TITLE: Functional metagenomic discovery of novel tetracycline and acarbose resistance genes from low biomass samples using METa assembly

Authors: Allman HM[1], Bernate EP[1], Franck E[1], Oliaro FJ[2], Hartmann EM[3], Crofts TS[1]

1. Department of Biomedical Sciences, Florida State University, Tallahassee, FL, USA
2. Conservation Research Department, John G. Shedd Aquarium, Chicago, IL, USA
3. Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL, USA

Corresponding author: Terence S. Crofts, tsc22a@fsu.edu

ABSTRACT

A significant challenge in the field of microbiology is the functional annotation of sequence novel genes from microbiomes. The increasing pace of sequencing technology development has made solving this challenge in a high-throughput manner even more important. Functional metagenomics offer a sequence-naïve and cultivation-independent solution. This forward genetics approach relies on the creation of functional metagenomic libraries (aka shotgun cloning) in which a microbial host such as *E. coli* is transformed with vectors containing metagenomic DNA fragments and, optimally, expresses any captured genes into a corresponding phenotype. These libraries can be screened or selected for a function of interest, such as antibiotic resistance, allowing the captured metagenomic DNA to be linked to a phenotype regardless of the sequence's novelty. Unfortunately, most methods for constructing functional metagenomic libraries require large input masses of metagenomic DNA, putting many sample types off limits to this toolset. Here, we show that our recently developed functional

metagenomic library preparation method, METa assembly, can be used to prepare useful libraries from much lower input DNA masses. Standard methods of functional metagenomic library preparation generally call for 5 μg to 60 μg of input metagenomic DNA. Here, we demonstrate that the threshold for input DNA mass can be lowered at least to 30.5 ng, a three-log decrease from prior art. These functional metagenomic libraries, prepared using between 30.5 ng and 100 ng of metagenomic DNA, nonetheless were sufficient to link three MFS efflux pumps to tetracycline resistance and capture two potential genes for degradation or resistance to the antidiabetic pharmaceutical acarbose. Our preparation of functional metagenomic libraries from aquatic samples and a model fecal swab demonstrate that METa assembly can be used to prepare functional metagenomic libraries from microbiomes that were previously incompatible with this approach. Functional metagenomic screens and selections are one of the few high-throughput methods that can link novel genes to functions and here we show that one of their significant drawbacks, a requirement for large amounts of metagenomic DNA, can now be overcome.

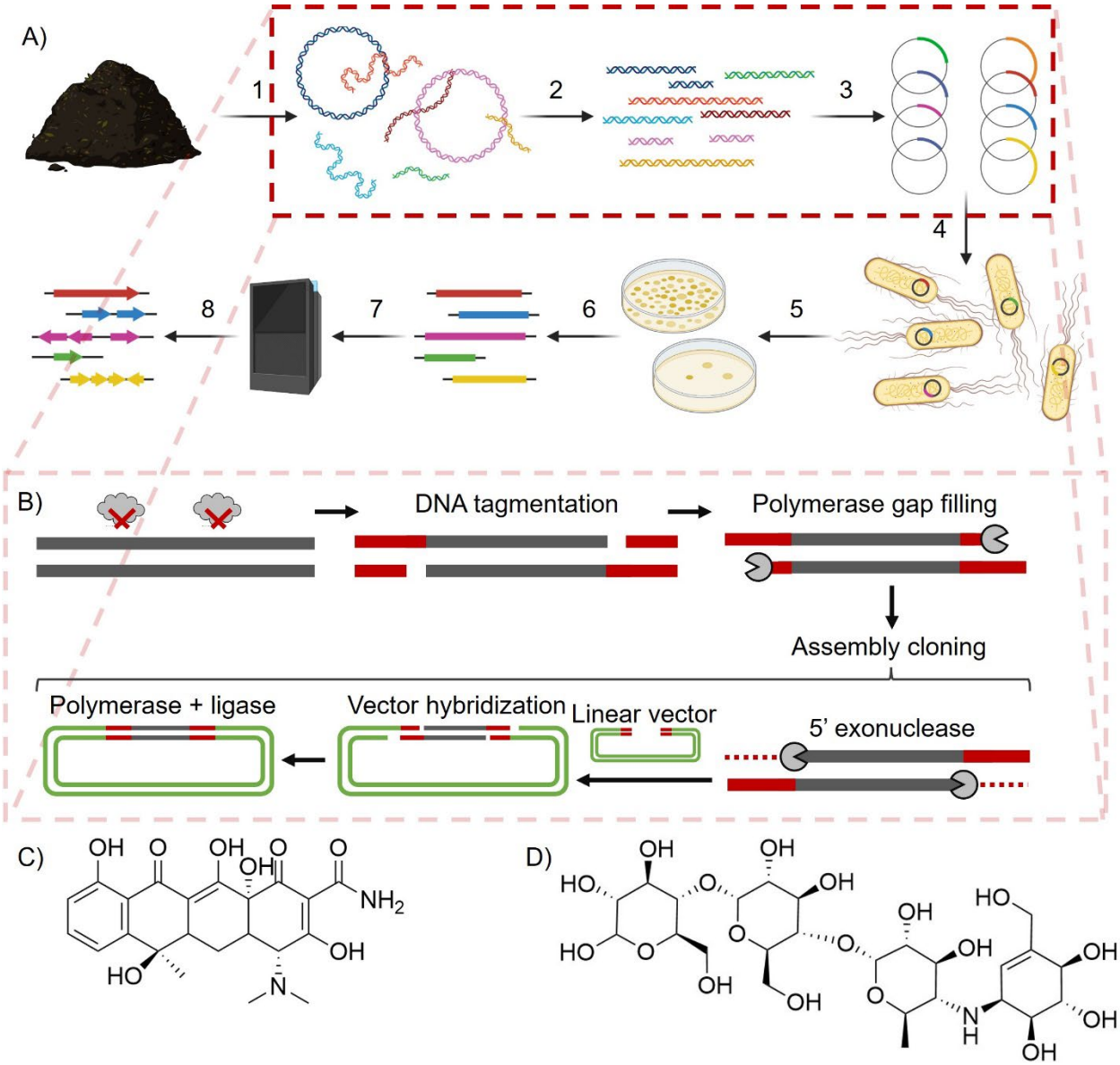

Figure 1. Functional metagenomics and METa assembly. A) General steps in the preparation and application of a functional metagenomic library: 1) Extraction of metagenomic DNA from a source microbiome. 2) Fragmentation of metagenomic DNA to preferred size range. 3) Packaging of inserts into expression vectors. 4) Transformation of host cells with vector library. 5) Screen or selection of functional metagenomic library for a phenotype of interest. 6) Collection of screened or selected metagenomic fragments. 7) Sequencing of selected inserts. 8) Open reading frame calling and annotation to identify potential genes underlying phenotypes of interest. B) Steps 2) and 3) above are modified in Mosaic ends tagmentation (METa) assembly. Fragmentation is achieved using Tn5 transposase tagmentation with mosaic end sequence oligos. Tagmented DNA is gap filled by polymerase and cloned into an expression vector with matching mosaic end sequences defining the cloning site using assembly cloning (*e.g.,* Gibson assembly). C) and D) The chemical structures of the antibiotic C) tetracycline and the diabetes pharmaceutical D) acarbose.

40 Microbiomes, assemblages of bacteria, fungi, viruses and other microorganisms, are
41 host to some of the highest levels of genetic diversity on the planet. Due to the difficulty in
42 culturing all but a select low percentage of organisms in microbiomes (often estimated to be
43 ~0.1% to 1%) current methods for studying microbiomes rely on DNA extracted from these
44 communities, metagenomic DNA[1–4]. High-throughput sequencing of metagenomic DNA provides
45 unrivaled insights into the taxonomic structure of a microbial community and can shed
46 significant light on its functional capacity as well[3]. However, even with the sequencing
47 revolution, the genetic diversity of microbiomes has outpaced our ability to accurately annotate
48 novel microbial genes. Complementing homology and structure-based methods for assigning
49 function to novel genes is the high-throughput, sequence- and cultivation-naïve method of
50 functional metagenomics. In this method, functional metagenomic libraries, sometimes referred
51 to as shotgun cloning libraries, are prepared by fragmenting metagenomic DNA and cloning
52 those fragments *en masse* into an expression microorganism. Once a library is prepared, the
53 theoretical goal is that any genes encoded within captured metagenomic DNA fragments are
54 expressed by the host cell to give a phenotype that can be selected or screened for[4–7] (**Figure**
55 **1A**).

