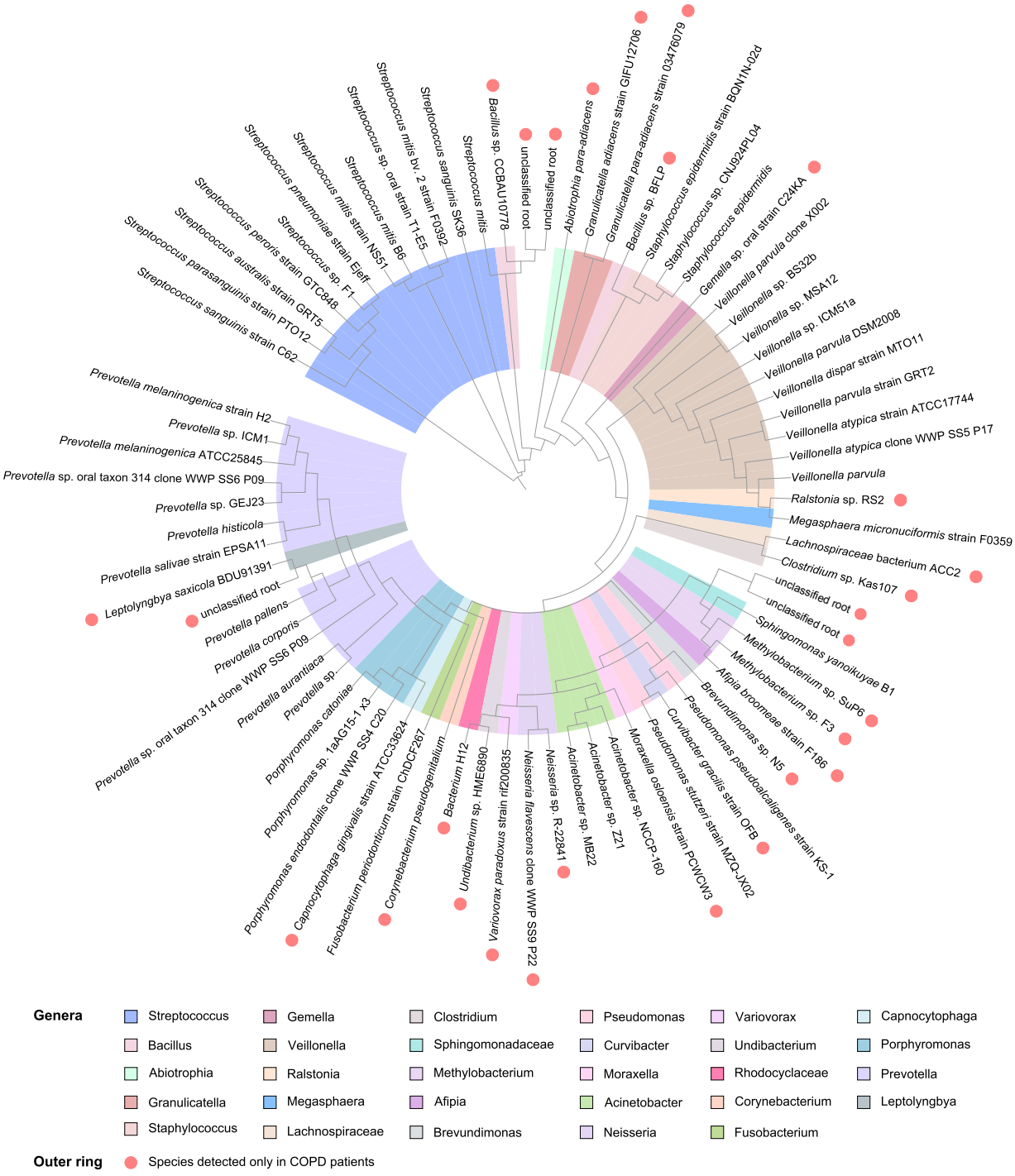

Figure 23: Phylogenetic tree ilustrating microbial community of lungs associated with COPD in stable state. Different genera are denoted with different colors with two exceptions: families *Rhodocyclaceae* and *Lachnospiraceae* represent the taxa of higher rank. Bacteria marked with red circles were detected only in COPD patients, whereas other microbes are the part of a core lung microbiome.

4.5 Principal component analysis of detected microbial communities

It has become apparent in recent years, that the adaptation of highly variative statistical tools from core data-analytic disciplines to microbial community analysis is needed. The major focus of research in this area is on accurate discrimination and classification of bacteria, and on understanding and describing the differences between communities.

PCA was performed with WEKA 3 software [70] with the correlation (standardized) option that first centers the TRFs and than divides each matrix entry by the standard deviation for each TRF, thus producing a correlation matrix. Firstly, all identified species were arranged to the genus level and organized in compliance with subject ID number and relative abundance (in %) in the community. Then four different classifiers (Naive Bayes, Multilayer perceptron, Support Vector Machine, and Random Forest) were applied in order to calculate the principal components of each class and design the classifier according to the projection of the data on the subspaces spanned by these principal components, corresponding to different classes (data not shown). The multilayer perceptron classifier delivered the best results, separating all data sets into two distinct groups with two exceptions (two healthy individual samples were identified as COPD positive). This classifier consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one.

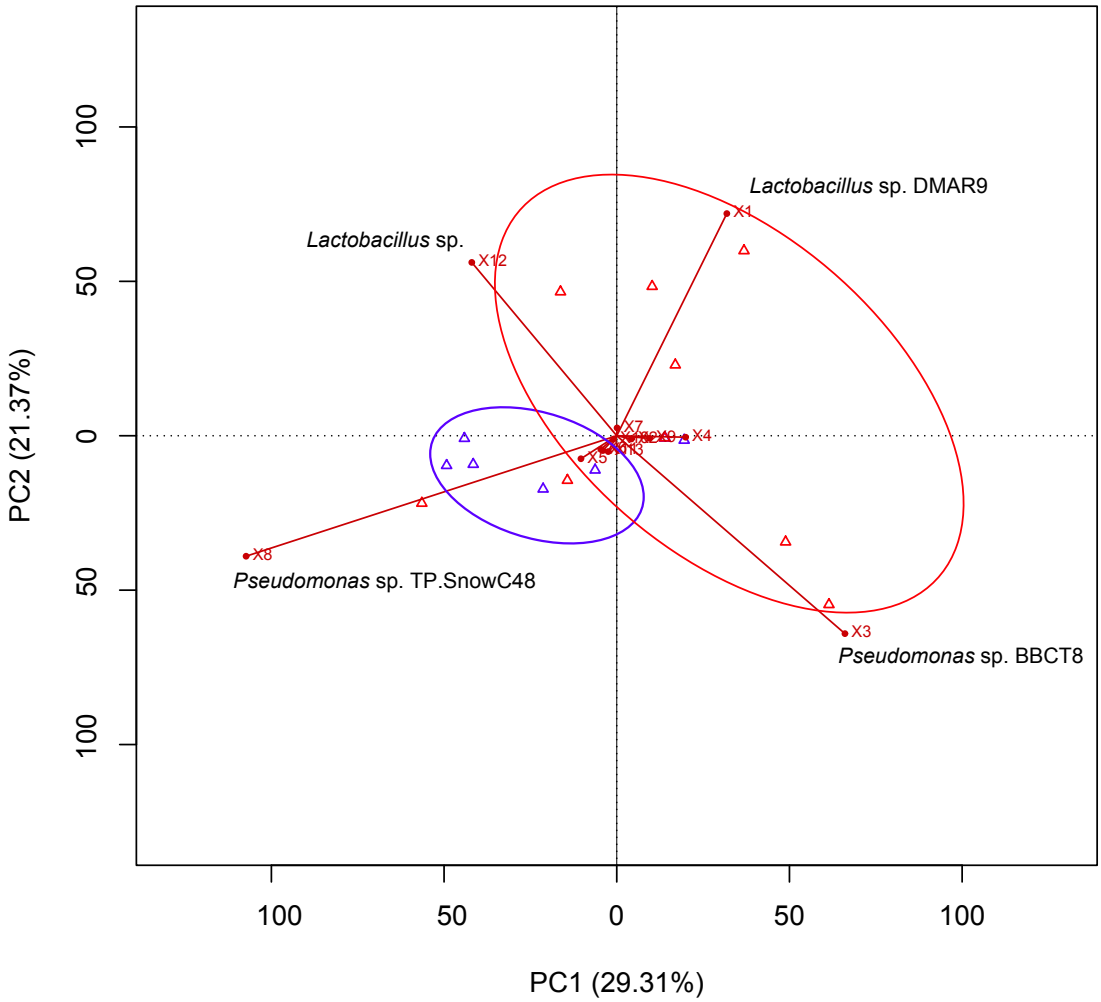

Figure 24: Biplot of the principal component analysis of the bacterial communities from bronchoalveolar lavages of the COPD patients and healthy individuals

Classification optimization yielded the results as shown in Table 12. Following phylotypes were used to build the classifier model based on our data set consisting of 9 T-RFLP analyzed BAL samples taken from healthy individuals and COPD patients respectively: *Neisseria*, *Corynebacterium*, *Staphylococcus*, *Bacillus*, *Mesorhizobium*, *Flavobacterium*, *Lysinibacillus*, *Lachnospiraceae*, *Streptomyces*, *Acinetobacter*, *Megasphaera*, *Pseudomonas*, *Clostridium*, *Sphingomonadales,* and *Clostridiaceae*.

Table 12: Classification results using a classifier and cross validation based on optimization

TP Rate	TP Rate	Precision	F-Measure	Class	Confusion Matrix
0.778	0	1	0.875	Healthy	a b <– classified as
					7 2 | a = healthy
					0 9 | b = COPD
1	0.222	0.818	0.9	COPD	

Out of 18 instances 16 (88.9 %) were correctly classified. With the application of Simple K-Means clustering we could further reduce the phylotypes retaining *Bacillus*, *Mesorhizobium*, *Flavobacterium*, *Streptomyces*, *Megasphaera* and *Clostridiaceae*. Given the TP and FP rates this genera combination is reliable for the COPD group. *Mesorhizobium* was only detected in samples of healthy individuals. Classification solely based on a prior feature selection via PCA did not create a good learning model as shown in Table 13.

Table 13: Classification results using a NaiveBayes classifier and cross validation based on prior feature selection

TP Rate	TP Rate	Precision	F-Measure	Class	Confusion Matrix
0.333	0.333	0.5	0.4	Healthy	a b <– classified as
					3 6 | a = healthy
					3 6 | b = COPD
0.667	0.667	0.5	0.571	COPD	

Different variance proportion cutoffs did not significantly change the results. Regarding *Pseudomonas* and *Lactobacillus* we used our PCA results to create a biplot using the R library (Figure 24).

4.6 *In situ* detection of bacteria

There is a core bacterial community in the lower region of the respiratory tract. But, it is important to emphasize that the methods which were used in the present study could deliver the information only about bacterial DNA present in BAL samples. The question remains whether do these DNA fragments correspond to the leaving bacterial cells colonizing the lung epithelium?

To clarify this question the FISH was preformed with universal eubacterial oligonucleotide probe EUB 338 on mice lung tissue sections. Figure 25 shows *S. aureus* cells hybridized with EUB 338. There is a clearly defined cocci-shaped pattern to the graph, and this can be taken to mean that whole cells seem to be stained.

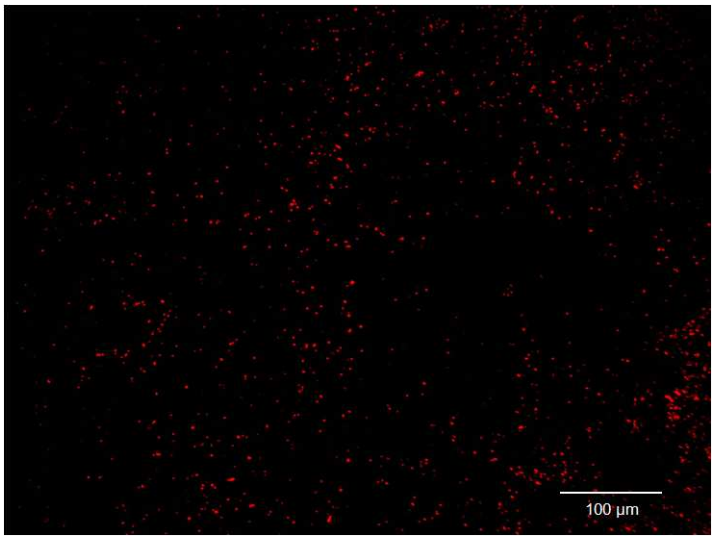

Figure 25: Positive control for FISH. Pure 24 h culture of *S. aureus* grown in nutrient broth was heat fixed and hybridized with the all-bacterium-specific EUB 338 probe. Spherical cells (red) were clearly to distinguishable in 10 examined light fields of each sample. All hybridizations were performed at 46 °C and wash step – at 48 °C. The glass slides with the stained samples were examined on a fluorescence research microscope Olympus BX 53® (Olympus America Inc., Melville, New York, USA) using cellSens® software control.

