Dissertation

Unravelling the microbiome: From human gut to spacecraft assembly cleanrooms

submitted by

MSc, BSc

Mina BASHIR

for the Academic Degree of

Doctor of Philosophy

(PhD)

at the

Medical University of Graz

Department of Internal Medicine

Division of Endocrinology and Diabetology

under the Supervision of

Univ.-Prof. Dr. med. univ. Thomas R. Pieber

2016

Statutory Declaration

I hereby declare that this thesis is my own original work and that I have fully acknowledged by name all of those individuals and organisations that have contributed to the research for this thesis. Due acknowledgement has been made in the text to all other material used. Throughout this thesis and in all related publications I followed the “Standards of Good Scientific Practice and Ombuds Committee at the Medical University of Graz“

Date: Signature:_____

(Mina BASHIR)

Acknowledgements

First of all, I'd like to thank Univ.-Prof. Dr. med. univ. Thomas Pieber, who never ceased to surprise me with his profound scientific knowledge, for his supervision and his infecting passion for science.

I'm also very grateful for all my colleagues of the Doctoral College of Molecular inflammation (DK-MOLIN) and the program itself for enabling me to participate at international conferences and to spend a research stay abroad. Being surrounded by positive and motivated people, helps through the struggle of a PhD. It was a pleasure working with you and I'm thankful for every single one.

Endoscopic procedures and biopsy acquisition were kindly performed by Christoph Högenauer and Patrizia Kump. 16S rRNA amplicon sequencing was partly performed under the supervision of Ingeborg Klymiuk and her team at the core facility molecular biology at the center of medical research, Graz. Isolation of Lamina propria mono nuclear cells and flow cytometry were kindly performed by Barbara Prietl and her team.

Cleanroom sampling and sample preparation was performed by Parag A. Vaishampayan and team and functional metagenomic sequencing was done at the United States Department of Energy (DOE) Joint Genome Institute (JGI) in Walnut Creek, CA, USA.

I also want to express my gratitude to all my colleagues at NASA Jet Propulsion Laboratory (JPL), Pasadena, CA, USA and Parag Vaishampayan, who kindly welcomed me at my research stay. You weren't just a great mentor, but also became a good friend. I really consider myself lucky to have had the pleasure of meeting all of you.

Ultimately, I want to thank my family and friends for their support and for making this PhD such a unique experience.

Table of Contents

Abbreviations and Definitions

Abbreviation	Definition
16S rRNA	16 Svedberg ribosomal ribonucleic acid
A. baumannii	*Acinetobacter baumannii*
A. lwoffii	*Acinetobacter lwoffii*
AC	ascending colon
ANOSIM	analyisis of similarity
AO	appendiceal orifice
BMI	body mass index
bp	base pairs
CAMP	cathalicidin antimicrobial peptide
cDNA	complementary DNA
cm	centimeter
COSPAR	Committee of Space Research
cRPMI-10	Roswell Park Memorial Institute media containing 10% fetal bovine serum
CTX	cholera toxin
CYP27B1	cytochrome P450 family 27 subfamily B member 1; 1-alpha-hydroxylase
DD	duodenum
DGGE	denaturing gradient gel electrophoresis
DI-MS	direct infusion mass-spectrometry
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EF-Tu	elongation factor thermo unstable
emPCR	emulsion PCR
ESKAPE	*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumanii, Pseudomonas aeruginosa*, and *Enterobacter species*
FDR	false discovery rate

fg	femtogram
FISH	fluorescence in situ hybridization
GA	gastric antrum
GC	gastric corpus
GC-MS	gas chromatography mass-spectrometry
GSE	ground support equippment
gyrB	gyrase B, topoisomerase
H. pylori	*Helicobacter pylori*
HCl	hydrogen chloride
HEPA	high-efficiency particulate air
IBD	inflammatory bowel disease
IBS	inflammatory bowel syndrome
ichip	isolation chip
IU	international units
JPL-SAF	Jet Propulsion Laboratory's spacecraft assembly facility
JSON	JavaScript Object Notation
kDa	kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
kg	kilogram
KO	KEGG orthologous
KSC-PHSF	Kennedy Space Center's Payload Hazardous Servicing Facility
L	liter
L. pneumophila	Legionella pneumophila
LB	Luria Broth
LC-MS	liquid chromatography mass-spectrometry
LMA-MTF	Lockheed Martin Aeronautics' Multiple Testing Facility
m	meter
MDA	multiple displacement amplification
mg	miligram
MID	multiplexing identifier
mL	mililiter
mm	milimeter

mM	milimol per liter
mmol	milimol
MPC	magnetic particle concentrator
mRNA	messenger RNA
MS	mass-spectrometry
MSL	Mars Science Laboratory
n	number
N50	length N for which 50% of all bases in the sequences are in a sequence of length
NaOH	Sodium hydroxide
NASA	National Aeronautics and Space Administration
ng	nanogram
NMR	nuclear magnetic resonance
OTU	operational taxonomic unit
PBS	phosphate buffer saline
PC	prinicple component
PCoA	priniciple coordinates analysis
PCR	polymerase chain reaction
PHX-A	sample taken from KSC-PHSF after assembly and testing of the Phoenix spacecraft
PHX-B	sample taken from KSC-PHSF before arrival of the Phoenix spacecraft
PHX-D	sample taken from KSC-PHSF during assembly and testing of the Phoenix spacecraft
pmol	picomol
PMA	propidium monoazide
PTH	parathyroid hormone
QIIME	quantitative insight into microbial ecology
RDP	ribosomal database project
SC	sigmoid colon
SD	standard deviation
SEM	standard error of the mean
TE	tris/EDTA

TI	terminal ileum
UV	ultraviolet
VDR	vitamin D receptor
µL	microliter

Abstract in German

Obwohl neue Technologien die Kultivierung einer Vielzahl von Mikroorganismen ermöglicht haben, lässt sich der Großteil der Mikroorganismen nicht in Laboren züchten. Dank kultivierungsunabhängigen Methoden, die sich neuer Sequenzierungsmethoden bemächtigen, kann man erste Einblicke in diese unbekannten Mikroökosysteme gewinnen. Diese Dissertation beschreibt eine 16S rRNA Amplikon basierte klinische Studie, die den Einfluss von oral verabreichtem Vitamin D_3 auf die mukosale Darmflora des Menschen untersucht und eine funktionelle Metagenomstudie mit speziellem Fokus auf Pathogene und deren Virulenzfaktoren in NASA Reinräumen, die dem Zusammenbau von Raumfahrzeugen dienen.

Vitamin D wirkt sich erwiesenermaßen positiv auf die menschliche Gesundheit aus. Die Auswirkungen von Vitamin D auf die Darmflora blieben jedoch bis dato unerforscht. Im Rahmen einer Interventionsstudie wurden 16 gesunde Freiwillige (7 Frauen, 9 Männer) vor und nach einer 8-wöchigen Vitamin D Kur endoskopiert um Veränderungen der Darnflora in 7 verschiedenen Regionen des Verdauungstraktes und Stuhl zu untersuchen. Wir sahen signifikante Veränderungen im oberen, jedoch keine Bedeutenden im unteren Verdauungstrakt oder Stuhl. Im oberen Verdauungstrakt konnten wir eine signifikante Reduktion von Gammaproteobakterien, wie z. B. *Pseudomonas* spp. and *Escherichia/Shigella* spp. und eine Zunahme in der bakteriellen Diversität beobachten. Der Darmflora modulierende Effekt von Vitamin D könnten zum Teil für den positiven Einfluss von Vitamin D mitverantwortlich sein. Die Tatsache, dass sich die Veränderungen auf den oberen Verdauungstrakt begrenzt hielten zeigt, dass externe Faktoren einen ausschließlich regionsspezifischen Effekt haben können, was bei zukünftigen klinischen Studien berücksichtigt werden sollte.

Um andere Himmelskörper nicht mit terrestrischen Mikroorganismen zu kontaminieren, werden Raumfahrzeuge in speziellen Reinräumen, die strengen Reinigungs- und Dekontaminierungskriterien entsprechen müssen zusammengebaut und getestet. Wir untersuchten ob diese Sicherheitsmaßnahmen zur Anreicherung von hartnäckigen Mikroorganismen, wie z. B. Pathogene führen. Proben wurden von drei geographisch getrennten Reinräumen während der Assemblierung von DAWN,

Phoenix und Mars Science Laboratory (MSL), bei denen es sich um NASA Raumfahrzeuge handelt, gesammelt. Zusätzliche Proben vor und nach der Assemblierung von Phoenix wurden ebenfalls gesammelt. Alle Proben wurden auf einer HiSeq 2500 Plattform schrotschusssequenziert. Proben die während dem Zusammenbau von Raumschiffen genommen wurden, waren sich trotz geographischer Trennung ähnlich in ihrer mikrobiellen Zusammensetzung und bestanden hauptsachlich aus *Acinetobacter* spp. Alle Proben enthielten Pathogene und Virulenzfaktoren. Obwohl der relative Anteil an Pathogenen während dem Zusammenbau niedriger war, konnten wir eine Akkumulation von Virulenzfaktoren im Vergleich zum Vor- und Nachher feststellen. Die verringerte phylogenetische und pathogene Diversität während des Raumfahrzeugzusammenbaus bestätigt dass die Dekontaminations- und Vorsichtsmaßnahmen effektiv und gut implementiert sind. Vier potentielle Pathogene, *Acinetobacter baumannii, Acinetobacter lwoffii, Escherichia coli* und *Legionella pneumophila,* und ihre dazugehörigen Virulenzfaktoren wurden in allen Proben gefunden. Die Ergebnisse dieser Studie könnten sich für andere geschlossene Räume wie Schulen, Krankenhäuser oder sogar der Internationalen Raumstation als nützlich erweisen.

Abstract in English

Although recent advances in microbiology have shed some light into the uncultivable biome, the majority of microorganisms still do not grow under laboratory conditions. Cultivation-independent techniques exploiting advances in next-generation sequencing are applied to sneak peek into the hidden microbial flora. This thesis describes a phylogeny based 16S rRNA amplicon sequencing study investigation the effects of an oral vitamin D_3 supplementation on the mucosa associated gut microbiome and functional metagenomics study with focus on pathogens and their corresponding virulence factors in spacecraft assembly cleanrooms.

Vitamin D is well-known for positively influencing human health but its impact on the human gut microbiome is still unknown. We performed an interventional pilot study, including 16 healthy volunteers (7 women, 9 men), who were endoscopically examined. A total of 7 sites, covering the major regions and stools were sampled before and after 8 weeks of vitamin D_3 supplementation. Vitamin D_3 supplementation

changed the gut microbiome in the upper gastrointestinal tract, but lower regions including colon and stool showed no major changes. Gammaproteobacteria, including *Pseudomonas* spp. and *Escherichia/Shigella* spp. decreased while bacterial richness increased significantly. Modulation of the gut microbiome of the upper gastrointestinal tract might explain vitamin D's positive influence on gastrointestinal diseases. The localized effects of vitamin D display distinct regional differences in the response of the GI-microbiome to confounders, which should be considered in future microbiome studies.

Strict planetary protection practices are implemented during spacecraft assembly to prevent inadvertent transfer of earth microorganisms to other planetary bodies. Therefore, spacecraft are assembled in cleanrooms, which undergo strict cleaning and decontamination procedures to reduce total microbial bioburden. We wanted to evaluate if these practices selectively favor survival and growth of hardy microorganisms, such as pathogens. Three geographically distinct cleanrooms were sampled during the assembly of three NASA spacecraft: DAWN, Phoenix, and the Mars Science Laboratory (Curiosity). For Phoenix cleanroom was sampled at three time points: before arrival of the spacecraft, during the assembly and testing, and after removal of the spacecraft. All samples were subjected to metagenomic shotgun sequencing on an Illumina HiSeq 2500 platform. Strict decontamination procedures had a greater impact on microbial communities than sampling location. Samples collected during spacecraft assembly were dominated by *Acinetobacter* spp. We found pathogens and virulence factors, which determine pathogenicity and in all the samples tested during this study. Though the relative abundance of pathogens was lowest during the Phoenix assembly, observed virulence factors were higher during assembly compared to before and after assembly, indicating a potential survival advantage. Decreased phylogenetic and pathogenic diversity indicates that decontamination and preventative measures were effective and well implemented. Four potential pathogens, *Acinetobacter baumannii, Acinetobacter lwoffii, Escherichia coli* and *Legionella pneumophila,* and their corresponding virulence factors were present in all cleanroom samples. The results of this study should be considered for microbial monitoring of enclosed environments such as schools, homes, hospitals and more isolated habitation such the International Space Station and future manned missions to Mars.

1 Introduction

1.1 The great plate count anomaly

The term “the great plate count anomaly” was first used by Staley *et a*l. in 1985 to describe the discrepancy between the number of different microorganisms extracted from natural environments that form viable colonies on agar plates and the number obtained by microscopy. To the great astonishment of early microbiologists, it turned out that the majority of bacteria, approximately 99 %, do not grow under laboratory conditions [1]. For decades, microbiologist have tried to tune several parameters of culturing, including various cofactors and chemicals, different temperatures, varying pressure, atmospheres and much more. Nevertheless, the majority of microorganisms were not cultivable. Multiple theories have been proposed to explain this phenomenon. Some microorganisms might grow extremely slow, such as one division every other year. As it would take a lifetime until a colony is formed, most researchers decided not to tackle this task.

Another theory suggests that bacteria only grow in their natural environment. Only if they are in the right environment, in close proximity to certain other microorganism, which produce unknown growth factors, they commit to division. Therefore, they avoid the risk of being killed by unfamiliar conditions or predators. In 2002 Kaeberlein et al. hypothesized that uncultivable microorganisms might grow in culture if provided with the chemical components of their natural environment. They had reasoned that if you cannot get microorganisms to grow in the lab, you can bring the lab to the wild. Consequently, they developed a diffusion chamber for *in-situ* cultivation, which separates an isolate from its natural environment by a semi-permeable membrane, allowing required growth factors to traverse and stimulate the growth. This method allows phenotypic and genomic characterization of isolated microorganisms with access to their naturally occurring growth factors. Exploiting this approach, scientists were able to restore up to 50 % of the microscopically observed microorganisms [2].

Although, this *in situ* cultivation method recovers significantly more microorganisms, it is laborious and cannot be used for high-throughput microbial isolation. An upgraded version of this diffusion chamber, called isolation chip (ichip) [3] allows for parallel isolation and growth of hundreds of microorganisms in one step. It has hundreds of

miniaturized diffusion chambers and uses microfluidics to get single cells per chamber [3]. Using ichip, a novel antibiotic, named Teixobactin which is able to inhibit the growth of multiple multi-resistant human pathogens, was found and characterized. Teixobactin has no detectable resistance so far. It interferes with cell wall synthesis by targeting lipid II, the precursor of peptidoglycan and lipid III, precursor of lipoteichoic acid [4].

The microbial trap, an additional version of *in-situ* cultivation chambers, [5, 6] is in contrast to the previous described chambers initially sterile and allows the growth of Actinomycetes or other hyphae forming microorganisms, as it's equipped with a semipermeable membrane with a pore diameter of 0.2 µm compared to 0.03 µm of the ichip. Any microorganisms captured must originate from cells moving into the trap. Recently, it has been applied to investigate the oral microbiome [7].

All these methods rely on providing isolates with their natural occurring growth factors, although their identity remains largely unknown. Performing co-culturing experiments with isolates from the same environment, iron chelating siderophores have been identified as one growth factor. One species which serves as a helper cell, provides the other cell with siderophores, enabling its growth.

Recently, Tanaka *et al.* reported the importance of media preparation on bacterial growth. When phosphate buffer and agar where simultaneously autoclaved, an increased proportion of hydrogen peroxide was found in the media killing the cells. Adding catalase on the center of the plate or separate autoclaving of agar and phosphate buffer abolished the toxic effect [8].

Although, recent advances of cultivation-dependent techniques have shed more light into the darkness of uncultivable microorganisms, there is still a considerably high fraction of uncultivable microorganisms. The unexplored fraction of the collective microorganisms of a given environment, known as microbiome offers a precious source for new drugs, therapeutics or multiple enzymes, which might be beneficial for various industrial sectors and thus for society [3]. Cultivation-independent methods, such as next-generation sequencing, enable scientists to get a more complete picture of a given microbial community without the need of labor and time consuming cultivation. The most popular methods rely on parallel high-throughput sequencing and characterization of isolated deoxyribonucleic acid (DNA) fragments.

1.2 Cultivation-independent techniques for microbial ecology

Cultivation-independent approaches aim at determining members of microbial communities using molecular biology techniques. A method which allows visual investigation of the microbial members of a given environment is fluorescence *in situ* hybridization (FISH) [9]. After formaldehyde-fixation, samples are hybridized with fluorescence labeled probes targeting highly conserved genes, such as topoisomerase (*gyrB*), elongation factor thermo unstable (EF-Tu; *tufA, tufB*) and small ribosomal subunits (5S, 16S, 18S and 23S rRNA genes) [10]. The most commonly used target is the 16S/18S rRNA gene, due to its high conservation (Figure 1) and the availability of universal as well as taxa specific probes [11, 12]. Subsequently, several washing steps are employed to minimize false positive signals and enable sample analysis in a microscope capable of exciting the used dye. Although FISH offers studying of microorganisms in their natural habitat such as biofilms, it is only suitable for low diversity communities [13].

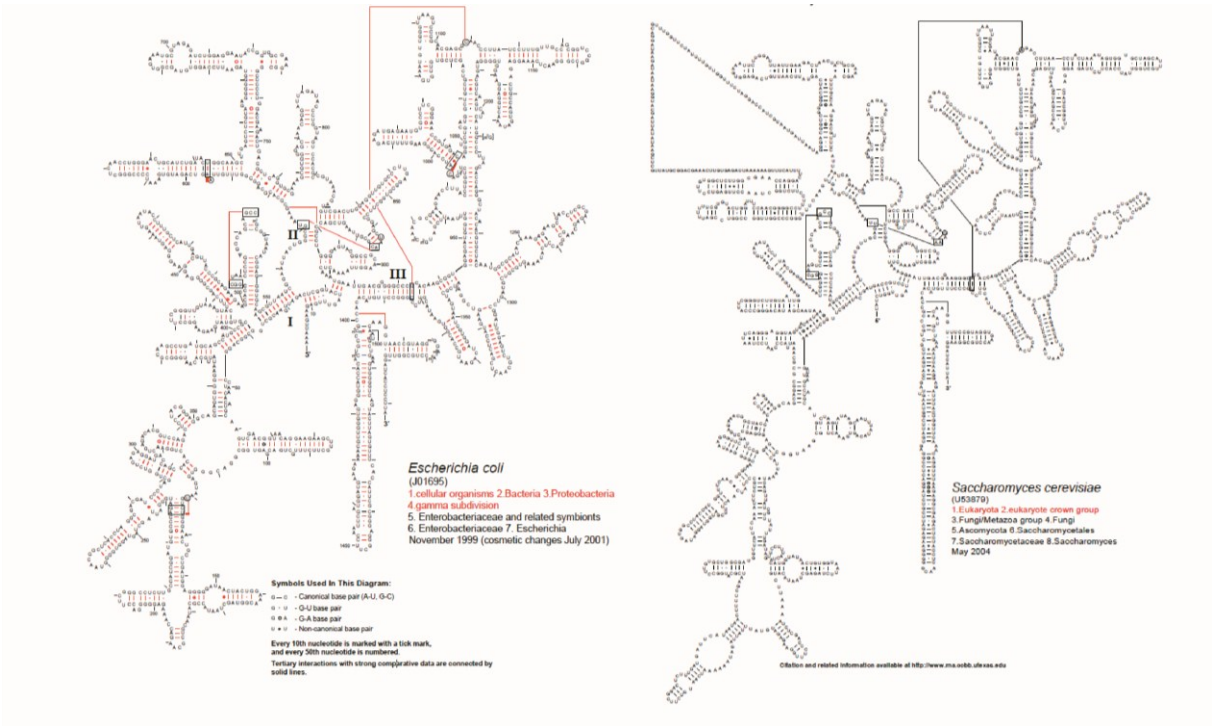

Figure 1: 16S (left) and 18S rRNA (right) secondary structure of *Escherichia coli* and *Saccharomyces cerevisiae*, respectively. Structure seems to be conserved throughout kingdoms. Figure adapted from the Comparative RNA website (http://www.rna.ccbb.utexas.edu)[10].

All succeeding methods are genome/gene centered and start with DNA extraction of sample containing a mixed microbial community. In case of Sanger sequencing, the 16S rRNA gene pool is amplified by polymerase chain reaction (PCR) using “universal” primers followed by cloning into a plasmid. A cloning strain is subsequently transformed with the constructs and positive clones are picked and subjected to Sanger sequencing [14, 15]. A sequencing reaction contains DNA polymerase units, all four deoxynucleotide triphosphates (dNTPs) combined with a little quantity of differently labelled dideoxynucleotide triphosphates (ddNTPs; one unique dye each), sequencing primers and the template. Whenever a ddNTP, which lacks the 3’-OH on the ribose backbone is incorporated, elongation is terminated. This reaction produces fragments with varying lengths which can be separated using capillary electrophoresis. When a dye is encountered, the molecule gets excited and a detector measures a signal corresponding to the wavelength of the given label [16]. Although this method is still the gold standard, it is highly labor- and cost inefficient. Moreover, unless using restriction enzyme-free cloning strategies, such as ET-, ligation independent cloning and AT-cloning, the use of restriction enzymes might introduce a bias by cleaving restriction sites within the targeted inserts.

Another popular technique to analyze even complex microbial populations is community fingerprinting by denaturing gradient gel electrophoresis (DGGE) [17]. After amplification of the gene of interest, mostly the 16S rRNA gene, a small amount of PCR product is loaded on a polyacrylamide gel containing a denaturing agent, such as urea and formamide [18]. Given the increasing denaturing conditions, DNA dehybridizes immediately into single strands. This method allows monitoring minor changes in similar fragments of the same length [19]. The drawbacks here are limited reproducibility and that only changes or differences in microbial populations can be determined. For identifying the different taxa, each varying band has to be excised from the gel and sequenced by Sanger sequencing, subsequently. This is not only laborious but also expensive.

PhyloChip, is a hybridisation based microarray assay, which can rapidly identify and quantify bacteria and archaea in a sample. It was developed at the Lawrence Berkley Laboratory and is owned by Affymetrix. rRNA or DNA can directly hybridize with immobilized probes on the chip surface. No prior amplification is necessary but can

be done if preferred. Its advantage is high inter-chip reproducibility and a high resolution taxonomic classification at subspecies level. Various studies using samples originating from diverse sources successfully utilized PhyloChip [20–22]. Like any other hybridization approach, this method suffers from false negative and positive signals and although the number of detectable species is constantly increasing, new or unknown species can not be discovered with PhyloChip. Although it is marketed as a cheap method for microbial community analysis, 750-1000 USD per chip is rather expensive (service quote 2016).

1.2.1 Next generation sequencing

A game changer was the introduction of next generation sequencing which allows parallel high-throughput sequencing for only a fraction of the costs of Sanger sequencing. Roche 454 pyrosequencing was the most popular sequencing method for microbial ecology given the long read length (up to 1kp with GS FLX+ according to manufacturer) [23–25]. Single DNA fragments are bound and amplified on a bead using emulsion PCR which get subsequently immobilized on a picotiter-plate. The picotiter-plate is serially flushed with only one nucleotide species and upon incorporation by a DNA polymerase, the produced pyrophosphate is enzymatically converted into a detectable light signal [26]. The biggest limitation of Roche 454 pyrosequencing, besides the struggle of an emulsion PCR and the high error rate (approx. 1 %) are homopolymers. Once the detector is saturated it cannot distinguish between 7 or more homopolymers [27]. Thus, reads derived from 454 pyrosequencing need to be denoised *in silico* before data interpretation which is a computationally intensive task [23]. Over the last decade Illumina kept on increasing read length (currently 2 x 300 bp on MiSeq V3 chemistry) while still retaining the high quality of the reads. Consequently, Roche 454 pyrosequencing was stepwise replaced by Illumina, specifically MiSeq, which is currently the most popular benchtop sequencer of microbiologists [28]. After addition of adapters using PCR or ligation, the library gets amplified on the flow cell through bridge-amplification [29], thus avoiding the need of the laborious emulsion PCR. Illumina follows the sequencing by synthesis strategy: The flow cell containing the clusters after bridge-amplification gets simultaneously flushed with all four dNTP, which are uniquely fluorescently labeled

each and contain a terminator. Therefore, only one nucleotide can be incorporated per strand. After imaging, the terminator is removed and the next cycle is initiated [29]. Although Illumina struggles with GC-rich regions resulting in substitutions, its error rate is only approx. 0.1 % and reads do not require denoising. Moreover one Illumina run produces 44-50 million paired 300 bp (13.2-15 Gb) reads for an average cost of 3000 USD compared to approx. 500,000 reads (35 Mb) for approx. 10,000 USD per run.

Another sequencer, Ion torrent is very similar to Roche 454 pyrosequencing but instead of detecting pyrophosphate, it senses the pH-shift upon nucleotide incorporation using a semiconductor [30]. Although it has a significantly higher throughput than Roche 454 pyrosequencing, it suffers from the same limitations (emulsion PCR, troubles with homopolymers, requires denoising) and an even higher error rate [31].

So-called single molecule sequencing aiming at lowering sequencing costs, increasing throughput and yielding longer reads are waiting in the wings for a while. Whereas all afore mentioned sequencing platforms are so-called wash and scan platforms, single molecule sequencers don’t pause for scanning the incorporated base and therefore allow live monitoring of the elongation process [32]. The first commercially available single molecule sequencing platform is Pacific Biosciences (PacBio) RS II. A DNA-polymerase is immobilized in a zero-mode waveguide (ZMW), which is small enough to allow the detection of the incorporation of a single nucleotide, in close proximity to a detector. All four differently labeled dNTPs are constantly available. The fluorescence labels are attached to the triphosphates of each nucleotide and diffuse away when the nucleotide is incorporated. Single molecule sequencers such as PacBio RS II do not require special adapters or amplification and therefore reduce library preparation, costs and possible sources of bias (e.g. primer and amplification bias) [32]. PacBio RS II can achieve read length up to 15 kb but suffer from high error rates of approximately 14 % (manufacturer). PacBio RS II generally produces fewer (500 – 1000 Mb per run) but significantly longer reads than next generation sequencing platforms [33], thus it is specifically interesting for applications such as de novo genome assembly, genome finishing and full length transcriptome analysis [34]. The successor of RSII is the Sequel. It is

significantly smaller, costs less than half the price (350,000 USD compared to 700,000 USD) and has a seven fold increase in throughput compared to its predecessor. PacBio RSII has been successfully used to profile microbial communities [35–37]. The reported error rates were 0.03 % after stringent filtering, error reduction and chimera removal which is comparable to what has been observed for Illumina MiSeq (0.02 %) and Roche 454 pyrosequencing (0.02%) generated data [27, 28]. The main limitation of using this platform remains accessibility. Illumina sequencers are almost ubiquitous in most research centers whereas only a hand full of European institutions offer PacBio RS II sequencing and none offers PacBio Sequel.

Other promising single molecule sequencers are advertised by Oxford Nanopore. Currently, the only commercially available sequencer is the MinION, which has the size of a USB thumb drive with impressive weight of 87 g. A DNA helicase is attached to a nanopore and an electric current is applied. The change in current is used to identify which base is currently passing through the nanopore. To date, only one study used the MinION for 16S rRNA sequencing and they received nearly full length reads which allowed the classification at species level [38]. Noteworthy, except for one study [35] microbial ecology studies using single molecule sequencers did not use real biological samples but investigated exclusively known mock communities. This one study concluded that not only sequence length determines the ability to classify sequences to the depth of a genus or species, but also sample type databases.

1.2.2 16S rRNA amplicon sequencing

The 16S rRNA gene is approximately 1,500 bp long and part of the small ribosomal subunit. It is the most commonly used marker gene in microbiome studies, given its highly conserved structure (Figure 1). Loss of function mutations, which for example alter secondary structure, would be lethal. In addition to conserved regions it has nine hypervariable regions [35, 36, 39] (Figure 2) which can be used to calculate distances between species and to build phylogenetic trees [40].

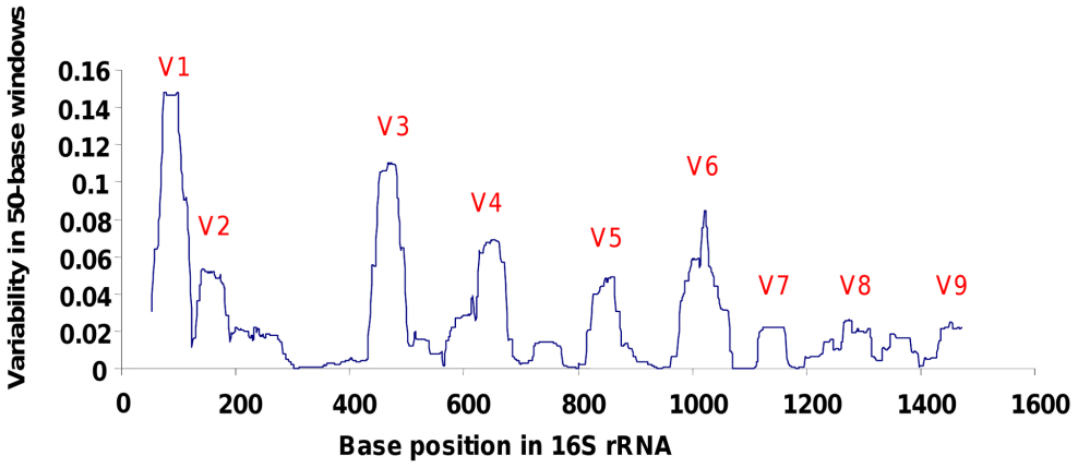

Figure 2: Hypervariable regions of the 16S rRNA gene. This graph is adapted from Bodilis et al. 2012 [39] **and shows variability in the alignment of 79 *Pseudomonas* type strains. Peaks denote the respective variable regions.**

Except for the aforementioned single molecule sequencers, a typical approach uses primers which anneal in constant regions, targeting only one to three variable regions, given the read length limitation of most sequencers. The use of so-called barcodes (Roche) and indices (Illumina), which are unique short sequences used for sample identification, allows simultaneous sequencing of multiple samples (multiplexing) in a single run. After sequencing, these unique sample identifiers are used to assign reads to samples (demultiplexing) [24].

Raw reads coming from sequencing platforms are full of errors. Every step, starting from sampling and DNA-extraction is prone to introduce bias and errors [41]. The most important step is sample acquisition. Sampling environment, equipment, operators and reagents have to be contaminant and DNA-free using UV-radiation or sodium hypochlorite (bleach). Currently, the only manufacturer, which guarantees DNA-free reagents, is Molzym. Sterile sampling is not sufficient, as free DNA would be targeted by the amplification primers as well. Amplification primers are also contributing to bias due limited target capacity. Based on nucleotide composition, each primer set will favour some molecules over others. Furthermore, there is the amplification bias due to error rates of DNA polymerases during the first amplification step, barcoding/indexing, emulsion PCR or cluster formation and the actual sequencing itself [41]. In classical genomics these errors are reduced by assembly and taking the consensus of all reads. This cannot be applied to amplicon sequencing, because there are no overlapping reads originating from the same

template which might complement each other, except for paired-end sequencing [35]. Paired-end sequencing sequences, as the name implies, one sequence from both directions. Thus, depending on the amplicon size and the sequencer, all bases in the overlapping regions are covered twice [28]. Errors lead to overestimation of diversity, thus impede a meaningful and reliable data interpretation. Stringent filtering and the use of various tools to reduce errors have to be applied. Chimeric sequences, which are generated during amplification also need to be detected and removed [42]. An important denoising step is to pre-cluster. Rare sequences are combined with more abundant sequences, based on a user defined threshold (allowing one mismatch per 100 bp as a rule of thumb). After all denoising steps, ambiguous reads (containing N characters) and reads which do not relate to amplicon size are excluded.

In order not to overestimate diversity, researcher agreed on clustering remaining high quality reads are clustered into operational taxonomic units (OTUs) using a similarity cut-off of 97 %, which is representing species level. A representative (QIIME software) or the consensus taxonomy of all reads of a given OTU (MOTHUR software) [43] is used to assign taxonomy to each OTU. Eventually, OTUs are presented in an OTU table which gives information about which OTUs are observed how often in a given sample. Remaining high quality reads can also be used to build phylogenetic trees and taxonomy is assigned to each branch. Given that amplicon sequences are using only a fraction of the full length 16S rRNA gene, taxonomy can only be reliably assigned to the depth of genus level [24].

