

1 **Manuscript title:** Mosaic Ends Tagmentation (METa) assembly for extremely efficient
2 construction of functional metagenomic libraries

3

4 **Running title:** Functional metagenomic libraries made by METa assembly

5

6 **Authors:** Terence S. Crofts^{a,#}, Alexander G. McFarland^b, and Erica M. Hartmann^b

7 ^{a,#} Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, 60208,
8 USA

9 ^b Department of Civil and Environmental Engineering, Northwestern University, Evanston,
10 Illinois, 60208, USA

11 [#] To whom correspondence should be addressed. Email: tcrofts@northwestern.edu

12

13 **Abstract word count:** 221

14 **Text word count:** asd

15

16 **ABSTRACT**

17 Functional metagenomic libraries, physical bacterial libraries which allow the high-
18 throughput capture and expression of microbiome genes, have been instrumental in the
19 sequence-naïve and cultivation-independent discovery of novel genes from microbial
20 communities. Preparation of these libraries is limited by their high DNA input requirement and
21 their low cloning efficiency. Here, we describe a new method, METa assembly, for extremely
22 efficient functional metagenomic library preparation. We apply tagmentation to metagenomic
23 DNA from soil and gut microbiomes to prepare DNA inserts for high-throughput cloning into

24 functional metagenomic libraries. The presence of mosaic end sequences in the resulting DNA
25 fragments synergizes with homology-based assembly cloning to result in a 300-fold increase in
26 library size compared to traditional blunt cloning based protocols. Compared to published
27 libraries prepared by state-of-the-art protocols we show that METa assembly is on average 23- to
28 270-fold more efficient and can be effectively used to prepare gigabase-sized libraries with as
29 little as 200 ng of input DNA. We demonstrate the utility of METa assembly to capture novel
30 genes based on their function by discovering novel aminoglycoside (26% amino acid identity)
31 and colistin (36% amino acid identity) resistance genes in soil and goose gut microbiomes.
32 METa assembly provides a streamlined, flexible, and efficient method for preparing functional
33 metagenomic libraries, enabling new avenues of genetic and biochemical research into low
34 biomass or scarce microbiomes.

35

36 **IMPORTANCE**

37 Medically and industrially important genes can be recovered from microbial communities
38 by high-throughput sequencing but are limited to previously sequenced genes and their relatives.
39 Cloning a metagenome *en masse* into an expression host to produce a functional metagenomic
40 library is a sequence-naïve and cultivation-independent method to discover novel genes. This
41 directly connects genes to functions, but the process of preparing these libraries is DNA greedy
42 and inefficient. Here we describe a library preparation method that is an order of magnitude more
43 efficient and less DNA greedy. This method is consistently efficient across libraries prepared
44 from cultures, a soil microbiome, and from a goose fecal microbiome and allowed us to discover
45 novel antibiotic resistance genes. This new library preparation method will potentially allow for
46 the functional metagenomic exploration of microbiomes that were previously off limits due to

47 their rarity or low microbial biomass, such biomedical swabs or exotic samples.

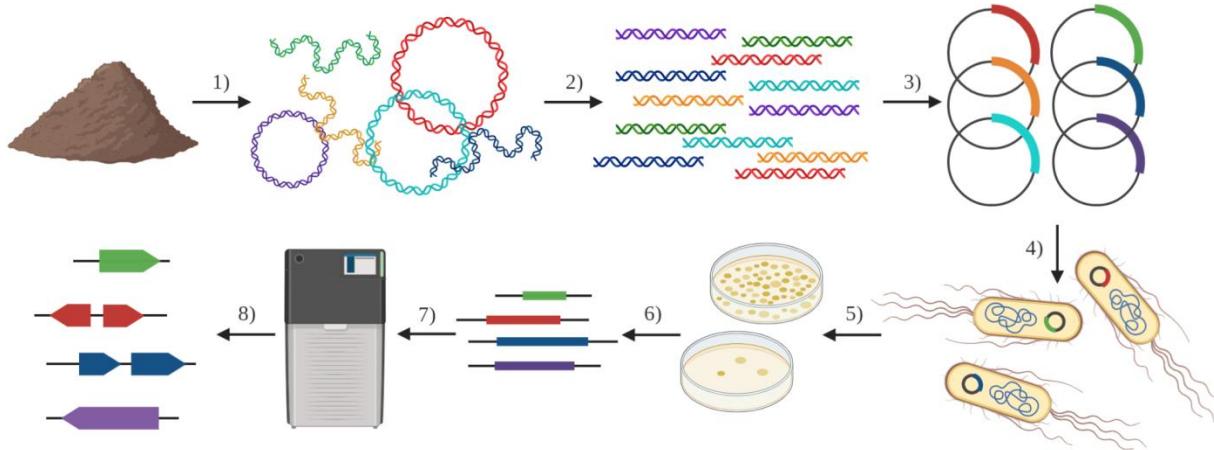
48

49 **INTRODUCTION**

50 The widespread adoption of high-throughput DNA sequencing technology has resulted in
51 a new and deserved appreciation for the genetic diversity present in microbial communities aka
52 microbiomes (1). Projects studying the chemical biology of microbiomes have shown that this
53 genetic diversity translates into enormous biochemical diversity (2–6). However, linking genes
54 from microbial community genetic material (the metagenome) to biochemical activity remains
55 difficult due to current limitations in gene prediction and annotation of specific activities. Direct
56 observation of biochemical function *in vitro* or phenotype *in vivo* remain the gold standards of
57 functional assignment as a result (7, 8).

58 One method that unites the culture- and sequence-independence of high throughput
59 sequencing with the functional observations that result from cloning and expression studies is
60 functional metagenomics. Functional metagenomics relies upon the construction of metagenomic
61 libraries in which a portion of a microbiome's metagenome is captured in a bacterial artificial
62 chromosome (BAC) or plasmid library and housed in an expression host, often *E. coli* (2)
63 (**Figure 1**). This technique allows function to be linked directly to genes without requiring
64 laboratory growth of the originating organisms or prior knowledge of target gene sequence.
65 Functional metagenomic libraries have been used to bioprospect for novel bioactive compounds
66 ((9) and reviewed in (3, 10)), novel enzymes of potential interest to industry (11, 12), and
67 enzymes useful in the production of biofuels (13, 14) (see also reviews (15–17)) and have been
68 created using metagenomic DNA from environments as varied as soils, adult and infant fecal
69 samples, sewage and waste-water effluent, and animal samples (18, 19, 28–35, 20–27). One

70 particularly successful application has been the identification of antimicrobial resistance genes
71 (7, 36) that would not have been identified by sequencing due to their low predicted amino acid
72 identity.



73
74 **Figure 1. Functional metagenomic library pipeline**

75 The general pipeline for the creation and use of functional metagenomic libraries to capture and
76 discover genes from metagenomes.

77 1) Extraction of metagenomic DNA from a microbiome (e.g. soil or fecal samples). 2)
78 Fragmentation of metagenomic DNA to desired size range (e.g. by sonication, restriction enzyme
79 digestion, or tagmentation). 3) Cloning of fragments into expression vectors following size
80 selection (e.g. by blunt ligation or homology-based assembly). 4) Transformation *en masse* of
81 vectors into an expression host (e.g. *E. coli* DH10B) to create functional metagenomic library. 5)
82 Functional selection or screen of library (e.g. on antibiotics to select for resistance). 6)
83 Amplification of selected inserts using vector-specific primers. 7) High-throughput sequencing
84 of selected metagenomic amplicons (e.g. by Illumina or PacBio technologies). 8) Annotation of
85 sequenced amplicons to link novel genes with selected/screened function (e.g. discovery of novel
86 aminoglycoside acetyltransferases). Figure created in BioRender.com.

87
88 The basic steps for creating a functional metagenomic library consist of metagenomic

89 DNA extraction, DNA fragmentation, cloning of fragments into a plasmid/BAC/etc., and
90 transformation of the plasmid library into an expression host (6, 36) (**Figure 1**). In many cases,
91 the libraries used in these experiments contain gene-sized inserts (*ca.* 1 kb to 10 kb) and make
92 use of sonication or other physical means to fragment metagenomic DNA into inserts followed
93 by blunt cloning into an expression vector (18, 21, 25, 26, 33, 37). These two steps, sonication
94 and blunt cloning (steps 2 and 3 in **Figure 1**), greatly lower the potential efficiency of functional

95 metagenomic library creation, necessitating multiple micrograms of input DNA mass.