Specificity tests under stringent FISH conditions using planktonically cultivated cells showed that the probe displayed the anticipated specificity. Cell permeabilization of paraformaldehyde-fixed samples by exposure to lysozyme proved to be absolutely necessary. Without lysozyme pretreatment the Gram-positive bacteria showed no fluorescence (data not shown). Another critical point appeared to be the RNase inhibitor. Its presence in the hybridization buffer was essential for getting a strong fluorescent signal.

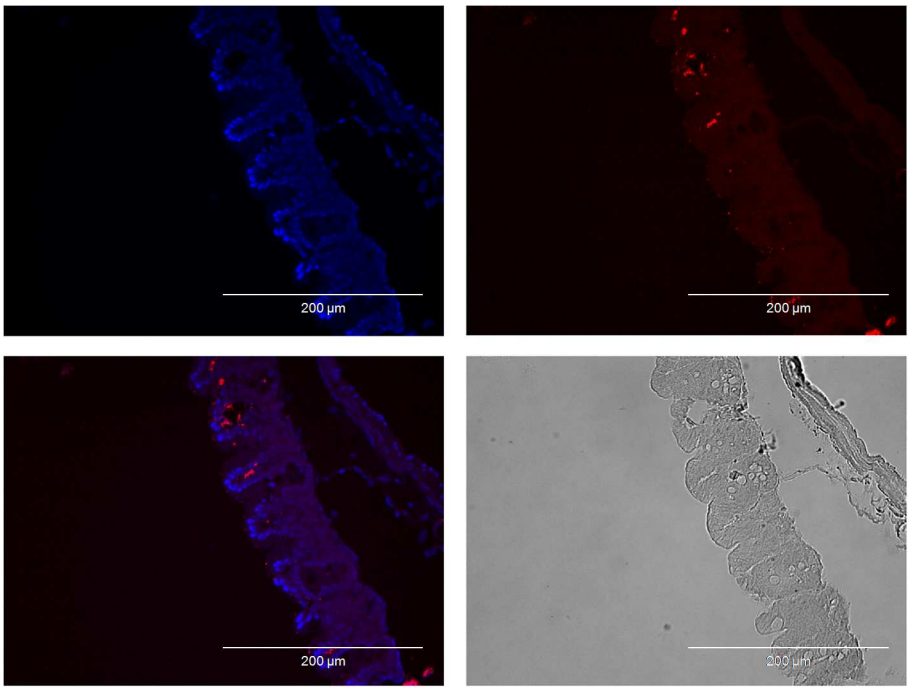

Figure 26: Overview of the microbial colonization pattern showing single bacteria localized under or on the epithelial layer. Some bacteria are seen as single cells but most bacteria occur in clusters. Representative data from the 10 paraffin-embedded mice trachea tissue sections (1 µm in thickness) of healthy mice are shown. Slides were dewaxed and hybridized with the all-bacterium-specific EUB 338 probe. Red (Alexa 647) represents bacteria and blue (DAPI) represents tracheal cells. All hybridizations were performed at 46 °C and wash step — at 48 °C. The glass slides with the stained samples were examined on a fluorescence research microscope Olympus BX 53® (Olympus America Inc., Melville, New York, USA) using cellSens® software control.

Figure 26 depicts that there were more bacteria under the epithelial layer than in the aeriferous space.

5 Discussion

The central aim that motivates this research is: to determine the compositional changes in pulmonary microbial communities linked with COPD and to extend the knowledge about the core lung microbiome. The evidence from this study suggests that a variety of factors relate to stable bacterial colonization in the lower respiratory tract and also to significant differences observed in COPD patients compared with healthy individuals.

5.1 Microorganisms can be detected in patients with AECOPD

Infections have an important role in the pathogenesis of COPD and AECOPD [151]. The identification of microorganisms in AECOPD is important to characterize the role of specific microorganisms in the disease process and potentially to develop antimicrobial strategies. Infection and colonization are strongly associated with stable, chronic COPD [150]. The role of microbes in AECOPD is complex. The acquisition of new bacterial strains from the environment appears to be one major factor in the initiation of exacerbations [149, 12]. Classically, bacteria have been identified by culture, viruses by culture, serology or immunofluorescence. In the past twenty years, NAT-based techniques have been developed to detect viruses and bacteria in respiratory secretions. These methods are very sensitive, however, they are limited due to relatively high costs, the need for specialized equipment, and problems with false-positive and -negative results [35]. PCR-based NAT has been applied to detect atypical bacterial [52] and viral pathogens [81] in AECOPD.

Sputum has routinely been used to determine the level of inflammatory markers and to cultivate microorganisms in patients with AECOPD. The use of spontaneous or induced sputum is limited by the difficulties of expectoration, low level of standardization and contamination by the upper airways or saliva. Bronchoalveolar lavage or protected specimen brush are invasive and generally not well tolerated by patients with AECOPD [170]. EBC could potentially offer an opportunity to analyse the microbiom of the lung. The use of EBC to monitor airway inflammation is based on the hypothesis that it contains aerosolized particles from the airway lining fluid. It is known that such factors as collection device, pattern of breathing, nasal and oral contamination can influence the obtained results [27]. Data that is not normalized cannot be compared from patient to patient and from study to study; this makes analysis of exhaled breath condensate less attractive [130]. A high degree of within-subject variability has been demonstrated for such EBC markers as pH value and leukotriene B4 in COPD patients [93]. The pH of EBC is significantly more variable in these patients than in healthy subjects. Although the detection of bacterial DNA in EBC was reported first in 1999, over the last ten years there have been only a limited number of publications confirm-

ing this statement. The majority of them present the negative results concerning the bacterial and viral PCR-based detection in cystic fibrosis, pulmonary tuberculosis and influenza patients [84, 59, 176]. The authors speculated about the possible reasons for this and concluded that one major problem might be decreased recovery of bacterial and viral nucleic acids in EBCs. *Mycobacterium tuberculosis* and perhaps other bacteria are too large to be transported as nonvolatile particles in EBC [84]. EBC has been used as the starting material for NAT-based techniques. There is only a limited amount of data on the detection of microorganisms based on NAT from EBC available [57, 84]. The current study shows that DNA and RNA can be detected in EBC [185]. We used the conventional PCR to identify bacteria (not the real-time quantitative PCR) since the dramatic differences in total DNA concentrations in EBC and sputum would not allow the comparison of the differences in obtained data. Moreover, change in bacterial load is unlikely to be an important mechanism for AECOPD [152]. Interestingly, a comparison with the results from sputum showed that the detection patterns did not correlate well. Several reasons could contribute to the divergent results: 1) The sources of sputum and EBC could be different and thus different microorganisms could have been sampled. The origin of EBC is not known, although the evidence suggests that it resembles not only the bronchial part of the lower respiratory tract but also the alveolar part. A recent comparison between EBCs and bronchoalveolar lavages concluded that the anatomical origin of the material is difficult to be determined and that the results of both methods were hardly comparable [83]. 2) The location of microbial components (i.e., the nucleic acids) might determine how easily they are included in the sample. While microorganisms are often located within sputum, an aerosilization process is necessary for detection in EBC. Many other factors could contribute here, including volatility, solubility, or charge of the molecule. 3) NA of specific microorganisms might be unstable in sputum or EBC. 4) EBC might be insensitive to lung pathogens and the positive results might reflect false-positive detection of bacterial of the standard microflora from the upper airways. *H. influenzae*, *S. aureus* and *M. catarrhalis* are found in the upper respiratory tract of healthy individuals.

The current study shows that bacteria and viruses are present in the airways of patients with AECOPD. It is well documented that exacerbation is associated with microorganisms, either with increased microbial load [179, 164], or with the acquisition of new bacterial strains [151]. In agreement with known data, we detected viruses and bacteria during AECOPD. One finding as a result of this data is that stable COPD patients are often also culture- or PCR-positive for these microorganisms [134, 150]. Our study revealed that many patients with AECOPD show the presence of NA from more than one microbial species indicative of co-infection. This is an interesting phenomenon because it is known that bacteria and viruses interact in the induction of an inflamma-

tory response [179, 127]. On the other hand ZETTERQUIST et al. have shown that EBC nitrite mainly originates in the pharyngo-oral tract and its increase in cystic fibrosis is possibly explained by a regional change in bacterial activity [186]. Identification of nitrate reducing bacteria in EBC will prove this hypothesis. Thus, the detection of bacterial DNA in samples should be taken into account by interpretation of results due to other biomarkers measured in EBC. EBC cannot be used to detect the viral pulmonary pathogens. Our data supports the results obtained from influenza patients [59].

In conclusion, the current study demonstrates that EBCs can be used to determine the presence of microbial NA in patients with AECOPD. This detection likely reflects the presence of the corresponding microorganism. Interestingly, the results from EBC and sputum do not correlate well. The study also reiterates that microbes have an important role in the induction of AECOPD.

5.2 Polymicrobial communities in the respiratory tract

It is generally agreed today that there is a core bacterial microbiome in the lower respiratory tract [56, 79, 75]. It demonstrates high individual diversity and detected species are often considered to be also the part of oral cavity standard microbiota. Species common to all sites in the mouth belong to the genera *Gemella*, *Granulicatella*, *Streptococcus*, and *Veillonella* [2]. Against this background, the central question that motivated this study is: whether the bacteria from the bronchial tree are present in these lower regions of the respiratory tract or whether the data represents contamination during the sampling procedure.

Bronchoscopy is a procedure performed with an instrument called the bronchoscope in order to inspect the airways of the lungs. The domain of interest for the bronchoscopist is that area of the respiratory tract found below the vocal cords. The bronchoscope is inserted into the airways, usually through the nose or mouth, or occasionally through a tracheostomy. This allows the practitioner to examine the patient's airways for abnormalities such as foreign bodies, bleeding, tumors, or inflammation. Specimens may be taken from inside the lungs. The construction of bronchoscopes ranges from rigid metal tubes with attached lighting devices to flexible optical fiber instruments with real-time video equipment.

In the present clinical trial the flexible fiberoptic bronchoscope was used. The flexible bronchoscope is longer and thinner than a rigid bronchoscope. It contains a fiberoptic system or photochip that transmits an image from the tip of the instrument to an eyepiece or video camera at the opposite end. The flexible bronchoscopy causes less discomfort for the patient and the procedure can be performed easily and safely under moderate sedation.

The oral cavity and upper airways the surface of the bronchoscope is covered with a variety of bacterial cells. In the present study, all the common oral genera were found in the COPD patient group (see Figure 23), but in healthy individuals their presence was limited only by two genera: *Veillonella* and *Streptococcus* (see Figure 22). On the other hand, according to Tables 10 and 11, the T-RFLP analysis could detect three oral genera in COPD patients (*Veillonella*, *Gemella,* and *Streptococcus*) and two genera (*Veillonella* and *Streptococcus*) in the healthy volunteer group. Several explanations for this fact have been offered. Detected oral microbial representatives made up not more than 9 % of the total bacterial community. Even if the contamination occurred, all the other ribotypes which were found actually referred to lower airway microbiome. Moreover, there is no strong evidence that oral bacteria cannot persist in alveolar regions of lungs as well. In the case of the above mentioned genera, the question of their including or excluding into the core pulmonary microbiome needs further investigation. The genus *Terrahaemophilus* was first detected in respiratory samples in a study focused on oral and respiratory bacterial species associated with ventilator-associated pneumonia [19]. But the authors could not provide any body of evidence, that these bacteria play a role in the pathogenesis of disease. In contrast to contamination hypothesis, the sequencing data shows that oral bacteria such as *E. saburreum* and *S. periodontii* typically could not be detected in COPD patient BAL samples. COENYE et al. in 2002 demonstrated the genus *Ralstonia* to chronically colonize CF patients by using genotyping of serial isolates recovered from the same patient [40]. In our study this genus was present only in COPD patients. Other genera included *Afipia*, *Brevundimonas*, *Curvibacter*, *Moraxella*, *Neisseria*, *Undibacterium*, *Corynebacterium*, *Capnocytophaga*, and *Leptolyngbia*.

