

 Open access • Posted Content • DOI:10.1101/2021.04.30.442084

Metagenomic investigation of the equine faecal microbiome reveals extensive taxonomic and functional diversity — [Source link](#)

[Rachel Gilroy](#), [Joy Leng](#), [Anuradha Ravi](#), [Evelien M. Adriaenssens](#) ...+7 more authors

Institutions: [Norwich Research Park](#), [University of Surrey](#), [Hebrew University of Jerusalem](#), [University of East Anglia](#)

Published on: 30 Apr 2021 - [bioRxiv](#) (Cold Spring Harbor Laboratory)

Topics: [Microbiome](#) and [Metagenomics](#)

Related papers:

- [Phylotyping and functional analysis of two ancient human microbiomes](#)
- [Multi-Method Characterization of the Human Circulating Microbiome.](#)
- [Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics.](#)
- [Defining the Human Microbiome](#)
- [Approaches to metagenomic classification and assembly](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/metagenomic-investigation-of-the-equine-faecal-microbiome-4wut6h78d7>

1 **Metagenomic investigation of the equine faecal**
2 **microbiome reveals extensive taxonomic and**
3 **functional diversity**

4

5 Rachel Gilroy¹, Joy Leng², Anuradha Ravi^{1,#}, Evelien M. Adriaenssens¹, Aharon Oren³,
6 David Baker¹, Roberto M. La Ragione², Christopher Proudman², Mark J. Pallen^{1,2,4*}

7

8 ¹Quadram Institute Bioscience, Norwich Research Park, Norwich, UK.

9 ²School of Veterinary Medicine, University of Surrey, Guildford, Surrey, UK.

10 ³The Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra
11 Campus, Jerusalem, Israel.

12 ⁴University of East Anglia, Norwich Research Park, Norwich, UK.

13 [#]Current address: Gemini Centre for Sepsis Research, Norwegian University of Science
14 and Technology, Trondheim, Norway

15

16 *Corresponding author

17 Email: mark.pallen@quadram.ac.uk

18

19

20

21 **Abstract**

22 **Background.** The horse plays crucial roles across the globe, including in horseracing, as
23 a working and companion animal and as a food animal. The horse hindgut microbiome
24 makes a key contribution in turning a high fiber diet into body mass and horsepower.
25 However, despite its importance, the horse hindgut microbiome remains largely
26 undefined. Here, we applied culture-independent shotgun metagenomics to thoroughbred
27 equine faecal samples to deliver novel insights into this complex microbial community.

28 **Results.** We performed metagenomic sequencing on five equine faecal samples to
29 construct 123 high- or medium-quality metagenome-assembled genomes from Bacteria
30 and Archaea. In addition, we recovered nearly 200 bacteriophage genomes. We
31 document surprising taxonomic and functional diversity, encompassing dozens of novel or
32 unnamed bacterial genera and species, to which we have assigned new *Candidatus*
33 names. Many of these genera are conserved across a range of mammalian gut
34 microbiomes.

35 **Conclusions.** Our metagenomic analyses provide new insights into the bacterial,
36 archaeal and bacteriophage components of the horse gut microbiome. The resulting
37 datasets provide a key resource for future high-resolution taxonomic and functional
38 studies on the equine gut microbiome.

39

40 **Introduction**

41 The horse has played a crucial role in human development and in the spread of human
42 populations (1). Domestication of the horse began at least 6,000 years ago and led to
43 diversification into numerous breeds, accompanied by significant biological changes (2). The
44 horse remains an important component of human society, with around 60 million horses
45 worldwide (3). Horses play roles as working animals across the globe, in transport,
46 agriculture or policing. Horse-riding and equine-assisted therapy provide health benefits,

47 while the horse remains an important food animal globally, with 5 million animals
48 slaughtered for food each year and horsemeat now in favour as a low-methane red-meat
49 alternative to beef (4). In the UK, there are around 374,000 horse-owning households and
50 horseracing is the second most attended sport in the country after football, contributing £4.7
51 billion to the UK economy (5).

52 As a foraging herbivore, the horse relies on a cellulose-rich diet of grass and
53 legumes. However, unlike cattle, horses have no rumen to digest complex carbohydrates.
54 Instead, they rely on hindgut fermentation: an efficient but enigmatic process—far less well
55 understood than ruminal digestion—that relies on a rich microbial community, the hindgut
56 microbiome, encompassing bacteria, archaea and viruses, together with fungi and other
57 eukaryotic microbes (6–8). This ecosystem plays key roles in nutrient assimilation and feed
58 conversion—effectively turning grass into horseflesh and horsepower. The horse gut also
59 acts as a reservoir of human and equine pathogens and of antimicrobial resistance (9).

60 Crucially, a range of diseases are known to be associated with disturbances in
61 hindgut microbial ecology, including foal diarrhoea, colitis, laminitis, colic and equine grass
62 sickness (10). Thus, by better understanding the equine hindgut microbiome, we stand to
63 inform interventions that can improve the health and welfare, performance, value and
64 longevity of horses.

65 Previous studies of the horse hindgut microbiome have documented a rich variety of
66 microorganisms (spanning phyla from all three domains of life) and have shown that the
67 taxonomic composition of this community varies with age, breed, disease status and has
68 changed during domestication (6, 7, 10–17). However, earlier studies have largely relied on
69 short-read meta-barcoding analyses of 16S rRNA gene sequences, which are limited in that
70 they fail to provide resolution down to the species or strain level, reveal nothing about the
71 population structures or functional repertoires of microbial species and fail to cover viruses
72 and eukaryotes. Thus, despite previous efforts—and drawing on comparisons with the
73 human microbiome, where new species are still being discovered (18, 19)—the horse
74 hindgut microbiome presents us with a vast unexplored landscape of taxonomic, ecological

75 and functional diversity, certain to encompass important, yet undiscovered roles. As in
76 studies of the human gut microbiome, faeces provide ready non-invasive access to the gut
77 contents. As part of the Alborada Well Foal study, a cohort study of foal gut microbial
78 development and health in later life, we applied shotgun metagenomics to five equine faecal
79 samples from 12-month-old thoroughbreds to expand our knowledge of this landscape.

80

81 **Materials and methods**

82 **Sample collection and storage**

83 Faecal samples were from five, 12-month-old Thoroughbred racehorses from the same
84 location in Ireland. Samples were collected as part of the Alborada Well Foal study, under
85 the University of Surrey's ethical review framework, project code: NERA-2017-007-SVM. All
86 horses were at pasture when sampled. 100 g of freshly evacuated faeces was collected from
87 each horse in sterile bijoux before immediate storage at 4°C on site at the stud. Once
88 shipped, faecal samples were aliquoted and stored at -80 °C until DNA extraction. Samples
89 were thawed and mixed before DNA extraction using the DNeasy PowerSoil kit (Qiagen),
90 following manufacturer's instructions. Extracted DNA was stored at -20 °C before further
91 analysis.

92

93 **Metagenomic sequencing and processing**

94 Illumina sequencing libraries were constructed as previously described by Ravi and
95 colleagues (2019) (20). Paired-end metagenomic sequencing was performed on the Illumina
96 NextSeq, before bioinformatic processing on the Cloud Infrastructure for Microbial
97 Bioinformatics (CLIMB) (21). Output reads (2x150bp) were assessed for quality using
98 FastQC v0.11.8 and then trimmed using Trimmomatic v0.36 configured to a minimum read
99 length of 40 (22, 23). All metagenomic samples described here can be accessed on the
100 Sequence Read Archive under BioProject ID PRJNA590977. Reads were aligned to the

101 horse genome (GCF_002863925.1) using Bowtie2 v2.3.4.1 (24), allowing removal of host
102 reads with SAMtools v1.3.1 (25).

103 Taxonomic profiling of sequencing reads was performed using Kraken 2 (26) to
104 search a microbial database built from archaeal, bacterial, fungal, protozoan, viral and
105 univec_core sequences in Refseq in January 2020. Bracken was used to estimate taxon
106 abundance from Kraken 2 profiles, accepting only those taxa with >1000 assigned reads
107 (27). Bracken-database files were generated using “bracken-build” on our microbial
108 database and visualised using Pavian (28).

109

110 **Metagenomic assembly and binning**

111 Host-depleted reads were assembled individually from each metagenomic sample with
112 MegaHIT (29), using kmer sizes 25,43,67,87 & 101, before assessing the quality of resulting
113 contiguous sequences (contigs) with anvi'o v7 (30). Filtered reads from each sample were
114 mapped against the associated assembly to provide an estimate of contig abundance using
115 Bowtie 2 (24). Resulting SAM files were converted to BAM files before being sorted and
116 indexed using SAMtools (25). Contig coverage depth was translated from each BAM file,
117 before separately binning contigs >1000 bp with MaxBin v2.2.6 (31) and CONCOCT v1.1.0
118 (32) and binning contigs >1500 bp with MetaBAT 2 v2.12.1 (33).

119 DAS Tool was applied to the output from all three bin predictors, generating a
120 catalogue of 196 bins from five samples (34). All bins were profiled against the BAM file for
121 their source metagenomic sample using the anvi'o 'anvi-profile' workflow (30). Using the
122 'anvi-interactive' tool, each bin was refined manually according to GC content, single copy
123 core gene (SCG) taxonomy and coverage as well as detection statistics. CheckM v1.0.11
124 (35) was used for quality assessment of all bins using the lineage_wf function. Bins showing
125 >50% completion and <10% contamination were assessed for quality score (defined as
126 estimated genome completeness score minus five times estimated contamination score)
127 (36). Bins with <70% completion and/or a quality score of <50 were categorised as low-
128 quality metagenome-assembled genomes (MAGs) (n=29); those with >70% completion,

129 <10% contamination and quality score >50 were categorised as medium-quality MAGs
130 (n=68) and those with >90% completion, <5% contamination and quality score >50 were
131 classified as high-quality MAGs (n=55).

132

133 **Taxonomic and phylogenetic profiling of MAGs**

134 Medium- and high-quality MAGs from all five samples were de-replicated at 95% average
135 nucleotide identity (ANI) with a default aligned fraction of >10% using dRep v2.0 (37), to
136 create a non-redundant species catalogue. Clustering at 99% ANI was used to identify a
137 non-redundant strain catalogue and select a representative MAG per strain. CompareM
138 v0.1.1 (38) was used to assign Average Amino-acid Identity (AAI) values followed by AAI
139 clustering at 60% to allow delineation at the genus level.

140 The Genome Taxonomy Database Toolkit (GTDB-Tk) v1.4.1 (39), the Contig
141 Annotation Tool (CAT/BAT) v5.2.3 (40) and ReferenceSeeker v1.4 (41) were used to
142 perform taxonomic assignment of representative MAGs at strain-level compared to the
143 'GTDB release 95', 'NCBI nr (2021-01-07)' and 'NCBI RefSeq release 201' databases,
144 respectively. Where taxonomic assignments differed between GTDB-Tk, CAT/BAT or
145 ReferenceSeeker, GTDB-Tk assignments took precedence. Only when no species-level
146 GTDB taxonomy was available did we adopt assignments according to CAT/BAT or
147 ReferenceSeeker (11% of assignments). Phylogeny for our final de-replicated catalogue of
148 MAGs was performed by aligning and concatenating a set of sixteen ribosomal protein
149 sequences (ribosomal proteins L1, L2, L3, L4, L5, L6, L14, L16, L18, L22, L24, S3, S8, S10,
150 S17 and S19), an approach previously used to reconstruct the tree of life (42). Ribosomal
151 sequences were extracted using anvio before alignment using MUSCLE v3.8.155 (43) and
152 refinement using trimAl v1.4 (44). A maximum-likelihood tree was constructed using
153 FastTree v2.1 (45). All novel metagenomic species were confirmed as monophyletic,
154 drawing on all publicly available genomes from the genus to which they had been assigned
155 by GTDB genus (with genomes retrieved from NCBI). Proteomes were predicted using
156 Prodigal v2.6.1 (46) before comparison against 400 universal marker proteins using

157 PhyloPhlAn v3.0.58 (47) in accordance with diamond v0.9.34 (48). Multiple sequence
158 alignment and subsequent refinement was performed using MAFFT v7.271 (49) and trimAl
159 v1.4 (44) before tree construction using FastTree v2.1 and RAxML v8.2.12 (45, 50). All trees
160 were subsequently visualised and manually annotated using iTol v5.7.

161

162 **Abundance and metabolic profiling of MAGs**

163 To estimate the proportion of reads within each BioSample represented by our final, de-
164 replicated MAG catalogue, contigs from the non-redundant MAG catalogue were
165 concatenated and filtered reads aligned back to this MAG database using Bowtie 2 (24).
166 Ordered BAM files were assessed using anvi'o (35) to calculate coverage statistics per-
167 contig, allowing the calculation of mean coverage across each assembled genome. Species
168 distribution analyses were conducted using the Vegan package in R (51) before visualisation
169 using ggplot2 (52).

170 Functional profiling of high- and medium-quality MAGs (n=123) was performed using
171 DRAM (Distilled and Refined Annotation of Metabolism) at a minimum contig length of
172 1000bp (53). Predicted amino-acid sequences identified by Prodigal in metagenome mode
173 (46) were searched against KOfam, Pfam, and CAZy databases. tRNA and rRNA
174 sequences were identified in MAGs using tRNAscan-SE (54) and Barrnap v0.9 (55),
175 respectively.

176

177 **Bacteriophage identification and characterisation**

178 VirSorter v1.0.5 (56) was applied to all contigs >5kb within each BioSample. Contig
179 sequences classified by VirSorter as Category 1 ("most confident") or Category 2 ("likely")
180 were considered for further analysis. Candidate bacteriophage sequences were assessed
181 for completeness and contamination, using CheckV v0.7.0 (57), retaining only the
182 sequences classified as "High-quality" (>90% completeness) or "complete". These
183 sequences were collated and de-replicated using rapid genome pairwise clustering at 95%
184 ANI with an aligned fraction of $\geq 70\%$ to generate a catalogue of bacteriophage genome

185 sequences. For dereplication clustering, all-vs-all genome comparisons were performed
186 using BLASTn before ANI based clustering using the ‘anicalc’ and ‘aniclust’ CheckV scripts
187 sequentially.

188 Bacteriophage contigs from the catalogue were used as queries in a BLASTn search
189 against the NCBI non-redundant nucleotide database (conducted on 21/12/2020) using an
190 e-value of $\leq 1e-5$. Only matches with a query cover $>50\%$ and percentage ID $>70\%$ were
191 selected as being significant. Initial taxonomic classification of phage genomes at order and
192 family level was performed using Demovir (58) against a viral subset of non-redundant
193 TrEMBL database with an e-value of $\leq 1e-5$. For each viral contig, individual coding
194 sequences were predicted using Prodigal (46), before concatenation for input into
195 vCONTACT2 v0.9.19 (59) for construction of a gene-sharing network incorporating a de-
196 replicated RefSeq database of reference prokaryotic virus genomes. The resulting network
197 was visualised using Cystoscape v3.8.0 (60).

198

199 **Results**

200 **Reference-based profiling documents microbial diversity**

201 Whole genome sequencing of five faecal samples derived from 12-month-old Thoroughbred
202 horses, each yielded >6 ng/ μ l DNA and collectively generated >280 million paired reads or
203 >84 Gbp of sequence data. Reads derived from the horse genome accounted for $<1\%$ of
204 reads from each sample (S1 Table). We initially analysed reads using the k-mer-based
205 program Kraken 2, followed by refined phylogenetic analysis via the allied program Bracken.
206 Such analyses revealed unexpected novelty and diversity in the equine faecal microbiome,
207 with $>65\%$ of sequence reads in each sample classified by Kraken as “unassigned”, i.e. from
208 unknown organisms (S2 Table). Assignable reads represented all three domains of life, as
209 well as viruses, although bacteria predominated, accounting for $>95\%$ of assigned reads.

210 Bacterial reads were predominantly assigned to the four phyla in the NCBI taxonomy
211 most commonly associated with animal gut microbiomes—Proteobacteria, Firmicutes,

212 Bacteroidetes and Actinobacteria. However, the Kraken 2 profiles also provided evidence of
213 over thirty additional bacterial phyla in this ecosystem. Many of these appear to be novel in
214 the context of the horse gut, including *Deinococcus-Thermus*, *Thermotogae* and the
215 *Candidatus* phylum Cloacimonetes (also called WWE1), which has been reported almost
216 exclusively from anaerobic fermenters and the aqueous environment (61, 62). However, as
217 this phylum has recently been detected in soil fertilised with manure from dairy cattle,
218 chickens and swine and has been implicated in anaerobic digestion of cellulose, it may play
219 important similar roles in the vertebrate gut (62, 63). Interestingly, in four of the five samples
220 more than a thousand reads were assigned to *Candidatus* Saccharibacteria, a phylum from
221 the candidate phyla radiation, which is home to bacteria that live as bacterial epibionts (64).
222 Reads assigned to eukaryotes provided evidence of budding yeasts and apicomplexan
223 parasites in these samples.

224 Remarkably, two samples show a very high relative abundance of reads assigned to
225 the genus *Acinetobacter* (20% and 9.4% of all reads or 57% and 34% of classified reads),
226 mirroring similar findings on two healthy horses in a previous study using 16S rRNA gene
227 sequences (65). Bracken assigns these reads to an implausible thirty-one species of
228 *Acinetobacter*, which is more likely to represent misassignment of reads rather than genuine
229 diversity within this genus in this context.

230

231 **Over a hundred newly named bacterial species**

232 We generated almost 200 non-redundant bins from single-sample assemblies using three
233 different approaches to binning. 123 bins represent medium- or high-quality metagenome-
234 assembled genomes (MAGs), 96 with ≥ 15 amino acid tRNAs (S3 and S4 Tables). Genome
235 sizes ranged from ~ 0.5 to 3.8 Mbp, while GC content ranged from 31% to 60%. De-
236 replication at 95% ANI clustered MAGs into 110 metagenomic species, spanning ten phyla
237 (Fig 1a). According to GTDB, around half (48%) of the metagenomic species belonged to
238 the *Bacteroidota*, while just over a third (35%) belonged to the *Firmicutes* (split by GTDB into
239 *Firmicutes*, *Firmicutes_A* and *Firmicutes_C*). Only twelve species of bacterial species from

240 the horse gut had been previously defined and delineated: eight with validly published Latin
241 binomials and four simply with alphanumerical designations assigned by GTDB (S5 Table).

242

243 **Fig 1. Taxonomic classification of 110 Metagenomic species derived from 5**

244 **metagenomic equine faecal samples. (A)** Depicted as a phylogenetic tree - where phylum,
245 as assigned by GTDB, is indicated by colour range. The tree was based upon an alignment
246 of 16 concatenated ribosomal proteins and constructed using FastTree. The final tree was
247 visualised and manually annotated using the online iTOLv5.7 tool. Phylum-level taxonomy is
248 described by branch colour according to GTDB designation (Phyla with an alphabetical suffix
249 have been collapsed). The presence (blue filled) or absence (hollow) of genes associated
250 with catalysing carbohydrate degradation (blue) or aiding in the metabolism of short chain
251 fatty acids (red) are reported in the associated binary plot. Hemicellulose substrates have
252 been clustered and highlighted in red text. **(B)** Average Nucleotide Identity (ANI) between
253 recovered MAGs and their closest representative within the GTDB database (release 95).
254 Only MAGs placed within a previously recognised genus, and whereby this taxonomic
255 assignment was inclusive of an ANI measurement, are shown. Individual plots are coloured
256 according to GTDB designated phylum, with phyla assigned an alphabetical suffix being
257 collapsed. A dotted line is placed at 95% ANI, representing the utilised species-level
258 boundary.

