METAGENOMICS IN ONE HEALTH–FROM STANDARDIZATION TO TARGETED APPLICATION

Dissertation
for the award of the degree

"Doctor of Philosophy" Ph.D. Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

within the doctoral program Biology
of the Georg-August University School of Science (GAUSS)

submitted by

Luisa Hallmaier-Wacker
from Berlin

Göttingen 2019

Thesis Committee:

Sascha Knauf, DVM, PhD, Work Group Neglected Tropical Diseases, Infection Biology Unit, Deutsches Primatenzentrum GmbH, Leibniz-Institute for Primate Research

PD Dr. Christian Roos, Primate Genetics Laboratory, Deutsches Primatenzentrum GmbH, Leibniz-Institute for Primate Research

Dr. Dietmar Zinner, Cognitive Ethology Laboratory, Deutsches Primatenzentrum GmbH, Leibniz-Institute for Primate Research

Members of the Examination Board:

Reviewer: Sascha Knauf, DVM, PhD

Second Reviewer: PD Dr. Christian Roos

Further members of the Examination Board:

Prof. Dr. Rolf Daniel, Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen

Dr. Alexander Hahn, Junior Research Group Herpesviruses, Deutsches Primatenzentrum GmbH, Leibniz-Institute for Primate Research

Prof. Dr. Stefanie Pöggeler, Genetics of Eukaryotic Microorganisms, Institute of Microbiology and Genetics, Georg-August University Göttingen

Dr. Dietmar Zinner, Cognitive Ethology Laboratory, Deutsches Primatenzentrum GmbH, Leibniz-Institute for Primate Research

Date of the oral examination: 10.05.2019

The more that you read,
the more things you will know.
The more that you learn,
the more places you'll go.

-Dr. Seuss

Contents

A GENERAL INTRODUCTION 1

1 DNA sequencing 1

2 Metagenomics 4

2.1 Targeted metagenomics 5

2.2 Shotgun metagenomics 7

3 One Health 8

4 Application of metagenomics in One Health 9

4.1 Pathogen detection in wildlife 10

4.2 Early detection of emerging infectious diseases 12

4.3 Examining entire microbial communities 12

5 Objective of this thesis 13

B PUBLICATIONS 15

1 Disease reservoirs: from conceptual frameworks to applicable criteria 16

2 The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis 22

2.1 Supplementary information 32

3 Lactation and menstruation shift the vaginal microbiota in captive rhesus monkeys to be more similar to the male urethral microbiota 37

3.1 Supplementary information 51

4 A Metataxonomic Tool to Investigate the Diversity of *Treponema* 59

4.1 Supplementary information 70

C GENERAL DISCUSSION 83

1 The importance of standardization 83

1.1 Standardizing terms for interdisciplinary communication 83

1.2 Establishing sequence thresholds for classification 84

1.3 Inclusion of appropriate controls 86

1.4 The importance of establishing sample handling 87

2 Application of targeted metagenomics in One Health 88

2.1 The importance of baseline data in microbiome analysis 88

2.2 Metataxonomic approach for the detection of *Treponema* 90

3 Nanopore sequencing in One Health 91

4 Concluding remarks 92

References 93

Summary 106

Acknowledgments 108

A GENERAL INTRODUCTION

Recent estimates show that Earth is colonized by as many as 1 trillion microbial species (Locey and Lennon, 2016). The vast majority of the microbial taxa remain undiscovered with less than 0.1% of microorganisms characterized to date (Locey and Lennon, 2016). Historically, microorganisms have been identified and detected using cultivation in the laboratory with (artificial) media (Handelsman, 2004). This classic approach substantially underestimates microbial diversity as many microorganisms cannot be cultured (Handelsman, 2004). The majority of the characterized microorganisms belong to only four phyla (*Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes*), highlighting the historic bias in bacterial detection (Rinke et al., 2013). The advent of sequencing and the subsequent introduction of metagenomics revolutionized microbial ecology as it bypassed the need for cultivation of microorganisms for characterization.

1 DNA sequencing

DNA sequencing, first introduced in 1977 by Sanger et al., is a process in which the precise order of the DNA nucleotides (Adenosine, Guanine, Cytosine and Thymine) is determined (Sanger et al., 1977). Due to its simplicity and high accuracy, polyacrylamide gel and later capillary gel electrophoresis-based Sanger sequencing (Swerdlow and Gesteland, 1990) became the most used sequencing technology and remains an important tool in diagnostics today (Heather and Chain, 2016). Automated Sanger sequencing allowed researchers to move from examining individual gene loci to a genome-wide approach (Hunkapiller et al., 1991). In 2001, the first human genome was published, costing a total of US$ 2.7 billion over 13 years of work (Lander et al., 2001; Venter et al., 2001). The increasing demand for low-cost sequencing technology led to the development of high-throughput sequencing (HTS) techniques, which became commercially available in 2005 (Margulies et al., 2005). HTS technologies are able to generate millions to billions of reads in a single sequencing run and thus drastically reduced the time and cost of DNA sequencing (Metzker, 2010). Different HTS instruments using distinct sequencing biochemistry were first introduced by Roche

(454) (Margulies et al., 2005), followed by Illumina (Solexa) (Bentley et al., 2008) and Life Technologies (ABI/ SOLiD) (Shendure and Ji, 2008). The read length, overall quality and throughput differed amongst these technologies, which have been continuously improved over time (Metzker, 2010). Today, the most successful and widely-adopted sequencing technology is lllumina's Sequencing by Synthesis (SBS) technology, implemented in several commercial HTS platforms including the MiSeq and HiSeq (Table 1; Buermans and Den Dunnen, 2014).

For the SBS process (Bentley et al., 2008), sequencing of DNA fragments takes places on a flowcell, a glass slide containing channels (also known as lanes). Prior to sequencing, the DNA fragment is ligated with adapters that covalently bind to the flowcell. The DNA library then binds randomly to the surface of the flowcell and clusters are formed by bridged amplification to intensify the detectable signal (Fig. 1A). Sequencing is then performed in cycles, where primers, polymerase, and four differently labeled dNTPs are added to the flowcell (Fig. 1B). The dNTPs are tagged with a fluorescent dye, which acts as a terminator blocking further polymerization. Following the addition of a single dNTP to each DNA strand, the flowcell is washed, the fluorescent signal is recorded and the terminator is removed (Fig. 1B). After each subsequent cycle, the Illumina software performs base calling, a process in which the DNA nucleotides are assigned to each measured florescent signal. Given that not all DNA molecules incorporate the dNTPs in each cycle, there is an accumulated effect that leads to lower signal intensities and a higher signal-to-noise ratio (reviewed in Fuller et al., 2009). Additional amplification creates sequencing errors, which lead to a decrease in sequence quality in each cycle and therefore limits the useful read length (reviewed in Fuller et al., 2009). Using the newest chemistry, the MiSeq has a maximum read length of 300 base pair (Buermans and Den Dunnen, 2014). However, the possibility of sequencing both ends of each fragment, also known as paired-end sequencing (Roach et al., 1995), allows for a larger versatility of the technique. The versatility of the HTS technology facilitates the comprehensive analysis of microbes today (Simon and Daniel, 2011).

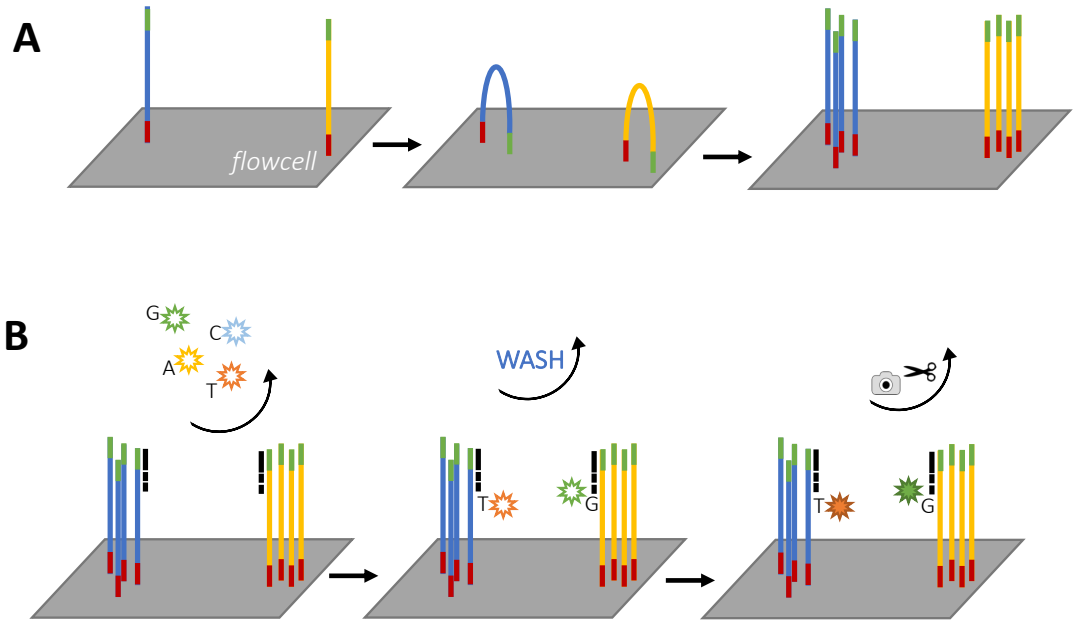


Figure 1. Key steps involved in Illumina's Sequencing by Synthesis (SBS) process (Bentley et al., 2008). Representation of (A) bridge PCR used to amplify signal in SBS and (B) basic mechanism of SBS process.

The recent introduction of long-read sequencing is challenging current HTS approaches (Table 1) (van Dijk et al., 2018). In 2011, PacBio released the first long-range sequencing instrument that uses SMRT (Single Molecule Real Time) technology (Eid et al., 2009). Then in 2014, Oxford Nanopore Technologies introduced nanopore sequencing (Manrao et al., 2012). SMRT and nanopore sequencing allow for real-time sequencing and generate reads from up to 10 kilobases to as many as 1 megabase (van Dijk et al., 2018). Both technologies are being continually developed to reduce the high error rate (depending on system between 3 and 15%) associated with the long reads and improve sample preparation protocols (Jain et al., 2017; Travers et al., 2010). As the technology becomes further developed, long-read sequencing will revolutionize genomic research, including the field of metagenomics (Xu and Zhao, 2018). In particular the portable and affordable MinION (Oxford Nanopore Technologies) makes HTS broadly accessible and applicable in new fields of studies (e.g. clinical applications) (Benítez-Páez et al., 2016; Judge et al., 2015; Quick et al., 2017).

Table 1. Pros and Cons of the select HTS platforms, based on Goodwin et al. (2016) and van Dijk et al. (2018).

Platform	Pros	Cons
Illumina (SBS)	• Low error rate (e.g. MiSeq: 0.1%) • Low cost per Gb • Large amounts of data • Versatile applications; low amount of starting material	• Long sequencing time (12-72 hrs) • Short read length (MiSeq: 300 bp) • High instrument cost • De-novo assembly difficult • No real-time data
PacBio	• Long read length up to 80 Kb • High accuracy (>99.999%) due to circular consensus sequencing technology • No problem with repeat regions, and different GC content	• High instrument cost • High cost per Gb • High error rate with single pass sequencing (~15%) • Large amount of stating material required
Oxford Nanopore	• Long read length up to 1 Mb • Portable, low-cost sequencers with USB connection • Real-time data analysis • Scalability; from small (e.g. MinION) to large and powerful (e.g. PromethION) sequencers	• High error rate (3-15%) including systematic errors with homopolymers • Large amount of stating material required • Quality of DNA dictates read length • Short flowcell shelf life

2 Metagenomics

The term metagenomics, first described in 1998, refers to the analysis of the complete genetic material obtained directly from complex environmental samples of different ecological niches (Handelsman et al., 1998). The collection of all organisms in a particular niche is known as the microbiota (Lederberg and McCray, 2001). With the decreasing costs of sequencing, metagenomics has become largely accessible and has thus been applied in a wide range of sample types, including detection of microbiota of different sites on the human body (e.g. human microbiome project (Turnbaugh et al., 2007)), deep cove aquatic vents (McMahon and Parnell, 2014) and office space environments (Adams et al., 2015). Regardless of sample type, the typical metagenomic workflow follows a straight-forward processing pipeline (Fig. 2). The initial two steps of the workflow, sample collection and DNA extraction, are crucial. Sample collection (including sample storage) and DNA extraction have been repeatedly shown to induce bias in metagenomic studies (Bai et al., 2012; Brooks et al., 2015; Choo et al., 2015; Kim et al., 2017). Hence, careful planning and appropriate controls are

essential, independent of downstream preparation (Kim et al., 2017; Salter et al., 2014). Sample preparation may differ for different sample types and thus microbial mock communities, composed of a known mixture of different microbes, have been used to study ideal conditions (Brooks et al., 2015).

Over the past two decades, metagenomic approaches have shifted from classical Sanger sequencing to HTS technologies. Today, microbial studies use both (1) targeted metagenomics (or metataxonomics (Marchesi and Ravel, 2015)), which relies on sequencing an amplified phylogenetic marker from multiple samples in a barcoded library and (2) shotgun metagenomics, where all extracted DNA is fragmented and sequenced without amplification (Buermans and Den Dunnen, 2014). The appropriate library preparation, sequencing protocol and data analysis depends on the selected metagenomic approach (Fig. 2).

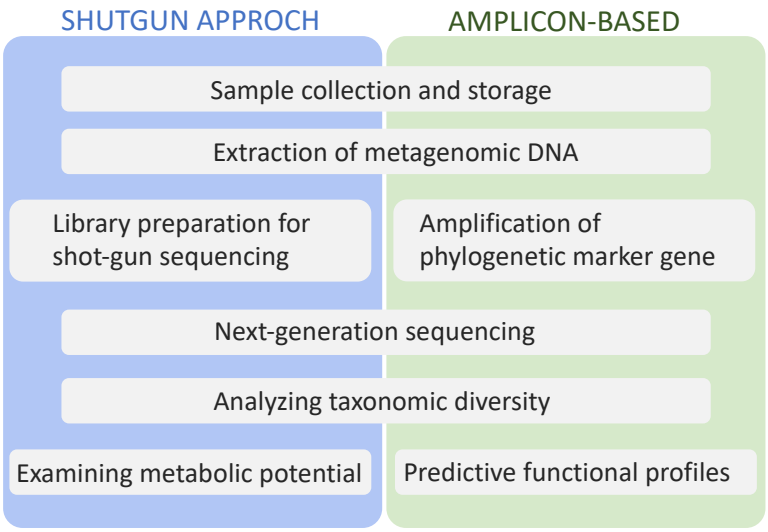


Figure 2. Sample processing pipeline in metagenomic analysis for shotgun and amplicon-based studies.

2.1 Targeted metagenomics

The targeted metagenomic approach, also referred to as metataxonomics (Marchesi and Ravel, 2015), relies on the sequencing of a phylogenetic marker gene to detect a wide range of microbes. An ideal phylogenetic marker gene for phylogenetic

reconstructions should be ubiquitous, aligned easily, and have an informative substitution rate (Patwardhan et al., 2014). The most common phylogenetic marker gene is the small subunit ribosomal RNA (e.g. 16S rRNA) (Amann et al., 1995; Pace, 1997). The 16S rRNA gene, which is approximately 1,500 base pairs in length, is most widely used for characterization of bacteria and archaea (Tringe and Hugenholtz, 2008). Pace et al. (1985) first proposed the usefulness of 16S rDNA Sanger sequencing as an approach to circumvent culture-based bacterial detection. The 16S rRNA gene has nine hypervariable regions interspersed amongst conversed sites (Van de Peer et al., 1996). The hypervariable regions uniquely differ between different types of bacterial and archaeal species, which allows for detection and taxonomic classification (Van de Peer et al., 1996). The conserved regions, on the other hand, are ideal primer binding sites for broad-range amplification of environmental samples (Baker et al., 2003). Sequencing of the 16S rDNA alone excludes the detection of eukaryotic microorganisms, which can have key roles in ecological niches (Bauer et al., 2018). For the detection of fungi and other eukaryote microbes, the internal transcribed spacer, large subunit ribosomal RNA, and small subunit ribosomal RNA (18S rRNA) have been used (Bik et al., 2012; Schoch et al., 2012). For the broadest detection of microbes, a combination of gene targets is often used on environmental samples (Fierer et al., 2007; Kim et al., 2017).

HTS technology does not allow for sequencing of the full-length marker gene (Buermans and Den Dunnen, 2014). Therefore, selecting the ideal variable region and most suitable primers is essential in library preparation of targeted metagenomics (Chakravorty et al., 2007). Not all variable regions have the same capacity of taxonomic classification and therefore studies have focused on identifying ideal targets for different sample types (Albertsen et al., 2015; Chakravorty et al., 2007; Walker et al., 2015). For example, the hypervariable region 4 (V4) of the 16S rRNA gene has become widely used, given that it is short yet provides enough information for reliable bacterial classification (Caporaso et al., 2011; Kozich et al., 2013). Illumina's dual-indexing amplification strategy allows for the simultaneous characterization of hundreds of different samples (Kozich et al., 2013). Additionally, the amplification steps during library preparation allows the detection of low abundance microbes (Salter et al., 2014;

Shah et al., 2011). However, amplification is also known to distort the microbial abundance and limit the detection of rare members of the community (Carlos et al., 2012; Jovel et al., 2016). Therefore, it is essential to minimize amplification cycles and be aware of the resulting bias of the targeted metagenomic approach (Ahn et al., 2012; Brooks et al., 2015; Salter et al., 2014). Despite some of its disadvantages, targeted metagenomics is cost-effective and thus often used to screen a large quantity of samples, i.e. from many patients or for a longitudinal study (Jovel et al., 2016; Lax et al., 2017).

2.2 Shotgun metagenomics

Shotgun metagenomic sequencing has been used to gain a more in-depth understanding of microbial communities (Dutilh et al., 2017; Norman et al., 2014). The library preparation of shotgun sequencing circumvents the need of a universal marker gene, as DNA is fragmented and analyzed directly from samples without amplification (Fig. 2; summarized by Quince et al. (2017)). The fragmentation of extracted DNA leads to random sampling and thus provides a comprehensive coverage of the microbial community without the bias of amplification (Thomas et al., 2012). The resulting sequence reads represent various locations within the collection of microbial genomes of both eukaryotic and prokaryotic organisms in the samples (Quince et al., 2017). Some DNA sequences can be used for taxonomic classification (e.g. the 16S rRNA gene), while others provide information about coding regions (Prakash and Taylor, 2012). The diverse set of reads allows for the detection of microbial genes that encode for different metabolic functions and thus provide insight into the functional potential of each microbiome (Prakash and Taylor, 2012; Thomas et al., 2012). However, to use a shotgun approach, a larger quantity of DNA is required for reliable results and selective enrichment is often needed to improve the ratio of host-to-microbial read output (Thomas et al., 2012). Additionally, metagenomic data from shotgun sequencing is complex and thus data analysis tools are being continuously developed to ease investigation (Dutilh et al., 2017; Lindgreen et al., 2016; Thomas et al., 2012). The ability to detect a wide range of microorganisms simultaneously makes shotgun sequencing a powerful tool that will grow in importance as cost of sequencing

decreases over time (Miller et al., 2013; Norman et al., 2014; Pallen, 2014; Thomas et al., 2012).

3 One Health

It is estimated that 61% of all human pathogens and 75% of emerging infectious diseases are of zoonotic origin (Taylor et al., 2001). Recently emerged zoonotic diseases, such as the Middle East respiratory syndrome caused by MERS-CoV (Zaki et al., 2012), Brucellosis caused by various *Brucella* species (Seleem et al., 2010), and bovine spongiform encephalopathy (BSE) a prion disease (Scott et al., 1999), have highlighted the interconnection between human and animal health (Jones et al., 2008). Additionally, it has become apparent that environmental changes such as climate change, continuous growth of human populations, and the reduction of wildlife habitats are driving forces in zoonotic disease emergence (Rabinowitz et al., 2013; Woolhouse and Gowtage-Sequeria, 2005). To emphasize this close relationship between human, animal and environmental health, the concept of One Health has emerged (Fig. 3) (Daszak et al., 2000). The American Veterinary Medical Association (2008) defines One Health as "the collaborative effort of multiple disciplines—working locally, nationally and globally—to attain optimal health of humans, animals, and the environment". This definition highlights that at the core of One Health is the interconnectedness of scientific fields such as infectious disease, toxicology, ecology, agriculture science, conservation, anthropology and social science. The creation of many initiatives (e.g. One Health Platform), scientific journals (e.g. Lancet Planetary Health) and One Health institutes demonstrates the rapid expansion of the One Health field over the last two decades (Cunningham et al., 2017). This concept has additionally been incorporated in public health policies and political declarations (e.g. UN Political Declaration on Antimicrobial Resistance) (Essack, 2018). For example, the World Health Organization (WHO), the World Organization for Animal Health (OIE), and the Food and Agriculture Organization of the United Nations (FAO) have formed a tripartite alliance and identified priority collaboration areas, such as fighting antimicrobial resistance (FAO/OIE/WHO Joint Scientific Consultation Writing Committee, 2011). Despite the increased awareness of the One Health approach amongst conservationists, ecologists

and veterinarians, many healthcare professions still fail to translate the One Health idea into action and do not incorporate the concept into current clinical research (Cunningham et al., 2017).

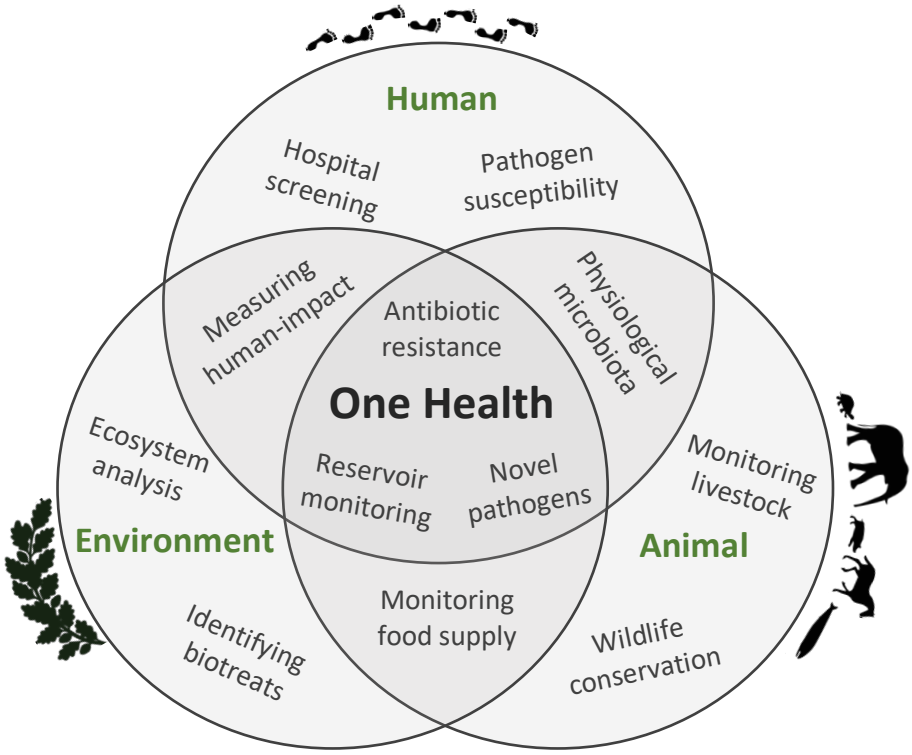


Figure 3. Applications of metagenomics within the triad of One Health. Examples of applications within and between each triad of One Health.

4 Application of metagenomics in One Health

The continual emergence of novel microorganisms in clinical and environmental settings challenges conventional surveillance efforts (Miller et al., 2013). Laboratory techniques must constantly evolve to detect new pathogens in a wide set of different sample types, ideally in real-time. A range of classical diagnostic methods have been used for decades, including microscopy (e.g. histological staining), culture-based analysis, immunoassays (e.g. detection of antigens from pathogen) and nucleic acid amplification tests (NAATs) (Miller et al., 2013). A way to ease the surveillance of various microbes is to use molecular techniques that are pathogen-agnostic, allowing for broad-range non-targeted detection (Miller et al., 2013). Metagenomics is a useful

approach that simplifies the detection of known and novel microbes in a wide range of clinical samples (Goldberg et al., 2015). Instead of focusing on a single assay for a specific pathogen, laboratories can now use a single metagenomic approach to identify a broad range of microorganisms (Pallen, 2014). The use of HTS technology also allows for the detection of pathogen genomes, which can be subsequently used to examine antibiotic resistance and to study disease outbreaks (Robinson et al., 2013). Figure 3 exemplifies the broad applicability of metagenomics in One Health. Despite many of its advantages, metagenomics is not a universal solution in One Health investigations as it only provides limited understanding of the functional and biological relevance of a microorganism (Prakash and Taylor, 2012). Additionally, issues with contamination, inherent bias within the technique (e.g. caused by different DNA extraction methods), and difficulties with storage and analysis of large datasets limit the applicability of metagenomic studies (Miller et al., 2013). It is important to keep in mind the study question and the available resources to evaluate the most suitable approach (e.g. amplicon HTS vs. shotgun metagenomics). In the following section, I will highlight three key applications of metagenomics in One Health.

4.1 Pathogen detection in wildlife

Pathogen detection in wildlife can be challenging due to the diversity of host species, the lack of established and validated assays and the often limited and low-quality sample material. Despite these challenges, screening wild animals is essential to identify disease reservoirs for humans or livestock (Cunningham et al., 2017; Haydon et al., 2002; Viana et al., 2014). Knowledge on the existence of a reservoir is central for disease prevention, management and surveillance. Metagenomic approaches allow for the characterization of microbes in wildlife and vectors of diseases (Qiu et al., 2014; Razzauti et al., 2015). It has been shown that using metagenomic tools, the tick salivary grads can be continually monitored to identify pathogens associated with livestock and human disease (Qiu et al., 2014). As vectors are often integral in cross-species transmission (Fig. 4A), monitoring the tick salivary grads can be an effective surveillance strategy. Continual epidemiological surveys of wildlife including metagenomic approaches are an asset for the identification, prevention and

subsequent control of emerging diseases at the human-livestock-wildlife interface (Kao et al., 2014; Razzauti et al., 2015).

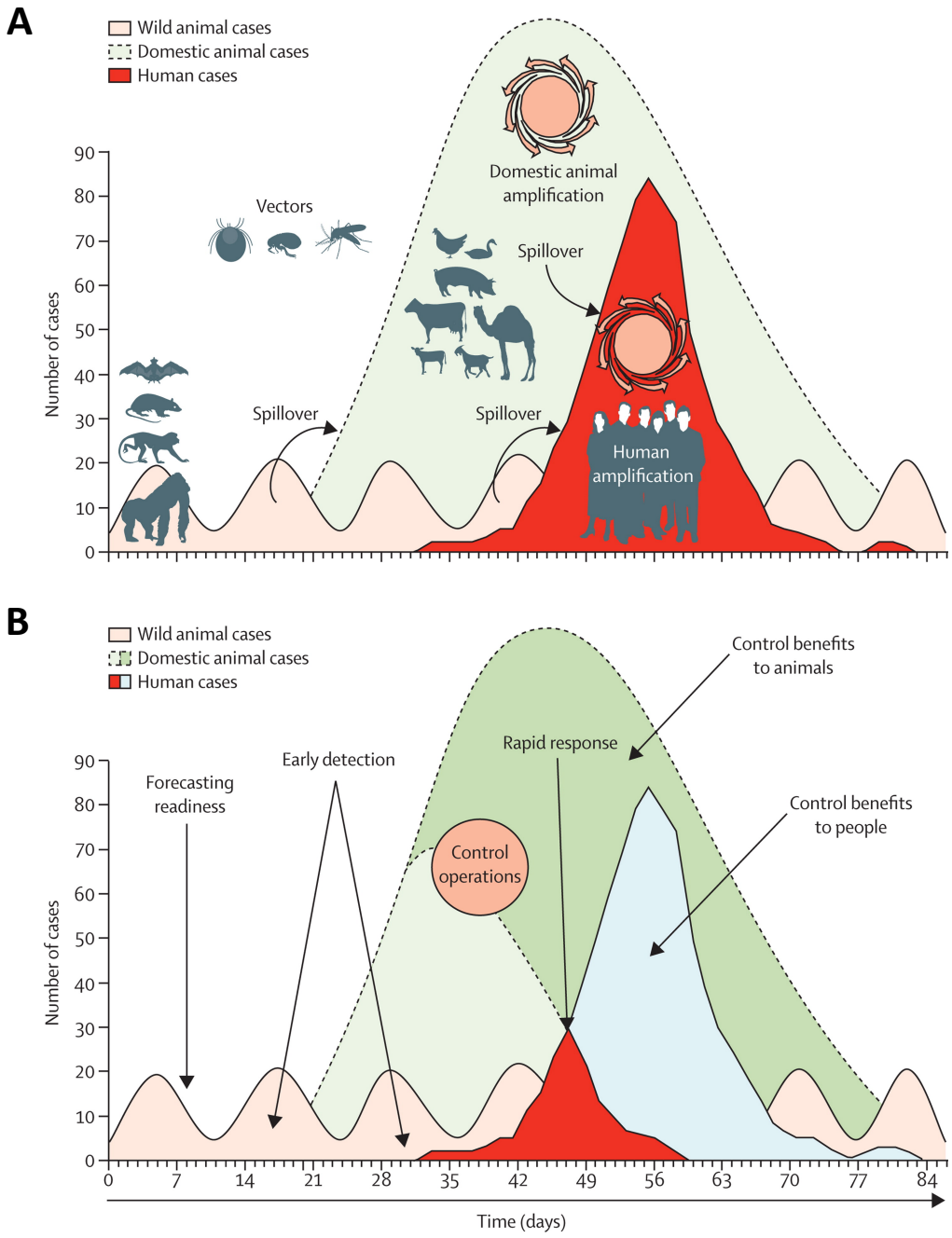


Figure 4. The benefit of surveillance, early detection, and rapid response on disease control. (A) Cyclical pathogen prevalence in wildlife (pink) can lead to spillovers into domesticated animals (light green) or humans (red). Spillover into domesticated animals can amplify the capacity of a pathogen to infect the human population. Vectors may aid in all cross-species transmission events. (B) Surveillance, early detection and rapid response reduce the disease incidence in both animals (dark green) and human population (light blue). Figure is a reprinted from Karesh et al. (2012).

4.2 Early detection of emerging infectious diseases

Early detection of zoonotic and anthropozoonosis threats is essential for rapid response to mitigate disease outbreaks (Fig. 4B; Karesh et al., 2012). However, early discovery of emerging infectious diseases can be difficult as the microorganisms are often well adapted with little to no clinical presentation in a resevoir population (Bäumler and Fang, 2013). Here, metagenomics allows for the detection of microorganisms that are unknown or unsuspected (Gardy and Loman, 2018; Pallen, 2014). For example, a large-scale study on rodents and small mammals in China used shotgun metagenomic sequencing to survey the range of viral families in 20 different provinces (Wu et al., 2018). The study examined the virus in the context of host taxonomy and geographic location, and identified several novel viruses and their evolutionary history (Wu et al., 2018). Continual HTS-based surveilance of zoonotic pathogens in their wildlife and livestock reservoir can be useful to detect early signs for possible spillover events and outbreaks (Fig. 4; Karesh et al., 2012; Wu et al., 2018). In order to continually surveil and discover viruses with zoonotic potential, the PREDICT program was launched by the US Agency of International Development in 2009 (Joly et al., 2016). Whether broad continual surveillance programs are successful at early detection and useful for mediating disease outbreaks is yet to be seen. However, it is unquestionable that information on the virome is beneficial for the early discovery of emerging infectious diseases (Carroll et al., 2018).

4.3 Examining entire microbial communities

The consequences of commensal microorganisms on the health of both animals and humans is often underestimated (Trinh et al., 2018). The diverse communities of bacteria, archaea and microbial eukaryotes that inhabit the various environments of humans and animals have been shown to impact the immune system (Belkaid and Hand, 2014; Round and Mazmanian, 2009), susceptibility to pathogenic microbes (Abt and Pamer, 2014; Buffie et al., 2015; Dennisonet al., 2014), and essential metabolic functions (De Menezes et al., 2011; Kau et al., 2011). Studying the microbiome in health and disease has provided an enhanced understanding of health consequences of non-pathogenic microbes that can be transmitted within the One Health triad (Trinh et al.,

2018). For example, cohabitation with animals, including pets (Song et al., 2013) and livestock (Kraemer et al., 2018), has been linked to a greater diversity in the gut, skin and nasal microbiota (Misic et al., 2015). A more diverse microbiota has been proposed to have a protective effect against inflammatory diseases, such as allergy (Pascal et al., 2018). Thus, restoring or altering the microbiota may provide a promising avenue for preventative therapy against different types of diseases (Pascal et al., 2018; Wang et al., 2018).

5 Objective of this thesis

Metagenomic approaches are valuable techniques in various different types of One Health investigations. The general aim of this thesis was to establish and validate targeted metagenomic tools to identify and study microbial communities in health and disease.

The first objective was to advance the standardization of terms and techniques in the rapidly expanding field of One Health. In Chapter B1, we focused on standardizing applicable criteria to determine a 'disease reservoir', a central concept for diseases of zoonotic and anthropozoonotic origin. Beyond standardizing terminology, we also determined sample handling conditions for metagenomic investigations of swab specimens in Chapter B2. The sampling regime was integral for the validation of all subsequent experimental studies.

The second objective was to examine different applications of targeted metagenomics. In Chapter B3, we studied the urogenital microbiota of rhesus monkeys (*Macaca mulatta*), a common translational animal model for infectious diseases, by using 16S rRNA gene sequencing. 16S rRNA sequencing provides a good overview of all bacterial species but often is not suitable for more detailed investigation. Therefore, to selectively examine the diversity of *Treponema*, we established an amplicon-based metataxonomic tool to investigate the diversity of this genus in Chapter B4.

B PUBLICATIONS

This section presents the following publications:

1 Hallmaier-Wacker LK, Munster VJ, Knauf S. 2017. Disease reservoirs: from conceptual frameworks to applicable criteria: Disease reservoir criteria. *Emerg Microbes Infect* 6: 1-5.

2 Hallmaier-Wacker LK, Lüert S, Roos C, Knauf S. 2018. The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Sci Rep* 8: 6292.

3 Hallmaier-Wacker LK, Lüert S, Roos C, Knauf S. 2019. Lactation and menstruation shift the vaginal microbiota in captive rhesus monkeys to be more similar to the male urethral microbiota. *Sci Rep* 9: 17399.

4 Hallmaier-Wacker LK, Lüert S, Gronow S, Spröer C, Overmann J, Buller N, Vaughan-Higgins RJ, Knauf S. 2019. A metataxonomic tool to investigate the diversity of Treponema. *Front Microbiol* 10: 2094.

Chapter B1

Disease reservoirs: from conceptual frameworks to applicable criteria

Luisa K. Hallmaier-Wacker[1,2], Vincent J. Munster[3], and Sascha Knauf[1]

Emerging microbes & infections **6**: 1-5 (2017) doi: 10.1038/emi.2017.65

[1] Neglected Tropical Diseases Work Group, Pathology Unit, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[2]Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[3]Virus Ecology Unit, Laboratory of Virology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rocky Mountain Laboratories, Hamilton, USA

Author contributions:

Conceived and designed the study: LHW, SK
Performed literature review: LHW
Manuscript preparation: LHW, VJM, SK

OPEN

Emerging Microbes & Infections (2017) 6, e79; doi:10.1038/emi.2017.65
www.nature.com/emi

REVIEW

Disease reservoirs: from conceptual frameworks to applicable criteria

Luisa K Hallmaier-Wacker[1,2], Vincent J Munster[3] and Sascha Knauf[1]

Central to the One Health approach and any disease eradication program is the question of whether a pathogen has a non-human reservoir. Despite well-established conceptual frameworks that define a reservoir of infection, empirical characterization of reservoirs often remains controversial, challenging and sometimes misleading. What is essentially missing are applicable requirements that standardize the use of the term 'reservoir of infection' across multiple disciplines. We propose an empirical framework, considering maintenance and feasible transmission of a pathogen, to standardize the acceptance of a disease reservoir across multiple disciplines. We demonstrate the intended use of these requirements by applying them to different diseases that are known to infect both humans and animals.
Emerging Microbes & Infections (2017) **6,** e79; doi:10.1038/emi.2017.65; published online 6 September 2017

Keywords: disease eradication; infection; infectious diseases; interdisciplinary; one health; multidisciplinary

A RESERVOIR NEEDS TO MAINTAIN THE PATHOGEN AND HAVE A FEASIBLE TRANSMISSION ROUTE

The high prevalence of infectious agents of zoonotic and anthropozoonotic origin pose a major health threat to both human and animal populations. A conceptual framework for understanding a reservoir of infection has been established through various studies that have emphasized different aspects of zoonotic diseases.[1–4] However, empirical characterization of reservoirs often remains controversial and challenging. The most applicable and accepted way to investigate and define a reservoir emphasizes the annotation of a target group (Figure 1), which is an explicitly defined population of interest in a dynamic and heterogeneous landscape (for example, humans at the livestock–wildlife–human interface).[4,5] According to Haydon *et al.*,[4] the target group is a matter of definition and may therefore be disconnected from the ecological reality. The target group provides a directionality to the study of a reservoir system. All other susceptible populations (non-target populations), which directly or indirectly connect epidemiologically to the target (Figures 1 and 2), can be part of the potential reservoir.[4] For a non-target population to be considered an accepted functional reservoir, maintenance of a single pathogen in the population needs to be shown in combination with a feasible transmission route between the target and non-target populations.[4]

Although the conceptual framework of a disease reservoir is already well-defined, applicable requirements for an evidence-based rejection or acceptance of a reservoir are currently missing. In particular, interdisciplinary standards on genetic and functional similarities of reservoir and human isolates of pathogens are nonexistent. Considering the increase in interdisciplinary research, we see the need to critically discuss and standardize the use of the term 'reservoir of infection' across different research fields to oppose the tendency of published scientific data to exaggerate positive results and hype certain areas of science.[6,7] Although we do not claim absolute standardization of empirical requirements to accept a reservoir across disciplines, we present a framework to serve as a basis for a pending discussion in the growing One Health community. The simplicity and functional orientation of the presented framework allows for straightforward application but does not negate more complex populations, as the same principles can be applied to multi-species systems and metapopulations (Figure 2).

According to the accepted definition of a reservoir proposed by Haydon *et al.*,[4] we discuss the requirements in two parts: the pathogen's maintenance in a potential population or community followed by a discussion on proof of a feasible transmission route. Although the two components are addressed separately, only together they demonstrate the existence of a functional reservoir.

