

1 **Evaluation of primer pairs for microbiome profiling across a food chain from soils to**
2 **humans within the One Health framework**

3

4 Wasimuddin¹, Klaus Schlaeppli², Francesca Ronchi³, Stephen L Leib¹, Matthias Erb², Alban
5 Ramette^{1,*}

6

7 ¹Institute for Infectious Diseases, University of Bern, Bern, Switzerland

8 ²Institute of Plant Sciences, University of Bern, Bern, Switzerland

9 ³Department for Biomedical Research, University of Bern, Inselspital, 3008 Bern, Switzerland

10

11 *Correspondence: Alban Ramette, University of Bern, Institute for Infectious Diseases,
12 Friedbühlstrasse 51, CH-3001 Bern, Switzerland. alban.ramette@ifik.unibe.ch

13

14 **Key words:** One Health, microbiome, 16S rRNA gene, high throughput sequencing, soil, plant
15 root, cow rumen, human gut

16 **Abstract**

17 The "One Health" framework emphasizes the ecological relationships between soil, plant,
18 animal and human health. Microbiomes play important roles in these relationships, as they
19 modify the health and performance of the different compartments and influence the transfer of
20 energy, matter and chemicals between them. Standardized methods to characterize
21 microbiomes along food chains are, however, currently lacking. To address this methodological

22 gap, we evaluated the performance of DNA extractions kits and commonly recommended
23 primer pairs targeting different hypervariable regions (V3-V4, V4, V5-V6, V5-V6-V7) of the
24 16S rRNA gene, on microbiome samples along a model food chain, including soils, maize roots,
25 cattle rumen, and cattle and human faeces. We also included faeces from gnotobiotic mice
26 colonized with defined bacterial taxa and mock communities to confirm the robustness of our
27 molecular and bioinformatic approaches on these defined low microbial diversity samples.
28 Based on Amplicon Sequence Variants, the primer pair 515F-806R led to the highest estimates
29 of species richness and diversity in all sample types and offered maximum diversity coverage
30 of reference databases in *in silico* primer analysis. The influence of the DNA extraction kits
31 was negligible compared to the influence of the choice of primer pairs. Comparing microbiomes
32 using 515F-806R revealed that soil and root samples have the highest estimates of species
33 richness and inter-sample variation. Species richness decreased gradually along the food chain,
34 with the lowest richness observed in human faeces. Primer pair choice directly influenced the
35 estimation of community changes (beta diversity) within and across compartments and may
36 give rise to preferential detection of specific taxa. This work demonstrates why a standardized
37 approach is necessary to analyse microbiomes within and between source compartments along
38 food chains in the context of the One Health framework.

39

40 **Introduction**

41 The "One Health" concept emphasizes the ecological relationships and interdependencies
42 between humans, plants, animals and environmental health (Destoumieux-Garzon et al., 2018).
43 Until recently, the One Health concept primarily focused on the origin and transfer of zoonotic
44 pathogens, vectors of pathogens and antibiotic resistance between interacting entities
45 (Destoumieux-Garzon et al., 2018). During the past decade, however, microbial communities
46 (or microbiomes) have been shown to play important roles in connecting the humans, plants,

47 animals and environment (van Bruggen et al., 2019). Thus, recommendations have been made
48 to extend the One Health concept to include the full breadth of microbes (Bell et al., 2018; Trinh
49 et al., 2018; van Bruggen et al., 2019). Adopting a microbiome perspective may strengthen the
50 One Health concept due to i) the vital services provided by microbiomes to overall ecosystem
51 health, ii) the importance of microbiome processes for the transfer of energy, matter and
52 chemicals between compartments along the food chain, and iii) the important contribution of
53 microbiomes to the health of the different hosts and compartments. However, methodological
54 challenges remain to adequately characterize and allow comparison of the different
55 microbiomes in order to track microbial transfer and to quantify the role of microbiomes in food
56 chain health (Trinh et al., 2018).

57 Widely established approaches enable quantifying the diversity and richness of microbiomes
58 with high resolution from diverse source compartments by sequencing the 16S rRNA marker
59 gene amplified by ‘universal’ primers (Fricker et al., 2019). In order to meet the sequence length
60 requirement of short-read sequencing technologies, various primer pairs have been designed to
61 amplify short hypervariable regions of the 16S rRNA gene. Both, choice of hypervariable
62 region of 16S and primer pair, influence the description of microbial diversity (Claesson et al.,
63 2010). Thus, care should be taken in choosing appropriate primer pairs, as limited taxa
64 coverage, over- or underrepresentation of taxa in a specific environment due to biases in primer
65 amplification could produce unreliable results (Claesson et al., 2010). Free-living microbial
66 communities such as those in soil and lake sediment may exhibit higher microbial richness than
67 host-associated communities such as animal gut microbiomes (Thompson et al., 2017).
68 Furthermore, among host associated communities, plant roots show higher microbial richness
69 than other plant and animal associated microbial communities (Thompson et al., 2017). Because
70 each source compartment is unique in terms of microbial richness and composition (Thompson
71 et al., 2017), in studies where microbiomes from different source compartments are

72 investigated, such as within the One Health framework, investigators should carefully select
73 primers to avoid methodological biases and to maximize the detection of taxa (Trinh et al.,
74 2018). To our knowledge, no study has systematically included diverse samples from different
75 compartments in a primer comparison experiment. Thus, the key information about choice of
76 primer pairs required to conduct One Health experiments is missing.

77 The prokaryotic primer pair 515F-806R, which was designed to detect both archaea and bacteria
78 by amplifying V4 region, is recommended by the Earth Microbiome Project and has been
79 extensively used to study soil microbiomes (Apprill et al., 2015;Parada et al., 2016;Walters et
80 al., 2016). Nevertheless, two recent studies that evaluated the best performing primer pair based
81 on taxa diversity coverage (Klindworth et al., 2013;Thijs et al., 2017), recommend the primer
82 pair 341F-805R, which amplifies the V3-V4 region, over other primer pairs. In the *Klindworth*
83 *et al.* study, 512 primer pairs were tested *in silico* against the SILVA v108 database (376,437
84 sequences) for amplification of archaeal and bacterial sequences. Yet, since that study 318,734
85 additional sequences have been added to the latest SILVA release v132, which almost doubled
86 the size of the SILVA database. Hence, the previously characterized primer coverages should
87 be re-examined using the enhanced current database. The study by *Thijs et al.* used both *in*
88 *silico*, as well laboratory experiments, to access the best primer pair but did not include the
89 515F-806R primer pair in their comparisons and performed rather shallow sequencing (454
90 pyrosequencing) of soil samples. Several studies target exclusively bacteria to answer specific
91 questions (Mazmanian et al., 2005;Hebbandi Nanjundappa et al., 2017), thus bacterial specific
92 primer pairs should be included as well in such comparisons. Two primer pairs; 799F-1193R
93 (V5-V6-V7) and 787F-1073R (V5-V6) have been preferably used in compartment specific
94 studies, primer pair 799F-1193R in plant system due to reduced amplification of plant organelle
95 DNA (Beckers et al. 2016) and 787F-1073R in mouse studies with limited diversity
96 microbiomes (Li et al., 2015;Hebbandi Nanjundappa et al., 2017). Furthermore, the Divisive

97 Amplicon Denoising Algorithm 2 (*DADA2*) introduced a model-based approach for identifying
98 sequencing errors without the need of constructing OTUs and at the same time for detecting
99 less false positives in comparison to earlier methods (Callahan et al., 2016). In light of the
100 advancement in high-throughput sequencing methods and in high-resolution analysis methods,
101 the choice of primer pairs should be re-examined in order to achieve higher taxonomic
102 coverage.

103 In this study, our objective was to evaluate the performance of four commonly used primer
104 pairs 787F-1073R, 799F-1193R, 515F-806R and 341F-805R for 16S amplicon sequencing
105 especially within a One Health framework by including microbial communities from different
106 source compartments along the human food chain; including samples from soil, plant, mouse,
107 cattle and humans. We then used the best performing method to gain first insights into the
108 commonalities and differences between the microbiomes along a model food chain.

109

110 **Material and Methods**

111 **Sample collection**

112 Samples were collected from four different source compartments with the aim to maximize the
113 heterogeneity within compartment in the experiment (Supplementary Table 1). Briefly, five soil
114 samples, each from a different soil type, six maize root samples from three different
115 geographical locations, five cow samples including three faeces and two rumen samples, six
116 human faeces samples from volunteers belonging to two couples, one child (3 years of age) and
117 one female sample, were collected. Additionally, faeces from a gnotobiotic mouse strain
118 colonized with defined microbial community (four bacteria of the Altered Schaedler's Flora
119 here referred as ASF.4: *Lactobacillus_acidophilus*_ASF360, *Lactobacillus_murinus*_ASF361,
120 *Clostridium_sp*_ASF500, *Bacteroides_distasonis*_ASF519) and a mock microbial community

121 DNA (8 bacterial + 2 yeasts species mixed in defined proportions) (ZymoBIOMICS Microbial
122 Community DNA standard, Zymo Research, USA) were included in the experiments. All
123 samples were stored at -80°C until further analysis.