Both, OTU tables and phylogenetic trees can be utilized for alpha- and beta diversity calculations. Alpha diversity informs about richness and evenness, the number of different species within a sample, and how equally species abundance is distributed, respectively. Evenness calculations are hindered, given that the 16S rRNA gene copy number varies from one to 15 copies per organism [39]. Therefore, intra-sample comparisons of different taxa are not valid, whereas inter-sample comparisons are.

Beta diversity informs how (dis)similar two or more samples are compared to each other [44]. To visualize beta diversity, multivariate statistics need to be applied for dimensional reduction. Assuming only three different OTUs are present in a given set of samples, they would simply be plotted one on x, y, and z axis, respectively. Close proximity represents similarity and a separation dissimilarity. Most biological samples

have more than three species and the human mind is not capable of imagining more than three dimensions. This is where multivariate statistics come into play. After calculating a distance matrix, which informs how dissimilar samples are compared to each other, principle coordinates are calculated using Eigenvectors, which aim at transferring variance in a decreasing manner to the coordinates [45].

One limitation which holds true for every genome centered approach is that a classical approach cannot distinguish between viable and dead proportion of a microbial community. Recently, several studies utilized propidium monoazide (PMA) to investigate the viable microbiome [46–48]. This intercalating dye is covalently bound to free DNA upon blue light exposure and masks it from the DNA polymerase. PMA is not able to penetrate intact membranes of living cells, thus only DNA originating from living cells gets amplified after DNA extraction [49].

1.2.3 Functional metagenomics

Unlike marker-gene centered approaches, functional metagenomics aim at identifying all genes present in a sample. This method provides taxonomic information of all four domains of life (bacteria, archaea, eukaryotes and viruses) present in a microbiome and their functional capacity [50]. It requires high throughput methods to increase the chance of targeting each DNA fragment in a sample also yielding sufficient coverage at the same time. To date, Illumina HiSeq has the highest throughput, thus it is the most commonly used sequencing platform [51, 52]. Nevertheless, functional metagenomics can technically be performed on all commercially available sequencing platforms. Similar to amplicon sequencing, reads can be sequenced from only one direction or from both (paired-end). After DNA extraction, the library gets fragmented mechanically or enzymatically followed by size selection, end repair, A-tailing, adapter ligation and indexing (see 2.2 for more details) [53]. Unlike marker-gene based approaches, in which specific taxa can be targeted, every DNA fragment in the sample will be subjected to sequencing. This might be problematic when investigating host associated microbial communities, such as plant or tissue biopsy microbiomes, because host DNA is prone to outnumber microbial communities simply by its genome size and amount. Selective lysis or fractionation can be

deployed to minimize host DNA during sample processing (e.g. particle filter for viruses or Molyzm for biopsies) [50].

It is significantly more expansive than marker based approaches and the bioinformatic analysis is more complex and requires access to large computing infrastructures or cloud computing. Unlike 16S rRNA amplicon sequencing analysis, there is no commonly used standardized pipeline. Reads retrieved from the sequencer can be assembled to gain bigger contigs or sometimes even to complete genomes. The bigger the contigs, the better the classification. Unfortunately only a small amount of reads are assemblable and the more complex the investigated community is, the worse the assembly [50]. One way to tackle this issue are hybrid assemblies where long reads from Roche 454 pyrosequencing, PacBio or Oxford Nanopore serve as a scaffold and short Illumina reads are used to reduce errors and improve coverage. In case of individual bacterial genomes this process has been automated to fully close a genome by a single click [54]. Recently, hybrid assembly have also been deployed to metagenomics [55, 56] but it is still not frequently used due to high sequencing cost. One limitation of an assembly based approach is that rare species might be suppressed as they cannot be assembled [50]. Assemblies can be used for gene prediction [57, 58] and the translated open reading frames can be compared to the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) non-redundant database [59].

Without the use of hybrid approaches, assembly results are often poor (no or few long contigs). Short unassembled reads are utilized for direct gene annotation [48] in those cases, but are more computationally intensive. Both assembled and unassembled reads most likely originating from the same or closely related genome are sorted into bins (binning). Binning can be performed exploiting GC-content or k-mer frequencies (compositional) of individual organisms or sequence similarity. Tools like Phylophytia [60], CONCOCT [61] and MetaPhlAn [62] use compositional binning, whereas MEGAN [63], MG-RAST [64] or IMG/M [65] use similarity based approaches. Some tools, such as PhymmBL [66] and MetaCluster [67] deploy both.

In general, short reads (e.g. 100 bp) do not provide a sufficient number of k-mers for a reliable composition based classification and thus similarity based tools should be used. However, they have significantly longer run times and may never even finish in

a reasonable amount of time. Unlike 16S rRNA amplicon sequencing, functional metagenomics yield a higher taxonomic resolution, thus providing species information.

Functional annotation of reads and assemblies rely on homology searches such as BLASTX [68] or DIAMOND [69] against annotated databases such as KEGG [70, 71], COG/KOG [72] and PFAM [73]. Although it is the most time and computational expensive task, only 20 % to 50 % of sequences can be functionally annotated [50]. After data annotation, metagenomes can be compared and alpha and beta diversity calculations can be applied to taxonomy and functions. Same as 16S rRNA amplicon sequencing, functional metagenomics do not provide information about whether the gene originated from viable or dead cells nor if it is actively expressed.

1.2.4 Metatranscriptomics

While functional metagenomics describe which organisms are present within a microbial community and their underlying functionality, metatranscriptomics is supposed to analyze their activity [74]. After RNA extraction, ribosomal RNAs, which may account to > 95 % of all RNAs, can be depleted using hybridization, such as RiboMinus (Thermo Fisher Scientifc, Waltham, MA, USA) kit to allow the detection of less abundant transcripts. This step is a potential source of bias, given that some messenger RNAs (mRNAs) might hybridize with the ribosomal probes as well [75]. Subsequently, mRNA can be either amplified with kits such as, MessageAmp II-Bacteria (Thermo Fisher Scientifc, Waltham, MA, USA) or directly reverse transcribed in complementary DNA (cDNA), which can subsequently be sequenced on the same platforms used for genomics [76]. The main challenge of metatranscriptomics is the assembly of short reads into complete transcripts, which can be used for quantifying gene expression [75]. New single molecule sequencers might help to overcome this bottleneck in the near future.

It’s the method of choice to investigate which genes are changing over time or to determine if gene expression is altered under different conditions. Its advantage also accounts for its limitation as it only provides expression and regulatory information (regulatory RNAs) at the time of sampling [76].

1.2.5 Metaproteomics

Transcripts can be silenced or can remain untranslated until certain conditions lead to a conformational change [77]. Metaproteomics aim at the identification of all translated proteins of a given microbial community. Unlike nucleic acids, proteins exist in different physical conformation and thus vary in their solubility [78]. Therefore, no extraction protocol is able to successfully extract all proteins of a given eco system. Fractionating samples based on their solubility might offer a more complete picture [79]. Isolated proteins might be separated using a two dimensional polyacrylamide gel electrophoresis followed by mass-spectrometry (MS) or liquid chromatography mass-spectrometry (LC-MS), which is more commonly used. MS is an analytical method which ionizes molecules followed by sorting of the ions based on the mass to charge ratio. The resulting spectra can be compared to existing databases for peptide identification [80].

1.2.6 Community metabolomics

Metabolomics is a highly sensitive method, which aims at generating a global profile of low molecular weight molecules, such as lipids, carbohydrates and amino acids of cells, organisms or microbial communities (metametabolomics) [81]. Serum and urinary metabolomics can be utilized to investigate the impact of the microbiome on the respective host [82]. For example, it offers the opportunity to monitor how microorganisms convert toxins and xenobiotics into valuable energy sources [83]. Metabolite extraction is the crucial step when it comes to metabolomics. When following an untargeted approach it is essential to extract as many metabolites as possible, whereas when following targeted metabolomics, the extraction method should maximize the metabolites of interest [84]. Analytical approaches to determine metabolite concentrations include nuclear magnetic resonance (NMR), LC-MS, gas chromatography mass-spectrometry (GC-MS) and direct infusion mass-spectrometry (DI-MS) [84]. Metabolomics provides an accurate snapshot of the actual physiological state [85]. Its high sensitivity is also the main limitation of this technique. Given that multiple factors, such as nutrition, general health status, and daytime are prominent

confounders, it is harder to pin observed changes or certain metabolites to a particular condition or disease.

All aforementioned meta-omics approaches are statistically challenging as the number of variables to be corrected for multiple-hypothesis testing is constantly increasing. This consequently weakens statistical power. A concise research question addressing a particular problem will reduce the need for multiple-hypothesis testing and empowers the significance of a study.

This thesis describes the results of two microbiome studies. A clinical trial, investigating the effects of a high dose of oral vitamin D_3 supplementation on the mucosa-associated microbiome, for which 16S rRNA amplicon sequencing on a Roche 454 Titanium pyrosequencing platform (best available method at that time) was used and a functional metagenomics study, aiming at evaluating the microbial composition of spacecraft assembly cleanroom with a special focus on pathogens and virulence factors.

1.3 Vitamin D and the human mucosa-associate microbiome

The human gut harbours approximately 10^{14} microorganisms, which outnumber the cells of the average human body by one order of magnitude. Collectively, these microorganisms are referred to as the gut microbiome. Recently, the human associated microbes have been subject of intensive research. These microscopic biochemical factories are interacting with its host on a myriad of levels, including energy acquisition, vitamin and cofactor production and detoxification of harmful compounds. Bacterial enzymes and metabolites interfere directly and indirectly with the human body beyond metabolism [86]. Bacterial products can affect distant organs such as liver, muscles or even he central nervous system, by converting amino acids, carbohydrates and other ingested substances into bioactive compounds, such as histamine or short chain fatty acids [87].

When it comes to energy acquisition, especially carbohydrate metabolism, the human genome is quite limited. By harbouring a healthy gut microbiome humans avoid the

need to code and express enzymes necessary for breaking down the multitude of various carbohydrates trespassing the human gut [88]. The human gut microbiome is also the main contributor to colonisation resistance against intestinal pathogens [89]. By occupying the intestinal surface, pathogens cannot successfully attach and induce diseases. The gut microbiome has already been shown to modulate the intestinal immune system [90] and impact host metabolism [91, 92].

To date, diet [93–95], host genetics [96], environmental circumstances [97], and drugs, such as antibiotics or laxatives [98, 99] have been proven to affect the composition of the gut microbiome. Impairment of gut homeostasis is referred to as dysbiosis. Dysbiosis of the human gut has been linked to many gastrointestinal diseases, such as inflammatory bowel diseases (IBD), intestinal infections, irritable bowel syndrome (IBS) and colon cancer [100–103] but also to non-intestinal diseases, such as diabetes mellitus, obesity, neurological disorders and autoimmune diseases [52].

Vitamin D is well known for its role on bone mineralization but has also major effects on the immune and the cardiovascular system as well as on host defence against pathogenic microorganisms [100, 104–106]. Vitamin D supplementation has been shown to be beneficial for maintaining health and also for improving treatment outcome of a number of diseases. A lack of UV-exposure and a diet lacking vitamin D-rich or fortified meals result in approximately half of the population to suffer from vitamin D deficiency (25-hydroxy-vitamin D level below 20 ng/mL) in many countries [107]. Higher vitamin D levels have been shown to decrease the risk of developing autoimmune diseases such as IBD, type 1 diabetes mellitus and rheumatic diseases, but also neoplasia and infections such as tuberculosis and hepatitis C [105]. Grant and Schuitemaker estimated that elevating vitamin D levels to and above 42 ng/mL would decrease disease rates in various cancers, cardiovascular diseases, diabetes mellitus, and infections by 10-50 % and the overall mortality rate by 18 % per year (Netherlands) [108].

The perception on how vitamin D affects human health has changed dramatically based on the findings that the vitamin D receptor (VDR) and the vitamin D activating enzyme CYP27B1 are expressed in many different cell types. VDR has its highest expression in $CD8^+$ T cells compared to other cells of the immune system [109].

Alongside to cells of the immune system, VDR and CYP27B1 can also be found in cells of the intestinal tract [107, 110], which indicates their prominent role in immune regulation. In addition to its immune modulating effects, vitamin D has shown to be involved in maintaining healthy gut homeostasis [111].

Many people are using vitamin D supplementation as part of their regular diet and some countries, such as the USA are fortifying milk products and juices. Although, vitamin D exerts multiple beneficial effects, its effects on the human intestinal microbiome still remain unknown. We hypothesized that some of the beneficial effects attributed to vitamin D might be mediated by the gastrointestinal microbiome. Therefore, we recruited sixteen healthy human volunteers for an interventional study to assess whether vitamin D_3, the most potent vitamin D alters the human gut microbiome. Most microbiome studies are limited to easily accessible stool samples, but studies have indicated that there can be a pronounced difference between mucosal and stool microbiomes [98, 112]. Therefore, we took a series of seven biopsies covering the major regions of the human gastrointestinal in addition to stools to investigate the mucosa-associated microbiome before and after eight weeks of oral vitamin D_3 supplementation.

1.4 New perspective on cleanrooms and human pathogens

Detection of signs of life on other planets is of particular interest for many of NASA's planetary missions. In order not to mistake earthborn microorganisms for unknown potential extraterrestrial life, planetary missions are subject to internationally accepted standards of planetary protection, established by the Committee of Space Research (COSPAR) [113]. Jet Propulsion Laboratory's Planetary Protection Group has undertaken huge efforts [114] to avoid inadvertent contamination of other planets with earthborn organisms and to minimize the bioburden on spacecraft. Spores are of particular interest, given their high resistance to multiple sterilization techniques, including radiation [115–117].

All spacecraft parts undergo extensive cleaning and sterilization steps, such as exposure to dry heat, vaporized hydrogen peroxide, radiation and alcohol on surfaces. Additional protocols to reduce the influx of particulate matter include daily

vacuuming and mopping of floors, high-efficiency particulate air (HEPA) filtration, regular replacement of tacky mats at all entry points, and strict gowning procedures. These precautions are routinely taken but with high frequency and stringency during the spacecraft assembly. All personnel that enter the cleanroom are required to put on cleanroom garments. This includes a full body suit, hair and beard nets, facemasks, additional head covering, gloves, shoe covers and cleanroom boots. These are necessary measures since humans are the major source of contamination in these environments [118, 119]. To monitor contamination levels, cleanrooms are regularly sampled for biological activity, particularly when spacecraft parts are being assembled and cleaned [116, 120].

Multiple sterilization methods are chosen, because there is no known method that can eradicate all microbes and is still compatible with spacecraft components. Only highly resistant microorganisms, such as spores, pathogens, and extremophiles, can overcome these strict decontamination procedures [121, 122]. Some microorganisms are even able to survive the harsh conditions of interstellar travel. Researchers placed spore-forming bacteria, isolated from cleanroom environment, outside the International Space Station for 18 months along with exposure to simulated Mars-like conditions, including atmospheric pressure and selective UV-radiation and some of them were still able to survive [115].

Pathogens might thrive in these environments perhaps due to their selective phenotypic characteristics, metabolic capabilities and reduced competition for scarce nutrients and niches. We were particularly interested in human pathogens, given that humans are the main source of contamination in cleanrooms [118, 119] and also because they are exposed to these constantly-evolving microbes. Most studies aiming at determining the microbiome of cleanrooms [21, 22, 46, 47], other indoor environments [123] or even the International Space Station [124] have used 16S rRNA amplicon sequencing. Although 16S rRNA amplicon sequencing has been prominently used in various studies [17, 125, 126], it lacks the capability to identify potential pathogens, as classification is only reliable down to genus level and does not allow detection of virulence factors. Previous functional metagenomic studies investigated pathogens in other indoor environments [127, 128] but this is the first

study, which focuses on the detection of pathogens as well as virulence factors in cleanrooms.

Our goal was to elucidate whether decontamination measures somehow lead to selection of hardy microorganisms, including pathogens, in the cleanroom and therefore posing a potential threat to human health. Three geographically distinct cleanrooms were sampled during the assembly of three NASA spacecraft: (i) Phoenix in Cape Canaveral, Florida, (ii) DAWN in Fort Worth, Texas, and (iii) Mars Science Laboratory (Curiosity) in Pasadena, California. Sample sets from Phoenix mission were collected from the cleanroom at three time points: before arrival of the spacecraft, during the assembly and testing of the Phoenix spacecraft, and after removal of the spacecraft from the facility. All samples were subjected to whole metagenome shotgun sequencing on an Illumina HiSeq 2500 platform. We screened for pathogens and virulence factors in order to determine pathogenicity.

2 Material & Methods

The thesis describes the results of two independent microbiome studies. The methods used in the particular studies are listed consecutively.

2.1 Effects of high doses of vitamin D_3 on mucosa-associated gut microbiome vary between regions of the human gastrointestinal tract

2.1.1 Study design

This study was designed as an interventional monocentric open pilot study. The inclusion criteria were: healthy volunteers, age between 18 and 40 years, body mass index (BMI) between 20 and 30 kg/m^2 and non-smoker. Hypercalcemia (>2.65 mmol/L), pregnancy, breast feeding, any disease which requires constant medication except for thyroid gland substitution, medication of any other clinical trial, a contraindication for Oleovit D_3 and complications during previous endoscopic procedures were defined as exclusion criteria. Employees of the Medical University of Graz were also not allowed to participate.

The study was approved by the local ethics committee (EK-Nr.: 24-024 ex 11/12) and was registered on the ClinicalTrials.gov (Identifier: NCT01538485). The study was conducted from January 2012 to August 2012 at the clinical Division of Endocrinology and Diabetology and the clinical Division of Gastroenterology and Hepatology of the Department of Internal Medicine at the Medical University of Graz. Principles of Good Clinical Practice, and ethical standards according to the Declaration of Helsinki and all later amendments were followed. All volunteers gave their written informed consent after being informed about the aim of the study, study procedure, and possible risks of participating in this study. Patient's characteristics are shown in Table 1.

Table 1: Study population characteristics

Parameter	Mean (± SD)
Study population size (n)	16 (9 men, 7 women)
Age	25 ± 4 years
Body Mass Index	23 ± 3 kg/m2
weight	69 ± 11 kg
height	172 ± 8

At the screening visit 1, anamneses, clinical examination, blood sampling, urine test, pregnancy test and education of gastroduodenoscopy and colonoscopy took place.

After baseline assessment, each volunteer received a weekly dose of 980 IU/kg bodyweight of vitamin D_3 (Oleovit D_3, Fresenius Kabi, Graz, Austria) for 4 weeks representing a daily dose of 140 IU/kg bodyweight, but maximal 68,600 IU per week in total (loading phase). A safety visit was set 4 weeks after colonoscopy where, among other parameters, serum calcium, vitamin D and CRP levels were determined. For the remaining 4 weeks, each volunteer received a weekly vitamin D_3 dose of 490 IU/kg bodyweight (daily dose of 70 IU/kg bodyweight), but maximal 34,300 IU per week in total. Four weeks after visit 4 at visit 5, a second gastroduodenoscopy was performed. Visit 6 was scheduled 1-10 days after visit 5 and a follow up visit 7 took place 2 ± 1 weeks after visit 6 (Table 2).

Table 2: Study flow chart

Visit number	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Visit (V) name	Screening, Baseline	Gastro	Colo.	Safety Visit	Gastro.	Colo.	Follow-up
Time point	0	+1-10 days after V1	+1-10 days after V2	four weeks after colo.	eight weeks after colo	+1-10 days after V5	two weeks (+/-1 week) after V6
Informed consent	x						
Study medication, Drug accountability			x	x	x	x	
Demography	x						
Medical history, illness	x						
Height, body weight	x						x
Physical examination	x						
Vital parameter	x	x	x	x	x	x	x
Blood test, serum profile	x				x		x
Urine test	x			x	x		x
Pregnancy test in case of women	x			x	x		x
Serum calcium, Vitamin D, CRP	x			x	x		x
Adverse effects	x	x	x	x	x	x	x
Other medication	x	x	x	x	x	x	
Gastroduodenoscopy		x			x		
Colonoscopy			x			x	
Mucosa sample for Immunology & Pathology		x	x		x	x	

Colo…colonoscopy
Gastro…gastroduodenoscopy

2.1.2 Sample collection and storage

Prior to oral vitamin D_3 supplementation, all volunteers underwent a gastroduodenoscopy followed by a colonoscopy on the next day. Stools were collected before bowel preparation for gastroduodenoscopy and immediately stored at −80 °C until DNA isolation. After eight weeks of oral vitamin D_3 supplementation, both, endoscopic procedures and stool sampling, were repeated. As a laxative, all volunteers received 2 litres of MOVIPREP® (Norgine, Marburg, Germany), a polyethylene glycol-based electrolyte solution for bowel preparation. During gastroduodenoscopy, biopsies were taken from the gastric corpus (GC), the gastric antrum (GA), and the descending part of the duodenum (DD) (Figure 1a). During colonoscopy, biopsies were taken from the terminal ileum (TI), the appendiceal orifice region (AO), the ascending colon (AC) at 10 cm distal to the ileocecal valve, and the sigmoid colon (SC) at 30 cm proximal to the anal canal (Figure 3). Two biopsies were taken from each region, immediately frozen and stored at −80 °C until DNA isolation, and two biopsies from each region were immediately placed on ice in complete RPMI media containing 10 % foetal calf serum (cRPMI-10), glutamine, and penicillin/streptomycin (Life Technologies, Darmstadt, Germany) for flow cytometry. Flow cytometry analysis was started within 1 h after biopsy acquisition. Endoscopies were kindly performed by Christoph Högenauer and Patrizia Kump. Flow cytometry was performed by Barbara Prietl and colleagues.

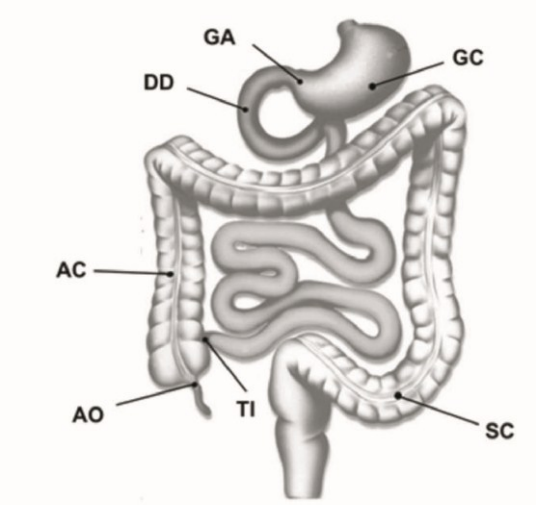

Figure 3: Sampling sites. Biopsied body regions (drawing generously provided by Tauschmann et al. 2013). GC gastric corpus, GA gastric antrum, DD duodenum, TI terminal ileum, AO appendiceal orifice, AC ascending colon, SC sigmoid colon. Figure modified from Bashir et al. 2015.

2.1.3 Isolation of lamina propria mononuclear cells

Gastrointestinal biopsies were immediately transferred to a dithiothreitol/EDTA solution (Sigma, Hamburg, Germany and Life Technologies, Darmstadt, Germany) and incubated for 15 minutes at 37°C. The samples were finely sliced and digested in a collagenase A solution (Roche, Mannheim, Germany) for 1 h at 37°C. The resulting cell suspension was passed through a 70 µm cell strainer (BD Biosciences, San Jose, CA, USA) and collected in a cRPMI-10 filled tube (Life Technologies, Darmstadt, Germany). After centrifugation, the pellet was resuspended in phosphate buffer saline (PBS; 137 mM NaCl, 2.7 mM KCl, 6 mM HPO_4^{2-}, 6 mM $H_2PO_4^-$, pH 7.4; Life Technologies, Darmstadt, Germany) and cell viability was confirmed by staining an aliquot with 0.4% trypan blue (Sigma, Hamburg, Germany) for microscopy.

2.1.4 Flow cytometry

Isolated mononuclear cells from lamina propria biopsies were stained for the expression of the surface markers CD3, CD4 and CD8. The staining was performed within one hour after the drawing of the biopsies. For the quantification of $CD3^+CD4^-CD8^+$ T cells ($CD8^+$ cytotoxic T cells) the following specific and monoclonal antibodies were used: anti-CD3 PerCP-Cy5.5 (eBiosciences, San Diego, CA, USA), anti-CD4 PE and anti-CD8 V500 (BD Biosciences, San Diego, CA, USA). All antibodies were used as recommended by the manufacturer. In detail, 5 µL anti-CD3 PerCP-Cy5.5, 20 µL anti-CD4 PE and 5 µL anti-CD8 V500 were used. After a 20 minute incubation period at room temperature cells were fixed by the addition of 2 mL lyse-fix solution (BD Biosciences, San Diego, CA, USA) followed by a centrifugation at 250 g for ten minutes. Finally, supernatant was discarded, cells were resuspended in 150 µL sterile PBS (Life Technologies, Darmstadt, Germany) and analyzed using a FACS CANTO II cytometer (BD, San Diego, CA, USA). Positive signals were defined by the use of unstained cell samples, isotype controls and the fluorochrome minus one (FMO) method. The exact gating strategy is given in the result part (Figure 8).

2.1.5 Vitamin D and safety parameter measurements

The vitamin D levels (calcidiol, 25-hydroxy-vitamin D) were measured using 25-Hydroxy Vitamin D EIA kit (IDS, Bolden, UK) with an intra- and inter-assay coefficient of variation of 5.6% and 6.4%, respectively according to manufacturer's instructions.

25 µL of each calibrator, control and sample were mixed with 1 mL of 25-D Biotin solution in borosilicate glasses, respectively followed by vortexing. Subsequently, 200 µL of each prepared sample, calibrator and controls were transferred to the antibody coated plate in duplicates. After sealing the plate, it was incubated at room temperature (18 °C to 25 °C) for two hours.

After incubation, supernatant was decanted sharply and each well was washed three times with 250 µL of Wash Buffer Solution. The plate was tapped firmly on an absorbent tissue to remove excess of Wash Buffer Solution. 200 µL of Enzyme Conjugate were added to each well. After sealing the plate, it was incubated incubation was carried out at room temperature for 30 minutes. Subsequently, the supernatant was removed and the plate was washed three times with 250 µL Wash Buffer Solution as described above.

Next, 200 µL of TMB substrate were added to each well and after sealing the plate it was again incubated at room temperature for 30 minutes. Subsequently, the reaction was stopped with 100 µL of Stop Solution (HCl) and the absorbance was measured at 450 nm in each well.

Safety parameters including serum calcium levels, urine calcium, urine calcium/creatinine ratio, complete blood cell count, serum phosphorus, serum albumin and parathyroid hormone (PTH) were assessed at baseline and every 4 weeks thereafter by standard laboratory methods.

2.1.6 DNA extraction

DNA extraction was performed semi-automatically using MagNA Pure LC DNA Isolation Kit III (bacteria, fungi) on a MagNA Pure LC 2.0 Instrument (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. In detail, an aliquot of approximately 180 mg of the individual stool samples or a single

mucosal biopsy were placed in MagnaLyser Green Beads tubes containing 500 µL of bacteria lysis buffer were subsequently homogenized using a MagnaLyser Instrument (Roche Diagnostics, Mannheim, Germany). After homogenization, 25 µL of lysozyme (enzyme cocktail II, 1 mg/mL) were added to each sample followed by an incubation at 37 °C for 30 minutes. Subsequently, 43.3 µL of proteinase K were added and samples were incubated over night at 65 °C on a shaking heat block.

250 µL of the supernatant were transferred to a MagNA Pure plate for automated DNA extraction on a bleached MagNA Pure LC 2.0 Instrument (Roche Diagnostics) according to the manufacturer's instructions. Elution volume was set to 100 µL and samples were eluted in elution buffer. Clean lab coats and bleached gloves were used and all preparations and manipulations were performed in a bleached and UV radiated laminar flow hood to minimize sample contamination. DNA concentration was determined photometrically using a SpectraMax Plus 384 (Molecular Devices, Sunnyvale, CA, USA) and then normalized to 10 ng/µL.

2.1.7 16S rRNA gene amplification and FLX sequencing

For library preparation 454 one-way read strategy (Lib-L kit, Primer A, Primer B; Roche 454 Life Science, Branford, CT, USA) amplifying the 16S rRNA hyper variable regions V1 to V2 with the target specific primers F27-AGAGTTTGATCCTGGCTCAG and R357-CTGCTGCCTYCCGTA (Table 3 and Table 4). FLX fusion primers with FLX adaptor, key and MID sequences (barcodes) were used for amplification (Table 5).

Table 3: amplification PCR components

Reagent	Volume (µL)
PCR-grade water	14.75
Fast Start High Fidelity Buffer (vial 2)	2.5
dNTPs (vial 6)	0.5
forward-amplification mrimer (10 pmol) MID1-30	1
High Fidelity Enzyme, 5 IU/µL (vial 1)	0.25
reverse-amplicon primer (10pmol)	1
DNA (10 ng/µL, biopsy or fecal)	5
Final volume	25

Table 4: amplification PCR program

Temperature	time (min)	
95°C	03:00	
95°C	00:45	30 x
55°C	00:45	
72°C	01:00	
72°C	07:00	
10°C	hold	

Table 5: Forward primers used for multiplexing

Name	Adapter	Key	MID/barcode	Target sequence
BSF8/20_1	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	ACGAGTGCGT	AGAGTTTGATCCTGGCTCAG
BSF8/20_2	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	ACGCTCGACA	AGAGTTTGATCCTGGCTCAG
BSF8/20_3	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	AGACGCACTC	AGAGTTTGATCCTGGCTCAG
BSF8/20_4	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	AGCACTGTAG	AGAGTTTGATCCTGGCTCAG
BSF8/20_5	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	ATCAGACACG	AGAGTTTGATCCTGGCTCAG
BSF8/20_6	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	ATATCGCGAG	AGAGTTTGATCCTGGCTCAG
BSF8/20_7	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	CGTGTCTCTA	AGAGTTTGATCCTGGCTCAG
BSF8/20_8	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	CTCGCGTGTC	AGAGTTTGATCCTGGCTCAG
BSF8/20_9	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TAGTATCAGC	AGAGTTTGATCCTGGCTCAG
BSF8/20_10	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TCTCTATGCG	AGAGTTTGATCCTGGCTCAG
BSF8/20_11	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TGATACGTCT	AGAGTTTGATCCTGGCTCAG
BSF8/20_12	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TACTGAGCTA	AGAGTTTGATCCTGGCTCAG
BSF8/20_13	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	CATAGTAGTG	AGAGTTTGATCCTGGCTCAG
BSF8/20_14	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	CGAGAGATAC	AGAGTTTGATCCTGGCTCAG
BSF8/20_15	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	ATACGACGTA	AGAGTTTGATCCTGGCTCAG
BSF8/20_16	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TCACGTACTA	AGAGTTTGATCCTGGCTCAG
BSF8/20_17	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	CGTCTAGTAC	AGAGTTTGATCCTGGCTCAG
BSF8/20_18	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TCTACGTAGC	AGAGTTTGATCCTGGCTCAG
BSF8/20_19	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TGTACTACTC	AGAGTTTGATCCTGGCTCAG
BSF8/20_20	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	ACGACTACAG	AGAGTTTGATCCTGGCTCAG
BSF8/20_21	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	CGTAGACTAG	AGAGTTTGATCCTGGCTCAG
BSF8/20_22	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TACGAGTATG	AGAGTTTGATCCTGGCTCAG
BSF8/20_23	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TACTCTCGTG	AGAGTTTGATCCTGGCTCAG
BSF8/20_24	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TAGAGACGAG	AGAGTTTGATCCTGGCTCAG
BSF8/20_25	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	TCGTCGCTCG	AGAGTTTGATCCTGGCTCAG
BSF8/20_26	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	ACATACGCGT	AGAGTTTGATCCTGGCTCAG
BSF8/20_27	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	ACGCGAGTAT	AGAGTTTGATCCTGGCTCAG
BSF8/20_28	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	ACTACTATGT	AGAGTTTGATCCTGGCTCAG
BSF8/20_29	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	ACTGTACAGT	AGAGTTTGATCCTGGCTCAG
BSF8/20_30	CCATCTCATCCCTGCGTGTCTCCGAC	TCAG	AGACTATACT	AGAGTTTGATCCTGGCTCAG

2.1.8 Amplicon purification and sequencing

15 to 30 µL of amplicons were purified on a denaturing HPLC on a WAVE apparatus (Transgenomic, Inc., Omaha, NE, USA) and eluted using a linear gradient (typically 12-17%) of acetonitrile in 0.1 M Triethylammoniumacetate (TEAA) over 10 min at 50 °C. Amplicon DNA was purified on NucleoFast® 96 PCR plates (Macherey-Nagel, Düren, Germany) by using a vacuum pump according to manufacturer's instructions and eluted in 30 µL elution buffer (QIAGEN, Hilden, Germany).