PROOF OF PATHOGEN MAINTENANCE IN A POTENTIAL RESERVOIR

Increases in technological advancements (for example, next-generation sequencing) and vast quantities of available data have not led to concrete applicable criteria when examining the capacity of a pathogen to be maintained in a population. Recognizing both the ethical limitations in regards to animal testing[8] and the advances in the molecular detection of pathogens, we propose the following criteria to demonstrate the maintenance of a pathogen in a population: (i) a high-genetic similarity of the pathogen found in the reservoir system, (ii) a high degree of functional similarity (infectivity and viability), and

[1]Work Group Neglected Tropical Diseases, Pathology Unit, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, Göttingen 37077, Germany; [2]Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, Göttingen 37077, Germany and [3]Virus Ecology Unit, Laboratory of Virology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rocky Mountain Laboratories, Hamilton, MT 59840, USA
Correspondence: S Knauf
E-mail: sknauf@dpz.eu
Received 21 March 2017; revised 20 June 2017; accepted 28 June 2017

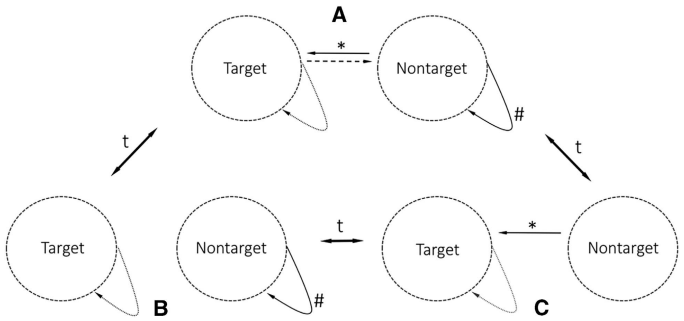


Figure 1 Three scenarios describing the dynamics of a simple reservoir system. (**A**) Pathogen maintenance in the non-target population and feasible transmission route towards the target population. Only this constellation fulfills the requirements of a functional reservoir system. (**B**) Pathogen maintenance in the non-target but no feasible transmission route towards the target population. This is a likely situation whether contact rates between the non-target and target populations are below the threshold. (**C**) No pathogen maintenance in the non-target, but a feasible transmission route exists. An example of the effect of a vaccination strategy in the non-target population. The dynamic of the system is indicated by arrows associated with a 't' (time factor). #Maintenance, *feasible transmission, solid arrows = obligatory, broken line = optional.

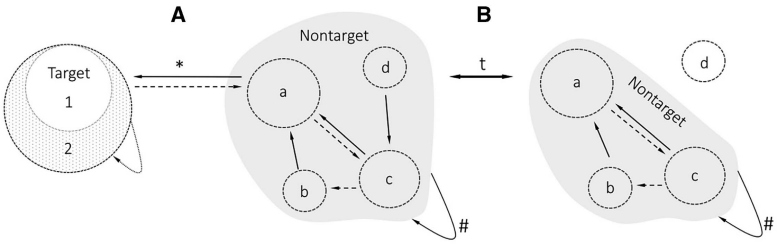


Figure 2 The simplicity and functional orientation of the presented framework allows for straightforward application but does not negate more complex populations. The same principles apply to multi-species systems and metapopulations. The defined target group may be adjusted based on interest and may therefore include metapopulations (targets 1 and 2). The non-target group increases in complexity due to the inclusion of multiple populations (a–d). (**A**) Similarly, to a simple reservoir system, all susceptible populations that connect to the target either (a) directly or (b–d) indirectly are part of the non-target population. (**B**) Temporal shifts in the ecological landscape of the non-target population may lead to the (d) exclusion of populations either due to lack of connectivity or susceptibility. The dynamic of the system is indicated by arrows associated with a 't' (time factor). #Maintenance, *feasible transmission, solid arrows = obligatory, broken line = optional.

(iii) a longitudinal approach that considers the factor of time (Table 1). Owing to the functional orientation of the requirements and for simplicity, all entities involved in the biological lifecycles of a parasite (for example, primary and intermediate hosts) should be considered a single functional unit. Appropriate sequence and functional analysis of a pathogen isolated multiple times from a potential reservoir should be required to prove that a pathogen is maintained in a population. The ability to quickly and cheaply sequence whole genomes has allowed for better genetic resolution.[49,50] Sequence data can be used to examine similarity in the pathogen between a potential reservoir and a target. However, mutation rates vary significantly between pathogens[51,52] and the threshold for sequence and functional similarity must be individually defined and accepted by the scientific community. A single-nucleotide difference can potentially result in a loss of infectivity, for example, when important invasion mechanisms are affected (receptor affinity). In bacteria, investigations can be further complicated by plasmids that can be exchanged and mutated over time.[53] A high amount of phylogenetic relatedness of pathogens isolated from the non-target and target populations does not provide sufficient evidence for the involvement of a pathogen and its ability to infect both groups. Importantly, DNA-based analyses only provide information on the functional potential of a pathogen and must not reflect the gene-expression within a host.[54] For example, the bacterium *Treponema paraluiscuniculi* (which causes syphilis in rabbits), is over 99% identical on the basis of the whole genome to the human pathogen *T. pallidum* (which causes human treponematosis), but does not infect humans.[55] As phylogenetic information fails to reflect the downstream effects of mutations, proof that a pathogen can proliferate in the potential reservoir is required.[56] Information on the transcriptome and proteome of bacteria or the phenotype of viruses are necessary to see the effect of mutations on pathogen viability.[57] There are different ways to test for the functional ability of a pathogen in different species. Owing to the ethical concerns, cell and tissue assays have been increasingly used in therapeutic research instead of animal models.[8] Although these assays are limited in their conclusiveness, they can provide important insight into the molecular mechanisms involved. For example, the failure to infect primary tissue culture from rhesus macaques with human immunodeficiency virus 1 (HIV-1) demonstrates that non-human primates were unlikely to act as a maintenance population (Table 1).[58] In some instances, for example, with uncultivable bacteria such as *Treponema pallidum*, it may be necessary to use animal models to examine the functionality of a pathogen within a potential reservoir species. Knowledge of the biology of the pathogen is essential to properly define a sequence and functional similarity threshold for a particular reservoir system.

Table 1 Applicable requirements that need to be fulfilled for the acceptance of a disease reservoir and their exemplary use in selected diseases that are known to infect humans and animals

Pathogen	Target	Non-target	Maintenance in NT			Feasible transmission route			Time factor	Refs.
			Main transmission route	High-genetic similarity	Functional similarity	Spatial and temporal connectivity	Pathogen involvement	Maintaining pathogen viability	Example of longitudinal study	
Influenza A virus (H1N1)	Human	Swine	Aerosol	X	X	X	X	X	9	10–12
MERS-Coronavirus	Human	Camel	Direct contact	X	X	X	X	(X)	13	14–16
Brucella melitensis (localized brucellosis)	Human	Sheep	Food-borne	X	X	X	X	X	17	18,19
Immunodeficiency virus	Human	NHP	Direct contact	(X)	NP	X	NP	NP	N/A	20
Treponema pallidum pertenue (yaws)	Human	NHP	Direct contact/vector	X	(X)	X	(X)	NP	N/A	21–23
Mycobacterium bovis (bovine tuberculosis)	Human	Cattle	Food-borne/aerosol	X	X	X	X	X	24	25,26
Rabies virus	Human	Fox	Bite	X	X	X	X	X	27	28–30
Echinococcus multilocularis (alveolar echinococcosis)	Human	Fox	Oral/fecal	X	X	X	X	X	31	32,33
Hantavirus	Human	Rodent	Aerosol	X	X	X	X	X	34	35
Ebola virus	Human	Bats	Contact/aerosol	X	NP	X	(X)	NP	N/A	36–38
Zika virus	Human	NHP	Vector	(X)	NP	X	X	NP	39	40–42
Borrelia burgdorferi (borreliosis)	Human	Wildlife	Vector	X	X	X	X	X	43	44–46
Yellow fever virus	Human	NHP	Vector	X	X	X	X	X	47	48

Abbreviations: not available, N/A; non-human primate, NHP; non-target, NT; not provided/no current evidence, NP; evidence, X; partial evidence, (X).
Classical reservoir systems fulfill all requirements proposed in this study.

When examining pathogen maintenance, a longitudinal approach is required to consider the dynamics of a potential reservoir system, including the influence of genetic variation in any given population. Defining a population that was infected at a single time point as a maintenance population for a pathogen is based on assumptions and is therefore speculative. Sero-prevalence surveys are an attractive way to detect the presence of pathogens in a population, as it indicates that an immunocompetent subject was in contact with the pathogen.[1] However, only longitudinal studies with adequate sampling regimes (multiple sampling) to test for antibodies against a pathogen can provide information on the timing or frequency of infection, both of which are important for reservoir studies.[59] Furthermore, cross-reactivity and erroneous assays can lead to false-positive results. For more diffuse reservoir systems, including multi-species compositions where the diversity of host susceptibility (at the individual, species or population level) protects against widespread infection (dilution effect),[60] a longer time frame must be applied. This guarantees a more accurate understanding of the maintenance within a population (for example, Ebola[36]).

PROOF OF FEASIBLE TRANSMISSION ROUTE

Maintenance of a pathogen in a population alone does not provide sufficient proof that a functional reservoir exists. A connection between the target and the non-target populations must be established;[4] otherwise the non-target population remains a maintenance population with the potential to be a reservoir. Therefore, the determination of a feasible and somewhat permanent transmission route between the non-target and target populations is key to identifying a reservoir system (Figure 1). For multi-species reservoir systems, the transmission route between the target and non-target populations may be indirect (Figure 2, connection between b and target), possibly incorporating different hierarchical levels of a non-target community.[4,61] The type of transmission route dictates the form of evidence needed to prove that a feasible transmission route exists between the reservoir and target. For simplicity, we define vectors as part of the transmission route, although under certain circumstances (for example, permanency or substantial amplification in the vector), they may act as part of the non-target community.[61] Four basic requirements need to be met to make a compelling argument for the existence of a feasible transmission route: (i) spatial (direct or indirect) and temporal connectivity between the reservoir system and the target population, (ii) pathogen involvement in this feasible transmission route, (iii) proof of viability of the pathogen during the proposed transmission route and (iv) a longitudinal approach that requires the isolation of a pathogen multiple times in a given transmission route (Table 1).

To prove the feasibility of a transmission route, direct or indirect spatial connectivity as well as temporal overlap between the non-target and target populations must be present. Connectivity measurements depend on the type of transmission route; for example, direct contact transmission requires overlapping territory. Computational tools can help determine the necessary overlap in a population by modeling the transmission across an affected population.[62] In addition to spatial and temporal overlap, the involvement of the pathogen in the particular transmission route needs to be shown, which again requires long-term field projects. In the case of Lyme disease caused by *Borrelia burgdorferi*, nucleic acids from the bacterium were detected in ticks using PCR.[44] However, the detection of DNA does not directly prove that transmission occurs. To gain further confidence that the transmission is feasible, it is therefore essential to show that the infectious organism remains viable during the proposed transmission

route.[45] This means that in addition to PCR detection, the viable pathogen needs to be isolated during a transmission event, where the measure of viability depends on the type of pathogen. In airborne transmission, for example, environmental factors such as size of droplets, UV light and humidity can greatly influence the transmissibility of a virus (as reviewed in Tang[63]). If the amount of viable and therefore infectious organisms is below the infectious dose, the particular transmission route is unfeasible. Without a feasible transmission route between target and non-target populations, no functional reservoir exists. Furthermore, to include all parts of a reservoir population, long-term investigations must focus on the transmission between the non-target and target groups as well as feasible transmission within the non-target community.[61] Unconnected maintenance host populations may become a future reservoir through temporal shifts of the ecosystem.

CHALLENGES OF IMPLEMENTATION

Biological systems are dynamic and can change over time (Figure 1). Single transmission events do not confirm a reservoir of infection (for example, HIV,[20] Table 1). It is therefore important to show continuity and persistence in both maintenance and transmission, which can only be achieved through multiple and adequately timed (field) investigations. Well-designed intervention studies can be used as quasi-experiments to study a reservoir of infection but should not be used as a stand-alone test for the existence of a reservoir.[1] Despite sufficient planning, the cause and effect of intervention studies are often difficult to determine[1,64] and the removal of a pathogen from a particular ecosystem may cause unanticipated effects. A negative outcome does not necessary indicate the lack of a reservoir or transmission route.[64,65] Instead, it can show that the intervention may have been incomplete or that the complexity of a reservoir is not entirely understood.

Pathogens must be studied in the context of natural ecosystems. The complexity of reservoir systems increases as multiple non-target populations interact as an ecological entity, which is influenced by factors such as competition, co-existence or predation.[66] Furthermore, the artificial environment in a laboratory, which is often used to study the susceptibility of a species, differs substantially from a natural setting.[67] The use of laboratory animals or cell- and tissue-based assays can be advantageous when studying pathogenicity, but it cannot solely contribute to the understanding of the epidemiology of a pathogen, which is largely impacted by variables such as genetic diversity, co-infection, cross-protective immunity and spatial connectivity. As a consequence, any epidemiological model requires additional information on the geographic range and the ecological landscape.[68] This includes population densities and functional profiles of species that are involved in the reservoir system.[60,69] The importance of sample size in field studies and animal experiments cannot be stressed enough as it greatly affects the efficacy of analysis, especially in reservoirs with low-frequency crossover events.

Neither laboratory experiments, nor intervention studies, nor epidemiological models alone can provide a full understanding of a natural reservoir of infection. Only the combination of methods that are based on established and validated species-specific assays and technically sound field investigations can provide confidence that the pathogen is maintained in a non-target population and that a feasible transmission route exists. This, however, requires the political will and financial support to conduct long-term One Health studies to explore diseases in their natural context.

CONCLUSION

The term 'disease reservoir' should be used carefully and only if there is convincing evidence demonstrating the maintenance and a feasible transmission route of a particular pathogen (Figure 1). We propose overarching requirements that must be fulfilled to provide ample proof that a reservoir exists (Table 1). Classical reservoir systems (for example, Lyme disease caused by *Borrelia burgdorferi*) fulfill all of the requirements proposed in this study, whereas some well-known diseases, such as Ebola, need further research until a reservoir system can be accepted (Table 1). For the pathogens without an accepted reservoir, the framework introduced in this study also indicates the outstanding questions that future research should focus on to investigate the presence of a reservoir system. A broader expert-based multidisciplinary discussion is needed to develop standards for the diversity of pathogens.

ACKNOWLEDGEMENTS

VJM is supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH). SK received funding from the German Research Foundation (DFG): KN1097/3-1 and KN1097/4-1. We would like to thank the reviewer for addressing important aspects that clearly improved the proposed framework.

1 Viana M, Mancy R, Biek R *et al*. Assembling evidence for identifying reservoirs of infection. *Trends Ecol Evol* 2014; **29**: 270–279.
2 Drexler JF, Corman VM, Müller MA *et al*. Bats host major mammalian paramyxoviruses. *Nat Commun* 2012; **3**: 796.
3 Ashford R. When is a reservoir not a reservoir?. *Emerg Infect Dis* 2003; **9**: 1495–1496.
4 Haydon DT, Cleaveland S, Taylor LH *et al*. Identifying reservoirs of infection: a conceptual and practical challenge. *Emerg Infect Dis* 2002; **8**: 1468–1473.
5 Dunning JB, Stewart DJ, Danielson BJ *et al*. Spatially explicit population models: current forms and future uses. *Ecol Appl* 1995; **5**: 3–11.
6 Caulfield T, Sipp D, Murry CE *et al*. Confronting stem cell hype. *Science* 2016; **352**: 776–777.
7 Vinkers CH, Tijdink JK, Otte WM. Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis. *BMJ* 2015; **351**: h6467–h6777.
8 Ferdowsian HR, Beck N. Ethical and scientific considerations regarding animal testing and research. *PLoS One* 2011; **6**: e24059.
9 Simon G, Larsen LE, Dürrwald R *et al*. European surveillance network for influenza in pigs: surveillance programs, diagnostic tools and Swine influenza virus subtypes identified in 14 European countries from 2010 to 2013. *PLoS One* 2014; **9**: e115815.
10 Brookes SM, Núñez A, Choudhury B *et al*. Replication, pathogenesis and transmission of pandemic (H1N1) 2009 virus in non-immune pigs. *PLoS One* 2010; **5**: e9068.
11 Steel J, Staeheli P, Mubareka S *et al*. Transmission of pandemic H1N1 influenza virus and impact of prior exposure to seasonal strains or interferon treatment. *J Virol* 2010; **84**: 21–26.
12 Maines TR, Jayaraman A, Belser JA *et al*. Transmission and pathogenesis of swine-origin 2009A (H1N1) influenza viruses in ferrets and mice. *Science* 2009; **325**: 484–487.
13 Müller MA, Meyer B, Corman VM *et al*. Presence of Middle East respiratory syndrome coronavirus antibodies in Saudi Arabia: a nationwide, cross-sectional, serological study. *Lancet Infect Dis* 2015; **15**: 559–564.
14 Mohd HA, Al-Tawfiq JA, Memish ZA. Middle East respiratory syndrome coronavirus (MERS-CoV) origin and animal reservoir. *Virol J* 2016; **13**: 87.
15 Van Doremalen N, Bushmaker T, Munster V. Stability of Middle East respiratory syndrome coronavirus (MERS-CoV) under different environmental conditions. *Euro Surveill* 2013; **18**: 20590.
16 Haagmans BL, van den Brand JM, Raj VS *et al*. An orthopoxvirus-based vaccine reduces virus excretion after MERS-CoV infection in dromedary camels. *Science* 2015; **351**: 77–81.
17 Dean AS, Crump L, Greter H *et al*. Global burden of human brucellosis: a systematic review of disease frequency. *PLoS Negl Trop Dis* 2012; **6**: e1865.
18 de Figueiredo P, Ficht TA, Rice-Ficht A *et al*. Pathogenesis and immunobiology of brucellosis: review of Brucella–Host Interactions. *Am J Pathol* 2015; **185**: 1505–1517.
19 Paixão TA, Roux CM, Den Hartigh AB *et al*. Establishment of systemic *Brucella melitensis* infection through the digestive tract requires urease, the type IV secretion system, and lipopolysaccharide O antigen. *Infect Immun* 2009; **77**: 4197–4208.
20 Gao F, Bailes E, Robertson DL *et al*. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* 1999; **397**: 436–441.

21 Knauf S, Dahlmann F, Batamuzi EK *et al*. Validation of serological tests for the detection of antibodies against *Treponema pallidum* in nonhuman primates. *PLoS Negl Trop Dis* 2015; **9**: e0003637.
22 Knauf S, Raphael J, Mitjà O *et al*. Isolation of *Treponema* DNA from necrophagous flies in a natural ecosystem. *EBioMedicine* 2016; **11**: 85–90.
23 Harper KN, Knauf S. *Treponema pallidum* infection in primates: clinical manifestations, epidemiology, and evolution of a stealthy pathogen. In: Brinkworth JF, Pechenkina K (eds). *Primates, Pathogens, and Evolution*. New York, NY: Springer, 2013: 189–219.
24 Walker TM, Lalor MK, Broda A *et al*. Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome sequences: an observational study. *Lancet Respir Med* 2014; **2**: 285–292.
25 Fitzgerald S, Kaneene J. Wildlife reservoirs of bovine tuberculosis worldwide: hosts, pathology, surveillance, and control. *Vet Pathol* 2013; **50**: 488–499.
26 Michel AL, Müller B, van Helden PD. *Mycobacterium bovis* at the animal-human interface: a problem, or not? *Vet Microbiol* 2010; **140**: 371–381.
27 Sidwa TJ, Wilson PJ, Moore GM *et al*. Evaluation of oral rabies vaccination programs for control of rabies epizootics in coyotes and gray foxes: 1995–2003. *J Am Vet Med Assoc* 2005; **227**: 785–792.
28 Rupprecht CE, Hanlon CA, Hemachudha T. Rabies re-examined. *Lancet Infect Dis* 2002; **2**: 327–343.
29 Hemachudha T, Laothamatas J, Rupprecht CE. Human rabies: a disease of complex neuropathogenetic mechanisms and diagnostic challenges. *Lancet Neurol* 2002; **1**: 101–109.
30 Chautan M, Pontier D, Artois M. Role of rabies in recent demographic changes in red fox (*Vulpes vulpes*) populations in Europe. *Mammalia* 2000; **64**: 391–410.
31 Schweiger A, Ammann RW, Candinas D *et al*. Human alveolar echinococcosis after fox population increase, Switzerland. *Emerg Infect Dis* 2007; **13**: 878–882.
32 Veit P, Bilger B, Schad V *et al*. Influence of environmental factors on the infectivity of *Echinococcus multilocularis* eggs. *Parasitology* 1995; **110**: 79–86.
33 Craig P. *Echinococcus multilocularis*. *Curr Opin Infect Dis* 2003; **16**: 437–444.
34 Kuenzi AJ, Douglass RJ, Bond CW *et al*. Long-term dynamics of Sin Nombre viral RNA and antibody in deer mice in Montana. *J Wildl Dis* 2005; **41**: 473–481.
35 Bi Z, Formenty PB, Roth CE. Hantavirus infection: a review and global update. *J Infect Dev Ctries* 2008; **2**: 3–23.
36 Leroy EM, Kumulungui B, Pourrut X *et al*. Fruit bats as reservoirs of Ebola virus. *Nature* 2005; **438**: 575–576.
37 Pourrut X, Kumulungui B, Wittmann T *et al*. The natural history of Ebola virus in Africa. *Microb Infect* 2005; **7**: 1005–1014.
38 Weingartl HM, Embury-Hyatt C, Nfon C *et al*. Transmission of Ebola virus from pigs to non-human primates. *Sci Rep* 2012; **2**: 811.
39 Roth A, Mercier A, Lepers C *et al*. Concurrent outbreaks of dengue, chikungunya and Zika virus infections-an unprecedented epidemic wave of mosquito-borne viruses in the Pacific 2012-2014. *Euro Surveill* 2014; **19**: 20929.
40 Favoretto S, Araujo D, Oliveira D *et al*. First detection of Zika virus in neotropical primates in Brazil: a possible new reservoir. *BioRxiv* 2016, doi: http://www.biorxiv.org/content/early/2016/04/20/049395.
41 Dick G. Zika virus (II). Pathogenicity and physical properties. *Trans R Soc Trop Med Hyg* 1952; **46**: 521–534.
42 Marano G, Pupella S, Vaglio S *et al*. Zika virus and the never-ending story of emerging pathogens and transfusion medicine. *Blood Transfus* 2016; **14**: 95–100.
43 Ogden NH, Bouchard C, Kurtenbach K *et al*. Active and passive surveillance and phylogenetic analysis of *Borrelia burgdorferi* elucidate the process of Lyme disease risk emergence in Canada. *Environ Health Perspect* 2010; **118**: 909–914.
44 Barbour AG, Maupin GO, Teltow GJ *et al*. Identification of an uncultivable *Borrelia* species in the hard tick *Amblyomma americanum*: possible agent of a Lyme disease-like illness. *J Infect Dis* 1996; **173**: 403–409.
45 Moriarty TJ, Norman MU, Colarusso P *et al*. Real-time high resolution 3D imaging of the lyme disease spirochete adhering to and escaping from the vasculature of a living host. *PLoS Pathog* 2008; **4**: e1000090.
46 Tilly K, Rosa PA, Stewart PE. Biology of infection with *Borrelia burgdorferi*. *Infect Dis Clin North Am* 2008; **22**: 217–234.
47 Almeida MA, Cardoso JDC, dos Santos E *et al*. Surveillance for yellow fever virus in non-human primates in Southern Brazil, 2001–2011: a tool for prioritizing human populations for vaccination. *PLoS Negl Trop Dis* 2014; **8**: e2741.
48 Monath TP. Yellow fever: an update. *Lancet Infect Dis* 2001; **1**: 11–20.
49 Roetzer A, Diel R, Kohl TA *et al*. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med* 2013; **10**: e1001387.
50 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* 1977; **74**: 5463–5467.
51 Dapp MJ, Heineman RH, Mansky LM. Interrelationship between HIV-1 fitness and mutation rate. *J Mol Biol* 2013; **425**: 41–53.
52 Taylor LH, Latham SM, Mark E. Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci* 2001; **356**: 983–989.
53 Okinaka R, Cloud K, Hampton O *et al*. Sequence, assembly and analysis of pX01 and pX02. *J Appl Microbiol* 1999; **87**: 261–262.
54 Miosge LA, Field MA, Sontani Y *et al*. Comparison of predicted and actual consequences of missense mutations. *Proc Natl Acad Sci* 2015; **112**: 5189–5198.
55 Šmajs D, Zobaníková M, Strouhal M *et al*. Complete genome sequence of *Treponema paraluiscuniculi*, strain Cuniculi A: the loss of infectivity to humans is associated with genome decay. *PLoS One* 2011; **6**: e20415.
56 Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 2008; **92**: 255–264.
57 Prakash T, Taylor TD. Functional assignment of metagenomic data: challenges and applications. *Brief Bioinform* 2012; **13**: 711–727.
58 Hofmann W, Schubert D, LaBonte J *et al*. Species-specific, postentry barriers to primate immunodeficiency virus infection. *J Virol* 1999; **73**: 10020–10028.
59 Hoye BJ, Munster VJ, Nishiura H *et al*. Reconstructing an annual cycle of interaction: natural infection and antibody dynamics to avian influenza along a migratory flyway. *Oikos* 2011; **120**: 748–755.
60 Roche B, Rohani P, Dobson AP *et al*. The impact of community organization on vector-borne pathogens. *Am Nat* 2012; **181**: 1–11.
61 Zeppelini CG, de Almeida AMP, Cordeiro-Estrela P. Zoonoses as ecological entities: a case review of plague. *PLoS Negl Trop Dis* 2016; **10**: e0004949.
62 Kao RR, Haydon DT, Lycett SJ *et al*. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. *Trends Microbiol* 2014; **22**: 282–291.
63 Tang JW. The effect of environmental parameters on the survival of airborne infectious agents. *J R Soc Interface* 2009; **6**: 737–746.
64 Dietze R, Barros GB, Teixeira L *et al*. Effect of eliminating seropositive canines on the transmission of visceral leishmaniasis in Brazil. *Clin Infect Dis* 1997; **25**: 1240–1242.
65 Ezenwa VO, Jolles AE. Opposite effects of anthelmintic treatment on microbial infection at individual versus population scales. *Science* 2015; **347**: 175–177.
66 Menge BA, Sutherland JP. Species diversity gradients: synthesis of the roles of predation, competition, and temporal heterogeneity. *Am Nat* 1976; **110**: 351–369.
67 McCafferty J, Mühlbauer M, Gharaibeh RZ *et al*. Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model. *ISME J* 2013; **7**: 2116–2125.
68 Lindenfors P, Nunn CL, Jones KE *et al*. Parasite species richness in carnivores: effects of host body mass, latitude, geographical range and population density. *Glob Ecol Biogeogr* 2007; **16**: 496–509.
69 Giardina AR, Schmidt KA, Schauber EM *et al*. Modeling the role of songbirds and rodents in the ecology of Lyme disease. *Can J Zool* 2000; **78**: 2184–2197.

Chapter B2

The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis

Luisa K. Hallmaier-Wacker[1,2], Simone Lüert[1,2], Christian Roos[2] and Sascha Knauf[1]

Scientific Reports **8**: 6292 (2018) doi: 10.1038/s41598-018-24573-y

[1] Neglected Tropical Diseases Work Group, Infection Biology Unit, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

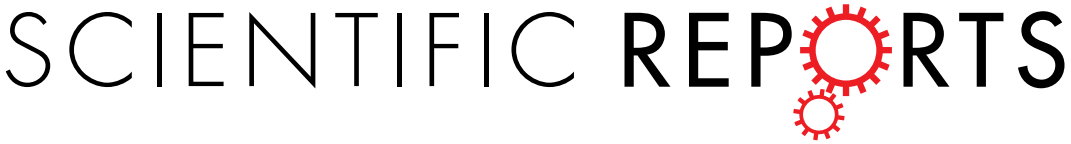
[2]Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

---

Author contributions:

Conceived and designed the study: LHW, CR, SK
Performed the laboratory work: LHW, SL
Analyzed the data: LHW, SK
Manuscript preparation: LHW, SL, CR, SK

www.nature.com/scientificreports



OPEN

The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis

Received: 30 November 2017
Accepted: 6 April 2018
Published online: 19 April 2018

Luisa K. Hallmaier-Wacker[1,2], Simone Lueert[1,2], Christian Roos[2] & Sascha Knauf[1]

Next-generation sequencing approaches used to characterize microbial communities are subject to technical caveats that can lead to major distortion of acquired data. Determining the optimal sample handling protocol is essential to minimize the bias for different sample types. Using a mock community composed of 22 bacterial strains of even concentration, we studied a combination of handling conditions to determine the optimal conditions for swab material. Examining a combination of effects simulates the reality of handling environmental samples and may thus provide a better foundation for the standardization of protocols. We found that the choice of storage buffer and extraction kit affects the detected bacterial composition, while different 16S rRNA amplification methods only had a minor effect. All bacterial genera present in the mock community were identified with minimal levels of contamination independent of the choice of sample processing. Despite this, the observed bacterial profile for all tested conditions were significantly different from the expected abundance. This highlights the need for proper validation and standardization for each sample type using a mock community and blank control samples, to assess the bias in the protocol and reduce variation across the datasets.

Microorganisms colonize various anatomical sites and play a crucial role in the balance of health and disease. The vaginal microbiome is known to maintain the health of women and thereby prevents urogenital diseases[1]. The advent of cultivation-independent molecular approaches, such as 16S rRNA amplicon sequencing, has allowed for a better understanding of the microbes that inhabit different biological niches. However, these powerful tools are not without important technical caveats that can lead to a distortion in the acquired data[2]. Such limitations have been well documented, and include sample collection, storage buffer, DNA extraction, amplification primers and methods, sequencing technology, and analysis techniques[3,4]. While it is impossible to negate all of these influences, it is important to understand the bias inherent in the analysis. Studies focusing on one or two technical limitations have made recommendations for improving the bias such as reducing the number of PCR cycles[5] or adding additional lysis pre-treatment[6].

DNA extraction, a critical step in culture-independent bacterial profiling, has been identified as a key driver of technical variation[3]. Most common studies on the microbiome of swab material use commercially available DNA extraction kits that vary in their lysis approach from mechanical to enzymatic treatment. Various studies have focused on technical variations in extraction kits, yet a field-wide consensus on sample extraction has not been reached[3,6–9]. Due to the large variety of microbiota and sample types, a single standard for all sample types is unlikely to be achieved. Despite the knowledge that the choice of extraction kit can have a significant effect on the results, there is often a lack of proper validation across sample types[3].

Similar to DNA extraction kits, the choice of sample storage buffer has been shown to influence the detected bacterial community[10–12]. The ideal storage choice largely depends on the available resources during sampling such as the availability of freezing conditions[11]. Selecting the optimal storage buffers is dependent upon its compatibility with all downstream analyses including the extraction method. Many studies, however, only focus on the effect of a single technical variation instead of examining the effect of different combination of storage buffer, DNA extraction kit, and amplification methods[2]. Studying a combination of effects mirrors the reality of sample handling more closely and may thus provide a better foundation for the standardization of sampling handling protocols prior to microbial analysis.

[1]Work Group Neglected Tropical Diseases, Infection Biology Unit, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077, Göttingen, Germany. [2]Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077, Göttingen, Germany. Correspondence and requests for materials should be addressed to S.K. (email: sknauf@dpz.eu)

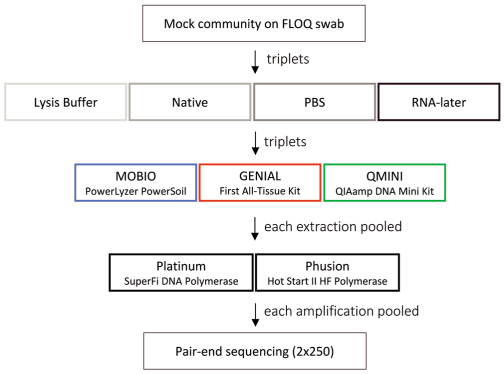


Figure 1. Outline of experimental design. A schematic showing the different treatment variables.

Extraction Method	Abbreviation	Lot #	Lysis type	Elution Volume
MOBIO PowerLyzer PowerSoil Kit	MOBIO	PL16C30	Mechanical, Column-based	100
GEN-IAL First All-Tissue Kit	GENIAL	0091.01	Enzymatic, Phenol-Chloroform	20
QIAamp DNA Mini Kit	QMINI	154035749	Enzymatic, Column-based	70

Table 1. Commercial extraction kits used in this study.

In this study, we used a mock community, composed of an even concentration of cells from 22 bacterial strains (19 genera), to assess the effect of storage buffers, extraction kits, and amplification methods (Fig. 1). Using a mock community to examine the effect of different sample handling conditions rather than environmental samples of unknown microbe composition is essential to be able to systematically compare the effects[3]. In addition to the use of a mock community, a blank control was included in all sample procedures to monitor any buffer, kit, or reagent specific contamination[13]. The aim of this study was to evaluate the performance of combinations of handling conditions commonly used in microbiome studies and to contribute to the ongoing debate on standardization in microbiome research.

Methods

Preparation of swab mock community samples. A cell mixture of 22 different bacterial strains at a concentration of 1×10^8 cells/mL of each organism (Microbial mock community, HM-280) in phosphate buffer saline (PBS) was obtained through Biodefense and Emerging Infectious Research (BEI) Resources, NIAID, NIH as part of the Human Microbiome Project (Manassas, USA; Supplementary Table S1). To simulate physiological conditions, 10 µl of mock community containing 1×10^6 cells/mL of each organism was added to a flocked swab (FLOQSwabs, Copan Improve Diagnostics, Brescia, Italy) and then placed in 500 µl of the respective storage buffer (Fig. 1). Four different storage buffers were used; PBS (PAN-Biotech GmbH, Aidenbach, Germany), a custom-made lysis buffer (10 mM Tris, pH 8.0, 0.1 M EDTA, pH 8.0 and 0.5% SDS), RNA-later (Thermo Fisher Scientific Inc., Waltham, MA, USA), and no buffer (native). A blank control swab sample was placed in each storage buffer without additive. All swab samples were frozen at −80 °C for one week prior to DNA extraction. Suitable precautions were taken during sample handling and processing to insure sterility during all procedures.

DNA Extraction methods. Three commercially available DNA extraction kits were used in this study to extract bacterial DNA from swab material stored in four different storage buffers (Table 1). Extraction was performed in triplets and the extracted DNA from each buffer was subsequently pooled prior to 16 S rRNA gene amplification. Processing of swab samples prior to DNA extraction is illustrated in Supplementary Fig. S1.

QIAamp DNA Mini Kit (QMINI). Samples were extracted using the QIAamp Mini Kit (Qiagen GmbH, Hilden, Germany) according to the standard protocol with minor modifications. Briefly, proteinase K (20 mg/µl) was added and the samples were incubated for 50 minutes at 56 °C. Then, AL buffer (Qiagen GmbH) and ethanol were added in the appropriate amount. The DNA from the lysate was subsequently purified using the spin columns provided by the manufacturer and eluted in 70 µl AVE buffer (Qiagen GmbH).

MOBIO PowerLyzer PowerSoil Kit (MOBIO). A maximum of 750 µl of swab lysate was added to the 0.1 mm PowerLyzer® Glass Bead Tube (Qiagen GmbH). DNA extraction was continued from step 2 as described in the MOBIO PowerLyzer PowerSoil Kit protocol (Qiagen GmbH). The DNA was eluted in a final volume of 100 µl of Solution C6 provided in the kit.

GEN-IAL First All-Tissue Kit (GENIAL). The first All-Tissue Kit (GEN-IAL, Troisdorf, Germany) was applied according to the manufacturer's protocol with minor modifications. Briefly, 5 µl proteinase K and 5 µl dithiothreitol (DTT) was added to the lysate and incubated at 65 °C for 60 min at 600 rpm in a thermomixer (Eppendorf, Hamburg, Germany). The lysate was purified according to the standard protocol and the DNA pellet was resuspended in 20 µl of C6 buffer (Qiagen GmbH).

16S rRNA gene amplification. For each pooled extraction, the V4 region of the 16 S ribosomal RNA (16 S rRNA) gene was amplified in triplets using the universal primers 515 F and 806 R adapted with linker regions and barcoded sequences used for dual-indexing[14]. Platinum SuperFi DNA Polymerase (Thermo Fisher Scientific) and the Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific) were both tested for amplification. Each PCR reaction consisted of 12.5 µl of 2x PCR master mix, 6 µl of Microbial DNA-Free water (Qiagen GmbH), 1.25 µl of each primer (0.5 mM each, Metabion, Martinsried, Germany) and 4 µl of template in a total reaction volume of 25 µl. PCR cycling conditions comprised of a pre-denaturation step of 30 s at 98 °C, followed by 30 cycles of 98 °C for 10 s, 55 °C for 15 s and 72 °C for 60 s, and a final 10 min extension step at 72 °C. For a selection of four samples, five additional cycles were added to the amplification procedure to examine if additional cycles may be favorable for samples with low concentrations. The amplicon triplets were pooled, purified using 0.7x AMPure XP beads (Beckman Coulter, Brea, USA) and quantified using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific). Amplicon integrity was verified for a representative number of 11 samples using a BioAnalyzer 2000 (Agilent, Palo Alto, USA) prior to pooling equimolar amounts (10 nM) of each amplicon for sequencing. For the blank samples, the maximum volume (5 µl) of sample was added to the library, as the concentrations prior to sequencing were below 10 nM. Illumina MiSeq. 2 × 250 bp paired-end sequencing (Illumina V2 chemistry) was performed in the Transcriptome and Genome Analysis Laboratory at the University of Göttingen[14]. All generated read files analyzed in this study were uploaded to the NCBI Sequence Read Archive (SRA) (SRP125723).

Mock community data processing and analysis. The sequencing reads were processed using the mothur software package (v.1.36.1)[15]. According to the MiSeq SOP[14], contigs were assembled, sequences trimmed, identical sequences merged, and chimeras removed (UCHIME[16]). Subsequently, sequences were aligned to the SILVA bacterial reference database[17]. Non-bacterial sequences, cross-sample singletons, and poorly aligned sequences were removed. The seq.error command was run for each mock sample in mothur and subsequently averaged to determine the error rate of the run. Due to low read numbers, blank control sample reads (control swabs containing no mock community) were removed from the dataset and analyzed separately. As subsampling is currently still an accepted method of normalization in microbial ecology[18], the reads of the remaining mock community samples were rarefied to 95,870 sequences/sample. A separate file with the theoretical sequence composition (actual) of the 22 bacterial strains of mock community was created and adjusted for the 16 S rRNA copy number (Supplementary Table S1) and normalized to the sequence count of the run (95,870 reads)[19]. After merging the actual (theoretical) mock community composition with the practically obtained sequences, the merged file was classified using the Bayesian classifier implemented in mothur[20]. Operational taxonomic units (OTUs) were assigned based on 97% sequence similarity and subsequently the alpha and beta diversity was analyzed. For alpha diversity, the richness (OTUs observed and Chao1) and community diversity (Inverse Simpson Metrix) was analyzed using the summary.single command in mothur. Additionally, the percentage of contaminant OTUs (OTUs that do not cluster to the theoretical mock community) was examined. Beta diversity was analyzed using Bray-Curtis dissimilarity index[21]. The dissimilarity matrix was visualized using nonmetric multidimensional scaling (NMDS) plots and Newick formatted dendrograms (visualized in FigTree v.1.4.2, http://tree.bio.ed.ac.uk/software/figtree/).

