bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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.

1 Description of *Schaedlerella arabinophila gen. nov.*, *sp. nov.*, a D-arabinose utilizing
2 bacterium isolated from feces of C57BL/6J mice and a close relative of *Clostridium* sp. ASF
3 502

4 Melissa Soh[a*], Sou Miyake[a*], Austin Lim[a], Henning Seedorf[a,b,#]

5

6 [a]Temasek Life Sciences Laboratory, 1 Research Link, Singapore 117604, Singapore
7 [b]Department of Biological Sciences, National University of Singapore, Singapore 117604,
8 Singapore

9

10

11

12

13

14

15

16

17

18 [#]Corresponding author: Henning Seedorf, henning@tll.org.sg

19 *authors contributed equally

20 Running title: Description of *Schaedlerella arabinophila*

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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.

21 **Abstract**

22 The use of gnotobiotics has gained large interest in recent years due to technological advances
23 that have revealed the importance of host-associated microbiomes for host physiology and
24 health. One of the oldest and most important gnotobiotics mouse model, the Altered Schaedler
25 Flora (ASF) has been used for several decades. ASF comprises eight different bacterial species,
26 which have been characterized to different extent, but only few are available through public
27 strain collections. Here, the isolation of a close relative to one of the less studied ASF strains,
28 *Clostridium* sp. ASF 502, is reported. Isolate TLL-A1, which shares 99.6% 16S rRNA gene
29 sequence identity with *Clostridium* sp. ASF 502, was obtained from feces of C57BL/6J mice
30 where is was detectable at a relative abundance of less than one percent. D-arabinose was used
31 as sole carbon source in the anaerobic cultivation medium. Growth experiments with TLL-A1
32 on different carbon sources and analysis of its ~6.5 gigabase genome indicate that TLL-A1
33 harbors a large gene repertoire to utilize different carbohydrates for growth. Comparative
34 genome analyses of TLL-A1 and *Clostridium* sp. ASF 502 reveal differences in genome
35 content between the two strains, in particular with regards to carbohydrate activating enzymes.
36 Based on physiology and genomic analysis it is proposed to name TLL-A1 to *gen. nov. sp. nov*
37 *Schaedlerella arabinophila* TLL-A1 (DSMZ 106076^T; KCTC 15657^T). The closely related
38 *Clostridium* sp. ASF 502 is proposed to be renamed to *Schaedlerella arabinophila* to reflect its
39 taxonomic standing and to keep ‘ASF 502’ as strain designation.

40

41

42

43

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

44 **Importance**

45 The Altered Schaedler Flora (ASF) remains an important tool to mechanistically investigate
46 host-microbe interactions in the mammalian alimentary tract. Extensively characterizing the
47 eight different bacterial strains, which are constituting ASF, has the potential to further increase
48 the definition of this widely used model microbial community and to enhance our
49 understanding of how individual microorganism interact with a host and/or how they may
50 affect its physiology. However, some of the ASF strains have unfortunately been lost or are not
51 easily accessible to the scientific community. The isolation and characterization of the here
52 described species, proposed to be named *Schaedlerella arabinophila,* which is closely related
53 to *Clostridium* species ASF502, may therefore be an important corner stone in further
54 improving the value of studies using ASF or other defined synthetic microbial communities
55 that require usage of autochthonous microorganisms in gnotobiotic mice.

56

57

58

59

60

61

62

63

64

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

65 **Introduction**

66 Microbiome studies have revealed extensive insights into the complex associations of
67 microorganisms with specific phenotypes of host physiology and health (1, 2). A major focus
68 of the studies remains on the mammalian intestinal tract, which is densely populated with
69 microorganisms from all three domains of life (3, 4). However, it is often not well established
70 whether an association of microbial taxa (or a collection thereof) with a host phenotype could
71 also equate to being the underlying cause. Establishing this causality still often relies on the
72 use of model organisms, such as the laboratory mouse, which offers several advantages: they
73 are well-characterized, can be genetically modified, and can be reared germ-free. The latter is
74 of particular interest as it allows to perform experiments with gnotobiotic animals (germ-free
75 animals that are associated with defined microbiota). The use of gnotobiotic mice was already
76 recognized in the middle of the last century when it became possible to rear germ-free mice
77 and to associate them with autochthonous microorganisms (5). One of the most prominent
78 examples from this era is the Altered Schaedler Flora (ASF). ASF is a defined microbial
79 community consisting of eight bacterial species, and has been used for a large number of studies
80 that investigate host-microbe interactions (6-8). ASF is an updated form of the 'Schaedler flora'
81 and was revised by Orcutt *et al*. in 1978 to standardize its use in gnotobiotic experiments (9,
82 10). At the time, the strain characterization and selection was mostly based on physiology and
83 microscopy from the cultivatable fraction of the enteric microbial community, given that
84 sequence-based characterization of microorganisms was still in its infancy. Subsequent
85 phylogenetic analyses of the 16S rRNA genes enabled additional taxonomic identification of
86 the strains, and thus the strains were named ASF 356 (*Clostridium* species), ASF 360
87 (*Lactobacillus intestinalis*), ASF 361 (*Lactobacillus murinus*), ASF 457 (*Mucispirillum*
88 *schaedleri*), ASF 492 (*Eubacterium plexicaudatum*), ASF 500 (*Pseudoflavonifractor* species),
89 ASF 502 (*Clostridium* species) and ASF 519 (*Parabacteroides goldsteinii*). See Dewhirst *et*

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

90 *al*. for more details on phylogenetic information of each ASF strain (11). Meanwhile, all eight
91 strains of ASF were genome sequenced (to draft quality) which have further characterized them
92 and have added value to using ASF as model community in gnotobiotics experiments (12).

93 However, the amount of information available on physiology and ecological role varies
94 greatly between each of the ASF strains, as some of them remain difficult to cultivate.
95 Detrimental to further characterization efforts is the fact that some of the original ASF strains
96 are not available (anymore) through public strain collections (11), and are therefore not easily
97 accessible. In some cases (e.g. ASF 492 (*Eubacterium plexicaudatum*), the original strain has
98 been lost (11). Nevertheless, ASF remains a valuable model, and close relatives have been
99 isolated for some of the lost strains. One recent example is the isolation and formal description
100 of *Mucispirillum schaedleri* ATCC BAA-1009^{T}, a close relative of ASF 457 (13). The type
101 strain has undergone additional characterizations that provide insights into the lifestyle of this
102 species in the mucus layer of the mouse intestinal tract (14), which may also have implications
103 for interpreting the role of ASF 457 in ASF and the use of ASF in general. For other strains,
104 such as ASF 502, less information is available. A qPCR-based study indicated that the
105 abundance of this particular strain can be highest of all eight ASF strains in ASF-harbouring
106 mice feces (15). The same study also revealed that the abundance level of the strain varied
107 considerably in the different sections of the intestinal tract, highlighting that this species may
108 have important roles in specific segments of the intestinal tract of ASF-harbouring mice (15).

109 Here, we present the isolation and formal description of a new genus and species,
110 *Schaedlerella arabinophila* TLL-A1. *S. arabinophila* TLL-A1 was isolated from feces of
111 C57BL/6J mouse and its 16S rRNA gene shares 99.6 percent sequence identity with that of
112 *Clostridium* species ASF 502. Concurrent to phenotypic and genomic characterization, 16S
113 rRNA-amplicon sequencing was performed to determine the relative abundance of *S.*
114 *arabinophilus* in healthy mouse fecal samples. Lastly, the genome of *S. arabinophila* was

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

115 sequenced and compared to the available draft genome of *Clostridium* species ASF 502 to
116 identify differences between the two strains in genome content.

117 **Results**

118 **Isolation and whole genome analysis of *S. arabinophila* TLL-A1**

119 *S. arabinophila* TLL-A1 was isolated as part of a larger cultivation effort that screened for
120 utilization of specific carbon sources by mouse intestinal microorganisms. For convenience,
121 the isolated strain TLL-A1 and the closest known relative *Clostridium* sp. ASF 502 (GenBank
122 ID GCA_000364245.1) will henceforth be referred to as TLL-A1 and ASF 502,
123 respectively.The colonies of TLL-A1 were initially obtained on solid medium containing 1.6
124 g per litre D-arabinose. The observed colonies appeared to be uniform in shape and colour. Ten
125 colonies were selected for growth in liquid medium and DNA was extracted from cultures. No
126 other microbial species were detected on clone libraries and the obtained strains (all identical)
127 had highest sequence identity to ASF 502. The purity of cultures was confirmed by serial
128 plating and selection of colonies before inoculating into larger liquid medium to obtain
129 sufficient cell mass for genome sequencing.

130 The gDNA was sequenced on PacBio RSII and raw reads were assembled by HGAP 4
131 with further error correction conducted by Pilon. The Pilon assembly yielded 3 contigs for a
132 total assembly size of 6,521,014 bp, with G+C content of 47.9%, N50 of 6,346,997, and an
133 average coverage of 107x across the genome. The large N50 is due to the three contigs having
134 uneven size (6.35 Mb, 0.16 Mb and 9.5 kb). The two shorter contigs were not identified as
135 plasmids with Plasmid Finder (16). Reads from MiSeq paired-end sequencing was used for
136 further error correction using Pilon (17). Because the assembly by SMRT analysis was already
137 of high quality, Pilon only corrected for 10 locations totalling 11 bases. The total assembled
138 genome was 6.52 Mb, comparable to 6.36 Mb draft genome published for *Clostridium* sp. ASF
139 502 (12). Circular display illustrates the three contigs, CDS on forward and reverse reads, non-

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

140 CDS features, AMR genes, transporters as well as GC content and GC skew (Figure 1). The
141 genome has an estimated completeness of 94.7% and contamination of 2.44% based on 333
142 marker genes conserved in Lachnospiraceae as identified by checkM (Table S1). The assembly
143 contain 6,724 CDSs; of which 59 were RNAS (6 rRNA and 53 tRNA genes), 313 repeat
144 regions, 23 and 22 CRISPR repeat and spacer regions, respectively. There are 3,225 genes
145 (48.0%) assigned putative functions and 3,499 hypothetical proteins. 3,632 are assigned as
146 FIGfam. Table S2 summarises the genome statistics. All annotations are publically available
147 online (www.patricbrc.org) under genome ID 97139.17.