Despite the fact that NAATs have made a major contribution to the improved etiological diagnosis of pulmonary disorders [110] there are still some limitations. Since only microbial DNA can be detected, there is a lack of information concerning important parameters such as: quantity or bacterial load and antibiotic sensitivity of isolated pathogens. Additionally, no isolates can be obtained for fingerprinting (by phenotypic or genotypic methods). Quantitative real-time PCR (qPCR) leads the field in providing an adequate solution for the problem of quantitative assessment of particular respiratory pathogens. But there is no suitable gene target in microbial genome for the qPCR design. Although there are some attempts to use the16 S rRNA gene for this purpose [56], unfortunately its copy number is a highly variable species-specific feature. The number of 16 S rRNA gene copies per genome can vary between 1 and 15 [100]. To normalize the qPCR data a standard curve using, for example, *Helicobacter hepaticus* DNA (a bacterium known to have only a single copy of the 16 S rRNA gene in its genome) can be constructed. But in this case only the total number in a sample 16 S rRNA gene copies can be evaluated. This provides no information about the total bacterial load.

It is also important to emphasize that detected bacterial DNA does not necessarily indicates the presence of living bacteria in the respiratory tract. Some studies showed that structural components of bacterial cells, such as LPS or peptidoglycan constantly activate pattern recognition receptors on epithelial cells, dendritic cells and macrophages, thus enabling the tolerance of the immune system and preventing constant inflammation [136]. This fact leads to the conclusion, that detection of bacterial DNA in respiratory samples plays an important role in providing of stimuli, which induce immune tolerance.

T-RFLP is a molecular approach that allows rapid comparison of the community structure and diversity of different origins. This technique has been used for assessing the diversity and structure of complex bacterial communities in various environments, mainly being soil and water. In the present study T-RFLP analysis was used to characterize and compare pulmonary microbiota present in BAL samples of 9 healthy subjects and 9 patients with COPD.

T-RFLP using the restriction endonuclease *Msp*I analysis could assign different peaks. But a number of restrictions of the present study and areas for future research should be mentioned. First of all, it is a well-known fact that multiple species can be predicted for the same TRF length and the identification of bacterial species can be improved by analysis of digests with multiple restriction enzymes. *Hha*I, *Msp*I and *Rsa*I were shown to have the best resolution rate in cutting the fragments of 16 S rRNA gene [99]. Since only one endonuclease was used to digest the PCR products, some gaps in community profiles could arise. By comparing the sequencing data with T-RFLP profiles some genera (*Haemophilus*, *Methylophilus*, *Propionibacterium*, *Terrahaemophilus*, *Eubacterium*, and *Centipeda* for healthy individual group; *Abiotrophia*, *Granulicatella*, *Methylobacterium*, *Afipia*, *Brevundimonas*, *Curvibacter*, *Moraxella*, *Variovorax*, *Undibacterium*, *Rhodocyclaceae*, *Capnocytophaga*, and *Leptolyngbya* for COPD patient cohort) are missing in generated profiles. This can be enterpreted in several ways; the resolution of *Msp*I is not high enough to distinguish between all the presenting genera. Or the copy number of 16 S rRNA gene is too low in this bacteria. In this case due to the low abundance they could not be read as separate peaks by the software and thus were excluded from the analysis.

Interestingly, *Mycoplasma* sp. and *Legionella* sp. were detected by T-RFLP, but not by sequencing. Since the method of DNA extraction did not differ in both approaches, it is rather difficult to find an explanation for this phenomenon. The used primer set (Ba27f and Ba907r) possibly did not target conservative regions in their 16 S rRNA gene.

There is a growing body of evidence that indicated that changes in pulmonary microbial communities are linked to different pathological states. These changes may play a key role in the pathogenesis and progression of airway obstruction and emphysema.

This study indicated that T-RFLP analysis is useful for the assessment of diversity of pulmonary microbiota and the rapid comparison of the community structure between subjects with and without COPD.

Because T-RFLP does not offer sufficient information to confirm the phylogenetic information of pulmonary microbial communities, the community structures were analyzed by sequencing. Sequencing is a procedure that is used to determine the order of nucleotides in a particular DNA fragment (16 S rRNA gene). The sequencing technique used for the present project relies on the Sanger dideoxy method, which was originally developed in 1977 by Frederick Sanger and permits analysis of the target DNA up to 1 kb from the site of primer hybridization. This procedure generates single-stranded DNA fragments terminating at each of four nucleotides (A, G, T, C), which are then differently labeled with fluorophores that emit light at different wavelengths upon excitation with a laser.

The directionality and antiparallel nature of all target DNA strands were considered before phylogenetic analysis (the plus strand delivered from bacterial isolates was used).

The T-RFLP data provided rapid information about individual changes in pulmonary microbial communities and revealed the presence of 9 core genera: *Prevotella*, *Sphingomonadaceae*, *Pseudomonas*, *Acinetobacter*, *Fusobacterium*, *Megasphaera*, *Veillonella*, *Staphylococcus*, and *Streptococcus*. Thus they could be well correlated with sequencing analysis. Despite some differences (such as detection of untypical pathogens), T-RFLP appeared to be a suitable tool for the pulmonary microbiome characteristic.

The sequencing allows the construction of multiple sequence alignments that are useful for depicting evolutionary relationships among and between genes from different bacteria. By rooting the phylogenetic trees all bacteria were classified into distinct clades (internal nodes) (Figure 22, 23). For examples, all 24 representatives of *Prevotella* sp. detected in healthy individual cohort formed one clade and are said to be monophyletic. A great deal of emphasis is placed on whether a set of taxa forms a clade and thus can be classified based on evolutionary ancestry. This is the most important advantage of sequencing compared with T-RFLP analysis. In study of STRESSMANN et al. T-RFLP analysis as well as sequencing were applied to study respiratory samples of cystic fibrosis patients [165]. The detection of species by clone sequencing that corresponded to T-RF band lengths in individual patient's profiles ranged from 13 % to 67 %.

5.3 Eukaryotic DNA blocks the PCR amplification of 16 S rRNA gene

It is a well-known fact that the presence of total eukaryotic DNA can inhibit the PCR amplification of bacterial nucleic acids. This could be due to competitive non-specific

hybridization of a large amount of human DNA with 16 S rRNA gene-specific primers [111]. Moreover, the inhibition may occur as a simple consequence of a decrease in the rate of diffusion of all macromolecular components of the reaction mixture in the presence of large amounts of long strands of host DNA. In the present work, the total DNA was used as a template for the PCR as well. The sensitivity of PCR using primers Ba27f / Ba907r was not affected by human DNA (see Figure 9). In the case of BAL samples, the used primer set could not provide sensitive enough results in the case of human and mice lung tissues (data not shown). There were unspecific by-products generated from eukaryotic DNA, which were observed.

Additionally, PCR-induced artifacts in microbial diversity studies remain an important issue in the search for patterns and the extent of microbial diversity. The basic types of PCR artifacts have been shown in controlled laboratory studies and can be divided into two categories: those resulting in sequence artifacts (PCR errors), and those skewing the distribution of PCR products due to unequal amplification (PCR bias) or cloning efficiency. Sequence artifacts may arise due to the formation of chimerical molecules, the formation of heteroduplex molecules, and *Taq* DNA polymerase error [3]. Thereafter the Phusion Hot Start proofreading DNA polymerase was used. It allowed the minimization of the sequence artifacts. To minimize chimeras, the smallest possible number of PCR amplification cycles should be carried out (e.g., until a band is barely visible on agarose gels). But due to the used sample type (total DNA containing eukaryotic part) it was not profitable to decrease the cycle number to less than 40.

By and large, the PCR conditions for amplification of 16 S rRNA gene fragment were optimized to possibly avoid the artifacts and bias without losing its sensitivity.

5.4 Sequencing: confidence and bias

A phylogenetic tree is considered to be well supported if it has passed statistical tests measuring significance. An accepted method used to assess confidence is called bootstrapping. In the most general sense, bootstrapping is a statistical procedure in which raw data are re-assembled numerous times so as to estimate model parameters such as the mean or variance of the set of data. Both phylogenetic trees presented in this work were bootstrapped 1000 times.

Nowadays, sequencing projects simply download all available genomic data from public databases, such as the NCBI GenBank, RDP or Greengenes. Although this approach gives a reasonable first approximation of the diversity within each species, there are problems associated with the marked sampling biases in public data. For instance, there is a particular overabundance of sequenced pathogen strains in comparison to their non-pathogenic relatives. Another issue is the variability of metadata associated

with each genome. The annotation of gene and other sequence features is often variable and dependent on the year of publication and the research group. Also, important information on strain provenance and phenotype is often unavailable. There are no accepted standards for the level of sequence read coverage necessary to produce a good alignment. Genome projects can vary considerably in the number of sequencing errors. Artifacts, such as frame shifts, can remove true coding sequences from consideration as part of a pan-genome.

Clones libraries construction approach has considerably increased our understanding of the role of bacteria in initiating, progression and spreading of different diseases [21, 36, 37, 43, 51, 63, 87, 105, 124, 189]. But databases used to classify the sequences still need to be improved.

5.5 Bacterial presence in lower airways is confirmed by FISH

Mouse tracheal cross-sections were used for FISH analysis. In the present study, the procedure for the *in situ* analysis of the distribution of bacteria in mice lung tissue samples was optimized for the first time. Key elements of the procedure are: adequate fixation, optimal permeabilization, and careful selection of the best combination between probe, fluorescence label, and formamide concentration. The major finding was that bacteria persist under the epithelial layer in the trachea.

The FISH procedure was performed on human and mice lung tissue sections (data not shown). Unfortunately, no clearly distinguishable positive signal could be detected. One possible explanation of this fact is that lung tissues are known to have an extremely high autofluorescence level. That's why a positive signal could just be overlapped by one produced by proteins in lung tissues. Use of new methods such as FISH affords experience in FISH microscopy and adherence to evaluation standards. FISH signals within complex eukaryotic tissues can generate a vast number of unspecific fluorescence phenomena [169].

M. catarrhalis and *H. influenzae* can be internalized in respiratory epithelial cells [158, 109]. This new aspect of the interplay of respiratory opportunistic pathogens may be essential to its chronic persistence inside the host.

6 Conclusion

The present study enhances understanding of the role of pulmonary microbial communities in health and disease. The 18 BAL samples obtained from healthy individuals and COPD patients were examined for the presence of a variety of bacteria. It appears that the lower respiratory tract is colonized by microbes, that can stably persist in it.

The T-RFLP approach illustrated the high qualitative and quantitative diversity of the microbiome structure in different individuals. Almost all studied samples demonstrated the presence of both common and unique ribotypes. No distinctive pattern in microbial composition could be seen either in healthy persons or in COPD patients. But, the presence of such genera as *Mycoplasma* (6 positive samples), *Pseudomonas* (5 positive samples) and *Staphylococcus* (5 positive samples) were highly correlated with disease. Patients with ID numbers 4 and 9 had probable monoinfection with *Pseudomonas* sp.. This led to the reduction of standard microbiota diversity. This fact can be used as a basis for a predictive model in using T-RFLP in rapid diagnostic tests. Interestingly, the presence of genus *Lactobacillus* completely correlated with the distribution of *Megasphaera* between healthy individuals and with the occurrence of *Mezorhizobium*, *Pseudomonas* and *Streptomyces*. The formation of complex microbial associations within the respiratory tract could provide the explanation for this phenomenon. Genera such as *Mycobacterium*, *Lysinibacillus*, *Gemella*, *Treponema*, *Streptobacillus*, and *Alkaliphilus* were characteristic of the COPD state. On the other hand, *Mesorhizobium*, *Microbacteriaceae*, *Micrococcus*, *Veillonelaceae*, *Rhizobium*, *Stenotrophomonas*, and *Lactococcus* presented exclusively in healthy individual cohort. Some significant quantitative changes were also found in the distribution of genus *Neisseria*. Its presence in lower amounts was characteristic of steady state. Increasement of *Neisseria* load was linked with COPD.

Sequencing data, supported by T-RFLP analysis, revealed the presence of core microbiome in the lower respiratory tract. The key genera included *Prevotella*, *Sphingomonas*, *Pseudomonas*, *Acinetobacter*, *Fusobacterium*, *Megasphaera*, *Veillonella*, *Staphylococcus*, and *Streptococcus*.

One of the main findings of this project is the detection of probiotic bacteria of the genera *Lactobacillus* and *Lactococcus* in respiratory samples. Due to their known role in the maintenance of the homeostasis, immunomodulatory and antiinflammatory activities; the results obtained can be used in the probiotic bacteria based drugs production.