2.1.9 Amplicon measurement

Amplicon DNA concentrations were determined in duplicates using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies, CA, USA) according to manufacturer's instructions. A 1:200 dilution of the stock PicoGreen dsDNA reagent and a concentration curve (range 0.244 ng/µL to 250 ng/µL; 1:2 dilution steps) were prepared in TE-buffer. 1 µL of sample was mixed with 99 µL of PicoGreen dsDNA reagent/TE working solution in a 96 well Polystyrol microplate for FLUOstar measuring (Greiner F-bottom plate). DNA concentration was determined on a FLUOstar Omega spectrophotometer (BMG Labtech, Ortenberg, Germany) with following settings:

- Excitation filter: 485P
- Emission filter: 520P
- Gain: 1300
- Optics: 3 mm liquid light guides
- Number of cycles: 1
- Number of flashes: 10
- Shake time: 60 s before reading cycle

After quantification, 30 barcode labeled amplicons were pooled equimolar and analyzed on a 2100 Bio Analyzer (Agilent Technologies, Waldbronn, Germany) using a DNA 7500 kit.

2.1.10 Emulsion PCR

Emulsion PCRs of the pooled samples were performed using the GS Titanium MV emPCR Kit (LibL) (Roche 454 Life Science) and the GS Titanium emPCR Breaking Kit MV (Roche 454 Life Science). First, the live amplification mix was prepared, vortexed for 5 seconds, and stored on 4 °C (Table 6).

Table 6: Live Amplification Mix

Reagent	Volume (µL)
Mol.Bio. Grade Water	2630
emPCR Additive	3000
5x Amplification Mix	1560
Amplification Primer	230
emPCR Enzyme Mix	400
PPiase	10
Total:	7830

Up to four times 230 µL (equals 6.8 x 10^6 beads) of DNA capture beads per reaction were transferred into a 1.5 mL tube and washed three times with 500 µL of 1x capture beads wash buffer TW. It is important to fully resuspend beads at each washing step. Samples were denatured at 95 °C for two minutes, placed on ice, and subsequently added to the beads. Next 875 µL of live amplification mix were added and the mix was vortexed for five seconds. Emulsion oil (2 mL) was premixed on a TissueLyser (QIAGEN, Hilden, Germany) at 28 Hz for two minutes. After addition of 1 mL of 1x Mock Amplification mix, tubes were inverted three times and premixed at 28 Hz for five minutes. Next the library/bead mix was added to the emulsion, inverted three times and mixed at 12 Hz for five minutes. 100 µL of the resulting emulsified emPCR amplification mixes were dispensed into a 96 well thermocycler plate placed in a thermocycler running the following program:

Table 7: emulsion PCR program

Temperatur	time (min)	
94°C	04:00	
94°C	00:30	40 x
58°C	04:30	
68°C	00:30	
10°C	hold	

Emulsion PCRs were collected in two 50 mL conical tube. Broken emulsion (two phases and cloudy) were excluded. Wells were washed three times with isopropyl alcohol to retrieve any residual beads. After vortexing, the 50 mL conical tubes were filled to a final volume of 40 mL with isopropyl alcohol; beads were resuspended followed by centrifugation (Pico/Fresco, Heraeus, Hanau, Germany) at 930 g for five minutes. After removal of the supernatant tubes were filled up to a final volume of 40 mL with Enhancing Fluid XT, resuspended and pelleted at 930 g for five minutes. Subsequently, two additional washing steps with isopropyl alcohol to a final volume of 40 mL with centrifugation at 930 g for five minutes were performed. The pellet was washed with ethanol and centrifuged at 930 g for five minutes. After another washing step with 40 mL Enhancing Fluid XT, a remaining volume of approximately 2 mL was transferred to a 1.5 mL tube. Beads were pelleted for 10 seconds on a benchtop centrifuge. Beads were washed twice with 500 µL of Enhancing Fluid XT and eventually resuspended in 500 µL of Enhancing Fluid XT.

Each 1.5 mL tube was incubated twice for 2 minutes with 1 mL of melt solution (125 µL 10 N NaOH + 9,875 µL PCR grade water) after supernatant removal. Subsequently, each tube was washed twice with 1000 µL of Annealing Buffer XT. The ssDNA/beads mix was resuspended in 500 µL of Annealing Buffer XT and an aliquot of 5 µL was counted on a Coulter Counter (Beckmann Coulter, Brea, CA, USA). Remaining beads were pelleted again and resuspended in 30 µL of Annealing Buffer XT followed by addition of 24 µL of enrichment primers. Enrichment primers were annealed to the bead-bound ssDNA fragments at 65 °C for five minutes followed by incubation on ice for two minutes. Subsequently, 500 µL of Enhancing Fluid XT were added and beads were washed twice with 500 µL of Enhancing Fluid XT as above and eventually resuspended in 800 µL of Enhancing Fluid XT.

Next, magnetic enrichment beads were vortexed for one minute and 80 µL of vortexed enrichment beads were pelleted using a magnetic particle concentrator (MPC). After supernatant removal beads were resuspended and washed twice with 1000 µL of Enhancing Fluid XT using the MPC and eventually resuspended in 320 µL of Enhancing Fluid XT.

80 µL of washed enrichment beads were added per 1.5 mL tube containing amplified DNA beads followed by rotation at ambient temperature (15 to 25 °C) for five

minutes. Subsequently, tubes were placed on the MPC for three to five minutes to pellet the paramagnetic enrichment. After removal of the milky supernatant enrichment beads were washed six to ten times with 1000 µL of Enhancing Fluid XT. The washed pellet was resuspended in 700 µL of melt solution and vortexed for five seconds. Tubes were placed back onto the MPC and the supernatant containing enriched DNA beads was transferred to a fresh 1.5 mL tube. The remaining pellet was again washed with 700 µL of melt solution and vortexed for five seconds. After placement of tubes back onto the MPC, supernatants were transferred to the corresponding 1.5 mL tubes. Paramagnetic enrichment beads were discarded. Enriched DNA beads were pelleted as above on a benchtop centrifuge and supernatant was removed. The pellet was washed three times with 500 µL of Annealing Buffer XT per collection tube and vortexed for five seconds in between. Beads were eventually resuspended in 200 µL of Annealing Buffer XT per collection tube and vortex for 5 seconds after addition of 48 µL of sequencing primers. Sequencing primers were annealed at 65 °C for five minutes followed by a prompt incubation on ice for two minutes. Beads were pelleted as above and washed three times with 1000 µL of Annealing Buffer XT with five seconds of vortexing in between. Eventually, each pellet was resuspended in 500 µL of Annealing Buffer XT.

Each tube containing an equimolar pool of 30 samples was loaded on a quarter of a PicoTiterPlate and was sequenced using the GS FLX Titanium Sequencing Kit XLR70 (Roche 454 Life Science) according to the manufacturer's instructions.

2.1.11 Sequencing data analysis

Raw sequences were pre-processed and filtered using MOTHUR v.1.31.2 according to the Schloss 454 standard operating procedure (see http://www.mothur.org/wiki/454_SOP) [27]. Given that errors are introduced during sequencing, especially in homopolymeric regions in case of Roche 454 pyrosequencing, raw flowgrams need to be clustered to reduce noise. Denoising was performed in MOTHUR using its implementation of PyroNoise [23]. Reads shorter than 200bp or containing more than 8 homopolymers were excluded using the screen.seqs() command. Remaining reads were aligned to the SILVA database [130,

131] which incorporates small ribosomal subunit sequences from all domains of life (excluding viruses). Remaining pyrosequencing errors were reduced with pre.cluster(diffs=2) [41], reads identified as chimeras by UCHIME [42] and non-bacterial sequences, such as archaea, mitochondria or chloroplasts were also excluded. Taxonomy was assigned by MOTHUR's implementation of the ribosomal database project (RDP)-classifier [132] followed by binning into phylotypes based on taxonomy, using a minimum probability cut-off of 80 %. As we were using short reads we only classified the reads down to the genus level. The shared file, MOTHUR's OTU-table, was subsequently converted into a JSON biom table and passed on to QIIME's v.1.7 core_diversity_analyses.py command using non-phylogenetic parameters [133]. The GC was subsampled to 1,178 sequences/sample, GA to 2,125, DD to 2,441, TI to 1,647, AO to 2,215, AC to 2,014, SC to 2,188 and stools to 1,788 sequences/sample. Core_diversity_analyses.py is a QIIME workflow generating, among others following outputs:

Alpha rarefaction curves - give insight into richness, diversity and sampling success.
Beta diversity principle coordinate analysis (PCoA) plots - inform whether microbial communities of multiple samples are similar or not.

taxa summary plots – collated taxa on all phylogenetic levels.

2.1.12 Statistical analysis

To determine whether phylotypes (e.g. phyla, classes, order, etc.) differed significantly between the two time points we used a paired *t* test since the study compared samples from the same patients. Subsequently, *P*-values were corrected for multivariate testing using the Benjamini-Hochberg false discovery rate (FDR) procedure, as Bonferroni correction is too conservative for sequencing data. Corrected *P*-values below 0.05 were considered statistically significant (**P*<.05; ***P*<.01; ****P*<.001). To determine whether the intervention had different effects in the biopsied regions we calculated a Bray-Curtis dissimilarity matrix and used an analysis of similarity (ANOSIM) to test for significantly different clustering. ANOSIM was also used to test whether mucosal samples and stools differ significantly. Phylotype richness was compared before and after supplementation by using alpha

rarefaction plots based on an observed OTU matrix and a non-parametric *t* test. The difference between $CD8^+$ T cells were tested with a paired *t* test. All values are given as mean ± SD, if not otherwise stated.

2.1.13 Agar diffusion test

Three mL of Luria Broth (LB) media were inoculated with non-pathogenic Gammaproteobacteria *Escherichia coli* J5, *Salmonella typhimurium*, and *Pseudomonas fluorescence* and grown at 37 °C aerobically under shaking conditions over night. The following day, 100 µL of each strain was plated on two Luria Broth agar plates. Three sterile filter plates were placed on each half of each plate followed by placing 20 µL of Oleovit D_3 or walnut oil (Oleovit D_3 solvent provided by the manufacturer) on each three filter plates. Agar plates were incubated lid up at 37 °C under aerobic and anaerobic conditions for two and five days, respectively.

2.2 New perspective on cleanrooms and human pathogens

2.2.1 Sampling locations

Multiple samples were collected from the floor of the Kennedy Space Center's Payload Hazardous Servicing Facility (KSC-PHSF), where the Phoenix spacecraft was assembled at three time points: before arrival of the Phoenix spacecraft (10 samples; PHX-B), during the assembly and testing of the Phoenix spacecraft (8 samples; PHX-D), and after removal of the spacecraft from the KSC-PHSF facility (10 samples; PHX-A). 10 samples from the Lockheed Martin Aeronautics' Multiple Testing Facility (LMA-MTF) floor were collected during the DAWN spacecraft assembly. Samples were collected from the Ground Support Equipment (GSE) at Jet Propulsion Laboratory's spacecraft assembly facility (JPL-SAF) during the Mars Science Laboratory (2 samples; MSL) spacecraft assembly. These three cleanroom facilities were certified at ISO 8 (3,520,000 particles >0.5 µm per m^3) level and maintained according to the standard cleaning practices. Before entering the cleanroom, personnel were required to take appropriate actions to minimize

contamination risk. Specific entry procedures varied depending on the presence or absence of mission hardware.

2.2.2 Sample collection

Each sample was collected from cleanroom floor or GSE by a wet surface sampling technique using Biological Sampling Kits (BiSKits, QuickSilver Analytics, Abingdon, MD, USA) and polyester wipes, respectively. In case of BiSKits, PBS was added to the manufacturer's provided buffer to a final volume of 30 mL and this solution was used to hydrate the sponge. In case of MSL a polyester wipe was soaked with PBS. The hydrated sponge or wipe was used to wipe one square meter first horizontally, then vertically and eventually in a diagonal fashion. After sampling the sponge was forcefully pushed back in the sampling device to squeeze the sample into the collection tube. The polyester wipe was squeezed into a sampling tube.

Samples from each sampling event were concentrated using Amicon Ultra-15 centrifugal filter tube (Millipore, Jaffrey, NH, Ultracel-50 membrane). The molecular mass cutoff of 50 kDa leads to concentration of microorganisms, including spores as well as nucleic acids with a size greater than 100 bp.

After transferring samples to Amicon Ultra-15 centrifugal filter tubes samples were centrifuged at 4 °C within a Sorvall RC-5B centrifuge (Thermo Scientific, Waltham, MA) and spun at 4,000 rpm for 10 min. Samples were eluted in approximately 500 µL of PBS. A comparable amount of sterile PBS was concentrated in a separate filter tube, serving as a negative control for each concentration/extraction.

2.2.3 DNA-extraction

DNA was extracted from each concentrated sample using bead beating and an automated DNA extraction instrument (AutoLyser A-2 DNA, Axcyte Genomics, Menlo Park, CA, USA) and pooled in a equimolar fashion. DNA samples were archived at -80 °C until further use. Negative controls such as field control (sampling device control), reagent control (during DNA extraction) at each step were collected. None of the negative controls had a sufficient DNA concentration for library preparation and

were thus not included in further downstream analysis. Sample processing was performed in a sodium hypochlorite (bleach) treated laminar flow hood in an ultra-clean environment. Single-use lab-coats, bleached gloves, hairnets and booties were used.

2.2.4 Multiple displacement amplification

Due to low DNA concentrations, samples were subjected to multiple displacement amplification (MDA) [134]. Each sample was divided into 1 ml aliquots, which were amplified via Multiple Displacement Amplification (MDA) using Repli-g single-cell whole genome amplification kit (QIAGEN, Hilden, Germany). All plastic ware and water were UV treated in a Stratalinker 2400 UV Crosslinker (Stratagene, La Jolla, CA) with 254-nm UV for 30 to 90 min on ice [135]. This represents a UV dose range of 5.7 to 17.1 J/cm^2 calculated by measuring the distance from inside the tubes to the light bulb (4 cm). Buffer and enzyme come pre-cleaned and do not require UV-radiation. MDA reaction was prepared following manufacturer protocol for single cells, scaling reaction volume down to 15 µl final volume and addition of Syto13 dye (Invitrogen, Carlsbad, CA, USA) for real-time monitoring. A standard curve, ranging from 5 to 500 fg of *Escherichia coli* genomic DNA served as a positive control. 96 well plate (LabCyte, Sunnyvale, CA, USA) was sealed with BioRad optical seal and placed in a BioRad CFX96 Real-time machine using the program described in Table 9. MDA reaction was stopped when sample amplification reached saturation or after four hours.

Table 8: MDA reaction according to Repli-g single-cell whole genome amplification kit instructions

Reagent	Volume (µL)
Lysis (1xDLB)	1.00
Stop	1.00
H2Osc	1.70
buffer	8.70
Phi-29 polymerase	0.60
Syto13 (1:120)	0.01
DNA	1.00
total	15.00

Table 9: Thermocycler program used for MDA

Temperatur	time (min)	
30 °C	10:00	23 x
29 °C	00:30*	
65 °C	10:00	
10 °C	05:00	
4 °C	hold	

***…reading**

2.2.5 Library Fragmentation

Amplified fractions of each sample were combined and the pooled DNA product (100 µL) was sheared using a Covaris E210 instrument (Covaris, Woburn, MA) set to:

- 10 % duty cycle
- Peak incident power (PIP): 450
- intensity 5
- 200 cycles per burst
- Shear time 80 seconds

These setting are supposed to have a target peak at 500 bp.

2.2.6 Size selection with AMPure XP Beads

50 µL of sheared DNA were combined with 50 µL of TE-buffer and 55 µL of AMPure XP beads (Beckman Coulter, Brea, CA, USA). After pulse vortexing for ten seconds samples were incubated at room temperature for five minutes. The DNA/beads mix was placed on a MPC until the supernatant was clear followed by transferring the supernatant into a fresh tube. 20 µL of AMPure XP Beads were added to the collected supernatant and mixed through pulse vortexing. The mix was incubated at room temperature for five minutes and subsequently placed on a MP until the liquid cleared. The supernatant was removed and the pellets were washed three times with 200 µL of freshly prepare 75 % ethanol for 30 seconds. After supernatant removal samples were placed with open lid in a thermocycler at 37 °C to dry the pellets. Pellets were resuspended in 53 µL of QIAGEN Elution Buffer, vortexed and incubated at room temperature for one minute. Tubes were placed on a MPC until

the supernatant cleared and 50 µL were transferred into fresh tubes. Insert size (500 bp) was validated on a TapeStation (Agilent, Santa Clara, CA, USA).

2.2.7 End repair

The library was prepared using Kapa Library preparation Kit (Kapa Biosystems, Wilmington, MA, USA). Sheared DNA generated through physical fragmentation, such as sonication, needs to be end-repaired, because these approaches produce a mixture of 5' overhangs, 3' overhangs and blunt ends. Reagents listed in Table 10 were prepared in a 1.5 mL tube and combined with 50 µL of sample. After pulse vortexing samples were incubated at 30 °C for 30 minutes.

Table 10: End repair cocktail

Reagent	Volume (µL)
Water	26
10x End Repair Buffer	9
End Repair Enzyme	5
total	40

Subsequently, 126 µL of AMPure beads XP were added, followed by pulse vortexing. The mix was incubated at room temperature for five minutes and subsequently placed on a MP until the liquid cleared. The supernatant was removed and the pellets were washed three times with 200 µL of freshly prepared 75 % ethanol for 30 seconds. After supernatant removal samples were place with open lid in a thermocycler at 37 °C to dry the pellets. DNA was eluted in 17.5 µL of QIAGEN Elution Buffer, vortexed and incubated at room temperature for one minute. Tubes were placed on a MPC until the supernatant cleared and 15 µL were transferred into clean PCR strip tubes.

2.2.8 A-tailing

A single adenine nucleotide was attached to the repaired blunt ends. Adapters have the complementary thymine nucleotide overhang. This approach minimizes

autoligation, chimera formation and increases the probability of subsequent adapter ligation.

Reagents listed in Table 11 were prepared in a 1.5 mL tube and subsequently combined with 15 µL of sample. After pulse vortexing, samples were incubated for 30 minutes at 30 °C followed by incubation at 70 °C for five minutes

Table 11: A-tailing cocktail

Reagents	Volume (µL)
Water	9
10x A-tailing Buffer	3
A-Tailing Enzyme	3
total	15

2.2.9 Adapter Ligation

Adapters, which are complementary to the sequences present at the flow cells for cluster formation need to be added to the sheared, end-repaired and A-tailed library.

Reagents listed in Table 12 were mixed in the order listed and combined with 30 µL of sample, mixed and incubated at 20 °C for 15 minutes.

Table 12: Adapter ligation cocktail

Reagents	Volume (µL)
5x Ligation Buffer	9
Ligase	5
Adapter (18 µM)	1
total	15

After ligation 45 µL of AMPure XP beads and 5 µL of QIAGEN Elution Buffer were added to the sample and pulse vortexed. The mix was incubated at room temperature for five minutes and subsequently placed on a MP until the liquid cleared. The supernatant was removed and the pellets were washed three times with 200 µL of freshly prepare 75 % ethanol for 30 seconds. After supernatant removal samples were place with open lid in a thermocycler at 37 °C to dry the pellets. DNA was eluted in 52 µL of QIAGEN Elution Buffer, vortexed and incubated at room

temperature for one minute. Tubes were placed on a MPC until the supernatant cleared and 50 µL were transferred into fresh PCR strip tubes. 45 µL of AMPure XP beads were added to the 50 µL transferred sample and pulse vortexed. The mix was incubated at room temperature for five minutes and subsequently placed on a MP until the liquid cleared. The supernatant was removed and the pellets were washed three times with 200 µL of freshly prepare 75 % ethanol for 30 seconds. After supernatant removal, samples were place with open lid in a thermocycler at 37 °C to dry the pellets. DNA was eluted in 25 µL of QIAGEN Elution Buffer, vortexed and incubated at room temperature for one minute. Tubes were placed on a MPC until the supernatant cleared and 23 µL were transferred into clean PCR strip tubes.

2.2.10 Indexing PCR

Next, a PCR which targeted the adapters was performed to enrich the template, eliminate adapter dimers, and to add indices and sequences complimentary to the primers on the flow cell. Reagents listed in Table 13 were prepared for each sample and placed on a thermocycler using the program shown in Table 14.

Table 13: Indexing PCR according to Illumina paired-end sample preparation protocol.

Reagent	Volume (µL)
DNA	10
Ultra Pure Water	13
PCR Primer PE 2.0	1
PCR Primer PE 1.0	1
Phusion DNA Polymerase (Finnzymes Oy)	25
total	50

Table 14: Indexing PCR program

Temperature	Time
98°C	00:30
98°C	00:40
65°C	00:30
72°C	00:30
72°C	05:00

PCRs were cleaned using QIAquick PCR Purification Kit according to manufacturer's instructions. 250 µL of Buffer PBI were added to each reaction and mixed. The mix was transferred to a QIAquick column, which had been placed in a 2 mL collection tube to bind DNA. Tubes were placed in a benchtop centrifuge and were centrifuged at 20,000 x g for 60 seconds. After discarding the flow-through, QIAquick columns were washed with 750 µL of Buffer PE by centrifugation at 20,000 x g for 60 seconds. Next, QIAquick column was placed in a fresh 1.5 mL tube and DNA was eluted in 30 µL of QIAGEN Elution Buffer by centrifugation at 20,000 x g for 60 seconds.

The concentration of the resulting Illumina-indexed libraries was again determined using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Libraries were pooled and normalized to a final concentration of 400 nM each, and the primary bands corresponding to the sizes were gel-purified and dissolved in 30 µL QIAGEN elution buffer. One flow-cell was generated from a pooled library, which was subsequently subjected to sequencing on an Illumina HiSeq2500 instrument (2x250 bp) in accordance with manufacturer-provided protocols. The raw sequence data, excluding human reads are available at IMG/M (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi).

2.2.11 Sequence data analysis

We started with a total of 15,001,132 paired reads for PHX-B, 14,654,014 for PHX-D, and 22,355,430 for PHX-A assembly. For MSL and DAWN we had 57,892,216 and 2,899,364 reads, respectively (Table 21).

FastQC v0.10.1 [136] was used to determine the base quality throughout the 250 bp HiSeq-generated paired-end reads. PEAR v0.9 [137] default parameters was used with default parameters to merge paired reads. Unmerged forward and reverse reads were retained. Merged and unmerged reads were processed using prinseq-lite v0.20.3 [138] with the following parameters: "-min_len 100 -trim_qual_right 20 -trim_qual_left 20 -trim_left 8." Adapter sequences and overrepresented sequences were identified with FastQC and were removed using Cutadapt v1.1 [139]. PhiX174 and a JGI-standard collection of potential contaminant genomes (Table 15) were

removed by mapping trimmed high-quality reads using BBMap short read aligner v31.18 [140] to the respective genomes. Any reads matching any of these contaminant genomes were removed from the dataset.

Table 15: List of contaminant genomes used to exclude common contaminants. Adapted from Bashir et al. 2015 [125]

gi|93352797|gb|CP000352.1| Cupriavidus metallidurans CH34, complete genome 5s, 16s, 23s masked
gi|160361034|gb|CP000884.1| Delftia acidovorans SPH-1, complete genome 5s, 16s, 23s masked
gi|253972022|gb|CP000819.1| Escherichia coli B str. REL606, complete genome 5s, 16s, 23s masked
NC_010473 gi|170079663|ref|NC_010473.1| Escherichia coli str. K-12 substr. DH10B, complete genome 5s, 16s, 23s masked
gi|206564770|gb|CP000964.1| Klebsiella pneumoniae 342, complete genome 5s, 16s, 23s masked
gi|57158257|dbj|AP006725.1| Klebsiella pneumoniae subsp. pneumoniae NTUH-K2044 DNA, complete genome 5s, 16s, 23s masked
gi|218888746|ref|NC_011770.1| Pseudomonas aeruginosa LESB58, complete genome 5s, 16s, 23s masked
gi|152983466|ref|NC_009656.1| Pseudomonas aeruginosa PA7, complete genome 5s, 16s, 23s masked
gi|110645304|ref|NC_002516.2| Pseudomonas aeruginosa PAO1 chromosome, complete genome 5s, 16s, 23s masked
gi|116048575|ref|NC_008463.1| Pseudomonas aeruginosa UCBPP-PA14, complete genome 5s, 16s, 23s masked
gi|104779316|ref|NC_008027.1| Pseudomonas entomophila L48 chromosome, complete genome 5s, 16s, 23s masked
gi|255961261|ref|NC_007492.2| Pseudomonas fluorescens Pf0-1, complete genome 5s, 16s, 23s masked
gi|229587578|ref|NC_012660.1| Pseudomonas fluorescens SBW25 chromosome, complete genome 5s, 16s, 23s masked
gi|70728250|ref|NC_004129.6| Pseudomonas fluorescens Pf-5 chromosome, complete genome 5s, 16s, 23s masked
gi|146305042|ref|NC_009439.1| Pseudomonas mendocina ymp, complete genome 5s, 16s, 23s masked
gi|148545259|ref|NC_009512.1| Pseudomonas putida F1 chromosome, complete genome 5s, 16s, 23s masked
gi|166857509|gb|CP000926.1| Pseudomonas putida GB-1, complete genome 5s, 16s, 23s masked
gi|26986745|ref|NC_002947.3| Pseudomonas putida KT2440 chromosome, complete genome 5s, 16s, 23s masked
gi|170719187|ref|NC_010501.1| Pseudomonas putida W619 chromosome, complete genome 5s, 16s, 23s masked
gi|146280397|ref|NC_009434.1| Pseudomonas stutzeri A1501, complete genome 5s, 16s, 23s masked
gi|71553748|gb|CP000058.1| Pseudomonas syringae pv. phaseolicola 1448A, complete genome 5s, 16s, 23s masked
gi|66043271|ref|NC_007005.1| Pseudomonas syringae pv. syringae B728a, complete genome 5s, 16s, 23s masked
gi|28856110|gb|AE016853.1| Pseudomonas syringae pv. tomato str. DC3000, complete genome 5s, 16s, 23s masked
gi|71733195|ref|NC_005773.3| Pseudomonas syringae pv. phaseolicola 1448A chromosome, complete genome 5s, 16s, 23s masked
gi|240863652|gb|CP001644.1| Ralstonia pickettii 12D chromosome 1, complete sequence 5s, 16s, 23s masked
gi|240867064|gb|CP001645.1| Ralstonia pickettii 12D chromosome 2, complete sequence 5s, 16s, 23s masked
gi|240868673|gb|CP001647.1| Ralstonia pickettii 12D plasmid pRp12D02, complete sequence
gi|187724002|gb|CP001068.1| Ralstonia pickettii 12J chromosome 1, complete sequence 5s, 16s, 23s masked
gi|30407127|emb|AL646052.1| Ralstonia solanacearum GMI1000 chromosome complete sequence 5s, 16s, 23s masked
gi|81239530|gb|CP000034.1| Shigella dysenteriae Sd197, complete genome 5s, 16s, 23s masked
gi|73854091|gb|CP000038.1| Shigella sonnei Ss046, complete genome 5s, 16s, 23s masked
gi|37509034|dbj|BA000037.2| Vibrio vulnificus YJ016 DNA, chromosome I, complete sequence 5s, 16s, 23s masked
gi|37595821|ref|NC_005128.1| Vibrio vulnificus YJ016 plasmid pYJ016, complete sequence

To generate the human DNA sequence free dataset, all remaining high-quality reads were mapped with BBMap short read aligner against the human genome GRCh38 (including mitochondrial DNA). All positive matches were removed from the dataset.

Both, datasets including and excluding human DNA sequences were compared to NCBI non-redundant database using DIAMOND BLASTX v0.7.1 [69] with default parameters. Results were imported to MEGAN v5.10.5 [63] (minimal bit score of 80 %; “minscore 80”) for taxonomic binning, functional assignments to KEGG pathways, and generation of rarefaction curves (phylogenetic diversity on genus level). After removal of unassigned and unclassified reads, taxonomy and KEGG pathways [70, 71] were visualized using Krona Tools v2.4 [141]. Taxonomic and metabolic diversity calculations were done in QIIME 1.9.1 [142] with all samples subsampled to the smallest sample size observed.

Virulence factors were identified by comparing contaminant- and human-DNA-sequence-free reads to the Microbial Virulence Database MvirDB [143] using DIAMOND BLASTX [69] with an 80 % sequence similarity cutoff and a maximum of target sequences of one. Classified sequences were searched for clinically relevant pathogens (http://www.bode-science-center.com/center/relevant-pathogens-from-a-z.html accessed on Dec 1 2015; Appendix 2).

2.2.12 Extraction of draft genomes

A co-assembly was performed with trimmed, human and contaminant DNA sequence free, merged, and unmerged reads from all Phoenix samples using Ray Meta [144]. Reads were mapped to the contigs from the co-assembly using BamM [145] to produce sorted indexed BAM-files. GroopM [146] was used to bin contigs to genomes and completeness and possible contaminations were assessed with CheckM [147]. One draft with high completeness (69 %) was compared to NCBI non-redundant database using DIAMOND BLASTX and results were binned using MEGAN to classify the individual contigs.

3 Results

3.1 Effects of high doses of vitamin D_3 on mucosa-associated gut microbiome vary between regions of the human gastrointestinal tract

Except for one volunteer who could not attend the second colonoscopy for personal reasons, but completed all other study procedures, all participants completed the entire study. Vitamin D levels, measured as 25-hydroxy-vitamin D_3 in serum, were increased significantly from 22.3 ± 13.1 ng/mL to 55.2 ± 13.3 ng/mL after eight weeks of oral vitamin D_3 supplementation. In all volunteers, all measured safety parameters, including urine and serum phosphate and calcium as well as PTH remained in the normal range and none of them were significantly changed. Even after the loading phase (double dose) the highest measured value 94.5 ng/mL was way below the toxic level of 150 ng/mL (Table 16).

We generated 1,003,488 high-quality 16S rRNA amplicon sequences with a range of 211-380 bp (mean 264 bp) covering mucosal and stool samples throughout this study. Five microbiome samples were excluded from the sequence data analysis due to poor coverage (less than 900 reads per sample). Three volunteers were diagnosed with a gastric *Helicobacter pylori* infection during routine clinical biopsy processing and this diagnosis was also confirmed by sequencing data analysis. In this subgroup of three volunteers, we found up to 90 % of all classified sequences to be assigned to *Helicobacter* spp. Therefore, this subgroup was analyzed separately. In total, 242 samples passed the stringent filtering criteria, resulting in 13 matched samples (before and after eight weeks of oral vitamin D_3 supplementation) representing the upper gastrointestinal tract (GC, GA, DD), 12 matched samples for the AC, 11 for the TI, AO and SC and eight matched stool samples (Figure 4).

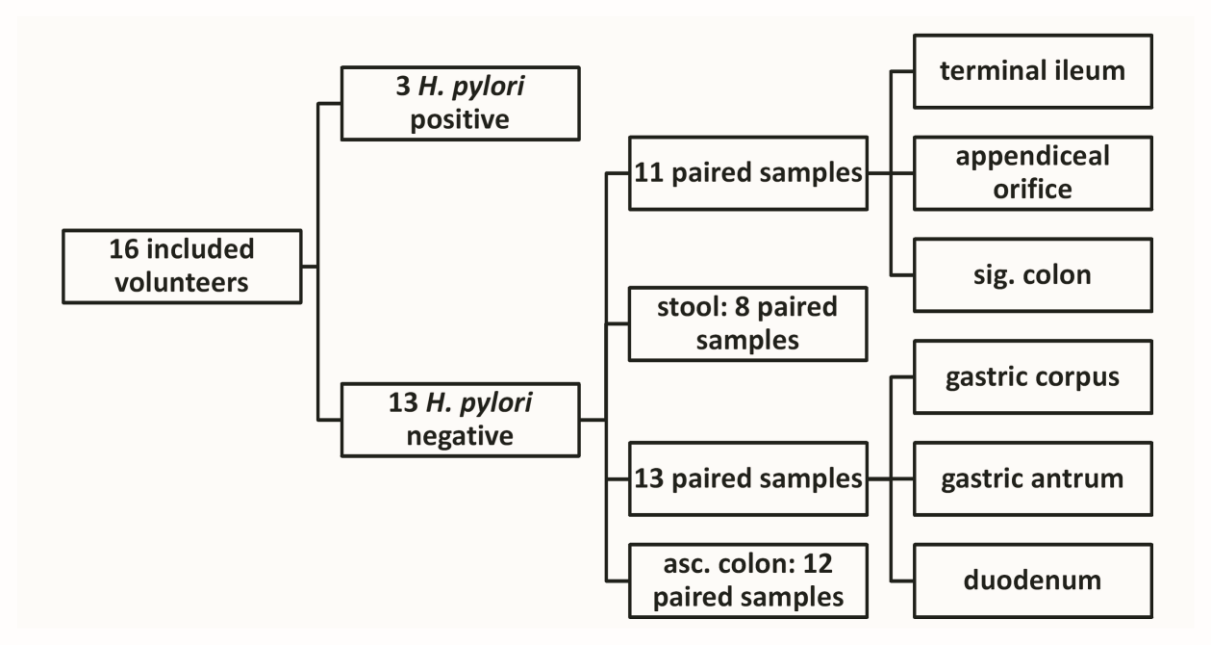

Figure 4: Paired samples passed quality filtering. A subgroup of three volunteers was analyzed separately due to a diagnosed gastric *Helicobacter pylori* infection. Figure modified from Bashir et al. 2015 [125].