148 **16S rRNA gene and whole-genome phylogeny of *S. arabinophila***

149 16S rRNA gene phylogeny of *S. arabinophila* strain TLL-A1 and closely related *Clostridium*
150 strain ASF 502 (GenBank ID GCA_000364245.1) show that two are closely related, with
151 *Clostridium hylemonae* and *Clostridium scindens* the closest known relatives outside the clade
152 (Figure 2A). This is in stark contrast to whole-genome phylogeny where TLL-A1 and ASF 502
153 are clustered with *C. glycyrrhizinilyticum* and *Ruminococcus* species (Figure 2B). Note that the
154 difference may be due to bias in the availability of 16S rRNA gene and whole genomes. At
155 16S rRNA gene level, TLL-A1 and ASF 502 shared 99.6% sequence identity. Both average
156 amino acid identity (AAI) and genome sequence-based delineation (GGDC) confirmed two
157 strains to have high probably of belonging to the same species (Two-way AAI: 94.46% (SD:
158 14.50%) from 3976 proteins; Probability that DDH > 70% = 92%, respectively). Given that the
159 two strains are considered the same species, we then opted to determine their phylogenetic
160 relationship with the closest sequenced relative, *Clostridium glycyrrhizinilyticum* JCM 13369
161 (NCBI accession no. NZ_BBAB00000000). GGDC showed that the probability that *C.*
162 *glycyrrhizinilyticum* and TLL-A1 belonged to the same species (i.e. DDH > 70%) was 0% (via
163 logistic regression). At 16S rRNA gene level, the two species shared 93.25% sequence identity
164 indicating that TLL-A1 and ASF 502 together likely represent a new genus. Thus, strains TLL-

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

165 A1 and ASF 502 were named *Schaedlerella arabinophila* TLL-A1 and *Schaedlerella*
166 *arabinophila* ASF 502, respectively (formal description below).

167 **Comparative genome analysis of *S. arabinophila* with *Clostridium* sp. ASF 502**

168 Genome comparison indicated similar general parameters for the two strains (Table S2). In
169 addition, overall differences in the genomes were compared using clusters of orthologous genes
170 (COGs). Of 4,226 and 4,257 COGs identified in TLL-A1 and ASF 502, 4081 were shared
171 between the two, which meant that 145 (3.43%) and 176 (4.13%) were unique to TLL-A1 and
172 ASF 502, respectively (Table S3 and Table S4). For the unique clusters, whist most of the
173 Swiss-Prot annotation remained unclassified, those given Gene Ontology (GO) terms were
174 grouped according to their molecular functions (Figure 3). TLL-A1 had higher occurrence of
175 unique genes related to cofactor binding (GO:0048037), electron carrier activity
176 (GO:0009055), hydrolase activity (GO:0016787), ion binding (GO:0043167), isomerase
177 activity (GO:0016853), metal cluster binding (GO:0051540), nucleotide binding
178 (GO:0000166), nutrient reservoir activity (GO:0045735), oxidoreductase activity
179 (GO:0016491), transferase activity (GO:0016740) and transporter activity (GO:0005215);
180 while ASF 502 had more of catalytic activity (GO:0003824), lyase activity (GO:0016829),
181 molecular function (GO:0003674), protein binding (GO:0005515) and signal transducer
182 activity (GO:0004871) genes. In particular, electron carrier activity (GO:0009055), metal
183 cluster binding (GO:0051540) and nutrient reservoir activity (GO:0045735) were exclusive to
184 TLL-A1 while catalytic activity (GO:0003824), lyase activity (GO:0016829) and protein
185 binding (GO:0005515) were only found in ASF 502 (Figure 3).

186 Carbohydrate Active EnZymes (CAZymes) are set of enzymes involved in the
187 synthesis, breakdown and transport of the carbohydrates. Given the difference observed in the
188 unique portions of the COGs, we further compared the CAZymes between the two genomes to
189 see if there are any clear differences in the carbohydrate-degrading enzymes. The strain TLL-

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

190 A1 genome encodes 168 CAZymes, which were classified as 124 glycoside hydrolases (GHs),
191 30 glycosyl transferases (GTs), 7 carbohydrate esterases (CEs) and 7 carbohydrate-binding
192 modules (CBMs), while ASF 502 genome consisted of 130 CAZymes, classified into 90 GHs,
193 28 GTs, 6 CEs and 6 CMBs (Figure 4). Family level classification of GHs showed both
194 genomes comprised of similar enzymes, except TLL-A1 had higher counts of GHs. β-
195 glucosidase/ β-galactosidase, responsible for cleaving oligosaccharides, were the most
196 abundant enzymes found in both genomes, comprising 7 families (GH1, 2, 3, 31, 35, 36, 42) in
197 varying number of genes. Other abundant GHs included GH78 (α-L-rhamnosidase (EC
198 3.2.1.40), 13 and 6 in TLL-A1 and ASF 502 respectively), GH13 (α-amylase (EC 3.2.1.1), 10
199 and 7), GH39 (α-L-iduronidase (EC 3.2.1.76), 9 and 5) and GH43 (β-xylosidase (EC 3.2.1.37),
200 8 and 5). Interestingly, there were only three unique GHs in either of the genomes, which were
201 GH130 (β-1,4-mannosylglucose phosphorylase (EC 2.4.1.281)) and GH 136 (lacto-N-
202 biosidase (EC 3.2.1.140)) found in TLL-A1 and GH20 (β-hexosaminidase (EC 3.2.1.52))
203 found in ASF 502.

204 Distribution of GTs were also similar between TLL-A1 and ASF 502 genomes. GT2
205 (cellulose synthase (EC 2.4.1.12) / chitin synthase (EC 2.4.1.16)) was the most abundant,
206 followed by GT4 (sucrose synthase (EC 2.4.1.13)), GT35 (glycogen or starch phosphorylase
207 (EC 2.4.1.1)) and GT51 (murein polymerase (EC 2.4.1.129)). In comparison to GHs and GTs,
208 CEs and CBMs were both less abundant, with only three families identified each (CE2, 4, 9;
209 CBM34, 48, 50).

210 **Physiological properties of *S. arabinophila strain* TLL-A1**

211 In terms of physical characteristics, TLL-A1 was found to be Gram-negative and rod shaped.
212 Doubling time under experimental condition was ~6.5h and cells reached a length of three to
213 five μm at the end of exponential/early stationary phase. Light microscopy did indicate

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

214 refractive structures at the terminal section in a smaller subset of observed cells. However,
215 spore staining was also not conclusive as only few cells seem to harbor potential spores.
216 The isolation medium for TLL-A1 contained D-arabinose as main carbon source, but
217 additional carbon sources (listed in the methods section) were tested as growth substrates for
218 this strain (see Table S5 for results). Overall, the growth experiments suggest that *S.*
219 *arabinophila* strain TLL-A1 can utilize a wide variety of different carbon sources.
220 Results from the characterizations performed at DSMZ are given below and in the
221 corresponding tables. Analysis of polar lipids indicates primary presence of phospholipids,
222 glycolipids and phosphoglycolipids (Figure S1). Analysis of cellular fatty acids indicated high
223 amounts of saturated C14 and C16 fatty acids (Table S6). API Rapid ID 32A revealed enzyme
224 activity for most tested carbohydrates, but not for any of the tested amino acids (Table S7). No
225 respiratory quinones were detected. Additionally, analysis of peptidoglycan total hydrolysate
226 contained the following amino acids: meso-diaminopimelic acid, alanine and glutamic acid
227 (additional proteinogenic amino acids were disregarded as protein contamination). In addition
228 to the three abovementioned amino acids, the peptidoglycan partial hydrolysate contained the
229 peptides L-Ala-D-Glu and Dpm-D-Ala. The peptidoglycan type of TLL-A1 was concluded to
230 be A1y meso-Dpm-direct.
231

232 **Abundance of *S. arabinophila* TLL-Al in the feces of mice**

233 DNA extracted from fecal pellets of four healthy male mice (strain C57BL/6J) were subjected
234 to amplicon sequencing to determine the relative abundance of TLL-A1 in the original source
235 of isolation. The gut microbiota of four 24-week old mice strain C57BL/6J were variable yet
236 consisted mainly of recently re-classified *Muribaculum* (formerly known as S24-7 group),
237 averaging 36.16% ±0.042 across four mice gut (Figure 5). Other abundant genera included
238 *Faecalibaculum* (11.44% ±0.078), *Lactobacillus* (8.95% ± 0.062), *Turicibacter* (8.59%

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

239 ±0.024) and a couple of Lachnospiraceae genera (6.28% ±0.015 and 4.12% ±0.016). *S.*
240 *arabinophila* belonged to the Lachnospiraceae UCG-006 clade, which ranged in abundance
241 from 0.87 – 0.11%. In order to know the relative abundance of TLL-A1, all sequences
242 belonging to UCG006 was picked out and re-classified with custom database that included only
243 UCG006 and TLL-A1 sequences. Sequences belonging to TLL-A1 were present in all four
244 mice, yet they were not very abundant; ranging from 6 - 23% of all UCG-006 picked out.
245