Modern microbiology is moving into the new “population genomics era”, when a typical microbial pathogen (1 Mb to 6 Mb) can be sequenced in just one day [38]. More than 1600 bacterial genomes are now available in publicly accessible genome sequence databases. This needs the reevaluation of the basic concepts in clinical diagnostics as well as in the fundamental field of microbiological research. Particularly, the

whole-genome sequencing projects have led to the idea of the species “pan-genome” [23, 86]. This means that each microbial genome consists of two principal parts: the core genome and the dispensable genome. The core genome is represented by the set of genes conserved across all strains in a species. It encodes the basic functions and features necessary for basic survival. The dispensable genome contributes to the diversity within the species, including niche adaptation, antibiotic resistance, virulence, transmissibility etc. In terms of pathogen detection and subtyping, discriminating the core genome is important for the identification of diagnostic targets of all bacteria in a taxon. The dispensable genome is helpful for identifying particular key genes that can be used to target only pathogenic strains. Important changes such as gene copy number alterations (through deletion, tandem duplication or site-specific recombination) can be registered by means of comparative population genomics. Knowing detailed information about the composition of microbiota in the lower respiratory tract will improve the therapeutical approaches of personalized medicine. Especially once it has been proved that microbiota in the lungs have individual differences.

The whole genome sequencing will become soon a routine epidemiological approach [117]. This data will deliver a fundamental new understanding of microbial genetics and ecology [168, 123]. For the routine praxis simultaneous detection of two or more genes of interest using novel microfluidic digital PCR will become more relevant [116].

Recently it was shown that three predominant enterotypes could be distinguished in the human population [15]. Future systemic studies may provide new links between the intestinal and respiratory microbiota. Defining the complexity of the human microbiome in health and disease will enhance the understanding of multiple pathological mechanisms and facilitate the development of novel diagnostic tools and therapeutic interventions.

Nomenclature

16 S rRNA gene	a section of prokaryotic DNA found in all bacteria and archaea
ALF	alveolar lining fluid
ASL	airway surface liquid
Bacteroid	symbiotic form of the nitrogen - fixing bacteria
COPD	chronic obstructive pulmonary disease
Diphteroid	pleomorphic non - motile, Gram - positive rods of the Corynebacterium and Propionibacterium genera
DNA	deoxyribonucleic acid
Endonuclease	enzyme which cleaves the phosphodiester bond within a polynucleotide chain
FAM	6-carboxy-fluorescein
FISH	fluorescence *in situ* hybridization
GOLD	the Global initiative for chronic Obstructive Lung Disease
LPS	lipopoligosaccharide
MAMPs	microbe - associated molecular patterns
NCBI	National Center for Biotechnology Information
NLR	NOD like receptor
OTU	operational taxonomic unit, a terminal node in phylogenetic analysis
PCA	principal component analysis
PCR	polymerase chain reaction
RDP	ribosomal database project
RLR	retinoic acid - inducible gene (RIG)-I-like receptor
ROS	reactive oxygen species

rRNA	ribosomal ribonucleic acid
T-RFLP	terminal restriction fragment length polymorphism
TLR	toll like receptor
TNF	tumor necrosis factor
TRF	terminal restriction fragment
Ub	ubiquitin
WHO	World Health Organization

Index

References

[1] Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Medical Communications Resources Inc, 2009. (Cited on pages 1, 2, and 3)

[2] J. A. Aas, B. J. Paster, L. N. Stokes, I. Olsen, and F. E. Dewhirst. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol*, 43(11):5721–5732, 2005. (Cited on pages 14 and 60)

[3] S. G. Acinas, R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol*, 71(12):8966–8969, 2005. (Cited on page 64)

[4] A. Aderem. Phagocytosis and the inflammatory response. *J Infect Dis*, 187:340–345, 2003. (Cited on page 6)

[5] S. Akira, S. Uematsu, and O. Takeuchi. Pathogen recognition and innate immunity. *Cell*, 124(4):783–801, 2006. (Cited on pages 4 and 5)

[6] M. Ali, E.P. Lillehoj, Y. Park, Y. Kyo, and K. C. Kim. Analysis of the proteome of human airway epithelial secretions. *Proteome Sci*, 9:4, 2011. (Cited on page 7)

[7] A. O. Aliprantis, R. B. Yang, M. R. Mark, S. Suggett, B. Devaux, J. D. Radolf, G. R. Klimpel, and P. Godowski. Cell activation and apoptosis by bacterial lipoproteins through toll-like receptor-2. *Science*, 285(5428):736–739, 1999. (Cited on page 5)

[8] R. Amann, B. M. Fuchs, and S. Behrens. The identification of microorganisms by fluorescence *in situ* hybridisation. *Curr Opin Biotechnol*, 12(3):231–236, 2001. (Cited on page 33)

[9] A. F. Andersson, M. Lindberg, H. Jakobsson, F. Baeckhed, P. Nyraen, and L. Engstrand. Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS One*, 3(7):2836, 2008. (Cited on page 15)

[10] A. Andoh, S. Sakata, Y. Koizumi, K. Mitsuyama, Y. Fujiyama, and Y. Benno. Terminal restriction fragment length polymorphism analysis of the diversity of fecal microbiota in patients with ulcerative colitis. *Inflamm Bowel Dis*, 13(8):955–962, 2007. (Cited on page 25)

[11] W. J. Ansorge. Next-generation DNA sequencing techniques. *New Biotechnology*, 25 (6)(4):195–203, 2009. (Cited on page 15)

[12] A. Anzueto, S. Sethi, and F. J. Martinez. Exacerbations of chronic obstructive pulmonary disease. *Proc Am Thorac Soc*, 4(7):554–564, 2007. (Cited on pages 4 and 58)

[13] K. Arakawa, Y. Kawai, K. Lioka, M. Tanioka, J. Nishimura, H. Kitazawa, K. Tsurumi, and T. Saito. Microbial community analysis of food-spoilage bacteria in commercial custard creams using culture-dependent and independent methods. *J Dairy Sci*, 91(8):2938–2946, 2008. (Cited on page 11)

[14] L. M. Armstrong, R. L. Andrew, K. M. Uppington, J. Robertson, L. R. Witherden, and T. D. Tetley. Expression of functional toll-like receptor-2 and -4 on alveolar epithelial cells. *Am J Respir Cell Mol Biol*, 31(2):241–245, 2004. (Cited on page 5)

[15] M. Arumugam, J. Raes, E. Pelletier, D. L. Paslier, T. Yamada, and P. Bork. Enterotypes of the human gut microbiome. *Nature*, 473(7346):174–180, 2011. (Cited on page 67)

[16] S. N. Aslam, G. Erbs, K. L. Morrissey, M. A. Newman, D. Chinchilla, T. Boller, A. Molinaro, R. W. Jackson, and R. M. Cooper. Microbe-associated molecular pattern (MAMP) signatures, synergy, size and charge: influences on perception or mobility and host defence responses. *Molecular Plant Pathology*, 10(3):375–387, 2009. (Cited on page 5)

[17] M. Axelson-Fisk. *Comparative gene finding: models, algorithms and implementation*. Springer-Verlag London Limited, 2010. (Cited on page 32)

[18] M. M. Ayers and P. K. Jeffery. Proliferation and differentiation in mammalian airway epithelium. *Eur Respir J*, 1(1):58–80, 1988. (Cited on page 6)

[19] F. K. Bahrani-Mougeot, B. J. Paster, S. Coleman, S. Barbuto, M. T. Brennan, J. Noll, T. Kennedy, P. C. Fox, and P. B. Lockhart. Molecular analysis of oral and respiratory bacterial species associated with ventilator-associated pneumonia. *J Clin Microbiol*, 45(5):1588–1593, 2007. (Cited on page 61)

[20] D. Banerjee, O. A. Khair, and D. Honeybourne. Impact of sputum bacteria on airway inflammation and health status in clinical stable COPD. *Eur Respir J*, 23(5):685–691, 2004. (Cited on pages 10 and 41)

[21] P. T. Becker, E. Egea, and I. Eeckhaut. Characterization of the bacterial communities associated with the bald sea urchin disease of the echinoid *Paracentrotus lividus*. *J Invertebr Pathol*, 98(2):136–147, 2008. (Cited on page 65)

[22] T. Bell, J. A. Newman, B. W. Silverman, S. L. Turner, and A. K. Lilley. The contribution of species richness and composition to bacterial services. *Nature*, 436(7054):1157–1160, 2005. (Cited on page 11)

[23] S. Bentley. Sequencing the species pan-genome. *Nat Rev Microbiol*, 7(4):258–259, 2009. (Cited on page 67)

[24] C. Bergeron and L. P. Boulet. Structural changes in airway diseases: characteristics, mechanisms, consequences, and pharmacologic modulation. *Chest*, 129(4):1068–1087, 2006. (Cited on page 2)

[25] E. M. Bik. The microbiotics of the human body. *Ned Tijdschr Geneeskd*, 152(12):668–673, 2008. (Cited on page 15)

[26] A. Bircan, M. Gokirmak, O. Kilic, O. Ozturk, and A. Akkaya. C-reactive protein levels in patients with chronic obstructive pulmonary disease: role of infection. *Med Princ Pract*, 17(3):202–208, 2008. (Cited on page 4)

[27] Z. L. Borrill, K. Roy, and D. Singh. Exhaled breath condensate biomarkers in COPD. *Eur Respir J*, 32(2):472–486, 2008. (Cited on page 58)

[28] O. G. Brakstad, K. Aasbakk, and J. A. Maeland. Detection of *Staphylococcus aureus* by polymerase chain reaction amplification of the nuc gene. *J Clin Microbiol*, 30(7):1654–1660, 1992. (Cited on page 23)

[29] C. J. Brown, M. Wong, C. C. Davis, A. Kanti, X. Zhou, and L. J. Forney. Preliminary characterization of the normal microbiota of the human vulva using cultivation-independent methods. *J Med Microbiol*, 56(Pt 2):271–276, 2007. (Cited on page 15)

[30] T. B. Brueck, W. M. Brueck, L. Z. Santiago-Vaezquez, P. J. McCarthy, and R. G. Kerr. Diversity of the bacterial communities associated with the azooxanthellate deep water octocorals *Leptogorgia minimata*, *Iciligorgia schrammi*, and *Swiftia exertia*. *Mar Biotechnol*, 9(5):561–576, 2007. (Cited on page 33)

[31] C. L. Campbell, D. L. Mummey, E. T. Schmidtmann, and W. C. Wilson. Culture-independent analysis of midgut microbiota in the arbovirus vector *Culicoides sonorensis* (Diptera: *Ceratopogonidae)*. *J Med Entomol*, 41(3):340–348, 2004. (Cited on page 25)

[32] L. A. Campbell, M. Perez Melgosa, D. J. Hamilton, C. C. Kuo, and J. T. Grayston. Detection of *Chlamydia pneumoniae* by polymerase chain reaction. *J Clin Microbiol*, 30(2):434–439, 1992. (Cited on page 23)

[33] M. J. Caplan, P. Seo-Mayer, and L. Zhang. Epithelial junctions and polarity: complexes and kinases. *Curr Opin.Nephrol.Hypertens.*, 17(5):506–512, 2008. (Cited on page 6)

[34] G. Caramori, M. Romagnoli, P. Casolari, C. Bellettato, G. Casoni, P. Boschetto, K. F. Chung, P. J. Barnes, I. M. Adcock, A. Ciaccia, L. M. Fabbri, and A. Papi. Nuclear localisation of p65 in sputum macrophages but not in sputum neutrophils during COPD exacerbations. *Thorax*, 58(4):348–351, 2003. (Cited on page 5)

[35] Y. R Chan and A. Morris. Molecular diagnostic methods in pneumonia. *Curr Opin Infect Dis*, 20(2):157–164, 2007. (Cited on page 58)

[36] J. Y. Chang, D. A. Antonopoulos, A. Kalra, A. Tonelli, W. T. Khalife, T. M. Schmidt, and W. B. Young. Decreased diversity of the fecal Microbiome in recurrent *clostridium* difficile-associated diarrhea. *J Infect Dis*, 197(3):435–438, 2008. (Cited on page 65)

[37] J. Chen, F. L. Fluharty, N. St-Pierre, M. Morrison, and Z. Yu. Technical note: Occurrence in fecal microbiota of genes conferring resistance to both macrolide-lincosamide-streptogramin B and tetracyclines concomitant with feeding of beef cattle with tylosin. *J Anim Sci*, 86(9):2385–2391, 2008. (Cited on page 65)

[38] P. E. Chen, K. M. Willner, A. Butani, S. Dorsey, M. George, A. Stewart, S. M. Lentz, C. E. Cook, A. Akmal, L. B. Price, P. S. Keim, A. Mateczun, T. N. Brahmbhatt, K. A. Bishop-Lilly, M. E. Zwick, T. D. Read, and S. Sozhamannan. Rapid identification of genetic modifications in *Bacillus anthracis* using whole genome draft sequences generated by 454 pyrosequencing. *PLoS One*, 5(8):e12397, 2010. (Cited on page 66)

[39] T. Coenye, J. Goris, T. Spilker, P. Vandamme, and J. J. LiPuma. Characterization of unusual bacteria isolated from respiratory secretions of cystic fibrosis patients and description of *Inquilinus limosus* gen. nov., sp. nov. *J Clin Microbiol*, 40(6):2062–2069, 2002. (Cited on page 51)