96 Like functional metagenomic libraries, shotgun sequencing libraries have, until recently,

97 relied largely upon physical methods for DNA fragmentation and have similarly required

98 substantial input DNA mass. In contrast, transposase enzymatic DNA fragmentation (38, 39)

99 (tagmentation, known commercially as Nextera) produces DNA fragments using transposase

100 enzymes that create mostly random (40) double stranded breaks in target DNA by insertion of

101 their mosaic end sequence oligo cargo (38). This method substantially decreases costs and input

102 DNA mass requirements (38–43) but has not been applied yet to the preparation of functional

103 metagenomic libraries. We hypothesized that tagmentation reactions could replace sonication in

104 the preparation of functional metagenomic libraries (**Figure 2A**), decreasing input DNA

105 requirements by greater than 10-fold compared to current methods (**Figure 2B**). Tagmentation

106 also results in incorporation of 19 bp transposase oligos on the 5' ends of both strands of

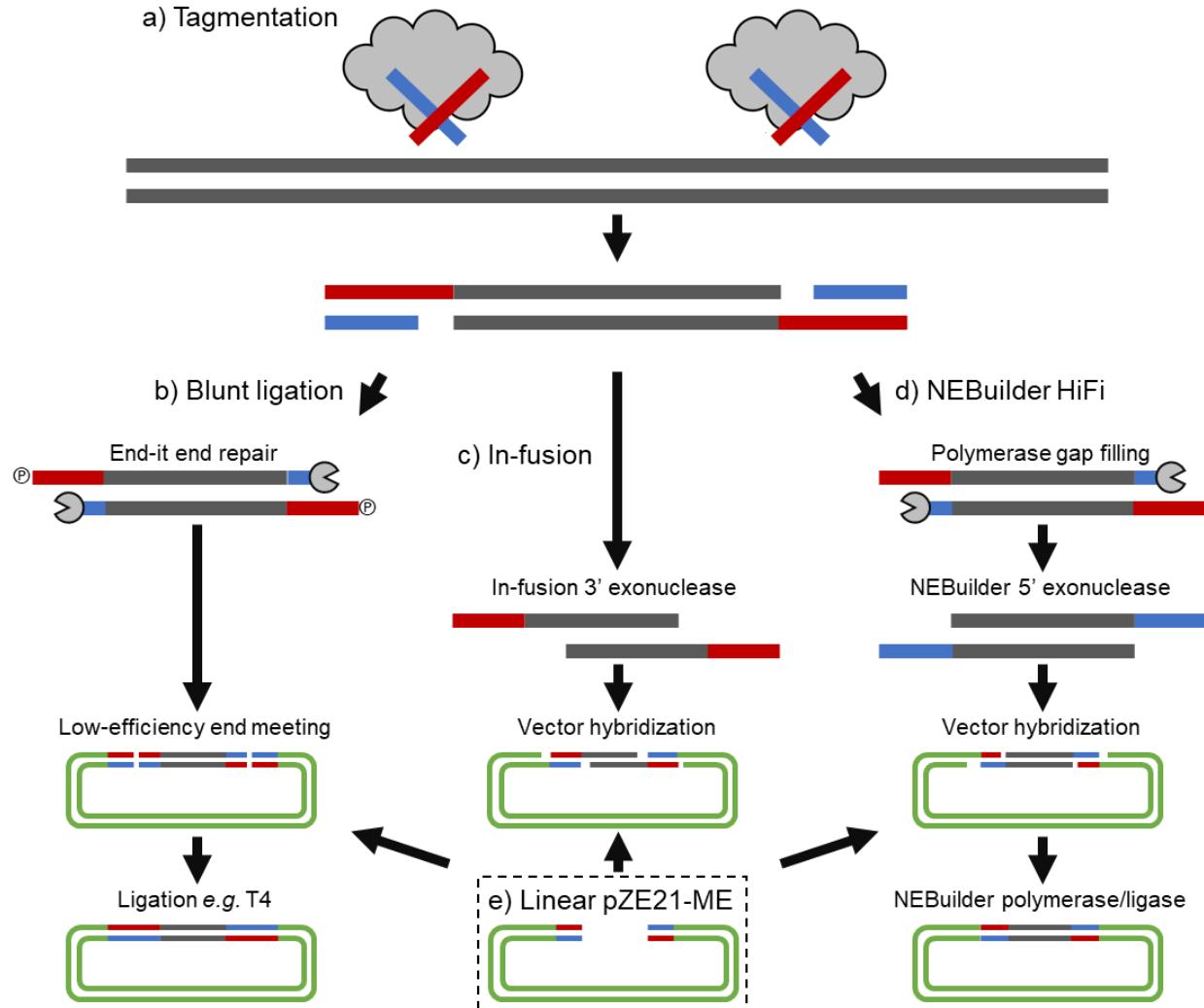
107 fragmented DNA. We hypothesized that incorporation of these sequences on the ends of inserts

108 could allow us to use homology-based DNA assembly protocols (*e.g.* Gibson assembly(44), *etc*)

109 in place of blunt ligation (**Figures 2C and 2D**). We hypothesized that incorporation of matching

110 sequences in an expression vector (**Figure 2E**) would allow the vector to capture inserts by

111 hybridization leading to significantly increased efficiency in library preparation.



112
113 **Figure 2. Blunt cloning protocol compared to METa assembly with NEBuilder HiFi or In-**
114 **Fusion**

115 a) Transposase enzyme fragments DNA with 5' mosaic end oligos. Inserts can be used as input
116 for all three methods. All three protocols are compatible with linear pZE21-ME vector prepared
117 by inverse PCR. b) Blunt cloning *via* end-repair and ligase. 5' overhangs must be resolved by
118 gap filling and phosphorylation using end-repair enzyme mixes. Blunt ended inserts can be
119 ligated to blunt ended vector. c) METa assembly *via* In-Fusion enzyme mix. In-Fusion 3'
120 exonuclease activity is directly compatible with transposase fragments. Single stranded DNA
121 overhangs on inserts and vector hybridize into a stable complex that can be transformed without
122 filling gaps or covalently sealing nicks. d) METa assembly *via* NEBuilder HiFi enzyme mix. 5'
123 overhangs must be resolved by DNA polymerase gap filling. NEBuilder HiFi enzyme mix
124 includes 5' exonuclease to create 3' overhangs which hybridize with target pZE21-ME. DNA
125 polymerase fills in gaps and ligase seals nicks. e) pZE21-ME is prepared and linearized by
126 inverse PCR and is compatible with all three DNA pipelines.

127
128 Here we report our testing of these hypotheses and the development, validation, and

129 application of a new general method for functional metagenomic library preparation that we are

130 calling Mosaic Ends Tagmentation assembly, or METa assembly. Our method takes advantage
131 of the unexpected synergy between tagmentation and assembly cloning to produce functional
132 metagenomic libraries with up to 270-fold more efficiency and 25-fold reduced input DNA mass
133 requirements compared to the current method for functional metagenomic library preparation.
134 METa assembly has the potential to greatly improve and expand the field bioprospecting,
135 catalyzing the discovery of novel microbial chemistry from genetically diverse microbiomes.
136

137 MATERIALS AND METHODS

138 Preparation and purification of transposase enzyme

139 Expression and purification of transposase enzyme was carried out based on
140 modifications of protocols published by Picelli *et al.* and Hennig *et al.* (41, 42). *E. coli* XL1 blue
141 carrying the pTXB1-Tn5 plasmid was a gift from Rickard Sandberg (Addgene plasmid #60240;
142 <http://n2t.net/addgene:60240>; RRID:Addgene_60240) and maintained as specified. The pTXB1-
143 Tn5 plasmid was recovered from an *E. coli* culture grown in LB supplemented with 100 µg/ml
144 carbenicillin (LB+CA100) via miniprep kit (New England Biolabs, cat#T1010S). The plasmid
145 was transformed into chemically competent *E. coli* BL21(DE3) cells (New England Biolabs,
146 cat#C2527I) following manufacturer recommendations and selected on LB+CA100 before
147 maintenance as a 15% glycerol stock at -80°C. A single colony of this strain was used to
148 inoculate 1 ml of LB+CA100 and incubated shaking overnight at 37°C. In the morning, the
149 saturated overnight culture was used to inoculate 1 L of Studier ZYM-5052 auto-induction media
150 (45) in a 2.8 L Fernbach flask. The culture was grown aerobically at 37°C until it began to turn
151 turbid by eye, approximately 3 hours, at which point the temperature was decreased to 20°C and
152 shaking maintained at 350 rpm overnight. Cells were collected by centrifugation at 8,000 rcf for

153 20 min at 4°C after OD₆₀₀ measurements suggested growth had plateaued at an OD₆₀₀ of ~3.72
154 AU. The resulting wet cell pellet weighed 8.08 g and a sample was analyzed by SDS-PAGE
155 (sodium dodecyl sulfate polyacrylamide gel electrophoresis) to verify induction of the ~75 kDa
156 Tn5-chitin binding domain fusion protein.

