Measuring the archaeome: detection and quantification of archaea signatures in the human body

Manuela R. Pausan[1], Cintia Csorba[1,+], Georg Singer[2], Holger Till[2], Veronika Schöpf[3,5], Elisabeth Santigli[4], Barbara Klug[4], Christoph Högenauer[1], Marcus Blohs[1], Christine Moissl-Eichinger[1,5,*]

[1] Department of Internal Medicine, Medical University of Graz, Graz, Austria

[2] Department of Pediatrics and Adolescent Surgery, Medical University of Graz, Graz, Austria

[3] Institute of Psychology, University of Graz, Graz, Austria

[4] Department of Dentistry and Maxillofacial Surgery, Medical University Graz, Graz, Austria

[5] BioTechMed, Graz, Austria

[+] Present address: AIT Austrian Institute of Technology GmbH, Tulln, Austria

[*] Corresponding author: christine.moissl-eichinger@medunigraz.at

Abstract

Due to their fundamentally different biology, archaea are consistently overlooked in conventional 16S rRNA gene amplicon surveys. Herein, we evaluate different methodological set-ups to detect and quantify archaea signatures in human samples (nose, oral, appendix, stool, and skin) using amplicon sequencing and quantitative PCR.

With our optimized protocol, we were able to increase the detection of archaeal RSVs from one (using a so-called "universal" approach) to 81 RSVs in a representative sample set. Moreover, we confirmed the presence of about 5% archaeal signatures in the human gut, but found, unexpectedly, an almost 1:1 ratio of archaeal to bacterial 16S rRNA genes in appendix and nose samples. This finding indicates a high prevalence of archaeal signatures in body regions thus far not analyzed for the presence of archaea using appropriate methods.

In order to assess the archaeome diversity and archaeal abundance, a specific archaea-targeting methodology is required, for which we propose two standard procedures. These methodologies might not only prove useful for analyzing the human archaeome in more detail, but could also be used for other holobionts' samples.

Introduction

The importance of microbial communities to human and environmental health motivates microbiome research to uncover their diversity and function. While the era of metagenomics and metatranscriptomics has begun, 16S rRNA gene amplicon sequencing still remains one of the most used methods to explore microbial communities, mainly due to the relatively low cost, the number of available pipelines for data analysis, and the comparably low computational power required.

It has been recognized that methodological issues in sample processing can significantly influence the outcome of microbiome studies, affecting comparability between different studies (Clooney et al., 2016; de la Cuesta-Zuluaga and Escobar, 2016) or leading to an over-and under-estimation of certain microbial clades (Eisenstein, 2018; Eloe-Fadrosh et al., 2016). For better comparability among different studies, standard operational procedures for sampling, storing samples, DNA extraction, amplification and analysis were set-up (e.g. the Earth Microbiome Project (Gilbert et al., 2014) and the Human Microbiome Project (Methé et al., 2012)). This includes the usage of so-called "universal primers" (Caporaso et al., 2012; Klindworth et al., 2013; Walters et al., 2016), to maximally cover the broadest prokaryotic diversity.

The human microbiome consists of bacteria, archaea, eukaryotes and viruses. The overwhelming majority of microbiome studies is bacteria-centric, but in recent years, awareness on eukaryotes (in particular fungi) and viruses has increased (Halwachs et al., 2017; Seed, 2014; Zou et al., 2016). However, most microbiome studies still remain blind for the human archaeome (Eisenstein, 2018; Moissl-Eichinger et al., 2018). A few of the underlying reasons for the under-representation of archaea in microbiome studies are (i) primer mismatches of the "universal primers" (Raymann et al., 2017), (ii) the sometimes too low abundance of the archaeal DNA in the studied samples, (iii) improper DNA extraction methods, and (iv) the incompleteness of the 16S rRNA gene reference databases. This is not only leading to an insufficient taxonomy assignment (unclassification) of archaea, but also, in the worst case, to the removal of archaeal signatures from the retrieved

datasets (Ding and Schloss, 2014; Fischer et al., 2016). Moreover, the clinical interest on archaea is minor, due to the fact that there are no known or proved archaeal pathogens yet (Gill et al., 2006).

Nevertheless, (methanogenic) archaea are part of the commensal microorganisms inhabiting the human body, being regularly detected in the oral cavity and the gastrointestinal tract (Chaudhary et al., 2015; Gaci et al., 2014; Horz and Conrads, 2011; Nkamga et al., 2017); in the latter they sometimes even outnumber the most abundant bacterial species (14%, (Tyakht et al., 2013);). Most human archaea studies use either cultivation or qPCR methods (Grine et al., 2017; Koskinen et al., 2017; van de Pol et al., 2017; Wampach et al., 2017) and only a few, 16S rRNA gene-based archaea-centric studies are available (Koskinen et al., 2017; Moissl-Eichinger et al., 2017). These new studies have shown that archaea are also present in the human respiratory tract (Koskinen et al., 2017) and on human skin in considerable amounts (Moissl-Eichinger et al., 2017; Probst et al., 2013). Furthermore, Koskinen et al. (2017) have shown for the first time that archaea reveal a body site specific pattern, similar to bacteria: the gastrointestinal tract being dominated by methanogens, the skin by Thaumarchaeota, the lungs by Woesearchaeota, and the nose archaeal communities being composed of mainly methanogens and Thaumarchaeota. Altogether, this indicates a substantial presence of archaea in some, or even all, human tissues.

As a logic consequence of our previous studies, we have started to optimize the detection and quantification methods of archaea as human commensals. We tested, *in silico* and experimentally, 27 different 16S rRNA gene targeting primer pair combinations suitable for NGS amplicon sequencing, to detect the archaeal diversity in samples from different body sites, including respiratory tract (nose samples), digestive tract (oral samples, stool and appendix specimens), and skin. Furthermore, we optimized qPCR protocols for quantifying the archaeal 16S rRNA gene to assess the bacteria/archaea ratios. Our results culminate in a proposed standard operating procedure for archaea diversity analysis and quantification in human samples.

Material and methods

Ethics statement

Research involving human material was performed in accordance with the Declaration of Helsinki and was approved by the local ethics committees (the Ethics Committee at the Medical University of Graz, Graz, Austria). (Bacterial) microbiome studies of some of the samples used in this study have already been published elsewhere (oral, nose, skin samples: (Klymiuk et al., 2016; Koskinen et al., 2018; Santigli et al., 2016). Details of the ethics approvals obtained are shown there. Appendix samples and stool samples have been obtained covered by the ethics votes: 25-469 ex12/13, and 27-151 ex 14/15.

Selection of samples and DNA extraction

Five representative samples types from various body sites including the respiratory tract (nose swabs), the digestive tract (oral biofilm, appendix biopsy and stool samples) and skin swabs were selected for the comparison of amplification-based protocols. All samples underwent pre-screening for archaea-positive qPCR and NGS signals with previously described protocols (Koskinen et al., 2017).

The nose swabs were obtained from healthy volunteers and were taken from the olfactory mucosa located at the ceiling of the nasal cavity using ultra minitip nylon flocked swabs (Copan, Brescia, Italy; n=7) (Koskinen et al., 2018). The oral samples have been obtained by standardized protocol for paper point sampling (Santigli et al., 2017) from healthy volunteers who participated in a microbiome study investigating the subgingival biofilm formation in children (n=7) (Santigli et al., 2016). Appendix samples have been obtained from collaboration partners at the Department of Pediatric and Adolescent Surgery and the Institute of Pathology, both Medical University of Graz (appendix specimens were obtained during pediatric appendectomies from either acute or ulcerous appendicitis; n=6). Stool samples have been obtained from healthy adult volunteers (n=5), and from one patient with above average methane production after metronidazole treatment (n=1; this

sample was used for comparing different amplification protocols). Skin samples were obtained from healthy adult volunteers from either the back (n=1; this sample was used for comparing different amplification protocols) or the left forearm, using BD Culture Swabs™ (Franklin Lakes, New Jersey, USA; n=7).

In all cases, the genomic DNA was extracted by a combination of mechanical and enzymatic lysis. However, depending on the sample type, different protocols were chosen: for the stool samples around 200mg of sample has been used for DNA extraction using the E.Z.N.A. stool DNA kit according to the manufacturer's instruction. The DNA from the appendix samples was obtained using the AllPrep DNA/RNA/Protein Mini Kit (QIAGEN), small pieces of cryotissue were homogenized before 3 times for 30s at 6500rpm using the MagNALyzer ® instrument (Roche Molecular Systems) with buffer RTL and β-mercaptoethanol (according to the manufacturer's instructions). For the nose and skin samples from the forearm, the DNA was extracted using the FastDNA Spin Kit (MP Biomedicals, Germany) according to the provided instructions. The DNA from the oral samples and from the skin samples from the back were isolated using the MagnaPure LC DNA Isolation Kit III (Bacteria, Fungi; Roche, Mannheim, Germany) as described by Santigli et al. (Santigli et al., 2016) and Klymiuk et al. (Klymiuk et al., 2016).

NOTE: Sample set 1 (one representative sample from each body site) was used to initially evaluate the primers and methods, whereas sample set 2 (7 nose samples, 7 oral samples, 6 appendices, 7 skin samples) was then used for assessing the archaeal diversity and quantity, based on the chosen, optimized protocol.

16S rRNA gene primer selection and pre-analysis *in silico* evaluation

Different primer pairs targeting the archaeal 16S rRNA gene region have been selected from recent publications (Klindworth et al., 2013; Koskinen et al., 2017). The main criteria for selection were: a. specificity for archaea *in-silico*, b. low or no amplification of eukaryotic DNA, and c. amplicon length between 150 to 300bp, suitable for NGS such as Illumina MiSeq. In addition, three "universal" primer

pairs (Caporaso et al., 2012; Klindworth et al., 2013; Walters et al., 2016) were tested in parallel to

determine their efficiency in detecting archaea in human samples. Full information on the selected

primer pairs is given in Table 1. *In silico* evaluation of the selected primer pairs has been performed

using the online tool TestPrime1.0 (Klindworth et al., 2013) and the non-redundant SILVA database

SSU132 (Quast et al., 2013). Two of the primers were also tested using TestProbe 3.0 (Klindworth et

al., 2013) and the SILVA database SSU132 to assess their individual coverage for the archaeal domain.

Table 1. Primer selection and results of the pre-analysis *in silico* evaluation of all primer pairs used. Coverage of archaea, bacteria and eukarya is given in percentages, depending on whether no or one mismatch was allowed. Designated "universal" primers (primer pairs 10-12) are indicated in bold letters.