[40] T. Coenye, P. Vandamme, and J. J. LiPuma. Infection by *Ralstonia species* in cystic fibrosis patients: identification of *R. pickettii* and *R. mannitolilytica* by polymerase chain reaction. *Emerg Infect Dis*, 8(7):692–696, 2002. (Cited on page 61)

[41] D. D. Creer, J. P. Dilworth, S. H. Gillespie, A. R. Johnston, S. L. Johnston, C. Ling, S. Patel, G. Sanderson, P. G. Wallace, and T. D. McHugh. Aetiological role of viral

and bacterial infections in acute adult lower respiratory tract infection (LRTI) in primary care. *Thorax*, 61(1):75–79, 2006. (Cited on page 4)

[42] S. W. Culman, H. G. Gauch, C. B. Blackwood, and J. E. Thies. Analysis of T-RFLP data using analysis of variance and ordination methods: a comparative study. *J Microbiol Methods*, 75(1):55–63, 2008. (Cited on page 35)

[43] T. Curran, W. A. Coulter, D. J. Fairley, T. McManus, J. Kidney, M. Larkin, J. E. Moore, and P. V. Coyle. Development of a novel DNA microarray to detect bacterial pathogens in patients with chronic obstructive pulmonary disease (COPD). *J Microbiol Methods*, 80(3):257–261, 2010. (Cited on pages 16 and 65)

[44] T. Curran, P. V. Coyle, T. E. McManus, J. Kidney, and W. A. Coulter. Evaluation of real-time PCR for the detection and quantification of bacteria in chronic obstructive pulmonary disease. *FEMS Immunol Med Microbiol*, 50(1):112–118, 2007. (Cited on page 4)

[45] Crisan D. *Hematopathology: genomic mechanisms of neoplastic diseases*. Springer Science+Business Media, 2010. (Cited on page 15)

[46] I. A. Darby. *In situ hybridization protocols*. Humana Press Inc., 2000. (Cited on page 34)

[47] A. Deli, D. Koutsioulis, V. E. Fadouloglou, P. Spiliotopoulou, S. Balomenou, S. Arnaouteli, M. Tzanodaskalaki, K. Mavromatis, M. Kokkinidis, and V. Bouriotis. LmbE proteins from *Bacillus cereus* are de-N-acetylases with broad substrate specificity and are highly similar to proteins in *Bacillus anthracis*. *FEBS J*, 277(13):2740–53, 2010. (Cited on pages 5 and 6)

[48] I. K. Demedts, K. R. Bracke, T. Maes, G. F. Joos, and G. G. Brusselle. Different roles for human lung dendritic cell subsets in pulmonary immune defense mechanisms. *Am J Respir Cell Mol Biol*, 35(3):387–393, 2006. (Cited on page 6)

[49] L. Dethlefsen, P. B. Eckburg, E. M. Bik, and D. A. Relman. Assembly of the human intestinal microbiota. *Trends Ecol Evol*, 21(9):517–523, 2006. (Cited on page 15)

[50] S. A. Di, G. Caramori, T. Oates, A. Capelli, M. Lusuardi, I. Gnemmi, F. Ioli, K. F. Chung, C. F. Donner, P. J. Barnes, and I. M. Adcock. Increased expression of nuclear factor-kappaB in bronchial biopsies from smokers and patients with COPD. *Eur Respir J*, 20(3):556–563, 2002. (Cited on page 5)

[51] D. Dicksved, J. Halfvarson, M. Rosenquist, G. Jaernerot, C. Tysk, J. Apajalahti, L. Engstrand, and J. K. Jansson. Molecular analysis of the gut microbiota of

identical twins with Crohn's disease. *ISME J*, 2(7):716–727, 2008. (Cited on page 65)

[52] B. M. Diederen, P. D. van der Valk, J. A. Kluytmans, M. F. Peeters, and R. Hendrix. The role of atypical respiratory pathogens in exacerbations of chronic obstructive pulmonary disease. *Eur Respir J*, 30(2):240–244, 2007. (Cited on page 58)

[53] J. Dunbar, S. White, and L. Forney. Genetic Diversity through the Looking Glass: Effect of Enrichment Bias. *Appl Environ Microbiol*, 63(4):1326–1331, 1997. (Cited on page 32)

[54] R. C. Evans E. W. Nester. Microbiology, 1998. (Cited on pages 8 and 11)

[55] M. Egert and M. W. Friedrich. Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure. *Appl Environ Microbiol*, 69(5):2555–2562, 2003. (Cited on page 27)

[56] J. R. Erb-Downward, D. L. Thompson, M. L. Han, C. M. Freeman, L. McCloskey, L. A. Schmidt, V. B. Young, G. B. Toews, J. L. Curtis, B. Sundaram, F. J. Martinez, and G. B. Huffnagle. Analysis of the lung microbiome in the "healthy" smoker and in COPD. *PLoS One*, 6(2):e16384, 2011. (Cited on pages 60 and 61)

[57] P. Fabian, J. J. McDevitt, W. M. Lee, E. A. Houseman, and D. K. Milton. An optimized method to detect influenza virus and human rhinovirus from exhaled breath and the airborne environment. *J Environ Monit*, 11(2):314–317, 2009. (Cited on pages 4 and 59)

[58] M. Fukata, A. S. Vamadevan, and M. T. Abreu. Toll-like receptors (TLRs) and Nod-like receptors (NLRs) in inflammatory disorders. *Semin Immunol*, 21 (6)(4):242–253, 2009. (Cited on page 4)

[59] K. S. George, M. E. Fuschino, K. Mokhiber, W. Triner, and S. D. Spivack. Exhaled breath condensate appears to be an unsuitable specimen type for the detection of influenza viruses with nucleic acid-based methods. *J Virol Methods*, 163(1):144–146, 2010. (Cited on pages 59 and 60)

[60] C. Gessner, R. Scheibe, M. Woetzel, S. Hammerschmidt, H. Kuhn, L. Engelmann, G. Hoheisel, A. Gillissen, U. Sack, and H. Wirtz. Exhaled breath condensate cytokine patterns in chronic obstructive pulmonary disease. *Respir Med*, 99(10):1229–1240, 2005. (Cited on page 4)

[61] A. T. Gewirtz, T. A. Navas, S. Lyons, and P. J. Godowski. Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *J Immunol*, 167 (6)(4):1882–1885, 2001. (Cited on page 5)

[62] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R. Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z. Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, H. O. Smith, and J. C. Venter. Creation of a bacterial bell controlled by a chemically synthesized genome. *Science*, 239(5987):52–6, 2010. (Cited on page 15)

[63] O. Goldenberg, S. Herrmann, G. Marjoram, M. Noyer-Weidner, G. Hong, S. Bereswill, and U. B. Gaebel. Molecular monitoring of the intestinal flora by denaturing high performance liquid chromatography. *J Microbiol Methods*, 68(1):94–105, 2007. (Cited on page 65)

[64] F. P. Gomez and R. Rodriguez-Roisin. Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines for chronic obstructive pulmonary disease. *Curr Opin Pulm Med*, 8(2):81–86, 2002. (Cited on pages 4 and 20)

[65] M. I. Gomez. Airway epithelial cell signaling in response to bacterial pathogens. *Pediatr Pulmonol*, 43 (6)(1):11–19, 2008. (Cited on page 6)

[66] V. Gontcharova, E. Youn, R. D. Wolcott, E. B. Hollister, T. J. Gentry, and S. E. Dowd. Black Box Chimera Check (B2C2): a windows-based software for batch depletion of chimeras from bacterial 16S rRNA gene datasets. *Open Microbiol J*, 4:47–52, 2010. (Cited on page 32)

[67] A. Guamer. Role of intestinal flora in health and disease. *Nutr Hosp*, 22:14–19, 2007. (Cited on page 15)

[68] L. Guillot, S. Medjane, K. Le-Barillec, V. Balloy, C. Danel, M. Chignard, and M. Si-Tahar. Response of human pulmonary epithelial cells to lipopolysaccharide involves Toll-like receptor 4 (TLR4)-dependent signaling pathways: evidence for an intracellular compartmentalization of TLR4. *J Biol Chem*, 279 (6)(4):2712–2718, 2004. (Cited on page 5)

[69] J. A. Guttman and B. B. Finlay. Tight junctions as targets of infectious agents. *Biochim Biophys Acta*, 1788(4):832–841, 2009. (Cited on page 6)

[70] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *SIGKDD Explorations*, 11:10–18, 2009. (Cited on page 53)

[71] A. R. Hauser. The type III secretion system of *Pseudomonas aeruginosa*: infection by injection. *Nat Rev Microbiol*, 7 (6)(9):654–665, 2009. (Cited on page 6)

[72] H. Hemmi, O. Takeuchi, T. Kawai, T. Kaisho, S. Sato, H. Sanjo, M. Matsumoto, K. Hoshino, H. Wagner, and K. Takeda. A Toll-like receptor recognizes bacterial DNA. *Nature*, 408 (6)(6813):740–745, 2000. (Cited on page 5)

[73] P. H. Hendolin, A. Markkanen, J. Ylikoski, and J. J. Wahlfors. Use of multiplex PCR for simultaneous detection of four bacterial species in middle ear effusions. *J Clin Microbiol*, 35(11):2854–2858, 1997. (Cited on page 23)

[74] C. Hess, C. Herr, C. Beisswenger, T. Zakharkina, R. M. Schmid, and R. Bals. Myeloid RelA regulates pulmonary host defense networks. *Eur Respir J*, 35(2):343–352, 2010. (Cited on page 5)

[75] M. Hilty, C. Burke, H. Pedro, P. Cardenas, A. Bush, C. Bossley, J. Davies, A. Ervine, L. Poulter, L. Pachter, M. F. Moffatt, and W. O. Cookson. Disordered microbial communities in asthmatic airways. *PLoS One*, 5(1):8578, 2010. (Cited on pages 9, 14, and 60)

[76] I. Horvath, J. Hunt, P. J. Barnes, K. Alving, A. Antczak, E. Baraldi, G. Becher, W. J. van Beurden, M. Corradi, R. Dekhuijzen, R. A. Dweik, T. Dwyer, R. Effros, S. Erzurum, B. Gaston, C. Gessner, A. Greening, L. P. Ho, J. Hohlfeld, Q. Joebsis, D. Laskowski, S. Loukides, D. Marlin, P. Montuschi, A. C. Olin, A. E. Redington, P. Reinhold, E. L. van Rensen, I. Rubinstein, P. Silkoff, K. Toren, G. Vass, C. Vogelberg, and H. Wirtz. Exhaled breath condensate: methodological recommendations and unresolved questions. *Eur Respir J*, 26(3):523–548, 2005. (Cited on pages 4 and 20)

[77] M. K. Hourfar, A. Themann, M. Eickmann, P. Puthavathana, T. Laue, E. Seifried, and M. Schmidt. Blood screening for influenza. *Emerg Infect Dis*, 13(7):1081–1083, 2007. (Cited on page 23)

[78] S. H. Huang, W. Zhou, and A. Jong. Focal point theory models for dissecting dynamic duality problems of microbial infections. *J Biomed Biotechnol*, 2008:856314, 2008. (Cited on page 7)

[79] Y. J. Huang, E. Kim, M. J. Cox, E. L. Brodie, R. Brown, J. P. Wiener-Kronish, and S. V. Lynch. A persistent and diverse airway microbiota present during chronic

obstructive pulmonary disease exacerbations. *OMICS*, 14(1):9–59, 2010. (Cited on pages 1, 16, and 60)

[80] J. R. Hurst, W. R. Perera, T. M. Wilkinson, G. C. Donaldson, and J. A. Wedzicha. Systemic and upper and lower airway inflammation at exacerbation of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 173(1):71–78, 2006. (Cited on page 3)

[81] A. F. Hutchinson, A. K. Ghimire, M. A. Thompson, J. F. Black, C. A. Brand, A. J. Lowe, D. M. Smallwood, R. Vlahos, S. Bozinovski, G. V. Brown, G. P. Anderson, and L. B. Irving. A community-based, time-matched, case-control study of respiratory viruses and exacerbations of COPD. *Respir Med*, 101(12):2472–2481, 2007. (Cited on page 58)

[82] H. Inoue, H. Nojima, and H. Okayama. High efficiency transformation of Escherichia coli with plasmids. *Gene*, 96(1):23–28, 1990. (Cited on page 29)

[83] A. S. Jackson, A. Sandrini, C. Campbell, S. Chow, P. S. Thomas, and D. H. Yates. Comparison of biomarkers in exhaled breath condensate and bronchoalveolar lavage. *Am J Respir Crit Care Med*, 175(3):222–227, Feb 2007. (Cited on pages 4 and 59)