157 The cell pellet was resuspended to 10% w/v in HEGX buffer composed of 20 mM
158 HEPES buffer pH 7.2, 0.8 M NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 10% v/v
159 glycerol, and 0.2% v/v triton X-100 and supplemented with 20 µM phenylmethylsulfonyl
160 fluoride (PMSF) as a protease inhibitor. Resuspended cells were lysed on ice by sonication using
161 a W-225 sonicator with 6 cycles of 1 min on and 1 min off at output 5, 50% duty. Insoluble
162 debris was removed by centrifugation at 15,000 rcf for 30 min at 4°C following which 2.1 ml of
163 neutralized 10% polyethyleneimine (Millipore Sigma, cat#P3143) was added to the decanted
164 supernatant dropwise while stirring at 4°C. The precipitated *E. coli* genomic DNA was removed
165 by centrifugation for 10 min at 9,000 rcf at 4°C.

166 The fusion protein was purified from the clarified supernatants by adding 20 ml of chitin
167 resin (New England Biolabs, cat#S6651S) and incubating with gentle rotation overnight at 4°C.
168 The resin was washed with approximately 400 ml of HEGX buffer, following which the drained
169 resin was added to 30 ml of HEGX buffer supplemented with 100 mM β-mercaptoethanol to
170 initiate intein cleavage of the Tn5 transposase from the chitin binding domain. Cleavage was
171 allowed to proceed with gentle rotation at 4°C for approximately 48 hr after which transposase
172 was collected from the resin by draining and saving the flow-through. The resin was washed with
173 2X dialysis buffer consisting of 100 mM HEPES pH 7.2, 0.2 M NaCl, 0.2 mM EDTA, 20% w/v
174 glycerol, 0.2% triton X-100, and 2 mM dithiothreitol (DTT) and washes were pooled with the
175 initial elution. The pooled eluates were concentrated and exchanged into 2X dialysis buffer using

176 an Amicon Ultra 15 ml 3,000 molecular weight cut-off filter (Millipore Sigma,
177 cat#UFC900324). This concentrate contained 1.85 mg/ml protein as determined by BCA assay
178 (Thermo Scientific, cat#23225) in *ca.* 5.67 ml 2X dialysis buffer. After adding 6.2 ml of glycerol
179 and 1.89 ml 2X dialysis buffer enzyme stocks were stored at -20°C as 500 µl aliquots. Final
180 transposase stocks contained approximately 763 ng/µl protein.

181 **Verification of transposase activity**

182 Enzyme activity was verified by observing degradation of a soil metagenomic DNA
183 extract. Mosaic end primers 5Phos_METagA1 and METagA2 (**Supplemental table 1**)
184 synthesized by Integrated DNA Technologies (IDT Inc.) were brought to a concentration of 100
185 µM in 50 mM NaCl, 40 mM Tris pH 8. Annealing was carried out by combining 10 µl aliquots
186 of each oligo and incubating in a thermocycler using the following settings: 5:00 at 95°C, cool to
187 65°C at 0.1°C/sec, hold at 65°C for 5:00, and cool to 4°C at 0.1°C/sec. Annealed oligos were
188 maintained as aliquots at -20°C until use. Annealed mosaic ends oligos were loaded into
189 transposases by adding 0.143 volumes of annealed oligos to one volume of 763 ng/µl transposase
190 stock and incubating at room temperature (*ca.* 23°C) for 1 hr.

191 As a test substrate, metagenomic DNA was extracted from a soil sample taken from the
192 Northwestern University campus (coordinates 42.05662, -87.674247) using a DNeasy PowerSoil
193 Kit (Qiagen, cat#12888-100). Quantification of DNA for all experiments was performed using a
194 QuantiFluor ONE dsDNA system (Promega, cat#E4871). Test fragmentation reactions were
195 performed by combining MilliQ water (to 20 µl final volume), 4 µl of 5X TAPS-DMF buffer (50
196 mM TAPS buffer pH 8.5, 25 mM MgCl₂, 50% v/v dimethylformamide), 1 µl of 50 ng/µl soil
197 metagenomic DNA, and 1 µl of water (control) or 1 µl of loaded transposase (corresponding to
198 665 ng). Reactions were incubated in a thermocycler for 7 min at 55°C at which point reactions

199 were quenched by addition of 5 μ l of 0.2% SDS (final concentration 0.05% SDS) and incubation
200 at 55°C for 5 min. For each 25 μ l quenched reaction (+/- transposase), 12.5 μ l was purified using
201 a silica column-based kit (New England Biolabs, cat#T1030S, polymerase chain reaction (PCR)
202 purification kit used for all future enzymatic clean-up steps) and eluted with 12.5 μ l water while
203 the remaining 12.5 μ l were not purified further. Next, 2.5 μ l of 6X loading dye was added to both
204 the un-purified reactions and purified reactions and samples were analyzed on an agarose gel.
205 Unless otherwise noted, all agarose gel experiments were conducted using ~0.7% agarose pre-
206 cast with SYBRSafe (Invitrogen, cat#S33102) following manufacturer's recommendations.
207 Conversion of the high molecular weight DNA smear into a low molecular weight smear
208 confirmed active transposase.

209 **Tagmentation of high molecular weight DNA to produce 1 kb to 10 kb fragments**

210 High molecular weight DNA for use as a test transposase substrate was obtained from
211 cultures of *Pseudomonas* sp. Strain PE-S1G-1 (referred to as ABC07) and *Pandoraea* sp. Strain
212 PE-S2T-3 (referred to as ABC10) (46–48). Each strain was grown in 1 ml of LB+CA100 at 30°C
213 for 48 hours and genomic DNA was extracted using a Quick-DNA High Molecular Weight kit
214 (Zymo Research, cat#D6060) according to the manufacturer's protocol, quantified, and
215 combined to give a 100 ng/ μ l stock solution with equal input by mass from each genome.

216 Tagmentation reactions were assembled in 20 μ l volumes in 1X TAPS-DMF buffer
217 containing 200 ng of DNA and assembled transposase in concentrations ranging from 0 ng
218 enzyme per ng DNA to 2 ng enzyme per ng DNA. Following incubation and quenching (see
219 above) 6X loading dye was added to each reaction and samples were analyzed by pulsed field gel
220 electrophoresis at 4°C using a Pippin Pulse power supply (Sage Science, cat#PPI0200) running
221 pre-set protocol #4: 16 hr, 75 V on a 0.75% TAE agarose gel. Following electrophoresis, the gel

222 was stained with SYBRSafe and visualized under UV light.

223 **Preparation of pZE21-ME vector**

224 The pZE21 plasmid (49) was used as a template for an inverse PCR reaction to replace
225 the multiple cloning site with tandem mosaic end sequences of AGATGTGTATAAGAGACAG-
226 CTGTCTCTTATACACATCT. The construct was amplified by 2-step PCR according to
227 manufacturer recommendations using Q5 high fidelity polymerase (New England Biolabs,
228 cat#M0494S) with primers 6469TSC and 6470TSC (**Supplemental table 1**). The vector product
229 was sized and purified from an agarose gel and circularized by end repair/phosphorylation
230 (Lucigen, cat#ER0720) and blunt ended ligation (Lucigen, cat#LK0750H) following
231 manufacturer instructions. The construct was transformed into *E. coli* DH10B (New England
232 Biolabs 10-beta, cat#C3020K), selected on LB supplemented with 50 µg/ml kanamycin
233 (LB+KA50), and frozen at -80°C as a 15% glycerol stock. Incorporation of the mosaic ends sites
234 23 bp downstream of the pZE21 ribosome binding site was confirmed by Sanger sequencing.