56 While functional metagenomics is a powerful method for linking novel genes to functions
57 without requiring the growth of non-model organisms in the lab, it has its own shortcomings.
58 These issues include the potential for gene toxicity, limited expression of foreign genes in model
59 library hosts such as *E. coli*, limited coverage of target metagenomes, and requirements for
60 large quantities of metagenomic DNA for preparation of the functional metagenomic library itself.
61 We recently reported the development of a new method for preparing small insert functional
62 metagenomic libraries (*i.e.,* 2 kb to 5 kb sized inserts) that we termed METa assembly[8]. This
63 functional metagenomic library preparation method relies on near-random tagmentation to
64 fragment input metagenomic DNA and an assembly-based cloning step to replace blunt ligation

65 (**Figure 1B**). We previously demonstrated that this method circumvents one shortcoming of
66 functional metagenomics, limited metagenomic coverage, by showing that METa assembly
67 libraries are more efficient at capturing DNA and can produce approximately 80-fold larger
68 libraries per input DNA mass. We also showed that METa assembly could be used to prepare
69 functional metagenomic libraries encoding tens of gigabase pairs (Gb) of captured DNA using
70 sub-µg inputs of metagenomic DNA, far lower than the standard 5 µg to 20 µg input[8,9].

71 Here we build on our earlier results by determining if METa assembly can be used to
72 prepare functional metagenomic libraries with significantly lower quantities of input DNA. We
73 found that we could consistently prepare functional metagenomic libraries starting with 50 ng or
74 less input DNA or even DNA extracted from a swab dipped into a standardized human fecal
75 sample. These libraries were large enough that when selected for resistance to the antibiotic
76 tetracycline and the anti-diabetic acarbose they yielded novel predicted proteins. These results
77 suggest that many more rare or low biomass microbiomes can be studied by functional
78 metagenomics than previously appreciated.

79

80 **METHODS**

81

82 **Strains and materials**

83 Routine cultivation of *E. coli* was performed at 37°C with shaking at 250 rpm for aeration
84 in the case of liquid cultures. Cultivation for cloning and starter cultures was performed using
85 lysogeny broth (Miller) (LB) with the addition of kanamycin at 50 µg/ml to maintain plasmids.
86 Antimicrobial susceptibility testing was performed using Cation-Adjusted Mueller-Hinton media
87 (Teknova, 101320-364). All *E. coli* strains used in this study are of the DH10B lineage and carry
88 plasmids derived from pZE21[8,10,11]. *E. coli* strains were maintained at -80°C in 50% glycerol in

89 LB. Transposase enzymes and mosaic ends oligos were previously prepared and stored at -
90 20°C. Standard chemicals used in this study were of high purity for molecular biology and
91 microbiology. Kanamycin sulfate (VWR Life Science, 75856-68), acarbose (Thermo Scientific
92 Chemical, 45908-0010), tetracycline (Aldon Corp, TT0070-5GR), ciprofloxacin hydrochloride
93 (Corning, 61-277-RG), azithromycin dihydrate (TCI, A2076), and florfenicol (Fisher Scientific,
94 F08115G) were stored at 4°C as powders. Working solutions were stored at -20°C as filtered
95 aqueous solutions at 50 mg/ml concentration (kanamycin and acarbose), 5 mg/ml in ethanol
96 (tetracycline), 10 mg/ml in alkaline water (ciprofloxacin), or 50 mg/ml in DMSO (azithromycin
97 and florfenicol).

98

99 **Preparation of 50 ng test functional metagenomic libraries**

100 To determine if agarose extraction could be used to purify and size-select DNA following
101 low mass tagmentation reactions, we used a DNA ladder (ThermoFisher Scientific 1 kb plus,
102 10787018) to mimic fragmented DNA. Mock tagmentation reactions were made with a final DNA
103 concentration of 50 ng/µl by combining 1 µl of diluted ladder DNA with 3 µl of autoclaved milliQ
104 water and 1 µl of 5X Taps tagmentation buffer (50 mM TAPS pH 8.5, 25 mM $MgCl_2$, 50% v/v
105 DMF in water)[12]. We added 0.25 µl of 1% sodium dodecyl sulfate (SDS) to the 5 µl mock
106 reactions to simulate reaction quenching followed by 1 µl of 6x DNA dye (15% Ficoll 400 w/v,
107 0.25% Orange G w/v in water) before loading on to a 0.7% agarose gel in TAE buffer containing
108 SybrSafe dye (ThermoFisher Scientific, S33102). DNA bands were visualized over blue light
109 and a clean razor was used to cut out gel portions containing bands greater than 2 kb in length.
110 DNA was extracted from the agarose gel fragments using a gel extraction kit (New England
111 Biolabs, T1020L) following manufacturer instructions with modification. Each agarose fragment
112 was weighed and incubated with four volumes of gel dissolving buffer for at least 10 minutes at
113 55°C. The dissolved agarose was then applied to the kit's spin column, eluted, then re-applied a

114 second time. The standard protocol was followed until the elution step where each column was
115 eluted twice sequentially with 6 µl of 55°C autoclaved milliQ water. DNA concentration was
116 measured using the Quant-It DNA quantification system (Invitrogen).

117 To test the preparation of functional metagenomic libraries using 50 ng of input
118 metagenomic DNA, we used a sample of previously extracted metagenomic DNA from a
119 Canada goose fecal pellet that was stored at -20°C[8]. For library preparation, we followed our
120 previously published METa assembly protocol[8]. Briefly, triplicate 5 µl tagmentation reactions
121 containing 50 ng of metagenomic DNA, 1x TAPS-DMF buffer, and 25 ng of in-house
122 transposase[12] loaded with mosaic end sequence oligos were incubated at 55°C for 7 minutes
123 followed by quenching by addition of SDS to reach 0.05% final concentration and incubation for
124 5 more minutes at 55°C. To each reaction, 2 µl of 6x loading dye was added and fragmented
125 metagenomic DNA in the range of 2 kb to 9 kb was extracted by gel excision as described
126 above. The 12 µl of size-selected metagenomic DNA fragments were gap-filled by the addition
127 of 12 µl of 2x Q5 DNA polymerase (New England Biolabs, M0492S) (pre-incubated at 98°C for
128 30 seconds) and held at 72°C for 15 minutes before purification by PCR and DNA cleanup kit
129 (New England Biolabs, T1030S) according to manufacturer instructions with the following
130 modifications: DNA dissolved in two volumes of binding buffer was applied to the spin column
131 and the flow-through was applied a second time. At elution, we used two sequential applications
132 of 6 µl of 55°C autoclaved milliQ water. Metagenomic DNA inserts were cloned into our plasmid
133 pZE21-ME using 2x NEBuilder HiFi assembly mix (New England Biolabs, E2621L) according to
134 manufacturer instructions (assuming an insert DNA mass of 15 ng and an average fragment
135 size of 2 kb). The triplicate libraries were purified by PCR and DNA cleanup kit and the full 12 µl
136 of purified library DNA was introduced into 25 µl of *E. coli* DH10B electrocompetent cells (New
137 England Biolabs, C3020K) by electroporation at 1.8 kV in 0.1 mm cuvettes. Actual voltage
138 reached and the electroporation time constant were recorded for all electroporation reactions to

139 monitor for failed or low efficiency transformations. All three libraries were rescued post-
140 transformation by inoculation into 1 ml of SOC outgrowth medium (New England Biolabs,
141 B9020S) pre-warmed to 37°C and incubated with aeration at 37°C for 1 hour.

142 Following recovery, 100 µl of 100-fold, 10,000-fold, and 1,000,000-fold diluted cultures
143 were plated on to LB agar with kanamycin and incubated overnight at 37°C to estimate the
144 number of unique clones resulting from each transformation. The remaining cells (~1 ml) were
145 inoculated into 50 ml of LB with kanamycin and incubated at room temperature overnight with
146 aeration to amplify the functional metagenomic library. Once cultures reached a 1 cm pathlength
147 optical density at 600 nm (OD_{600}) of between 0.6 and 1.0, cells were collected by centrifugation
148 at 4,000 rcf for 7 minutes at 4°C. Pellets were resuspended in 10 ml of LB containing kanamycin
149 and 15% glycerol and stored at -80°C as 1 ml aliquots. The triplicate functional metagenomic
150 library sizes were determined using the Dantas lab calculator (http://
151 dantaslab.wustl.edu/LibSizeCalc/) considering unique cell count, determined above, average
152 metagenomic insert size, and the proportion of clones containing an empty plasmid. This
153 proportion and the average insert size were determined by performing colony PCR as
154 previously described on 9 clones from each replicate library[8].

155

156 **Preparation of an aquatic functional metagenomic library from a Shedd Aquarium**
157 **microbiome**

158 Preparation of a functional metagenomic library from aquatic metagenomic DNA began
159 with a previously extracted metagenomic DNA sample. The initial sample was collected from the
160 Shedd Aquarium in Chicago in the Underwater Beauty exhibit in May, 2019 and metagenomic
161 DNA was extracted[13]. A functional metagenomic library was prepared from 50 ng of DNA as
162 described above and evaluated by colony counting and colony PCR (n=14 colonies) to

163 determine library size. After preparation of the functional metagenomic library, quantification of
164 the input DNA by Quant-It revealed that only 30.5 ng of metagenomic DNA had been used as
165 input for this library and this corrected DNA mass was used in determining library preparation
166 efficiency.