Primer pair	Name	Primer name, according to (Klindworth et al., 2013)	Sequence (5' -> 3')	0 mismatch			1 mismatch		
				Archaea	Bacteria	Eukarya	Archaea	Bacteria	Eukarya
1	344F	S-D-Arch-0344-a-S-20	ACGGGGYGCAGCAGGCGCGA	46.1%	0.0%	0.0%	68.6%	0.0%	0.0%
	915R	S-D-Arch-0911-a-A-20	GTGCTCCCCCGCCAATTCCT						
2	349F	S-D-Arch-0349-a-S-17	GYGCASCAGKCGMGAAW	71.8%	0.0%	0.0%	87.8%	0.0%	0.0%
	915R	S-D-Arch-0911-a-A-20	GTGCTCCCCCGCCAATTCCT						
3	344F	S-D-Arch-0344-a-S-20	ACGGGGYGCAGCAGGCGCGA	51.5%	0.0%	0.0%	73.0%	0.0%	0.0%
	1041R	S-D-Arch-1041-a-A-18	GGCCATGCACCWCCTCTC						
4	349F	S-D-Arch-0349-a-S-17	GYGCASCAGKCGMGAAW	71.2%	0.0%	0.0%	90.0%	0.0%	0.0%
	1041R	S-D-Arch-1041-a-A-18	GGCCATGCACCWCCTCTC						
5	519F	S-D-Arch-0519-a-S-15	CAGCMGCCGCGGTAA	79.3%	0.0%	0.0%	93.7%	0.0%	0.0%
	1041R	S-D-Arch-1041-a-A-18	GGCCATGCACCWCCTCTC						
6	344F	S-D-Arch-0344-a-S-20	ACGGGGYGCAGCAGGCGCGA	48.3%	0.0%	0.0%	71.3%	0.0%	0.0%
	806R	S-D-Arch-0786-a-A-20	GGACTACVSGGGTATCTAAT						
7	349F	S-D-Arch-0349-a-S-17	GYGCASCAGKCGMGAAW	75.2%	0.0%	0.0%	91.1%	0.0%	0.0%
	806R	S-D-Arch-0786-a-A-20	GGACTACVSGGGTATCTAAT						
8	519F	S-D-Arch-0519-a-S-15	CAGCMGCCGCGGTAA	85.6%	6.8%	0.0%	95.2%	90.9	0.1%
	806R	S-D-Arch-0786-a-A-20	GGACTACVSGGGTATCTAAT						
9	349F	S-D-Arch-0349-a-S-17	GYGCASCAGKCGMGAAW	79.3%	0.0%	0.0%	92.8%	0.0%	0.1%
	519R	S-D-Arch-0519-a-A-16	TTACCGCGGCKGCTG						
10	**519F**	S-D-Arch-0519-a-S-15	CAGCMGCCGCGGTAA	88.9%	88.8%	0.6%	95.3%	95.4%	1.2%
	785R	S-D-Bact-0785-b-A-18	TACNVGGGTATCTAATCC						
11	**515F**	515F-original	GTGCCAGCMGCCGCGGTAA	52.9%	86.8%	0.0%	94.6%	95.0%	0.3%
	806uR	806R-original	GGACTACHVGGGTWTCTAAT						
12	**515FB**	515F-modified	GTGYCAGCMGCCGCGGTAA	85.7%	87.7%	0.0%	95.4%	95.1%	1.4%
	806RB	806R-modified	GGACTACNVGGGTWTCTAAT						

PCR and library preparation

For archaea-targeting PCR, a nested approach was chosen to increase the specificity for archaea and to avoid the formation of primer dimers caused by the tag, necessary for Illumina sequencing, attached to the primers (Koskinen et al., 2017; Peng et al., 2015).

In addition to the nested approach, a standard PCR was performed with three different universal primer pairs, and one archaeal primer pair for comparative reasons, and to test if a universal approach is capable to cover archaea in human samples in sufficient depth. All primer combinations (in total 27) used for the PCR reactions are provided in Table 2.

For the first PCR, each reaction was performed in a final volume of 20 µl containing: TAKARA Ex Taq® buffer with MgCl$_2$ (10 X; Takara Bio Inc., Tokyo, Japan), primers 500 nM, BSA (Roche Lifescience, Basel, Switzerland) 1 mg/ml, dNTP mix 200 µM, TAKARA Ex Taq® Polymerase 0.5 U, water (Lichrosolv®; Merck, Darmstadt, Germany), and DNA template (1-50 ng/µl).

After the first PCR, the resulting amplicons were purified to remove primer remnants. This purification was performed with three different kits to compare the different yields and efficiencies, namely MinElute PCR Purification kit (Qiagen; Hilden, Germany), Monarch® PCR & DNA Cleanup Kit (5 µg) (New England Biolabs GmbH; Ipswich, USA), or innuPREP DOUBLEpure Kit (Analytik Jena, Germany) as indicated in Table 2. The purified PCR product was eluted in 10 µl water (Lichrosolv®; Merck, Darmstadt, Germany).

Two µl of the resulting, purified PCR products were transferred into a subsequent 2nd PCR containing the following mixture: TAKARA Ex Taq® buffer with MgCl$_2$ (10 X; Takara Bio Inc., Tokyo, Japan), primers 500 nM, BSA (Roche Lifescience, Basel, Switzerland) 1 mg/ml, dNTP mix 200 µM, TAKARA Ex Taq® Polymerase 0.5 U, and water (Lichrosolv®; Merck, Darmstadt, Germany) up to a volume of 25 µL.

Table 2 displays all primer pair combinations used for the first and the second PCR of the nested approach and the "universal" PCR. If not indicated otherwise (in brackets), the first PCR was followed by a purification of the PCR product by the MinElute PCR Purification kit (QIAGEN) kit. n.a.: not applicable.

PCR #	Primer combination 1st PCR	Primer combination 2nd PCR
PCR21	349F-915R	Illu 349F-Illu519R
PCR22		Illu 519F-Illu785R
PCR23		Illu 519F-Illu806R
PCR31	344F-1041R	Illu 349F-Illu519R
PCR33		Illu 519F-Illu785R
PCR34		Illu 519F-Illu806R
PCR41	349F-1041R	Illu 349F-Illu519R
PCR42		Illu 519F-Illu785R
PCR43		Illu 519F-Illu806R
PCR61	349F-806R	Illu 349F-Illu519R
PCR62		Illu 519F-Illu785R
PCR63		Illu 519F-Illu806R
PCR71	519F-1041R	Illu 519F-Illu785R
PCR72		Illu 519F-Illu806R
PCR81	519F-806R	Illu 519F-Illu785R
PCR82		Illu 519F-Illu806R
PCR91	344F-519R	Illu 349F-Illu519R
PCRQ1	344F-915R (QIAGEN)	Illu 349F-Illu519R
PCRQ3		Illu 519F-Illu785R
PCRQ4		Illu 519F-Illu806R
PCRM1	344F-915R (NEB Monarch)	Illu 349F-Illu519R
PCRM3		Illu 519F-Illu785R
PCRM4		Illu 519F-Illu806R
PCRA1	344F-915R (Analytik Jena)	Illu 349F-Illu519R
PCRA3		Illu 519F-Illu785R
PCRA4		Illu 519F-Illu806R
PCRQ5	344F-806R (QIAGEN)	Illu 349F-Illu519R
PCRQ6		Illu 519F-Illu785R
PCRQ7		Illu 519F-Illu806R
PCRM5	344F-806R (NEB Monarch)	Illu 349F-Illu519R
PCRM6		Illu 519F-Illu785R
PCRM7		Illu 519F-Illu806R
PCR8-Uni	n.a.	Illu 515F-Illu806uR
PCR9-Uni		Illu 515FB-Illu806RB
PCR10		Illu 519F-Illu806R
PCR11-Uni		Illu 519F-Illu785R

0

1

2

The PCR cycling conditions are listed in Table 3, according to the primer pairs used. For all primer pairs, annealing temperatures were either determined experimentally by gradient PCR or adopted from literature information.

Table 3: PCR conditions. For denaturation, annealing and elongation the corresponding time and temperature is given.

Target	Archaea (16S rRNA gene)			"Universal" (16S rRNA gene)	
(Nested) PCR, round	1°	1°	2°	1°	1°
Primer pair	344F / 915R 349F / 915R 344F / 806R 349F / 806R 519F / 806R	344F / 1041R 349F / 1041R 519F / 1041R	All Illumina tagged primer pairs	Illu519F /Illu806R Illu519F/Illu785R	Illu515F/Illu806uR Illu515FB/Illu806RB
Initial denaturation	2', 95°C	5', 95°C	5', 95°C	5', 95°C	3', 94°C
Denaturation	30'', 96°C (first 10 cycl.), 25'' 94°C	30'', 94°C	40'', 95°C	40'', 95°C	45", 94°C
Annealing	30'', 60°C	45'', 56°C	2', 63°C	2', 63°C	1', 50°C
Elongation	1', 72°C	1', 72°C	1', 72°C	1',72°C	1' 30", 72°C
Final elongation	10', 72°C	10', 72°C	10', 72°C	10', 72°C	10', 72°C
No. of cycles	25	25	30	40	40

Sample set 2 was amplified using the primer combination 344F-1041R/519F-806R (Table 2). For the first PCR, each reaction was performed in a final volume of 20 µl as described above. After the first PCR, the PCR products were purified using Monarch® PCR & DNA Cleanup Kit (5 µg; New England Biolabs GmbH). For the second PCR, the final volume was 25 µl, as described above, only the volume of the DNA template varied: 2 µl purified PCR product for stool and nose samples, 4 µl for all other samples.

Next generation sequencing, bioinformatics and statistical analyses

Amplicons were sequenced at the ZMF Core Facility Molecular Biology in Graz, Austria, using the Illumina MiSeq platform (Klymiuk et al., 2016). The MiSeq amplicon sequence data was deposited in the European Nucleotide Archive under the study accession number PRJEB27023.

The data processing of the obtained MiSeq sequence data was performed using the open source package DADA2 (Divisive Amplicon Denoising Algorithm; (Callahan et al., 2016)) as described previously (Mora et al., 2016). Shortly, the DADA2 turns paired-end fastq files into merged, denoised, chimera-free, and inferred sample sequences called ribosomal sequence variants (RSVs). The taxonomic affiliations were determined using SILVA v128 database as the reference database (Quast et al., 2013). In the resulting RSV table, each row corresponds to non-chimeric inferred sample sequence with a separate taxonomic classification. RSV tables are given in Supplementary Tables 1 (a-c) and 2 (available on request).

Negative controls (extraction controls and no-template controls) were included during PCR amplification. The RSVs overlapping the negative controls and samples were either subtracted or completely removed from the data sets. RSVs detected in the negative controls are provided in Supplementary Table 3 (available on request).

Processing of sequencing data was performed using the in-house Galaxy set-up (Klymiuk et al., 2016) and subsequent statistical analyses were performed in R version 3.4.3 (R Core Team, 2013). Alpha diversity was calculated using the Shannon diversity. In order to identify differences between the archaeal diversity, Wilcoxon Rank Test was performed. The diversity of the archaeal communities within sample set 2 was determined using two diversity matrices (Shannon and richness). Analysis of variance (ANOVA) was performed to test for differences in the archaeal diversity based on the body location. Principal Coordinates Analysis (PCoA) based on Bray-Curtis distances was used to visualize differences between the samples from different body site. Redundancy discrimination analysis (RDA) was used to analyze the association between archaeal community composition and the body site location. RDA, alpha diversity and PCoA analysis were performed using Calypso Version 8.62 (Zakrzewski et al., 2016). The RSV tables obtained were used to summarize taxon abundance at different taxonomic levels. The taxonomic profiles obtained at the genus level were used to generate bar graphs for all samples.