124

125 **Bacterial DNA extraction and 16S rRNA gene amplicon sequencing**

126 Samples were homogenized by bead beating at 50Hz for four minutes using a TissueLyser LT
127 (QIAGEN, Germany). Genomic bacterial DNA was extracted from all samples by using
128 DNeasy PowerSoil Pro kit (QIAGEN, Germany) according to the manufacturer's instructions.
129 As samples from diverse source compartments were included in the planned experiment, we
130 additionally extracted bacterial DNA using kits, which are generally used for the particular
131 source compartment, in order to examine the DNA extraction kit effect on source compartment
132 microbiome (Knauth et al., 2013;Lim et al., 2018). We extracted soil and root samples with the
133 NucleoSpin Soil DNA extraction kit (Macherey-Nagel, Germany) according to the
134 manufacturer's instructions (Knauth et al., 2013). Likewise, mouse faecal samples were
135 extracted with the QIAamp DNA FAST Stool Mini Kit (QIAGEN, Germany) (Lim et al., 2018)
136 following manufacturer's instructions, however an additional step of lysozyme treatment was
137 added as reported previously (Mamantopoulos et al., 2017).

138 Four primer pairs, namely 787F-1073R, 799F-1193R, 515F-806R and 341F-805R, were used
139 to amplify the V5-V6, V5-V6-V7, V4, V3-V4 hypervariable regions of the 16S rRNA gene,
140 respectively (Table1). Forward primers and reverse primers carried overhang adapters (5'
141 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-Forward primer, 5'
142 GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-Reverse primer) for compatibility
143 with Illumina index and sequencing adapters. A two-round amplification process was used to
144 amplify the DNA samples, while reducing dimer formation, which is often the problem in multi-

145 primer, multi-template PCR, especially with primers containing long overhang regions (Kalle
146 et al., 2014). Amplicon PCR reactions were carried out using the Faststart PCR system (Roche,
147 Switzerland). The 25- μ l PCR mix was composed of 3 ng/ μ l DNA, 1 \times FastStart PCR grade
148 nucleotide mix buffer without MgCl₂, 4.5 nM MgCl₂, 200 μ M each of PCR grade nucleotides,
149 0.05 U/ μ l Fast Start Taq DNA Polymerase, 400 nM target-specific primers, 5% DMSO and 9
150 μ l of PCR certified water. PCR cycling conditions consisted of an initial activation step at 95°
151 C for 3 min, followed by 32 cycles with denaturation at 95° C for 30 s, annealing at 62° C for
152 30 s, extension at 72° C for 30 s and final extension at 72° C for 10 min. PCR products were
153 subsequently purified using SPRI based size selection (Beckman Coulter Genomics, USA) and
154 quantified using Qubit 2.0 Fluorometer. Equal amount of first round purified PCR products
155 were used as templates for the second round indexing PCR using Nextera XT Index kit
156 (Illumina USA). Briefly, 50 μ l of reaction mix consisted of 5 μ l of first round PCR product (2.5
157 ng/ μ l), 5 μ l of Nextera XT Index Primer 1, 5 μ l of Nextera XT Index Primer 2, 25 μ l of MyFi
158 Mix (2x) (Bioline, Meridian Bioscience, France) and 10 μ l of PCR certified water. Indexing
159 PCRs cycling conditions were according to standard Illumina 16S metagenomic sequencing
160 library preparation protocol (<https://support.illumina.com/>). Second-round amplicon libraries
161 were purified using SPRI based size selection (Beckman Coulter Genomics, USA) and
162 quantified using Fragment Analyzer (Agilent, USA). The final pooled libraries were paired-end
163 sequenced (2x300 cycles) in a single run on Illumina MiSeq at the NGS platform of University
164 of Bern (www.ngs.unibe.ch). Negative controls were included in both the DNA extraction (no
165 DNA template added) and 16S PCR amplification (with PCR certified water) to test for
166 contamination. No noticeable DNA contamination of the negative controls after PCR
167 amplification was observed during quantification using Qubit 2.0 fluorometer and by Fragment
168 Analyzer.

169

Primer Name	Reference	Direction	Sequence 5'-3'	Length [bp]	Region	Amplicon size	Domain universality*				Coverage weighted score**			Coverage weighted score (pair)**			
							A+B	A+E	B+E	A+B+E	A	B	E	A	B	E	
787F (P1)	Sundquist et al., 2007	forward	ATTAGATACCCYGGTAGTCC	20	V5-V6	286	+					96.02	97.89	0.98	49.14	96.91	0
1073R (P1)	Sundquist et al., 2007	reverse	ACGAGCTGACGACARCCATG	20								50.58	98.91	0			
799F (P2)	Chelius and Triplett, 2001	forward	AACMGGATTAGATACCCCKG	19	V5-V6-V7	394	+					85.56	87.59	0.39	0	86.04	0
1193R (P2)	Bodenhausen et al., 2013	reverse	ACGTCATCCCCACCTTCC	18								0	97.92	0.01			
515F (P3)	Parada et al. 2015	forward	GTGYCAGCMGCCGCGGTAA	19	V4	291				+		97.79	98.48	94.66	96.39	96.2	16.63
806R (P3)	Apprill et al. 2015	reverse	GGACTACNVGGGTWTCTAAT	20						+		98.43	97.5	17.28			
341F (P4)	Herlemann et al 2011	forward	CCTACGGGNGGCWGCAG	17	V3-V4	444	+					84.56	98.35	2.87	83.59	96.69	0.12
805R (P4)	Herlemann et al 2011	reverse	GACTACHVGGGTATCTAATCC	21						+		98.4	98.17	0.85			

*Domain universality is assigned based on 80% domain coverage criteria for at least two domains.

***In silico* primer coverage weighted score was predicted using primer-prospector analysis in this study

Table 1. Details of the primers used in the current study and their *in silico* evaluation

170 **Bioinformatics**

171 Demultiplexed reads without barcodes and adapters were received as output from the
172 sequencing centre. All subsequent analyses were performed within the R environment (R
173 version 3.5.1, R Development Core Team 2011). For data pre-processing, we followed the
174 *DADA2* pipeline (version 1.10; Callahan et al., 2016) by adjusting parameters to each of the
175 four primer pair datasets. For each primer pair dataset, reads were trimmed from both ends
176 based on quality profile, error rates were learned from the data using the parametric error model
177 as implemented in *DADA2*. After denoising and merging, chimeric sequences (bimera) were
178 removed from the datasets by following the ‘consensus’ method as implemented in *DADA2*.
179 The final table thus consisted of a tabulation of number of occurrences of non-chimeric
180 amplicon sequence variants (ASVs; i.e. sequence differing by as little as one nucleotide) in each
181 sample. Taxonomy assignments of representative ASVs were performed using the naïve
182 Bayesian classifier method with the latest SILVA v132 non-redundant (NR) database. SILVA
183 database was chosen because it also contains eukaryotic sequences, which would be helpful to
184 determine whether the primer pairs also amplify eukaryotic sequences. Species level
185 assignment was done by exact matching (100% identity) of ASVs with database sequences, as
186 previously recommended (Edgar, 2018). *Phyloseq* (version 1.24.2, McMurdie and Holmes,
187 2013) package was used for further data processing, and ASVs belonging to chloroplast,
188 mitochondria, and unassigned ASVs at phylum level were removed from the dataset. We
189 merged ASV read abundance profiles based on their phylum, genus and species level
190 assignments to analyse the microbiome diversity across the four datasets produced by the
191 different primer sets.

192

193 **Alpha, beta diversity analysis and differential species abundance**

194 We investigated the effects of DNA extraction kits, primer pair, and source compartment on
195 microbial diversity for each sample by using two different alpha diversity indices (number of
196 observed species and Shannon) after rarefying the data to 10,300 sequences per sample using
197 *Phyloseq*. To analyse the association of DNA extraction kits, primer pair and source
198 compartment with these alpha diversity matrices, we performed General Linear Modelling
199 (GLM) by using the *lme4* package in R (Bates et al., 2015). We included primer pairs (787F-
200 1073R, n= 48; 799F-1193R, n= 46; 515F-806R, n= 46 and 341F-805R, n= 47), source
201 compartment (mock; n= 20, mouse; n= 39, soil; n= 38, root; n= 48, cow; n= 20 and human; n=
202 22), DNA extraction kit (source specific kits; n=61, Powersoil; n=126) and interaction between
203 primer pair and source compartment (primer pair*source compartment) in the model as
204 explanatory variables for each alpha diversity metric table.

205 Beta diversity analyses were based on calculated Jaccard and Bray-Curtis dissimilarity matrices
206 after rarefying the data to 10,300 sequences per sample by using *Phyloseq*. The permutational
207 multivariate analysis of variance (PERMANOVA) was employed as implemented in the *adonis*
208 function of the *vegan* package (version 2.5-2; Oksanen et al., 2019) to test the significance of
209 the differences in community composition with 999 permutations. For both beta diversity
210 metrics, we similarly included DNA extraction kit, primer pairs, source compartment, and
211 interaction between primer pair and source compartment (primer pair*source compartment) in
212 the models as explanatory variables. To visualize patterns of separation between different
213 sample categories, non-metric multidimensional scaling (NMDS) plots were prepared based on
214 the Bray-Curtis dissimilarity coefficient. To understand whether choice of primer pair reflects
215 true shift in microbial community composition or differential spread (dispersion) of data points
216 from their group centroid, we assessed the multivariate homogeneity of group dispersions
217 (variances) using the *betadisper* function of the *vegan* package.