167

168 **Functional metagenomic selection for antibiotic resistance**

169 The aquatic functional metagenomic library was selected for resistance to the antibiotics
170 tetracycline, ciprofloxacin, and florfenicol. Cation-Adjusted Mueller-Hinton agar plates containing
171 50 µg/ml kanamycin were prepared to contain each antibiotic at the following concentrations
172 previously established to be inhibitory to *E. coli*[14]: 8 µg/ml tetracycline, 0.5 µg/ml ciprofloxacin,
173 and 8 µg/ml florfenicol. An aliquot of the functional metagenomic library was thawed from -80°C
174 storage and a volume of suspension calculated to contain approximately 10-fold more total cells
175 than the unique cell count was plated on each antibiotic and incubated overnight at 37°C. A
176 parallel series of agar plates were plated with a similar cell count of *E. coli* cells containing only
177 empty vectors as a control.

178 Following incubation, 15 colonies each from the tetracycline and florfenicol selections
179 were collected and re-streaked onto LB agar plates containing kanamycin and either
180 tetracycline or florfenicol to obtain pure cultures of resistant strains. Plasmids were extracted
181 from each strain by miniprep kit (New England Biolabs, T1010L) and whole plasmid sequencing
182 was performed by Plasmidsaurus using Oxford Nanopore Technology with custom analysis and
183 annotation. The metagenomic inserts were identified by the presence of flanking mosaic end
184 sequences (5'- agatgtgtataagagacag-3') and clustered on the Clustal Omega webserver[15] to de-
185 replicate metagenomic fragments that were captured in the selection more than once. The
186 metagenomic inserts targeted for further study (TET1, TET3, TET13) were analyzed by the

187 MetaGeneMark program[16] to identify potential open reading frames which were subjected to
188 analysis by Basic Local Alignment Search Tool (BLAST)[17,18] against the NCBI non-redundant
189 database and The Comprehensive Antibiotic Resistance Database (CARD)[19,20]. The annotated
190 metagenomic DNA fragments were made into schematics using ApE plasmid editor[21] and
191 Microsoft PowerPoint.

192

193 **Preparation of a functional metagenomic library from a simulated fecal swab**

194 An aliquot of the ZymoBIOMICS fecal reference standard (Zymo Research, D6323) was
195 used as a model source for a fecal microbiome swab functional metagenomic library. An aliquot
196 of fecal standard was thawed from -80°C storage and a sterile swab was dipped into the tube
197 and gently wrung out against the tube's side as it was removed. Metagenomic DNA was
198 extracted from the swab using a PowerSoil Pro kit (Qiagen, 47014). The swab was agitated in
199 800 µl of C1 solution to release material and DNA extraction was performed following
200 manufacturer instructions with the eluted DNA reduced in volume using a vacuum concentrator.
201 A functional metagenomic library was prepared by METa assembly as above, with the
202 tagmentation reaction scaled to match the 100 ng of metagenomic DNA available from the
203 extraction. Library size was evaluated by colony count and colony PCR as above (n=14
204 colonies).

205

206 **Functional metagenomic selection for acarbose modifying genes**

207 To determine conditions appropriate for a functional metagenomic selection for acarbose
208 modifying genes, we performed an agar dilution assay on minimal media plates containing
209 maltose as the sole carbon and energy source with acarbose concentrations ranging from 0.5
210 µg/ml to 256 µg/ml in two-fold jumps[22]. The agar contained 0.4% w/v maltose in a modified

211 MDG minimal medium[23] consisting of 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM
212 Na_2SO_4, 2 mM $MgSO_4$, 0.2x trace metals solution (Teknova, T1001) and was supplemented with
213 50 µg/ml leucine *E. coli* DH10B strains are auxotrophic for leucine. Each agar plate was
214 inoculated with 100 µl of a 0.5 McFarland unit suspension of *E. coli* DH10B harboring an empty
215 pZE21-ME plasmid. Plates were incubated at 37°C for roughly 48 hours then photographed to
216 record growth.

217 A functional metagenomic selection for acarbose modifying genes was carried out as
218 above with the Zymo fecal library plated to carry an estimated ten-fold excess of cells over
219 unique clones. The functional metagenomic library was washed once in modified MDG media
220 with no carbon added by pelleting at 8,000 rcf for 5 minutes, discarding the supernatant, and
221 resuspending in no carbon media. The washed library was then plated on a minimal media agar
222 plate containing 0.4% maltose as the sole carbon source supplemented with 50 µg/ml
223 kanamycin and 256 µg/ml acarbose. In parallel, a suspension of a control *E. coli* strain with
224 empty vector was washed as above and plated on maltose minimal media agar with acarbose at
225 a similar predicted titer. The plates were incubated at 37°C overnight. A weather-related power
226 outage and campus closure resulted in the plates incubating at room temperature for multiple
227 days. Eight colonies from the Zymo fecal library plate that were judged to be substantially larger
228 than background colonies on the empty plasmid control plate were picked, re-streaked to obtain
229 pure cultures, and processed by miniprep kit to obtain their plasmids for sequencing as above.
230 The plasmid sequences were analyzed as above and the most prevalent sequence (ACA1, the
231 S-layer homology domain-containing protein) and a sequence with homology to α-amylase
232 family glycosyl hydrolase (ACA2) were chosen for further study. The predicted amino acid
233 sequence of ACA1 was further submitted to the Contrastive Learning-Enabled Enzyme
234 Annotation (CLEAN) webserver for potential annotation[24].

235

236 **Phylogenetic analyses**

237 For phylogenetic analysis of the putative tetracycline efflux pumps (TET1, TET3, and
238 TET13), we downloaded CARD amino acid sequences[19,20] for the ontology term “major
239 facilitator superfamily (MFS) antibiotic efflux pump” and added the functionally-selected efflux
240 pump amino acid sequences to the FASTA file. For analysis of the predicted acarbose
241 resistance proteins (ACA1 and ACA7), we compared them to sequences for the *Klebsiella*
242 *grimontii* Apg enzyme and a *Raoultella ornithinolytica* α-glycosidase from the same
243 manuscript[25], additional reported acarbose degradation enzymes (maltogenic amylases from
244 *Thermus* sp. IM6501, *Geobacillus strearothermophilus*, and a glycoside hydrolase from
245 *Lactiplantibacillus plantarum*)[26–29], as well as acarbose phosphotransferases AcbK
246 (*Actinoplanes* sp. SE50/110), GacK (*Streptomyces glaucescens*), CatK (*Streptomyces*
247 *coelicoflavus* ZG0656), and Mak1 (uncultured bacterium)[30]. Phylogenetic trees were prepared
248 by aligning the protein sequences using the Clustal Omega web server with default settings[15]
249 and preparing a maximum likelihood phylogenetic tree from the alignment using IQ-TREE[31,32]
250 with the “Find best and apply” substitution model, standard 100 replicates bootstrap and
251 otherwise default settings. The resulting trees were visualized using the Interactive Tree of Life
252 display and annotation tool[33].

253

254 **Microbroth dilution antimicrobial susceptibility testing**

255 The antimicrobial resistance of *E. coli* strains expressing metagenomic fragments
256 encoding TET1, TET3, and TET13 were quantified by microbroth dilution assay[22,34]. Briefly, the
257 corresponding strains and control *E. coli* carrying an empty plasmid were streaked for single
258 colonies on LB agar with kanamycin. Approximately 10 colonies for each strain were suspended
259 in 2 ml of MH media with kanamycin and brought to an OD_{600} of between 0.1 and 0.15 to

260 correspond to a 0.5 McFarland standard density, then diluted 100-fold in MH broth with
261 kanamycin. We prepared a 96-well plate (Costar, 3370) to contain 50 µl of MH broth with 50
262 µg/ml kanamycin and 2-fold increasing gradients of tetracycline, chloramphenicol, or
263 azithromycin at twice the final target concentration and added 50 µl of the diluted bacterial
264 suspensions with each strain assayed in quadruplicate. The 96-well plates were sealed with a
265 Breathe-Easy membrane (MilliporeSigma, Z380059) and incubated at 37°C with shaking for 20
266 to 24 hours. Culture density was recorded as OD_{600} with a 1 cm pathlength correction using an
267 Epoch 2 plate reader (BioTek) and concentration-response curves were fitted to four parameter
268 Hill equations in GraphPad Prism 10.2.3 (GraphPad Software, La Jolla, CA, USA) to determine
269 50% inhibitory concentration (IC_{50}) values. Differences between IC_{50} values for control *E. coli*
270 and the tested strains (TET1, TET3, TET13) were evaluated for significance in GraphPad Prism
271 using Brown-Forsythe ANOVA tests with multiple comparison correction using the Dunnett T3
272 method.

