Lisa Oberauner-Wappis, Bakk.Biol. MSc.

Metagenomics: New insights into bioresources of extreme environments

DISSERTATION

Zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften

eingereicht an der

Technischen Universität Graz

Betreuerin

Univ.-Prof. Dipl.-Biol. Dr. rer. nat. Gabriele Berg

Institut für Umweltbiotechnologie

Graz, Juli 2014

Table of Contents

Abstract 1

Zusammenfassung 3

1. Introduction 5

  1.1. Metagenomic-based analyses of microbial communities 5

    1.1.1. Detection of the previous uncultured majority 5

    1.1.2. Common metagenomic methods and bioinformatic strategies 6

  1.2. Extreme environments and their microbial life 8

    1.2.1.Indoor microbiome – influence and implications to human health 8

    1.2.2. Functional potential of the *Sphagnum*-dominated bog ecosystem 10

2. Methodical approach 13

3. Objectives of the work 14

4. Results and discussion 15

  4.1. Analysis of the bacterial communities associated with the indoor hospital environment 15

  4.2. Analysis of taxonomic and functional composition within the *Sphagnum* microbiome 18

  4.3. Conclusion 20

5. References 22

6. Publication 26

  The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing

7. Book Chapter 43

  Complex indoor communities: Bacterial life under extreme conditions in clean rooms and intensive care units

8. Manuscript I 50

  The *Sphagnum* microbiome supports greatly bog ecosystem functioning under extreme conditions

9. Manuscript II 87

  Mining for bioactive compounds: analysis of NRPS and PKS genes in the Sphagnum-bog metagenome

10. Curriculum Vitae 104

11. Publication List 105

12. Acknowledgements 107

13. Statutory Declaration 108

Abstract

Microorganisms are the most abundant organisms on Earth representing a major reservoir of genetic diversity. For a long time, analysis of microorganisms was limited by cultivation techniques but now metagenomics, a DNA/RNA-based approach, was developed to enable a comprehensive analysis of microbial communities in different ecosystems. In this study, diverse metagenomic methods were applied to detect the taxonomic and functional diversity of two environments, which provide extreme conditions for microbial communities: i) indoor and ii) outdoor microbiomes.

The composition of the indoor microbiome, especially in intensive care units (ICUs) of hospitals, plays an important role in everyday human health. Amplicon pyrosequencing revealed seven different bacterial phyla: *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Firmicutes*, *Nitrospira* and *Proteobacteria*. At genus level, the amplicon library of the ICU included 405 genera, where 76 of them were reaching 1% relative abundance. Species associated with the outside environment, taxa closely related to potential human pathogens as well as beneficials were detected within the hospital environment. Network and Principal coordinate plot analysis based on amplicon libraries showed significant different bacterial area profiles for floors, medical devices and workplaces. However, identical molecular fingerprints offered bacterial similarities and indicated a transmission of microorganisms between sampling sites. Interestingly, only 2.5% of bacterial diversity was detected by the currently used standard cultivation approach compared to our 454-pyrosequencing results. This shows the essential integration of DNA-based approaches in monitoring systems.

The outdoor microbial communities of *Sphagnum*-dominated bogs, which are also characterised by extreme exterior conditions, are of importance in carbon storage and stabilization of climate world-wide. The role of *Sphagnum*-associated microbial communities in the bog ecosystem, which was assessed by applying an Illumina-based approach followed by *de novo* assembly and MG-RAST annotation, revealed that the *Sphagnum* microbiome carries an essential genetic potential for sustainable functioning in association with the host plants. Highly abundant subsystems responsible for oxidative and drought stress, genetic exchange, repair or resistance realised a plasticity-stability balance. Multiple interactions among each other and plants were indicated by diverse genes necessary for quorum sensing, biofilm formation and nutrient exchange. *Sphagnum* mosses are colonised by highly diverse microbial communities. Moreover, 16S rDNA analysis within the metagenomic approach indicated a higher structural diversity than PCR-dependent techniques, where *Proteobacteria* (65.8%) followed by *Acidobacteria* (11.4%), *Actinobacteria* (5.6%) were the dominant phyla. An inter-environmental comparison revealed that the *Sphagnum* microbiome harbours highly specific genetic features that distinguish it significantly from comparable microbiomes of higher plants and peat soils. It is well known that the *Sphagnum* microbiome produces bioactive secondary metabolites and provides a huge reservoir of chemically diverse natural products.

Due to the high diversity and the extraordinary metabolic capacity of *Sphagnum*-associated communities, the *Sphagnum* metagenome was selected for a deeper analysis. *Sphagnum* moss-associated bacteria were screened for presence of non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs). Both enzymes as well as their hybrids were successfully detected in the microbiome by PCR-screening and *in silico* analysis. Screening of a metagenomic fosmid library revealed the presence of gene sequences displaying high homology to genera of *Pseudomonas* and *Pectobacterium*. Simultaneous Illumina-based *in silico* analysis resulted in 328 NRPS, 456 PKS as well as 57 of their hybrids, where a high number of sequences shared significant similarity to genera of *Streptomyces*, *Paenibacillus, Mycobacteria* and *Lysobacter* within the *Sphagnum* microbiome.

Overall, it was demonstrated that extreme environments are colonised by specific microbial communities. Comparative metagenomic approaches clarify the unique positions of these ecosystems as promising bio-resource and serve as a source for biotechnological applications.

Zusammenfassung

Mikroorganismen sind die am häufigsten vorkommenden Organismen weltweit und besitzen ein großes Reservoir an genetischer Vielfalt. Durch kultivierungs-abhängige Methoden war die Analyse dieser Mikroorgansimen lange Zeit limitiert. „Metagenomics“, eine DNA/RNA basierte Methode, wurde entwickelt, um die mikrobielle Gemeinschaft in ihren unterschiedlichsten Ökosystemen zu erforschen. In dieser Arbeit wurde die taxonomische und funktionelle Diversität zweier Habitate untersucht, die extreme Bedingungen für Mikroorgansimen darstellen: i) indoor und ii) outdoor Mikrobiome.

Die mikrobielle Gemeinschaft im indoor-Bereich, speziell auf Intensivstationen in Krankenhäusern, spielt eine sehr wichtige Rolle für die menschliche Gesundheit. Mittels Amplicon Sequenzierung wurden sieben unterschiedliche bakterielle Phyla auf der Intensivstation identifiziert: *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Firmicutes*, *Nitrospira* und *Proteobacteria*. Auf Gattungsebene wurden insgesamt 405 Bakterien ermittelt, von denen 76 die 1%-Häufigkeitsgrenze erreichten. In der Intensivstation wurden bakterielle Arten, die aus der Umwelt ins Krankenhaus mittransportiert wurden, potentielle Humanpathogene und zahlreichen Nützlingen, sogenannte „beneficials“, gefunden. Netzwerk- und Principal coordinate Analysen zeigten signifikante Unterschiede in der bakteriellen Gemeinschaft von Bodenproben, medizinischen Geräten und Arbeitsflächen. Zusätzlich zeigten identische Fingerprints der Bakterien deutlich eine Übertragung von Mikroorganismen zwischen den Probenahmestellen. Interessant war auch der Vergleich der 454-Pyrosequenzierung mit der angewandten kultivierungs-abhängigen Methode, bei der nur 2,5% der bakteriellen Diversität erfasst werden konnte. Diese Ergebnisse zeigten deutlich, dass DNA-basierte Methoden essentiell für die Bestimmung bakterieller Gemeinschaften in Überwachungssystemen sind.

Die mikrobielle Gemeinschaft von *Sphagnum*-dominierten Moorlandschaften, ebenfalls charakterisiert durch extreme Bedingungen, ist sehr wichtig für den globalen Kohlenstoffkreislauf und für die Klimaentwicklung weltweit. Durch Illumina Sequenzierung und anschließender MG-RAST-Auswertung wurde für das *Sphagnum*-Mikrobiom ein großes genetisches Potential gezeigt, welches in Assoziation mit den Wirtspflanzen essentielle Funktionen für Nachhaltigkeit aufweist.

Häufig vorkommende Subsysteme die für oxidativen Stress, Trockenstress, genetischen Austausch, Reparatur oder Resistenz verantwortlich sind, machen eine Plastizität-Stabilität-Balance erkennbar. Multiple Wechselwirkungen innerhalb dieser Subsysteme und zwischen Pflanzen wurden von verschiedensten Genen bekräftigt, die für Quorum Sensing, Biofilmbildung und Nährstoffaustausch notwendig sind. *Sphagnum*-Moose sind von sehr unterschiedlichen mikrobiellen Gemeinschaften kolonisiert. Eine auf dem Metagenom-Datensatz basierende 16S rRNA Analyse zeigte eine im Vergleich mit PCR-abhängigen Techniken höhere strukturelle Diversität, bei der *Proteobacteria* (65,8%) am häufigsten vorkamen, gefolgt von *Acidobacteria* (11,4%) und *Actinobacteria* (5,6%). Das *Sphagnum*-Mikrobiom zeigte spezifische, genetische Eigenschaften die sich signifikant von

vergleichbaren Mikrobiomen höherer Pflanzen und Torfboden unterschieden. Das *Sphagnum*-Mikrobiom ist bekannt für seine Sekundärmetaboliten und stellt ein großes Reservoir an chemisch verschiedenen Naturprodukten dar.

Aufgrund der bekannten hohen Diversität und der außergewöhnlichen metabolischen Kapazität von *Sphagnum*-assoziierten Gemeinschaften, wurde das *Sphagnum* Metagenom für weitere Analysen gewählt. Dabei wurden die assoziierten Bakterien auf das Vorkommen von non-ribosomal peptide synthetases (NRPSs) und polyketide synthases (PKSs) untersucht. Mittels PCR-screening und einer *in silico* Analyse wurden beide Enzyme und auch ihre Hybride erfolgreich identifiziert. Beim Screening einer Fosmid-Bibliothek wurden Sequenzen mit einer hohen Homologie zu den Bakterien der Gattung *Pseudomonas* und *Pectobacterium* festgestellt. Die *in silico* Auswertung erfasste 328 NRPS, 456 PKS und 57 Hybride, von denen eine große Sequenzanzahl den Gattungen *Streptomyces*, *Paenibacillus*, *Mycobacteria* und *Lysobacter* zugeordnet werden konnte.

Allumfassend wurde gezeigt, dass extreme Habitate durch ihre spezifischen mikrobiellen Gemeinschaften gekennzeichnet sind. Vergleichbare Metagenom-Analysen verdeutlichten die einzigartige Position dieser Ökosysteme als vielversprechende Bioressource und dienten als wichtige Quelle für biotechnologische Anwendungen.

1. Introduction

1.1. Metagenomic-based analyses of microbial communities

Microbes are the most abundant organisms on Earth and play an important role in ecosystems such as soil, water, and air right up to indoor environment. The total number of microbial cells on Earth is estimated to be $4 \cdot 10^{30}$-$6 \cdot 10^{30}$, comprising more than 10^6 different species (distinct taxonomic groups based on gene sequence analysis) within more than 70 phyla (Whitman *et al*., 1998; Curtis *et al*., 2002; Pace *et al*., 2009). It is widely known that the genomes of microorganisms represent a major pool of genetic diversity (Whitman *et al*., 1998; Ferrer *et al*., 2009) as well as a largely untapped reservoir of novel enzymes and metabolic capabilities of uncultured species (Rappe and Giovannoni, 2003). New possibilities of metagenomics analysis bypass the need for isolation or cultivation of microorganisms.

1.1.1. Detection of the previous uncultured majority

Microbiology was changing during the last 25 years altering microbiologists' view of microorganisms and how to study them. Traditional microbiologists were always dependent on culture-based methods for the identification of microbes in environmental samples. The challenge to identify and characterize uncultured organisms began 1960 to mid-1980s, where scientists considered that cultured microorganisms did not represent the whole microbial world (Staley and Konopka, 1985; Handelsman, 2004). "The plate count anomaly", the discrepancy in the microbial number between dilution plating and microscopy was the main indicator for the uncultured world of microbes and rethinking in microbial sciences (Staley and Konopka, 1985). Such observations demonstrated that in natural samples less than one cell in a thousand produces a colony (Torsvik *et al*., 1990). From now on, non-culture based approaches have been developed and used for wide-ranging analysis of different communities in a microbial environment.

The first milestone of metagenomic analysis was set by an idea from Pace (1985) and colleagues to propose the direct cloning of environmental DNA. The next technical scientific breakthrough was the development of the PCR technology and the design of primers that can be used to amplify entire genes. The 16S rRNA gene was defined as a marker for taxonomic analysis (Woese, 1987) and used as a tool for bacterial diversity analysis (Schmidt, 1991) changing radically the understanding of the microbial world. Further technical developments have usher in a new metagenomic era (Handelsman, 2004; Chistoserdova, 2010), where total

DNA from environmental samples (eDNA) was applied for direct sequencing. This direct sequence analysis of eDNA is currently considered the most utilized method for assessing the structure of an environmental microbial community. The term "metagenomics" has become acquainted by Handelsman and his group within their study of natural products from soil microbes (Handelsman, 2004). Next generation sequencing methods and bioinformatic tools are able to detect a significant amount of novel species as well as functional genes and metabolic pathways.

Metagenomics, referred also as community genomics, environmental genomics or population genomics, is a powerful tool for comparing and exploring the ecology. In nature, polymicrobial interactions occur between bacteria, fungi, viruses or archaea. Current metagenomic methods allow the detection and exploitation of the taxonomic and also the metabolic diversity within these microbial communities and are an aspiring field of research compared to other methods.

1.1.2. Common metagenomic methods and bioinformatic strategies

One of the primary methods was to construct metagenomic libraries by isolation of high quality DNA that was suitable for cloning and covered their microbial diversity (Simon and Daniel, 2001). Preparing such a library is highly time-consuming, including the main step of ligation of restriction-digested or blunt-ended metagenomic DNA into vectors. As example, large insert and particularly fosmids have been very common for metagenomic studies due to their high cloning efficiency and the improved stability in *Escherichia coli* (Ghai *et al*., 2010). Fosmids are large insert cloning vectors with the capabilities of holding up to 40 kb inserts of contiguous genomic sequences from microorganisms without requiring prior cultivation (Shizuya *et al*., 1992; Simon and Daniel, 2011). Sequence-based and function-based screenings were used to identify genes of interest within a metagenomic sample. The following Sanger sequencing (Sanger *et al*., 1977) has almost exclusively been carried out and was the most applied method during this time. Moreover, the majority of biomolecules is derived from metagenomic libraries which have been constructed from temperate soil samples (Lorenz and Eck, 2005; Sjöling and Cowan, 2008).

Over the past ten years, alternative sequencing platforms have become widely available and genome sequencing capabilities have expanded exponentially. Recently, several analysing methods including shotgun sequencing have been used in metagenomic studies (Metzker, 2010). The next generation technologies of 454/Roche and Illumina/Solexa systems are nowadays extensively applied in metagenomic research, even in this work. The 454/Roche system is based on individual and parallel pyrosequencing and the Illumina/Solexa technology is a sequencing-by-synthesis process running on a HiSeq instrument. At the moment, the lower costs and recent success in its application to metagenomics make the Illumina technology an increasingly popular choice compared to 454-pyrosequencing approaches (Thomas *et al.*, 2012). Many bioinformatic analysing tools are now available for assembly, binning and annotation of metagenomic datasets (Neelakanta and Sultana, 2013). These new methods generate an enormous amount of datasets and it is evermore important to deposit these large datasets into databases. Services, like MG-RAST (Meyer *et al.*, 2008) as a prominent database source, are public available for large-scale metagenomic analysis. Such databases simplify the handling and analysis of taxonomic and functional composition of microbial environments.

Hitherto, metagenomics has provided significant information into the microbial community. Analyses using high throughput sequencing or library construction have been very important for describing microbial structure and functionality in different ecosystems and for identifying novel genes. Studies from several habitats from arctic tundra, marine environment to animals have yielded microbial enzymes with potential for biocatalytic applications (Adrio and Demain, 2014). Especially the soil ecosystem is an important reservoir for the discovery of novel microbial enzymes. Thus, many microbial enzymes could be identified from various metagenomic soils studies as well as from extreme environments like amylases, cellulases, esterases, lipases, proteases or xylanases (Uchiyama and Miyazaki, 2009; Liszka *et al.,* 2012; Lee and Lee, 2013; Adrio and Demain, 2014). In addition to the novel enzymes, environments are rich sources of a variety of small molecules with bioactivities, such as antibiotics and other pharmaceutically applicable activities including polyketides, turbomycins, glycopeptides or cyanobactins (Iqbal *et al.*, 2012). Looking forward, new approaches in the engineering of enzymes (e.g. data mining) by using computational design methods are getting more important to identify possible candidates for further characterisation (Delmont *et al.*, 2011; Thomas *et al.*, 2012).

To sum up, the applications of next generation sequencing seem almost endless. Metagenomics is an important tool to address fundamental questions of microbial ecology, evolution and diversity. These new technologies and applications allow us to access the majority of uncultured microorganisms to get new insights into still unknown bioresources.

1.2. Extreme environments and their microbial life

Microorganisms exist in different environments with high diversity all around the world. They have been discovered in habitats like soil, water, alkaline and hot springs, deserts as well as indoor habitats (Simon and Daniel, 2011; Liszka *et al.*, 2012). These environments provide extreme conditions for life and their colonised microbes, which have adapted to temperatures, drought, humidity, pressures or alkalinity/acidity. Microorganisms must have special enzymes that function under strict conditions to survive in their environment. These enzymes found in such environments (e.g. high/low temperature) are typically more tolerant of other conditions (e.g. organic solvents); thus, naturally occurring robust enzymes can be used or evolved for use in a variety of harsh environments (Liszka *et al.*, 2012). Liszka and his colleagues (2012) also reported that many industrially relevant enzymes have been isolated from organisms growing at high temperature, high salt concentration, or in environments contaminated with organic solvents, that are significant challenges and limitations in bioprospecting for extremophilic enzymes. Microorganisms, which are able to adapt and colonise in harsh environments comprise special properties of comprehensive interest for biological scientists as well as for biotechnological industry. This study is focused mainly on detecting of microbial composition ant its potential within the habitat of two extreme environments: an indoor microbiome and a *Sphagnum* dominated Alpine bog.

1.2.1. Indoor microbiome – influence and implications to human health

The majority of our lifetime is spent indoor such as home, work place or public buildings, but our knowledge of microbial diversity is limited. This provides new habitats and residence to numerous microbial communities comprising many individual bacterial species. Over the last years, the study of indoor environments has raised more and more interest. Latest cultivation-based studies detected potentially pathogenic and allergenic indoor microorganisms (Täubel *et al.*, 2009; Yamamoto *et al.*, 2011), but less is known about the real indoor microbial diversity. Recently, next generation sequencing methods gave us the possibility to get a deep

and valuable insight of the indoor microbiome, revealing a generally high prokaryotic diversity including diverse bacterial, archaeal and fungal phyla (Flores *et al*., 2011, 2013; Moissl-Eichinger, 2011; Hewitt *et al*., 2012, 2013; Kembel *et al*., 2012; Dunn *et al*., 2013; Kelley and Gilbert, 2013; Meadow *et al*., 2013).

Indoor microbiomes originate mainly from the human skin, soils, pets or from outside air flow. All of these sources contain potential human pathogens as well as beneficial bacteria interacting with their host in a positive way (Flores *et al*., 2011; Kembel *et al*., 2012; Meadow *et al*., 2013). Good examples for indoor environments under extreme conditions are the intensive care unit (ICU) in hospitals, or clean rooms with nutrient-poor, dry and detergents exposed conditions (Fig. 1). Humans are the major source of contamination in these extreme environments. Transmission of skin microbiota through contact between surfaces and humans leads to a rapid spreading among individuals. Due to strict sanitation protocols, many hardy extremophiles can survive in these oligotrophic conditions over a long period of time (Kramer *et al*., 2006). Furthermore, also airborne bacterial communities influenced by ventilation, occupancy, and outdoor air source have serious effects to humans (Meadow *et al*., 2013). Kembel *et al*. (2012) was the first group analysing patent rooms and found a strong correlating effect between architecture and ventilation. Thus, indoor microbial communities are an important part of everyday human health. Hence, it is not surprising, that they are part of human-associated bacteria and can change with climatic conditions (Fierer *et al*., 2008) due to the high emission rate of up to 10^6 bacteria per person-hour (Qian *et al*., 2012). Sources of these airborne and passed bacteria of built environments are not well known and it is a big challenge especially in ICUs and clean rooms. Particularly in the ICU, sanitation protocols are

Figure 1 Analysed indoor environment of intensive care unit (ICU) and spacecraft assembly clean room.

stricter comparing to other indoor areas, but many patients in hospitals develop so called nosocomial infections, which lead to serveral diseases and even cause death (Vincent *et al*., 1995; Plowman, 2000). Therefore, hospital surfaces are very often an overlooked reservoir for bacteria (Hota, 2004; Gastmeier *et al*., 2005; Kramer *et al*., 2006).

Indoor environments are new fields of research and gain more and more attention. New sequencing techniques and bioinformatic possibilities allow studying the indoor microbiome to understand ecological interaction between humans and microorganisms including their beneficials.

1.2.2. Functional potential of the Sphagnum-dominated bog ecosystem

Bog ecosystems belong to the oldest vegetation forms on earth, where *Sphagnum*-dominated peatlands represent one of the most extensive types of Northern wetlands (Dedysh, 2011). They cover with four million km^2 approximately 3% of the earth's surface with a high value for biodiversity conservation, as reservoir of fresh water and play an extraordinary role in carbon sequestration to profit human´s welfare as well as our world climate (Succow and Joosten, 2001; Raghoebarsing *et al*., 2005; Dise, 2009). In spite of their age, these long-existing ecosystems are extremely sensitive to change abiotic factors connected with climate change (Belyea and Malmer, 2004; Dise, 2009). Mosses of the genus *Sphagnum* are consisting of approximately 300 different species commonly occurring worldwide and form the dominant component of bog vegetation (Daniels and Eddy, 1985). *Sphagnum* bogs are unique environments for several plants and animals, even though they can be regarded as an extreme habitat for microorganisms. They are characterized by high acidity (pH 3.5–5.0), low temperature and water saturation together with extremely low concentration of mineral nutrients (Richardson *et al*., 1978). It is also known that *Sphagnum* mosses are able to change their environments: living Sphagna have extraordinarily high cation exchange capacity and therefore, acidify their environment by exchanging tissue-bound protons for basic cations in surrounding water (Soudzilovskaia *et al*., 2010).

In this study, we were concentrated on microbial communities associated with the bryophyte species of the genus *Sphagnum magellanicum*. Generally, *S. magellanicum* grows in ombrotrophic (precipitation-derived nutrition) to weakly minerotrophic (supplied by ground-waters), acidic and relatively dry sites, and forms broad carpets or hummocks. The colour of

gametophytes varies from pale green to red depending on their light exposure. Figure 2 shows the peatland and a single plant of *S. magellanicum* from the Alpine bog Pirker Waldhochmoor (N46°37’38.66’’ E14°26’5.66’’) which were analysed within this study. These plant leaves are highly specialised. They form a particular tissue of living, chlorophyll-containing chlorocytes and dead cell content-free hyalocytes, which are responsible for their high water holding capacity. The unique morphology of *Sphagnum* gametophytes was studied for microbial colonization patterns. Fluorescent *in situ* hybridisation (FISH) coupled with Confocal laser scanning microscopy (CLSM) observation of *S. magellanicum* leaves revealed colonisation of the outer surface and inner hyaline cells (Fig. 2). *Sphagnum* mosses are characterised by a specific and diverse microbial community during their whole lifecycle, where the immense bacterial diversity was transferred via the sporophyte to the gametophyte explaining the high specificity over long distances (Bragina *et al*., 2012a). Furthermore, peat mosses are colonized by highly diverse bacterial communities with antagonistic and plant-growth promoting activities and they also produce bioactive secondary metabolites influencing microbial colonisation (Zhu *et al*., 2006; Opelt *et al*., 2007a, 2007b).

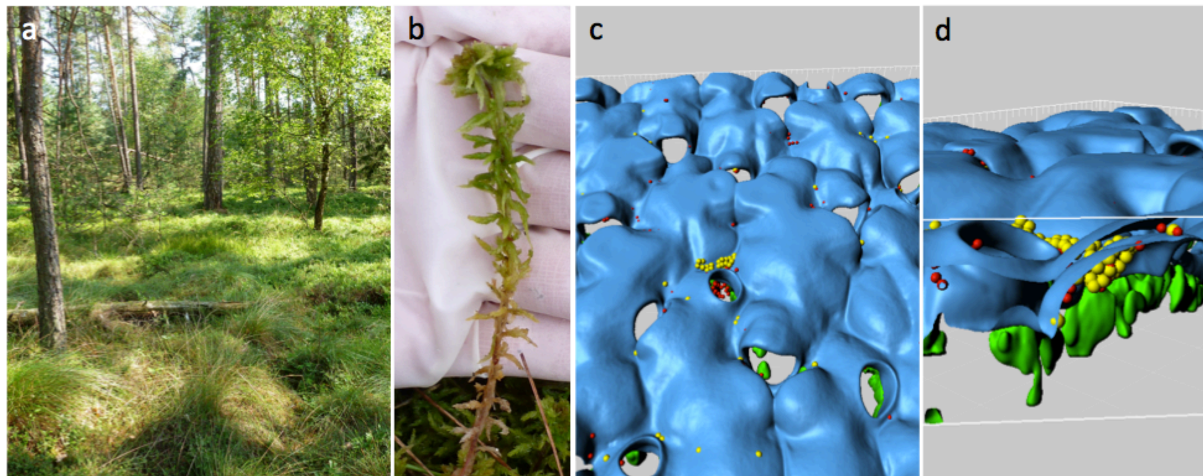


Figure 2 A typical bog complex from Pirker Waldhochmoor in Carinthia/Austria (a) composed of characteristic single plants of *Sphagnum magellanicum* (b). Fluorescent *in situ* hybridisation (FISH) of *S. magellanicum* leaves showed colonisation of bacteria on the outer surface (c) as well as on the inner cells (d); blue: cell walls of *Sphagnum* cells; green: chlorophyll-containing *Sphagnum* chlorocytes; yellow: *Alphaproteobacteria*; red: other bacteria. Images were acquired by confocal laser scanning microscopy (CLSM) and processed by 3D computer reconstruction using Imaris7.0. Scale bar = 20 µm (c) and 10 µm (d).

Recently it has been reported that the moss-ecosystem is a huge reservoir for the discovery of novel microbial enzymes and comprises a high antagonistic potential (Opelt and Berg, 2004). Therefore, we are focused on the analysis of taxonomic diversity, distribution and genomic context of gene clusters including hot spots for antibiotic resistance genes, which are relevant for secondary metabolism, e.g. polyketide synthases (PKSs) or non-ribosomal peptides (NRPSs) within the *Sphagnum* microbiome.

2. Methodical approach

In the past decade, the use of metagenomic approaches has increased exponentially and the ability to sequence has become accessible to research all over the world. Figure 3 demonstrates our applied methods within this work. Samples from two extreme habitats were taken and analysed in different ways depending on posing of their question. The first trial was conducted on several areas in an intensive care unit (ICU), where the profiling of the community structure was achieved by amplified ribosomal DNA restriction analysis (ARDRA) and BOX fingerprinting (Berg *et al*., 2002). Deeper insights of bacterial contribution were gained by 454-pyrosequencing of 16S rRNA genes. Additional principal coordinate analysis (PCoA) created with QIIME (Caporaso *et al*., 2010) and clustering network utilisation (Smoot *et al*., 2011) confirmed the distinct profiles between different areas in the ICU using both strategies. The *Sphagnum* microbiome was selected for unravelling the functional diversity and for deeper analysis of bacterial composition within the peat bog. An Illumina-based metagenomic approach followed by *de novo* assembly and MG-RAST annotation (Meyer *et al*., 2008) revealed specific biochemical pathways and adaptive strategies within the moss metagenome. Furthermore, screening of antibiotic producing genes were done by PCR approach with designed primer pairs and by *in silico* analysis. Subsequently, the dominant bacterial taxa were visualised on *Sphagnum* cells by fluorescent *in-situ* hybridisation (FISH) coupled with confocal laser scanning microscopy (CLSM) and computer-assisted reconstructing.

These methods supplied us new taxonomic and functional insights into our two extreme habitats by combining conventional cultivation techniques with next generation sequencing methods.

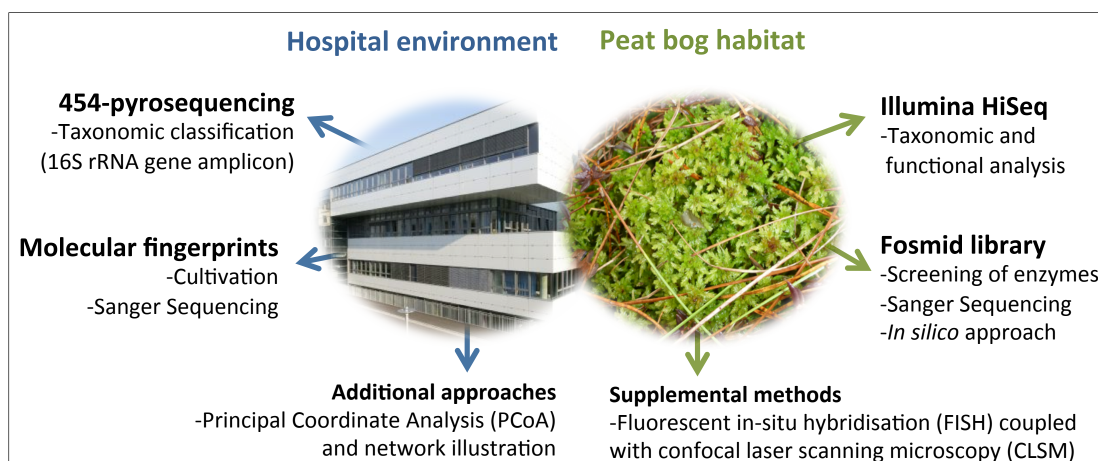


Figure 3 Overview of used methods applied on two extreme habitats within this study.

3. Objectives of the work

The overall aim of this study was to develop biological tools to apply sequence-based metagenomic studies, to establish metagenomic libraries and the screening potentially antibiotic producing genes.

The first purpose of the work was to analyse the bacterial community composition from the indoor environment of the ICU at the Department of Internal Medicine at the University Hospital in Graz, Austria. Samples from three general areas (floors, medical devices, work-places) were compared using two different approaches, cultivation-dependent as well as cultivation-independent methods. The most abundant taxa were identified across the three sampling areas and visualized using principal coordinate analysis (PCoA) and a profile clustering network illustration. Detected microbial communities and their influence to humans by diverse abiotic and biotic factors in indoor environments were reviewed.

Based on the knowledge that *Sphagnum* mosses comprise a high bacterial diversity, a metagenomic analysis was applied using Illumina HiSeq sequencing and annotation via MG-RAST (Meyer *et al*., 2008) annotation. Priority was placed on detection of the functional diversity with special focus on their ecological specialisation and comparison with other published metagenomes. In addition, screening antimicrobial enzymes (NRPS and PKS) involved in the synthesis of natural products was performed by different approaches. In this way the identification of gene clusters involved in the biosynthesis of secondary metabolites by moss-associated bacteria was pursued.

Using conventional cultivation techniques and high-throughput methods, this work provides a wealth of information on microbial and functional diversity in extreme environments and expands the understanding of microbial ecology within the analysed habitats.

4. Results and discussion

4.1. Analysis of the bacterial communities associated with the indoor hospital environment

Buildings, especially intensive care units (ICU) and clean rooms are complex ecosystems consisting of microorganisms interacting with each other and their environment. Altogether, 34 surface samples obtained from three general areas (floors, medical devices, work-places) inside the ICU of the University Hospital in Graz, Austria, were studied and analysed by a multifaceted approach.

Structure and diversity of bacterial community profiles

The composition of microbial communities within the ICU were analysed by 454-pyrosequencing and included seven different phyla. Proteobacteria (64%) was the most abundant phylum across all samples followed by *Acidobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria, Firmicutes* and*, Nitrospira.* The amplicon library comprised sequences of 405 genera, where 76 of them were reaching 1% of relative abundance. Although the surfaces in the ICU were characterized by highly diverse bacterial communities and they were actually reduced in comparison with other indoor environments such as living, patient, class or rest rooms (Kembel *et al*., 2012). Similar to other indoor microbial communities, these communities were partially colonized by human-associated bacteria. While it is impossible to predict the pathogenicity of a strain based on 16S rRNA sequences, the proportion of bacteria identified as those genera or species closely related to human pathogens was very high. They are known for their facultative pathogenic and nosocomial character, e.g. *Acinetobacter*, *Stenotrophomonas*, *Burkholderia*, *Flavobacterium*, *Propionibacterium*, *Pseudomonas*, *Staphylococcus* and *Escherichia*/*Shigella* (Clayton et al., 2006; Das *et al*., 2011). Moreover, Gram-negative pathogens were also identified e.g. *E. coli*, *Klebsiella*, *Pseudomonas*, *Serratia*, *Enterobacter*, *Edwardsiella*, *Proteus* and *Chryseobacterium*.

Differences between the bacterial compositions across different areas

Beta diversity of the bacterial communities within the ICU revealed clear distinctions between bacterial populations among the three areas. PCoA and network illustration indicated that the floor-associated bacterial communities formed clusters distinct from devices, while the analysed samples from workplaces and devices were similar. A comparison of the relative

abundances at genus level indicated that *Pseudomonas* and *Propionibacterium* were clearly most abundant on all sampling sites. On the floor, the most frequently present genus was *Acinetobacter* (24%) among the other commonly found genera *Propionibacterium*, *Pseudomonas*, *Staphylococcus*, and *Streptococcus*. The dominant genera on devices and workplaces were *Pseudomonas*, *Novosphingobium*, *Burkholderia*, *Bradyrhizobium* and *Propionibacterium*. The most abundant genus on floors, *Acinetobacter*, was less present in these two areas. Although bacterial communities in the ICU could be effectively differentiated, connections and transmissions were also detected by these two approaches.

Comparative cultivation-dependent approach

Comparative samples were taken from 10 sampling sites of defined positions on devices and workplaces. A total of 130 isolates obtained from contact plates were characterized by molecular fingerprint methods. Representative strains were partially sequenced and determined by their 16S rRNA. The genera *Aerococcus*, *Arthrobacter*, *Bacillus*, *Corynebacterium*, *Kocuria*, *Micrococcus*, *Paenibacillus*, *Planomicrobium*, *Roseomonas* and *Staphylococcus* were detected, where the majority of isolates and were identified as species belonging to the genus *Staphylococcus*. With the exception of the *Roseomonas mucosa* strain, only Gram-positive species, e.g. *Staphylococcus* sp., were found, which are known as causal agents of nosocomial infections with diverse resistances against antibiotics (Uçkay *et al.*, 2009).

Transmission of the hospital-associated microbiome

The bacterial communities from three general area floors, medical devices, and workplaces were characterized by a specific and distinct composition. Skin-associated genera (*Propionibacterium*, *Corynebacterium*, *Staphylococcus*, *Streptococcus*, *Bradyrhizobium*) (Grice and Segre, 2011) were highly abundant on medical devices and working surfaces, which was expected considering that they are frequently touched by hands of hospital staff: a typical hand surface harbored on average more than 150 unique species-level bacterial phylotypes (Fierer *et al.*, 2008). Interestingly, genera of *Burkholderia* and *Bradyrhizobium* were most abundant on devices and workplaces and are both originally plant-associated genera. Transmission of bacteria, e.g. staphylococci, from one site to the other was also detected by comparing molecular fingerprints of the isolated strains. Deposition of bacterial strains from identical or similar sources, e.g. by personal staff, can also explain this finding. Surface sanitation is an often overlooked, yet crucial component of transmission (Otter and

French, 2009), which should be considered more in sanitation protocols. Another interesting point was that the indoor hospital environment also contained plant-associated taxa that can undergo bivalent interactions with humans by causing facultative infections (Berg *et al*., 2005; 2009). Their origin and function in hospital environments is still unclear. A possibility of transmission could be the transport via pollen into the hospital environment due to the detection of pollen as a vector for specific plant-associated bacteria (Fürnkranz *et al*., 2012). In addition to air conditioning, the investigated ICU was also window-ventilated, which has been known to result in an increased abundance of chloroplast DNA than in exclusively mechanically ventilated rooms (Kembel *et al*., 2012).

Comparison between 16S pyrosequencing and standard cultivation

Culture-dependent identification was compared with 16S rRNA gene 454-pyrosequencing analysis. While operational taxonomic units taken from the amplicon libraries were affiliated with 405 different genera (76 genera ≥ 1% of relative abundance), standard cultivation obtained only 10 bacterial genera corresponding to 2.5% of the total bacterial diversity. Although it is well-known that cultivation-dependent techniques capture only a small part of the microbiome (Staley and Konopka, 1985; Schleifer, 2004), an unexpected high difference between the bacterial diversity was found using both methods – standard cultivation and amplicon sequencing. While in the 16S rRNA gene amplicon library the amount of Gram-positive and Gram-negative bacteria was nearly the same, we detected almost exclusively Gram-positive bacteria by cultivation. With the exception of *Roseomonas mucosa* known to be associated with bacteremia and other human infections (Christakis *et al*., 2006), cultivation failed to capture the Gram-negative spectrum. However, it must be considered that pyrosequencing based on DNA may also detect DNA from non-living and living bacteria.

- A detailed representation of the results is given in publication "The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing" and in the book chapter “Complex indoor communities: Bacterial life under extreme conditions in clean rooms and intensive care units”.

4.2. Analysis of taxonomic and functional composition within the *Sphagnum* microbiome

Sphagnum-dominated bogs are unique habitats for a lot of plants and animals, even though they form an extreme habitat (low pH, temperature, low nutrient concentrations, etc.) for microorganisms. *Sphagnum magellanicum* plants were sampled in an Alpine bog in Austria and were analysed by the different metagenomic approaches.