218 In order to identify the species accountable for differences in grouping by primer pair, we
219 employed a negative binomial model-based approach available in the *DESeq2* package in R
220 (Love et al., 2014). Wald tests were performed and only species remaining significant ($p < 0.01$)
221 after the Benjamini–Hochberg correction were retained.

222

223 ***In silico* primer analysis**

224 We estimated the primer pair's predicted coverage and mismatches to the target database with
225 *PrimerProspector* (Walters et al., 2011). For this purpose, we used the latest SILVA v132 NR
226 16S rRNA gene database with 695,171 sequences. Primers weighted scores were calculated
227 with *PrimerProspector*'s default formula. Additional penalty score of 3.00 was given, if the final
228 3' base of primer had a mismatch with its target sequence. Lower value of weighted score
229 suggests better primer performance, whereas values above 0 suggest poor performance of
230 primer pairs (Walters et al., 2011). Predicted coverage of the primer pairs was calculated at
231 domain and phylum levels. We attributed domain universality to a primer sequence at a
232 stringency criterion of 80%, i.e. only when a primer sequence showed 80% or more coverage
233 of at least two taxonomic domains.

234

235 **Results**

236 **Read output and taxa distribution**

237 The Illumina MiSeq sequencing of different regions of 16S rRNA gene amplified by four
238 primer pairs generated roughly 17 million raw reads in total with on average 84,883 reads per
239 sample. Total number of raw reads differed according to primer pair, with highest number of
240 reads associated with 787F-1073R (5,408,844 reads) followed by 515F-806R (4,898,357

241 reads), 799F-1193R (3,687,401 reads) and 341F-805R (3,151,874 reads). Different proportions
242 of reads were filtered out at each step of quality filtering, denoising, merging and chimera
243 removing for each primer pair (Supplementary Table 2) and in the end, the highest number of
244 reads was retained for 515F-806R (72.6%), followed by 799F-1193R (51.8%), 341F-805R
245 (49.5%), and 787F-1073R (37.7%). In terms of read taxonomic classification at the domain
246 level, no reads were found to belong to Eukaryota for all primer pairs, however, distribution of
247 reads assigned to Bacteria and Archaea and for Chloroplast and Mitochondria differed
248 according to primer pair used (Supplementary Table 3). From four diverse source compartments
249 together with the two low diversity controls (mouse and mock bacterial community), in total
250 43 different phyla were observed after removing chloroplast and mitochondria sequences with
251 half (n=21) of the phyla commonly identified by all primer pairs (Figure 1a). The highest
252 number of phyla (n=42) was detected by the primer pair 515F-806R; yet the latter did not detect
253 the *Caldiserica* phylum, which was only detected by 341F-805R in a single soil sample S2B
254 with only two reads. Identified phyla differed in terms of relative abundance across source
255 compartments but also among primer pairs within a compartment (Figure 1b). As most of the
256 ASVs were not assigned at the species level, the genus level was chosen as a common
257 taxonomic level for dataset comparison across primer pairs. At genus level, a total of 955 unique
258 genera were identified, with only 348 commonly identified by all four primer pairs (Figure 1c)
259 and the largest number (n=696) identified by primer pair 515F-806R.

260

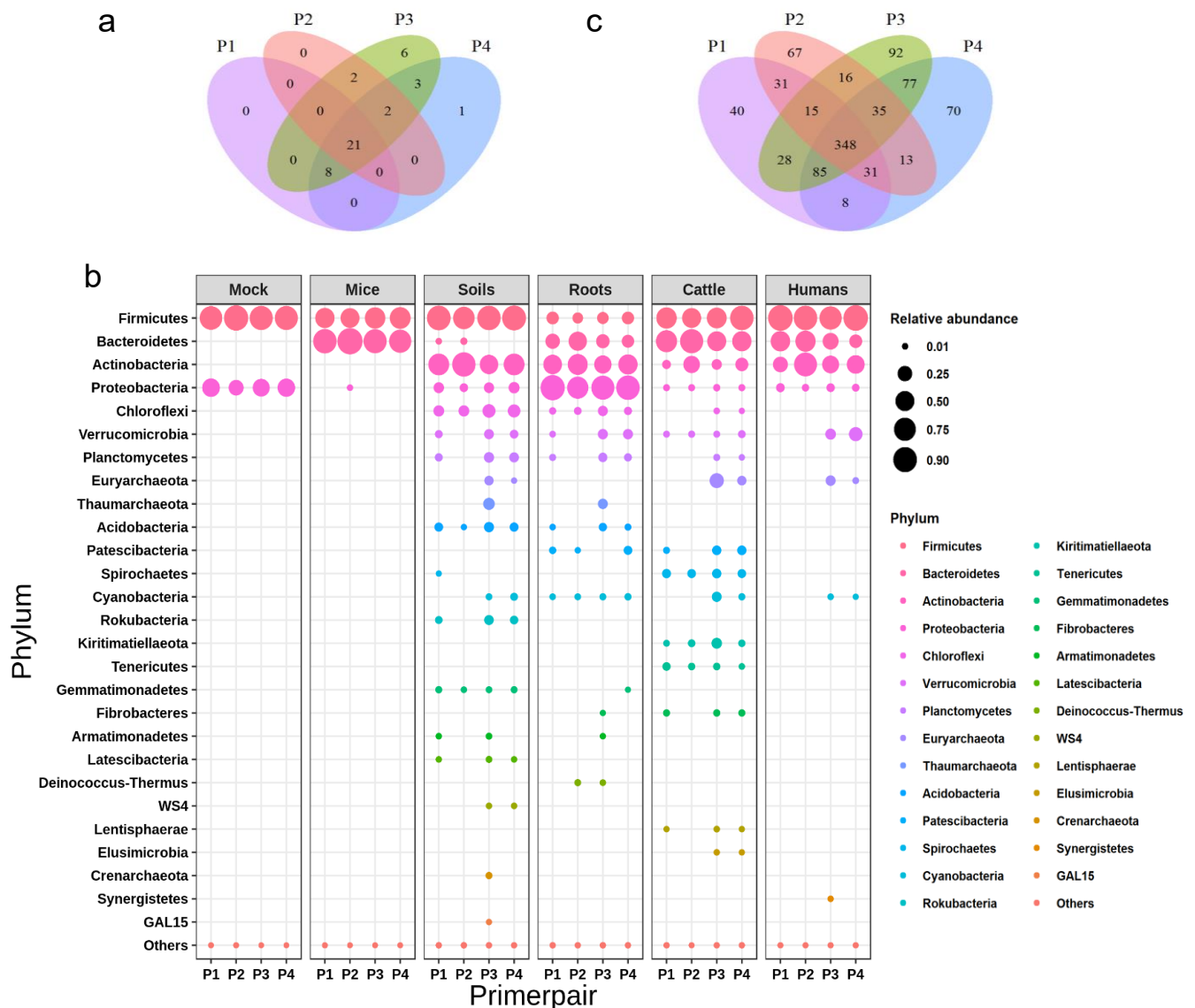


Figure 1. Comparative evaluation of primer pairs based on number and relative abundance of assigned taxa. Number of taxa shared and uniquely identified by primer pairs (787F-1073R (P1), 799F-1193R (P2), 515F-806R (P3) and 341F-805R (P4)) at (a.) phylum and (c.) genus level. Out of total 43 different phyla and 955 genera, 21 phyla and 348 genera are commonly identified by all four primer pairs. (b.) Bubble plot showing relative abundance of phylum by each primer pair and across source compartments. Size of the circles are proportional to relative abundance of phylum. Phyla are arranged according to total decreasing relative abundance and rarer phyla (<0.01%) are jointly included as ‘Others’.

261 With the mock community samples, we observed consistent performance of all primer pairs:
 262 out of eight bacterial species with negligible impurity (<0.01% foreign microbial DNA
 263 according to supplier), all eight bacterial species were recovered, however maximally assigned
 264 at genus level (Supplementary Figure 1). Only with primer pair 787F-1073R and 799F-1193R,

265 we observed an additional taxon *Parabacteroids*, which was present in few (five) samples with
266 <20 reads, which could be due to contamination or impurity. Removing rare ASVs from the
267 dataset is recommended based on the number of samples in which they are present or by total
268 count, because *DADA2* may be more sensitive to low amount of contamination (Caruso et al.,
269 2019). However, we avoided using such filtering of rarer ASVs as in our dataset each sample
270 is almost unique in terms of microbiome (due to the choice of heterogeneous samples in each
271 compartment) and each primer pair dataset can be differentially influenced by such filtering
272 parameter. Most reads from all primer pairs were successfully assigned at genus level for the
273 mock community, however small proportions of taxonomically unassigned ASVs (at genus
274 level) were present (Supplementary Figure 1). With respect to genus relative abundances, the
275 primer pair 799F-1193R was noted to be biased towards *Bacillus* with more than half of reads
276 assigned to this genus, and thus detected relatively fewer reads of all other genera
277 (Supplementary Figure 1).