Statistical comparison of sequence data. To evaluate and compare the type of extraction and amplification method used, the values of the alpha or beta diversity measurement were pooled for each variable (e.g. the buffer type). The statistical significance of the pooled data was analyzed in GraphPad Prism 6 (GraphPad software, La Jolla, CA, USA). In case of normal distribution (Kolmogorov-Smirnov normality test), the parametric paired two-tailed students t-test was used for comparison. In all other cases the non-parametric Wilcoxon matched-pairs signed rank test was used. For multiple comparisons, a one-way ANOVA with Bonferroni's multiple comparisons test was applied. Differences in community structure between storage buffers and extraction methods were tested using analysis of molecular variance (AMOVA) in mothur[22]. Non-metric multidimensional scaling (NMDS) plot of Bray-Curtis dissimilarities and UPGMA-clustered dendrograms (Bray-Curtis) were used to visualize data points. Parsimony (mothur) hypothesis testing was performed to test whether the differential clustering of the PBS samples in the dendrograms was significant[23]. Differences in the 30 most abundant OTUs were assessed using the metastats command in mothur[24] and p-values for differences in individual OTUs were corrected for multiple comparisons using Bonferroni correction. Values of $p < 0.05$ were considered statistically significant.

Results

The pooled library (n = 28 mock samples, n = 36 blank/control samples) produced 12,968,125 16 S rRNA sequence reads, of which 9,920,805 reads were retained after quality control (77%). A total of 8,974,393 sequences, with a mean read count of 249,288 reads per sample, were retained after the sequences corresponding to the blank control samples were removed. After rarefying to 95,870 sequences per sample, *de novo* OTU picking returned 228 OTUs, of which 19 OTUs corresponding to the mock community make up more than 99% of the pooled community. The average error rate of the run was found to be 0.040% (±0.004).

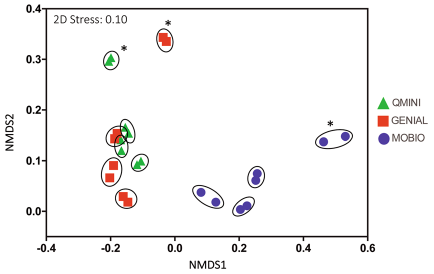


Figure 2. Clustering of samples amplified with two different polymerases on a non-metric multidimensional scaling (NMDS) plot of the Bray-Curtis dissimilarities. Points are colored by applied extraction kit. The encircled pairs correspond to a single sample where each data point represents one 16 S rRNA amplification with Phusion Hot Start II High-Fidelity DNA Polymerase and the another with Platinum SuperFi DNA Polymerase. Sample pairs labeled with * were stored in PBS.

Effect of different amplification method. The choice of polymerase (Platinum SuperFi DNA polymerase vs. Phusion Hot Start II High-Fidelity DNA polymerase) was not found to significantly change the number of observed OTUs ($p = 0.08$ [paired t-test] or Inverse Simpson index, $p = 0.48$, [paired t-test]). Furthermore, pairwise comparison of the Bray-Curtis dissimilarity between the two polymerases yielded only small variations (maximum difference 0.076, Supplementary Table S2) indicating near identical bacterial community profile for a single sample (Fig. 2). Since the results indicate that these two applied high-fidelity polymerases do not significantly impact the observed microbial diversity, we pooled the data from the two polymerases for identical sample for the analyses of buffer and extraction kit choice. The addition of five cycles in 16 S rRNA gene amplification shows only a minor impact on the detected bacterial composition when tested on MOBIO extractions (Supplementary Fig. S2a). There was, however, a significant increase of the number of OTUs detected with additional cycles ($p = 0.029$, Supplementary Fig. S2b), indicating that lower cycle numbers are favorable.

Effect of storage buffer. The effect of the four storage buffer (lysis buffer, native, PBS or RNA-later) on the alpha diversity was assessed based on OTU richness (identified absolute number of taxa) and evenness (Inverse Simpson index). The choice of storage buffer had no significant influence on the OTU richness of the swab samples ($p = 0.158$ [ANOVA], Fig. 3a), nor the overall evenness. However, PBS treated samples that were extracted with MOBIO, detected a lower evenness compared to all other treatment conditions (Wilcoxon test, Fig. 3b).

Pairwise AMOVA of Bray-Curtis dissimilarity showed that the storage buffer choice had a significant impact on the community structure ($p = 0.004$, AMOVA). A dendrogram of the Bray-Curtis dissimilarity shows that the PBS stored samples clustered separately from the other buffer types which was confirmed by parsimony analysis ($p = 0.001$, Fig. 3c). To examine which OTUs drive the differential clustering, we examined the read count for each OTU. Four bacterial OTUs corresponding to *Neisseria*, *Pseudomonas*, *Porphyromonas* and *Helicobacter* are significantly different in the PBS stored samples for all extraction kits (Fig. 3d–g). These results indicated that PBS buffer significantly alters single OTUs as well as the overall bacterial composition compared to all other storage buffers, independent of extraction kit choice. The bacterial profile of the blank control samples indicated that this effect is not caused by a buffer specific contamination as there appears to be no obvious buffer or kit specific profile (Supplementary Fig. S3).

Effect of extraction method. Richness, both the observed number of OTUs and Choa1, were analyzed to see the effect of the extraction kit choice on the alpha diversity. Pairwise comparison showed no significant effect on OTU richness between the different extraction kits ($p = 0.893$ [ANOVA], Table 2). In general, all extraction kits detect a higher OTU richness compared to the expected richness of the mock community (Table 2). In addition to assessing richness, evenness was analyzed using the Inverse Simpson index. The evenness of the samples extracted using MOBIO was significantly lower compared to the QMINI and GEN-IAL extractions ($p = 0.008$, $p = 0.023$, Wilcoxon test, Table 2). The evenness did not significantly vary between QMINI and GEN-IAL. Yet, the mean ($\pm$SEM) observed evenness (5.21 ± 0.08) was significantly lower than the expected evenness of the mock community (18.3). The same five OTUs, *Enterococcus, Neisseria, Escherichia, Pseudomonas*, and *Bacillus* dominate the bacterial profile independent of extraction kit choice (Fig. 3c).

Pairwise AMOVA of Bray-Curtis dissimilarity indicated that the extraction kit choice significantly impacted the community structure ($p = 0.001$, AMOVA). To assess which extraction kit more accurately represents the bacterial community structure, a theoretical ideal mock community (actual) composition was created for comparison (see methods for details). In the ideal scenario, the experimental data would be identical to the actual composition and there would be no Bray-Curtis dissimilarity. To assess the extraction kits, Bray-Curtis dissimilarity was calculated between the observed and actual mock community for each sample (Fig. 4). The samples extracted with the same commercial kit were grouped in a boxplot and pairwise comparison was performed. The QMINI kit produced a significantly better representation of the bacterial community compared to all other kits tested (paired t-test, all $p < 0.01$, Fig. 4). On the contrary, the MOBIO kit performed significantly poorer than all

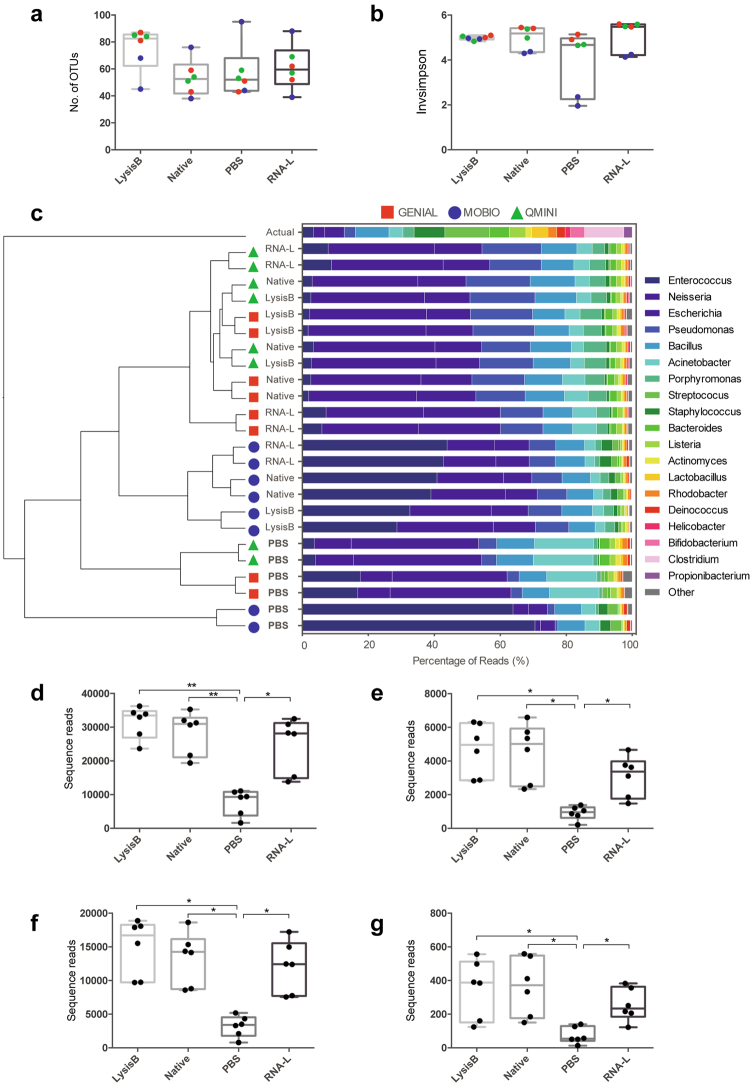


Figure 3. PBS stored samples significantly distort individual OTUs and cluster separately from other buffer types. Boxplots (median ± range) of (**a**) the number of OTUs and (**b**) the Inverse of the Simpson index for each buffer type. (**c**) UPGMA clustering on Bray-Curtis dissimilarities including taxa plots showing the relative abundance of OTUs in percentage of reads. Differential clustering of PBS to all other buffers was found to be significant (parsimony test, $p = 0.001$) (**d**–**g**) Individual bacterial OTUs are significantly underrepresented for PBS-stored samples. Number of sequence reads for OTUs corresponding to (**d**) *Neisseria*, (**e**) *Pseudomonas*, (**f**) *Porphyromonas*, and (**g**) *Helicobacter*. (Wilcoxon test, $*p < 0.05$, $**p < 0.01$).

other tested kits (all $p < 0.01$, Fig. 4). Overall, all the extraction kits distort the bacterial profile compared to the expected bacterial composition of the mock community (Fig. 4).

Discussion

We compared a variety of storage buffers, extraction kits, and amplification methods to examine which combination of handling conditions best represents the microbial diversity of an even mock community (Fig. 1). Different combinations of factors that most closely resemble the reality of sample handling were analyzed to facilitates the

Extraction Method	Observed OTUs	Chao1	InvSimpson
MOBIO PowerLyzer PowerSoil Kit	62.88 ± 8.38	69.79 ± 10.31	3.9 ± 0.40
GEN-IAL First All-Tissue Kit	59.75 ± 5.82	66.02 ± 7.51	5.3 ± 0.09
QIAamp DNA Mini Kit	64.00 ± 4.87	78.04 ± 5.94	5.1 ± 0.13
Actual/Expected Mock Community	22	22	18.3

Table 2. Alpha diversity measurements (mean ± SEM) for each of the DNA extraction kits (n = 8).

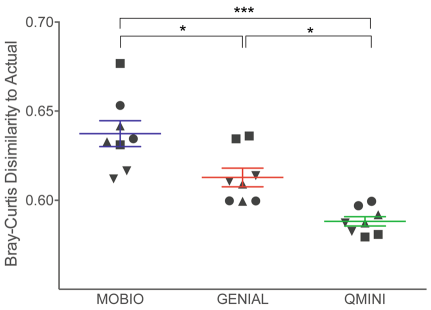


Figure 4. Bray-Curtis dissimilarity between observed and expected strain proportion for each of the tested extraction methods. The expected strain proportion (actual) was generated for comparison and represents the theoretically composition of the mock community (see methods for detail). The pair-wise proportions (expected to observed) from samples extracted with the same commercial kit were grouped in a single boxplot (mean ± SEM). Symbols illustrate different buffer types (■ PBS, ▼RNA-later, ▲native, ●lysis buffer) (Paired t-test, *$p < 0.05$, ***$p < 0.001$).

establishment of standards for the analyses of microbial compositions in swab samples. We show that the choice of storage buffer and extraction kit affects the detected bacterial composition, while different amplification methods had only a minor effect.

Using a mock community, four storage buffers were tested that have been previously used in various studies[25–28]. All samples in this study were frozen at −80 °C rapidly after collection. The samples stored in RNA-later, lysis buffer and native performed similarly to each other and revealed a similar detected bacterial diversity (Fig. 3). Samples stored in RNA-later have been previously reported to decrease DNA purity, lower DNA extraction yields, and to significantly alter the microbial diversity compared to native frozen samples[10,29]. This, however, was not observed in our study. It is likely, that this reflects differences in the sample material (microbes on swab vs. fecal samples) as it has been observed that fecal samples are harder to disperse evenly in RNA-later which may affect the storage and extraction efficiency[10]. Interestingly, compared to the other buffer types, swabs stored in PBS show an altered bacterial composition. There is no indication of a PBS buffer specific contamination profile in the blank samples that could explain this differential clustering. Moreover, PBS buffer in combination with the MOBIO extraction kit detected a lower evenness, which indicates that PBS seems to be particularly incompatible with certain extraction kits. PBS is a balanced salt solution that maintains pH, osmotic balance and is therefore frequently used as a wash buffer in cell and tissue culture. PBS storage has been recommended by manufacturers protocols and has been previously used when examining various extraction kits[12,30]. Other studies examining the effect of different storage conditions have not tested PBS despite its use in DNA extraction from swab material[6,10–12]. It is not clear what properties of PBS effect the mock community differently from other storage buffers. Due to its properties, the buffer may stabilize certain cell types and therefore create a different bacterial profile. Interestingly, despite the different bacterial profile, the PBS samples perform similarly to the other buffer types when comparing them to the mock community. This indicates that the choice of buffer can affect the bacterial profile and specific OTUs, but does not lead to a significantly worse representation of the bacterial community. Our findings support the notion that standardization in sample collection and handling is essential to allow comparison of data within a study[31]. Additionally, field-wide standardization across handling protocols is vital for each sample type, so that cross-study comparisons become possible.

All extraction methods used in this study identify all 19 OTUs present in the mock microbial community (22 bacterial strains of 19 genera, Supplement Table S1). However, all kits detected a higher richness compared to the actual richness of the mock control. A low concentration of mock community (approximately 1×10^7 cells/mL of each organism) was used in this study to simulate the expected bacterial amount in vaginal or oral swab samples[32]. Therefore, it was not surprising that additional OTUs were detected[13]. However, 99% of the pooled library clusters into 19 OTUs which correspond to the bacteria in the mock community. This indicates that the additionally detected OTUs correspond to a small fraction of sequence reads and may therefore be a result of contamination. This study in combination with previous work suggests that the expected biomass of

vaginal and oral swab samples is sufficient for amplicon-based microbial detection without the need of additional target enrichment[13]. The use of a mock microbial community in this study allowed for direct assessment of the extraction kit performance. This comparison indicated that QMINI provides the best representation of the bacterial community when compared to MOBIO and GENIAL. Using a mock community, Yuan *et al.* also found that an altered version of QMINI provided the best bacterial profile[6]. A study using oral swabs confirmed that QMINI extracts DNA with significantly greater yield and good quality compared to other extraction kits[2]. This is in contrast to previous studies on fecal and soil samples, which found that MOBIO most effectively extracts microbial DNA of various bacterial strains[33]. These reported differences in optimal extraction kit may be due to the differences in sample type. The overall bacterial DNA and exogenous material (e.g. fiber) differs substantially between fecal and swab material[34]. Standardization of the extraction kit may thus only be appropriate within each sample type.

In this study, we find that the choice between the two polymerases and the addition of five cycles in amplification of the 16S rRNA gene did not have a significant effect on the bacterial community structure (Fig. 2). Contrary to our findings, Wu *et al.* report that the choice of polymerase had an effect on the microbial community structure, however, the two polymerases that were tested had considerable differences in the fidelity (20 times and 4 times higher than *Taq*)[35]. The two hot-start polymerases used in our study, had significantly higher fidelity (100 times and 52 times higher compared to *Taq*) and are both recommended for NGS applications by the manufacturers. This may likely explain the lack of observable differences. Unlike polymerase choice, which had no effect on the detected evenness or richness, the addition of five PCR cycles to the amplification method led to an overestimation of the bacterial richness. Previous studies have already suggested that this increase is due to an upsurge of chimeric structures with increased cycle numbers[3,5,35]. This supports the notion that lower cycles numbers are favorable for amplicon sequencing[5].

All tested conditions in this study lead to a distortion of the bacterial community structure compared to the expected bacterial mock composition (Fig. 4). *Enterococcus, Neisseria, Escherichia*, and *Pseudomonas* dominated the detected profile in our study, while other bacteria genera such as *Lactobacillus* were underrepresented. Knowledge of which genera are underestimated in the detected bacterial profile (e.g. *Lactobacillus*) is essential to properly estimate the bias when studying certain bacterial communities (e.g., the vaginal microbiome). In a recent study using the same mock community, the bacterial profile resembled the one detected in our study, indicating that the observed distortion is most likely not due to laboratory or kit specific contamination[3,13]. Instead, the bias could be attributed to a variety of factors that were not examined in this study, such as differential susceptibility of bacteria to lysis[6]. To increase lysis efficiency of a broader spectrum of bacteria, enzymatic pre-treatment has been studied as a potential solution, with mixed results[6,36,37]. Another potential cause for the observed bias is the use of primers for 16 S rRNA gene amplification. Although these are universal, amplification may favour certain bacterial strains thus creating bias in the analysis[38,39]. Shotgun metagenomics has been proposed as a solution as it negates some of the bias caused by the amplification, however, this technique does not negate all of technical caveats as storage and extraction kit choice can still have a major impact on the results[3,40]. Continual improvement to the sample handling conditions for both amplicon sequencing and shotgun metagenomics using mock communities is therefore essential.

Conclusion

For now, investigators should standardize the sample handling methods for each sample type as consistency among sample collection, sample storage and sample processing is able to significantly reduce variation. Preliminary tests on specific sample types should be used to ensure that the comparative analysis is as accurate as possible. Caution is, however, warranted when drawing conclusions about the relative abundance of bacterial populations in a single sample and when combining data for meta-analyses.

References

1. Sobel, J. D. Is there a protective role for vaginal flora? *Curr. Infect. Dis. Rep.* **1**, 379–383 (1999).
2. Vesty, A., Biswas, K., Taylor, M. W., Gear, K. & Douglas, R. G. Evaluating the Impact of DNA Extraction Method on the Representation of Human Oral Bacterial and Fungal Communities. *PLoS One* **12**, e0169877, https://doi.org/10.1371/journal.pone.0169877 (2017).
3. Brooks, J. P. *et al.* The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.* **15**, 66, https://doi.org/10.1186/s12866-015-0351-6 (2015).
4. V Wintzingerode, F., Göbel, U. B. & Stackebrandt, E. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* **21**, 213–229 (1997).
5. Ahn, J.-H., Kim, B.-Y., Song, J. & Weon, H.-Y. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *J. Microbiol.* **50**, 1071–1074 (2012).
6. Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z. & Forney, L. J. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* **7**, e33865, https://doi.org/10.1371/journal.pone.0033865 (2012).
7. Mackenzie, B. W., Waite, D. W. & Taylor, M. W. Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Front. Microbiol.* **6**, 130, https://doi.org/10.3389/fmicb.2015.00130 (2015).
8. Kennedy, N. A. *et al.* The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One* **9**, e88982, https://doi.org/10.1371/journal.pone.0088982 (2014).
9. Willner, D. *et al.* Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS One* **7**, e34605, https://doi.org/10.1371/journal.pone.0034605 (2012).
10. Dominianni, C., Wu, J., Hayes, R. B. & Ahn, J. Comparison of methods for fecal microbiome biospecimen collection. *BMC Microbiol.* **14**, 103, https://doi.org/10.1186/1471-2180-14-103 (2014).
11. Choo, J. M., Leong, L. E. & Rogers, G. B. Sample storage conditions significantly influence faecal microbiome profiles. *Sci. Rep.* **5**, 16350, https://doi.org/10.1038/srep16350 (2015).
12. Bai, G. *et al.* Comparison of storage conditions for human vaginal microbiome studies. *PLoS One* **7**, e36934, https://doi.org/10.1371/journal.pone.0036934 (2012).

13. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87, https://doi.org/10.1186/s12915-014-0087-z (2014).
14. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
15. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
16. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
17. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
18. Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27, https://doi.org/10.1186/s40168-017-0237-y (2017).
19. Větrovský, T. & Baldrian, P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* **8**, e57923, https://doi.org/10.1371/journal.pone.0057923 (2013).
20. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
21. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
22. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
23. Schloss, P. D. & Handelsman, J. Introducing TreeClimber, a test to compare microbial community structures. *Appl. Environ. Microbiol.* **72**, 2379–2384 (2006).
24. White, J. R., Nagarajan, N. & Pop, M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* **5**, e1000352, https://doi.org/10.1371/journal.pcbi.1000352 (2009).
25. Perez, G. I. P. *et al.* Body site is a more determinant factor than human population diversity in the healthy skin microbiome. *PLoS One* **11**, e0151990, https://doi.org/10.1371/journal.pone.0151990 (2016).
26. Yildirim, S. *et al.* Primate vaginal microbiomes exhibit species specificity without universal Lactobacillus dominance. *The ISME journal* **8**, 2431–2444 (2014).
27. Pyles, R. B. *et al.* Cultivated vaginal microbiomes alter HIV-1 infection and antiretroviral efficacy in colonized epithelial multilayer cultures. *PLoS One* **9**, e93419, https://doi.org/10.1371/journal.pone.0093419 (2014).
28. Chehoud, C. *et al.* Complement modulates the cutaneous microbiome and inflammatory milieu. *Proceedings of the National Academy of Sciences* **110**, 15061–15066 (2013).
29. Hale, V. L., Tan, C. L., Knight, R. & Amato, K. R. Effect of preservation method on spider monkey (Ateles geoffroyi) fecal microbiota over 8 weeks. *J. Microbiol. Methods* **113**, 16–26 (2015).
30. Abusleme, L., Hong, B.-Y., Dupuy, A. K., Strausbaugh, L. D. & Diaz, P. I. Influence of DNA extraction on oral microbial profiles obtained via 16S rRNA gene sequencing. *J. Oral Microbiol.* **6**, https://doi.org/10.3402/jom.v6.23990 (2014).
31. Stulberg, E. *et al.* An assessment of US microbiome research. *Nature microbiology* **1**, 15015, https://doi.org/10.1038/nmicrobiol.2015.15 (2016).
32. Virtanen, S., Kalliala, I., Nieminen, P. & Salonen, A. Comparative analysis of vaginal microbiota sampling using 16S rRNA gene analysis. *PLoS One* **12**, e0181477 (2017).
33. İnceoğlu, Ö., Hoogwout, E. F., Hill, P. & van Elsas, J. D. Effect of DNA extraction method on the apparent microbial diversity of soil. *Appl. Environ. Microbiol.* **76**, 3378–3382 (2010).
34. Sender, R., Fuchs, S. & Milo, R. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* **14**, e1002533, https://doi.org/10.1371/journal.pbio.1002533 (2016).
35. Wu, J.-Y. *et al.* Effects of polymerase, template dilution and cycle number on PCR based 16S rRNA diversity analysis using the deep sequencing method. *BMC Microbiol.* **10**, 255, https://doi.org/10.1186/1471-2180-10-255 (2010).
36. Gill, C., van de Wijgert, J. H., Blow, F. & Darby, A. C. Evaluation of Lysis Methods for the Extraction of Bacterial DNA for Analysis of the Vaginal Microbiota. *PLoS One* **11**, e0163148, https://doi.org/10.1371/journal.pone.0163148 (2016).
37. Rosenthal, R., Blundell, J. & Perkins, H. Strain-related differences in lysozyme sensitivity and extent of O-acetylation of gonococcal peptidoglycan. *Infect. Immun.* **37**, 826–829 (1982).
38. Soergel, D. A., Dey, N., Knight, R. & Brenner, S. E. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *The ISME journal* **6**, 1440–1444 (2012).
39. Hamady, M. & Knight, R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res.* **19**, 1141–1152 (2009).
40. Jovel, J. *et al.* Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front. Microbiol.* **7**, https://doi.org/10.3389/fmicb.2016.00459 (2016).

Acknowledgements

The authors thank the Biodefense and Emerging Infectious Research (BEI) Resources, NIAID, NIH for providing the cells from Microbial Mock Community (Even, HM-280) as part of the Human Microbiome Project. We thank Dr. Dietmar Zinner, Uwe Schönmann and Dr. Angela Noll (German Primate Center) for their guidance and general support. We thank the staff of the Transcriptome and Genome Analysis Laboratory at the University of Göttingen for their assistance in optimizing the sequencing run.

Author Contributions

The study was designed by L.H.W., C.R., and S.K. Laboratory work was conducted at the German Primate Center and performed by L.H.W. and S.L. Data were analyzed by L.H.W. and S.K. All authors (L.H.W., S.L., C.R. and S.K.) contributed to the manuscript preparation.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-24573-y.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter B2

Supplemental information:

Table S1. Organisms in the microbial mock community HM-280

Figure S1. Processing of swab samples prior to DNA extraction

Table S2. Bray-Curtis Dissimilarity between identical samples amplified with either Platinum SuperFi DNA Polymerase or the Phusion Hot Start II High-Fidelity DNA Polymerase

Figure S2. Five additional cycles ('c') significantly altered the number of detected OTUs but did not alter the bacterial composition

Figure S3. Bacterial profile of blank samples for each buffer and kit used in our study

The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis

Luisa K Hallmaier-Wacker[1,2], Simone Lueert[1,2], Christian Roos,[2] Sascha Knauf,[1*]

Supplementary Information:

Table S1: Organisms in the microbial mock community HM-280

Organism	16S rRNA gene copy number
Acinetobacter baumannii, strain 5377	5
Actinomyces odontolyticus, strain 1A.21	2
Bacillus cereus, strain NRS 248	12
Bacteroides vulgatus, strain NTC 11154	7
Bifidobacterium adolescentis, strain E194a	5
Clostridium beijerinckii, strain NCIMB 8052	14
Deinococcus radiodurans, strain R1 (smooth)	3
Enterococcus faecalis, strain OG1RF	4
Escherichia coli, strain MG1655	7
Helicobacter pylori, strain 26695	2
Lactobacillus gasseri, strain 63AM	6
Listeria monocytogenes, strain EGDe	6
Neisseria meningitides, strain MC58	4
Porphyromonas gingivalis, strain 2561	4
Propionibacterium acnes, strain KPA171202	3
Pseudomonas aeruginosa, strain PAO1-LAC	4
Rhodobacter sphaeroides, strain ATH 2.4.1	3
Staphylococcus aureus, strain TCH959	6
Staphylococcus epidermidis, FDA strain PCI 1200	5
Streptococcus agalactiae, strain 2603 V/R	7
Streptococcus mutans, strain UA159	5
Streptococcus pneumoniae, strain TIGR4	4

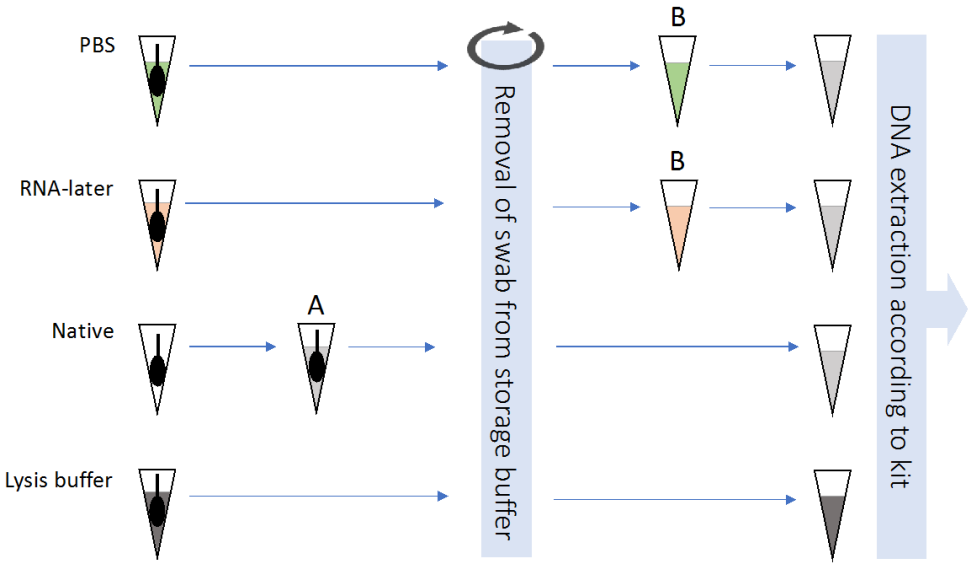


Figure S1: Processing of swab samples prior to DNA extraction. (A) A kit-specific lysis buffer was added to the native swab sample, which was briefly vortexed to ensure even distribution of the buffer. Subsequently, the bottom of the swab-containing tubes (PBS, RNA-later, native and custom-made lysis buffer) was pierced by a sterile needle (20G) to allow for the separation of the swab material from the respective storage buffer. Separation was forced by centrifugation (2,500g for 1 minute at room temperature). (B) Prior to DNA extraction, RNA-later and PBS buffer were centrifuged at 11,000g for 5 minutes at 4 °C. The supernatant was removed and replaced by kit-specific lysis buffer.

Table S2: Bray-Curtis Dissimilarity between identical samples amplified with either Platinum SuperFi DNA Polymerase or the Phusion Hot Start II High-Fidelity DNA Polymerase.

Kit	Buffer	Bay-Curtis Dissimilarity
MOBIO	Lysis buffer	0.064
	Native	0.039
	PBS	0.076
	RNA-later	0.032
GENIAL	Lysis buffer	0.028
	Native	0.036
	PBS	0.027
	RNA-later	0.025
QMINI	Lysis buffer	0.062
	Native	0.076
	PBS	0.019
	RNA-later	0.044

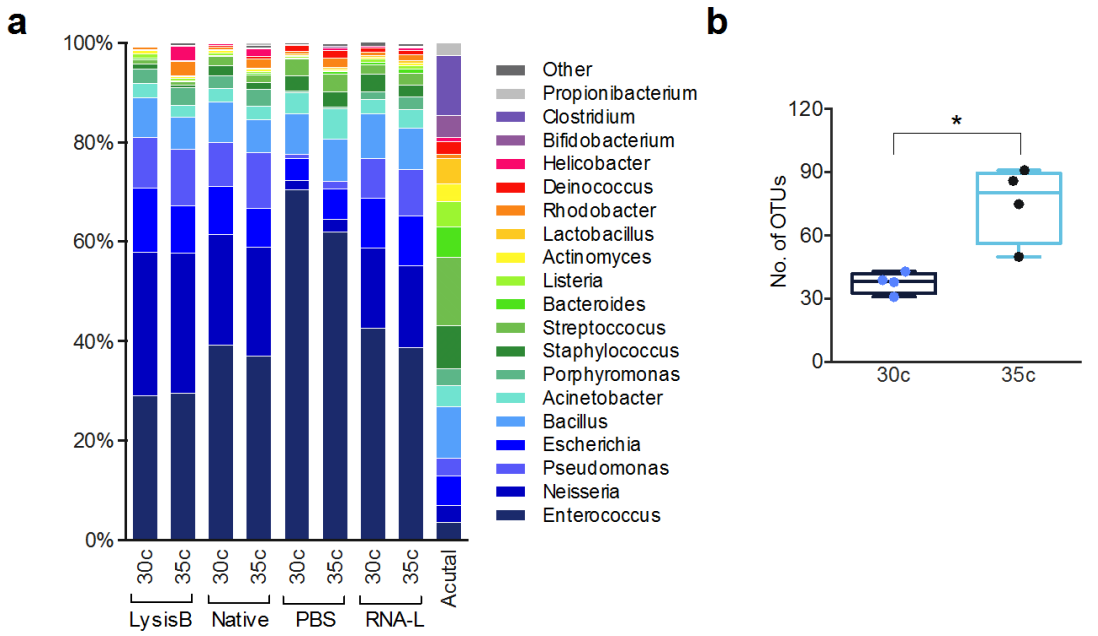


Figure S2: **Five additional cycles ('c') significantly altered the number of detected OTUs but did not alter the bacterial composition**. (a) Taxa plots showing the relative abundance of OTUs identified in percentage of reads. Actual refers to the predicted bacterial composition of the mock community. (b) Boxplots (median ± range) of the number of OTUs in each of the samples plotted by cycle number ('c'). (Wilcoxon t-test, *$p<0.05$)

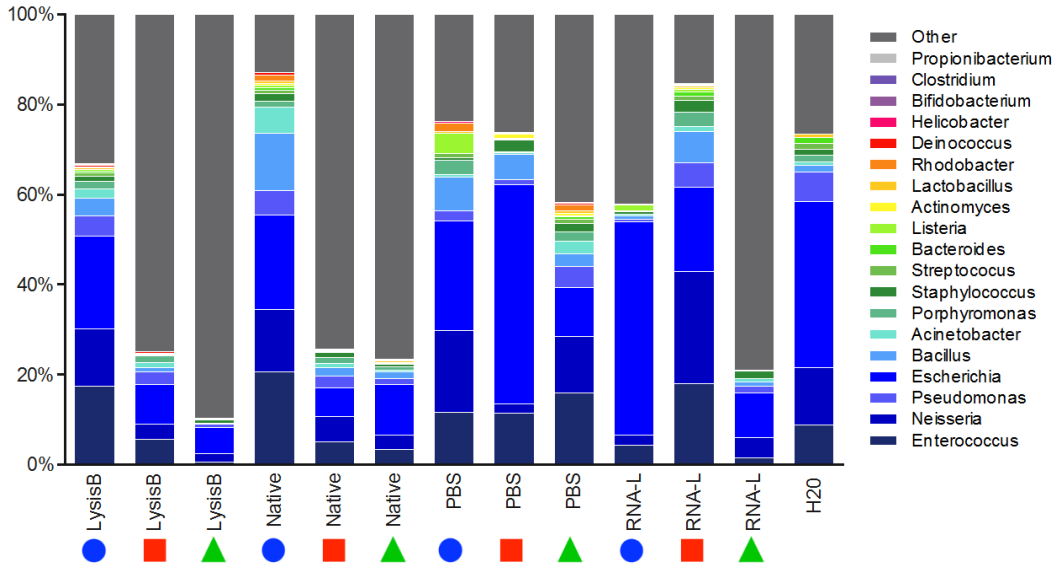


Figure S3: Bacterial profile of blank samples for each buffer and kit used in our study. A total of 946,412 sequences corresponded to the blank sample, with a mean read count of 12,727 reads per sample. Taxa plots showing the relative abundance of OTUs identified in percentage of reads. Symbols illustrate different extraction kits (●MOBIO, ■ GENIAL, ▲QMINI)

Chapter B3

Lactation and menstruation shift the vaginal microbiota in captive rhesus monkeys to be more similar to the male urethral microbiota

Luisa K. Hallmaier-Wacker[1,2], Simone Lüert[1,2], Christian Roos[2,3] and Sascha Knauf[1,4]

Scientific Reports **9**: 17399 (2019) doi: 10.1038/s41598-019-53976-8

[1]Neglected Tropical Diseases Work Group, Infection Biology Unit, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[2]Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[3]Gene Bank of Primates, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[4]Division of Microbiology and Animal Hygiene, Georg-August-University, Göttingen, Germany

---

Author contributions:

Conceived and designed the study: LHW, CR, SK
Performed sample collection: LHW, SL, SK
Conduced laboratory work: LHW, SL
Analyzed the data: LHW, SK
Manuscript preparation: LHW, SL, CR, SK

www.nature.com/scientificreports

SCIENTIFIC REPORTS
nature research

OPEN

Lactation and menstruation shift the vaginal microbiota in captive rhesus monkeys to be more similar to the male urethral microbiota

L. K. Hallmaier-Wacker[1,2], S. Lüert[1,2], C. Roos[2,3] & S. Knauf[1,4*]

The vaginal microbiota of nonhuman primates differs substantially from humans in terms of *Lactobacillus* abundance, overall taxonomic diversity, and vaginal pH. Given these differences, it remains unclear in what way the nonhuman primate genital microbiota protects against pathogens, in particular sexually transmitted infections. Considering the effect that microbiota variations can have on disease acquisition and outcome, we examined endogenous and exogenous factors that influence the urogenital microbiota of male and female captive rhesus monkeys. The male urethral (n = 37) and vaginal (n = 194) microbiota of 11 breeding groups were examined in a cross-sectional study. During lactation and menstruation, the vaginal microbiota becomes significantly more diverse and more similar to the microbes observed in the male urethra. Group association and cage-mate (sexual partners) relationships were additionally associated with significant differences in the urogenital microbiota. Our results demonstrate that microbiota considerations are necessary in order to make informed selection of nonhuman primates as translational animal models.

In recent years there has been an increased interest in the microbiota of nonhuman primates (NHPs) for evolutionary, experimental, and conservation purposes. However, microbiota considerations are currently not used to refine and reduce experiments with NHPs, despite increasing evidence that the microbiota in humans can influence disease progression (reviewed by[1]). Of the NHP animal models, the Asian rhesus monkey (*Macaca mulatta*) and long-tailed macaque (*Macaca fascicularis*) are the most extensively utilized species[2–4]. In laboratory settings, rhesus monkeys cycle year-round, have a reproductive cycle that is similar to that of humans and experience similar changes in the hormonal levels during sexual cycle, pregnancy and post-partum[5–7]. Therefore, the vagina of rhesus monkeys has been used to model the human vaginal epithelium and study sexual transmitted infections (STIs)[8]. For example, rhesus monkeys have been extensively used to study the disease acquisition and outcome of simian-/human immunodeficiency virus (SIV/HIV)[3,9]. In a study on SIV susceptibility, estrogen treatment in rhesus monkeys protected female rhesus monkeys from the sexually transmitted infection[3]. Smith *et al.* proposed that not just the thickening of the vaginal epithelium but also a potential change in vaginal microenvironment may have led to the observed effect under the influence of high estrogen levels[3].