246 Discussion

247 The initial aim of the study was to identify murine intestinal bacteria that have the capacity to
248 utilize D-arabinose for growth. The cultivation experiments led to the isolation of a microbial
249 isolate, named *Schaedlerella arabinophila* TLL-A1, which shared highest sequence identity
250 with *Clostridium* species ASF 502, renamed here *Schaedlerella arabinophila* ASF 502. The
251 isolation approach and characteristics of *S. arabinophila*, genomic differences between the two
252 *S. arabinophila* strains, as well as the abundance of *S. arabinophila* in the mouse intestinal
253 tract are discussed in the following.
254

255 Isolation and characteristics of *Schaedlerella arabinophila* strain TLL-A1

256 Intestinal microorganisms are known to be efficient utilizers of a wide variety of different
257 carbohydrates (18-20). The pentose arabinose is in this regard of particular interest as L-
258 arabinose is more common in nature than D-arabinose, opposed to most other monosaccharides
259 where the D-configuration is more common than the L-configuration. The use of D-arabinose
260 as sole carbon source in solid medium resulted in growth of TLL-A1 in two independent
261 experiments, with no other microbial species detected from the resulting colonies. Although
262 the total number of selected colonies may be too small to be conclusive about how selective
263 the medium is towards TLL-A1 isolation, the results indicate that either the growth conditions

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

264 are only suitable for TLL-A1, or that the diversity of D-arabinose utilizers in the murine gut is
265 low. The exact mechanism of D-arabinose utilization by TLL-A1 remains to be elucidated
266 through follow-up analysis, but it is has been shown for other organisms, such as *E. coli,* that
267 growth may occur via a L-fucose degradation pathway (21). Considering that fucose is a
268 common component of host glycans (22), together with the finding that the *S. arabinophila*
269 genomes encode for a large repertoire of CAZymes, it seems plausible that *S. arabinophila*
270 may be able to utilize a similar pathway for growth. Additional characterization of the growth
271 properties of *S. arabinophila* revealed that this species is able to utilize not only D-arabinose,
272 but also a wide variety of other carbohydrates, which may indicate that this species is a versatile
273 commensal in the murine intestine that may be able thrive on carbohydrates from different
274 dietary sources.
275 The TLL-A1 genome also revealed the presence of sporulation genes (48 annotated
276 genes related to sporulation). TLL-A1 cultures were therefore inspected for spores, but the
277 results were inconclusive: initial light microscopy and spore staining revealed a few potential
278 spores, but only in small numbers (less than five percent) of observed cells. Exact growth
279 conditions that promote sporulation, if any, are currently unknown for *S. arabinophila* DSMZ
280 106076^{T}, but would further help to confirm the capability of this strain to form spores.
281

282 **Differences between *S. arabinophila* TLL-A1 and ASF 502**

283 Comparative genome analyses of *S. arabinophila* strains TLL-A1 and ASF 502 revealed a
284 number of distinct differences between the two microbial strains. Some of these predicted
285 functional differences are apparent from the CAZyme repertoire. The genome of strain TLL-
286 A1 harbors a substantially higher number of these CAZyme-encoding genes across all four
287 different categories (CBMs, CEs, GHs, and GTs) than the stain ASF 502. However, the biggest
288 numerical difference between the two strains exists in glycosyl hydrolases. While both strains

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

289 share most of the detected glycosyl hydrolases (Figure 4), the enrichment for some of the GH
290 in either strain may represent a specific adaptation to an ecological niche. Overall, these
291 differences and those for other predicted molecular functions indicate that the strains may have
292 originated from a common ancestor a long time ago. There is currently limited information on
293 how the genetic repertoire of individual intestinal bacterial strains, such as *Bacteroides*
294 *thetaiotaomicron* (23), or entire microbial communities change over time. Considering that
295 ASF has been in use for multiple decades and that the comprising strains may have undergone
296 thousands of replications in some mouse colonies, it may seem plausible to assume that
297 substantial genetic changes may have occurred over time. It also needs be noted that strain ASF
298 502 was isolated from CD-1 mice (derived from Swiss mice) (24), while strain TLL-A1 was
299 isolated from C57BL/6J (derived from C57BL) (25). Both mouse strains (including their
300 progenitors) are genetically distinct, and have been established for over 80 years. This time
301 period may therefore also need to be taken into consideration for a potential evolutionary
302 separation of the two *Schaedlerella* strains as the genetic differences in the microbial strains
303 may represent adaptations to a specific host. Ultimately, it may not be possible to fully tease
304 out the factors that led to genomic differences of two *S. arabinophila* strains, especially
305 considering that the association of their common ancestor with host mice is likely to be older
306 than the history of mice as laboratory model itself. Isolating and characterizing additional *S.*
307 *arabinophila* strains may help to provide insights into the diversity of this clade, with the
308 resulting pangenome revealing the metabolic potential and flexibility of *S. arabinophila*. In a
309 broader context, understanding strain-specific differences in mouse-associated microorganisms
310 may also increase our understanding of how human-associated microorganisms may adapt to
311 different human individuals and different organs within them.

312 **Relative abundance of *S. arabinophila* and implications for ASF**

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

313 Amplicon sequencing was performed on DNA extracted from mouse fecal samples that served
314 as inoculum for the isolation of *S. arabinophila* TLL-A1. The amplicon data reveals that *S.*
315 *arabinophila* is only present at low relative abundance (less than one percent) in the analysed
316 samples. It needs to be noted that mice were maintained on a standard chow rich in fibre, which
317 generally favours Bacteroidetes over Firmicutes (a phylum *S. arabinophila* belongs to) (26).
318 The latter may be more abundant on other diets with higher caloric intake, such as the western
319 diet (26). It may therefore be misleading to put too much emphasis on the observed relative
320 abundance in this study, as intestinal microbial communities are able to respond quickly and
321 drastically to dietary changes (27). However, it may still be worthwhile to take these
322 observations into consideration when working with ASF.

323 ASF has served as a valuable synthetic microbial community for a long time, despite
324 some known limitations (28, 29). While ASF 502 may reach high relative abundances in ASF-
325 harbouring mice fed a standard chow (15), it is also apparent from this study that ASF 502 is
326 not a major component of conventionally-raised laboratory mouse microbiota on a
327 polysaccharide rich diet. For future studies it may therefore be necessary to enhance or alter
328 the Altered Schaedler Flora again, taking into consideration also the different experimental
329 conditions, such as dietary changes for the animals. Recent cultivations efforts have resulted in
330 the cultivation of 100 murine intestinal microbes (30). While many more mouse intestinal
331 microorganisms remain to be (re-) cultured, this collection may serve as an important tool to
332 increase the diversity and complexity of autochthonous synthetic microbial communities in
333 mice. In addition, protocols and bioinformatics pipelines, e.g. COPRO-seq (31), have been
334 developed to analyse defined synthetic microbial communities. Combining these tools may
335 help to meet the growing need to develop synthetic communities of autochthonous
336 microorganism that realistically mirror the composition of the conventional mouse microbiota
337 and that allow to study the specific interactions between hosts and microorganisms (32, 33).

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

338 In conclusion, this study provides a detailed description of *Schaedlerella arabinophila*,
339 a representative of a new bacterial genus from the intestinal tract of mice and a close relative
340 of *Clostridium* species ASF 502. The results highlight the need to further investigate intra-
341 species differences of intestinal bacteria and how host-strain specific adaptations may have
342 shaped the genomes of gut commensals.
343
344 Description of *Schaedlerella* gen. nov.
345 Schaedlerella (Schaed.ler.el.la; N.L. fem. n. named after Russel W. Schaedler, one of the
346 pioneering researchers in the study of host-microbiota interactions). The genus represent two
347 strains of a species that are closely related to *Ruminococcus* species and *Clostridium*
348 *glycyrrhizinilyticum*. Cells are uniform, rod-shaped (2-5 µm long), Gram positive and obligate
349 anaerobe. Saccharoclastic chemo-organotrophs. Peptidoglycan is A1γ meso-Dpm-direct type.
350 GC-content inferred from the genomes of strains TLL-A1 and ASF 502 is 47.9%. The type
351 species is *Schaedlerella arabinophila.*
352
353 Description of *S. arabinophila* sp. nov. *arabinophila* (a.ra.bi.no.phi.la), N.L. fem. –arabin -
354 natural rubber that has high content of arabinose, adj. philus -liking). Grows on sucrose, D-
355 cellubiose, D-maltose, D-melibiose, D-raffinose, A-lactose, D-mannitol, Glycogen (type II),
356 tannic acids, dextrin, maltodextrin, D-fructose, D-glucose, D-xylose, D-galactose, L-fucose,
357 D-ribose, meso-erythritol, N-acetyl-D-glucosamine, N-acetyl-neuramine, D-arabinose, L-
358 arabinose. No growth was observed (within 15 days) on ethanolamine, choline, starch,
359 cellulose, chitosan, pyruvate. The type strain is DSMZ 106076^T / KCTCC 15657^T.
360
361 **Materials and Methods**
362 **Isolation of *S. arabinophila***