278

279 **Choice of primer pair influences alpha and beta diversity**

280 We used General Linear Modelling (GLM) to test whether differences in alpha diversity
281 estimates (either number of observed species, or Shannon) between samples could be explained
282 by DNA extraction kit, primer pair, source compartment and the interaction between primer
283 pair and source compartment (primer pair*source compartment). Significant effects were
284 observed for primer pair (observed number of species, $p<0.001$; Shannon, $p<0.001$) and source
285 compartment (observed number of species, $p<0.001$; Shannon, $p<0.001$; Supplementary Table
286 4) on tested alpha diversity indices. However, no effect of DNA extraction kit (observed number
287 of species; $p=0.846$, Shannon; $p=0.188$) was observed on both tested alpha diversity indices
288 (Supplementary Table 4, Figure 2). Interaction between primer-pair and source compartment
289 showed no significant effect on observed species ($p=0.171$), but marginal effect on Shannon

290 (p=0.042; Supplementary Table 4, Figure 2). Out of the four diverse source compartments, soil
291 showed the highest microbial diversity, while the mouse and mock community samples (with
292 defined bacterial strains) as expected showed the lowest (Figure 2). Similarly, primer pair 515F-
293 806R revealed highest microbial diversity and 799F-1193R lowest (Figure 2).

294 To determine whether choice of primer pair influences microbial community composition, we
295 calculated two beta diversity metrics (Jaccard and Bray-Curtis) and included again DNA
296 extraction kit, primer pair, source compartment and the interaction primer pair*source
297 compartment as explanatory variables in PERMANOVA models. We observed a significant
298 effect of primer pair (Jaccard: $R^2=0.024$, $p=0.001$; Bray-Curtis: $R^2=0.021$, $p=0.001$) and source
299 compartment (Jaccard: $R^2=0.486$, $p=0.001$; Bray-Curtis: $R^2=0.624$, $p=0.001$), but also
300 interaction primer pair*source compartment (Jaccard: $R^2=0.106$, $p=0.001$; Bray-Curtis:
301 $R^2=0.081$, $p=0.001$) on microbial beta diversity estimates (Figure 3, Supplementary Table 5).
302 However, no effect of the DNA extraction kit (Jaccard: $R^2=0.003$, $p=0.052$; Bray-Curtis:
303 $R^2=0.002$, $p=0.142$) on the microbial community composition was observed (Supplementary
304 Table 5). Homogeneity of dispersion analysis for primer pairs using *betadisper* function
305 suggested true homogeneity in dispersions (Jaccard: $p=0.831$; Bray-Curtis: $p=0.616$;
306 Supplementary Figure 2).

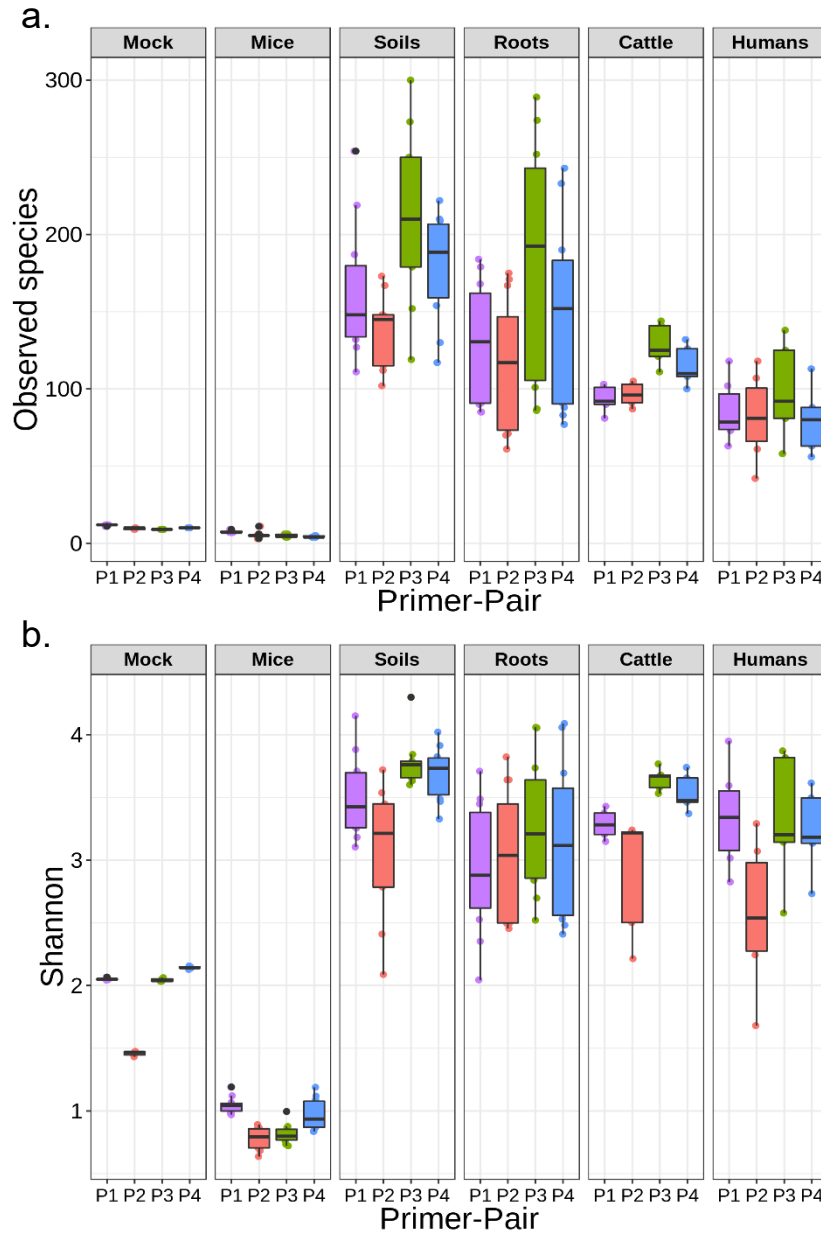


Figure 2. Box plots showing microbial alpha diversity of source compartments revealed by each primer pair. Difference in (a.) number of observed species and (b.) Shannon diversity of each source compartment by the primer pairs used (787F-1073R (P1), 799F-1193R (P2), 515F-806R (P3) and 341F-805R (P4)). For most source compartments, primer pair 515F-806R (P3) reflected the highest microbial diversity and 799F-1193R (P2) the lowest. Soil samples showed the highest alpha diversity and gnotobiotic mouse samples with defined colonized bacteria the least.