[84] R. Jain, C. A. Schriever, L. H. Danziger, S. H. Cho, and I. Rubinstein. The IS6110 repetitive DNA element of *Mycobacterium tuberculosis* is not detected in exhaled breath condensate of patients with active pulmonary tuberculosis. *Respiration*, 74(3):329–333, 2007. (Cited on pages 4 and 59)

[85] P. Joppa, D. Petrasova, B. Stancak, and R. Tkacova. Systemic inflammation in patients with COPD and pulmonary hypertension. *Chest*, 130(2):326–333, 2006. (Cited on page 4)

[86] S. J. Joseph and T. D. Read. Bacterial population genomics and infectious disease diagnostics. *Trends Biotechnol*, 28(12):611–618, 2010. (Cited on page 67)

[87] T. Kawanami, K. Fukuda, K. Yatera, T. Kido, C. Yoshii, H. Taniguchi, and M. Kido. Severe pneumonia with Leptotrichia sp. detected predominantly in bronchoalveolar lavage fluid by use of 16S rRNA gene sequencing analysis. *J Clin Microbiol*, 47(2):496–498, 2009. (Cited on pages 13 and 65)

[88] D. Kelly, J. I. Campbell, T. P. King, G. Grant, E. A. Jansson, A. G. P. Coutts, S. Pettersson, and S. Conway. Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA. *Nat Immunol*, 5(1):104–112, 2004. (Cited on page 7)

[89] H. Kessler. *Molecular diagnostics of infectious diseases*. Walter de Gruyter GmbH & Co, 2010. (Cited on page 12)

[90] O. A. Khair, R. J. Davies, and J. L. Devalia. Bacterial-induced release of inflammatory mediators by bronchial epithelial cells. *Eur Respir J*, 9(9):1913–1922, 1996. (Cited on page 6)

[91] R. Knippers. *Molekulare Genetik*. Georg Thieme Verlag, 9 edition, 2006. (Cited on page 28)

[92] A.-R. Koczulla, S. Noeske, C. Herr, R. A. Jorres, H. Rummelt, C. Vogelmeier, and R. Bals. Acute and chronic effects of smoking on inflammation markers in exhaled breath condensate in current smokers. *Respiration*, 79(1):61–67, 2010. (Cited on page 4)

[93] K. Kostikas, M. Gaga, G. Papatheodorou, T. Karamanis, D. Orphanidou, and S. Loukides. Leukotriene B4 in exhaled breath condensate and sputum supernatant in patients with COPD and asthma. *Chest*, 127(5):1553–1559, 2005. (Cited on page 58)

[94] A. Kumar, H. Wu, L. S. Collier-Hyams, J. M. Hansen, T. Li, K. Yamoah, Z. Q. Pan, D. P. Jones, and A. S. Neish. Commensal bacteria modulate cullin-dependent signaling via generation of reactive oxygen species. *EMBO J*, 26 (6)(21):4457–4466, 2007. (Cited on page 7)

[95] H. Kumar, T. Kawai, and S. Akira. Toll-like receptors and innate immunity. *Biochem Biophys Res Commun*, 388 (6)(4):621–625, 2009. (Cited on page 5)

[96] Y. Lai, A. Di Nardo, T. Nakatsuji, A. Leichtle, Y. Yang, A. L. Cogen, Z. R. Wu, L. V. Hooper, R. R. Schmidt, S. von Aulock, K. A. Radek, C. M. Huang, and A. F. Ryan. Commensal bacteria regulate Toll-like receptor 3-dependent inflammation after skin injury. *Nat Med*, 15 (6)(12):1377–1382, 2009. (Cited on page 6)

[97] J. D. Lambeth. NOX enzymes and the biology of reactive oxygen. *Nat Rev Immunol*, 4(3):181–189, 2004. (Cited on page 7)

[98] K. Larsson. Aspects on pathophysiological mechanisms in COPD. *J Intern Med*, 262(3):311–340, 2007. (Cited on page 20)

[99] H. K. Lee, H. R. Kim, A. Mengoni, and D. H. Lee. Modified T-RFLP methods for taxonomic interpretation of T-RF. *J Microbiol Biotechnol*, 18(4):624–630, 2008. (Cited on page 62)

[100] Z. M.-P. Lee, C. Bussema, and T. M. Schmidt. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res*, 37:489–493, 2009. (Cited on page 61)

[101] P. Leophonte. Infections in chronic obstructive pulmonery disease. *Bull Acad Natl Med*, 188(1):47–64, 2004. (Cited on page 3)

[102] I. Letunic and P. Bork. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res*, 2011. (Cited on page 32)

[103] F. Li, M. A. J. Hullar, and L. W. Lampe. Optimization of terminal restriction fragment polymorphism (TRFLP) analysis of human gut microbiota. *J Microbiol Methods*, 68(2):303–311, 2007. (Cited on page 25)

[104] W. T. Liu, T. L. Marsh, H. Cheng, and L. J. Forney. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol*, 63(11):4516–4522, 1997. (Cited on page 25)

[105] P. McKenna, C. Hoffmann, N. Minkah, P. P. Aye, A. Lackner, Z. Liu, C. A. Lozupone, M. Hamady, R. Knight, and F. D. Bushman. The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS Pathog*, 4(2):e20, 2008. (Cited on page 65)

[106] R. Medzhitov. Recognition of microorganisms and activation of the immune response. *Nature*, 449(7164):819–826, 2007. (Cited on page 5)

[107] C. Miravitlles, C. Herr, I. Ferrarotti, R. Jardi, F. Rodriguez-Frias, M. Luisetti, and R. Bals. Laboratory testing of individuals with severe alpha1-antitrypsin deficiency in three European centres. *Eur Respir J*, 35(5):960–968, 2010. (Cited on page 2)

[108] H. Miyamoto, H. Yamamoto, K. Arima, J. Fujii, K. Maruta, K. Izu, T. Shiomori, and S. Yoshida. Development of a new seminested PCR method for detection of *Legionella* species and its application to surveillance of legionellae in hospital cooling tower water. *Appl Environ Microbiol*, 63(7):2489–2494, 1997. (Cited on pages 4 and 23)

[109] P. Morey, V. Cano, P. Marti-Lliteras, A. Lopez-Gomez, V. Regueiro, C. Saus, J. Bengoechea, and J. Garmendia. Evidence for a non-replicative intracellular stage of nontypable *Haemophilus influenzae* in epithelial cells. *Microbiology*, 157(1):234–250, 2011. (Cited on page 65)

[110] D. R. Murdoch. Molecular genetic methods in the diagnosis of lower respiratory tract infections. *APMIS*, 112(11-12):713–727, 2004. (Cited on page 61)

[111] E. Navarro, J. Escribano, J. Fernandez, and J. Solera. Comparison of three different PCR methods for detection of *Brucella* spp. in human blood samples. *FEMS Immunol Med Microbiol*, 34(2):147–151, 2002. (Cited on page 64)

[112] A. S. Neish, A. T. Gewirtz, H. Zeng, A. N. Young, M. E. Hobert, V. Karmali, and A. S. Rao. Prokaryotic regulation of epithelial responses by inhibition of IkappaB-alpha ubiquitination. *Science*, 289(5484):1560–1563, 2000. (Cited on page 7)

[113] P. D. N'Guessan, S. Hippenstiel, M. O. Etouem, J. Zahlten, W. Beermann, D. Lindner, B. Opitz, M. Witzenrath, S. Rosseau, N. Suttorp, and B. Schmeck. *Streptococcus pneumoniae* induced p38. *Am J Physiol Lung Cell Mol Physiol*, 290(6):1131–1138, 2006. (Cited on page 5)

[114] H. Nielsen. *Microbial ecology af activated sludge.* IWA Publishing, 2010. (Cited on page 12)

[115] R. O'Donnell, D. Breen, S. Wilson, and R. Djukanovic. Inflammatory cells in the airways in COPD. *Thorax*, 61(5):448–454, 2006. (Cited on page 2)

[116] E. A. Ottesen, J. W. Hong, S. R. Quake, and J. R. Leadbetter. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science*, 314(5804):1464–1467, 2006. (Cited on page 67)

[117] M. J. Pallen and B. W. Wren. Bacterial pathogenomics. *Nature*, 449(7164):835–842, 2007. (Cited on page 67)

[118] A. Paradiso, S. Tommasi, R. Pinto, G. E. Carpagnano, and M. P. Foschino-Barbaro. Exhaled breath condensate is not suitable to detect EGFR somatic mutations. *Eur Respir J*, 32(4):1126–1127, 2008. (Cited on page 4)

[119] P. D. Patrick and J. Handelsman. Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Appl Environ Microbiol*, 72(10):6773–6779, 2006. (Cited on page 33)

[120] G. P. Patrinos and W. Ansorge. *Molecular diagnostics.* Elsevier Academic Press, 2005. (Cited on page 11)

[121] J. F. Petrosino, S. Highlander, R. A. Luna, R. A. Gibbs, and J. Versalovic. Metagenomic pyrosequencing and microbial identification. *Clin Chem*, 55(5):856–866, 2009. (Cited on page 11)

[122] I. R. Poxton. Molecular techniques in the diagnosis and management of infectious diseases: do they have a role in bacteriology? *Med Princ Pract*, 14:20–26, 2005. (Cited on page 11)

[123] G. A. Preidis and J. Versalovic. Targeting the human microbiome with antibiotics, probiotics, and prebiotics: gastroenterology enters the metagenomics era. *Gastroenterology*, 136(6):2015–2031, 2009. (Cited on page 67)

[124] D. Preza, I. Olsen, J. A. Aas, T. Willumsen, B. Grinde, and B. J. Paster. Bacterial profiles of root caries in elderly patients. *J Clin Microbiol*, 46(6):2015–2021, 2008. (Cited on page 65)

[125] J. Qin, R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J. M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Dore, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, P. Bork, S. D. Ehrlich, and J. Wang. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65, 2010. (Cited on page 15)

[126] E. Rakhimova, L. Wiehlmann, A. L. Brauer, S. Sethi, T. F. Murphy, and B. Tuemmler. *Pseudomonas aeruginosa* population biology in chronic obstructive pulmonary disease. *J Infect Dis*, 200(12):1928–1935, 2009. (Cited on page 3)

[127] A. J. Ratner, E. S. Lysenko, M. N. Paul, and J. N. Weiser. Synergistic proinflammatory responses induced by polymicrobial colonization of epithelial surfaces. *Proc Natl Acad Sci U S A*, 102(9):3429–3434, 2005. (Cited on page 60)

[128] R. C. Read, R. Wilson, A. Rutman, V. Lund, H. C. Todd, A. P. Brain, P. K. Jeffery, and P. J. Cole. Interaction of nontypable *Haemophilus influenzae* with human respiratory mucosa in vitro. *J Infect Dis*, 163(3):549–558, 1991. (Cited on page 11)

[129] M. Rescigno, M. Urbano, B. Valzasina, M. Francolini, G. Rotta, R. Bonasio, F. Granucci, J. P. Kraehenbuhl, and P. Ricciardi-Castagnoli. Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nat Immunol*, 2(4):361–367, 2001. (Cited on page 6)

[130] T. H. Risby. Further discussion on breath condensate analysis. *Am J Respir Crit Care Med*, 167(10):1301–1302, 2003. (Cited on page 58)

[131] M. Ritter, D. Mennerich, A. Weith, and P. Seither. Characterization of Toll-like receptors in primary lung epithelial cells: strong impact of the TLR3 ligand poly(I:C) on the regulation of Toll-like receptors, adaptor proteins and inflammatory response. *J Inflamm*, 2 (6):16–16, 2005. (Cited on page 5)

[132] G. B. Rogers, M. P. Carroll, D. J. Serisier, P. M. Hockey, G. Jones, and K. D. Bruce. Characterization of bacterial community diversity in cystic fibrosis lung infections by use of 16S ribosomal DNA terminal restriction fragment length polymorphism profiling. *J Clin Microbiol*, 42(11):5176–5183, 2004. (Cited on page 13)

[133] G. B. Rogers, T. W. Daniels, A. Tuck, M. P. Carroll, G. J. Connett, G. J. David, and K. D. Bruce. Studying bacteria in respiratory specimens by using conventional and molecular microbiological approaches. *BMC Pulm Med*, 9:14, 2009. (Cited on page 11)

[134] G. Rohde, A. Wiethege, I. Borg, M. Kauth, T. T. Bauer, A. Gillissen, A. Bufe, and G. Schultze-Werninghaus. Respiratory viruses in exacerbations of chronic obstructive pulmonary disease requiring hospitalisation: a case-control study. *Thorax*, 58(1):37–42, 2003. (Cited on page 59)