Many studies have laid the groundwork in characterizing the genital microenvironment of various species of captive and wild NHPs[2,10–12]. Unlike the vaginal microbiota of humans, which is often dominated by a single *Lactobacillus* species[13], NHPs, including rhesus monkeys, harbor a diverse set of vaginal microbes[2,11]. In humans, the acidic nature of the vaginal environment (pH $\leq$ 4.5) protects women against STIs[14]. The vaginal microbiota of NHPs on the other hand has a low abundance of *Lactobacillus* (<2% of microbiota), a near neutral vaginal pH and is instead colonized by a diverse set of bacterial genera including *Sneathia*, *Aerococcus*, *Prevotella, Porphyromonas, Fusobacterium and Atopobium*[2,11,12,15]. Considering these differences, it currently remains unclear in what way the vaginal microenvironment of rhesus monkeys protects against infectious diseases. Additionally, despite increasing evidence that sexual exposures can alter the composition of the human genital microbiota[16,17], the urethral

[1]Work Group Neglected Tropical Diseases, Infection Biology Unit, Deutsches Primatenzentrum GmbH, Leibniz Institute for Primate Research, Goettingen, Germany. [2]Primate Genetics Laboratory, Deutsches Primatenzentrum GmbH, Leibniz Institute for Primate Research, Goettingen, Germany. [3]Gene Bank of Primates, Deutsches Primazentrum GmbH, Leibniz Institute for Primate Research, Goettingen, Germany. [4]Division of Microbiology and Animal Hygiene, Georg-August-University, Goettingen, Germany. *email: sknauf@dpz.eu

	Female (n = 194)	Male (n = 37)
Mean age, years	10.0 ± 4.9	7.9 ± 5.6
n Geriatric (>19 years)	10	1
n Adult (5–19 years)	168	17
n Juvenile (<5 years)	16	19
Breeding groups	11	10
Vaginal pH, mean ± SD	6.4 ± 0.7 (n = 138)	N/A
% EVC*	76.3% (n = 148)	N/A
n Phase 1	44†	N/A
n Phase 2	57†	N/A
n Phase 3	47†	N/A
% Lactating	30.4 (n = 59)	N/A
% Breeder males	N/A	40.5 (n = 15)

Table 1. Characteristics of sampled rhesus monkeys in this study. *Cytology phases classification described in detail in Methods and Supplementary Table S2; †The subdivision of sexual cycle phases by lactating status can be seen in Table S2.

microbiota of male NHPs remains largely uncharacterized. A better understanding of factors that influence the rhesus monkey genital microbiota of both male and female animals in health and disease is thus warranted.

In this study, we investigate the genital microbiota of a large breeding colony of rhesus monkeys at the German Primate Center. To identify endogenous and exogenous factors that influence the microbiota, we examined the genital microbiota in the breeding colonies in the context of age, breeding group association, social rank, body mass, and long-term health status. We studied both, the vaginal microbiota of female and the urethral microbiota of male rhesus monkeys which has not been done in previous studies. We are therefore able to compare bacterial composition between male and female animals in a single cohort of rhesus monkeys and found that during menstruation and lactation the vaginal microbiota shifts to be more similar to the male urethral microbiota.

Results

We examined the vaginal microbiota of 194 female rhesus monkeys and the urethral microbiota of 37 male rhesus monkeys housed at the breeding facility of the German Primate Center (data file S1). The mean age, number of breeding groups, and other characteristics of the sampled animals are shown in Table 1. The V4 region of the bacterial 16S rRNA gene was selected as its informative, short-read length from the Illumina MiSeq platform allows full overlap of paired-end reads and thus higher confidence in the sequence data. Sequencing of the V4 region of the rhesus monkey samples generated a total of 14,571,505 unfiltered reads with a mean read count of 48,630 reads per sample (± 15,461 SD) after quality filtering. Sequences were rarified to 11,371 sequences per sample and clustered into operational taxonomic units (OTUs) based on the 97% similarity threshold. We first examined the microbiota of the male urethra and vagina separately and then compared composition similarities between male and female animals.

Appropriate control samples and a mock community (Microbial mock community, HM-280, Biodefense and Emerging Infectious Research (BEI) Resources, Manassas, USA) were included in the sequencing run. Using the mock community, the observed error rate for the run was found to be 0.036%. The collected control samples showed that contamination was highest during sample collection procedures, while controls taken during amplification procedures in the laboratory yielded only minimal read counts (Fig. S1a). Taxa plots of control samples that were taken during sampling at two different breeding units, show that the relative abundance of contaminant OTUs was similar (Fig. S1b).

The vaginal microbiota is significantly altered during lactation and menstruation. We investigated the vaginal microbiota (n = 194), vaginal pH (n = 138), and sexual cycle phase (n = 148) of clinically-healthy, reproductively-active rhesus monkeys housed in eleven breeding groups (Table 1). None of the females showed signs of pregnancy, as defined by transabdominal palpation. At the time of sampling 30.4% of the animals were lactating. In addition to lactation, we characterized the sexual cycle phase using exfoliative vaginal cytology (EVC) (see Methods, Table S1). The sexual cycle phase (P1-P3) of non-lactating females were evenly distributed with 35.6% in an ovulatory phase (P1), 41.3% in an intermediate phase (P2) and 23.1% in a menstruation-like phase (P3) (Table S1). For lactating females, 52.3% of the animals were in a menstruation-like phase, 31.8% in an intermediate phase, and 15.9% in the ovulatory phase (Table S1). For the purpose of the microbiota analysis, lactation status and sexual cycle phases were analyzed independently.

Overall, a mean of 219.8 ± 160.7 (unless otherwise stated all values are given in mean ± SD) OTUs were observed in the vaginal microbiota of the rhesus monkeys. The most abundant genus was *Prevotella* with a mean abundance of 20.5 ± 16.4%. Different OTUs were identified as *Prevotella*, indicating that a diverse set of species from this genus were present (Fig. 1). *Porphyromonas* (9.5 ± 9.9%), *Streptobacillus* (9.1 ± 13.4%) and an unclassified genus of the family *Ruminococcaceae* (9.5 ± 7.3) were the other dominant taxa in the otherwise diverse community (Fig. 1).

Lactation status and sexual cycle phase strongly correlated with the OTU richness (identified absolute number of taxa) and evenness (inverse Simpson index; Fig. 2). Lactating females had a significantly higher OTU richness ($p \leq 0.0001$ [Mann-Whitney t-test]) and the bacterial taxa were significantly more evenly distributed ($p \leq 0.0001$

Figure 1. Heatmap of the relative abundance of microbial taxa identified in the vaginal microbiota of rhesus monkeys in multiple breeding groups. Ward linkage clustering of samples based on the composition and relative abundance of the 20 most abundant OTUs in the vaginal microbiota. Lactation status (1: lactating and 0: non-lactating), group association (A–K) and EVC (sexual cycle phases) (P1: ovulatory phase, P2: intermediate stage, P3: menstruation-like and NT: not tested) of each sample are shown beside the heatmap. C1 and C2 indicate two main clusters in the ward linkage clustering. See Table 1 for sample size composition.

[Mann-Whitney t-test] than in non-lactating females (Fig. 2a,b). Similarly, animals in a menstruation-like (Phase 3, Table S1) sexual cycle phase had a significantly higher OTU richness ($p \leq 0.0001$ [Kruskal-Wallis test]) and were significantly more evenly distributed ($p \leq 0.001$ [Kruskal-Wallis test]) than animals in the ovulatory (Phase 1) or intermediate phase (Phase 2; Fig. 2c,d). A heatmap of the relative abundance of the 20 most common OTUs shows that lactating animals and animals in menstruation-like sexual phase clustered separately from other animals (Fig. 1). Vaginal bacterial communities from these animals clustered prominently in cluster 1 (C1) and are characterized by different bacterial taxa than the cluster 2 (C2; Fig. 1). Of the ten most abundant OTUs, *Provella*, *Mobiluncus*, *Porphyromonas* and an unclassified genus of the family *Ruminococcaceae* were significantly different in the lactating and menstruation-like animals (Fig. S2a,b). These cluster differences were confirmed by significant differences in the unweighted UniFrac distances, which are visualized on the principal coordinates plot along axis 1 (23.1%) (Fig. 3a,b). Pairwise AMOVA confirmed that the differential clustering of lactating and menstruating-like animals resulted in significantly different community structures ($p \leq 0.001$).

In order to examine an additional functional variable of the vaginal microbiota, we tested the vaginal pH at the time of sampling using pH-indicator paper. The mean overall vaginal pH of the sampled animals was found to be 6.4 ± 0.7 (Table 1). The vaginal pH of lactating females (6.8 ± 0.5) was significantly higher than that of non-lactating females (6.3 ± 0.7) ($p \leq 0.0001$ [Mann-Whitney t-test], Fig. S3a). Similarly, animals in menstruation-like sexual phase (7.0 ± 0.5) had a higher vaginal pH compared to individuals in the other sexual cycle phase (P1: 6.0 ± 0.6, P2: 6.3 ± 0.7, $p \leq 0.0001$ [Kruskal-Wallis test], Fig. S3b).

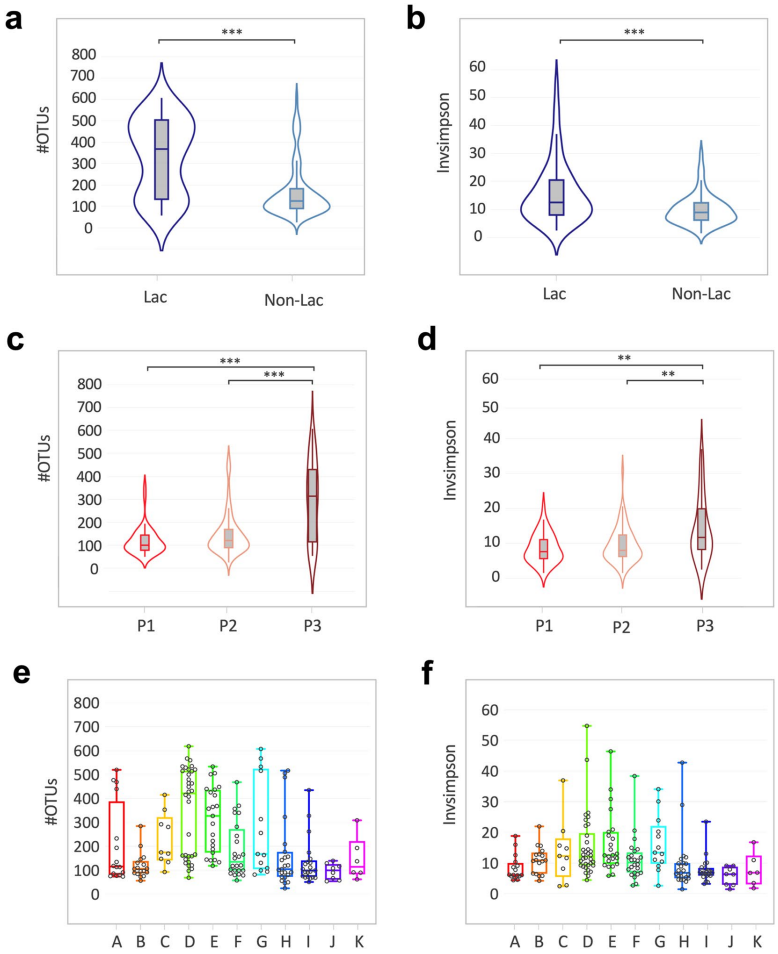


Figure 2. Alpha diversity measurements for the vaginal microbiota of female rhesus monkeys. Violin plots of the observed OTUs and InvSimpson index clustered based on (**a**,**b**) lactation status (Mann-Witney t-test, ***$p \leq 0.0001$) and (**c**,**d**) sexual cycle phases (P1: ovulatory phase, P2: intermediate stage, P3: menstruation-like) (Kruskal-Wallis test, **$p \leq 0.001$, ***$p \leq 0.0001$). (**e**,**f**) Boxplots (median ± range) of the observed OTUs and InvSimpson index clustered by breeding groups (groups association: A-K). See Table 1 for sample size composition.

Breeding groups influence the vaginal microbiota. Aside from lactation status and cycle phase, we examined if the vaginal microbiota is shaped by age and breeding group. Animals were subdivided into juveniles (<5 years), adults (5–19 years), and geriatric (>19 years). Age was not found to have an influence on alpha diversity (Fig. S4) and beta diversity ($p = 0.155$ [AMOVA; Fig. S5a). The group association, in contrast, significantly correlate with the OTU richness ($p \leq 0.0001$ [Kruskal-Wallis test], Fig. 2e) and evenness ($p \leq 0.0001$ [Kruskal-Wallis test], Fig. 2f). The breeding group association of each sample can be seen on the heatmap of the 20 most abundant OTUs (Fig. 1). Additionally, we observed a significant difference in unweighted UniFrac distances when considering all breeding groups ($p \leq 0.001$ [AMOVA; Fig. S5b). Pairwise comparisons of alpha and beta diversity measurements between individual breeding groups was, however, not significant for all tested groups.

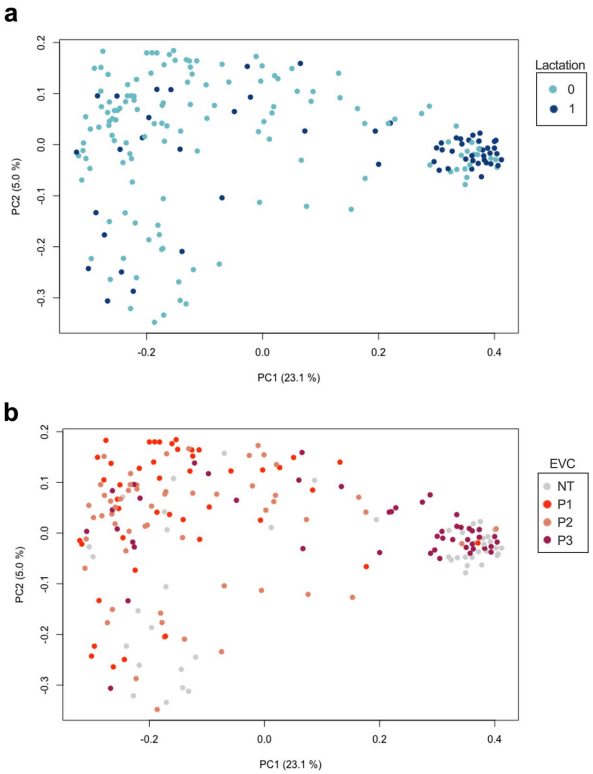


Figure 3. The vaginal microbiota of menstruating-like and lactating females clusters separately. Principal coordinates analysis (PCoA) of vaginal samples colored by (**a**) lactation status (lactating (dark blue, 1) and non-lactating (light blue, 0) and (**b**) sexual cycle status (P1: ovulatory phase (red), P2: intermediate stage (pink), P3: menstruation-like (dark red) and NT: not tested (gray)). Distances between samples were calculated using the unweighted UniFrac metrics. See Table 1 for sample size composition. Fig. S5 shows the corresponding PCoA plot classified by group association and age classification.

Influence of age in adult male rhesus monkeys. We characterized the urethral microbiota of clinically-healthy, reproducing male rhesus monkeys housed in ten different breeding groups (Table 1). Breeding groups are composed of a single adult male (breeding male) and juvenile off-spring males (≤5 years) which are removed from the breeding groups upon reaching adulthood. Overall, the urethral microbiota of male rhesus monkeys is composed of a diverse community of microbes, with a mean of 481.3 ± 127.0 OTUs observed. On a phylum level, *Firmicutes* (54.1 ± 8.3%), *Bacteroidetes* (25.5 ± 9.3%), *Proteobacteria* (9.0 ± 6.7%) and *Actinobacteria* (6.6 ± 3.2%) made up 95% of the identified sequences. The four dominant phyla were present in all 37 samples. On the genus level, the bacterial community is diverse with no single dominating OTU (Fig. 4). The most abundant genus in the male rhesus monkey urethra was *Prevotella* with a mean abundance of 14.4 ± 9.7% followed by *Porphyromonas* (7.5 ± 6.6%) and *Ezakiella* (7.3 ± 6.8%).

We examined if the urethral microbiota of the adult, breeding male differed from the microbiota of juvenile males. Each breeding group had a single adult breeding male with the exception of group H (n = 4) and I (n = 3), which were further divided into subgroups within a single housing unit. Adult males neither differed from other males in the OTU richness (p = 0.145, [Mann-Whitney test], Fig. 4a) nor the evenness (p = 0.453 [Mann-Whitney test], Fig. 4b). Pairwise AMOVA of unweighted UniFrac distances found that being the breeding male in a group had no effect on community structure (p = 0.123). A dendrogram of unweighted UniFrac distances shows that the adult animals did not cluster separately from other animals (parsimony analysis, p = 0.768; Fig. 4c). OTU richness and evenness measurements of each breeding group are shown in Fig. S6. We note here, that the sample size of male animals in each breeding group were low (n = 1 to 4 animals). Therefore, statistical analysis was not performed to examine breeding group differences.

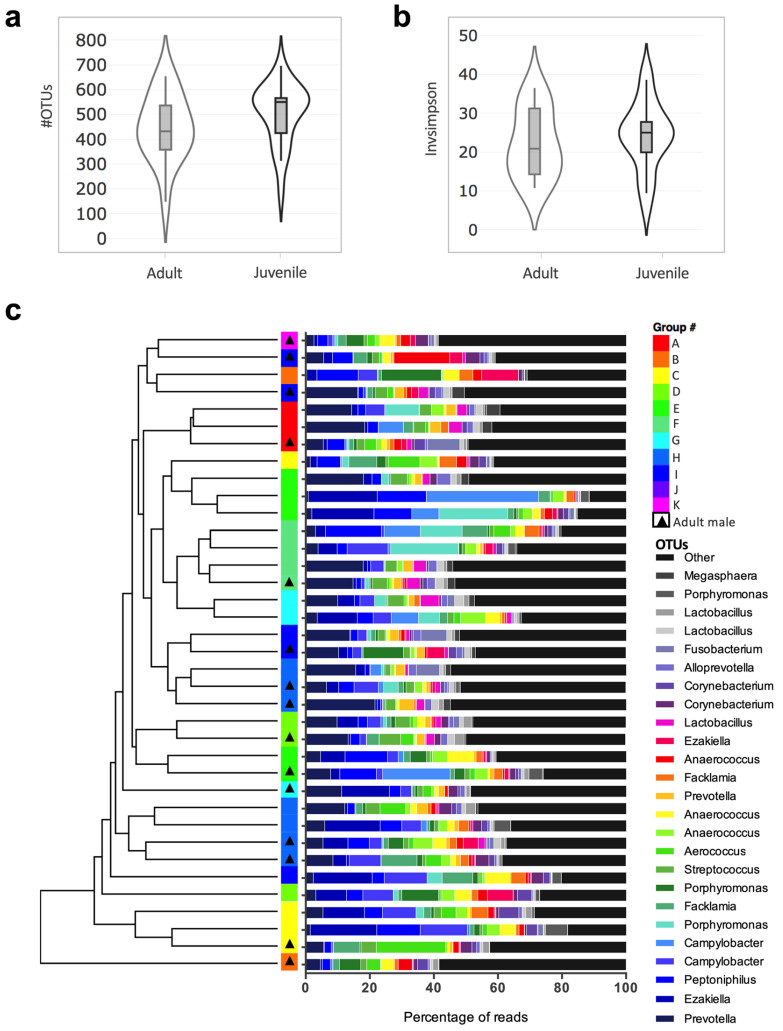


Figure 4. The male urethral microbiota of adult and juvenile rhesus monkeys does not differ significantly. Violin plots of (**a**) the number of OTUs and (**b**) the inverse Simpson index for adult and juvenile males (Mann-Witney t-test). (**c**) UPGMA clustering on unweighted UniFrac including taxa plots showing the relative abundance of the 25 most abundant OTUs in percentage of reads. Genus-level bacterial classification of OTUs shown legend with the percent of sequences that classified with each genus. Group are shown accordingly and adult (breeding male within each group) is indicated. See Table 1 for sample size composition.

Lactating and menstruating female have a more similar microbiota to the male urethra. On the phylum level, *Firmicutes* and *Bacteroidetes* dominated both microbiotas, making up 80.1 ± 6.9% in the male urethra and 69.9 ± 17.8% in the vagina. Yet, *Fusobacteria*, the third most abundant phylum in the vagina (14.4 ± 16.1%), only made up 1.9 ± 3.3% in the male urethra. The most abundant genus across the dataset for both, male and female genital microbiota was *Prevotella*. Several OTUs cluster into this bacterial genus and the most abundant *Prevotella* OTUs (mean abundance of 6.0 ± 8.7%) was found in 226 out of 231 animals. Other *Prevotella* OTUs were less abundant and only dominant in some samples (Fig. 1: vaginal microbiota and Fig. 4: male urethral microbiota).

To further examine similarities between the vagina and male urethra, overall OTU richness and evenness was compared. As we previously observed a significant difference in alpha and beta diversity of the vaginal microbiota based on lactation status and sexual cycle phase, these variables were plotted separately (Fig. 5). The male urethra had a significantly higher OTU abundance compared to non-lactating and non-menstruation-like (ovulatory and intermediate phase) animals ($p \leq 0.0001$ [Kruskal-Wallis test], Fig. 5a,b). Contrary, menstruation-like (P3) and lactating female rhesus monkeys showed no significant difference in the number of OTUs compared to the male urethra microbiota ($p \leq 0.05$ [Kruskal-Wallis test], Fig. 5a,b). Similarly, inverse Simpson index measurements were significantly different between males and non-lactating and non-menstruation-like (ovulatory (P1) and intermediate phase (P2)) animals ($p \leq 0.0001$ [Kruskal-Wallis test], Fig. 5c,d). Inverse Simpson index measurements were not significantly different between males and lactating and menstruation-like animals ($p \leq 0.05$ [Kruskal-Wallis test], Fig. 5c,d). To examine, if the trend in the alpha diversity could be observed in the overall bacterial composition, pairwise unweighted UniFrac distances were calculated between the male urethral microbiota and the vaginal microbiota. More similar microbiotas resulted in a smaller calculated UniFrac distances and vise versa. The UniFrac distances were grouped in violin plots based on either lactation status (Fig. 5e) or sexual cycle phase (Fig. 5f). We found that the bacterial composition of the vaginal microbiota of lactating and menstruating-like animals (P3) was significantly more similar to that of the male urethra microbiota (Fig. 5e,f).

Cage-mates are more similar in their genital microbiota. In order to assess if sexual contact shapes the genital microbiota of the captive rhesus monkeys, pairwise unweighted UniFrac distances were calculated between the male urethral microbiota of the adult male in each group and the vaginal microbiota of adult females. Females and males of the same breeding group were considered cage-mates and thus potential sexual partners. UniFrac distances were grouped into violin plots as 'cage-mates' or from 'other breeding groups' (no sexual contact possible) (Fig. 6). Cage-mates were found to be significantly more similar in the bacterial composition compared to non-cage-mates ($p \leq 0.0001$ [Mann-Whitney test], Fig. 6). As we observed a significant difference in lactation status and sexual cycle phase, these variables were additionally plotted in separate paired-violin plots to examine cage-mate differences for each group (Fig. S7a,b). Cage-mates were found to be significantly more similar in the bacterial composition for lactation, menstruation-like and ovulatory phase animals (Fig. S7a,b). Cage-mate similarity was not observed for the non-lactating group or for animals in an intermediate sexual phase (P2; Fig. S7a,b).

Discussion

Considering the effect that microbiota variation can have on disease acquisition and outcome[1], we examined endogenous and exogenous factors that influence the urogenital microbiota of captive rhesus monkeys. The inclusion of optimal negative controls (Fig. S1) and the relatively large sample size strengthen the study. However, based on our cross-sectional study design we were limited in drawing causal relationships between factors and variations in the genital microbiota. Nevertheless, our results urge for the inclusion of microbiota analysis in the selection and experimental use of rhesus monkeys as indicated by the differences between the vaginal microbiota during lactation and sexual cycles phases.

We showed that during lactation and menstruation the bacterial composition shifts towards a more diverse community (Figs. 1–3). As reported previously, we confirmed that the mean vaginal pH of rhesus monkeys (6.4 ± 0.7) is significantly higher than that found in humans (Fig. S3)[2]. Instead of the *Lactobacillus*-dominance observed in women (reviewed by[18]), the vaginal microbiota of captive and wild NHPs harbor a more diverse set of bacteria (Fig. 1)[2,11,12,15]. Our study shows that in captive rhesus monkeys the already diverse bacterial community shifts to an even more diverse and significantly different bacterial composition during lactation and the menstruation-like phase (Figs. 2–3). While a previous study on captive baboons (*Papio anubis*) found no difference in the vaginal microbiota of cycling females[15], a recent study in wild baboons (*Papio cynocephalus*) reported that the ovarian cycle phase and the reproductive state shaped the vaginal microbiota[12]. Both of these studies used visual assessment of perivulvar swellings to determine the sexual cycle phase[12,15]. Inconsistent classification of these phases in the two studies in combination with low sample sizes may explain the difference in the outcome of both studies. Instead of using perivulvar swellings, we performed vaginal exfoliative cytology to classify the animals into three sexual cycle phases (Table S1)[19–21]. Vaginal exfoliative cytology reflects the current state of the vaginal epithelium and therefore serves as a reliable marker for the sexual cycle phase[19–21]. The even distribution of all three sexual cycle phases in non-lactating rhesus monkeys (Table S1) is indicative of a healthy reproductive community. Using cytology as a marker of sexual cycle phase, this study supports Miller *et al.*'s finding that ovarian cycle phase (menstruation-like) and reproductive state (lactation) shifts the vaginal microbiota in NHPs[12]. Similar changes have been reported in temporal and cross-section studies in women during menstruation and post-partum[13,22], where it has been demonstrated that the vaginal microbiota shifted from a *Lactobacillus*-dominant state towards a more diverse bacterial composition[13,22]. Despite the remarkable differences in bacterial species composition of the rhesus monkeys and human vaginal microbiota, it is interesting that similar factors (e.g. hormonal changes) seem to influence the vaginal microbiota. This is supported by our finding that the observed changes in bacterial vaginal diversity in the rhesus monkeys coincide with changes in the pH, a functional measurement of the vaginal ecosystem. Whether the observed variance in the bacterial diversity and the difference in pH is physiologically relevant cannot be determined in this study. However, none of the animals had any clinical manifestations of vaginitis, which supports our notion that the bacterial variation and the pH differences observed in this study is within physiologically range.

It has been proposed that hormonal fluctuations during the sexual cycle, pregnancy and post-partum shape the vaginal microbiota (reviewed by[18]). Both lactation and menstruation are marked by hormonal changes in the vagina, which may be indirect or direct driving factors for the shift in vaginal microbiota observed in this study (Figs. 1–3). Studies on SIV susceptibility in rhesus monkeys have shown that hormone treatment can lead to an

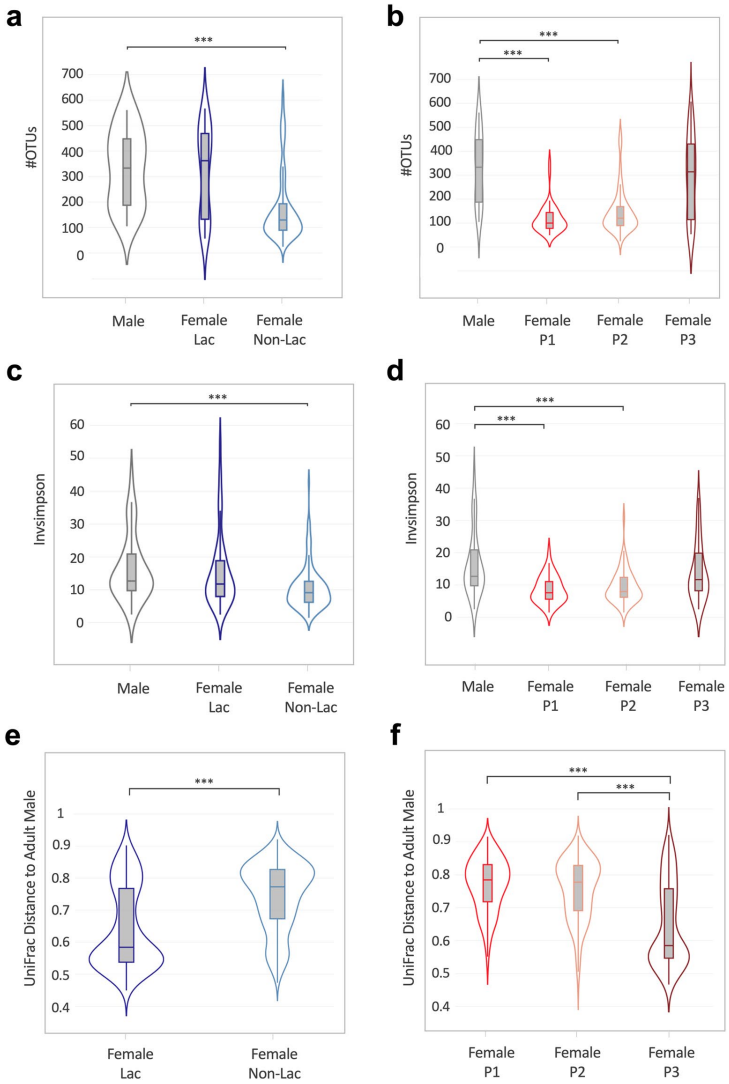


Figure 5. The vaginal microbiota of lactating and menstruating females is more similar to the urethral microbiota. Violin plots of the number of OTUs (**a**) in the male urethra and vaginal microbiota of lactating and non-lactating females and (**b**) in the male urethra and vaginal microbiota of females in three sexual cycle phases (P1: ovulatory phase, P2: intermediate stage, P3: menstruation-like) (Kruskal-Wallis test, ***$p \leq 0.0001$). Violin plots of the inverse Simpson index for (**c**) male urethra and vaginal microbiota of lactating and non-lactating females and (**d**) male urethra and vaginal microbiota of females in three sexual cycle phases (Kruskal-Wallis test, ***$p \leq 0.0001$). (**e,f**) Violin representations showing unweight UniFrac distance of each adult female to the adult male in each group. Data is plotted by (**e**) lactation status (Kruskal-Wallis test, ***$p \leq 0.0001$) and (**f**) sexual cycle phase (Mann-Witney t-test, ***$p \leq 0.0001$). See Table 1 for sample size composition.

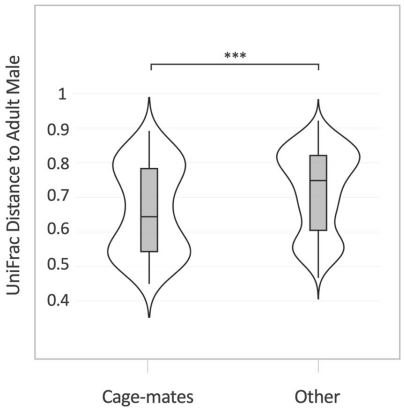


Figure 6. Comparison between the male urethral microbiota and female vaginal microbiota of cage-mates. Violin plots showing unweight UniFrac distance of the adult vaginal microbiota to the urethral microbiota of the adult males. Data is plotted separately for female animals in the same breeding group (cage-mates) and females of other breeding groups (other) (Mann-Witney t-test ***$p \leq 0.0001$). For further subdivision by lactation status and sexual cycle phase see Fig. S7.

altered susceptibility[3,9]. During high levels of estrogen, changes in the vaginal epithelium, including changes in vaginal microenvironment, may have a protective effect[3]. Further investigations are necessary to examine the causal relationship between hormone levels, changes in the NHP vaginal microbiota, and susceptibility to pathogens. However, it has become clear, that a more holistic understanding on host-pathogens interactions is required for the interpretation of animal experiments as host factors can influence the microbiota and vice versa (reviewed by[1]).

We examined the male urethral microbiota of the rhesus monkeys to further compare the genital microbiota of females and males in a single breeding unit. To our knowledge, there has been no studies on the urethral microbiota of wild or captive NHPs to date. Four bacterial phyla, *Firmicutes, Bacteroidetes*, *Proteobacteria*, and *Actinobacteria*, compose the majority of identified sequences in the urethra. On the phylum level, the urethral microbiota of the male rhesus monkeys were similar to that reported in humans with *Firmicutes* making up the largest proportion[23]. In our male animals, notable urethral taxa include *Prevotella*, *Porphyromonas*, and *Ezakiella*, have all been previously associated with the urinary tract microbiota of adult men[23–25]. *Prevotella* has been previously detected in the genital microbiota of healthy female rhesus monkeys indicating that this genus plays a residential role in the rhesus monkeys' genital microbiota[10]. In humans, some species of *Prevotella* have been associated with disease states (e.g., bacterial vaginosis[26]) while other species can be found in clinically healthy women (e.g., post-partum[22]). Identifying the specific species of *Prevotella* and their functional role in NHPs may be important to further understand the vaginal microbiota of NHPs. For the urethral microbiota, it is difficult to compare the prevalence of *Prevotella* in the male rhesus monkey to other studies, as there is currently no consensus on the core urethral microbiota, even in humans[25]. As a result, large scale investigations need to be performed to study the male urogenital microbiota including factors that influence this unique ecosystem in health and disease[25].

It has been hypothesized that sexual exposures can alter the composition of the genital microbiota[13,16,25]. A recent study on sexual partners with bacterial vaginosis (BV), showed that women with BV were significantly more similar to the urogenital microbiota of their partner[17]. To test if sexual contact affected the genital microbiota of NHPs, we first examined if adult, breeding males in captive rhesus monkeys had different urethral microbiota from juvenile males. Breeding groups in this study contained a single adult male, who monopolized the cage-mates in estrus. We found that the age of male rhesus macaques did not shape the urethral microbiota (Fig. 4). This may be due to the fact that juvenile rhesus monkeys already engage in socio-sexual mounting as a form of play[27]. Sexual history in healthy adolescent men has been reported to be a possible determinant of the urogenital microbiota[16]. Known sexually transmitted bacteria and taxa associated with the urethral tract of adult men[23], were observed rarely in adolescent men[16]. To further study the effect of sexual contact, we examined the similarity of the genital microbiota in cage-mates (adult male to adult females in the same breeding group). We found that overall, cage-mates were significantly more similar to each other compared to non-cage-mates (Fig. 6). When subdividing cage-mate by lactation status and sexual cycle, the observed cage-mate effect was not seen for non-lactation and intermediate sexual phase animals (Fig. S7a,b). This may be due to an inappropriate subsampling of these two groups. For example, the intermediate sexual phase classification used in the EVC may represent both, proliferative phase and secretory phase, and is therefore an oversimplification. This highlights the limitation of this cross-sectional study in assessing cage-mate similarities. A controlled temporal study is necessary to examine the effect of sexual contract in NHP breeding groups.

NHPs can be an advantages model to further examine microbiota similarities in sexual partners (in health and disease) as sexual contact is easily observed and controlled.

A surprising finding of our study was that independent of breeding group association, the bacterial composition of lactating monkeys and/or those in the menstruation-like sexual phase were more similar to the urethral microbiota than the non-lactating/non-menstruating animals (Fig. 5). As females in a menstruation-like sexual phase are less attractive to male rhesus monkeys, we presume that the similarity is not caused by recent sexual contact. A possible explanation for this finding is that the altered hormonal state allows otherwise more-suppressed bacteria to dominate the microbiota. To understand the cause of the vaginal microbiota shift towards the male urethra microbiota, controlled temporal experiments in NHPs would be necessary. Interestingly, a temporal study in humans has shown that the vaginal microbiota post-partum shifts towards the gut microbiota[22]. The study was able to show that the shift towards the gut microbiota persisted for multiple months and was independent of delivery method (vaginal vs. caesarean). These findings support the notion that during changes in the genital ecosystem (e.g., shifts in hormones or delivery), the vagina is more susceptible to 'foreign' bacteria. This potentially altered susceptibility should be carefully considered when performing vaginal inoculations in NHPs for future experiments (e.g., HIV).

We found that breeding groups can have an effect on the vaginal microbiota (Fig. 2). Breeding group similarities could be influenced by various factors including host genetics[28], differences in group size or cage effects[29]. Many of these factors could not be properly examined in this study and require planned and controlled animal experiments. In mice, it has been shown that animals kept in the same cage become more similar in microbiota composition over time[29]. This effect could be studied in captive NHPs by examining microbiota changes in various ecological niches (genital, skin, fecal) during cage transfers. A cage effect in NHPs could have major implications for the use of NHPs as translational animal models. A better understanding of the NHP microbiota could therefore refine animal selection for animal experiments where a higher standardization can lead to reduced animal numbers[3,9]. The inclusion of appropriate controls in microbiota studies cannot be stressed enough[30]. Especially low abundance microbiotas like the urethral microbiota are vulnerable to contaminations during sampling and laboratory analysis[31]. The inclusion of blank control samples, especially at the site of sample collection, is essential and should be understood as Good Laboratory and Scientific Practice (Fig. S1). Only well-planned and controlled microbiota studies on NHPs will provide a better understanding of factors that influence microbiotas of NHPs.

Methods

Ethical statement. All samples included into this study were obtained from clinically healthy rhesus monkeys that underwent the mandatory annual health check at the German Primate Center between June 2016 and May 2017. Animals were not purposely immobilized to collect samples for this study. Swabs were taken as part of a routine annual health monitoring and tuberculosis screening. Animal were short-term immobilized by trained veterinarians who checked and documented the general health condition of each individual. Sampling included the collection of blood, oral and genital swab samples. The use of the samples was reviewed and approved by the animal welfare and ethics committee of the German Primate Center (EC No. 1–16). All work steps involving the handling of live animals followed the rules of 'Good Veterinary Practice'.

Study design and animals. Urethral swabs of 37 male and vaginal swabs of 194 female rhesus monkeys were collected. A cross-sectional study design was applied. Samples from apparently pregnant individuals, clinically diseased animals, or animals that received medical treatment within the last 6 months were excluded from analysis. Moreover, we excluded samples from animals below the age of three. Data file S1 provides a detailed overview on the samples analyzed in this study as well as the respective NCBI Sequence Read Archive numbers. Lot numbers for consumables were kept consistent and are reported in the Supplementary Material (Table S2).