235 **Comparison of METa assembly using In-fusion or NEBuilder HiFi assembly to blunt
236 ligation**

237 Triplicate assembly/ligation tests were carried out to determine the feasibility of using
238 mosaic end tags as targets for homology-mediated assembly of metagenomic cloning reactions.
239 Soil metagenomic DNA (see above) was used as input DNA in a bulk transposase reaction
240 containing 477.4 µl autoclaved MilliQ water, 124 µl freshly prepared 5X TAPS-DMF buffer,
241 15.5 µl of 290 ng/µl metagenomic DNA (7.25 ng DNA per µl final), and 3.1 µl of 665 ng/µl
242 assembled transposase (0.46 ng transposase per ng metagenomic DNA) in 620 µl total volume.
243 The bulk reaction was incubated at 55°C for 7 minutes and quenched by addition of 3.1 µl of
244 10% SDS (final concentration 0.05% SDS) followed by incubation at 55°C for 5 minutes. The

245 bulk reaction was purified and concentrated by PCR purification kit according to manufacturer's
246 directions for targeting fragments >2 kb in length. Tracking dye was added to the elution and the
247 entire sample was loaded onto an agarose gel and run at 70V for 120 minutes alongside a λ DNA
248 BstEII digest ladder (New England Biolabs, cat#N3014S). Prior to running, the electrophoresis
249 apparatus and gel tray were washed with milliQ water, soaked in 10% bleach for 15 minutes, and
250 rinsed with milliQ water. The bleaching process and use of λ DNA ladder are both necessary to
251 prevent contamination of the metagenomic DNA fragments by foreign DNA that can be
252 mistakenly incorporated during blunt ligation cloning (37). Fragments between ~1 kb and ~8 kb
253 were size selected by excision with a clean razor blade and DNA was purified from the agarose
254 by a silica column-based gel extraction kit (New England Biolabs, cat#T1020S, used in all gel
255 extraction steps) with ~500 mg of agarose used per purification column. For downstream cloning
256 reaction calculations, the average insert size was assumed to be 4 kb. Triplicate In-Fusion,
257 NEBuilder, and blunt ligation reactions all used this pool of inserts as their DNA inputs.

258 In-Fusion assembly reactions were prepared in triplicate containing 175 ng of DNA
259 fragments, 50 ng of pZE21-ME vector (a ~2:1 insert to vector ratio), 4 µl of 5X In-Fusion premix
260 (Takara Bio, cat#638909), and milliQ water up to 20 µl. Reactions were held at 50°C for 15
261 minutes then transferred to ice according to manufacturer's recommendations.

262 DNA for NEBuilder HiFi assembly reactions was prepared by combining 5 µl of Q5 high
263 fidelity polymerase previously heated to 98°C for 30 seconds with 5 µl of DNA fragments (600
264 ng) and incubated at 72°C for 15 minutes to fill in potential 5' overhangs. DNA fragments were
265 purified by silica column as before. Triplicate NEBuilder HiFi assembly reactions were prepared
266 according to manufacturer's protocols containing ~2:1 insert to vector as follows: 10 µl of 2X
267 NEBuilder HiFi master mix (New England Biolabs, cat#E2621S), 158.7 ng of DNA fragments,

268 45.2 ng pZE21-ME, and milliQ water to a volume of 20 μ l. Reactions were incubated at 50°C for
269 15 minutes then transferred on to ice.

270 Preparation of a metagenomic DNA library by blunt ligation was based on a modified
271 version of the Dantas lab protocol (21, 37). DNA fragments for blunt cloning were repaired
272 using an End-It DNA End-Repair Kit (Lucigen, cat#ER0720) according to the manufacturer's
273 protocol, with the single 50 μ l reaction containing 600 ng of DNA fragments, 5 μ l each of 10X
274 End-Repair buffer, dNTP mix, and ATP mix, and 1 μ l of End-Repair Enzyme mix in milliQ
275 water. The reaction was held at room temperature (*ca.* 23°C) for 45 minutes then purified by
276 silica column as before. Triplicate blunt ligation reactions were prepared with each reaction
277 containing 115 ng of end-repaired DNA, 21.9 ng of pZE21-ME (for a ~5:1 insert to vector ratio),
278 5 μ l of 2X Blunt/TA Ligase master mix (New England Biolabs, cat#M0367S), and milliQ water
279 up to 10 μ l total volume. Blunt ligation was carried out at room temperature (*ca.* 23°C) for 45
280 minutes, as suggested by the ligase manufacturer, then the reactions were transferred to ice.

281 All nine reactions (three techniques, each in triplicate) were de-salted and purified using
282 silica column kits for DNA fragments >2 kb and eluted in 10 μ l of 50°C milliQ water. *E. coli*
283 DH10B cells were made electrocompetent following a modified protocol (50) and tested to have
284 a moderate transformation efficiency of 4.7×10^8 cfu/ μ g of pUC19 DNA. Nine 45 μ l aliquots of
285 homemade competent cells stored at -80°C were thawed and, sequentially, combined with 10 μ l
286 of purified DNA in a 0.1 cm Gene Pulser cuvette (Bio-Rad, cat#1652083) and electroporated on
287 an Electroporator 2510 instrument (Eppendorf) at 1.8 kV with default settings of 10 μ F
288 capacitance and 600 Ω resistance (note, this results in a similar τ constant to standard settings on
289 other instruments of 25 μ F capacitance and 200 Ω resistance). Cells were immediately rescued
290 with 1 ml of room temperature (*ca.* 23°C) SOC media (Takara Bio) and recovered by incubating

291 with shaking for 1 hour at 37°C.

292 Following recovery, 50 µl of each transformation reaction was plated onto LB+KA50
293 agar at 50-fold, 500-fold, and 5,000-fold dilution (equivalent to plating 1 µl, 0.1 µl, and 0.01 µl
294 of the 1 ml recovery sample) and incubated overnight at 37°C. Colony counts from each titer
295 plate were taken and used to estimate the post-recovery population density as colony forming
296 units per ml (cfu/ml). To estimate average insert size and the rate of successful insert capture,
297 individual colonies were picked into 120 µl of milliQ water and used as template for 15 µl
298 colony PCR as follows: 0.3 µl each of 10 µM primers 6463TSC and 6464TSC (**Supplemental**
299 **table 1**), 1.5 µl of colony suspension, 7.5 µl of OneTaq Quick-Load 2X master mix (New
300 England Biolabs, cat#M0486), and milliQ water to 15 µl. The reactions were incubated in a
301 thermocycler with the following program: 5:00 at 94°C followed by 25 cycles of 0:30 at 94°C,
302 0:45 at 62°C, and 8:00 at 68°C, followed by 5:00 at 68°C and storage at 4°C. Reactions were
303 analyzed on agarose gel and average insert size was calculated by comparison to a DNA ladder.
304 The proportion of colonies with an insert was estimated by taking the proportion of reactions
305 returning a >500 bp product (500 bp being the expected amplicon size for a vector backbone
306 only reaction) over the total number of successful reactions. The library size for each assembly
307 or cloning reaction was calculated using the following equation (cfu is calculated from cfu/ml *
308 1 ml total recovery volume):

$$\text{Library size in Gb} = \frac{\text{cfu} * \text{proportion of colonies with inserts} * \text{average insert size}}{1E9 \text{ bp per Gb}}$$

309 Following library size estimation, each library was normalized to the quantity of insert DNA
310 used for the cloning step (*i.e.* In-Fusion, NEBuilder HiFi, or blunt ligation step) for direct
311 comparison across techniques in the form of Gb library/ng insert DNA.

312 **Repeat comparison of NEBuilder HiFi METa assembly to blunt ligation**

313 To confirm the efficacy of the NEBuilder HiFi METa assembly protocol, we prepared
314 triplicate libraries by METa assembly or blunt ligation using as input mixed genomic DNA from
315 *Pseudomonas* sp. PE-S1G-1 and *Pandoraea* sp. PE-S2T-3 (prepared above). A 1 ml
316 fragmentation reaction consisting of 1X TAPS-DMF buffer, 10 µg mixed genomic DNA (10 ng
317 input DNA per µl final volume), and 5 µg assembled transposase (0.5 ng of transposase per ng of
318 input DNA) was incubated, quenched, concentrated by PCR purification kit, and size selected
319 and purified from agarose gel as before, targeting DNA fragments between ~1 kb and ~8 kb.