Taxonomic diversity and spatial structure of the S. magellanicum *microbiome*

Partial 16S rRNA genes (in total 7,318 reads) were obtained from metagenomic sequences to characterise the structure of bacterial communities. At phylum level, the majority of reads were assigned to Proteobacteria (65.8%) followed by Acidobacteria (11.4%), Actinobacteria (5.6%), Bacteroidetes (4.2%) and Verrucomicrobia (2.0%). Other analysed reads was distributed among 13 bacterial phyla which notably contained Planctomycetes. The taxonomic hits distribution of metagenomic sequences with predicted protein coding regions and ribosomal rRNA genes revealed highly similar dominant patterns to the 16S rRNA genes data. Within the reads assigned to domain Bacteria (61,528,765 sequences), dominant portion was composed of Proteobacteria (61.9%), Acidobacteria (13.1%), Actinobacteria (8.3%), Bacteroidetes (4.2%), and Verrucomicrobia (3.0%). The minor fraction of functional bacterial reads was distributed among 16 phyla that were not covered by partial 16S rRNA genes. This approach allowed a deep analysis of the 16S rRNA gene diversity without PCR-based bias. Although the dominant bacterial taxa detected using Illumina sequencing were similar to those revealed by PCR-dependent approaches (Bragina *et al*., 2012a), their relative abundance considerably differed. As such, a low number of Planctomycetes 16S rRNA genes was observed that contrasts with their relatively high abundance in the Northern peat bogs and Arctic peat soils (Serkebaeva *et al*., 2013; Tveit *et al*., 2013).

Colonisation patterns of Sphagnum *gametophytes analysed by FISH-CLSM*

Fluorescence *in situ* hybridization (FISH) and confocal laser scanning microscopy (CLSM) techniques were used to visualise the most abundant bacterial patterns in *S. magellanicum* gametophytes. In general, *Sphagnum* mosses are characterised by unique morphology that distinguishes them from other bryophytes (Daniels and Eddy, 1985). Especially, *Sphagnum* leaves are composed of a single-layer cell net of photosynthetic chlorocytes and dead hyalocytes, which contain large pores. By applying FISH-CLSM approach, it was demonstrated that hyalocytes of moss leaves serve as a main colonisation compartment for bacteria. One of the most abundant bacterial taxa – Alphaproteobacteria – represented up to

31.9% of the detected bacterial cells that coincided with its relative abundance in metagenomic datasets (30.2%).

Functional analysis and genetic potential of the metagenomic moss microbiome

The *Sphagnum* microbiome carries essential genetic potential for sustainable functioning in association with the host plants. To elucidate this profound diversity, a framework in the form of plasticity-stability-interaction that integrates genetic signatures of symbiosis (Gilbert *et al*., 2012) was developed within the plant-microbe biocoenosis. Specifically, the moss metagenome contained a relatively high number of mobile elements which were also found in the metagenomes of symbiotic bacterial consortia and considered to play an important role in the evolution of bacterial genomes for symbiosis with their hosts (Ochman and Moran, 2001; Thomas *et al*., 2010). Furthermore, *Sphagnum* mosses belong to the poikilohydric plants that undergo repetitive desiccation and oxidative stress (Daniels and Eddy, 1985; Scheibe and Beck, 2011). Due to the high diversity and abundance of genes responsible for the oxidative stress response in the studied metagenome, it is proposed that the bacterial capacity to tolerate oxidative stress may determine the effective and stable colonisation of the *Sphagnum* mosses. In regards to interaction traits, vegetation in peatland ecosystems is strongly limited by nitrogen availability and therefore requires prokaryotic associates for nitrogen supply (Rydin and Jeglum, 2006). Since Granhall and Hofsten (1976) observed nitrogen-fixing symbiotic Cyanobacteria in *Sphagnum* for the first time, diazotrophic communities of Sphagna have been characterised by a high taxonomic diversity and shown to transfer fixed nitrogen to the host plants (Bragina *et al*., 2012b, 2013; Berg *et al*., 2013). Overall, we provided evidence that the *Sphagnum* microbiome carries essential genetic potential for sustainable functioning in association with the host plants and within the peatland ecosystem.

Comparison of different metagenomes

An inter-environmental comparison (PCoA plot) demonstrated that the moss microbiome is distinct from microbial communities of higher plants and peat soils by its genetic context. This difference indicates the specific interactions established between *Sphagnum* mosses and their microbiome. Previous research proposed that the *Sphagnum* microbiome intimately cooperated with the host plants via nutrient supply and defence against pathogens (Raghoebarsing et al., 2005; Opelt *et al*., 2007b; Bragina *et al*., 2013), but Illumina sequencing of the moss metagenome obtained a much higher functional diversity than previously reported.

Screening of NRPS and PKS genes within the moss microbiome

Plants and their inhabit microorganisms provide an immense reservoir of chemically diverse natural products with potential biological activity. Prominent antibiotic producing gene clusters are non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs). Employing several metagenomic approaches (Screening of a fosmid library and *in silico* analysis) NRPS and PKS genes could be detected within the *Sphagnum* moss-associated bacteria. Degenerated PCR primers were employed successfully to amplify NRPS and PKS gene sequences for screening of the metagenomic fosmid library. The sequences retrieved showed high homology to the gene sequences of the genera *Pseudomonas* and *Pectobacterium*. Parallel *in silico* Illumina-based metagenomic analysis identified 328 NRPS, 456 PKS as well as 57 of their hybrid genes. Genera of *Streptomyces*, *Paenibacillus, Mycobacteria* and *Lysobacter* were the most detected sequences within the *Sphagnum* microbiome. The widespread appearance of NRPS and PKS gene clusters across the phyla Proteobacteria, Actinobacteria and Firmicutes has been reported (Wanga *et al*., 2014). Intriguingly, mixed/hybrid NRPS-PKS genes were also present within the moss microbiome dataset, where all sequences were allocated to the phylum *Proteobacteria.* Despite technical limitations (Teeling and Glöckner, 2012) of next generation sequencing methods and following *in silico* analysis, metagenomic screening is a successful approach in the discovery of novel biomolecules using for biotechnological applications.

- A detailed representation of the results is given in manuscript I "The *Sphagnum* microbiome supports greatly bog ecosystem functioning under extreme conditions" and in manuscript II “Metagenomic analysis of NRPS and PKS genes within the *Sphagnum* microbiome”.

4.3. Conclusion

The results showed that the microbial diversity in the enclosed ICU is altered and partially reduced compared to the outdoor moss environment. An unexpected high diversity from the bacterial communities with bacteria closely related to human pathogens was found in both environments, but also taxa known for their beneficial interaction with eukaryotes. This make necessary to think about new cleaning and hygiene strategies in indoor environments especially in hospitals, where the existing measurements often promote multi-resistant pathogens instead of supporting beneficials. The plant-associated bacteria, e.g. from

Sphagnum bogs, could act as counterparts against pathogens within the microbial ecosystem. In this respect it has been shown, that plants provide beneficial bacteria for indoor rooms leading to positively influence in human health (Berg *et al*., 2014).

Altogether, both environments harbour a significant diversity of uncultured bacteria. Next generation sequencing applications are important tools to address fundamental questions of microbial ecology, evolution and diversity. These new metagenomic technologies and applications allowed accessing the majority of uncultured microorganisms to get new insights into still unknown bioresources.

5. References

Adrio JL and Demain AL. (2014). Microbial Enzymes: Tools for Biotechnological Processes. Biomolecules 2014, 4, 117-139.

Belyea LR and Malmer N. (2004). Carbon sequestration in peatland: patterns and mechanisms of response to climate change. Glob Change Biol 10: 1043–1052.

Berg G, Roskot N, Steidle A, Eberl L, Zock A and Smalla K. (2002). Plant-dependent genotypic and phenotypic diversity of antagonistic rhizobacteria isolated from different Verticillium host plants. Appl Environ Microbiol. 68(7):3328-38.

Berg G, Eberl L and Hartmann A. (2005).The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. Environ Microbiol 7, 1673–1685

Berg, G. (2009). Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. Appl Microbiol Biotechnol 84, 11–18

Berg G, Zachow C, Müller H, Philipps J and Tilcher R. (2013). Next-generation bio-products sowing the seeds of success for sustainable agriculture. Agronomy, 3, 648–656.

Berg G, Mahnert A and Moissl-Eichinger C. (2014). Beneficial effects of plant-associated microbes on indoor microbiomes and human health?. Front. Microbiol.

Bragina A, Maier S, Berg C, Müller H, Chobot V, Hadacek F and Berg G. (2012a). Similar diversity of Alphaproteobacteria and nitrogenase gene amplicons on two related sphagnum mosses.
Front Microbiol. 2012 Jan 10;2:275.

Bragina A, Berg C, Cardinale M, Shcherbakov A, Chebotar V and Gabriele Berg G. (2012b). Sphagnum mosses harbour highly specific bacterial diversity during their whole lifecycle. The ISME Journal, 1–12.

Bragina A, Berg C, Müller H, Moser D and Berg G. (2013).Insights into functional bacterial diversity and its effects on Alpine bog ecosystem functioning. Sci Rep. 2013;3:1955.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J and Knight R. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7, 335–336.

Christakis GB, Perlorentzou S, Alexaki P, Megalakaki A and Zarkadis IK. (2006). Central line-related bacteraemia due to Roseomonas mucosa in a neutropenic patient with acute myeloid leukaemia in Piraeus, Greece. J Med Microbiol 55, 1153–1156

Chistoserdova L. (2010). Recent progress and new challenges in metagenomics for biotechnology. Biotechnol Lett. 32:1351-1359.

Clayton JJ, Baig W, Reynolds GW and Sandoe JA. (2006). Endocarditis caused by Propionibacterium species: a report of three cases and a review of clinical features and diagnostic difficulties. J Med Microbiol. 55(Pt 8):9817.

Curtis TP, Sloan WT and Scannell JW. (2002). Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci USA. 99:10494-10499.

Daniels RE and Eddy A. (1985). Handbook of Europaen Sphagna. Aberystwyth: Cambrian News.

Das I, Lambert P, Hill D, Noy M, Bion J and Elliott T. (2011). Carbapenem-resistant Acinetobacter and role of curtains in an outbreak in intensive care units. J Hosp Infection 50, 110–114

Dedysh SN. (2011). Cultivating uncultured bacteria from north- ern wetlands: knowledge gained and remaining gaps. Front. Micro- biol. 2:184. doi: 10.3389/fmcib.2011. 00184

Delmont TO, Malandain C, Prestat E, Larose C, Monier JM, Simonet P and Vogel TM. (2011). Metagenomic mining for microbiologists. ISME J.5(12):1837-43.

Dise NB. (2009). Environmental science. Peatland re- sponse to global change. Science 326: 810–811.

Dunn RR, Fierer N, Henley JB, Leff JW and Menninger H. (2013). Home life: factors structuring the bacterial diversity found within and between homes. PLoS ONE 8:e64133. doi: 10.1371/journal.pone.0064133

Ferrer M, Beloqui A, Timmis KN and Golyshin PN. (2009). Metagenomics for mining new genetic resources of microbial communities. J Mol Microbiol Biotechnol. 16:109-123.

Fierer N, Liu Z, Rodrıguez-Hernandez M, Knight R, Henn M and Hernandez M.T. (2008) Short-term temporal variability in airborne bacterial and fun- gal populations. Appl. Environ. Micro- biol., 74, 200–207.

Flores GE, Bates ST, Caporaso JG, Lauber CL, Leff JW, Knight R, et al. (2013). Diversity, distribution and sources of bacteria in residen- tial kitchens. Environ. Microbiol. 15, 588–596.

Flores GE, Bates ST, Knights D, Lauber CL, Stombaugh J, Knight R, et al. (2011). Microbial biogeography of public restroom surfaces. PLoS ONE 6:e28132. doi: 10.1371/journal.pone.0028132

Fürnkranz M, Lukesch B, Müller H, Huss H, Grube M and Berg G. (2012). Microbial diversity inside pumpkins: microhabitat-specific communities display a high antagonistic potential against phytopathogens. Microb Ecol. 63(2):418-28.

Gastmeier P, Schwab F, Bärwolff S, Rüden H and Grundmann H. (2005). Correlation between the genetic diversity of nosocomial pathogens and their survival time in intensive care units. J. Hosp. Infect. 62, 181186.

Ghai R, Martin-Cuadrado AB, Molto AG, Heredia IG, Cabrera R, Martin J, Verdú M, Deschamps P, Moreira D, López-García P, Mira A and Rodriguez-Valera F. (2010). Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. ISME J. 9:1154-66.

Gilbert SF, Sapp J, Tauber AI. (2012). A symbiotic view of life: we have never been individuals. The Quarterly Review of Biology, 87, 325–341.

Granhall U and Hofsten AV. (1976). Nitrogenase activity in relation to intracellular organisms in Sphagnum mosses. Physiologia Plantarum, 36, 88–94.

Grice EA and Segre JA. (2011). The skin microbiome. Nature Rev Microbiol 9, 244–253.

Handelsman J. (2004). Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev. 68:669–85.

Hewitt KM, Gerba CP, Maxwell SL and Kelley ST. (2012). Office space bacterial abundance and diversity in three metropolitan areas. PLoS ONE 7:e37849. doi: 10.1371/journal.pone.0037849

Hewitt KM, Mannino FL, Gonzalez A, Chase JH, Caporaso JG, Knight R, et al. (2013). Bacterial diversity in two neonatal intensive care units (NICUs). PLoS ONE 8:e54703.

Hota B. (2004). Contamination, disinfection, and cross-colonization: are hospital surfaces reservoirs for nosocomial infection? Clin. Infect. Dis. 39, 1182–1189. doi: 10.1086/424667

Iqbal HA, Feng Z and Brady SF. (2012). Biocatalysts and small molecule products from metagenomic studies. Curr Opin Chem Biol.16:109-116.

Kelley ST and Gilbert JA. (2013). Studying the microbiology of the indoor environment. Genome Biology. 14:202

Kembel SW, Jones E, Kline J, Northcutt D, Stenson J, Womack AM, Bohannan BJ, Brown GZ and Green JL. (2012). Architectural design influences the diversity and structure of the built environment microbiome. ISME J. (8):1469-79.

Kramer A, Schwebke I and Kampf G. (2006). How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. BMC Infect Dis. 6:130.

Lee MH and Lee SW. (2013). Leecorresponding author Bioprospecting Potential of the Soil Metagenome: Novel Enzymes and Bioactivities. Genomics Inform. 11(3): 114–120.

Liszka MJ, Clark ME, Schneider E and Clark DS. (2012). Nature Versus Nurture: Developing Enzymes That Function Under Extreme Conditions. Annu. Rev. Chem. Biomol. Eng. 3:77–102

Lorenz P and Eck J. (2005). Metagenomics and industrial applications. Nature Reviews Microbiology 3, 510-516

Metzker ML. (2010). Sequencing technologies-the next generation. Nat Rev Genet. 11:31–46.

Meadow JF, Altrichter AE, Kembel SW, Kline J, Mhuireach G, Moriyama M, et al. (2013). Indoor airborne bacterial communities are influ- enced by ventilation, occupancy, and outdoor air source. Indoor Air 24, 41–48.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J and Edwards RA. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 19;9:386.

Moissl-Eichinger C. (2011). Archaea in artificial envi- ronments: their presence in global spacecraft clean rooms and impact on planetary protection. ISME J. 5, 209–219. doi: 10.1038/ismej.2010.124

Neelakanta G and Sultana H. (2013). The use of metagenomic approaches to analyze changes in microbial communities. Microbiol Insights. 6:37-48.

Ochman H and Moran NA. (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science, 292, 1096–1098.

Opelt and Berg. (2004). Diversity and Antagonistic Potential of Bacteria Associated with Bryophytes from Nutrient-Poor Habitats of the Baltic Sea Coast. Appl Environ Microbiol. 70(11):6569-79.

Opelt K, Berg C, Schönmann S, Eberl L and Berg G. (2007a). High specificity but contrasting biodiversity of Sphagnum-associated bacterial and plant communities in bog ecosystems independent of the geographical region. The ISME Journal, **1**, 502–516.

Opelt K, Chobot V, Hadacek F, Schönmann S, Eberl L and Berg G. (2007b). Investigation of the structure and function of bacterial communities associated with Sphagnum mosses. Environmental Microbiology, **91**, 2795–2809.

Otter JA and French GL. (2009). Survival of nosocomial bacteria and spores on surfaces and inactivation by hydrogen peroxide vapor. J Clin. Microbiol 47, 205–207

Pace NR, Stahl DA, Lane DJ, and Olsen GJ. (1985). The analysis of natural microbial populations by ribosomal RNA sequences. Am. Soc. Microbiol. News 51:4-12.

Pace NR. (2009). Mapping the tree of life: progress and prospects. Microbiol Mol Biol Rev. 73:565-576.

Plowman R. (2000). The socioeconomic burden of hospital acquired infection. Euro Surveill. 5(4):49-50.

Qian J, Hospodsky D, Yamamoto N, Nazaroff WW and Peccia J. (2012). Size-resolved emission rates of airborne bacteria and fungi in an occupied class-room. Indoor Air, 22, 339–351.

Raghoebarsing AA, Smolders AJP, Schmid MC, Rijpstra WIC, Wolters-Arts M, Derksen JM et al. (2005). Methanotrophic symbionts provide carbon for photosynthesis in peat bogs. Nature 436: 1153–1156.

Rappe MS and Giovannoni SJ. (2003). The uncultured microbial majority. Annu Rev Microbiol. 57:369-394.

Richardson C, Tilton D, Kadlec J, Chamie J and Wentz W. (1978). NY, USA: Academic Press 1978, pp. 217–241, 44, 14969-14976

Rydin H and Jeglum J. (2006). Nutrients, light and temperature. In: The biology of peatlands, pp. 164–186. Oxford University Press, Oxford.

Sanger F, Nicklen S and Coulson AR. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci, USA. 74:5463-5467.

Scheibe R and Beck E. (2011). Drought, desiccation, and oxidative stress. In: Plant desiccation tolerance (eds Lüttge U, Beck E, Bartels D), pp. 209–232. Springer-Verlag, Berlin, Heidelberg.

Schleifer K. (2004). Microbial diversity: facts, problems and prospects. Systematic Applied Microbiology 27, 3-9.

Schmidt TM, DeLong EF and Pace NR. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol. 173:4371-4378.

Serkebaeva YM, Kim Y, Liesack W and Dedysh SN. (2013). Pyrosequencing-based assessment of the Bacteria diversity in surface and subsurface peat layers of a Northern wetland, with focus on poorly studied phyla and candidate divisions. PLoS One, 8, e63994.

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y and Simon M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc Natl Acad Sci, USA. 89:8794-8797.

Simon C and Daniel R. (2011). Metagenomic analyses: past and future trends. Appl Environ Microbiol. (4):1153-61.

Sjöling S and Cowan DA. (2008). Metagenomics: Microbial Community Genomes Revealed. Psychrophiles: from Biodiversity to Biotechnology, pp 313-332

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 27(3):431-2.

Soudzilovskaia NA, Cornelissen JHC, During HJ, van Logtestijn RSP, Lang Si and Aerts R. (2010). Similar cation exchange capacities among bryophyte species refute a presumed mechanism of peatland acidification. Ecology, 91, 2716–2726.

Staley JT and Konopka A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol. 1985;39:321-46.

Succow M and Joosten H. (2001). Landschaftsökologische Moorkunde. Schweizerbart`sche Verlagsbuchhandlung, Stuttgart, Germany.

Täubel M, Rintala H, Pitkäranta M, Paulin L, Laitinen S, Pekkanen J, et al. (2009). The occupant as a source of house dust bacteria. J. Allergy Clin. Immunol. 124, 834–840.

Teeling H and Glöckner FO. (2012). Current opportunities and challenges in microbial metagenome analysis--a bioinformatic perspective. Brief. Bioinform. doi: 10.1093/bib/bbs039.

Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A et al. (2010) Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. The ISME Journal, 4, 1557–1567.

Thomas T, Gilbert J and Meyer F. (2012). Metagenomics - a guide from sampling to data analysis. Microb Inform Exp. 2(1):3.

Torsvik V, Goksoyr J and Daae FL. (1990). High diversity in DNA of soil bacteria. Appl Environ Microbiol. 56:782-787.

Tveit A, Schwacke R, Svenning MM and Urich T. (2013). Original carbon transformation in high-Arctic peat soils: key functions and microorganisms. The ISME Journal, 7, 299–311.

Uchiyama T and Miyazaki K. (2009). Functional metagenomics for enzyme discovery: challenges to efficient screening. Current Opinion in Biotechnology. 20:616–622

Uçkay I, Pittet D, Vaudaux P, Sax H, Lew D and Waldvogel F. (2009). Foreign body infections due to Staphylococcus epidermidis. Ann Med 41, 109–119.

Vincent JL, Bihari DJ, Suter PM, Bruining HA, White J, Nicolas-Chanoin MH, Wolff M, Spencer RC and Hemmer M. (1995). The prevalence of nosocomial infection in intensive care units in Europe. Results of the European Prevalence of Infection in Intensive Care (EPIC) Study. EPIC International Advisory Committee. JAMA. 274(8):639-44.

Wanga H, Fewera DP, Holmb L, Rouhiainena L and Sivonena K. (2014). Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. PNAS, doi: 10.1073/pnas.1401734111.

Whitman WB, Coleman DC and Wiebe WJ. (1998). Prokaryotes: the unseen majority. Proc Natl Acad Sci USA. 95:6578-6583.

Woese CR. (1987). Bacterial evolution. Microbiol Rev. 51:221-271.

Yamamoto N, Shendell DG and Peccia J. (2011). Assessing allergenic fungi in house dust by floor wipe sampling and quantitative PCR. Indoor Air 21, 521–530. doi: 10.1111/j.1600- 0668.2011.00732.x

Zhu, R.-L., D. Wang, L. Xu, R.-P. Shi, J. Wang and M. Zheng. (2006). Antibacterial activity in extracts of some bryophytes from China and Mongolia. The Journal of the Hattori Botanical Laboratory 100: 603-615.

6. Publication

Nature Scientific Reports (2013) **3**,1413; DOI:10.1038/srep01413

The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing

Lisa Oberauner[1,2], Christin Zachow[1,2], Stefan Lackner[3], Christoph Högenauer[4], Karl-Heinz Smolle[4] & Gabriele Berg[2]

[1]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Graz, Austria,
[2]Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria,
[3]Center for Medical Research, Medical University of Graz, Austria,
[4]Department of Internal Medicine, Medical University of Graz, Austria

SCIENTIFIC REPORTS

SUBJECT AREAS:
ENVIRONMENTAL MICROBIOLOGY
BIODIVERSITY
PATHOGENS
COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

Received
9 November 2012

Accepted
22 February 2013

Published
11 March 2013

Correspondence and requests for materials should be addressed to G.B. (gabriele.berg@tugraz.at)

The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing

Lisa Oberauner[1,2], Christin Zachow[1,2], Stefan Lackner[3], Christoph Högenauer[4], Karl-Heinz Smolle[4] & Gabriele Berg[2]

[1]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Graz, Austria, [2]Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria, [3]Center for Medical Research, Medical University of Graz, Austria, [4]Department of Internal Medicine, Medical University of Graz, Austria.

Indoor microbial communities play an important role in everyday human health, especially in the intensive care units (ICUs) of hospitals. We used amplicon pyrosequencing to study the ICU microbiome and were able to detect diverse sequences, in comparison to the currently used standard cultivation technique that only detected 2.5% of the total bacterial diversity. The phylogenetic spectrum combined species associated with the outside environment, taxa closely related to potential human pathogens, and beneficials as well as included 7 phyla and 76 genera. In addition, *Propionibacterium* spp., *Pseudomonas* spp., and *Burkholderia* spp. were identified as important sources of infections. Despite significantly different bacterial area profiles for floors, medical devices, and workplaces, similarities by network analyses and strains with identical molecular fingerprints were detected. This information will allow for new assessment of public health risks in ICUs, help create new sanitation protocols, and further our understanding of the development of hospital-acquired infections.

The majority of our life time is spent in indoor environments, but little is known about the bacterial communities with which we share indoors. Recently, the application of next generation sequencing techniques has allowed new insight into indoor microbial communities. In general, they are characterized by a high prokaryotic diversity and are comprised of diverse bacterial and archaeal phyla[1–4]. Indoor microbiomes originate mainly from the human skin or from outside air, and have even been known to include extremophiles. Furthermore, all of them contain potential human pathogens, but also beneficial bacteria that are characterized by a positive interaction with their host[1,3]. Kembel *et al.*[3] were the first to analyze patient rooms and find a strong effect from both architecture and ventilation. In contrast to the majority of indoor environments, rooms in hospitals and especially intensive care units (ICUs) are routinely monitored[5]. Standard cultivation techniques, such as contact plates, are commonly used to monitor the microbial burden. However, culture collections contain a restricted spectrum and only a very small proportion of the total bacteria as already described in 1985 by Staley and Konopka[6]. A comparison of the bacterial diversity obtained by standard monitoring and next generation sequencing techniques in ICUs has unfortunately not yet been published. We hypothesize that the ICU microbiome is characterized by a much higher bacterial diversity and abundance than is currently thought.

Indoor microbial communities are an important component of everyday human health[3,7], and are even partially composed of human-associated bacteria[1] due to the high emission rate of up to 10^6 bacteria per person-hour[8]. In ICUs, sanitation protocols are stricter than in other areas of the hospital, yet many patients treated in ICUs are infected with hospital-acquired "nosocomial infections" often due to an underlying severe disease[9,10]. Moreover, these nosocomial infections remain among the leading causes of death in hospitals of developed countries. For example, they are a significant cause of morbidity and mortality in the United States; 1.7 million infections resulting in 99,000 deaths were reported in 2002[11]. In Europe, the risk for nosocomial infections for patients in ICUs is reported as 45%[9]. Hospital surfaces are often overlooked reservoirs for this bacteria[12–14], thus new sanitation standards are needed to drastically reduce this risk for hospital-acquired infections[15]. New sequencing techniques will allow for a greater understanding of whole ICU bacterial communities, including their beneficials, and contribute to a new perspective on hospital sterility.

The objective of this work was to analyse the structure of bacterial communities from the ICU of the Department of Internal Medicine at the University Hospital in Graz/Austria using a comparative approach

between currently used standard cultivation of the ICU and 16S rRNA gene amplicon sequencing. Altogether, 34 surface samples obtained from three general areas (floors, medical devices, workplaces) inside the ICU were studied. 16S rRNA gene amplicons and isolates were identified and compared by a principal coordinate analysis. In addition, network analysis using Cytoscape 2.8 software[16] was performed to identify the most abundant taxa and to compare their abundance across the three sampling areas.

Results

Structure and diversity of bacterial community profiles. From all surface samples of the three areas within the ICU - the floor environment (A: 5 samples), devices (B: 11 samples) and workplaces (C: 8 samples) - visualized in Fig. 1 - a high number of amplicons were obtained and sequenced. In total, the raw dataset of all 24 samples contained 356,571 sequences. After trimming, the final operational taxonomic unit (OTU) table consisted of 308,440 sequences. The docking station (MID53) contained the highest (16,137) and the floor of the patient room after cleaning (MID35) the fewest (5,321) amount of sequences respectively. Due to the different number of sequences among samples, the data was normalized to 5,321 sequences.

The composition of microbial communities included 7 different bacterial phyla: *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Firmicutes*, *Nitrospira* and *Proteobacteria*. *Proteobacteria* (64%) was the most abundant phylum across all samples. At genus level, differences between the communities of the various areas were detected (Fig. 2). The amplicon library of the ICU included sequences of 405 genera, 76 of them were reaching 1% of relative abundance. Only a minor part of sequences belonging to members of the genera *Acinetobacter*, *Bradyrhizobium*, *Burkholderia*, *Delftia*, *Enhydrobacter*, *Propionibacterium*, *Pseudomonas*, *Serratia*, *Staphylococcus* and *Streptococcus* were retrieved from surfaces of all three areas. Figure 2 also indicates that devices contained a greater variety of bacteria (23 genera) than the workplaces (15 genera) and the floor (8 genera). An overlap between the detected genera of the three areas was also observed; the highest was found between devices/workplaces (12 genera) and followed by floor/workplaces (4) and floor/devices (3). A comparison of the relative abundances confirmed differences between the three main areas (Fig. 3). Most notably, *Pseudomonas* and *Propionibacterium* were clearly most abundant on all sampling sites. On the floor (A), the most frequently present genus was *Acinetobacter* (24%) among the other commonly found genera *Propionibacterium, Pseudomonas, Staphylococcus*, and *Streptococcus*. The dominant genera on devices (B) and workplaces (C) were *Pseudomonas* (4% in B, 7% in C), *Novosphingobium* (10% in B, 5% in C), *Burkholderia* (14% in B, 15% in C), *Bradyrhizobium* (16% in B, 17% in C) and *Propionibacterium* (7% in B, 5% in C). The most abundant genus on floors, *Acinetobacter,* was less present in these two areas (3% in B, 2% in C). *Chryseobacterium*, *Janthinobacterium*, *Legionella*, *Methylobacterium* and *Shigella* were minimal on devices and workplaces. *Corynebacterium* was only present in some floor samples and workplaces, whereas *Serratia* was measured in low numbers in several samples of all areas. *Gemella*, *Flavobacterium* and *Stenotrophomonas* were only detected on several devices, while *Bacillus*, *Granulicatella* and *Nitrospira* were all observed in relatively high abundances. Taxonomic classification of each sampling site is shown in Fig. S1.

To determine richness and diversity, OTUs were identified at genetic distances of 3% (species level), 5% (genus level) and 20% (phylum level) by using quality sequences with a read length of $\geq$ 150 bp per sample. At 20% sequence divergence, most rarefaction curves showed saturation, indicating that the surveying effort covered almost the full extent of taxonomic diversity at this genetic distance (Fig. S2). Comparison of the rarefaction analyses with the number of OTUs determined by the Chao1 richness estimator revealed that 83% to 100% (20% genetic distance) of the estimated

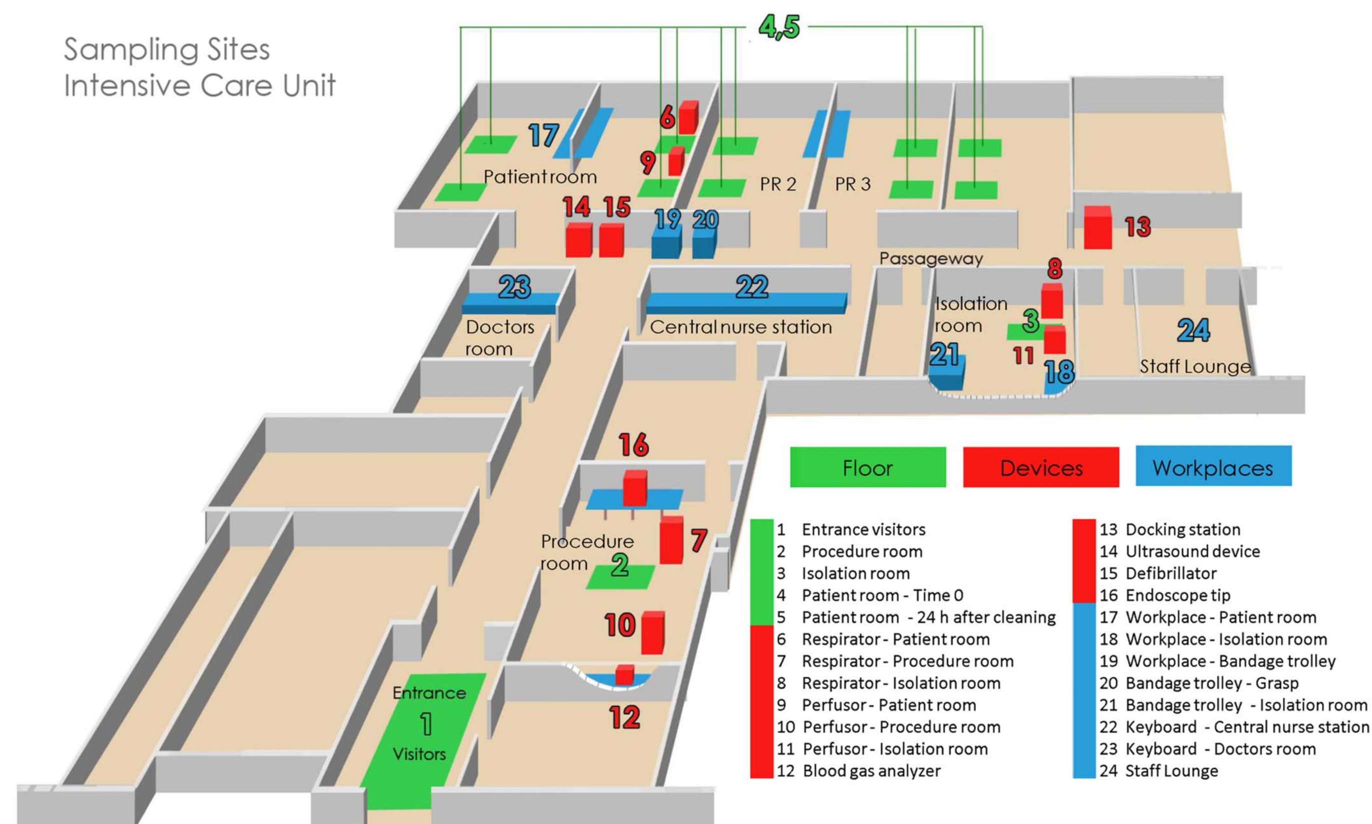


Figure 1 | Cartoon illustration showing each sampling site of the intensive care unit (ICU). Origin of the single samples is explained in Table 1.

Floor

Actinobacteria: *Cellulomonas, Micrococcus, Nocardioides;* **Firmicutes:** *Lachnospiraceae Incertae Sedis, Subdoligranulum;* **Proteobacteria:** *Defluviicoccus, Enterobacter, Rhodovulum*

Bacteroidetes: *Niastella, Terrimonas;* **Proteobacteria:** *Sphingomonas*

Devices

Acidobacteria: Gp3, Gp7, *Kocuria, Propionicimonas,* **Bacteroidetes:** *Cloacibacterium, Flavobacterium;* **Firmicutes:** *Bacillus, Gemella, Granulicatella, Paenibacillus;* **Nitrospirae:** *Nitrospira;* **Proteobacteria:** *Actinobacillus, Aquabacterium, Brevundimonas, Byssovorax, Massilia, Paucibacter, Phyllobacterium, Rhodoplanes, Simplicispira, Sphingobium, Stenotrophomonas, Variovorax*

Actinobacteria: *Propionibacterium;* **Firmicutes:** *Staphylococcus, Streptococcus;* **Proteobacteria:** *Acinetobacter, Bradyrhizobium, Burkholderia, Delftia, Enhydrobacter, Pseudomonas, Serratia*

Actinobacteria: *Corynebacterium,* **Firmicutes:** *Anaerococcus, Facklamia, Lactobacillus*

Acidobacteria: Gp2; **Bacteroidetes:** *Chryseobacterium;* **Firmicutes:** *Lactococcus;* **Proteobacteria:** *Acidovorax, Caedibacter, Curvibacter, Janthinobacterium, Legionella, Mesorhizobium, Methylobacterium, Novosphingobium, Shigella*

Workplaces

Acidobacteria: Gp11, Gp6, *Curtobacterium, Microbacterium;* **Bacteroidetes:** *Sphingobacterium;* **Firmicutes:** *Cohnella, Enterococcus, Leuconostoc, Peptoniphilus, Peptostreptococcus, Veillonella;* **Proteobacteria:** *Herbaspirillum, Pigmentiphaga, Psychrobacter, Sphingopyxis*

Figure 2 | Schematic drawing showing the detected bacterial genera found in the three different areas (floor, devices and workplaces) in the intensive care unit (ICU). Overlaps between the facilities are indicated by the arrangement of the circles.

taxonomic richness was detected. At 3% and 5% genetic distance, the rarefaction curves were not saturated and the richness estimators indicated that 45% to 78% and 47% to 84% of the estimated richness were recovered. As a result, we did not survey the full extent of taxonomic diversity at these genetic distances, but a substantial fraction of the bacterial diversity within individual samples was assessed at species and genus level. The Shannon index of diversity (H′) was determined for all samples (Table 2). The highest bacterial diversity at a genetic distance of 3% was found on workplaces (3.46), followed by the floor (3.14) and devices (3.0). The Shannon index of each sampling site ranged from 2.29 to 4.64; the surfaces of the bandage trolley (MID38) and workplaces of the patient room (MID37) showed the highest diversity.

Differences between the bacterial community profiles. Using the software package QIIME[17], the final OTU table for principal coordinate analysis (PCoA) comprised 3,925 OTUs and was distributed into 556 OTUs represented by more than 10 sequences. Beta diversity of the bacterial communities within the ICU revealed clear distinctions between bacterial populations among the three areas. Floor-associated bacterial communities formed clusters distinct from devices (Fig. 4), while the analyzed samples from workplaces and devices were similar. Samples from workplaces of the patient room (MID37) and bandage trolley (MID38) were significantly differentiated from those of the other workplaces and were closer to samples from the floor. Furthermore, the structure of the bacterial community found on the docking station (MID53) was completely distinct from other communities.

To gain better insight into the differences of the three areas, we applied a profile clustering network analysis (Fig. 5). This profile obtained by a Cytoscape network analysis showed the most abundant 40 OTUs and highlighted the relative distribution and abundances. *Acinetobacter* was the most abundant and ubiquitous bacterial genus with dominant occurrence on the floor. In addition, *Bradyrhizobium* and *Burkholderia* were among the dominant genera of all areas.