Swab sample collection. Immobilized female rhesus monkeys were placed in dorsal recumbency and the area around the vulva was cleaned using 70% ethanol. To facilitate sampling, a sterile silicon tube, 15 mm diameter and 40 mm length, was used to avoid swab contamination with skin or fecal material. A flocked swab (FLOQSwabs, Copan Improve Diagnostics, Brescia, Italy) was moistened using a single drop of sterile physiological saline solution (WDT eG, Garbsen, Germany) and was subsequently inserted midway into the vaginal canal. Subsequently the swab was rotated 20-times on the dorsal wall before it was gently removed and transferred into 500 µl of custom-made lysis buffer (10 mM Tris, pH 8.0, 0.1 M EDTA, pH 8.0 and 0.5% SDS). Samples were kept on ice until transported to the inhouse laboratory facilities where they were stored at −80 °C[30].

An additional swab was collected to perform an EVC. Briefly, the swab was rolled onto a microscope glass slide after which it was allowed to air-dry. Slides were then stained with a Romanowski stain (Diff-Quik) and subsequently examined under the microscope by two independent investigators[19]. Cytological scoring was performed as previously described by McLennan *et al.*[20]. The maturation index was calculated by counting 100 representative epithelial cells, which were scored according to their cell type. Briefly, parabasal cells were assigned a value of 0, intermediate cells a value of 0.5, and superficial cells a value of 1. Based on the cumulative maturation score, the animals were categorized into three stages (ovulatory phase (P1), intermediate phase (P2), and menstruation-like phase (P3); see Supplementary Table 2).

The vaginal pH was measured using a swab which was inserted midway into the vagina and then rolled onto a pH-indicator paper (Merck & Co., Kenilworth, New Jersey). The vaginal pH was scored by two independent researchers following the manufacturer's instructions using a scale ranging from 5.5 to 9.0.

Immobilized male rhesus monkeys were placed in ventral recumbency and sampled for urethral swabs. A minitip FLOQ swabs (Copan Improve Diagnostics) was moistened using sterile physiological saline solution and

subsequently inserted 1–2 cm into the urethra of the animal. Subsequent handling of the samples was identical to the procedure described for vaginal swab samples.

Suitable precautions were taken during sample collection to avoid microbial contamination. As a sample collection control, a FLOQ swab with a single drop of sterile physiological saline solution was immediately transferred into a 500 µl custom-made lysis buffer at the breeding facility at the time of sampling.

DNA extraction. We used the QIAamp DNA Mini Kit (Qiagen GmbH, Hilden, Germany) to extract bacterial DNA. This kit was previously validated for microbial analysis of swab material[30]. Briefly, proteinase K (50 mg/µl) was added and the samples were incubated overnight at 56 °C at 600 rpm (Thermomix comfort, Eppendorf, Hamburg, Germany). Appropriate amounts of AL buffer (Qiagen GmbH) and ethanol were added. The DNA was subsequently purified from the lysate using the spin columns provided in the kit. Extracted DNA was eluted in 75µl Microbial DNA-Free water (Qiagen GmbH). Suitable precautions were taken during sample handling and processing in the laboratory to limit microbial contamination and maintain consistency during all procedures. The order of sample processing was randomized to avoid handling bias. As a laboratory analysis collection control, a FLOQ swab was transferred into a 500 µl custom-made lysis buffer under the DNA extraction bench at the time the rhesus monkey samples were handled.

16S ribosomal RNA gene sequencing. The universal primers 515 F and 806 R, which were adapted with linker regions and barcode sequences, were used to amplify the V4 region of the 16S ribosomal RNA (16S rRNA) gene[32]. Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, Massachusetts), which has been previously validated for the use in microbiota studies[30], was used to amplify each sample in triplets. PCR reactions consisted of 12.5µl of 2x PCR master mix, 8µl of Microbial DNA-Free water (Qiagen GmbH), 1.25µl of each primer (0.5 mM each, Metabion, Steinkirchen, Germany) and 2µl of template in a total reaction volume of 25µl. PCR cycling conditions comprised of a pre-denaturation step of 30 s at 98 °C, followed by 30 cycles of 98 °C for 10 s, 55 °C for 15 s and 72 °C for 60 s, as well as a final 10 min extension step at 72 °C. A blank control (Microbial DNA-Free water) and a mock control sample (Microbial mock community, HM-280, Biodefense and Emerging Infectious Research (BEI) Resources, Manassas, Virginia) were included in 16S rRNA gene amplification. The amplicon triplets were pooled, purified using 0.7x AMPure XP beads (Beckman Coulter, Brea, California), and quantified using the Qubit 3.0 Fluorometer (Thermo Fisher Scientific). Subsequently, we verified the amplicon integrity for a representative number of eleven samples using the BioAnalyzer 2000 (Agilent, Santa Clara, California). Equimolar amounts (10 nM) of sample amplicon and maximum volume of control samples (5µl) were pooled prior to sequencing. Illumina MiSeq. 2 × 250 bp paired-end sequencing (V2 chemistry, Illumina, San Diego, California) was performed in the Transcriptome and Genome Analysis Laboratory at the University of Goettingen in accordance with published guidelines[32].

Data processing and analysis. The sequencing reads were processed using the mothur software package (v.1.39.5)[33]. According to the MiSeq SOP[33], contigs were assembled, sequences were quality filtered, and PCR artifacts were removed. The SILVA bacterial reference database[34] was used to align the sequences and OTUs were assigned based on 97% sequence similarity. Cross-sample singletons and poorly aligned sequences were removed. The seq.error command was used to determine the error rate and the mock community was eliminated from the dataset. Due to low read numbers, control sample reads were excluded from the dataset and analyzed separately.

To examine differences in the microbial community structure, alpha (species richness within a single sample) and beta diversities (microbial community diversity between samples) were calculated. As alpha diversity measurements, we determined the number of observed OTUs and calculated the inverse Simpson Metrix using the summary.single command in mothur. Beta diversity was determined using unweighted UniFrac metrics[35]. The dissimilarity matrix was visualized using Principal Coordinates Analysis (PCoA) and a Newick formatted dendrogram (visualized in FigTree v.1.4.2, http://tree.bio.ed.ac.uk/software/figtree/). ClustVis tool (https://biit.cs.ut.ee/clustvis/) was used to create a heatmap of the relative abundance of bacterial taxa[36]. Violin plots (R package plot.ly) and box plots (GraphPad Prism 6) were used to visualize data points for different variables.

Statistical analysis. The statistical significance of the pooled data was analyzed in GraphPad Prism 6 (GraphPad software) and in R (v3.4.3;[37]) using the package vegan (version 2.5;[38]). Whenever appropriate, we tested for normality distribution of the data using the Kolmogorov-Smirnov normality test. The significance in alpha diversity and pair-wise beta diversity between two or more groups was tested using the non-parametric Mann-Whitney-U or Kruskal-Wallis tests including correction for multiple testing using Dunn's post hoc tests. Differences in community structure based on age of animals, group association, and lactation status was tested using analysis of molecular variance (AMOVA, 1,000 permutations) in mothur[39]. PCoA plots of unweighted UniFrac metrics and UPGMA-clustered dendrograms (unweighted UniFrac metrics) were used to visualize data points. Differences in the ten most abundant OTUs in vaginal samples were assessed using the metastats command in mothur[40]. p-values for differences in individual OTUs were corrected for multiple comparisons using Bonferroni correction. Values of $p \leq 0.05$ were considered statistically significant.

Data availability

All generated read files have been deposited in the NCBI Sequence Read Archive under the accession number SRP184988. Detailed information about the samples is provided in Data file S1 in the supplemental material.

Received: 18 April 2019; Accepted: 7 November 2019;
Published online: 22 November 2019

References

1. Wang, B., Yao, M., Lv, L., Ling, Z. & Li, L. The human microbiota in health and disease. *Engineering* **3**, 71–82 (2017).
2. Spear, G. T. *et al.* Identification of rhesus macaque genital microbiota by 16S pyrosequencing shows similarities to human bacterial vaginosis: implications for use as an animal model for HIV vaginal infection. *AIDS Res. Hum. Retroviruses* **26**, 193–200, https://doi.org/10.1089/aid.2009.0166 (2010).
3. Smith, S. M., Baskin, G. B. & Marx, P. A. Estrogen protects against vaginal transmission of simian immunodeficiency virus. *J. Infect. Dis.* **182**, 708–715 (2000).
4. Haus, T. *et al.* Genome typing of nonhuman primate models: implications for biomedical research. *Trends Genet.* **30**, 482–487, https://doi.org/10.1016/j.tig.2014.05.004 (2014).
5. Bielert, C. *et al.* Mating in the rhesus monkey (*Macaca mulatta*) after conception and its relationship to oestradiol and progesterone levels throughout pregnancy. *J. Reprod. Infertil.* **46**, 179–187 (1976).
6. Gordon, T. P. Reproductive behavior in the rhesus monkey: social and endocrine variables. *Am. Zool.* **21**, 185–195 (1981).
7. Wilson, M., Walker, M., Pope, N. & Gordon, T. Prolonged lactational infertility in adolescent rhesus monkeys. *Biol. Reprod.* **38**, 163–174 (1988).
8. King, B. F. Ultrastructure of the nonhuman primate vaginal mucosa: epithelial changes during the menstrual cycle and pregnancy. *J. Ultrastruct. Res.* **82**, 1–18 (1983).
9. Marx, P. A. *et al.* Progesterone implants enhance SIV vaginal transmission and early virus load. *Nat. Med.* **2**, 1084–1089 (1996).
10. Chen, Z. *et al.* Diversity of macaque microbiota compared to the human counterparts. *Sci. Rep.* **8**, 15573, https://doi.org/10.1038/s41598-018-33950-6 (2018).
11. Yildirim, S. *et al.* Primate vaginal microbiomes exhibit species specificity without universal *Lactobacillus* dominance. *ISME J.* **8**, 2431–2444, https://doi.org/10.1038/ismej.2014.90 (2014).
12. Miller, E. A., Livermore, J. A., Alberts, S. C., Tung, J. & Archie, E. A. Ovarian cycling and reproductive state shape the vaginal microbiota in wild baboons. *Microbiome* **5**, 8, https://doi.org/10.1186/s40168-017-0228-z (2017).
13. Gajer, P. *et al.* Temporal dynamics of the human vaginal microbiota. *Sci. Transl. Med* **4**, 132ra52, https://doi.org/10.1126/scitranslmed.3003605 (2012).
14. Breshears, L. M., Edwards, V. L., Ravel, J. & Peterson, M. L. *Lactobacillus crispatus* inhibits growth of *Gardnerella vaginalis* and *Neisseria gonorrhoeae* on a porcine vaginal mucosa model. *BMC Microbiol.* **15**, 276, https://doi.org/10.1186/s12866-015-0608-0 (2015).
15. Uchihashi, M. *et al.* Influence of age, reproductive cycling status, and menstruation on the vaginal microbiome in baboons (*Papio anubis*). *Am. J. Primatol.* **77**, 563–578, https://doi.org/10.1002/ajp.22378 (2015).
16. Nelson, D. E. *et al.* Bacterial communities of the coronal sulcus and distal urethra of adolescent males. *PloS One* **7**, e36298, https://doi.org/10.1371/journal.pone.0036298 (2012).
17. Zozaya, M. *et al.* Bacterial communities in penile skin, male urethra, and vaginas of heterosexual couples with and without bacterial vaginosis. *Microbiome* **4**, 16, https://doi.org/10.1186/s40168-016-0161-6 (2016).
18. Farage, M. A., Miller, K. W. & Sobel, J. D. Dynamics of the vaginal ecosystem—hormonal influences. *Infect. Dis. Res. Treat.* **3**, 1–15, https://doi.org/10.4137/IDRT.S3903 (2010).
19. Knauf, S., Batamuzi, E., Mätz-Rensing, K., Leendertz, F. & Wehrend, A. Exfoliative vaginal cytology: a diagnostic tool for sexual cycle stages in nonhuman primates. *Tanzania Veterinary Journal* **26**, 1–6 (2009).
20. MacLennan, A. H. & Wynn, R. M. Menstrual cycle of the baboon: I. clinical features, vaginal cytology and endometrial histology. *Obstet. Gynecol.* **38**, 350–358 (1971).
21. Weinbauer, G. F. *et al.* Physiology and endocrinology of the ovarian cycle in macaques. *Toxicol. Pathol.* **36**, 7S–23S, https://doi.org/10.1177/0192623308327412 (2008).
22. DiGiulio, D. B. *et al.* Temporal and spatial variation of the human microbiota during pregnancy. *Proc. Natl. Acad. Sci. USA* **112**, 11060–11065, https://doi.org/10.1073/pnas.1502875112 (2015).
23. Nelson, D. E. *et al.* Characteristic male urine microbiomes associate with asymptomatic sexually transmitted infection. *PloS One* **5**, e14116, https://doi.org/10.1371/journal.pone.0014116 (2010).
24. Sabat, A. J. *et al.* Targeted next-generation sequencing of the 16S-23S rRNA region for culture-independent bacterial identification-increased discrimination of closely related species. *Sci. Rep.* **7**, 3434, https://doi.org/10.1038/s41598-017-03458-6 (2017).
25. Whiteside, S. A., Razvi, H., Dave, S., Reid, G. & Burton, J. P. The microbiome of the urinary tract—a role beyond infection. *Nat. Rev. Urol.* **12**, 81, https://doi.org/10.1038/nrurol.2014.361 (2015).
26. Pybus, V. & Onderdonk, A. B. Evidence for a commensal, symbiotic relationship between *Gardnerella vaginalis* and *Prevotella bivia* involving ammonia: Potential significance for bacterial vaginosis. *J. Infect. Dis.* **175**, 406–413 (1997).
27. Brown, G. R. & Dixson, A. F. The development of behavioural sex differences in infant rhesus macaques (*Macaca mulatta*). *Primates* **41**, 63–77 (2000).
28. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014).
29. McCafferty, J. *et al.* Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model. *ISME J.* **7**, 2116–2125, https://doi.org/10.1038/ismej.2013.106 (2013).
30. Hallmaier-Wacker, L. K., Lueert, S., Roos, C. & Knauf, S. The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Sci. Rep.* **8**, 6292, https://doi.org/10.1038/s41598-018-24573-y (2018).
31. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87, https://doi.org/10.1186/s12915-014-0087-z (2014).
32. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
33. Schloss, P. D. *et al.* Introducing mothur: Open-source, platform-independent, community- supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
34. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
35. Lozupone, C. & Knight, R. UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
36. Metsalu, T. & Vilo, J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **43**, W566–W570 (2015).
37. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available from, https://www.R-project.org (2018).
38. Oksanen, J., *et al.* vegan: Community ecology package. Available from, https://CRAN. R-project.org/package=vegan (2018).
39. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
40. White, J. R., Nagarajan, N. & Pop, M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* **5**, e1000352, https://doi.org/10.1371/journal.pcbi.1000352 (2009).

Acknowledgements
The authors thank the Biodefense and Emerging Infectious Research (BEI) Resources, NIAID, NIH for providing the cells from Microbial Mock Community (Even, HM-280) as part of the Human Microbiome Project. We thank Tamara Becker, Annette Schrod, Annette Husung, Melina Urh, and all the animal caretakers for their assistance during sample collection. Additionally, we thank Uwe Schönmann and Dietmar Zinner (Deutsches Primatenzentrum GmbH) for their guidance and general support. We thank the staff of the Transcriptome and Genome Analysis Laboratory at the Georg-August-University of Goettingen for their assistance in optimizing the sequencing run.

Author contributions
L.H.W., C.R. and S.K. designed the study. Sample collection was performed by L.H.W., S.L. and S.K. Laboratory work was conducted at the German Primate Center and performed by L.H.W. and S.L. Data were analyzed by L.H.W. and S.K. All authors (L.H.W., S.L., C.R. and S.K.) contributed to the manuscript preparation.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-53976-8.

Correspondence and requests for materials should be addressed to S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter B3

Supplemental information:

Figure S1. Bacterial profile of control samples included in this study

Figure S2. Individual bacterial OTUs are significantly different in the vaginal microbiota of lactating and menstruating female monkeys

Figure S3. Vaginal pH was significantly different in lactating and menstruating animals

Figure S4. Alpha diversity measurements for vaginal microbiota of rhesus monkeys grouped by age composition

Figure S5. Principal coordinates analysis of vaginal samples separated by (a) age comparison and (b) group association

Figure S6. Alpha diversity measurements for urethral microbiota of male rhesus monkeys grouped by group composition

Figure S7. Cage-mate comparison between adult male and female genital microbiota subdivided by lactation status and sexual cycle phase

Table S1. Maturating value and phase classification for exfoliative vaginal cytology used in this study

Table S2. Reagents and kits used in this study with lot numbers and suppliers

Supplementary Figures:

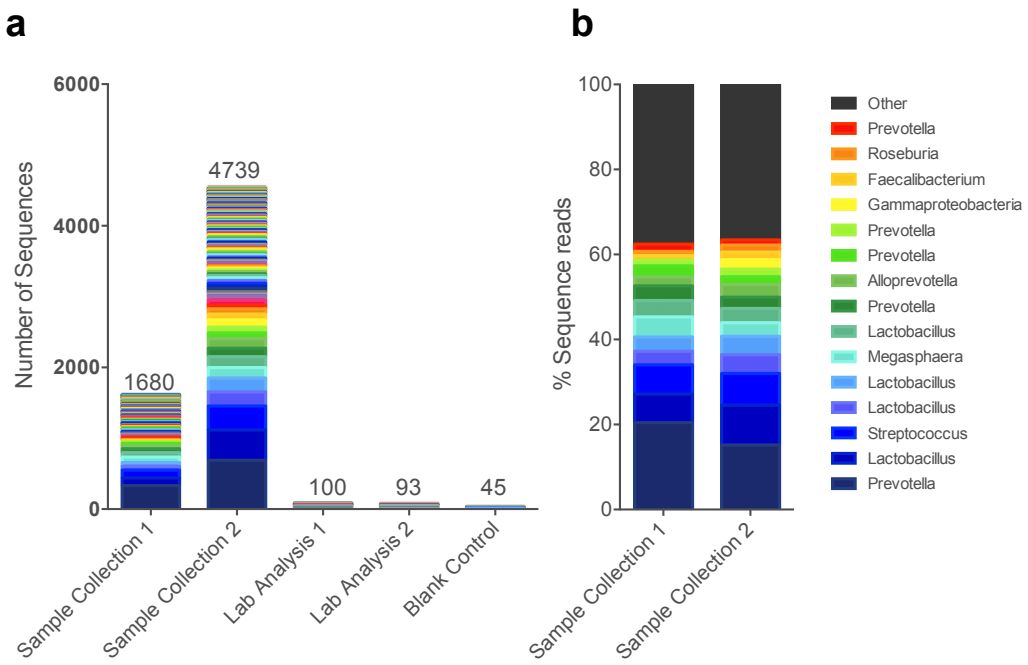


Fig. S1: Bacterial profile of control samples included in this study. (**a**) A total of 6,657 sequences corresponded to the control samples included during sample collection, lab analysis and 16S rRNA amplification (blank control). (**b**) Taxa plots showing the relative abundance of the 15 most abundant OTUs in percentage of reads for control samples taken during sample collection at two different breeding groups (sample collection 1: group A and sample collection 2: group E).

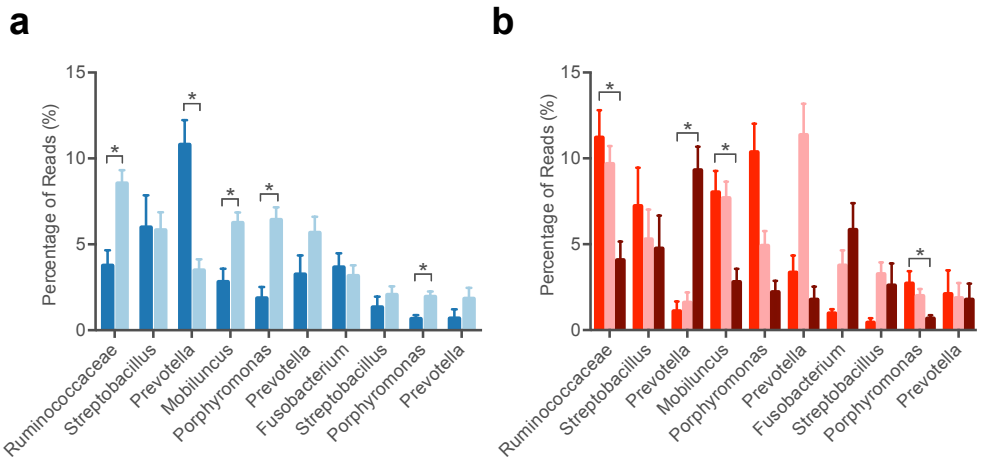


Fig. S2: Individual bacterial OTUs are significantly different in the vaginal microbiota of lactating and menstruating female monkeys. Percentage of sequence reads (mean ± SD) for the ten most abundant OTUs in the vaginal microbiota sorted by (**a**) lactation status (lactating (dark blue) and non-lactating (light blue) and (**b**) sexual cycle phases (P1: ovulatory phase (red), P2: intermediate stage (pink), P3: menstruation-like (dark red)). Differences in the ten most abundant OTUs were assessed using the metastats command in mothur (*p≤0.001; For (**b**) significance is shown only if P3 was significantly different for both P1 and P2).

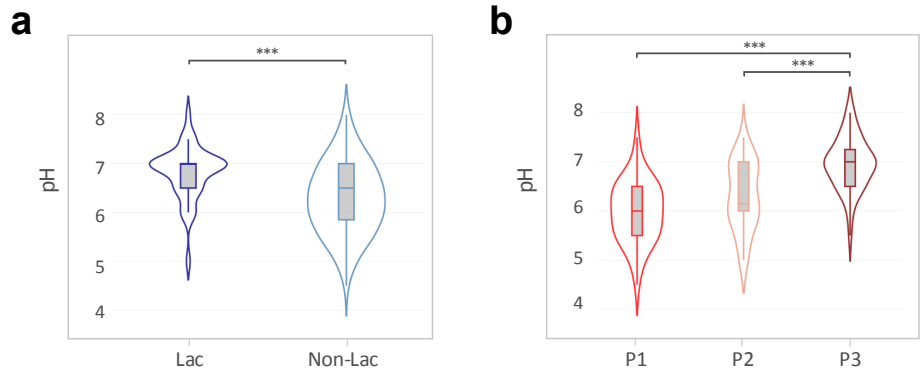


Fig. S3: Vaginal pH was significantly different in lactating and menstruating animals. Violin plots of the vaginal pH grouped by (**a**) lactation status (Mann-Witney t-test ***p≤0.0001) and (**b**) sexual cycle phases (P1: ovulatory phase, P2: intermediate stage, P3: menstruation-like; Kruskal-Wallis test ***p≤0.0001).

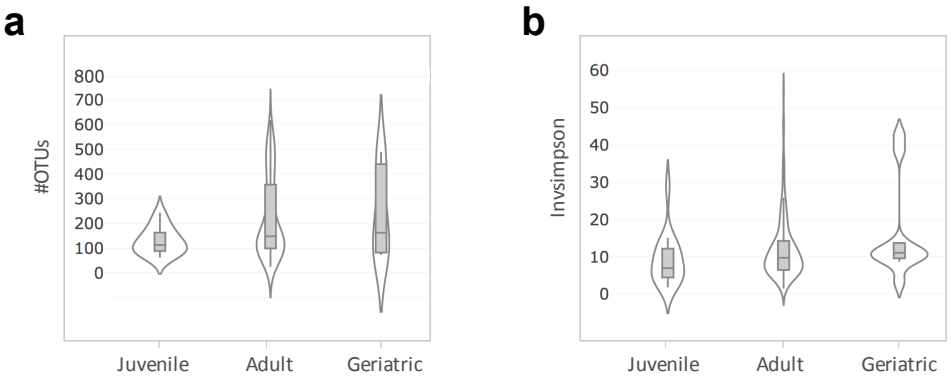


Fig. S4: Alpha diversity measurements for vaginal microbiota of rhesus monkeys grouped by age composition. Violin plots of the (**a**) observed OTUs and (**b**) InvSimpson index (Kruskal-Wallis test). See Table 1 for sample size composition and age classification.

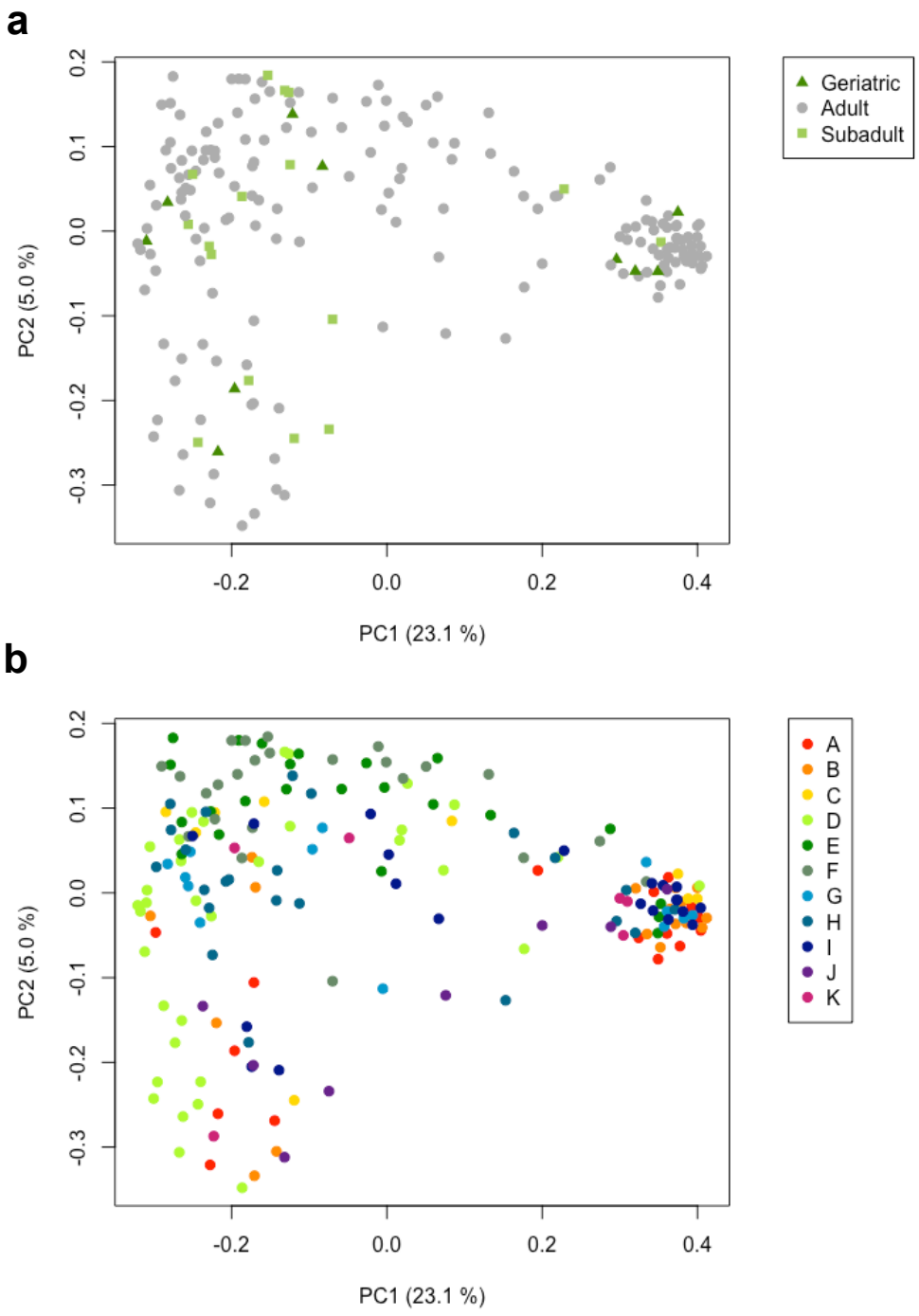


Fig. S5: Principal coordinates analysis of vaginal samples separated by (a) age comparison and (b) group association. Distances between samples were calculated using the unweighted UniFrac metrics. Figure legend shows age groups and group association (groups: A-K). See Table 1 for sample size composition and age classification.

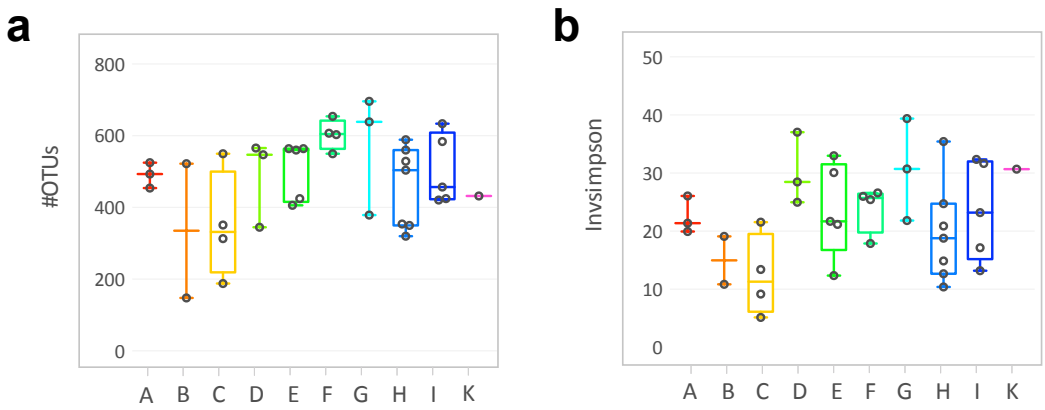


Fig. S6: Alpha diversity measurements for urethral microbiota of male rhesus monkeys grouped by group composition. Boxplots (median ± range) of the (**a**) observed OTUs and (**b**) InvSimpson index clustered.

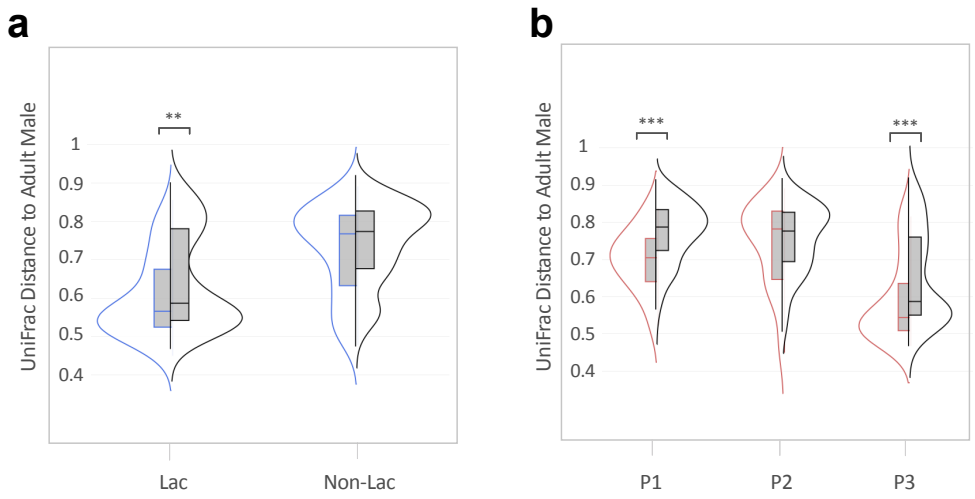


Fig. S7: Cage-mate comparison between adult male and female genital microbiota subdivided by lactation status and sexual cycle phase. Paired violin representations showing unweight UniFrac Distance of the adult vaginal microbiota to the urethral microbiota of the adult males. Data is sorted by (**a**) lactating status and (**b**) sexual cycle phase (P1: ovulatory phase, P2: intermediate stage, P3: menstruation-like). The colored (blue/red) violin plots correspond to the unweight UniFrac distance between the adult male and females in the same cage, while black plots correspond to the unweighted UniFrac distance between the alpha male and females from other breeding groups (no sexual contact possible) (Mann-Witney t-test ***p≤0.0001).

Supplementary Tables:

Table S1: Maturating value and phase classification for exfoliative vaginal cytology used in this study

Phase Assignment	Maturation value	Sexual cycle phases	Cell composition	% Lac (n= 44)	% Non-Lac (n=104)
Phase 1	95-100	Ovulatory	Dominated by superficial cells	15.9%	35.6%
Phase 2	45-95	Intermediate	Intermediate, parabasal & basal cells	31.8%	41.3%
Phase 3	>45	Menstruation-like*	Diverse including cellular debris & blood cells (erythrocytes and leucocytes)	52.3%	23.1%

*Menstruation-like: this phase represents both menstruation and postpartum amenorrhea

Table S2: Reagents and kits used in this study with lot numbers and suppliers

Reagent	Supplier	Lot Number
FLOQSwabs, regular	Copan Improve Diagnostics	F5RM00
FLOQ swabs, mini-tip	Copan Improve Diagnostics	8H0D00
Microtube SafeSeal, 1.5 ml	Sarstedt AG & Co.	7080311
DNA LoBind Tubes, 1.5 ml	Eppendorf AG	G171343G
DNA LoBind Tubes, 2.0 ml	Eppendorf AG	D157963O
The QIAamp Mini Kit	Qiagen GmbH	157033520
Microbial DNA-Free water	Qiagen GmbH	JE01
Phusion HS II HF DNA Polymerase	Thermo Fisher Scientific	00607540
AMPure XP beads	Beckman Coulter	16909300

Chapter B4

A Metataxonomic Tool to Investigate the Diversity of *Treponema*

Luisa K. Hallmaier-Wacker[1,2], Simone Lüert[1,2], Sabine Gronow[3], Cathrin Spröer[3], Jörg Overmann[3], Rebecca J. Vaughan-Higgins[4], Sascha Knauf[1]

Frontiers in Microbiology **10**: 2094 (2019) doi: 10.3389/fmicb.2019.02094

[1]Neglected Tropical Diseases Work Group, Infection Biology Unit, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[2]Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[3]Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

[4]Department of Microbiology, Braunschweig University of Technology, Braunschweig, Germany

[5]Animal Pathology – Bacteriology Laboratory, Department of Primary Industries and Regional Development, South Perth, WA, Australia

[6]Department of Conservation Medicine, College of Veterinary Medicine, School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia

---

Author contributions:

Conceived and designed the study: LHW, SK
Conduced laboratory work: LHW, SL, SG, CS, SK
Analyzed the data: LHW, SL, SK
Contributed DNA from the Gilbert's potoroo: NB, RVH
Contributed DNA samples of spirochetes for the mock sample: SG, CS, JO
Manuscript preparation: LHW, SL, SG, CS, JO, RVH, SK

frontiers in Microbiology

METHODS
published: 10 September 2019
doi: 10.3389/fmicb.2019.02094

A Metataxonomic Tool to Investigate the Diversity of *Treponema*

Luisa K. Hallmaier-Wacker[1,2], Simone Lüert[1,2], Sabine Gronow[3], Cathrin Spröer[3], Jörg Overmann[3,4], Nicky Buller[5], Rebecca J. Vaughan-Higgins[6] and Sascha Knauf[1†]*

[1] Neglected Tropical Diseases Work Group, Infection Biology Unit, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany, [2] Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany, [3] Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, [4] Department of Microbiology, Braunschweig University of Technology, Braunschweig, Germany, [5] Animal Pathology – Bacteriology Laboratory, Department of Primary Industries and Regional Development, South Perth, WA, Australia, [6] Department of Conservation Medicine, College of Veterinary Medicine, School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia

OPEN ACCESS

Edited by:
Yasir Muhammad,
King Abdulaziz University,
Saudi Arabia

Reviewed by:
Naruo Nikoh,
The Open University of Japan, Japan
Yuki Chan,
The University of Hong Kong,
Hong Kong

***Correspondence:**
Sascha Knauf
sknauf@dpz.eu

†Present address:
Sascha Knauf,
Division of Microbiology and Animal Hygiene, University of Göttingen,
Göttingen, Germany

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 13 June 2019
Accepted: 26 August 2019
Published: 10 September 2019

Citation:
Hallmaier-Wacker LK, Lüert S, Gronow S, Spröer C, Overmann J, Buller N, Vaughan-Higgins RJ and Knauf S (2019) A Metataxonomic Tool to Investigate the Diversity of Treponema. Front. Microbiol. 10:2094. doi: 10.3389/fmicb.2019.02094

The genus *Treponema* contains a number of human and animal pathogenic as well as symbiotic bacteria that are found in vastly different anatomical and environmental habitats. Our understanding of the species range, evolution, and biology of these important bacteria is still limited. To explore the diversity of treponemes, we established, validated, and tested a novel metataxonomic approach. As the informative nature of the hypervariable regions of the 16S rRNA gene differ, we first analyzed each variable region independently. Considering the *in silico* results obtained, we established and validated the sequencing of the V4-region of the 16S rRNA gene using known mixtures of *Treponema* species as well as a selected number of clinical samples. The metataxonomic approach was able to identify *Treponema* to a near-species level. We demonstrate that using a spirochete-specific enrichment, our method is applicable to complex microbial communities and large variety of biological samples. The metataxonomic approach described provides a useful method to unravel the full diversity and range of *Treponema* in various ecosystems.

Keywords: metagenomics, metataxonomics, one health, spirochete, 16S rRNA, *Treponema*, marsupial, *Potorous*

INTRODUCTION

Spirochaetes, a phylum of spiral-shaped bacteria, range from pathogenic (e.g., *Treponema pallidum*) to symbiotic (e.g., *Sphaerochaeta coccoides*) to free-living (e.g., *Exilispira thermophile*) species (Paster, 2001). The ability of spirochetes to inhabit vastly anatomical and ecological habitats is remarkable and indicates a high diversity of the bacterial members of this phylum (Paster, 2001). Until recently, spirochetes were predominantly discovered and subsequently characterized using cultivation, microscopical, or serological approaches. These techniques make it difficult and sometimes impossible to characterize not-yet-cultivated species, to identify species in multiple-spirochete infections, or to discover commensal microbes. The advent of cultivation-independent molecular techniques (e.g., nucleic acid amplification technology) has allowed for a broader detection of *Treponema* in different biological niches. To date, the 16S rRNA phylogenetic marker gene has been particularly instrumental in the detection of *Treponema* diversity (Pace, 1997). Based on defined similarity thresholds the 16S rRNA sequences can be grouped into phylotypes. For

example, using a clonal 16S rRNA gene library and subsequent Sanger sequencing, the termite (*Reticulitermes flavipes*) gut was found to harbor more than 67 different treponemal phylotypes (Lilburn et al., 1999) and the human oral cavity up to 23 different treponemal clusters (Choi et al., 1994).

More recently, 16S rRNA gene-based metataxonomic studies have moved from clonal libraries to high-throughput sequencing approaches. Single hypervariable regions of the 16S rDNA have been used to examine different microbiomes (Kozich et al., 2013) and have identified treponemes in many ecological niches (Hong et al., 2012; Klitgaard et al., 2014; Rodriguez-R et al., 2015; Clayton et al., 2018; Hicks et al., 2018). For example, the gut microbiome of wild western lowland gorillas (*Gorilla gorilla gorilla*) (Hicks et al., 2018) and other nonhuman primates (Clayton et al., 2018) harbors multiple operational taxonomic units (OTUs) corresponding to the genus *Treponema*. Yet, conventional data analysis pipelines used in microbiome studies still do not allow for species-level characterization (Schloss et al., 2009; Caporaso et al., 2010). Taxonomic classification for many bacterial genera is restricted by the limited sequence differences in the 16S rRNA gene (Wang et al., 2007). Rossi-Tamisier et al. (2015) showed that spirochetes, in particular *Treponema* and *Spirochaeta*, have an exceptionally large variability in the 16S rRNA gene. For *Treponema*, only 2.1% of the analyzed 16S rRNA sequences fell within the recommended similarity threshold (95–98.7%) (Rossi-Tamisier et al., 2015).