Table 16: Measured safety parameters. No significant change in any of the assessed safety parameters could be detected. The only significant result was an elevated vitamin D level (calcidiol, 25-hydroxy vitamin D) as expected. Adapted from Bashir et al. 2015 [125]

	before				safety visit				after eight weeks of vitamin D_3 supplementation			
volunteer	**Calcidiol [ng/mL]**	**Calcium [mmol/L]**	**Ca/Creatinine ratio**	**PTH [pg/mL]**	**Calcidiol [ng/mL]**	**Calcium [mmol/L]**	**Ca/Creatinine ratio**	**PTH [pg/mL]**	**Calcidiol [ng/mL]**	**Calcium [mmol/L]**	**Ca/Creatinine ratio**	**PTH [pg/mL]**
301	16.1	2.4	0.14	33.6	76.1	2.23	0.06	34.9	51.4	2.25	0.2	45
302	17.8	2.5	0.4	62.4	64.1	2.41	0.23	43.5	48	2.47	0.29	75.8
303	8.6	2.36	0.3	38.1	67.6	2.35	0.11	47.5	54.1	2.37	0.06	32.9
304	18.1	2.3	0.21	41.8	73	2.45	0.1	22.4	66.6	2.51	0.04	23
305	16.4	2.57	0.54	21.3	44.3	2.43	0.35	29.9	33.1	2.5	0.24	34.7
306	7.2	2.58	0.06	35.9	64.3	2.42	0.06	34.6	55.5	2.43	0.11	34.5
307	46.1	2.41	0.73	24	74.3	2.39	0.19	30.8	65.5	2.38	0.26	20.4
308	42.4	2.48	0.14	45.8	94.5	2.33	0.12	11.8	59.1	2.37	0.2	27.8
309	24.5	2.43	0.18	38.1	53.7	2.6	0.15	25.3	43.9	2.46	0.49	21.8
310	19.6	2.28	0.31	43.2	58.9	2.36	0.34	59	42.2	2.35	0.05	40.3
313	11.2	2.38	0.31	48.6	77.3	2.51	0.07	42.6	71.9	2.35	0.08	19
315	8.9	2.41	0.13	36.2	41.8	2.47	0.22	30.1	38.7	2.54	0.21	28.9
316	44.7	2.39	0.06	27.9	85.6	-	-	28.9	73.8	2.39	0.26	24.1
317	20.7	2.43	0.25	64	64.9	2.34	0.29	39.5	50.1	2.41	0.12	23.2
318	37.1	2.55	0.09	48.7	70.1	2.45	0.1	68.4	79.3	2.55	0.22	53.3
319	16.7	2.5	0.3	22.8	66.3	2.48	0.2	23.8	50.2	2.52	0.31	30.8
mean	**22.3**	**2.4**	**0.26**	**39.5**	**67.3**	**2.4**	**0.17**	**35.8**	**55.2**	**2.4**	**0.20**	**33.5**
P-value									**<0.0001**	**0.7531**	**0.2447**	**0.1381**

3.1.1 A walk through the phylogenetic landscape of the human gastrointestinal tract

The mucosa associated human microbiome as well as the stool microbiome were dominated by two phyla, namely Bacteroidetes and Firmicutes (Table 17).

Table 17: Dominating phyla. Adapted from Bashir et al. 2015 [125]

	Bacteroidetes	Firmicutes
GC (%)	36.30 ± 7.56	36.93 ± 9.34
GA (%)	37.46 ± 13.44	35.94 ± 12.92
DD (%)	32.06 ± 9.75	38.43 ± 14.48
TI (%)	58.44 ± 8.47	36.52 ± 8.40
AO (%)	57.99 ± 7.23	38.29 ± 8.07
AC (%)	57.86 ± 7.69	38.20 ± 7.57
SC (%)	58.16 ± 7.92	36.50 ± 8.26
Stool (%)	62.87 ± 9.66	31.43 ± 9.90

Values are given as mean ± SD.

Interestingly, the upper gastrointestinal tract (GC, GA, and DD) had significantly more Proteobacteria, Actinobacteria, TM7, and Fusobacteria compared to lower gastrointestinal tract (TI, AO, AC, SC) and stool samples ($P < 0.0001$) (Table 18 and Table 19). This is probably reflecting the oral microbial community, which is considered as the main source for the human gut microbiome. Interestingly, we found more unclassified bacterial phyla in the lower gastrointestinal tract compared to the upper one. The microbial communities of the TI and the AO region were more similar to the large intestine (AC and SC) and stool rather than the upper gastrointestinal tract such as the DD, which belong to the small intestines. The microbiome of stool samples was more similar to the colon (AC and SC) than to the upper gastrointestinal tract (Figure 5). A list of the mean relative abundance of genera we found in the investigated regions of the human gastrointestinal tract is provided in Appendix 1. Due to easy access to stool samples, they are often the sole source investigated in microbiome studies. In our study we found that the stool microbiome significantly differs from the upper gastrointestinal tract (ANOSIM on genus level $P < 0.001$). The respective genera are listed in Table 20. The microbiome of the lower gastrointestinal tract (TI, AO, AC, and SC) was not significantly different from the stool microbiome (ANOSIM on genus level $P = 0.105$).

Table 18: Mean relative abundances of phyla before and after eight weeks of vitamin D_3 supplementation in the different regions of the gastrointestinal tract. Adapted from Bashir et al 2015 [125]

	gastric corpus		gastric antrum		duodenum		terminal ileum		appendiceal orifice		ascending colon		sigmoid colon		stools	
Taxon	**before**	**8 wks**	**before**	**8 wks**	**before**	**8 wks**	**before**	**8 wks**	**before**	**8 wks**	**before**	**8 wks**	**before**	**8 wks**	**before**	**8 wks**
Acidobacteria	0.0002	0	0.0006	0	0.0005	0	0	0	0	0	0	0	0.0001	0	0	0
Actinobacteria	0.0332	0.0348	0.029	0.0652	0.0478	0.0469	0.002	0.0018	0.0014	0.0024	0.0012	0.0022	0.0016	0.0017	0.0007	0.0033
Bacteroidetes	0.3039	**0.4305***	0.3109	0.4276	0.3156	**0.4611***	0.5675	0.5018	0.575	0.5089	0.5804	0.5087	0.5757	0.515	0.6057	0.5671
Deinococcus-Thermus	0	0	0.0003	0.0001	0.0002	0	0	0	0	0	0	0	0	0	0	0
Firmicutes	0.3235	0.2815	0.2974	0.2661	0.3671	0.3338	0.3877	0.4506	0.3874	0.4465	0.3824	0.4461	0.3727	0.4278	0.3235	0.3154
Fusobacteria	0.0385	0.0641	0.0336	0.04	0.0319	0.0442	0.0013	0.0003	0.0004	0.0003	0.0002	0.0003	0.0008	0.0002	0.0003	0.0006
Proteobacteria	0.2869	**0.1686***	0.3155	**0.1797***	0.2208	0.0908	0.0271	0.0286	0.0228	0.0265	0.0211	0.0253	0.0337	0.0349	0.0437	0.0886
SR1	0.0006	0.004	0.0007	0.0034	0.0008	0.0067	0	0	0	0	0	0	0	0	0	0
Spirochaetes	0.0016	0.0018	0.0004	0.0019	0.0015	0.0016	0	0	0	0	0	0	0	0	0	0
TM7	0.0053	0.0106	0.0046	0.0121	0.005	0.0069	0.0002	0	0	0.0002	0.0001	0.0002	0	0.0001	0	0
Tenericutes	0.0004	0.0001	0	0.0007	0.0005	0.0003	0	0	0	0	0	0	0	0	0	0
Verrucomicrobia	0	0	0	0	0	0	0	0.0001	0	0.0001	0.0002	0	0	0	0	0.0003
unclassified	0.0058	0.004	0.0068	0.003	0.0081	0.0076	0.0143	0.0168	0.0128	0.0152	0.0144	0.0173	0.0152	0.0203	0.0259	0.0246

***Significantly affected phyla (*P*<.05).**
All other phyla were not significantly affected.

Table 19: Comparing lower and upper GI's mean relative abundances of phyla. Adapted from Bashir et al. 2015 [125]

Taxon	upper GI*	lower GI+	FDR cor. *P*-value
Actinobacteria	0.040398764	0.001199235	5.49E-23
Bacteroidetes	0.351007792	0.583232274	1.72E-20
Fusobacteria	0.038526838	0.000673727	1.05E-18
Proteobacteria	0.182012102	0.026746975	1.15E-15
TM7	0.005354578	8.08E-05	4.65E-13
Firmicutes	0.372404336	0.368892662	0.86930121
unclassified	0.007422402	0.019079958	0.00203004

***Upper GI tract (GC; GA, DD)**
+Lower GI tract (TI, AO; AC; SC, stool)

Table 20: Comparison of upper gastrointestinal tract* and stools by mean abundance of genera. Adapted from Bashir et al. 2015 [125].

Test-Statistic	FDR corr. *P*	upper mean*	stool mean	taxonomy
-7.183462401	0.01490	154.26	554.77	Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides
5.651249219	0.01490	203.23	3.46	Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus
-4.830345776	0.01490	9.97	64.77	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiracea incertae sedis
-6.143188217	0.01490	10.90	68.08	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Faecalibacterium
3.958454884	0.01490	106.44	0.00	Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas
4.948102884	0.01490	57.51	0.08	Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Veillonella
4.113187014	0.01490	28.23	0.08	Fusobacteria;Fusobacteria;Fusobacteriales;Fusobacteriaceae;Fusobacterium
-4.302171678	0.01490	8.74	33.38	unclassified;unclassified;unclassified;unclassified;unclassified
-4.176815788	0.01490	2.72	21.77	Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Parabacteroides
3.99461268	0.01490	20.67	0.00	Firmicutes;Bacilli;Lactobacillales;Carnobacteriaceae;Granulicatella
-5.167064331	0.01490	1.33	36.62	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;unclassified
3.439037558	0.01490	12.03	0.00	Actinobacteria;Actinobacteria;Actinomycetales;Propionibacteriaceae;Propionibacterium
-5.48115343	0.01490	0.85	13.31	Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes
-4.037206985	0.01490	3.59	12.62	Firmicutes;Clostridia;Clostridiales;unclassified;unclassified
-4.141343241	0.01490	0.82	14.54	Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae;Clostridium XI
3.787443987	0.01490	9.41	0.00	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Oribacterium
-4.119199737	0.01490	0.59	7.38	Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella
4.355840504	0.01490	6.36	0.00	Firmicutes;Clostridia;Clostridiales;Eubacteriaceae;Eubacterium
-4.030836398	0.01490	1.31	17.46	Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Dialister
3.624964669	0.01490	4.38	0.00	Proteobacteria;Gammaproteobacteria;Pasteurellales;Pasteurellaceae;Haemophilus
-4.8907214	0.01490	0.59	5.85	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillibacter
-5.875120889	0.01490	0.00	0.69	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Butyricicoccus
-2.514474228	0.01490	0.00	2.15	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Flavonifractor
-3.562145536	0.02636	28.28	82.00	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;unclassified
-4.118569754	0.02636	0.31	2.00	Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter
-3.241018618	0.02636	0.05	1.08	Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae;unclassified

3.354133462	0.03316	30.82	2.69	Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;unclassified
3.445449499	0.03316	14.82	0.08	Actinobacteria;Actinobacteria;Actinomycetales;Actinomycetaceae;Actinomyces
3.375292365	0.03316	6.10	0.00	Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Solobacterium
3.273238834	0.03316	5.18	0.08	Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Megasphaera
-3.508232077	0.03316	0.03	0.85	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ruminococcus
3.838655167	0.03916	6.31	0.00	TM7;TM7 class;TM7 order;TM7 family;TM7 genus incertae sedis
3.525059741	0.03916	6.38	0.00	Proteobacteria;Gammaproteobacteria;Pasteurellales;Pasteurellaceae;unclassified
-3.407845789	0.03916	0.08	1.69	Proteobacteria;Betaproteobacteria;Burkholderiales;Sutterellaceae;Parasutterella
-3.066883499	0.03916	0.10	1.15	Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Holdemania
2.623603767	0.04393	19.08	0.00	Firmicutes;Bacilli;Bacillales;Bacillales Incertae Sedis XI;Gemella
3.280163193	0.04393	13.64	0.00	Fusobacteria;Fusobacteria;Fusobacteriales;Leptotrichiaceae;Leptotrichia
-3.194837818	0.04393	0.59	9.08	Firmicutes;Negativicutes;Selenomonadales;Acidaminococcaceae;Phascolarctobacterium
3.511184366	0.04393	3.18	0.00	Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae;Campylobacter

***Upper GI tract (GC; GA, DD)**

To investigate which phyla were affected by our intervention, we compared the relative abundance of all phyla before and after eight weeks of oral vitamin D_3 supplementation. We detected a significant reduction of Proteobacteria in the GC (P = 0.0348) and in the GA (P = 0.0258), while Bacteroidetes increased in the GC (P = 0.0013) and DD (P = 0.0013) (Figure 5 and Table 18). Fusobacteria, Actinobacteria, and Firmicutes were not significantly altered by eight weeks of oral vitamin D_3 supplementation in any of the investigated regions of the gastrointestinal tract.

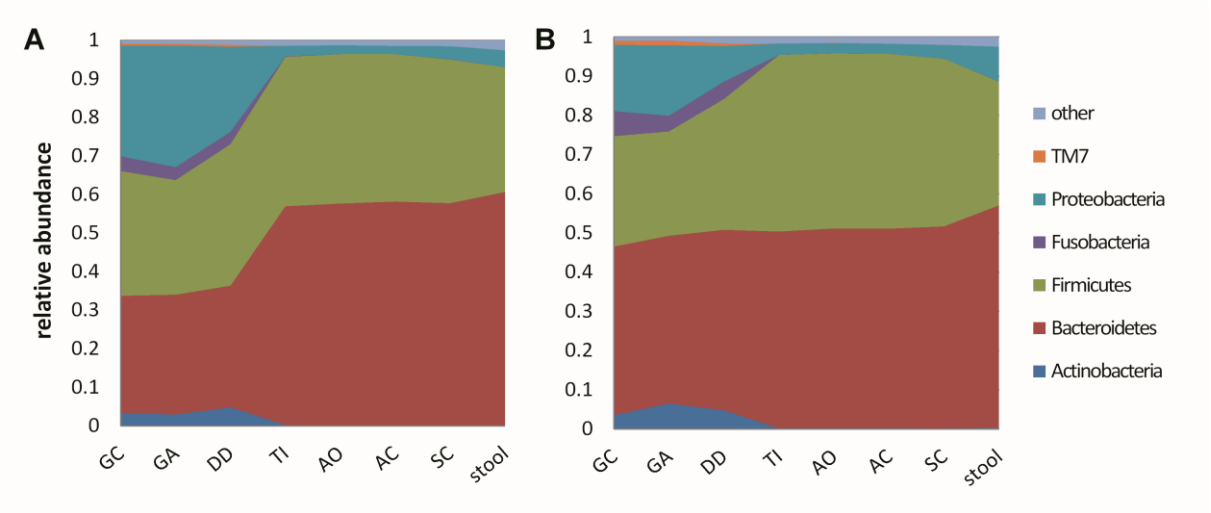

Figure 5: Mean relative abundance of the phylogenetic landscape before (A) and after (B) eight weeks of vitamin D_3 supplementation in 13 volunteers on the phylum level. A decline in Proteobacteria and an increase in Bacteroidetes in the upper gastrointestinal tract (GC, GA, and DD) was observed. Lower regions, such as TI, AO, AC, and SC and stools did not change on the phylum level. Gastric corpus (GC), gastric antrum (GA), duodenum (DD), terminal ileum (TI), appendiceal orifice (AO), ascending colon (AC), and sigmoid colon (SC). Adapted from Bashir et al. 2015 [125].

Beta-diversity analysis based on Bray-Curtis dissimilarity metric, visualized by PCoA plots showing the (dis)similarity of multiple samples when compared to each other. We detected that eight weeks of oral vitamin D_3 supplementation lead to a significant change in the microbial communities of the upper gastrointestinal tract. Samples taken before our intervention from the GC, GA and DD clearly differ from samples taken after eight weeks of oral vitamin D_3 supplementation. The microbial community in stool samples or in lower regions of the gastrointestinal tract (TI, AO, AC and SC) were not significantly affected, given that samples were not clustering separately (Figure 6).

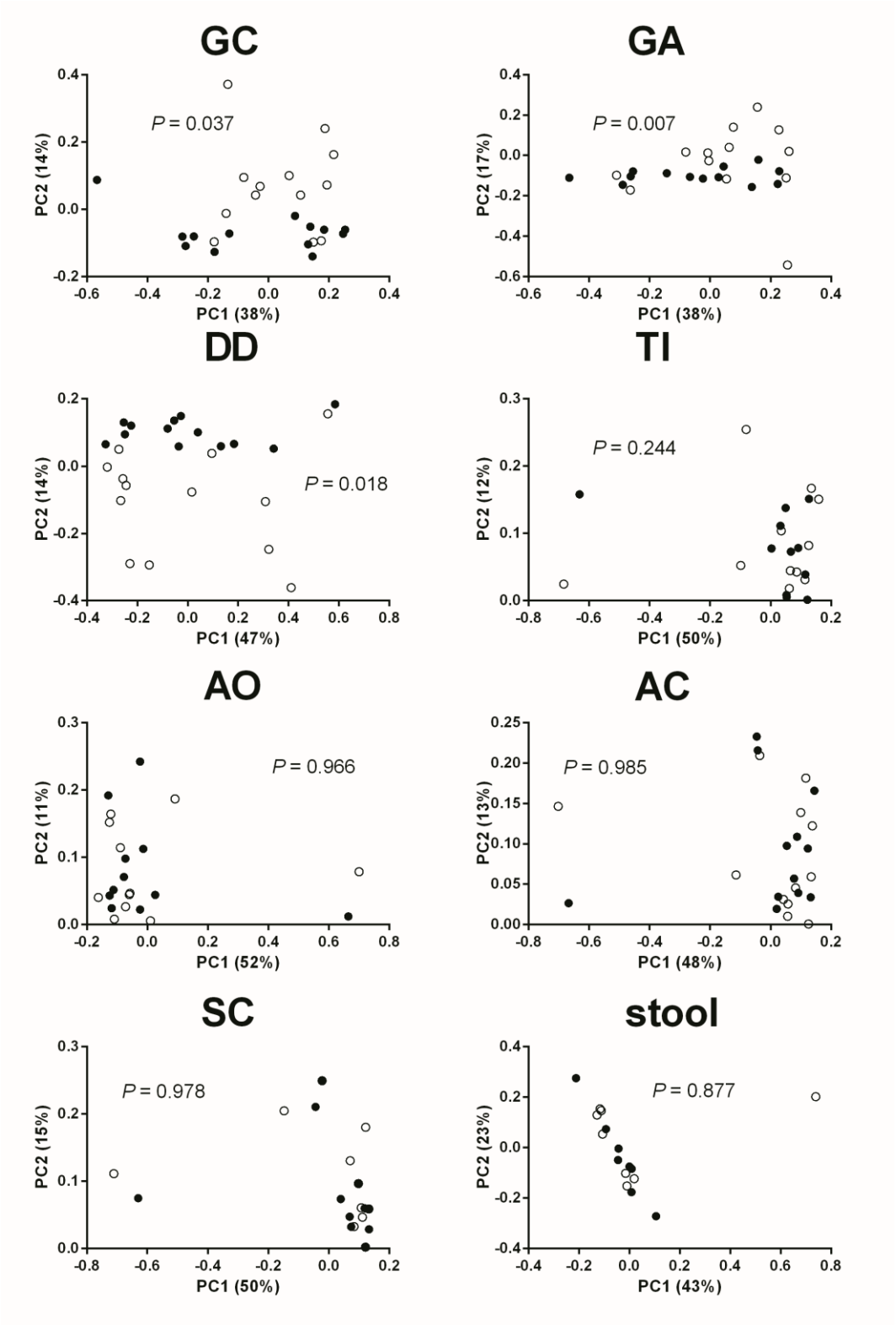

Figure 6: Vitamin D_3 supplementation leads to a significant shift in the bacterial community profile in the upper gastrointestinal tract, but not in the lower gastrointestinal tract or in stool. PCoA plots were based on Bray-Curtis dissimilarity metric on the genus level. We see a distinct separation of samples taken before and after eight weeks of vitamin D_3 supplementation. ANOSIM was used to test if these changes are statistically significant. White dots represent samples taken before and black dots represent samples taken after eight weeks of vitamin D_3 supplementation. Gastric corpus (GC), gastric antrum (GA), duodenum (DD), terminal ileum (TI), appendiceal orifice (AO), ascending colon (AC), and sigmoid colon (SC). Adapted from Bashir et al. 2015 [125].

3.1.2 Vitamin D_3 reduces the class of Gammaproteobacteria

To assess whether vitamin D_3 affects all Proteobacteria and Bacteroidetes or only a particular phylotype, we investigated the changes on all taxonomic levels, starting at phylum, down to genus. Only the class of Gammaproteobacteria (GC, *P* = 0.045; GA *P* = 0.034; DD *P* = 0.066), including *Pseudomonas* spp. (GC, *P* = 0.0280; GA *P* = 0.0061; DD *P* = 0.0316) and *Escherichia/Shigella* spp. (GC, *P* = 0.0022; GA *P* = 0.0078; DD *P* = 0.0190) showed a significant reduction while other Proteobacteria such as *Bradyrhizobium* spp. (Alphaproteobacteria, GA *P* = 0.0003; DD *P* = 0.0016) and *Sulfurospirillum* spp. (Epsilonproteobacteria, (GC, *P* = 0.033) showed a significant increase. Not only Gammaproteobacteria were declining after eight weeks of vitamin D_3 supplementations. *Ralstonia* spp. (Betaproteobacteria) was also significantly reduced (*P* = 0.0342) while the relative abundance of Alphaproteobacteria was significantly higher after our intervention (*P* = 0.0295) in the GA (Figure 7). Besides the effects on Proteobacteria, oral vitamin D_3 supplementation resulted in a significant reduction of *Lactococcus* spp. (*P* = 0.0434) and *Variovorax* spp. (*P* = 0.0164) in the GC. *Actinomyces* spp. increased significantly (*P* = 0.0175) in the GC, while *Alkalibacterium* spp. (*P* = 0.0394) increased in the GA, only. The relative abundance of *Leucobacter* spp. (*P* = 0.0289) decreased and *Janthinobacterium* spp. (*P* = 0.0318) increased in the DD.

Compared to the upper gastrointestinal tract, observed changes in the lower GI tract were minor. In the TI *Roseburia* spp. were significantly increased (*P* = 0.0421) and *Peptostreptococcus* spp. were decreased (*P* = 0.0423), both representing the phylum Firmicutes. We also found a decline of an unclassified *Clostridia* genus in the AO region (*P* = 0.0438). Colonic mucosal samples showed no significant changes in any phylotype, whereas in stools the class of Betaproteobacteria (*P* = .0234) decreased and *Actinomyces* spp. (*P* = 0.0073) increased in relative abundance (Figure 7).

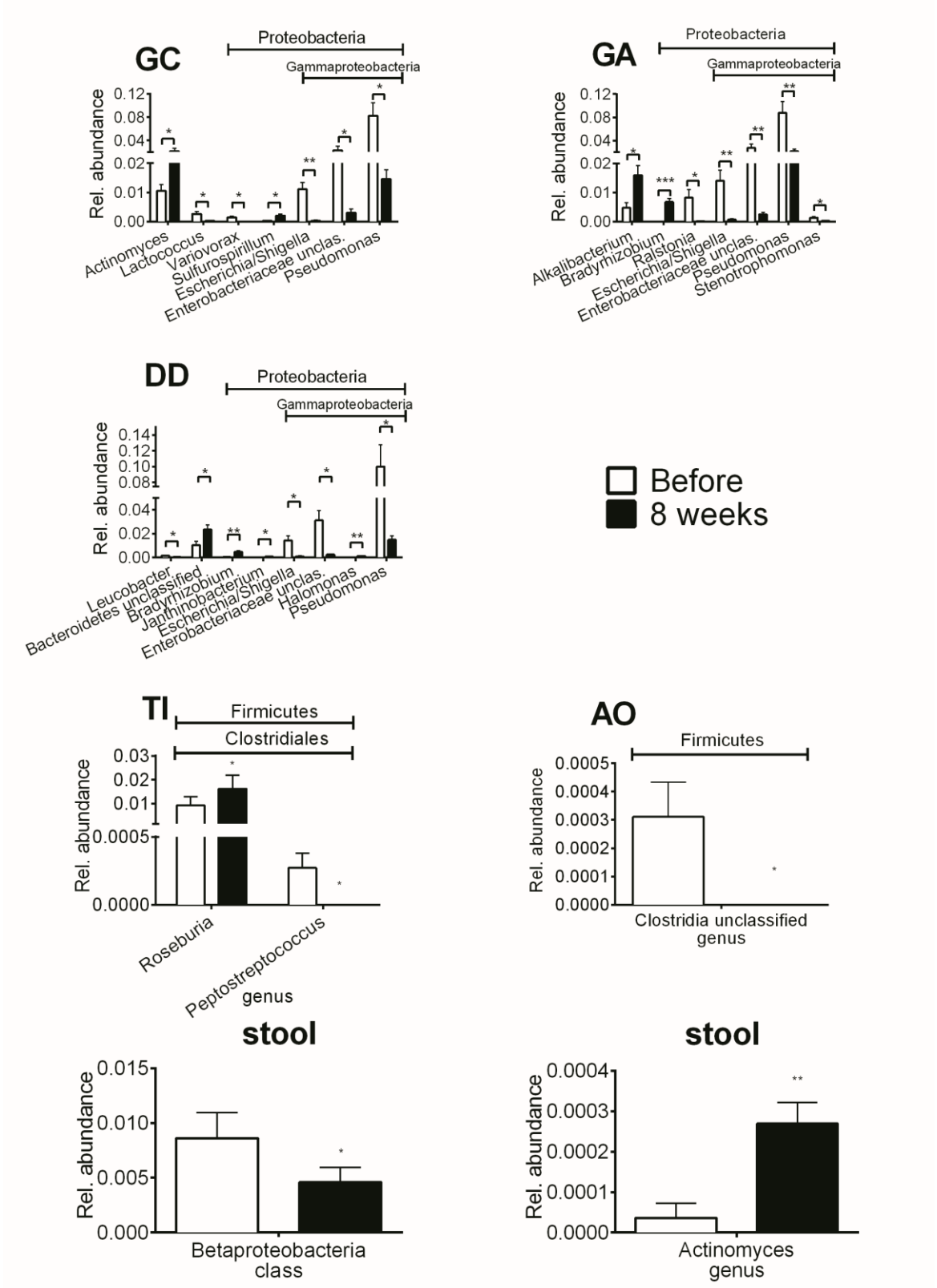

Figure 7: Significantly affected phylotypes (genera & class) after eight weeks of oral vitamin D_3 supplementation. No phylotype on any level was affected in colon (AC, SC). Samples taken before vitamin D_3 supplementation are shown in white and samples taken after supplementation are shown in black. Gastric corpus (GC), gastric antrum (GA), duodenum (DD), terminal ileum (TI), appendiceal orifice (AO), *P*<.05=*; *P*<.01=; *P*<.001=***; n=13, bars = mean ± SEM. A paired *t* test was used to calculate *P*-values, which were FDR-corrected for multiple testing. Adapted from Bashir et al. 2015** [125].

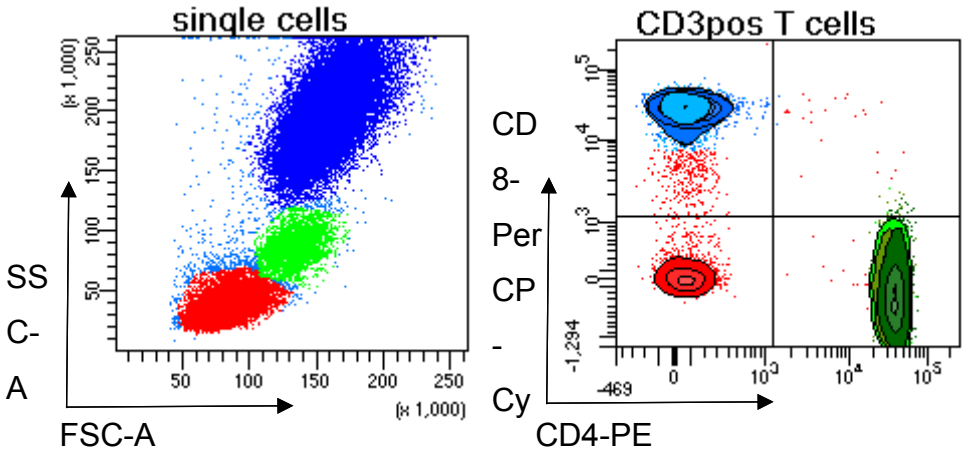

Figure 8: Gating strategy for flow cytometry. Green populations represent $CD4^+$ T cells, $CD8^+$ T cells are shown in blue and $CD4^-CD8^-$ lamina propria lymphocytes are shown in red. Figure was generated by Barbara Prietl.

Taken together, our data indicate that vitamin D_3 supplementation significantly alters Gammaproteobacteria in the upper GI tract, whereas other GI regions seem to be less affected. $CD8^+$ T cells may contribute to the decline of Gammaproteobacteria. Flow cytometry analysis of $CD8^+$ T cell counts showed a significant increase in the AO (P = 0.0213) and a trend towards increased $CD8^+$ T cells in the other regions after eight weeks of vitamin D_3 supplementation (Figure 9).

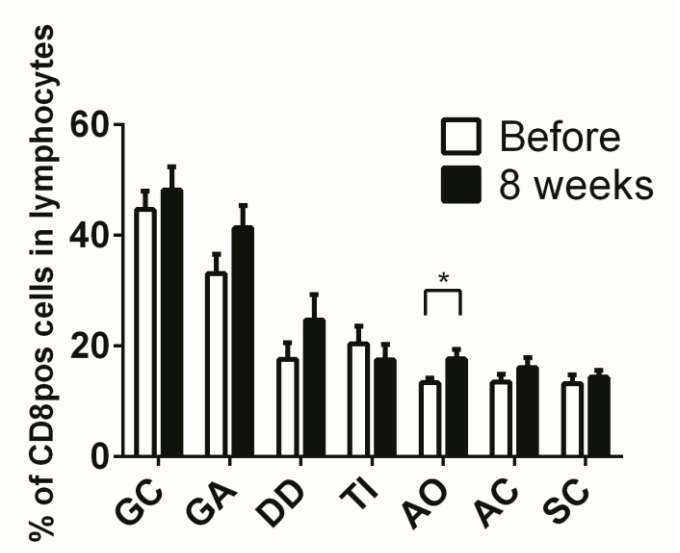

Figure 9: Changes of mucosal $CD8^+$ T cell counts after vitamin D_3 supplementation. Percentage of $CD8^+$ T cells of all lymphocytes isolated from mucosal biopsies (GC, P = 0.5168; GA, P = 0.1317; DD, P = 0.2058; TI, P = 0.5012; AO, P = 0.0315; AC, P = 0.2476; SC, P = 0.5773). Samples taken before vitamin D_3 supplementation are shown in white and samples taken after supplementation are shown in black. Gastric corpus (GC), gastric antrum (GA), duodenum (DD), terminal ileum (TI), appendiceal orifice (AO), ascending colon (AC), and sigmoid colon (SC). P<.05=*; n=13, bars = mean ± SEM. A paired t test was used to calculate P-values. Adapted from Bashir et al. 2015 [125].