A phylogenetic tree was constructed with the obtained archaeal RSVs from the universal approach, the archaeal primer pair 519F-806R, and from the archaeal specific primer pair combination 344F-1041R/519F-806R. The alignment was performed using the SILVA SINA (Pruesse et al., 2012) and the 5 most closely related available sequences (neighbors) were downloaded together with the aligned sequences. All sequences were cropped to the same length and used to construct a tree based on maximum-likelihood algorithm using MEGA7 (Kumar et al., 2016), using a bootstrap value of 500. The Newick output was further processed with iTOL interactive online platform (Letunic and Bork, 2007).

Quantitative PCR

To determine the optimal qPCR approach for assessing the absolute number of archaeal 16S rRNA gene copies in relevant samples, three different procedures were tested, two based on SYBR, one on TaqMan chemistry. In addition, two bacteria targeting approaches, one based on SYBR, one on TaqMan chemistry, were evaluated in parallel to determine the ratio of bacteria: archaea in these samples.

The used primers are given in Table 4 and the *in silico* evaluation results thereof are given in Table 5.

Table 4: Primers used for qPCR: approach, target, primer names, primer sequences, and according references.

Approach and target	Name	Primer name, according to (Klindworth et al., 2013)	Sequence (5'-3')	Reference
SYBR (Archaea)	344aF	S-D-Arch-0344-a-S-20	ACGGGGYGCAGCAGGCGCGA	(Raskin et al., 1994)
	517uR	S-*-Univ-0517-a-A-15	GWATTACCGCGGCKGCTG	(Amann et al., 1995)
SYBR (Archaea)	806aF	S-D-Arch-0787-a-S-20	ATTAGATACCCSBGTAGTCC	(Raskin et al., 1994)
	958aR	S-D-Arch-0958-a-A-19	YCCGGCGTTGAMTCCAATT	(DeLong, 1992)
TaqMan (Archaea)	349aF	S-D-Ar-0349-aS17_F	GYGCASCAGKCGMGAAW	(Takai and Horikoshi, 2000)
	806aR	S-D-Arch-0786-a-A-20	GGACTACVSGGGTATCTAAT	(Takai and Horikoshi, 2000)
	516aF	Arch516F_FAM	TGYCAGCCGCCGCGGTAAHACCVGC	(Takai and Horikoshi, 2000)
SYBR (Bacteria)	338bF	S-D-Bact-0337-a-S-20	ACTCCTACGGGAGGCAGCAG	(el Fantroussi et al., 1999)
	517uR	S-*-Univ-0517-a-A-15	GWATTACCGCGGCKGCTG	(Amann et al., 1995)
TaqMan (Bacteria)	331bF	S-D-Bact-0340-a-S-19	TCCTACGGGAGGCAGCAGT	(Hunter et al., 2002)
	797R	S-D-Bact-0781-a-A-26	GGACTACCAGGGTATCTAATCCTGTT	(Hunter et al., 2002)
	528R	Bact_HEX	CGTATTACCGCGGCTGCTGGCAC	(Hunter et al., 2002)

Table 5: *In silico* evaluation of qPCR primers and TaqMan probes.

Name	Primer name according to (Klindworth et al., 2013)	Sequence (5' -- 3')	0 mismatch			1 mismatch		
			Archaea	Bacteria	Eukarya	Archaea	Bacteria	Eukarya
344aF	S-D-Arch-0344-a-S-20	ACGGGGYGCAGCAGGCGCGA	28.9%	0.0%	0.0%	71.2%	0.0%	0.0%
517uR	S-*-Univ-0517-a-A-15	GWATTACCGCGGCKGCTG						
806aF	S-D-Arch-0787-a-S-20	ATTAGATACCCSBGTAGTCC	45.6%	0.0%	0.0%	81.8%	0.0%	0.0%
958aR	S-D-Arch-0958-a-A-19	YCCGGCGTTGAMTCCAATT						
349aF	S-D-Ar-0349-aS17_F	GYGCASCAGKCGMGAAW	75.2%	0.0%	0.0%	91.1%	0.0%	0.0%
806aR	S-D-Arch-0786-a-A-20	GGACTACVSGGGTATCTAAT						
338bF	S-D-Bact-0337-a-S-20	ACTCCTACGGGAGGCAGCAG	0.0%	90.5%	0.0%	0.0%	94.9%	0.1%
517uR	S-*-Univ-0517-a-A-15	GWATTACCGCGGCKGCTG						
331bF	S-D-Bact-0340-a-S-19	TCCTACGGGAGGCAGCAGT	0.0%	75.3%	0.0%	0.0%	82.8%	0.0%
797R	S-D-Bact-0781-a-A-26	GGACTACCAGGGTATCTAATCCTGTT						
Bact 528R_HEX (probe)		CGTATTACCGCGGCTGCTGGCAC	0.0%	79.1%	0.0%	0.4%	95.8%	0.8%
Arch 516F_FAM (probe)		TGYCAGCCGCCGCGGTAAHACCVGC	84.5%	0.0%	0.0%	96.2%	0.0%	1.2%

For SYBR-based qPCR, the reaction mix contained: 1x SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad, Hercules, USA), 300 nM of forward and reverse primer, gDNA template, water (Lichrosolv®; Merck, Darmstadt, Germany). For TaqMan-based qPCR, the reaction mix contained: 1x SsoAdvanced™ Universal Probes Supermix (Bio-Rad, Hercules, USA), 800 nM of forward and reverse primer, 200 nM of FAM-marked probe for archaea and HEX-marked probe for bacteria, gDNA template, water (Lichrosolv®; Merck, Darmstadt, Germany).

The qPCR was performed using the CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, Hercules, USA). The qPCR conditions used are given in Table 6.

Table 6: Quantitative PCR conditions. For denaturation, annealing and elongation the corresponding time and temperature is given. n.a.: not applicable.

Target and primers		QPCR conditions						Standard curves	
Target gene	primer forward/reverse	Initial denaturation	Denaturation	Annealing	Elongation	No. of cycles	Melting Curve	Efficiency	R2
16S rRNA Bacteria	338bF-517uR	15', 95°C	15", 94°C	30", 60°C	40", 72°C	40	60-95°C	106.4	0.949
	331bF - 797R (Probe 528bR)	10', 95°C	25", 95°C	2', 60°C		40	n.a.	94.9	0.97
16S rRNA Archaea	344aF-517uR	15', 95°C	15", 94°C	45", 60°C	30", 72°C	40	60-95°C	104.5	0.972
	806aF-957aR	15', 95°C	15", 94°C	30", 53°C	40", 72°C	40	60-95°C	101.5	0.983
	349aF-806aR (Probe 515aF)	10', 95°C	25", 95°C	5', 60°C		40	n.a.	92.5	0.837

Crossing point (Cq) values were determined using the regression method within the Bio-Rad CFX Manager software version 3.1. Absolute copy numbers of bacterial and archaeal 16S rRNA genes were calculated using the Cq values and the reaction efficiencies based on standard curves obtained from defined DNA samples from *Nitrososphaera viennensis* and *Escherichia coli* (Probst et al., 2013). The qPCR efficiency and R^2 values of the standard curves can be found in Table 6 (last two columns) and were in range of 92.5% to 106.40% and the R^2 values were between 0.837 and 0.983, the archaeal TaqMan approach having a low R^2 value of only 0.837.

Detection limits were defined based on the average Cq values of non-template controls (triplicates) and the corresponding standard curves of the positive controls. The detection limits were variable, depending on the primer pair used; herein, the detection limit is defined as the last positive signal before the signal of the negative control, in order to exclude false-positive results. For archaea detection the limit was 480 copies/µl when the primer pair 806aF-958aR was used and $7.46*10^4$ copies/µl with the primer pair 344aF-517uR, respectively, and 186 copies/µl for 349aF-806aR. For bacterial 16S rRNA genes, the detection limit was $6.45*10^3$ copies/µl for the primer pair 338bF-517uR that was used for SYBR and 508 copies/µl for the primer pair 331bF-797R that was used for TaqMan. The archaeal (806aF-958aR) and bacterial (338bF-517uR) qPCR was then used to determine the absolute copy numbers of 16S rRNA genes for sample set 2. The crossing points (Cq) were determined using the single threshold in Bio-Rad CFX Manager software version 3.1. The absolute copy numbers were calculated as mentioned above. The qPCR efficiency was 108.6% for archaea,

and 90.9% for Bacteria, respectively and R^2 values were 0.976 for archaea and 0.929 for Bacteria. The

detection limits were 715 copies/µl for archaea, and 7.06E+04 copies/µl for Bacteria.

All qPCR reactions have been performed in triplicates. Only samples with positive results in 2 out of 3

or 3 out of 3 replicates were considered for further analysis.

Results

Primer pairs were evaluated with respect to the following characteristics: i) high *in silico* specificity for archaeal 16S rRNA genes and an amplicon length of 150 to 300 bp, suitability for NGS and quantitative PCR, ii) *in vitro* capability to amplify diverse archaeal 16S rRNA genes from a variety of human specimens, and iii) *in vitro* suitability for qPCR with stringent requirements in efficiency and specificity.

Besides archaea-specific primer pairs, two widely used "universal" primers (515F-806uR original; 515FB-806RB modified; (Caporaso et al., 2012; Walters et al., 2016)) were evaluated all along to assess the potential of "universal" primers to display archaeal diversity associated with the human body.

Most archaea-targeting primers reveal good coverage in silico

A total of 12 different primer pairs were evaluated *in silico* (Table 1) using sample set 1. Most primer pairs showed high coverage for the archaeal domain ranging from 46% to 89% and revealed a high domain-specificity (8 of 12 primer pairs without matches outside of the archaeal domain). When one mismatch was allowed, the coverage increased to values from 68% to 95%.

One designated archaeal primer pair was found to target additionally sequences of the bacterial and eukaryotic domain when one mismatch was allowed, namely primer pair 519F-806R, with a coverage of the bacterial domain > 90%.

We further evaluated the detailed coverage of the primer pairs for specific archaeal phyla and genera of particular interest in human archaeome studies: Euryarchaeota, Thaumarchaeota, and Wosearcheota, as well as *Nitrososphaera, Methanobrevibacter*, *Methanosphaera* and *Methanomassiliicoccus*. For all subsequent *in silico* analyses we allowed one mismatch.

All primer pairs revealed a high coverage for the Euryarchaeota phylum (in total >90%), for genera *Methanobrevibacter* (between 94.6% and 98.9%) and *Methanomassiliicoccus* (between 92.9% and

100%), while the coverage for *Methanosphaera* was below 90% for most primer pairs except for 519F-806R and 349F-519R (Table 7).

The coverage of the Thaumarchaeota phylum depended on the primer pair used. Most analyses that included the primer 344F showed a low *in silico* coverage for Thaumarchaeota (below 30%) while all other primer pair combinations revealed a high coverage of this phylum (>90%; Table 7). The coverage for *Nitrososphaera* in particular varied between 86.9% and 94.4%. The class Wosearchaeia showed variable coverage between 65.2% and 89.5%.