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

363 *S. arabinophila* TLL-A1 was isolated from freshly collected feces of C57BL/6J mice.
364 Experimental procedures involving animals were performed according to the approved IACUC
365 protocol TLL-16–016. Mouse fecal material was homogenized and serially diluted prior to
366 plating. The medium composition for isolation and maintenance of *S. arabinophila* are as
367 follows (weight given in grams per litre of distilled water): Eight grams sodium acetate (Sigma-
368 Aldrich, St. Louis, Missouri, USA), 0.2g magnesium sulfate heptahydrate (Sigma-Aldrich,
369 Bangalore, India), 0.25g ammonium sulfate (Sigma-Aldrich, Tokyo, Japan), 0.69g potassium
370 phosphate dibasic (Sigma-Aldrich, St. Louis, Missouri, USA), 0.5g yeast extract (Oxoid Ltd.,
371 Basingstoke, United Kingdom), 0.5ml of 0.2% resazurin (Sigma, St. Louis, Missouri, USA),
372 7.5g D-arabinose (Alfa Aesar, Lancashire, United Kingdom), ten ml of trace mineral
373 supplement (ATCC, Manassas, Virginia, USA), and one ml of vitamin supplement (ATCC,
374 Manassas, Virginia, USA). The pH is adjusted to pH7.0 using 37% fuming hydrochloric acid
375 (Merck, Darmstadt, Germany). The prepared medium was filter sterilised and placed into the
376 anaerobic chamber. Filter-sterilized 50 mg of sodium dithionite (Sigma-Aldrich, Munich,
377 Germany) and three grams of sodium bicarbonate (Sigma-Aldrich, St. Louis, Missouri, USA)
378 were further added to the medium at least one day prior to usage. Solid medium, used for
379 streaking and picking out individual colonies to ensure pure culture, was prepared by
380 combining autoclaved 4% agar in distilled water with 500 ml of two-fold concentrated filtered
381 liquid medium. *S. arabinophila* was maintained in the above mentioned liquid medium. 0.5ml
382 of bacteria culture is inoculated into ten ml of fresh medium every two days. Cultures were
383 kept in an anaerobic Balch tube in 37°C incubator. For long-term preservation cultures were
384 stored in 20% glycerol at -80˚C.
385

386 **Growth characteristics of *S. arabinophila***

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

387 Arabinose was omitted from the aforementioned cultivation medium to investigate growth of
388 *S. arabinophila* on different carbon sources. The following carbon sources were tested: sucrose
389 (Nacalai Tesque, Kyoto, Japan), D-cellubiose (Sigma, Dorset, England, United Kingdom), D-
390 maltose monohydrate (Sigma-Aldrich, St. Louis, Missouri, USA), ethanolamine (Sigma-
391 Aldrich, St. Louis, Missouri, USA), choline chloride (Sigma-Aldrich, Shanghai, China), D-
392 melibiose (Sigma-Aldrich, Bratislava, Slovakia), D-raffinose pentahydrate (Sigma-Aldrich,
393 Shanghai, China), L-Lactic acid sodium (Sigma, Buchs, Switzerland), A-lactose monohydrate
394 (Sigma, Munich, Germany), D-mannitol (Sigma-Aldrich, Shanghai, China), glycogen Type II
395 (Sigma, Tokyo, Japan), soluble starch (Sigma-Aldrich, Munich, Germany), tannic acid (Sigma-
396 Aldrich, Shanghai, China), A-cellulose bioreagent (Sigma, St. Louis, Missouri, USA), chitosan
397 (Aldrich, Reykjavik, Iceland), dextrin 10 (Sigma, Shanghai, China), maltodextrin (Aldrich, St.
398 Louis, Missouri, USA), D-fructose (Sigma, St. Louis, Missouri, USA), D-glucose (Sigma,
399 Lyon, France), D-Xylose (Sigma-Aldrich, Shanghai, China), D-galactose (Sigma-Aldrich,
400 Munich, Germany), L-fucose (Sigma, Bratislava, Slovakia), D-ribose (Sigma, Shanghai,
401 China), pyruvic acid sodium (Sigma, Tokyo, Japan), meso-erythritol (Sigma, Shanghai,
402 China), N-acetyl-D-glucosamine (Sigma, Shanghai, China), and N-acetylneuraminic acid
403 (Sigma-Aldrich, St. Louis, Missouri, USA). Each carbon source was added at a final
404 concentration of 10mM except for polymers that were added at 5g/L
405 To prepare the *S. arabinophila* for inoculation into the various carbon source medium, two sets
406 of four ml of overnight *S. arabinophila* culture was each inoculated into 80 ml of base medium
407 in sterile glass bottles. Cultures were kept in a 37°C incubator overnight before inoculation into
408 the various carbon source medium. Three replicates of each type of medium were made. One
409 ml of two days post-inoculation culture of *S. arabinophila* growing in base medium was
410 inoculated into each of the 20 ml of various carbon source medium. Optical density was
411 visually classified every 24 hours for fifteen days. On the sixth and fifteenth day post

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

412 inoculation, Optical density was measured using Ultrospec 8000 5706 V1.0.1, at wavelength
413 600.0nm, bandwidth of 1nm, integration time of one second. For the cultures with sufficient
414 growth, gDNA of the resultant cultures were extracted. The gDNA was then sequenced to
415 ensure that the growth observed was that of *S. arabinophila*.

416 **Gram and spore stains**

417 Gram reactivity of *S. arabinophila* was investigated using Gram's crystal violet solution,
418 Gram's iodine solution, Gram's safranin solution (Sigma-Aldrich, St. Louis, Missouri, USA)
419 and a decoloriser solution of 1:1 ethanol (Fisher Scientific U.K. Limited, Leicester, United
420 Kingdom) to acetone (Merck Specialities Private Limited, Mumbai, India). Protocol described
421 by Sigma-Aldrich Gram Staining Kit was used. Spore stain was conducted on *S. arabinophila*
422 using Schaeffer & Fulton`s Spore Stain A (Malachite green 50 g/l in water) and Schaeffer &
423 Fulton`s Spore Stain B (Safranin O 5 g/l in water) (Sigma-Aldrich, Buchs, Switzerland). Spore
424 stain protocol used was as described by Sigma-Aldrich Schaeffer and Fulton Spore Stain Kit.
425 The stained cells were imaged using a widefield upright microscope. The camera used was
426 Leica microscope camera DFC7000T and cooled colour camera for LASX. The microscope
427 used is the Axioplan 2 imaging upright motorised microscope. Leica application suite X
428 3.0.1.15878 from Leica microsystems was used to capture and process the images.

429 **Characterization of cellular properties**

430 The following characterizations of *S. arabinophila* were performed at DSMZ, Braunschweig,
431 Germany: Analysis of polar lipids, analysis of respiratory quinones, analysis of the cellular
432 fatty acids, peptidoglycan structure, BioMérieux API 50 CH (bioMérieux, Craponne France),
433 BioMérieux API Rapid ID 32 A (bioMérieux, Craponne France), and identification of whole
434 cell sugars.

435 **DNA extraction and sequencing**

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

436 Liquid culture of *S. arabinophila* was grown to optical density of 1.80 at a wavelength of
437 600nm. 50 ml culture was centrifuged (Beckman Coulter rotor JA10, 6000g, 20min, 4°C) to
438 pellet the cells out of the media and subject to genomic DNA extraction (modified after (34))
439 using two rounds of phenol-chloroform purification. The extracted gDNA was stored in -20 °C
440 until further use. Ten μg of prepared genomic DNA was purified with AMPure XP magnetic
441 beads and quality checked with Nano drop and Qubit fluorometer (Invitrogen, Carlsbad, CA,
442 USA) and subsequently sent for PacBio RSII Single Molecule and Illumina MiSeq paired-end
443 (251x251bp) sequencing at Singapore Centre for Environmental Life Sciences Engineering
444 (SCELSE) in Nanyang Technological University.

445 **Genome assembly and annotation**

446 Raw PacBio sequencing reads were *de novo* assembled with SMRT analysis software (via
447 HGAP 4) with standard parameters except the estimated genome size was set as 6.5 Mb
448 according to the genome of ASF 502 (12). Reads from MiSeq paired-end sequencing were used
449 for further error correction using Pilon (17).

450 As an initial quality check step of the assembly, RNAmmer (35) was used to annotate
451 and verify the RNA genes, of which the 16S rRNA gene was further used for phylogenetic tree
452 construction. In addition, CheckM (36) was used to assess genome completeness, percentage
453 contamination, as well as finding out missing single-copy marker genes. Because the polished
454 genome consisted of three contigs of which two were short (6.35 Mb, 0.16 Mb and 9.5 kb),
455 Plasmid Finder (16) was used to assess if any were plasmids. Genomes were annotated using
456 RASTtk (37) online through PATRIC server (38).

457 **Phylogeny**

458 16S rRNA gene and whole genome phylogenies were constructed in order to compare the
459 evolutionary placement of strain TLL-A1. The 16S rRNA gene was extracted from the polished
460 genome using RNAmmer as described earlier, aligned with SINA (39) and phylogeny

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

461 constructed by RAxML within ARB (40). Whole genome phylogeny was constructed in
462 PATRIC using conserved protein sequences via RAxML. Two genomes were further compared
463 by average amino acid identity (AAI) (41) and *in silico* DNA-DNA hybridization (by genome
464 sequence-based delineation (GGDC)) (42) to determine if the two strains belong to the same
465 species. Typically, genomes are considered to belong to the same species for AAAI values
466 above 95% or over 70% DNA-DNA hybridization.