3.1.3 Vitamin D_3 does not inhibit the growth of Gammaproteobacteria directly

To evaluate if vitamin D_3 itself inhibits the growth of Gammaproteobacteria, we performed an agar diffusion test with three non-pathogenic laboratory strains, namely *E. coli* J5, *S. typhimurium*, and *P. fluorescence* under aerobic and anaerobic conditions. None of the tested strains showed any growth inhibition due to vitamin D_3 or its solvent, neither under aerobic nor anaerobic condition. All strains were even growing on the filter plates soaked with vitamin D_3 or solvent (Figure 10).

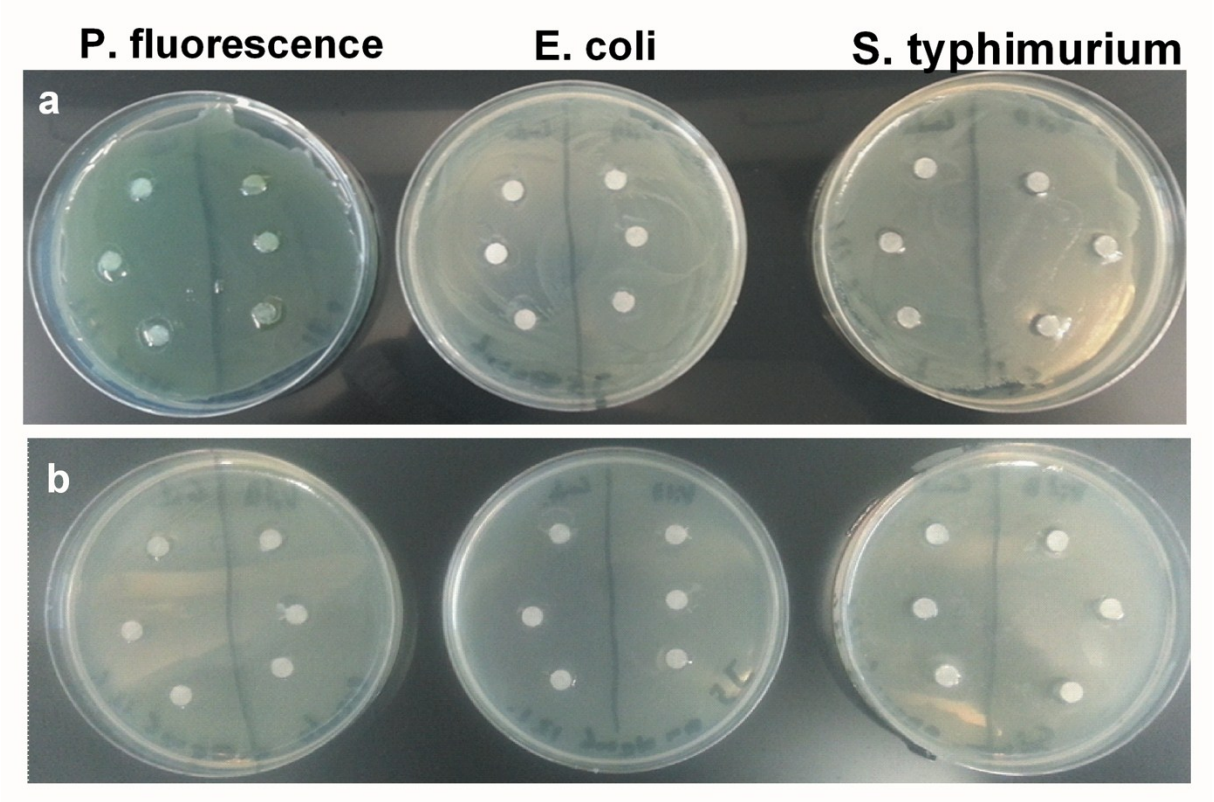

Figure 10: Agar diffusion test to determine growth inhibition of vitamin D_3 on selected laboratory Gammaproteobacteria. Plates were incubated under (a) aerobic and (b) anaerobic conditions. On each plate the three left and right filter plates were soaked with walnut oil (Oleovit D_3 solvent) and Oleovit D_3, respectively. None of the tested strains showed a zone of inhibition.

3.1.4 Vitamin D_3 increases phylotype richness in the gastric antrum and duodenum

Rarefaction analysis plots showed that samples were well covered, but also indicate that not every species has been detected, given that none of the curves plateaued. Besides sufficient coverage we also observed a significantly increased phylotype richness in the GA (P = 0.004) and a trend towards an increased phylotype richness

in the DD (*P* = 0.06). No change in phylotype richness was observed in the GC (*P* = 0.933), TI (*P* = 0.327), AO (*P* = 0.778), AC (*P* = 0.758), SC (*P* = 0.951) and stool *(P* = 0.313) (Figure 11).

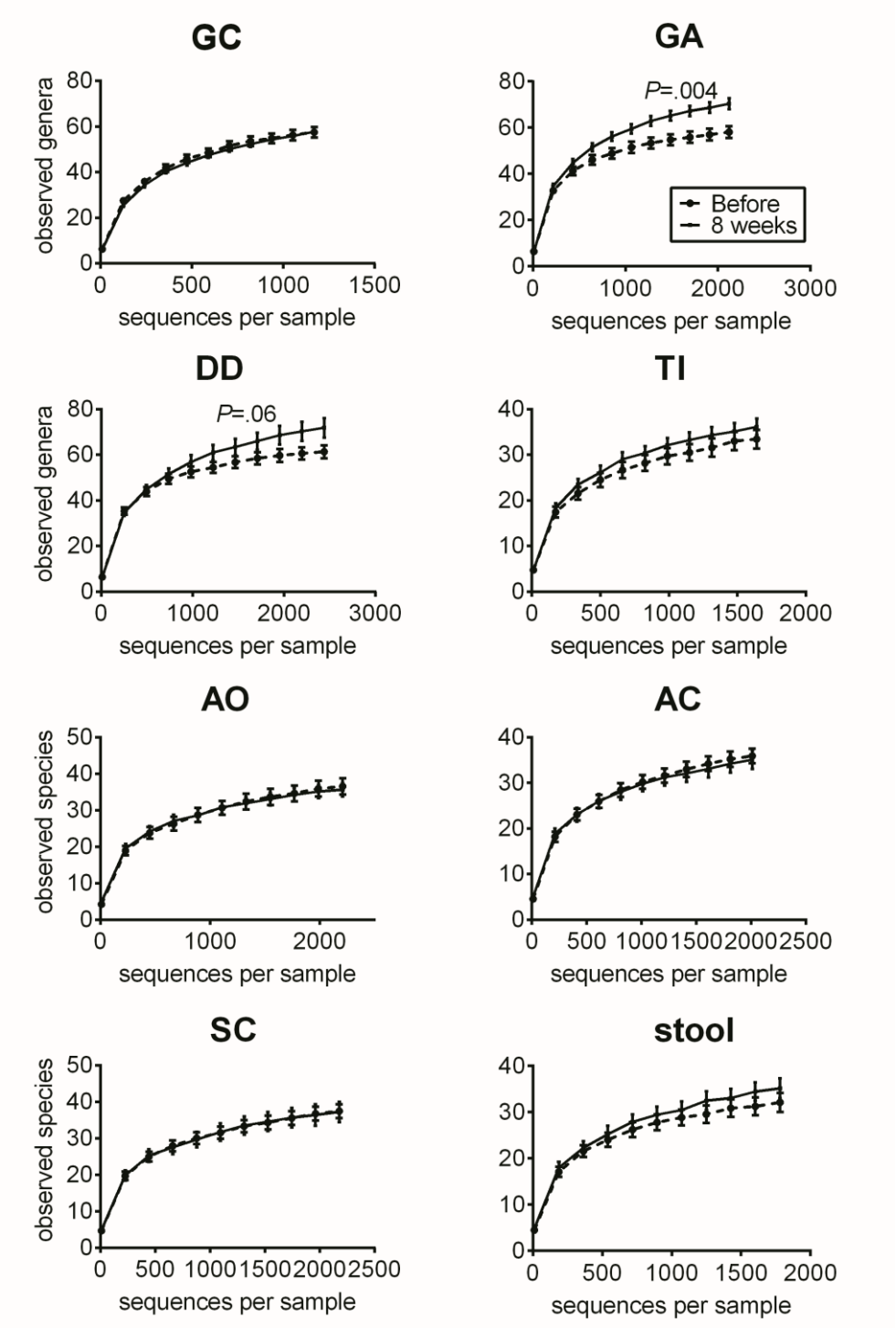

Figure 11: Vitamin D_3 supplementation increases phylotype richness in the gastric antrum and duodenum; Alpha rarefaction curves of all investigated regions of the gastrointestinal tract and stool. The number of observed species is plotted against the number of sequences per sample. We see an increase in richness in the gastric antrum (*P* = 0.004) and a trend in the duodenum (*P* = 0.06) but not in the GC (*P* = 0.933), TI (*P* = 0.327), AO (*P* = 0.778), AC (*P* = 0.758), SC (*P* = 0.951) and stool (*P* = 0.313). Rarefaction curves also indicate a good coverage, but show that not every phylotype has been detected. Gastric corpus (GC), gastric antrum (GA), duodenum (DD), terminal ileum (TI), appendiceal orifice (AO), ascending colon (AC), and sigmoid colon (SC). Adapted from Bashir et al. 2015 [125].

3.1.5 Vitamin D_3 reduces relative abundance of *Helicobacter pylori*

Sequencing data analysis and routine biopsy-processing indicated *H. pylori* infection in three volunteers. 97.72 % of all classified sequences of volunteer Y were identified as *Helicobacter* spp. and the lowest relative *Helicobacter* spp. abundance observed in the GA before eight weeks of oral vitamin D_3 supplementation was 86.63 %. All three volunteers showed an overall decline of *Helicobacter* spp. after eight weeks of oral vitamin D_3 supplementation. We observed a trend towards a decline of *Helicobacter* spp. in the GA of all 3 volunteers, but changes in the GC and DD were inconsistent. The biggest reduction was observed in the GA of volunteer Z, where *Helicobacter* spp. abundance declined from 86.63 % to 4.11 % (Figure 12).

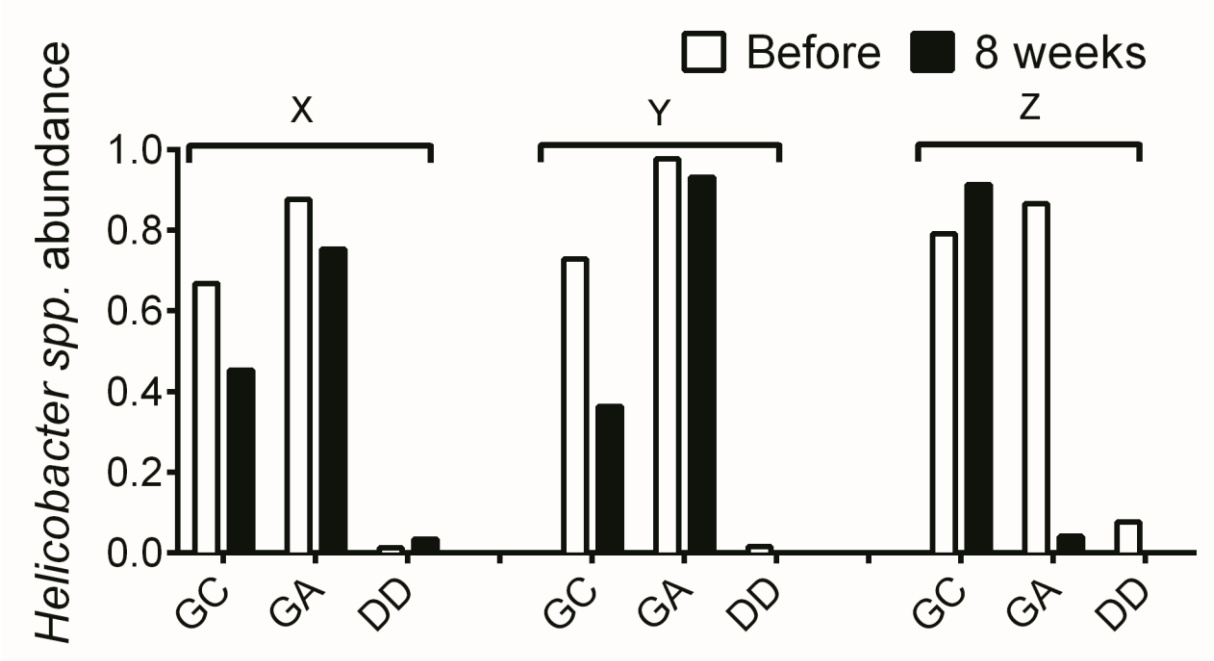

Figure 12: Relative *Helicobacter* spp. abundance in three *H. pylori* positive volunteers.
Each volunteer is shown separately as the group consists of only three volunteers. In general we see a decline of *Helicobacter* spp. abundance after eight weeks of oral vitamin D_3 supplementation, except for volunteer X's duodenum and volunteer Z's gastric corpus. (GC) gastric corpus, (GA) gastric antrum, (DD) duodenum. Adapted from Bashir et al. 2015 [125].

3.2 New perspective on cleanrooms and human pathogens

3.2.1 Phylogenetic diversity of cleanroom samples

Alpha rarefaction curves indicate, beside sufficient sampling efforts, that the bioburden is drastically lower during the actual spacecraft assemblies (PHX-D,

DAWN and MSL) compared to before or after. This might be due to the very strict gowning, cleaning and sterilization procedures which were obviously well executed, and highly effective. MSL had the highest sampling depth but lowest bioburden (Figure 13A) as GSEs undergo stringent cleaning procedures and are exposed to less handling and human contact compared to the cleanroom floors. Interestingly, microbial community profiles during active spacecraft assembly (PHX-D, DAWN, and MSL) were more similar to each other than to samples from individual locations (Figure 13B). *Moraxellaceae* was the dominating bacterial family, accounting for 83 %, 73 % and 62 % of all classified sequences for PHX-D, DAWN and MSL, respectively. The majority of all *Moraxellaceae* (94 % to 100 %) were *Acinetobacter* spp. (Figure 14), representing the most dominating taxa during spacecraft assembly.

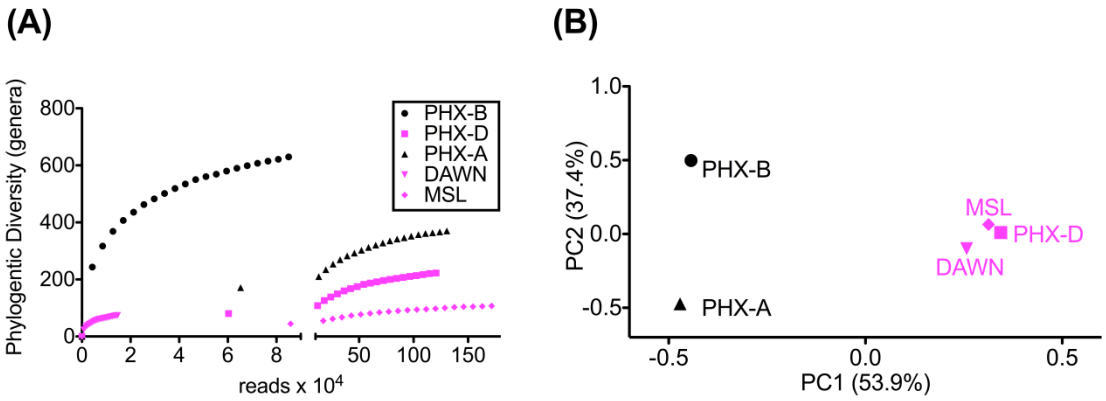

Figure 13: Diversity is lower during assembly. (A) Rarefaction curves of samples taken during assembly (PHX-D, DAWN, MSL) show less genera at the same sample size compared to samples taken before (PHX-B) and after (PHX-A) assembly. (B) Principle coordinates analysis on genus taxonomic level based on a Bray-Curtis dissimilarity matrix. In fact, samples taken during spacecraft assembly exhibit a similar community profile although they were separated by hundreds of miles.

In general, bacteria were the most dominant kingdom present in all tested cleanrooms with 63 % to >99 % of all classified sequences. Archaea and viruses on the other hand accounted for less than 0.1 % relative abundance combined. Interestingly, archaea were only present in KSC-PHSF with the highest abundance before arrival of the phoenix spacecraft (PHX-B). Nevertheless, even the highest abundance was below 0.01 % (Figure 15). Surprisingly, the amount of potentially human DNA was minimal. Only 0.04 % to 2 % of all sequences were classified as

primates: PHX-B 2 %, PHX-D 0.08 %, PHX-A 0.2 %, DAWN 0.05 %, and MSL 0.2 % (Figure 15).

In PHX-B eukaryotes made 36 % of all classified sequences. Most of these sequences (22 % of total; 60 % of eukaryotes) belong to the class of arthropods, such as insects and arachnids. In all other samples less than 0.1 % of all classified sequences were arthropods. Probably arthropod sequences originated from free DNA associated with dust particles considering that no living spiders or insects are present in any cleanrooms. In MSL all eukaryotic sequences were assigned to craniate. Fungi were also not prominent in any of the evaluated cleanrooms. The fungal abundance ranged from 0.0008 % (MSL) to 1 % (PHX-B) (Figure 15).

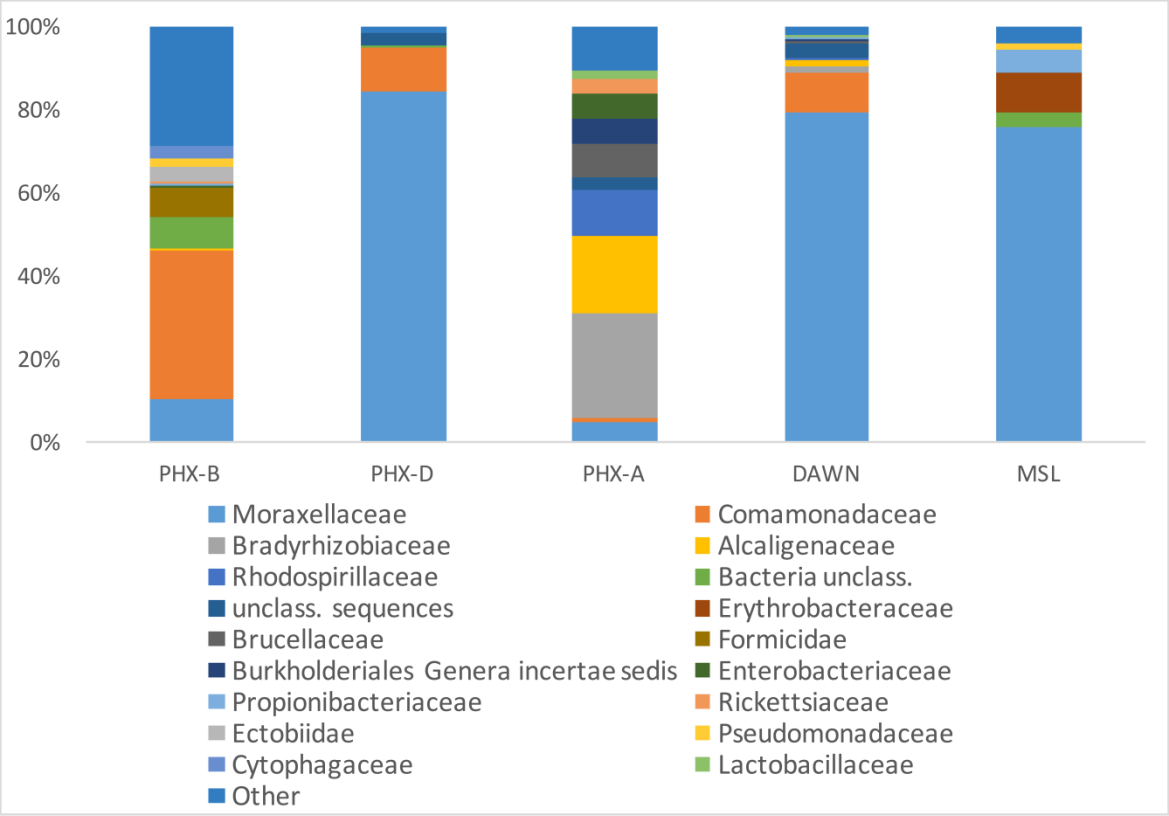

Figure 14: Moraxellaceae dominate cleanroom during spacecraft assembly. Relative abundance of taxa on family level. Human, viral and archaeal impact was minimal. All taxa with a collective abundance of equal or less than 2 % from all samples were combined in "Other".

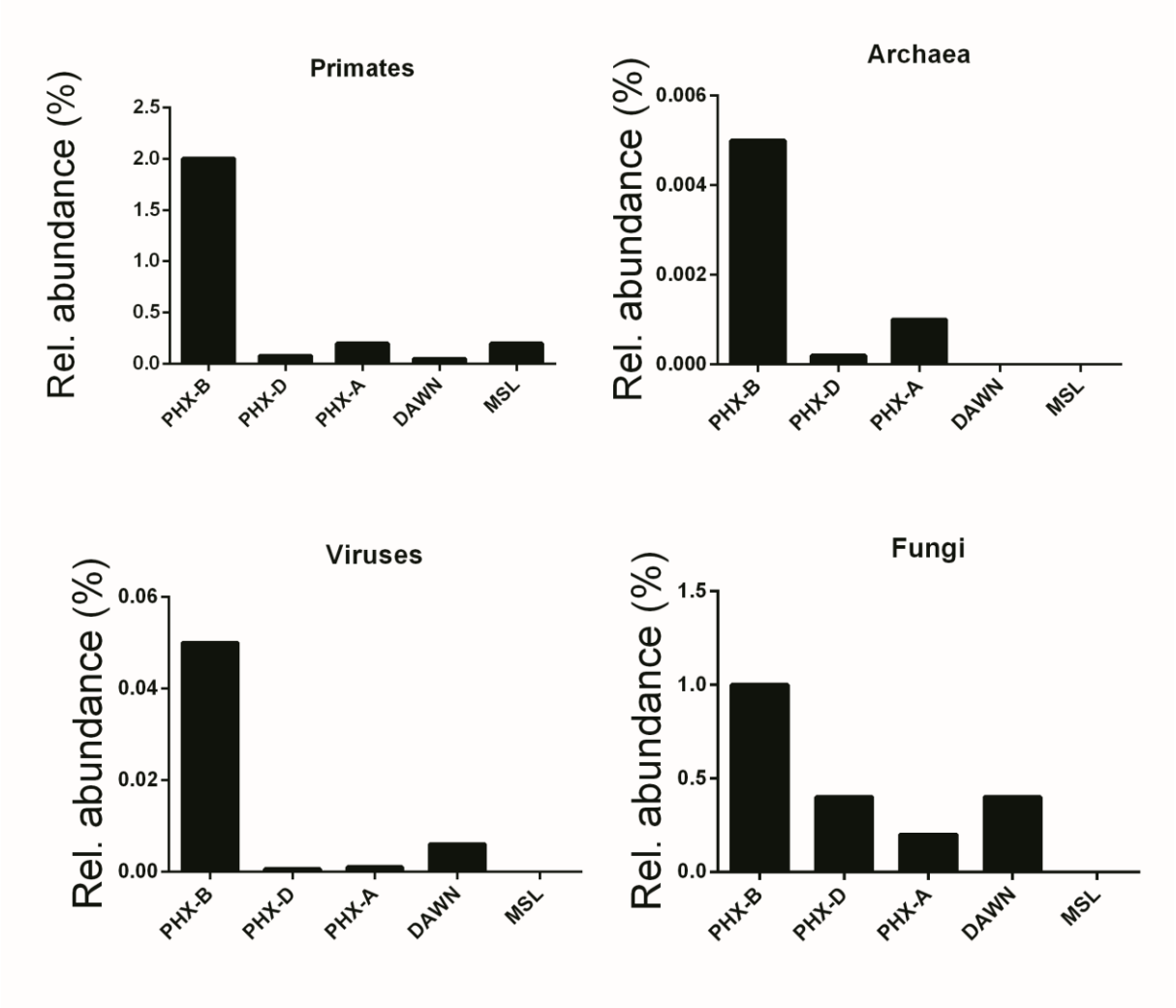

Figure 15: Non-bacterial sequence abundance. PHX-B had the highest non-bacterial load in all investigated taxa. Interestingly, archaea were only present in KSC-PHSF with the highest abundance before arrival of the phoenix spacecraft (PHX-B). Nevertheless, even the highest abundance was below 0.01 %.

3.2.2 Metabolic diversity during spacecraft assembly

Functional assignment resulted in 13,360 KEGG orthologous (KO) for PHX-B, 24,916 KOs for PHX-D, 298,350 KOs for PHX-A, 557 for DAWN and 664,699 for MSL (Table 21). As evident from Figure 16, the majority of the functional classification was assigned to metabolism (PHX-B 67 %, PHX-D 67 %, PHX-A 90 %, DAWN 75 %, MSL 29 %; Figure 16).

Although the percentage of sequences assigned to metabolism did not differ much across samples (Figure 16), a higher metabolic diversity was observed during spacecraft assembly compared to before or after (Table 21).

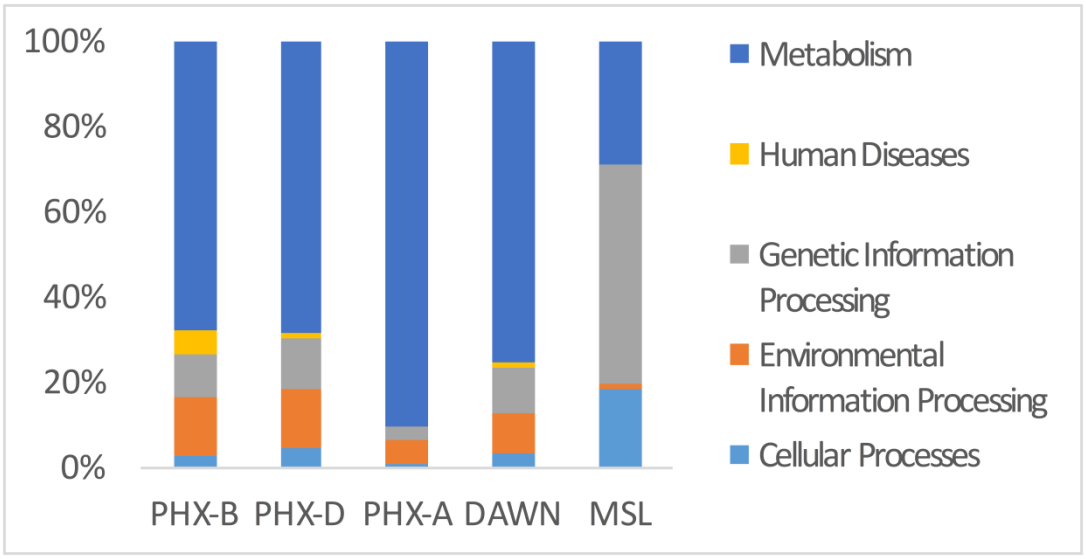

Figure 16: Metabolic genes cover majority of functional classifications. KEGG Pathway analysis of PHX-B, PHX-D, and PHX-A assembly, DAWN and MSL. During assembly of PHX-D and DAWN, a bigger fraction of all classified sequences has been assigned to metabolism.

We also found that the metabolism of pantothenate and coenzyme A is higher during assembly (PHX-D 4 %, DAWN 2 %) compared to PHX-B or PHX-A, 0.2 % respectively. Nevertheless, no function associated with pantothenate and coenzyme A was found in MSL during assembly. 52 % of all functional classification from MSL was assigned to Holliday junction DNA helicase RuvB (genetic information processing; replication and repair). In PHX-B, 9 % were assigned to genetic information processing, 12 %, in PHX-D 3 %, in PHX-A 11 % in DAWN and 52 % in MSL (Figure 16).

Table 21: Data statistics: Number of reads per sample starting with raw reads coming from the sequencer until final taxonomic and functional classification.

Sequences	PHX-B	PHX-D	PHX-A	DAWN	MSL
Raw paired reads	15,001,132	14,654,014	22,355,430	2,899,364	57,892,216
Passed quality filter	10,760,642	11,889,258	16,338,684	166,392	34,615,498
KEGG assignment	13,360	24,916	298,350	557	664,699
With tax class.	174,622	1,328,890	2,903,271	17,306	7,652,616
Metabolic diversity	145.8	188.5	42.9	119.8	5.5
Observed genera	396	36	104	82	25

3.2.3 Potential pathogens in cleanroom samples

After taxonomic classification we selectively screened the classified binned sequences of all our samples for clinically relevant pathogens (Appendix 2). In total we found 52 different human pathogens in all cleanrooms, responsible for various diseases, ranging from gastrointestinal to the nervous system. 30 different pathogens were detected in PHX-B, 18 in PHX-D, 34 in PHX-A, 10 in DAWN and 11 in MSL (Table 22). Strikingly, four potential pathogens, namely *Acinetobacter baumannii, Acinetobacter lwoffii, Escherichia coli* and *Legionella pneumophila*, were detected in all cleanrooms even though they are geographically separate. Besides these four species present in all samples, we detected potential pathogens that were exclusive to KSC-PHSF during all three time points; *Bacillus cereus, Burkholderia pseudomallei, Enterobacter cloacae, Enterococcus faecalis, Listeria monocytogenes, Pseudomonas aeruginosa, Staphylococcus aureus* and *Staphylococcus epidermidis*. In case of PHX-D, 83 % of all classified reads were identified as *Acinetobacter* spp. which is listed as a clinically relevant pathogen (Appendix 2). Additionally, we found a decreased pathogenic diversity during the actual spacecraft assembly (Table 22) in KSC-PHSF.

Table 22: Pathogenic diversity is lowest during assembly: pathogens found in the different cleanroom samples.

	Pathogen	PHX-B	PHX-D	PHX-A	DAWN	MSL
Urinary tract infection (catheter-associated or otherwise)	***Acinetobacter baumannii****	+	+	+	+	+
	Acinetobacter lwoffii*	+	+	+	+	+
	Bacillus cereus	+	+	+	-	-
	Brevundimonas diminuta	+	+	-	+	-
	Burkholderia cepacia	+	-	-	-	-
	Enterobacter aerogenes	-	-	+	-	-
	Enterobacter cloacae	+	+	+	-	-
	Enterococcus faecalis	+	+	+	+	-
	Enterococcus faecium	+	-	-	-	-
	Escherichia coli*	+	+	+	+	+
	Klebsiella pneumoniae	+	-	+	-	+
	Providencia stuartii	-	-	+	-	-
	Pseudomonas aeruginosa	+	+	+	-	+
	Serratia marcescens	-	+	-	-	-
	Staphylococcus haemolyticus	+	-	+	-	-
	Staphylococcus saprophyticus	-	-	+	-	-

	Stenotrophomonas maltophilia	+	+	-	-	+
	Sum	**13**	**10**	**12**	**5**	**6**
pneumonia, respiratory disease or infections	***Acinetobacter baumannii****	+	+	+	+	+
	Brevundimonas diminuta	+	+	-	+	-
	Burkholderia cepacia	+	-	-	-	-
	Chlamydia psittaci	-	-	+	-	-
	Coxiella burnetii	+	-	-	-	-
	Enterobacter aerogenes	-	-	+	-	-
	Klebsiella pneumoniae	+	-	+	-	+
	Legionella pneumophila*	+	+	+	+	+
	Serratia marcescens	-	+	-	-	-
	Streptococcus pneumoniae	-	-	+	-	-
	Streptococcus pyogenes	-	-	+	-	-
	Histoplasma capsulatum	+	-	-	-	-
	Sum	**7**	**4**	**7**	**3**	**3**
Meninges, central and peripheral nervous system	*Bacillus cereus*	+	+	+	-	-
	Clostridium botulinum	-	-	+	-	-
	Haemophilus influenzae	-	-	+	-	-
	Leptospira interrogans	-	-	+	-	-
	Neisseria meningitidis	-	-	+	+	+
	Polyomavirus	+	-	-	-	-
	Streptococcus pyogenes	-	-	+	-	-
	Burkholderia pseudomallei	+	+	+	-	-
	Sum	**3**	**2**	**7**	**1**	**1**
Cardiovascular (sepsis, endocarditis)	*Candida parapsilosis*	-	-	+	-	-
	Enterococcus hirae	+	-	-	-	-
	Escherichia coli*	+	+	+	+	+
	Haemophilus influenzae	-	-	+	-	-
	Neisseria meningitidis	-	-	+	+	+
	Pseudomonas aeruginosa	+	+	+	-	+
	Serratia marcescens	-	+	-	-	-
	Staphylococcus capitis	+	-	+	-	-
	Staphylococcus epidermidis	+	+	+	+	-
	Staphylococcus lugdunensis	-	-	+	-	-
	Streptococcus pneumoniae	-	-	+	-	-
	Streptococcus pyogenes	-	-	+	-	-
	Trichinella spiralis	+	-	-	-	-
	Sum	**6**	**4**	**10**	**3**	**3**
Gastrointestinal (gastroenteritis, stomach ulcers, diarrhoea)	*Campylobacter coli/jejuni*	+				
	Corynebacterium ulcerans	+	-	-	-	-
	Helicobacter pylori	-	+	-	-	-
	Listeria monocytogenes	+	+	+	-	-
	Providencia stuartii	-	-	+	-	-

	Vibrio cholerae	-	+	+	+	+
	Sum	**2**	**3**	**3**	**1**	**1**
Skin, wound or surgical opening	*Burkholderia pseudomallei*	+	+	+	-	-
	Candida parapsilosis	-	-	+	-	-
	Enterobacter cloacae	+	+	+	-	-
	Pediculus humanus corporis	+	-	-	-	+
	Staphylococcus aureus	+	+	+	-	+
	Staphylococcus haemolyticus	+	-	+	-	-
	Staphylococcus lugdunensis	-	-	+	-	-
	Streptococcus pyogenes	-	-	+	-	-
	Sum	**5**	**3**	**7**	**0**	**2**
Typhus	*Orientia tsutsugamushi*	+	-	-	-	-
	Rickettsia prowazekii	+	-	+	-	-
Lyme Disease	*Borrelia burgdorferi group*	+	-	-	-	-
Glanders, Malleus	*Burkholderia mallei*	-	-	+	+	-
Gonorrhoea	*Neisseria gonorrhoeae*	-	-	+	-	-
Pathogens associated with more than two diseases	*Alcaligenes faecalis*	-	+	+	-	-
	Peritonis, meningitis, otitis media, appendicitis, blood stream infection					
	Bacillus anthracis	-	-	+	-	-
	Anthrax (pulmonary, cutaneous and gastrointestinal)					
	Trichinella spiralis	+	-	-	-	-
	Aedema, urticaria, meningitis, encephalitis, myocarditis, and pneumonia					
Sum		**30**	**18**	**34**	**10**	**11**

+...present
-...not present
***...pathogens found in all cleanroom samples**

3.2.4 Pathogens and corresponding virulence factors in cleanrooms

Virulence factors are features which distinguish pathogens from commensals or symbionts [148]. We found that the fraction of sequences identified as virulence factors increased over time in KSC-PHSF (Phoenix, Table 23), although overall diversity was lower during assembly (Figure 13A). LMA-MTF (DAWN) had approximately half the virulence factor fraction compared to KSC-PHSF, but JPL-SAF (MSL), which was sampled from GSEs, had approximately 20 times less virulence factors compared to Phoenix.