Table 7: *In silico* analysis of the coverage of chosen primer pairs for specific archaeal taxa of interest. For primer full names and sequences, please refer to Table 1.

primer pair	Name	Euryarchaeota				Thaumarchaeota		Nanoarchaeota
		total	*Methano-brevibacter*	*Methano-sphaera*	*Methano-massiliicoccus*	total	*Nitrososphaera*	(Wosearchaeia)
1	344F	90.80%	95.30%	82.20%	100%	20.60%	87.60%	66.40%
	915R							
2	349F	91.50%	95.30%	84.20%	100%	92%	89.70%	70.30%
	915R							
3	344F	90.80%	94.60%	79.40%	100%	20.70%	89.00%	67.90%
	1041R							
4	349F	91.50%	94.60%	79.40%	100%	96.40%	92.30%	74.30%
	1041R							
5	519F	95%	97.80%	85.40%	92.90%	96.60%	90.60%	83%
	1041R							
6	344F	92.30%	95.50%	82.60%	100%	23.30%	88%	65.20%
	806R							
7	**349F**	93.20%	95.60%	84.20%	100%	96.50%	90.10%	72.90%
	806R							
8	519F	96.60%	98.90%	90%	95%	96.70%	89.50%	83.10%
	806R							
9	349F	93.60%	95.80%	90.70%	95%	98%	94.40%	83.10%
	519R							
10	519F	96.50%	98.60%	89.60%	95%	96.20%	87.80%	87.60%
	785R							
11	515F	96.20%	98.60%	89.60%	95%	94.70%	86.90%	89.50%
	806R							
12	515FB	96.20%	98.60%	89.60%	95%	96.50%	89%	89.50%
	806RB							

As the archaeal primer 344F has often been used for detecting archaea in a variety of environmental samples (Fontana et al., 2016; Zhang et al., 2014), we took a closer look on its coverage capacity using the TestProbe 3.0 (Klindworth et al., 2013) and the SILVA database SSU132 (Quast et al., 2013). Overall, the primer revealed 73.2% coverage of the archaeal domain. The *in silico* results showed a high coverage of the Euryarchaeota phylum (93.8%) and the genera within, especially *Methanobrevibacter* with 96.1%, *Methanosphaera* with 89.9% and *Methanomassiliicoccus* with 100%. It also revealed a good coverage for Wosearchaeia with 74.6%, but showed, despite a high coverage for the genus *Nitrososphaera* (93.6%), a generally low coverage of the Thaumarchaeota phylum with only 24%, indicating a potentially low capacity for studies with thaumarchaeotal diversity in focus.

Another primer that we analyzed in more detail was primer 519F, also known as S-D-Arch-0519-a-S-15. As the sequence of this primer (5' - CAGCMGCCGCGGTAA - 3') overlaps with the sequence of the "universal" primer S-*-Univ-0519-a-S-18 (5' - CAGCMGCCGCGGTAATWC - 3'), we were interested to compare their coverages.

As expected, the results from the *in silico* analysis indicated that the primer S-D-Arch-0519-a-S-15 targets Bacteria (coverage 98%), archaea (coverage 98.2%) and Eukarya (coverage 96.4%). The universal primer S-*-Univ-0519-a-S-18 has a similar coverage and specificity for the three domains of life: Bacteria (coverage 97.5%), archaea (coverage 96.4%), and eukarya (coverage 95.6%). Considering our *in silico* results, the primer S-D-Arch-0519-a-S-15 cannot be used to target archaea specifically and should be re-named to S-D-Univ-0519-a-S-15.

As most selected archaea-targeting primers revealed a good coverage of the archaeal domain in general, all primer pairs were used for subsequent wet-lab experiments.

Archaeal community composition varies according to the used primer pairs and universal primers fail to detect the archaeal diversity

Herein we sought to identify the optimal primer pair for amplicon sequencing of the archaeomes in human samples. For this, we selected five representative sample types from different body sites: nose (upper nasal cavity), oral (subgingival sites), stool and appendix specimens, and skin (back) (sample set 1). The stool sample represented the positive control and served as a natural mock community.

Next generation sequencing was performed, after a two-step nested PCR (for archaea) or a single-step PCR ("universal" target). The nested PCR approach was selected based on the reasons given in the Materials and Methods section. In brief, the first PCR was intended to select the archaeal community of interest, the second to further amplify the archaeal signal.

The results obtained after amplification, NGS and data analysis based on DADA2 algorithm (Callahan et al., 2016; Koskinen et al., 2017) are summarized in Supplementary Table 4 (a-c) (available on request), which includes the number of reads and observed ribosomal sequence variants (RSVs) obtained for all samples covering the three domains, Archaea, Bacteria, and Eukarya (plus unclassified taxa).

The use of universal primers (primer pair 515F-806uR, 515FB-806RB and 519F-785R) in the PCR reaction resulted in reads that were classified mainly within the bacterial domain with almost no reads classified within the archaea, confirming our previous observations (Koskinen et al., 2017). In fact, when the two universal primer pairs (515F-806uR original and 515FB-806RB) were compared regarding the archaeal domain, only primer pair 515F-806uR allowed the detection of only one RSV being classified within the archaea and from only one sample, the stool sample (Supplementary Table 1a) (available on request).

Universal primer pair 519F-785R yielded slightly better results, allowing the detection of three different archaeal RSVs from two different samples: *Methanobrevibacter* and *Methanosphaera* in the

stool sample, and one RSV from the nose sample, classified within the Thaumarchaeota phylum. Very similar results (detection of the same methanoarchaeal signatures in the stool sample, and one thaumarchaeal signature in the oral sample instead of the nose sample) were obtained from primer pair 519F-806R, which was originally described to be archaea-specific, but revealed wide coverage of the bacterial and archaeal domain (>90%, when one mismatch allowed) *in silico* (see previous chapter).

The obtained archaeal RSVs from the universal approaches were used to construct a phylogenetic tree to identify whether the universal primer pairs allow the detection of the same RSVs or closely related RSVs in the analyzed samples (Fig. 1). In addition, the RSVs obtained from the archaeal specific primer pair combination 344F-1041R/519F-806R were included for comparison. This approach allowed the detection of 20 RSVs in the nose, 19 RSVs in the oral, one RSV in the appendix, 3 RSVs in the stool, and 39 RSVs in the skin sample. For the stool sample, the RSVs obtained from the universal and archaeal specific approach grouped together in clades, either within *Methanobrevibacter* or *Methanosphaera* clade (Fig. 1). The RSV from the nose sample detected through the universal approach grouped separately to the RSVs identified using the archaeal specific primers. The oral RSV from the universal approach grouped also separately from the RSVs obtained with the archaeal specific primers, the archaeal specific primers did not detect any Thaumarchaeota in the oral sample, only RSVs that were grouped within Woesearchaeota.

Fig. 1: Phylogenetic tree based on the retrieved RSVs from the universal approach, archaeal approach with primer 519F-806R or from the PCR based on the primer pair combination 344F-1041R/519F-806R as indicated in colors as an outermost circle (legend "Primer combinations"). The inner circle represents the body site from where the RSVs were identified (see legend). Reference sequences from the SILVA database are shown without label. The branches of the tree were colored according to the phyla, blue: Woesearchaeota, green: Euryarchaeota, and orange: Thaumarcheota. In the Euryarchaeota the different shades of green indicate the different clades, dark green: *Methanosphaera*, lighter green: *Methanobrevibacter*, green: *Methanobacterium*.

Overall, 10 out of 24 primer pair combinations allowed the detection of archaeal signatures in all analyzed samples (Supplementary Table 1a; provided on request). But all 24 primer pair combinations were able to identify archaeal reads in at least one of the sample types analyzed, for example all primer pair combinations detected archaeal RSVs in the stool sample; the number of RSVs, however, varied according to the used primer pair combination from 1 RSV to 7 RSVs.

Depending on the used primer pair, the archaeal community composition was found to be highly variable (Fig. 2). We observed that the detected variation in the archaeal composition was due to the used primer pair in the first PCR, the primer pair used to select the communities, while the second

PCR and primer pair enhanced the signal of the first PCR (Fig. 2). It shall be mentioned that for the second PCR only three different primer pairs have been used, 349F-519R, 519F-785R and the 519F-806R, of which the first two primer pairs had been used before to explore archaeal communities in human samples (Koskinen et al., 2017) and in confined habitats (Mora et al., 2016).
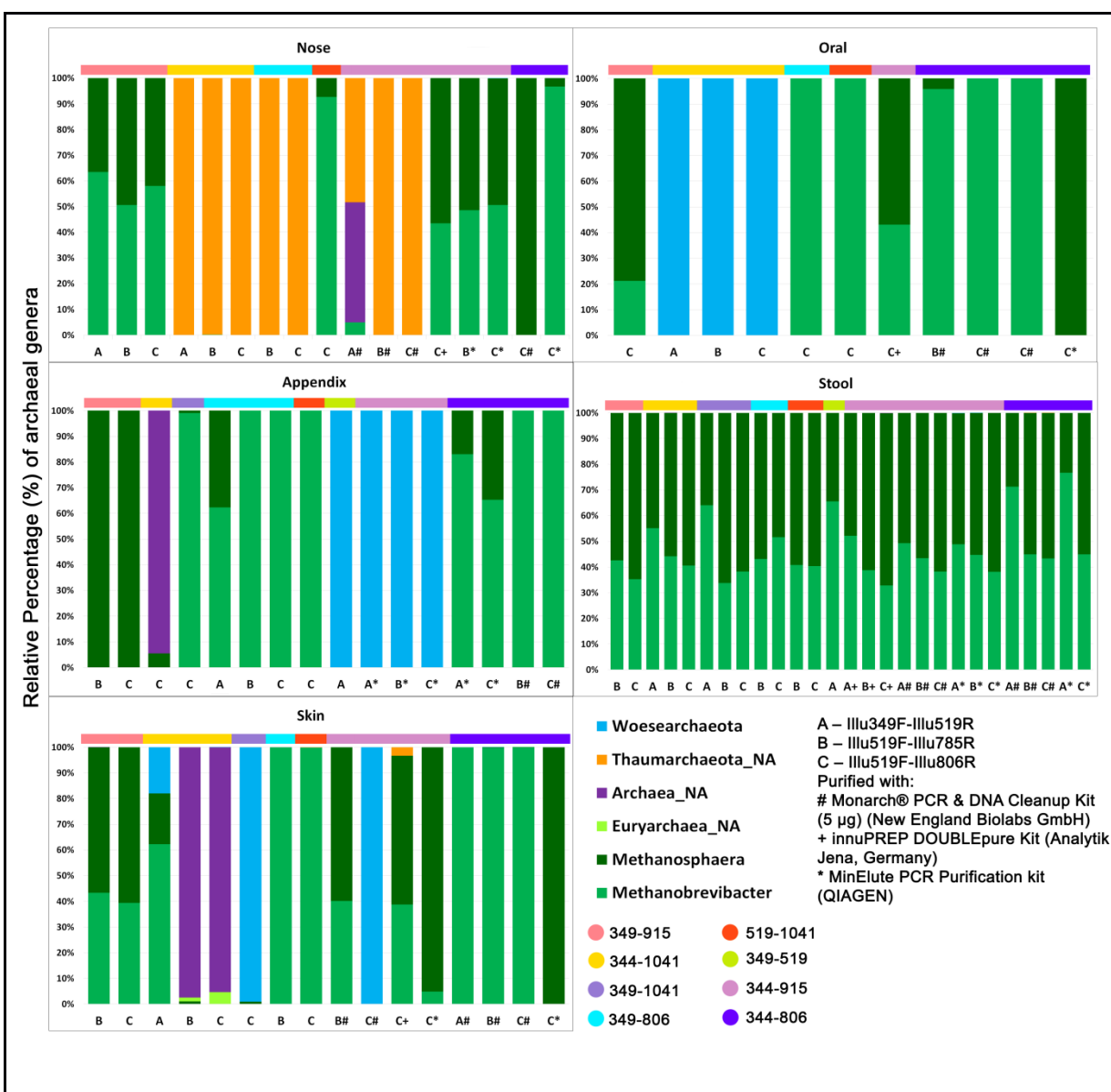
Fig. 2: Bar charts illustrating the diversity of archaeal communities on genus level in dependence of primer pair combinations. Primer combinations are illustrated via the colored bars above the columns (first primer pair), and by letters below the columns (second primer pair). In addition, information is given on the different purification kits that have been used between the two steps of the nested PCR. If no information is given next to the letter, MinElute PCR Purification kit (QIAGEN) has been used to purify the PCR product.