To explore the range and diversity of *Treponema*, we established, validated, and tested a newly designed spirochete-specific metataxonomic approach that utilizes the 16S rRNA gene. Based on the known variability of the 16S rRNA gene, we hypothesized that a single hypervariable region of this gene provides a good target for a metagenomics-based assay to examine the diversity of *Treponema* in various biological sample types.

MATERIALS AND METHODS

In silico Analysis

Based on the nomenclature of Bergey's Manual of Systematic Bacteriology, we selected all bacterial species that are classified within the phylum *Spirochaetes* (Paster, 2001). Subsequently, a representative 16S rRNA gene sequence corresponding to each *Spirochaetes* bacterial species was retrieved from the GenBank database[1]. Where possible, sequences were chosen with maximal length and no ambiguous bases. Sequences shorter than 1,250 base pairs and/or containing more than two ambiguous bases were not included in the dataset even if no other sequence of the bacterial species was available (**Supplementary Table S1**).

The Perl-based high-throughput software tool V-Xtractor was used to locate the hypervariable regions (V2–V8) of the 16S rRNA sequences using Hidden Markov Models (Hartmann et al., 2010). Subsequently, the sequences of each variable region were analyzed using the mothur software package (v.1.41.1) (Schloss et al., 2009). In an initial step, identical sequences were removed using the unique.seq command. Then, the SILVA bacterial reference database (Quast et al., 2012) was utilized to align the sequences [align.seqs command using kmer searching (8mers) and Needleman–Wunsch pairwise alignment method]. OTU clustering was performed for distance threshold ranging from 0.01 to 0.10 at increments of 0.01 (cluster.split command with the OptiClust algorithm) (Westcott and Schloss, 2017).

[1]https://www.ncbi.nlm.nih.gov

Spirochete Mock Community

The spirochete mock community comprised an equal mixture of 19 strains of the phylum *Spirochaetes*. Single bacterial DNA isolates were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). DNA from rabbit inoculated *T. pallidum* subsp. *pertenue* strain Gauthier (referred to as *T. pallidum* throughout the manuscript) was obtained from David Šmajs, Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic. The 19 *Spirochaetes* species which were used in this study, including the cultivation method, DSMZ reference number, 16S rRNA gene copy number, genome size, and NCBI reference, are shown in **Supplementary Table S2**.

The DNA of the cultured spirochetes obtained from DSMZ was quantified using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific). *T. pallidum* DNA, due to the rabbit background DNA from *in vivo* inoculation experiments, was quantified using an established TaqMan PCR (qPCR) targeting the *polA* gene with slight modifications as described previously (Knauf et al., 2018). Based on the DNA content, genome size and 16S rRNA gene copy number, the 19 spirochetes were mixed together at equimolar (even) ribosomal RNA operon counts per organism. The final spirochete mock community contained 100,000 16S rDNA copies/µl of each species. All dilutions were made using Microbial DNA-Free water (Qiagen GmbH). Suitable precautions were taken during all sample handling and processing to avoid microbial contamination.

Treponema Mock Communities

In addition to the spirochete mock community, we created three bacterial DNA validation sets to evaluate the intra-metagenomic assays performance. *T. pallidum* DNA was quantified using TaqMan PCR as described above. For the first validation set, the stock of *T. pallidum* (50,000 16S rRNA copies) was used to make a 10-fold dilution series. The dilutions of *T. pallidum* DNA were subsequently mixed with bacterial DNA contained no *Spirochaetes* [Microbial Mock Community, HM-280, Biodefense and Emerging Infectious Research (BEI) Resources, Manassas, VA, United States] (**Supplementary Table S3**). The second validation set was a mixture of *T. pallidum* and *T. denticola* in different ratios (**Supplementary Table S4**). The final ratios of the *T. pallidum* to *T. denticola* were 1:100, 1:10, 1:1, 10:1, and 100:1. The third validation set was a 10-fold serial dilution series of *T. pallidum* starting at 50,000 copies of 16S rRNA gene. Dilutions for all validation sets were made using Microbial DNA-Free water (Qiagen GmbH).

Spirochete 16S Ribosomal RNA Gene Enrichment

Spirochete-selective primers were used to enrich spirochetal DNA (Dewhirst et al., 2010). The primers F24 (5′-GAGTTTGATYMTGGCTCAG-3′) and M98 (5′-GTTACGACTTCACCCYCCT-3′) were used to amplify a ∼1,450 bp fragment of the 16S rRNA gene covering the V1–V9 region. This first PCR step was performed in triplicates using the Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific), which has been validated for the use in microbiome studies (Hallmaier-Wacker et al., 2018). PCR reactions consisted of 12.5 μl of 2× PCR master mix, 9.5 μl of Microbial DNA-Free water (Qiagen GmbH), 1.0 μl of each primer (0.5 mM each, Metabion), and 1 μl of template in a total reaction volume of 25 μl. PCR cycling conditions comprised of a pre-denaturation step of 30 s at 98°C, followed by either 20 or 35 cycles of 98°C for 10 s, 57°C for 15 s and 72°C for 120 s, and a final 10 min extension step at 72°C. A 16S rRNA amplification control sample (blank controls; Microbial DNA-Free water) was included. Subsequently, PCR triplicates were pooled before library preparation.

Analysis of the V4 Region of the 16S rRNA Gene After an Initial Spirochete-Specific Amplification Step

A pre-test to re-amplify the V3, V4, and V6 regions was performed to identify the most suitable variable regions. The V4 region was selected, as the V3 and V6 region primers demonstrated technical issues to evenly amplify the variable regions. A modular, two-step PCR process was used to specifically re-amplify the V4-region of the 16S rRNA gene and prepare the samples for sequencing on the MiSeq platform. In the first step, the V4 region of the 16S rRNA gene was amplified using TruSeq adaptor-tailed universal primers 515F and 806R. The primers 515F-TruSeq (5′-ACACTCTTTCCCTCCACGACGCTCTTCCGCTCTGTGTGCCAGCMGCCGCGGTAA-3′) and 806R-TruSeq (5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCCGGACTACHVGGGTWTCTAAT-3′) were composed of the universal primer targeting the V4 region (Caporaso et al., 2011) followed by a linker and the TruSeq adaptor (Illumina, Inc.). Amplification was performed in triplicates and each 25.0 μl reaction contained 1.0 μl of PCR product of the enrichment step, 12.5 μl of 2× Phusion Hot Start II High-Fidelity PCR Master Mix (Thermo Fisher Scientific), 9.5 μl of Microbial DNA-Free water (Qiagen GmbH), and 1.0 μl of each V4-targeting 16S primer (0.5 mM each, Metabion). The cycling conditions were as follows: a pre-denaturation step of 30 s at 98°C, followed by 20 cycles of 98°C for 10 s, 55°C for 15 s and 72°C for 60 s, and a final 10 min extension step at 72°C. To monitor contamination, the blank control of the enrichment step was included as a 16S rDNA amplification control.

In the second-step PCR reaction, sample-specific Illumina indices and flow cell adapters were added in an indexing reaction. Illumina i7 and i5 indices were added to each amplicon using the indexing primer P5 (5′-AATGATACGGCGACCACCGAG ATCTACAC-[i5-INDEX]-ACACTCTTTCCCTACACGACGCTC-3′) and indexing primer P7 (5′-CAAGCAGAAGACGGCATACGAGAT-[i7-INDEX]-GTGACTGGAGTTCAGAC GTGT-3′). Amplification was performed in a 50.0 μl reaction containing 2.0 μl of PCR product from the first-step, 25.0 μl of 2× KAPA HiFi HotStart ReadyMix (KAPA Biosystems), 21.0 μl of Microbial DNA-Free water (Qiagen GmbH), and 1.0 μl of each Truseq index primer (0.5 mM each, Metabion). The cycling conditions were as follows: a pre-denaturation step of 3 min at 98°C, followed by eight cycles of 98°C for 20 s, 62°C for 30 s and 72°C for 30 s, and a final 5 min extension step at 72°C. To monitor overall contamination, the blank control of the first-step PCR reaction was included as a 16S rRNA gene amplification control.

V4-Region 16S rDNA Amplification Without an Initial Spirochete-Specific Amplification Step

For comparison the initial enrichment PCR was not performed on a sample of the spirochete mock community and a sample of validation set 1 (5,000 16S rDNA copies of *T. pallidum*). For these two samples the first-step V4-targeting PCR reaction was performed for additional 15 cycles (total of 35 cycles). A 16S rRNA gene amplification control was included for this altered procedure. All other conditions were kept identical.

Applications to Clinical Samples

The applicability of the metataxonomic approach was tested on extracted DNA from genital swabs of Gilbert's potoroo (*Potorous gilbertii*), a small marsupial found in Western Australia (Vaughan et al., 2009). For more information on sample processing see the **Supplementary Materials**.

MiSeq Library Preparation and Pooling

After the indexing reaction, all amplicons were purified using 0.7× AMPure XP beads (Beckman Coulter), and quantified using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific) (**Supplementary Table S5**). The amplicon integrity was verified for a representative number of four samples using the BioAnalyzer 2000 (Agilent). Equimolar amounts (2 nM) of sample amplicons were pooled. For samples with <2 nM concentration, the maximum volume (5 μl) was pooled prior to sequencing. The Transcriptome and Genome Analysis Laboratory at the University of Goettingen performed the Illumina MiSeq 2 × 250 bp paired-end sequencing (Illumina V2 chemistry) run.

Data Processing and Analysis

Raw reads were processed using the mothur software package (version 1.41.1) (Schloss et al., 2009). Initial pre-processing and quality control were performed in accordance with the MiSeq SOP (Schloss et al., 2009). Briefly, paired-end reads were assembled using the make.contigs command. Subsequently, the screen.seqs command was used to trim sequences and filter out any sequences with ambiguous base calls. Identical trimmed sequences (unique.seq command) were aligned (align.seqs command) to the SILVA bacterial

TABLE 1 | Identifiable *in silico* OTUs for the different genus within the phylum of *Spirochaetes* at a 97% threshold.

Variable region	Total OTU	*Treponema* (*n* = 28)[#]	*Sphaerochaeta* (*n* = 4)	*Spirochaeta* (*n* = 16)	*Leptospira* (*n* = 21)	*Exilispira* (*n* = 1)	*Leptonema* (*n* = 1)	*Borrelia* (*n* = 30)	*Brachyspira* (*n* = 14)	*Spironema* (*n* = 1)
V2	69	25	3	14	11	1	1	5	8	1
V3	50	25	2	12	4	1	1	3	1	1
V4	53	24	3	13	3	1	1	4	3	1
V5	47	22	3	12	2	1	1	2	3	1
V6	53	24	4	12	6	1	1	2	3	1
V7	34	12	2	11	3	1	1	2	1	1
V8	47	23	2	11	5	1	1	1	2	1

[#]*(n) indicates the number of unique representative sequences in the in silico.fasta file for each genus.*

reference database (Quast et al., 2012). Poorly aligned sequences, chimeras [chimera.uchime command; UCHIME algorithm (Edgar et al., 2011)], and other erroneous non-bacterial sequences (remove.lineage command) were removed. The remaining sequences were classified using a Bayesian classifier implemented in mothur and OTUs were assigned based on a distance threshold of 0.03.

For the species-level classification, *Treponema*-classified sequences were extracted from the dataset using the get.lineage command. Using the *Treponema* sequence data in the *in silico* fasta file (**Supplementary Table S1**), a database was created using the create.database command. Using this database, the taxonomy of the filtered sequences was assigned using the classify.otu command.

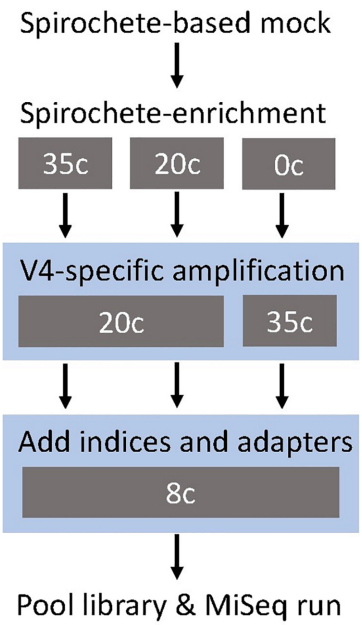


FIGURE 1 | Study design of the metataxonomic assay targeting the V4-region of the 16S rRNA gene. The gray boxes show the tested cycle (c) conditions for each step. The blue shading indicates the modular two-step library preparation.

Data Availability

All generated read files have been deposited in the NCBI Sequence Read Archive under the accession number PRJNA541286.

RESULTS

In silico Analysis of the Information Content of the V2–V8 Regions of the 16S rRNA Gene

We analyzed each hypervariable region (V2–V8) of the 16S rRNA gene for its potential to distinguish nine bacterial genera that make up the phylum *Spirochaetes*. In total, we analyzed the information content of the variable regions of 114 representative sequences *in silico* (**Supplementary Table S1**). Hypervariable regions V2–V8 were able to distinguish the nine bacterial genera at a similarity threshold of 97% (**Table 1**). The V2 region identified the largest total number of OTUs (n = 69) compared to all other tested regions (**Table 1**). Overall, the least number of OTUs were identified in the genera *Leptospira*, *Borrelia*, and *Brachyspira*. For the genus *Treponema* on the other hand, all variable regions with the exception of V7 were able to detect a high number of distinct OTUs (**Table 1**). Regions V2 and V3 were both able to detect 25 OTUs at a threshold of 97% in the *in silico* dataset containing 28 unique representative sequences. To examine the robustness of the *in silico* results for *Treponema*, we examined the identifiable OTUs at threshold cutoffs ranging from 90 to 99% (**Supplementary Figure S1**). For V2, V3, and V4 regions at 90% similarity threshold, >15 OTUs are distinguishable in the *Treponema* genus (**Supplementary Figure S1**).

V4-Region 16S rDNA Amplification of the Spirochete Mock Community

We tested three different amplification conditions targeting the V4-region of the 16S rRNA gene (**Figures 1**, **2**). All 16S rRNA gene amplification conditions were able to identify all seven

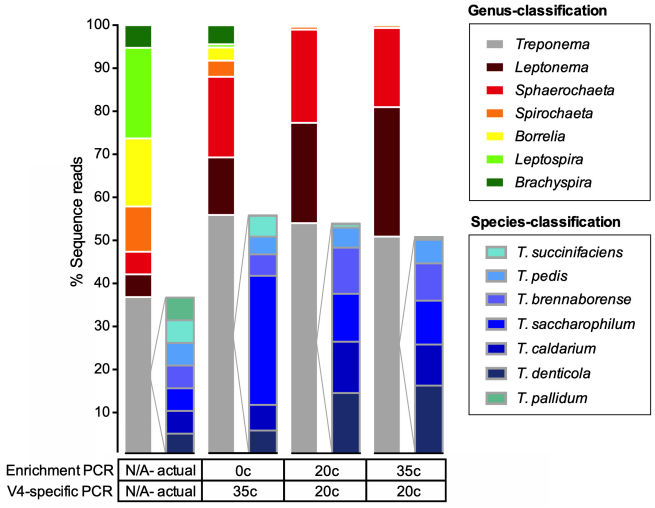


FIGURE 2 | Actual and observed proportions of a spirochete mock community to a genus and species-level. Relative abundance of OTUs in percentage of reads for a spirochete microbial mock community amplified using multiple methods. The spirochete enrichment and V4-region-specific PCR cycle numbers are shown below the taxa plots. The expected strain proportion (actual) of the spirochete mock community represents the theoretical composition and was corrected for differential copy number of the 16S rRNA gene. Species-level classification only shown for the genus *Treponema*.

genera that were included in the mock community samples (**Figure 2**). The amplification condition without spirochete-specific 16S rRNA gene enrichment differed less from the actual mixing proportion than the samples which were enriched (**Figure 2**). In all conditions, *Treponema*, *Leptonema*, and *Sphaerochaeta* were preferentially detected. For the spirochete enriched samples (20 cycles and 35 cycles), *Borrelia*, *Leptospira*, and *Brachyspira* made up <1% of the detected sequence reads (**Figure 2**). In addition to the genus-level identification, we classified the *Treponema* sequences on a species-level using a *Treponema*-specific database of the V4-region. At 97% similarity threshold, the database contains 24 OTUs of which 21 OTUs correspond to single species and 3 OTUs correspond to species-clusters (*denticola*-, *medium*-, and the *pallidum*-cluster) (**Supplementary Table S6**). All 16S rDNA amplification conditions were able to identify all seven species of *Treponema* in the mock community (**Figure 2**). Independent of amplification conditions, sequences corresponding to *T. pallidum*-cluster were amplified less efficiently.

Amplification control samples (blank samples) were included for each amplification method to test for contamination during the amplification process. The blank sample from the 20-cycle enrichment had the lowest amplicon quantity before sequencing as well as the lowest overall corresponding number of sequences reads (**Supplementary Figure S2** and **Supplementary Table S5**). Compared to the blank control enriched for 20 cycles, the control of the enrichment for 35 cycles had a 100× fold increase in sequence reads (**Supplementary Figure S2**). Despite an overall lower cycle count, the sequence reads corresponding to the nonenriched sample were as high as for the control enriched for 35 cycles. Unlike the enriched samples (20 cycle and 35 cycle), which detected minimal *Treponema* in the blank sample (<10 sequence reads), the non-enriched control sample detected 6,364 reads of *Treponema* (**Supplementary Figure S2**).

Intra-Metagenomic Assays Performance

The validation sample sets were used to assess the efficiency of the spirochete enrichment amplification, the effect of competition between two species, and the detection limit of the amplicon sequencing approach. The first validation set was a mixture of different concentrations of *T. pallidum* with a microbial mock community (HM-280). **Figure 3** shows that both 20 cycle- and 35 cycle-enrichment steps significantly improve the detection of *T. pallidum* compared to the unenriched sample at 5,000 16S rRNA gene copies of *T. pallidum*. As the input DNA of *T. pallidum* decreases, the dilution effect between *T. pallidum* to microbial mock community HM-280 can be visualized clearly (**Figure 3**). Four to six 16S rRNA gene copies of *T. pallidum* were detected for 35 cycle- (44,141 sequence reads) and 20 cycle-enrichment (2,224 sequence reads).

The second validation set was a mixture of *T. pallidum* and *T. denticola* in different ratios (see the section "Materials and Methods" for details). For all ratios, *T. denticola* outcompeted *T. pallidum* in detected sequence reads (**Table 2**). However, both species of *Treponema* were detected at all tested ratios (**Table 2**). The third validation set was a 10-fold serial dilution series of *T. pallidum*. Sequence reads were >9,000

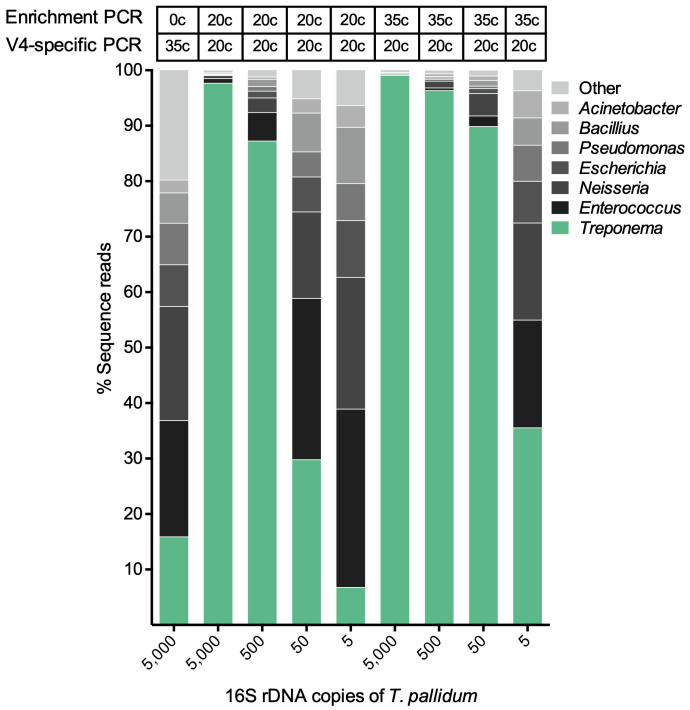


FIGURE 3 | The effect of spirochete enrichment on the detection limit of *T. pallidum*. Relative abundance of OTUs in percentage of reads for a *T. pallidum*-spiked microbial mock community (HM-280). The theoretically calculated amount of spike in *T. pallidum* ranging from 5 to 5,000 16S rDNA copies is shown along the *x*-axis. The biologically effective range of *T. pallidum* may differ from theoretically calculated amount (e.g., 5 copies = 4–6 copies). Cycle number for the spirochete enrichment and V4-region-specific PCR is shown above the taxa plots.

reads down to 500 16S rRNA gene copies of *T. pallidum* (**Figure 4**). At 50 16S rRNA gene copies, *T. pallidum* was detectable but overall sequence reads were markedly decreased (13,453 sequence reads). For the final two dilutions, total read numbers were <1,500 and *T. pallidum* sequence detection was analogous to the blank control (<10 sequence reads) (**Figure 4**).

Applications to Clinical Samples (Gilbert's Potoroo)

We examined samples from four Gilbert's potoroo which had been found to harbor a *Treponema* infection (Vaughan et al., 2009). Using the amplicon sequencing technique, we identified a *Treponema* species in all four clinical samples (**Supplementary Figure S3**). Sequence reads corresponding to the *Treponema* made up >75% of the total read count for Gilbert's potoroo samples No. 2–4 (**Supplementary Figure S3**). The *Treponema* sequences clustered into a single OTU, which cannot be identified using the *Treponema*-specific V4-region database at a 97% threshold identity.

TABLE 2 | Relative abundance of OTUs in percentage of reads for different proportions of *T. pallidum* and *T. denticola* 16S rRNA gene.

Ratio of *T. pallidum* and *T. denticola* 16S rRNA gene	% sequence reads for *T. pallidum* (read count)	% sequence reads for *T. denticola* (read count)
1:100	0.1 (37)	99.9 (81,032)
1:10	0.3 (350)	99.7 (112,402)
1:1	3.0 (3,108)	97.0 (100,507)
10:1	26.6 (31,077)	73.4 (85,750)
100:1	77.3 (74,179)	22.7 (21,809)

DISCUSSION

We used the *in silico* analysis to predict the informative nature of each 16S rDNA hypervariable region for different spirochetes. Our findings expand on the results of Rossi-Tamisier et al. (2015), indicating that three genera of spirochetes have low interspecies sequence similarity with the 16S rRNA gene, which makes it a suitable gene target for identification (**Table 1**). The

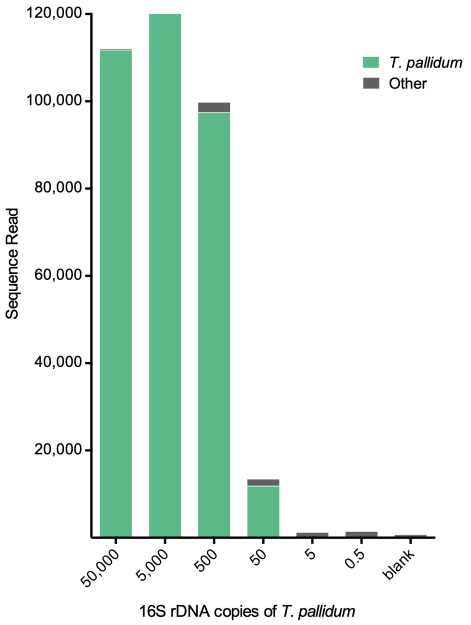


FIGURE 4 | Detection limit of the metagenomic approach for *Treponema*. Total sequence reads resulting from different input amounts of *T. pallidum*. Displayed amounts of *T. pallidum* (50,000–0.5 16S rDNA copies) represent theoretically calculated amounts. Biologically effective range may differ from theoretically calculated amount (e.g., 0.5 copies = 0–2 copies). Blank control represents the 16S rDNA amplification control using microbial DNA-free water as input. For these samples 20 cycles of enrichment PCR was followed by 20 cycles of V4-region-specific PCR (for more detail see the section "Materials and Methods").

V3 and V4 region have been previously described for their discriminatory power (Chakravorty et al., 2007; Yang et al., 2016; Graspeuntner et al., 2018). In a study of 110 bacterial species, V2 and V3 were the most suitable candidates (Chakravorty et al., 2007). Considering phylogenetic resolution, the variable regions 4, 5, and 6 have been previously identified as prime targets (Yang et al., 2016; Graspeuntner et al., 2018). Overall, the *in silico* analysis provided good initial data to efficiently design the *in vitro* experiments. It is, however, important to note that technical caveats of NGS sequencing must be considered prior to the selection of the most appropriate region (Kozich et al., 2013). For example, for the Illumina MiSeq Platform, paired-end sequencing can currently cover 300 base pairs. Considering the overall error rate of this platform [∼0.1–0.01% per base, depending on the data-filtering scheme (Meacham et al., 2011; Loman et al., 2012)], the ideal read length for a metataxonomic approach allows for full overlap of the two pair-end reads (Kozich et al., 2013). Based on our *in silico* results and pre-test using different primers, we selected the V4 region of the 16S rRNA gene for further *in vitro* testing.

A spirochete mock community of known species composition allowed for the systematic comparison between the different amplification methods (**Figure 1**; Brooks et al., 2015). Independent of the amplification method, our metagenomic approach was able to detect all seven genera of spirochetes in the mock community, as well as all species of *Treponema* (**Figure 2**). However, not all spirochetes were detected equally well with all amplification procedures (**Figure 2**). The spirochete-specific enrichment step, which was included for a better detection of spirochetes, led to the distortion of the actual proportions and favored *Treponema*, *Sphaerochaeta*, and *Leptonema* (**Figure 2**). The distortion of the bacterial profiles due to preferential amplification of multi-template PCR is a known phenomenon and a major limitation of 16S rRNA gene amplification that results from sub-optimal primer binding (Polz and Cavanaugh, 1998; Brooks et al., 2015; Hallmaier-Wacker et al., 2018). It has been previously shown that this distortion effect is not significantly influenced with decreasing the number of amplification cycles (Acinas et al., 2005; Sipos et al., 2007; Wu et al., 2010; Brooks et al., 2015). Similarly, our results did not remarkably change with an increased number of enrichment cycles (20 cycles vs. 35 cycles; **Figure 2**). Nevertheless, the use of unnecessary cycles should be avoided as it can lead to formation of unwanted side products such as chimeras (Ahn et al., 2012), as well as a higher risk of overamplifying reads that originate from contamination (blank controls; **Supplementary Figure S2**) (Salter et al., 2014).

To examine the benefits of the spirochete enrichment PCR (20 cycles and 35 cycles) on the detection limit of analysis, we tested the metataxonomic approach on mock communities that simulate bacterial proportions found in clinical samples (**Supplementary Table S3**). For these samples, the enrichment step critically improved the detection of *Treponema* at low copy numbers, thus indicating that enrichment is a useful tool for samples with low spirochete numbers (<5,000 16S rDNA copies) (**Figure 3**). We showed that five 16S rRNA gene copies of *T. pallidum* were detectable in a sample with 20 other bacterial species (even bacterial mock HM-280). Using serial dilutions, we were able to detect as little as 50 16S rRNA gene copies of *T. pallidum* using 20 cycle enrichment amplification (**Figure 4**). These data indicate a sensitivity of our assay that is comparable to standard TaqMan qPCR and which outcompetes the conventional 16S rRNA clonal approach (Leslie et al., 2007). Obtaining a high detection limit using a clonal approach is both time consuming and resource intensive (Leigh et al., 2010). On the other hand, Sanger sequence analysis of clone libraries provide greater phylogenetic resolution due to an increased read length, covering the full 16S rRNA gene (Leigh et al., 2010). The complex microbial communities present in many clinical samples is a frequent challenge in diagnostics and in these sample the occurrence of multi-*Treponema* species is not uncommon. For example, in oral syphilis patients, *T. pallidum* can be found in combination with *T. denticola* (Scott and Flint, 2005). We therefore tested the effect of competing species by simulating a co-infection of *T. pallidum* and *T. denticola* (**Supplementary Table S4**). Overall, the metataxonomic approach underestimated the ratio

of *T. pallidum* in the samples (**Table 2**). Amplicon sequencing was, however, sufficient to identify both species of *Treponema* at all tested ratios (**Table 2**). We note here that the metataxonomic approach does not accurately represent absolute abundance of different species (Widder et al., 2016). The used primers may have a significant effect in distorting tested ratios and thus alternative primer should be designed and evaluated for specific research questions. Additionally, quantitative techniques such as qPCR, flow cytometry, or fluorescence *in situ* hybridization (FISH) may be superior for evaluating known competing species (e.g., *T. pallidum and T. denticola*) (Props et al., 2017). Moreover, the metataxonomic approach should not be used for defining novel bacterial species even if species-level clustering is possible using the 16S rRNA sequence information (e.g., *Treponema*) (Tindall et al., 2010). Instead, 16S rRNA gene amplicon sequencing provides a qualitative view on the diversity of treponemes within a given DNA sample. For example, we used the metataxonomic approach to examine clinical samples of the Gilbert's potoroo that have been previously found to harbor a *Treponema* infection (Vaughan et al., 2009). As the potoroo clinical samples were associated with a polymicrobial environment and infection was believed to be chronic (Vaughan et al., 2009), we performed a 35 cycle-enrichment in order to detect low concentrations of spirochetes. We identified a single *Treponema* species in all tested samples of the four Gilbert's potoroos, which currently remains unclassified at a species level. The high percentage of *Treponema* in the detected samples (>75%; **Supplementary Figure S3**) indicates that the amplicon method is applicable for clinical samples and guides subsequent approaches that aim to fully characterize the discovered *Treponema* species. The results from the metataxonomic approach can be used to select most promising samples for whole-genome analysis (WGS), as well as provide a preliminary understanding of the possible phylogeny, which may assist in reference-based assembly (Wyres et al., 2014). Importantly, further WGS sequences of known and unknown *Treponema* are crucial to study the evolution and epidemiology of this ancient group of bacteria and to enhance future shotgun metagenomic studies. Currently, there is only a limited number of whole-genome sequences of *Treponema*, in particular the non-pathogenic species, due to the difficulty to culture many of the species [e.g., from the termite gut (Paster et al., 1996)].

CONCLUSION

We showed that the V4 region of the 16S rRNA gene is a valuable target to explore the diversity of *Treponema* in various biological sample types. To monitor the quality of each sequencing run, it is essential to including relevant controls with all clinical samples. When applied appropriately, the presented modular metataxonomic approach is broadly applicable as it requires only small amounts of bacterial DNA for the detection of a broad range of *Treponema* species.

DATA AVAILABILITY

The datasets generated for this study can be found in the NCBI Sequence Read Archive accession number PRJNA541286.

AUTHOR CONTRIBUTIONS

LH-W and SK conceived and designed the study. LH-W, SL, SG, CS, and SK performed the experiments in the laboratory. LH-W, SL, and SK analyzed the data. NB and RV-H contributed DNA from the Gilbert's potoroos. SG, CS, and JO contributed DNA samples of spirochetes for the mock sample. All authors contributed to the writing of the manuscript, read, reviewed, and approved the final manuscript.

FUNDING

SK received funding to conduct parts of this study by the German Research Foundation (DFG KN1097/3-2 and KN1097/7-1). The funders had no role in any part of this study.

ACKNOWLEDGMENTS

We thank the Biodefense and Emerging Infectious Research (BEI) Resources, NIAID, NIH for providing the cells from Microbial Mock Community (Even, HM-280) as part of the Human Microbiome Project. We thank David Šmajs of the Department of Biology, Faculty of Medicine at the Masaryk University for providing DNA from the *T. pallidum* subsp. *pertenue* strain Gauthier. Additionally, we thank Christian Roos (German Primate Center) and Fabian Ludewig (Transcriptome and Genome Analysis Laboratory at the University of Göttingen) for their assistance in optimizing the sequencing run. Finally, we thank Simone Severitt and Carola Berg (both DSMZ) for excellent technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02094/full#supplementary-material

REFERENCES

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M. F. (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* 71, 8966–8969. doi: 10.1128/AEM.71.12.8966-8969.2005

Ahn, J.-H., Kim, B. Y., Song, J., and Weon, H. Y. (2012). Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *J. Microbiol.* 50, 1071–1074. doi: 10.1007/s12275-012-2642-z

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., et al. (2015). The truth about metagenomics: quantifying and

counteracting bias in 16S rRNA studies. *BMC Microbiol.* 15:66. doi: 10.1186/s12866-015-0351-6

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4516–4522. doi: 10.1073/pnas.1000080107

Chakravorty, S., Helb, D., Burday, M., Connell, N., and Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* 69, 330–339. doi: 10.1016/j.mimet.2007.02.005

Choi, B. K., Paster, B. J., Dewhirst, F. E., Göbel, U. B. (1994). Diversity of cultivable and uncultivable oral spirochetes from a patient with severe destructive periodontitis. *Infect. Immun.* 62, 1889–1895.

Clayton, J. B., Gomez, A., Amato, K., Knights, D., Travis, D. A., Blekhman, R., Knight, R., et al. (2018). The gut microbiome of nonhuman primates: lessons in ecology and evolution. *Am. J. Primatol.* 30:e22867. doi: 10.1002/ajp.22867

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W.-H., et al. (2010). The human oral microbiome. *J. Bacteriol.* 192, 5002–5017. doi: 10.1128/JB.00542-10

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Graspeuntner, S., Loeper, N., Künzel, S., Baines, J. F., and Rupp, J. (2018). Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. *Sci. Rep.* 8:9678. doi: 10.1038/s41598-018-27757-8

Hallmaier-Wacker, L. K., Lueert, S., Roos, C., and Knauf, S. (2018). The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Sci. Rep.* 8:6292. doi: 10.1038/s41598-018-24573-y

Hartmann, M., Howes, C. G., Abarenkov, K., Mohn, W. W., and Nilsson, R. H. (2010). V-Xtractor: an open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16S/18S) ribosomal RNA gene sequences. *J. Microbiol. Methods* 83, 250–253. doi: 10.1016/j.mimet.2010.08.008

Hicks, A. L., Lee, K. J., Couto-Rodriguez, M., Patel, J., Sinha, R., Guo, C., et al. (2018). Gut microbiomes of wild great apes fluctuate seasonally in response to diet. *Nat. Commun.* 9:1786. doi: 10.1038/s41467-018-04204-w

Hong, P. Y., Li, X., Yang, X., Shinkai, T., Zhang, Y., Wang, X., et al. (2012). Monitoring airborne biotic contaminants in the indoor environment of pig and poultry confinement buildings. *Environ. Microbiol.* 14, 1420–1431. doi: 10.1111/j.1462-2920.2012.02726.x

Klitgaard, K., Nielsen, M. W., Ingerslev, H.-C., Boye, M., and Jensen, T. K. (2014). Discovery of bovine digital dermatitis-associated Treponema spp. in the dairy herd environment by a targeted deep-sequencing approach. *Appl. Environ. Microbiol.* 80, 4427–4432. doi: 10.1128/AEM.00873-14

Knauf, S., Lüert, S., Šmajs, D., Strouhal, M., Chuma, I. S., Frischmann, S., et al. (2018). Gene target selection for loop-mediated isothermal amplification for rapid discrimination of *Treponema pallidum* subspecies. *PLoS Negl. Trop. Dis.* 12:e0006396. doi: 10.1371/journal.pntd.0006396

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi: 10.1128/AEM.01043-13

Leigh, M. B., Taylor, L., Neufeld, J. D. (2010). "Clone libraries of ribosomal RNA gene sequences for characterization of bacterial and fungal communities," in *Handbook of hydrocarbon and Lipid Microbiology*, ed. K. N. Timmis (Berlin: Springer) 3969–3993. doi: 10.1007/978-3-540-77587-4_310

Leslie, D. E., Azzato, F., Karapanagiotidis, T., Leydon, J., and Fyfe, J. (2007). Development of a real-time PCR assay to detect Treponema pallidum in clinical specimens and assessment of the assay's performance by comparison with serological testing. *J. Clin. Microbiol.* 45, 93–96. doi: 10.1128/JCM.01578-06

Lilburn, T. G., Schmidt, T. M., Breznak, J. A. (1999). Phylogenetic diversity of termite gut spirochaetes. *Environ. Microbiol.* 1, 331–345. doi: 10.1046/j.1462-2920.1999.00043.x

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., et al. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 30, 434–439. doi: 10.1038/nbt.2198

Meacham, F., Boffelli, D., Dhahbi, J., Martin, D. I., Singer, M., and Pachter, L. (2011). Identification and correction of systematic error in high-throughput sequence data. *BMC Bioinformatics* 12:451. doi: 10.1186/1471-2105-12-451

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740. doi: 10.1126/science.276.5313.734

Paster, B. J. (2001). "Phylum XV. *Spirochaetes*," in *Bergey's Manual of Systematic Bacteriology*, 2nd Edn, Vol. 4. eds N. R. Krieg, J. T. Staley, D. R. Brown, B. P. Hedlund, B. J. Paster, N. L. Ward, et al. (New York, NY: Springer) 471–566.

Paster, B. J., Dewhirst, F. E., Cooke, S. M., Fussing, V., Poulsen, L. K., and Breznak, J. A. (1996). Phylogeny of not-yet-cultured spirochetes from termite guts. *Appl. Environ. Microbiol.* 62, 347–352.

Polz, M. F., and Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* 64, 3724–3730.