273

274 **Acarbose growth inhibition**

275 The growth of a control *E. coli* strain and strains carrying the ACA1- or ACA7-containing
276 metagenomic DNA fragments were recorded as OD_{600} culture densities following 24 hours of
277 aerobic growth at 37°C using a BioTek Epoch 2 plate reader. The three strains to be tested were
278 prepared as above to give a 100-fold dilution of 0.5 McFarland suspension in modified MDG
279 minimal media without a carbon source added. We prepared a 96-well plate where each row
280 contained 50 µl of a different media condition (all wells containing kanamycin, acarbose
281 concentrations are twice the final concentration): MH ± 512 µg/ml acarbose, modified MDG with
282 0.4% glucose ± 512 µg/ml acarbose, modified MDG with 0.4% maltose ± 512 µg/ml acarbose,
283 and modified MDG with no carbon added. The cultures were inoculated in quadruplicate for
284 each condition by adding 50 µl of bacterial suspension in modified MDG no carbon to the above

285 media. The plate was sealed with a Breathe-Easy membrane and incubated with shaking at
286 37°C for 24 hours before recording growth. For each media condition, final OD_{600} values were
287 compared pairwise for each strain with or without acarbose in each media, and significance was
288 determined in GraphPad Prism as above.

289

290 **Sequence availability**

291 Sequence files for TET1, TET3, TET13, ACA1, and ACA7 can be found as supplemental files 1
292 through 5.

293

294 **RESULTS**

295

296 **METa assembly can reproducibly prepare Gb sized libraries with 50 ng of input**
297 **metagenomic DNA**

298 In our manuscript describing METa assembly, we successfully prepared functional
299 metagenomic libraries using metagenomic DNA input quantities of 200 ng to 300 ng,
300 approximately 25- to 67-fold less than is commonly used to prepare small insert libraries[8,9]. An
301 advantage of our tagmentation-based fragmentation approach over physical fragmentation (*e.g.,*
302 *via* sonication) is its ready scalability to lower volumes and therefore lower input DNA masses.
303 However, we suspected that DNA losses during purification and fragment size-selection by gel
304 excision would make sub-200 ng libraries impractical. Our published METa assembly protocol
305 tagments metagenomic DNA at a concentration of 10 ng/µl, meaning a 50 ng DNA tagmentation
306 reaction would have a 5 µl volume where accurate pipetting starts to become difficult. To test the
307 practicality of extracting useful DNA fragments from this size reaction, we diluted 50 ng of a

Table 1. Summary statistics of functional metagenomic libraries

Library	Input DNA	DNA mass (ng)	Ave insert size (kb)	Proportion	Library size (Gb)	Efficiency (Gb/µg)	Selections[1]
50 ng goose 1	Goose gut microbiome	50	2.48	1.00	0.89	17.8	n/a
50 ng goose 2[2]	Goose gut microbiome	50	2.62	1.00	0.10	2.09	n/a
50 ng goose 3	Goose gut microbiome	50	3.13	1.00	2.46	49.3	n/a
Aquatic	Shedd aquarium	30.5	2.77	0.88	0.72	23.6	TET, CIP, FFC
Fecal	Human fecal reference	100.8	2.70	0.90	1.48	1.48	ACA

[1]no selection (n/a), tetracycline (TET), ciprofloxacin (CIP), florfenicol (FFC), acarbose (ACA)

[2]Replicate sparked during electroporation

308 DNA ladder in tagmentation buffer in triplicate to a concentration of 10 ng/µl to make mock 5 µl
309 tagmentation reactions. The triplicate mock fragmented samples were loaded on to an agarose
310 gel for excision of bands ranging from approximately 2 kb to 10 kb (**Supplemental figure 1A**).
311 Surprisingly to us, this resulted in 'fragmented DNA' quantities averaging 12.7 ng (± 2.7
312 standard deviation). Given that the total mass of the DNA bands 2 kb and higher was 16 ng, this
313 represented a 79% yield and suggested that size-selection of fragmented DNA from a 50 ng
314 tagmentation reaction could be feasible (**Supplemental figure 1B**).

315 Next, we tested this explicitly by preparing a 50 ng metagenomic DNA functional
316 metagenomic library in triplicate. Using previously extracted goose fecal metagenomic DNA as
317 input[8], we performed three tagmentation reactions and used gel excision to size-select for DNA
318 fragments between 2 kb and 6.5 kb (**Supplemental figure 2A**). These fragments were used to
319 prepare three functional metagenomic libraries by METa assembly. During the electroporation
320 step, replicates 1 and 3 reached full voltage (1.8 kV) and had time constants of 4.6 and 4.8
321 seconds respectively. Replicate 2 reported a voltage of 1.65 kV and a 1.0 second time constant,
322 consistent with the sample 'sparking'. Across the three libraries we found an average insert size
323 of 2.75 kb with no empty vectors (**Supplemental figure 2B**), an average library size of 1.15 Gb,
324 and a library preparation efficiency of 23 Gb/µg of input DNA. Notably, these averages include
325 replicate 2 with a dramatically smaller library size (100 Mb compared to 890 Mb and 2.5 Gb)
326 (**Table 1**). If replicate 2 is excluded, the average library size for the 50 ng functional
327 metagenomic libraries increases to 1.68 Gb with an efficiency of 33.6 Gb/µg.

328

329 **METa assembly preparation of a functional metagenomic library from a limited, low**
330 **biomass microbiome**

331 Two areas where input mass is limiting for classic functional metagenomic library
332 preparation are 1) when the microbiome of interest is lower in biomass (*e.g.,* aquatic sources as
333 opposed to soil or fecal samples) and 2) when the microbiome of interest is not readily available
334 for re-sampling. We decided to test low-input mass METa assembly on a microbiome that fulfills
335 both limitations. In a previous project, metagenomic DNA was extracted from the Underwater
336 Beauty exhibit at the Shedd Aquarium in Chicago, Illinois following filtration capture of planktonic
337 cells[13]. We used a sample of previously extracted metagenomic DNA to prepare a functional
338 metagenomic library. Tagmentation was performed on 30.5 ng of metagenomic DNA and
339 fragments from 2 kb to 10 kb were size selected by gel excision. Following transformation,
340 random colonies from the functional metagenomic library were used as template for PCR to
341 amplify captured metagenomic inserts to determine library statistics. We found that this aquatic
342 functional metagenomic library captured 720 Mb of metagenomic DNA with an average insert
343 size of 2.77 kb and with 12% of plasmids not containing an insert. The library preparation
344 efficiency was calculated to be 23.6 Gb/µg (**Table 1**).

345 To evaluate the usefulness of this functional metagenomic library, we used it to
346 investigate antimicrobial resistance against three antibiotics that are commonly used in
347 aquaculture: tetracycline (**Figure 1C**), florfenicol, and ciprofloxacin[35,36]. The selection on
348 ciprofloxacin did not yield any resistant colonies but colonies were recoverable from both the
349 tetracycline and florfenicol plates. As an initial investigation, fifteen colonies from the tetracycline
350 selection were collected and submitted for whole plasmid sequencing. We dereplicated
351 repeated inserts, which resulted from plating a ten-fold excess of total cells over unique cells,
352 resulting in three *E. coli* strains, TET1, TET3, and TET13. The metagenomic DNA captured in
353 these strains had top nucleotide BLAST hits to *Lentibacter algarum* strain SH36 (95% coverage,