[135] A. Rosell, E. Monso, N. Soler, F. Torres, J. Angrill, G. Riise, R. Zalaca, J. Morera, and A. Torres. Microbiologic determinants of exacerbation in chronic obstructive pulmonary disease. *Arch Intern Med*, 165(8):891–897, 2005. (Cited on page 4)

[136] J. L. Round, S. M. Lee, J. Li, G. Tran, B. Jabri, T. A. Chatila, and S. K. Mazmanian. The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science*, 332(6032):974–977, 2011. (Cited on page 62)

[137] K. Rudi, M. Zimonja, P. Trosvik, and T. Naes. Use of multivariate statistics for 16S rRNA gene analysis of microbial communities. *Int J Food Microbiol*, 120(1-2):95–99, 2007. (Cited on page 12)

[138] E. R. Sanders and J. H. Miller. *I, microbiologist: a discovery-based course in microbial ecology and molecular evolution.* ASM Press, 2010. (Cited on page 32)

[139] Y. Sanz, E. Saenchez, M. Marzotto, M. Calabuig, S. Torriani, and F. Dellaglio. Differences in faecal bacterial communities in coeliac and healthy children as detected by PCR and denaturing gradient gel electrophoresis. *FEMS Immunol Med Microbiol*, 51(3):562–568, 2007. (Cited on page 1)

[140] S. Sauer, A. Freiwald, T. Maie, M. Kube, R. Reinhardt, M. Kostrzewa, and K. Geider. Classification and identification of bacteria by mass spectrometry and computational analysis. *PLoS One*, 3(7):e2843, 2008. (Cited on page 16)

[141] S. Sauer and M. Kliem. Mass spectrometry tools for the classification and identification of bacteria. *Nat Rev Microbiol*, 8(1):74–82, 2010. (Cited on pages 15 and 16)

[142] L. Sbordone and C. Bortolaia. Oral microbial biofilms and plaque-related diseases: microbial communities and their role in the shift from oral health to disease. *Clin Oral Investig*, 7(4):181–188, 2003. (Cited on page 1)

[143] J. Schellenberg, M. G. Links, J. E. Hill, T. J. Dumonceaux, G. A. Peters, S. Tyler, T. B. Ball, A. Severini, and F. A. Plummer. Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. *Appl Environ Microbiol*, 75(9):2889–2898, 2009. (Cited on page 15)

[144] P. D. Schloss and J. Handelsman. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol*, 71(3):1501–1506, 2005. (Cited on page 33)

[145] U. M. E. Schoette, Z. Abdo, S. J. Bent, C. Shyu, C. J. Williams, J. D. Pierson, and L. J. Forney. Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Appl Microbiol Biotechnol*, 80(3):365–380, 2008. (Cited on page 13)

[146] A. J. Scupham, T. J. Patton, E. Bent, and D. O. Bayles. Comparison of the cecal microbiota of domestic and wild turkeys. *Microb Ecol*, 56(2):322–331, 2008. (Cited on page 12)

[147] S. Sethi. Bacterial infection and the pathogenesis of COPD. *Chest*, 117(5):286–291, 2000. (Cited on page 3)

[148] S. Sethi. Infection as a comorbidity of COPD. *Eur Respir J*, 35(6):1209–1215, 2010. (Cited on pages 2, 3, and 10)

[149] S. Sethi, N. Evans, B. J. B. Grant, and T. F. Murphy. New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. *N Engl J Med*, 347(7):465–471, 2002. (Cited on page 58)

[150] S. Sethi, J. Maloney, L. Grove, C. Wrona, and C. S. Berenson. Airway inflammation and bronchial bacterial colonization in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 173(9):991–998, 2006. (Cited on pages 58 and 59)

[151] S. Sethi and T. F. Murphy. Infection in the pathogenesis and course of chronic obstructive pulmonary disease. *N Engl J Med*, 359(22):2355–2365, 2008. (Cited on pages 58 and 59)

[152] S. Sethi, R. Sethi, K. Eschberger, P. Lobbins, X. Cai, B. J. B. Grant, and T. F. Murphy. Airway bacterial concentrations and exacerbations of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 176(4):356–361, 2007. (Cited on page 59)

[153] Q. Sha, A. Q. Truong-Tran, J. R. Plitt, L. A. Beck, and R. P. Schleimer. Activation of airway epithelial cells by toll-like receptor agonists. *Am J Respir Cell Mol Biol*, 31(3):358–364, 2004. (Cited on page 5)

[154] S. Sharma, R. Brousseau, and S. Kasatiya. Detection and confirmation of *Mycoplasma pneumoniae* in urogenital specimens by PCR. *J Clin Microbiol*, 36(1):277–280, 1998. (Cited on page 23)

[155] J. Shendure and H. Ji. Next-generation DNA sequencing. *Nat Biotechnol*, 26(10):1135–1145, 2008. (Cited on page 15)

[156] J. Shendure, R. D. Mitra, C. Varma, and G. M. Church. Advanced sequencing technologies: methods and goals. *Nat Rev Genet.*, 5(5):335–344, 2004. (Cited on page 15)

[157] Y. Shi, L. Chen, J. Tong, and C. Xu. Preliminary characterization of vaginal microbiota in healthy Chinese women using cultivation-independent methods. *J Obstet Gynaecol Res*, 35(3):525–532, 2009. (Cited on page 15)

[158] H. Slevogt, J. Seybold, K. N. Tiwari, A. C. Hocke, C. Jonatat, S. Dietel, S. Hippenstiel, B. B. Singer, S. Bachmann, N. Suttorp, and B. Opitz. *Moraxella catarrhalis* is internalized in respiratory epithelial cells by a trigger-like mechanism and initiates a TLR2- and partly NOD1-dependent inflammatory immune response. *Cell Microbiol*, 9(3):694–707, 2007. (Cited on pages 4 and 65)

[159] G. Soong, D. Parker, M. Magargee, and A. S. Prince. The type III toxins of *Pseudomonas aeruginosa* disrupt epithelial barrier function. *J Bacteriol*, 190(8):2814–2821, 2008. (Cited on page 3)

[160] S. Specter. *Rapid detection of infectious agents.* Plenum Press, 1998. (Cited on page 12)

[161] B. Stecher, R. Robbiani, A. W. Walker, A. M. Westendorf, M. Barthel, M. Kremer, S. Chaffron, M. I. Macpherson, J. Buer, J. Parkhill, G. Dougan, C. von Mering, and W. D. Hardt. *Salmonella enterica* serovar typhimurium exploits inflammation to compete with the intestinal microbiota. *PLoS Biol*, 5(10):2177–2189, 2007. (Cited on page 1)

[162] K. Steiling, J. Ryan, J. S. Brody, and A. Spira. The field of tissue injury in the lung and airway. *Cancer Prev Res*, 1(6):396–403, 2008. (Cited on page 3)

[163] R. Stockley. *Chronic obstructive pulmonary disease*. Blackwell Publishing Ltd, 2007. (Cited on pages 10 and 11)

[164] R. A. Stockley, C. O'Brien, A. Pye, and S. L. Hill. Relationship of sputum color to nature and outpatient management of acute exacerbations of COPD. *Chest*, 117(6):1638–1645, 2000. (Cited on page 59)

[165] F. A. Stressmann, G. B. Rogers, E. R. Klem, A. K. Lilley, S. H. Donaldson, T. W. Daniels, M. P. Carroll, N. Patel, B. Forbes, R. C. Boucher, M. C. Wolfgang, and K. D. Bruce. Analysis of the bacterial communities present in lungs of patients with cystic fibrosis from American and British centers. *J Clin Microbiol*, 49(1):281–291, 2011. (Cited on page 63)

[166] E. Stsepetova, E. Sepp, K. Julge, E. Vaughan, M. Mikelsaar, and W. M. Vos. Molecularly assessed shifts of *Bifidobacterium* ssp. and less diverse microbial communities are characteristic of 5-year-old allergic children. *FEMS Immunol Med Microbiol*, 51(2):260–269, 2007. (Cited on page 1)

[167] A. Suau, R. Bonnet, M. Sutren, J. J. Godon, G. R. Gibson, M. D. Collins, and J. Dor. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol*, 65(11):4799–4807, 1999. (Cited on page 12)

[168] S. J. Ho Sui, A. Fedynak, W. W. Hsiao, M. G. Langille, and F. S. Brinkman. The association of virulence factors with genomic islands. *PLoS One*, 4 (6)(12):8094, 2009. (Cited on page 67)

[169] A. Swidsinski. Standards for bacterial identification by fluorescence textitin situ hybridization within eukaryotic tissue using ribosomal rRNA-based probes. *Inflamm Bowel Dis*, 12(8):824–6, 2006. (Cited on page 65)

[170] A. Sykes, P. Mallia, and S. L. Johnston. Diagnosis of pathogens in exacerbations of chronic obstructive pulmonary disease. *Proc Am Thorac Soc*, 4(8):642–646, 2007. (Cited on pages 4 and 58)

[171] J. Treimo, G. Vegarud, T. Langsrud, S. Marki, and K. Rudi. Total bacterial and species-specific 16S rDNA micro-array quantification of complex samples. *J Appl Microbiol*, 100(5):985–998, 2006. (Cited on page 16)

[172] R. M. Tuder, S. M. Janciauskiene, and I. Petrache. Lung disease associated with alpha1-antitrypsin deficiency. *Proc Am Thorac Soc*, 7(6):381–386, 2010. (Cited on page 2)

[173] A. C. van de Pol, A. M. van Loon, T. F. W. Wolfs, N. J. G. Jansen, M. Nijhuis, E. K. Breteler, R. Schuurman, and J. W. A. Rossen. Increased detection of respiratory syncytial virus, influenza viruses, parainfluenza viruses, and adenoviruses with real-time PCR in samples from patients with respiratory symptoms. *J Clin Microbiol*, 45(7):2260–2262, 2007. (Cited on page 23)

[174] R. E. Vance and R. R. Isberg. Patterns of pathogenesis: discrimination of pathogenic and nonpathogenic microbes by the innate immune system. *Cell Host Microbe*, 6(1):10–21, 2009. (Cited on page 5)

[175] S. B. Veeramachaneni and S. Sethi. Pathogenesis of bacterial exacerbations of COPD. *COPD*, 3(2):109–115, 2006. (Cited on page 3)

[176] C. Vogelberg, T. Hirsch, A. Roesen-Wolff, M.-L. Kerkmann, and W. Leupold. *Pseudomonas aeruginosa* and *Burkholderia cepacia* cannot be detected by PCR in the breath condensate of patients with cystic fibrosis. *Pediatr Pulmonol*, 36(4):348–352, 2003. (Cited on page 59)

[177] J. A. Wedzicha. Exacerbations: etiology and pathophysiologic mechanisms. *Chest*, 121(5):136S–141S, 2002. (Cited on page 2)

[178] D.G. Wilkinson. *In situ hybridization*. Oxford University Press, 1999. (Cited on page 33)

[179] T. M. A. Wilkinson, J. R. Hurst, W. R. Perera, M. Wilks, G. C. Donaldson, and J. A. Wedzicha. Effect of interactions between lower airway bacterial and rhinoviral infection in exacerbations of COPD. *Chest*, 129(2):317–324, 2006. (Cited on pages 59 and 60)

[180] M. Wilson. *Microbial inhabitants of humans*. Cambridge University Press, 2004. (Cited on pages 1, 8, and 9)

[181] P. L. Worthen, C. J. Gode, and J. Graf. Culture-independent characterization of the digestive-tract microbiota of the medicinal leech reveals a tripartite symbiosis. *Appl Environ Microbiol*, 72(7):4775–4781, 2006. (Cited on page 11)

[182] M. Xiang and J. Fan. Pattern recognition receptor-dependent mechanisms of acute lung injury. *Mol Med*, 16(1-2):69–82, 2010. (Cited on page 5)

[183] J. Xiong. *Essential Bioinformatics*. Cambridge University Press, 2006. (Cited on page 49)

[184] T. Yamamoto, X. Zhou, C. J. Williams, A. Hochwalt, and L. J. Forney. Bacterial populations in the vaginas of healthy adolescent women. *J Pediatr Adolesc Gynecol*, 22(1):11–18, 2009. (Cited on page 15)

[185] T. Zakharkina, A.-R. Koczulla, O. Mardanova, A. Hattesohl, and R. Bals. Detection of microorganisms in exhaled breath condensate during acute exacerbations of COPD. *Respirology*, 16(6):932–938, 2011. (Cited on page 59)

[186] W. Zetterquist, H. Marteus, P. Kalm-Stephens, E. Nes, L. Nordvall, M. Johannesson, and K. Alving. Oral bacteria–the missing link to ambiguous findings of exhaled nitrogen oxides in cystic fibrosis. *Respir Med*, 103(2):187–193, 2009. (Cited on page 60)