259

260 Two of the species with validly published names, *Ligilactobacillus hayakitensis*
261 (synonym *Lactobacillus hayakitensis*) (66) and *Limosilactobacillus equigenerosi* (synonym
262 *Lactobacillus equigenerosi*) (67), have been previously cultured from the faeces of
263 thoroughbred racehorses and are thought to be positively associated with equine intestinal
264 health (68). Similarly, the species *Streptococcus equinus* was named in the early twentieth
265 century after its association with horse dung and has been repeatedly isolated from this
266 source (69, 70). Another of the named species found among our MAGs, *Treponema*

267 *succinifaciens*, has been reported from the equine gut by 16S studies (71), but ours
268 represents the first report of a genome from this species in this setting.

269 The recently named species *Acinetobacter lanii* (72) has been isolated from the
270 Tibetan wild ass *Equus kiang*, but our MAG represents the first report of an association
271 between this species and the domesticated horse. Similarly, although the genus
272 *Phascolarctobacterium* is known to inhabit the horse gut (15, 16), here we provide the first
273 evidence of a specific link between the horse and the species *P. succinatutens*, previously
274 found in human and pig faeces (73, 74). Our metagenomic species provide the first report in
275 the horse of the species *Pseudomonas lundensis* (first isolated from meat, but now
276 recognised as an emerging pathogen of humans (75, 76)) and of *Akkermansia glycaniphila*,
277 previously isolated from the faeces of the reticulated python (77).

278 Among our metagenomic species, eighty-five represent new candidate species within
279 fifty bacterial genera previously delineated by GTDB (Fig 1b). All fifty of these genera occur
280 in the gut microbiota of at least one additional mammalian host species. Twelve of our
281 metagenomic species that could be assigned only to the level of family fell into ten clusters
282 (delineated at 60% AAI) representing novel candidate genera from seven different families
283 (S6 Table). The archaeal genus *Methanocorpusculum* is thought to play a role in methane
284 production in the equine gut (78). Here, we have delineated a novel species from this
285 ecosystem: *Candidatus Methanocorpusculum equi*.

286 Building on our recent efforts with the chicken gut microbiome and with the
287 automated creation of well-formed Latin names, we have created *Candidatus* names
288 (abbreviated as *Ca.*) for all the unnamed taxa revealed by our metagenomic analyses (Table
289 1). We also created Latin names for species and genera recognised by GTDB, but
290 previously assigned only alphanumeric designations. For taxa found only in the horse, we
291 created names that incorporated Greek or Latin roots for this host (e.g. *Ca. Equimonas*).
292 However, if searches of the GTDB and NCBI databases suggested that genera had
293 representatives in other gut microbiomes, we opted for names that specified gut or faeces as
294 habitat (e.g. *Ca. Limimonas*).

295

296 **Table 1. Protologues for newly named *Candidatus* genera and species.**

297 Protologues for new *Candidatus* taxa identified by analysis of metagenome-assembled
298 genomes from equine faeces.

299

300 **Description of *Candidatus Alistipes equi* sp. nov.**

301 *Candidatus Alistipes equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

302 A bacterial species identified by metagenomic analyses. This species includes all bacteria
303 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
304 species to which we have assigned the MAG ID E3_MB2_80 and which is available via
305 NCBI BioSample SAMN18472495. The GC content of the type genome is 40.8 % and the
306 genome length is 2.08 Mbp.

307

308 **Description of *Candidatus Apopatocola* gen. nov.**

309 *Candidatus Apopatocola* (A.po.pa.to'cola. Gr. masc. n. *apopatos*, dung; N.L. masc./fem.
310 suffix *-cola*, an inhabitant; N.L. fem. n. *Apopatocola* a microbe associated with faeces)

311 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
312 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
313 from the type species *Candidatus Apopatocola equi*. This is a new name for the GTDB
314 alphanumeric genus UBA738, which is found in diverse mammalian guts. This genus has
315 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
316 *Oscillospirales* and to the family *Oscillospiraceae*.

317

318 **Description of *Candidatus Apopatocola equi* sp. nov.**

319 *Candidatus Apopatocola equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

320 A bacterial species identified by metagenomic analyses. This species includes all bacteria
321 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
322 species to which we have assigned the MAG ID E1_MB2_75 and which is available via

323 NCBI BioSample SAMN18472466. The GC content of the type genome is 59.6 % and the
324 genome length is 1.56 Mbp.

325

326 **Description of *Candidatus Apopatosoma* gen. nov.**

327 *Candidatus Apopatosoma* (A.po.pa.to.so'ma. Gr. masc. n. *apopatos*, dung; Gr. neut. n.
328 *soma*, a body; N.L. neut. n. *Apopatosoma*, a microbe associated with faeces)

329 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
330 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
331 from the type species *Candidatus Apopatosoma equi*. This is a new name for the GTDB
332 alphanumeric genus CAG-724, which is found in diverse mammalian guts. This genus has
333 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
334 *Oscillospirales* and to the family CAG-272.

335

336 **Description of *Candidatus Apopatosoma equi* sp. nov.**

337 *Candidatus Apopatosoma equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

338 A bacterial species identified by metagenomic analyses. This species includes all bacteria
339 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
340 species to which we have assigned the MAG ID E2_100 and which is available via NCBI
341 BioSample SAMN18472471. The GC content of the type genome is 49.5 % and the genome
342 length is 1.54 Mbp.

343

344 **Description of *Candidatus Apopatosoma intestinale* sp. nov.**

345 *Candidatus Apopatosoma intestinale* (in.tes.ti.na'le. N.L. neut. adj. *intestinale*, pertaining to
346 the intestines)

347 A bacterial species identified by metagenomic analyses. This species includes all bacteria
348 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
349 species to which we have assigned the MAG ID E5_133 and which is available via NCBI
350 BioSample SAMN18472535. This is a new name for the alphanumeric GTDB species

351 sp003524145, which is found in diverse mammalian guts. The GC content of the type
352 genome is 53.8 % and the genome length is 1.55 Mbp.

353

354 **Description of *Candidatus Apopatousia* gen. nov.**

355 *Candidatus Apopatousia* (A.po.pat.ou's.ia. Gr. masc. n. *apopatos*, dung; Gr. fem. n. *ousia*,
356 an essence; N.L. fem. n. *Apopatousia*, a microbe associated with faeces)

357 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
358 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
359 from the type species *Candidatus Apopatousia equi*. This is a new name for the GTDB
360 alphanumeric genus UBA9845, which is found in diverse mammalian guts. This genus has
361 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
362 *Christensenellales* and to the family *UBA1242*.

363

364 **Description of *Candidatus Apopatousia equi* sp. nov.**

365 *Candidatus Apopatousia equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

366 A bacterial species identified by metagenomic analyses. This species includes all bacteria
367 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
368 species to which we have assigned the MAG ID E5_MB2_6 and which is available via NCBI
369 BioSample SAMN18472550. The GC content of the type genome is 31.9 % and the genome
370 length is 0.57 Mbp.

371

372 **Description of *Candidatus Blautia equi* sp. nov.**

373 *Candidatus Blautia equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

374 A bacterial species identified by metagenomic analyses. This species includes all bacteria
375 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
376 species to which we have assigned the MAG ID E4_MB2_89 and which is available via
377 NCBI BioSample SAMN18472531. GTDB has assigned. This species to a genus marked
378 with an alphabetical suffix. However, as this genus designation cannot be incorporated into a

379 well-formed binomial, in naming. This species, we have used the current validly published
380 name for the genus. The GC content of the type genome is 48 % and the genome length is
381 2.14 Mbp.

382

383 **Description of *Candidatus Caballimonas* gen. nov.**

384 *Candidatus Caballimonas* (Ca.bal.li.mo'nas. L. masc. n. *caballus*, a horse; L. fem. n. *monas*,
385 a monad; N.L. fem. n. *Caballimonas*, a microbe associated with horses)

386 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
387 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
388 from the type species *Candidatus Caballimonas caccae*. This genus has been assigned by
389 GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
390 *Christensenellales* and to the family *Borkfalkiaceae*.

391

392 **Description of *Candidatus Caballimonas caccae* sp. nov.**

393 *Candidatus Caballimonas caccae* (cac'cae. Gr. fem. n. *kakke*, faeces; N.L. gen. n. *caccae*,
394 of faeces)

395 A bacterial species identified by metagenomic analyses. This species includes all bacteria
396 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
397 species to which we have assigned the MAG ID E3_31 and which is available via NCBI
398 BioSample SAMN18472486. The GC content of the type genome is 34.9 % and the genome
399 length is 0.91 Mbp.

400

401 **Description of *Candidatus Cacconaster* gen. nov.**

402 *Candidatus Cacconaster* (Cac.co.nas'ter. Gr. fem. n. *kakke*, dung; Gr. masc. n. *naster*, an
403 inhabitant; N.L. masc. n. *Cacconaster*, a microbe associated with faeces)

404 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
405 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
406 from the type species *Candidatus Cacconaster caballi*. This is a new name for the GTDB

407 alphanumeric genus Bact-11, which is found in diverse mammalian guts. This genus has
408 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
409 *Bacteroidales* and to the family *UBA932*,

410

411 **Description of *Candidatus Cacconaster caballi* sp. nov.**

412 *Candidatus Cacconaster caballi* (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

413 A bacterial species identified by metagenomic analyses. This species includes all bacteria
414 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
415 species to which we have assigned the MAG ID E2_MB2_69 and which is available via
416 NCBI BioSample SAMN18472478. The GC content of the type genome is 50.7 % and the
417 genome length is 1.38 Mbp.

418

419 **Description of *Candidatus Cacconaster equi* sp. nov.**

420 *Candidatus Cacconaster equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

421 A bacterial species identified by metagenomic analyses. This species includes all bacteria
422 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
423 species to which we have assigned the MAG ID E1_MB2_89 and which is available via
424 NCBI BioSample SAMN18472469. The GC content of the type genome is 48.5 % and the
425 genome length is 1.65 Mbp.

426

427 **Description of *Candidatus Cacconaster equifaecalis* sp. nov.**

428 *Candidatus Cacconaster equifaecalis* (e.qui.fae.ca'lis. L. masc. n. *equus*, a horse; N.L.

429 masc. adj. *faecalis*, faecal; N.L. masc. adj. *equifaecalis*, associated with the faeces of

430 horses)

431 A bacterial species identified by metagenomic analyses. This species includes all bacteria
432 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
433 species to which we have assigned the MAG ID E5_MB2_108 and which is available via

434 NCBI BioSample SAMN18472541. The GC content of the type genome is 51.7 % and the
435 genome length is 1.71 Mbp.

436

437 **Description of *Candidatus Cacconaster merdequi* sp. nov.**

438 *Candidatus Cacconaster merdequi* (merd.e'qui. L. fem. n. *merda*, faeces; L. masc. n. *equus*,
439 a horse; N.L. gen. n. *merdequi*, associated with the faeces of horses)

440 A bacterial species identified by metagenomic analyses. This species includes all bacteria
441 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
442 species to which we have assigned the MAG ID E5_MB2_33 and which is available via

443 NCBI BioSample SAMN18472547. The GC content of the type genome is 49 % and the
444 genome length is 1.90 Mbp.

445

446 **Description of *Candidatus Cacconaster scatequi* sp. nov.**

447 *Candidatus Cacconaster scatequi* (scat.e'qui. Gr. neut. n. *skor*, *skatos*, dung; L. masc.
448 n. *equus*, a horse; N.L. gen. n. *scatequi*, associated with the faeces of horses)

449 A bacterial species identified by metagenomic analyses. This species includes all bacteria
450 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
451 species to which we have assigned the MAG ID E3_MB2_97 and which is available via

452 NCBI BioSample SAMN18472499. The GC content of the type genome is 50.6 % and the
453 genome length is 1.90 Mbp.

454

455 **Description of *Candidatus Cacconaster stercorequi* sp. nov.**

456 *Candidatus Cacconaster stercorequi* (ster.cor.e'qui. L. masc. n. *stercus*, *stercoris*, dung; L.
457 masc. n. *equus*, a horse; N.L. gen. n. *stercorequi*, associated with the faeces of horses)

458 A bacterial species identified by metagenomic analyses. This species includes all bacteria
459 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
460 species to which we have assigned the MAG ID E4_MB2_17 and which is available via

461 NCBI BioSample SAMN18472518. The GC content of the type genome is 54.5 % and the
462 genome length is 1.83 Mbp.

463

464 **Description of *Candidatus Chryseobacterium enterohippi* sp. nov.**

465 *Candidatus Chryseobacterium enterohippi* (en.te.ro.hip'pi. Gr. neut. n. *enteron*, gut, bowel,
466 intestine; Gr. masc./fem. n. *hippos*, a horse; N.L. gen. n. *enterohippi*, associated with the
467 horse gut)

468 A bacterial species identified by metagenomic analyses. This species includes all bacteria
469 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
470 species to which we have assigned the MAG ID E1_189 and which is available via NCBI
471 BioSample SAMN18472455. The GC content of the type genome is 34.3 % and the genome
472 length is 2.05 Mbp.

473

474 **Description of *Candidatus Colenecus* gen. nov.**

475 *Candidatus Colenecus* (Col.en.e'cus. L. neut. n. *colon*, large intestine; N.L. masc. n. *enecus*,
476 an inhabitant; N.L. masc. n. *Colenecus*, a microbe associated with the large intestine)

477 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
478 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
479 from the type species *Candidatus Colenecus caballi*. This is a new name for the GTDB
480 alphanumeric genus UBA1179, which is found in diverse mammalian guts. This genus has
481 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
482 *Bacteroidales* and to the family *Bacteroidaceae*.

483

484 **Description of *Candidatus Colenecus caballi* sp. nov.**

485 *Candidatus Colenecus caballi* (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

486 A bacterial species identified by metagenomic analyses. This species includes all bacteria
487 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
488 species to which we have assigned the MAG ID E3_160 and which is available via NCBI

489 BioSample SAMN18472483. The GC content of the type genome is 49.7 % and the genome
490 length is 2.25 Mbp.

491

492 **Description of *Candidatus Colicola* gen. nov.**

493 *Candidatus Colicola* (Co.li.co'la. L. neut. n. *colon*, large intestine; N.L. masc./fem. suffix –
494 *cola*, an inhabitant; N.L. fem. n. *Colicola*, a microbe associated with the large intestine)

495 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
496 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
497 from the type species *Candidatus Colicola* *caballi*. This is a new name for the GTDB
498 alphanumeric genus RF16, which is found in diverse mammalian guts. This genus has been
499 assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
500 *Bacteroidales* and to the family *Paludibacteraceae*.

501

502 **Description of *Candidatus Colicola* *caballi* sp. nov.**

503 *Candidatus Colicola* *caballi* (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

504 A bacterial species identified by metagenomic analyses. This species includes all bacteria
505 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
506 species to which we have assigned the MAG ID E1_MB2_58 and which is available via
507 NCBI BioSample SAMN18472465. The GC content of the type genome is 46.6 % and the
508 genome length is 1.45 Mbp.

509

510 **Description of *Candidatus Colicola* *caccequi* sp. nov.**

511 *Candidatus Colicola* *caccequi* (cacc.e'qui. Gr. fem. n. *kakke*, faeces; L. masc. n. *equus*, a
512 horse; N.L. gen. n. *caccequi*, associated with the faeces of horses)

513 A bacterial species identified by metagenomic analyses. This species includes all bacteria
514 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
515 species to which we have assigned the MAG ID E4_134 and which is available via NCBI

516 BioSample SAMN18472502. The GC content of the type genome is 44.2 % and the genome
517 length is 1.71 Mbp.

518

519 **Description of *Candidatus Colicola coprequi* sp. nov.**

520 *Candidatus Colicola coprequi* (copr.e'qui. Gr. fem. n. *kopros*, dung; L. masc. n. *equus*, a
521 horse; N.L. gen. n. *coprequi*, associated with the faeces of horses)

522 A bacterial species identified by metagenomic analyses. This species includes all bacteria
523 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
524 species to which we have assigned the MAG ID E2_MB2_30 and which is available via
525 NCBI BioSample SAMN18472476. The GC content of the type genome is 46.1 % and the
526 genome length is 1.53 Mbp.

527

528 **Description of *Candidatus Colicola equi* sp. nov.**

529 *Candidatus Colicola equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

530 A bacterial species identified by metagenomic analyses. This species includes all bacteria
531 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
532 species to which we have assigned the MAG ID E1_186 and which is available via NCBI
533 BioSample SAMN18472454. The GC content of the type genome is 44.4 % and the genome
534 length is 2.05 Mbp.

535

536 **Description of *Candidatus Colicola faecequi* sp. nov.**

537 *Candidatus Colicola faecequi* (faec.e'qui. L. fem. n. *faex*, *faeces*, *dregs*; L. masc. n. *equus*, a
538 horse; N.L. gen. n. *faecequi*, associated with the faeces of horses)

539 A bacterial species identified by metagenomic analyses. This species includes all bacteria
540 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
541 species to which we have assigned the MAG ID E4_MB2_124 and which is available via
542 NCBI BioSample SAMN18472515. The GC content of the type genome is 52.3 % and the
543 genome length is 1.86 Mbp.

544

545 **Description of *Candidatus Colimonas* gen. nov.**

546 *Candidatus* Colimonas (Co.li.mo'nas. L. neut. n. *colon*, large intestine; L. fem. n. *monas*, a
547 monad; N.L. fem. n. *Colimonas*, a microbe associated with the large intestine)

548 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
549 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
550 from the type species *Candidatus* Colimonas fimequi. This is a new name for the GTDB
551 alphanumeric genus UBA1191, which is found in diverse mammalian guts. This genus has
552 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
553 *Peptostreptococcales* and to the family *Anaerovoracaceae*.

554

555 **Description of *Candidatus Colimonas fimequi* sp. nov.**

556 *Candidatus* Colimonas fimequi (fim.e'qui. L. masc. n. *firmus*, dung; L. masc. n. *equus*, a
557 horse; N.L. gen. n. *fimequi*, associated with the faeces of horses)

558 A bacterial species identified by metagenomic analyses. This species includes all bacteria
559 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
560 species to which we have assigned the MAG ID E4_13 and which is available via NCBI
561 BioSample SAMN18472501. The GC content of the type genome is 44.3 % and the genome
562 length is 1.70 Mbp.

563

564 **Description of *Candidatus Colimorpha* gen. nov.**

565 *Candidatus* Colimorpha (Co.li.mor'pha. L. neut. n. *colon*, large intestine; Gr. fem. n. *morphe*,
566 a form, shape; N.L. fem. n. *Colimorpha*, a microbe associated with the large intestine)

567 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
568 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
569 from the type species *Candidatus* Colimorpha merdihippi. This is a new name for the GTDB
570 alphanumeric genus UBA1711, which is found in diverse mammalian guts. This genus has

571 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
572 *Bacteroidales* and to the family *P3*.

573

574 **Description of *Candidatus Colimorpha enterica* sp. nov.**

575 *Candidatus Colimorpha enterica* (en.te'ri.ca. Gr. neut. n. *enteron*, gut, bowel, intestine; L.
576 fem. adj. suff. *-ica*, pertaining to; N.L. fem. adj. *enterica*, pertaining to intestine)

577 A bacterial species identified by metagenomic analyses. This species includes all bacteria
578 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
579 species to which we have assigned the MAG ID E3_60 and which is available via NCBI
580 BioSample SAMN18472488. This is a new name for the alphanumeric GTDB species
581 sp000433515, which is found in diverse mammalian guts. The GC content of the type
582 genome is 52.3 % and the genome length is 1.43 Mbp.