467 **Metabolic reconstruction and detection of potential pathogenic activity**

468 Further downstream analysis, including genome visualization, comparison and metabolic
469 construction, was conducted in PATRIC (38). The genome visualization was conducted using
470 circular display where genome contiguity, CDS, non-CDS features, antimicrobial resistance
471 (AMR) genes, transporters, GC content and GC skew were displayed. Further metabolic
472 potential of TLL-A1 was determined by PATRIC interface, specifically looking into
473 presence/absence of key enzymes in some instances. In brief, major central metabolism genes
474 (i.e. Embden-Meyerhof-Parnas pathway, citric acid cycle, pentose phosphate pathway, NADH-
475 dehydrogenase, cytochrome C, cytochrome C oxidase, ATP synthase genes) were checked.
476 Potential genomic islands (GI) were visualized using IslandViewer4 (43) and gene annotation
477 for major GI regions were checked.

478 **Genome Comparison**

479 The assembly was compared for similarities and differences with the previously sequenced
480 draft genome *Clostridium* sp. ASF 502. OrthoVenn (44) was used for comparisons and
481 annotation of orthologous gene clusters. Given the similarity of the two genomes, COGs that
482 were unique to each genome were further compared according to gene ontology (GO)
483 classification. Additionally, carbohydrate-active enzymes (CAZymes), important for
484 carbohydrate breakdown, were compared between the two genomes. CAZymes were annotated
485 via dbCAN2 meta server (45) and the major class of CAZy checked against the online database

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

486 (46). Note that only CAZymes identified with >2 methods (out of HMMER (E-Value < 1e-15,
487 coverage > 0.35; http://hmmer.org/), DIAMOND (E-Value < 1e-102; (47)) and Hotpep
488 (Frequency > 2.6, Hits > 6; (48)) were considered, as recommended by the authors.

489 **DNA extraction and library preparation for amplicon sequencing**

490 Samples were transferred to sterile bead-beating vials containing ~0.7g zirconium beads and
491 200µl of 20% SDS. 282µl Buffer A (see Rius *et al* for composition (49)), 268µl Buffer PM,
492 550µl phenol/chloroform/isoamyl alcohol (25:24:1, pH 8) were added. The pelleted cells were
493 subject to phenol-chloroform-based extraction, after combined mechanical and chemical lysis
494 of cells (49). The samples were quantified and diluted to 40 ng/µl per sample and stored in -20
495 °C for further use.

496 The extracted DNA was processed according to the Earth Microbiome Project
497 (http://www.earthmicrobiome.org/protocols-and-standards/16s/) as described in Thompson *et*
498 *al* [13] except a dual-indexing strategy was used as described by (50) to reduce the number of
499 barcoded primers required. PCR was conducted with QIAGEN Taq MasterMix (CAT NO
500 201445, QIAGEN, Germany) in triplicates (plus one negative control per sample) under the
501 condition: 94 °C for 3 mins; 35 cycles of 94 °C for 45 secs, 50 °C for 60 °C, 72 °C for 90 secs;
502 72 °C for 10 mins; 4 °C hold. The amplicons were quantified with Quant IT Picogreen (CAT
503 NO P7589, Thermo Fisher Scientific, USA), pooled together in equimolar concentrations and
504 sequenced at Singapore Centre for Environmental Life Sciences Engineering (SCELSE) at
505 Nanyang Technological University using Illumina MiSeq paired-end chemistry (251x251bp).
506 The output raw sequences were analysed with MOTHUR v1.39.1 (51) according to the
507 standard MiSeq protocol (52). The sequence pairs were merged, de-multiplexed and quality-
508 filtered accordingly. Furthermore, because Lachnospiraceae UCG-006 had the highest
509 sequence identity to *S. arabinophila* at 16S rRNA level, all sequences that were classified as
510 Lachnospiraceae UCG-006 were picked out and reclassified with a modified database that

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

511 contained the *S. arabinophila* 16S rRNA gene (annotated by RNAmmer from the genome
512 sequenced in this study) to determine the relative abundance of the latter. Visualization of
513 processed data was conducted by R project following tutorials provided in
514 https://joey711.github.io/phyloseq/index.html. Packages used include: ape (53), dplyr (54),
515 ggdendro (55), ggplot2 (56), ggpubr (57), gplots (58), grid (59), gridExtra (60), Heatplus (61),
516 pheatmap (6), phyloseq (62), plyr (63), RColorBrewer (64), reshape2 (65), sfsmisc (66), vegan
517 (67) and viridis (68).

518

519 **Accession numbers:**

520 The demultiplexed, pair-matched amplicon sequences are deposited in NCBI SRA under
521 accession number PRJNA498133 . The assembled genome of *Schaedlerella arabinophila*
522 TLL-A1 is deposited in GenBank under the accession number PRJNA497957 .

523

524 **Acknowledgments**

525 We thank Daniela Moses at SCELSE Singapore for performing the PacBio sequencing, and
526 Amit Anand and Muhammad Khairillah Bin Nanwi at TLL biocomputing for bioinformatics
527 support. This work was funded by Temasek Life Sciences Laboratory.

528 **References**

529 1. **Shreiner AB, Kao JY, Young VB.** 2015. The gut microbiome in health and in disease. Current
530 opinion in gastroenterology **31:**69.
531 2. **Wade WG.** 2013. The oral microbiome in health and disease. Pharmacological research
532 **69:**137-143.
533 3. **Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE,**
534 **Relman DA.** 2005. Diversity of the human intestinal microbial flora. science **308:**1635-1638.
535 4. **Lukeš J, Stensvold CR, Jirků-Pomajbíková K, Parfrey LW.** 2015. Are human intestinal
536 eukaryotes beneficial or commensals? PLoS pathogens **11:**e1005039.
537 5. **Gordon HA, Pesti L.** 1971. The gnotobiotic animal as a tool in the study of host microbial
538 relationships. Bacteriological reviews **35:**390.

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

539 6. **Biggs MB, Medlock GL, Moutinho TJ, Lees HJ, Swann JR, Kolling GL, Papin JA.** 2017.
540 Systems-level metabolism of the altered Schaedler flora, a complete gut microbiota. The
541 ISME journal **11:**426.
542 7. **Bouskra D, Brézillon C, Bérard M, Werts C, Varona R, Boneca IG, Eberl G.** 2008. Lymphoid
543 tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis.
544 Nature **456:**507.
545 8. **Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-y M, Glickman JN,**
546 **Garrett WS.** 2013. The microbial metabolites, short-chain fatty acids, regulate colonic Treg
547 cell homeostasis. Science **341:**569-573.
548 9. **Orcutt R, Gianni F, Judge R.** 1987. Development of an "Altered Schaedler Flora" for NCI
549 gnotobiotic rodents. Microecol Ther **17**.
550 10. **Schaedler RW, Dubos R, Costello R.** 1965. The development of the bacterial flora in the
551 gastrointestinal tract of mice. Journal of Experimental Medicine **122:**59-66.
552 11. **Dewhirst FE, Chien C-C, Paster BJ, Ericson RL, Orcutt RP, Schauer DB, Fox JG.** 1999.
553 Phylogeny of the defined murine microbiota: altered Schaedler flora. Applied and
554 environmental microbiology **65:**3287-3292.
555 12. **Wannemuehler MJ, Overstreet A-M, Ward DV, Phillips GJ.** 2014. Draft genome sequences
556 of the altered schaedler flora, a defined bacterial community from gnotobiotic mice.
557 Genome announcements **2:**e00287-00214.
558 13. **Robertson BR, O'Rourke JL, Neilan BA, Vandamme P, On SL, Fox JG, Lee A.** 2005.
559 Mucispirillum schaedleri gen. nov., sp. nov., a spiral-shaped bacterium colonizing the mucus
560 layer of the gastrointestinal tract of laboratory rodents. International journal of systematic
561 and evolutionary microbiology **55:**1199-1204.
562 14. **Loy A, Pfann C, Steinberger M, Hanson B, Herp S, Brugiroux S, Neto JCG, Boekschoten MV,**
563 **Schwab C, Urich T.** 2017. Lifestyle and horizontal gene transfer-mediated evolution of
564 Mucispirillum schaedleri, a core member of the murine gut microbiota. Msystems **2:**e00171-
565 00116.
566 15. **Sarma-Rupavtarm RB, Ge Z, Schauer DB, Fox JG, Polz MF.** 2004. Spatial distribution and
567 stability of the eight microbial species of the altered schaedler flora in the mouse
568 gastrointestinal tract. Applied and environmental microbiology **70:**2791-2800.
569 16. **Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller**
570 **Aarestrup F, Hasman H.** 2014. In silico detection and typing of plasmids using PlasmidFinder
571 and plasmid multilocus sequence typing. Antimicrob Agents Chemother **58:**3895-3903.
572 17. **Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,**
573 **Wortman J, Young SK, Earl AM.** 2014. Pilon: an integrated tool for comprehensive microbial
574 variant detection and genome assembly improvement. PLoS One **9:**e112963.
575 18. **El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B.** 2013. The abundance and
576 variety of carbohydrate-active enzymes in the human gut microbiota. Nature reviews
577 Microbiology **11:**497.
578 19. **Flint HJ, Scott KP, Duncan SH, Louis P, Forano E.** 2012. Microbial degradation of complex
579 carbohydrates in the gut. Gut microbes **3:**289-306.
580 20. **Martens EC, Koropatkin NM, Smith TJ, Gordon JI.** 2009. Complex glycan catabolism by the
581 human gut microbiota: the Bacteroidetes Sus-like paradigm. Journal of Biological
582 Chemistry:jbc. R109. 022848.
583 21. **LeBlanc DJ, Mortlock RP.** 1971. Metabolism of D-arabinose: a new pathway in Escherichia
584 coli. Journal of bacteriology **106:**90-96.
585 22. **Becker DJ, Lowe JB.** 2003. Fucose: biosynthesis and biological function in mammals.
586 Glycobiology **13:**41R-53R.
587 23. **Peterson DA, Planer JD, Guruge JL, Xue L, Downey-Virgin W, Goodman AL, Seedorf H,**
588 **Gordon JI.** 2015. Characterizing the interactions between a naturally-primed