320 NEBuilder HiFi assembly input DNA fragments were modified as before in triplicate
321 reactions (PCR overhang filling by Q5 polymerase). Inserts for blunt ligation cloning were also
322 modified in triplicate reactions (end repair by End-It kit). For both sets of triplicate reaction each
323 replicate used 300 ng of size-selected DNA fragments as input. NEBuilder HiFi assembly was
324 performed as before, with triplicate reactions containing 40 ng of inserts and 20 ng of pZE21-ME
325 linear vector, and parallel triplicate sham reactions containing milliQ water in place of inserts.
326 Blunt ligation reactions were prepared to follow established functional metagenomic library
327 cloning protocols (21, 37). Triplicate reactions were prepared containing 1.5 µl of Fast-Link 10X
328 ligation buffer, 0.75 µl ATP solution (10 mM), 100 ng of insert fragments, 40 ng of linearized
329 pZE21-ME, 1 µl of Fast-Link DNA ligase (Lucigen, cat#LK0750H), and milliQ water up to 15
330 µl total volume. In parallel, another triplicate set of sham reactions were prepared with milliQ
331 water replacing the corresponding volume of insert DNA. Blunt ligation reactions were
332 incubated at room temperature (*ca.* 23°C) overnight then heat inactivated at 70°C for 15 minutes.

333 All 12 reactions (two techniques with triplicate insert reactions and triplicate sham
334 reactions each) were purified and de-salting by silica column kit. For each reaction, the entire 10
335 µl milliQ water elution was electroporated into a 25 µl aliquot of commercial 10-beta

336 electrocompetent *E. coli* DH10B cells (advertised transformation efficiency of 2×10^{10} cfu/ μ g
337 DNA), immediately rescued in 1 ml of 37°C SOC outgrowth medium (New England Biolabs,
338 cat#B9020S) and incubated with shaking at 37°C for 1 hour. Following recovery, 100 μ l of 100-
339 fold (10^2), 10,000-fold (10^4), and 1,000,000-fold (10^6) diluted cultures were plated onto
340 LB+KA50 plates overnight at 37°C. The remaining stocks of the sham reactions were discarded,
341 while the remaining stocks of the insert reactions were individually used to inoculate 50 ml of
342 LB+KA50 media and shaken O/N at 18°C before being transferred to a 37°C shaking incubator
343 the next day to amplify the libraries to an OD₆₀₀ of between 0.6 and 1.0 AU.

344 Library size was estimated as above, by counting the titer plates to find the number of
345 colony forming units present post-electroporation and by using colony PCR to determine the
346 proportion of cells containing an insert and to determine the average insert size. The libraries
347 amplified in 50 ml of LB+KA50 were concentrated by centrifugation at 4,000 rcf for 7 minutes
348 and resuspended to 10 ml in 15% glycerol in LB+KA50 media. The concentrated libraries were
349 aliquoted 1 ml at a time into cryovials for storage at -80°C.

350 **Assembly of a large soil microbiome metagenomic library by METa assembly**

351 The previously extracted and purified soil metagenomic DNA was used as input for a 5
352 μ g fragmentation reaction performed in 1X TAPS-DMF buffer with 10 ng/ μ l final metagenomic
353 DNA concentration and 0.5 ng of annealed transposase per ng of metagenomic DNA. The
354 reaction was incubated at 55°C for 7 minutes then quenched for 5 minutes at 55°C by adding
355 SDS to 0.05%. The reaction was purified and concentrated by silica column clean-up as before
356 and eluted twice with 6 μ l of 55°C elution buffer. To the full elution was added 1 μ l of 6X
357 tracking dye and 2 μ l of glycerol before loading onto an agarose gel and running at 75V for 2 hr
358 alongside a BstEII digest of λ DNA ladder. Fragments were size selected by excision from

359 agarose gel, targeting inserts between 1 kb and 6 kb in length and purified by silica column and
360 eluted twice with 25 µl elution buffer heated to 55°C. Fragments overhangs were filled by PCR
361 with 50 µl of Q5 2X master mix for 15 min at 72°C, purified by silica column, eluted twice with
362 6 µl of 55°C water and quantified. Out of a total of 1.122 µg of available purified inserts two
363 libraries were assembled. The first test library used 175 ng of inserts (~0.14 pmol assuming 2 kb
364 average size) assembled into 100 ng of pZE21-ME (~0.07 pmol) using NEBuilder Hifi enzyme
365 mix as before in a 20 µl total volume. The second library used the remaining inserts in a scaled
366 up reaction volume of 54 µl total maintaining the recommended 0.2 pmol of total fragments per
367 20 µl of reaction. The assembly reactions were de-salted by silica column purification, eluted
368 twice with 6 µl 55°C water (the 175 ng assembly) or eluted twice with 25 µl 55°C water (the bulk
369 reaction), and electroporated into 25 µl (the 175 ng assembly) or 125 µl (the bulk assembly) of
370 NEB 10-beta electrocompetent *E. coli* cells at 1.8 kV. The libraries were amplified overnight in
371 LB+KA50 and stored in 1 ml glycerol stocks as before. Library sizes were quantified by colony
372 counting and colony PCR. Colony PCR was performed using the high processivity polymerase
373 Q5, with each 10 µl colony PCR reaction containing 0.5 µl each of primers TSC6463 and
374 TSC6464, 5 µl of 2X Q5 polymerase, 4 µl of water. Template consisted of a sterile toothpick
375 touched to a colony and dipped into the reaction solution. PCR reactions were carried out by
376 heating to 98°C for 3:00, followed by 25 cycles of 98°C for 0:10 and 72°C for 4:00, followed by
377 holding at 72°C for 5:00.

378 **METa assembly of soil or goose fecal metagenomic libraries using limited input DNA**

379 A library to test the limits of METa assembly was prepared using a 10 kb DNA amplicon
380 as input. To generate the input DNA for this assembly a 50 µl Q5 PCR reaction using template
381 DNA and primers from a Phusion HiFi amplification control (Thermo Scientific, cat#F553S)

382 was run as follows: 25 µl of Q5 2X hot start master mix, 6.25 µl of 4 µM primers P1 and P2 and
383 1 µl of template DNA (Thermo Scientific, cat#F553S), and milliQ water up to 50 µl.
384 Amplification was carried out according to manufacturer's protocols with a T_A of 65°C. The size
385 and purity of the 10 kb product was verified by agarose gel electrophoresis, purified by silica
386 column and used as substrate in a 200 ng, 20 µl tagmentation reaction (10 ng DNA per µl
387 reaction, 0.5 ng assembled Tn5 per ng input DNA). Size selection, overhang filling, and
388 purification were performed as before. Fragments (59.4 ng) were assembled into pZE21-ME
389 vector (30 ng) by NEBuilder HiFi assembly master mix (total volume 20 µl), purified, and
390 transformed into electrocompetent 10-beta *E. coli* cells to produce a library for quantitation by
391 colony counting and colony PCR.

392 A soil functional metagenomic library was prepared as above, with 250 ng of previously
393 extracted soil metagenomic DNA used as input in a 25 µl tagmentation reaction. The reaction
394 was purified by silica column and fragments between 2 kb and 6 kb in length were purified by
395 extraction from agarose gel. Overhang filling, purification and assembly by NEBuilder HiFi
396 master mix (35.2 ng of inserts, 21 ng of linearized pZE21-ME, 20 µl total volume) were
397 performed as before followed by assembly purification and transformation into electrocompetent
398 10-beta *E. coli*. The library was quantified by colony counting and colony PCR and amplified
399 overnight as above followed by storage at -80°C in 10 x 1 ml aliquots.

400 To prepare a goose fecal microbiome library a freshly voided adult Canada goose (*Branta*
401 *canadensis*) (**Supplemental figure 11A**) fecal pellet was collected from the Northwestern
402 University campus (coordinates 42.056084, -87.670661). Fecal pellet collection was approved by
403 the Northwestern University Institutional Animal Care and Use Committee (IACUC) under
404 protocol EC20-0252. Within 15 minutes of collection, 474 mg of fecal pellet was used as input

405 for metagenomic DNA extraction using a DNeasy PowerSoil Kit (Qiagen, cat#12888-100)
406 following modifications for goose microbiome extraction suggested by *Cao et. al* 2020 (51).
407 Briefly, these consisted of incubating the sample suspended in Qiagen buffer CD1 at 65°C for 10
408 minutes followed by incubation at 95°C for 10 min followed by the kit manufacturer's protocol.