583

584 **Description of *Candidatus Colimorpha merdhippi* sp. nov.**

585 *Candidatus Colimorpha merdhippi* (mer.di.hip'pi. L. fem. n. *merda*, faeces; Gr. masc./fem. n.
586 *hippos*, a horse; N.L. gen. n. *merdhippi*, associated with the faeces of horses)

587 A bacterial species identified by metagenomic analyses. This species includes all bacteria
588 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
589 species to which we have assigned the MAG ID E1_90 and which is available via NCBI
590 BioSample SAMN18472457. The GC content of the type genome is 48.5 % and the genome
591 length is 3.11 Mbp.

592

593 **Description of *Candidatus Colimorpha onthhippi* sp. nov.**

594 *Candidatus Colimorpha onthhippi* (on.tho.hip'pi. Gr. masc. n. *onthos*, dung; Gr. masc./fem.
595 n. *hippos*, a horse; N.L. gen. n. *onthhippi*, associated with the faeces of horses)

596 A bacterial species identified by metagenomic analyses. This species includes all bacteria
597 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
598 species to which we have assigned the MAG ID E5_36 and which is available via NCBI

599 BioSample SAMN18472537. The GC content of the type genome is 46.2 % and the genome
600 length is 2.04 Mbp.

601

602 **Description of *Candidatus Colimorpha pelethequi* sp. nov.**

603 *Candidatus* Colimorpha pelethequi (pe.leth.e'qui. Gr. masc. n. *pelethos*, dung; L. masc.
604 n. *equus*, a horse; N.L. gen. n. *pelethequi*, associated with the faeces of horses)

605 A bacterial species identified by metagenomic analyses. This species includes all bacteria
606 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
607 species to which we have assigned the MAG ID E5_MB2_81 and which is available via
608 NCBI BioSample SAMN18472551. The GC content of the type genome is 46.7 % and the
609 genome length is 2.38 Mbp.

610

611 **Description of *Candidatus Colinaster* gen. nov.**

612 *Candidatus* Colinaster (Co.li.nas'ter. L. neut. n. *colon*, large intestine; Gr. masc. n. *naster*, an
613 inhabitant; N.L. masc. n. *Colinaster* a microbe associated with the large intestine)

614 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
615 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
616 from the type species *Candidatus* Colinaster scatohippi. This is a new name for the GTDB
617 alphanumeric genus UBA1712, which is found in diverse mammalian guts. This genus has
618 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
619 *Lachnospirales* and to the family *Lachnospiraceae*.

620

621 **Description of *Candidatus Colinaster scatohippi* sp. nov.**

622 *Candidatus* Colinaster scatohippi (sca.to.hip'pi. Gr. neut. n. *skor*, *skatos*, dung; Gr.
623 masc./fem. n. *hipposa* horse; N.L. gen. n. *scatohippi*, associated with the faeces of horses)

624 A bacterial species identified by metagenomic analyses. This species includes all bacteria
625 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
626 species to which we have assigned the MAG ID E4_MB2_45 and which is available via

627 NCBI BioSample SAMN18472524. The GC content of the type genome is 38.7 % and the
628 genome length is 2.18 Mbp.

629

630 **Description of *Candidatus Coliplasma* gen. nov.**

631 *Candidatus Coliplasma* (Co.li.plas'ma. L. neut. n. *colon*, large intestine; Gr. neut. n. *plasma*,
632 a form; N.L. neut. n. *Coliplasma*, a microbe associated with the large intestine)

633 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
634 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
635 from the type species *Candidatus Coliplasma caballi*. This is a new name for the GTDB
636 alphanumeric genus UBA1752, which is found in diverse mammalian guts. This genus has
637 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
638 *Oscillospirales* and to the family CAG-382.

639

640 **Description of *Candidatus Coliplasma caballi* sp. nov.**

641 *Candidatus Coliplasma caballi* (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

642 A bacterial species identified by metagenomic analyses. This species includes all bacteria
643 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
644 species to which we have assigned the MAG ID E3_MB2_28 and which is available via
645 NCBI BioSample SAMN18472492. The GC content of the type genome is 54.8 % and the
646 genome length is 1.41 Mbp.

647

648 **Description of *Candidatus Coliplasma equi* sp. nov.**

649 *Candidatus Coliplasma equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

650 A bacterial species identified by metagenomic analyses. This species includes all bacteria
651 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
652 species to which we have assigned the MAG ID E3_142 and which is available via NCBI
653 BioSample SAMN18472481. The GC content of the type genome is 49.7 % and the genome
654 length is 1.52 Mbp.

655

656 **Description of *Candidatus Colisoma* gen. nov.**

657 *Candidatus Colisoma* (Co.li.so'ma. L. neut. n. *colon*, large intestine; Gr. neut. n. *soma*, a
658 body; N.L. neut. n. *Colisoma*, a microbe associated with the large intestine)

659 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
660 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
661 from the type species *Candidatus Colisoma equi*. This is a new name for the GTDB
662 alphanumeric genus UBA1067, which is found in diverse mammalian guts. This genus has
663 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
664 *RFP12* and to the family *UBA1067*.

665

666 **Description of *Candidatus Colisoma equi* sp. nov.**

667 *Candidatus Colisoma equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

668 A bacterial species identified by metagenomic analyses. This species includes all bacteria
669 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
670 species to which we have assigned the MAG ID E4_MB2_14 and which is available via
671 NCBI BioSample SAMN18472517. The GC content of the type genome is 60 % and the
672 genome length is 2.52 Mbp.

673

674 **Description of *Candidatus Colivicinus* gen. nov.**

675 *Candidatus Colivicinus* (Co.li.vi'ci.nus. L. neut. n. *colon*, large intestine; N.L. masc. n.
676 *vicinus*, a neighbour; N.L. masc. n. *Colivicinus*, a microbe associated with the large intestine)

677 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
678 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
679 from the type species *Candidatus Colivicinus equi*. This is a new name for the GTDB
680 alphanumeric genus UBA636, which is found in diverse mammalian guts. This genus has
681 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
682 *Erysipelotrichales* and to the family *Erysipelotrichaceae*.

683

684 **Description of *Candidatus Colivicinus equi* sp. nov.**

685 *Candidatus Colivicinus equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

686 A bacterial species identified by metagenomic analyses. This species includes all bacteria
687 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
688 species to which we have assigned the MAG ID E4_MB2_36 and which is available via
689 NCBI BioSample SAMN18472522. The GC content of the type genome is 31.9 % and the
690 genome length is 1.69 Mbp.

691

692 **Description of *Candidatus Colivivens* gen. nov.**

693 *Candidatus Colivivens* (Co.li.vi'vens. L. neut. n. *colon*, large intestine; N.L. masc./fem. pres.
694 part. *vivens*, living; N.L. fem. n. *Colivivens*, a microbe associated with the large intestine)

695 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
696 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
697 from the type species *Candidatus Colivivens caballi*. This is a new name for the GTDB
698 alphanumeric genus UBA1786, which is found in diverse mammalian guts. This genus has
699 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
700 *Bacteroidales* and to the family *Bacteroidaceae*.

701

702 **Description of *Candidatus Colivivens caballi* sp. nov.**

703 *Candidatus Colivivens caballi* (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

704 A bacterial species identified by metagenomic analyses. This species includes all bacteria
705 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
706 species to which we have assigned the MAG ID E3_198 and which is available via NCBI
707 BioSample SAMN18472484. The GC content of the type genome is 47.7 % and the genome
708 length is 2.55 Mbp.

709

710 **Description of *Candidatus Colivivens equi* sp. nov.**

711 *Candidatus Colivivens equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

712 A bacterial species identified by metagenomic analyses. This species includes all bacteria
713 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
714 species to which we have assigned the MAG ID E1_MB2_52 and which is available via
715 NCBI BioSample SAMN18472463. The GC content of the type genome is 38.2 % and the
716 genome length is 2.64 Mbp.

717

718 **Description of *Candidatus Colousia* gen. nov.**

719 *Candidatus Colousia* (Col.ou's.ia. L. neut. n. *colon*, large intestine; Gr. fem. n. *ousia*, an
720 essence; N.L. fem. n. *Colousia*, a microbe associated with the large intestine)

721 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
722 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
723 from the type species *Candidatus Colousia faecequi*. This is a new name for the GTDB
724 alphanumeric genus SFVR01, which is found in diverse mammalian guts. This genus has
725 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
726 *Bacteroidales* and to the family *Paludibacteraceae*.

727

728 **Description of *Candidatus Colousia faecequi* sp. nov.**

729 *Candidatus Colousia faecequi* (faec.e'qui. L. fem. n. *faex*, *faeces*, *dregs*; L. masc. n. *equus*,
730 a horse; N.L. gen. n. *faecequi*, associated with the faeces of horses)

731 A bacterial species identified by metagenomic analyses. This species includes all bacteria
732 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
733 species to which we have assigned the MAG ID E3_MB2_91 and which is available via
734 NCBI BioSample SAMN18472498. The GC content of the type genome is 47.1 % and the
735 genome length is 1.67 Mbp.

736

737 **Description of *Candidatus Comamonas equi* sp. nov.**

738 *Candidatus Comamonas equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

739 A bacterial species identified by metagenomic analyses. This species includes all bacteria
740 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
741 species to which we have assigned the MAG ID E2_118 and which is available via NCBI
742 BioSample SAMN18472472. The GC content of the type genome is 59.2 % and the genome
743 length is 2.60 Mbp.

744

745 **Description of *Candidatus Copronaster* gen. nov.**

746 *Candidatus Copronaster* (Co.pro.nas'ter. Gr. fem. n. *kopros*, dung; Gr. masc. n. *naster*, an
747 inhabitant; N.L. masc. n. *Copronaster*, a microbe associated with faeces)

748 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
749 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
750 from the type species *Candidatus Copronaster equi*. This is a new name for the GTDB
751 alphanumeric genus CAG-488, which is found in diverse mammalian guts. This genus has
752 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
753 *Oscillospirales* and to the family *Acutalibacteraceae*.

754

755 **Description of *Candidatus Copronaster equi* sp. nov.**

756 *Candidatus Copronaster equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

757 A bacterial species identified by metagenomic analyses. This species includes all bacteria
758 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
759 species to which we have assigned the MAG ID E5_MB2_59 and which is available via
760 NCBI BioSample SAMN18472549. The GC content of the type genome is 39.1 % and the
761 genome length is 1.98 Mbp.

762

763 **Description of *Candidatus Cryptobacteroides aphodequi* sp. nov.**

764 *Candidatus Cryptobacteroides aphodequi* (aph.od.e'qui. Gr. fem. n. *aphodos*, dung; L. masc.
765 n. *equus*, a horse; N.L. gen. n. *aphodequi*, associated with the faeces of horses)

766 A bacterial species identified by metagenomic analyses. This species includes all bacteria
767 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
768 species to which we have assigned the MAG ID E3_MB2_98 and which is available via
769 NCBI BioSample SAMN18472500. The GC content of the type genome is 54.9 % and the
770 genome length is 1.48 Mbp.

771

772 **Description of *Candidatus Cryptobacteroides caccocaballi* sp. nov.**

773 *Candidatus Cryptobacteroides caccocaballi* (cac.co.ca.bal'li. Gr. fem. n. *kakke*, faeces; L.
774 masc. n. *caballus*, a horse; N.L. gen. n. *caccocaballi*, associated with the faeces of horses)

775 A bacterial species identified by metagenomic analyses. This species includes all bacteria
776 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
777 species to which we have assigned the MAG ID E4_MB2_58 and which is available via
778 NCBI BioSample SAMN18472527. The GC content of the type genome is 51.8 % and the
779 genome length is 2.22 Mbp.

780

781 **Description of *Candidatus Cryptobacteroides choladohippi* sp. nov.**

782 *Candidatus Cryptobacteroides choladohippi* (cho.la.do.hip'pi. Gr. fem. n. *kholas*, *kholados*,
783 guts; Gr. masc./fem. n. *hipposa* horse; N.L. gen. n. *choladohippi*, associated with the horse
784 gut)

785 A bacterial species identified by metagenomic analyses. This species includes all bacteria
786 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
787 species to which we have assigned the MAG ID E1_MB2_55 and which is available via
788 NCBI BioSample SAMN18472464. The GC content of the type genome is 54.2 % and the
789 genome length is 2.24 Mbp.

790

791 **Description of *Candidatus Cryptobacteroides equifaecalis* sp. nov.**

792 *Candidatus* Cryptobacteroides equifaecalis (e.qui.fae.ca'lis. L. masc. n. *equus*, a horse; N.L.
793 masc. adj. *faecalis*, faecal; N.L. masc. adj. *equifaecalis*, associated with the faeces of
794 horses)

795 A bacterial species identified by metagenomic analyses. This species includes all bacteria
796 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
797 species to which we have assigned the MAG ID E4_MB2_98 and which is available via
798 NCBI BioSample SAMN18472533. The GC content of the type genome is 52.5 % and the
799 genome length is 1.61 Mbp.

800

801 **Description of *Candidatus* Cryptobacteroides faecihippi sp. nov.**

802 *Candidatus* Cryptobacteroides faecihippi (fae.ci.hip'pi. L. fem. n. *faex*, *faeces*, *dregs*; Gr.
803 masc./fem. n. *hipposa* horse; N.L. gen. n. *faecihippi*, associated with the faeces of horses)

804 A bacterial species identified by metagenomic analyses. This species includes all bacteria
805 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
806 species to which we have assigned the MAG ID E1_MB2_112 and which is available via
807 NCBI BioSample SAMN18472461. The GC content of the type genome is 54.8 % and the
808 genome length is 2.25 Mbp.

809

810 **Description of *Candidatus* Cryptobacteroides fimicaballi sp. nov.**

811 *Candidatus* Cryptobacteroides fimicaballi (fi.mi.ca.bal'li. L. masc. n. *fimius*, dung; L. masc.
812 n. *caballus*, a horse; N.L. gen. n. *fimicaballi*, associated with the faeces of horses)

813 A bacterial species identified by metagenomic analyses. This species includes all bacteria
814 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
815 species to which we have assigned the MAG ID E3_MB2_135 and which is available via
816 NCBI BioSample SAMN18472490. The GC content of the type genome is 51 % and the
817 genome length is 1.33 Mbp.

818

819 **Description of *Candidatus* Cryptobacteroides onthequi sp. nov.**

820 *Candidatus* Cryptobacteroides onthequi (onth.e'qui. Gr. masc. n. *onthos*, dung; L. masc.
821 n. *equus*, a horse; N.L. gen. n. *onthequi*, associated with the faeces of horses)
822 A bacterial species identified by metagenomic analyses. This species includes all bacteria
823 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
824 species to which we have assigned the MAG ID E1_MB2_10 and which is available via
825 NCBI BioSample SAMN18472459. The GC content of the type genome is 53.4 % and the
826 genome length is 2.96 Mbp.

827

828 **Description of *Candidatus* Cryptobacteroides onthocaballi sp. nov.**

829 *Candidatus* Cryptobacteroides onthocaballi (on.tho.ca.bal'li. Gr. masc. n. *onthos*, dung; L.
830 masc. n. *caballus*, a horse; N.L. gen. n. *onthocaballi*, associated with the faeces of horses)
831 A bacterial species identified by metagenomic analyses. This species includes all bacteria
832 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
833 species to which we have assigned the MAG ID E3_MB2_147 and which is available via
834 NCBI BioSample SAMN18472491. The GC content of the type genome is 52 % and the
835 genome length is 1.47 Mbp.

836

837 **Description of *Candidatus* Egerieousia equi sp. nov.**

838 *Candidatus* Egerieousia equi (e'qui. L. gen. masc. n. *equi*, of a horse)
839 A bacterial species identified by metagenomic analyses. This species includes all bacteria
840 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
841 species to which we have assigned the MAG ID E4_MB2_106 and which is available via
842 NCBI BioSample SAMN18472513. The GC content of the type genome is 46.4 % and the
843 genome length is 1.92 Mbp.

844

845 **Description of *Candidatus* Enterousia merdequi sp. nov.**

846 *Candidatus* Enterousia merdequi (merd.e'qui. L. fem. n. *merda*, faeces; L. masc. n. *equus*, a
847 horse; N.L. gen. n. *merdequi*, associated with the faeces of horses)

848 A bacterial species identified by metagenomic analyses. This species includes all bacteria
849 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
850 species to which we have assigned the MAG ID E3_MB2_90 and which is available via
851 NCBI BioSample SAMN18472497. The GC content of the type genome is 33.9 % and the
852 genome length is 0.74 Mbp.

853

854 **Description of *Candidatus Enterousia onthequi* sp. nov.**

855 *Candidatus* Enterousia onthequi (onth.e'qui. Gr. masc. n. *onthos*, dung; L. masc. n. *equus*, a
856 horse; N.L. gen. n. *onthequi*, associated with the faeces of horses)

857 A bacterial species identified by metagenomic analyses. This species includes all bacteria
858 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
859 species to which we have assigned the MAG ID E5_MB2_19 and which is available via
860 NCBI BioSample SAMN18472546. The GC content of the type genome is 38.8 % and the
861 genome length is 0.88 Mbp.

862

863 **Description of *Candidatus Enterousia scatequi* sp. nov.**

864 *Candidatus* Enterousia scatequi (scat.e'qui. Gr. neut. n. *skor*, *skatos*, dung; L. masc.
865 n. *equus*, a horse; N.L. gen. n. *scatequi*, associated with the faeces of horses)

866 A bacterial species identified by metagenomic analyses. This species includes all bacteria
867 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
868 species to which we have assigned the MAG ID E5_MB2_120 and which is available via
869 NCBI BioSample SAMN18472543. The GC content of the type genome is 39.9 % and the
870 genome length is 0.76 Mbp.

871

872 **Description of *Candidatus Equadaptatus* gen. nov.**

873 *Candidatus* Equadaptatus (Equ.a.dap.ta'tus. L. masc. n. *equus*, a horse; L. masc. perf. part.
874 *adaptatus*, adapted to; N.L. masc. n. *Equiadaptatus*, a microbe associated with horses)

875 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
876 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
877 from the type species *Candidatus Equadaptatus faecalis*. This genus has been assigned by
878 GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order *Synergistales* and
879 to the family *Synergistaceae*.

880

881 **Description of *Candidatus Equadaptatus faecalis* sp. nov.**

882 *Candidatus Equadaptatus faecalis* (fae.ca'lis. N.L. masc. adj. *faecalis*, faecal)

883 A bacterial species identified by metagenomic analyses. This species includes all bacteria
884 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
885 species to which we have assigned the MAG ID E4_60 and which is available via NCBI
886 BioSample SAMN18472510. The GC content of the type genome is 48.4 % and the genome
887 length is 1.60 Mbp.

888

889 **Description of *Candidatus Equibacterium* gen. nov.**

890 *Candidatus Equibacterium* (E.qui.bac.te'ri.um. L. masc. n. *equus*, a horse; L. neut. n.
891 *bacterium*, a bacterium; N.L. neut. n. *Equibacterium* a microbe associated with horses)

892 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
893 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
894 from the type species *Candidatus Equibacterium intestinale*. This genus has been assigned
895 by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order *Bacteroidales*
896 and to the family *UBA932*.