Props, R., Kerckhof, F.-M., Rubbens, P., De Vrieze, J., Sanabria, E. H., Waegeman, W., et al. (2017). Absolute quantification of microbial taxon abundances. *ISME J.* 11, 584–587. doi: 10.1038/ismej.2016.117

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Rodriguez-R, L. M., Overholt, W. A., Hagan, C., Huettel, M., Kostka, J. E., and Konstantinidis, K. T. (2015). Microbial community successional patterns in beach sands impacted by the Deepwater Horizon oil spill. *ISME J.* 9, 1928–1940. doi: 10.1038/ismej.2015.5

Rossi-Tamisier, M., Benamar, S., Raoult, D., and Fournier, P. E. (2015). Cautionary tale of using 16S rRNA gene sequence similarity values in identification of human-associated bacterial species. *Int. J. Syst. Evol. Microbiol.* 65, 1929–1934. doi: 10.1099/ijs.0.000161

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12:87. doi: 10.1186/s12915-014-0087-z

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Scott, C., and Flint, S. R. (2005). Oral syphilis—re-emergence of an old disease with oral manifestations. *Int. J. Oral Maxillofac. Surg.* 34, 58–63. doi: 10.1016/j.ijom.2004.01.029

Sipos, R., Székely, A. J., Palatinszky, M., Révész, S., Márialigeti, K., and Nikolausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiol. Ecol.* 60, 341–350. doi: 10.1111/j.1574-6941.2007.00283.x

Tindall, B. J., Rosselló-Móra, R., Busse, H. J., Ludwig, W., and Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266. doi: 10.1099/ijs.0.016949-0

Vaughan, R. J., Warren, K. S., Mills, J. S., Palmer, C., Fenwick, S., Monaghan, C. L., et al. (2009). Hematological and serum biochemical reference values and cohort analysis in the Gilbert's potoroo (*Potorous gilbertii*). *J. Zoo Wildl. Med.* 40, 276–288. doi: 10.1638/2008-0058.1

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Westcott, S. L., and Schloss, P. D. (2017). OptiClust, an improved method for assigning amplicon-based sequence data to operational taxonomic units. *mSphere* 2:e00073-17. doi: 10.1128/mSphereDirect.00073-17

Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., et al. (2016). Challenges in microbial ecology: building predictive understanding of

community function and dynamics. *ISME J.* 10, 2557–2568. doi: 10.1038/ismej.2016.45

Wu, J. Y., Jiang, X. T., Jiang, Y. X., Lu, S. Y., Zou, F., and Zhou, H. W. (2010). Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. *BMC Microbiol.* 10:255. doi: 10.1186/1471-2180-10-255

Wyres, K. L., Conway, T. C., Garg, S., Queiroz, C., Reumann, M., Holt, K., et al. (2014). WGS analysis and interpretation in clinical and public health microbiology laboratories: what are the requirements and how do existing tools compare? *Pathogens* 3, 437–458. doi: 10.3390/pathogens3020437

Yang, B., Wang, Y., and Qian, P. Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* 17:135. doi: 10.1186/s12859-016-0992-y

Conflict of Interest Statement: SG, CS, and JO were employed by the Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Chapter B4

Supplemental information:

Supplementary Material and Methods - Processing of Gilbert´s potoroo samples

Table S1. Spirochetes (with and without taxonomic validation*) and corresponding NCBI Accession numbers used in *in silico* determination

Table S2. Organisms included in the spirochete mock community

Table S3. Bacterial mixture of validation set 1

Table S4. Bacterial mixture of validation set 2

Table S5. Qubit measurements of spirochete mock community, *Treponema* validation sets (Val1-3) and blank controls included in this study

Table S6. *In silico* results for identifiable OTUs at a 97% threshold for the V4-region of the 16S rRNA gene of the *Treponema* genus

Figure S1. *In silico* results for identifiable OTUs at different threshold cut-offs for the *Treponema* genus

Figure S2. Bacterial profile of control samples included in this study.

Figure S3. Sequence read counts of Gilbert´s potoroo samples

1 Supplementary Material and Methods

Processing of Gilbert´s potoroo samples. We obtained previously extracted DNA from clinical samples that were taken for a different study (1). The collection of the samples was reviewed and approved by the Murdock University Animal Ethics Committee (W1082/04). No animals were handled for this study. Please refer to Vaughan-Higgins *et al*. for details on sampling procedures and DNA extraction. The metataxonomic assay was performed for the Gilbert´s potoroo samples using 35-cycles enrichment PCR followed by two-step Truseq library preparation. Data analysis including species-level classification for *Treponema* was performed as described for the mock communities.

2 Supplementary Tables

Table S1. Spirochetes (with and without taxonomic validation*) and corresponding NCBI Accession numbers used in *in silico* determination. *The taxonomic standing of each spirochete species can be cross-checked using the NCBI GenBank Accession Number and Integrated Taxonomic Information System.

Genus	Species	GenBank Number
Borrelia	*afzelii*	JX888452
Borrelia	*americana*	NR_116166
Borrelia	*anserina*	U42284.1
Borrelia	*bavariensis*	NR_074854
Borrelia	*burgdorferi*	AB091823
Borrelia	*carolinensis*	EU085416
Borrelia	*coriaceae*	NR_114544
Borrelia	*crocidurae*	GQ358200
Borrelia	*duttonii*	AF107366
Borrelia	*garinii*	D67018
Borrelia	*hermsii*	EU203150
Borrelia	*hispanica*	GQ202264

Borrelia	*japonica*	L46696
Borrelia	*lonestari*	AY166715
Borrelia	*mayonii*	KP972468
Borrelia	*miyamotoi*	AB904793
Borrelia	*parkeri*	AF307100
Borrelia	*persica*	HQ610931
Borrelia	*recurrentis*	AF107361
Borrelia	*sinica*	NR_024713
Borrelia	*spielmanii*	AM182231
Borrelia	*tanukii*	D67023
Borrelia	*theileri*	KF569941
Borrelia	*turcica*	NR_024820
Borrelia	*turdi*	D67024
Borrelia	*turicatae*	AY934610
Borrelia	*valaisiana*	EU135596
Borrelia	*yangtzensis*	NR_145665
Brachyspira	*aalborgi*	AF200693
Brachyspira	*alvinipulli*	JF430707
Brachyspira	*canis*	HM450994
Brachyspira	*corvi*	EU819070
Brachyspira	*hyodysenteriae*	NR_044764
Brachyspira	*innocens*	NR_044763
Brachyspira	*intermedia*	KR809388
Brachyspira	*murdochii*	KR809386
Brachyspira	*muridarum*	GU189376
Brachyspira	*muris*	GU189383
Brachyspira	*pilosicoli*	AB120008
Brachyspira	*pulli*	KR809387
Brachyspira	*rattus*	GU189374
Brachyspira	*suanatina*	DQ473578
Exilispira	*thermophila*	NR_041644
Leptonema	*illini*	JQ988853
Leptospira	*alexanderi*	JQ988836.1
Leptospira	*biflexa*	JQ988840.1
Leptospira	*borgpetersenii*	AY995716
Leptospira	*broomii*	Y19243
Leptospira	*fainei*	JQ988851.1
Leptospira	*idonii*	AB721966.1
Leptospira	*inadai*	AY631896.1
Leptospira	*interrogans*	AY995726
Leptospira	*kirschneri*	EF536998

Leptospira	*kmetyi*	AB279549.1
Leptospira	*licerasiae*	NR_044310
Leptospira	*mayottensis*	NR_134067.1
Leptospira	*meyeri*	HQ709385.1
Leptospira	*noguchii*	EU349496
Leptospira	*santarosai*	JQ988838.1
Leptospira	*terpstrae*	NR_115294.1
Leptospira	*vanthielii*	NR_115297.1
Leptospira	*weilii*	JQ988839.1
Leptospira	*wolbachii*	AY631879.1
Leptospira	*wolffii*	KC662454.1
Leptospira	*yanagawae*	MG979779.1
Sphaerochaeta	*associata*	NR_145842.1
Sphaerochaeta	*coccoides*	NR_042260
Sphaerochaeta	*globosa*	NR_114608
Sphaerochaeta	*pleomorpha*	NR_114609
Spirochaeta	*africana*	NR_026302
Spirochaeta	*alkalica*	NR_026301
Spirochaeta	*americana*	NR_028820
Spirochaeta	*asiatica*	NR_026300
Spirochaeta	*aurantia*	FR749896.1
Spirochaeta	*cellobiosiphila*	NR_044505
Spirochaeta	*dissipatitropha*	AY995150.1
Spirochaeta	*halophila*	NR_044756.2
Spirochaeta	*isovalerica*	FR749931.1
Spirochaeta	*lutea*	HG965770.2
Spirochaeta	*perfilievii*	AY337318
Spirochaeta	*psychrophila*	NR_134185.1
Spirochaeta	*smaragdinae*	NR_027585
Spirochaeta	*taiwanensis*	AY735103.1
Spirochaeta	*thermophila*	CP002903.1
Spirochaeta	*xylanolyticus*	AY735097
Spironema	*culicis*	AF166259.1
Treponema	*amylovorum*	JN713358
Treponema	*azotonutricium*	NR_074168
Treponema	*berlinense*	NR_042797.1
Treponema	*brennaborense*	NR_029348
Treponema	*bryantii*	AB849328
Treponema	*caldarium*	NC_015732.1
Treponema	*denticola*	KC415235
Treponema	*isoptericolens*	NR_042486.1

Treponema	*lecithinolyticum*	GU420631
Treponema	*maltophilum*	X87140
Treponema	*medium*	JN713397
Treponema	*pallidum*	NC_021179
Treponema	*paraluisleporidarum* ecovar Cuniculus	NC_015714
Treponema	*paraluisleporidarum* ecovar Lepus	JX899416
Treponema	*parvum*	AF302939.1
Treponema	*pectinovorum*	GU562449.1
Treponema	*pedis*	KP063170
Treponema	*phagedenis*	FJ004921
Treponema	*porcinum*	NR_042942
Treponema	*primitia*	NC_015578
Treponema	*putidum*	NR_027189
Treponema	*saccharophilum*	M71238
Treponema	*socranskii*	AB015892
Treponema	*stenostreptum*	NR_113042
Treponema	*succinifaciens*	NR_074755.1
Treponema	*vincentii*	AY119690
Treponema	*zioleckii*	DQ065758
Treponema	*zuelzerae*	NR_104797

Table S2. Organisms included in the spirochete mock community. * Catalogue number of the German Collection of Microorganisms and Cell Cultures (DSMZ).

DSM	Genus	Species	Cultivation	Genome size (Mbp)	16S rDNA copies	NCBI Reference
10508	*Borrelia*	*afzelii*	pure culture	0.905	2	NC_018887.1
5251	*Borrelia*	*hermsii*	pure culture	0.923	1	NZ_CP014349.1
4680	*Borrelia*	*burgdorferi*	pure culture	0.911	1	NC_001318.1
105803	*Brachyspira*	*hyodysenteriae*	pure culture	3.041	1	NZ_CP015910.2
21528	*Leptonema*	*illini*	pure culture	4.521	1	NZ_AHKT00000000.1
21526	*Leptospira*	*kirschneri*	pure culture	4.409	1	NZ_AHMN00000000.2
21537	*Leptospira*	*meyeri*	pure culture	4.188	1	NZ_AKXE00000000.1
21521	*Leptospira*	*terpstrae*	pure culture	4.092	2	NZ_AOGW00000000.2
21534	*Leptospira*	*broomii*	pure culture	4.395	3	NZ_AHMO00000000.2
22777	*Sphaerochaeta*	*globosa*	pure culture	3.316	4	NC_015152.1
6578	*Spirochaeta*	*thermophila*	pure culture	2.560	2	NC_017583.1
8902	*Spirochaeta*	*africana*	pure culture	3.286	3	NC_017098.1
12168	*Treponema*	*brennaborense*	pure culture	3.056	4	NC_015500.1
14222	*Treponema*	*denticola*	pure culture	2.843	2	NC_002967.9
18691	*Treponema*	*pedis*	pure culture	2.889	2	NC_022097.1
2985	*Treponema*	*saccharophilum*	pure culture	3.454	1	NZ_AGRW00000000.1
2489	*Treponema*	*succinifaciens*	pure culture	2.732	4	NC_015385.1
7334	*Treponema*	*caldarium*	pure culture	3.239	3	NC_015732.1
N/A	*Treponema*	*pallidum*	*in vivo*	1.139	2	NZ_CP003679.1

Table S3. Bacterial mixture of validation set 1. * Microbial mock community, HM-280 (Biodefense and Emerging Infectious Research (BEI) Resources, Manassas, USA) contains no *Spirochaetes*. # Effective biological range is 4-6 copies of 16S rDNA.

Mock community HM-280 * (ng/µl)	Concentration of *T. pallidum (*copies of 16S rDNA)
0.50	5,000
0.50	500
0.50	50
0.50	5 #

Table S4. Bacterial mixture of validation set 2.

Concentration of *T. pallidum (*copies of 16S rDNA)	Concentration of *T. denticola (*copies of 16S rDNA)	Final ratio of *T. pallidum* to *T. denticula*
500	50,000	1:100
5,000	50,000	1:10
50,000	50,000	1:1
50,000	5,000	10:1
50,000	500	100:1

Table S5. Qubit measurements of spirochete mock community, *Treponema* validation sets (Val1-3) and blank controls included in this study. * *TP*: theoretically computed *T. pallidum* 16Sr RNA gene copies in each sample; # OFR: out of range ≤ 0.2 ng/µl.

Sample	Enrichment PCR (cycles)	V4-specific PCR (cycles)	Qubit Average (ng/µl)
Spirochete mock	0	35	12.8
Spirochete mock	20	20	14.9
Spirochete mock	35	20	16.8
Val1 (5,000 copies TP)*	0	35	12.8
Val1 (5,000 copies TP)	20	20	2.5
Val1 (500 copies TP)	20	20	3.3
Val1 (50 copies TP)	20	20	0.43
Val1 (5 copies TP)	20	20	0.45
Val1 (5,000 copies TP)	35	20	9.4
Val1 (500 copies TP)	35	20	12.0
Val1 (50 copies TP)	35	20	10.6
Val1 (5 copies TP)	35	20	29.3
Val2 (1:100)	20	20	2.0
Val2 (1:10)	20	20	6.8
Val2 (1:1)	20	20	4.9
Val2 (10:1)	20	20	8.2
Val2 (100:1)	20	20	2.4
Val3 (50,000 copies TP)	20	20	3.4
Val3 (5,000 copies TP)	20	20	8.3
Val3 (500 copies TP)	20	20	1.7
Val3 (50 copies TP)	20	20	0.14
Val3 (5 copies TP)	20	20	OFR#
Val3 (0.5 copies TP)	20	20	OFR
Blank control	0	35	1.1
Blank control	20	20	OFR
Blank control	35	20	0.3

Table S6. *In silico* results for identifiable OTUs at a 97% threshold for the V4-region of the 16S rRNA gene of the *Treponema* genus.

OTU	Species association	OTU Representation
1	*T. amylovorum*	Single species
2	*T. azotonutricium*	Single species
3	*T. berlinense*	Single species
4	*T. brennaborense*	Single species
5	*T. bryantii*	Single species
6	*T. caldarium*	Single species
7	*T. denticola, T. putidum*	*Denticola*-cluster
8	*T. isoptericolens*	Single species
9	*T. lecithinolyticum*	Single species
10	*T. maltophilum*	Single species
11	*T. medium, T. vincentii*	*Medium*-cluster
12	*T. pallidum, T. paraluisleporidarum*	*Pallidum*-cluster
13	*T. pectinovorum*	Single species
14	*T. pedis*	Single species
15	*T. parvum*	Single species
16	*T. phagedenis*	Single species
17	*T. porcinum*	Single species
18	*T. primitia*	Single species
19	*T. saccharophilum*	Single species
20	*T. socranskii*	Single species
21	*T. stenostreptum*	Single species
22	*T. succinifaciens*	Single species
23	*T. zioleckii*	Single species
24	*T. zuelzerae*	Single species

3 Supplementary Figures

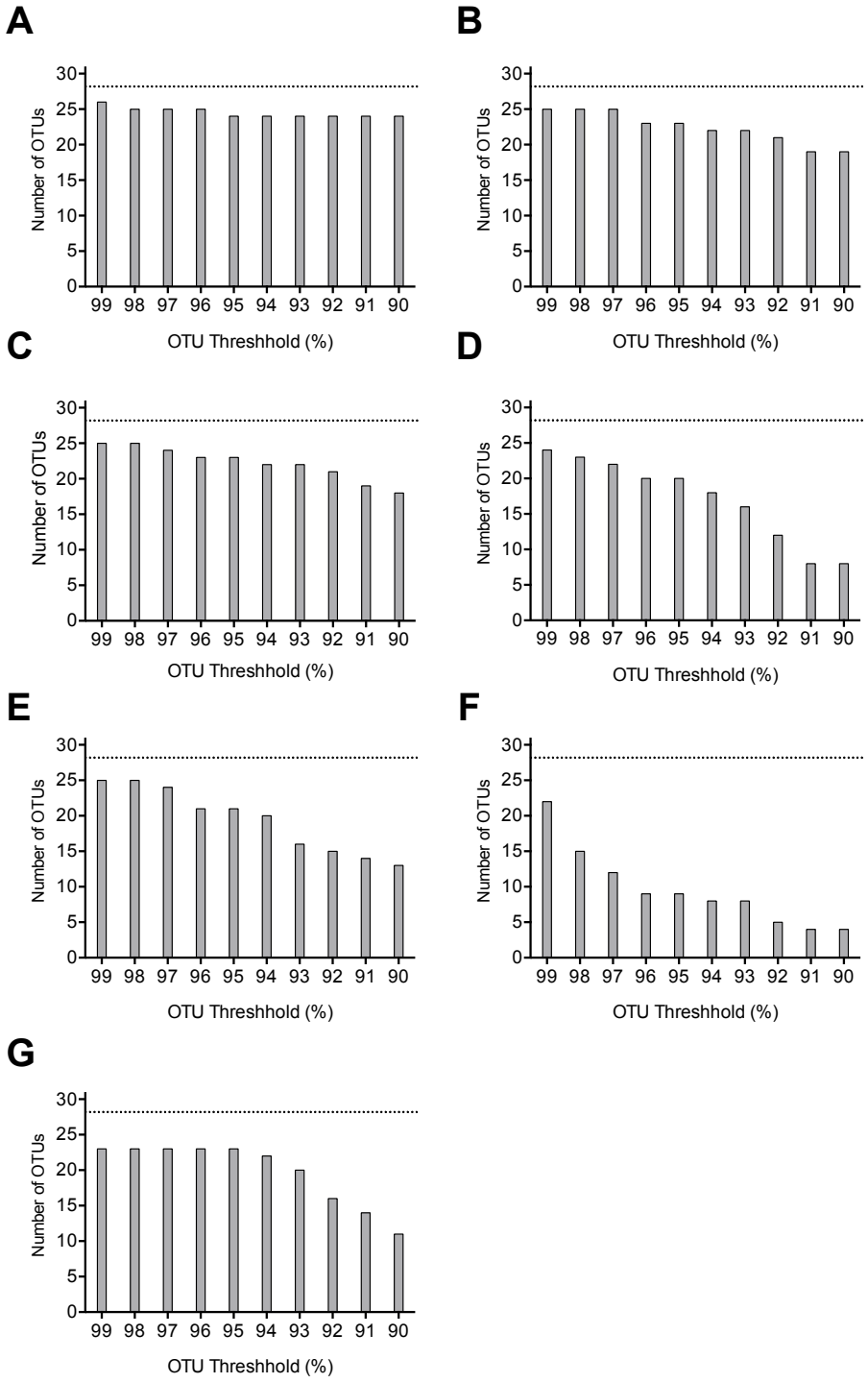


Fig. S1. *In silico* results for identifiable OTUs at different threshold cut-offs for the *Treponema* genus. Gray bar graphs show the number of OTUs detected at thresholds ranging from 99% to 90% similarity for 16S rDNA region (A) V2, (B) V3, (C) V4, (D) V5, (E) V6, (F) V7 and (G) V8. The dashed line on the plots represents the number of defined *Treponema* species in *in silico* dataset.

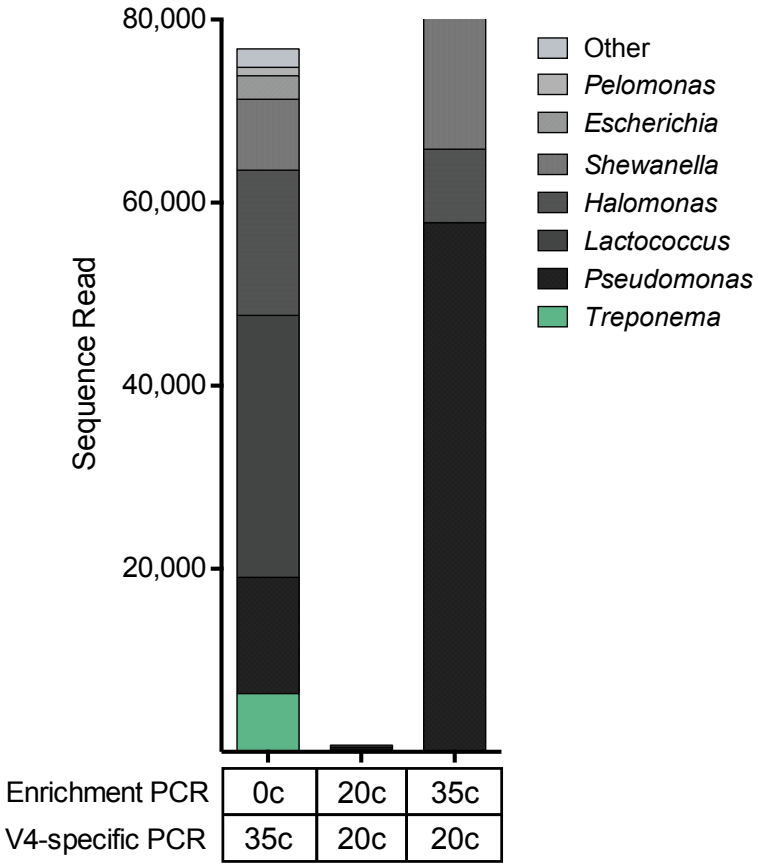


Fig. S2. Bacterial profile of control samples included in this study. Total sequence reads corresponding to 16S rRNA amplification control samples (blank controls) in this study. The spirochete enrichment and V4-region specific PCR cycle number are shown below the plot.

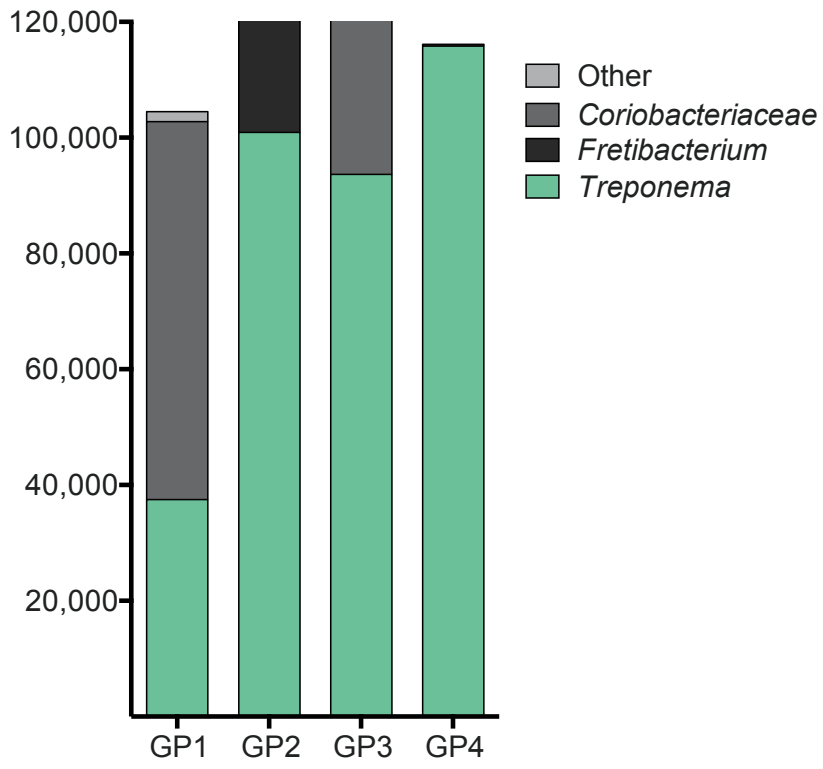


Fig. S3. Sequence read counts of Gilbert´s potoroo samples. Total sequence reads corresponding to four clinical samples. For the clinical samples, 35-cycles of spirochete enrichment PCR was followed by 20-cycles of V4-region specific PCR (for more detail see the methods).

4 Supplementary References

1. Vaughan RJ, Warren KS, Mills JS, Palmer C, Fenwick S, Monaghan CL, Friend AJ. 2009. Hematological and serum biochemical reference values and cohort analysis in the Gilbert's potoroo (*Potorous gilbertii*). J Zoo Wildl Med 40:276-288.

C GENERAL DISCUSSION

1 The importance of standardization

The One Health field is rapidly expanding, leading to growing interest in the multi-disciplinary approach and an increasing number of specialized 'One Health' journals (Osterhaus and MacKenzie, 2016). As the One Health field is highly interdisciplinary with many cross-continental collaborations and large-scale consortiums (Cunningham et al., 2017), standardization is an important topic of discussion. This thesis highlights the importance of standardizing terms, sequence thresholds, best practices, and appropriate controls in order to produce reliable results and avoid misunderstandings amongst the scientific community and the general public (Marchesi and Ravel, 2015).

1.1 Standardizing terms for better interdisciplinary communication

Standardization of terms and definitions is essential for effective interdisciplinary communication and an important tool to align future research goals (Strehlow, 1993). The use of the term 'parasite' illustrates the dissonance that differences in vocabulary can cause between researchers, the medical community, and the general public. For researchers, the definition of parasite includes all forms of life that live in a harmful host-dependent relationship, including animals, plants, fungi, bacteria, and viruses (Lucius et al., 2017). However, for the medical community and the general public, a parasite is a common term used to describe only parasitic protozoa, worms (helminths), and arthropods (Callaway and Cyranoski, 2015; Lucius et al., 2017). This example highlights the need to not only standardize terms but also provide appropriate applications and support activity that encourages a common vocabulary amongst different stakeholders (Strehlow, 1993). Attempts to standardize terminology have been reported for multiple fields, especially in medicine where clinically meaningful categorizations are often essential for disease treatment (Frassetto et al., 2007; Marchesi and Ravel, 2015; Raebel et al., 2013). In Chapter B1, we focused on defining applicable requirements to standardize the term 'reservoir of disease', a central concept in One Health (Hallmaier-Wacker et al., 2017). Despite many conceptual

frameworks for this term (Ashford, 2003; Drexler et al., 2012; Haydon et al., 2002; Viana et al., 2014), the characterization of reservoirs remains controversial, challenging and often driven by the tendency of science to hype certain areas of research (Caulfield et al., 2016; Vinkers et al., 2015). In light of this, we deemed it necessary to move away from the theoretical concept of disease reservoirs, towards well-structured applicable criteria (Hallmaier-Wacker et al., 2017). Furthermore, we applied the new framework to known zoonotic diseases to illustrate the intended use and avoid misunderstanding. We believe that the further standardization of vocabulary will facilitate effective communication across scientific disciplines, counteract the tendency of science to overemphasize positive results, and aid future collaborations within One Health.

1.2 Establishing sequence thresholds for classification

As highlighted in Chapter B1, there is a need to define and establish thresholds of sequence similarity for microorganisms. Mutation rates vary significantly amongst pathogens (Dapp et al., 2013; Matic et al., 1997) and thus current sequence thresholds often misclassify microorganisms, especially when comparing pathogens in multiple hosts and environments (potential reservoir system) (Hallmaier-Wacker et al., 2017). At the core of defining sequence thresholds is a fundamental discussion of how to classify microorganisms (Cohan, 2002; Rosselló-Mora and Amann, 2001; Staley, 2006). Cell-culture used to rely on observable morphological traits between bacterial cultures for taxonomic classification (Fig. 1A; Rosselló-Mora and Amann, 2001). Molecular technology, such as 16S rRNA sequencing (Fig. 1B-C), multi-locus strain typing (MLST; Fig. 1D) and whole genome sequencing (WGS; Fig. 1E), have allowed for DNA-based classification, yet have not resolved the ongoing debate (Fraser et al., 2009; Konstantinidis et al., 2006). For 16S rRNA-based classification, threshold cutoffs (97% vs. 99%; Fig. 1B-C) for classification have been discussed. However, it has been shown that any selected cutoff value is valid for less than 50% of studied microorganisms (Rossi-Tamisier et al., 2015). Rossi-Tamisier et al. (2015) suggested that threshold values need to be defined separately for each bacterial genus. To negate the issue of taxonomic classification, 16S rRNA-based metagenomic studies (e.g. Chapter B3) have relied on binning sequences into operational taxonomic units (OTU) (Blaxter et al.,

2005; Schloss et al., 2009). Recent studies have suggested that OTU binning thresholds need to be reconsidered (Edgar, 2018) and alternative dynamic clustering per taxonomic family has been proposed (Mysara et al., 2017). The large sequence variation in the 16S rRNA gene in some bacterial families can be advantageous for metagenomic detection, as shown in Chapter B4. The high variability in the genus *Treponema* allowed for near species-level classification, but also highlights the current discrepancies between sequence thresholds. The issue of defining appropriate sequence thresholds is not limited to 16S rRNA data but has also been reported for MLST and whole genome approaches (Konstantinidis et al., 2006). For molecular-based techniques, it is important to further address this challenge with pragmatic theoretical solutions as well as novel bioinformatic approaches.

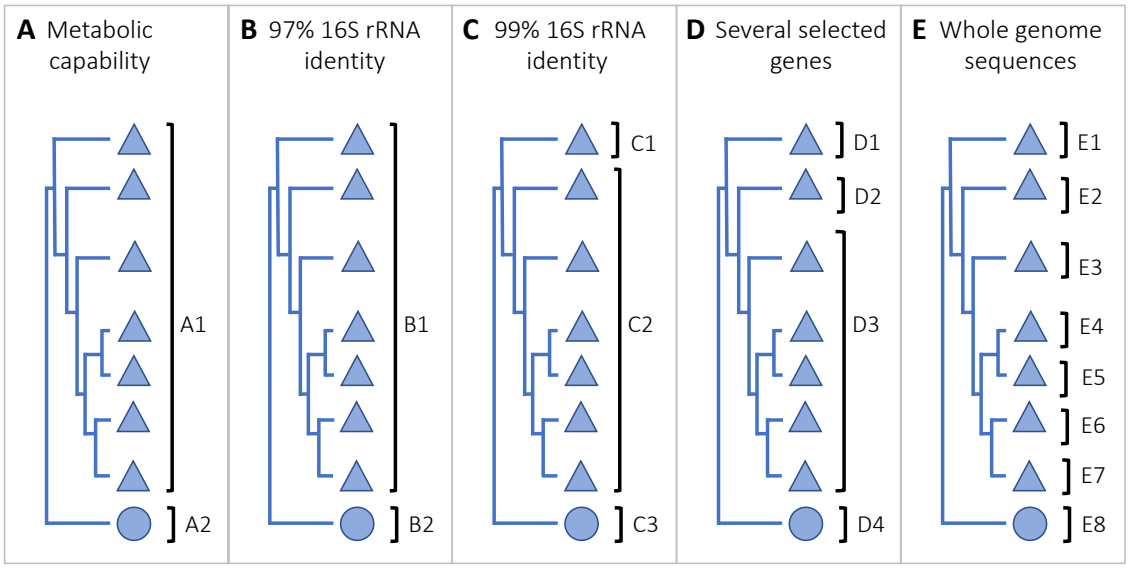


Figure 1. Bacterial classification using different criteria. At the tip of each phylogenetic tree, the geometric shapes represent different isolated bacteria. The circle represents the outgroup and is metabolically significantly different from the triangles. The basis of classification (A-E) results in a different division of the bacterial isolates (geometric shapes), which is indicated by the vertical bar and binning label (e.g. A1). Classification is based on (A) obvious metabolic differences (square vs. circle), (B) 16S rRNA gene (threshold: 97%), (C) 16S rRNA gene (threshold: 99%), (D) several selected genes that are highly divergent (e.g. MLST), and (E) whole genome sequences. The figure was originally published by Kopac and Cohan (2011) and was modified for this thesis.

1.3 Inclusion of appropriate controls

Before using any metagenomic approach, it is important to understand the technique's inherent bias (Brooks et al., 2015). To examine this bias and optimize existing methods, microbial mock communities composed of a known mixture of different microorganisms have been generated and are now commercially available (Jumpstart Consortium Human Microbiome Project Data Generation Working Group, 2012). In this thesis, microbial mock communities were included as positive controls in all presented experimental studies (Chapter B2-4). In Chapter B2, the cell-based microbial mock community allowed us to screen sampling, extraction and amplification procedures (Hallmaier-Wacker et al., 2018). Similarly, in Chapter B4, the DNA-based spirochete mock community was instrumental to establish and validate the detection technique (Hallmaier-Wacker *et al*., 2019b). Using environmental samples of unknown microbe composition would not have allowed us to systematically compare different experimental conditions (Brooks et al., 2015; Kim et al., 2017). In addition to mock samples, negative controls (i.e. blank samples) are essential to monitor sampling and laboratory contamination (Salter et al., 2014). If the microbial biomass is low, it can be difficult to distinguish the reads resulting from contamination from the actual microbiota reads (Kim et al., 2017). For example, a unique placenta microbiota was identified by multiple studies (Aagaard et al., 2014; Amarasekara et al., 2015; Antony et al., 2015; Zheng et al., 2015), however a recent study questioned these results by showing that the microbiota of the placenta was similar to the negative controls for six human specimens (Lauder et al., 2016). Well-planned studies with appropriate controls and sample sizes are necessary to address if the placenta harbors a unique microbiota (Kim et al., 2017). Whenever applicable, we included appropriate blank controls, reported lot numbers of reagents and randomized sample processing to closely monitor and avoid contamination (Chapter B2-4). We believe that inclusion of appropriate positive and negative controls should be considered good laboratory practice in metagenomic studies and thus mandatory for publication.

1.4 The importance of establishing sample handling

Establishing and validating methods is an essential but often overlooked aspect of metagenomic studies. Sample handling and processing regimes for different sample types should be informed by laboratory-based validation studies (Quince et al., 2017). It has been shown that inadequate sample collection and processing can lead to misinterpretation of data and false conclusions (Laurence et al., 2014; Salter et al., 2014). Chapter B2 focuses on establishing appropriate sample handling methods for swab material, a key sample type for many studies (Hallmaier-Wacker et al., 2018). This chapter highlights that the ideal sample handling conditions can greatly vary between different sample types (e.g. fecal vs. swab type). Independent of sequencing approach, the ideal conditions need to be defined for the entire library preparation process, including sample collection (Choo et al., 2015), DNA extraction (Wagner Mackenzie et al., 2015) and library amplification (Ahn et al., 2012; Jones et al., 2015). Previous studies have identified that distortion is particularly driven by DNA extraction and amplification (Acinas et al., 2005; Brooks et al., 2015; Wagner Mackenzie et al., 2015). While PCR amplification can be circumvented using shotgun metagenomic sequencing (Miller et al., 2013), DNA extraction affects all types of metagenomic approaches (Brooks et al., 2015; Jovel et al., 2016). Extracting a representative microbial community is challenging due to large differences in the abundance of genetic content (e.g. host vs. microorganisms) (Kleiner et al., 2015), as well as differences in the susceptibility to lysis conditions (Vesty et al., 2017; Yuan et al., 2012). Different lysis treatments have been studied but there is currently no ideal method for all sample types (Carrigg et al., 2007; Wagner Mackenzie et al., 2015; Yuan et al., 2012). To allow for cross-study comparisons, handling conditions should be standardized as much as possible therefore reducing the variation across different datasets (Hallmaier-Wacker et al., 2018). The handling conditions established in Chapter B2 informed our sampling protocol at the German Primate Center (Chapter B3), as well as the processing of the spirochete mock community (Chapter B4). Additionally, to ensure the comparability between datasets, sampling efforts of ongoing studies (e.g. *Treponema* field investigations; Chapter B4) have incorporated the established handling practices for swab material.

2 Application of targeted metagenomics in One Health

Metagenomics is a powerful tool and has the potential to significantly aid different kinds of One Health investigations (Miller et al., 2013). However, reliable infrastructure, laboratory expertise, and bioinformatic understanding are necessary to produce reliable metagenomic data (Miller et al., 2013). Metagenomic studies should be well planned and not applied as a one-size-fits-all solution. For example, for diagnostics, classical pathogen specific methods (e.g. immunoassay) should not be negated as optimal solutions for targeted, rapid, on-site testing (Miller et al., 2013). Metagenomics can, however, aid diagnostics by providing avenues for novel pathogen detection (Pallen, 2014) and validating unexpected negative test results (Marks et al., 2017). Nevertheless, the detection of a pathogen using metagenomics does not provide information about disease progression or biological relevance (Miller et al., 2013). Results must be interpreted with caution and causality (e.g. between disease and presents of microorganism) should be carefully examined in well-controlled experiments (Fritz et al., 2013). Investigations need to critically assess which methods are ideal for answering the particular research question. In this thesis, I highlighted two applications of targeted metagenomics in the One Health field. Chapters B3 and B4 use sequencing of a single hypervariable region of the 16S rRNA gene to examine bacterial diversity, which cannot be assessed using classical detection methods. Both studies were conducted under controlled laboratory conditions and thus provide important baseline data for future One Health investigations.

2.1 The importance of baseline data in microbiome analysis

Studying the microbiome in health and diseases has provided a better understanding of the role that symbiotic microbes can play within the One Health triad (Trinh et al., 2018). 16S rRNA-based analyses have been performed for a wide range of samples and the importance of the human and animal-associated microbiota has become evident (Bahrndorff et al., 2016; Turnbaugh et al., 2007). Susceptibility to pathogenic microbes can be greatly influenced by the microbiota (Abt and Pamer, 2014; Buffie et al., 2015; Dennison et al., 2014). These findings have implications not only for disease ecology

but also for the use of translational animal models (Kostic et al., 2013). In Chapter B3, we examined the genital microbiota of male and female rhesus monkeys (Hallmaier-Wacker et al., 2019a). Rhesus monkeys are commonly used translational animal models for infectious diseases, including sexually transmitted infections (Haus et al., 2014; Spear et al., 2010). For example, rhesus monkeys have been used to model human coinfections with syphilis (*T. pallidum* subsp. *pallidum*) and simian immunodeficiency virus (SIV) (Marra et al., 1992). In Chapter B3, our data show that endocrine-regulated processes (lactation/menstruation), cage-mate (sexual partners) relationships and group associations can significantly impact the urogenital microbiota (Hallmaier-Wacker et al., 2019a). Since these differences likely influence biomedical experiments, we urge for increased awareness of microbiota considerations of translational animal models. Just like genetic differences guide the selection of animals for biomedical experiments today (Haus et al., 2014; Nguyen and Xu, 2008), microbiota consideration should influence future selection of translational animal models. Continual improvement of the selection of non-human primates (NHPs) is essential to refine and reduce the required number of animals (Prescott et al., 2017). In addition to important data on captive rhesus monkeys, Chapter B3 also provides baseline data for One Health investigations on wild NHPs in health and disease. Future research should focus on the causal relationships between the microbiota and pathogens during health and disease. For example, a recent study demonstrated that cardiac disease in zoo-housed western lowland gorillas (*Gorilla gorilla gorilla*) was associated with a change in the gut microbiota (Krynak et al., 2017). The authors suggest that *Spirochaetes*, *Proteobacteria* and *Firmicutes* were markers of cardiac disease (Krynak et al., 2017), without showing a direct causal relationship between the two factors. Caution is warranted in making premature assumptions about relationships, given that changes in the microbiota may be indirect effects of disease (e.g. changes in diet (David et al., 2014)). Controlled clinical trials in humans (e.g. clinical trials for fecal transplants (Rossen et al., 2015)) and translational animal models (e.g. germ-free mice (Bendtsen et al., 2015)) are key resources to examine causality in microbiota shifts.