To further explore the influence of the primer pair selection on the archaeal community composition, the alpha diversity was calculated using the Shannon index (Fig. 3). For this analysis, we excluded the results obtained from the second primer pair 349F-519R as most samples herein (except stool samples) yielded less than 500 reads. To compare the alpha diversity within the results obtained we performed a Wilcoxon Rank Test to identify potential differences in the primer combination performances.

The highest archaeal diversity could be detected with the primer combination 344F-1041R/519F-806R (PCR34); this result was found to be significantly higher ($p < 0.05$) compared to PCR 33 (344F-1041R/519F-785R), PCR Q7 (344F-806R/519F-806R) and PCR M7 (344F-806R/519F-806R; see Table 2 and Fig. 3), whereas no other significant differences could be detected.

Fig. 3: Shannon index indicates the diversity received from different PCR approaches. The results have been plotted and grouped according to the first PCR used and the statistical significance (p-value <0.05; Wilcoxon Rank Test) is indicated by *. p-value < 0.05 for: C (344F-1041R) vs B (344F-1041R), C (344F-1041R) vs C (344F-806R QIAGEN) and C (344F-1041R) vs C (344F-806R BioLabs).
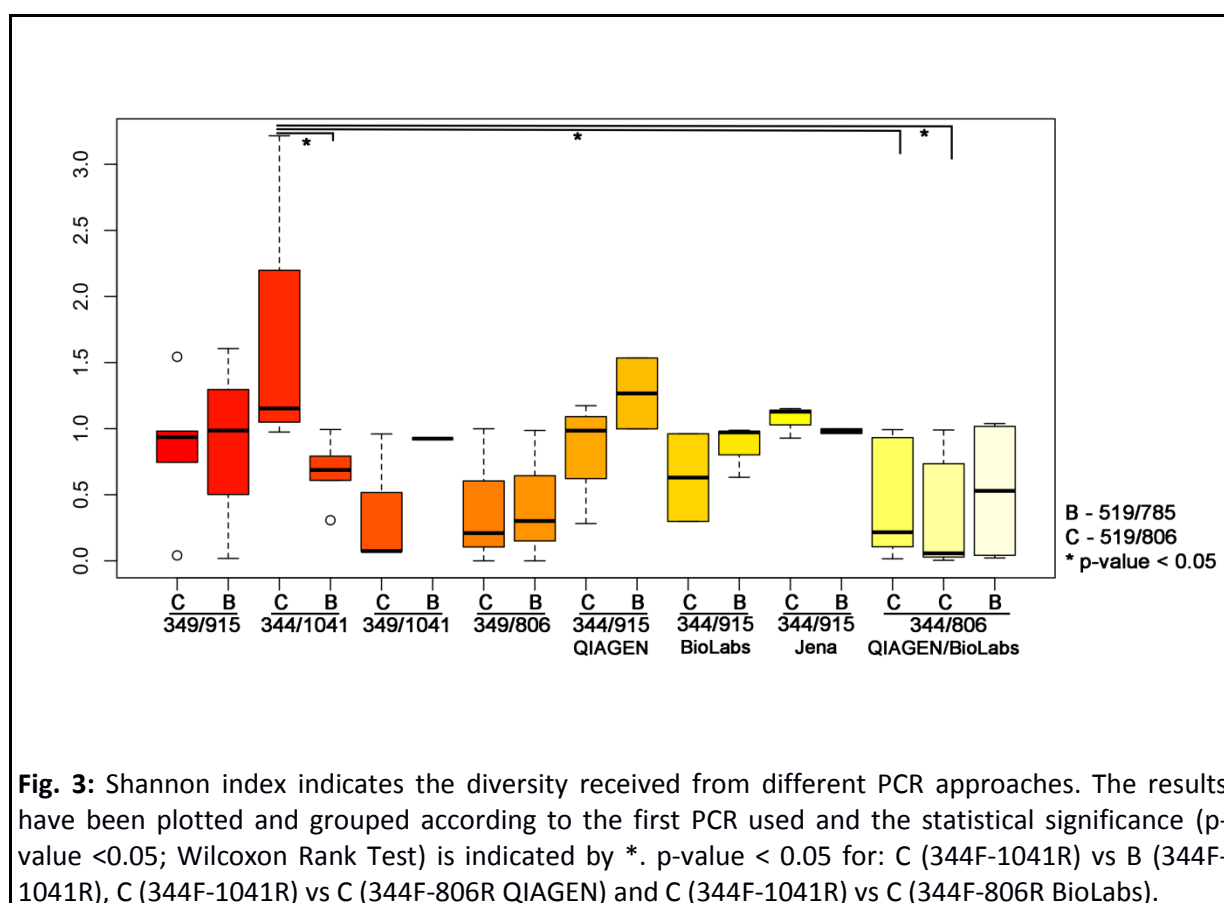
According to the comparison of the alpha diversity of the archaeal communities between the different primer pair combinations, we recommend the use of the nested approach with the primer

pair 344F-1041R in the first PCR, followed by a second PCR with the primers 519F-806R for studying and exploring the archaeal communities in human samples.

The primer combination with superior performance revealed a broad archaeal diversity in stool, appendix, nose, oral and skin samples

To further test and validate the use of the primer pair combination 344F-1041R/519F-806R for studying the archaeal communities within human samples, we selected additionally samples from the same body sites: nose (n=5), oral (n=6), appendix (n=5), stool (n=5), and skin (n=7) (sample set 2).

Our selected PCR approach allowed the detection of archaea in all samples investigated with an average of 102,366 reads and 8 observed RSVs for the nose, 56,480 reads and 35 observed RSVs for oral, 46,022 reads and 8 observed RSVs for the appendix, 93,948 reads and 4 observed RSVs for the stool sample, and 76,001 reads and 30 observed RSVs for the skin samples.

A summary of the number of archaeal, bacterial and eukaryotic reads/RSVs can be found in Supplementary Table 5 (provided on request). The results were plotted to indicate the archaeal communities present at genus level in the analyzed samples (Fig. 4).

Fig. 4: Bar chart displaying the different archaeal genera detected in different human samples using the superiorly performing primer combination 344F-1041R/519F-806R.
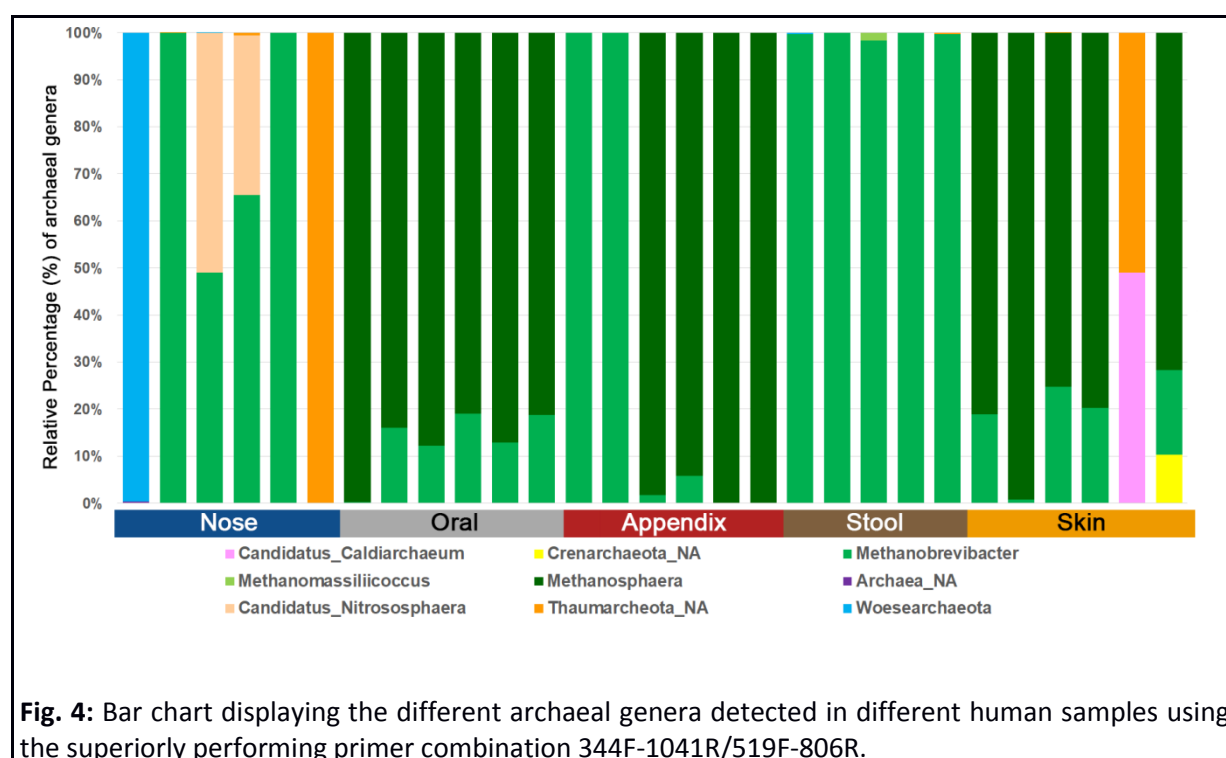
We further characterized the archaeal community information with respect to alpha and beta diversity. Depending on the body site a significant difference (p-value < 0.05) could be shown for alpha (Shannon index and richness) and beta diversity (PCoA and RDA) (Fig 5). Our results confirm the findings that archaeal communities are body site specific (Koskinen et al., 2017).

Notably, the stool samples revealed the overall lowest diversity of Archaea, with only 3-5 identified archaeal RSVs, while skin and oral samples contained a higher diversity, with 5 to 49 RSVs found in the skin samples and 14 to 49 RSVs in the oral samples.