897

898 **Description of *Candidatus Equibacterium intestinale* sp. nov.**

899 *Candidatus Equibacterium intestinale* (in.tes.ti.na'le.N.L. neut. adj. *intestinale*, pertaining to
900 the intestines)

901 A bacterial species identified by metagenomic analyses. This species includes all bacteria
902 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the

903 species to which we have assigned the MAG ID E5_MB2_82 and which is available via
904 NCBI BioSample SAMN18472552. The GC content of the type genome is 52.3 % and the
905 genome length is 1.76 Mbp.

906

907 **Description of *Candidatus Equicaccousia* gen. nov.**

908 *Candidatus* Equicaccousia (E.qui.cacc.ou's.ia. L. masc. n. *equus*, a horse; Gr. fem. n. *kakke*,
909 faeces; Gr. fem. n. *ousia*, an essence; N.L. fem. n. *Equicaccousia*, a microbe associated
910 with horse faeces)

911 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
912 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
913 from the type species *Candidatus* Equicaccousia limhippi. This is a new name for the GTDB
914 alphanumeric genus UMGS1279, which is found in diverse mammalian guts. This genus has
915 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
916 *Oscillospirales* and to the family *Acutalibacteraceae*.

917

918 **Description of *Candidatus Equicaccousia limhippi* sp. nov.**

919 *Candidatus* Equicaccousia limhippi (li.mi.hip'pi. L. masc. n. *limus*, dung; Gr. masc./fem. n.
920 *hipposa* horse; N.L. gen. n. *limhippi*, of horse dung)

921 A bacterial species identified by metagenomic analyses. This species includes all bacteria
922 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
923 species to which we have assigned the MAG ID E2_98 and which is available via NCBI
924 BioSample SAMN18472475. The GC content of the type genome is 44.9 % and the genome
925 length is 1.15 Mbp.

926

927 **Description of *Candidatus Equicola* gen. nov.**

928 *Candidatus* Equicola (E.qui'co.la. L. masc. n. *equus*, a horse; N.L. masc./fem. suffix *-cola*,
929 an inhabitant; N.L. fem. n. *Equicola*, a microbe associated with horses)

930 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
931 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
932 from the type species *Candidatus Equicola stercoris*. This genus has been assigned by
933 GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order *Bacteroidales* and
934 to the family *Bacteroidaceae*.

935

936 **Description of *Candidatus Equicola faecalis* sp. nov.**

937 *Candidatus Equicola faecalis* (fae.ca'lis. N.L. fem. adj. *faecalis*, faecal)

938 A bacterial species identified by metagenomic analyses. This species includes all bacteria
939 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
940 species to which we have assigned the MAG ID E4_176 and which is available via NCBI
941 BioSample SAMN18472505. The GC content of the type genome is 44.8 % and the genome
942 length is 2.09 Mbp.

943

944 **Description of *Candidatus Equicola stercoris* sp. nov.**

945 *Candidatus Equicola stercoris* (ster'co.ris. L. gen. masc. n. *stercoris*, of dung)

946 A bacterial species identified by metagenomic analyses. This species includes all bacteria
947 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
948 species to which we have assigned the MAG ID E3_MB2_38 and which is available via
949 NCBI BioSample SAMN18472493. The GC content of the type genome is 42 % and the
950 genome length is 1.75 Mbp.

951

952 **Description of *Candidatus Equihabitans* gen. nov.**

953 *Candidatus Equihabitans* (E.qui.ha'bi.tans. L. masc. n. *equus*, a horse; L. masc./fem. pres.
954 part. *habitans*, an inhabitant; N.L. fem. n. *Equihabitans*, a microbe associated with horses)

955 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
956 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
957 from the type species *Candidatus Equihabitans merdae*. This genus has been assigned by

958 GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order *Lachnospirales*
959 and to the family *Lachnospiraceae*.

960

961 **Description of *Candidatus Equihabitans merdae* sp. nov.**

962 *Candidatus Equihabitans merdae* (mer'dae. L. gen. fem. n. *merdae*, of faeces)

963 A bacterial species identified by metagenomic analyses. This species includes all bacteria
964 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
965 species to which we have assigned the MAG ID E4_98 and which is available via NCBI
966 BioSample SAMN18472512. The GC content of the type genome is 47 % and the genome
967 length is 1.86 Mbp.

968

969 **Description of *Candidatus Equimonas* gen. nov.**

970 *Candidatus Equimonas* (E.qui.mo'nas. L. masc. n. *equus*, a horse; L. fem. n. *monas*, a
971 monad; N.L. fem. n. *Equimonas*, a microbe associated with horses)

972 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
973 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
974 from the type species *Candidatus Equimonas enterica*. This genus has been assigned by
975 GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order *Bacteroidales* and
976 to the family *Bacteroidaceae*.

977

978 **Description of *Candidatus Equimonas enterica* sp. nov.**

979 *Candidatus Equimonas enterica* (en.te'ri.ca. Gr. neut. n. *enteron*, gut, bowel, intestine; L..
980 fem. adj. suff. *-ica*, pertaining to; N.L. fem. adj. *enterica*, pertaining to intestine)

981 A bacterial species identified by metagenomic analyses. This species includes all bacteria
982 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
983 species to which we have assigned the MAG ID E1_145 and which is available via NCBI
984 BioSample SAMN18472453. The GC content of the type genome is 55.9 % and the genome
985 length is 1.85 Mbp.

986

987 **Description of *Candidatus Equimonas faecalis* sp. nov.**

988 *Candidatus* Equimonas faecalis (fae.ca'lis. N.L. fem. adj. *faecalis*, faecal)

989 A bacterial species identified by metagenomic analyses. This species includes all bacteria
990 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
991 species to which we have assigned the MAG ID E1_115 and which is available via NCBI
992 BioSample SAMN18472452. The GC content of the type genome is 55.6 % and the genome
993 length is 2.59 Mbp.

994

995 **Description of *Candidatus Equinaster* gen. nov.**

996 *Candidatus* Equinaster (E.qui.nas'ter. L. masc. n. *equus*, a horse; Gr. masc. n. naster, an
997 inhabitant; N.L. masc. n. *Equinaster*, a microbe associated with horses)

998 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
999 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1000 from the type species *Candidatus* Equinaster intestinalis. This genus has been assigned by
1001 GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order *Oscillospirales*
1002 and to the family *Acutalibacteraceae*.

1003

1004 **Description of *Candidatus Equinaster intestinalis* sp. nov.**

1005 *Candidatus* Equinaster intestinalis (in.tes.ti.na'lis. N.L. masc. adj. *intestinalis*, pertaining to
1006 the intestines)

1007 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1008 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1009 species to which we have assigned the MAG ID E3_MB2_43 and which is available via
1010 NCBI BioSample SAMN18472494. The GC content of the type genome is 43.4 % and the
1011 genome length is 1.50 Mbp.

1012

1013 **Description of *Candidatus Equispira* gen. nov.**

1014 *Candidatus* Equispira (E.qui.spi'ra. L. masc. n. *equus*, a horse; Gr. fem. n. *speira*, a coil,
1015 helix; N.L. fem. n. *Equispira*, a helical microbe associated with horses)
1016 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1017 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1018 from the type species *Candidatus* Equispira faecalis. This genus has been assigned by
1019 GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order *Lachnospirales*
1020 and to the family *Lachnospiraceae*.

1021

1022 **Description of *Candidatus* Equispira faecalis sp. nov.**

1023 *Candidatus* Equispira faecalis (fae.ca'lis. N.L. fem. adj. *faecalis*, faecal)
1024 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1025 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1026 species to which we have assigned the MAG ID E5_MB2_109 and which is available via
1027 NCBI BioSample SAMN18472542. The GC content of the type genome is 39.2 % and the
1028 genome length is 2.33 Mbp.

1029

1030 **Description of *Candidatus* Faecinaster gen. nov.**

1031 *Candidatus* Faecinaster (Fae.ci.nas'ter. L. fem. n. *faex*, *faecis*, dregs; Gr. masc. n. *naster*, an
1032 inhabitant; N.L. masc. n. *Faeцинaster*, a microbe associated with faeces)
1033 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1034 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1035 from the type species *Candidatus* Faecinaster equi. This is a new name for the GTDB
1036 alphanumeric genus UBA6382, which is found in diverse mammalian guts. This genus has
1037 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1038 *Bacteroidales* and to the family *Bacteroidaceae*.

1039

1040 **Description of *Candidatus* Faecinaster equi sp. nov.**

1041 *Candidatus* Faecinaster equi (e'qui. L. gen. masc. n. *equi*, of a horse)

1042 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1043 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1044 species to which we have assigned the MAG ID E3_MB2_9 and which is available via NCBI
1045 BioSample SAMN18472496. The GC content of the type genome is 37.2 % and the genome
1046 length is 3.36 Mbp.

1047

1048 **Description of *Candidatus Fiminaster gen. nov.***

1049 *Candidatus Fiminaster* (Fi.mi.nas'ter. L. neut. n. *fimum*, dung; Gr. masc. n. *naster*, an
1050 inhabitant; N.L. masc. n. *Fiminaster*, a microbe associated with faeces)

1051 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1052 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1053 from the type species *Candidatus Fiminaster equi*. This is a new name for the GTDB
1054 alphanumeric genus UBA3207, which is found in diverse mammalian guts. This genus has
1055 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1056 *RFN20* and to the family *CAG-826*.

1057

1058 **Description of *Candidatus Fiminaster equi sp. nov.***

1059 *Candidatus Fiminaster equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1060 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1061 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1062 species to which we have assigned the MAG ID E4_MB2_69 and which is available via
1063 NCBI BioSample SAMN18472528. The GC content of the type genome is 34.5 % and the
1064 genome length is 0.89 Mbp.

1065

1066 **Description of *Candidatus Flavobacterium equi sp. nov.***

1067 *Candidatus Flavobacterium equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1068 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1069 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the

1070 species to which we have assigned the MAG ID E2_MB2_6 and which is available via NCBI
1071 BioSample SAMN18472477. The GC content of the type genome is 37.7 % and the genome
1072 length is 2.17 Mbp.

1073

1074 **Description of *Candidatus Hippenecus* gen. nov.**

1075 *Candidatus Hippenecus* (Hipp.en.e'cus. Gr. masc./fem. n. *hippos*, a horse; N.L. masc. n.
1076 *enecus*, an inhabitant; N.L. masc. n. *Hippenecus* a microbe associated with horses)

1077 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1078 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1079 from the type species *Candidatus Hippenecus merdae*. This genus has been assigned by
1080 GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order *Lachnospirales*
1081 and to the family *Lachnospiraceae*.

1082

1083 **Description of *Candidatus Hippenecus merdae* sp. nov.**

1084 *Candidatus Hippenecus merdae* (mer'dae. L. gen. fem. n. *merdae*, of faeces)

1085 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1086 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1087 species to which we have assigned the MAG ID E3_87 and which is available via NCBI
1088 BioSample SAMN18472489. The GC content of the type genome is 52.7 % and the genome
1089 length is 1.11 Mbp.

1090

1091 **Description of *Candidatus Hippobium* gen. nov.**

1092 *Candidatus Hippobium* (Hip.po'bi.um. Gr. masc./fem. n. *hippos*, a horse; Gr. masc. n. *bios*,
1093 life; N.L. neut. n. *Hippobium*, a microbe associated with horses)

1094 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1095 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1096 from the type species *Candidatus Hippobium faecium*. This genus has been assigned by

1097 GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order *UBA5829* and to
1098 the family *UBA5829*.

1099

1100 **Description of *Candidatus Hippobium faecium* sp. nov.**

1101 *Candidatus* Hippobium faecium (fae'ci.um. L. fem. n. *faex*, *dregs*; L. gen. pl. n. *faecium*, of
1102 the *dregs*, of *faeces*)

1103 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1104 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1105 species to which we have assigned the MAG ID E3_206 and which is available via NCBI
1106 BioSample SAMN18472485. The GC content of the type genome is 39.1 % and the genome
1107 length is 2.12 Mbp.

1108

1109 **Description of *Candidatus Holdemanella equi* sp. nov.**

1110 *Candidatus* Holdemanella equi (e'qui. L. gen. masc. n. *equi*, of a horse)

1111 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1112 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1113 species to which we have assigned the MAG ID E4_26 and which is available via NCBI
1114 BioSample SAMN18472508. The GC content of the type genome is 35.6 % and the genome
1115 length is 2.44 Mbp.

1116

1117 **Description of *Candidatus Kurthia equi* sp. nov.**

1118 *Candidatus* Kurthia equi (e'qui. L. gen. masc. n. *equi*, of a horse)

1119 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1120 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1121 species to which we have assigned the MAG ID E1_MB2_88 and which is available via
1122 NCBI BioSample SAMN18472468. The GC content of the type genome is 35.7 % and the
1123 genome length is 3.58 Mbp.

1124

1125 **Description of *Candidatus Limimonas gen. nov.***

1126 *Candidatus Limimonas* (Li.mi.mo'nas. L. masc. n. *limus*, dung; L. fem. n. *monas*, a monad;

1127 N.L. fem. n. *Limimonas*, a microbe associated with faeces)

1128 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1129 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1130 from the type species *Candidatus Limimonas coprohippi*. This is a new name for the GTDB
1131 alphanumeric genus UBA1227, which is found in diverse mammalian guts. This genus has
1132 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1133 *Oscillospirales* and to the family *Acutalibacteraceae*.

1134

1135 **Description of *Candidatus Limimonas coprohippi sp. nov.***

1136 *Candidatus Limimonas coprohippi* (co.pro.hip'pi. Gr. fem. n. *kopros*, dung; Gr. masc./fem.

1137 n. *hippos*, a horse; N.L. gen. n. *coprohippi*, associated with the faeces of horses)

1138 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1139 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1140 species to which we have assigned the MAG ID E1_MB2_82 and which is available via
1141 NCBI BioSample SAMN18472467. The GC content of the type genome is 40.5 % and the
1142 genome length is 1.33 Mbp.

1143

1144 **Description of *Candidatus Limimonas egerieequi sp. nov.***

1145 *Candidatus Limimonas egerieequi* (e.ge.ri.e.e'qui. L. fem. n. *egeries*, dung; L. masc.

1146 n. *equus*, a horse; N.L. gen. n. *egerieequi*, associated with the faeces of horses)

1147 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1148 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1149 species to which we have assigned the MAG ID E5_MB2_129 and which is available via
1150 NCBI BioSample SAMN18472544. The GC content of the type genome is 41.8 % and the
1151 genome length is 1.70 Mbp.

1152

1153 **Description of *Candidatus Limimorpha caballi* sp. nov.**

1154 *Candidatus* Limimorpha caballi (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

1155 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1156 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1157 species to which we have assigned the MAG ID E5_119 and which is available via NCBI
1158 BioSample SAMN18472534. The GC content of the type genome is 48.3 % and the genome
1159 length is 2.76 Mbp.

1160

1161 **Description of *Candidatus Limimorpha equi* sp. nov.**

1162 *Candidatus* Limimorpha equi (e'qui. L. gen. masc. n. *equi*, of a horse)

1163 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1164 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1165 species to which we have assigned the MAG ID E1_MB2_99 and which is available via
1166 NCBI BioSample SAMN18472470. The GC content of the type genome is 45.1 % and the
1167 genome length is 2.72 Mbp.

1168

1169 **Description of *Candidatus Liminaster* gen. nov.**

1170 *Candidatus* Liminaster (Li.mi.nas'ter. L. masc. n. *limus*, dung; Gr. masc. n. *naster*, an
1171 inhabitant; N.L. masc. n. *Liminaster*, a microbe associated with faeces)

1172 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1173 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1174 from the type species *Candidatus* Liminaster caballi. This is a new name for the GTDB
1175 alphanumeric genus UBA3663, which is found in diverse mammalian guts. This genus has
1176 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1177 *Bacteroidales* and to the family UBA3663.

1178

1179 **Description of *Candidatus Liminaster caballi* sp. nov.**

1180 *Candidatus* Liminaster caballi (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

1181 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1182 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1183 species to which we have assigned the MAG ID E4_95 and which is available via NCBI
1184 BioSample SAMN18472511. The GC content of the type genome is 50.1 % and the genome
1185 length is 2.94 Mbp.

1186

1187 **Description of *Candidatus Merdinaster* gen. nov.**

1188 *Candidatus Merdinaster* (Mer.di.nas'ter. L. fem. n. *merda*, dung; Gr. masc. n. *naster*, an
1189 inhabitant; N.L. masc. n. *Merdinaster*, a microbe associated with faeces)

1190 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1191 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1192 from the type species *Candidatus Merdinaster equi*. This is a new name for the GTDB
1193 alphanumeric genus UBA7050, which is found in diverse mammalian guts. This genus has
1194 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1195 *Lachnospirales* and to the family *Lachnospiraceae*,

1196

1197 **Description of *Candidatus Merdinaster equi* sp. nov.**

1198 *Candidatus Merdinaster equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1199 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1200 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1201 species to which we have assigned the MAG ID E4_MB2_128 and which is available via
1202 NCBI BioSample SAMN18472516. The GC content of the type genome is 40.7 % and the
1203 genome length is 1.95 Mbp.

1204

1205 **Description of *Candidatus Methanocorpusculum equi* sp. nov.**

1206 *Candidatus Methanocorpusculum equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1207 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1208 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the

1209 species to which we have assigned the MAG ID E2_MB2_79 and which is available via
1210 NCBI BioSample SAMN18472479. The GC content of the type genome is 50.2 % and the
1211 genome length is 1.15 Mbp.

1212

1213 **Description of *Candidatus Minthenecus* gen. nov.**

1214 *Candidatus* Minthenecus (Minth.en.e'cus. Gr. masc. n. *minthos*, dung; N.L. masc. n. *enecus*,
1215 an inhabitant; N.L. masc. n. *Minthenecus*, a microbe associated with faeces)

1216 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1217 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1218 from the type species *Candidatus* Minthenecus merdequi. This is a new name for the GTDB
1219 alphanumeric genus SFVR01, which is found in diverse mammalian guts. This genus has
1220 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1221 *Bacteroidales* and to the family *Paludibacteraceae*.

1222

1223 **Description of *Candidatus Minthenecus merdequi* sp. nov.**

1224 *Candidatus* Minthenecus merdequi (merd.e'qui. L. fem. n. *merda*, faeces; L. masc. n. *equus*,
1225 a horse; N.L. gen. n. *merdequi*, associated with the faeces of horses)

1226 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1227 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1228 species to which we have assigned the MAG ID E5_MB2_18 and which is available via
1229 NCBI BioSample SAMN18472545. The GC content of the type genome is 42.5 % and the
1230 genome length is 1.80 Mbp.

1231

1232 **Description of *Candidatus Minthocola* gen. nov.**

1233 *Candidatus* Minthocola (Min.tho'co.la. Gr. masc. n. *minthos*, dung; N.L. masc./fem. suffix -
1234 *cola*, an inhabitant; N.L. fem. n. *Minthocola*, a microbe associated with faeces)

1235 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1236 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain

1237 from the type species *Candidatus Minthocola equi*. This is a new name for the GTDB
1238 alphanumeric genus UBA3774, which is found in diverse mammalian guts. This genus has
1239 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1240 *Lachnospirales* and to the family *Lachnospiraceae*.

1241

1242 **Description of *Candidatus Minthocola equi* sp. nov.**

1243 *Candidatus Minthocola equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1244 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1245 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1246 species to which we have assigned the MAG ID E5_MB2_38 and which is available via
1247 NCBI BioSample SAMN18472548. The GC content of the type genome is 45.2 % and the
1248 genome length is 1.20 Mbp.