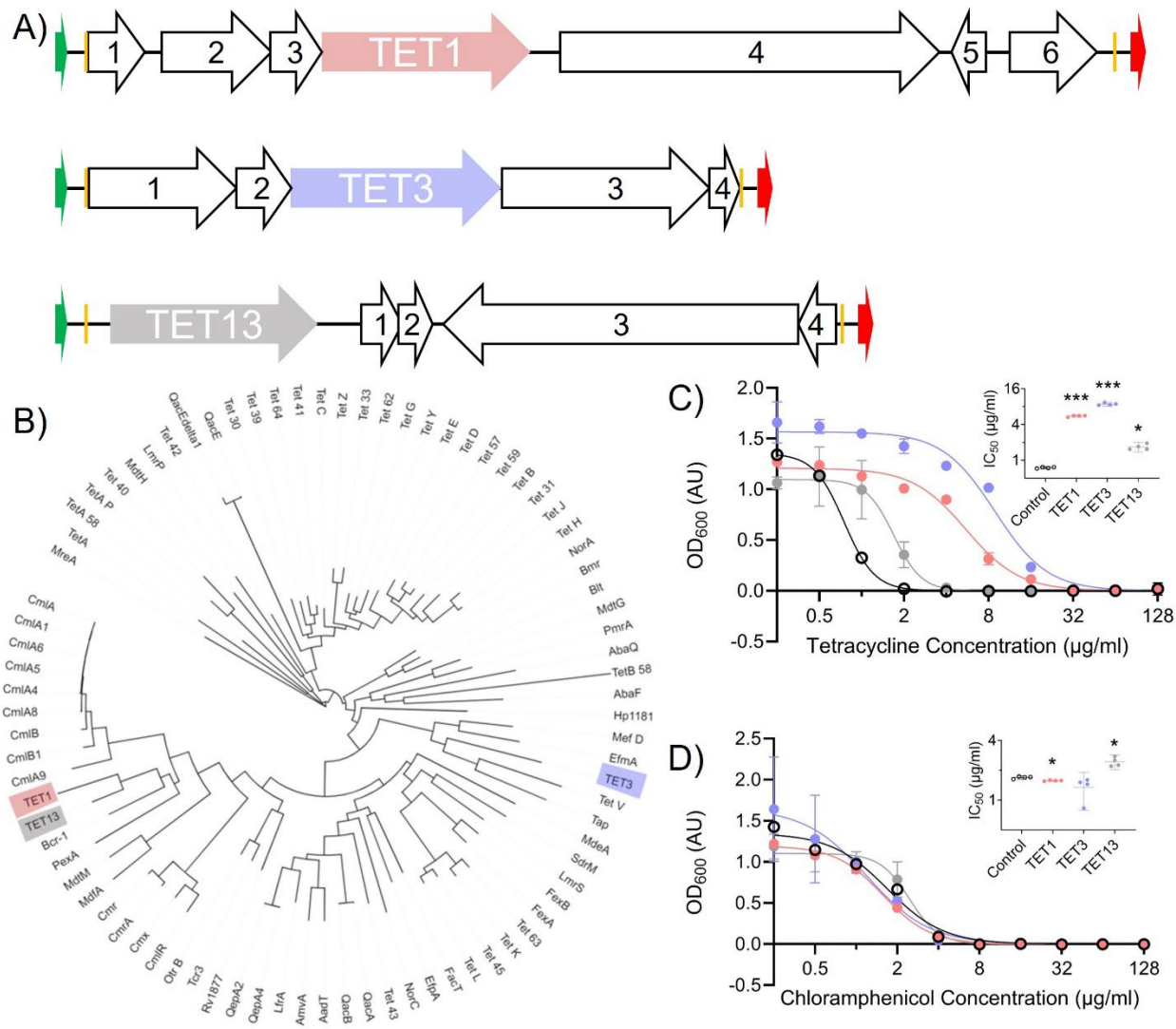

Figure 2. Tetracycline resistance in the Shedd aquarium aquatic microbiome. A) Gene schematics of the TET1, TET3, and TET13 metagenomic DNA fragments with predicted efflux pumps highlighted (TET1 red, TET3 blue, TET13 gray). Other predicted open reading frame are shown as empty arrows and plasmid backbone markers are highlighted as follows: Promoter (green), mosaic end sequence (orange), terminator (red). B) Maximum likelihood phylogenetic tree of CARD MFS efflux pumps with predicted TET1, TET3, and TET13 amino acid sequences highlighted. Microbroth dilution assays performed un quadruplicate for C) tetracycline or D) chloramphenicol resistance with calculated 50% inhibition concentration (IC_{50}) values graphed as insets. For statistical comparisons: * $p<0.05$, ** $p<0.005$, and *** $p<0.0005$.

82.09% identity), *Legionella pneumophila* strain NY23 (30% coverage, 66.87% identity), and *Saccharophagus degradans* strain FZY0027 (32% coverage, 70.26% identity), respectively.

We annotated the metagenomic DNA fragments captured in these three strains (**Supplemental figures 3A, 3B, and 3C**) and found them to each contain predicted MFS family

358 efflux pumps with low full-length amino acid identity (<30%) to their closest match in The
359 Comprehensive Antibiotic Resistance Database (CARD)[19,20] (**Figure 2A**). When placed on a
360 phylogenetic tree alongside known antibiotic resistance efflux pumps, TET1 and TET13 cluster
361 with each other and a variety of efflux pumps with associated chloramphenicol resistance. TET3
362 instead clusters with known macrolide and tetracycline efflux pumps (**Figure 2B**). The
363 resistance of all three strains to tetracycline, chloramphenicol, and azithromycin were measured
364 by microbroth dilution. All three strains showed significantly increased resistance to tetracycline,
365 with TET1 and TET3 showing growth at 8 µg/ml tetracycline or higher (**Figure 2C**). The *E. coli*
366 strain expressing the TET3 metagenomic fragment also showed a statistically significant
367 reduction in chloramphenicol susceptibility, while the strain corresponding to TET1 showed a
368 significantly increased susceptibility to the antibiotic (**Figure 2D**). None of the three strains
369 showed significant differences in azithromycin resistance compared to the control strain
370 (**Supplemental figure 3D**).

371

372 **METa assembly preparation of a functional metagenomic library from a fecal swab**

373 Another area where the ability to prepare a functional metagenomic library using minimal
374 DNA resources could be beneficial is within clinical settings. For example, fecal swabs provide
375 access to the medically important gut microbiome, but extraction of metagenomic DNA from this
376 sample type falls far below what classic functional metagenomic library preparation methods call
377 for[14,37]. To test the possibility of medical swabs being used as input for METa assembly, we
378 turned to the ZymoBIOMICS fecal reference with TruMatrix technology fecal standard. Each
379 aliquot of this standard is expected to contain 10 mg (wet weight) of human fecal material in 1
380 ml of DNA/RNA shield preservation buffer. We sampled this microbiome by dipping a sterile
381 swab into a thawed tube of the material followed by DNA extraction from the swab, resulting in
382 collection of 100.8 ng of DNA. We prepared and evaluated a functional metagenomic library by

383 METa assembly as above and found it to contain an average insert size of 2.7 kb, 10% empty
384 vectors, for a total expected library size of 1.48 Gb and a library preparation efficiency of 14.7
385 Gb/µg (**Table 1**).

386 The ZymoBIOMICS fecal standard has already been sequenced and annotated for
387 known antibiotic resistance genes. As a result, we chose to investigate a different function in the
388 human gut microbiome, pharmaceutical metabolism. Specifically, we investigated modification
389 of acarbose (**Figure 1D**), a glucosidase inhibitor used in the treatment of diabetes[38,39]. Previous
390 work has established that acarbose can be metabolized by bacteria in the human gut
391 microbiome through modification by phosphorylation[30] and degradation[25–29,29,40] and that
392 acarbose can prevent growth of *E. coli* when it is cultured with maltose as a sole carbon
393 source[41]. This suggested to us that a functional metagenomic approach could be used to
394 investigate acarbose metabolism. We validated this by plating *E. coli* in a lawn on minimal
395 media agar plates containing maltose as a sole carbon and energy source with acarbose
396 concentrations ranging from 0.5 µg/ml to 256 µg/ml. We observed confluent growth until an
397 acarbose concentration of 256 µg/ml at which point almost no colonies were observed
398 (**Supplemental figure 4A**). In contrast, we confirmed growth in the presence of 256 µg/ml
399 acarbose when *E. coli* was plated on minimal media containing glucose as a carbon source
400 (**Supplemental figure 4B**).

401 Based on these observations, we performed a functional metagenomic selection for
402 growth on maltose in the presence of acarbose. An aliquot of the Zymo fecal functional
403 metagenomic library and a similar titer of *E. coli* containing an empty vector were plated on to
404 minimal media plates containing maltose as the sole carbon source and acarbose at a
405 concentration of 256 µg/ml. Inclement weather resulted in the selections incubating at room
406 temperature for multiple days after which both plates showed a lawn of very small colonies.
407 However, the agar plate that was seeded with the functional metagenomic library contained

408 several larger colonies, the eight largest of which were collected for further study
409 (**Supplemental figure 4C**).

410 Following whole plasmid sequencing of these eight colonies, we found that four
411 contained identical inserts to each other and refer to the representative strain as ACA1. Protein
412 BLAST analysis of the single predicted open reading frame of ACA1 annotated it as an S-layer
413 homology domain-containing protein (**Figure 3A**). We used the CLEAN enzyme function
414 annotation tool[24] to suggest a potential function. Interestingly, E.C. families 3.2.1.8 and 3.2.1.4
415 were suggested as potential hits and both belong to the glycosidase family of enzymes. Analysis
416 of the ACA1 metagenomic DNA fragment by nucleotide BLAST failed to find a high-quality hit,
417 with the top predicted match (12% coverage, 77.18% identity) being to a bacteriophage and the
418 top predicted bacterial match (3% coverage, 90.74% identity) being to an *Intestinibacillus*
419 species. Plasmid sequencing of one of the other colonies (termed ACA7) found a metagenomic
420 insert predicted to encode three open reading frames (**Supplemental figure 5A**), one of which
421 was predicted to encode an α-amylase family glycosyl hydrolase gene (**Figure 3A**). Unlike the
422 enigmatic ACA1, the nucleotide sequence associated with ACA7 had a high-quality hit on
423 BLAST, *Alistipes onderdonkii* susbspecies *vulgaris* (100% coverage, 97.18% identity).

424 To orient the predicted effectors of ACA1 and ACA7 we placed them on a maximum
425 likelihood phylogenetic tree built on the sequences of known acarbose degrading enzymes and
426 known acarbose phosphotransferase (or kinase) enzymes (**Figure 3B**). As expected, these two
427 functions, glycan degradation and phosphorylation, segregate on the phylogenetic tree. The
428 predicted α-amylase ACA7 clusters with the known acarbose degrading enzymes while ACA1
429 does not appear to cluster with either enzyme group; its exact function is currently under
430 investigation.