According to the statistical analysis, 330 out of 3,925 examined OTUs showed significant differences between the floor environment (A) and devices (B), and 336 between devices (B) and workplaces (C). A comparison of the floor environment (A) and the workplaces (C) resulted in a statistically significant difference between species for

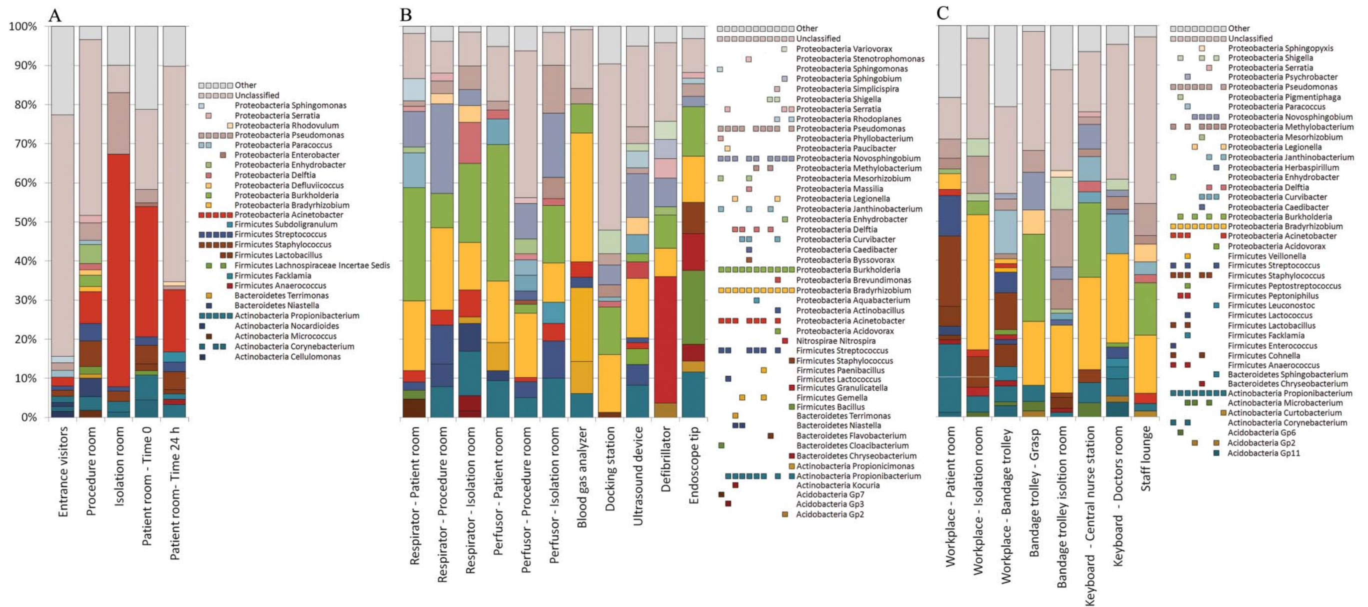


Figure 3 | The bacterial communities of the floor environment (A), medical devices (B) and workplaces (C). Relative clone composition of genera was determined by pyrosequencing of 16S rRNA from metagenomic DNA extracted from the intensive care unit. The identification of the closest strain based on 16S rRNA sequence similarity was achieved using the web server SnoWMAn 1.11 (http://snowman.genome.tugraz.at). Phylogenetic groups accounting for 1% of all quality sequences are summarized in the artificial group *Other*. Multi-colored charts at the legend are shown for each genus and sample correspondingly.

155 OTUs. Statistical analysis of data is shown in Tab. S1. Most notably, in numerous samples of the floor environment (up to 59%) and in some samples from devices and workplaces, the relative abundance of *Unclassified* was particularly high. Comparison of the relative abundances indicated a correlation of Cyanobacteria at phylum level with the *Unclassified* at genus level. Additional BLAST analysis of the *Unclassified* detected that most sequences classified as Cyanobacteria-like are chloroplast sequences that originate from

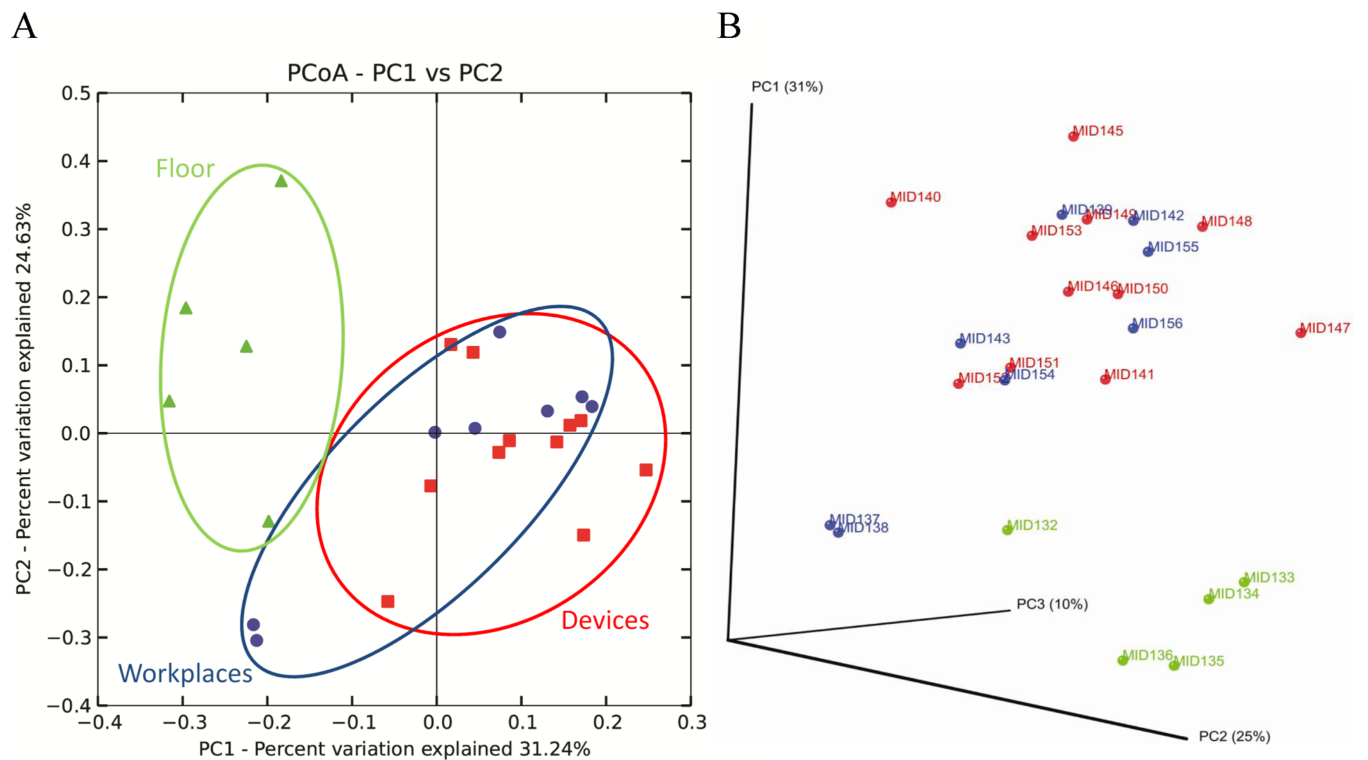


Figure 4 | Bacterial communities associated with different areas of the intensive care unit by principal coordinate analysis (PCoA). Two- (A) and three-dimensional (B) PCoA plot based on the weighted UniFrac distance matrix. Percentage of the diversity distribution explained by each axes is indicated on the figure. Samples associated with the floor (green triangle), medical devices (red rectangle) and workplaces (blue points) are shown as single points.

Figure 5 | Profile clustering Cytoscape network visualize the 40 most abundant OTUs across the floor (green), medical devices (red) and workplaces (blue) of the ICU. Node sizes correspond to the mean relative abundance between the three sampling areas where two (three) sampling groups feature the same relative abundance over the respective samples and the full nodes are subdivided into two (three) areas. A comparative node (grey) indicates the size of a node that would represent 100 reads in a sample.

Pinaceae and from other plant components. Sequences were affiliated to species of conifer in the Pinaceae family containing *Abies* sp., *Larix* sp., *Picea* sp. or *Pinus* sp. with maximal identity of 99% and originated from *Pinus* pollen from outside air. Abundance of chloroplast sequences is shown in Tab. S2.

Comparison between standard cultivation and 16S pyrosequencing. Comparative samples were taken from 10 sampling sites of defined positions on devices and workplaces (Table 1). The colonies showed a low diversity in their morphology and colour (white approx. 80%; the others yellow and red). The highest number of colonies (512) was found on the keyboard in the central nurse station, but only two colonies were detected on the workplace of the patient room. The number of colony forming units (CFUs) of all sampling sites is listed in Table 1.

A total of 130 isolates obtained from contact plates were characterized by ARDRA (amplified ribosomal RNA gene restriction analysis using *Hha*I) and divided into 36 ARDRA groups at a cutoff level of 85%. Representative strains of each group were partially sequenced and identified by their 16S rRNA; the genera *Aerococcus*, *Arthrobacter*, *Bacillus*, *Corynebacterium*, *Kocuria*, *Micrococcus*, *Paenibacillus*, *Planomicrobium*, *Roseomonas* and *Staphylococcus* were identified. Twelve ARDRA groups include the majority of isolates and were identified as species belonging to the genus *Staphylococcus*. In contrast, 21 groups were represented by only one isolate. To analyze the genotypic diversity within the ARDRA groups at population level, BOX-PCR patterns of the whole bacterial genome were used. At 80% similarity, 24 unique populations were differentiated (Table 3). With the exception of the *Roseomonas mucosa* strain (ARDRA group 34), only Gram-positive species were found; 49 of them were identified as *Staphylococcus epidermidis* and *Staphylococcus hominis*. Other *Staphylococcus* sequences were classified as *S. auricularis*, *S. caprae*, *S. cohnii, S. haemolyticus*, *S. lugdunensis* and *S. warneri*. Less retrieved sequences belonged to *Micrococcus antarcticus, M. luteus, Arthrobacter agilis, Bacillus circulans, B. idriensis* and *Kocuria palustris.* Furthermore, a few sequences were identified as *Aerococcus urinaeequi*, *Bacillus aerophilus*, *B. frigoritolerans*, *B. herbersteinensis*, *B. simplex*, *Corynebacterium propinquum*, *Kocuria rosea*, *Paenibacillus barcinonensis*, *Planomicrobium koreense* and *Roseomonas mucosa*. Interestingly, the keyboard of the central nurse station showed the highest number of CFUs (512) containing six different species. In comparison, the respirator from the patient room contained only 15 colonies, but nine different species were identified. Isolates with identical and similar BOX patterns were detected on nearly all sampling sites (Fig. 6) indicating transmission between them or deposition of bacteria from identical or similar sources.

Culture-dependent identification was compared with 16S rRNA gene 454-pyrosequencing analysis. While OTUs taken from the

Table 1 | Locations and characteristics of sampling points for cultivation-independent and -dependent methods of the intensive care unit (ICU)

(A) Samples for cultivation-independent analysis

Sampling location	Description	Sampling tool[1]	Sample ID	MID
Entrance visitors	Floor	BiSKit	F1	MID32
Procedure room	Floor	BiSKit	F2	MID33
Isolation room	Floor	BiSKit	F3	MID34
Patient room - Time 0	Floor	BiSKit	F4	MID35
Patient room - 24 h after cleaning	Floor	BiSKit	F5	MID36
Respirator - Patient room	Device	Nylon flocked swab	D6	MID45
Respirator - Procedure room	Device	Nylon flocked swab	D7	MID46
Respirator - Isolation room	Device	Nylon flocked swab	D8	MID47
Perfusor - Patient room	Device	Nylon flocked swab	D9	MID48
Perfusor - Procedure room	Device	Nylon flocked swab	D10	MID49
Perfusor - Isolation room	Device	Nylon flocked swab	D11	MID50
Blood gas analyzer	Device	Nylon flocked swab	D12	MID51
Docking station	Device	Nylon flocked swab	D13	MID53
Ultrasound device	Device	Nylon flocked swab	D14	MID41
Defibrillator	Device	Nylon flocked swab	D15	MID40
Endoscope tip	Device	Nylon flocked swab	D16	MID52
Workplace - Patient room	Working Surface	BiSKit	W17	MID37
Workplace - Isolation room	Working Surface	BiSKit	W18	MID54
Workplace - Bandage trolley	Working Surface	BiSKit	W19	MID38
Bandage trolley - Grasp	Working Surface	Nylon flocked swab	W20	MID42
Bandage trolley in isolation room	Working Surface	Nylon flocked swab	W21	MID43
Keyboard - Central nurse station	Working Surface	Nylon flocked swab	W22	MID55
Keyboard - Doctors room	Working Surface	Nylon flocked swab	W23	MID56
Staff Lounge	Working Surface	Nylon flocked swab	W24	MID39

(B) Samples for cultivation-dependent analysis

Sampling location	Description	Sampling tool[2]	Sample ID	CFU
Workplace - Patient room	Working Surface	Contact test	WS 42	2
Workplace - Bandage trolley	Working Surface	Contact test	WB 43	21
Respirator - Patient room	Device	Contact test	RS 44	15
Perfusor - Patient room	Device	Contact test	PS 45	210
Keyboard - Central nurse station	Device	Contact test	KO 46	512
Keyboard - Doctors room	Device	Contact test	KD 47	48
Blood gas analyzer	Device	Contact test	BG 48	72
Docking station	Device	Contact test	DS 49	87
Ultrasound device - Control panel	Device	Contact test	UU 50	6
Ultrasound device - Probe	Device	Contact test	UU 51	25

Abbreviations: BiSKit, Biological Sampling Kit; MID, Multiplex Identifier Sequence; CFU, Colony Forming Units per sample.
[1,2] Respective sample areas were 1 m^2 for BiSKits, 25 cm^2 for Nylon flocked swabs and for 10 cm^2 contact tests.

amplicon libraries were affiliated with 405 different genera (76 genera ≥ 1% of relative abundance), standard cultivation obtained only 10 bacterial genera corresponding to 2.5% of the total bacterial diversity. Most of the reference sequences of isolates presented an exact match with the pyrosequencing data, but some MIDs did not reach 1% of the relative abundance. Complete linkage clustering indicated that sequences of *Micrococcus luteus* (Sequence ID: 43/6; max. identity: 92%), *Corynebacterium propinquum* (49/11; 99%) and *Bacillus aerophilus* (43/16; 79%) were not present in the cultivation-independent data because their similarity values were lower (Table 3) in comparison to the other sequences.

Discussion

In this study we found a much higher diversity of bacterial communities in the ICU by using the 16S pyrosequencing approach than compared to the standard cultivation technique. Distinct profiles between the floor environment, medical devices and workplaces were found using both strategies. However, various ubiquitous taxa as well as genotypically identical strains were frequently observed.

Although the surfaces in ICUs were characterized by highly diverse bacterial communities, they were actually reduced in comparison with other indoor environments such as living, patient, class or rest rooms[3]. Similar to other indoor microbial communities, these communities were partially colonized by human-associated bacteria. Although it is impossible to predict the pathogenicity of a strain based on 16S rRNA sequences, the proportion of bacteria identified as those genera/species closely related to human pathogens was very high. They are known for their facultative pathogenic and nosocomial character, e.g. *Acinetobacter, Stenotrophomonas*, *Burkholderia, Flavobacterium, Propionibacterium, Pseudomonas, Staphylococcus* and *Escherichia/Shigella*[18,19]. Humans are not only the most important dispersal vectors for bacteria inside rooms[11]; their bacterial fingerprint represents a unique mix of bacteria including pathogens[20]. Therefore, patients in the ICU may have contributed to this high proportion of potential pathogens. To evaluate this hypothesis, we compared our data with the infections acquired by patients during this time (February-May 2011) in the ICU and found several overlaps which could potentially confirm the potential pathogenic character of several surface-associated bacteria. Altogether, from 101 bacterial infections, the majority was caused by *Staphylococcus* (40; *S. aureus, S. epidermidis* and spec. div.). However, Gram-negative pathogens were also identified, e.g. according to their abundance: *E. coli, Klebsiella, Pseudomonas, Serratia, Enterobacter, Edwardsiella, Proteus* and *Chryseobacterium*. Conversely, we found bacterial genera,

Table 2 | Species richness estimates obtained at 3%, 5% and 20% genetic dissimilarity from pyrosequencing of 16S rRNA from metagenomic DNA extracted from the floor, medical devices and workplaces

	Shannon index (H′)[a]			Rarefaction[b] (No. of OTUs)			Chao1[c] (No. of OTUs)			Coverage (%)		
Dissimilarity cut-off[d]	3%	5%	20%	3%	5%	20%	3%	5%	20%	3%	5%	20%
Entrance visitors	4.09	3.64	2.24	778.46	553.66	97.97	1713.009	1035.833	121.100	45.5	53.5	80.9
Procedure room	3.03	2.73	2.08	198.9	129.94	17	361.303	223.261	17.000	55.1	58.2	100.0
Isolation room	2.29	2.04	1.24	347.8	250.87	34	778.055	525.405	35.500	44.7	47.8	95.8
Sick chamber - Time 0	3.54	3.02	2.23	571.68	412.8	56.99	1234.414	740.217	63.429	46.3	55.8	89.9
Sick chamber - 24 h after cleaning	2.74	2.40	1.60	200.9	131.94	24	354.270	222.048	26.000	56.7	59.4	92.3
Respirator - Patient room	2.86	2.51	1.80	163.94	98.97	13	210.923	117.913	13.000	77.8	84.0	100.0
Respirator - Procedure room	2.81	2.60	1.95	181.92	95.98	12	294.857	116.313	12.000	61.7	82.5	100.0
Respirator - Isolation room	3.07	2.70	2.05	167.93	94.97	14	233.022	122.750	17.000	72.1	77.4	82.4
Perfusor - Patient room	2.97	2.45	1.58	153.94	90.96	11	247.261	203.875	11.000	62.3	44.6	100.0
Perfusor - Procedure room	3.44	3.13	2.14	224.89	143.94	15	399.000	230.667	15.000	56.4	62.4	100.0
Perfusor - Isolation room	2.89	2.59	1.90	147.93	80.98	10	247.400	99.056	10.000	59.8	81.8	100.0
Blood gas analyzer	2.90	2.58	1.82	147.94	84.97	10	226.792	109.800	10.000	65.3	77.4	100.0
Docking station	3.10	2.81	2.06	242.9	160.95	28	397.895	220.111	29.500	61.1	73.1	94.9
Ultrasound device	3.11	2.90	1.90	181.92	119.96	16	297.161	154.440	16.000	61.2	77.7	100.0
Defibrillator	2.78	2.38	1.88	154.93	102.97	14	224.000	121.207	14.000	69.2	85.0	100.0
Endoscope tip	3.00	2.66	1.85	173.92	100.95	16	356.045	181.571	17.500	48.9	55.6	91.4
Workplace - Patient room	4.11	3.73	2.44	425.83	301.91	42	653.586	386.875	42.250	65.2	78.1	99.4
Workplace - Isolation room	3.05	2.77	1.75	169.93	101.97	14	266.871	146.400	14.000	63.7	69.7	100.0
Workplace - Bandage trolley	4.64	4.24	2.82	572.74	404.85	52.99	1073.039	579.690	63.500	53.4	69.9	83.5
Bandage trolley - Grasp	2.94	2.55	1.61	136.95	67.98	9	192.682	83.833	9.000	71.1	81.1	100.0
Bandage vehicle isoltion room	3.57	3.29	2.32	295.86	193.93	27	539.886	299.556	27.000	54.8	64.8	100.0
Keyboard - Central nurse station	3.34	3.04	1.76	181.94	111.97	15	235.318	137.143	15.000	77.3	81.7	100.0
Keyboard - Doctors room	3.45	3.14	2.03	197.92	118.96	16	306.243	153.167	16.000	64.7	77.7	100.0
Staff Lounge	2.60	2.26	1.50	141.95	74.98	7	224.650	117.857	7.000	63.2	63.6	100.0

[a]a higher number indicates more diversity.
[b]the results from the rarefaction analyses are also depicted in Figure 4.
[c]nonparametric richness estimator based on the distribution of singletons and doubletons.
[d]Clusters were obtained with 3%, 5% and 20% dissimilarity cut-offs which correspond to the taxonomic level of species, genera and phyla.

e.g. *Burkholderia*, *Pseudomonas*, *Lactobacillus*, or *Methylobacterium* which contain plant-associated taxa that can also undergo bivalent interactions with humans. Although they can cause facultative infections in those with certain predispositions, they can also live in symbiosis with plants or can be used as pro- and prebiotics for both plants and humans[21–24]. Their origin and function in hospital environments is still unclear. However, one method of transmission could be the transport via pollen into the hospital environment due to the detection of pollen as a vector for specific plant-associated bacteria[25]. We found a high proportion of chloroplast sequences from *Pinus* pollen - *Pinus* trees were among the most common plants outside - in the floor environment and on several devices. These sequences were also detected and discussed in other pyrosequencing-based studies[1,3]. In addition to air conditioning, the investigated ICU was also window-ventilated, which has been known to result in an increased abundance of chloroplast DNA than in exclusively mechanically ventilated rooms[3].

The bacterial communities from three general area floors, medical devices, and workplaces were characterized by a specific and distinct composition. Skin-associated genera (*Propionibacterium*, *Corynebacterium*, *Staphylococcus*, *Streptococcus*, *Bradyrhizobium*)[26] were highly abundant on medical devices and working surfaces, which was expected considering that they are frequently touched by hands of hospital staff: a typical hand surface harbored on average more than 150 unique species-level bacterial phylotypes[20]. Interestingly, genera of *Burkholderia* and *Bradyrhizobium* were most abundant on devices and workplaces and are both originally plant-associated genera with the potential to fix nitrogen, but are also able to colonize hospital water supplies and surfaces[27]. In contrast, the floor environment contained genera that are typically distributed in environmental sources, such as soil or water. *Acinetobacter* was the dominant genus of the floor sample but was also present on almost all devices. During recent years, antibiotic-resistant *Acinetobacter* infections have become an increasingly common nosocomial problem[28,29]. Another emerging nosocomial pathogen *Clostridium difficile* was fortunately not detected in our amplicon libraries. Due to its long-living spores, this bacterium is often found and can cause large-scale outbreaks of nosocomial diarrhea[5].

Although bacterial communities in the ICU could be effectively differentiated, connections and transmissions were also detected by principal coordinate- and network analyses. For example, samples of the bandage trolley (MID38) and the workplace in the patient room (MID37) are located between clusters of devices and floor. This bandage trolley is used the whole day in several rooms by different persons of the staff. Additionally, the workplace of the patient room is frequently in contact with the hospital staff, which can explain the transfer. In addition, these two sampling sites were characterized by the highest bacterial diversity (Shannon diversity indices: MID38 4.64; MID37 4.11). Transmission of bacteria from one site to the other was also found comparing molecular fingerprints of the isolated strains. For example, similar BOX fingerprints of staphylococci were identified on nearly all sampling sites. However, deposition of bacterial strains from identical or similar sources, e.g. by personal staff, can also explain this finding. Surface sanitation is an often overlooked, yet crucial component of transmission[15], which should be considered more in sanitation protocols. No differences were observed between the bacterial communities of the isolation room for patients who were temporarily isolated with the risk of spreading an infectious disease or were severely immunocompromised and the other patient rooms. This observation is most likely because the isolation room was not in use at the time of sampling.

Although it is well-known that cultivation-dependent techniques capture only a small part of the microbiome[6,30], we found an unexpected high difference between the bacterial diversity using both

Table 3 | Cultivation dependent ARDRA (amplified ribosomal RNA gene restriction analysis; cutoff level of 85%) and BOX analysis (cutoff level of 85%) of 130 isolates

ARDRA group	BOX group	Sampling site	Sample-ID	Closest NCBI database match	Accession No.	Max. identity
1	1	KD	47/8	*Micrococcus luteus*	NR_037113.1	99%
2	1	DS	49/11	*Corynebacterium propinquum*	NR_037038.1	99%
3	1	WB	43/24	*Roseomonas mucosa*	NR_028857.1	99%
4	1	DS	49/16	*Staphylococcus hominis*	NR_041323.1	99%
4	2	DS	49/9	*Staphylococcus hominis*	NR_041323.1	99%
4	3	DS	49/1	*Staphylococcus hominis*	NR_041323.1	99%
4	4	KD	47/15	*Staphylococcus caprae*	NR_024665.1	100%
4	4	DS	49/13	*Staphylococcus caprae*		
4	5	UU	51/12	*Staphylococcus epidermidis*	NR_036904.1	99%
4	5	UU	51/9	*Staphylococcus epidermidis*		
4	5	KD	47/12	*Staphylococcus epidermidis*		
4	5	UU	51/2	*Staphylococcus epidermidis*		
4	5	KD	47/10	*Staphylococcus epidermidis*		
4	5	UU	51/10	*Staphylococcus epidermidis*		
4	5	KD	47/16	*Staphylococcus epidermidis*		
4	5	KD	47/5	*Staphylococcus epidermidis*		
4	5	UU	51/17	*Staphylococcus epidermidis*		
4	5	DS	49/14	*Staphylococcus epidermidis*		
4	5	UU	51/18	*Staphylococcus epidermidis*		
4	5	UU	50/1	*Staphylococcus epidermidis*		
4	5	UU	51/13	*Staphylococcus epidermidis*		
4	5	DS	49/7	*Staphylococcus epidermidis*		
4	5	UU	51/4	*Staphylococcus epidermidis*		
4	5	UU	51/14	*Staphylococcus epidermidis*	NR_036904.1	99%
4	5	KD	47/14	*Staphylococcus epidermidis*		
4	6	KD	47/17	*Staphylococcus hominis*	NR_041323.1	99%
4	7	KD	47/13	*Staphylococcus caprae*	NR_024665.1	99%
4	8	DS	49/10	*Staphylococcus epidermidis*	NR_036904.1	100%
4	9	DS	49/6	*Staphylococcus warneri*	NR_025922.1	100%
4	10	DS	49/4	*Staphylococcus hominis*		
4	10	UU	50/3	*Staphylococcus hominis*		
4	10	UU	50/2	*Staphylococcus hominis*		
4	10	UU	50/5	*Staphylococcus hominis*	NR_041323.1	98%
4	10	DS	49/15	*Staphylococcus hominis*		
4	11	DS	49/3	*Staphylococcus caprae*	NR_024665.1	100%
4	12	DS	49/2	*Staphylococcus cohnii*	NR_036902.1	100%
4	13	DS	49/8	*Staphylococcus epidermidis*	NR_036904.1	100%
4	14	UU	51/6	*Staphylococcus epidermidis*	NR_036904.1	100%
4	15	UU	50/4	*Staphylococcus hominis*	NR_041323.1	98%
5	1	KD	47/11	*Staphylococcus epidermidis*	NR_036904.1	99%
6	1	BG	48/18	*Staphylococcus epidermidis*	NR_036904.1	100%
7	1	KO	46/17	*Staphylococcus warneri*	NR_025922.1	100%
7	2	KO	46/21	*Staphylococcus epidermidis*	NR_036904.1	100%
7	3	KO	46/18	*Staphylococcus hominis*		
7	3	KO	46/19	*Staphylococcus hominis*	NR_036956.1	99%
8	1	KO	46/20	*Staphylococcus epidermidis*		
8	1	KO	46/22	*Staphylococcus epidermidis*	NR_036904.1	99%
9	1	KD	47/1	*Staphylococcus epidermidis*	NR_036904.1	98%
9	1	BG	48/5	*Staphylococcus epidermidis*		
9	1	BG	48/1	*Staphylococcus epidermidis*		
9	2	BG	48/10	*Staphylococcus epidermidis*	NR_036904.1	100%
9	3	BG	48/7	*Staphylococcus epidermidis*	NR_036904.1	100%
9	4	BG	48/6	*Staphylococcus haemolyticus*	NR_036955.1	99%
9	5	BG	48/11	*Staphylococcus hominis*	NR_041323.1	99%
9	5	BG	48/12	*Staphylococcus hominis*		
9	5	BG	48/9	*Staphylococcus hominis*		
9	5	BG	48/13	*Staphylococcus hominis*		
9	6	BG	48/3	*Staphylococcus lugdunensis*	NR_024668.1	100%
10	1	BG	48/14	*Staphylococcus hominis*	NR_041323.1	99%
10	2	BG	48/15	*Staphylococcus hominis*	NR_041323.1	99%
11	1	BG	48/2	*Planomicrobium koreense*		
11	1	BG	48/8	*Planomicrobium koreense*	NR_025011.1	100%
12	1	UU	51/16	*Bacillus idriensis*	NR_043268.1	99%
12	1	UU	51/3	*Bacillus idriensis*	NR_043268.1	99%
13	1	KO	46/23	*Paenibacillus barcinonensis*	NR_042272.1	99%
14	1	UU	51/15	*Staphylococcus auricularis*	NR_036897.1	99%
14	2	UU	51/7	*Staphylococcus auricularis*	NR_036897.1	99%
15	1	KD	47/3	*Arthrobacter agilis*	NR_026198.1	98%

Table 3 | Continued

ARDRA group	BOX group	Sampling site	Sample-ID	Closest NCBI database match	Accession No.	Max. identity
16	1	UU	51/8	*Bacillus herbersteinensis*	NR_042286.1	96%
17	1	WS	42/2	*Staphylococcus hominis*	NR_036956.1	99%
17	2	RS	44/11	*Staphylococcus haemolyticus*	NR_036955.1	99%
17	3	KO	46/7	*Staphylococcus warneri*	NR_025922.1	100%
17	4	WB	43/18	*Staphylococcus hominis*		
17	4	PS	45/5	*Staphylococcus hominis*		
17	4	PS	45/8	*Staphylococcus hominis*		
17	4	WB	43/13	*Staphylococcus hominis*	NR_041323.1	99%
17	5	WB	43/17	*Staphylococcus epidermidis*		
17	5	KO	46/16	*Staphylococcus epidermidis*		
17	5	WB	43/2	*Staphylococcus epidermidis*		
17	5	KO	46/3	*Staphylococcus epidermidis*		
17	5	KO	46/15	*Staphylococcus epidermidis*		
17	5	RS	44/3	*Staphylococcus epidermidis*	NR_036904.1	100%
17	6	KO	46/6	*Staphylococcus epidermidis*	NR_036904.1	99%
17	7	PS	45/6	*Staphylococcus epidermidis*	NR_036904.1	100%
17	8	PS	45/7	*Staphylococcus hominis*	NR_041323.1	99%
17	9	PS	45/3	*Staphylococcus hominis*	NR_041323.1	98%
17	10	KO	46/13	*Staphylococcus hominis*		
17	10	KO	46/2	*Staphylococcus hominis*		
17	10	KO	46/11	*Staphylococcus hominis*	NR_041323.1	98%
17	11	KO	46/14	*Staphylococcus lugdunensis*	NR_024668.1	100%
17	12	RS	44/18	*Staphylococcus hominis*		
17	12	KO	46/10	*Staphylococcus hominis*		
17	12	RS	44/12	*Staphylococcus hominis*		
17	12	RS	44/2	*Staphylococcus hominis*		
17	12	PS	45/11	*Staphylococcus hominis*		
17	12	PS	45/4	*Staphylococcus hominis*		
17	12	WS	42/1	*Staphylococcus hominis*		
17	12	RS	44/1	*Staphylococcus hominis*	NR_036956.1	99%
17	12	WB	43/12	*Staphylococcus hominis*		
17	12	WB	43/1	*Staphylococcus hominis*		
17	13	KO	46/9	*Staphylococcus warneri*	NR_025922.1	99%
18	1	WB	43/10	*Staphylococcus epidermidis*	NR_036904.1	100%
18	2	RS	44/17	*Staphylococcus epidermidis*	NR_036904.1	97%
19	1	RS	44/10	*Staphylococcus caprae*	NR_024665.1	99%
20	1	WB	43/4	*Kocuria rosea*		
20	1	WB	43/5	*Kocuria rosea*	NR_044871.1	99%
21	1	WB	43/23	*Bacillus circulans*	NR_042726.1	99%
21	2	WB	43/7	*Bacillus circulans*	NR_042726.1	99%
21	3	WB	43/16	*Bacillus aerophilus*	GU339271.1	79%
22	1	RS	44/14	*Bacillus frigoritolerans*		
22	1	RS	44/4	*Bacillus frigoritolerans*	NR_042639.1	100%
22	2	PS	45/10	*Bacillus simplex*	NR_042136.1	100%
23	1	BG	48/17	*Staphylococcus hominis*	NR_041323.1	99%
24	1	RS	44/16	*Micrococcus luteus*		
24	1	RS	44/5	*Micrococcus luteus*	NR_037113.1	97%
24	2	WB	43/6	*Micrococcus luteus*	NR_037113.1	92%
24	3	WB	43/3	*Micrococcus luteus*	NR_037113.1	99%
25	1	RS	44/8	*Arthrobacter agilis*	NR_026198.1	100%
26	1	DS	49/5	*Micrococcus luteus*	NR_037113.1	98%
26	2	UU	51/1	*Micrococcus luteus*	NR_037113.1	97%
27	1	KD	47/4	*Kocuria palustris*	NR_026451.1	98%
28	1	BG	48/16	*Micrococcus antarcticus*	NR_025285.1	99%
29	1	BG	48/4	*Micrococcus antarcticus*	NR_025285.1	99%
30	1	WB	43/9	*Micrococcus antarcticus*	NR_025285.1	99%
31	1	KD	47/7	*Micrococcus luteus*	NR_037113.1	99%
32	1	RS	44/9	*Micrococcus antarcticus*	NR_025285.1	99%
33	1	KD	47/9	*Micrococcus luteus*	NR_037113.1	99%
34	1	RS	44/13	*Kocuria palustris*	NR_026451.1	99%
35	1	PS	45/9	*Aerococcus urinaeequi*	NR_043443.1	99%
36	1	KO	46/4	*Micrococcus luteus*	NR_037113.1	99%

methods – standard cultivation and amplicon sequencing. While in the 16S rRNA gene amplicon library the amount of Gram-positive and Gram-negative bacteria was nearly the same, we detected almost exclusively Gram-positive bacteria by cultivation. With the exception of *Roseomonas mucosa* known to be associated with bacteremia and other human infections[31], cultivation failed to capture the Gram-negative spectrum. This similar proportion of Gram-negative and Gram-positive bacteria as in our libraries was also found in the bacterial infections acquired during this time in the ICU. While the Gram-positive spectrum comprised mainly *Staphylococcus* and

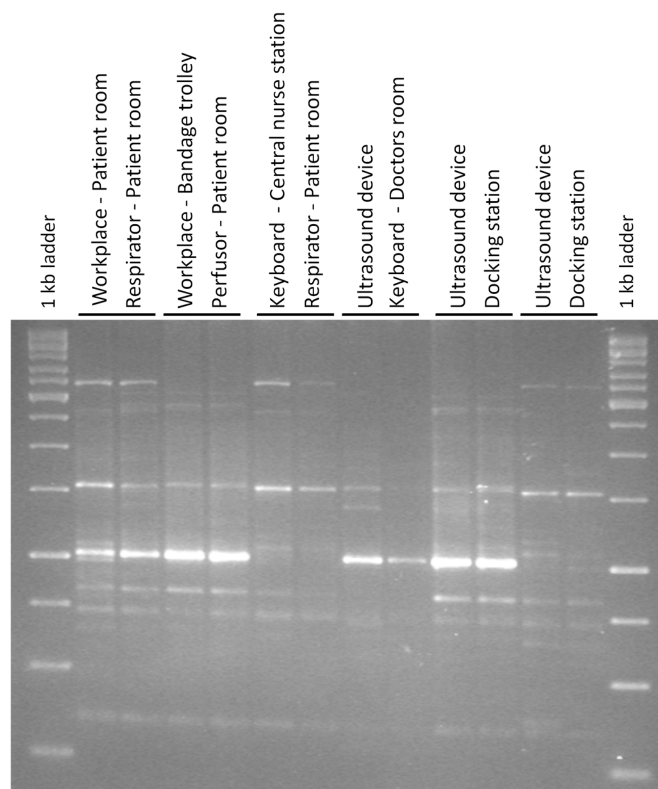


Figure 6 | BOX PCR fingerprints of genetically very similar *Staphylococcus* species isolated from nearly all sampling sites. The similarity of BOX patterns between them was more than 80%.

Enterococcus species, the Gram-negative pathogens were characterized by a higher taxonomic diversity. The Gram-positive diversity was well represented in the isolate collection, where the majority of the obtained cultivation-based sequences belonged to *Staphylococcus (S. epidermidis, S. hominis, S. auricularis*, *S. caprae, S. cohnii*, *S. haemolyticus*, *S. lugdunensis, S. warneri)*, which are already described as the most common bacteria in hospitals. Several of them such as *S. epidermidis, S. haemolyticus* or *S. warneri* emerged as causal agents of nosocomial infections with diverse resistances against antibiotics[32]. Interestingly, identical 16S rRNA gene sequences were found for isolates as well as in the amplicon library. Moreover, all sequences from isolates could be found in the amplicon libraries. However, due to the fact that the cultivation of Gram-negatives ultimately failed, new standard protocols should be developed to assess the overall diversity. For this cultivation, we used CASO agar plates that are synonymous to Tryptic Soy Agar (TSA) and Soybean Casein Digest Agar (CSA) suggested by both the European (EP) and United States Pharmacopoeia (USP). While it is difficult to explain why colonies of the well-cultivable bacteria genera such as *Burkholderia* and *Pseudomonas* were not isolated from plates, our cultivation results were highly similar to those obtained from the weekly routine monitoring and our 16S amplicon library was comparable with other pyrosequencing based studies[1–4]. However, we must consider that pyrosequencing based on DNA may also detect DNA from non-living and living bacteria. Light-activated ethidium monoazide or propidium monoazide can help to remove free extracellular DNA from environmental samples in the future[33].

In support of our hypothesis, we found an unexpected high diversity from the bacterial communities with bacteria closely related to human pathogens, but also taxa known for their beneficial interaction with eukaryotes. By using 16S pyrosequencing and corresponding network analysis, we have the tools to evaluate existing sanitation concepts. We suggest that the whole bacteria community should be considered in the assessment of these concepts because diversity within these communities often correlates with the ecosystem function of disease suppression[34]. If this can be proved for indoor microbiomes, the demand to "Stop killing beneficial bacteria" by Blaser[35] should be expanded.