1249

1250 **Description of *Candidatus Minthomonas* gen. nov.**

1251 *Candidatus Minthomonas* (Min.tho.mo'nas. Gr. masc. n. *minthos*, dung; L. fem. n. *monas*, a
1252 monad; N.L. fem. n. *Minthomonas*, a microbe associated with faeces)

1253 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1254 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1255 from the type species *Candidatus Minthomonas equi*. This is a new name for the GTDB
1256 alphanumeric genus CAG-831, which is found in diverse mammalian guts. This genus has
1257 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1258 *Bacteroidales* and to the family *UBA932*.

1259

1260 **Description of *Candidatus Minthomonas equi* sp. nov.**

1261 *Candidatus Minthomonas equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1262 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1263 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1264 species to which we have assigned the MAG ID E5_18 and which is available via NCBI

1265 BioSample SAMN18472536. The GC content of the type genome is 47.6 % and the genome
1266 length is 1.36 Mbp.

1267

1268 **Description of *Candidatus Minthosoma* gen. nov.**

1269 *Candidatus Minthosoma* (Min.tho.so'ma. Gr. masc. n. *minthos*, dung; Gr. neut. n. *soma*, a
1270 body; N.L. neut. n. *Minthosoma*, a microbe associated with faeces)

1271 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1272 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1273 from the type species *Candidatus Minthosoma caballi*. *This* is a new name for the GTDB
1274 alphanumeric genus UBA4334, which is found in diverse mammalian guts. This genus has
1275 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1276 *Bacteroidales* and to the family *Bacteroidaceae*

1277

1278 **Description of *Candidatus Minthosoma caballi* sp. nov.**

1279 *Candidatus Minthosoma caballi* (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

1280 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1281 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1282 species to which we have assigned the MAG ID E5_9 and which is available via NCBI

1283 BioSample SAMN18472539. The GC content of the type genome is 44.2 % and the genome
1284 length is 3.21 Mbp.

1285

1286 **Description of *Candidatus Minthosoma equi* sp. nov.**

1287 *Candidatus Minthosoma equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1288 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1289 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1290 species to which we have assigned the MAG ID E4_MB2_18 and which is available via

1291 NCBI BioSample SAMN18472519. The GC content of the type genome is 44.1 % and the
1292 genome length is 3.51 Mbp.

1293

1294 **Description of *Candidatus Minthousia* gen. nov.**

1295 *Candidatus* Minthousia (Minth.ou's.ia. Gr. masc. n. *minthos*, dung; Gr. fem. n. *ousia*, an
1296 essence; N.L. fem. n. *Minthousia*, a microbe associated with faeces)

1297 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1298 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1299 from the type species *Candidatus* Minthousia equi. This is a new name for the GTDB
1300 alphanumeric genus UBA4293, which is found in diverse mammalian guts. This genus has
1301 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1302 *Bacteroidales* and to the family *Bacteroidaceae*.

1303

1304 **Description of *Candidatus Minthousia equi* sp. nov.**

1305 *Candidatus* Minthousia equi (e'qui. L. gen. masc. n. *equi*, of a horse)

1306 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1307 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1308 species to which we have assigned the MAG ID E4_55 and which is available via NCBI
1309 BioSample SAMN18472509. The GC content of the type genome is 42.9 % and the genome
1310 length is 2.61 Mbp.

1311

1312 **Description of *Candidatus Mogibacterium equifaecale* sp. nov.**

1313 *Candidatus* Mogibacterium equifaecale (e.qui.fae.ca'le. L. masc. n. *equus*, a horse; N.L.
1314 neut. adj. *faecale*, faecal N.L. neut. adj. *equifaecale*, associated with the faeces of horses)

1315 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1316 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1317 species to which we have assigned the MAG ID E4_MB2_51 and which is available via
1318 NCBI BioSample SAMN18472526. The GC content of the type genome is 43.4 % and the
1319 genome length is 1.39 Mbp.

1320

1321 **Description of *Candidatus Mogibacterium onthequi* sp. nov.**

1322 *Candidatus* Mogibacterium onthequi (onth.e'qui. Gr. masc. n. *onthos*, dung; L. masc.

1323 n. *equus*, a horse; N.L. gen. n. *onthequi*, associated with the faeces of horses)

1324 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1325 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1326 species to which we have assigned the MAG ID E4_MB2_84 and which is available via
1327 NCBI BioSample SAMN18472530. The GC content of the type genome is 43.8 % and the
1328 genome length is 1.61 Mbp.

1329

1330 **Description of *Candidatus Mogibacterium scatequi* sp. nov.**

1331 *Candidatus* Mogibacterium scatequi (scat.e'qui. Gr. neut. n. *skor*, *skatos*, dung; L. masc.

1332 n. *equus*, a horse; N.L. gen. n. *scatequi*, associated with the faeces of horses)

1333 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1334 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1335 species to which we have assigned the MAG ID E4_MB2_90 and which is available via
1336 NCBI BioSample SAMN18472532. The GC content of the type genome is 45.2 % and the
1337 genome length is 1.45 Mbp.

1338

1339 **Description of *Candidatus Mycoplasma equi* sp. nov.**

1340 *Candidatus* Mycoplasma equi (e'qui. L. gen. masc. n. *equi*, of a horse)

1341 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1342 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1343 species to which we have assigned the MAG ID E4_MB2_29 and which is available via
1344 NCBI BioSample SAMN18472521. The GC content of the type genome is 31.4 % and the
1345 genome length is 0.64 Mbp.

1346

1347 **Description of *Candidatus Onthonaster* gen. nov.**

1348 *Candidatus* Onthonaster (On.tho.nas'ter. Gr. masc. n. *onthos*, dung; Gr. masc. n. *naster*, an
1349 inhabitant; N.L. masc. n. *Onthonaster*, a microbe associated with faeces)

1350 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1351 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1352 from the type species *Candidatus* Onthonaster equi. *This* is a new name for the GTDB
1353 alphanumeric genus YIM-102668, which is found in diverse mammalian guts. This genus
1354 has been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the
1355 order *Flavobacteriales* and to the family *Weeksellaceae*

1356

1357 **Description of *Candidatus* Onthonaster equi sp. nov.**

1358 *Candidatus* Onthonaster equi (e'qui. L. gen. masc. n. *equi*, of a horse)

1359 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1360 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1361 species to which we have assigned the MAG ID E1_98 and which is available via NCBI
1362 BioSample SAMN18472458. *This* is a new name for the alphanumeric GTDB species
1363 sp003687725, which is found in diverse mammalian guts. The GC content of the type
1364 genome is 31.1 % and the genome length is 2.30 Mbp.

1365

1366 **Description of *Candidatus* Phascolarctobacterium caballi sp. nov.**

1367 *Candidatus* Phascolarctobacterium caballi (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

1368 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1369 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1370 species to which we have assigned the MAG ID E4_135 and which is available via NCBI
1371 BioSample SAMN18472503. The GC content of the type genome is 39.4 % and the genome
1372 length is 1.56 Mbp.

1373

1374 **Description of *Candidatus* Phascolarctobacterium equi sp. nov.**

1375 *Candidatus* Phascolarctobacterium equi (e'qui. L. gen. masc. n. *equi*, of a horse)

1376 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1377 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1378 species to which we have assigned the MAG ID E2_44 and which is available via NCBI
1379 BioSample SAMN18472473. The GC content of the type genome is 46.7 % and the genome
1380 length is 0.93 Mbp.

1381

1382 **Description of *Candidatus Physcocola* gen. nov.**

1383 *Candidatus Physcocola* (Phys.co'co.la. Gr. fem. n. *physke*, the colon; N.L. masc./fem. suffix
1384 *-cola*, an inhabitant; N.L. fem. n. *Physcocola*, a microbe associated with the large intestine)

1385 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1386 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1387 from the type species *Candidatus Physcocola equi*. This is a new name for the GTDB
1388 alphanumeric genus UBA4345, which is found in diverse mammalian guts. This genus has
1389 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1390 *Bacteroidales* and to the family *Paludibacteraceae*.

1391

1392 **Description of *Candidatus Physcocola equi* sp. nov.**

1393 *Candidatus Physcocola equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1394 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1395 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1396 species to which we have assigned the MAG ID E4_MB2_42 and which is available via
1397 NCBI BioSample SAMN18472523. The GC content of the type genome is 43.3 % and the
1398 genome length is 2.99 Mbp.

1399

1400 **Description of *Candidatus Physcosoma* gen. nov.**

1401 *Candidatus Physcosoma* (Phys.co.so'ma. Gr. fem. n. *physke*, the colon; Gr. neut. n. *soma*, a
1402 body; N.L. neut. n. *Physcosoma*, a microbe associated with the large intestine)

1403 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1404 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1405 from the type species *Candidatus Physcosoma equi*. This is a new name for the GTDB
1406 alphanumeric genus UBA5920, which is found in diverse mammalian guts. This genus has
1407 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1408 *Sphaerochaetales* and to the family *Sphaerochaetaceae*.

1409

1410 **Description of *Candidatus Physcosoma equi* sp. nov.**

1411 *Candidatus Physcosoma equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1412 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1413 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1414 species to which we have assigned the MAG ID E4_160 and which is available via NCBI
1415 BioSample SAMN18472504. The GC content of the type genome is 49.1 % and the genome
1416 length is 2.06 Mbp.

1417

1418 **Description of *Candidatus Physcousia* gen. nov.**

1419 *Candidatus Physcousia* (Physc.ou'si.a. Gr. fem. n. *physke* the colon; Gr. fem. n. *ousia*, an
1420 essence.e; N.L. fem. n. *Physcousia*, a microbe associated with the large intestine)

1421 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1422 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1423 from the type species *Candidatus Physcousia caballi*. This is a new name for the GTDB
1424 alphanumeric genus UBA4372, which is found in diverse mammalian guts. This genus has
1425 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1426 *Bacteroidales* and to the family *Bacteroidaceae*.

1427

1428 **Description of *Candidatus Physcousia caballi* sp. nov.**

1429 *Candidatus Physcousia caballi* (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

1430 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1431 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1432 species to which we have assigned the MAG ID E4_MB2_73 and which is available via
1433 NCBI BioSample SAMN18472529. The GC content of the type genome is 50.5 % and the
1434 genome length is 3.81 Mbp.

1435

1436 **Description of *Candidatus Physcousia equi* sp. nov.**

1437 *Candidatus Physcousia equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1438 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1439 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1440 species to which we have assigned the MAG ID E4_MB2_112 and which is available via
1441 NCBI BioSample SAMN18472514. The GC content of the type genome is 52.4 % and the
1442 genome length is 2.43 Mbp.

1443

1444 **Description of *Candidatus Prevotella equi* sp. nov.**

1445 *Candidatus Prevotella equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1446 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1447 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1448 species to which we have assigned the MAG ID E4_23 and which is available via NCBI
1449 BioSample SAMN18472507. The GC content of the type genome is 44.5 % and the genome
1450 length is 3.45 Mbp.

1451

1452 **Description of *Candidatus Ruminococcus equi* sp. nov.**

1453 *Candidatus Ruminococcus equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1454 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1455 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1456 species to which we have assigned the MAG ID E3_41 and which is available via NCBI
1457 BioSample SAMN18472487. GTDB has assigned. This species to a genus marked with an

1458 alphabetical suffix. However, as this genus designation cannot be incorporated into a well-
1459 formed binomial, in naming. This species, we have used the current validly published name
1460 for the genus. The GC content of the type genome is 39.9 % and the genome length is 1.73
1461 Mbp.

1462

1463 **Description of *Candidatus Scatohabitans* gen. nov.**

1464 *Candidatus Scatohabitans* (Sca.to.ha'bi.tans. Gr. neut. n. *skor*, *skatos*, dung; L. masc./fem.
1465 pres. part. *habitans*, an inhabitant; N.L. fem. n. *Scatohabitans* a microbe associated with
1466 faeces)

1467 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1468 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1469 from the type species *Candidatus Scatohabitans aphodohippi*. This is a new name for the
1470 GTDB alphanumeric genus C941, which is found in diverse mammalian guts. This genus
1471 has been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the
1472 order *Bacteroidales* and to the family *Muribaculaceae*.

1473

1474 **Description of *Candidatus Scatohabitans aphodohippi* sp. nov.**

1475 *Candidatus Scatohabitans aphodohippi* (aph.o.do.hip'pi. Gr. fem. n. *aphodos*, dung; Gr.
1476 masc./fem. n. *hippos*, a horse; N.L. gen. n. *aphodohippi*, associated with the faeces of
1477 horses)

1478 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1479 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1480 species to which we have assigned the MAG ID E3_0 and which is available via NCBI
1481 BioSample SAMN18472480. The GC content of the type genome is 50 % and the genome
1482 length is 2.49 Mbp.

1483

1484 **Description of *Candidatus Scatohabitans fimicaballi* sp. nov.**

1485 *Candidatus Scatohabitans fimicaballi* (fi.mi.ca.bal'li. L. masc. n. *fimius*, dung; L. masc.
1486 n. *caballus*, a horse; N.L. gen. n. *fimicaballi*, associated with the faeces of horses)
1487 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1488 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1489 species to which we have assigned the MAG ID E4_193 and which is available via NCBI
1490 BioSample SAMN18472506. The GC content of the type genome is 48.1 % and the genome
1491 length is 2.46 Mbp.

1492

1493 **Description of *Candidatus Scatohabitans limicaballi* sp. nov.**

1494 *Candidatus Scatohabitans limicaballi* (li.mi.ca.bal'li. L. masc. n. *limus*, dung; L. masc.
1495 n. *caballus*, a horse; N.L. gen. n. *limicaballi*, associated with the faeces of horses)
1496 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1497 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1498 species to which we have assigned the MAG ID E2_8 and which is available via NCBI
1499 BioSample SAMN18472474. The GC content of the type genome is 50.4 % and the genome
1500 length is 3.16 Mbp.

1501

1502 **Description of *Candidatus Scatonaster* gen. nov.**

1503 *Candidatus Scatonaster* (Sca.to.nas'ter. Gr. neut. n. *skor*, *skatos*, dung; Gr. masc. n. *naster*,
1504 an inhabitant; N.L. masc. n. *Scatonaster* a microbe associated with faeces)
1505 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1506 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1507 from the type species *Candidatus Scatonaster coprocaballi*. This is a new name for the
1508 GTDB alphanumeric genus Firm-16, which is found in diverse mammalian guts. This genus
1509 has been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the
1510 order *Saccharofermentanales* and to the family *Saccharofermentanaceae*.

1511

1512 **Description of *Candidatus Scatonaster coprocaballi* sp. nov.**

1513 *Candidatus* Scatonaster coprocaballi (co.pro.ca.bal'li. Gr. fem. n. *kopros*, dung; L. masc.
1514 n. *caballus*, a horse; N.L. gen. n. *coprocaballi*, associated with the faeces of horses)
1515 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1516 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1517 species to which we have assigned the MAG ID E5_MB2_10 and which is available via
1518 NCBI BioSample SAMN18472540. The GC content of the type genome is 46.9 % and the
1519 genome length is 2.23 Mbp.

1520

1521 **Description of *Candidatus* Scybalocola gen. nov.**

1522 *Candidatus* Scybalocola (Scy.ba.lo'co.la. Gr. neut. n. *skybalon*, dung; N.L. masc./fem. suffix
1523 *-cola*, an inhabitant; N.L. fem. n. *Scybalocola* a microbe associated with faeces)
1524 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1525 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1526 from the type species *Candidatus* Scybalocola fimicaballi. This is a new name for the GTDB
1527 alphanumeric genus UBA1723, which is found in diverse mammalian guts. This genus has
1528 been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the order
1529 *Bacteroidales* and to the family *Paludibacteraceae*.

1530

1531 **Description of *Candidatus* Scybalocola fimicaballi sp. nov.**

1532 *Candidatus* Scybalocola fimicaballi (fi.mi.ca.bal'li. L. masc. n. *fimius*, dung; L. masc.
1533 n. *caballus*, a horse; N.L. gen. n. *fimicaballi*, associated with the faeces of horses)
1534 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1535 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1536 species to which we have assigned the MAG ID E1_25 and which is available via NCBI
1537 BioSample SAMN18472456. This is a new name for the alphanumeric GTDB species
1538 sp002317115, which is found in diverse mammalian guts. The GC content of the type
1539 genome is 41.7 % and the genome length is 3.11 Mbp.

1540

1541 **Description of *Candidatus Scybalousia* gen. nov.**

1542 *Candidatus Scybalousia* (Scy.bal.ou's.ia. Gr. neut. n. *skybalon*, dung; Gr. fem. n. *ousia*, an
1543 essence; N.L. fem n. *Scybalousia*, a microbe associated with faeces)

1544 A bacterial genus identified by metagenomic analyses. The genus includes all bacteria with
1545 genomes that show $\geq 60\%$ average amino acid identity (AAI) to the genome of the type strain
1546 from the type species *Candidatus Scybalousia scybalohippi*. This is a new name for the
1547 GTDB alphanumeric genus Phil12, which is found in diverse mammalian guts. This genus
1548 has been assigned by GTDB-Tk v1.3.0 working on GTDB Release 05-RS95 (39, 79) to the
1549 order *Bacteroidales* and to the family *P3*.

1550

1551 **Description of *Candidatus Scybalousia scybalohippi* sp. nov.**

1552 *Candidatus Scybalousia scybalohippi* (scy.ba.lo.hip'pi. Gr. neut. n. *skybalon*, dung; Gr.
1553 masc./fem. n. *hippos*, a horse; N.L. gen. n. *scybalohippi*, associated with the faeces of
1554 horses)

1555 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1556 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1557 species to which we have assigned the MAG ID E3_144 and which is available via NCBI
1558 BioSample SAMN18472482. The GC content of the type genome is 35.4 % and the genome
1559 length is 2.63 Mbp.

1560

1561 **Description of *Candidatus Stomatobaculum equi* sp. nov.**

1562 *Candidatus Stomatobaculum equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1563 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1564 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1565 species to which we have assigned the MAG ID E1_MB2_36 and which is available via
1566 NCBI BioSample SAMN18472462. GTDB has assigned. This species to a genus marked
1567 with an alphabetical suffix. However, as this genus designation cannot be incorporated into a
1568 well-formed binomial, in naming. This species, we have used the current validly published

1569 name for the genus. The GC content of the type genome is 44.2 % and the genome length is
1570 1.44 Mbp.

1571

1572 **Description of *Candidatus Treponema caballi* sp. nov.**

1573 *Candidatus Treponema caballi* (ca.bal'li. L. gen. masc. n. *caballi*, of a horse)

1574 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1575 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1576 species to which we have assigned the MAG ID E1_106 and which is available via NCBI
1577 BioSample SAMN18472451. GTDB has assigned. This species to a genus marked with an
1578 alphabetical suffix. However, as this genus designation cannot be incorporated into a well-
1579 formed binomial, in naming. This species, we have used the current validly published name
1580 for the genus. The GC content of the type genome is 47.1 % and the genome length is 2.91
1581 Mbp.

1582

1583 **Description of *Candidatus Treponema equi* sp. nov.**

1584 *Candidatus Treponema equi* (e'qui. L. gen. masc. n. *equi*, of a horse)

1585 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1586 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1587 species to which we have assigned the MAG ID E4_MB2_46 and which is available via
1588 NCBI BioSample SAMN18472525. GTDB has assigned. This species to a genus marked
1589 with an alphabetical suffix. However, as this genus designation cannot be incorporated into a
1590 well-formed binomial, in naming. This species, we have used the current validly published
1591 name for the genus. The GC content of the type genome is 44.3 % and the genome length is
1592 1.79 Mbp.