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

589 immunoglobulin A and its conserved Bacteroides thetaiotaomicron species-specific epitope
590 in gnotobiotic mice. Journal of Biological Chemistry:jbc. M114. 633800.
591 24. **Chia R, Achilli F, Festing MF, Fisher EM.** 2005. The origins and uses of mouse outbred stocks.
592 Nature genetics **37:**1181.
593 25. **Morse HC.** 1978. Origins of inbred mice. Academic Press.
594 26. **Jumpertz R, Le DS, Turnbaugh PJ, Trinidad C, Bogardus C, Gordon JI, Krakoff J.** 2011.
595 Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient
596 absorption in humans–. The American journal of clinical nutrition **94:**58-65.
597 27. **David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin**
598 **AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ.** 2014. Diet rapidly and
599 reproducibly alters the human gut microbiome. Nature **505:**559.
600 28. **Norin E, Midtvedt T.** 2010. Intestinal microflora functions in laboratory mice claimed to
601 harbor a "normal" intestinal microflora. Is the SPF concept running out of date? Anaerobe
602 **16:**311-313.
603 29. **Wymore Brand M, Wannemuehler MJ, Phillips GJ, Proctor A, Overstreet A-M, Jergens AE,**
604 **Orcutt RP, Fox JG.** 2015. The altered Schaedler flora: continued applications of a defined
605 murine microbial community. ILAR journal **56:**169-178.
606 30. **Lagkouvardos I, Pukall R, Abt B, Foesel BU, Meier-Kolthoff JP, Kumar N, Bresciani A,**
607 **Martínez I, Just S, Ziegler C, Brugiroux S, Garzetti D, Wenning M, Bui TPN, Wang J,**
608 **Hugenholtz F, Plugge CM, Peterson DA, Hornef MW, Baines JF, Smidt H, Walter J,**
609 **Kristiansen K, Nielsen HB, Haller D, Overmann J, Stecher B, Clavel T.** 2016. The Mouse
610 Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and
611 functional potential of the gut microbiota. Nature microbiology **1:**16131.
612 31. **McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, Henrissat B,**
613 **Oozeer R, Cools-Portier S, Gobert G, Chervaux C, Knights D, Lozupone CA, Knight R, Duncan**
614 **AE, Bain JR, Muehlbauer MJ, Newgard CB, Heath AC, Gordon JI.** 2011. The impact of a
615 consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and
616 monozygotic twins. Science translational medicine **3:**106ra106-106ra106.
617 32. **Chung H, Pamp SJ, Hill JA, Surana NK, Edelman SM, Troy EB, Reading NC, Villablanca EJ,**
618 **Wang S, Mora JR, Umesaki Y, Mathis D, Benoist C, Relman DA, Kasper DL.** 2012. Gut
619 immune maturation depends on colonization with a host-specific microbiota. Cell **149:**1578-
620 1593.
621 33. **Seedorf H, Griffin NW, Ridaura VK, Reyes A, Cheng J, Rey FE, Smith MI, Simon GM,**
622 **Scheffrahn RH, Woebken D.** 2014. Bacteria from diverse habitats colonize and compete in
623 the mouse gut. Cell **159:**253-266.
624 34. **Sambrook J, Fritsch EF, Maniatis T.** 1989. Molecular cloning: a laboratory manual. Cold
625 spring harbor laboratory press.
626 35. **Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW.** 2007. RNAmmer:
627 consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research **35:**3100-
628 3108.
629 36. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW.** 2015. CheckM: assessing
630 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
631 Genome research:gr. 186072.186114.
632 37. **Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B,**
633 **Pusch GD, Shukla M, Thomason III JA, Stevens R, Vonstein V, Wattam AR, Xia F.** 2015.
634 RASTtk: a modular and extensible implementation of the RAST algorithm for building custom
635 annotation pipelines and annotating batches of genomes. Scientific reports **5:**8365.
636 38. **Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T,**
637 **Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ,**
638 **Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V,**
639 **Warren A, Xia F, Yoo H, Stevens RL.** 2017. Improvements to PATRIC, the all-bacterial

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

640 Bioinformatics Database and Analysis Resource Center. Nucleic Acids Research **45:**D535-
641 D542.
642 39. **Pruesse E, Peplies J, Glöckner FO.** 2012. SINA: accurate high-throughput multiple sequence
643 alignment of ribosomal RNA genes. Bioinformatics **28:**1823-1829.
644 40. **Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi**
645 **S, Jobb G, Foerster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S,**
646 **Jost R, König A, Liss T, Lüßmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A,**
647 **Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH.** 2004. ARB: a software
648 environment for sequence data. Nucleic acids research **32:**1363-1371.
649 41. **Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM.** 2007. DNA–
650 DNA hybridization values and their relationship to whole-genome sequence similarities.
651 International journal of systematic and evolutionary microbiology **57:**81-91.
652 42. **Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M.** 2013. Genome sequence-based species
653 delimitation with confidence intervals and improved distance functions. BMC bioinformatics
654 **14:**60.
655 43. **Bertelli C, Laird MR, Williams KP, Group SFURC, Lau BY, Hoad G, Winsor GL, Brinkman FS.**
656 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets.
657 Nucleic acids research **45:**W30-W35.
658 44. **Wang Y, Coleman-Derr D, Chen G, Gu YQ.** 2015. OrthoVenn: a web server for genome wide
659 comparison and annotation of orthologous clusters across multiple species. Nucleic Acids
660 Research **43:**W78-W84.
661 45. **Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y.** 2018. dbCAN2: a
662 meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research
663 **46:**W95-W101.
664 46. **Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B.** 2014. The
665 carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research **42:**D490-
666 D495.
667 47. **Buchfink B, Xie C, Huson DH.** 2014. Fast and sensitive protein alignment using DIAMOND.
668 Nature methods **12:**59.
669 48. **Busk PK, Pilgaard B, Lezyk MJ, Meyer AS, Lange L.** 2017. Homology to peptide pattern for
670 annotation of carbohydrate-active enzymes and prediction of function. BMC bioinformatics
671 **18:**214.
672 49. **Rius A, Kittelmann S, Macdonald K, Waghorn G, Janssen P, Sikkema E.** 2012. Nitrogen
673 metabolism and rumen microbial enumeration in lactating cows with divergent residual feed
674 intake fed high-digestibility pasture. Journal of dairy science **95:**5024-5034.
675 50. **Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, Ravel J.** 2014. An improved
676 dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq
677 platform. Microbiome **2:**6.
678 51. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,**
679 **Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.**
680 2009. Introducing mothur: open-source, platform-independent, community-supported
681 software for describing and comparing microbial communities. Applied and environmental
682 microbiology **75:**7537-7541.
683 52. **Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD.** 2013. Development of a dual-
684 index sequencing strategy and curation pipeline for analyzing amplicon sequence data on
685 the MiSeq Illumina sequencing platform. Applied and environmental microbiology**:**AEM.
686 01043-01013.
687 53. **Paradis E, Schliep K, Schwartz R.** 2018. ape 5.0: an environment for modern phylogenetics
688 and evolutionary analyses in R. Bioinformatics **1:**3.
689 54. **Wickham H, Francois R, Henry L, Müller K.** 2015. dplyr: A grammar of data manipulation. R
690 package version 04 **3**.

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

691 55. **de Vries A, Ripley B.** 2013. Ggdendro: tools for extracting dendrogram and tree diagram plot
692 data for use with ggplot. R package version 01-12, Retrieved from http://CRAN R-project
693 org/package= ggdendro.
694 56. **Wickham H.** 2016. ggplot2: elegant graphics for data analysis. Springer.
695 57. **Kassambara A.** 2017. ggpubr:"ggplot2" based publication ready plots. R package version 01
696 **6**.
697 58. **Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler**
698 **M, Magnusson A, Moeller S.** 2009. gplots: Various R programming tools for plotting data. R
699 package version **2:**1.
700 59. **Murrell P.** 2016. R graphics. CRC Press.
701 60. **Auguie B.** 2012. gridExtra: functions in Grid graphics. R package version 09 **1**.
702 61. **Ploner A.** 2012. Heatplus: Heatmaps with row and/or column covariates and colored
703 clusters. R package version **2**.
704 62. **McMurdie PJ, Holmes S.** 2013. phyloseq: an R package for reproducible interactive analysis
705 and graphics of microbiome census data. PloS one **8:**e61217.
706 63. **Wickham H.** 2009. plyr: Tools for splitting, applying and combining data. R package version
707 01 **9:**651.
708 64. **Neuwirth E.** 2011. Package 'RColorBrewer'. CRAN 2011-06-17 08: 34: 00. Apache License
709 2.0,
710 65. **Wickham H.** 2012. reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package. R
711 Package Version 1.
712 66. **Maechler M.** 2012. sfsmisc: Utilities from seminar fuer Statistik ETH Zurich. R package
713 version**:**1.0-20.
714 67. **Oksanen J, Blanchet FG, Kindt R, Legendre P, O'hara R, Simpson GL, Solymos P, Stevens**
715 **MHH, Wagner H.** 2010. vegan: Community Ecology Package. R package version 1.17-2.
716 http://cran r-project org> Acesso em **23:**2010.
717 68. **Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, Perry T, Kao D, Mason AL,**
718 **Madsen KL, Wong GK-S.** 2016. Characterization of the gut microbiome using 16S or shotgun
719 metagenomics. Frontiers in microbiology **7:**459.