409 Tagmentation was performed as above (10 ng input DNA per μ l volume, 0.5 ng of
410 assembled transposase per ng of input DNA) using 300 ng of input DNA. The tagmentation
411 reaction was quenched, purified, and loaded onto a 0.6% agarose gel as before. DNA fragments
412 between *ca.* 1.7 kb and 6.3 kb were collected and purified by gel extraction as above, eluting
413 twice with 6 μ l of 55°C water. To the 12 μ l elution was added 12 μ l of 2X Q5 master mix
414 previously held at 98°C for 30 seconds and the gap-filling reaction was held at 72°C for 15
415 minutes before purification and elution, resulting in 60.5 ng of blunt ended DNA fragments. All
416 60.5 ng of insert DNA (~0.05 pmol assuming 2 kb average size) were added to a 20 μ l
417 NEBuilder HiFi assembly reaction with 35 ng of linear pZE21-ME vector (~0.025 pmol) and
418 held at 50°C for 15 minutes followed by column purification and elution twice with 6 μ l of 55°C
419 water. The purified assembly reaction was transformed into electrocompetent *E. coli*, rescued,
420 amplified, titered for cfu count, and subjected to colony PCR to determine average insert length
421 and library size as above.

422 **Extraction of plasmid DNA from un-selected libraries**

423 Functional metagenomic library stocks prepared from strains ABC07 and ABC10 (see
424 above) were plated in triplicate on LB+KA50 agar to determine cfu/ml concentrations. Each
425 triplicate library was plated again on LB+KA50 agar plates with volumes calculated to result in
426 ~1000 colonies plated, resulting in an average of ~600 colonies per plate following overnight
427 incubation at 37°C. Colonies on each plate were collected by addition of 750 μ l of LB to the

428 plate and resuspending colonies using a cell spreader and removal of the media. This process was
429 repeated, resulting in ~1 ml of turbid bacterial culture which was subsequently used as input for
430 plasmid purification by miniprep kit (New England Biolabs, cat#T1010S). The pooled plasmid
431 library DNA for each library was eluted in 35 µl of elution buffer and quantified, resulting in an
432 average concentration of 58 ng/µl miniprep.

433 **Functional metagenomic selection for antibiotic resistance**

434 Libraries to be tested were first plated on LB+KA50 agar plates to determine cfu/ml titer
435 following which the volume of frozen stock necessary to provide 10-fold coverage of unique
436 inserts in each library was calculated. This volume, brought up to 100 µl with LB media if
437 necessary, was plated on Mueller-Hinton II cation adjusted agar plates containing 50 µg/ml
438 kanamycin (MH+KA50) and another selective antibiotic depending on the library. Triplicate
439 ABC07/ABC10 libraries prepared by METa assembly or blunt ligation were plated on
440 MH+KA50 supplemented with 1,000 µg/ml (1 mg/ml) penicillin G sodium salt (Fisher
441 Scientific, cat#AAJ6303214). The 162 Gb soil library was plated on MH+KA50 supplemented
442 with 64 µg/ml nourseothricin sulfate (Dot Scientific, cat#DSN51200-1) and the 300 ng goose
443 fecal pellet library was plated on MH+KA50 supplemented with either 8 µg/ml tetracycline
444 (Fisher Scientific, cat#AAJ6171406) or 4 µg/ml colistin sulfate (Fisher Scientific,
445 cat#AAJ6091503). Following overnight selection of plates at 37°C resistant colonies were
446 collected as slurries and plasmids extracted and purified as above. Antibiotic concentrations were
447 chosen based on literature precedent (*i.e.* 1 mg/ml penicillin (47, 48), 8 µg/ml tetracycline (37), 4
448 µg/ml colistin (35), and 64 µg/ml nourseothricin (52)). Complete growth inhibition of *E. coli*
449 DH10B with empty pZE21-ME vector was confirmed for each antibiotic after plating a similar
450 high density lawn and incubating at 37°C overnight.

451 **Amplicon sequencing of functional metagenomic libraries**

452 Plasmid minipreps from un-selected and antibiotic selected libraries were used as
453 templates for PCR reactions targeting vector inserts. Seven 100 µl PCR reactions, one for each
454 library, were performed containing 50 µl of Q5 2X master mix, 39 µl water, 5 µl each of 10 µM
455 primers 6463TSC and 6464 TSC, and 1 µl of library miniprep corresponding to between 5.4 ng
456 and 8.6 ng of DNA. Reactions were run in a thermocycler using the following settings: holding
457 at 98°C for 30 seconds followed by 16 cycles of 98°C for 10 seconds and 72°C for 4 minutes
458 followed by holding at 72°C for 5 minutes. Reactions were purified by silica column and each
459 column was eluted twice with 20 µl of 55°C water and quantified. Insert amplicon integrity was
460 verified by running 100 ng of purified DNA from each reaction on a 0.8% agarose gel stained
461 with SYBRSafe.

462 Samples were shipped overnight to the University of Illinois at Urbana-Champaign Roy
463 J. Carver Biotechnology Center as 30 µl aliquots containing 500 ng of each reaction. Samples
464 were used as input for library preparation and sequencing on the PacBio Sequel II platform at the
465 center as follows. Amplicons were ligated to barcoded adaptors using the Barcoded Overhang
466 Adaptors Kit (Pacific Biosciences, CA). The barcoded amplicons were normalized to the
467 estimated number of unique inserts, based off of colony counts, and pooled. The pooled
468 amplicons were used as input for a SMRTBell Express Template Prep kit 2.0 (Pacific
469 Biosciences) to prepare the sequencing library. The library was quantitated by Qubit fluorometer,
470 and DNA fragment size and quality were confirmed on a Fragment Analyzer (Agilent, CA). The
471 library was sequenced on a SMRTcell 8M on a PacBio Sequel II with 20 hs movie time. Circular
472 consensus analysis was performed on the resulting BAM file using SMRTLink V8.0 using the
473 following parameters: ccs --min-length 500 --max-length 12000 --min-passes 3 --min-rq 0.99.

474 Demultiplexing was performed with lima (Pacific Biosciences) using default parameters.

475 **Analysis of functional metagenomic library sequencing**

476 Read coverage and evenness of the ABC07/ABC10 libraries were determined for each
477 genome individually. Long reads were mapped to either ABC07 (ASM217990v1) or ABC10
478 (ASM217996v1) assemblies using minimap2 v2.17-p94 with default settings, converting
479 resulting SAM files to BAM files with samtools v1.9 (53) using default settings, and calculating
480 the average read coverage over a 1,000 bp window using pileup.sh from the BBmap v38.86
481 suite. For visualization of coverage, a rolling average was produced using the rollmean()
482 function from the Zoo R package v1.8.8 (54).

483 To identify potential β -lactamase genes in the ABC07 (ASM217990v1) and ABC10
484 (ASM217996v1) genomes the amino acid fasta files were downloaded and individually

485 submitted to the Comprehensive Antibiotic Resistance Database (CARD) Resistance Gene
486 Identifier (RGI) webserver (<https://card.mcmaster.ca/analyze/rgi>) (55) for prediction of all
487 resistance genes using parameters ‘Protein sequence’, ‘Perfect, Strict and Loose’, ‘Include
488 nudge’, and ‘High quality/coverage’. The resulting dataframe was filtered for antibiotic
489 resistance genes belonging to the β -lactamase AMR Gene Family that had ‘antibiotic
490 inactivation’ as its resistance mechanism to generate a list of ABC07 or ABC10-specific β -
491 lactamase amino acid sequences. For each functional metagenomic approach and selection
492 method, the following was performed on all reads that mapped to either ABC07 or ABC10.

493 Open reading frames (ORFs) were identified using Prokka v1.14.6 (56) with parameters –norrna,
494 --notrna, --noanno, and --fast. Resulting ORFs for each combination were clustered with
495 mmseqs2 linclust (57) with parameters --min-seq-id 0.95 and -c 0.95. Predicted β -lactamases
496 were searched against the set of representative ORFs using an e-value cutoff of 10^{-6} . Afterwards,

497 a single best hit determined by bitscore, identity, and coverage was kept. A hit was considered a
498 true β -lactamase if it had $\geq 95\%$ coverage and identity.

499 Functional annotation of the soil and goose fecal microbiome libraries was carried out by
500 ORF searching and clustering all metagenomic long reads using Prokka and mmseqs2 linclust as
501 previously described. Representative ORFs were submitted to the CARD RGI portal to search
502 for ARGs using the previously stated parameters. Non-resistance functional annotations were
503 obtained by submitting representative ORFs to eggNOG-mapper v2 ([http://eggnog-](http://eggnog-mapper.embl.de/)
504 [mapper.embl.de/](http://eggnog-mapper.embl.de/)) (58). Antibiotic resistance genes and functional annotations were then mapped
505 back to all identified ORFs using a custom script. Counts of antibiotic resistance genes, antibiotic
506 resistance gene ontology, superfamilies, and drug class were extracted from the CARD RGI
507 output.