Methods

Experimental design and sampling procedure. Samples were taken from selected surface areas and devices of the intensive care unit (ICU, Department of Internal Medicine, University hospital) in Graz/Austria. The intensive care unit contains 15 beds, including one isolation unit for severe immunocompromised patients. In this ICU the critically ill patients from all internal medicine subspecialities as well as neurologic patients are treated. All sampling locations and their characteristics are given in Table 1. During sampling, all employees and devices of the ICU were in full operation and eight large surface samples (1 m^2) were performed by using biological sampling kits (BiSKits; QuickSilver Analytics, Abingdon, MD, USA). For wet sampling of 1 m^2, BiSKits were premoistened with the manufacturer-provided sterile buffer[36] and the selected area was wiped in three different directions while rotating the sampling device[2]. Afterwards, samples were stored and chilled (4 to 8°C) during transportation, and frozen immediately at −70°C upon arrival at the laboratory.

Nylon flocked swabs (MicroRheologics, Copan, Brescia, Italy) were used to take samples of 16 devices. For sampling, swabs were moistened briefly in a 15 ml Falcon test tube containing 2.5 ml of sterile water[37]. After sampling of provided surfaces (5 × 5 cm), the swabs were broken into the Falcon test tubes and were kept cool (4 to 8°C). In the laboratory, microorganisms were extracted by vortexing (5 seconds) and sonication (84 W; 35 kHz; Sonorex super DK) for 120 seconds. Lastly, all samples were stored at −70°C. In addition, samples for cultivation were taken using CASO agar plates (Heipha Diagnostika, Eppelheim, Germany). In our study, ten contact tests of these CASO plates were performed according to the guidelines of the quarterly controls of the ICU. Samples were taken on defined positions and incubated at room temperature (RT) for 4 days. Colony forming units (CFU) were counted and 130 isolates were selected and subcultured on CASO agar. The isolates were purified and stored at −70°C in a nutrient broth (NB) containing 50% glycerol.

Total community DNA isolation. UV sterilized Amicon Ultra-15 filters (cutoff 50 kDa; Millipore GmbH, Schwalbach, Germany) were used for concentration of the sampling liquid from BiSKits. The resulting BiSKit suspension of each sampling liquid from swabs were subjected to DNA extraction using the modified XS buffer method[2]: XS buffer (2 x) was freshly prepared as follows: (20 ml stock solution): 1 M Tris/HCl (pH 7.4) (4 ml); 7 M ammonium acetate (4.56 ml); 250 mM ethylene diamine tetraacetic acid (3.2 ml); 10% sodium dodecyl sulfate (w/v) (4 ml); potassium ethyl xanthogenate (0.4 g); PCR-grade water (4.99 ml). For completely dissolving the xanthogenate, the buffer was incubated at 65°C for 15 min. Starting with 1 ml sample, in total, 1 ml of 2 × XS buffer was added, and the mixture was stirred gently (short vortex). After an incubation of 2 h at 65°C, and mixing by hand every 30 min, the suspension was vortexed for 10 seconds. The tube was placed on ice for 10 min and centrifuged afterwards (100 g, 5 min, 4°C). The supernatant was transferred into a PhaseLock Gel tube (Eppendorf, Hamburg, Germany), and an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added. The suspension was mixed gently and centrifuged (2000 g, 5 min, 15°C). The aqueous layer was transferred into a new tube. To precipitate DNA, the same volume of cold 100%isopropanol and 1/10 volume of 4 M ammonium acetate was added and gently mixed. After incubation at −20°C overnight, the suspension was centrifuged at 13500 g at 4°C for 30 min. The (invisible) pellet was washed with 1 ml 70% ethanol (ice cold) and centrifuged (13500 g, 30 min, 4°C). The pellet was then dried completely and dissolved in 20 µl PCR-grade water.

16S rRNA gene amplicon pyrosequencing. For pyrosequencing, the 16S rRNA gene was amplified in a nested PCR approach with the universal bacterial primer set 27f (5′-GAGTTTGATCMTGGCTCAG-3′) and 1492r (5′-AAGGAGGTGATCCANCCRCA-3′). The PCR reaction mixture (10 µl) contained 1 × Taq&Go, 0.25 µM of each primer and 1 µl of template DNA (95°C, 5 min; 30 cycles of 95°C, 30 s; 57°C, 30 s; 72°C, 90 s; and elongation at 72°C, 5 min). In a second PCR, 1 µl of the amplicon was used. 16S rRNA gene sequences were amplified by using the forward primer Unibac-II-515f (5′-GTGCCAGCAGCCGC-3) containing the 454-pyrosequencing adaptors (MIDs) and the reverse primer UnibacII927r_454 (5′-CCCGTCAATTYMTTTGAGTT-3′). Sequences of MIDs are listed in Tab. S3. The reaction mixture for the second PCR (60 µl) contained 1 × Taq&Go, 0.25 µM of each primer and 6 µl of the PCR product solution (95°C, 5 min; 32 cycles of 95°C, 20 s; 54°C, 15 s; 72°C, 30 s; and elongation at 72°C, 10 min). PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, USA). The partial 16S rRNA genes were sequenced using a Roche GS FLX+ 454 pyrosequencer (GATC Biotech, Konstanz, Germany).

Sequences that were shorter than 150 bp in length or of low quality were removed from the pyrosequencing-derived data sets using the pipeline initial process of the RDP pyrosequencing pipeline (http://pyro.cme.msu.edu)[38]. Due to the different number of sequences among samples, the data was normalized. The webserver SnoWMAn 1.11 (http://snowman.genome.tugraz.at)[39] for taxonomic-based analysis was used with the following settings: analysis type: BLAT pipeline; reference database: greengenes_24-Mar-2010; rarefaction method: RDP; taxonomy: RDP; confidence

threshold: 80%; include taxa covering more than: 1%. For rarefaction analysis, operational taxonomic units (OTUs) were clustered with 3% (species level), 5% (genus level) and 20% (phylum level) dissimilarity cut-offs[40,41]. Rarefaction curves were constructed by using the tools aligner, complete linkage clustering, and rarefaction of the RDP pyrosequencing pipeline. Shannon[42] and Chao1 indices[43] were calculated with the complete linkage clustering data. Principal coordinate analysis (PCoA) plots were generated using the open source software package QIIME (http://qiime.sourceforge.net), which allows analysis of high-throughput community sequencing data[17]. Network analysis was performed to visualize the most abundant taxa and to compare their abundance across the three sampling areas. The relative abundance of single OTUs was calculated for each sample and were used to generate an average value for each of the three areas. The open source software Cytoscape 2.8[16] was employed to visualize the 40 most abundant OTUs based on total read counts. To differentially detect abundant microbial clusters between the three areas clusters with ≥10 sequences were explored using Metastats web interface[44]. Bacterial populations of the ICU dataset were examined using a combination of the nonparametric t-test, exact Fisher's test, and the false discovery rate with 1000 permutations. P-values were determined for each cluster correspondingly.

Phylogenetic analysis of the isolates. Based on isolated colony DNA, Amplified ribosomal RNA gene restriction analysis (ARDRA) using the restriction enzyme HhaI (MP Biomedicals, Eschwege, Germany) was performed to cluster isolates with similar band pattern into genotypic groups according to Berg et al.[45]. Isolates with similar ARDRA patterns were clustered in one group and analyzed performing BOX-PCR fingerprinting. BOX-PCR was done using the BOX_A1R primer (5′-CTACGGCAAGGCGACGCTGACG-3′) as described by Rademaker and de Bruijn[46]. Computer-assisted evaluation of ARDRA- and BOX-PCR generated fingerprints were made using the GelCompar II software (version 5.1; Applied Math, Kortrijk, Belgium). Chosen isolates were identified by partial 16S rRNA gene sequencing at the sequencing core facility in Microsynth AG, Balgach, Switzerland. Obtained sequences were aligned using the NCBI sequence database and the BLAST algorithm. The 16S rRNA gene sequences of the isolates were deposited in the NCBI nucleotide sequence database under accession numbers HE962211-HE962235.

1. Flores, G. E. *et al.* Microbial biogeography of public restroom surfaces. *PLoS ONE* **6**, e28132 (2011).
2. Moissl-Eichinger, C. Archaea in artificial environments: Their presence in global spacecraft clean rooms and impact on planetary protection. *The ISME J* **5**, 209–219 (2011).
3. Kembel, S. W. *et al.* Architectural design influences the diversity and structure of the built environment microbiome. *The ISME Journal* **6**, 1469–1479 (2012).
4. Hewitt, K. M., Gerba, C. P., Maxwell, S. L. & Kelley, S. T. Office space bacterial abundance and diversity in three metropolitan areas. *PLoS ONE* **7**, e37849 (2012).
5. Daneman, N., Stukel, T. A., Ma, X., Vermeulen, M. & Guttmann, A. Reduction in *Clostridium difficile* infection rates after mandatory hospital public reporting: findings from a longitudinal cohort study in Canada. *PLoS Medicine* **9**, e1001268 (2012).
6. Staley, J. T. & Konopka, A. Measurements of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* **39**, 321–346 (1985).
7. Aundel, A. V., Sterling, E. M., Biggin, J. H. & Sterling, T. D. Indirect health effects of relative humidity in indoor environments. *Environ Health Perspect* **65**, 351–361 (1986).
8. Qian, J., Hospodsky, D., Yamamoto, N., Nazaroff, W. W. & Peccia, J. Size-resolved emission rates of airborne bacteria and fungi in an occupied class room. *Indoor Air* **22**, 339–5 (2012).
9. Vincent, J. L. *et al.* The prevalence of nosocomial infection in intensive care units in Europe. Results of the European Prevalence of Infection in Intensive Care (EPIC) Study. EPIC International Advisory Committee. *JAMA* **274**, 639–644 (1995).
10. Plowman, R. The socioeconomic burden of hospital acquired infection. *Euro Surveill* **5**, 49–50 (2000).
11. Klevens, R. M. *et al.* Estimating health care-associated infections and deaths in U.S. hospitals, 2002. *Public Health Rep* **122**, 160–166 (2007).
12. Hota, B. Contamination, disinfection, and cross-colonization: Are hospital surfaces reservoirs for nosocomial infection? *Clin Infectious Diseases* **39**, 1182–1189 (2004).
13. Gastmeier, P., Schwab, F., Bärwolff, S., Rüden, H. & Grundmann, H. Correlation between the genetic diversity of nosocomial pathogens and their survival time in intensive care units. *Journal of Hospital Infection* **62**, 181–186 (2005).
14. Kramer, A., Schwebke, I. & Kampf, G. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infect Diss* **6**, 130 (2006).
15. Otter, J. A. & French, G. L. Survival of nosocomial bacteria and spores on surfaces and inactivation by hydrogen peroxide vapor. *J Clin. Microbiol* **47**, 205–207 (2009).
16. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics.* **27**, 431–432 (2011).
17. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).
18. Clayton, J. J., Baig, W., Reynolds, G. W. & Sandoe, J. A. T. Endocarditis caused by *Propionibacterium* species: a report of three cases and a review of clinical features and diagnostic difficulties. *J Med Microbiol* **55**, 981–987 (2006).
19. Das, I. *et al.* Carbapenem-resistant *Acinetobacter* and role of curtains in an outbreak in intensive care units. *J Hosp Infection* **50**, 110–114 (2002).
20. Fierer, N., Hamady, M., Lauber, C. L. & Knight, R. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci U S A* **105**, 17994–17999 (2008).
21. Berg, G., Eberl, L. & Hartmann, A. The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environ Microbiol* **7**, 1673–1685 (2005).
22. Berg, G. Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. *Appl Microbiol Biotechnol* **84**, 11–18 (2009).
23. Ryan, R. P. *et al.* The versatility and adaptation of bacteria from the genus *Stenotrophomonas*. *Nat Rev Microbiol* **7**, 514–25 (2009).
24. Lai, C. C. *et al.* Infections caused by unusual *Methylobacterium* species. *J. Clin Microbiol* **49**, 3329–3331 (2011).
25. Fürnkranz, M. *et al.* Microbial diversity inside pumpkins: microhabitat-specific communities display a high antagonistic potential against phytopathogens. *Microb Ecol* **63**, 418–428 (2012).
26. Grice, E. A. & Segre, J. A. The skin microbiome. *Nature Rev Microbiol* **9**, 244–253 (2011).
27. Rutala, W. A. & Weber, D. J. Water as a reservoir of nosocomial pathogens. *Infect Control Hosp Epidemiol* **18**, 609–616 (1997).
28. Carey, R. B., Banerjee, S. N. & Srinivasan, A. Multidrug-resistant *Acinetobacter* infections, 1995–2004. *Presented at the 46th Interscience Conference on Antimicrobial Agents and Chemotherapy*, San Francisco (2006).
29. Munoz-Price, L. S. & Weinstein, R. A. Acinetobacter infection. *New England Journal of Medicine* **358**, 1271–81 (2008).
30. Schleifer, K. Microbial diversity: facts, problems and prospects. *Systematic Applied Microbiology* **27**, 3–9 (2004).
31. Christakis, B. B., Perlorentzou, S., Alexaki, P., Megalakaki, A. & Zarkadis, I. K. Central line-related bacteraemia due to *Roseomonas mucosa* in a neutropenic patient with acute myeloid leukaemia in Piraeus, Greece. *J Med Microbiol* **55**, 1153–1156 (2006).
32. Uçkay, I. *et al.* Foreign body infections due *to Staphylococcus epidermidis*. *Ann Med* **41**, 109–119 (2009).
33. Wagner, A. O., Malin, C., Knapp, B. A. & Illmer, P. Removal of free extracellular DNA from environmental samples by ethidium monoazide and propidium monoazide. *Appl Environ Microbiol.* **74**, 2537–2539 (2008).
34. Van Elsas, J. D. *et al.* Microbial diversity determines the invasion of soil by a bacterial pathogen. *Proc Natl Acad Sci U S A* **109**, 1159–1164 (2012).
35. Blaser, M. Antibiotic overuse: Stop the killing of beneficial bacteria. *Nature* **476**, 393–394 (2011).
36. Buttner, M. P. *et al.* Evaluation of the Biological Sampling Kit (BiSkit) for large-area surface sampling. *Appl Environ Microbiol* **70**, 7040–7045 (2004).
37. Probst, A., Facius, R., Wirth, R. & Moissl-Eichinger, C. Validation of a nylon-flocked-swab protocol for efficient recovery of bacterial spores from smooth and rough surfaces. *Appl Environ Microbiol* **76**, 5148–5158 (2010).
38. Cole, J. R. *et al.* The ribosomal database project: improved alignments and new tools for rRNA analysis *Nucleic Acids Res.* **37**, D141–D145 (2009).
39. Stocker, S. *et al.* SnoWMAn: High-throughput phylotyping, analysis and comparison of microbial communities. *Under revision* (2011).
40. Hur, I. & Chun, J. A method for comparing multiple bacterial community structures from 16S rDNA clone library sequences. *J Microbiol* **42**, 9–13 (2004).
41. Schloss, P. D. & Handelsman, J. Toward a census of bacteria in soil. *PLoS Comput Biol.* **2**, 786–793 (2006).
42. Shannon, C. E. The mathematical theory of communication. 1963. *MD Comput* **14**, 306–317 (1997).
43. Chao, A. & Bunge, J. Estimating the number of species in a sochastic abundance model. *Biometrics* **58**, 531–539 (2002).
44. White, J. R., Nagarajan, N. & Pop, M. Statistical methods for detecting deferentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* **5**, e1000352 (2009).
45. Berg, G. *et al.* Plant-dependent genotypic and phenotypic diversity of antagonistic rhizobacteria isolated from different Verticillium host plants. *Appl Environ Microbiol* **68**, 3328–3338 (2002).
46. Rademaker, J. L. W. & De Bruijn, F. J. Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer-assisted pattern analysis. In: Caetano-Anolle′s G, Gresshoff PM, eds. DNA markers: protocols, applications and overviews. New York: John Wiley & Sons. pp. 151–171 (1997).

Acknowledgements

We thank Anastasia Bragina (Graz) and Christine Moissl-Eichinger (Regensburg) for fruitful discussions and valuable technical support. Furthermore, we would also like to thank Meg Starcher (Graz/Washington) for English revision of the manuscript. This work has been supported by a grant to G.B. from the Federal Ministry of Economy, Family and Youth (BMWFJ), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol and ZIT - Technology

Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG.

Author contributions

Conceived and designed the experiments: G.B., C.H., K.H.S. Performed the experiments: L.O., C.Z. Analyzed the data: L.O., S.L., G.B. Wrote the first draft of the manuscript: L.O., G.B. Contributed to the writing of the manuscript: L.O., C.H., G.B., I.C.M.J.E. criteria for authorship read and met: L.O., C.Z., S.L., C.H., K.H.S., G.B. Agree with manuscript results and conclusions: L.O., C.Z., S.L., C.H., K.H.S., G.B.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/

How to cite this article: Oberauner, L. *et al*. The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing. *Sci. Rep.* **3**, 1413; DOI:10.1038/srep01413 (2013).

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing

Lisa Oberauner[1,2], Christin Zachow[1,2], Stefan Lackner[3], Christoph Högenauer[4], Karl-Heinz Smolle[4], and Gabriele Berg[2]

[1]*Austrian Centre of Industrial Biotechnology GmbH, Graz, Austria*
[2]*Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria*
[3]*Center for Medical Research, Medical University of Graz, Austria*
[4]*Department of Internal Medicine, Medical University of Graz, Austria*

Figure S1 Taxonomic classification of each sampling site. Pyrosequencing reads are classified at genus level with a confidence threshold of 80%. Groups below 1% of relative abundance are included in *Other*.

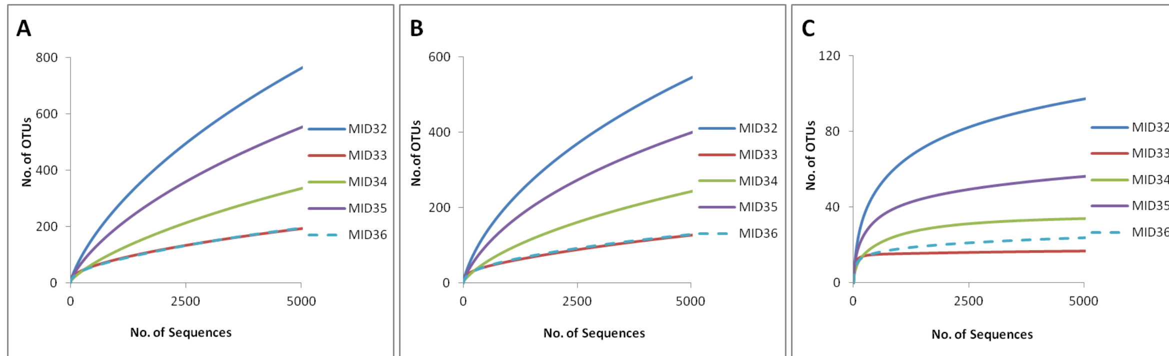
Figure S2 Rarefaction curves indicating the observed number of OTUs calculated by employing the tools Aligner, Complete Linkage Clustering, and Rarefaction of the RDP pyrosequencing pipeline. OTUs are shown at the 3% (A), 5% (B) and 20% (C) genetic distance levels of different areas of the intensive care unit. Origin of the single samples is explained in Table 1.

Table S1 Differentially abundant genera between the floor environment (A), medical devices (B) and workplaces (C) reaching more than 500 sequences over all samples.

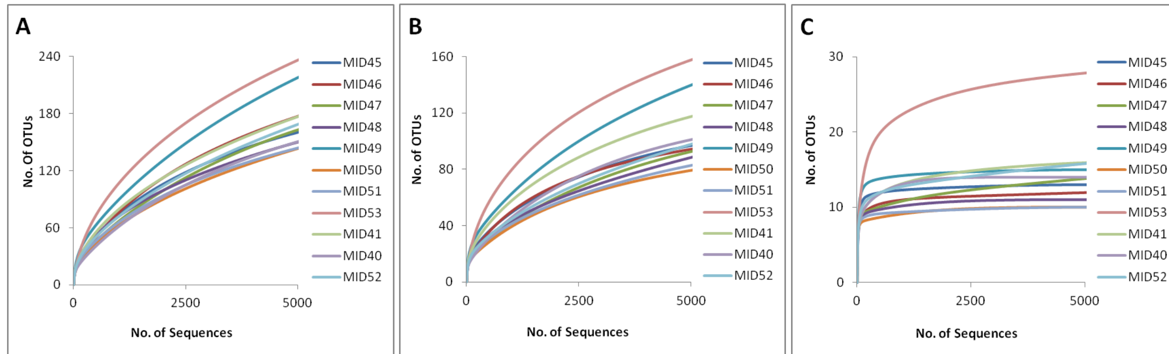
Table S2 Relative abundance (%) of chloroplast sequences of each sampling site.

Table S3 Multiplex identifier sequences (MIDs) for the 454 pyrosequencing approach.

Floor



Devices



Workplaces

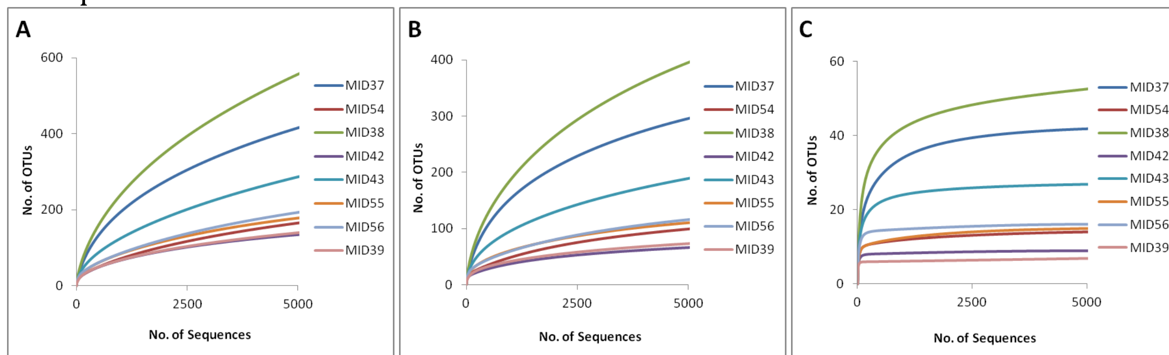


Figure S2

Table S1 Differentially abundant genera between the floor environment (A), medical devices (B) and workplaces (C) reaching more than 500 sequences over all samples.

OTU IDs	Genera	Abundance over all samples	p-values[a]		
			A vs. B	B vs. C	A vs. C
OTU_00174	*Bradyrhizobium*	7694	**2.8689E-05**	0.353646354	**0.00014644**
OTU_00533	*Burkholderia*	6521	**0.000456967**	0.246753247	0.054878613
OTU_00002	*Propionibacterineae*	5059	**0.047780738**	0.813186813	0.284078998
OTU_00096	*Acinetobacter*	4727	**0.023625**	0.421578422	**0.022315992**
OTU_02384	*Erythrobacter*	3890	**0.002245902**	0.086913087	**0.043998073**
OTU_00101	*Afipia*	3701	**0.012995902**	0.43956044	**0.035188825**
OTU_00314	*Rhizomicrobium*	2616	0.379436475	0.51048951	0.252589595
OTU_00137	*Acinetobacter*	2403	**0.016057377**	0.115884116	**0.009996146**
OTU_00042	*Staphylococcus*	2099	**0.020868852**	0.121878122	0.756938343
OTU_00431	*Burkholderia*	2055	**0.000127049**	0.13986014	**0.049526012**
OTU_00008	*Streptococcus*	2035	0.052063525	0.53046953	0.284695568
OTU_00224	*Bradyrhizobium*	1892	**0.003879098**	0.707292707	**0.000522158**
OTU_01890	*Bradyrhizobium*	1833	**0.00010041**	0.518481518	**0.00022736**
OTU_12890	*Escherichia/Shigella*	1670	0.181452869	0.433566434	0.068115607
OTU_01727	*Escherichia/Shigella*	1593	0.168870902	0.424575425	0.055165703
OTU_01045	*Afipia*	1473	0.217569672	0.946053946	**0.014853565**
OTU_31338	*Nitrospira*	1455	0.536790984	**0.000999001**	0.374400771
OTU_04520	*Pseudomonas*	1429	0.162887295	0.161838162	0.95761079
OTU_00711	*Ferruginibacter*	1369	0.982668033	0.355644356	0.540578035
OTU_04146	*Legionella*	1060	0.132870902	0.667332667	0.099693642
OTU_00237	*Undibacterium*	1052	0.365502049	0.998001998	0.092398844
OTU_10647	*Delftia*	1021	0.319342213	0.591408591	0.803315992
OTU_24267	*Bacillus*	1005	0.394139344	0.160839161	1
OTU_25444	*Paenibacillus*	970	0.393112705	0.160839161	1
OTU_05224	*Pseudomonas*	917	0.184190574	0.155844156	0.164208092
OTU_00005	*Staphylococcus*	818	**0.009719262**	0.231768232	0.986217726
OTU_02221	*Burkholderia*	813	0.296094262	0.822177822	0.367957611
OTU_20941	*Janthinobacterium*	813	0.114452869	0.7002997	0.092878613
OTU_14159	*Massilia*	710	**0.048922131**	0.140859141	0.615148362
OTU_00039	*Xenophilus*	709	0.533518443	0.842157842	0.52916185
OTU_02065	*Acinetobacter*	703	**0.01917418**	0.622377622	**0.015743738**
OTU_00230	*Gemella*	694	0.136678279	0.400599401	0.816662813
OTU_00036	*Bradyrhizobium*	675	0.346375	0.905094905	0.074570328
OTU_00148	*Pseudomonas*	674	0.619243852	**0.045954046**	**0.034040462**
OTU_01429	*Oligotropha*	635	**0.004297131**	0.356643357	**0.036510597**
OTU_00060	*Methylobacterium*	631	0.325415984	**0.027972028**	**0.000657033**
OTU_10693	*Yersinia*	610	0.369706967	0.197802198	0.935067437
OTU_01761	*Curvibacter*	592	0.148961066	0.513486513	0.124462428
OTU_00003	*Paracoccus*	569	**0.01375**	**0.000999001**	0.579549133
OTU_15711	*Gp6*	542	1	0.48951049	0.374400771
OTU_02237	*Bradyrhizobium*	538	**3.27869E-05**	0.316683317	**0.000181118**
OTU_11303	*Novosphingobium*	525	0.520284836	0.177822178	0.384615385

[a] p-values <0.05 are shown bold

Table S2 Relative abundance (%) of chloroplast sequences of each sampling site.

Entrance visitors	58.84
Procedure room	39.73
Isolation room	2.20
Sick chamber - Time 0	13.89
Sick chamber - Time 24 h	55.63
Respirator - Sick chamber	0.04
Respirator - Procedure room	0.00
Respirator - Isolation room	0.00
Perfusor - Sick chamber	0.00
Perfusor - Procedure room	0.00
Perfusor - Isolation room	0.00
Blood gas analysis device	0.00
Docking station	24.43
Ultrasound unit	0.04
Defibrillator	0.00
Endoscope	0.00
Workplace - Sick chamber	5.36
Workplace - Isolation room	0.06
Workplace - Bandage vehicle	15.37
Bandage vehicle - Grasp	4.59
WP - Bandage vehicle isoltion room	0.47
Keyboard - Office	1.92
Keyboard - Doctors room	0.00
Social room	0.02

Table S3 Multiplex identifier sequences (MIDs).

MID	BarcodeSequence	LinkerPrimerSequence
MID32	AGTACGCTAT	GTGCCAGCAGCCGC
MID33	ATAGAGTACT	GTGCCAGCAGCCGC
MID34	CACGCTACGT	GTGCCAGCAGCCGC
MID35	CAGTAGACGT	GTGCCAGCAGCCGC
MID36	CGACGTGACT	GTGCCAGCAGCCGC
MID45	TACACACACT	GTGCCAGCAGCCGC
MID46	TACACGTGAT	GTGCCAGCAGCCGC
MID47	TACAGATCGT	GTGCCAGCAGCCGC
MID48	TACGCTGTCT	GTGCCAGCAGCCGC
MID49	TAGTGTAGAT	GTGCCAGCAGCCGC
MID50	TCGATCACGT	GTGCCAGCAGCCGC
MID51	TCGCACTAGT	GTGCCAGCAGCCGC
MID53	TCTATACTAT	GTGCCAGCAGCCGC
MID41	TGACGTATGT	GTGCCAGCAGCCGC
MID40	TGTGAGTAGT	GTGCCAGCAGCCGC
MID52	ACAGTATATA	GTGCCAGCAGCCGC
MID37	ACGCGATCGA	GTGCCAGCAGCCGC
MID54	ACTAGCAGTA	GTGCCAGCAGCCGC
MID38	AGCTCACGTA	GTGCCAGCAGCCGC
MID42	AGTATACATA	GTGCCAGCAGCCGC
MID43	AGTCGAGAGA	GTGCCAGCAGCCGC
MID55	AGTGCTACGA	GTGCCAGCAGCCGC
MID56	CGATCGTATA	GTGCCAGCAGCCGC
MID39	CGCAGTACGA	GTGCCAGCAGCCGC

7. Book Chapter

Encyclopedia of Metagenomics (2013) DOI 10.1007/978-1-4614-6418-1_322-2

Complex Indoor Communities: Bacterial Life Under Extreme Conditions in Clean Rooms and Intensive Care Units

Lisa Oberauner[a,b]*, Alexander Mahnert[b]*, Anastasia Bragina[b] and Gabriele Berg[b]

[a]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Graz, Austria
[b]Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria

Encyclopedia of Metagenomics
DOI 10.1007/978-1-4614-6418-1_322-2
@ Springer Science+Business Media New York 2013

Complex indoor communities: bacterial life under extreme conditions in clean rooms and intensive care units

Lisa Oberauner[1,2*], Alexander Mahnert[2*], Anastasia Bragina[2], and Gabriele Berg[2]
[1]*Austrian Centre of Industrial Biotechnology (ACIB GmbH), Graz, Austria*
[2]*Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria*
*Both authors contributed equally

Correspondence: Gabriele Berg
Telephone: +43 316 873 8310, Fax: +43 316 873 8819,
E-mail: gabriele.berg@tugraz.at

Running title: Complex bacterial indoor communities
Submitted to:
Encyclopedia of Metagenomics edited by Karen E. Nelson (SPRINGER)

Synonyms
Microbiome of built environments

Definition
Indoor microbiomes are communities of microorganisms that inhabit the interior of built environments and are influenced by complex a-biotic (e.g. climate, geographic location, building architecture and maintenance) and biotic factors (human and animals/pets dynamics, greenery status, etc.).

Introduction to Indoor Microbiome
Although microbes have often been recognized as pathogens, it is now well established that the majority of host-bacterial interactions are symbiotic (Blaser 2011). This partnership is based on molecular signaling to mediate beneficial outcomes for both microbes and their hosts. This relationship between microbial diversity and host health was shown not only for plants and soils, but also for animals and humans (Keesing et al. 2010). Despite the fact that the majority of our lifetime is spent in indoor environments such as the home, work place, or public buildings (Fig. 1, Table 1), our knowledge of microbial diversity inside buildings is limited. We are not alone in these indoor environments: they provide new habitats and residence to numerous microbial communities comprising possibly hundreds of individual bacterial and fungal species. The most recent cultivation-based studies analyzed potential indoor pathogens with an emphasis on allergenic microorganisms (Yamamoto et al. 2011), yet little is known about the real microbial diversity indoors that has adapted to nutrient-poor, extreme conditions and are composed of only a small cultivable fraction of microbes. The indoor microbiome should be continuously explored with special focus on the beneficial microbial inhabitants.

Recently, the application of next-generation sequencing (NGS) techniques has provided new insights into indoor microbial communities (Fig. 2). In general, they are characterized by a high prokaryotic diversity and comprise diverse bacterial and archaeal phyla (Flores et al. 2011, 2013; Moissl-Eichinger 2011; Hewitt et al. 2012, 2013; Kembel et al. 2012, Kelley and Gilbert 2013). Indoor environments are also characterized by a specifically adapted fungal microbiome with an atypical building composition unlike those shown for bacteria (Pitkäranta et al. 2008). In addition, fungi are able to grow indoors when water is available (Zalar et al. 2011). Indoor microbiomes originate mainly from human skin, pets, or outside air, and are even known to include extremophiles. Furthermore, all of them can contain potential human pathogens in addition to beneficial bacteria that are characterized

DOI 10.1007/978-1-4614-6418-1_322-2
@ Springer Science+Business Media New York 2013

by a positive interaction with their host (Flores et al. 2011; Kembel et al. 2012). Kembel et al. (2012) were the first to analyze patient rooms and find a strong impact from both architecture and ventilation. Similarly, other factors influencing the indoor diversity are of geographic and climatic origin (Hewitt et al. 2012). Furthermore, two different types of microbial communities live indoors: airborne and surface-associated organisms. Airborne microbes — bacteria, fungi or microscopic algae — are scattered and can travel long distances such as in the wind or in clouds before returning to the ground. Surface-associated microbes, however, tend to form biofilms. Despite the studies concerning indoor microbial communities published within the last two years in which molecular microbial ecology methods were applied, the majority of microbial co-inhabitants in our built environments and their dynamics are still unknown.

The Impact of Indoor Microbiome on Human Health

Indoor microbial communities are an important component of everyday human health. They are partially composed of human-associated bacteria (Fierer et al. 2008) due to the high emission rate of up to 10^6 bacteria per person per hour as reported from genome copies measured in the air from individual persons (Qian et al. 2012). As a result, many patients in hospitals and especially in intensive care units (ICUs) develop hospital-acquired "nosocomial infections" that compound their underlying severe disease (Plowman et al. 2000). Moreover, these nosocomial infections remain among the leading causes of death in hospitals of developed countries, as the risk for nosocomial infections for patients in European ICUs, for example, was reported as 45% (Plowman et al. 2000). Hospital surfaces are often overlooked reservoirs for these bacteria (Kramer et al. 2006). Therefore, new sanitation standards are needed to drastically reduce the risk for these hospital-acquired infections. Apart from hospitals, indoor microorganisms affect human health as allergenic agents as well (Hanski et al. 2012). Indoor microorganisms are also involved in the development of the Sick building syndrome (SBS) which causes symptoms such as sensory irritation of the eyes, nose, and throat, neurotoxic or general health problems, skin irritation, nonspecific hypersensitivity reactions, and odor and taste sensations.

Bacterial Communities in Intensive Care Units

In contrast to the majority of indoor environments, rooms in hospitals and especially intensive care units (ICUs) are routinely monitored (Fig. 1, Table 1) (Hewitt et al. 2013). However, this monitoring is based on cultivation and not DNA sequencing. As the tendency is now shifting, 16S rRNA gene amplicon pyrosequencing was used to study the ICU microbiome in comparison with the currently used standard cultivation technique (Oberauner et al. 2013). Only 2.5% of the total bacterial diversity was detected using cultivation; however all sequences were represented in the sequencing libraries. The phylogenetic spectrum comprised 7 phyla and 76 genera and combined species associated with the outside environment, taxa closely related to potential human pathogens, and beneficials. Namely, *Propionibacterium, Pseudomonas*, and *Burkholderia* were identified as important sources of infection (Fig. 2). Despite significantly different bacterial area profiles for floors, medical devices, and workplaces, network analysis and molecular fingerprints were used to show similarities and evidence for the transmission of strains, respectively. This information allows for a new assessment of public health risks in ICUs and will help to create new sanitation protocols to better understand the development of hospital-acquired infections.

Encyclopedia of Metagenomics
DOI 10.1007/978-1-4614-6418-1_322-2
@ Springer Science+Business Media New York 2013

Figure 1 Illustrations of built environments. **a**) bedroom (private room); **b**) office (public room); **c**) intensive care unit (ICU); **d**) spacecraft assembly clean room.

Bacterial Communities in Clean Rooms

Clean rooms are established facilities that have been involved in various industrial processes since the 1940s (Fig. 1). Whereas first applications were situated in the areas of micro-technology and optics, today these fields are enhanced by the production of semi-conductors, applications in medical, pharmaceutical, and food engineering, as well as spacecraft assembly among many other industrial branches. Clean rooms are categorized into different clean room classes by the amount of particles present. Several standards and guidelines define limits for particle numbers of various sizes. For the DIN EN-ISO 14644-1 classification, the ISO classes 1-6 correspond to the number of particles (10-10^6) per m^3 with 0.1-0.2 µm in size. The amount of these particles is controlled via filters, airflow rate, pressure, humidity and temperature. Despite stringent cleaning and maintenance, clean rooms used for

Table 1 Studies analysing indoor environment microbiomes and parameters.

Indoor environment	Classification	Human dynamic	Maintenance	Monit oring	Materials	Associated microbiome (references)
Public buildings	moderate	high (day), moderate (night)	standard, mechanical ventilated	moder ate	polymers, textiles, wood	Flores et al. 2011 Hewitt et al. 2012 Qian et al. 2012
Private buildings	spare	moderate (day), high (night)	diverse, window ventilated	low	organic, wood, textiles, polymers	Flores et al. 2013 Dunn et al. 2013
Intensive care units (ICUs)	strict	high (day and night)	standard, mechanical ventilated, frequently cleaned, use of disinfectants, very sanitary	contro lled	polymers, metals, textiles	Hewitt et al. 2013 Oberauner et al. 2013 http://hospitalmicrobiome .com/
Clean rooms	strict	minor (day and night)	cleaning with alkaline reagents, controlling of particles, airflow, humidity, temperature, mechanical ventilated	strict	polymer, metals	La Duc et al. 2007 Moissl et al. 2007 Moissl-Eichinger et al. 2011 Vaishampayan et al. 2013

spacecraft assembly are not devoid of microorganisms and many hardy extremophiles can survive in these oligotrophic conditions as previously reported (Table 1) (La Duc et al. 2007; Moissl et al. 2007, Moissl-Eichinger 2011). Due to planetary protection regulations, a peculiar monitoring of biological contaminants (bioburden) and characterization of these microbial populations in such a well-maintained extremely low-biomass environment must be followed at each step of the assembly process. Most of the standard assays are based on cultivation dependent methods; however there has been a trend in recent studies to also include cultivation independent methods including metagenomic approaches (Vaishampayan et al. 2013).