1593

1594 **Description of *Candidatus Treponema equifaecale* sp. nov.**

1595 *Candidatus Treponema equifaecale* (e.qui.fae.ca'le. L. masc. n. *equus*, a horse; N.L. neut.
1596 adj. *faecale*, faecal; N.L. neut. adj. *equifaecale*, associated with the faeces of horses)

1597 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1598 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1599 species to which we have assigned the MAG ID E4_MB2_2 and which is available via NCBI
1600 BioSample SAMN18472520. GTDB has assigned. This species to a genus marked with an
1601 alphabetical suffix. However, as this genus designation cannot be incorporated into a well-
1602 formed binomial, in naming. This species, we have used the current validly published name
1603 for the genus. The GC content of the type genome is 40.2 % and the genome length is 2.81
1604 Mbp.

1605

1606 **Description of *Candidatus Treponema merdequi* sp. nov.**

1607 *Candidatus Treponema merdequi* (merd.e'qui. L. fem. n. *merda*, faeces; L. masc. n. *equus*,
1608 a horse; N.L. gen. n. *merdequi*, associated with the faeces of horses)

1609 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1610 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1611 species to which we have assigned the MAG ID E5_50 and which is available via NCBI
1612 BioSample SAMN18472538. GTDB has assigned. This species to a genus marked with an
1613 alphabetical suffix. However, as this genus designation cannot be incorporated into a well-
1614 formed binomial, in naming. This species, we have used the current validly published name
1615 for the genus. The GC content of the type genome is 35.8 % and the genome length is 2.70
1616 Mbp.

1617

1618 **Description of *Candidatus Treponema scatequi* sp. nov.**

1619 *Candidatus Treponema scatequi* (scat.e'qui. Gr. neut. n. *skor*, *skatos*, dung; L. masc.
1620 n. *equus*, a horse; N.L. gen. n. *scatequi*, associated with the faeces of horses)

1621 A bacterial species identified by metagenomic analyses. This species includes all bacteria
1622 with genomes that show $\geq 95\%$ average nucleotide identity (ANI) to the type genome for the
1623 species to which we have assigned the MAG ID E1_MB2_111 and which is available via
1624 NCBI BioSample SAMN18472460. GTDB has assigned. This species to a genus marked

1625 with an alphabetical suffix. However, as this genus designation cannot be incorporated into a
1626 well-formed binomial, in naming this species, we have used the current validly published
1627 name for the genus. The GC content of the type genome is 38.4 % and the genome length is
1628 2.31 Mbp.

1629

1630

1631 **A newly named class within the *Armatimonadetes***

1632 One of our MAGs—and the associated metagenomic species, which we have called *Ca.*
1633 *Hippobium faecium*—was assigned to an unnamed family within the recently named phylum
1634 *Armatimonadetes* (also called *Armatimonadota*; previously known as OP10) (80). Scrutiny of
1635 the NCBI database in March 2021 reveals that no genome assemblies and only twenty-five
1636 16S gene sequences linked to this phylum originate from the vertebrate gut, which include
1637 sequences from cattle (81), pigs (82–84), catfish (85) and migratory passerines.

1638 Phylogenetic analysis placed the new species on a deep branch within the
1639 *Armatimonadetes*, clustering with a solitary bioreactor-derived metagenome-assembled
1640 genome UBA5829 (GCA_002431715.1) assigned by GTDB to its own class, order, family
1641 and genus, all given the same alphanumeric designation UBA5829. Sequence comparisons
1642 between the genomes of *Ca. H. faecium* UBA5829 report an AAI of 51% (S7 Table), which
1643 suggests that they sit within the same class, which we propose should be named *Ca.*

1644 *Hippobiia*.

1645

1646 **Distribution and metabolism**

1647 Our de-replicated high- and medium-quality MAGs account for 18% ($\pm 5\%$) of our host-
1648 depleted metagenomic reads. Distribution analysis identified 17 species present at $\geq 1x$
1649 coverage in all samples, spanning four bacterial phyla and the archaea (Fig 2a and S8
1650 Table). No species were present at $\geq 10x$ coverage in all samples. Species quantification
1651 shows a steady incline in the cumulative number of species identified with each consecutive
1652 sample (Fig 2b).

1653

1654 **Fig 2. Distribution and metabolism of equine microbial genomes. (A).** Heat map

1655 depicting the abundance of 110 MAGs across 5 metagenomic samples. Metagenomic

1656 species have been annotated with their taxonomic class and species assignment. All data

1657 were Log₁₀ transformed with Blue colour depicting species of low abundance and Red

1658 showing high abundance. **(B)** Species accumulation curve based on coverage data for 110

1659 MAGs over 5 metagenomic samples. **(C)** Proportion of functional CAZyme classes

1660 encoded by metagenomic assemblies. Species have been ordered according to GTDB-tk

1661 assigned phylum, with functional classes depicted by bar colour.

1662

1663 We created a catalogue of 228,125 genes from our medium- and high-quality MAGs.

1664 All 123 MAGs encoded known carbohydrate-active enzymes (CAZymes), with an average of

1665 69 CAZymes per genome (S9 Table). Most (>70%) metagenomic species with a higher-than

1666 average repertoire of CAZymes belonged to the *Bacteroidota*. Of the ~8,500 CAZyme genes

1667 reported, most were associated with classes devoted to assembly (glycosyltransferases [GT]

1668 29%) and breakdown (glycoside hydrolases [GH] 51%) of carbohydrate complexes, with far

1669 fewer from other groups of CAZymes; being the polysaccharide lyases (PL) and

1670 carbohydrate esterases (CE) alongside two further non-enzymatic groups being the

1671 carbohydrate-binding modules (CBM) and the auxiliary activities (AA). (Fig 2c). Recovery of

1672 93 classes of glycoside hydrolases from the equine gut mirrors similar enzymes in the sheep

1673 rumen linked to fibre degradation (86). Over half of our equine MAGs encode CAZymes with

1674 presumed involvement in degradation of hemi-cellulose (58%), cellulose (51%) or pectin or

1675 soluble fibre (>60%).

1676

1677 **Many novel bacteriophage genomes**

1678 The program VirSorter classified 2,500 contigs as “highly likely” or “likely” to originate from

1679 bacteriophages (S10 Table). Of these, 190 bacteriophage genomes were identified as “high-

1680 quality” (n=181) or “complete” (n=9) after de-replication (Fig 3a). However, as none showed

1681 close identity to known viral sequences, they all represent novel bacteriophage species.
1682 Genome sizes ranged from 5 kb to 145 kb, including 42 genomes ranging from 5 to 15 kb in
1683 length. Using the viral taxonomy tool Demovir, we could assign 150 of these new phages to
1684 known viral families. An additional 29 could be assigned to taxonomically informative viral
1685 clusters, based on similarities between predicted proteins from our contigs and proteins from
1686 the viral component of the Refseq94 database (S11 Table). Just under half (n=14) of these
1687 viral clusters contained at least one reference genome, thus expanding the known diversity
1688 of four viral families (Fig 3b).

1689

1690 **Fig 3. Bacteriophage analysis of equine faecal samples from five thoroughbred**
1691 **horses. (A)** CheckV quality tiers versus contig length (provided as Log₁₀ values). **(B)** Protein
1692 sharing network of 190 High-quality or Complete phage genomes assembled from five
1693 equine faecal metagenomes and compared against a de-replicated RefSeq database of
1694 reference prokaryotic virus genomes. Each node represents a viral genome, with node
1695 colour depicting source sample and node size scales according to metagenome contig
1696 length. Grey nodes depict reference genomes, with no size scaling shown. Network edges
1697 indicate statistically significant relationships between the protein profiles of respective viral
1698 genomes. Annotation has been provided to highlight viral clusters of interest. **(C)** Upset plot
1699 of phage genomes shared between or specific to source faecal sample, set colour is defined
1700 by sample. Each bar represents the number of phage genomes described within the given
1701 samples.

1702

1703 Almost all of our viral genomes represented tailed dsDNA phages from the order
1704 *Caudovirales* (87) and could be sub-classified into the families *Siphoviridae* (73%),
1705 *Podoviridae* or the newly delineated *Schitoviridae* (88) (21%) and *Myoviridae* (6%). Seven
1706 genomes were assigned to ssDNA viruses from the family *Microviridae*, four of which cluster
1707 as part of the subfamily *Gokushovirinae*. Weak connections of three viral genomes to a viral
1708 cluster of *Obolenskivirus*, whose known members all infect *Acinetobacter* sp., likely indicates

1709 the presence of novel bacteriophage genomes predating on the prominent population of
1710 bacterial *Acinetobacter* within the equine hind-gut. Present within the viral cluster network
1711 but notably absent within our bacteriophage catalogue included the model *Escherichia coli*
1712 phages T4 (*Tevenvirinae*) and T7 (*Studiervirinae*) or the *Mycobacterium* infecting
1713 actinophages. We observed several novel viral clusters comprising only genomes
1714 assembled in this study, which could be classified as the first representatives of new horse
1715 hindgut-associated phage families. Based on the proteome comparisons (Fig 3b), we predict
1716 at least three new families.

1717 Over three quarters of the recovered phage genomes were found at >1x coverage in
1718 just a single sample (Fig 3c). Only one phage was found in all five samples, with coverage
1719 ranging from 1.9x - 29x and forming a viral cluster with Lactococcus phage P087 of the
1720 family *Siphoviridae*.

1721

1722 **Discussion**

1723 Compared to the human gut, the microbiology of the horse gut remains largely unexplored.
1724 Here, we deliver new insights into this important ecosystem while also showcasing the
1725 advantages of shotgun metagenomics in providing catalogues of genes and genome
1726 sequences that take us well beyond what can be achieved using 16S ribosomal RNA gene
1727 sequences. Exploration of just five faecal samples allowed discovery of—and recovery of
1728 genomes from—over 100 new bacterial and archaeal species and nearly 200
1729 bacteriophage genomes, substantially increasing the known microbial diversity of this
1730 environment. Deposition of genomes from these species into publicly available databases
1731 will underpin all future studies, improving the quality of reference-based taxonomic
1732 assignments.

1733 While the limited scope of this study means it cannot hope to provide a
1734 comprehensive view of taxonomic diversity within the horse gut, it gives us a tantalizing
1735 glimpse of the richness that awaits us when such approaches are rolled out more widely,

1736 particularly as integration of long-read sequencing into metagenomics brings the promise of
1737 genome assemblies rivaling those from cultured isolates (89–91). Just as the horse allowed
1738 humans to explore new external landscapes, new sequencing and bioinformatics
1739 approaches will allow us to explore the inner world of the equine gut microbiome.

1740

1741 **Funding**

1742 The researchers gratefully acknowledge the support of the Biotechnology and
1743 Biological Sciences Research Council (BBSRC). RG, AR and MJP are supported by
1744 the Quadram Institute Bioscience BBSRC-funded Strategic Programme: Microbes in
1745 the Food Chain (project no. BB/R012504/1) and its constituent
1746 project BBS/E/F/000PR10351 (Theme 3, Microbial Communities in the Food Chain)
1747 and by the Medical Research Council CLIMB grant (MR/L015080/1), EMA was funded by the
1748 BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1
1749 and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356. JL and
1750 CP would like to extend further thanks to the Alborada Trust
1751 (<http://www.alboradatrust.com>) as part of their Alborada Well Foal study. The funders
1752 had no role in study design, data collection and analysis, decision to publish, or
1753 preparation of the manuscript.

1754

1755 **Acknowledgements**

1756 The authors would like to thank all horse owners, stud farms and vets for facilitating the
1757 collection of horse faecal samples. The authors would also like to thank the core
1758 bioinformatics team working at Quadram Institute Bioscience for their help in sample and
1759 data processing.

1760

1761 **Author Contributions**

1762 RG analysed the data, prepared figures and/or tables, authored or reviewed
1763 drafts of the paper, and approved the final draft. JL performed the experiments, authored or
1764 reviewed drafts of the paper, and approved the final draft. AR analysed the data, authored or
1765 reviewed drafts of the paper, and approved the final draft. EMA analysed the data, authored
1766 or reviewed drafts of the paper, and approved the final draft. AO analysed the data,
1767 prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the
1768 final draft. DB performed the experiments, authored or reviewed drafts of the paper, and
1769 approved the final draft. RML & CP conceived and designed the experiments, authored or
1770 reviewed drafts of the paper, and approved the final draft. MJP conceived and designed the
1771 experiments, analysed the data, prepared figures and/or tables, authored or reviewed drafts
1772 of the paper, and approved the final draft

1773

1774 **Data Availability**

1775 Data are available on the Sequence Read Archive at BioProject ID PRJNA590977 and
1776 Figshare: <https://doi.org/10.6084/m9.figshare.14268095>

1777

1778 **References**

- 1779 1. Roberts A. Tamed: ten species that changed our world. Random House; 2017.
- 1780 2. Fages A, Hanghøj K, Khan N, Gaunitz C, Seguin-Orlando A, Leonardi M, et al.
1781 Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time
1782 Series. *Cell*. 2019;177: 1419–1435.e31.
- 1783 3. Clarkson N. World horse population likely to be over 60 million, figures suggest. 2017.
1784 <https://www.horsetalk.co.nz/2017/07/10/world-horse-population-60-million/>
- 1785 4. Belaunzaran X, Bessa RJB, Lavín P, Mantecón AR, Kramer JKG, Aldai N. Horse-meat
1786 for human consumption — Current research and future opportunities. *Meat Science*.
1787 2015;108: 74–81.

- 1788 5. Association BET. Market Information. 2019. <http://www.beta-uk.org/pages/industry->
1789 information/market-information.php
- 1790 6. Costa MC, Weese JS. Understanding the Intestinal Microbiome in Health and Disease.
1791 Vet Clin North Am Equine Pract. 2018;34: 1–12.
- 1792 7. Julliand V, Grimm P. Horse species symposium: The microbiome of the horse hindgut:
1793 History and current knowledge. Journal of Animal Science. 2016;94: 2262–2274.
- 1794 8. Santos AS, Rodrigues MA, Bessa RJ, Ferreira LM, Martin-Rosset W. Understanding
1795 the equine cecum-colon ecosystem: current knowledge and future perspectives.
1796 Animal. 2011;5: 48–56.
- 1797 9. Maddox TW, Clegg PD, Williams NJ, Pinchbeck GL. Antimicrobial resistance in
1798 bacteria from horses: Epidemiology of antimicrobial resistance. Equine Vet J. 2015;47:
1799 756–765.
- 1800 10. Leng J, Proudman C, Darby A, Blow F, Townsend N, Miller A, et al. Exploration of the
1801 Fecal Microbiota and Biomarker Discovery in Equine Grass Sickness. J Proteome Res.
1802 2018;17: 1120–1128.
- 1803 11. Massacci FR, Clark A, Ruet A, Lansade L, Costa M, Mach N. Inter-breed diversity and
1804 temporal dynamics of the faecal microbiota in healthy horses. J Anim Breed Genet.
1805 2020;137: 103–120.
- 1806 12. O’ Donnell MM, Harris HM, Jeffery IB, Claesson MJ, Younge B, O’ Toole PW, et al.
1807 The core faecal bacterial microbiome of Irish Thoroughbred racehorses. Lett Appl
1808 Microbiol. 2013;57: 492–501.
- 1809 13. Promsopone B, Morishita TY, Aye PP, Cobb CW, Veldkamp A, Clifford JR. Evaluation
1810 of an avian-specific probiotic and Salmonella typhimurium-Specific antibodies on the
1811 colonization of Salmonella typhimurium in broilers. J Food Prot. 1998;61: 176–180.
- 1812 14. Stewart HL, Pitta D, Indugu N, Vecchiarelli B, Engiles JB, Southwood LL.
1813 Characterization of the fecal microbiota of healthy horses. Am J Vet Res. 2018;79:
1814 811–819.
- 1815 15. Edwards JE, Shetty SA, van den Berg P, Burden F, van Doorn DA, Pellikaan WF, et al.

- 1816 Multi-kingdom characterization of the core equine fecal microbiota based on multiple
1817 equine (sub)species. *Anim Microbiome*. 2020;2: 6.
- 1818 16. Metcalf JL, Song SJ, Morton JT, Weiss S, Seguin-Orlando A, Joly F, et al. Evaluating
1819 the impact of domestication and captivity on the horse gut microbiome. *Sci Rep*.
1820 2017;7: 15497.
- 1821 17. Leng J, Walton G, Swann J, Darby A, La Ragione R, Proudman C. “Bowel on the
1822 Bench”: Proof of Concept of a Three-Stage, In Vitro Fermentation Model of the Equine
1823 Large Intestine. *Appl Environ Microbiol*. 2019;86: e02093–19.
- 1824 18. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new
1825 genomic blueprint of the human gut microbiota. *Nature*. 2019;568: 499–504.
- 1826 19. Forster SC, Kumar N, Anonye BO, Almeida A, Viciani E, Stares MD, et al. A human gut
1827 bacterial genome and culture collection for improved metagenomic analyses. *Nat*
1828 *Biotechnol*. 2019;37: 186–192.
- 1829 20. Ravi A, Halstead FD, Bamford A, Casey A, Thomson NM, van Schaik W, et al. Loss of
1830 microbial diversity and pathogen domination of the gut microbiota in critically ill
1831 patients. *Microb Genom*. 2019;5: e000293.
- 1832 21. Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, Poplawski R, et al. CLIMB
1833 (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the
1834 medical microbiology community. *Microb Genom*. 2016;2: e000086.
- 1835 22. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2019.
1836 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- 1837 23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
1838 data. *Bioinformatics*. 2014;30: 2114–2120.
- 1839 24. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*.
1840 2012;9: 357–359.
- 1841 25. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
1842 Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25: 2078–2079.
- 1843 26. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2.