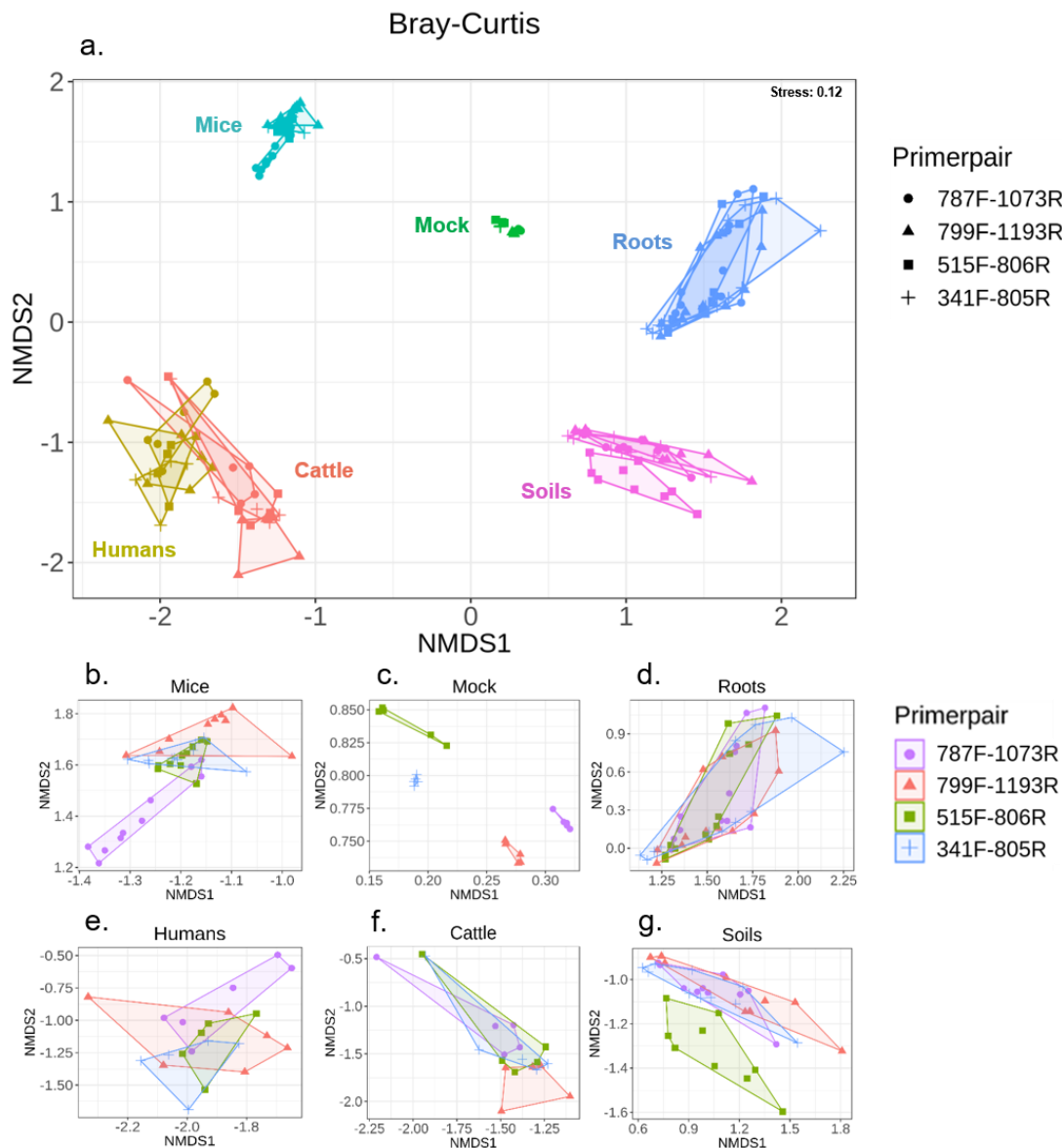
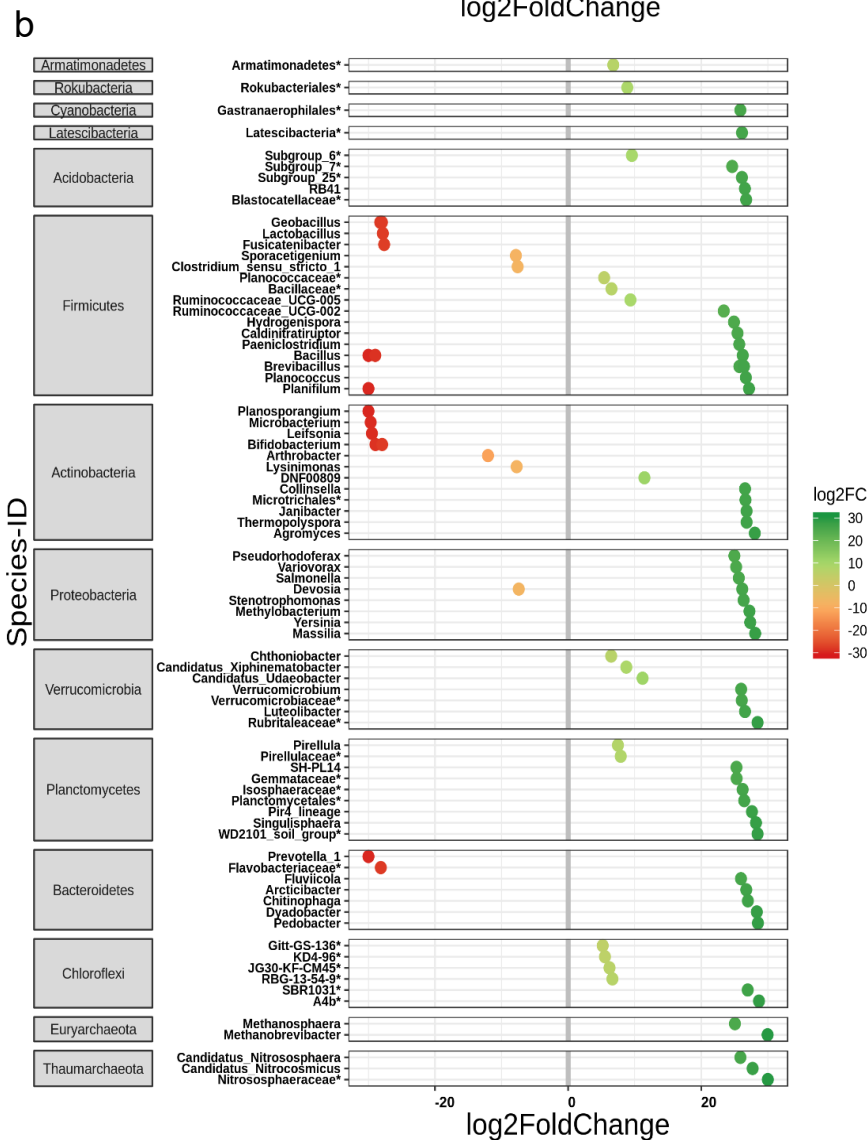
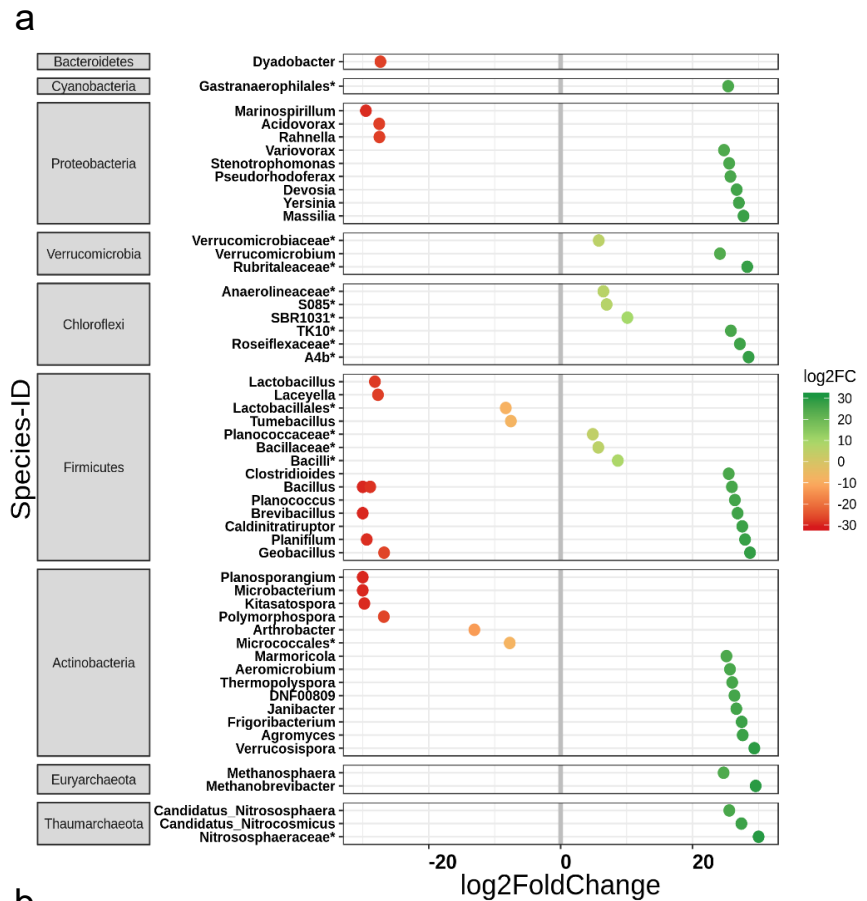


Figure 3. NMDS plots showing microbial beta diversity of source compartments based on Bray-Curtis distance metric. NMDS plots reflecting differences in microbial community composition of (a.) all source compartments together and for each source compartment (extracted values from the main NMDS plot); (b.) mice, (c.) mock, (d.) roots, (e.) humans, (f.) cattle, (g.) soils based on Bray-Curtis distance metric calculated using four primer pairs.

307 **Primer pairs are associated with differential abundance of bacterial species**

308 To determine whether the choice of primer pair influences the ability to detect differences in
 309 relative species abundance, we employed negative binomial-based Wald tests. We compared
 310 species abundances obtained with 515F-806R, as it detected highest number of species, against

311 the species abundances obtained with all other primer pairs (787F-1073R, 799F-1193R, 341F-
312 805R). We identified 58 bacterial species that differed significantly ($p < 0.01$) in abundance
313 between primer pairs 515F-806R and 787F-1073R, with 19 (32.8%) and 39 (67.2%) showing
314 significant decrease and increase in mean relative abundance, respectively (Figure 4a). Highest
315 number of differential abundant species belonged to the phylum Firmicutes (19 species)
316 followed by Actinobacteria (14 species). Five members of Archaea (3 species;
317 Thaumarchaeota, 2 species; Euryarchaeota) showed higher abundance with 515F-806R than
318 with 787F-1073R (Figure 4a). Analysing the primer pair 515F-806R against 799F-1193R
319 revealed 86 bacterial species, significantly different in abundance, with 19 (22.1%) and 67
320 (77.9%) species showed significant decrease and increase in mean relative abundance,
321 respectively (Figure 4b). Bacterial species showing differential abundance were mainly from
322 the phylum Firmicutes (21 species), followed by Actinobacteria (18 species). Similar to
323 comparison between 515F-806R and 787F-1073R, five members of Archaea (3 species,
324 Thaumarchaeota; 2 species, Euryarchaeota) showed higher abundance with 515F-806R than
325 with 799F-1193R (Figure 4b). Likewise, 25 bacterial species showed significant difference in
326 abundance between primer pair 515F-806R and 341F-805R, whereas 9 (36.0%) and 16 (64.0%)
327 showed significant decrease and increase in mean relative abundance, respectively (Figure 4c).
328 Similar to previous comparisons, highest number of bacterial species showing differential
329 abundance were found for Firmicutes (11 species) and archaeal members (3 species) were
330 higher in abundance with 515F-806R. Overall, among all the comparisons, primer pair 515F-
331 806R revealed higher number of taxa with increased abundance compared to other primer pairs.
332 Among all the taxa showing significant differential abundance, the majority of species were
333 from phyla Firmicutes and Actinobacteria. Archaeal taxa also showed higher abundance using
334 the 515F-806R primer pair than when using other primer pairs (Figure 4).