Bacterial communities in clean rooms for spacecraft assembly at the EADS facility in Friedrichshafen (Germany) and at the NASA Jet Propulsion Laboratory (CA, USA) were investigated in a joint project. Floor samples were studied using cultivation-dependent (mesophiles/oligotrophs, alkaliphiles/alkalitolerants and facultative anaerobes) and -independent assays [ATP and propidium monoazide (PMA)] to measure microbial burden (Vaishampayan et al. 2013). When samples were

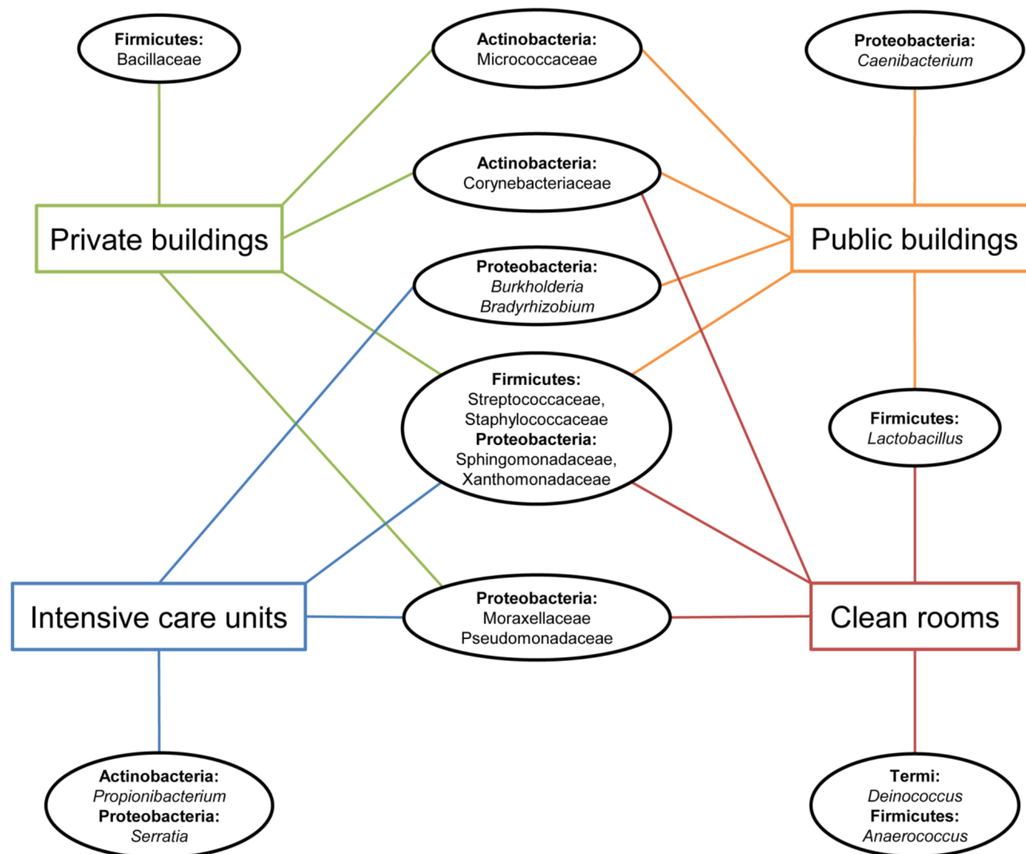


Figure 2 Overview of typical and dominant bacterial groups in the built environments. Schematic chart represents occurrence of the bacterial inhabitants indoors. Bacterial families and genera (black ellipses) are arranged according to their phylum affiliation (bold) and are connected to certain types of the built environments (colored squares). This image was compiled from information in table 1 and is not a holistic representation.

treated with PMA prior to DNA extraction, the chemical intercalated with DNA from dead microbes, thus disabling PCR amplification (Wagner et al. 2008). The PMA-treated (viable microbes) and untreated (total microbes) portions were analyzed using qPCR and 16S rRNA gene amplicon deep

DOI 10.1007/978-1-4614-6418-1_322-2
@ Springer Science+Business Media New York 2013

sequencing to estimate bioburden and measure viable microbial diversity, respectively. Overall, the clean room floors contained less total and viable microbial burden when measured by any assay than the adjacent servicing area locations. Hence, stringent maintenance and cleaning reduced the viable microbial population in the clean room by 1-2 logs .This reiterates the fact that the proper maintenance of the NASA Jet Propulsion Laboratory (JPL) spacecraft assembly clean room floors might have removed substantial number of microbial cells, but some selective microbial populations were able to survive under these clean conditions. The application of ATP-assays and PMA-qPCRs are both suitable to target the viable microbial population. However, the deep sequencing analysis in combination with a prior PMA treatment showed that viable microbial diversity also exists in the clean room and not only in the servicing area as expected. While Proteobacteria and Firmicutes were the dominant bacterial phyla (Fig. 2), Archaea and fungi were also detected as most microbes seem to be introduced by humans. In addition, a metagenomic approach targeting various genes is planned at JPL to reveal the presence of active functional microbial species. Results of this study will enable scientists to accurately track the true viable microbial population and perform accurate risk assessment of microbial contamination to the assembled products in the clean room environment.

Summary and Conclusions

Indoor microbiomes are complex microbial ecosystems influenced by diverse a-biotic and biotic factors. Indoor microbes originate from humans, pets, in and outdoor plants, dust, and soil, yet altogether every individual leaves a significant signature within his or her built environment as a result of unique microbiomes and activities. Advances driven by novel high-throughput technologies (e.g. next-generation sequencing) have completely altered our perspective on the microbiology of built environments. Therefore, these techniques should also be used for evaluation of standard maintenance in clean rooms and validation of clean room products, but also for evaluation of our hygiene standards in hospitals. Overall, the indoor microbiome plays an important role for human health and contains not only pathogens, but also a substantial proportion of beneficials which should be ultimately maintained.

Acknowledgement

Bacterial communities of the clean rooms for spacecraft assembly at the EADS facility in Friedrichshafen (Germany) were studied in collaboration with the planetary protection group of Dr. Christine Moissl-Eichinger at the University of Regensburg (Germany) and at the NASA Jet Propulsion Laboratory (CA, USA) in collaboration with the biotechnology and planetary protection group of Dr. Kasthuri Venkateswaran at site. We thank all group members for excellent cooperation. Furthermore, we would also like to thank Meg Starcher (Graz, Austria / Washington, DC, USA) for English revision of the manuscript. This work has been supported by a grant to G.B. from the Federal Ministry of Economy, Family and Youth (BMWFJ), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol and ZIT - Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG.

References

1. Blaser M. Antibiotic overuse: Stop the killing of beneficial bacteria. Nature. 2011; 476: 393–394.
2. Dunn RR, Fierer N, Henley JB, et al. Home Life: Factors Structuring the Bacterial Diversity Found within and between Homes. PLoS ONE. 2013;8(5): e64133.
3. Fierer N, Hamady M, Lauber CL, Knight R. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. PNAS. 2008; 105: 17994–17999.

4. Flores GE, Bates ST, Knights D, et al. Microbial biogeography of public restroom surfaces. PLoS ONE. 2011; 6: e28132.
5. Flores GE, Bates ST, Caporaso JG, et al. Diversity, distribution and sources of bacteria in residential kitchens. Environ Microbiol. 2013; 15: 588–596.
6. Hanski I, von Hertzen L, Fyhrquist N, et al. Environmental biodiversity, human microbiota, and allergy are interrelated. PNAS. 2012; 109: 8334–8339.
7. Hewitt KM, Gerba CP, Maxwell SL, Kelley ST. Office space bacterial abundance and diversity in three metropolitan areas. PLoS ONE. 2012; 7: e37849.
8. Hewitt KM, Mannino FL, Gonzalez A, et al. Bacterial Diversity in Two Neonatal Intensive Care Units (NICUs). PLoS ONE. 2013; 8: e54703.
9. Keesing F, Belden LK, Daszak P, et al. Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature. 2010; 468: 647–652.
10. Kelley ST, Gilbert JA. Studying the microbiology of the indoor environment. Genome Biol. 2013; 14: 202.
11. Kembel SW, Jones E, Kline J, et al. Architectural design influences the diversity and structure of the built environment microbiome. ISME J. 2012; 6: 1469–1479.
12. Kramer A, Schwebke I, Kampf G. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. BMC Infect Dis. 2006; 6: 130.
13. La Duc MT, Dekas A, Osman S, et al. Isolation and Characterization of Bacteria Capable of Tolerating the Extreme Conditions of Clean Room Environments. Appl Environ Microbiol. 2007; 73: 2600–2611.
14. Moissl C, Osman S, La Duc MT, et al. Molecular bacterial community analysis of clean rooms where spacecraft are assembled. FEMS Microbiol Ecol. 2007; 61: 509–521.
15. Moissl-Eichinger C. Archaea in artificial environments: their presence in global spacecraft clean rooms and impact on planetary protection. ISME J. 2011; 5: 209–219.
16. Oberauner L, Zachow C, Lackner S, et al. The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing. Sci Rep. 2013; 3; 1413.
17. Pitkäranta M, Meklin T, Hyvärinen A, et al. Analysis of fungal flora in indoor dust by ribosomal DNA sequence analysis, quantitative PCR, and culture. Appl Environ Microbiol. 2008; 74: 233–244.
18. Plowman R. The socioeconomic burden of hospital acquired infection. Euro Surveill. 2000; 5: 49–50.
19. Qian J, Hospodsky D, Yamamoto N, et al. Size-resolved emission rates of airborne bacteria and fungi in an occupied class room. Indoor Air. 2012; 22: 339–351.
20. Vaishampayan P, Probst AJ, La Duc MT, et al. New perspectives on viable microbial communities in low-biomass cleanroom environments. ISME J. 2013; 7: 312–324.
21. Wagner AO, Malin C, Knapp BA, Illmer P. Removal of free extracellular DNA from environmental samples by ethidium monoazide and propidium monoazide. Appl Environ Microbiol. 2008; 74: 2537–2539.
22. Yamamoto N, Shendell DG, Peccia J. Assessing allergenic fungi in house dust by floor wipe sampling and quantitative PCR. Indoor Air. 2011;21,521–530.
23. Zalar P, Novak M, De Hoog GS, et al. Dishwashers—a man-made ecological niche accommodating human opportunistic fungal pathogens. Fungal Biol. 2011;115: 997–1007.

8. Manuscript I

Molecular Ecology (2014), resubmitted

The *Sphagnum* microbiome supports greatly bog ecosystem functioning under extreme conditions

Anastasia Bragina[1]*†, Lisa Oberauner-Wappis[1,2]†, Christin Zachow[1], Bettina Halwachs[2,3], Gerhard G. Thallinger[2,3], Henry Müller[1] and Gabriele Berg[1]

[1]Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria
[2]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Graz, Austria
[3]Institute for Genomics and Bioinformatics, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria

The *Sphagnum* microbiome supports greatly bog ecosystem functioning under extreme conditions

Anastasia Bragina[1*†], Lisa Oberauner-Wappis[1,2†], Christin Zachow[1], Bettina Halwachs[2,3], Gerhard G. Thallinger[2,3], Henry Müller[1] and Gabriele Berg[1]

[1]*Institute of Environmental Biotechnology, Graz University of Technology, Petersgasse 12, 8010 Graz, Austria*

[2]*Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, 8010 Graz, Austria*

[3]*Institute for Genomics and Bioinformatics, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria*

[†]Equal contributors

Keywords: *Sphagnum* moss, plant microbiome, bog ecosystem, Illumina-based metagenomics, FISH-CLSM

[*]Correspondence: Anastasia Bragina

Address: 6530 Institute of Environmental Biotechnology, Graz University of Technology, 8010 Petersgasse 12/I, Graz, Austria

Fax: +43 316 873 8819

E-mail: anastasia.krikovtseva@tugraz.at

Running title: Diversity in *Sphagnum* bogs

Abstract

Sphagnum-dominated bogs represent an extremely unique yet widely distributed type of terrestrial ecosystem, and strongly contribute to global biosphere functioning. *Sphagnum* is colonised by highly diverse microbial communities, but less is known about their function. We identified a high functional diversity within the *Sphagnum* microbiome applying an Illumina-based metagenomic approach followed by *de novo* assembly and MG-RAST annotation. An inter-environmental comparison revealed that the *Sphagnum* microbiome harbours highly specific genetic features that distinguish it significantly from microbiomes of higher plants and peat soils. The differential traits especially support ecosystem functioning by a symbiotic life style under poikilohydric and ombrotrophic conditions. To realise a plasticity-stability balance, we found highly abundant subsystems responsible to cope with oxidative and drought stresses, to exchange (mobile) genetic elements, and genes that encode for resistance to detrimental environmental factors, repair, and self-controlling mechanisms. Multiple interactions with each other and plants were also found to play a crucial role as indicated by diverse genes necessary for biofilm formation, interaction via quorum sensing and nutrient exchange. A high proportion of genes involved in nitrogen cycle and recycling of organic material supported the role of bacteria for nutrient supply. 16S rDNA analysis indicated a higher structural diversity than PCR-dependent techniques. Altogether, the diverse *Sphagnum* microbiome has the ability to support the life of the host plant and the entire ecosystem under changing environmental conditions. Beyond this, the moss microbiome presents a promising bio-resource for environmental biotechnology – with respect to novel enzymes or stress-protecting bacteria.

46 ## Introduction

47 Bog ecosystems belong to the oldest vegetation forms on earth; they have a high value for biodiversity
48 conservation, are a reservoir for fresh water, and play an extraordinary role in carbon sequestration to
49 benefit both human welfare and our world climate (Succow & Joosten 2001; Raghoebarsing *et al*.
50 2005; Dise 2009). However, these long-existing ecosystems are extremely sensitive to changing
51 abiotic factors connected with climate change (Strack 2008; Dise 2009). For example, degraded
52 peatlands release their stored carbon in the form of greenhouse gases, and drainage of peat soils results
53 in CO_2 and N_2O global emissions of 2–3 Gt CO_2-eq per year (Joosten & Couwenberg 2009). Mosses
54 of the genus *Sphagnum* are among the most abundant and cosmopolitan in bog vegetation in the
55 Northern hemisphere and greatly contribute to both global carbon turnover and global climate (Jassey
56 *et al*. 2011). The ecological significance of bogs is directly related to the physical, morphological, and
57 chemical characteristics of *Sphagnum* peat mosses, which belong to the poikilohydric plants that
58 undergo repetitive desiccation and oxidative stress (Daniels & Eddy 1985). Moreover, *Sphagnum*
59 mosses are able to change their environments: living Sphagna have extraordinarily high cation
60 exchange capacity and therefore acidify their environment by exchanging tissue-bound protons for
61 basic cations in surrounding water (Soudzilovskaia *et al*. 2010).

62 Recently, the plant microbiome was identified as one of the key determinants of plant health
63 and productivity (rev. in Berg *et al*. 2013b; Bulgarelli *et al.* 2013; Philippot *et al*. 2013). *Sphagnum*
64 mosses are interesting models to study plant, especially phyllosphere microbiomes, and an enormous
65 associated bacterial diversity was already detected (Raghoebarsing *et al*. 2005; Opelt *et al*. 2007a;
66 Larmola *et al*. 2010; Kip *et al*. 2011; Putkinen *et al*. 2012). Since this phylogenetically old plants have
67 no roots, the leaf-associated bacteria fulfil important functions such as nutrient supply and pathogen
68 defence for moss growth and health (Opelt *et al*. 2007b). Host specificity of moss-associated
69 microbiomes was detected independent of geographic region at both structural and functional levels
70 (Bragina *et al*. 2011, 2013). Additionally, the degree of host specificity varied between distant and
71 closely related moss species and corresponded to spectra of secondary metabolites produced by plants
72 (Bragina *et al*. 2012). Moreover, environmental factors such as acidity and nutrient richness were
73 defined as the main ecological drivers for microbial diversity, and that plant specificity of functional
74 bacterial groups is determined by their role within the ecosystem (Bragina *et al*. 2013). A core
75 microbiome not only contained mostly potential beneficials, but was also shared between the moss
76 generations and transferred within the spore capsules that emphasize the importance of the
77 microbiome for mosses as the oldest phylogenetic land plants on earth (Bragina *et al*. 2012). Although
78 this extraordinarily high diversity of the *Sphagnum* microbiome is now well-studied, less is known
79 about its functional diversity. Omics-technologies significantly contribute to a functional

80 understanding of microbial ecosystems (Gilbert *et al*. 2011) but very little is known for plants (Knief
81 *et al*. 2012; Sessitsch *et al*. 2012).

82 The objective of this study was to unravel the functional diversity associated with *Sphagnum*
83 mosses. We applied an Illumina-based metagenomic approach based on four independent replicates,
84 and through *de novo* assembly and MG-RAST annotation we revealed specific biochemical pathways
85 and adaptive strategies within the moss metagenome (Meyer *et al*. 2008). We analysed the *Sphagnum*
86 microbiome with a special focus on plasticity, stability, and interactions, and performed a comparison
87 with other published metagenomes of plants, peat soils, as well as aquatic systems to discover unique
88 features and potential differences.

89

90 Methods

91 *Sampling procedure*

92 For this metagenomic study, we selected peat moss *Sphagnum magellanicum* BRID. (section
93 *Sphagnum*), a typical and widespread vegetation component of the acidic peat bogs (Daniels & Eddy
94 1985) illustrated in Fig. S1 (Supporting information). Gametophyte samples of *S. magellanicum* were
95 collected from the Alpine bog Pirker Waldhochmoor (N46°37'38.66'' E14°26'5.66'') in Austria in
96 December 2011. Four independent replicates consisting of the living moss plants were collected from
97 the sampling points situated at a distance of at least 150 m. The collected samples were placed into
98 sterile plastic bags, cooled (4 to 8°C), and transported to the laboratory.

99

100 *Total community DNA isolation*

101 To isolate the total community DNA of the *S. magellanicum* microbiome, 200 g of each sample were
102 transferred into Stomacher bags (20 g/bag) and supplied with 0.85% NaCl solution (50 ml/bag). The
103 diluted samples were shaken in a Stomacher laboratory blender (BagMixer, Interscience, St. Nom,
104 France) for 3 min. To remove the plant debris, the suspension was subsequently strained through two
105 sieves (500 µm and 63 µm) and the resulting liquid was centrifuged in 50 ml tubes at low speed (5,000
106 g, 5 min, 4°C). The supernatant was discarded and the pellets were resuspended in 1.5 ml 0.85% NaCl.
107 After centrifugation at high speed (10,000 g, 20 min, 4°C), the obtained pellets were stored at -70°C.
108 The total community DNA was extracted using the FastDNA Spin Kit for Soil (BIO 101, Carlsbad,
109 USA) according to the manufacturer's protocol. Finally, DNA aliquots from all samples were pooled
110 together and sent to Eurofins MWG Operon (http://www.eurofinsgenomics.eu/) for Illumina
111 sequencing.

112

113

114

115 *Sequencing and bioinformatic analysis*

116 The sequencing was performed with an Illumina HiSeq 2000 system (2 x 100 bp). Prior to sequencing,
117 the total community DNA was split into two aliquots. The first aliquot was sequenced untreated and
118 the second aliquot underwent a normalisation treatment that allowed removal of the most dominant
119 sequence patterns for deeper ecological analysis. The normalisation was achieved through one cycle of
120 denaturation and re-association of the DNA, followed by separation of the re-associated ds-DNAs
121 from the remaining ss-DNAs (normalised DNA) by passing the mixture over a hydroxylapatite
122 column. After hydroxylapatite chromatography, the ss-DNAs were sequenced according to the
123 Eurofins MWG Operon protocol. The generated paired-end reads of the normalised metagenome were
124 *de novo* assembled using the CLC Genomic Workbench version 4.5.1 (CLC Bio, Aarhus, Denmark)
125 and default settings. The complete metagenome, which resulted from untreated DNA sequencing, was
126 used for abundance-based analyses, while normalised metagenome was used to study ecologically-
127 relevant functional patterns.

128 The functional composition of the *S. magellanicum* microbiome was analysed using the
129 Metagenomic RAST (MG-RAST) server (Meyer *et al*. 2008). For this purpose, both the complete and
130 the normalised metagenomes were uploaded to the server and annotated using hierarchical
131 classification with default parameters: SEED subsystems as an annotation source, a maximum e-value
132 of e^{-5}, a minimum identity of 60 %, and a minimum alignment length of 15 measured in aa for protein
133 and bp for RNA databases. For the normalised metagenome, sequences from the single subsystems
134 were aligned against a non-redundant protein sequences (nr) database using BLASTx algorithm to
135 check their affiliation. Distribution of the functional subsystems within the normalised metagenome
136 was visualised using Krona plot (Ondov *et al*. 2011). Enzymes involved in nitrogen metabolism from
137 the complete and normalised metagenomes were visualised using KEEG mapper tool of the MG-
138 RAST server with default parameters.

139 The inter-environmental comparison of the complete *S. magellanicum* metagenome with
140 publicly available metagenomes was performed using the principal coordinate analysis (PCoA) tool of
141 the MG-RAST server. Relevant publicly available metagenomes obtained from peat soils, freshwater
142 habitats, plant tissues, and human bodies are summarised in Table S1 (Supporting information). PCoA
143 analysis was performed for the metagenomic datasets that were annotated using hierarchical
144 classification with default parameters. For each dataset, sequence counts were normalised and scaled
145 according to the algorithm, which is specified at the MG-RAST server
146 (http://blog.metagenomics.anl.gov/howto/mg-rast-analysis-tools/). The distance matrix for PCoA
147 analysis was calculated using Bray-Curtis as a distance metric (Bray & Curtis 1957). The inter-
148 environmental comparison of the metagenomes was expanded by constructing a heatmap of the
149 complete *S. magellanicum*, higher plant, and peat soil metagenomes and their functional subsystems

150 using the MG-RAST heatmap tool. The selected metagenomes (Table S2, Supporting information)
151 were grouped using complete linkage clustering with Bray-Curtis distance. For these metagenomes,
152 Kolmogorov-Smirnov test (Massey 1951) was applied on the raw abundances to test probability
153 distributions of each subsystem (Table S2, Supporting information). Scale normalisation factors were
154 calculated to scale the raw library sizes prior to significance analysis. To make the count data ready for
155 linear modelling, raw counts were transformed using the voom function (Law *et al*. 2014). The
156 probability distribution of each group was visualised before and after data transformation using
157 density plots (Fig. S2, Supporting information). Changes of the subsystems included in the heatmap
158 between the different groups were assessed by statistical analysis using the linear modelling approach
159 implemented by the R Bioconductor package limma (version 3.16.8) (Smyth 2004). Significance
160 analysis within limma was performed by the moderated t-statistic, which was computed for each probe
161 and each contrast. To account for multiple comparisons, p-values were adjusted by the method
162 described by Benjamini and Hochberg (1995). Adjusted p-values of less than 0.05 were considered as
163 statistically significant.

164 The taxonomic structure of the *Sphagnum*-associated bacterial community was determined on
165 the basis of 16S rRNA genes derived from total metagenomic quality reads of the complete
166 metagenome. Prior to taxonomic assignment, reads that comprised exclusively partial 16S rRNA
167 genes were extracted after alignment to references of the whole 16S rRNA gene by a homology-based
168 approach using BLASTn algorithm. Only reads which consisted of 16S rRNA gene sequences
169 covering a length between 80 and 100 bp were retained and processed using QIIME pipeline with
170 default parameters (release 1.7.0, Caporaso *et al*. 2010). In detail, sequence clustering was performed
171 at 97% similarity using UCLUST algorithm and a pre-designed taxonomy map (Edgar 2010)
172 implemented in the QIIME workflow *pick_open_reference_otus.py*. Taxonomic assignment of
173 representative sequences was done using RDP naïve Bayesian rRNA classifier (Wang *et al*. 2007)
174 based on the reference database Greengenes release 13_5 (DeSantis *et al*. 2006). In addition,
175 taxonomic hits distribution was deduced from the complete metagenome for both the sequences with
176 predicted protein coding regions and ribosomal rRNA genes using all reference databases available at
177 the MG-RAST server.
178

179 ### *Fluorescent* in situ *hybridisation and confocal laser scanning microscopy*

180 Single gametophytes of *S. magellanicum* were fixed with 4% paraformaldehyde/phosphate buffered
181 salt (3:1, v/v) and stained by in-tube FISH (Grube *et al*. 2009). The samples were consequently
182 hybridised with rRNA-targeting probes (genXpress, Wiener Neudorf, Austria) specific for
183 Alphaproteobacteria (ALF968) (Loy *et al*. 2007) and with a set of universal bacterial probes
184 (EUB338/EUB338II/EUB338III) (Amann *et al*. 1990; Daims *et al*. 1999). Hybridisation was carried

185 out at 41°C using hybridisation buffer with 35% and 15% formamide, respectively. Negative control
186 was hybridised with non-target NON-EUB probe (Amann *et al*. 1990) at the same stringency
187 conditions applied for the positive FISH probes. Confocal laser scanning microscopy (CLSM) was
188 performed with a Leica TCS SPE confocal microscope (Leica Microsystems, Mannheim, Germany) as
189 previously described (Bragina *et al*. 2012) followed by volume rendering of confocal stacks and three-
190 dimensional modelling using the software Imaris 7.3 (Bitplane, Zurich, Switzerland).
191

192 Results

193 *The* Sphagnum *metagenomic dataset*

194 Illumina HiSeq 2x100 paired-end sequencing resulted in 172,590,841 reads (41.8 Gbps in total) and
195 141,411,216 reads (32.0 Gbps) from the untreated and the normalised metagenomic DNA of
196 *Sphagnum* moss, respectively (Table S3, Supporting information). *De novo* assembly of the
197 normalised metagenome yielded 1,115,029 scaffolded contigs totalling 558,360,453 bps with an
198 average length of 501 bps. For both metagenomes, MG-RAST statistical analysis revealed that all
199 sequences passed the quality control (QC) pipeline. Out of the complete dataset, 153,819,621 (89%)
200 sequences produced a total of 151,683,238 predicted protein coding regions, while the normalised and
201 assembled dataset contained 1,075,645 (97%) sequences that encoded 1,411,717 predicted protein
202 coding regions, correspondingly. Based on their best e-value scores (Fig. S3, Supporting information),
203 SEED subsystems were selected as an annotation source for functional analysis of the moss
204 metagenome (Overbeek *et al*. 2005). The subsystems approach allowed us to precisely assign
205 metagenomic sequences to the groups with known or hypothetical biological functions with the
206 exception of clustering-based and miscellaneous categories.

207 Within the complete metagenome, the most dominant subsystems represented carbohydrates
208 and protein metabolism (amino acids and protein metabolism) as the most important biochemical
209 processes for all forms of life (Fig. 1). Sub-dominant subsystems contained metagenomic sequences
210 that encode pathways for biological monomers (nucleoside and nucleotides), more complex
211 biochemical compounds (cofactors, vitamins, prosthetic groups, pigments; aromatic compounds; fatty
212 acids, lipids and isoprenoids), and structural elements such as the cell wall and capsule. Subsystems
213 corresponding to environmental information processing such as membrane transport, stress responses,
214 virulence, disease, and defence followed. Among the sub-dominant subsystems, several subsystems
215 were crucial for processing genetic information in- and outside the cells (DNA and RNA metabolism;
216 phages, prophages, transposable elements, plasmids). Subsystems responsible for single chemical
217 element cycling (N, S, P, K, Fe) comprised a minor portion of all subsystems with the highest relative
218 abundance for sulfur metabolism. Genetic features that characterise cellular processes were irregularly
219 distributed within the annotated metagenome and found in the sub-dominant subsystems of cell

regulation and signalling, cell division and cycle, in the minor subsystems of motility and chemotaxis, and dormancy and sporulation.

Taxonomic diversity and spatial structure of the S. magellanicum *microbiome*

A total of 7,318 reads containing partial 16S rRNA genes were obtained from metagenomic sequences to characterise the structure of bacterial communities (Fig. 2). At phylum level, the majority of reads were assigned to Proteobacteria (65.8%) followed by Acidobacteria (11.4%), Actinobacteria (5.6%), Bacteroidetes (4.2%) and Verrucomicrobia (2.0%). The remaining portion of the classified reads was distributed among 13 bacterial phyla which notably contained Planctomycetes. At class level, Alphaproteobacteria and Betaproteobacteria were the most abundant taxa among the phylum Proteobacteria, while Gammaproteobacteria represented a sub-dominant taxon. The classes Acidobacteria, Actinobacteria, and Sphingobacteria dominated the phyla Acidobacteria, Actinobacteria, and Bacteroidetes, respectively.

The taxonomic hits distribution of metagenomic sequences with predicted protein coding regions and ribosomal rRNA genes (Fig. S4, Supporting information) revealed highly similar dominant patterns to the 16S rRNA genes data. Within the reads assigned to domain Bacteria (61,528,765 sequences), dominant portion was composed of Proteobacteria (61.9%), Acidobacteria (13.1%), Actinobacteria (8.3%), Bacteroidetes (4.2%), and Verrucomicrobia (3.0%). The minor fraction of functional bacterial reads was distributed among 16 phyla that were not covered by partial 16S rRNA genes. Consequently, FISH and CLSM techniques were used to visualise the most abundant bacterial patterns in *S. magellanicum* gametophytes. In general, *Sphagnum* mosses are characterised by unique morphology that distinguishes them from other bryophytes (Daniels & Eddy, 1985). Especially, *Sphagnum* leaves are composed of a single-layer cell net of photosynthetic chlorocytes and dead hyalocytes, which contain large pores. By applying FISH-CLSM approach, we demonstrated that hyalocytes of moss leaves serve as a main colonisation compartment for bacteria (Fig. S5, Supporting information). One of the most abundant bacterial taxa – Alphaproteobacteria – represented up to 31.9% of the detected bacterial cells that coincided with its relative abundance in metagenomic datasets (30.2%).

Unique plant-microbe biocoenosis assessed using comparative metagenomics

To study the specificity of the *Sphagnum* microbiome, the complete *S. magellanicum* metagenome was compared with publicly available metagenomes accessible through MG-RAST. We selected metagenomes obtained from peat soils, freshwater habitats, and plant tissues most relevant to the moss metagenome and metagenomes obtained from human bodies as outgroups to all tested environmental metagenomes (Table S1, Supporting information). PCoA analysis showed that the *Sphagnum*

255 metagenome has a distinct position outside all examined groups (Fig. 3). On the PCoA biplot, the
256 closest group of metagenomes originated from higher plants, such as rice, clover, soybean, and thale
257 cress (*Arabidopsis thaliana*). The heatmap and statistical analyses revealed a high specificity for the
258 functional traits that underlie the *Sphagnum*-microbe biocoenosis (Fig. S6, Table S4, Supporting
259 information). Statistical analysis resulted in a significant difference ($p < 0.05$) for 106 functional
260 groups that were differentially abundant between the *S. magellanicum* and higher plant metagenomes,
261 of which 51 groups were significantly enriched in the moss metagenome. Comparison of moss and
262 peat soils metagenomes revealed 37 differentially abundant functional groups, of which 20 groups
263 were significantly enriched in the moss metagenome. In conclusion, we demonstrated that the
264 *Sphagnum* microbiome harbours highly specific genetic features that distinguish it from microbial
265 communities of higher plants and peat soils.
266
267 *Functional versatility of the moss metagenome*
268 Functional subsystems were further studied in terms of plasticity, stability, and interaction as main
269 maintenance strategies of the *Sphagnum*-microbe biocoenosis (Table 1). For this purpose, we analysed
270 the normalised and assembled metagenomic dataset that comprised 657,466 sequences assigned to
271 certain functional subsystems of SEED database (Fig. S7, Supporting information). Regarding
272 plasticity traits, we detected highly abundant subsystems responsible for genetic exchange: i)
273 temperate bacteriophages (prophages) and their genetic transfer element (GTA) analogues; ii)
274 plasmids likely involved in natural competence; and iii) type IV pili and conjugative transport
275 systems. Genetic attributes of microbiome stability were found in subsystems that encode for
276 resistance to environmental factors, repair, and self-controlling mechanisms. For instance, we
277 identified a set of pathways that contribute to the oxidative stress response and DNA repair. These
278 subsystems encode enzymatic responses of the cells and damage illumination caused by the oxidative
279 stress. Notably, the highest diversity was observed among subsystems essential for bacterial
280 interaction within the microbiome. In particular, quorum sensing was represented by autoinducer-2
281 (AI-2), acyl homoserine lactones (AHLs), and gamma-butyrolactones signalling pathways. Other
282 mechanisms coupled with biofilm formation were the production of surface adhesins and extracellular
283 polysaccharides, motility, and chemotaxis. In order to restrict our search of bacterial interactions based
284 on the food web, we focused on the nitrogen cycle as one of the bottlenecks in bog ecosystems.
285 Interestingly, subsystems of nitrogen acquisition and conservation strongly prevailed over subsystems
286 of nitrogen release and efflux from the ecosystem (Fig. S8, Supporting information). Moreover,
287 bacterial protein degradation, which corresponds to organic nitrogen recycling (mineralisation), was
288 shown among highly abundant subsystems.
289

290 ## Discussion

291 We revealed a highly versatile genetic potential by analysing our metagenomic dataset for the
292 *Sphagnum* moss microbiome: a model for the first and phylogenetically oldest land plants. Moreover,
293 we found unique functional traits in terms of plasticity, stability, and microbe-microbe and plant-
294 microbe interactions.

295 By employing an inter-environmental comparison, we demonstrated that the moss microbiome
296 is distinct from microbial communities of higher plants and peat soils by its genetic context. This
297 difference indicates the specific interactions established between *Sphagnum* mosses and their
298 microbiome. Previous research proposed that the *Sphagnum* microbiome intimately cooperated with
299 the host plants via nutrient supply and defence against pathogens (Raghoebarsing et al. 2005; Opelt *et*
300 *al*. 2007b; Bragina *et al*. 2013), but Illumina sequencing of the moss metagenome obtained a much
301 higher functional diversity than previously reported. To elucidate this profound diversity, we
302 developed a framework in the form of plasticity-stability-interaction that integrates genetic signatures
303 of symbiosis (Gilbert *et al*. 2012) within the plant-microbe biocoenosis (Fig. 4). Specifically, the moss
304 metagenome contained a relatively high number of mobile elements which were also found in the
305 metagenomes of symbiotic bacterial consortia and considered to play an important role in the
306 evolution of bacterial genomes for symbiosis with their hosts (Ochman & Moran 2001; Thomas *et al*.
307 2010). Furthermore, *Sphagnum* mosses belong to the poikilohydric plants that undergo repetitive
308 desiccation and oxidative stress (Daniels & Eddy 1985; Scheibe & Beck 2011). Due to the high
309 diversity and abundance of genes responsible for the oxidative stress response in the studied
310 metagenome, we proposed that the bacterial capacity to tolerate oxidative stress may determine the
311 effective and stable colonisation of the *Sphagnum* mosses. In regards to interaction traits, vegetation in
312 peatland ecosystems is strongly limited by nitrogen availability and therefore requires prokaryotic
313 associates for nitrogen supply (Rydin & Jeglum 2006). Since Granhall and Hofsten (1976) observed
314 nitrogen-fixing symbiotic Cyanobacteria in *Sphagnum* for the first time, diazotrophic communities of
315 Sphagna have been characterised by a high taxonomic diversity and shown to transfer fixed nitrogen
316 to the host plants (Bragina *et al*. 2011, 2013; Berg *et al*. 2013a). In the current study, we observed and
317 determined the entire nitrogen turnover of the moss microbiome. The pathways for bacterial nitrogen
318 acquisition and conservation strongly prevailed over those for nitrogen release and efflux processes
319 within the metagenome. Overall, we provided evidence that the *Sphagnum* microbiome carries
320 essential genetic potential for sustainable functioning in association with the host plants and within the
321 peatland ecosystem.

322 This metagenome study provided also new insights into the taxonomic diversity of the
323 *Sphagnum*-associated microbiome. Our approach allowed for a deep analysis of the 16S rRNA gene
324 diversity without PCR-based bias. Although the dominant bacterial taxa detected using Illumina

325 sequencing were similar to those revealed by PCR-dependent approaches (Bragina *et al*. 2012), their
326 relative abundance considerably differed. As such, we observed a low number of Planctomycetes 16S
327 rRNA genes that contrasts with their relatively high abundance in the Northern peat bogs and Arctic
328 peat soils (Serkebaeva *et al*. 2013; Tveit *et al*. 2013). Despite these differences, we were able to prove
329 the dominance of Alphaproteobacteria in *Sphagnum* microbiome by FISH-CLSM analysis.
330 Furthermore, the microbiome composition was complemented with several bacterial classes and
331 candidate divisions through sequencing of the metagenome that were not observed in previous studies
332 (Bragina *et al*. 2012; Serkebaeva *et al*. 2013). Additionally, we found evidence for the dominant
333 functional groups (subsystems) of Proteobacteria, which were reported as the most abundant nitrogen-
334 fixing bacteria in *Sphagnum* mosses (Bragina *et al*. 2011, 2013). Moreover, genes that encode for
335 autoinducers produced by Proteobacteria for quorum sensing (Miller & Bassler 2001) were shown
336 among the dominant functional groups of the *S. magellanicum* metagenome.
337 For the interpretation of metagenomic data several limitations have to be considered
338 (Committee on Metagenomics 2007; Thomas *et al*. 2012). For example, high-throughput sequencing
339 of the metagenome provides only a partial DNA sampling, which however might have to be used to
340 predict general features rather than analyse the total functional diversity of the sample (Prakash &
341 Taylor 2012). Furthermore, automatic *in silico* annotation is characterised by a relatively high error
342 rate and disregards proteins of unknown function as well (Teeling & Glöckner 2012). However,
343 through the combination of the newly discovered genetic features and knowledge of ecological
344 ontology of the samples, we can cautiously interpret the metagenomic data in terms of microbiome
345 biodiversity and functioning. For the inter-environmental comparison, we used publically available
346 and *S. magellanicum* metagenomes that were generated using Roche 454 and Illumina technologies,
347 respectively. Although these technologies vary in sequencing depth and reads length, they provide
348 comparable view of the sampled communities (Luo *et al*. 2012). Despite this fact, this technical source
349 of error cannot be completely excluded. Moreover, sampling strategies, DNA isolation procedure and
350 libraries preparation can be potential confounding factors of the analysis.