Fig. 5: Alpha (a; Shannon index and richness) and beta diversity (b; PCoA and RDA) analyses of the obtained archaeal community information, based on primer combination 344F-1041R/519F-806R.
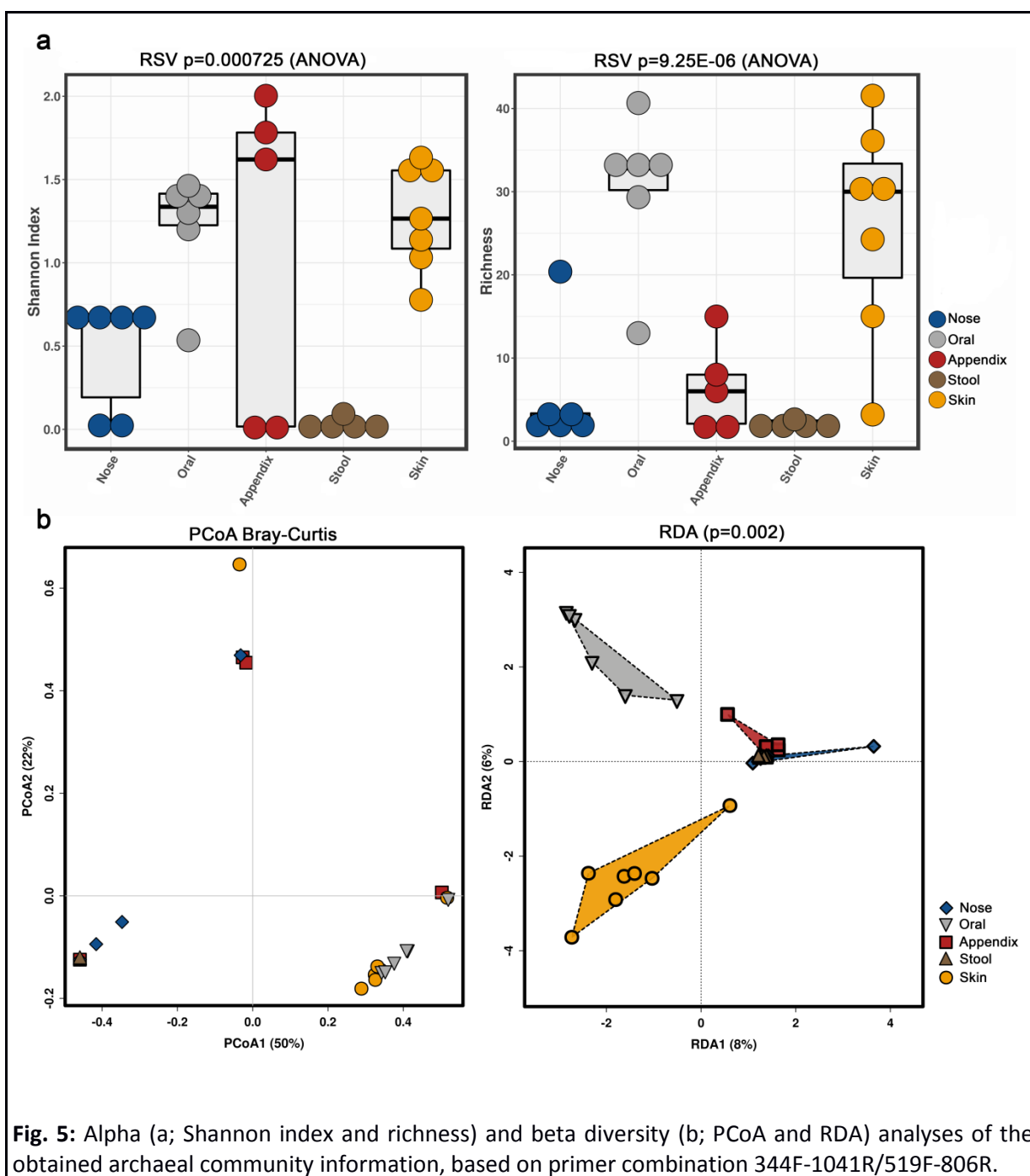
Optimization of quantitative PCR for determination of the archaea:bacteria ratio

The ratio of archaea vs. bacteria in human tissues and samples is widely unknown and can only be inadequately assessed from amplicon-based studies or metagenomics. It was our goal to set-up a suitable qPCR-based methodology to determine the ratios of bacterial and archaeal 16S rRNA gene contents in diverse samples.

Based on literature review, we selected three different set-ups for the quantification of archaeal 16S rRNA genes, and two different set-ups for the quantification of bacterial 16S rRNA genes. We included both chemistries, namely TaqMan and SYBR (Table 4). The same samples (sample set 1) that have been used for amplicon sequencing (nose, oral, appendix, stool and skin) have been tested for the different qPCR set-ups.

In silico analyses of the qPCR primers indicate a low coverage for archaea when no mismatch is considered, below 50% for two of the primers, but when one mismatch is considered the coverage increases from 71.2% to 91.1%. The coverage of the bacterial primers ranged from 75.3% to 90.5% when zero mismatches were considered and from 82.8% to 95.8% when one mismatch was allowed (Table 5).

For quantification of archaeal 16S rRNA genes, three different approaches were investigated, two SYBR approaches using two different primers pair: 806aF-958aR and 344aF-517uR, and one TaqMan approach using the primers pair 349aF-806aR and the probe (515F-FAM) (Hunter et al., 2002; Probst et al., 2013; Takai and Horikoshi, 2000).

All primer combinations showed qPCR results within the set limits for efficiency (between 90% - 110%). However, the SYBR protocols proved superior compared to the TaqMan approach, as this methodology allowed the quantification of archaeal signatures in only one out of five samples, namely the stool sample (Fig. 6a).

The SYBR 806aF-958aR allowed the quantification of archaeal 16S rRNA genes in 3 (stool, nose and appendix) out of 5 samples (Fig. 6a). However, no archaeal signatures could be detected in the oral and skin samples as signals were below the detection limit of 480 copies/µl.

The SYBR approach based on primer pair 344aF-517uR allowed the quantification of archaea in 4 out of 5 samples (Fig. 6a), with visible bands on the electrophoresis gel for stool, nose and appendix samples. Due to the high detection limit, we consider that this approach does not allow the detection of archaeal signals in samples with low biomass and the results might be overestimated.

Therefore, we suggest the use of the SYBR approach based on the primers pair 806aF-958aR for optimally detecting archaeal signatures within human samples. With this method we detected around 10^8 copies of archaeal genes in the stool sample, 10^5 copies in the nose sample and 10^6 copies in the appendix sample. Moreover, it was the most conservative approach, with the lowest chance to overestimate the archaeal abundance.

For the quantification of bacterial 16S rRNA genes, we used two different approaches, a SYBR approach using the primers pair: 338bF-517uR and a TaqMan approach using the primers pair: 331bF-797R and the probe 528bR labeled with HEX. Both methods allowed an adequate quantification of the bacterial load in all samples (Fig. 6b). The highest bacterial load was found in the stool samples, with around 10^{11} copies/total DNA (200mg sample), followed by the nose samples containing around 10^7 copies/total DNA. The appendix bacterial load was around 10^6 copies/ total DNA, and the oral sample had the lowest bacterial load with only 10^5 copies/total DNA. For the skin samples the absolute number is reported on the area sampled since the DNA could not be quantified using the Qubit HS detection method (Fig. 6).

Based on our experiments and their results, we propose to use primer pair 806aF-958aR for archaea-targeted SYBR qPCR.