508 Mobile genetic elements syntenic to antibiotic resistance genes were identified based on
509 keyword searches of the eggNOG ‘annotation’ column using the search terms “transposase”,
510 “conjugative”, “phage”, “integrase”, “replication”, and “recombinase”. ORFs that matched these
511 search terms were further verified using UniProt (<https://www.uniprot.org/>) (59). Plots were
512 made using R package ggplot v3.3.2 (60) and gene region visualizations were created using R
513 package ggenes v0.4.0.

514 Streptothricin acetyltransferase gene family phylogenetic analysis was performed by
515 extracting the five amino acid sequences for streptothricin resistance enzymes from the CARD
516 database and using these as input to run NCBI blastp (61) against the NR database (run
517 November, 2020). Sequence hits were filtered for $\geq 25\%$ amino acid identity and $\geq 70\%$
518 alignment length to match CARD workflow. Replicates were removed from the combined
519 sequences followed by clustering using CD-Hit (62) at a 90% identity threshold. The amino acid

520 sequences of three representative putative streptothricin acetyltransferase enzymes from the
521 nourseothricin selection were added to the sequence file and all were aligned using mafft v7.471
522 (63) with parameters –thread 8 –localpair –maxiterate 1000. An approximate maximum
523 likelihood phylogenetic tree was calculated using FastTree v2.1.11 (64) with parameter -wag.
524 The resulting phylogenetic tree was visualized with ggtree (65).

525 **Literature search for comparable library statistics**

526 A literature search was carried out on the National Library of Medicine using PubMed
527 using the search terms “functional metagenomics” OR “metagenomic libraries” on September
528 30th, 2020. The results were sorted by publication date and the 125 most recent publications were
529 manually searched for functional metagenomic library preparation details including insert size
530 (small or large), input DNA mass, and total library size. An additional three publications known
531 to contain these details were appended to the 125 publications to give a total of 128 publications
532 searched. Off-topic publications, publications that do not describe the preparation of new
533 libraries, and publications with insufficient methods details were removed leaving six suitable
534 publications: Campbell *et al.* (2020), Gasparrini *et al.* (2019), Kintses *et al.* (2019), Marathe *et*
535 *al.* (2018), Gibson *et al.* (2016), and Moore *et al.* (2013) (21, 28, 35, 66–68). Library efficiency
536 was determined by normalization of reported library size to reported or best estimate input DNA
537 mass.

538

539 **RESULTS**

540 **Strategies to improve functional metagenomic library preparation efficiency**

541 Our first strategy to increase functional metagenomic library preparation efficiency was
542 to replace sonication-based fragmentation with transposase-based fragmentation (**Figure 2A**). In

543 theory, this would allow for lower input DNA mass and remove expensive capital equipment
544 from the blunt ligation cloning protocol (**Figure 2B**).

545 Our second strategy to increase efficiency was to replace blunt ligation cloning with
546 homology-based seamless assembly cloning, taking advantage of the fact that tagmentation-
547 produced fragments would all have known DNA sequence on their ends. One commercial
548 assembly protocol that we hypothesized to be compatible with tagmented DNA fragments was
549 In-Fusion cloning from Takara (69). This method is optimized for overlap regions with lengths
550 between 12 bp and 21 bp and is therefore compatible with homology consisting of 19 bp mosaic
551 end sequences. In-Fusion uses 3' exonuclease activity to produce single stranded DNA 5'
552 overhangs, allowing treated inserts and vector to hybridize. In-fusion assembly does not result in
553 a covalently closed circular DNA product but instead retains gaps and nicks that are resolved
554 post-transformation in the recipient cell (**Figure 2C**).

555 A second assembly option that we hypothesized to be compatible was NEBuilder HiFi
556 DNA assembly from New England Biolabs, which functions similarly to Gibson assembly (44).
557 This method requires overlap regions with melting temperatures greater than 50°C which is
558 compatible with the 19 bp mosaic end sequence (an estimated melting temperature of 52°C by
559 2AT+4GC rule). NEBuilder HiFi and Gibson assembly use 5' exonucleases to produce 3'
560 overhangs. Because tagmentation results in covalent addition of mosaic end oligos to only the 5'
561 ends of DNA fragments, 5' exonuclease activity would effectively erase the mosaic end
562 sequence from the inserts. Nextera tagmentation protocols overcome this obstacle by including a
563 brief DNA polymerase gap filling reaction that would be applicable to our protocol as well. The
564 resulting DNA would be mixed with the NEBuilder HiFi assembly master mix, resulting in 3'
565 overhangs able to hybridize to complementary sequences on linearized vector. Following

566 hybridization, NEBuilder HiFi assembly master mix polymerase and ligase fill gaps and ligate
567 nicks, respectively, to produce a covalently sealed construct for transformation (**Figure 2D**).

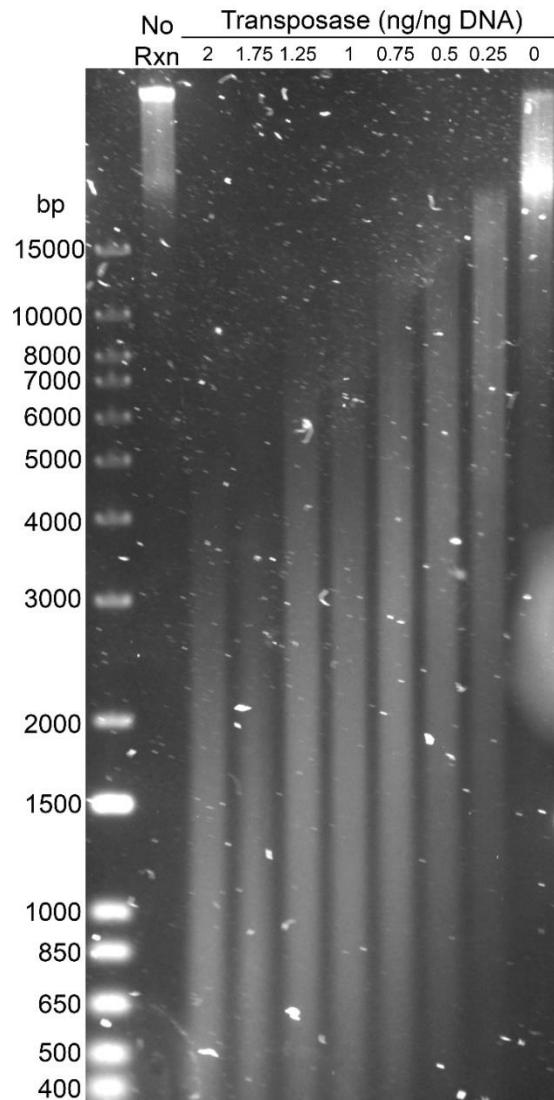
568

569 **Expression and purification of transposase enzyme and 1 kb to 10 kb DNA tagmentation**

570 We began our experiments with the goal of increasing DNA efficiency during
571 metagenomic library preparation by adapting the tagmentation work flow used in shotgun
572 sequencing library protocols (38, 39). This was aided by the availability of non-commercial
573 protocols for the preparation of the transposase reagent (41, 42). We modified these published
574 protocols for expression and purification of Tn5 transposase enzyme, most notably in our use of
575 auto-induction media (45) and other expression conditions. Our expression system responded
576 well to auto-induction (**Supplemental figure 1A**) with one liter of culture yielding 8 g of wet
577 cell mass and we obtained approximately 10.5 mg of well purified enzyme (**Supplemental**
578 **figure 1B**) in good agreement with reported yields of 15 mg per liter of culture (41). We verified
579 the activity of our transposase by observing its ability to convert metagenomic DNA into low
580 molecular weight fragments <500 bp in size (**Supplemental figure 1C**).

581 Next, we investigated if transposase could be used to prepare metagenomic or genomic
582 DNA in fragments capable of containing bacterial open reading frames (ORFs, roughly 1 kb to 5
583 kb in length). We used high molecular weight genomic DNA (measured at ~70 kb) isolated from
584 two penicillin-catabolizing bacteria, ABC07 (*Pseudomonas* sp. PE-S1G-1) and ABC10
585 (*Pandoraea* sp. PE-S2T-3) (46–48) as input for these test reactions. While tagmentation is
586 usually used to create ~200 bp DNA fragments for sequencing on the Illumina platform we were
587 able to alter this by adjusting the ratio of transposase to input DNA (similarly, others have
588 successfully used another enzyme used to prepare shotgun sequencing libraries, fragmentase, to

589 make functional metagenomic libraries (67)). Using 0.5 ng of transposase per ng of target DNA
590 yielded a fragmentation pattern centered around 2.5 kb (**Figure 3**).