351 Besides the importance of the *Sphagnum* microbiome for ecosystem function in association
352 with the host plants – seen as meta-organisms - and within the peatland ecosystem, this microbiome
353 presents a promising bio-resource for environmental biotechnology. For example, drought resistance is
354 one of the major challenges for sustainable agriculture influenced by climate change (Berg *et al*.
355 2013b). Stress-protecting bacteria that have co-evolved in association with the poikilohydric
356 *Sphagnum* moss can contribute to solve these problems as already shown by Zachow *et al*. (2013).
357
358

359 **Acknowledgements**

360 We thank Christian Berg (Graz) and Wilfried R. Franz (Klagenfurt) for their bryological expertise,
361 sampling and helpful discussions. We would also like to acknowledge Meg Starcher
362 (Graz/Washington) for English revision of the manuscript. We are grateful to Massimiliano Cardinale
363 (Graz) for assistance with computer reconstruction of FISH-CLSM data. This work has been
364 supported by a grant to G.B. from the Federal Ministry of Economy, Family and Youth (BMWFJ), the
365 Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion
366 Agency SFG, the Standortagentur Tirol and ZIT – Technology Agency of the City of Vienna through
367 the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG.

368

369 **References**

370 Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF (2004) Divergence and redundancy of the 16S rRNA in genomes with multiple *rrn*
371 operons. *Journal of Bacteriology*, **186**, 2629–2635.

372 Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA (1990) Combination of 16S rRNA-targeted oligonucleotide probes
373 with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology*, **56,** 1919–1925.

374 Barak JD, Jahn CE, Gibson DL, Charkowski AO (2007) The role of cellulose and O-antigen capsule in the colonization of plants by
375 *Salmonella enterica. Molecular Plant-Microbe Interactions*, **20**, 1083–1091.

376 Benjamini Y, Hichberg Y (1995) Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the*
377 *Royal Statistical Society: Series B (Statistical Methodology)*, **57**, 289–300.

378 Berg A, Danielsson Å, Svensson B (2013a) Transfer of fixed-N from N-fixing cyanobacteria associated with the moss *Sphagnum riparum*
379 results in enhanced growth of the moss. *Plant and Soil*, **362**, 271–278.

380 Berg G, Zachow C, Müller H, Philipps J, Tilcher R (2013b) Next-generation bio-products sowing the seeds of success for sustainable
381 agriculture. *Agronomy*, **3**, 648–656.

382 Bragina A, Berg C, Cardinale M, Shcherbakov A, Chebotar V, Berg G (2012) *Sphagnum* mosses harbor highly specific bacterial diversity
383 during their whole lifecycle. *The ISME Journal,* **6**, 802–813.

384 Bragina A, Berg C, Müller H, Moser D, Berg G (2013) Insights into functional bacterial diversity and its effect on Alpine bog ecosystem
385 functioning. *Scientific Reports*, **3**, 1955.

386 Bragina A, Maier S, Berg C, Müller H, Chobot V, Hadacek F *et al*. (2011) Similar diversity of Alphaproteobacteria and nitrogenase gene
387 amplicons on two related *Sphagnum* mosses. *Frontiers in Microbiology*, **2**, 275.

388 Bray JR, Curtis JT (1957) An ordination of upland forest communities of southern Wisconsin. *Ecological Monographs*, **27**, 325–349.

389 Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P (2013) Structure and functions of the bacterial microbiota
390 of plants. *Annual Review Plant Biology*, **64,** 807–38.

391 Caporaso JG, Kuczynski J, Strombauh J, Bittinger K, Bushman FD, Costello EK *et al*. (2010) QIIME allow analysis of high-throughput
392 community sequencing data. *Nature Methods*, **7**, 335–336.

393 Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brüssow H (2003) Phage as lateral gene transfer. *Current Opinion in*
394 *Microbiology*, **6**, 414–424.

395 Committee on Metagenomics, National Research Council (2007) *Data management and bioinformatics challenges of metagenomics. The*
396 *new science of metagenomics. Revealing the secrets of our microbial planet*, pp. 85–97. The National Academies Press,
397 Washington, D.C.

398 D'aes J, De Maeyer K, Pauwelyn E, Höfte M (2010) Biosurfactants in plant-*Pseudomonas* interactions and their importance to biocontrol.
399 *Environmental Microbiology Reports*, **2**, 359–372.

400 Daims H, Brühl A, Amann R, Schleifer KH, Wagner M (1999) The domain-specific probe EUB338 is insufficient for the detection of all
401 bacteria: development and evaluation of a more comprehensive probe set. *Systematic and Applied Microbiology*, **22,** 434–444.

402 Daniels RE, Eddy A (1985) *Handbook of European Sphagna*. Institute of Terrestrial Ecology, Aberystwyth, UK.

403 DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al*. (2006) Greengenes, a chimera checked 16S rRNA gene database
404 and workbench compatible with ARB. *Applied and Environmental Microbiology*, **72**, 5069–5072.
405 Dise NB (2009) Environmental science. Peatland response to global change. *Science*, **326**, 810–811.
406 Dorhan T, Fuqua C (2007) Biofilm formation by plant-associated bacteria. *Annual Review of Microbiology*, **61**, 401–422.
407 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.
408 Gilbert JA, O'Dor R, King N, Vogel TM (2011) The importance of metagenomic surveys to microbial ecology: or why Darwin would have
409 been a metagenomic scientist. *Microbial Informatics and Experimentation*, **1**, 5.
410 Gilbert SF, Sapp J, Tauber AI (2012) A symbiotic view of life: we have never been individuals. *The Quarterly Review of Biology*, **87**,
411 325–341.
412 Granhall U, Hofsten AV (1976) Nitrogenase activity in relation to intracellular organisms in *Sphagnum* mosses. *Physiologia Plantarum*, **36**,
413 88–94.
414 Grauman PL, Marahiel MA (1998) A superfamily of proteins that contain a cold-shock domain. *Trends in Biochemical Sciences*, **23**,
415 286–290.
416 Grube M, Cardinale M, de Castro JV Jr, Müller H, Berg G (2009) Species-specific functional and structural diversity of bacterial
417 communities in lichen symbioses. *The ISME Journal*, **3**, 1105–1115.
418 Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167–170.
419 Jassey VEJ, Gilbert D, Binnet P, Toussaint M, Chiapusioa G (2011) Effect of temperature gradient on *Sphagnum fallax* and its associated
420 microbial communities: a study under controlled conditions. *Canadian Journal of Microbiology*, **57**, 226–235.
421 Joosten H, Couwenberg J (2009) *Are emissions reductions from peatlands MRV- able?* Wetlands International, Ede, Netherlands.
422 Kato JY, Funa N, Watanabe H, Ohnishi Y, Horinouchi S (2006) Biosynthesis of gamma-butyrolactone autoregulators that switch on
423 secondary metabolism and morphological development in *Streptomyces*. *Proceedings of the National Academy of Sciences*, **104**,
424 2378–2383.
425 Kip N, Ouyang W, van Winden J, Raghoebarsing A, van Niftrik L, Pol A *et al*. (2011) Detection, isolation and characterization of
426 acidophilic methanotrophs from *Sphagnum* mosses. *Applied and Environmental Microbiology*, **77**, 5643–5654.
427 Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, Wassmann R *et al*. (2012) Metaproteogenomic analysis of microbial communities
428 in the phyllosphere and rhizosphere of rice. *The ISME Journal*, **6,** 1378–90.
429 Lang GS, Zhaxybayeva O, Beatty JT (2012) Gene transfer agents: phage-like elements of genetic exchange. *Nature Reviews Microbiology*,
430 **10**, 472–482.
431 Larmola T, Tiottila ES, Tiirola M, Nykänen H, Martikainen PJ, Yrjälä K *et al*. (2010) The role of *Sphagnum* mosses in the methane cycling
432 of a boreal mire. *Ecology*, **91**, 2356–2365.
433 Law CW, Chen Y, Shi W, Smyth GK (2014) Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome*
434 *Biology*, **15**, R29.

435 Loy A, Maixner F, Wagner M, Horn M (2007) probeBase - an online resource for rRNA-targeted oligonucleotide probes: new features 2007.
436 *Nucleic Acids Research*, **35,** D800–D804.

437 Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT (2012) Direct comparison of Illumina vs. Roche 454 sequencing technologies
438 on the same microbial community DNA sample. *PLoS ONE*, **7**, e30087.

439 Massey FJ Jr (1951) The Kolmogorov-Smirnov test for goodness of fit. *The Journal of American Statistical Association*, **46**, 68–78.
440 Mäki M, Renkonen R (2004) Biosynthesis of 6-deoxyhexose glycans in bacteria. *Glycobiology*, **14**, 1R–15R.
441 Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al*. (2008) The metagenomics RAST server – a public resource for the
442 automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, **9**, 386.
443 Miller MB, Bassler BL (2001) Quorum sensing in bacteria. *Annual Review in Microbiology*, **55**, 165–199.
444 Nadell CD, Xavier JB, Foster KR (2009) The sociobiology of biofilms. *FEMS Microbiology Ecology*, **33**, 206–224.
445 Ochman H, Moran NA (2001) Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science*, **292**, 1096–1098.
446 Ondov BD, Bergman NH, PhillippyAM (2011) Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, **12**, 385.

447 Opelt K, Berg C, Schönmann S, Eberl L, Berg G (2007a) High specificity but contrasting biodiversity of *Sphagnum*-associated bacterial and
448 plant communities in bog ecosystems independent of the geographical region. *The ISME Journal*, **1**, 502–516.
449 Opelt K, Chobot V, Hadacek F, Schönmann S, Eberl L, Berg G (2007b) Investigation of the structure and function of bacterial communities
450 associated with *Sphagnum* mosses. *Environmental Microbiology*, **91**, 2795–2809.
451 Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M *et al*. (2005) The subsystems approach to genome annotation and
452 its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, **33**, 5691–5702.
453 Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH (2013) Going back to the roots the microbial ecology of the rhizosphere.
454 *Nature Reviews Microbiology*, **11**, 789−799.
455 Prakash T, Taylor TD (2012) Functional assignment of metagenomic data: challenges and applications. *Briefings in Bioinformatics*, **13**,
456 711–727.
457 Putkinen A, Larmola T, Tuomivirta T *et al*. (2012) Water dispersal of methanotrophic bacteria maintains functional methane oxidation in
458 *Sphagnum* mosses. *Frontiers in Microbiology*, **3**, 15.
459 Raghoebarsing AA, Smolders AJP, Schmid MC, Rijpstra WIC, Wolters-Arts M, Derksen J *et al*. (2005) Methanotrophic symbionts provide
460 carbon for photosynthesis in peat bogs. *Nature*, **436**, 1153–1156.
461 Rydin H, Jeglum J (2006) Nutrients, light and temperature. In: *The biology of peatlands*, pp. 164–186. Oxford University Press, Oxford.
462 Sessitsch A, Hardoim P, Weilharter A, Krause A, Woyke T, Mitter B (2012) Functional characteristics of an endophyte community
463 colonizing rice roots as revealed by metagenomic analysis. *Molecular Plant-Microbe Interactions*, **25**, 28–36.
464 Scheibe R, Beck E (2011) Drought, desiccation, and oxidative stress. In: *Plant desiccation tolerance* (eds Lüttge U, Beck E, Bartels D), pp.
465 209–232. Springer-Verlag, Berlin, Heidelberg.
466 Schimel JP, Bennett J (2004) Nitrogen mineralization: challenges of a changing paradigm. *Ecology*, **85**, 591–602.
467 Serkebaeva YM, Kim Y, Liesack W, Dedysh SN (2013) Pyrosequencing-based assessment of the Bacteria diversity in surface and
468 subsurface peat layers of a Northern wetland, with focus on poorly studied phyla and candidate divisions. *PLoS One*, **8**, e63994.
469 Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical*
470 *Applications in Genetics and Molecular Biology*, **3**, article 3.
471 Soudzilovskaia NA, Cornelissen JHC, During HJ, van Logtestijn RSP, Lang Si, Aerts R (2010) Similar cation exchange capacities among
472 bryophyte species refute a presumed mechanism of peatland acidification. *Ecology*, **91**, 2716–2726.
473 Strack M (2008) *Peatlands and climate change*. International Peat Society, Jyväskylä, Finnland.
474 Succow M, Joosten H (2001) *Landschaftsökologische Moorkunde*. Schweizerbart`sche Verlagsbuchhandlung, Stuttgart, Germany.
475 Teeling H, Glöckner FO (2012) Current opportunities and challenges in microbial metagenome analysis - a bioinformatic perspective.
476 *Briefings in Bioinformatics*, **13**, 728–742.
477 The Human Microbiome Project Consortium (2012) A framework for human microbiome research. *Nature*, **486**, 215–221.
478 Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A *et al*. (2010) Functional genomic signatures of sponge bacteria reveal
479 unique and shared features of symbiosis. *The ISME Journal*, **4**, 1557–1567.
480 Thomas T, Gilbert J, Meyer F (2012) Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, **2**,
481 3.
482 Tveit A, Schwacke R, Svenning MM, Urich T (2013) Original carbon transformation in high-Arctic peat soils: key functions and
483 microorganisms. *The ISME Journal*, **7**, 299–311.
484 Van Melderen L (2010) Toxin-antitoxin systems: why so many, what for? *Current Opinion in Microbiology*, **13**, 781–785.
485 Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial
486 taxonomy. *Applied and Environmental Microbiology,* **73**, 5261–5267.
487 Wiegert T (2010) Regulated intramembrane proteolysis in bacterial transmembrane signaling. In: *Bacterial signaling* (eds Krämer R, Jung
488 K), pp. 229–246. Willey-VCH, Weinheim.
489 Zachow C, Müller H, Tilcher R, Donat C, Berg G (2013) Catch the best: novel screening strategy to select stress protecting agents for crop
490 plants. *Agronomy* **3,** 794–815.
491
492
493

494 ## Author contributions

495 G.B., H.M. and C.Z. designed research. G.B. collected samples. L.O.-W. and C.Z. performed research,
496 B.H. and G.G.T. assembled contigs and performed statistical analysis. A.B. and H.M. analysed the
497 bioinformatic data. A.B., L.O.-W., H.M., and G.B. wrote the manuscript. All authors discussed results
498 and commented on the manuscript at all stages.

499

500 ## Data accessibility

501 Partial 16S rRNA gene sequences from the complete *S. magellanicum* metagenome and partial gene
502 sequences encoding for nitrogen metabolism from the complete and normalised metagenomes were
503 deposited in the DRYAD repository (http://datadryad.org/pages/repository) under the accession
504 identifier doi:10.5061/dryad.9r816.

505

506 **Tables**

507 **Table 1** Ecologically relevant functional subsystems of *S. magellanicum* metagenome.

Category	Function	Level 2 subsystems	Abundance, seq.[2]	Details	References
Plasticity	Genetic exchange	Phages, prophages[1]	15,076	Temperate bacteriophages (prophages) responsible for horizontal gene transfer	Canchaya *et al*. 2003
		Protein and nucleoprotein secretion system, type IV[1]	12,187	Plasmids carrying type IV secretion system genes, type IV pili and conjugative transport systems	–
		Genetic transfer agents (GTA)[1]	2,047	Phage-like elements in Bacteria	Lang *et al*. 2012
Stability	Stress tolerance	Oxidative stress[1]	15,355	Glutathione-, mycothiol-, rubrerythrin-mediated reactions etc.	–
		Cold shock	1,207		
	Repair	DNA repair	7,967	Base excision and mismatch repair, non-homologous end-joining, homologous recombination and SOS-response systems	–
	Resistance	Resistance to antibiotics and toxic compounds	18,512	Cobalt-zinc-cadmium resistance, multidrug resistance efflux pumps	–
		CRISPs	652	Clustered regularly interspace short palindromic repeats (CRISPRs) – resistance to exogenous genetic elements	Horvath & Barrangou 2010
	Self-control	Programmed cell death ant toxin-antitoxin systems	5,045	Various toxin-antitoxin (programmed cell death) systems: PhD-Doc, YdcED, MazEF etc.	Van Melderen 2010
Interaction	Motility and chemotaxis	Flagellar motility in Prokaryota[1]	7,788	Flagellar biosynthesis proteins and transcription initiation factors	–
		Social motility and nonflagellar swimming in bacteria[1]	307	Rhamnolipids (biosurfactants) in *Pseudomonas*	D'aes *et al*. 2010
	Quorum sensing, biofilm formation and signalling	Quorum sensing and biofilm formation[1]	7,313	Biofilm adhesion biosynthesis, autoinducer-2 and acyl homoserine lactone biosynthesis and processing, symbiotic colonisation and sigma-dependent biofilm formation etc.	Nadell *et al*. 2009

		Capsular and extracellular polysaccharides	6,087	Biosynthesis of rhamnolysated glycans	Mäki & Renkonen 2004
		Proteolytic pathway	3,341	Regulatory intramembrane proteolysis in Bacteria[3]	Wiegert 2010
		Bacterial cytostatics, differentiation factors and antibiotics[1]	1,726	γ-butyrolactones and other morphogens	Kato *et al*. 2006
	Attachment	Adhesion[1]	2,465	Adhesins from non-pathogenic bacteria[3]	Dorhan & Fuqua 2007
		Desiccation stress[1]	562	O-antigen capsule important for plants colonisation (Yih family proteins)	Barak *et al*. 2007
	Nutrition (N-cycling)	Nitrogen metabolism,–	14,751	Dominant: nitrate and nitrite ammonification, nitrogen fixation and ammonium assimilation; Minor: dissimilatory nitrite reductase, allantoin utilization, nitric oxide synthase, cyanate hydrolysis, denitrification and nitrilase	Rydin & Jeglum 2006
		Protein degradation[1]	9,114	Proteolysis in bacteria, eukaryotic and bacterial proteasomes	Schimel & Bennet 2004

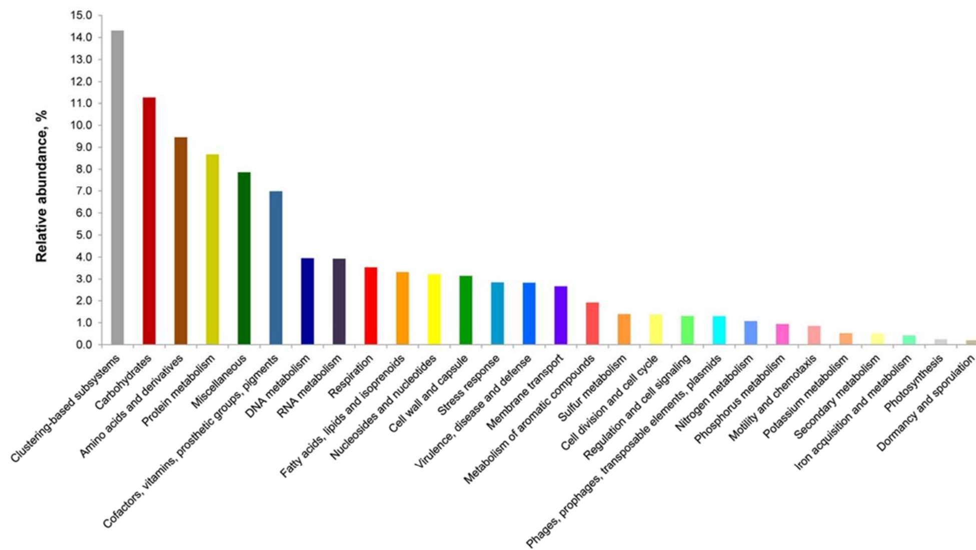
508 [1]Differentialy abundant subsystems show statistically significant difference ($p < 0.05$) between metagenomes of *S. magellanicum*, higher plants and/or peat soils.

509 [2]Sequence abundances correspond to the normalised metagenome that accounts for 657,455 assembled metagenomic sequences.

510 [3]According to the BLASTx alignment.

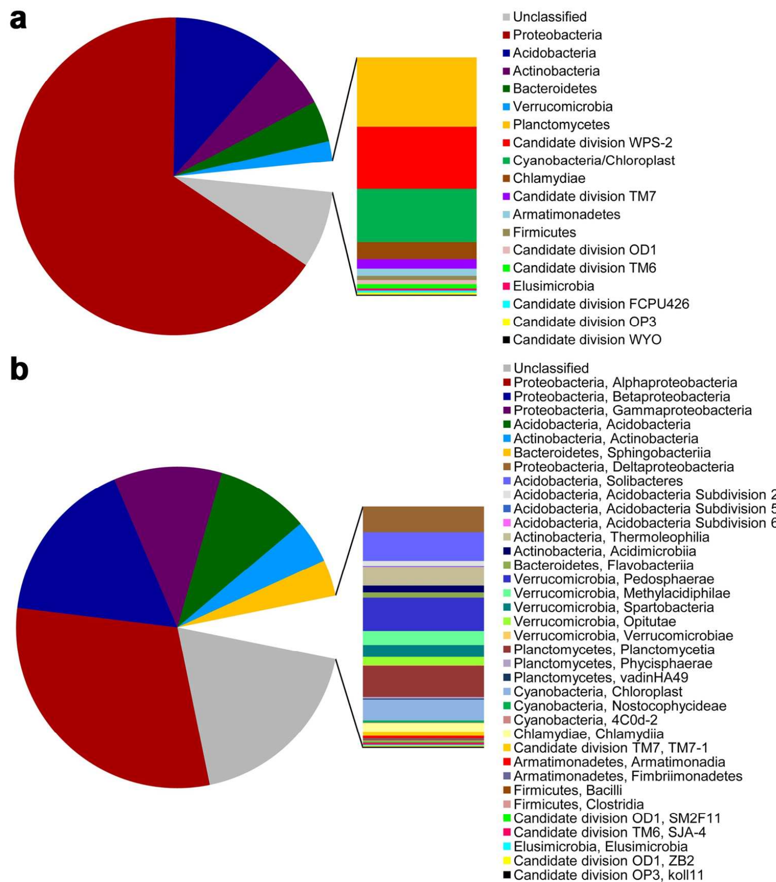
511 # Figure legends

512



513

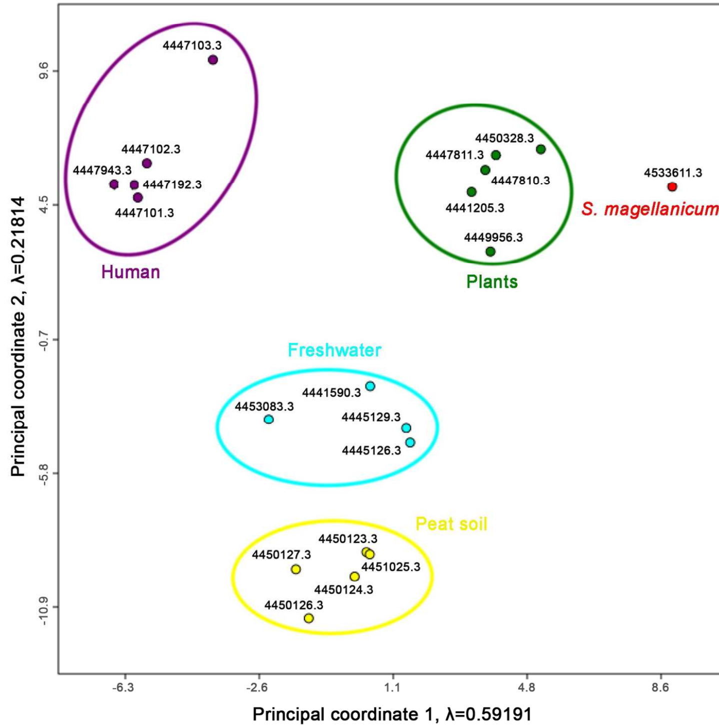
514 **Figure 1** Functional composition of the complete *S. magellanicum* metagenome. Distribution of 35,702,611 metagenomic
515 sequences annotated using functional subsystems of SEED database with max. e-value cut-off of e^{-5}, min. identity cut-off of
516 60%, and min. alignment length of 15 aa (protein annotations) or bp (rRNA annotations). All functional groups are shown at
517 the subsystems level 1.



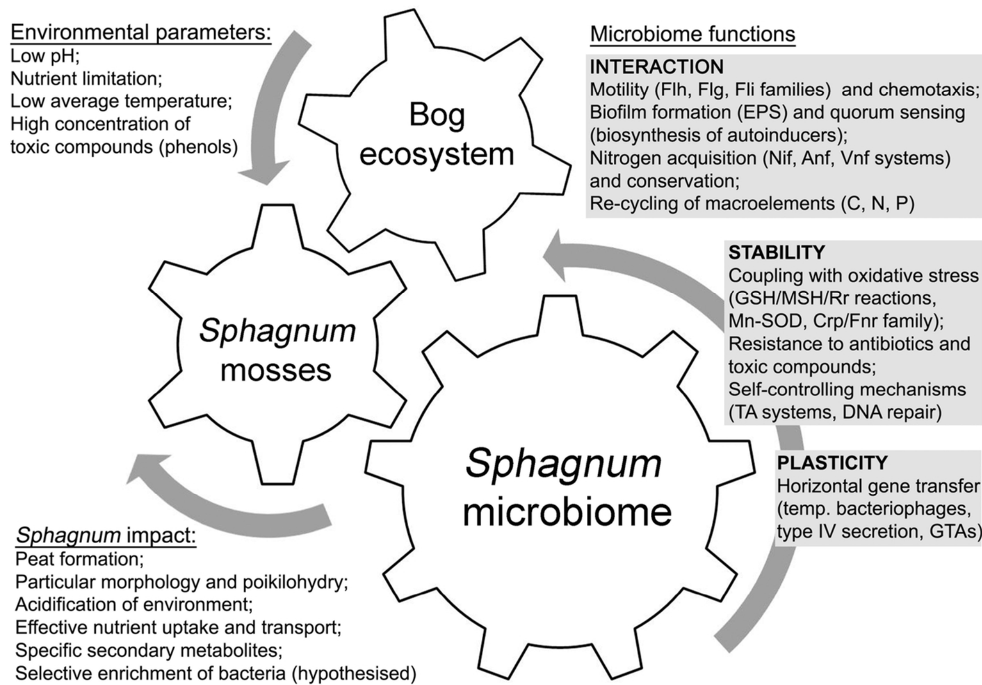
518

519 **Figure 2** Taxonomic composition of *S. magellanicum*-associated bacterial community. 16S rRNA gene sequences were
520 retrieved from the complete metagenome and classified using RDP-classifier with a confidence threshold of 80%. Pie charts
521 represent relative abundance of bacterial taxa at phylum (**a**) and class (**b**) level. Taxa below 1% of relative abundance are
522 shown as separate bar charts.

523



524
525 **Figure 3** Inter-environmental comparison of the complete *S. magellanicum* and publicly available metagenomes by principal
526 coordinate analysis (PCoA). PCoA biplot is based on the Bray-Curtis distance matrix of metagenomes that were assigned to
527 functional subsystems of SEED database and normalised. Single metagenomes are shown by coloured dots with MG-RAST
528 identical numbers (IDs) and grouped according to biome types (coloured ellipses). Eigenvalues (λ) correspond to variation
529 explained by each principal coordinate respectively.
530



531
532 **Figure 4** Model for *Sphagnum*-microbiome biocoenosis. Interaction, stability, and plasticity traits of microbiome were
533 deduced from metagenomic sequences that were annotated using functional subsystems of SEED database. Examples in the
534 brackets are the most pronounced and differentially abundant genetic signatures. Abbreviations: EPS: extracellular
535 polysaccharides; GSH: glutathione; MSH: mycothiol; Rr: rubrerythrin; TA: toxin-antitoxin; GTAs: genetic transfer agents.
536

537 ## Supporting information

538 **Table S1** Description of publicly available metagenomes used in this study.

539 **Table S2** Summary of the Kolmogorov-Smirnov test results. The results are shown for the complete *S.*
540 *magellanicum*, higher plants, and peat soils metagenomes that were tested for normal distribution using the one-
541 sample Kolmogorov-Smirnov-Test and subsystems raw count values.

542 **Table S3** Overview of sequencing data after CLC genomic workbench *de novo* assembly and MG-RAST
543 analysis.

544 **Table S4** Statistical data for differentially abundant functional subsystems of the complete *S. magellanicum*,
545 higher plants and peat soils metagenomes. The table is provided as a separate file (SuppInfo_Table_S4) in Excel
546 format.

547 **Figure S1** This micrograph shows *S. magellanicum* plants (red) as the dominant vegetation component and
548 higher plants such as *Calluna vulgaris* and *Andromeda polifolia* in the bog ecosystem.

549 **Figure S2** Density plots of the statistically analysed metagenomes. Density plots show count data distribution of
550 the complete moss, higher plants, and peat soils metagenomes before and after data transformation using the
551 voom method. The number of subsystems (N) and bandwidth are specified for each plot respectively.

552 **Figure S3** Annotation of the normalised *S. magellanicum* metagenome using various databases. The graph
553 displays the number of predicted coding regions (features) that were annotated using protein databases, protein
554 databases with functional hierarchy information, and ribosomal RNA databases. The bars are coloured by e-
555 value range corresponding to reliability of annotation.

556 **Figure S4** Taxonomic hits distribution of the complete *S. magellanicum* metagenome. Pie charts shows
557 distribution of taxonomic domains (**a**) and bacterial phyla (**b**) for the metagenomic sequences with predicted
558 protein coding regions and ribosomal rRNA genes. The annotation data is based on all reference databases used
559 by MG-RAST.

560 **Figure S5** Colonisation pattern of *S. magellanicum* microbiome visualised by FISH-CLSM. Images show branch
561 leaves of *S. magellanicum* hybridised with Alphaproteobacteria-specific and universal bacterial probes. Yellow:
562 Alphaproteobacteria; red: other bacteria; green: *Sphagnum* chlorocytes; blue: moss cell walls. These images
563 were acquired by confocal laser scanning microscopy and processed by volume rendering of confocal stacks (**a**)
564 and three-dimensional modelling (**b**) using Imaris 7.3. Scale bar = 30 µm.

565 **Figure S6** Functional heatmap of the complete *S. magellanicum*, higher plants. and peat soils metagenomes. The
566 heatmap shows the relation between single metagenomes (horizontal) and their functional subsystems at level 2
567 (vertical). The single metagenomes were grouped using complete linkage clustering with Bray-Curtis as distance
568 metric. The analysis is based on the normalised and scaled sequence counts that are represented by red (low
569 abundance) to green (high abundance) range as specified by the legend.

570 **Figure S7** Dominant functional groups of the normalised *S. magellanicum* metagenome. Interactive Krona plot
571 shows distribution of 657,455 metagenomic sequences annotated using functional subsystems of SEED database
572 with max. e-value cut-off of e^{-5}, min. identity cut-off of 60%, and min. alignment length of 15 aa (protein
573 annotations) or bp (rRNA annotations). All functional groups are shown at the subsystems level 1, which is close
574 to the chart center. Functional groups below 1% and 0.5% of relative abundance at level 2 and level 3
575 respectively are specified as 'minor groups'. The figure is provided as a separate file (SuppInfo_Figure_S7) in
576 HTML format.

577 **Figure S8** KEGG map of enzymes involved in nitrogen metabolism in the *S. magellanicum* metagenomes. The
578 map shows enzyme hits that were detected in the complete (blue) and normalised (red) *S. magellanicum*
579 metagenomes using KEGG mapper of the MG-RAST server with default parameters. The numbers within
580 coloured and transparent blocks specify enzyme entries in the KEGG database. The legend shows enzyme names
581 and hit numbers in each metagenome.

1 **Supporting information**

2 **Table S1** Description of publicly available metagenomes used in this study.

MG-RAST ID	Biome	Habitat	Country	Sequencing method	# of sequences	Alpha-diversity[1]
4450123.3	Soil	Peat soil	Norway	454	108,237	762.552
4450124.3	Soil	Peat soil	Norway	454	89,359	ND
4450125.3	Soil	Peat soil	Norway	454	103,366	729.747
4450126.3	Soil	Peat soil	Norway	454	101,781	773.351
4450127.3	Soil	Peat soil	Norway	454	136,679	786.969
4450328.3	Plant-associated	Rice phyllosphere	Philippines	454	2,213,945	582.844
4449956.3	Plant-associated	Rice rhizosphere	Philippines	454	1,026,982	844.453
4447810.3	Plant-associated	*A. thaliana* phyllosphere	Spain	454	1,098,311	470.804
4447811.3	Plant-associated	Clover phyllosphere	Switzerland	454	1,028,814	421.794
4441205.3	Plant-associated	Soybean phyllosphere	Switzerland	454	523,769	381.345
4453083.3	Freshwater	Freshwater	Australia	454	409,743	1030.286
4441590.3	Freshwater	Lake water	Panama	unspecified	296,355	785.380

4445126.3	Freshwater	Lake bottom mud	Canada	454	256,849	585.456
4445129.3	Freshwater	Lake bottom mud	Canada	454	335,705	525.289
4447101.3	Human-associated	Oral cavity	Spain	454	295,072	299.147
4447102.3	Human-associated	Oral cavity	Spain	454	244,881	217.806
4447103.3	Human-associated	Oral cavity	Spain	454	464,594	ND
4447192.3	Human-associated	Oral cavity	Spain	454	204,218	217.793
4447943.3	Human-associated	Oral cavity	Spain	454	4,447,943	222.364

3 Abbreviations: ND, not determined.

4 [1]Alpha-diversity is estimated by Shannon diversity index (H') for annotations at the species level. The species-level annotations are from all the

5 annotation source databases used by MG-RAST.

Table S2 Summary of the Kolmogorov-Smirnov test results. The results are shown for the complete *S. magellanicum*, higher plants, and peat soils metagenomes that were tested for normal distribution using the one-sample Kolmogorov-Smirnov test and subsystems raw count values.

Metagenomes	Number of subsystems	p-value	Statistic	Alternative
Peat soils	990	< 2.2e-16	0.9368	two-sided
Higher plants	990	< 2.2e-16	0.9744	two-sided
S. magellanicum	198	< 2.2e-16	0.9949	two-sided

11 **Table S3** Overview of sequencing data after CLC genomic workbench *de novo* assembly and

12 MG-RAST analysis.

De novo assembly of the normalised metagenome			
	Count (bp)	Average length (bp)	Total (bp)
Reads	141,411,216	100	14,141,121,600
Matched	34,618,656	100	3,461,865,600
Not matched	106,792,560	100	10,679,256,000
Contigs	1,115,029	501	558,360,453
Contigs min. size	71		
Contigs max. size	121,226		
Contigs N50	557		
Reads in pairs	21,446,968	295.53	
Broken paired reads	13,171,688	100	

MG-RAST statistical analysis		
Type of metagenome	Complete	Normalised/assembled
MG-RAST ID	4533611.3	4516652.3
Upload: sequences count	172,590,841	1,115,029
Upload: mean GC content (%)	56 ± 13	55 ± 9
Artificial duplicate reads: sequence count	0	0
Post QC: bp count	17,299,825,549	558,360,453
Post QC: sequences count	172,590,841	1,115,029
Post QC: mean sequence length (bp)	100 ± 4	500 ± 667
Post QC: mean GC content (%)	56 ± 13	55 ± 9
Processed: predicted protein features	151,683,238	1,411,717

Processed: predicted rRNA features	24,532,966	95,617

13 Abbreviations: QC, quality control.

14

5

15 **Table S4** Statistical data for differentially abundant functional subsystems of the complete *S.*
16 *magellanicum*, higher plants, and peat soils metagenomes. The table is provided as a separate
17 file (SuppInfo_Table_S4) in Excel format.