a

Archaea

b
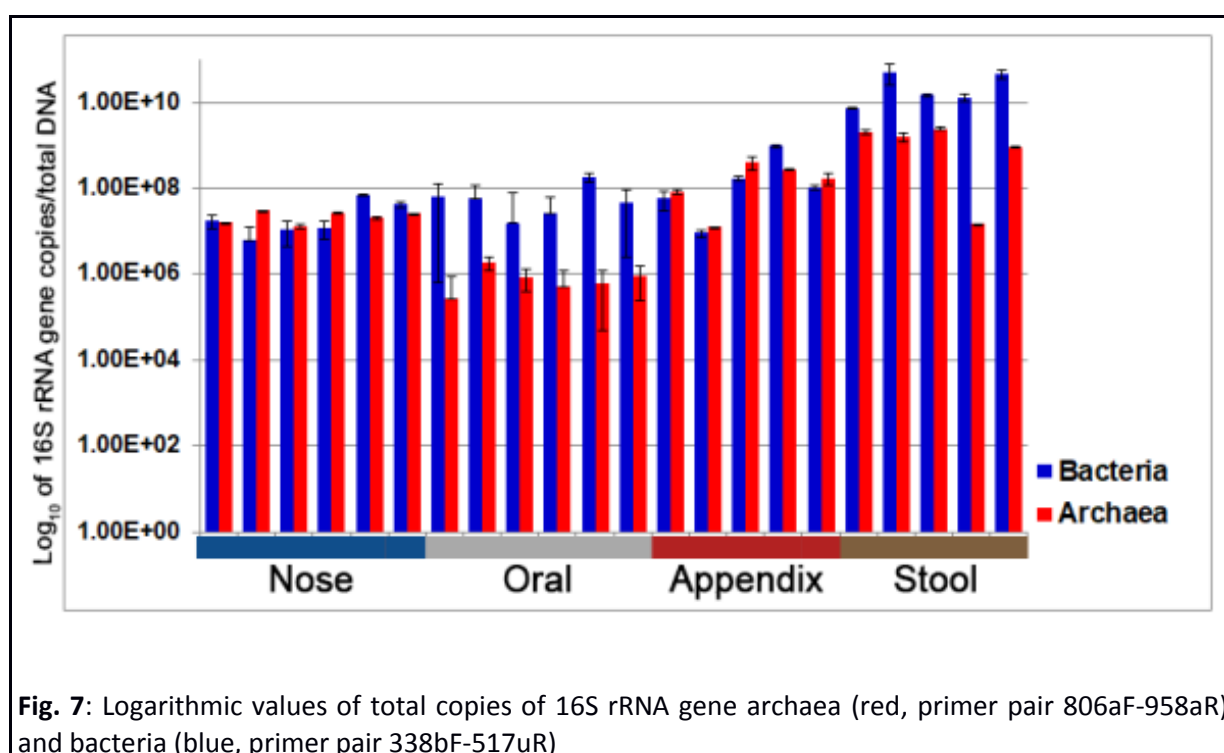
Bacteria

c
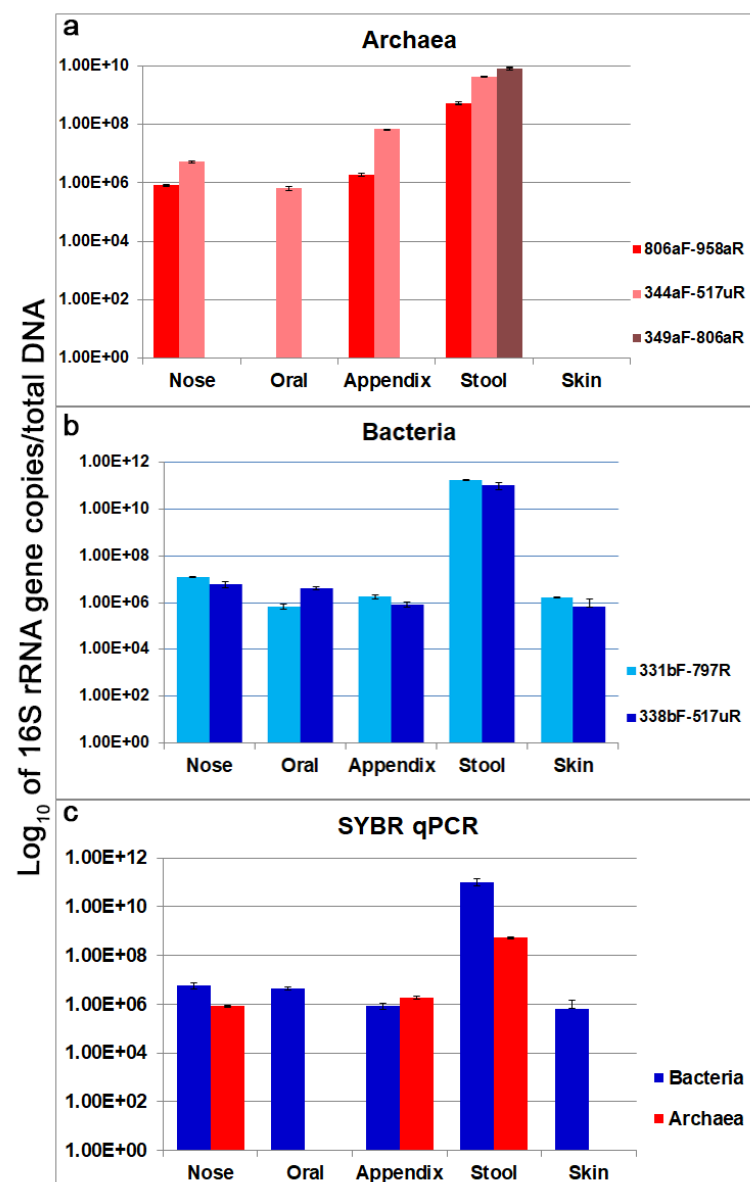
SYBR qPCR

Fig. 6. Logarithmic values of total copies of 16S rRNA gene for a. archaea, b. bacteria per total DNA, except for the skin samples where the results are per area sampled. c. the results from bacterial SYBR qPCR based on 338bF-517uR and archaeal SYBR qPCR with primers 806aF-958aR are plotted. The ratio between bacterial and archaeal 16S rRNA genes (determined on three technical replicates) varied depending on the sample.

Optimized quantitative PCR protocols reveal the presence of up to 50% archaeal signatures in certain human samples

The selected qPCR approach for archaea has been further tested using a set of different human samples (sample set 2). We selected nose (n=5), oral (n=6), appendix (n=5) and stool (n=5) samples to determine the absolute number of archaeal 16S rRNA genes and to assess the ratio between bacterial and archaeal 16S rRNA genes. Skin samples could not be included in the experiment due to their low DNA content.

For the archaeal qPCR we used the SYBR approach based on primer pair 806aF-958aR, and for the bacteria we used the SYBR qPCR approach with the primer pair 338bF-517uR.

The results are plotted in Fig. 7. The ratio between bacterial and archaeal 16S rRNA genes was determined using the average for each sample type. For stool samples the ratio was 20:1 (0.1 to 21.3% archaeal signatures). For nose and appendix, the ratio was around 1:1 (21.8 to 70.7 % archaeal signatures for appendix, and 22.8 to 82.8% for nose), and for oral samples the ratio was 77:1 (0.3 to 5.3 % archaeal signatures).

Fig. 7: Logarithmic values of total copies of 16S rRNA gene archaea (red, primer pair 806aF-958aR) and bacteria (blue, primer pair 338bF-517uR)

Discussion

Up to now, little it is known about the composition of the healthy human archaeome, or the real bacteria/archaea ratios in certain body parts. It is unknown, whether archaeal communities are affected by dysbiosis or human disease, or how we acquire these microorganisms after birth, although several studies have shown that archaea are present in the first year of life (Palmer et al., 2007; Wampach et al., 2017). Additionally, it is largely unexplored, how archaeal communities interact/communicate with other commensal microorganisms inhabiting the human body. Furthermore, there still remains the most burning question, if there are really no archaeal pathogens. Facing these numerous unsolved mysteries, we argue that more studies are needed with respect to the human archaeome. For these, however, standardized protocols are required, which are effective enough to reliably assess archaeal diversity and abundance based on 16S rRNA gene signatures.

To address the need for archaea-targeted amplicon method for NGS in human samples, we herein tested 12 different primers previously described in literature (Klindworth et al., 2013), in 27 primer pair combinations and evaluated their performance using *in silico* and experimental approaches on five different human sample types.

Despite their overall good *in silico* results, the three universal primer pairs tested failed to assess the archaeal diversity in the experiments. Two of these primer pairs represent the most used universal primers for amplicon sequencing methods (Caporaso et al., 2012; Walters et al., 2016), resulting in the detection of one (515F-806uR) or zero archaeal RSVs (515FB-806RB) in five sample types that evidentially possessed a variety of archaeal signatures. This was particularly intriguing, as the presence of archaeal signatures in the appendix and nose sample was confirmed by qPCR, with a ratio of 1:1 and 8:1 bacteria: archaea, respectively (sample set 1).

The reasons for the failure of the universal primers to detect Archaea are unclear; however, it seems bacterial signatures outcompete archaeal signatures, just due to slightly better primer matches, depending on the diversity within the sample.

Furthermore, an archaeal primer pair (519F-806R) that has been used before for amplicon sequencing (Siles et al., 2018) detected only a small proportion of the archaeal diversity in the analyzed samples, but the same primer pair performed better when used in a nested PCR together with the primer pair 344F-1041R for the first PCR.

Nested PCR has been shown to improve sensitivity and specificity and are useful for suboptimal DNA samples (Bomberg et al., 2003; Vissers et al., 2009). Based on our experience in the past (Koskinen et al., 2017), other reports (De Vrieze et al., 2018), and due to the fact that all attempts to use Illumina-tagged archaeal primers to directly identify archaeal 16S rRNA genes in human samples failed, we kept to this approach for the archaeal diversity assessment.

We used a combination of an archaea-specific first PCR (9 different primer combinations) and two archaeal specific and one universal primer pair, resulting in 24 different approaches (Table 2).

Notably, although the primer pair combinations 344F-915R/349F-519R and 344F-915R/519F-785R had been used earlier to detect archaeal signatures in human samples and confined environments (Koskinen et al., 2017) (Mora et al., 2016), our study revealed that when the second PCR contained the Illumina-tagged primers 349F-519R, almost no reads apart from the stool samples were retrieved (Suppl. Table 4a; available on request).

Ten out of the 24 different primer combinations allowed the detection of archaeal signatures in all analyzed samples (sample set 1). The results of two of the primer pair combinations were outstanding regarding the number of reads and observed RSVs identified in each sample (Supplementary Table 4a; available on request), namely primer pair 344F-1041R/519F-806R and 344F-1041R/519F-785R. The comparison of the alpha diversity (based on Shannon index) indicated that the archaeal diversity uncovered with the primer pair 344F-1041R/519F-806R was significantly higher than the one obtained with the primer pair combination 344F-1041R/519F-785R (Fig. 3), which was thus considered superior.

To further test and validate the use of the primer pair 344F-1041R/519F-806R, we selected 29 samples from different body sites (nose, oral, appendix, stool, skin; sample set 2), resulting in overall 85 archaeal RSVs from 6 different phyla. We were able to confirm body-site specificity through PCoA and RDA analysis (Koskinen et al., 2017), with the gastrointestinal tract (stool and appendix samples) being dominated by euryarchaeal communities, the oral samples dominated by archaeal communities from the Euryarchaeota phylum but different from the ones found in the gastrointestinal tract and the nose dominated by Euryarchaeota and Thaumarchaeota signatures. The skin revealed a mix of Euryarchaeota, Thaumarchaeota, Aenigmarchaeota, and, in very low amounts also Crenarcheota, confirming previous results (Koskinen et al., 2017; Moissl-Eichinger et al., 2017; Tsai et al., 2016).

According to the obtained results we recommend the use of the primer pair combination 344F-1041R/519F-806R to identify and characterize archaeal communities within human samples, even though the second primer pair 519F-806R is a universal primer pair according to the *in silico* results. Although this led to retrieval of not only archaeal reads, but also reads classified within Bacteria and Eukarya which had to be filtered bioinformatically, this procedure proved superior to all the other primer pairs tested in identifying archaeal signatures in the analyzed samples.

Another issue we addressed in this study is the need for a quantitative assay to determine the ratio of archaeal and bacterial signatures in human samples. Published qPCR methodology focuses on specific taxa, such as *Methanobrevibacter*, *Methanosphaera* and *Methanomassiliicoccus* (for more details see Table S1 in (Koskinen et al., 2017)), or has not yet been fully evaluated for various human samples (Hunter et al., 2002; Probst et al., 2013; Takai and Horikoshi, 2000). Our results revealed, that two qPCR approaches (SYBR 344aF-517uR and TaqMan 331bF-797R with probe 528bR) were unsuitable, due to the high detection limit, low efficiency, extraordinary long run times and potential over-estimation of the archaeal signatures. Therefore, we recommend the use of a SYBR approach based on primer pair 806aF-958aR, as this method allowed the quantification of archaeal 16S rRNA

genes in stool, appendix and nose samples, and had a low detection limit of 480 copies 16S rRNA genes/ μl.

In conclusion, we have shown that the choice of the archaeal primer pair influences substantially the perspective of the obtained archaeal community in the analyzed samples. Therefore, for future comparisons between studies focused on exploring and characterizing the archaeal community in human samples using amplicon sequencing approach, it should be considered to make use of the same, standardized methodology. For this we recommend the use of a nested approach with the primer pair 344f-1041R for the first PCR, followed by a second PCR with the primer pair 519F-806R. Furthermore, for quantifying the number of archaeal 16S rRNA gene copies we recommend the use of the SYBR approach based on the primer pair 806aF-958aR.

Conclusions

The optimized and evaluated protocol for archaeal signature detection and quantification can now be used for all human samples and might also be useful for samples from other environments and holobionts, such plants or animals.

Acknowledgements

References

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews* 59, 143–169. doi:10.1016/j.jip.2007.09.009.

Bomberg, M., Jurgens, G., Saano, A., Sen, R., and Timonen, S. (2003). Nested PCR detection of Archaea in defined compartments of pine mycorrhizospheres developed in boreal forest humus microcosms. *FEMS Microbiology Ecology* 43, 163–171. doi:10.1111/j.1574-6941.2003.tb01055.x.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods* 13, 581–3. doi:10.1038/nmeth.3869.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal* 6, 1621–1624. doi:10.1038/ismej.2012.8.

Chaudhary, P. P., Gaci, N., Borrel, G., O'Toole, P. W., and Brugère, J.-F. (2015). Molecular methods for studying methanogens of the human gastrointestinal tract: current status and future directions. *Applied Microbiology and Biotechnology* 99, 5801–5815. doi:10.1007/s00253-015-6739-2.

Clooney, A. G., Fouhy, F., Sleator, R. D., O' Driscoll, A., Stanton, C., Cotter, P. D., et al. (2016). Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis. *PLOS ONE* 11, e0148028. doi:10.1371/journal.pone.0148028.

de la Cuesta-Zuluaga, J., and Escobar, J. S. (2016). Considerations For Optimizing Microbiome Analysis Using a Marker Gene. *Frontiers in Nutrition* 3, 26. doi:10.3389/fnut.2016.00026.