Table 23: Accumulation of virulence factors over time: total number of virulence factors and hits normalized to hits per million reads found in cleanrooms.

Sample	Total MvirDB hits	MvirDB hits/Mio reads
PHX-B	252	23
PHX-D	5662	476
PHX-A	12703	777
DAWN	39	234
MSL	615	18

To evaluate if pathogens could be assigned to their corresponding virulence factors, we screened for virulence factors that are specifically associated with the pathogens found in our samples. We found 14 different virulence factors which correspond to the classified pathogens in PHX-B, 48 for PHX-D and 41 for PHX-A. 9 different virulence factors were found to correspond with the classified pathogens in DAWN, and 6 were found for the MSL mission.

We were particularly interested in detecting virulence factors of the four pathogens, *Acinetobacter baumannii, Acinetobacter lwoffii, Escherichia coli* and *Legionella pneumophila*, which were found in all geographically separated cleanrooms. We found a *Acinetobacter* sp. specific aminoglycoside 6`-N-acetyltransferase Iv and tetA, which are kanamycin B and tetracycline resistance genes in PHX-D. Moreover, we found adeABC, which is an *Acinetobacter baumannii* specific multidrug efflux pump and beta-lactamase TEM-1 (Appendix 3).

We found that the abundance of virulence factors with associated pathogens increased over time (Table 24). Although virulence factors diversity did not change over time, we observed a trend towards increased pathogens with associated virulence factors (Table 24, pathogenic diversity). Again, MSL had the smallest pathogenic diversity.

Table 24: Virulence factors and their corresponding pathogens.

	PHX-B	PHX-D	PHX-A	DAWN	MSL
Sum	24	2867	5458	16	501
Sum norm.	2	241	334	96	15
Virulence diversity	14	48	41	9	6
Virulence diversity norm.	1	4	2	54	0.2
Pathogenic diversity*	3	6	11	3	2

norm: normalized to counts per million reads

*** number of pathogens with ≥ 1 corresponding virulence factors**

3.2.5 Isolation of draft genome

Since all individual Phoenix samples were retrieved from the same cleanroom, a co-assembly (Table 25) is increasing the probability of covering bigger proportions of single genomes. Simply by the fact that missing or underrepresented genomic region in one sample might be complemented by the collection of multiple samples. Therefore, we performed a co-assembly with all Phoenix samples and extracted contigs into draft genomes using GroopM. All bins were analyzed with CheckM. Most bins were less than 42 % complete and were not further analyzed, except for one bin. This bin was identified as *Bradyrhizobium* spp. and shows 68.99 % completeness (Table 26).

Table 25: Co-assembly descriptives for contigs > 500 bp

	sequences
Number	39,863
Total length	40,928,550
Average	1,026
N50	1,032
Median	699
Largest	50,749

Table 26: Description of isolated draft genome

	isolate
Marker lineage	g__Bradyrhizobium (UID3699)
Likely chimeric	False
length (bp)	5,691,754
# seqs	962
GC mean ± SD	0.65 ± 0.04
Completeness	68.99 %
Contamination	0.87 %
Strain heterogeneity	60

4 Discussion

4.1 Effects of high doses of vitamin D_3 on mucosa-associated gut microbiome vary between regions of the human gastrointestinal tract

Our study clearly demonstrates the effects of an eight week high-dose oral vitamin D_3 supplementation on the upper gastrointestinal tract of the human gut microbiome. We observed a reduction of opportunistic pathogens and an increase in phylotype richness.

The effects of vitamin D_3 on the human gut microbiome have not been studied, yet. So far, the effects of a vitamin D supplementation on the gut microbiome have only been studied in murine colitis model. The active form of vitamin D, 1,25-hydroxy-vitamin D, protected these mice from dextran sodium sulfate–induced colitis, which is a common model of IBD and colitis of was significantly more severe in VDR and CYP27B1 knockout mice [149]. The microbial composition of the mouse gut is, however substantially different from the human microbiome [150], any interventional effects that are found in mouse studies need to be confirmed in humans.

Biopsies offer the best access to monitor changes in the mucosa-associated microbiome in adequate detail. By sampling the major regions of the human gastrointestinal tract, we found a decrease in Proteobacteria mostly in the mucosa of the upper gastrointestinal tract together with an overall increase in phylotype richness. In particular, the class of Gammaproteobacteria decreased significantly, including genera of typical opportunistic pathogens, such as *Pseudomonas* spp. and *Escherichia/Shigella* spp. Mucosal samples of the lower gastrointestinal tract and stool samples displayed only minor changes which supports recent findings that the mucosa-associated microbiome varies significantly from stools [98]. Only biopsies offer the chance to monitor changes in the mucosa-associated microbiome in adequate detail.

Due to easy access to stool and the abundance of sample, stool is often the sole source used for the investigation of the human intestinal microbiome [93, 94, 101, 102, 151]. Some studies have already focused on rectal/colonic mucosal samples

[98, 103, 152] to investigated external influences, such as the effects of diet [93, 94, 153, 154], physical exercise [151] or drugs [99] and have identified factors associated with certain diseases like IBD, IBS or obesity [93, 103, 152]. Other interventional studies using therapeutics which aim at altering the microbiome (e.g. prebiotics, omega-3 fatty acids, etc.) were not able to find significant changes in stool samples [155, 156] since some effects might be limited to the upper gastrointestinal tract. Our study design with comprehensively sampling the majority of the human gastrointestinal tract allowed us to show that the major effects of vitamin D on the human gastrointestinal microbiome were only prominent in the upper gastrointestinal tract. Using stool samples only, we would have missed this conclusion. Future microbiome study designs need to be aware that some interventions such as vitamin D supplementation can have significantly pronounced effects on different regions of the gastrointestinal tract. Moreover, stool appears to not always be the best surrogate to monitor these changes.

In general, the assessment of the microbial composition of samples from the gastrointestinal system is highly affected by sample processing, such as sample storage, DNA extraction, primer pairs used, and high inter-individual variation [157]. The Bacteroidetes to Firmicutes ratio described in the literature varies noticeably. Some studies state a microbial composition dominated by Firmicutes [112, 158, 159] while others describe similar ratios as observed in our study [157, 160, 161]. One possible explanation for receiving different results in these studies is perhaps that freezing of stool samples before DNA extraction can lead to an increase in the Bacteroidetes to Firmicutes ratio [162].

We could confirm the variation in microbial composition along the human gastrointestinal tract observed by Zhang et al., who found a similar bacterial distribution pattern in the ileum and colon of healthy humans [163].

The microbiome along the gastrointestinal tract is strongly affected by different niche factors, such as pH, oxygen concentration, tension and nutrients. Microbiome data do not necessarily reflect the anatomical categorisation. In our study, the microbial composition in the duodenum was more similar to the stomach rather than to the terminal ileum, although both are part of small intestines. The terminal ileum was more similar to colon and stool samples. This reflects the fact that different regions of

the human gastrointestinal tract differ in function, permeability, type and abundance of immune cells [129, 164].

Given that the most prominent changes were exclusively detected for Gammaproteobacteria residing in the upper gastrointestinal tract, we performed an agar diffusion test to examine vitamin D_3 toxicity on Gammaproteobacteria. As we could not observe any zone of inhibition, neither under aerobic nor under anaerobic conditions, we concluded that vitamin D_3 does not directly interfere with the growth of Gammaproteobacteria. However, we cannot exclude that any products of vitamin D_3 or its interaction with the immune system exert the reduction of this particular bacterial class.

As part of the first line of resistance against foodborne diseases $CD8^+$ T cells are significantly more abundant in the stomach (gastric corpus and antrum) compared to lower gastrointestinal regions [129]. The decrease of Gammaproteobacteria exclusively found in the mucosa of the upper gastrointestinal tract might have been driven by an increase of $CD8^+$ T cells, we observed after eight weeks of oral vitamin D_3 supplementation. These cells show the highest VDR expression levels compared to other cells of the immune system [109]. This advocates that $CD8^+$ T cell function is also regulated by vitamin D. High levels of vitamin D have been linked to a decline in naïve $CD8^+$ T cells and an increase in cytotoxic $CD8^+$ T cells [165]. We found a trend towards increased $CD8^+$ T cell numbers in all investigated regions of the human gastrointestinal tract after eight weeks of vitamin D_3 supplementation, but we did not examine the different $CD8^+$ T cell subpopulations in the course of this study. Mucosal $CD8^+$ T cells stimulated by vitamin D_3 could contribute for the effects we saw on Gammaproteobacteria together with unknown and known effects of vitamin D, such as the induction of antimicrobial peptides (e.g. CAMP; cathelicidin antimicrobial peptide [166]) and the stimulation of certain cytokine production by mucosal dendritic cells [167]. Moreover, mouse $CD8^+$ T cells [168] and human activated T cells [169] have been shown to express 1α-hydroxylase (CYP27B1), the enzyme responsible for the production of hormonal active 1,25-hydroxy vitamin D. 1,25-hydroxy vitamin D alters cytokine production pattern in human $CD8^+$ T cells [170] and is known to act as a stop signal for inflammation [171, 172]. In addition, $CD8^+$ T cells are also capable of

directly killing harmed or infected proinflammatory host cells, thus resulting in a reduced inflammatory environment [173].

Under inflammatory condition, opportunistic pathogens, such as *Escherichia/Shigella* spp. or *Pseudomonas* spp. can outcompete commensal bacteria [174], because they are better adapted to inflammation by evolutionary means. By reducing inflammation, this competitive advantage is abated, which in return allows beneficial commensals such as Bacteroidetes to outcompete opportunistic pathogens, resulting in increased phylotype richness which we found in our study.

Increase bacterial richness could be specifically beneficial for gastrointestinal diseases, such as non-alcoholic steatohepatitis or IBD which is characterized by lower bacterial richness, decreased levels of vitamin D and an overgrowth of pathogens, mostly Gammaproteobacteria [102, 161, 175–179]. Various studies have described beneficial effects of a vitamin D in patients suffering from IBD [180] which could be to some extent be explained by restoration of gut microbiome homeostasis. These studies have reported a reduction of opportunistic pathogens such as Enterobacteriaceae, which belong to the class of Gammaproteobacteria [102, 176–178, 181]. Metagenomic sequencing like16S rRNA amplicon sequencing used in this study provides only relative abundances. Therefore we cannot actually differentiate between an increase in Bacteroidia and a decrease in Gammaproteobacteria. We hypothesized that a decrease in Gammaproteobacteria is more likely, given that we found several genera of Gammaproteobacteria to be reduced, whereas no single genus of Bacteroidia was significantly changed after eight weeks of oral vitamin D_3 supplementation.

Another surprising finding was that vitamin D_3 supplementation lead to a decrease in overall abundance of *Helicobacter* spp. in the *H. pylori* positive subgroup. These data support the finding that CYP27B1 knockout mice unable to produce 1,25-hydroxy vitamin D show a significantly higher relative abundance of Helicobacteriaceae compared to wild type mice [149]. In these knockout mice 1,25-hydroxy vitamin D supplementation resulted in a reduction of Helicobacteriaceae levels. Recently, Hosoda et al. elucidated the mechanisms responsible for the vitamin D action on *H. pylori*. They reported that vitamin D_3, which is a structurally instable compound, decomposes into vitamin D decomposition products. A certain decomposition product

was able to interact with dimyristoyl-phosphatidylethanolamine, the most abundant phosphatidylethanolamine in *H. pylori,* subsequently lysing the pathogen. Other commensal bacteria, including Gammaproteobacteria, such as *E. coli*, *Pseudomonas aeruginosa* and *Salmonella typhimurium* were not affected by this decomposition product [182]. Interestingly, *H. pylori* itself induces increased expression of the VDR [183].

Dosing of vitamin D_3 in our study was higher than current recommendation [184] and resulted in mean 25-hydroxy vitamin D levels of 55.2 ng/mL, which is still significantly below toxic levels (>150 ng/mL). The safety of such a high vitamin D supplementation has previously been described in a clinical trial [185]. No study participant reported any adverse effects and none of the evaluated safety parameters, including serum and urine calcium and phosphate levels were significantly changed during eight weeks of a high-dose of oral vitamin D_3 supplementation. Nevertheless, we would not suggest to generally administer such high vitamin D_3 doses for longer time periods to humans before a complete evaluation of long term risks and benefits in randomized interventional trials. Despite the relatively small study cohort, mainly because of the invasive procedures, and a high degree of inter-individual variation of the intestinal microbiome, we could observe a significant modulatory effect of vitamin D_3 in the upper gastrointestinal tract.

The increased bacterial richness and the pronounced reduction of Gammaproteobacteria, which included typical opportunistic pathogens support the positive effect of a high dose oral vitamin D_3 supplementation on the human gut microbiome. This might partly explain the effects of vitamin D treatment in patients suffering from IBD or bacterial infections and may inspire future studies to investigate the effects of vitamin D in these patients. The fact that the vitamin D only affected the microbiome of the upper gastrointestinal tract should be considered in future microbiome studies and encourages the assessment of the upper gastrointestinal tract in addition to stool and colonic mucosal samples.

4.2 New perspective on cleanrooms and human pathogens

In this study, we demonstrated for the first time the presence of pathogens and their corresponding virulence in spacecraft assembly cleanrooms. Our approach allowed us not only to prove the presence of pathogens in the spacecraft assembly cleanroom, but also their associated virulence factors. Most studies investigating the cleanroom microbiome have only used 16S rRNA amplicon sequencing [46, 47]. For example, the archived samples from KSC-PHSF during the Phoenix mission used in this study have previously been described using a cultivation based [186] and cultivation independent technique [120]. On one hand, cultivation based techniques offer a very limited insight into the wide spectrum of microbial diversity, given that most microorganisms are not cultivable, while 16S rRNA amplicon sequencing on the other hand shows a more broad picture. However, it does not allow a reliable phylogenetic classification below genus level and does not provide any information regarding virulence factors and potential pathogenicity.

We observed that cleanroom samples are dominated by bacteria as reported previously (Weinmaier et al., 2015). In contrast to previous studies, which found substantially more human, archaeal and viral sequences in cleanrooms (Moissl-Eichinger, 2011; Weinmaier et al., 2015), we found significantly less of each taxon in all cleanrooms tested in this study. These previous studies have sampled uncontrolled gowning area and ISO-8 cleanrooms, where no active spacecraft assembly was ongoing. Moreover, each cleanroom is unique, because of factors such as geographical location [187], assembly activities, different decontamination procedures and most importantly, different workers which represent the main source of contamination.

We saw an increased metabolic diversity in samples collected from cleanrooms during spacecraft assembly. Cleanrooms are sometimes referred to as extreme environments; not only due to strict decontamination procedures, but also due to the lack of nutrients, water and cofactors [116, 186]. Since there are very few resources to rely on in an area that is maintained to be uninhabitable, any microbes able to survive here would have to synthesize all necessary factors themselves. Sterilization procedures and gowning requirements are even stricter during assembly, making it even harder for microorganisms to survive. Strict gowning protocols and increased

frequency of cleaning lead to a reduced import of human associated microbes despite high human activities in the cleanroom during assembly. This might also explain lower phylogenetic diversity during Phoenix spacecraft assembly compared to before or after assembly. In addition to the decreased phylogenetic diversity, pathogenic diversity was also lower during spacecraft assembly. A considerable amount, in case of MSL more that 50%, of all reads was assigned to genetic information processing. This highlights the importance of genetic information processing, including DNA repair in such a harsh environment. Surprisingly, microbial profiles during assembly were very similar. Although DAWN and MSL samples were collected from geographically distinct locations, they were more similar to PHX-D than PHX-B or PHX-A. This suggests that decontamination procedures have a bigger effect on the cleanroom microbiome than location. Taken together, these results show that decontamination and gowning measures were not only sufficient, but also well executed.

Most virulence factors are organized in so-called pathogenicity-islands [188]. Commensals can turn into pathogens by acquiring a pathogenicity-island through phages or horizontal gene transfer. For example, wild type *Vibrio cholerae* are not able to cause deadly diarrhea. Only upon infection by the CTX prophage, they acquire a pathogenicity island coding for virulence factors, such as the cholera toxin and pili needed for the recognition of the respective host and disease induction [148]. Therefore, virulence factor detection is the only reliable method to identify pathogens.

Acinetobacter baumannii, Acinetobacter lwoffii, Escherichia coli and Legionella pneumophila were found in all samples, although samples were collected from three geographically distinct sites. These prevalent pathogens have to be very resistant to overcome all the cleaning and decontamination procedures. *Acinetobacter* spp., such as *A. baumannii* and *A. lwoffii* are non-fastidious and can rely on a single energy source from different substrates [189]. They are resistant to radiation [190] and several disinfectants and can survive in a wide range of temperatures [191] and even low pH. These features might explain why *Acinetobacter* spp. represent the most dominating species during spacecraft assembly in this study. *Acinetobacter* spp. have also been reported in high abundance in cleanrooms in previous studies [46, 120]. *A. baumannii* has been isolated from water and soil [192], but it has also been

found in other hostile environments such as intensive care units. Although *A. baumannii* is not pathogenic for healthy individuals, it can be an opportunistic pathogen in immunocompromised patients. Hence, it is one of the ESKAPE (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumanii, Pseudomonas aeruginosa,* and *Enterobacter species*) pathogens [193], which are multidrug-resistant bacteria responsible for the majority of nosocomial infections [194].

We found *A. baumannii* specific beta-Lactamase TEM-1, AdeABC and another cation/multidrug efflux pump which might be responsible for *A. baumannii's* resistance against all decontamination measures. AdeABC alone is responsible for resistance to aminoglycosides, tetracyclines, erythromycin, chloramphenicol, trimethoprim, fluoroquinolones, some beta-lactams, and also recently tigecycline since they have been described as substrates for this multidrug efflux pump [195]. We did not detect *A. lwoffii* associated virulence factors in our data set. MvirDB has only three *A. lwoffii* (formerly known as *Acinetobacter calcoaceticus*) associated virulence factors (two beta-lactamases and a chloramphenicol acetyl transferase). Nevertheless, the presence of this opportunistic pathogen in all our sample collection from locations separated by hundreds of miles, its resistant features, and our finding that *Acinetobacter* spp. were dominating in all three locations during assembly suggests that A. *lwoffii* and *A. baumannii* are actually viable in the spacecraft cleanroom environment. *L. pneumophila*, another pathogen present in all three distinct locations, is the causative agent of the Legionnaires' disease [196], with symptoms such as fever, chills, and coughing. We found Legionella secretion pathway protein E (LspE), which is part of a type II secretion system required for full virulence and environmental persistence [197]. In addition, other *L. pneumophila* associated virulence factors, such as the catalase-peroxidase KatB and superoxide dismutase were present, potentially explaining why this species is resistant to hydrogen peroxide treatment, one of the decontamination procedures. The last potential pathogen we found in all cleanrooms was *E. coli*. Although we have found several virulence factors such as transposases and antimicrobial resistance genes, we cannot confirm whether or not this specific *E. coli* is a pathogen, given that more and more antimicrobial resistance genes are being found in commensal *E. coli* [198–200]. Although we think that the four pathogens found in all geographically separated

cleanrooms are alive, given their resistant features, we are not able to determine if the classified taxa and functions originate from intact living or dead cells. In an ongoing study we're including propidium monoazide staining, enabling us to differentiate between sequences coming from intact live and dead microorganisms.

Interestingly, virulence factor abundance increased over time, despite lower phylogenetic diversity during assembly. We have concluded that virulence factors may provide a survival advantage in this very hostile environment [189]. Multidrug efflux pumps might be pumping out harmful compounds before they are able to execute their deadly effect [201]. We also found pathogens not belonging to the bacterial kingdom; such as *Candida parapsilosis*, a fungus, which plays an important role in wound and tissue sepsis of immunocompromised patients and makes up to 15% of all *Candida* infections.

Besides detecting pathogens in cleanrooms, we were able to extract a *Bradyrhizobium* draft genome from the Phoenix datasets. Given that, we had only three similar samples from the same location, the draft genome was only 69% complete. In a future study, we will include more samples from the same habitat to extract a more complete draft or even a full genome. These genomes might serve as a reference for future sample return missions, which aim at detecting extraterrestrial life. If we find DNA in extraterrestrial samples, we could exclude all to sequences mapping to this reference to confirm that it's actually non-earthborne.

One limitation of this study is the low biomass in cleanroom samples, due to the repeated strict cleaning and decontamination practices. MDA was necessary to acquire DNA concentrations sufficient for library preparation. MDA can introduce bias by favoring some DNA fragments over others [202]. Therefore, some microorganisms might not have been detected in our approach, while others might be overrepresented. Although we performed stringent quality filtering of our reads, it is impossible to get rid of all errors and biases.

Humans spend most of their lives indoors [203]. Recent studies have speculated that human microbiome is the major contributor to the overall indoor microbiome [204]. Stringent cleaning and maintenance practices in highly controlled indoor environments such as cleanrooms, hospitals and intensive care units may lead to a relative increase of human pathogens in these environments. This may have serious

impact on health of the inhabitants, specifically infants. Studying the impact of disinfectants and other sterilisation techniques in indoor environments might be of particular interest for hospitals, but also for caring parents. Moreover, monitoring pathogens and virulence factors in these indoor environments may prevent diseases such as nosocomial infections and sustain human health.

The results of this study could be used to develop fast and cost-efficient tests [205] to detect the presence of specific pathogens or their virulence factors in enclosed environments such as public transport, pharmaceutical cleanrooms, hospitals and intensive care units. This study has broadened our understanding of the role of pathogens in such highly controlled environments and should be considered for microbial monitoring of the ISS during sustained presence of humans in space and future manned missions to Mars.

5 Bibliography

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

2. Kaeberlein T, Lewis K, Epstein SS (2002) Isolating “uncultivable” microorganisms in pure culture in a simulated natural environment. Science 296:1127–1129

3. Nichols D, Cahoon N, Trakhtenberg EM, Pham L, Mehta A, Belanger A, Kanigan T, Lewis K, Epstein SS (2010) Use of ichip for high-throughput in situ cultivation of “uncultivable” microbial species. Appl Environ Microbiol 76:2445–50

4. Ling LL, Schneider T, Peoples AJ, et al (2015) A new antibiotic kills pathogens without detectable resistance. Nature 517:455–459

5. Gavrish E, Bollmann A, Epstein S, Lewis K (2008) A trap for in situ cultivation of filamentous actinobacteria. J Microbiol Methods 72:257–62

6. Epstein S, Lewis K, Nichols D, Gavrish E (2010) New approaches to microbial isolation. In: Man. Ind. Microbiol. Biotechnol. pp 3–12

7. Sizova M V, Hohmann T, Hazen A, Paster BJ, Halem SR, Murphy CM, Panikov NS, Epstein SS (2012) New approaches for isolation of previously uncultivated oral bacteria. Appl Environ Microbiol 78:194–203

8. Tanaka T, Kawasaki K, Daimon S, Kitagawa W, Yamamoto K, Tamaki H, Tanaka M, Nakatsu CH, Kamagata Y (2014) A hidden pitfall in the preparation of agar media undermines microorganism cultivability. Appl Environ Microbiol 80:7659–66

9. Amann R, Fuchs BM (2008) Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. Nat Rev Microbiol 6:339–348

10. Cannone JJ, Subramanian S, Schnare MN, et al (2002) The comparative RNA

web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics 3:2

11. Ramm J, Lupu A, Hadas O, Ballot A, Rücker J, Wiedner C, Sukenik A (2012) A CARD-FISH protocol for the identification and enumeration of cyanobacterial akinetes in lake sediments. FEMS Microbiol Ecol 82:23–36

12. Moissl-Eichinger C (2011) Archaea in artificial environments: their presence in global spacecraft clean rooms and impact on planetary protection. ISME J 5:209–19

13. Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–69

14. Chandler DP, Brockman FJ, Fredrickson JK (1997) Use of 16S rDNA clone libraries to study changes in a microbial community resulting from ex situ perturbation of a subsurface sediment Understanding microbial community structure and dynamics in natural environments is a fundamental component of microbial. 20:

15. Stackebrandt E, Pukall R, Ulrichs G, Rheims H (1999) Analysis of 16S rDNA clone libraries: part of the big picture. Proc 8th Int Symp Microb Ecol Microb Biosyst new Front Atl Canada Soc Microb Ecol Halifax, Nov Scotia, Canada 1–9

16. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–7

17. Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S (2007) Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl Environ Microbiol 73:278–88

18. McAuliffe L, Ellis RJ, Lawes JR, Ayling RD, Nicholas RAJ (2005) 16S rDNA PCR and denaturing gradient gel electrophoresis; a single generic test for detecting and differentiating Mycoplasma species. J Med Microbiol 54:731–9

19. Fischer SG, Lerman LS (1979) Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis. Cell 16:191–200

20. Krebs JE, Vaishampayan P, Probst AJ, Tom LM, Marteinsson VT, Andersen GL, Venkateswaran K (2014) Microbial Community Structures of Novel Icelandic Hot Spring Systems Revealed by PhyloChip G3 Analysis. Astrobiology 14:229–240

21. Moissl-Eichinger C, Auerbach AK, Probst AJ, et al (2015) Quo vadis? Microbial profiling revealed strong effects of cleanroom maintenance and routes of contamination in indoor environments. Sci Rep 5:9156

22. La Duc MT, Osman S, Vaishampayan P, Piceno Y, Andersen G, Spry JA, Venkateswaran K (2009) Comprehensive census of bacteria in clean rooms by using DNA microarray and cloning methods. Appl Environ Microbiol 75:6559–67

23. Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. Nat Methods 6:639–641

24. Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT (2014) Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 9:e93827

25. Gosalbes MJ, Durbán A, Pignatelli M, Abellan JJ, Jiménez-Hernández N, Pérez-Cobas AE, Latorre A, Moya A (2011) Metatranscriptomic approach to analyze the functional human gut microbiota. PLoS One 6:e17447

26. Voelkerding K V., Dames SA, Durtschi JD (2009) Next-generation sequencing:from basic research to diagnostics. Clin Chem 55:641–658

27. Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One 6:e27310

28. 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 79:5112–20

29. Bentley DR, Balasubramanian S, Swerdlow HP, et al (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–9

30. Rusk N (2010) Torrents of sequence. Nat Methods 8:44–44

31. Erguner B, Ustek D, Sagiroglu MS, Ergüner B, Üstek D, Sağıroğlu MŞ (2015) Performance Comparison of Next Generation Sequencing Platforms *. IEEE 2015:6453–6456

32. Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. Hum Mol Genet 19:R227–40

33. Quail M a, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341

34. Thompson JF, Milos PM (2011) The properties and applications of single-molecule DNA sequencing. Genome Biol 12:217

35. Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK (2016) Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. PeerJ 4:e1869

36. Fichot EB, Norman RS (2013) Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. Microbiome 1:10

37. Singer E, Bushnell B, Coleman-Derr D, et al (2016) High-resolution phylogenetic microbial community profiling. ISME J. doi: 10.1038/ismej.2015.249

38. 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

39. Bodilis J, Nsigue-Meilo S, Besaury L, Quillet L (2012) Variable copy number, intra-genomic heterogeneities and lateral transfers of the 16S rRNA gene in Pseudomonas. PLoS One 7:e35647

40. Konstantinidis KT, Tiedje JM (2007) Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microbiol 10:504–9

41. Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ Microbiol 12:1889–1898

42. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200

43. Westcott SL, Schloss PD (2015) De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. PeerJ 3:e1487

44. Whittaker RH (1960) Vegetation of the Siskiyou Mountains, Oregon and California. Ecol Monogr 30:407

45. Zuur AF, Ieno EN, Smith GM (2007) Analysing Ecological Data. doi: 10.1007/978-0-387-45972-1

46. Mahnert A, Vaishampayan P, Probst AJ, Auerbach A, Moissl-Eichinger C, Venkateswaran K, Berg G (2015) Cleanroom Maintenance Significantly Reduces Abundance but Not Diversity of Indoor Microbiomes. PLoS One 10:e0134848

47. Vaishampayan P, Probst AJ, La Duc MT, Bargoma E, Benardini JN, Andersen GL, Venkateswaran K (2013) New perspectives on viable microbial communities in low-biomass cleanroom environments. ISME J 7:312–24

48. Weinmaier T, Probst AJ, La Duc MT, Ciobanu D, Cheng J-F, Ivanova N, Rattei

T, Vaishampayan P (2015) A viability-linked metagenomic analysis of cleanroom environments: eukarya, prokaryotes, and viruses. Microbiome 3:62

49. Nocker A, Sossa KE, Camper AK (2007) Molecular monitoring of disinfection efficacy using propidium monoazide in combination with quantitative PCR. J Microbiol Methods 70:252–60

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

51. Qin J, Li R, Raes J, et al (2010) A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59–65

52. Qin J, Li Y, Cai Z, et al (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490:55–60

53. Tringe SG, Rubin EM (2005) Metagenomics: DNA sequencing of environmental samples. Nat Rev Genet 6:805–14

54. Bashir A, Klammer AA, Robins WP, et al (2012) A hybrid approach for the automated finishing of bacterial genomes. Nat Biotechnol 30:701–7

55. Tsai Y-C, Conlan S, Deming C, Segre JA, Kong HH, Korlach J, Oh J (2016) Resolving the Complexity of Human Skin Metagenomes Using Single-Molecule Sequencing. MBio 7:e01948–15

56. Frank JA, Pan Y, Tooming-Klunderud A, Eijsink VGH, McHardy AC, Nederbragt AJ, Pope PB (2015) Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. bioRxiv. doi: 10.1101/026922

57. Salzberg SL, Delcher AL, Kasif S, White O (1998) Microbial gene identification using interpolated Markov models. Nucleic Acids Res 26:544–8

58. Foissac S, Gouzy J, Rombauts S, Mathe C, Amselem J, Sterck L, de Peer Y V., Rouze P, Schiex T (2008) Genome Annotation in Plants and Fungi: EuGene as a Model Platform. Curr Bioinform 3:87–97

59. Morgan XC, Huttenhower C (2012) Chapter 12: Human microbiome analysis. PLoS Comput Biol 8:e1002808

60. McHardy AC, Martín HG, Tsirigos A, Hugenholtz P, Rigoutsos I (2007) Accurate phylogenetic classification of variable-length DNA fragments. Nat Methods 4:63–72

61. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C (2014) Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146

62. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods 9:811–814

63. Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome Res 377–386

64. Meyer F, Paarmann D, D'Souza M, et al (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386

65. Markowitz VM, Ivanova NN, Szeto E, et al (2008) IMG/M: a data management and analysis system for metagenomes. Nucleic Acids Res 36:D534–8

66. Brady A, Salzberg SL (2009) Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. Nat Methods 6:673–6

67. Leung HCM, Yiu SM, Yang B, Peng Y, Wang Y, Liu Z, Chen J, Qin J, Li R, Chin FYL (2011) A robust and accurate binning algorithm for metagenomic sequences with arbitrary species abundance ratio. Bioinformatics 27:1489–95

68. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–10

69. Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60

70. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

71. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 42:D199–205

72. Tatusov RL, Fedorova ND, Jackson JD, et al (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41

73. Finn RD, Coggill P, Eberhardt RY, et al (2015) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:gkv1344

74. Lim YW, Schmieder R, Haynes M, et al (2012) Metagenomics and metatranscriptomics: Windows on CF-associated viral and microbial communities. J Cyst Fibros. doi: 10.1016/j.jcf.2012.07.009

75. Xiong X, Frank DN, Robertson CE, et al (2012) Generation and Analysis of a Mouse Intestinal Metatranscriptome through Illumina Based RNA-Sequencing. PLoS One 7:e36009

76. Abram F (2015) Systems-based approaches to unravel multi-species microbial community functioning. Comput Struct Biotechnol J 13:24–32

77. Grossman AD, Straus DB, Walter WA, Gross CA (1987) Sigma 32 synthesis can regulate the synthesis of heat shock proteins in Escherichia coli. Genes Dev 1:179–84

78. Wilmes P, Bond PL (2006) Metaproteomics: Studying functional gene expression in microbial ecosystems. Trends Microbiol 14:92–97

79. Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, Shah M, Hettich RL, Banfield JF (2005) Community proteomics of a natural microbial biofilm. Science 308:1915–20

80. Hettich RL, Pan C, Chourey K, Giannone RJ (2013) Metaproteomics : Harnessing the Power of High Performance Mass Spectrometry to Identify the Suite of Proteins That Control Metabolic Activities in Microbial Communities.