431 To verify that ACA1 and ACA7 counteract the inhibitory activity of acarbose on *E. coli*
432 growth on maltose, we compared the growth of the two corresponding strains to an empty
433 plasmid control *E. coli* strain under a variety of conditions. When cultured in Mueller-Hinton
434 broth, the addition of acarbose resulted in a decreased final optical density of the control *E. coli*,
435 potentially reflecting the mixed carbon sources available in that media. In contrast, the ACA1
436 and ACA7 strains showed no effect on addition of acarbose **(Supplemental figure 5B**). In

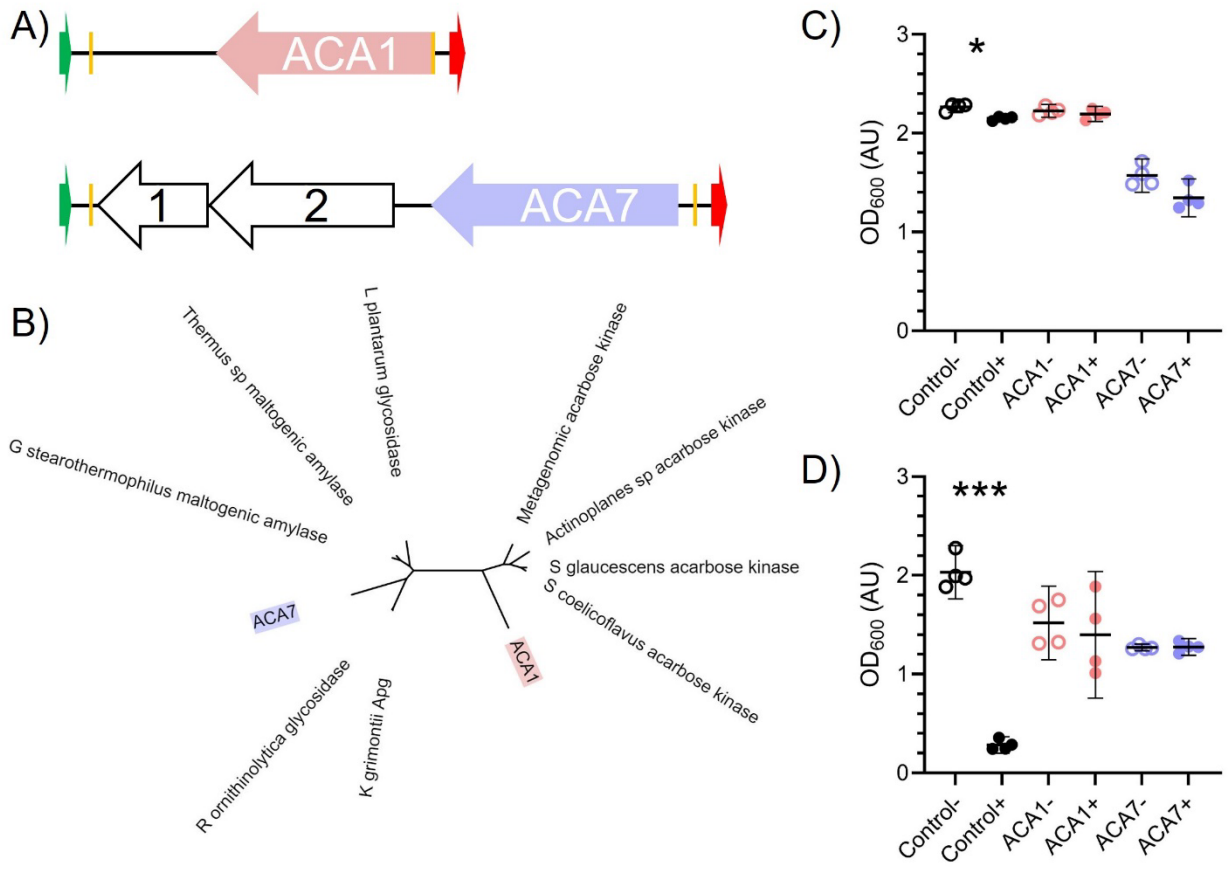

Figure 3. Predicted acarbose degradation enzymes from the human gut microbiome. A) Gene schematics of the ACA1 and ACA7 metagenomic DNA fragments with the predicted effector genes highlighted (ACA1 red and ACA7 blue). Predicted genes captured on the same DNA fragment are shown as empty arrows. Plasmid backbone elements are: Promoter (green), mosaic end sequence (orange), terminator (red). B) Maximum likelihood phylogenetic tree of the predicted proteins for ACA1 (red) and ACA7 (blue) in the context of known acarbose degrading or modifying enzymes (*i.e.,* kinases). C) and D) Culture densities with or without addition of 256 µg/ml acarbose after 24 hours of growth in minimal media with C) 0.4% glucose or D) 0.4% maltose as carbon sources. Pairwise statistical comparisons for the effect of acarbose addition were made for each strain and culture condition: * $p<0.05$, ** $p<0.005$, and *** $p<0.0005$.

437 minimal media with glucose as a carbon source, all three strains showed essentially no change
438 in growth in the presence of acarbose, although we observed a minor but statistically significant
439 decrease in the OD_{600} reached by the control strain (**Figure 3C**). Finally, when grown in minimal
440 media with maltose as a carbon source, the growth of the empty plasmid *E. coli* strain is nearly
441 entirely inhibited in the presence of acarbose while ACA1 and ACA7 show no difference in
442 growth on the addition of acarbose to the media (**Figure 3D**).

443

444 **DISCUSSION**

445 Here we report that our METa assembly method for preparing functional metagenomic
446 libraries can be extended to preparing libraries from low biomass and otherwise precious
447 microbiome samples where metagenomic DNA would otherwise be limiting. As an illustration of
448 this potential, we prepared modestly-sized functional metagenomic libraries from metagenomic
449 DNA extracted from an aquarium filtrate and from a simulated fecal swab. We further
450 demonstrated the utility of the resulting functional metagenomic libraries by applying them to
451 identify likely tetracycline efflux pumps and novel acarbose degradation genes.

452 A key result of our experiments is that using METa assembly to prepare functional
453 metagenomic libraries allows us move the minimal metagenomic DNA input to approximately
454 two orders of magnitude lower than previous standard protocols[9,14]. We show that, across our
455 experiments, 50 ng or lower input metagenomic DNA masses can be used to prepare functional
456 metagenomic libraries encoding near or low Gb libraries (**Table 1**). We propose, and
457 demonstrate, that this opens new sample types to analysis by functional metagenomic
458 selections or screens. These include samples that are intrinsically of low biomass, either due to
459 difficulty in collection or their rarity (*e.g.,* aquatic metagenomes, but also historically banked
460 microbiomes or samples from remote regions). These also include microbiomes sampled using

461 medical swabs. While our example used Zymo fecal reference material, the mass of the
462 metagenomic DNA extracted from the swab and used to prepare a functional metagenomic
463 library was noticeably less than what has been reported in extractions from other fecal swabs
464 (*i.e.,* ~100 ng vs ~370 ng). Functional metagenomic selections have been proposed to be useful
465 for medical applications[42], and our results support that.

466 Our previous smallest METa assembly input, 200 ng, resulted in a functional
467 metagenomic library encoding approximately 13.5 Gb with a preparation efficiency of ~68
468 Gb/µg[8]. Here, our 50 ng, 30.5 ng, and 100.8 ng libraries averaged an efficiency of between 20.5
469 Gb/µg and 24.1 Gb/µg, depending on the inclusion or exclusion of the library that sparked
470 during transformation, suggesting that lower input DNA masses may come with decreased
471 library preparation efficiency (**Table 1**). It is possible that this is due to DNA losses during the
472 multiple purification steps involved in library preparation. Nevertheless, these levels of efficiency
473 still perform better than our previously estimated average of 1 Gb/µg using blunt ligation-based
474 techniques[8].

475 While the functional metagenomic libraries prepared here are smaller than our previous
476 high-mass input libraries, their usefulness is demonstrated by their capture of genes for
477 tetracycline efflux pumps and acarbose degradation (**Figures 2 and 3**). The three studied
478 predicted tetracycline efflux pumps show homology to previously identified MFS efflux pumps,
479 with the predicted amino acid sequences having near complete coverage and between 65% and
480 almost 100% identity to hits in the NCBI database. In contrast, they have only low identity to
481 CARD hits, with full length alignments showing only 28% to 29% identity. It is common for
482 proteins identified through functional metagenomic selections for antibiotic resistance to have
483 high identity to known hits, but for those proteins to not previously have been identified as
484 having antibiotic resistance potential[43]. Tetracycline efflux pumps are found within many
485 environments[44–46] but the substrate specificity of novel transporters makes the identification of

486 potential resistance genes from environmental microbiomes difficult. This is illustrated by the
487 contrast between the apparent clustering of the TET1 and TET13 MFS pumps with amphenicol
488 efflux and TET3 with macrolide efflux (**Figure 2B**) and their respective lack of resistance to
489 chloramphenicol (**Figure 2D**) or azithromycin (**Supplemental figure 3D**). Our work highlights
490 the important role that functional metagenomic selections can play in filling this knowledge gap.