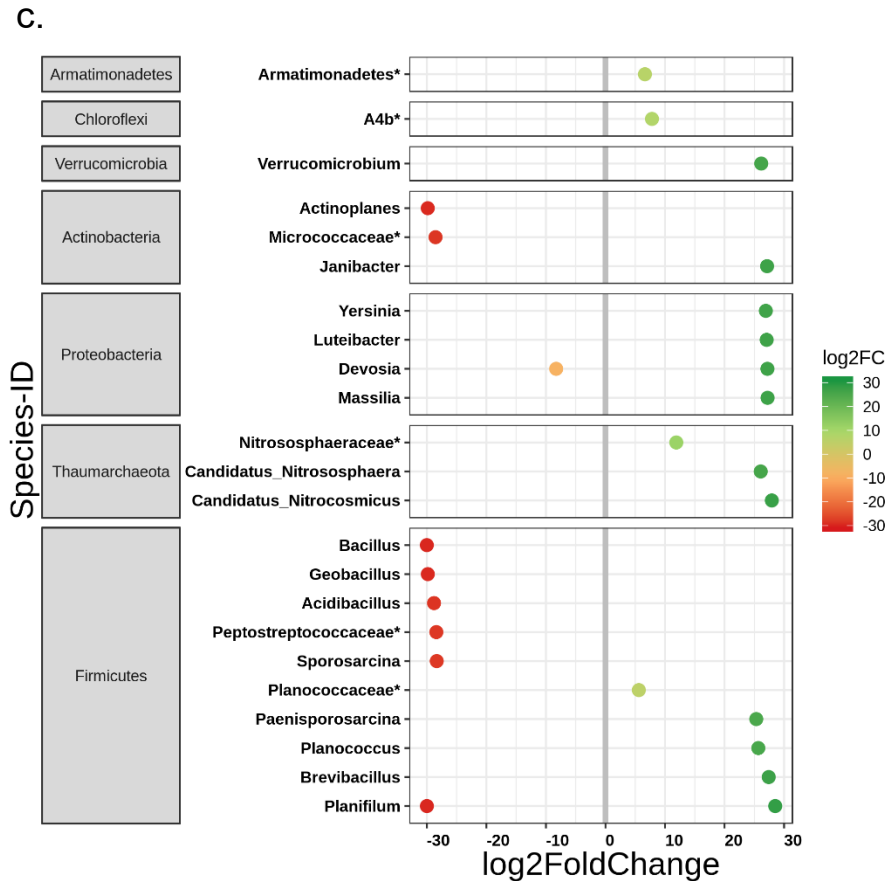


Figure 4. Differential abundance of species according to primer pair used. Shown are species that differ in their mean abundance in relation to the primer pair used. Primer pair 515F-806R was compared against (a.) 787F-1073R, (b.) 799F-1193R and (c.) 341F-805R. The values indicate a log₂-fold (log₂FC) decrease (red; using other primer pairs) or increase (green, using 515F-806R primer pair) in relative abundance of species. Species are arranged according to increasing values of log₂-fold change and grouped according to their respective phylum. The highest possible taxonomic assignment (maximal to the genus level) is shown for each species. In all the comparisons, more species show increased abundance using 515F-806R than using any other primer pair.

335 *In silico* primer pair evaluation

336 In order to evaluate the primer pairs *in silico*, weighted scores of primer matches were
 337 calculated for each primer pair against the latest SILVA v132 NR 16S rRNA gene database.
 338 Weighted score for individual primers was best (lowest) for 515F (0.09 ± 0.0006), followed by
 339 341F (0.27 ± 0.0008), 806R (0.30 ± 0.0009), 1073R (0.43 ± 0.0012), 787F (0.51 ± 0.0015),
 340 805R (0.62 ± 0.0019), 1193R (0.79 ± 0.0019) and 799F (0.89 ± 0.0022). Overall best score

341 among the primer pairs was for 515F-806R (0.19), followed by 341F-805R (0.45), 787F-1073R
342 (0.47) and 799F-1193R (0.84).

343 *In silico* primer coverage was predicted both for individual primers and pairs thereof at the
344 domain (Archaea, Bacteria, Eukaryota) and phylum levels. At the domain level, primer
345 coverage score was best for primer 515F, which covered all three domains (>90% coverage),
346 whereas primer 1193R was found to be bacteria specific (Table 1). In terms of domain
347 universality based on 80% coverage criteria, only primer 515F was found to be “universal” to
348 all three domains, whereas 1073R and 1193R were bacteria specific. Likewise, both 515F-806R
349 and 341F-805R were predicted universal for Bacteria and Archaea and two other tested primer
350 pairs (787F-1073R, 799F-1193R) were found to be more bacteria specific (Table 1). Detailed
351 coverage of primer pairs at domain and bacterial phylum levels are shown in Supplementary
352 Table 6.

353

354 **Discussion**

355 The inclusion of microbiomes as connecting links between trophic levels has been proposed in
356 the One Health framework (van Bruggen et al., 2019). Before examining microbial transfers
357 from one source compartment to another and implications for health at each level,
358 methodological challenges for detecting and characterising microbiomes from trophic levels
359 along the food chain must be met (Trinh et al., 2018). To overcome such limitations and to
360 achieve a standard microbiome analysis approach, we investigated microbiomes from diverse
361 source compartments along the food chain, using four commonly available primer pairs
362 targeting different regions of 16S rRNA gene. As far as we know, this is the first integrated
363 analysis of microbiomes along a food chain. We observed that among the tested primers all
364 were explicitly targeting the bacterial domain and no reads belonging to Eukaryota were found.

365 Primer pairs 515F-806R (P3) and 341F-805R (P4) performed better than others and also
366 detected archaea as expected. Specifically, we recommend primer pair 515F-806R (P4) for One
367 Health studies as it recovers the highest bacterial diversity both *in silico* and in samples from
368 diverse source compartments along the food chain.

369 All the four tested primer pairs performed well and recovered the expected bacterial diversity
370 in mock and mouse gut microbiomes. Nevertheless, we also noticed few rarer taxa in mock
371 community samples (neglectable level of noise; <0.01%), which could be due to the fact that
372 we did not perform taxa prevalence/abundance based filtering in our analyses because, each
373 sample being unique in terms of microbiome composition in our study, such filtering would
374 influence the four primer dataset differentially. As *DADA2* may be more sensitive to low
375 amount of such contamination (Caruso et al., 2019), we recommend performing such filtering
376 steps. We found that primer pair choice may significantly influence bacterial alpha diversity,
377 which is supported by earlier studies that compared various primer pairs amplifying 16S rRNA
378 gene (Beckers et al., 2016;Thijs et al., 2017). Highest microbial diversity was revealed by
379 primer pair 515F-806R using two alpha diversity indices (number of observed species,
380 Shannon). Our result is in contrast to previous studies (Klindworth et al., 2013;Thijs et al.,
381 2017) where the primer pair 341F-805R showed the highest microbial diversity. Since the *in*
382 *silico* study by *Klindworth et al.*, the SILVA database nearly doubled in size following the
383 inclusion of new sequences, and current primer coverage of the new database with existing
384 primer pairs was up to now unknown. Meanwhile, there have also been several improvements
385 made to the design of the forward and reverse primers of the 515F-806R pair in order to cover
386 previously undetected taxa and to reduce biases against Crenarchaeota/Thaumarchaeota (Apprill
387 et al., 2015;Parada et al., 2016;Walters et al., 2016). Primer pair 515F-806R was not included
388 in the comparative study by *Thijs et al.* because they intended to find additional suitable primer
389 pairs for soil microbiome studies other than existing 515F-806R. Similar highest alpha diversity

390 results for 515F-806R primer pair in comparison to other primer pairs were obtained in a recent
391 study (Chen et al., 2019), however, limited to human gut microbiome. Primer pair 515F-806R
392 has been the recommended primer pair for microbiome studies by the Earth Microbiome Project
393 and has successfully been used in studies from other source compartments (Parada et al. 2016;
394 Apprill et al. 2015; Walters et al. 2015), but, to the best of our knowledge, has never been tested
395 with several compartments in the same study. We could not demonstrate a significant effect of
396 extraction kit on overall alpha diversity, studies showing no effect or significant effect of DNA
397 extraction kit on microbial alpha diversity are available in the literature (Hallmaier-Wacker et
398 al., 2018;Fiedorova et al., 2019;Mattei et al., 2019).

399 The origin of the microbiota samples explained the majority of differences in alpha diversity
400 patterns. Such differences in diversity were expected as each source compartment possesses its
401 own microbial signature (Thompson et al., 2017;Ikeda-Ohtsubo et al., 2018;Reese and Dunn,
402 2018). Although, our study was focused on an agricultural food chain, the obtained alpha
403 diversity results using 515F-806R primer pair are comparable with Earth Microbiome Project's
404 (EMP) findings, where similar primer pair was used to amplify diverse free-living and host-
405 associated microbial communities (Thompson et al., 2017). We observed high microbial
406 richness in free-living microbial community (i.e. in soil) compared to host-associated microbial
407 communities (i.e. gut microbiome) similar to observations in the EMP. However, a notable
408 exception was observed in the EMP, where plant roots showed highest microbial richness
409 compared to all other host associated or free-living studied compartments (Thompson et al.,
410 2017). We did not observe such patterns and among the studied source compartments, soil
411 samples showed highest microbial richness in our study. Such discrepancy could arise due to
412 differences in sample types (for example different species of plants), sampled root region
413 (whole root processed in our experiment whereas the microbial rich rhizosphere was
414 investigated in EMP) but also due to the fact that in EMP, root samples were collected from

415 only two locations compare to worldwide equal distribution of collected soil samples, which
416 showed a large microbial richness gradient (Thompson et al., 2017). In our study each source
417 compartment exhibited unique microbiome, and only ~17% of phyla were shared among
418 microbial communities from different source compartments (soils, plant roots, cattle and
419 humans) along the food chain. Four phyla; Firmicutes, Bacteroidetes, Actinobacteria and
420 Proteobacteria dominate microbial communities along the model food chain, as each of them
421 represent among major phyla in individual compartment specific studies (Hacquard et al.,
422 2015;Ikeda-Ohtsubo et al., 2018). Furthermore, two archaeal phyla Euryarchaeota and
423 Thaumarchaeota, only detected by primer pairs 515F-806R and 341F-805R, showed overall
424 high abundance along the food chain. Species richness decreased gradually across the food
425 chain, with the lowest richness observed in human faeces.