18

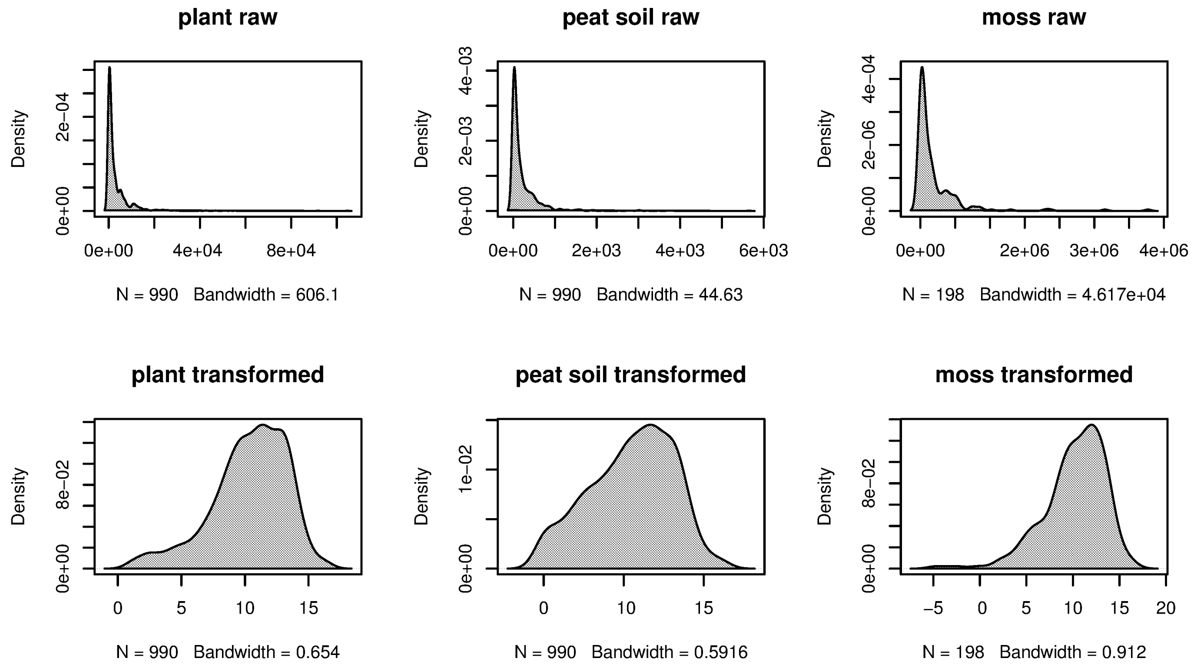
19

20 **Figure S1** This micrograph shows *S. magellanicum* plants (red) as the dominant vegetation

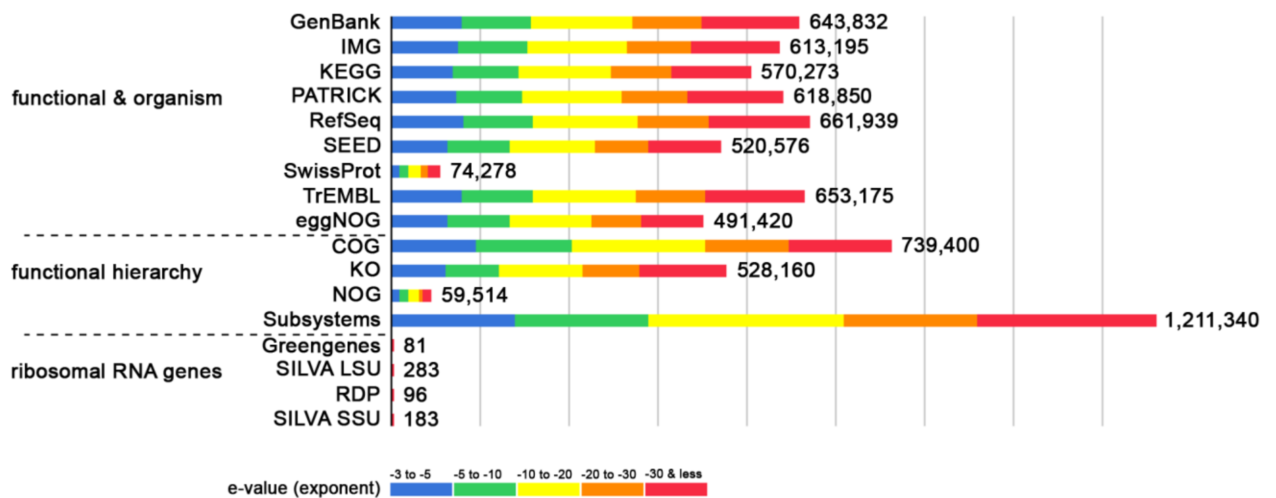
21 component and higher plants such as *Calluna vulgaris* and *Andromeda polifolia* in the bog

22 ecosystem.

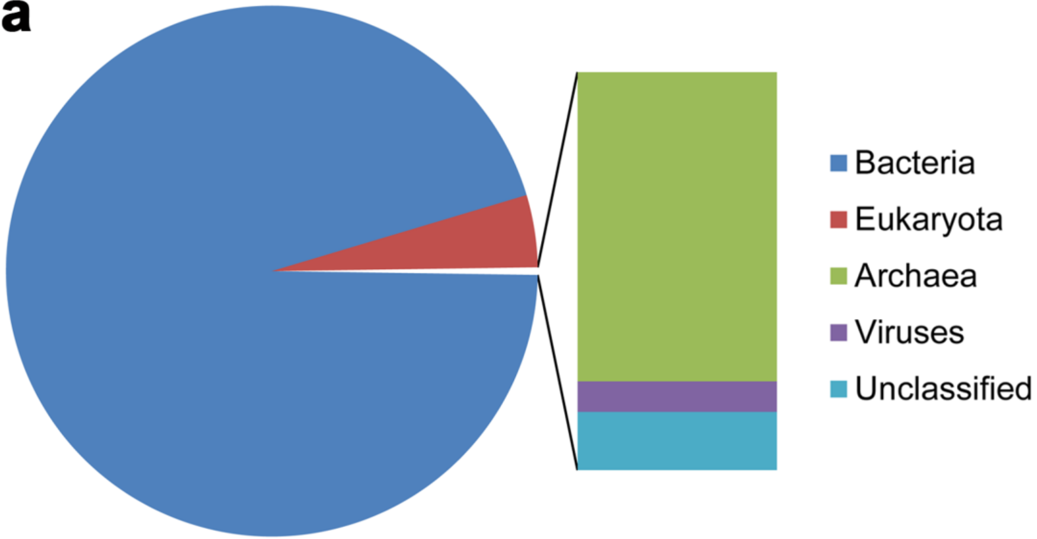
23



24
25 **Figure S2** Density plots of the statistically analysed metagenomes. Density plots show count
26 data distribution of the complete moss, higher plants, and peat soils metagenomes before and
27 after data transformation using the voom method. The number of subsystems (N) and
28 bandwidth are specified for each plot respectively.
29



30
31 **Figure S3** Annotation of the normalised *S. magellanicum* metagenome using various
32 databases. The graph displays the number of predicted coding regions (features) that were
33 annotated using protein databases, protein databases with functional hierarchy information,
34 and ribosomal RNA databases. The bars are coloured by e-value range corresponding to
35 reliability of annotation.
36



37

38 **Figure S4** Taxonomic hits distribution of the complete *S. magellanicum* metagenome. Pie
39 charts shows distribution of taxonomic domains (**a**) and bacterial phyla (**b**) for the
40 metagenomic sequences with predicted protein coding regions and ribosomal rRNA genes.
41 The annotation data is based on all reference databases used by MG-RAST.
42

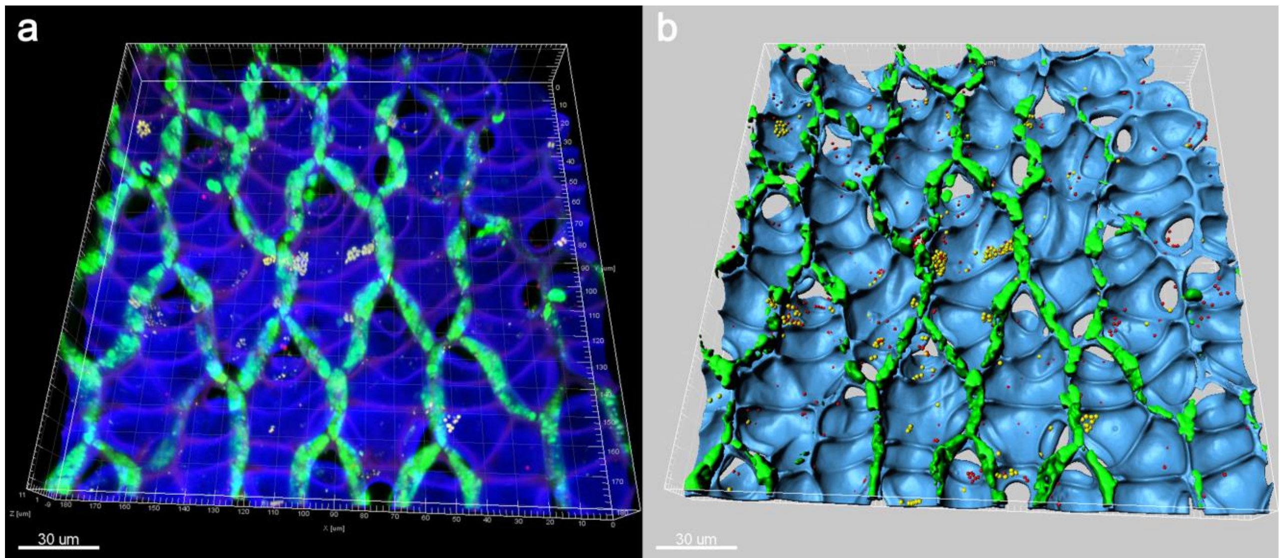


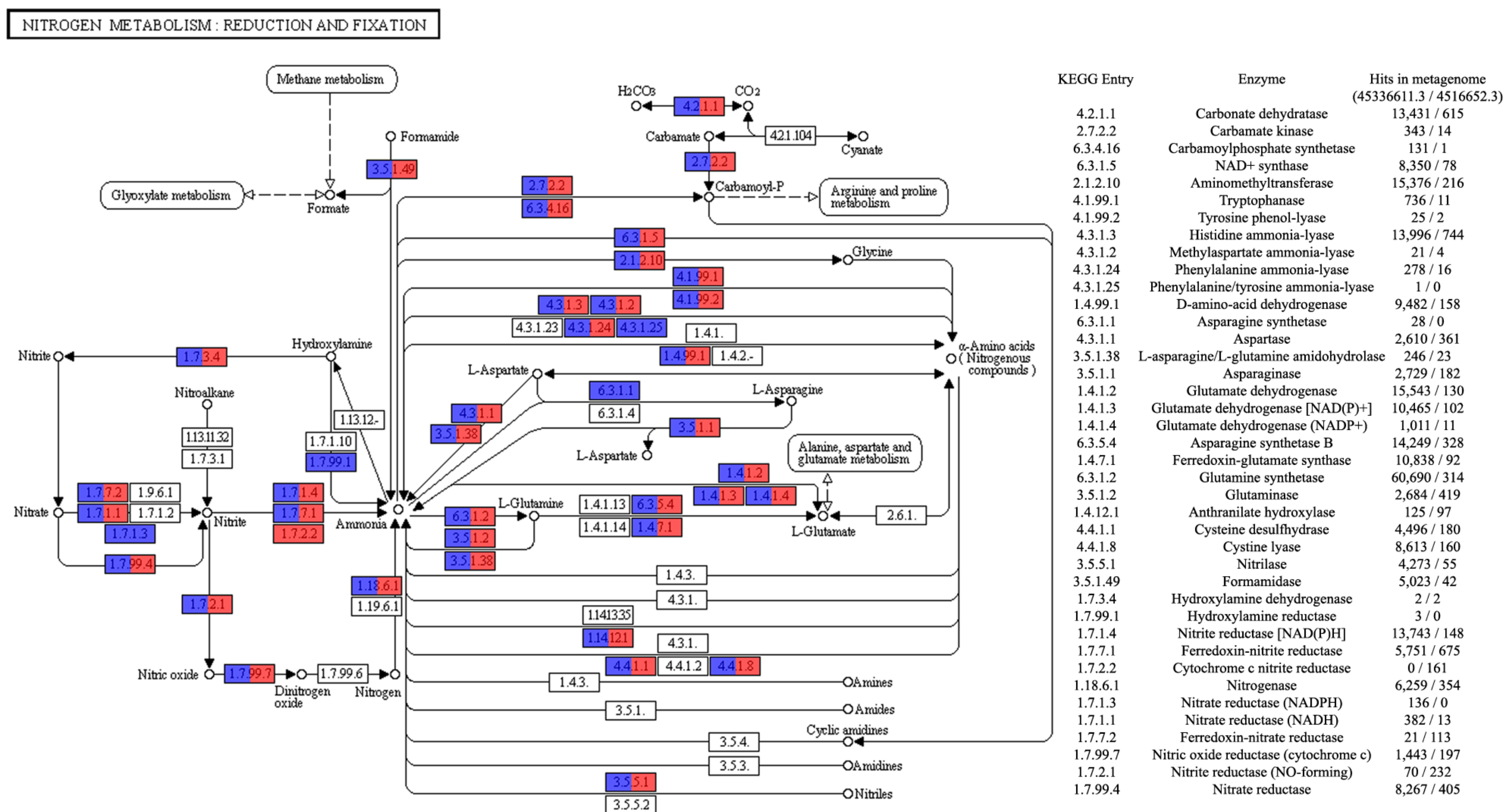
Figure S5 Colonisation pattern of *S. magellanicum* microbiome visualised by FISH-CLSM. Images show branch leaves of *S. magellanicum* hybridised with Alphaproteobacteria-specific and universal bacterial probes. Yellow: Alphaproteobacteria; red: other bacteria; green: *Sphagnum* chlorocytes; blue: moss cell walls. These images were acquired by confocal laser scanning microscopy and processed by volume rendering of confocal stacks (**a**) and three-dimensional modelling (**b**) using Imaris 7.3. Scale bar = 30 µm.

51

13

52 **Figure S6** Functional heatmap of the complete *S. magellanicum*, higher plants, and peat soils
53 metagenomes. The heatmap shows the relation between single metagenomes (horizontal) and
54 their functional subsystems at level 2 (vertical). The single metagenomes were grouped using
55 complete linkage clustering with Bray-Curtis as distance metric. The analysis is based on the
56 normalised and scaled sequence counts that are represented by red (low abundance) to green
57 (high abundance) range as specified by the legend.
58

59 **Figure S7** Dominant functional groups of the normalised *S. magellanicum* metagenome.
60 Interactive Krona plot shows distribution of 657,455 metagenomic sequences annotated using
61 functional subsystems of SEED database with max. e-value cut-off of e^{-5}, min. identity cut-off
62 of 60%, and min. alignment length of 15 aa (protein annotations) or bp (rRNA annotations).
63 All functional groups are shown at the subsystems level 1, which is close to the chart centre.
64 Functional groups below 1% and 0.5% of relative abundance at level 2 and level 3
65 respectively are specified as 'minor groups'. The figure is provided as a separate file
66 (SuppInfo_Figure_S7) in HTML format.



KEGG Entry	Enzyme	Hits in metagenome (45336611.3 / 4516652.3)
4.2.1.1	Carbonate dehydratase	13,431 / 615
2.7.2.2	Carbamate kinase	343 / 14
6.3.4.16	Carbamoylphosphate synthetase	131 / 1
6.3.1.5	NAD+ synthase	8,350 / 78
2.1.2.10	Aminomethyltransferase	15,376 / 216
4.1.99.1	Tryptophanase	736 / 11
4.1.99.2	Tyrosine phenol-lyase	25 / 2
4.3.1.3	Histidine ammonia-lyase	13,996 / 744
4.3.1.2	Methylaspartate ammonia-lyase	21 / 4
4.3.1.24	Phenylalanine ammonia-lyase	278 / 16
4.3.1.25	Phenylalanine/tyrosine ammonia-lyase	1 / 0
1.4.99.1	D-amino-acid dehydrogenase	9,482 / 158
6.3.1.1	Asparagine synthetase	28 / 0
4.3.1.1	Aspartase	2,610 / 361
3.5.1.38	L-asparagine/L-glutamine amidohydrolase	246 / 23
3.5.1.1	Asparaginase	2,729 / 182
1.4.1.2	Glutamate dehydrogenase	15,543 / 130
1.4.1.3	Glutamate dehydrogenase [NAD(P)+]	10,465 / 102
1.4.1.4	Glutamate dehydrogenase (NADP+)	1,011 / 11
6.3.5.4	Asparagine synthetase B	14,249 / 328
1.4.7.1	Ferredoxin-glutamate synthase	10,838 / 92
6.3.1.2	Glutamine synthetase	60,690 / 314
3.5.1.2	Glutaminase	2,684 / 419
1.4.12.1	Anthranilate hydroxylase	125 / 97
4.4.1.1	Cysteine desulfhydrase	4,496 / 180
4.4.1.8	Cystine lyase	8,613 / 160
3.5.5.1	Nitrilase	4,273 / 55
3.5.1.49	Formamidase	5,023 / 42
1.7.3.4	Hydroxylamine dehydrogenase	2 / 2
1.7.99.1	Hydroxylamine reductase	3 / 0
1.7.1.4	Nitrite reductase [NAD(P)H]	13,743 / 148
1.7.7.1	Ferredoxin-nitrite reductase	5,751 / 675
1.7.2.2	Cytochrome c nitrite reductase	0 / 161
1.18.6.1	Nitrogenase	6,259 / 354
1.7.1.3	Nitrate reductase (NADPH)	136 / 0
1.7.1.1	Nitrate reductase (NADH)	382 / 13
1.7.7.2	Ferredoxin-nitrate reductase	21 / 113
1.7.99.7	Nitric oxide reductase (cytochrome c)	1,443 / 197
1.7.2.1	Nitrite reductase (NO-forming)	70 / 232
1.7.99.4	Nitrate reductase	8,267 / 405

Figure S8 KEGG map of enzymes involved in nitrogen metabolism in the *S. magellanicum* metagenomes. The map shows enzyme hits that were detected in the complete (blue) and normalised (red) *S. magellanicum* metagenomes using KEGG mapper of the MG-RAST server with default parameters. The numbers within coloured and transparent blocks specify enzyme entries in the KEGG database. The legend shows enzyme names and hit numbers in each metagenome.

9. Manuscript II

In preparation

Mining for bioactive compounds: analysis of NRPS and PKS genes in the Sphagnum-bog metagenome

Lisa Oberauner-Wappis[1,2], Christina Andrea Müller[1,2], Christin Zachow[2], Henry Müller[2], Peyman Alavi[2], Greg Amos[3], Elizabeth M. Wellington[3] and Gabriele Berg[2]

[1]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Graz, Austria
[2]Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria
[3]School of Life Sciences, University of Warwick, Coventry, United Kingdom

Mining for bioactive compounds: analysis of NRPS and PKS genes in the *Sphagnum*-bog metagenome

Lisa Oberauner-Wappis[1,2], Christina Andrea Müller[1,2], Christin Zachow[2], Henry Müller[2], Peyman Alavi[2], Greg Amos[3], Elizabeth M. Wellington[3] and Gabriele Berg[2*]

[1]Austrian Centre of Industrial Biotechnology (ACIB GmbH), Petersgasse 14, 8010 Graz, Austria

[2]Institute of Environmental Biotechnology, Graz University of Technology, Petersgasse 12, 8010 Graz, Austria

[3] School of Life Sciences, University of Warwick, Gibbet Hill Campus, Coventry CV4 7AL, United Kingdom

[*]Correspondence: Gabriele Berg

Address: 6530 Institute of Environmental Biotechnology, Graz University of Technology, 8010 Petersgasse 12/I, Graz, Austria

E-mail: gabriele.berg@tugraz.at

Running title: NRPS/PKS screening of the *Sphagnum* microbiome

Abstract

Bog ecosystems are one of the oldest vegetation forms harbouring a specific microbial community and high functional diversity. *Sphagnum*-associated bacteria produce diverse bioactive substances, and within the *Sphagnum* metagenome many genes indicate a rich secondary metabolism. Non-ribosomal peptide synthetases (NRPS) and Polyketide synthases (PKS) are main biosynthetic systems involved in the synthesis of important biologically active compounds produced by microorganisms. *Sphagnum*-associated microorganisms were used to discover such NRPS and PKS genes within the moss metagenome, where degenerated PCR primers were employed successfully to amplify NRPS and PKS gene sequences for screening of the metagenomic fosmid library (90,000 clones). The sequences retrieved displayed high homology to the gene sequences of the genera *Pseudomonas* and *Pectobacterium*. In parallel, using an *in silico* Illumina-based metagenomic approach followed by *de novo* assembly and BLASTx analysis 328 NRPS, 456 PKS as well as 57 of their hybrid gene clusters were identified. Most of the sequences were ascribed to genera of *Streptomyces*, *Paenibacillus, Mycobacteria* and *Lysobacter* within the *Sphagnum* microbiome. This study highlights the role of the bog ecosystem as a potential source for detection of secondary compounds and enzymes with a serves as a source for biotechnological applications.

Keywords: *Sphagnum* moss, fosmid library, *in silico* analysis, NRPS/PKS

Introduction

Microbial communities are able to colonize all types of environments (soils, oceans, atmosphere), but also other organisms, the human body and plants (Turner *et al*., 2013). The plant microbiome has gained attention in the last years as important player for plant health and agricultural productivity (Berg *et al*., 2013; Bulgarelli *et al*., 2013; Berendsen *et al*., 2012). Mosses are a major group of land plants colonised by microorganisms, which are adapted to the abiotic conditions of bog ecosystems. These ecosystems are a unique extreme habitat displaying high acidity, low temperature and water saturation, together with extremely low concentrations of mineral nutrients (Richardson *et al*., 1978). Especially *Sphagnum* bogs reflect their vast importance because of the approved role in the global carbon cycle (Succow and Joosten, 2001; Raghoebarsing *et al*., 2005) and is therefore been used globally as an indicator of climate change (Whinam and Copson, 2006; Granath *et al*., 2009).

The role of Sphagnum mosses as an important model for examining the plant-microbe interactions as well as the ecology of plant-associated bacteria has been reported (Raghoebarsing *et al*., 2005; Putkinen *et al*., 2012). Moreover, moss-associated microorganisms, beneficial and pathogenic, living in the peat are well studied (Opelt and Berg, 2004; Opelt *et al*., 2007). Particularly *Sphagnum* mosses are characterized by a highly functional diversity and a specific but diverse microbial community during their whole lifecycle (Bragina *et al*., 2012, 2014). Plant-associated bacterial communities fulfil important functions in cooperation with their host, promoting plant growth by an enhanced nutrient supply and showing antagonistic activity against plant pathogens (Opelt *et al*., 2007ab). In fact, high abundance of functional systems responsible for oxidative and drought stress, genetic exchange, repair and resistance were detected recently by a novel metagenomic analysis of the *Sphagnum* microbiome (Bragina *et al*., 2014). Additionally, Bragina and colleagues (2014) reported multiple interactions among the microbiome and the host plant, indicated by the occurrence of diverse genes necessary for quorum sensing, biofilm formation and nutrient exchange. Plants and their inhabit microorganisms provide an enormous reservoir of chemically diverse natural products with potent biological activity. It is known, that *Sphagnum* species produce bioactive secondary metabolites influencing their microbial colonisation (Basile *et al*., 1999; Zhu et al., 2006; Opelt *et al*., 2007a). The culture-dependent analysis of endo- and ectophytic bacterial populations revealed that the *Sphagnum* moss harboured a high proportion of antifungal as well as a lower proportion of antibacterial isolates (Opelt *et al*., 2007a). The major fraction (97 %) of microbial communities associated to *Sphagnum* mosses belong, however, to non-cultivable forms (Bragina *et al*., 2012). Therefore, the huge potential of the microbiome for biotechnological and biomedical applications remains mostly unexplored.

Prominent classes of natural products from microbial origins (e.g. antibiotics, antifungal and anticancer agents) are synthesized by large multi-modular enzymes, i.e. non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) or hybrids thereof (Walsh, 2008). Both NRPS

83 and PKS genes are organized in a modular assembly line fashion for production of complex chemical
84 scaffolds from simple metabolite building blocks (e.g. acyl-CoAs for polyketides, as well as
85 proteinogenic and nonproteinogenic amino acids for non-ribosomal peptides) (Walsch, 2008). Eeach
86 catalytic module consists of a set of enzymatic domains responsible for one round of chain elongation
87 and a variable set of modifications on each intermediate (Staunton and Weissman, 2001; Marahiel,
88 2009). Because of the structural and functional similarities between elements of each class, NRPS and
89 PKS can form mixed assembly lines, so called hybrid gene clusters (Du *et al*., 2001, Boettger and
90 Hertweck, 2013).
91
92 New metagenomic approaches have been developed, permitting the assessment and exploitation of the
93 taxonomic as well as functional diversity of microbial communities (Ferrer *et al*., 2009; Gilbert and
94 Dupont, 2011). The discovery of new biocatalyst for production of natural active compounds can be
95 accomplished through screening of metagenomic libraries. Different metagenomic applications were
96 recently applied for detection of NRPS/PKS gens of bacterial communities in soil (Parsley *et al*.,
97 2010) and marine environments (Hodges *et al*., 2012). PCR-based screening of metagenomic libraries
98 is a successful approach for the discovery of novel genes or gene clusters based on sequence
99 similarities to previously known enzymes (Ayuso-Sacido and Genilloud, 2005).
100
101 In this study, our aim was to identify sequences assigned to NRPSs and PKSs genes in the *Sphagnum*-
102 associated bacteria, allowing new insights into potentially novel synthetic pathways and biocatalysts.
103 We applied a PCR-amplification screening employing a fosmid clone library in *E. coli*. Additionally,
104 we performed an *in silico* analysis to mine for PKS, NRPS and hybrids thereof in the moss
105 metagenome. Our results demonstrate how sequence-based screenings can be used to detect NRPS and
106 PKS genes involved in the biosynthesis of secondary metabolites within the *Sphagnum* microbiome.
107

108 Methods

109 *Sampling and total community DNA isolation*

110 Sampling of *S. magellanicum* gametophytes and DNA extraction procedure were performed as
111 reported in Bragina *et al*. (2014). To construct a fosmid library for PCR-based screening analysis, total
112 community DNA was extracted using the Meta-G-Nome™ DNA Isolation Kit (Epicentre, Madison,
113 WI, USA) according to the manufacturer's protocol. Metagenomic DNA was randomly sheared to
114 fragment sizes approximately 40 kb and was directly used for the construction of the fosmid library.
115

116 *Metagenomic fosmid library generation and PCR screening*

117 A moos microbiome metagenomic fosmid library was constructed using the CopyControl Fosmid
118 Library Production Kit (EPICENTRE Biotechnologies, Madison, WI, USA) as described by the
119 manufacturer's instructions. The genomic DNA of approximately 40 kb was ligated into the

120 CopyControl pCC2FOS™ vector (EPICENTRE) and packaged with MaxPlax Lambda Packaging
121 Extracts (EPICENTRE). The titer of the packaged phage particles was determined using *E. coli*
122 EPI300-T1 cells (EPICENTRE) as host. The fosmid library was spread onto LB agar plates containing
123 12.5 µg/ml chloramphenicol and incubated at 37°C overnight. In total, 9500 clones were transferred to
124 96-well microtiter plates (MTP) containing 150 µl of liquid LB medium with chloramphenicol
125 (12.5µg/ml) using sterile tooth picks, each plate consisting of ninety-five different clones and one
126 negative control (only medium). The cultures in MTP were allowed to grow at 37°C overnight by
127 shaking at 225 rpm and finally stored at -70°C after addition of glycerol to a final concentration of 25
128 % (v/v) to each well. Restriction digestion analysis was performed with the enzyme BamHI. For the
129 PCR screening, 10 clones were pooled together for a total of 10 MTP-pools in LB medium (12.5µg/ml
130 chloramphenicol). The pooled MTPs were cultivated under the conditions described above upon
131 addition of Fosmid Autoinduction Solution (2 µl/ml, EPICENTRE) to induce high copy number and
132 were finally diluted 1:2 with ddH$_2$O. A denaturation step of 15 min at 99°C and a centrifugation at
133 4000 rpm for 5 min of the 10 MTP pools were performed in order to make the fosmid DNA accessible
134 for the PCR screening with the three degenerate primer pairs PKS, NRPS1 and NRPS2 (Tab. 1). A
135 standard PCR reaction (25 µl) contained 1x Taq 2xMaster Mix (12.5 µl, New England Biolabs,
136 Ipswich/UK), 0.4 µM of each primer (1 µl degenerated primer, Table 1; Sigma-Aldrich, St
137 Louis/USA), ddH$_2$O (4,25 µl), 5% (v/v) DMSO (1.25 µl) and 5 µl of pooled template DNA (95°C, 5
138 min; 35 cycles of 95°C, 1 min; 49°C, 1 min for degPKS or 57°C, 1 min for degNRPS1/NRPS2; 68°C,
139 1 min; and elongation at 68°C, 10 min). PCR probes were subsequently analyzed by 2%
140 agarose/TAE gel electrophoresis. Localization of positive clones was achieved by repetition of the
141 PCR as described above, employing in this case the 10 single clones from the previously identified
142 positive MTP-pool.
143
144 *Sequencing of fosmid clones*
145 Single fosmid clones identified as a positive hit during PCR screening were amplified with non-
146 degenerated PCR primers (Tab. 1) employing the above mentioned PCR program. The resulting PCR
147 products were purified using a Wizard® SV 96 PCR Clean-Up System (Promega) and sent for Sanger
148 sequencing at the sequencing core facility LGC Genomics (Berlin, Germany). Nucleotide BLAST
149 analysis of the obtained sequences was performed using the NCBI BLASTx program (Altschul *et al*.,
150 1997) against a comprehensive, nonredundant collection of microbial genome sequences. In order to
151 analyse PKS/NRPS sequences, a search database nucleotide collection (nr/nt) was performed using the
152 BLASTx algorithm.
153
154 *In silico analysis and bioinformatic strategy*
155 The metagenomic dataset of the *Sphagnum*-microbiome (Bragnina *et al*., 2014) was employed for data
156 mining towards NRPS and PKS genes. The generated paired-end reads of both datasets were quality

filtered using the PRINSEQ software (http://prinseq.sourceforge.net) according to the manual. Untreated and normalized datasets were pooled and the generated mixed dataset was *de novo* assembled using the SOAPdenovo2 program (http://soap.genomics.org.cn) and default settings. BLASTx analysis was performed using the scaffolds, contigs or read sequences as query. A minimum identity of 60% was used as threshold. Multiple alignment (ClustalW algorithm) of sequences showing highest identity (>60%: 25 NRPS, 25 PKS, 10 hybrids) and the Neighbour Joining method were applied to create a phylogenetic tree using the program MEGA6.06 (http://www.megasoftware.net) and default settings.

Results

Fosmid library screening

For construction of our fosmid library we used 30 ng/µl of *Sphagnum* DNA, with an estimated size of 3,8 Gb and 90,000 recombinant fosmid clones on agar plates. From this collection 9,500 clones were employed for the screening of NRPS and PKS genes by PCR amplification in MTP using three different degenerated oligonucleotide primer pairs (NRPS1, NRPS2 and PKS). The primers were design to specifically bind to specific domains of the enyzmes (e.g. the phosphopantetheine attachment site) (Amos and Wellington, in prep). In total, 25 NRPS1, 33 NRPS2 and 52 PKS wells containing each a pool of ten clones gave a positive amplification result. A second round of screening of the corresponding single clones resulted in 11 NRPS1, 26 NRPS2 and 25 PKS potentially positive hits. Selected clones were subjected to rescreening by PCR with non-degenerated primer pairs. Finally, PCR products of 2 NRPS1, 4 NRPS2 and 8 PKS distinct positive fosmid clones were sent for sequencing. From the first annotation of the 14 fosmid sequences using BLASTx analysis, only three clone sequences (3-F3, 3-H3 and 7-B9) were assigned to the expected enzymes (Tab. 2), where the other 11 sequences could identify genes corresponding to other enzyme classes. These three clones show high homology to NRPS (83% protein sequence similarity for clone 3-F3, 82% for 3-H3 and 72% for 7-B9) as well as Peptide synthase genes (100% similarity for clone 3-F3 and 92% for clone 3-H3) belonging to the genera *Pseudomonas* and *Pectobacterium* (Tab. 2).

Metagenomic NGS dataset and bioinformatic analysis

The Illumina HiSeq 2x100 paired-end sequencing of *Sphagnum* moss consists of a raw dataset of 17,323 Mbp (pair-number: 86,617,475) and 14,141 Mbp of normalised metagenomic DNA (pair-number: 70,705,608) (Bragina *et al*., 2014). *De novo* assembly of the pooled metagenome yielded 1,062,181 scaffolded contigs featuring a total size of 188,233,190 with an average length of 183 bps (Table S1). BLASTx analysis of the metagenomic dataset revealed that both, NRPS and PKS genes are present in the moss microbiome. Without cut-off settings, the datasets consist of 328 NRPS and 456 PKS genes, where 21 of the NRPSs and 36 of the PKSs could be identified as hybrid gene clusters. Employing a minimum identity of 60 % as a threshold 39 NRPS and 356 PKS genes as well

as 11 NRPS and 27 PKS hybrids were identified in each dataset. Sequences displaying the highest homology (25 NRPS, 25 PKS and 10 hybrid genes) were selected for BLASTx analysis (Tab. S2). To gain a better overview of the three gene clusters found in the metagenomic moss database, a phylogenetic tree of the respective NRPS and PKS sequences as well as their hybrids was generated (Fig. 1). Most notably, we could detect an unexpected high difference between the bacterial diversity within the NRPS dataset. Many NRPS sequences are closely related to members of the genera *Streptomyces*, *Paenibacillus*, *Pseudomonas*, *Burkholderia*, *Bradyrhizobium*, *Actinoplanes* and *Acidobacterium*. Obtained hits were also assigned to the genera *Microcystis*, *Coleofasciculus*, *Gluconacetobacter*, and *Photobacterium*. Two Sequences of uncultured bacteria could also be detected with more than 70 % identity to NRPS genes (Tab. S2). In comparison to NRPS hits, the PKS analysis was characterized by a specific and distinct bacterial composition. Genera of *Mycobacterium* and *Streptomyces* were clearly most abundant in the PKS dataset. Further obtained PKS-sequences were identified as genera of *Burkholderia*, *Bordetella*, *Amycolatopsis*, *Saccharothrix* and *Salinispora*. BLASTx alignment of NRPS and PKS genes also resulted in the detection of bacteria containing sequences of hybrid gene clusters. The majority of these hybrid-gene sequences was affiliated to genera of *Lysobacter* and followed by *Pseudomonas*, *Ralstonia* and *Stigmatella*.

Discussion

Metagenomic analyses are multiple possibilities for the detection of microorganisms and the corresponding enzymes involved in the biosynthesis of secondary metabolites. Non-ribosomal peptide synthetases (NRPS) and Polyketide synthases (PKS) are such enzymes, synthesising diverse groups of natural products with complex chemical structures with an enormous pharmaceutical potential. In this study, we used two strategies to identify NRPS and PKS genes within moss-associated bacteria.

Specific primer pairs were designed and applied for identification of NRPS and PKS genes through PCR-amplification screening of a moss-metagenomic library. This PCR-screening could identify two NRPS1, four NRPS2 and eight PKS distinct positive fosmid clones. From those only three fosmid clones could be assigned as NRPS genes closely related to the genera *Pseudomonas* and *Pectobacterium*. It has been reported that novel NRPS systems were successfully detected in *Pseudomonas* by new PCR-screening methods (Rokni-Zadeh *et al*., 2011). In this study BLASTx analyses from two out of three fosmid clones (3-F3, 3-H3; identified by primer pair NRPS1) revealed the presence of the predicted genes with significant similarity to peptide synthases. The remaining fosmid clones identified by this method were assigned to other enzyme classes that are clearly related to the synthetic machinery of polyketides and non-ribosomal peptides, such as methyltransferases, reductases or thioesterases (Fischbach and Walsh, 2006).