- 1844 Genome Biol. 2019;20: 257.
- 1845 27. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance
1846 in metagenomics data. PeerJ Computer Science. 2017;3: e104.
- 1847 28. Breitwieser FP, Salzberg SL. Pavian: Interactive analysis of metagenomics data for
1848 microbiomics and pathogen identification. BioRxiv.
1849 2016;<https://www.biorxiv.org/content/10.1101/084715v1.full>:
- 1850 29. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, et al. MEGAHIT v1.0: A fast
1851 and scalable metagenome assembler driven by advanced methodologies and
1852 community practices. Methods. 2016;102: 3–11.
- 1853 30. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an
1854 advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3: e1319.
- 1855 31. Wu S, Li T, Niu H, Zhu Y, Liu Y, Duan Y, et al. Effects of glucose oxidase on growth
1856 performance, gut function, and cecal microbiota of broiler chickens. Poultry Science.
1857 2019;98: 828–841.
- 1858 32. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning
1859 metagenomic contigs by coverage and composition. Nat Methods. 2014;11: 1144–
1860 1146.
- 1861 33. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive
1862 binning algorithm for robust and efficient genome reconstruction from metagenome
1863 assemblies. PeerJ. 2019;7: e7359.
- 1864 34. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of
1865 genomes from metagenomes via a dereplication, aggregation and scoring strategy.
1866 Nat Microbiol. 2018;3: 836–843.
- 1867 35. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing
1868 the quality of microbial genomes recovered from isolates, single cells, and
1869 metagenomes. Genome Research. 2015;25: 1043–1055.
- 1870 36. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al.
1871 Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the

- 1872 tree of life. *Nature Microbiology*. 2017;2: 1533–1542.
- 1873 37. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic
1874 comparisons that enables improved genome recovery from metagenomes through de-
1875 replication. *The ISME journal*. 2017
- 1876 38. Parks D. CompareM. 2020. <https://github.com/dparks1134/CompareM>
- 1877 39. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify
1878 genomes with the Genome Taxonomy Database. *Bioinformatics*. 2019;36: 1925–1927.
- 1879 40. von Meijenfeldt FAB, Arkhipova K, Cambuy DD, Coutinho FH, Dutilh BE. Robust
1880 taxonomic classification of uncharted microbial sequences and bins with CAT and
1881 BAT. *Genome Biol*. 2019;20: 217.
- 1882 41. Schwengers O, Hain T, Chakraborty T, Goesmann A. ReferenceSeeker: rapid
1883 determination of appropriate reference genomes. *Journal of Open Source Software*.
1884 2020;5: 1994.
- 1885 42. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new
1886 view of the tree of life. *Nat Microbiol*. 2016;1: 16048.
- 1887 43. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high
1888 throughput. *Nucleic Acids Res*. 2004;32: 1792–1797.
- 1889 44. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated
1890 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:
1891 1972–1973.
- 1892 45. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees
1893 for large alignments. *PLoS One*. 2010;5: e9490.
- 1894 46. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal:
1895 prokaryotic gene recognition and translation initiation site identification. *BMC*
1896 *Bioinformatics*. 2010;11: 119.
- 1897 47. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, et al. Precise
1898 phylogenetic analysis of microbial isolates and genomes from metagenomes using
1899 PhyloPhlAn 3.0. *Nat Commun*. 2020;11: 2500.

- 1900 48. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND.
1901 Nat Methods. 2015;12: 59–60.
- 1902 49. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple
1903 sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:
1904 3059–3066.
- 1905 50. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1906 large phylogenies. Bioinformatics. 2014;30: 1312–1313.
- 1907 51. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan:
1908 Community Ecology Package. R package version 2.5–6. 2019. 2019. [https://cran.r-](https://cran.r-project.org/web/packages/vegan/index.html)
1909 [project.org/web/packages/vegan/index.html](https://cran.r-project.org/web/packages/vegan/index.html)
- 1910 52. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag; 2016.
- 1911 53. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa S, Solden LM, et al. DRAM
1912 for distilling microbial metabolism to automate the curation of microbiome function.
1913 Nucleic Acids Research. 2020;48: 8883–8900.
- 1914 54. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences.
1915 Gene Prediction. Springer; 2019. pp. 1–14.
- 1916 55. Seemann T. Barnap. 2018. <https://github.com/tseemann/barnap>
- 1917 56. Rouger A, Moriceau N, Prévost H, Remenant B, Zagorec M. Diversity of bacterial
1918 communities in French chicken cuts stored under modified atmosphere packaging.
1919 Food Microbiol. 2018;70: 7–16.
- 1920 57. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial
1921 genomic data. PeerJ. 2015;3: e985.
- 1922 58. Ryan F. Demovir. 2019. <https://github.com/feargalr/Demovir>
- 1923 59. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al.
1924 Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-
1925 sharing networks. Nature Biotechnology. 2019;37: 632–639.
- 1926 60. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a
1927 software environment for integrated models of biomolecular interaction networks.

- 1928 Genome research. 2003;13: 2498–2504.
- 1929 61. Calusinska M, Goux X, Fossépré M, Muller EEL, Wilmes P, Delfosse P. A year of
1930 monitoring 20 mesophilic full-scale bioreactors reveals the existence of stable but
1931 different core microbiomes in bio-waste and wastewater anaerobic digestion systems.
1932 Biotechnol Biofuels. 2018;11: 196.
- 1933 62. Limam RD, Chouari R, Mazeas L, Wu TD, Li T, Grossin-Debattista J, et al. Members of
1934 the uncultured bacterial candidate division WWE1 are implicated in anaerobic digestion
1935 of cellulose. Microbiologyopen. 2014;3: 157–167.
- 1936 63. Laconi A, Mughini-Gras L, Tolosi R, Grilli G, Trocino A, Carraro L, et al. Microbial
1937 community composition and antimicrobial resistance in agricultural soils fertilized with
1938 livestock manure from conventional farming in Northern Italy. Sci Total Environ.
1939 2021;760: 143404.
- 1940 64. Cross KL, Campbell JH, Balachandran M, Campbell AG, Cooper SJ, Griffen A, et al.
1941 Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. Nat
1942 Biotechnol. 2019;37: 1314–1321.
- 1943 65. Costa MC, Arroyo LG, Allen-Vercoe E, Stämpfli HR, Kim PT, Sturgeon A, et al.
1944 Comparison of the fecal microbiota of healthy horses and horses with colitis by high
1945 throughput sequencing of the V3-V5 region of the 16S rRNA gene. PLoS One. 2012;7:
1946 e41484.
- 1947 66. Morita H, Shiratori C, Murakami M, Takami H, Kato Y, Endo A, et al. *Lactobacillus*
1948 *hayakitensis* sp. nov., isolated from intestines of healthy thoroughbreds. International
1949 journal of systematic and evolutionary microbiology. 2007;57: 2836.
- 1950 67. Endo A, Roos S, Satoh E, Morita H, Okada S. *Lactobacillus equigenerosi* sp. nov., a
1951 coccoid species isolated from faeces of thoroughbred racehorses. International Journal
1952 of Systematic and Evolutionary Microbiology. 2008;58: 914–918.
- 1953 68. Morita H, Nakano A, Shimazu M, Toh H, Nakajima F, Nagayama M, et al. *Lactobacillus*
1954 *hayakitensis*, *L. equigenerosi* and *L. equi*, predominant lactobacilli in the intestinal flora
1955 of healthy thoroughbreds. Animal Science Journal. 2009;80: 339–346.

- 1956 69. Andrewes FW, TJ H. A study of the streptococci pathogenic for man. *Lancet*. 1906;ii:
1957 708–713.
- 1958 70. Smith DG, Shattock PM. The serological grouping of *Streptococcus equinus*. *J Gen*
1959 *Microbiol*. 1962;29: 731–736.
- 1960 71. Daly K, Stewart CS, Flint HJ, Shirazi-Beechey SP. Bacterial diversity within the equine
1961 large intestine as revealed by molecular analysis of cloned 16S rRNA genes. *FEMS*
1962 *microbiology ecology*. 2001;38: 141–151.
- 1963 72. Zhu W, Dong K, Yang J, Lu S, Lai XH, Pu J, et al. *Acinetobacter lanii* sp. nov.,
1964 *Acinetobacter shaoyimingii* s p. nov. and *Acinetobacter wanghuae* sp. nov., isolated
1965 from faeces of *Equus kiang*. *Int J Syst Evol Microbiol*. 2021;71:
1966 10.1099/ijsem.0.004567.
- 1967 73. Wang W, Hu H, Zijlstra RT, Zheng J, Gänzle MG. Metagenomic reconstructions of gut
1968 microbial metabolism in weanling pigs. *Microbiome*. 2019;7: 48.
- 1969 74. Watanabe Y, Nagai F, Morotomi M. Characterization of *Phascolarctobacterium*
1970 *succinatutens* sp. nov., an asaccharolytic, succinate-utilizing bacterium isolated from
1971 human feces. *Appl Environ Microbiol*. 2012;78: 511–518.
- 1972 75. Molin G, Ternström A, Ursing J. Notes: *Pseudomonas lundensis*, a New Bacterial
1973 Species Isolated from Meat. *International Journal of Systematic and Evolutionary*
1974 *Microbiology*. 1986;36: 339–342.
- 1975 76. Scales BS, Erb-Downward JR, Falkowski NR, LiPuma JJ, Huffnagle GB. Genome
1976 Sequences of 12 *Pseudomonas lundensis* Strains Isolated from the Lungs of Humans.
1977 *Genome Announc*. 2018;6: e01461–17.
- 1978 77. Ouwkerk JP, Aalvink S, Belzer C, de Vos WM. *Akkermansia glycaniphila* sp. nov., an
1979 anaerobic mucin-degrading bacterium isolated from reticulated python faeces. *Int J*
1980 *Syst Evol Microbiol*. 2016;66: 4614–4620.
- 1981 78. Murru F, Fliegerova K, Mura E, Mrázek J, Kopečný J, Moniello G. A comparison of
1982 methanogens of different regions of the equine hindgut. *Anaerobe*. 2018;54: 104–110.
- 1983 79. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. A

- 1984 complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol.*
1985 2020;38: 1079–1086.
- 1986 80. Tamaki H, Tanaka Y, Matsuzawa H, Muramatsu M, Meng XY, Hanada S, et al.
1987 *Armatimonas rosea* gen. nov., sp. nov., of a novel bacterial phylum, *Armatimonadetes*
1988 phyl. nov., formally called the candidate phylum OP10. *Int J Syst Evol Microbiol.*
1989 2011;61: 1442–1447.
- 1990 81. Jeong JY, Park HD, Lee KH, Weon HY, Ka JO. Microbial community analysis and
1991 identification of alternative host-specific fecal indicators in fecal and river water
1992 samples using pyrosequencing. *J Microbiol.* 2011;49: 585–594.
- 1993 82. Li M, Monaco MH, Wang M, Comstock SS, Kuhlenschmidt TB, Fahey GC, et al.
1994 Human milk oligosaccharides shorten rotavirus-induced diarrhea and modulate piglet
1995 mucosal immunity and colonic microbiota. *ISME J.* 2014;8: 1609–1620.
- 1996 83. Mu C, Yang Y, Su Y, Zoetendal EG, Zhu W. Differences in Microbiota Membership
1997 along the Gastrointestinal Tract of Piglets and Their Differential Alterations Following
1998 an Early-Life Antibiotic Intervention. *Front Microbiol.* 2017;8: 797.
- 1999 84. Poppel MT, Skiebe E, Laue M, Bergmann H, Ebersberger I, Garn T, et al.
2000 *Acinetobacter equi* sp. nov., isolated from horse faeces. *International Journal of*
2001 *Systematic and Evolutionary Microbiology.* 2016;66: 881–888.
- 2002 85. Shangong W, Tianheng G, Yingzhen Z, Weiwei W, Yingyin C, Guitang W. Microbial
2003 diversity of intestinal contents and mucus in yellow catfish (*Pelteobagrus fulvidraco*).
2004 *Aquaculture.* 2010;303: 1 – 7.
- 2005 86. He B, Jin S, Cao J, Mi L, Wang J. Metatranscriptomics of the Hu sheep rumen
2006 microbiome reveals novel cellulases. *Biotechnol Biofuels.* 2019;12: 153.
- 2007 87. Babenko VV, Millard A, Kulikov EE, Spasskaya NN, Letarova MA, Konanov DN, et al.
2008 The ecogenomics of dsDNA bacteriophages in feces of stabled and feral horses.
2009 *Comput Struct Biotechnol J.* 2020;18: 3457–3467.
- 2010 88. Wittmann J, Turner D, Millard AD, Mahadevan P, Kropinski AM, Adriaenssens EM.
2011 From Orphan Phage to a Proposed New Family-the Diversity of N4-Like Viruses.

- 2012 Antibiotics (Basel). 2020;9: 663.
- 2013 89. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from
2014 microbiomes using nanopore sequencing. Nat Biotechnol. 2020;38: 701–707.
- 2015 90. Nicholls SM, Quick JC, Tang S, Loman NJ. Ultra-deep, long-read nanopore
2016 sequencing of mock microbial community standards. Gigascience. 2019;8:
- 2017 91. Xie H, Yang C, Sun Y, Igarashi Y, Jin T, Luo F. PacBio Long Reads Improve
2018 Metagenomic Assemblies, Gene Catalogs, and Genome Binning. Front Genet.
2019 2020;11: 516269.

2020

2021 **Supporting information**

2022 **Additional File 1. (xlsx.)**

2023 **S1 Table. Sequence Summaries.** Summaries of sequencing data from 5 metagenomic
2024 samples sourced from equine faeces from BioProject PRJNA590977

2025 **S2 Table. Read-based taxonomic analysis.** Bracken read based relative abundance
2026 values for 5 equine faecal samples from BioProject PRJNA590977.

2027 **S3 Table. Metagenome Assembled Genome statistics.** CheckM genome statistics for
2028 MAG catalogue following assembly of metagenomic reads from 5 horse faecal samples. For
2029 High and medium quality sequences, clusters at 95% and 99% ANI have been detailed.

2030 **S4 Table. tRNA presence in high and medium quality MAGs.** tRNA presence across
2031 MAG catalogue following assembly of metagenomic reads from 5 horse faecal samples.

2032 **S5 Table. MAG taxonomic assignments.** Taxonomic analysis for de-replicated
2033 metagenomic species according to GTDB (release 95), CAT/BAT (NCBI nr database) and
2034 ReferenceSeeker (RefSeq database). Newly assigned Latin binomials have been provided
2035 where appropriate.

2036 **S6 Table. AAI analysis for novel genera.** Average amino acid identity (AAI) scores for all
2037 genomes of novel genera as determined by CompareM.

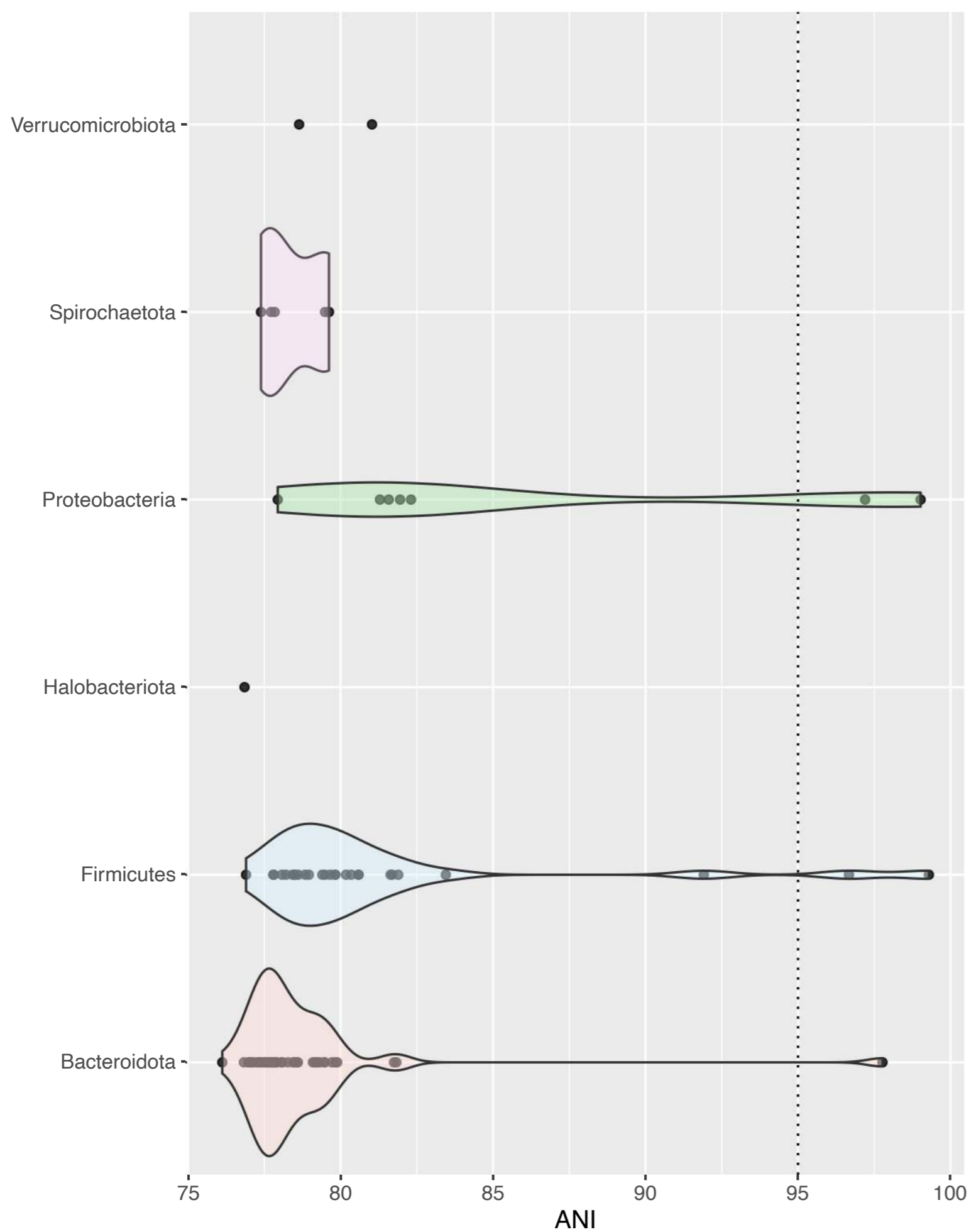
2038 **S7 Table. AAI characterisation for novel *Armatimonadota* species.** Reference
2039 sequences associated with relevant *Armatimonadota* species according to GTDB and NCBI
2040 used for phylogenomic and AAI analysis against our query *Armatimonadota* sequence.
2041 **S8 Table. Distribution analysis of recovered MAGs.** Coverage statistics for 110
2042 metagenomic species recovered from 5 metagenomic samples derived from equine faeces.
2043 **S9 Table. Functional annotation of recovered MAGs.** Presence of genes associated with
2044 CAZyme function across 123 MAGs recovered from metagenomic reads of 5 horse faecal
2045 samples.
2046 **S10 Table. Genome and quality analysis of with recovered phage sequences.** CheckV
2047 summary statistics of all VirSorter Category 1 and Category 2 phages >5 kb in length
2048 derived from metagenomic assemblies of horse faecal samples. For all High and medium
2049 quality phage sequences, further detail of taxonomic annotation and sequences coverage
2050 have been provided.
2051 **S11 Table. Protein based clustering of high quality or complete phage sequences.**
2052 vCONTACT2 output for de-replicated catalogue of 190 phage genomes derived from equine
2053 faecal samples and classified as 'High-quality' or 'Complete' by CheckV
2054

a.