720

721

722

723

724

725

726

727

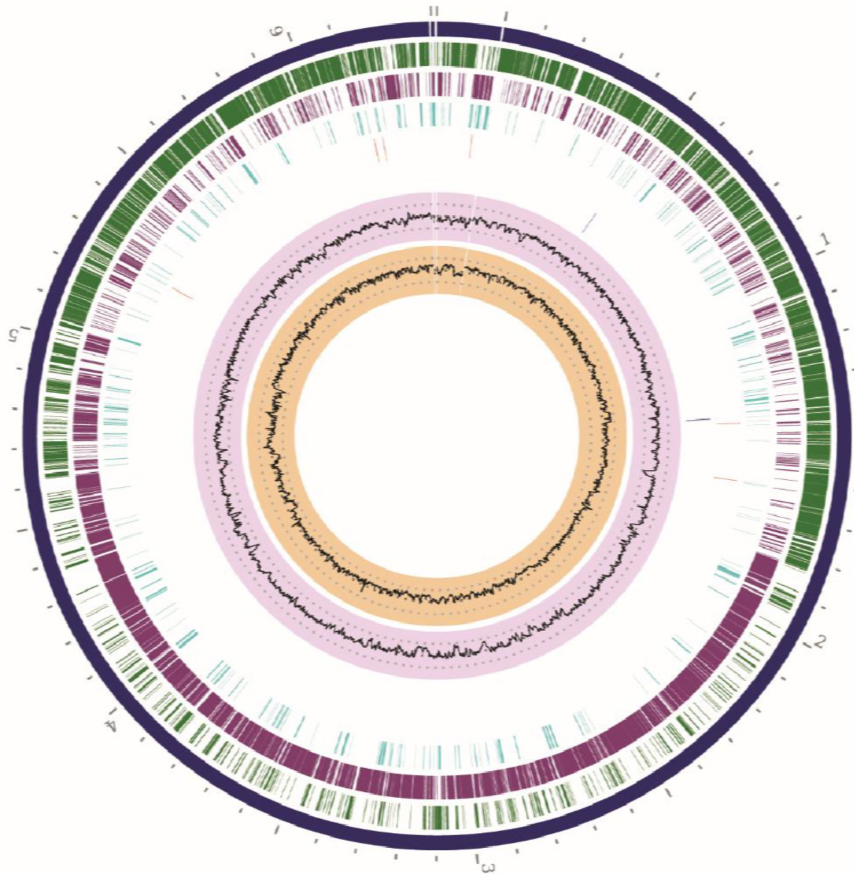
728

729

730

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

731 **Figures**



732

733 **Figure 1.** Circular display of *Schaedlerella arabinophila* strain TLL-A1 genome displaying
734 (from outside to in) contiguity, CDS on forward and reverse reads, non-CDS features, AMR
735 genes, transporters as well as GC content and GC skew.

736

737

738

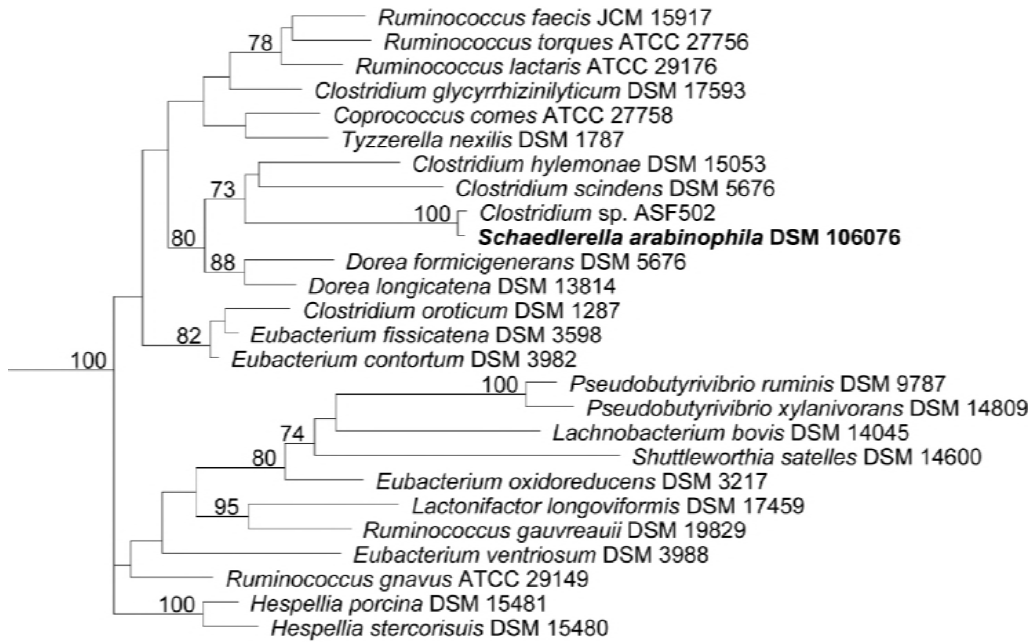
739

740

741

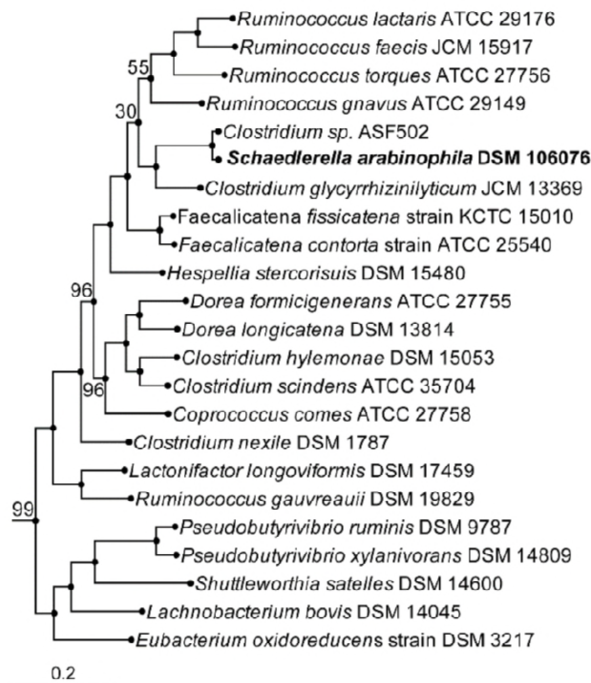
bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

742 A



743

744 B



745

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

746 **Figure 2. Phylogenetic trees showing the relationship of *Schaedlerella arabinophila* to**
747 **closest related bacteria.** A) RaxML 16S rRNA gene phylogeny of *Schaedlerella arabinophila*
748 strain TLL-A1 (in bold) and related organisms. The sequences were SINA-aligned (Pruesse et
749 al., 2012) and the tree was constructed in ARB. B) Whole-genome phylogeny of *Schaedlerella*
750 *arabinophila* and related genomes. The tree was constructed in PATRIC with RaxML
751 algorithm.