426 As observed for alpha diversity, beta diversity, as measured by Jaccard and Bray-Curtis
427 dissimilarity indices, was significantly influenced by the choice of primer pairs; a finding that
428 is supported by previous studies (Thijs et al., 2017;Chen et al., 2019). We observed no
429 significant difference among primer pairs in beta-dispersion, suggesting that observed
430 difference in beta diversity metrics among primer pairs are due to true difference in microbial
431 community and not due to differential dispersion from centroids for each primer pair. As
432 evidenced for alpha diversity, strong effect of source compartment on beta diversity is also
433 expected as each source compartment harbours compositionally different sets of bacteria in
434 different proportions (Hacquard et al., 2015;Thompson et al., 2017). Additionally, we found
435 the factor interaction "primer pair*source compartment" to be significant in our model, thus
436 indicating that some primer pairs may be better at revealing changes in microbial community
437 composition than others for specific source compartment: For example, 515F-806R was
438 associated with better spread of data points for soil samples in comparison to other primer pairs
439 in the NMDS plot. Nevertheless, interpretation of beta diversity results is not as straightforward

440 as for alpha diversity, because larger difference between samples, for example, can give overall
441 high value for beta diversity, but this could arise due to limited, and thus differential, detection
442 of taxa by a primer pair. Furthermore, no significant effect of extraction kit on beta diversity
443 was observed in our analysis.

444 To investigate the observed differences in alpha and beta diversity using different primer pairs,
445 we compared the relative abundances of bacterial species between the primer pair datasets.
446 Overall, we found the largest number of differentially abundant taxa (86 species) when
447 comparing primer pair 515F-806R with 799F-1193R, than comparing either 787F-1073R (58
448 species) or 341F-805R (25 species). Larger differences in case of 799F-1193R and 787F-1073R
449 could be due to the fact that they amplify different hypervariable regions, V5-V6-V7 and V5-
450 V6, respectively, as compared to the V4 region amplified by 515F-806R. Such discrepancies in
451 abundance and taxa assignment of reads originating from different hypervariable regions of 16S
452 rRNA gene have been reported previously (Claesson et al., 2010). In all comparisons performed
453 in our analysis, higher proportion of bacterial species showed increased abundance when
454 profiled with primer pair 515F-806R, thus reflecting the better performance of the latter primer
455 pair over the three other primer pairs in detecting changes in taxon abundance. Many of the
456 differential abundant species, which showed decrease or increase in all the comparisons
457 belonged to Firmicutes and Actinobacteria, which are dominant phyla of microbiome
458 community in different source compartments. Many species showing higher representation in
459 515F-806R primer pair in comparison with other primer pairs belonged to archaeal phyla,
460 namely Euryarchaeota and Thaumarchaeota. Overall, a total of 78,659 (2.24%) reads were
461 observed for these two phyla using primer pair 515F-806R as compared to 1,552 (0.09%) reads
462 using the other prokaryotic primer pair 341F-805R. Euryarchaeota is a highly diverse archaeal
463 phylum with predominating methanogenic species found from marine water to soil, plant root,
464 cattle and human gut and it plays functionally important roles in each source compartment

465 (Moissl-Eichinger et al., 2018). In humans, they are noted to be highly heritable and can be
466 present in up to 95.7% of individuals in diverse studies (Lurie-Weinberger and Gophna, 2015).
467 Their loss in ruminants can result in loss of energy and their role in obesity was suggested
468 (Lurie-Weinberger and Gophna, 2015). Similarly, Thaumarchaeota commonly found in marine
469 water, where they suggested to be keystone members of microbial community but also in
470 diverse soil types and in association with plants (Brochier-Armanet et al., 2012;Taffner et al.,
471 2019). Members of Thaumarchaeota are primarily known as ammonia oxidizers but also
472 members with unknown energy metabolism (Ren et al., 2019). Considering the ubiquity and
473 importance of Euryarchaeota and Thaumarchaeota in different source compartments, they could
474 be important features in between-compartment studies and thus selection of prokaryotic primer
475 pairs such as 515F-806R is highly recommended to amplify archaeal and bacterial taxa
476 conjointly.

477 Our *in silico* analysis found 515F to be the most ‘universal’ among all tested primers, a finding
478 similarly reported by *Klindworth et al.* Primer pair 515F-806R gave the best (lowest) weighted
479 score followed by 341F-805R. In terms of coverage, we found both primer pairs 515F-806R
480 and 341F-805R to be prokaryote specific with 515F-806R providing better coverage of archaea
481 (96.39%) as compared to 341F-805R (83.59%), confirming our experimental observation,
482 where higher proportion of reads assigned to archaea by primer pair 515F-806R than 341F-
483 805R. However, primer pair 341F-805R was mentioned to be the best primer pair in terms of
484 overall coverage by *Klindworth et al.* who used a smaller version of the SILVA database. Also,
485 since then both the forward and reverse primers of the 515F-806R pair were made more general,
486 i.e. wobble bases were added to both primers, and the reverse primer is now 6 base pair shorter
487 in order to be less specific (Apprill et al., 2015;Parada et al., 2016;Walters et al., 2016).

488 In conclusion, we performed both laboratory and *in silico* tests to compare four commonly
489 available 16S primer pairs in order to assess microbial community coverage across diverse

490 source compartments along the food chain. Overall, we observed that the choice of primer pair
491 can significantly influence microbial alpha and beta diversity and can identify differential taxa
492 abundance. We showed that primer pair 515F-806R provides greater depth and taxa coverage
493 as compared to other tested primer pairs using samples from different compartments, but also
494 provides higher database coverage performance *in silico*. To conclude, we recommend
495 including general prokaryotic primer pair such as 515F-806R to also recover archaea in One
496 Health studies, due to their important roles in diverse systems. With this information, the
497 methodological bottleneck concerning the choice of primer pair to adequately reflect microbial
498 diversity within and between each source compartment can be addressed. This information will
499 help identify and characterize the importance of microbiomes from heterogeneous origins
500 within a One Health framework.

501

502 **Acknowledgements**

503 This research was supported by the University of Bern through the Interfaculty Research
504 Cooperation One Health (www.onehealth.unibe.ch). We are grateful to all the participating
505 laboratories for providing source compartment specific samples. We thank the Interfaculty
506 Bioinformatics Unit (IBU) Linux cluster from University of Bern and Institute for Infectious
507 Diseases for providing computing facilities and Stefan Neuenschwander for assistance with
508 software installation.

509

510 **Author's contributions**

511 ME, KS and AR were involved in the development of the conceptual framework for this study.
512 AR and W conceived the study. KS and FR provided extracted DNA samples of soil/root and
513 mouse, respectively. W carried out laboratory experiments, bioinformatics and statistical

514 analyses. W wrote and AR edited the MS. All authors commented and approved the final
515 version of the MS.

516

517 **Data accessibility**

518 The individual microbial 16S rRNA gene sequences are available under BioProject ID
519 PRJNA576426.

520

521 **Conflict of interest**

522 The authors declare no conflict of interests.

523

524 **Contribution to the field**

525 Microbiomes not only provide essential services at each trophic level of the food chain i.e. soil,
526 plant, cattle, human, but was recently recognised as a potential connecting links between trophic
527 levels. Therefore, there is a need to track microbiome composition and transfer across diverse
528 source compartments to understand microbial influence on the health of each compartment
529 along the food chain. Yet before being able to compare compartments appropriately,
530 methodological challenges to characterise microbial communities from diverse source
531 compartments must be addressed. In that respect, we investigated the influence of different
532 extraction kits and choice of 16S rRNA gene primer pairs on microbial community
533 characterisation for the first time on diverse source compartments along the food chain. We
534 observed that choice of primer pair can significantly influence microbiome characterisation.
535 We make the recommendation of specific primer pair, 515F-806R which can cover highest
536 microbial community diversity and providing best comparability between source

537 compartments. Comparing microbiomes using 515F-806R revealed that soil and root samples
538 have the highest estimates of species richness and inter-sample variation. Species richness
539 decreased gradually along the food chain. Such knowledge has broader relevance, in making
540 informed decisions concerning the choice of primer pair for microbiome researchers,
541 irrespective of their studied source compartment and specifically can serve as a guide for those
542 planning cross-compartment studies.