2.2 Metataxonomic approach for the detection of *Treponema*

Treponema are often found at high numbers in microbiome analysis but generally remain undiscussed and unclassified (Clayton et al., 2018). For example, the gastrointestinal tract of termites has been found to harbour more than 67 treponemal phylotypes (Lilburn et al., 1999), of which only a fraction has been cultured and further characterized (Dröge et al., 2008; Graber et al., 2004; Lilburn et al., 2001). The large diversity of *Treponema* and the difficulty to culture many (Orth et al., 2010) have further hindered investigations and characterizations. In Chapter B4, we establish and validate a metataxonomics approach to examine the diversity of *Treponema* (Hallmaier-Wacker *et al*., 2019b). Previous studies have expressed interest in the diversity of oral spirochetes and have used 16S rRNA clonal sequencing approaches to estimate the diversity (Choi et al., 1996; Dewhirst et al., 2010). Compared to clonal library approach, metataxonomics is sensitive and high-throughput and thus easily applicable to a wide spectrum of sample types. In Chapter B4, we demonstrate one clinical application of the metataxonomic approach by examining genital samples from Gilbert's potoroo, which were previously described to harbour a *Treponema* infection (Vaughan et al., 2009). We were able to obtain near species-level classification for the *Treponema* (Rossi-Tamisier et al., 2015), which indicated that the spirochete in the Gilbert's potoroo did not cluster within any known *Treponema* species (Hallmaier-Wacker *et al*., 2019b). Previously, data from the 16S rDNA sequences, in combination with simple microscopy, have been used to characterize new species of *Treponema* (Nordhoff et al., 2005). However, 16S rRNA-based data should only provide a starting point for further characterizations and should not be used alone for novel species identification (Tindall et al., 2010). In order to gain a more comprehensive understanding of the isolated *Treponema*, WGS is currently being conducted. Results from the metataxonomics approach can be used when choosing sample for WGS and also yield a preliminary understanding which may contribution to reference-based assembly of whole genome sequences (Wyres et al., 2014). In addition to WGS, further sampling efforts and well-planned investigations should be performed in order to further examine the pathogenic potential and disease progression in the Gilbert's potoroo (Vaughan, 2008). In order to examine the *Treponema* infection within the One

Health triad, future studies should focus on examining other species (including marsupials) and environments in the Australian ecosystem.

3 Nanopore sequencing in One Health

All experiential studies in this thesis used the Illumina platform, which is the current standard for metagenomic studies. The Illumina platform provides accurate reads (i.e. low error rate) and large amounts of data, but is limited by sequence length, turnaround time and the necessity to pool samples to reduce per-sample cost (Buermans and Den Dunnen, 2014; Fuller et al., 2009; Ma et al., 2017). Oxford Nanopore Technologies has released a small, portable, inexpensive sequencing platform called the MinION, which has tremendous potential to change the applicability of metagenomics (Laver et al., 2015; Mikheyev and Tin, 2014; Xu et al., 2018). Proof of concept studies using cell culture, bacterial mock communities and clinical samples with high pathogen loads have shown the benefits and applications of real-time data analysis (Benítez-Páez et al., 2016; Benítez-Páez & Sanz, 2017; Bradley et al., 2015; Quick et al., 2017). For example, multiple studies have used real-time data analysis to examine antibiotic resistance of bacterial species (Judge et al., 2015; Stakenborg et al., 2005) and have shown that complete real-time resistance profiles can be determined after 10 hours of sequencing (Cao et al., 2016). However, the current nanopore technology still has limitations in terms of error-rate, required sample quality, and data analysis (Ma et al., 2017). Many of these challenges are being addressed (Cao et al., 2016; Xu et al., 2018) and further development of nanopore technology will make it a robust and superior alternative to the current SBS sequencing methods (Lyon et al., 2018; Minei et al., 2018; Voskoboinik et al., 2018). The long read lengths (up to 10kb in length) will allow for a much more detailed understanding of entire microbial communities and thus provide us with a far more detailed understanding of microbial community structures (Xu et al., 2018). Therefore, personalized medicine for infectious diseases, targeted antibiotics treatment, microbiota monitoring for early disease prevention, and immediate pathogen screening in wildlife, may all become reality in the near future (Ku and Roukos, 2013; Lemon et al., 2012; Miller et al., 2013).

4 Concluding remarks

This study examined targeted metagenomic approaches in the One Health field. I highlighted the importance of standardization of terms, inclusion of appropriate controls and the use of established sample handling conditions. Without standardization, communications and collaboration in the interdisciplinary One Health field will not be effective at producing meaningful metagenomic data for cross-study comparisons. Based on the established standardized procedures, I demonstrated the applicability of target metagenomics in two studies addressing different aspects within the One Health triad. The urogenital microbiota and *Treponema* metataxonomic study were conducted under controlled laboratory conditions and thus provide important baseline data for future investigations. Further characterization of various microbiotas with a focus on *Treponema* diversity will provide informative data about microorganism-host relationships not just in primates and Australian marsupials, but also in various other hosts and ecosystems. Continually adapting current metagenomic applications to new forms of standardization and novel sequencing technology (e.g. MinION) is crucial to further improve our understanding of the microbial world.

References (Chapter A and C)

Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. 2014. The placenta harbors a unique microbiome. *Sci Transl Med* 6: 237ra265.

Abt MC, Pamer EG. 2014. Commensal bacteria mediated defenses against pathogens. *Curr Opin Immunol* 29: 16-22.

Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* 71: 8966-8969.

Adams RI, Bateman AC, Bik HM, Meadow JF. 2015. Microbiota of the indoor environment: a meta-analysis. *Microbiome* 3: 49.

Ahn JH, Kim BY, Song J, Weon HY. 2012. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *J Microbio* 50: 1071-1074.

Albertsen M, Karst SM, Ziegler AS, Kirkegaard RH, Nielsen PH. 2015. Back to basics–the influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. *PLoS One* 10: e0132783.

Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59: 143-169.

Amarasekara R, Jayasekara RW, Senanayake H, Dissanayake VH. 2015. Microbiome of the placenta in pre-eclampsia supports the role of bacteria in the multifactorial cause of pre-eclampsia. *J Obstet Gynaecol Res* 41: 662-669.

American Veterinary Medical Association. 2008. One health: a new professional imperative. One health initiative task force: final report. Available at: http://www.avma.org/onehealth/default.asp.

Antony KM, Ma J, Mitchell KB, Racusin DA, Versalovic J, Aagaard K. 2015. The preterm placental microbiome varies in association with excess maternal gestational weight gain. *Am J Obstet Gynecol* 212: 653.e1–653.16.

Ashford R. 2003. When is a reservoir not a reservoir?. *Emerg Infect Dis* 9: 1495-1496.

Bahrndorff S, Alemu T, Alemneh T, Lund Nielsen J. 2016. The microbiome of animals: implications for conservation biology. *Int J Genomics* 2016: 5304028.

Bai G, Gajer P, Nandy M, Ma B, Yang H, Sakamoto J, Blanchard MH, Ravel J, Brotman RM. 2012. Comparison of storage conditions for human vaginal microbiome studies. *PLoS One* 7: e36934.

Baker GC, Smith JJ, Cowan DA. 2003. Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55: 541-555.

Bäumler A, Fang FC. 2013. Host specificity of bacterial pathogens. *Cold Spring Harb Perspect Med* 3: a010041.

Bauer MA, Kainz K, Carmona-Gutierrez D, Madeo F. 2018. Microbial wars: Competition in ecological niches and within the microbiome. *Microb Cell* 5: 215-219.

Belkaid Y, Hand TW. 2014. Role of the microbiota in immunity and inflammation. *Cell* 157: 121-141.

Bendtsen KM, Fisker L, Hansen AK, Hansen CH, Nielsen DS. 2015. The influence of the young microbiome on inflammatory diseases—lessons from animal studies. *Birth Defects Res C Embryo Today* 105: 278-295.

Benítez-Páez A, Portune KJ, Sanz Y. 2016. Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION™ portable nanopore sequencer. *Gigascience* 5: 4.

Benítez-Páez A, Sanz Y. 2017. Multi-locus and long amplicon sequencing approach to study microbial diversity at species level using the MinION™ portable nanopore sequencer. *Gigascience* 6: gix043.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456: 53-59.

Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK. 2012. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol Evol* 27: 233-243.

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E. 2005. Defining operational taxonomic units using DNA barcode data. *Philos Trans R Soc Lond B Biol Sci* 360: 1935-1943.

Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, De Cesare M, et al. 2015. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun* 6: 10063.

Brooks JP, Edwards DJ, Harwich MD, Rivera MC, Fettweis JM, Serrano MG, Reris RA, Sheth NU, Huang B, Girerd P, et al. 2015. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol* 15: 66.

Buermans HP, Den Dunnen JT. 2014. Next generation sequencing technology: advances and applications. *Biochim Biophys Acta* 1842: 1932-1941.

Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, et al. 2015. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* 517: 205-208.

Callaway E, Cyranoski D. 2015. Anti-parasite drugs sweep Nobel prize in medicine 2015. *Nature News* 526: 174-175.

Cao MD, Ganesamoorthy D, Elliott AG, Zhang H, Cooper MA, Coin LJ. 2016. Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinION TM sequencing. *Gigascience* 5: 32.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS* 108: 4516-4522.

Carlos N, Tang YW, Pei Z. 2012. Pearls and pitfalls of genomics-based microbiome analysis. *Emerg Microbes Infect* 1: e45.

Carrigg C, Rice O, Kavanagh S, Collins G, O'Flaherty V. 2007. DNA extraction method affects microbial community profiles from soils and sediment. *Appl Microbiol Biotechnol* 77: 955-964.

Carroll D, Daszak P, Wolfe ND, Gao GF, Morel CM, Morzaria S, Pablos-Méndez A, Tomori O, Mazet JA. 2018. The global virome project. *Science* 359: 872-874.

Caulfield T, Sipp D, Murry CE, Daley GQ, Kimmelman J. 2016. Confronting stem cell hype. *Science* 352: 776-777.

Chakravorty S, Helb D, Burday M, Connell N, Alland D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* 69: 330-339.

Choi BK, Wyss C, Göbel UB. 1996. Phylogenetic analysis of pathogen-related oral spirochetes. *J Clin Microbiol* 34: 1922-1925.

Choo JM, Leong LE, Rogers GB. 2015. Sample storage conditions significantly influence faecal microbiome profiles. *Sci Rep* 5: 16350.

Clayton JB, Gomez A, Amato K, Knights D, Travis DA, Blekhman R, Knight R, Leigh S, Stumpf R, Wolf T, Glander KE. 2018. The gut microbiome of nonhuman primates: Lessons in ecology and evolution. *Am J Primatol* 80: e22867.

Cohan FM. 2002. What are bacterial species?. *Annu Rev Microbiol* 56: 457-487.

Cunningham AA, Daszak P, Wood JL. 2017. One Health, emerging infectious diseases and wildlife: two decades of progress? *Philos Trans R Soc Lond B Biol Sci* 372: 20160167.

Dapp MJ, Heineman RH, Mansky LM. 2013. Interrelationship between HIV-1 fitness and mutation rate. *J Mol Biol* 425: 41-53.

Daszak P, Cunningham AA, Hyatt AD. 2000. Emerging infectious diseases of wildlife--threats to biodiversity and human health. *Science* 287: 443-449.

David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, et al. 2014. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505: 559-563.

De Menezes AB, Lewis E, O'Donovan M, O'Neill BF, Clipson N, Doyle EM. 2011. Microbiome analysis of dairy cows fed pasture or total mixed ration diets. *FEMS Microbiol Ecol* 78: 256-265.

Dennison NJ, Jupatanakul N, Dimopoulos G. 2014. The mosquito microbiota influences vector competence for human pathogens. *Curr Opin Insect Sci* 3: 6-13.

Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. 2010. The human oral microbiome. *J Bacteriol* 192: 5002-5017.

Drexler JF, Corman VM, Müller MA, Maganga GD, Vallo P, Binger T, Gloza-Rausch F, Cottontail VM, Rasche A, Yordanov S, et al. 2012. Bats host major mammalian paramyxoviruses. *Nat Commun* 3: 796.

Dröge S, Rachel R, Radek R, König H. 2008. *Treponema isoptericolens* sp. nov., a novel spirochaete from the hindgut of the termite *Incisitermes tabogae*. *Int J Syst Evol Microbiol* 58: 1079-1083.

Dutilh BE, Reyes A, Hall RJ, Whiteson KL. 2017. Virus discovery by metagenomics: The (im) possibilities. *Front Microbiol* 8: 1710.

Edgar RC. 2018. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* 34: 2371-2375.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323: 133-138.

Essack, SY. 2018. Environment: the neglected component of the One Health triad. *Lancet Planet Health* 2: e238-e239.

FAO/OIE/WHO Joint Scientific Consultation Writing Committee. 2011. Influenza and other emerging zoonotic diseases at the human-animal interface. Proceedings of the FAO/OIE/WHO Joint Scientific Consultation, 27-29 April 2010, Verona (Italy).

Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, et al. 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl Environ Microbiol* 73: 7059-7066.

Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323: 741-746.

Frassetto LA, Lanham-New SA, Macdonald HM, Remer T, Sebastian A, Tucker KL, Tylavsky FA. 2007. Standardizing terminology for estimating the diet-dependent net acid load to the metabolic system. *J Nutr* 137: 1491-1492.

Fritz JV, Desai MS, Shah P, Schneider JG, Wilmes P. 2013. From meta-omics to causality: experimental models for human microbiome research. *Microbiome* 1: 14.

Fuller CW, Middendorf LR, Benner SA, Church GM, Harris T, Huang X, Jovanovich SB, Nelson JR, Schloss JA, Schwartz DC, Vezenov DV. 2009. The challenges of sequencing by synthesis. *Nature Biotechnol* 27: 1013-1023.

Gardy JL, Loman NJ. 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet* 19: 9-20.

Goldberg B, Sichtig H, Geyer C, Ledeboer N, Weinstock GM. 2015. Making the leap from research laboratory to clinic: challenges and opportunities for next-generation sequencing in infectious disease diagnostics. *mBio* 6: e01888-15.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17: 333-351.

Graber JR, Leadbetter JR, Breznak JA. 2004. Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl Environ Microbiol* 70: 1315-1320.

Hallmaier-Wacker LK, Munster VJ, Knauf S. 2017. Disease reservoirs: from conceptual frameworks to applicable criteria. *Emerg Microbes Infect* 6: e79.

Hallmaier-Wacker LK, Lüert S, Roos C, Knauf S. 2018. The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Sci Rep* 8: 6292.

Hallmaier-Wacker LK, Lüert S, Roos C, Knauf S. 2019a. Lactation and menstruation shift the vaginal microbiota in captive rhesus monkeys to be more similar to the male urethral microbiota. *Sci Rep* 9: 17399.

Hallmaier-Wacker LK, Lüert S, Gronow S, Spröer C, Overmann J, Buller N, Vaughan-Higgins RJ, Knauf S. 2019b. A metataxonomic tool to investigate the diversity of Treponema. *Front Microbiol* 10: 2094.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5: R245-R249.

Handelsman J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol* 68: 669-685.

Haus T, Ferguson B, Rogers J, Doxiadis G, Certa U, Rose NJ, Teepe R, Weinbauer GF, Roos C. 2014. Genome typing of nonhuman primate models: implications for biomedical research. *Trends Genet* 30: 482-487.

Haydon DT, Cleaveland S, Taylor LH, Laurenson MK. 2002. Identifying reservoirs of infection: a conceptual and practical challenge. *Emerg Infect Dis* 8: 1468-1473.

Heather JM, Chain B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics* 107: 1-8.

Hunkapiller T, Kaiser RJ, Koop BF, Hood L. 1991. Large-scale and automated DNA sequence determination. *Science* 254: 59-67.

Jain M, Tyson JR, Loose M, Ip CL, Eccles DA, O'Grady J, Malla S, Leggett RM, Wallerman O, Jansen HJ, et al. 2017. MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9.0 chemistry. *F1000Research* 6: 760.

Joly D, Johnson CK, Goldstein T, Anthony SJ, Karesh W, Daszak P, Wolfe N, Murray S, Mazet J. 2016. The first phase of PREDICT: Surveillance for emerging infectious zoonotic diseases of wildlife origin (2009-2014). *Int J Infect Dis* 53: 31-32.

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008. Global trends in emerging infectious diseases. *Nature* 451: 990-993.

Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, Fabani MM, Seguritan V, Green J, Pride DT, et al. 2015. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *PNAS* 112: 14024-14029.

Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, Perry T, Kao D, Mason AL, Madsen KL, Wong GK. 2016. Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front Microbiol* 7: 459.

Judge K, Harris SR, Reuter S, Parkhill J, Peacock SJ. 2015. Early insights into the potential of the Oxford Nanopore MinION for the detection of antimicrobial resistance genes. *J Antimicrob Chemother* 70: 2775-2778.

Jumpstart Consortium Human Microbiome Project Data Generation Working Group. 2012. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS One* 7: e39315.

Kao RR, Haydon DT, Lycett SJ, Murcia PR. 2014. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. *Trends Microbiol* 22: 282-291.

Karesh WB, Dobson A, Lloyd-Smith JO, Lubroth J, Dixon MA, Bennett M, Aldrich S, Harrington T, Formenty P, Loh EH, et al. 2012. Ecology of zoonoses: natural and unnatural histories. *Lancet* 380: 1936-1945.

Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. 2011. Human nutrition, the gut microbiome and the immune system. *Nature* 474: 327-336.

Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, Lauder A, Sherrill-Mix S, Chehoud C, Kelsen J, et al. 2017. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome* 5: 52.

Kleiner M, Hooper LV, Duerkop BA. 2015. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* 16: 7.

Konstantinidis KT, Ramette A, Tiedje JM. 2006. The bacterial species definition in the genomic era. *Philos Trans R Soc Lond B Biol Sci* 361: 1929-1940.

Kopac S, Cohan FM. 2011. A theory-based pragmatism for discovering and classifying newly divergent bacterial species. In: Tibayrenc M (ed). Genetics and Evolution of Infectious Disease. London: Elsevier, 21-41.

Kostic AD, Howitt MR, Garrett WS. 2013. Exploring host–microbiota interactions in animal models and humans. *Genes Dev* 27: 701-718.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79: 5112-5120.

Kraemer JG, Ramette A, Aebi S, Oppliger A, Hilty M. 2018. Influence of pig farming on the human nasal microbiota: key role of airborne microbial communities. *Appl Environ Microbiol* 84: e02470-02417.

Krynak KL, Burke DJ, Martin RA, Dennis PM. 2017. Gut microbiome composition is associated with cardiac disease in zoo-housed western lowland gorillas (*Gorilla gorilla gorilla*). *FEMS Microbiol Lett* 364: fnx149.

Ku CS, Roukos DH. 2013. From next-generation sequencing to nanopore sequencing technology: paving the way to personalized genomic medicine. *Expert Rev Med Devices* 10: 1-6.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, Leite R, Elovitz MA, Parry S, Bushman FD. 2016. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* 4: 29.

Laurence M, Hatzis C, Brash DE. 2014. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLos One* 9: e97876.

Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. 2015. Assessing the performance of the Oxford nanopore technologies MinION. *Biomol Detect Quantif* 3: 1-8.

Lax S, Sangwan N, Smith D, Larsen P, Handley KM, Richardson M, Guyton K, Krezalek M, Shogan BD, Defazio J, Flemming I. 2017. Bacterial colonization and succession in a newly opened hospital. *Sci Transl Med* 9: eaah6500.

Lederberg J, McCray AT. 2001. 'Ome sweet 'omics—a genealogical treasury of words. *Scientist* 15: 8-10.

Lemon KP, Armitage GC, Relman DA, Fischbach MA. 2012. Microbiota-targeted therapies: an ecological perspective. *Sci Transl Med* 4: 137rv135.

Lilburn TG, Kim KS, Ostrom NE, Byzek KR, Leadbetter JR, Breznak JA. 2001. Nitrogen fixation by symbiotic and free-living spirochetes. *Science* 292: 2495-2498.

Lilburn TG, Schmidt TM, Breznak JA. 1999. Phylogenetic diversity of termite gut spirochaetes. *Environ Microbiol* 1: 331-345.

Lindgreen S, Adair KL, Gardner PP. 2016. An evaluation of the accuracy and speed of metagenome analysis tools. *Sci Rep* 6: 19233.

Locey KJ, Lennon JT. 2016. Scaling laws predict global microbial diversity. *PNAS* 113: 5970-5975.

Lucius R, Loos-Frank B, Lane RP, Poulin R, Roberts C, Grencis RK. 2017. The biology of parasites. Weinheim: Wiley VCH Verlag.

Lyon WJ, Smith ZK, Geier B, Baldwin JC, Starr CR. 2018. Evaluating an Upper Respiratory Disease Panel on the Portable MinION Sequencer. *bioRxiv*: 436600.

Ma X, Stachler E, Bibby K. 2017. Evaluation of Oxford nanopore MinION sequencing for 16S rRNA microbiome characterization. *bioRxiv:* 099960.

Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, Pavlenok M, Niederweis M, Gundlach JH. 2012. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol* 30: 349-353.

Marchesi JR, Ravel J. 2015. The vocabulary of microbiome research: a proposal. *Microbiome* 3: 31.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.

Marks M, Fookes M, Wagner J, Butcher R, Ghinai R, Sokana O, Sarkodie YA, Lukehart SA, Solomon AW, Mabey DC, Thomson N. 2017. Diagnostics for Yaws Eradication: Insights From Direct Next-Generation Sequencing of Cutaneous Strains of *Treponema pallidum*. *Clin Infect Dis* 66: 818-824.

Marra CM, Handsfield HH, Kuller L, Morton WR, Lukehart SA. 1992. Alterations in the course of experimental syphilis associated with concurrent simian immunodeficiency virus infection. *J Infect Dis* 165: 1020-1025.

Matic I, Radman M, Taddei F, Picard B, Doit C, Bingen E, Denamur E, Elion J. 1997. Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* 277: 1833-1834.

McMahon S, Parnell J. 2014. Weighing the deep continental biosphere. *FEMS Microbiol Ecol* 87: 113-120.

Metzker ML. 2010. Sequencing technologies—the next generation. *Nature Rev Genet* 11: 31-46.

Mikheyev AS, Tin MM. 2014. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour* 14: 1097-1102.

Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P. 2013. Metagenomics for pathogen detection in public health. *Genome Med* 5: 81.

Minei R, Hoshina R, Ogura A. 2018. De novo assembly of middle-sized genome using MinION and Illumina sequencers. *BMC Genomics* 19: 700.

Misic AM, Davis MF, Tyldsley AS, Hodkinson BP, Tolomeo P, Hu B, Nachamkin I, Lautenbach E, Morris DO, Grice EA. 2015. The shared microbiota of humans and companion animals as evaluated from *Staphylococcus* carriage sites. *Microbiome* 3: 2.

Mysara M, Vandamme P, Props R, Kerckhof FM, Leys N, Boon N, Raes J, Monsieurs P. 2017. Reconciliation between operational taxonomic units and species boundaries. *FEMS Microbiol Ecol* 93: fix029.

Nguyen D, Xu T. 2008. The expanding role of mouse genetics for understanding human biology and disease. *Dis Model Mech* 1: 56-66.

Nordhoff M, Taras D, Macha M, Tedin K, Busse HJ, Wieler LH. 2005. *Treponema berlinense* sp. nov. and *Treponema porcinum* sp. nov., novel spirochaetes isolated from porcine faeces. *Int J Syst Evol Microbiol* 55: 1675-1680.

Norman JM, Handley SA, Virgin HW. 2014. Kingdom-agnostic metagenomics and the importance of complete characterization of enteric microbial communities. *Gastroenterology* 146: 1459-1469.

Orth R, O'Brien-Simpson N, Dashper S, Walsh K, Reynolds E. 2010. An efficient method for enumerating oral spirochetes using flow cytometry. *J Microbiol Methods* 80: 123-128.

Osterhaus A, MacKenzie J. 2016. The 'One Health'journal: filling a niche. *One Health* 2: 18.

Pace NR. 1985. Analyzing natural microbial populations by rRNA sequences. *ASM News* 51: 4-12.

Pace NR. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276: 734-740.

Pallen MJ. 2014. Diagnostic metagenomics: potential applications to bacterial, viral and parasitic infections. *Parasitology* 141: 1856-1862.

Pascal M, Perez-Gordo M, Caballero T, Escribese MM, Longo MN, Luengo O, Manso L, Matheu V, Seoane E, Zamorano M, Labrador M. 2018. Microbiome and allergic diseases. *Front Immunol* 9: 1584.

Patwardhan A, Ray S, Roy A. 2014. Molecular markers in phylogenetic studies—a review. *J Phylogen Evol Biol* 2: 131.

Prakash T, Taylor TD. 2012. Functional assignment of metagenomic data: challenges and applications. *Brief Bioinform* 13: 711-727.

Prescott MJ, Langermans JA, Ragan I. 2017. Applying the 3Rs to non-human primate research: Barriers and solutions. *Drug Discov Today Dis Models* 23: 51-56.

Qiu Y, Nakao R, Ohnuma A, Kawamori F, Sugimoto C. 2014. Microbial population analysis of the salivary glands of ticks; a possible strategy for the surveillance of bacterial pathogens. *PLoS One* 9: e103961.

Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, et al. 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 12: 1261-1276.

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 35: 833-844.

Rabinowitz PM, Kock R, Kachani M, Kunkel R, Thomas J, Gilbert J, Wallace R, Blackmore C, Wong D, Karesh W, et al. 2013. Toward proof of concept of a One Health approach to disease prediction and control. *Emerg Infect Dis* 19: e130265.

Raebel MA, Schmittdiel J, Karter AJ, Konieczny JL, Steiner JF. 2013. Standardizing terminology and definitions of medication adherence and persistence in research employing electronic databases. *Med Care* 51: S11-21.

Razzauti M, Galan M, Bernard M, Maman S, Klopp C, Charbonnel N, Vayssier-Taussat M, Eloit M, Cosson JF. 2015. A comparison between transcriptome sequencing and 16S metagenomics for detection of bacterial pathogens in wildlife. *PLoS Negl Trop Dis* 9: e0003929.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499: 431-437.

Roach JC, Boysen C, Wang K, Hood L. 1995. Pairwise end sequencing: a unified approach to genomic mapping and sequencing. *Genomics* 26: 345-353.

Robinson ER, Walker TM, Pallen MJ. 2013. Genomics and outbreak investigation: from sequence to consequence. *Genome Med* 5: 36.

Rosselló-Mora R, Amann R. 2001. The species concept for prokaryotes. *FEMS Microbiol Rev* 25: 39-67.

Rossen NG, Fuentes S, van der Spek MJ, Tijssen JG, Hartman JH, Duflou A, Löwenberg M, van den Brink GR, Mathus-Vliegen EM, de Vos WM, et al. 2015. Findings from a randomized controlled trial of fecal transplantation for patients with ulcerative colitis. *Gastroenterology* 149: 110-118.e114.

Rossi-Tamisier M, Benamar S, Raoult D, Fournier PE. 2015. Cautionary tale of using 16S rRNA gene sequence similarity values in identification of human-associated bacterial species. *Int J Syst Evol Microbiol* 65: 1929-1934.

Round JL, Mazmanian SK. 2009. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 9: 313-323.

Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12: 87.

Sanger F, Nicklen S, Coulson AR. 1997. DNA sequencing with chain-terminating inhibitors. *PNAS* 74: 5463-5467.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75: 7537-7541.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *PNAS* 109: 6241-6246.

Scott MR, Will R, Ironside J, Nguyen HO, Tremblay P, DeArmond SJ, Prusiner SB. 1999. Compelling transgenetic evidence for transmission of bovine spongiform encephalopathy prions to humans. *PNAS* 96: 15137-15142.

Seleem MN, Boyle SM, Sriranganathan N. 2010. Brucellosis: a re-emerging zoonosis. *Vet Microbiol* 140: 392-398.

Shah N, Tang H, Doak TG, Ye Y. 2011. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pac Symp Biocomput*: 165-176.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nature Biotechnol* 26: 1135-1145.

Simon C, Daniel R. 2011. Metagenomic analyses: past and future trends. *Appl Environ Microbiol* 77: 1153-1161.

Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielny S, et al. 2013. Cohabiting family members share microbiota with one another and with their dogs. *eLife* 2: e00458.

Spear GT, Gilbert D, Sikaroodi M, Doyle L, Green L, Gillevet PM, Landay AL, Veazey RS. 2010. Identification of rhesus macaque genital microbiota by 16S pyrosequencing shows similarities to human bacterial vaginosis: implications for use as an animal model for HIV vaginal infection. *AIDS Res Hum Retroviruses* 26: 193-200.

Stakenborg T, Vicca J, Butaye P, Maes D, Minion FC, Peeters J, De Kruif A, Haesebrouck F. 2005. Characterization of in vivo acquired resistance of *Mycoplasma hyopneumoniae* to macrolides and lincosamides. *Microb Drug Resist* 11: 290-294.

Staley JT. 2006. The bacterial species dilemma and the genomic–phylogenetic species concept. *Philos Trans R Soc Lond B Biol Sci* 361: 1899-1909.

Strehlow RA. 1993. Standardizing terminology for better communication: practice, applied theory, and results (Vol. 1166). Philadelphia: ASTM International.

Swerdlow H, Gesteland R. 1990. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Res* 18: 1415-1419.

Taylor LH, Latham SM, Woolhouse ME. 2001. Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci* 356: 983-989.

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

Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P. 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 60: 249-266.

Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res* 38: e159.

Tringe SG, Hugenholtz P. 2008. A renaissance for the pioneering 16S rRNA gene. *Curr Opin Microbiol* 11: 442-446.

Trinh P, Zaneveld JR, Safranek S, Rabinowitz P. 2018. One Health Relationships Between Human, Animal, and Environmental Microbiomes: A Mini-Review. *Front Public Health* 6: 235.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* 449: 804-810.

Van de Peer Y, Chapelle S, De Wachter R. 1996. A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Res* 24: 3381-3391.

Van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. 2018. The third revolution in sequencing technology. *Trends Genet* 34: 666-81.

Vaughan RJ. 2008. A health and disease study of Australia's most critically endangered mammal, the Gilbert's potoroo (*Potorous gilbertii*). PhD Thesis, Veterinary Science Division of Conservation Medicine, Murdoch University. Available from: http://researchrepository.murdoch.edu.au

Vaughan RJ, Warren KS, Mills JS, Palmer C, Fenwick S, Monaghan CL, Friend AJ. 2009. Hematological and serum biochemical reference values and cohort analysis in the Gilbert's potoroo (*Potorous gilbertii*). *J Zoo Wildl Med* 40: 276-288.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. 2001. The sequence of the human genome. *Science* 291: 1304-1351.

Vesty A, Biswas K, Taylor MW, Gear K, Douglas RG. 2017. Evaluating the Impact of DNA Extraction Method on the Representation of Human Oral Bacterial and Fungal Communities. *PLoS One* 12: e0169877.

Viana M, Mancy R, Biek R, Cleaveland S, Cross PC, Lloyd-Smith JO, Haydon DT. 2014. Assembling evidence for identifying reservoirs of infection. *Trends Ecol Evol* 29: 270-279.

Vinkers CH, Tijdink JK, Otte WM. 2015. Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis. *BMJ* 351: h6467.

Voskoboinik L, Motro U, Darvasi A. 2018. Facilitating complex DNA mixture interpretation by sequencing highly polymorphic haplotypes. *Forensic Sci Int Genet* 35: 136-140.

Wagner Mackenzie B, Waite DW, Taylor MW. 2015. Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Front Microbiol* 6: 130.

Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 2015. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome* 3: 26.

Wang JW, Kuo CH, Kuo FC, Wang YK, Hsu WH, Yu FJ, Hu HM, Hsu PI, Wang JY, Wu DC. 2018. Fecal microbiota transplantation: Review and update. *J Formos Med Assoc* 118: S21-S31.

Woolhouse ME, Gowtage-Sequeria S. 2005. Host range and emerging and reemerging pathogens. *Emerg Infect Dis* 11: 1842-1847.

Wu Z, Lu L, Du J, Yang L, Ren X, Liu B, Jiang J, Yang J, Dong J, Sun L, Zhu Y. 2018. Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. *Microbiome* 6: 178.

Wyres K, Conway T, Garg S, Queiroz C, Reumann M, Holt K, Rusu L. 2014. WGS analysis and interpretation in clinical and public health microbiology laboratories:

what are the requirements and how do existing tools compare?. *Pathogen* 3: 437-458.

Xu Y, Lewandowski K, Lumley S, Pullan S, Vipond R, Carroll M, Foster D, Matthews PC, Peto T, Crook D. 2018. Detection of viral pathogens with multiplex Nanopore MinION sequencing: be careful with cross-talk. *Front Microbiol* 9: 2225.

Xu Y, Zhao F. 2018. Single-cell metagenomics: challenges and applications. *Protein Cell* 9: 501-510.

Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* 7: e33865.

Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367: 1814-1820.

Zheng J, Xiao X, Zhang Q, Mao L, Yu M, Xu J. 2015. The placental microbiome varies in association with low birth weight in full-term neonates. *Nutrients* 7: 6924-6937.

Summary

This thesis is meant to further the applicability of metagenomics in One Health by discussing key forms of standardization, as well as targeted applications. Metagenomics is a powerful pathogen-agnostic tool, allowing for broad-range non-targeted detection of microorganisms. This is especially useful for application in One Health investigations to detect known and unknown microorganisms in animals, humans and the environment. In this thesis, I highlight the benefits and limitations of metagenomics in One Health and urge for the continual adaptation to new forms of standardization, evolving technology, and novel areas of research.

Standardization of terms and techniques is a key challenge in One Health investigations as the field is highly interdisciplinary with many cross-continental collaborations and large-scale consortiums. This thesis addresses two forms of standardization to improve effective communication and interdisciplinary exchange. First, it highlights the need for an empirical framework to determine a 'disease reservoir', a central topic for pathogens of zoonotic or anthropozoonotic origin (Chapter B1). To exemplify the use of the proposed criteria and avoid misunderstandings, the established framework is applied to a variety of known zoonotic diseases. In addition to standardizing terminology, this thesis establishes appropriate handling conditions for swab samples for metagenomic studies (Chapter B2). Here, I emphasize the need for appropriate controls (blank controls and mock communities), standardized procedures and continual monitoring of contamination. The presented handling conditions were used for all subsequent procedures (Chapter B3 and B4) to ensure compatibility between datasets.

Based on the standardized procedures, this thesis demonstrates the applicability of target metagenomics in two studies investigating different aspects of the One Health triad. In both studies, the bacterial diversity was examined by sequencing a single hypervariable region of the 16S rRNA gene under controlled laboratory conditions and thus provide important baseline data for future investigations. The first study focused on examining the urogenital microbiota of rhesus monkeys (*Macaca mulatta*), a commonly-used translational animal model

(Chapter B3). Endocrine-regulated processes, cage-mate relationships and group associations were found to impact the urogenital microbiota of rhesus monkeys at the German Primate Center. Based on the observed plasticity of the urogenital microbiota, an increased awareness of microbiota considerations of translational animal models is essential. Additionally, the identified microbiota provides baseline data for One Health investigations on wild non-human primates in health and disease. Here, *Treponema* species have been previously identified in multiple taxa of wild non-human primates. To further explore the diversity of these medically and ecologically-relevant bacteria, the second application of metagenomics in this thesis focused on establishing and validating a metataxonomic tool to identify *Treponema* (Chapter B4). The incorporation of a spirochete-specific enrichment step and the modular amplicon approach allowed us to detect *Treponema* to a near species-level using only a small amount of DNA. The presented data from the in-silico and in-vitro experiments using mock communities and clinical samples provided confidence in the applicability of the metataxonomic approach. Further characterization of different microbiota with a focus on *Treponema* will provide informative data about microorganism-host relationships in various animals and ecosystems within the One Health triad.

Acknowledgments

I would like to thank Dr. Sascha Knauf for giving me the opportunity to work in this fascinating field. Thank you for your continual encouragement, for showing me Lake Manyara National Park and for always supporting me to go to conferences around the world. Working with you has made me a passionate scientist, a more critical thinker and a better teacher.

I would also like to thank my thesis committee members, PD Dr. Christian Roos and Dr. Dietmar Zinner, for their instrumental feedback and continual advice.

I am very thankful to all my colleagues both in Germany and Tanzania, working in an international and interdisciplinary team has been wonderful. First and foremost, to Simone Lüert for her technical advice, moral support and the continual flow of snacks during my PhD. Simone, I truly admire your dedication to our research. Also, thank you to Felipa, Chuma, Clara, Lena, and Elaine for keeping me company and giving me advice throughout my time at the German Primate Center.

I am grateful to Uwe Schönmann for his general support. A special thank you to Dr. Tamara Becker, Dr. Annette Schrod, Annette Husung, Wolfgang Henkel, Melina Urh and all the animal caretakers for their help during sampling at the German Primate Center. Thank you to Daniel Reckel and Dr. Angela Noll for your help with command line and fixing numerous bugs.

Thank you to all the collaborators and co-authors that have made these projects possible. Thank you to Dr. Vincent Munster, PD Dr. Sabine Gronow, Dr. Cathrin Spröer, Prof. Dr. Jörg Overmann, and Dr. Rebecca Vaughan-Higgins. I would also like to thank Fabian Ludewig and Dr. Gabriela Salinas of the Transcriptome and Genome Analysis Laboratory for their assistance in optimizing all the sequencing runs.

Thank you to Dr. Jonas Bart for clarifying all the GAUSS regulations and organizing all the helpful workshops for graduate students. Thank you to everyone who proofread parts of this thesis, especially Shally Yuan for the language corrections.

Thank you to my roommates, Lena, Kathinka, Alex, Jan, Marina, Thiemo, Marc, Manisha +1, for always pointing out Option D, for the mushroom fights in the kitchen, and for the summer BBQs. Thank you, Flavia, for showing me the Göttingen jungle, Rodolfo for introducing me to Brazilian children's songs and Yi for striking me out that one time. You are amongst the most precious things I gained from this experience.

Thank you to my family for their unconditional love, inspiration, and support. To Julian for making me laugh, to Hiltrudis for always listening and to Frank for being proud of me. Finally, to Ramiro, thank you for encouraging me to follow a dream of mine and for your trust, friendship, and love.