Tree scale: 1

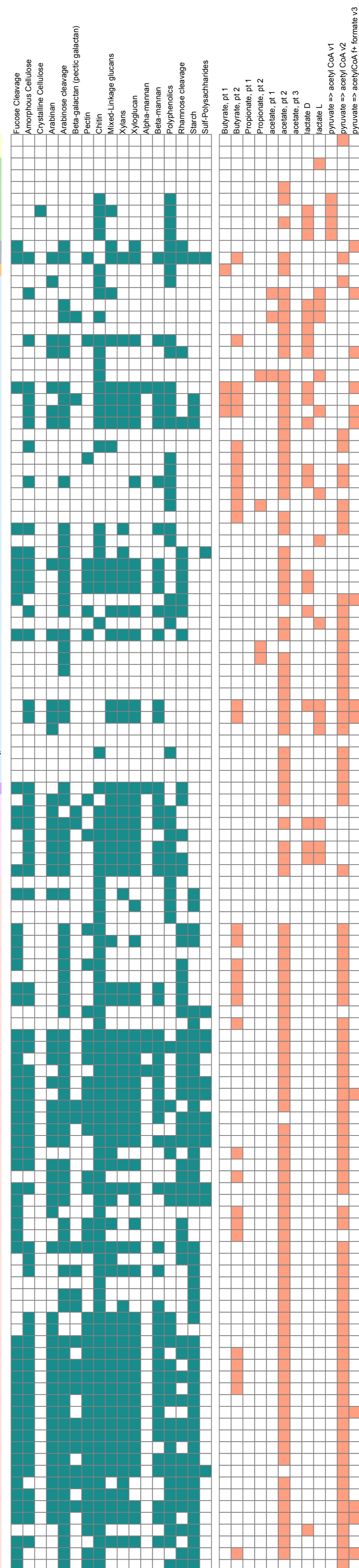
b.

bioRxiv preprint doi: <https://doi.org/10.1101/2021.04.30.442084>; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

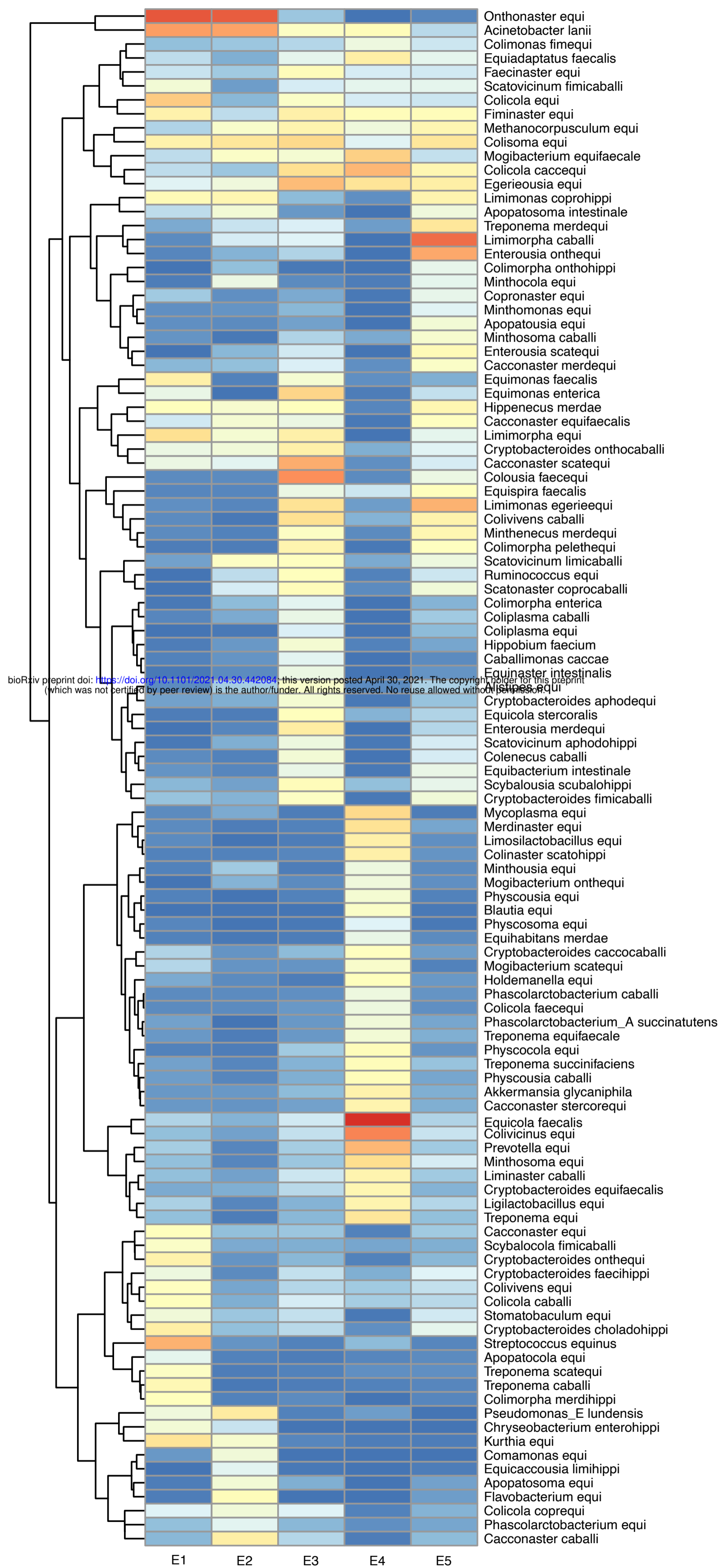


Polymer Degradation

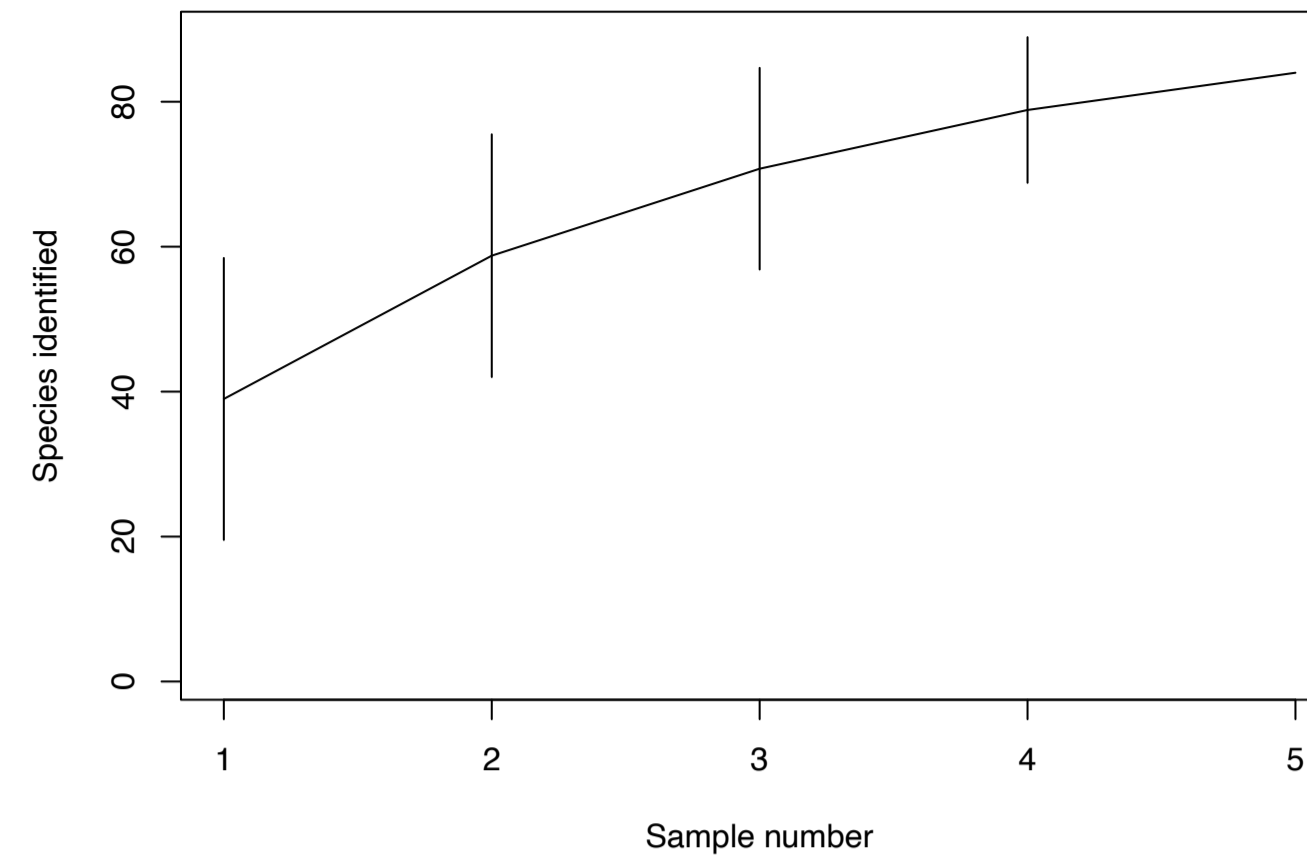
SCFA Production



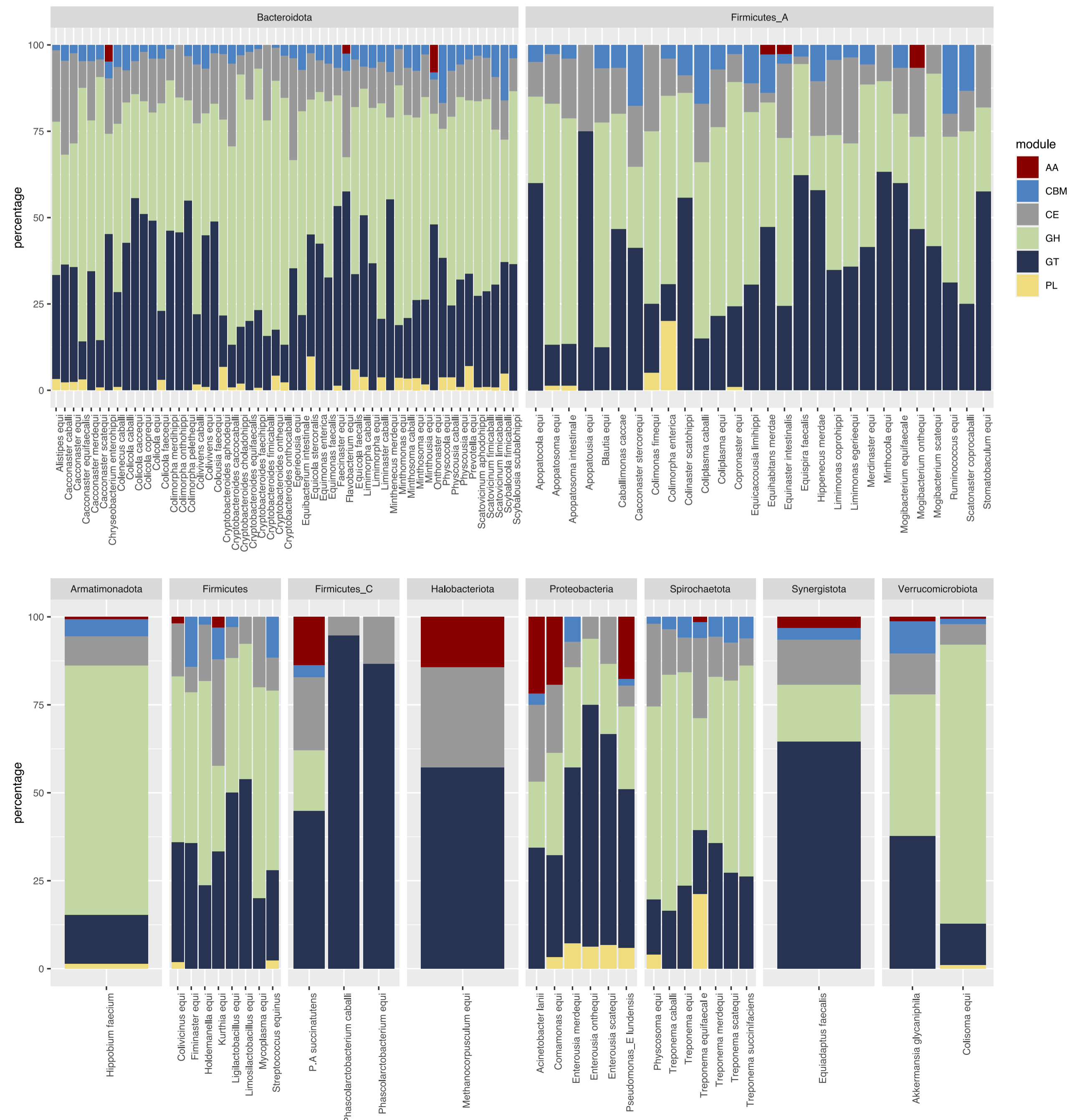
a.



b.



c.

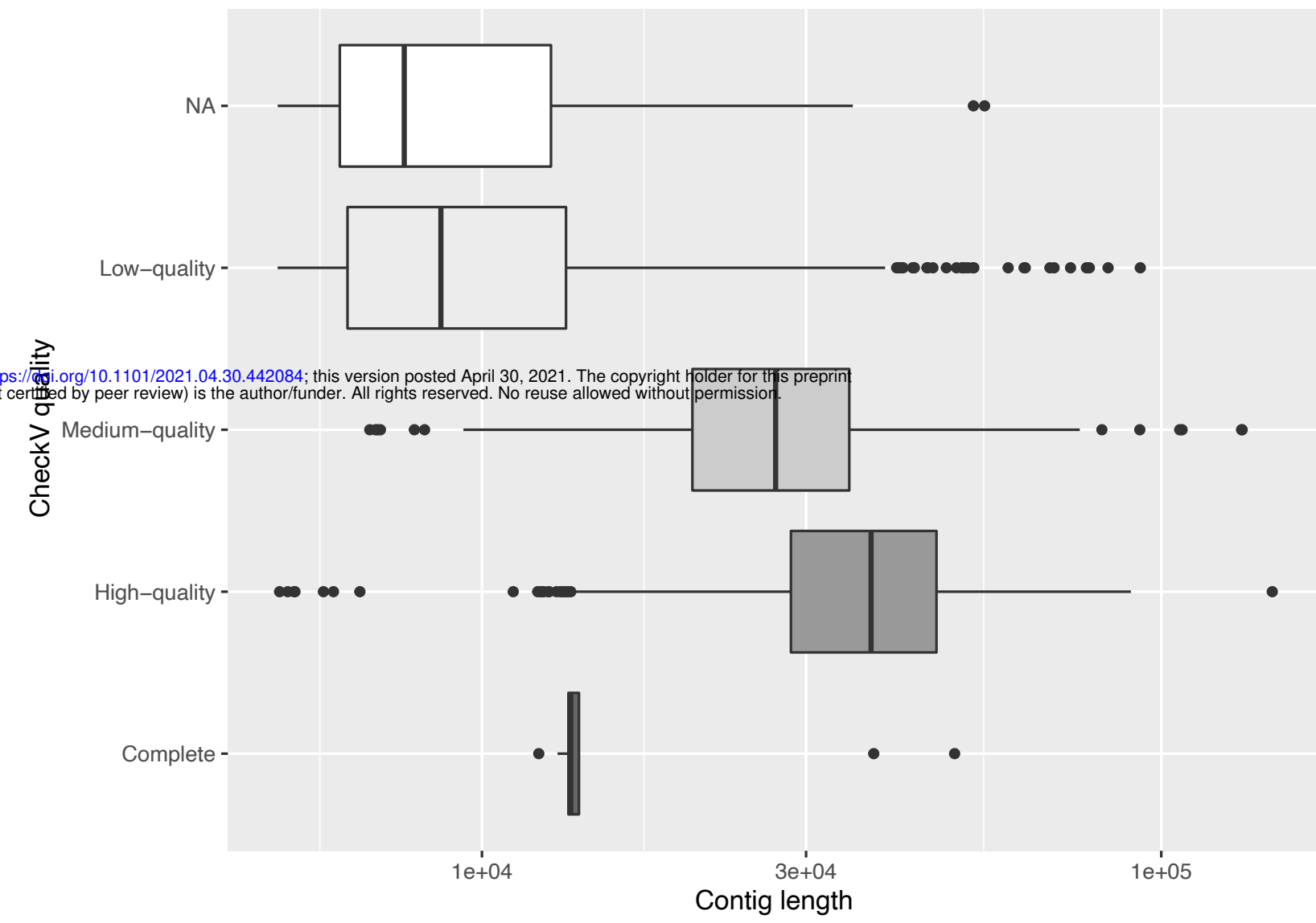


bioRxiv preprint doi: <https://doi.org/10.1101/2021.04.30.442084>; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

a.

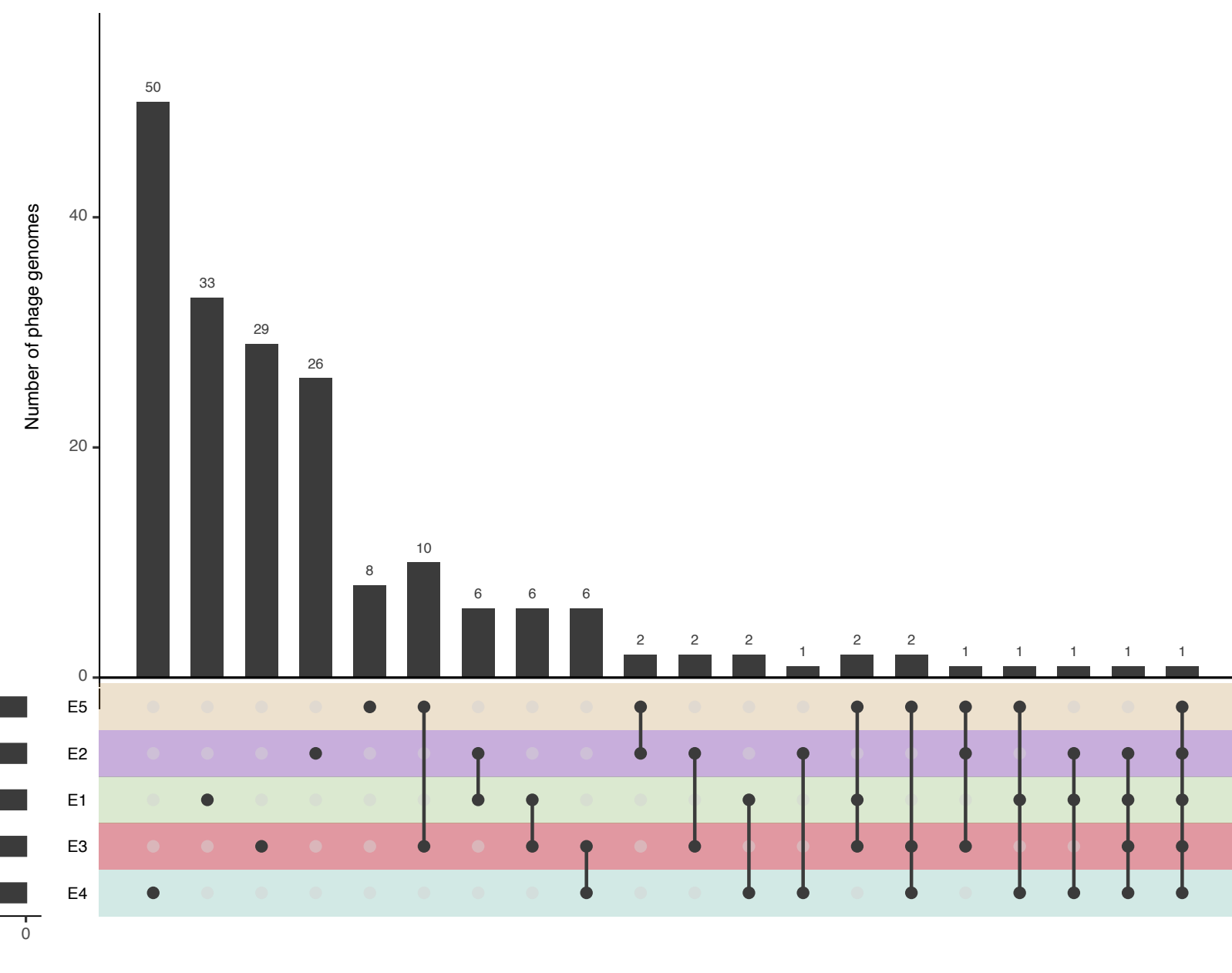
checkV quality

- Complete
- High-quality
- Medium-quality
- Low-quality
- NA



bioRxiv preprint doi: <https://doi.org/10.1101/2021.04.30.442084>; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

c.



b.