491 Similarly, one of the predicted proteins from the acarbose/maltose selection, ACA7, has
492 very high coverage and identity to known proteins in the NCBI database that lack an acarbose-
493 modification annotation. On the other hand, the predicted protein from ACA1 only showed
494 homology proteins with limited annotation as an S-layer homology domain-containing protein.
495 Interestingly, while ACA7 appears to clearly cluster with known acarbose degrading proteins (a
496 class which expanded again recently following a functional metagenomic screen for glycan
497 breakdown[40]), ACA1 does not (**Figure 3B**). Instead, ACA1 may represent a new acarbose
498 modifying enzyme. When ACA1 is compared against the CARD database by protein BLAST, the
499 top hit is a chloramphenicol acetyltransferase with an E-value of 0.09. We are currently
500 investigating the possibility that ACA1 may have transferase activity.

501 While we see the opening of low biomass functional metagenomic libraries as a
502 significant advancement, limitations to the method remain. First, this method relies on *E. coli*
503 cells as the host organism for the functional metagenomic library and therefore is subject to
504 potential limits on what metagenomic genes can be successfully expressed by this host[47]. In
505 theory this can be surmounted using alternative hosts[48–50], and the added efficiency of METa
506 assembly may be more important when using strains with lower transformation efficiency than
507 commercial *E. coli* strains. Second, there are potentially diminishing returns to preparing smaller
508 functional metagenomic libraries even if their preparation efficiency increases. However, the
509 libraries prepared here were shown to still be useful by their capture of potentially novel
510 antibiotic efflux pumps and acarbose degradation enzymes even without covering more than a

511 fraction of the genetic diversity found in many microbiomes. Tagmentation reactions are
512 regularly performed on metagenomic DNA samples with 5 ng or less mass[12,51,52] suggesting that
513 the limit for functional metagenomic libraries could be pushed even further.

514 In conclusion, while high-throughput DNA sequencing has helped circumvent the great
515 plate count anomaly and has revolutionized microbiome research, methods for accurately
516 annotating novel genes have not managed to keep pace. Machine-learning tools will likely be
517 able to help with this issue but will also require large amounts of high-quality sequence-function
518 correlation data. Functional metagenomics is well placed to supply this data and, in the
519 meantime, can make significant strides in narrowing the sequence annotation gap. We propose
520 that the fulfillment of this potential by functional metagenomics is significantly aided by the ability
521 to make the necessary libraries more efficiently and using less input DNA, a precondition that
522 we demonstrate is reached here.

523

524 **Acknowledgements**

525 CRediT contributions: Conceptualization – TSC; Data curation –TSC; Formal analysis – HMA,
526 TSC; Funding acquisition – EMH, TSC; Investigation – HMA, EPB, EF, TSC; Project
527 administration – TSC Resources – FJO EMH, TSC; Supervision – EMH, TSC; Visualization –
528 HMA, EPB, TSC; Writing - original draft – TSC; Writing - reviewing and editing – HMA, EPB,
529 FJO, EMH, TSC. We would like to thank Chrissy Cabay, William Van Bonn, and James Clark for
530 their help in acquiring the Shedd Aquarium sample and Jinglin Hu for the metagenomic DNA.
531 This work was supported by a grant from the Florida State University Council for Research and
532 Creativity and by start-up funds from the Florida State University College of Medicine.

533

534 **References**

1. Staley, James T; Konopka, A., and Staley Allan, J.T.K. (1985). Microorganisms in Aquatic and Terrestrial Habitats. Annual review of microbiology *39*, 321–346.

2. Boolchandani, M., D'Souza, A.W., and Dantas, G. (2019). Sequencing-based methods and resources to study antimicrobial resistance. Nature Reviews Genetics *20*, 356–370. https://doi.org/10.1038/s41576-019-0108-4.

3. Dick, G. (2018). Genomic Approaches in Earth and Environmental Sciences (John Wiley & Sons, Ltd) https://doi.org/10.1002/9781118708231.

4. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & biology *5*, R245-9. https://doi.org/10.1016/S1074-5521(98)90108-9.

5. Crofts, T.S., Gasparrini, A.J., and Dantas, G. (2017). Next-generation approaches to understand and combat the antibiotic resistome. Nature reviews. Microbiology *15*, 422–434. https://doi.org/10.1038/nrmicro.2017.28.

6. Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D., and Charles, T.C. (2015). Current and future resources for functional metagenomics. Front. Microbiol. *6*. https://doi.org/10.3389/fmicb.2015.01196.

7. Taupp, M., Mewis, K., and Hallam, S.J. (2011). The art and design of functional metagenomic screens. Current opinion in biotechnology *22*, 465–472. https://doi.org/10.1016/j.copbio.2011.02.010.

8. Crofts, T.S., McFarland, A.G., and Hartmann, E.M. (2021). Mosaic Ends Tagmentation (METa) Assembly for Highly Efficient Construction of Functional Metagenomic Libraries. mSystems *6*, e0052421. https://doi.org/10.1128/mSystems.00524-21.

9. Mahmud, B., Boolchandani, M., Patel, S., and Dantas, G. (2023). Functional Metagenomics to Study Antibiotic Resistance. Methods Mol Biol *2601*, 379–401. https://doi.org/10.1007/978-1-0716-2855-3_21.

10. Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Research *25*, 1203–1210. https://doi.org/10.1093/nar/25.6.1203.

11. Crofts, T.S., Sontha, P., King, A.O., Wang, B., Biddy, B.A., Zanolli, N., Gaumnitz, J., and Dantas, G. (2019). Discovery and Characterization of a Nitroreductase Capable of Conferring Bacterial Resistance to Chloramphenicol. Cell chemical biology *26*, 559-570.e6. https://doi.org/10.1016/j.chembiol.2019.01.007.

12. Picelli, S., Björklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome research *24*, 2033–2040. https://doi.org/10.1101/gr.177881.114.

13. Hu, J., Hellgeth, N., Cabay, C., Clark, J., Oliaro, F.J., Van Bonn, W., and Hartmann, E.M. (2022). Towards understanding microbial degradation of chloroquine in large saltwater systems. Science of the Total Environment *807*, 150532. https://doi.org/10.1016/j.scitotenv.2021.150532.

14. Boolchandani, M., Patel, S., and Dantas, G. (2017). Antibiotics P. Sass, ed. (Springer New York) https://doi.org/10.1007/978-1-4939-6634-9.

577 15. Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan,
578 N., Kolesnikov, A., and Lopez, R. (2022). Search and sequence analysis tools services from
579 EMBL-EBI in 2022. Nucleic Acids Research *50*, W276–W279.
580 https://doi.org/10.1093/nar/gkac240.

581 16. Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in
582 metagenomic sequences. Nucleic acids research *38*, e132.
583 https://doi.org/10.1093/nar/gkq275.

584 17. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman,
585 D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search
586 programs. Nucleic Acids Research *25*, 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

587 18. Schaffer, A.A. (2001). Improving the accuracy of PSI-BLAST protein database searches
588 with composition-based statistics and other refinements. Nucleic Acids Research *29*, 2994–
589 3005. https://doi.org/10.1093/nar/29.14.2994.

590 19. Alcock, B.P., Huynh, W., Chalil, R., Smith, K.W., Raphenya, A.R., Wlodarski, M.A.,
591 Edalatmand, A., Petkau, A., Syed, S.A., Tsang, K.K., et al. (2023). CARD 2023: expanded
592 curation, support for machine learning, and resistome prediction at the Comprehensive
593 Antibiotic Resistance Database. Nucleic Acids Research *51*, D690–D699.
594 https://doi.org/10.1093/nar/gkac920.

595 20. Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A.,
596 Huynh, W., Nguyen, A.L.V., Cheng, A.A., Liu, S., et al. (2020). CARD 2020: Antibiotic
597 resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids
598 Research *48*, D517–D525. https://doi.org/10.1093/nar/gkz935.

599 21. Davis, M.W., and Jorgensen, E.M. (2022). ApE, A Plasmid Editor: A Freely Available
600 DNA Manipulation and Visualization Program. Frontiers in Bioinformatics *2*, 1–15.
601 https://doi.org/10.3389/fbinf.2022.818619.

602 22. Wiegand, I., Hilpert, K., and Hancock, R.E.W. (2008). Agar and broth dilution methods to
603 determine the minimal inhibitory concentration (MIC) of antimicrobial substances. Nature
604 Protocols *3*, 163–175. https://doi.org/10.1038/nprot.2007.521.

605 23. Studier, F.W. (2005). Protein production by auto-induction in high density shaking
606 cultures. Protein expression and purification *41*, 207–234.
607 https://doi.org/10.1016/j.pep.2005.01.016.

608 24. Yu, T., Cui, H., Li, J.C., Luo, Y., Jiang, G., and Zhao, H. (2023). Enzyme function
609 prediction using contrastive learning. Science *379*, 1358–1363.
610 https://doi.org/10.1126/science.adf2465.

611 25. Tian, J., Li, C., Dong, Z., Yang, Y., Xing, J., Yu, P., Xin, Y., Xu, F., Wang, L., Mu, Y., et al.
612 (2023). Inactivation of the antidiabetic drug acarbose by human intestinal microbial-mediated
613 degradation. Nat Metab *5*, 896–909. https://doi.org/10.1038/s42255-023-00796-w.