De Vrieze, J., Pinto, A. J., Sloan, W. T., and Ijaz, U. Z. (2018). The active microbial community more accurately reflects the anaerobic digestion process: 16S rRNA (gene) sequencing as a predictive tool. *Microbiome* 6, 63. doi:10.1186/s40168-018-0449-9.

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proceedings of the National Academy of Sciences of the United States of America* 89, 5685–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1608980 [Accessed May 16, 2018].

Ding, T., and Schloss, P. D. (2014). Dynamics and associations of microbial community types across the human body. *Nature* 509, 357–360. doi:10.1038/nature13178.

Eisenstein, M. (2018). Microbiology: making the best of PCR bias. *Nature Methods* 15, 317–320. doi:10.1038/nmeth.4683.

el Fantroussi, S., Verschuere, L., Verstraete, W., and Top, E. M. (1999). Effect of phenylurea herbicides on soil microbial communities estimated by analysis of 16S rRNA gene fingerprints and community-level physiological profiles. *Applied and environmental microbiology* 65, 982–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10049851 [Accessed May 16, 2018].

Eloe-Fadrosh, E. A., Ivanova, N. N., Woyke, T., and Kyrpides, N. C. (2016). Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nature Microbiology* 1, 15032. doi:10.1038/nmicrobiol.2015.32.

Fischer, M. A., Güllert, S., Neulinger, S. C., Streit, W. R., and Schmitz, R. A. (2016). Evaluation of 16S rRNA Gene Primer Pairs for Monitoring Microbial Community Structures Showed High

Reproducibility within and Low Comparability between Datasets Generated with Multiple Archaeal and Bacterial Primer Pairs. *Frontiers in microbiology* 7, 1297. doi:10.3389/fmicb.2016.01297.

Fontana, A., Patrone, V., Puglisi, E., Morelli, L., Bassi, D., Garuti, M., et al. (2016). Effects of geographic area, feedstock, temperature, and operating time on microbial communities of six full-scale biogas plants. *Bioresource Technology* 218, 980–990. doi:10.1016/J.BIORTECH.2016.07.058.

Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W., and Brugère, J. F. (2014). Archaea and the human gut: New beginning of an old story. *World Journal of Gastroenterology* 20, 16062–16078. doi:10.3748/wjg.v20.i43.16062.

Gilbert, J. A., Jansson, J. K., and Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC Biology* 12, 69. doi:10.1186/s12915-014-0069-1.

Gill, S., Pop, M., DeBoy, R., and Eckburg, P. (2006). Metagenomic analysis of the human distal gut microbiome. *Science* 312, 1355–1359. doi:10.1126/science.1124234.Metagenomic.

Grine, G., Boualam, M. A., and Drancourt, M. (2017). Methanobrevibacter smithii, a methanogen consistently colonising the newborn stomach. *European Journal of Clinical Microbiology & Infectious Diseases* 36, 2449–2455. doi:10.1007/s10096-017-3084-7.

Halwachs, B., Madhusudhan, N., Krause, R., Nilsson, R. H., Moissl-Eichinger, C., Högenauer, C., et al. (2017). Critical Issues in Mycobiota Analysis. *Frontiers in Microbiology* 8, 180. doi:10.3389/fmicb.2017.00180.

Horz, H.-P., and Conrads, G. (2011). Methanogenic *Archaea* and oral infections – ways to unravel the black box. *Journal of Oral Microbiology* 3, 5940. doi:10.3402/jom.v3i0.5940.

Hunter, N., Nadkarni, M. A., Jacques, N. A., and Martin, F. E. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* 148, 257–266. doi:10.1099/00221287-148-1-257.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* 41, e1–e1. doi:10.1093/nar/gks808.

Klymiuk, I., Bambach, I., Patra, V., Trajanoski, S., and Wolf, P. (2016). 16S Based Microbiome Analysis from Healthy Subjects' Skin Swabs Stored for Different Storage Periods Reveal Phylum to Genus Level Changes. *Frontiers in microbiology* 7, 2012. doi:10.3389/fmicb.2016.02012.

Koskinen, K., Pausan, M. R., Perras, A. K., Beck, M., Bang, C., Mora, M., et al. (2017). First insights into the diverse human archaeome: Specific detection of Archaea in the gastrointestinal tract, lung, and nose and on skin. *mBio* 8. doi:10.1128/mBio.00824-17.

Koskinen, K., Reichert, J. L., Hoier, S., Schachenreiter, J., Duller, S., Moissl-Eichinger, C., et al. (2018). The nasal microbiome mirrors and potentially shapes olfactory function. *Scientific Reports* 8, 1296. doi:10.1038/s41598-018-19438-3.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33, 1870–1874. doi:10.1093/molbev/msw054.

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128. doi:10.1093/bioinformatics/btl529.

Methé, B. A., Nelson, K. E., Pop, M., Creasy, H. H., Giglio, M. G., Huttenhower, C., et al. (2012). A framework for human microbiome research. *Nature* 486, 215–221. doi:10.1038/nature11209.

Moissl-Eichinger, C., Pausan, M., Taffner, J., Berg, G., Bang, C., and Schmitz, R. A. (2018). Archaea Are Interactive Components of Complex Microbiomes. *Trends in Microbiology* 26, 70–85. doi:10.1016/j.tim.2017.07.004.

Moissl-Eichinger, C., Probst, A. J., Birarda, G., Auerbach, A., Koskinen, K., Wolf, P., et al. (2017). Human age and skin physiology shape diversity and abundance of Archaea on skin. *Submitted*.

Mora, M., Perras, A., Alekhova, T. A., Wink, L., Krause, R., Aleksandrova, A., et al. (2016). Resilient microorganisms in dust samples of the International Space Station—survival of the adaptation specialists. *Microbiome* 4, 65. doi:10.1186/s40168-016-0217-7.

Nkamga, V. D., Henrissat, B., and Drancourt, M. (2017). Archaea: Essential inhabitants of the human digestive microbiota. *Human Microbiome Journal* 3, 1–8. doi:10.1016/j.humic.2016.11.005.

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A., and Brown, P. O. (2007). Development of the Human Infant Intestinal Microbiota. *PLoS Biology* 5, e177. doi:10.1371/journal.pbio.0050177.

Peng, Q., Vijaya Satya, R., Lewis, M., Randad, P., and Wang, Y. (2015). Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC genomics* 16, 589. doi:10.1186/s12864-015-1806-8.

Probst, A. J., Auerbach, A. K., and Moissl-Eichinger, C. (2013). Archaea on Human Skin. *PLoS ONE* 8, e65388. doi:10.1371/journal.pone.0065388.

Pruesse, E., Peplies, J., and Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics (Oxford, England)* 28, 1823–9. doi:10.1093/bioinformatics/bts252.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41, D590–D596. doi:10.1093/nar/gks1219.

R Core Team (2013). R: A language and environment for statistical computing. Available at: http://www.r-project.org/.

Raskin, L., Stromley, J. M., Rittmann, B. E., and Stahl, D. A. (1994). Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Applied and environmental microbiology* 60, 1232–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7517128 [Accessed February 20, 2017].

Raymann, K., Moeller, A. H., Goodman, A. L., and Ochman, H. (2017). Unexplored Archaeal Diversity in the Great Ape Gut Microbiome. *mSphere* 2, e00026-17. doi:10.1128/mSphere.00026-17.

Santigli, E., Koller, M., and Klug, B. (2017). Oral Biofilm Sampling for Microbiome Analysis in Healthy Children. *Journal of Visualized Experiments*. doi:10.3791/56320.

Santigli, E., Trajanoski, S., Eberhard, K., and Klug, B. (2016). Sampling Modification Effects in the Subgingival Microbiome Profile of Healthy Children. *Frontiers in microbiology* 7, 2142. doi:10.3389/fmicb.2016.02142.

Seed, P. C. (2014). The human mycobiome. *Cold Spring Harbor perspectives in medicine* 5, a019810. doi:10.1101/cshperspect.a019810.

Siles, J. A., Öhlinger, B., Cajthaml, T., Kistler, E., and Margesin, R. (2018). Characterization of soil

bacterial, archaeal and fungal communities inhabiting archaeological human-impacted layers at Monte Iato settlement (Sicily, Italy). *Scientific Reports* 8, 1903. doi:10.1038/s41598-018-20347-8.

Takai, K., and Horikoshi, K. (2000). Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Applied and environmental microbiology* 66, 5066–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11055964 [Accessed February 20, 2017].

Tsai, Y. C., Conlan, S., Deming, C., Nisc Comparative Sequencing Program, Segre, J. A., Kong, H. H., et al. (2016). Resolving the complexity of human skin metagenomes using single-molecule sequencing. *mBio*. doi:10.1128/mBio.01948-15.

Tyakht, A. V., Kostryukova, E. S., Popenko, A. S., Belenikin, M. S., Pavlenko, A. V., Larin, A. K., et al. (2013). Human gut microbiota community structures in urban and rural populations in Russia. *Nature Communications* 4, 2469. doi:10.1038/ncomms3469.

van de Pol, J. A. A., van Best, N., Mbakwa, C. A., Thijs, C., Savelkoul, P. H., Arts, I. C. W., et al. (2017). Gut Colonization by Methanogenic Archaea Is Associated with Organic Dairy Consumption in Children. *Frontiers in microbiology* 8, 355. doi:10.3389/fmicb.2017.00355.

Vissers, E. W., Bodelier, P. L. E., Muyzer, G., and Laanbroek, H. J. (2009). A nested PCR approach for improved recovery of archaeal 16S rRNA gene fragments from freshwater samples. *FEMS Microbiology Letters* 298, 193–198. doi:10.1111/j.1574-6968.2009.01718.x.

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., et al. (2016). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems* 1, e00009-15. doi:10.1128/mSystems.00009-15.

Wampach, L., Heintz-Buschart, A., Hogan, A., Muller, E. E. L., Narayanasamy, S., Laczny, C. C., et al. (2017). Colonization and Succession within the Human Gut Microbiome by Archaea, Bacteria, and Microeukaryotes during the First Year of Life. *Frontiers in Microbiology* 8, 738. doi:10.3389/fmicb.2017.00738.

Zakrzewski, M., Proietti, C., Ellis, J. J., Hasan, S., Brion, M.-J., Berger, B., et al. (2016). Calypso: a user-friendly web-server for mining and visualizing microbiome–environment interactions. *Bioinformatics*, btw725. doi:10.1093/bioinformatics/btw725.

Zhang, Y., Chen, L., Sun, R., Dai, T., Tian, J., Liu, R., et al. (2014). Effect of wastewater disposal on the bacterial and archaeal community of sea sediment in an industrial area in China. *FEMS Microbiology Ecology* 88, 320–332. doi:10.1111/1574-6941.12298.

Zou, S., Caler, L., Colombini-Hatch, S., Glynn, S., and Srinivas, P. (2016). Research on the human virome: where are we and what is next. *Microbiome* 4, 32. doi:10.1186/s40168-016-0177-y.