[187] M. Zhang, Q. Li, X-Y. Zhang, X. Ding, D. Zhu, and X. Zhou. Relevance of lower airway bacterial colonization, airway inflammation, and pulmonary function in the stable stage of chronic obstructive pulmonary disease. *Eur J Clin Microbiol Infect Dis*, 29(12):1487–1493, 2010. (Cited on page 10)

[188] M. Zhang, B. Liu, Y. Zhang, H. Wei, Y. Lei, and L. Zhao. Structural shifts of mucosa-associated *Lactobacilli* and *Clostridium leptum* subgroup in patients with ulcerative colitis. *J Clin Microbiol*, 45(2):496–500, 2007. (Cited on page 1)

[189] X. Zhou, C. J. Brown, Z. Abdo, C. C. Davis, M. A. Hansmann, P. Joyce, J. A. Foster, and L. J. Forney. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. *ISME J*, 1(2):121–133, 2007. (Cited on page 65)

7 Attachment

- Patient information sheet (in German)
- Patient consent form (in German)
- Personnel records sheet (in German)
- Publications and presentations
- Academic teachers
- Acknowledgements
- Statement of original authorship (Ehrenwörtliche Erklärung)

Patientenaufklärung, COPD Bakteria, Version HOM 5 - 2010

Homburg, am 16 / 06 / 2010

Patienteninformation zur Studie

„Untersuchung des Einflusses chronischer bakterieller Kolonisation auf den Krankheitsverlauf in Patienten mit Asthma und chronisch obstruktiver Lungenerkrankung (COPD)"

Studienleiter: Prof. Dr. Dr. Robert Bals an der Klinik für Innere Medizin,

Schwerpunkt Pneumologie des Universitätsklinikums des Saarlandes

Kirrbergerstr. 1, 66421 Homburg

Sehr geehrte Patientin, sehr geehrter Patient,

Dieses Informationsblatt soll dazu diesen, Sie über die bevorstehende Studie zu informieren und Sie um Ihre Teilnahme zu bitten. Bevor Sie sich entscheiden, an dieser Studie teilzunehmen, lesen Sie bitte dieses Informationsblatt sorgfältig durch. Es beschreibt Einzelheiten der Studie und was von Ihnen bei einer Teilnahme der Studie erwartet werden würde.

Worum geht es bei dieser Studie ?

Patienten mit Asthma oder der chronisch obstruktiven Lungenerkrankung (COPD) sind oft von Infektion der Atemwege betroffen. Auch führen diese Infektionen dazu, dass sich die Erkrankung verschlechtert.

Im Rahmen der Studie soll untersucht werden, ob die Atemwege von Patienten mit Asthma oder COPD mit Bakterien besiedelt sind. Dies soll mit Ergebnissen von Personen verglichen werden, die nicht an Asthma oder COPD erkrankt sind. Dazu sollen neue Methoden verwendet werden, da bisher verwendete Verfahren nicht empfindlich genug sind.

Was kommt auf Sie zu ?

Aus medizinischen Gründen werden Sie sicCurriculum vitaeh einer Bronchoskopie mit Lungenspülung unterziehen. Das überschüssige Material, das nicht für die Diagnostik verwendet wird, verwenden wir für die Studie. Zur Untersuchung der Entzündungsaktivität im Körper ist es notwendig, dass Sie einen Fragebogen ausfüllen. Weiterhin benötigen wir die Daten einer aktuellen Lungenfunktion.

Welche Nebenwirkungen sind möglich ?
Alle Maßnahmen sind risikofrei oder risikoarm.

Können Sie aus dieser Studie wieder ausscheiden ?
Die Teilnahme an der Studie ist freiwillig. Sie können jederzeit und ohne Angabe von Gründen Ihre Einwilligung zurückziehen und die Studie abbrechen, ohne dass Ihnen Nachteile bei der späteren medizinischen Versorgung entstehen.

Wie vertraulich werden die Daten behandelt ?
Die im Rahmen der Studie aufgezeichneten Daten sowie die Untersuchungsergebnisse werden ohne Ihre persönlichen Daten, d.h. anonymisiert, ausgewertet. Eine Offenlegung oder Übermittlung Ihrer Daten findet nicht statt. Beim Umgang mit den erhobenen Daten werden alle Grundsätze des Datenschutzes gewahrt.

Falls Sie Interesse haben, erklären wir Ihnen das Vorgehen gerne im Detail. Wir teilen Ihnen auch gerne die Ergebnisse unserer Untersuchungen mit.

Falls Sie weitere Fragen haben sollten, können Sie jederzeit den Studienleiter, Herrn Prof. Dr. Dr. Robert Bals, erreichen (Tel. 06841 16 2600).

Prof. Dr. med. Dr. rer. nat. Robert Bals
Klinik für Innere Medizin V - Pneumologie
Tel. +49 (0) 6421 286 4994

Patienteneinwilligung, COPD Bakteria

Version vom 2 - 2006

Einwilligung

Ich, _____ , geb. am _____

wohnhaft in _____ _____ _____
PLZ Ort Telefon

wurde von Herrn / Frau _____ über Ziel und Ablauf der Studie:

„Untersuchung des Einflusses chronischer bakterieller Kolonisation auf den Krankheitsverlauf in Patienten mit Asthma und chronisch obstruktiver Lungenerkrankung (COPD)“

umfassend aufgeklärt. Die Aufklärung wurde mir zusätzlich schriftlich in Form einer Patienteninformation übergeben. Ich habe den Inhalt der Aufklärung verstanden und zur Kenntnis genommen, dass ich bei weiteren Fragen jederzeit meinen behandelnden Arzt oder den Studienleiter (Dr. Dr. Robert Bals - an der Klinik für Innere Medizin, Schwerpunkt Pneumologie des Universitätsklinikums des Saarlandes, 66421 Homburg, Tel +49 (0) 64841 16 23600) ansprechen kann. Des Weiteren ist mir bekannt, dass ich die Untersuchung jederzeit, ohne Nennung von Gründen und ohne persönliche Nachteile, beenden kann. Ich werde dies meinem behandelnden Arzt sofort mitteilen. Mir ist bekannt, dass für diese Studie keinerlei personenbezogene Daten gesammelt oder gespeichert werden, die Rückschlüssen über meine Person erlauben würden. Hiermit erkläre ich meine Einwilligung zur freiwilligen Teilnahme an der Untersuchung.

_____ _____
Ort, Datum Unterschrift des Patienten

_____ _____
Ort, Datum Aufklärender Arzt

Erfassungsbogen 16s Heidelberg

Seite 1, 10-10-2007

Erfassungsbogen:

Nummer

Raucher

Ja _____ Packungsjahre

Nein

COPD

Ja _____ Erstdiagnose

Nein

Infektionsparameter

CRP:

Leukozyten:

Lungenfunktion

FEV 1:

VC:

FVC:

TLC:

SR:

Blutgase

pH:

pO_2:

pCO_2:

Weitere Diagnosen

Aktuelle Medikation

Durchgeführte / geplante Operation

Publications and presentations

Parts of this work have been revealed in the following publications:

1. T. Zakharkina, E. Heinzel, J. Pauling, J. Baumbach, and R. Bals. Standard lung microbiota in healthy individuals and COPD patients (in preparation)

2. T. Zakharkina and R. Bals. The Role of Pathogenic and Commensal Bacteria in Airway Epithelial Barrier Homeostasis (in preparation)

3. Zakharkina T., Koczulla R., Mardanova O., Hattesohl A., and Bals R. Detection of microorganisms in exhaled breath condensate during acute exacerbations of COPD. *Respirology*, 16: 932–8, 2011

4. C. Herr, T. Greulich, R. A. Koczulla, S. Meyer, T. Zakharkina, M. Branscheidt, R. Eschmann, and R. Bals. The role of vitamin D in pulmonary disease: COPD, asthma, infection, and cancer. *Respir Res*, 18: 12–31, 2011

5. C. Hess, C. Herr, C. Beisswenger, T. Zakharkina, R. M. Schmid, and R. Bals. Myeloid RelA regulates pulmonary host defense networks. *Eur Respir J*, 35: 343–52, 2010

Key presentations (related to this work):

1. 51st Congress of the German Society for Pneumology and Ventilatory Support (Hannover, Germany, March, 2010)

2. American Thoracic Society International Conference (San Diego, USA, May, 2009)

3. European Respiratory Society Annual Congress (Vienna, Austria, September, 2009)

4. 50th Congress of the German Society for Pneumology and Ventilatory Support (Mannheim, Germany, March, 2009)

5. Annual Meeting of the Cell Biology Section of the German Society for Pneumology and Ventilatory Support (Freiburg, Germany, November, 2008)

6. Summer School “Pathogen-Host Interactions at Cellular Barriers” (Muenster, Germany, June, 2008)

Academic teachers

Odesa

Andrievsky
Aphonin
Buchtiyarov
Burdenyuk
Dzhurtubaeva
Elinskaya
Galkin
Haustova
Ivanitsa
Kivganov
Kolomyichuk
Korovin
Lisyutin
Medvedeva
Panchenko
Petrov
Philippova
Prosyanyuk
Rakhimova
Rusakova
Stoilovsky
Zaporozhchenko

Marburg

Bals
Bauer
Fehrenbach
Garn
Gemsa
Lohoff
Renz
Vogelmeier

Homburg

Bals
Baumbach
Gortner
Heinzel
Herrmann
Tschernig

Acknowledgements

This thesis would not have been possible unless my supervisor Prof. Dr. Dr. Robert Bals (Faculty of Medicine, University of Saarland) supported me from the preliminary to the concluding level enabling me to develop an understanding of the subject. It was a great pleasure to work in his group: Prof. Dr. Dr. Bals created a friendly atmosphere for producing and publishing scientific results. Thank you for being a great mentor!

My deepest gratitude goes to Dr. Christian Herr (Faculty of Medicine, University of Saarland) for many productive scientific discussions and critical appraisal of my ideas.

I thank Dr. Dr. Olaf Pinkenburg (Institute for Immunology, Philipps University of Marburg) for providing the protocol for obtaining ultra competent *E. coli* cells, for excellent and skillful assistance with many of cloning experiments.

I owe my deepest gratitude to PD Dr. Werner Liesack (Max-Planck-Institute for Terrestrial Microbiology, Marburg) for sharing of protocol for fluorescence *in situ* hybridization. Discussions with Prof. Dr. Michael W. Friedrich (Faculty of Biology, University of Bremen) were most helpful during the optimization of the T-RFLP procedure.

I also would like to make a special reference to Akira Hattesohl (Faculty of Medicine, Philipps University of Marburg) and Josch Pauling (Max-Planck-Institute for Informatics, Saarbruecken). Thank you for your immense support from the very beginning, for your help with software usage and statistical evaluation of results. Suggestions of Dr. Siarhei Khirevich (Faculty of Chemistry, Philipps University of Marburg) in typewriting were very helpful. I'd like to thank also Mary Neville for the proofreading of this thesis.

Many thanks also to Claudia Trier and Johannes Sierigk and many more colleagues from the Biomedical Research Center in Marburg for contribution to the publication but also for hosting cheerful lunch breaks.

I wish to express my sincerest thanks to all the people that supported me in the lab, especially Dr. Elke Heinzel, Lena Arnold, Annette Püchner, Nicole Reuter, Thomas Damm and many more. Thank you for being both a great help and entertaining company!

Finally, the deepest thanks go to my dear family and to my friends for always supporting, encouraging and loving me. Without you, this dissertation would not have been possible.

Homburg/Saar, ...

Tetyana Zakharkina

Erklärung gemäß § 7 Abs. 1 Nr. 4

Ich erkläre hiermit an Eides statt, dassich die vorliegende Arbeit ohne un zulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten presonen in der jeweils beschriebenen Weise unentgeltlich geholfen:

1. Prof. Dr. Dr. Bals — general support in all experiments
2. PD Dr. Werner Liesack — fluorescence *in situ* hybridization
3. Dr. Dr. Olaf Pinkenburg — competent *E. coli* cells
4. Dr. Ali Önder Ildirim — mice lung tissue sections
5. Dr. Elke Heinzel — T-RFLP analysis
6. Josch Pauling — principal component analysis

Weitere Personen waren an der inhaltlich-materiallen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater / innen oder anderer Personen) in Anspruch genommen. Außer den Angegebenen hat niemand von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form in einem anderen Verfahren zur Erlangung des Doktorgrades einer anderen Prüfungsbehörde vorgelegt.

Ich versichere an Eides statt, dass ich nach besten Wissen die Wahrheit gesagt und nichts verschwiegen habe.

Vor Aufnahme der vorstehenden Versicherung an Eides Statt wurde ich über die Bedeutung einer eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung belehrt.

Ort, Datum

Tetyana Zakharkina