614 26. Kim, T.-J., Kim, M.-J., Kim, B.-C., Kim, J.-C., Cheong, T.-K., Kim, J.-W., and Park, K.-H.
615 (1999). Modes of Action of Acarbose Hydrolysis and Transglycosylation Catalyzed by a
616 Thermostable Maltogenic Amylase, the Gene for Which Was Cloned from a *Thermus* Strain.
617 Appl Environ Microbiol *65*, 1644–1651. https://doi.org/10.1128/AEM.65.4.1644-1651.1999.

618 27. Cheong, K., Kim, T., Yoon, J., Park, C., Lee, T., Kim, Y., Park, K., and Kim, J. (2002).
619 Catalytic activities of intracellular dimeric neopullulanase on cyclodextrin, acarbose and
620 maltose. Biotech and App Biochem *35*, 27–34. https://doi.org/10.1042/BA20010052.

621 28. Cha, H., Yoon, H., Kim, Y., Lee, H., Kim, J., Kweon, K., Oh, B., and Park, K. (1998).
622 Molecular and enzymatic characterization of a maltogenic amylase that hydrolyzes and
623 transglycosylates acarbose. European Journal of Biochemistry *253*, 251–262.
624 https://doi.org/10.1046/j.1432-1327.1998.2530251.x.

625 29. Jang, M.-U., Kang, H.-J., Jeong, C.-K., Kang, Y., Park, J.-E., and Kim, T.-J. (2018).
626 Functional expression and enzymatic characterization of Lactobacillus plantarum
627 cyclomaltodextrinase catalyzing novel acarbose hydrolysis. J Microbiol. *56*, 113–118.
628 https://doi.org/10.1007/s12275-018-7551-3.

629 30. Balaich, J., Estrella, M., Wu, G., Jeffrey, P.D., Biswas, A., Zhao, L., Korennykh, A., and
630 Donia, M.S. (2021). The human microbiome encodes resistance to the antidiabetic drug
631 acarbose. Nature *600*. https://doi.org/10.1038/s41586-021-04091-0.

632 31. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., and Minh, B.Q. (2016). W-IQ-TREE: a
633 fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res *44*, W232–
634 W235. https://doi.org/10.1093/nar/gkw256.

635 32. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von
636 Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for
637 Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution *37*, 1530–1534.
638 https://doi.org/10.1093/molbev/msaa015.

639 33. Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the
640 phylogenetic tree display and annotation tool. Nucleic Acids Research, gkae268.
641 https://doi.org/10.1093/nar/gkae268.

642 34. Committee, E., Testing, S., Microbiology, C., and Escmid, I.D. (2003). Determination of
643 minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution. Clinical
644 Microbiology and Infection *9*, ix–xv. https://doi.org/10.1046/j.1469-0691.2003.00790.x.

645 35. Schar, D., Klein, E.Y., Laxminarayan, R., Gilbert, M., and Van Boeckel, T.P. (2020).
646 Global trends in antimicrobial use in aquaculture. Sci Rep *10*, 21878.
647 https://doi.org/10.1038/s41598-020-78849-3.

648 36. Bondad-Reantaso, M.G., MacKinnon, B., Karunasagar, I., Fridman, S., Alday-Sanz, V.,
649 Brun, E., Le Groumellec, M., Li, A., Surachetpong, W., Karunasagar, I., et al. (2023). Review
650 of alternatives to antibiotic use in aquaculture. Reviews in Aquaculture *15*, 1421–1451.
651 https://doi.org/10.1111/raq.12786.

652 37. Biehl, L.M., Garzetti, D., Farowski, F., Ring, D., Koeppel, M.B., Rohde, H., Schafhausen,
653 P., Stecher, B., and Vehreschild, M.J.G.T.G.T. (2019). Usability of rectal swabs for
654 microbiome sampling in a cohort study of hematological and oncological patients. PLOS ONE
655 *14*, e0215428. https://doi.org/10.1371/journal.pone.0215428.

656 38. Whang, A., Nagpal, R., and Yadav, H. (2019). Bi-directional drug-microbiome
657 interactions of anti-diabetics. EBioMedicine *39*, 591–602.
658 https://doi.org/10.1016/j.ebiom.2018.11.046.

659 39. Wehmeier, U.F., and Piepersberg, W. (2004). Biotechnology and molecular biology of
660 the ?-glucosidase inhibitor acarbose. Applied Microbiology and Biotechnology *63*, 613–625.
661 https://doi.org/10.1007/s00253-003-1477-2.

662 40. Nasseri, S.A., Lazarski, A.C., Lemmer, I.L., Zhang, C.Y., Brencher, E., Chen, H.-M., Sim,
663 L., Panwar, D., Betschart, L., Worrall, L.J., et al. (2024). An alternative broad-specificity

664 pathway for glycan breakdown in bacteria. Nature. https://doi.org/10.1038/s41586-024-
665 07574-y.

666 41. Brunkhorst, C., Andersen, C., and Schneider, E. (1999). Acarbose, a
667 Pseudooligosaccharide, Is Transported but Not Metabolized by the Maltose-Maltodextrin
668 System of *Escherichia coli*. J Bacteriol *181*, 2612–2619.
669 https://doi.org/10.1128/JB.181.8.2612-2619.1999.

670 42. van der Helm, E., Imamovic, L., Hashim Ellabaan, M.M., van Schaik, W., Koza, A., and
671 Sommer, M.O.A. (2017). Rapid resistome mapping using nanopore sequencing. Nucleic
672 acids research *45*, e61. https://doi.org/10.1093/nar/gkw1328.

673 43. Sommer, M.O.A., Dantas, G., and Church, G.M. (2009). Functional characterization of
674 the antibiotic resistance reservoir in the human microflora. Science (New York, N.Y.) *325*,
675 1128–1131. https://doi.org/10.1126/science.1176950.

676 44. Hartmann, E.M., Hickey, R., Hsu, T., Betancourt Román, C.M., Chen, J., Schwager, R.,
677 Kline, J., Brown, G.Z., Halden, R.U., Huttenhower, C., et al. (2016). Antimicrobial Chemicals
678 Are Associated with Elevated Antibiotic Resistance Genes in the Indoor Dust Microbiome.
679 Environmental science & technology *50*, 9807–9815.
680 https://doi.org/10.1021/acs.est.6b00262.

681 45. Riesenfeld, C.S., Goodman, R.M., and Handelsman, J. (2004). Uncultured soil bacteria
682 are a reservoir of new antibiotic resistance genes. Environmental Microbiology *6*, 981–989.
683 https://doi.org/10.1111/j.1462-2920.2004.00664.x.

684 46. Leclercq, S.O., Wang, C., Zhu, Y., Wu, H., Du, X., Liu, Z., and Feng, J. (2016). Diversity
685 of the Tetracycline Mobilome within a Chinese Pig Manure Sample. Applied and
686 environmental microbiology *82*, 6454–6462. https://doi.org/10.1128/AEM.01754-16.

687 47. Gabor, E.M., Alkema, W.B.L., and Janssen, D.B. (2004). Quantifying the accessibility of
688 the metagenome by random expression cloning techniques. Environmental Microbiology *6*,
689 879–886. https://doi.org/10.1111/j.1462-2920.2004.00640.x.

690 48. Van Schaik, J., Li, Z., Cheadle, J., and Crook, N. (2023). Engineering the Maize Root
691 Microbiome: A Rapid MoClo Toolkit and Identification of Potential Bacterial Chassis for
692 Studying Plant–Microbe Interactions. ACS Synth. Biol. *12*, 3030–3040.
693 https://doi.org/10.1021/acssynbio.3c00371.

694 49. Apjok, G., Számel, M., Christodoulou, C., Seregi, V., Vásárhelyi, B.M., Stirling, T.,
695 Eszenyi, B., Sári, T., Vidovics, F., Nagrand, E., et al. (2023). Characterization of antibiotic
696 resistomes by reprogrammed bacteriophage-enabled functional metagenomics in clinical
697 strains. Nat Microbiol *8*, 410–423. https://doi.org/10.1038/s41564-023-01320-2.

698 50. Lam, K.N., Martens, E.C., and Charles, T.C. (2018). Developing a Bacteroides System
699 for Function-Based Screening of DNA from the Human Gut Microbiome. mSystems *3*, 1–13.
700 https://doi.org/10.1128/mSystems.00195-17.

701 51. Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M., and Kishony, R.K.
702 (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. PLoS
703 ONE *10*, 1–15. https://doi.org/10.1371/journal.pone.0128036.

704 52. Rinke, C., Low, S., Woodcroft, B.J., Raina, J., Skarshewski, A., Le, X.H., Butler, M.K.,
705 Stocker, R., Seymour, J., Tyson, G.W., et al. (2016). Validation of picogram- and femtogram-
706 input DNA libraries for microscale metagenomics. PeerJ *4*, e2486.
707 https://doi.org/10.7717/peerj.2486.

708