752

753

754

755

756

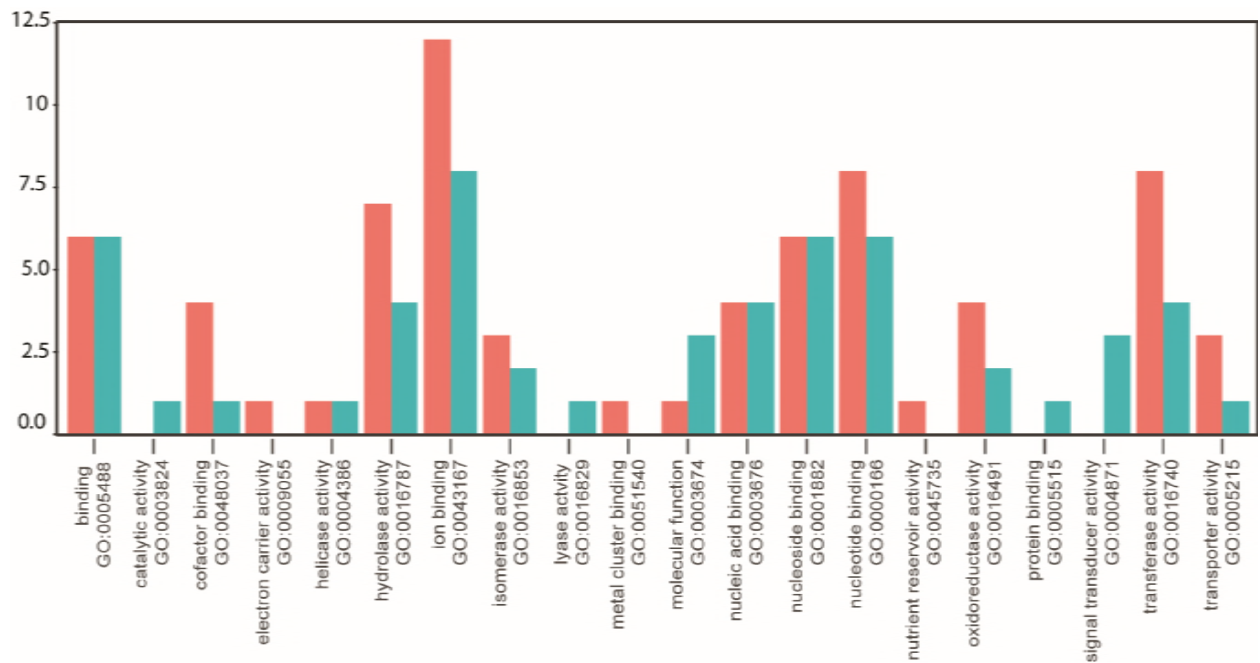
757

758

759

760

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

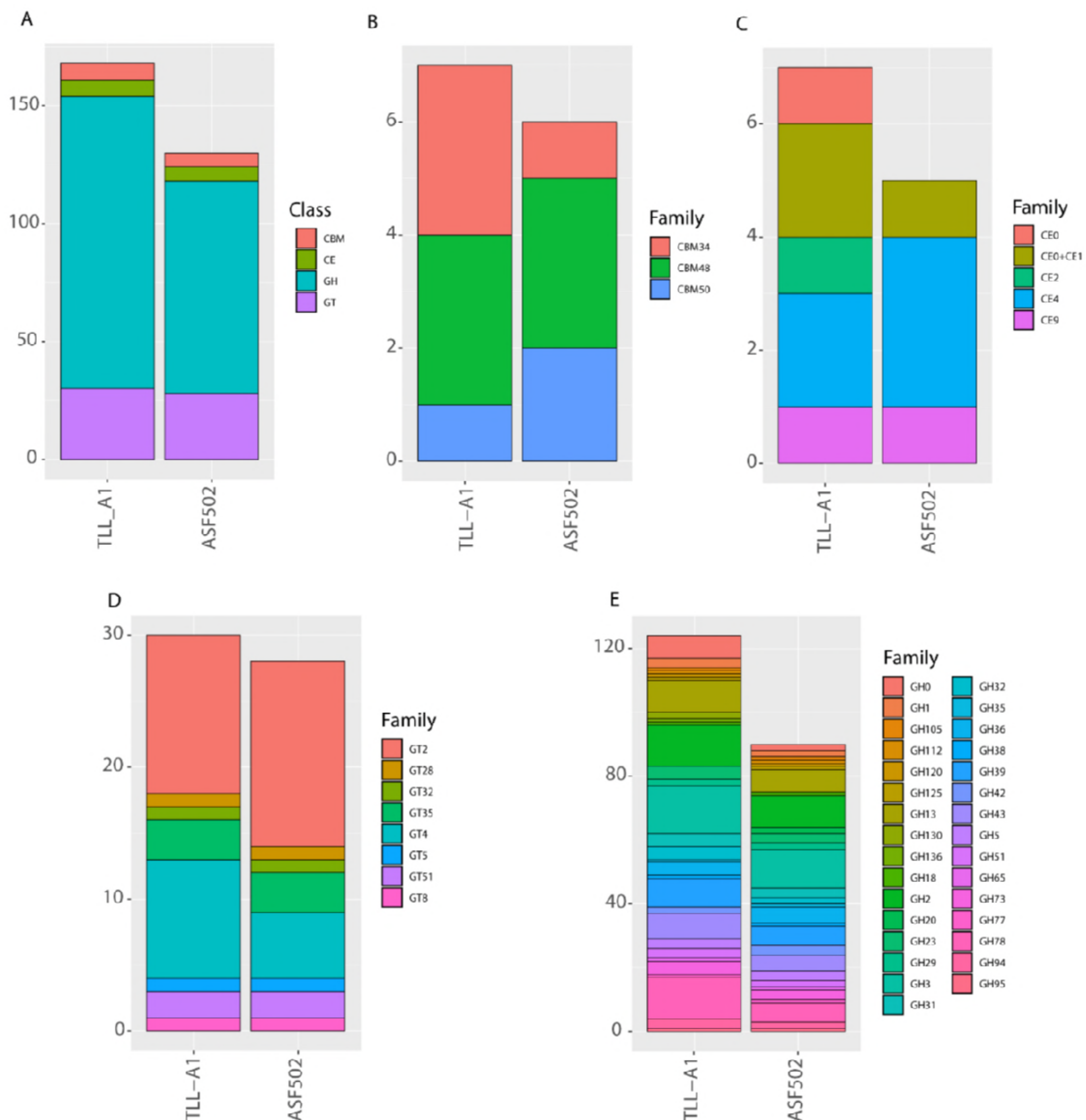


761

762 **Figure 3. Differences in predicted molecular function between *Schaedlerella arabinophila***
763 **strains TLL-A1 and ASF 502.** Shown are counts of Swiss-Prot annotations of unique COGs
764 in *Schaedlerella arabinophila* TLL A1 (red) and ASF 502 (green) classified into Gene
765 Ontology (GO) terms according to molecular function.

766

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.



767

768 **Figure 4. Comparative analysis of carbohydrate Active EnZymes (CAZymes) identifies**
769 **differences between *Schaedlerella arabinophila* strain TLL-A1 and ASF 502 in the**
770 **predicted capability to process carbohydrates.** A) Overall comparison of CAZyme category
771 counts between strain *S. arabinophila* strain TLL-A1 and ASF 502. B) Family level
772 classification of CBMs in *Schaedlerella arabinophila* and ASF 502. C) Family level
773 classification of CEs in *Schaedlerella arabinophila* and ASF 502. D) Family level
774 classification of GTs in *Schaedlerella arabinophila* and ASF 502. E) Family level classification
775 of GHs in *Schaedlerella arabinophila* and ASF 502. Only CAZymes identified with >2

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.

776 methods (out of HMMER (E-Value < 1e-15, coverage > 0.35; http://hmmer.org/), DIAMOND
777 (E-Value < 1e-102; (47)) and Hotpep (Frequency > 2.6, Hits > 6; (48)) were considered, as
778 recommended by the authors.

779

780

781

782

783

784

785

786

787

788

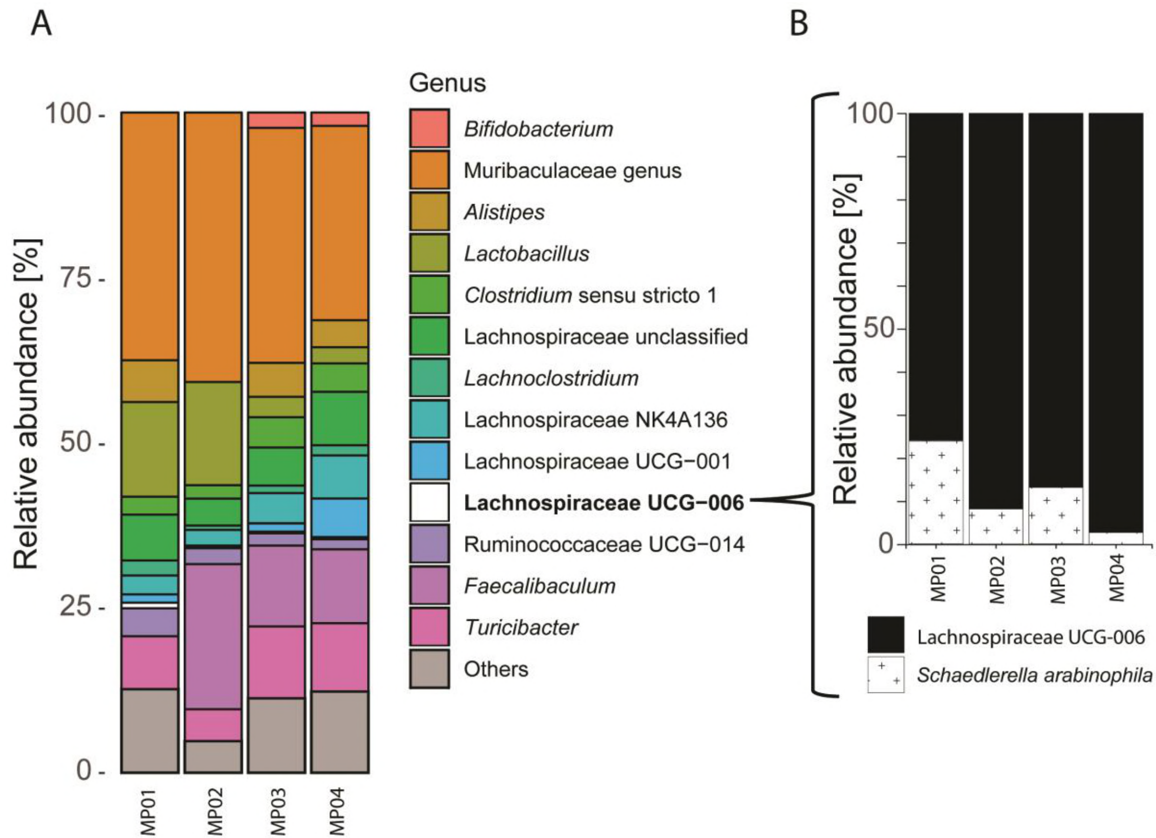
789

790

791

792

bioRxiv preprint doi: https://doi.org/10.1101/496661; this version posted December 17, 2018. 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.



793

794 **Figure 5. Analysis of relative abundance of *S. arabinophila* in the fecal microbiota of**
795 **conventionally-raised C57BL/6J mice.** A) Genus-level taxonomy composition of four
796 healthy mice fecal microbiota (strain C57BL/6J). Genera that have a relative abundance of less
797 than 1% within the sample are classified as others. B) Relative abundance of *Schaedlerella*
798 *arabinophila* within Lachnospiraceae UCG-006, closest match in terms of 16S rRNA gene
799 from Silva database. Fecal samples from same mice were used as inoculum for isolation of *S.*
800 *arabinophila*

801