543
544 **References**

- 545
546 Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU
547 rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton.
548 *Aquatic Microbial Ecology* 75, 129-137.
- 549 Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models
550 Using lme4. *Journal of Statistical Software* 67, 1-48.
- 551 Beckers, B., Op De Beeck, M., Thijs, S., Truyens, S., Weyens, N., Boerjan, W., et al. (2016).
552 Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere
553 Bacterial Microbiomes in Metabarcoding Studies. *Front Microbiol* 7, 650.
- 554 Bell, V., Ferrao, J., Pimentel, L., Pintado, M., and Fernandes, T. (2018). One Health, Fermented
555 Foods, and Gut Microbiota. *Foods* 7, 195.
- 556 Bodenhausen, N., Horton, M.W., and Bergelson, J. (2013). Bacterial communities associated
557 with the leaves and the roots of *Arabidopsis thaliana*. *PLoS One* 8, e56329.
- 558 Brochier-Armanet, C., Gribaldo, S., and Forterre, P. (2012). Spotlight on the Thaumarchaeota.
559 *ISME J* 6, 227-230.
- 560 Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P.
561 (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat*
562 *Methods* 13, 581-583.
- 563 Caruso, V., Song, X., Asquith, M., and Karstens, L. (2019). Performance of Microbiome
564 Sequence Inference Methods in Environments with Varying Biomass. *mSystems* 4,
565 e00163-00118.
- 566 Chelius, M.K., and Triplett, E.W. (2001). The Diversity of Archaea and Bacteria in Association
567 with the Roots of *Zea mays* L. *Microb Ecol* 41, 252-263.
- 568 Chen, Z., Hui, P.C., Hui, M., Yeoh, Y.K., Wong, P.Y., Martin, C., et al. (2019). Impact of
569 Preservation Method and 16S rRNA Hypervariable Region on Gut Microbiota Profiling.
570 *mSystems* 4, e00271-00218.
- 571 Claesson, M.J., Wang, Q., O'sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., et al. (2010).
572 Comparison of two next-generation sequencing technologies for resolving highly
573 complex microbiota composition using tandem variable 16S rRNA gene regions.
574 *Nucleic Acids Res* 38, e200.
- 575 Destoumieux-Garzon, D., Mavingui, P., Boetsch, G., Boissier, J., Darriet, F., Duboz, P., et al.
576 (2018). The One Health Concept: 10 Years Old and a Long Road Ahead. *Front Vet Sci*
577 5, 14.
- 578 Edgar, R.C. (2018). Updating the 97% identity threshold for 16S ribosomal RNA OTUs.
579 *Bioinformatics* 34, 2371-2375.

- 580 Fiedorova, K., Radvansky, M., Nemcova, E., Grombirikova, H., Bosak, J., Cernochova, M., et
581 al. (2019). The Impact of DNA Extraction Methods on Stool Bacterial and Fungal
582 Microbiota Community Recovery. *Front Microbiol* 10, 821.
- 583 Fricker, A.M., Podlesny, D., and Fricke, W.F. (2019). What is new and relevant for sequencing-
584 based microbiome research? A mini-review. *J Adv Res* 19, 105-112.
- 585 Hacquard, S., Garrido-Oter, R., Gonzalez, A., Spaepen, S., Ackermann, G., Lebeis, S., et al.
586 (2015). Microbiota and Host Nutrition across Plant and Animal Kingdoms. *Cell Host*
587 *Microbe* 17, 603-616.
- 588 Hallmaier-Wacker, L.K., Lueert, S., Roos, C., and Knauf, S. (2018). The impact of storage
589 buffer, DNA extraction method, and polymerase on microbial analysis. *Sci Rep* 8, 6292.
- 590 Hebbandi Nanjundappa, R., Ronchi, F., Wang, J., Clemente-Casares, X., Yamanouchi, J.,
591 Sokke Umeshappa, C., et al. (2017). A Gut Microbial Mimic that Hijacks Diabetogenic
592 Autoreactivity to Suppress Colitis. *Cell* 171, 655-667 e617.
- 593 Herlemann, D.P., Labrenz, M., Jurgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F.
594 (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the
595 Baltic Sea. *ISME J* 5, 1571-1579.
- 596 Ikeda-Ohtsubo, W., Brugman, S., Warden, C.H., Rebel, J.M.J., Folkerts, G., and Pieterse,
597 C.M.J. (2018). How Can We Define "Optimal Microbiota?": A Comparative Review of
598 Structure and Functions of Microbiota of Animals, Fish, and Plants in Agriculture.
599 *Front Nutr* 5, 90.
- 600 Kalle, E., Kubista, M., and Rensing, C. (2014). Multi-template polymerase chain reaction.
601 *Biomol Detect Quantif* 2, 11-29.
- 602 Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013).
603 Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-
604 generation sequencing-based diversity studies. *Nucleic Acids Res* 41, e1.
- 605 Knauth, S., Schmidt, H., and Tippkötter, R. (2013). Comparison of commercial kits for the
606 extraction of DNA from paddy soils. *Lett Appl Microbiol* 56, 222-228.
- 607 Li, H., Limenitakis, J.P., Fuhrer, T., Geuking, M.B., Lawson, M.A., Wyss, M., et al. (2015).
608 The outer mucus layer hosts a distinct intestinal microbial niche. *Nat Commun* 6, 8292.
- 609 Lim, M.Y., Song, E.J., Kim, S.H., Lee, J., and Nam, Y.D. (2018). Comparison of DNA
610 extraction methods for human gut microbial community profiling. *Syst Appl Microbiol*
611 41, 151-157.
- 612 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
613 dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.
- 614 Lurie-Weinberger, M.N., and Gophna, U. (2015). Archaea in and on the Human Body: Health
615 Implications and Future Directions. *PLoS Pathog* 11, e1004833.
- 616 Mamantopoulos, M., Ronchi, F., Van Hauwermeiren, F., Vieira-Silva, S., Yilmaz, B., Martens,
617 L., et al. (2017). Nlrp6- and ASC-Dependent Inflammasomes Do Not Shape the
618 Commensal Gut Microbiota Composition. *Immunity* 47, 339-348 e334.
- 619 Mattei, V., Murugesan, S., Al Hashmi, M., Mathew, R., James, N., Singh, P., et al. (2019).
620 Evaluation of Methods for the Extraction of Microbial DNA From Vaginal Swabs Used
621 for Microbiome Studies. *Front Cell Infect Microbiol* 9, 197.
- 622 Mazmanian, S.K., Liu, C.H., Tzianabos, A.O., and Kasper, D.L. (2005). An
623 immunomodulatory molecule of symbiotic bacteria directs maturation of the host
624 immune system. *Cell* 122, 107-118.
- 625 Mcmurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive
626 analysis and graphics of microbiome census data. *PLoS One* 8, e61217.
- 627 Moissl-Eichinger, C., Pausan, M., Taffner, J., Berg, G., Bang, C., and Schmitz, R.A. (2018).
628 Archaea Are Interactive Components of Complex Microbiomes. *Trends Microbiol* 26,
629 70-85.

- 630 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2019).
631 vegan: Community Ecology Package. Version 2.5-6.
- 632 Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016). Every base matters: assessing small
633 subunit rRNA primers for marine microbiomes with mock communities, time series and
634 global field samples. *Environ Microbiol* 18, 1403-1414.
- 635 R Development Core Team, R. 2011. R: A Language and Environment for Statistical
636 Computing. *R Foundation for Statistical Computing*. [https://doi.org/10.1007/978-3-](https://doi.org/10.1007/978-3-540-74686-7)
637 540-74686-7.
- 638 Reese, A.T., and Dunn, R.R. (2018). Drivers of Microbiome Biodiversity: A Review of General
639 Rules, Feces, and Ignorance. *MBio* 9, e01294-01218.
- 640 Ren, M., Feng, X., Huang, Y., Wang, H., Hu, Z., Clingenpeel, S., et al. (2019). Phylogenomics
641 suggests oxygen availability as a driving force in Thaumarchaeota evolution. *ISME J*
642 13, 2150-2161.
- 643 Sundquist, A., Bigdeli, S., Jalili, R., Druzin, M.L., Waller, S., Pullen, K.M., El-Sayed, Y.Y.,
644 Taslimi, M.M., Batzoglou, S., and Ronaghi, M. (2007). Bacterial flora-typing with
645 targeted, chip-based Pyrosequencing. *BMC Microbiol* 7, 108.
- 646 Taffner, J., Cernava, T., Erlacher, A., and Berg, G. (2019). Novel insights into plant-associated
647 archaea and their functioning in arugula (*Eruca sativa* Mill.). *J Adv Res* 19, 39-48.
- 648 Thijs, S., Op De Beeck, M., Beckers, B., Truyens, S., Stevens, V., Van Hamme, J.D., et al.
649 (2017). Comparative Evaluation of Four Bacteria-Specific Primer Pairs for 16S rRNA
650 Gene Surveys. *Front Microbiol* 8, 494.
- 651 Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., et al. (2017).
652 A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457-
653 463.
- 654 Trinh, P., Zaneveld, J.R., Safranek, S., and Rabinowitz, P.M. (2018). One Health Relationships
655 Between Human, Animal, and Environmental Microbiomes: A Mini-Review. *Front*
656 *Public Health* 6, 235.
- 657 Van Bruggen, A.H.C., Goss, E.M., Havelaar, A., Van Diepeningen, A.D., Finckh, M.R., and
658 Morris, J.G., Jr. (2019). One Health - Cycling of diverse microbial communities as a
659 connecting force for soil, plant, animal, human and ecosystem health. *Sci Total Environ*
660 664, 927-937.
- 661 Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., et al.
662 (2016). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal
663 Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems*
664 1, e00009-00015.
- 665 Walters, W.A., Caporaso, J.G., Lauber, C.L., Berg-Lyons, D., Fierer, N., and Knight, R. (2011).
666 PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase
667 chain reaction primers. *Bioinformatics* 27, 1159-1161.