In silico analysis revealed that both, NRPS and PKS genes are present in the moss microbiome. In fact, the BLASTx analysis showed that there is a clear difference in bacterial diversity between NRPS and PKS genes in the studied microbiome. The composition of microbial communities

231 derived from NRPS related sequences comprised 4 different bacterial phyla belonging to
232 Proteobacteria (genera *Bradyrhizobium*, *Pseudomonas*, *Burkholderia*, *Photobacterium*,
233 *Gluconacetobacter*), Actinobacteria, (*Actinoplanes*, *Streptomyces*), Firmicutes (*Paenibacillus*) and
234 Cyanobacteria (*Coleofasciculus*, *Microcystis*). The widespread occurrence of the NRPS and PKS
235 genetic machinery across the phyla Proteobacteria, Actinobacteria and Firmicutes has been reported
236 (Wanga *et al*., 2014). Cyanobacteria are also a rich source of structurally diverse oligopeptides that are
237 mostly synthesized by NRPSs. In *Microcystis*, a common cyanobacteria genus, various bioactive
238 peptides have been identified (Welker and Dohren, 2006). Compared to the NRPS results, the PKS
239 sequences were assigned to the phylum *Actinobacteria* (> 90 % sequence similarity) including the
240 majority of genera *Mycobacterium* and *Streptomyces*. These two *Actinobacteria* are also prominent for
241 their ability to produce bioactive natural products including NRPSs and PKSs (Chen *et al*., 2012).
242 Interestingly, mixed/hybrid NRPS-PKS gene clusters were also present within the moss microbiome
243 dataset. All sequences were assigned to the phylum *Proteobacteria*, being the most abundant genus
244 *Lysobacter*, which is emerging as a promising source of new bioactive natural products (Xie *et al*.,
245 2011), followed by *Stigmatella*, *Pseudomonas* and *Ralstonia*. It has been reported, that there is a
246 limited number of such mixed NRPS/PKS clusters existing in microorganisms (Garcia *et al*., 2012).
247 Notably, a crucial point for discussion is that after assembly of our metagenomic data, the
248 contigs and scaffolds comprised relatively short sequences resulting in short alignment lengths
249 between 23 and 71 bp (see Supplementary Table S2). Unfortunately, short or error-prone sequences
250 are a commonly encountered problem and the *de novo* assembly of DNA sequences is one of the
251 biggest limitation factors within metagenomic analyses (Teeling and Glöckner, 2012). However, next
252 generation sequencing represents a huge opportunity to gain new insights into bacterial composition
253 and their function in plant microbiomes (Mendes *et al*., 2013; Knief, 2014).
254 The two applied methodological strategies, the screening of a moss metagenomic fosmid
255 library and the *in silico* analysis of the corresponding NGS Illumina HiSeq sequence database,
256 provides an excellent overview of the composition of NRPS and PKS gene clusters within the
257 *Sphagnum* microbiome. The PCR-amplification based metagenomic screening is a successful
258 approach for the discovery of novel gene sequences and their application in biotechnological
259 processes.
260
261
262 **Acknowledgements**
263 We thank Gerhard G. Thallinger (Graz), Cornelia Rainer (Graz) and Barbara Fetz (Graz) for helpful
264 discussions and valuable technical support. This work has been supported by a grant to G.B. from the
265 Federal Ministry of Economy, Family and Youth (BMWFJ), the Federal Ministry of Traffic,
266 Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the

267 Standortagentur Tirol and ZIT - Technology Agency of the City of Vienna through the COMET-
268 Funding Program managed by the Austrian Research Promotion Agency FFG.
269

270 **References**

271 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ. (1997). *Gapped BLAST and PSI-*
272 *BLAST: a new generation of protein database search programs*. Nucleic Acids Res. 25:3389-3402.
273 Ayuso-Sacido A. and Genilloud O. (2005). *New PCR Primers for the Screening of NRPS and PKS-I Systems in*
274 *Actinomycetes: Detection and Distribution of These Biosynthetic Gene Sequences in Major Taxonomic Groups*. Microb Ecol.
275 49(1):10-24.
276 Basile A, Giordano S, López-Sáez JA and Cobianchi RC. (1999). *Antibacterial activity of pure flavonids isolated from*
277 *mosses*. Phytochemistry 52:1479-1482.
278 Berendsen RL, Pieterse CMJ, Bakker PAHM. (2012). *The rhizosphere microbiome and plant health*. Trends Plant Sci. 17
279 478–48610.1016/j.tplants.2012.04.001
280 Berg G, Zachow C, Müller H, Philipps J, Tilcher R. (2013). *Next-generation bio-products sowing the seeds of success for*
281 *sustainable agriculture. Agronomy*, 3, 648–656.
282 Boettger D and Hertweck C. (2013) *Molecular diversity sculpted by fungal PKS-NRPS hybrids*. Chembiochem 14: 28–42.
283 Bragina A, Berg C, Cardinale M, Shcherbakov A, Chebotar V and Gabriele Berg G. (2012). *Sphagnum mosses harbour*
284 *highly specific bacterial diversity during their whole lifecycle*. The ISME Journal, 1–12.
285 Bragina A, Oberauner-Wappis L, Zachow C, Halwachs B, Thallinger GG, Müller H and Berg G. (2014). *The Sphagnum*
286 *microbiome supports greatly bog ecosystem functioning under extreme conditions*. Molecular Ecology. Submitted.
287 Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E and Schulze-Lefert P. (2013). *Structure and functions of the*
288 *bacterial microbiota of plants*. Annu. Rev. Plant Biol. 64 807–838 10.1146/annurev-arplant-050312-120106
289 Chen Y, Ntai I, Ju KS, Unger M, Zamdborg L, Robinson S, Doroghazi J, Labeda D, Metcalf W, and Kelleher N. (2012). *A*
290 *Proteomic Survey of Nonribosomal Peptide and Polyketide Biosynthesis in Actinobacteria*. J Proteome Res. 11(1):85–94.
291 doi:10.1021/pr2009115.
292 Du L, Sánchez C and Shen B. (2001). *Hybrid Peptide–Polyketide Natural Products: Biosynthesis and Prospects toward*
293 *Engineering Novel Molecules*. Metab Eng 3: 78-95. doi:10.1006/mben.2000.0171. PubMed: 11162234.
294 Fischbach MA and Walsh CT. (2006). *Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic,*
295 *machinery, and mechanisms.* Chem Rev. 106(8):3468-96.
296 Garcia I, Vior NM, Braña AF, González-Sabin J, Rohr J et al. (2012). *Elucidating the Biosynthetic Pathway for the*
297 *Polyketide-Nonribosomal Peptide Collismycin A: Mechanism for Formation of the 2, 2′ -bipyridyl Ring*. Chem Biol 19:
298 399-413. doi:10.1016/j.chembiol.2012.01.014. PubMed: 22444595.
299 Gilbert JA and Dupont CL. (2011). *Microbial Metagenomics: Beyond the Genome*. Ann Rev Mar Sci. 3:347-71.
300 Granath G, Wiedermann MM and Strengbom J. (2009). *Physiological responses to nitrogen and sulphur addition and raised*
301 *temperature in Sphagnum balticum*. Oecologia 161, 481–490.
302 Hodges TW, Slattery M and Olson JB. (2012). *Unique actinomycetes from marine caves and coral reef sediments provide*
303 *novel PKS and NRPS biosynthetic gene clusters*. Mar. Biotechnol. 14:270–280.
304 Knief C. (2014). *Analysis of plant microbe interactions in the era of next generation sequencing technologies*. Front Plant
305 Sci. 21;5:216.
306 Marahiel MA. (2009). *Working outside the protein-synthesis rules: insights into nonribosomal peptide synthesis*. J Pept Sci
307 15: 799–807. doi: 10.1002/psc.1183.
308 Mendes, R., Garbeva, P., and Raaijmakers, J. M. (2013). *The rhizosphere microbiome: significance of plant beneficial, plant*
309 *pathogenic, and human pathogenic microorganisms*. FEMS Microbiol. Rev. 37, 634–663.
310 Opelt and Berg. (2004). *Diversity and Antagonistic Potential of Bacteria Associated with Bryophytes from Nutrient-Poor*
311 *Habitats of the Baltic Sea Coast*. Appl Environ Microbiol. 70(11):6569-79.
312 Opelt K, Berg C, Schönmann S, Eberl L and Berg G. (2007a). *High specificity but contrasting biodiversity of Sphagnum-*
313 *associated bacterial and plant communities in bog ecosystems independent of the geographical region*. The ISME Journal, 1,
314 502–516.
315 Opelt K, Chobot V, Hadacek F, Schönmann S, Eberl L and Berg G. (2007b*). Investigation of the structure and function of*
316 *bacterial communities associated with Sphagnum mosses*. Environmental Microbiology, 91, 2795–2809.
317 Parsley LC, Consuegra EJ, Kakirde KS, Land AM, Harper WF and Liles MR. (2010). *Identification of diverse antimicrobial*
318 *resistance deter- minants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial*
319 *assemblage*. Appl. Environ. Microbiol. 76, 3753–3757.
320 Raghoebarsing AA, Smolders AJP, Schmid MC, Rijpstra WIC, Wolters-Arts, Derksen JM, Jetten MSM, Schouten S, Damsté
321 JSS, Lamers LPM, Roelofs JGM, Op den Camp HJM and Strous M. (2005). *Methanotrophic symbionts provide carbon for*
322 *pho- tosynthesis in peat bogs.* Nature 436, 1153–1156.
323 Putkinen A, Larmola T, Tuomivirta T et al. (2012). *Water dispersal of methanotrophic bacteria maintains functional*
324 *methane oxidation in Sphagnum mosses*. Frontiers in Microbiology, 3, 15.
325 Richardson CJ, Tilton DL, Kadlec .A, et al. (1978). *Nutrient dynamics of northern wetland ecosystems. In Freshwater*
326 *Wetlands – Ecological Processes and Management Potential*. Good, R.E., Whigham, D.F., and Simpson, R.L. (eds). New
327 York, NY, USA: Academic Press, pp. 217–241.
328 Rokni-Zadeh H, Mangas-Losada A and De Mot R. (2011). *PCR detection of novel non-ribosomal peptide synthetase genes in*
329 *lipopeptide-producing Pseudomonas*. Microb Ecol. 2011 Nov;62(4):941-7. doi: 10.1007/s00248-011-9885-9.

330 Staunton J and Weissman KJ. (2001). *Polyketide biosynthesis: a millennium review*. Nat Prod Rep 18: 380–416. doi:
331 10.1039/a909079g.
332 Succow M and Joosten H. (2001). *Landschaftsökologische Moorkunde*. Schweizerbart`sche Verlagsbuchhandlung, Stuttgart,
333 Germany.
334 Teeling H and Glöckner FO. (2012). Current opportunities and challenges in microbial metagenome analysis--a
335 bioinformatic perspective. Brief. Bioinform. doi: 10.1093/bib/bbs039.
336 Turner TR, James EK and Poole PS. (2013). *The plant microbiome*. Genome Biol. 14, 209–219. doi: 10.1186/gb-2013-14-6-
337 209
338 Walsh CT. (2008). *The chemical versatility of natural-product assembly lines*. Acc Chem Res. 41(1):4-10.
339 Wanga H, Fewera DP, Holmb L, Rouhiainena L and Sivonena K. (2014). *Atlas of nonribosomal peptide and polyketide*
340 *biosynthetic pathways reveals common occurrence of nonmodular enzymes*. PNAS, doi: 10.1073/pnas.1401734111.
341 Welker M andvon Dohren H. (2006). *Cyanobacterial peptides – nature's own combinatorial biosynthesis*. FEMS Microbiol
342 Rev 30: 530–563. doi: 10.1111/j.1574-6976.2006.00022.
343 Whinam J and Copson G. (2006). *Sphagnum moss: an indicator of climate change in the sub-Antarctic*. Polar Record 42: 43-
344 49.
345 Xie Y, Wright S, Shen Y, and Du L. (2011). *Bioactive natural products from Lysobacter*. Nat Prod Rep. Nov 8, 2012;
346 29(11): 1277–1287. doi: 10.1039/c2np20064c.
347 Zhu, R.-L., D. Wang, L. Xu, R.-P. Shi, J. Wang and M. Zheng. (2006). *Antibacterial activity in extracts of some bryophytes*
348 *from China and Mongolia*. The Journal of the Hattori Botanical Laboratory 100: 603-615.
349

350 # Figures and Tables

351

Table 1 Oligonucleotide primer-pairs for PCR amplification of NRPS and PKS genes

Degenerate primer-pairs

Target gene	Primername		Sequence(5'-3')	Amplicon size	Source
Nonribosomal peptide-synthetase	NRPS1deg	F1	CGC TGA CCC CCA ACG GNA ARB TNY A	180	Amos and Wellington, in prep
		R	GGA GTG GCC GCC CAR NYB RAA RAA		
	NRPS2deg	F2	CGC GCG CAT GTA CTG GAC NGG NGA YYT	420	
		R	GGA GTG GCC GCC CAR NYB RAA RAA		
Polyketide synthase	PKSdeg	F	GGC AAC GCC TAC CAC ATG CAN GGN YT	350	Amos and Wellington, in prep
		R	GGT CCG CGG GAC GTA RTC NAR RTC		

Non-degenerate primer-pairs

Target gene	Primername		Sequence(5'-3')	Amplicon size	Source
Nonribosomal peptide-synthetase	NRPS1ndeg	F1	CGC TGA CCC CCA ACG G	180	Amos and Wellington, in prep
		R	GGA GTG GCC GCC CA		
	NRPS2ndeg	F2	CGC GCG CAT GTA CTG GAC	420	
		R	GGA GTG GCC GCC CA		
Polyketide synthase	PKSndeg	F	GGC AAC GCC TAC CAA ATG CA	350	Amos and Wellington, in prep
		R	GGT CCG CGG GAC GTA		

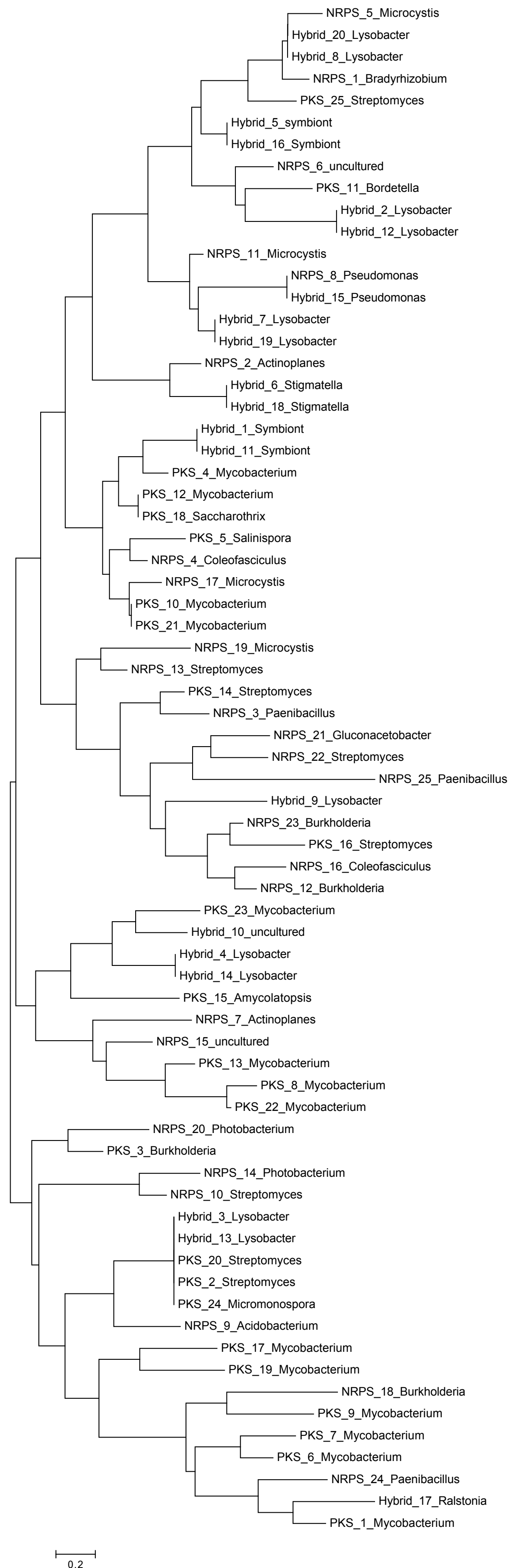
352

Table 2 Sequence alignments of NRPS (Nonribosomal peptide synthetase) and PKS (Polyketide synthases) hits obtained through BLASTx analysis of the moss metagenomic fosmid library

Primer	Clone ID	Closest Blastx hit[a]	Nearest neighbour	Accession No.	Max score	E-value	Max. ident.
NRPS1	3-F3	Peptide synthase	*Pseudomonas fluorescens*	WP_003208543.1	57	8,00E-08	100%
		Non-ribosomal peptide synthetase	*Pseudomonas chlororaphis*	WP_009049996.1	46,6	3,00E-05	83%
NRPS1	3-H3	Peptide synthetase	*Pseudomonas extremaustralis*	WP_010564295.1	76,6	1,00E-14	92%
		Non-ribosomal peptide synthetase	*Pseudomonas chlororaphis*	WP_009049996.	68,9	3,00E-13	82%
NRPS2	7-B9	Non-ribosomal peptide synthetase	*Pectobacterium carotovorum*	WP_010309813.1	197	2,00E-55	72%

353 [a] BLASTx search against the non-redundant protein sequences database at NCBI employing the translated nucleotide query

354



355

356 **Figure 1** The nearest neighbour-joining tree was generated using MEGA6.06 from sequences of bacteria
357 containing NRPS, PKS and hybrid gene clusters derived from metagenomic *in silico* and BLASTx analysis.

358 ## Supplementary information

Table S1 Overview of quality analysis and *de novo* statistics from the *Sphagnum* moss microbiome database (untreated (M_raw) and normalized (M_norm) datasets)

Raw read statistics (Bragina *et al*., 2014)					
Sample	*Pair number*	*Data amount (Mbp)*	*Mean quality score*	*Q30 percent*	
M_raw	86,617,475	17,323	32.98	83.69	
M_norm	70,705,608	14,141	33.64	85.75	
Filtered reads					
Sample	*Pair number*	*Data amount (Mbp)*	*Data amount (percent)*		
M_raw	63,675,767	13,681	78.98%		
M_norm	54,658,254	11,613	82.12%		
De novo statistics of the samples separately and pooled					
Sample	*Scaffolds*	*Mean size*	*Median size*	*N50*	*Total size*
M_raw	851,653	191	129	215	155,084,100
M_norm	396,151	251	188	305	93,793,736
M_pool	1,062,181	183	126	199	188,223,190

359

360

361

362

363

364

365

366

367

368

369

370

371 **Table S2** BLASTx hits for NRPS, PKS and hybrid gene clusters.

Tab S2 Blastx analysis of Nonribosomal peptide synthetase (NRPS), Polyketide synthase (PKS) genes an their hybrid cluster genes.

Nb.	Hit title	Closest Blastx hit	Query	Query length	Hit	Hit lenght	Alignment lenght	Bitscore	E-value	Identity [%]
1	Dimodular nonribosomal peptide synthetase (fragment)	Bradyrhizobium sp. STM 3843	C22031558	155	gi|496258497|ref|WP_008971882.1|	621	51	93,2	1,00E-20	86,27
2	Nonribosomal peptide synthetase, partial	Actinoplanes sp. MN07-A0330	C20459845	100	gi|358030585|dbj|BAL15202.1|	216	32	58,9	9,00E-10	84,38
3	Nonribosomal peptide synthetase subunit	Paenibacillus alvei	C21963473	149	gi|491697684|ref|WP_005549937.1|	4179	49	94	1,00E-20	83,67
4	Nonribosomal peptide synthetase adenylation domain	Coleofasciculus chthonoplastes PCC 7420	scaffold52044	224	gi|57638579|gb|AAW55327.1|	334	24	48,9	4,00E-05	83,33
5	Nonribosomal peptide synthetase	Microcystis aeruginosa K-139	C20570539	100	gi|223717788|dbj|BAH22762.1|	1112	33	62	5,00E-10	81,82
6	Nonribosomal peptide synthetase	uncultured bacterium	C22285759	194	gi|228015632|gb|ACP50608.1|	252	28	46,6	1,00E-04	78,57
7	Nonribosomal peptide synthetase, partial	Actinoplanes sp. MN07-A0362	C21651348	130	gi|358030621|dbj|BAL15220.1|	225	28	45,8	1,00E-04	78,57
8	Nonribosomal peptide synthetase	Pseudomonas sp. CMR12a	C21513819	124	gi|384254960|gb|AFH75321.1|	6466	41	72,8	2,00E-13	78,05
9	Nonribosomal peptide synthetase	Acidobacterium capsulatum ATCC 51196	C20505486	100	gi|225872804|ref|YP_002754261.1|	3102	26	46,6	1,00E-04	76,92
10	Dimodular nonribosomal peptide synthetase	Streptomyces griseoflavus	C21991020	152	gi|491070256|ref|WP_004931880.1|	1360	50	96,3	1,00E-21	76
11	Nonribosomal peptide synthetase, partial	Microcystis aeruginosa NIES-89	C21011967	108	gi|223717803|dbj|BAH22774.1|	83	24	45,1	4,00E-05	75
12	Nonribosomal peptide synthetase	Burkholderia gladioli BSR3	C22329147	205	gi|330821053|ref|YP_004349915.1|	4527	65	99,8	3,00E-22	73,85
13	Nonribosomal peptide synthetase for virginiamycin S	Streptomyces virginiae	C20318812	100	gi|347800593|dbj|BAK86398.1|	4920	30	46,6	1,00E-04	73,33
14	Nonribosomal peptide synthetase	Photobacterium damselae subsp. piscicida	C21354317	118	gi|87251741|emb|CAJ76917.1|	1979	36	57	5,00E-08	72,22
15	Nonribosomal peptide synthetase	uncultured bacterium	C20615495	100	gi|59149847|gb|AAW84165.1|	338	32	46,2	1,00E-04	71,88
16	Nonribosomal peptide synthetase adenylation domain	Coleofasciculus chthonoplastes PCC 7420	C22107944	164	gi|57638577|gb|AAW55326.1|	332	28	46,6	1,00E-04	71,43
17	Nonribosomal peptide synthetase C	Microcystis aeruginosa	C20791725	103	gi|488879470|ref|WP_002791695.1|	4713	32	52,8	1,00E-06	68,75
18	Nonribosomal peptide synthetase	Burkholderia gladioli BSR3	C20582015	100	gi|330821053|ref|YP_004349915.1|	4527	32	50,8	4,00E-06	68,75
19	Nonribosomal peptide synthetase	Microcystis aeruginosa K-139	C20683544	101	gi|223717790|dbj|BAH22764.1|	4732	30	46,6	1,00E-04	66,67
20	Nonribosomal peptide synthetase	Photobacterium damselae subsp. piscicida	C21748862	135	gi|87251742|emb|CAJ76918.1|	3996	42	49,3	2,00E-05	66,67
21	Nonribosomal peptide synthetase DhbF	Gluconacetobacter diazotrophicus PAl 5	C21595578	127	gi|148530016|gb|ABQ82214.1|	238	36	45,4	1,00E-04	66,67
22	Nonribosomal peptide synthetase	Streptomyces griseoviridis	C21687905	131	gi|511513407|gb|AGN74876.1|	2638	43	57	5,00E-08	65,12
23	Nonribosomal peptide synthetase	Burkholderia gladioli BSR3	C21619288	128	gi|330821053|ref|YP_004349915.1|	4527	42	53,5	1,00E-06	64,29
24	Nonribosomal peptide synthetase	Paenibacillus alvei TS-15	C21455397	121	gi|528194692|gb|EPY04934.1|	2448	39	61,2	2,00E-09	64,1
25	Nonribosomal peptide synthetase NRPS	Paenibacillus elgii B69	C21433628	120	gi|337272965|gb|AEI70245.1|	2369	36	47	1,00E-04	63,89

372

373

374

Nb.	Hit title	Closest Blastx hit	Query	Query length	Hit	Hit lenght	Alignment lenght	Bitscore	E-value	Identity [%]
1	Polyketide synthase, partial	Mycobacterium fortuitum	C20578593	100	gi|489978573|ref|WP_003881704.1|	3606	32	65.5	4,00E-11	96.88
2	Polyketide synthase	Streptomyces sp. ID05-A0332	C22185416	174	gi|238768112|dbj|BAH67927.1|	366	31	62.4	5,00E-10	96.77
3	Putative type I polyketide synthase WcbR, partial	Burkholderia thailandensis	C21073822	110	gi|497599280|ref|WP_009913464.1|	339	27	51.2	2,00E-06	96.30
4	Polyketide synthase	Mycobacterium tuberculosis CAS/NITR204	C21940154	147	gi|494698207|ref|YP_007963710.1|	299	48	99.0	8,00E-24	95.83
5	Type I polyketide synthase	Salinispora arenicola	scaffold152789	295	gi|269931547|gb|ACZ54293.1|	224	22	49.3	3,00E-05	95.45
6	Polyketide synthase Pks12 required for biosynthesis of mannosyl-beta-1-phosphomycoketide (MPM)	Mycobacterium canettii CIPT 140070010	C21626375	128	gi|433631164|ref|YP_007264792.1|	4151	42	84.0	3,00E-17	95.24
7	Polyketide synthase	Mycobacterium sp. H4Y	scaffold75089	318	gi|495538454|ref|WP_008263033.1|	1773	56	114	2,00E-40	94.64
8	Polyketide synthase, partial	Mycobacterium xenopi	C21205994	113	gi|490017621|ref|WP_003920151.1|	1851	37	67.8	9,00E-12	94.59
9	Polyketide synthase	Mycobacterium sp. 155	C20999025	108	gi|517431742|ref|WP_018602684.1|	3667	35	68.6	5,00E-12	94.29
10	Polyketide synthase, partial	Mycobacterium avium	C20427117	100	gi|518514357|ref|WP_019684564.1|	1175	33	59.7	3,00E-09	93.94
11	Polyketide synthase	Bordetella bronchiseptica MO149	C20489966	100	gi|410420013|ref|YP_006900462.1|	450	33	62.8	2,00E-10	93.94
12	Polyketide synthase	Mycobacterium marinum M	C20603737	100	gi|183983026|ref|YP_001851317.1|	4187	33	62.0	7,00E-10	93.94
13	Polyketide synthase	Mycobacterium sp. 141	C21991990	152	gi|518815781|ref|WP_019971735.1|	3704	49	96.7	2,00E-21	93.88
14	Type I polyketide synthase, partial	Streptomyces marokkonensis	scaffold8910	204	gi|514826036|gb|AGO59064.1|	422	32	59.7	4,00E-13	93.75
15	Polyketide synthase	Amycolatopsis alba	scaffold158166	325	gi|522123480|ref|WP_020634689.1|	1397	14	31.2	3,00E-15	92.86
16	Polyketide synthase	Streptomyces rapamycinicus	C21379967	119	gi|526996467|ref|WP_020871223.1|	1665	39	73.9	7,00E-14	92.31
17	Polyketide synthase	Mycobacterium avium	scaffold123904	204	gi|489972699|ref|WP_003875898.1|	3679	38	73.9	1,00E-15	92.11
18	Polyketide synthase	Saccharothrix espanaensis DSM 44229	C20603737	100	gi|433606522|ref|YP_007038891.1|	2646	25	47.0	9,00E-05	92.00
19	Polyketide synthase	Mycobacterium ulcerans Agy9	C20960147	107	gi|118617785|ref|YP_906117.1|	4191	35	73.2	1,00E-13	91.43
20	Polyketide synthase	Streptomyces sp. ID05-A0065	C22185416	174	gi|238767302|dbj|BAH67642.1|	351	23	47.0	1,00E-04	91.30
21	Polyketide synthase Pks13	Mycobacterium ulcerans Agy99	C20427117	100	gi|118620000|ref|YP_908332.1|	1783	33	57.0	4,00E-08	90.91
22	Polyketide synthase	Mycobacterium intracellulare]	C20435239	100	gi|497642231|ref|WP_009956415.1|	4177	33	61.2	1,00E-09	90.91
23	Polyketide synthase	Mycobacterium sp. 155	C20687864	101	gi|517431742|ref|WP_018602684.1|	3667	33	60.1	4,00E-09	90.91
24	Macrolide type polyketide synthase	Micromonospora sagamiensis	C20689238	101	gi|5442289|gb|AAD43312.1|AF144052_1	243	33	61.2	2,00E-10	90.91
25	Polyketide synthase, partial	Streptomyces griseolus	C21277327	115	gi|238765884|dbj|BAH66933.1|	268	22	45.4	1,00E-04	90.91

375

376

377

378

Nb.	Hit title	Closest Blastx hit	Query	Query length	Hit	Hit lenght	Alignment lenght	Bitscore	E-value	Identity [%]
1H	Mixed type I polyketide synthase/ nonribosomal peptide synthetase	Symbiont bacterium of Paederus fuscipes	C21039697	109	gi|44662946|gb|AAS47564.1|	8601	23	49,3	2,00E-05	91,3
2H	Hybrid polyketide synthase and nonribosomal peptide synthetase	Lysobacter enzymogenes	scaffold3282	221	gi|119656129|gb|ABL86391.1|	3123	25	48,5	1,00E-04	88
3H	Hybrid polyketide synthase and nonribosomal peptide synthetase	Lysobacter enzymogenes	C22367360	217	gi|119656129|gb|ABL86391.1|	3123	71	126	2,00E-31	87,32
4H	Hybrid polyketide synthase and nonribosomal peptide synthetase	Lysobacter enzymogenes	C21704231	132	gi|119656129|gb|ABL86391.1|	3123	38	66,6	3,00E-11	84,21
5H	Mixed type I polyketide synthase/ nonribosomal peptide synthetase	Symbiont bacterium of Paederus fuscipes	C21477059	122	gi|44662946|gb|AAS47564.1|	8601	40	63,9	2,00E-10	82,5
6H	Mixed trans-AT type I polyketide synthase/nonribosomal peptide synthetase	Stigmatella aurantiaca Sg a15	C21365113	118	gi|359329654|emb|CCA89326.1|	12313	27	49,7	2,00E-05	81,48
7H	Nonribosomal peptide synthetase - polyketide synthase hybrid	Lysobacter lactamgenus	C21520411	124	gi|82503185|gb|ABB80392.1|	5049	41	72	3,00E-13	80,49
8H	Nonribosomal peptide synthetase - polyketide synthase hybrid	Lysobacter lactamgenus	C20934910	106	gi|82503185|gb|ABB80392.1|	5049	34	63,2	3,00E-10	79,41
9H	Nonribosomal peptide synthetase - polyketide synthase hybrid	Lysobacter lactamgenus	C21198394	113	gi|82503185|gb|ABB80392.1|	5049	37	62,4	6,00E-10	78,38
10H	Hybrid trans-AT polyketide synthase - nonribosomal peptide synthetase	Uncultured bacterium psy1	C21993510	152	gi|282554962|gb|ADA82585.1|	12645	30	47,8	1,00E-04	63,33
11H	Mixed type I polyketide synthase/ nonribosomal peptide synthetase	symbiont bacterium of Paederus fuscipes	C21039697	109	gi|44662946|gb|AAS47564.1|	8601	23	49.3	2,00E-05	91.30
12H	Hybrid polyketide synthase and nonribosomal peptide synthetase	Lysobacter enzymogenes	scaffold3282	221	gi|119656129|gb|ABL86391.1|	3123	25	48.5	1,00E-04	88.00
13H	Hybrid polyketide synthase and nonribosomal peptide synthetase	Lysobacter enzymogenes	C22367360	217	gi|119656129|gb|ABL86391.1|	3123	71	126	2,00E-31	87.32
14H	Hybrid polyketide synthase and nonribosomal peptide synthetase	Lysobacter enzymogenes	C21704231	132	gi|119656129|gb|ABL86391.1|	3123	38	66.6	3,00E-11	84.21
15H	Thioesterase of type I polyketide synthase or non-ribosomal peptide synthase like protein, partial	Pseudomonas sp. GM30	C21513819	124	gi|495244949|ref|WP_007969711.1|	643	25	50.1	9,00E-06	84.00
16H	Mixed type I polyketide synthase/ nonribosomal peptide synthetase	symbiont bacterium of Paederus fuscipes	C21477059	122	gi|44662946|gb|AAS47564.1|	8601	40	63.9	2,00E-10	82.50
17H	Polyketide synthase/non ribosomal peptide synthetase protein	Ralstonia solanacearum IPO1609	C20889452	105	gi|207739085|ref|YP_002257478.1|	2313	34	60.5	2,00E-09	82.35
18H	Mixed trans-AT type I polyketide synthase/nonribosomal peptide synthetase	Stigmatella aurantiaca Sg a15	C21365113	118	gi|359329654|emb|CCA89326.1|	12313	27	49.7	2,00E-05	81.48
19H	Nonribosomal peptide synthetase - polyketide synthase hybrid	Lysobacter lactamgenus	C21520411	124	gi|82503185|gb|ABB80392.1|	5049	41	72.0	3,00E-13	80.49
20H	Nonribosomal peptide synthetase - polyketide synthase hybrid	Lysobacter lactamgenus	C20934910	106	gi|82503185|gb|ABB80392.1|	5049	34	63.2	3,00E-10	79.41

10. Curriculum Vitae

Personal

Name	Lisa Oberauner-Wappis
Date of Birth	September 29th 1983
Place of Birth	Spittal/Drau, Austria
Marital status:	Married
Citizenship	Austrian

Education

01/2011 – 08/2014	PhD studies, Institute of Environmental Biotechnology, Graz University of Technology, Graz, Austria PhD in Natural Sciences 08/2014
10/2008 – 12/2010	Master studies in Molecular Microbiology at Karl-Franzens-University and Graz University of Technology within the NAWI Graz project Master of Science (MSc.) 12/2010 Master Thesis "Survival of human-associated microbes on clean room textiles" at the Institute of Environmental Biotechnology/TU Graz under the supervision of Univ.-Prof. Dr.rer.nat. Gabriele Berg
10/2004 – 09/2008	Bachelor studies in Biology at Leopold-Franzens-University in Innsbruck/Austria Bakkalaureat Biologie (Bakk. Biol.) 09/2008
10/2002 – 09/2004	Study of Medicine at Leopold-Franzens-University in Innsbruck
09/1994 – 06/2002	Academic High School, Spittal/Drau

Career

11/2010 – 04/2014	Research assistant, Austrian Centre of Industrial Biotechnology (ACIB GmbH), Graz, Austria
01/2010 – 10/2010	Student Research Fellow, Research Center Pharmaceutical Engineering (RCPE), Graz, Austra

11. Publications

Patent specification

Berg, G.; Liebminger, S.; Oberauner, L.; Klein, T.; Stampf, R.: Clean room device with movable electromagnetic radiation source. WO2012085250 22.12.11

Berg, G.; Liebminger, S.; Oberauner, L.; Klein, T.; Stampf, R.: Photodynamic control of microbial growth on surfaces. WO2012085257 22.12.10

Article in a professional journal

Bragina, A.*; Oberauner-Wappis, L.*; Zachow, C.; Halwachs, B.; Thallinger G.G.; Müller, H.; Berg, G.: The S*phagnum* microbiome supports greatly bog ecosystem functioning under extreme conditions. - in: *Molecular Ecology* (2014); Submitted

Oberauner, L.*; Mahnert, A.*; Bragina, A.; Berg, G.: Complex indoor communities: bacterial life under extreme conditions in clean rooms and intensive care units. - in: Encyclopedia of Metagenomics. (2013); In Press

Oberauner, L.; Zachow, C.; Lackner, S.; Högenauer, C.; Smolle, K.-H.; Berg, G.: The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing. - in: Nature Scientific reports, 3:1413 doi:10.1038/srep01413 (2013)

Liebminger, S.; Aichner, M.; Oberauner, L.; Fürnkranz, M.; Cardinale, M.; Berg, G.: A new Textile-Based Approach to Assess the Antimicrobial Activity of Volatiles. - in: Textile research journal 82, S. 484 - 491 (2012)

Aichner, M.; Oberauner, L.; Liebminger, S.; Fürnkranz, M.; Berg, G.: Volatile organic compounds of plant-associated bacteria to reduce microbial contamination on clean room textile. - in: IOBC/WPRS Bulletin (2010)

Academic Thesis

Survival of human-associated microbes on clean room textiles. - NAWI Graz, Graz University of Technology, Master Thesis (2010)

Lecture within an academic meeting

New insights into microbial communities in intensive care units. - in: ISME 14 - 14th International Symposium on Microbial Ecology. Copenhagen (2012)

Metagenomics to explore the ecology and biotechnological potential of the Alpine bog microbiome. - in: 4th Annual ÖGMBT Meeting. Graz (2012)

Poster used within an academic meeting

Oberauner, L.; Zachow, C.; Lackner, S.; Högenauer, C.; Smolle, K.-H.; Berg, G.: The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing. - in: 5th Annual ÖGMBT Meeting. Innsbruck (2013)

Halwachs, B.; Oberauner-Wappis, L.; Müller, H.; Berg, G.; Thallinger, G.: Sequencing strategies for metagenomes in enzyme discovery. - in: ACIB Science Days 2013. Graz (2013)

Bragina, A.; Oberauner-Wappis, L.; Halwachs, B.; Thallinger, G.; Berg, C.; Müller, H.; Berg, G.: Unique functional diversity within Sphagnum peat bogs. - in: 2nd Thünen Symposium on Soil Metagenomics. Braunschweig (2013)

Oberauner, L.; Müller, H.; Zachow, C.; Thallinger, G.; Bragina, A.; Berg, G.: Metagenomics to explore function and diversity of the Alpine bog microbiome. - in: The 5th Congress of European Microbiologists (FEMS 2013). Leipzig (2013)

Oberauner, L.; Zachow, C.; Lackner, S.; Högenauer, C.; Smolle, K.-H.; Berg, G.: The ignored diversity: complex bacterial communities in an intensive care unit revealed by 16S pyrosequencing. - in: ISME 14 - 14th International Symposium on Microbial Ecology. Copenhagen (2012)

Liebminger, S.; Cernava, T.; Oberauner, L.; Fürnkranz, M.; Cardinale, M.; Berg, G.: Strategies for the management of surface microbial communities based on volatile compounds and visible light. - in: 14th International Symposium on Microbial Ecology (ISME-14). Copenaghen, Danemark (2012)

Liebminger, S.; Oberauner, L.; Zachow, C.; Berg, G.: Innovative concepts for the management of microbial communities in clean rooms. - in: 3th Annual ÖGMBT Meeting. Salzburg (2011)

Liebminger, S.; Aichner, M.; Oberauner, L.; Berg, G.: How much sterility is possible? New concepts for clean room environments based on plant-associated bacteria. - in: 11th Conference on Bacterial Genetics and Ecology - BAGECO11. Corfu-Greece (2011)

Suzzi, D.; Fraser, S.; Schinagl, G.; Radl , S.; Oberauner, L.; Aichner, M.; Cardinale, M.; Khinast, J.; Berg, G.: Innovative Concepts for Personnel Locks in Clean Room Technology. - in: Central European Symposium on Pharmaceutical Technology. Graz (2010)

Liebminger, S.; Cardinale, M.; Oberauner, L.; Moschner, C.; Berg, G.: Survival of Human-Associated and Environmental Bacteria on Clean Room Textiles. - in: International Microscopy Conference. Rio de Janeiro (2010)

12. Acknowledgements

I would like to express my sincere gratitude to all who contributed to the success of this thesis. First, it is with immense gratitude that I acknowledge the great support and continuous guidance of my supervisor Prof. Dr. Gabriele Berg for giving me the opportunity for this doctoral thesis and providing my research visit abroad in England and financial backing.

My acknowledgement is addressed to the Austrian Centre of Industrial Biotechnology (ACIB GmbH) for the financial support of my doctorate study.

Prof. Dr. Martin Grube from the Institute of Plant Sciences of the Karl-Franzens-University is gratefully acknowledged for his contribution as a member of the thesis committee.

I want to thank our project partners Dr. Karl Heinz Smolle, Dr. Christoph Högenauer and Dipl.-Ing. Stefan Lackner from the Graz Medical University as well as Dr. Gerhard Thallinger and Dipl.-Ing. Bettina Halwachs from the Graz University of Technology for their scientific support and fascinating discussions.

Thanks also to Dr. Christian Berg (Karl-Franzens-University) and Dr. Wilfried Franz (University of Klagenfurt) for their assistance during sampling.

I am grateful to Prof. Dr. Liz Wellington from the Warwick University/UK for the scientific research stay in her working group and Greg Amos for his help and instructions in the laboratory.

I owe my deepest gratitude to Dr. Christin Zachow, Dr. Anastasia Bragina, Dr. Henry Müller and Dr. Christina Müller (Graz University of Technology) for their excellent support in practical and theoretical matters.

Special thanks to Barbara, Conny, Monika, Meli, Tanja and Tobija for their technical help during laboratory work.

Greatest thanks also go to Martina, Armin, Tomi, Alex, Matthias, Stefan and all other colleagues and friends from the Institute of Environmental Biotechnology for the pleasant working atmosphere and funny times we had together.

Most importantly, I would like to thank my husband Georg, my whole family and above all my parents for their great support throughout my study.

13.

EIDESSTATTLICHE ERKLÄRUNG

AFFIDAVIT

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzen Quelle wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladenen Textdokument ist mit der vorliegenden Dissertation identisch.

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral dissertation.

_______________________ _______________________________________

Date / Datum Unterschrift / Signatare