Anal Chem 85:4203–4214

81. Llewellyn C a, Sommer U, Dupont CL, Allen AE, Mark R (2015) Using community metabolomics as a new approach to discriminate marine microbial particulate organic matter in the western English Channel. Prog Oceanogr 1–13

82. Devaraj S, Hemarajata P, Versalovic J (2013) The human gut microbiome and body metabolism: implications for obesity and diabetes. Clin Chem 59:617–28

83. Nicholson JK, Holmes E, Wilson ID (2005) Gut microorganisms, mammalian metabolism and personalized health care. Nat Rev Microbiol 3:431–8

84. Roberts LD, Souza AL, Gerszten RE, Clish CB (2012) Targeted Metabolomics. In: Curr. Protoc. Mol. Biol. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp 1–34

85. Tang J (2011) Microbial Metabolomics. Curr Genomics 12:391–403

86. Nicholson J, Holmes E, Kinross J (2012) Host-gut microbiota metabolic interactions. Science 336:1262–1268

87. O'Shea EF, Cotter PD, Stanton C, Ross RP, Hill C (2012) Production of bioactive substances by intestinal bacteria as a basis for explaining probiotic mechanisms: bacteriocins and conjugated linoleic acid. Int J Food Microbiol 152:189–205

88. Neish AS (2009) Microbes in gastrointestinal health and disease. Gastroenterology 136:65–80

89. Lawley TD, Walker AW (2013) Intestinal colonization resistance. Immunology 138:1–11

90. Hooper LV, Littman DR, Macpherson AJ (2012) Interactions between the microbiota and the immune system. Science 336:1268–73

91. Gravitz L (2012) Microbiome: The critters within. Nature 485:12–13

92. Tilg H, Moschen AR (2014) Microbiota and diabetes: an evolving relationship.

Gut 0:1513–1521

93. Halmos EP, Christophersen CT, Bird AR, Shepherd SJ, Gibson PR, Muir JG (2014) Diets that differ in their FODMAP content alter the colonic luminal microenvironment. Gut 0:1–8

94. David L a, Maurice CF, Carmody RN, et al (2014) Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–63

95. Matijašić BB, Obermajer T, Lipoglavšek L, Grabnar I, Avguštin G, Rogelj I (2014) Association of dietary type with fecal microbiota in vegetarians and omnivores in Slovenia. Eur J Nutr 53:1051–1064

96. Turnbaugh PJ, Hamady M, Yatsunenko T, et al (2009) A core gut microbiome in obese and lean twins. Nature 457:480–4

97. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman D a (2012) The application of ecological theory toward an understanding of the human microbiome. Science 336:1255–62

98. Gorkiewicz G, Thallinger GG, Trajanoski S, Lackner S, Stocker G, Hinterleitner T, Gülly C, Högenauer C (2013) Alterations in the colonic microbiota in response to osmotic diarrhea. PLoS One 8:e55817

99. Dethlefsen L, Relman DA (2011) Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A 108:4554–4561

100. Ananthakrishnan AN, Cagan A, Gainer VS, et al (2014) Higher plasma vitamin D is associated with reduced risk of Clostridium difficile infection in patients with inflammatory bowel diseases. Aliment Pharmacol Ther 39:1136–42

101. Gagliani N, Hu B, Huber S, Elinav E, Flavell R a (2014) The fire within: microbes inflame tumors. Cell 157:776–83

102. Manichanh C, Borruel N, Casellas F, Guarner F (2012) The gut microbiota in IBD. Nat Rev Gastroenterol Hepatol 9:599–608

103. Gevers D, Kugathasan S, Denson L a, et al (2014) The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 15:382–392

104. Kovach MA, Ballinger MN, Newstead MW, Zeng X, Bhan U, Yu F, Moore BB, Gallo RL, Standiford TJ (2012) Cathelicidin-related antimicrobial peptide is required for effective lung mucosal immunity in Gram-negative bacterial pneumonia. J Immunol 189:304–11

105. Holick MF (2004) Vitamin D: importance in the prevention of cancers, type 1 diabetes, heart disease, and osteoporosis. Am J Clin Nutr 79:362–371

106. Ananthakrishnan AN, Cagan A, Gainer VS, et al (2013) Normalization of plasma 25-hydroxy vitamin D is associated with reduced risk of surgery in Crohn's disease. Inflamm Bowel Dis 19:1921–7

107. Holick MF (2008) The vitamin D deficiency pandemic and consequences for nonskeletal health: mechanisms of action. Mol Aspects Med 29:361–368

108. Grant WB, Schuitemaker GE (2010) Health benefits of higher serum 25-hydroxyvitamin D levels in The Netherlands. J Steroid Biochem Mol Biol 121:456–458

109. Veldman CM, Cantorna MT, DeLuca HF (2000) Expression of 1,25-dihydroxyvitamin D(3) receptor in the immune system. Arch Biochem Biophys 374:334–338

110. Battault S, Whiting SJ, Peltier SL, Sadrin S, Gerber G, Maixent JM (2013) Vitamin D metabolism, functions and needs: from science to health claims. Eur J Nutr 52:429–441

111. Cantorna MT (2012) Vitamin D, multiple sclerosis and inflammatory bowel disease. Arch Biochem Biophys 523:103–106

112. Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, Croitoru K, Moreno-Hagelsieb G, Neufeld JD (2011) Bacterial biogeography of the human digestive tract. Sci Rep 1:170

113. National Research Council (2006) Preventing the Forward Contamination of

Mars. doi: 10.17226/11381

114. (1999) NASA Policy Directive (NPD) 8020.7G. In: NASA Online Dir. Inf. Syst. Libr. http://nodis3.gsfc.nasa.gov/displayDir.cfm?Internal_ID=N_PD_8020_007G_&page_name=main&search_term=8020.7.

115. Vaishampayan P a., Rabbow E, Horneck G, Venkateswaran KJ (2012) Spores for a Prolonged Period of Time in Real Space Conditions. Astrobiology 12:487–497

116. La Duc MT, Dekas A, Osman S, Moissl C, Newcombe D, Venkateswaran K (2007) Isolation and Characterization of Bacteria Capable of Tolerating the Extreme Conditions of Clean Room Environments. Appl Environ Microbiol 73:2600–2611

117. Venkateswaran K, Kempf M, Chen F, Satomi M, Nicholson W, Kern R (2003) Bacillus nealsonii sp. nov., isolated from a spacecraft-assembly facility, whose spores are gamma-radiation resistant. Int J Syst Evol Microbiol 53:165–72

118. La Duc M, Kern R, Venkateswaran K, La Duc MT (2004) Microbial monitoring of spacecraft and associated environments. Microb Ecol 47:150–8

119. Probst AJ, Auerbach AK, Moissl-Eichinger C (2013) Archaea on Human Skin. PLoS One 8:e65388

120. Vaishampayan P, Osman S, Andersen G, Venkateswaran K (2010) High-density 16S microarray and clone library-based microbial community composition of the Phoenix spacecraft assembly clean room. Astrobiology 10:499–508

121. Derecho I, McCoy KB, Vaishampayan P, Venkateswaran K, Mogul R (2014) Characterization of Hydrogen Peroxide-Resistant Acinetobacter Species Isolated during the Mars Phoenix Spacecraft Assembly. Astrobiology 14:837–47

122. Ghosh S, Zhang P, Li YQ, Setlow P (2009) Superdormant spores of Bacillus

species have elevated wet-heat resistance and temperature requirements for heat activation. J Bacteriol 191:5584–5591

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

124. Checinska A, Probst AJ, Vaishampayan P, et al (2015) Microbiomes of the dust particles collected from the International Space Station and Spacecraft Assembly Facilities. Microbiome 3:50

125. Bashir M, Prietl B, Tauschmann M, Mautner SI, Kump PK, Treiber G, Wurm P, Gorkiewicz G, Högenauer C, Pieber TR (2015) Effects of high doses of vitamin D3 on mucosa-associated gut microbiome vary between regions of the human gastrointestinal tract. Eur J Nutr. doi: 10.1007/s00394-015-0966-2

126. Stadlbauer V, Leber B, Lemesch S, et al (2015) Lactobacillus casei Shirota Supplementation Does Not Restore Gut Microbiota Composition and Gut Barrier in Metabolic Syndrome: A Randomized Pilot Study. PLoS One 10:e0141399

127. Afshinnekoo E, Meydan C, Levy S, Mason CE (2015) Geospatial Resolution of Human and Bacterial Diversity with City-Scale Metagenomics. Cell Syst 1:1–15

128. Tringe SG, Zhang T, Liu X, et al (2008) The airborne metagenome in an indoor urban environment. PLoS One 3:e1862

129. Tauschmann M, Prietl B, Treiber G, Gorkiewicz G, Kump P, Högenauer C, Pieber TR (2013) Distribution of CD4(pos) -, CD8(pos) - and regulatory T cells in the upper and lower gastrointestinal tract in healthy young subjects. PLoS One 8:e80362

130. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–596

131. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO

(2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188–7196

132. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–5267

133. Caporaso JG, Kuczynski J, Stombaugh J, et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336

134. Dean FB, Hosono S, Fang L, et al (2002) Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A 99:5261–5266

135. Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D, Clingenpeel S, Malmstrom R, Stepanauskas R, Cheng JF (2011) Decontamination of MDA reagents for single cell whole genome amplification. PLoS One 6:2–6

136. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

137. Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620

138. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–4

139. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12

140. Bushnell B (2014) BBMap short read aligner, and other bioinformatics tools.

141. Ondov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12:385

142. Caporaso JG, Kuczynski J, Stombaugh J, et al (2011) QIIME allows analysis of high-throughput community sequencing data. 7:335–336

143. Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T (2007) MvirDB - A microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res 35:391–394

144. Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J (2012) Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biol 13:R122

145. Rabosky DL (2014) Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One 9:e89543

146. Imelfort M, Parks D, Woodcroft BJ, Dennis P, Hugenholtz P, Tyson GW (2014) GroopM: an automated tool for the recovery of population genomes from related metagenomes. PeerJ 2:e603

147. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM : assessing the quality of microbial genomes recovered from isolates , single cells , and metagenomes. Genome Res 25:1043–1055

148. Das B, Bischerour J, Barre F-X (2011) Molecular mechanism of acquisition of the cholera toxin genes. Indian J Med Res 133:195–200

149. Ooi J, Li Y, Rogers C, Cantorna M (2013) Vitamin D Regulates the Gut Microbiome and Protects Mice from Dextran Sodium Sulfate–Induced Colitis. J Nutr 143:1679–1686

150. Chung H, Pamp SJ, Hill JA, et al (2012) Gut immune maturation depends on colonization with a host-specific microbiota. Cell 149:1578–93

151. Clarke SF, Murphy EF, O'Sullivan O, et al (2014) Exercise and associated dietary extremes impact on gut microbial diversity. Gut 0:1–8

152. Carroll IM, Ringel-Kulka T, Keku TO, Chang Y-H, Packey CD, Sartor RB, Ringel Y (2011) Molecular analysis of the luminal- and mucosal-associated intestinal microbiota in diarrhea-predominant irritable bowel syndrome. Am J Physiol Gastrointest Liver Physiol 301:G799–807

153. White ANJ, Ng V, Spain CV, Johnson CC, Kinlin LM, Fisman DN (2009) Let the sun shine in: effects of ultraviolet radiation on invasive pneumococcal disease

risk in Philadelphia, Pennsylvania. BMC Infect Dis 9:196

154. Wu GD, Chen J, Hoffmann C, et al (2011) Linking long-term dietary patterns with gut microbial enterotypes. Science 334:105–8

155. Rajkumar H, Mahmood N, Kumar M, Varikuti SR, Challa HR, Myakala SP (2014) Effect of probiotic (VSL#3) and omega-3 on lipid profile, insulin sensitivity, inflammatory markers, and gut colonization in overweight adults: a randomized, controlled trial. Mediators Inflamm 2014:348959

156. Urwin HJ, Miles EA, Noakes PS, Kremmyda L-S, Vlachava M, Diaper ND, Godfrey KM, Calder PC, Vulevic J, Yaqoob P (2014) Effect of salmon consumption during pregnancy on maternal and infant faecal microbiota, secretory IgA and calprotectin. Br J Nutr 111:773–84

157. Huttenhower C, Gevers D, Knight R, et al (2012) Structure, function and diversity of the healthy human microbiome. Nature 486:207–214

158. Hu J, Nomura Y, Bashir A, Fernandez-Hernandez H, Itzkowitz S, Pei Z, Stone J, Loudon H, Peter I (2013) Diversified microbiota of meconium is affected by maternal diabetes status. PLoS One 8:e78257

159. De Angelis M, Piccolo M, Vannini L, Siragusa S, De Giacomo A, Serrazzanetti DI, Cristofori F, Guerzoni ME, Gobbetti M, Francavilla R (2013) Fecal Microbiota and Metabolome of Children with Autism and Pervasive Developmental Disorder Not Otherwise Specified. PLoS One 8:e76993

160. Davenport M, Poles J, Leung JM, Wolff MJ, Abidi WM, Ullman T, Mayer L, Cho I, Loke P (2014) Metabolic alterations to the mucosal microbiota in inflammatory bowel disease. Inflamm Bowel Dis 20:723–31

161. Wong VW-S, Tse C-H, Lam TT-Y, et al (2013) Molecular characterization of the fecal microbiota in patients with nonalcoholic steatohepatitis--a longitudinal study. PLoS One 8:e62885

162. Bahl MI, Bergström A, Licht TR (2012) Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by

downstream quantitative PCR analysis. FEMS Microbiol Lett 329:193–197

163. Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, Ma ZS, Shi P (2014) Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. ISME J 8:881–93

164. Davis GR, Santa Ana CA, Morawski SG, Fordtran JS (1982) Permeability characteristics of human jejunum, ileum, proximal colon and distal colon: results of potential difference measurements and unidirectional fluxes. Gastroenterology 83:844–50

165. Hwang YG, Hsu H, Lim FC, Wu Q, Yang P, Fisher G, Hunter GR, Mountz JD (2013) Increased vitamin D is associated with decline of naïve, but accumulation of effector, CD8 T cells during early aging. Adv Aging Res 2:72–80

166. Borella E, Nesher G, Israeli E, Shoenfeld Y (2014) Vitamin D: a new anti-infective agent? Ann N Y Acad Sci 1317:76–83

167. Bartels LE, Bendix M, Hvas CL, Jørgensen SP, Agnholt J, Agger R, Dahlerup JF (2014) Oral vitamin D3 supplementation reduces monocyte-derived dendritic cell maturation and cytokine production in Crohn's disease patients. Inflammopharmacology 22:95–103

168. Ooi JH, McDaniel KL, Weaver V, Cantorna MT (2014) Murine CD8+ T cells but not macrophages express the vitamin D 1α-hydroxylase. J Nutr Biochem 25:58–65

169. Sigmundsdottir H, Pan J, Debes GF, Alt C, Habtezion A, Soler D, Butcher EC (2007) DCs metabolize sunlight-induced vitamin D3 to "program" T cell attraction to the epidermal chemokine CCL27. Nat Immunol 8:285–293

170. Willheim M, Thien R, Schrattbauer K, et al (1999) Regulatory effects of 1alpha,25-dihydroxyvitamin D3 on the cytokine production of human peripheral blood lymphocytes. J Clin Endocrinol Metab 84:3739–3744

171. Kreutz M, Andreesen R, Krause SW, Szabo A, Ritz E, Reichel H (1993) 1,25-

dihydroxyvitamin D3 production and vitamin D3 receptor expression are developmentally regulated during differentiation of human monocytes into macrophages. Blood 82:1300–1307

172. Fritsche J, Mondal K, Ehrnsperger A, Andreesen R, Kreutz M (2003) Regulation of 25-hydroxyvitamin D3-1 alpha-hydroxylase and production of 1 alpha,25-dihydroxyvitamin D3 by human dendritic cells. Blood 102:3314–3316

173. Murphy K (2011) T cell-mediated cytotoxicity. In: Janeway CA, Travers JP, Walport M, Shlomchik MJ (eds) Janeway's Immunobiol., 8th ed. Garland Science, New York, p 888

174. Stecher B, Robbiani R, Walker AW, et al (2007) Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. PLoS Biol 5:2177–89

175. Pappa H, Gordon C, Saslowsky T (2006) Vitamin D status in children and young adults with inflammatory bowel disease. Pediatrics 118:1950–1961

176. Jellbauer S, Raffatellu M (2014) An intestinal arsonist: pathobiont ignites IBD and flees the scene. Gut 63:1034–5

177. Mukhopadhya I, Hansen R, El-Omar EM, Hold GL (2012) IBD-what role do Proteobacteria play? Nat Rev Gastroenterol Hepatol 9:219–30

178. Frank DN, Amand ALS, Feldman RA, Boedeker EC, Harpaz N, Pace NR, St Amand AL (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A 104:13780–13785

179. Jørgensen SP, Hvas CL, Agnholt J, Christensen LA, Heickendorff L, Dahlerup JF (2013) Active Crohn's disease is associated with low vitamin D levels. J Crohns Colitis 7:e407–13

180. Jørgensen SP, Agnholt J, Glerup H, Lyhne S, Villadsen GE, Hvas CL, Bartels LE, Kelsen J, Christensen L a, Dahlerup JF (2010) Clinical trial: vitamin D3 treatment in Crohn's disease - a randomized double-blind placebo-controlled

study. Aliment Pharmacol Ther 32:377–83

181. Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z, Järnerot G, Tysk C, Jansson JK, Engstrand L (2010) A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. Gastroenterology 139:1844–1854.e1

182. Hosoda K, Shimomura H, Wanibuchi K, Masui H, Amgalanbaatar A, Hayashi S, Takahashi T, Hirai Y (2015) Identification and characterization of a vitamin D_3 decomposition product bactericidal against Helicobacter pylori. Sci Rep 5:8860

183. Guo L, Chen W, Zhu H, Chen Y, Wan X, Yang N, Xu S, Yu C, Chen L (2014) Helicobacter pylori Induces Increased Expression of the Vitamin D Receptor in Immune Responses. Helicobacter 19:37–47

184. Ross AC, Manson JE, Abrams SA, et al (2011) The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. J Clin Endocrinol Metab 96:53–58

185. Prietl B, Treiber G, Mader JK, Hoeller E, Wolf M, Pilz S, Graninger WB, Obermayer-Pietsch BM, Pieber TR (2013) High-dose cholecalciferol supplementation significantly increases peripheral $CD4^+$ Tregs in healthy adults without negatively affecting the frequency of other immune cells. Eur J Nutr 53:751–759

186. Ghosh S, Osman S, Vaishampayan P, Venkateswaran K (2010) Recurrent isolation of extremotolerant bacteria from the clean room where Phoenix spacecraft components were assembled. Astrobiology 10:325–35

187. Moissl C, Osman S, La Duc MT, Dekas A, Brodie E, DeSantis T, Desantis T, Venkateswaran K (2007) Molecular bacterial community analysis of clean rooms where spacecraft are assembled. FEMS Microbiol Ecol 61:509–21

188. Schmidt H, Hensel M (2004) Pathogenicity Islands in Bacterial Pathogenesis Pathogenicity Islands in Bacterial Pathogenesis. Clin Microbiol Rev 17:14–56

189. Rathinavelu S, Zavros Y, Merchant JL (2003) Acinetobacter lwoffii infection and

gastritis. Microbes Infect 5:651–657

190. Firstenberg-Eden R, Rowley DB, Shattuck GE (1980) Factors affecting inactivation of Moraxella-Acinetobacter cells in an irradiation process. Appl Environ Microbiol 40:480–5

191. Firstenberg-Eden R, Rowley DB, Shattuck E (1980) Thermal inactivation and injury of Moraxella-Acinetobacter cells in ground beef. Appl Environ Microbiol 39:159–64

192. Yeom J, Shin J-H, Yang J-Y, Kim J, Hwang G-S (2013) (1)H NMR-based metabolite profiling of planktonic and biofilm cells in Acinetobacter baumannii 1656-2. PLoS One 8:e57730

193. Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B, Bartlett J (2009) Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America. Clin Infect Dis 48:1–12

194. Rice LB (2008) Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens : No ESKAPE. 197:1079–1081

195. Wieczorek P, Sacha P, Hauschild T, Zórawski M, Krawczyk M, Tryniszewska E (2008) Multidrug resistant Acinetobacter baumannii--the role of AdeABC (RND family) efflux pump in resistance to antibiotics. Folia Histochem Cytobiol 46:257–67

196. Shevchuk O, Jäger J, Steinert M (2011) Virulence properties of the legionella pneumophila cell envelope. Front Microbiol 2:74

197. Hales LM, Shuman HA (1999) Legionella pneumophila contains a type II general secretion pathway required for growth in amoebae as well as for secretion of the Msp protease. Infect Immun 67:3662–6

198. Kaesbohrer a., Schroeter a., Tenhagen B a., Alt K, Guerra B, Appel B (2012) Emerging antimicrobial resistance in commensal escherichia coli with public health relevance. Zoonoses Public Health 59:158–165

199. Tadesse D a., Zhao S, Tong E, Ayers S, Singh A, Bartholomew MJ, McDermott

PF (2012) Antimicrobial drug resistance in Escherichia coli from humans and food animals, United States, 1950-2002. Emerg Infect Dis 18:741–749

200. Wasyl D, Hoszowski A, Zając M, Szulowski K (2013) Antimicrobial resistance in commensal Escherichia coli isolated from animals at slaughter. Front Microbiol 4:221

201. Yoon E-J, Courvalin P, Grillot-Courvalin C (2013) RND-type efflux pumps in multidrug-resistant clinical isolates of Acinetobacter baumannii: major role for AdeABC overexpression and AdeRS mutations. Antimicrob Agents Chemother 57:2989–95

202. Direito SOL, Zaura E, Little M, Ehrenfreund P, Röling WFM (2014) Systematic evaluation of bias in microbial community profiles induced by whole genome amplification. Environ Microbiol 16:643–657

203. Höppe P, Martinac I (1998) Indoor climate and air quality. Review of current and future topics in the field of ISB study group 10. Int J Biometeorol 42:1–7

204. Lax S, Smith DP, Hampton-Marcell J, et al (2014) Longitudinal analysis of microbial interaction between humans and the indoor environment. Science 345:1048–52

205. Craw P, Mackay RE, Naveenathayalan A, Hudson C, Branavan M, Sadiq ST, Balachandran W (2015) A Simple, Low-Cost Platform for Real-Time Isothermal Nucleic Acid Amplification. Sensors (Basel) 15:23418–30

6 Appendix

Appendix 1: Mean relative abundance of genera in investigated gastrointestinal regions. Adapted from Bashir et al. 2015 [125]

Taxon	GC (%)	GA (%)	DD (%)	TI (%)	AO (%)	AC (%)	SC (%)	stool(%)
Acidobacteria;Acidobacteria Gp16	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Acidobacteria;Acidobacteria Gp2	0.02	0.04	0.00	0.00	0.00	0.00	0.01	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Actinomycetaceae;Actinomyces	0.93	1.44	1.41	0.04	0.00	0.02	0.01	0.01
Actinobacteria;Actinobacteria;Actinomycetales;Actinomycetaceae;Mobiluncus	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Actinomycetaceae;unclassified	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Cellulomonadaceae;Tropheryma	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Corynebacteriaceae;Corynebacterium	0.04	0.02	0.12	0.00	0.00	0.01	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Corynebacteriaceae;unclassified	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Dermabacteraceae;Brachybacterium	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Dermacoccaceae;Dermacoccus	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Dietziaceae;Dietzia	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Intrasporangiaceae;Janibacter	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Microbacteriaceae;Amnibacterium	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Microbacteriaceae;Curtobacterium	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Microbacteriaceae;Leifsonia	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Microbacteriaceae;Leucobacter	0.14	0.15	0.13	0.00	0.00	0.00	0.01	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Microbacteriaceae;Microbacterium	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Microbacteriaceae;Zimmermannella	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Microbacteriaceae;unclassified	0.07	0.01	0.09	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Micrococcaceae;Kocuria	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Micrococcaceae;Micrococcus	0.07	0.00	0.03	0.00	0.00	0.01	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Micrococcaceae;Rothia	0.43	0.30	0.67	0.01	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Micrococcaceae;unclassified	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Mycobacteriaceae;Mycobacterium	0.01	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Nakamurellaceae;Humicoccus	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Nocardiaceae;Rhodococcus	0.18	0.12	0.37	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Nocardioidaceae;Nocardioides	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Nocardioidaceae;unclassified	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Actinobacteria;Actinobacteria;Actinomycetales;Propionibacteriaceae;Microlunatus	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Propionibacteriaceae;Propionibacterium	1.06	0.74	1.27	0.01	0.03	0.01	0.02	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Propionibacteriaceae;unclassified	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Pseudonocardiaceae;Pseudonocardia	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;Streptomycetaceae;Streptomyces	0.03	0.00	0.01	0.00	0.01	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Actinomycetales;unclassified;unclassified	0.02	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Bifidobacteriales;Bifidobacteriaceae;Bifidobacterium	0.03	0.00	0.00	0.01	0.00	0.00	0.00	0.01
Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Asaccharobacter	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01
Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Atopobium	0.65	0.49	0.53	0.06	0.00	0.01	0.00	0.01
Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Collinsella	0.00	0.03	0.00	0.06	0.04	0.04	0.06	0.01
Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Coriobacterium	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Eggerthella	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02
Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Olsenella	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.00
Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;unclassified	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00
Actinobacteria;Actinobacteria;unclassified;unclassified;unclassified	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Ohtaekwangia	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides	11.47	12.39	15.42	46.63	47.19	46.89	45.81	47.09
Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella	0.07	0.01	0.07	0.45	0.73	0.56	0.65	0.63
Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Butyricimonas	0.00	0.00	0.01	0.10	0.16	0.05	0.10	0.06
Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter	0.06	0.01	0.01	0.26	0.23	0.15	0.35	0.17
Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter	0.00	0.03	0.04	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Parabacteroides	0.16	0.35	0.18	1.54	1.95	1.51	1.61	1.85
Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Porphyromonas	1.14	0.75	0.72	0.00	0.01	0.00	0.02	0.00
Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Tannerella	0.04	0.05	0.03	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;unclassified	0.00	0.00	0.05	0.13	0.11	0.10	0.05	0.05
Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Hallella	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Paraprevotella	0.00	0.01	0.01	0.06	0.08	0.12	0.14	0.17
Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella	19.60	19.45	12.96	4.34	3.75	4.58	5.55	9.62
Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;unclassified	0.15	0.35	0.07	0.20	0.13	0.23	0.24	0.12
Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes	0.00	0.06	0.16	0.53	0.59	0.54	0.76	1.13
Bacteroidetes;Bacteroidia;Bacteroidales;unclassified;unclassified	0.50	0.87	0.56	0.94	1.01	0.83	1.06	0.24
Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae;Capnocytophaga	0.10	0.08	0.04	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae;Chryseobacterium	0.07	0.18	0.21	0.00	0.00	0.01	0.00	0.00
Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae;Cloacibacterium	0.24	0.12	0.05	0.00	0.00	0.00	0.00	0.00

Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae;Elizabethkingia	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae;Flavobacterium	0.06	0.04	0.18	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae;unclassified	0.18	0.16	0.10	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Sphingobacteria;Sphingobacteriales;Chitinophagaceae;unclassified	0.01	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Dyadobacter	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;Hymenobacter	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Sphingobacteria;Sphingobacteriales;Cytophagaceae;unclassified	0.04	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Sphingobacteria;Sphingobacteriales;Sphingobacteriaceae;Pedobacter	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;Sphingobacteria;Sphingobacteriales;Sphingobacteriaceae;Sphingobacterium	0.08	0.03	0.01	0.00	0.00	0.00	0.00	0.00
Bacteroidetes;unclassified;unclassified;unclassified;unclassified	1.55	2.74	1.08	1.90	1.97	1.83	1.12	0.49
Deinococcus-Thermus;Deinococci;Deinococcales;Deinococcaceae;Deinococcus	0.00	0.01	0.02	0.00	0.01	0.00	0.00	0.00
Deinococcus-Thermus;Deinococci;Thermales;Thermaceae;Thermus	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Bacillales;Bacillaceae_1;Anoxybacillus	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.00
Firmicutes;Bacilli;Bacillales;Bacillaceae_1;Bacillus	0.05	0.03	0.03	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Bacillales;Bacillaceae_1;Geobacillus	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Bacillales;Bacillales_Incertae_Sedis_XI;Gemella	1.41	1.51	1.93	0.14	0.02	0.01	0.01	0.00
Firmicutes;Bacilli;Bacillales;Paenibacillaceae_1;Paenibacillus	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Bacillales;Planococcaceae;Planococcaceae_incertae_sedis	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Bacillales;Staphylococcaceae;Macrococcus	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Bacillales;Staphylococcaceae;Staphylococcus	0.20	0.14	0.45	0.00	0.00	0.01	0.00	0.00
Firmicutes;Bacilli;Bacillales;Thermoactinomycetaceae_1;Thermoactinomyces	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Lactobacillales;Aerococcaceae;Abiotrophia	0.02	0.03	0.08	0.00	0.00	0.00	0.01	0.00
Firmicutes;Bacilli;Lactobacillales;Carnobacteriaceae;Alkalibacterium	0.77	0.59	0.80	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Lactobacillales;Carnobacteriaceae;Carnobacterium	0.01	0.04	0.40	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Lactobacillales;Carnobacteriaceae;Dolosigranulum	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Lactobacillales;Carnobacteriaceae;Granulicatella	1.76	1.51	2.00	0.10	0.04	0.01	0.02	0.00
Firmicutes;Bacilli;Lactobacillales;Carnobacteriaceae;unclassified	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Lactobacillales;Enterococcaceae;Enterococcus	0.04	0.03	0.03	0.00	0.00	0.00	0.00	0.01
Firmicutes;Bacilli;Lactobacillales;Enterococcaceae;Vagococcus	0.00	0.07	0.01	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus	0.13	0.11	0.17	0.00	0.00	0.00	0.02	0.00
Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Pediococcus	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Lactococcus	0.30	0.13	0.35	0.00	0.00	0.00	0.00	0.01
Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus	18.79	15.20	17.77	0.79	0.17	0.14	0.16	0.29
Firmicutes;Clostridia;Clostridiales;Clostridiaceae_1;Clostridium_sensu_stricto	0.06	0.05	0.05	0.00	0.01	0.00	0.03	0.27

Firmicutes;Clostridia;Clostridiales;Clostridiaceae_1;unclassified	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.03
Firmicutes;Clostridia;Clostridiales;Clostridiales_Incertae_Sedis_XI;Finegoldia	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Firmicutes;Clostridia;Clostridiales;Clostridiales_Incertae_Sedis_XI;Parvimonas	0.00	0.04	0.05	0.00	0.00	0.00	0.01	0.00
Firmicutes;Clostridia;Clostridiales;Clostridiales_Incertae_Sedis_XI;Peptoniphilus	0.05	0.02	0.02	0.00	0.00	0.00	0.00	0.00
Firmicutes;Clostridia;Clostridiales;Clostridiales_Incertae_Sedis_XIII;Anaerovorax	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00
Firmicutes;Clostridia;Clostridiales;Clostridiales_Incertae_Sedis_XIII;Mogibacterium	0.14	0.05	0.10	0.01	0.00	0.00	0.00	0.00
Firmicutes;Clostridia;Clostridiales;Clostridiales_Incertae_Sedis_XIII;unclassified	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.01
Firmicutes;Clostridia;Clostridiales;Eubacteriaceae;Anaerofustis	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Firmicutes;Clostridia;Clostridiales;Eubacteriaceae;Eubacterium	0.46	0.50	0.66	0.01	0.01	0.00	0.00	0.00
Firmicutes;Clostridia;Clostridiales;Eubacteriaceae;unclassified	0.08	0.05	0.04	0.00	0.00	0.00	0.00	0.00
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Anaerostipes	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Blautia	1.19	0.50	0.34	3.98	3.69	4.05	4.21	1.10
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Butyrivibrio	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Catonella	0.05	0.06	0.04	0.00	0.00	0.00	0.00	0.00
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Clostridium_XlVa	0.08	0.04	0.09	0.69	0.53	0.48	0.64	0.25
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Clostridium_XlVb	0.00	0.03	0.00	0.04	0.01	0.05	0.04	0.05
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Coprococcus	0.05	0.26	0.08	1.07	1.06	1.06	0.98	0.29
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Dorea	0.06	0.16	0.04	0.19	0.14	0.18	0.18	0.14
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiracea_incertae_sedis	0.62	0.86	1.06	7.58	6.95	8.83	7.33	5.50
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Moryella	0.18	0.12	0.09	0.01	0.00	0.00	0.00	0.00
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Oribacterium	0.91	0.70	0.79	0.02	0.00	0.00	0.00	0.00
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Roseburia	0.09	0.51	0.30	0.87	1.26	0.98	0.83	0.82
Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;unclassified	1.47	3.01	2.72	10.02	10.90	10.35	11.37	6.96
Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae;Clostridium_XI	0.05	0.07	0.08	0.35	0.39	0.34	0.29	1.23
Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae;Filifactor	0.07	0.08	0.02	0.00	0.00	0.00	0.00	0.00
Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae;Peptostreptococcus	0.14	0.17	0.19	0.04	0.00	0.01	0.01	0.00
Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae;unclassified	0.00	0.01	0.00	0.02	0.00	0.01	0.00	0.09
Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Butyricicoccus	0.00	0.00	0.00	0.06	0.07	0.14	0.07	0.06
Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Clostridium_IV	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.03
Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Faecalibacterium	0.34	1.18	1.26	6.44	7.12	6.92	6.29	5.78
Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Flavonifractor	0.00	0.00	0.00	0.06	0.06	0.07	0.04	0.18
Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillibacter	0.01	0.10	0.03	0.21	0.12	0.17	0.16	0.50
Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Pseudoflavonifractor	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.07
Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ruminococcus	0.01	0.00	0.00	0.01	0.01	0.00	0.03	0.07

Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Subdoligranulum	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.03
Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;unclassified	0.03	0.22	0.09	0.42	0.60	0.39	0.61	3.11
Firmicutes;Clostridia;Clostridiales;unclassified;unclassified	0.35	0.26	0.30	0.52	0.49	0.48	0.54	1.07
Firmicutes;Clostridia;unclassified;unclassified;unclassified	0.00	0.01	0.03	0.01	0.07	0.02	0.00	0.03
Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Bulleidia	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Catenibacterium	0.00	0.03	0.00	0.01	0.02	0.03	0.02	0.05
Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Clostridium_XVIII	0.03	0.04	0.01	0.55	0.62	0.69	0.40	0.07
Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Coprobacillus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Erysipelothrix	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Erysipelotrichaceae_incertae_sedis	0.01	0.30	0.02	1.57	1.41	1.31	1.16	1.04
Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Holdemania	0.00	0.00	0.03	0.10	0.03	0.03	0.03	0.10
Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Solobacterium	0.61	0.46	0.48	0.07	0.01	0.01	0.01	0.00
Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Turicibacter	0.00	0.00	0.03	0.07	0.01	0.01	0.08	0.03
Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;unclassified	0.08	0.20	0.40	0.69	0.73	0.73	0.58	0.82
Firmicutes;Negativicutes;Selenomonadales;Acidaminococcaceae;Acidaminococcus	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Firmicutes;Negativicutes;Selenomonadales;Acidaminococcaceae;Phascolarctobacterium	0.01	0.10	0.04	0.28	0.40	0.33	0.41	0.77
Firmicutes;Negativicutes;Selenomonadales;Acidaminococcaceae;unclassified	0.05	0.10	0.10	0.01	0.00	0.00	0.00	0.00
Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Allisonella	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01
Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Dialister	0.10	0.13	0.10	0.13	0.24	0.20	0.20	1.48
Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Megamonas	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Megasphaera	0.50	0.40	0.42	0.04	0.00	0.00	0.00	0.01
Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Mitsuokella	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Selenomonas	0.16	0.08	0.15	0.00	0.00	0.00	0.00	0.00
Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Veillonella	5.52	5.15	3.97	0.26	0.22	0.14	0.18	0.01
Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Veillonellaceae_genus_incertae_sedis	0.01	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Zymophilus	0.01	0.01	0.02	0.00	0.00	0.00	0.00	0.00
Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;unclassified	0.14	0.24	0.03	0.20	0.15	0.11	0.10	0.10
Firmicutes;unclassified;unclassified;unclassified;unclassified	0.11	0.05	0.03	0.25	0.37	0.27	0.25	0.15
Fusobacteria;Fusobacteria;Fusobacteriales;Fusobacteriaceae;Fusobacterium	2.58	2.31	2.31	0.16	0.05	0.03	0.09	0.01
Fusobacteria;Fusobacteria;Fusobacteriales;Leptotrichiaceae;Leptotrichia	1.45	1.35	0.68	0.00	0.00	0.00	0.01	0.00
Fusobacteria;Fusobacteria;Fusobacteriales;Leptotrichiaceae;Sneathia	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fusobacteria;Fusobacteria;Fusobacteriales;Leptotrichiaceae;Streptobacillus	0.05	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Fusobacteria;Fusobacteria;Fusobacteriales;unclassified;unclassified	0.24	0.29	0.26	0.00	0.00	0.00	0.00	0.00

Proteobacteria;Alphaproteobacteria;Caulobacterales;Caulobacteraceae;Brevundimonas	0.07	0.01	0.05	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Caulobacterales;Caulobacteraceae;Caulobacter	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;Aurantimonadaceae;Aurantimonas	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;Beijerinckiaceae;Chelatococcus	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium	0.10	0.00	0.02	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;unclassified	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;Brucellaceae;Daeguia	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;Brucellaceae;Ochrobactrum	0.00	0.02	0.05	0.00	0.00	0.01	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Devosia	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium	0.02	0.11	0.05	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium	0.03	0.00	0.02	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhizobiales;unclassified;unclassified	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Haematobacter	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Paracoccus	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;unclassified	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Novosphingobium	0.00	0.04	0.29	0.00	0.00	0.00	0.01	0.00
Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Sphingobium	0.03	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Sphingomonas	0.04	0.01	0.28	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;unclassified	0.06	0.01	0.02	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;Sphingomonadales;unclassified;unclassified	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Alphaproteobacteria;unclassified;unclassified;unclassified	0.00	0.00	0.07	0.00	0.00	0.00	0.01	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Alcaligenaceae;Achromobacter	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Ralstonia	0.87	0.82	1.03	0.00	0.00	0.00	0.01	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiales_incertae_sedis;Aquabacterium	0.00	0.02	0.01	0.00	0.00	0.00	0.01	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiales_incertae_sedis;Ideonella	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiales_incertae_sedis;Tepidimonas	0.15	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiales_incertae_sedis;unclassified	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Acidovorax	0.03	0.04	0.04	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Comamonas	0.07	0.08	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Delftia	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Diaphorobacter	0.03	0.00	0.08	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Pelomonas	0.06	0.12	0.16	0.00	0.01	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Schlegelella	0.06	0.12	0.10	0.00	0.00	0.00	0.00	0.00

Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Variovorax	0.20	0.12	0.16	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;unclassified	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Herbaspirillum	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Janthinobacterium	0.16	0.16	0.00	0.00	0.01	0.01	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Massilia	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae;Undibacterium	0.04	0.04	0.10	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Burkholderiales;Sutterellaceae;Parasutterella	0.00	0.00	0.02	0.12	0.18	0.10	0.18	0.14
Proteobacteria;Betaproteobacteria;Burkholderiales;Sutterellaceae;Sutterella	0.05	0.08	0.06	0.98	0.70	0.45	0.86	0.28
Proteobacteria;Betaproteobacteria;Burkholderiales;Sutterellaceae;unclassified	0.00	0.00	0.00	0.06	0.13	0.16	0.10	0.12
Proteobacteria;Betaproteobacteria;Burkholderiales;unclassified;unclassified	0.03	0.01	0.01	0.04	0.01	0.05	0.07	0.03
Proteobacteria;Betaproteobacteria;Hydrogenophilales;Hydrogenophilaceae;Hydrogenophilus	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Hydrogenophilales;Hydrogenophilaceae;Petrobacter	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Hydrogenophilales;Hydrogenophilaceae;Thiobacillus	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae;Aquitalea	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae;Eikenella	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae;Kingella	0.12	0.00	0.02	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae;Neisseria	0.50	0.46	0.37	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae;Simonsiella	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae;unclassified	0.07	0.07	0.17	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Azospira	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Quatrionicoccus	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Betaproteobacteria;unclassified;unclassified;unclassified	0.00	0.03	0.00	0.11	0.16	0.12	0.16	0.12
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Bilophila	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio	0.00	0.04	0.00	0.01	0.01	0.01	0.00	0.01
Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;unclassified	0.00	0.01	0.00	0.01	0.02	0.01	0.01	0.01
Proteobacteria;Deltaproteobacteria;unclassified;unclassified;unclassified	0.00	0.00	0.02	0.02	0.01	0.01	0.05	0.01
Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae;Arcobacter	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae;Campylobacter	0.31	0.29	0.21	0.01	0.00	0.01	0.00	0.00
Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae;Sulfurospirillum	0.03	0.10	0.03	0.00	0.00	0.00	0.01	0.00
Proteobacteria;Epsilonproteobacteria;Campylobacterales;Helicobacteraceae;Helicobacter	0.48	0.07	0.16	0.00	0.02	0.02	0.01	0.03
Proteobacteria;Gammaproteobacteria;Aeromonadales;Aeromonadaceae;unclassified	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Aeromonadales;Succinivibrionaceae;Succinivibrio	0.00	0.07	0.00	0.04	0.03	0.00	0.08	0.28
Proteobacteria;Gammaproteobacteria;Alteromonadales;Alteromonadaceae;Alishewanella	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Cardiobacteriales;Cardiobacteriaceae;Cardiobacterium	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00

Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Buttiauxella	0.03	0.01	0.03	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Citrobacter	0.00	0.00	0.00	0.01	0.02	0.02	0.04	0.10
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Enterobacter	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Escherichia_Shigella	1.11	1.38	1.48	0.93	0.72	0.66	0.84	1.28
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Klebsiella	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Kluyvera	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Morganella	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Serratia	0.03	0.12	0.08	0.01	0.00	0.00	0.01	0.00
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Yersinia	0.01	0.03	0.02	0.00	0.00	0.00	0.01	0.00
Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;unclassified	2.02	2.72	3.11	0.10	0.11	0.10	0.12	0.23
Proteobacteria;Gammaproteobacteria;Oceanospirillales;Halomonadaceae;Chromohalobacter	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Oceanospirillales;Halomonadaceae;Halomonas	0.03	0.07	0.03	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Pasteurellales;Pasteurellaceae;Actinobacillus	0.45	0.12	0.12	0.06	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Pasteurellales;Pasteurellaceae;Aggregatibacter	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Pasteurellales;Pasteurellaceae;Haemophilus	0.48	0.33	0.31	0.10	0.07	0.07	0.08	0.00
Proteobacteria;Gammaproteobacteria;Pasteurellales;Pasteurellaceae;unclassified	0.50	0.46	0.67	0.11	0.06	0.06	0.07	0.00
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Moraxellaceae;Acinetobacter	0.04	0.08	0.07	0.00	0.00	0.00	0.00	0.05
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Moraxellaceae;Alkanindiges	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Moraxellaceae;Enhydrobacter	0.05	0.00	0.02	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Moraxellaceae;Moraxella	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Moraxellaceae;Psychrobacter	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Cellvibrio	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas	8.40	8.84	9.87	0.13	0.07	0.18	0.20	0.00
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Rhizobacter	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;unclassified	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Xanthomonadales;Sinobacteraceae;Nevskia	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Lysobacter	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas	0.09	0.10	0.22	0.00	0.01	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Thermomonas	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;unclassified	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Proteobacteria;Gammaproteobacteria;unclassified;unclassified;unclassified	0.01	0.03	0.01	0.00	0.00	0.00	0.00	0.00
Proteobacteria;unclassified;unclassified;unclassified;unclassified	0.00	0.00	0.00	0.06	0.08	0.07	0.18	0.07
SR1	0.07	0.09	0.10	0.00	0.00	0.00	0.00	0.00
Spirochaetes;Spirochaetes;Spirochaetales;Spirochaetaceae;Treponema	0.20	0.05	0.14	0.00	0.00	0.00	0.00	0.00

TM7	0.57	0.54	0.49	0.02	0.00	0.01	0.01	0.00
Tenericutes;Mollicutes;Mycoplasmatales;Mycoplasmataceae;Mycoplasma	0.05	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales;Verrucomicrobiaceae;Akkermansia	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
Bacteria;unclassified;unclassified;unclassified;unclassified;unclassified	0.67	0.83	0.73	1.66	1.51	1.67	1.82	2.83

Appendix 2: List of clinically relevant pathogens used for pathogen search

Acinetobacter baumannii, Acinetobacter lwoffii, Acinetobacter spp., Actinomycetes, Adenovirus, *Aeromonas spp., Alcaligenes faecalis, Alcaligenes spp./Achromobacter spp., Alcaligenes xylosoxidans*, Arbovirus, *Aspergillus spp.,* Astrovirus*, Bacillus anthracis, Bacillus cereus, Bacteriodes fragilis, Bartonella quintana, Bordetella pertussis, Borrelia burgdorferi, Borrelia recurrentis, Brevundimonas diminuta, Brevundimonas vesicularis, Brucella spp., Burkholderia cepacia, Burkholderia mallei, Burkholderia pseudomallei, Campylobacter jejuni / coli, Candida albicans, Candida krusei, Candida parapsilosis,* Chikungunya virus, CHIKV*, Chlamydia pneumoniae, Chlamydia psittaci, Chlamydia trachomatis, Citrobacter spp., Clostridium botulinum, Clostridium difficile, Clostridium perfringens, Clostridium tetani,* Coronavirus*, Corynebacterium diphtheriae, Corynebacterium pseudotuberculosis, Corynebacterium spp., Corynebacterium ulcerans, Coxiella burnetii, Coxsackievirus,* Crimean-Congo haemorrhagic fever virus*, Cryptococcus neoformans, Cryptosporidium hominis, Cryptosporidium parvum, Cyclospora cayetanensis,* Cytomegalovirus*,* Dengue virus, Ebola virus, Echovirus*, Entamoeba histolytica, Enterobacter aerogenes, Enterobacter cloacae, Enterococcus faecalis, Enterococcus faecium, Enterococcus hirae, Epidermophyton spp.,* Epstein-Barr virus, EBV*, Escherichia coli,* Foot-and-mouth disease virus, FMDV, *Francisella tularensis, Giardia lamblia, Haemophilus influenzae,* Hantavirus*, Helicobacter pylori, Helminths,* Hepatitis A virus, HAV, Hepatitis B virus, HBV, Hepatitis C virus, HCV, Hepatitis D virus, Hepatitis E virus, Herpes simplex virus, HSV, *Histoplasma capsulatum,* Human enterovirus 71, Human herpesvirus 6, HHV-6, Human herpesvirus 7, HHV-7, Human herpesvirus 8, HHV-8, Human immunodeficiency virus, HIV, Human metapneumovirus, Human papillomavirus, Influenza virus, *Klebsiella granulomatis, Klebsiella oxytoca, Klebsiella pneumoniae,* Lassa virus, *Leclercia adecarboxylata, Legionella pneumophila, Leishmania spp.,*

Leptospira interrogans, Leuconostoc pseudomesenteroides, Listeria monocytogenes, Marburg virus, Measles virus*, Micrococcus luteus, Microsporum spp., Molluscipoxvirus, Morganella spp.,* Mumps virus, *Mycobacterium chimaera Myco, Mycobacterium leprae Myco, Mycobacterium tuberculosis, Mycoplasma pneumoniae, Neisseria meningitidis, Neisseria gonorrhoeae,* Norovirus, *Orientia tsutsugamushi, Pantoea agglomerans,* Parainfluenza virus, Parvovirus*, Pediculus humanus capitis, Pediculus humanus corporis, Plasmodium spp., Pneumocystis jiroveci,* Poliovirus, Polyomavirus*, Proteus mirabilis, Proteus vulgaris, Providencia rettgeri, Providencia stuartii, Pseudomonas aeruginosa, Pseudomonas spp.,* Rabies virus*, Ralstonia spp.,* Respiratory syncytial virus, RSV*, Rhinovirus, Rickettsia prowazekii, Rickettsia typhi,* Rotavirus, Rubella virus*, Salmonella enteritidis, Salmonella paratyphi, Salmonella spp., Salmonella typhimurium, Sarcoptes scabiei,* Sapovirus*, Serratia marcescens, Shigella sonnei, Sphingomonas species, Staphylococcus aureus, Staphylococcus capitis, Staphylococcus epidermidis, Staphylococcus haemolyticus, Staphylococcus hominis, Staphylococcus lugdunensis, Staphylococcus saprophyticus, Stenotrophomonas maltophilia, Streptococcus pneumoniae, Streptococcus pyogenes, Streptococcus spp.,* TBE virus*, Toxoplasma gondii, Treponema pallidum, Trichinella spiralis, Trichomonas vaginalis, Trichophyton spp., Trichosporon spp., Trypanosoma brucei gambiense, Trypanosoma brucei rhodesiense, Trypanosoma cruzi,* Vaccinia virus, Varicella zoster virus, Variola virus*, Vibrio cholerae,* West Nile virus, Yellow fever virus*, Yersinia enterocolitica, Yersinia pestis, Yersinia pseudotuberculosis*, Zika virus

Appendix 3: Virulence factors matching the classified pathogens

Virulence factor	PHX-B	PHX-D	PHX-A	DAWN	MSL
114|vfid|115|vsiid|114|ssid|transposase [Escherichia coli]	1	130	5	0	0
13827|vfid|22101|vsiid|42154|ssid|ORF B protein [Escherichia coli]	1	0	0	0	330
13876|vfid|22199|vsiid|42301|ssid|putative reverse transcriptase [Escherichia coli]	1	0	0	0	0
14026|vfid|22499|vsiid|42751|ssid|putative lysil-tR0 synthetase LysU [Escherichia coli]	1	135	0	0	0
16399|vfid|37388|vsiid|58225|ssid|hypothetical protein c5198 [Escherichia coli CFT073]	6	0	0	0	0
16400|vfid|37390|vsiid|58227|ssid|transposase insD [Escherichia coli CFT073]	1	0	0	0	0

16653|vfid|37896|vsiid|58733|ssid|DLP12 prophage; predicted murein endopeptidase [Escherichia coli str K-12 substr MG1655]	1	0	0	0	0
16692|vfid|37974|vsiid|58811|ssid|unknown protein encoded by IS2 [Escherichia coli O157:H7 EDL933]	2	0	0	0	0
16718|vfid|38026|vsiid|58863|ssid|putative resolvase [Escherichia coli O157:H7 EDL933]	1	0	0	0	0
558|vfid|559|vsiid|558|ssid|transposase TnpA [Escherichia coli]	1	0	5217	0	0
7787|vfid|9499|vsiid|11548|ssid|alginate o-acetyltransferase AlgI [Pseudomonas aeruginosa PAO1]	2	0	1	0	0
8173|vfid|10259|vsiid|12686|ssid|endopeptidase Clp ATP-binding chain C [Listeria monocytogenes EGD-e]	5	7	11	0	0
961|vfid|962|vsiid|961|ssid|transposase [Pseudomonas aeruginosa]	1	33	0	0	0
13271|vfid|20471|vsiid|39961|ssid|aldehyde dehydrogenase [Vibrio cholerae O1 biovar El Tor str N16961]	0	1	1	0	0
136|vfid|137|vsiid|136|ssid|TnpA [Escherichia coli]	0	265	7	0	0
13942|vfid|22331|vsiid|42499|ssid|hypothetical protein [Escherichia coli]	0	167	0	0	0
13943|vfid|22333|vsiid|42502|ssid|hypothetical protein [Escherichia coli]	0	12	0	0	0
13944|vfid|22335|vsiid|42505|ssid|hypothetical protein [Escherichia coli]	0	3	0	0	0
14086|vfid|22620|vsiid|42932|ssid|bifunctional enterobactin receptor/adhesin protein [Escherichia coli CFT073]	0	4	0	0	0
14168|vfid|22786|vsiid|43180|ssid|ClpV1 [Pseudomonas aeruginosa PAO1]	0	5	0	0	0
16423|vfid|37436|vsiid|58273|ssid|putative iron-regulated outer membrane virulence protein [Escherichia coli CFT073]	0	2	0	0	0
16472|vfid|37534|vsiid|58371|ssid|hypothetical protein c2399 [Escherichia coli CFT073]	0	11	0	0	0
16608|vfid|37806|vsiid|58643|ssid|IS30 transposase [Escherichia coli str K-12 substr MG1655]	0	1	0	0	0
16612|vfid|37814|vsiid|58651|ssid|putative amino acid/amine transport protein [Escherichia coli K12]	0	2	0	0	0
16712|vfid|38014|vsiid|58851|ssid|hypothetical protein Z5879 [Escherichia coli O157:H7	0	2	0	0	0

EDL933]					
16850|vfid|38289|vsiid|59126|ssid|putative phage inhibition, colicin resistance and tellurite resistance protein [Escherichia coli O157:H7 EDL933]	0	6	6	0	0
16853|vfid|38295|vsiid|59132|ssid|putative phage inhibition, colicin resistance and tellurite resistance protein [Escherichia coli O157:H7 EDL933]	0	1	0	0	0
16856|vfid|38301|vsiid|59138|ssid|hypothetical protein Z1608 [Escherichia coli O157:H7 EDL933]	0	3	0	0	0
16877|vfid|38343|vsiid|59180|ssid|putative urease accessory protein G [Escherichia coli O157:H7 EDL933]	0	52	6	0	0
16882|vfid|38353|vsiid|59190|ssid|urease subunit gamma [Escherichia coli O157:H7 EDL933]	0	2	2	0	0
21131|vfid|52490|vsiid|73357|ssid|aminoglycoside 6`-N-acetyltransferase Iv [Acinetobacter sp 631]	0	1	0	0	0
21358|vfid|52928|vsiid|73795|ssid|AdeA membrane fusion protein [Acinetobacter baumannii]	0	17	0	0	0
21359|vfid|52930|vsiid|73797|ssid|AdeA membrane fusion protein [Acinetobacter baumannii]	0	5	0	0	0
21360|vfid|52932|vsiid|73799|ssid|AdeB [Acinetobacter baumannii]	0	369	0	0	0
21361|vfid|52934|vsiid|73801|ssid|AdeB RND protein [Acinetobacter baumannii]	0	3	0	0	0
21362|vfid|52936|vsiid|73803|ssid|AdeC outer membrane protein [Acinetobacter baumannii]	0	1	0	0	0
22924|vfid|55916|vsiid|76783|ssid|cation/multidrug efflux pump [Acinetobacter baumannii]	0	188	7	0	0
24250|vfid|58435|vsiid|79302|ssid|RND efflux transporter [Pseudomonas aeruginosa PA7]	0	55	19	0	0
24251|vfid|58437|vsiid|79304|ssid|RND efflux transporter [Pseudomonas aeruginosa UCBPP-PA14]	0	3	0	0	0
24958|vfid|59824|vsiid|80691|ssid|Putative tetA efflux pump [Acinetobacter baumannii]	0	9	0	0	0
452|vfid|453|vsiid|452|ssid|TnpA [Escherichia coli]	0	66	3	0	0

7250|vfid|8385|vsiid|10303|ssid|Biopolymer transport exbB protein [Escherichia coli CFT073]	0	6	6	0	0
7402|vfid|8690|vsiid|10608|ssid|twitching motility protein PilT [Pseudomonas aeruginosa PAO1]	0	66	0	0	0
7403|vfid|8692|vsiid|10610|ssid|twitching motility protein PilU [Pseudomonas aeruginosa PAO1]	0	22	0	0	0
7404|vfid|8694|vsiid|10612|ssid|twitching motility protein PilG [Pseudomonas aeruginosa PAO1]	0	4	0	0	0
7407|vfid|8700|vsiid|10618|ssid|twitching motility protein PilJ [Pseudomonas aeruginosa PAO1]	0	9	0	0	0
7412|vfid|8710|vsiid|10628|ssid|type 4 fimbrial biogenesis protein PilB [Pseudomonas aeruginosa PAO1]	0	6	0	0	0
7416|vfid|8718|vsiid|10636|ssid|two-component response regulator PilR [Pseudomonas aeruginosa PAO1]	0	1	0	0	0
7424|vfid|8734|vsiid|10652|ssid|Type 4 fimbrial biogenesis outer membrane protein PilQ precursor [Pseudomonas aeruginosa PAO1]	0	1	0	0	0
7813|vfid|9551|vsiid|11626|ssid|ferripyoverdine receptor [Pseudomonas aeruginosa PAO1]	0	1080	0	0	1
7827|vfid|9581|vsiid|11672|ssid|general secretion pathway protein E [Pseudomonas aeruginosa PAO1]	0	60	0	0	0
8350|vfid|10572|vsiid|13135|ssid|chaperonin GroEL [Legionella pneumophila subsp. pneumophila str Philadelphia 1]	0	13	0	0	0
8399|vfid|10670|vsiid|13282|ssid|type II protein secretion ATPase LspE [Legionella pneumophila subsp. pneumophila str Philadelphia 1]	0	7	0	0	0
8459|vfid|10790|vsiid|13462|ssid|catalase-peroxidase KatB [Legionella pneumophila subsp. pneumophila str Philadelphia 1]	0	8	1	0	0
8460|vfid|10792|vsiid|13465|ssid|superoxide dismutase [Legionella pneumophila subsp. pneumophila str Philadelphia 1]	0	1	0	0	0
864|vfid|865|vsiid|864|ssid|tnpR [Pseudomonas aeruginosa]	0	22	0	0	0

14055|vfid|22557|vsiid|42838|ssid|hypothetical protein c3561 [Escherichia coli CFT073]	0	0	4	0	0
16173|vfid|36936|vsiid|57773|ssid|putative tail component of prophage [Escherichia coli CFT073]	0	0	1	0	0
16179|vfid|36948|vsiid|57785|ssid|putative tail component of prophage [Escherichia coli CFT073]	0	0	2	0	0
16196|vfid|36982|vsiid|57819|ssid|cold shock-like protein cspI [Escherichia coli CFT073]	0	0	95	0	0
16616|vfid|37822|vsiid|58659|ssid|IS1 transposase B [Escherichia coli str K-12 substr MG1655]	0	0	1	0	0
16625|vfid|37840|vsiid|58677|ssid|ornithine carbamoyltransferase 2, chain F; CP4-6 prophage [Escherichia coli str K-12 substr MG1655]	0	0	1	0	0
16721|vfid|38032|vsiid|58869|ssid|hypothetical protein Z5888 [Escherichia coli O157:H7 EDL933]	0	0	1	0	0
21217|vfid|52650|vsiid|73517|ssid|kanamycin nucleotidyltransferase [Staphylococcus aureus subsp aureus Mu50]	0	0	1	0	0
22715|vfid|55533|vsiid|76400|ssid|bleomycin resistance protein [Staphylococcus aureus subsp aureus N315]	0	0	4	0	0
24255|vfid|58445|vsiid|79312|ssid|COG0841: Cation/multidrug efflux pump [Pseudomonas aeruginosa C3719]	0	0	2	0	0
24268|vfid|58471|vsiid|79338|ssid|multidrug efflux protein [Pseudomonas aeruginosa PA7]	0	0	5	0	0
384|vfid|385|vsiid|384|ssid|nonmetallocarbapenamase regulator [Enterobacter cloacae]	0	0	2	0	0
480|vfid|481|vsiid|480|ssid|Beta-lactamase [Escherichia coli]	0	0	2	0	0
7698|vfid|9291|vsiid|11222|ssid|glycerol ester hydrolase [Staphylococcus aureus subsp aureus MW2]	0	0	1	0	0
788|vfid|789|vsiid|788|ssid|SHV-5a [Klebsiella pneumoniae]	0	0	1	0	0
7931|vfid|9789|vsiid|11984|ssid|UDP-glucose pyrophosphorylase [Streptococcus pyogenes M1 GAS]	0	0	11	0	0
794|vfid|795|vsiid|794|ssid|ES-beta-lactamase [Escherichia coli]	0	0	4	0	0

800|vfid|801|vsiid|800|ssid|ES-beta-lactamase [Klebsiella pneumoniae]	0	0	1	0	0
8175|vfid|10263|vsiid|12692|ssid|ATP-dependent Clp protease proteolytic subunit [Listeria monocytogenes EGD-e]	0	0	2	0	0
832|vfid|833|vsiid|832|ssid|beta-lactamase TEM-1 [Acinetobacter baumannii]	0	0	18	0	0
8350|vfid|10572|vsiid|13135|ssid|chaperonin GroEL [Legionella pneumophila subsp pneumophila str Philadelphia 1]	0	0	2	0	0
8460|vfid|10792|vsiid|13465|ssid|superoxide dismutase [Legionella pneumophila subsp pneumophila str Philadelphia 1]	0	0	1	0	0
912|vfid|913|vsiid|912|ssid|class A beta-lactamase TEM-106 [Escherichia coli]	0	0	1	0	0
927|vfid|928|vsiid|927|ssid|inhibitor-resistant beta-lactamase TEM-79 [Escherichia coli]	0	0	1	0	0
959|vfid|960|vsiid|959|ssid|beta-lactamase class A [Klebsiella pneumoniae]	0	0	1	0	0
973|vfid|974|vsiid|973|ssid|beta-lactamase [Enterobacter aerogenes]	0	0	1	0	0
114|vfid|115|vsiid|114|ssid|transposase.[Escherichia.coli]	0	0	0	1	0
136|vfid|137|vsiid|136|ssid|TnpA.[Escherichia.coli]	0	0	0	2	0
13942|vfid|22331|vsiid|42499|ssid|hypothetical.protein.[Escherichia.coli]	0	0	0	1	0
21360|vfid|52932|vsiid|73799|ssid|AdeB.[Acinetobacter.baumannii]	0	0	0	3	0
452|vfid|453|vsiid|452|ssid|TnpA.[Escherichia.coli]	0	0	0	1	0
7403|vfid|8692|vsiid|10610|ssid|twitching.motility.protein.PilU.[Pseudomonas.aeruginosa .PAO1]	0	0	0	1	0
7407|vfid|8700|vsiid|10618|ssid|twitching.motility.protein.PilJ.[Pseudomonas.aeruginosa. PAO1]	0	0	0	1	0
7813|vfid|9551|vsiid|11626|ssid|ferripyoverdine.receptor[Pseudomonas.aeruginosaPAO1]	0	0	0	5	0
864|vfid|865|vsiid|864|ssid|tnpR.[Pseudomonas.aeruginosa]	0	0	0	1	0
13606|vfid|21659|vsiid|41491|ssid|unknown [Escherichia coli]	0	0	0	0	84
13828|vfid|22103|vsiid|42157|ssid|ORF A protein [Escherichia coli]	0	0	0	0	78
13879|vfid|22205|vsiid|42310|ssid|ORF B protein [Escherichia coli]	0	0	0	0	7
14021|vfid|22489|vsiid|42736|ssid|transposase Orf B, insertion element IS100 [E. coli]	0	0	0	0	1