591
592 **Figure 3. Transposase to target DNA ratios for 1 kb to 10 kb insert fragmentation**
593 High molecular weight genomic DNA (no rxn = no reaction) was incubated with transposase at
594 concentrations from 0 ng/ng of DNA up to 2 ng/ng DNA and analyzed by pulsed field agarose
595 gel electrophoresis.

596
597 **METa assembly using In-fusion or NEBuilder HiFi assembly compared to blunt ligation**

598 We hypothesized that inserts prepared by tagmentation could be compatible with both
599 blunt ligation and assembly-based methods of functional metagenomic library preparation. To
600 test this hypothesis, we extracted high molecular weight DNA from a soil sample and

601 fragmented it by tagmentation, resulting in fragments between approximately 1 kb and 8 kb in
602 length (**Supplemental figure 2**). We purified the fragmented DNA by excision from an agarose
603 gel and used the resulting pool of mosaic end 5' tagged fragments as input for three library
604 preparation methods, described above (**Figure 2**), in triplicate. DNA fragment preparation began
605 with a single aliquot of 4.5 µg of metagenomic DNA which dropped to 3.175 µg following
606 tagmentation and PCR purification (~30% loss), and dropping to 2.145 µg of DNA following
607 size selection and gel purification (~30% step loss, ~50% loss overall). After cloning inserts into
608 vector (by enzymatic assembly or ligation), we transformed the entirety of each purified reaction
609 by electroporation into homemade competent *E. coli* cells. We plated dilutions of the recovered
610 cells to determine titer. The following day, we used colony PCR to find the average insert size
611 and proportion of colonies with an insert (as opposed to empty vectors) (**Supplemental figure**
612 **3**).

613 METa assembly using In-Fusion premix (**Figure 2C**) resulted in recovery of only a
614 single colony on a 100-fold dilution plate and we did not pursue this strategy any further. We
615 instead focused on comparing libraries created *via* blunt ligation to libraries created by METa
616 assembly with NEBuilder HiFi. Blunt ligation cloning (**Figure 2B**) resulted in, on average,
617 dozens of colonies on agar plates spread with 100-fold diluted transformation culture
618 corresponding to 1.6×10^4 cfu, 4.6×10^4 cfu, and 5.2×10^4 cfu for each replicate. In comparison,
619 NEBuilder HiFi mediated METa assembly (**Figure 2D**) agar plates had uncountable colonies at
620 100-fold dilutions and ~100 colonies on petri dishes plated with 10,000-fold dilutions,
621 corresponding to 1.3×10^6 cfu, 9.8×10^5 cfu, and 4.9×10^5 cfu for each replicate. When we
622 normalized these plate counts to how much insert DNA was used during triplicate cloning
623 reactions we found that the NEBuilder HiFi reactions resulted in ~17.5-fold higher cfu/ng

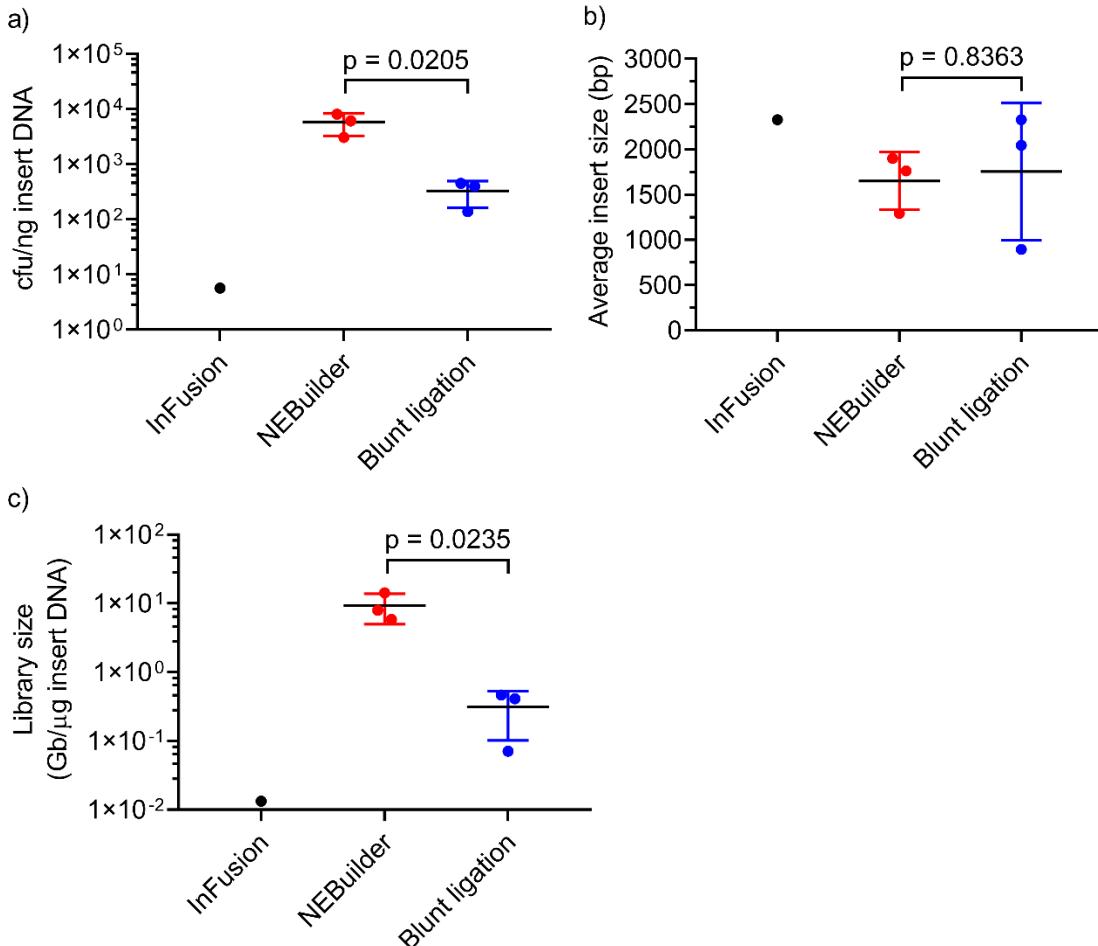
624 assemblies and an associated *p*-value of 0.0205 (**Figure 4A**). Average insert length did not
625 appear to differ between the two methods, with the average NEBuilder HiFi assembly insert
626 being 0.94 times as large as the average blunt ligation insert (**Figure 4B**). NEBuilder HiFi
627 assembly reactions resulted in significantly larger libraries (~30-fold larger, *p* = 0.0235) (**Figure**
628 **4C**), in part because of their higher titer (**Figure 4A**), but also because all of the colonies assayed
629 contained inserts whereas approximately half of the blunt ligation colonies carried empty vectors
630 or insert DNA <200 bp in length despite both methods using the same linearized vector DNA
631 (**Supplemental figure 3**).

632

633 **NEBuilder HiFi METa assembly compared to blunt ligation**

634 Because of the failure of our In-Fusion mediated METa assembly reactions, we focused
635 on NEBuilder HiFi mediated METa assembly reactions which will be referred to simply as
636 METa assembly from this point forward. The apparent much higher efficiency of METa
637 assembly compared to traditional blunt ligation in constructing metagenomic libraries (**Figure 4**)
638 led us to seek replication of our findings. Our approach was broadly similar to that detailed
639 above, with two changes. First, instead of using soil metagenomic DNA as input to tagmentation,
640 we combined and tagmented (**Supplemental figure 4**) genomic DNA of two bacterial strains of
641 interest to us: *Pseudomonas* sp. Strain PE-S1G-1 (ABC07) and *Pandoraea* sp. Strain PE-S2T-3
642 (ABC10). These previously sequenced strains (46) have been shown to be capable of using the
643 antibiotic penicillin as their sole carbon source (48) *via* a pathway likely initiated by a β -
644 lactamase enzyme (47). Creation of a functional mixed genomic library from these strains would
645 allow us to characterize penicillin resistance genes in a sequence naïve manner and test previous
646 annotations and findings. Second, instead of transforming METa assembled or blunt ligation

647 cloned libraries into homemade electrocompetent cells we purchased electrocompetent *E. coli*
648 cells with much higher transformation efficiency. These are routinely used in functional
649 metagenomic library preparation and result in larger libraries (37).

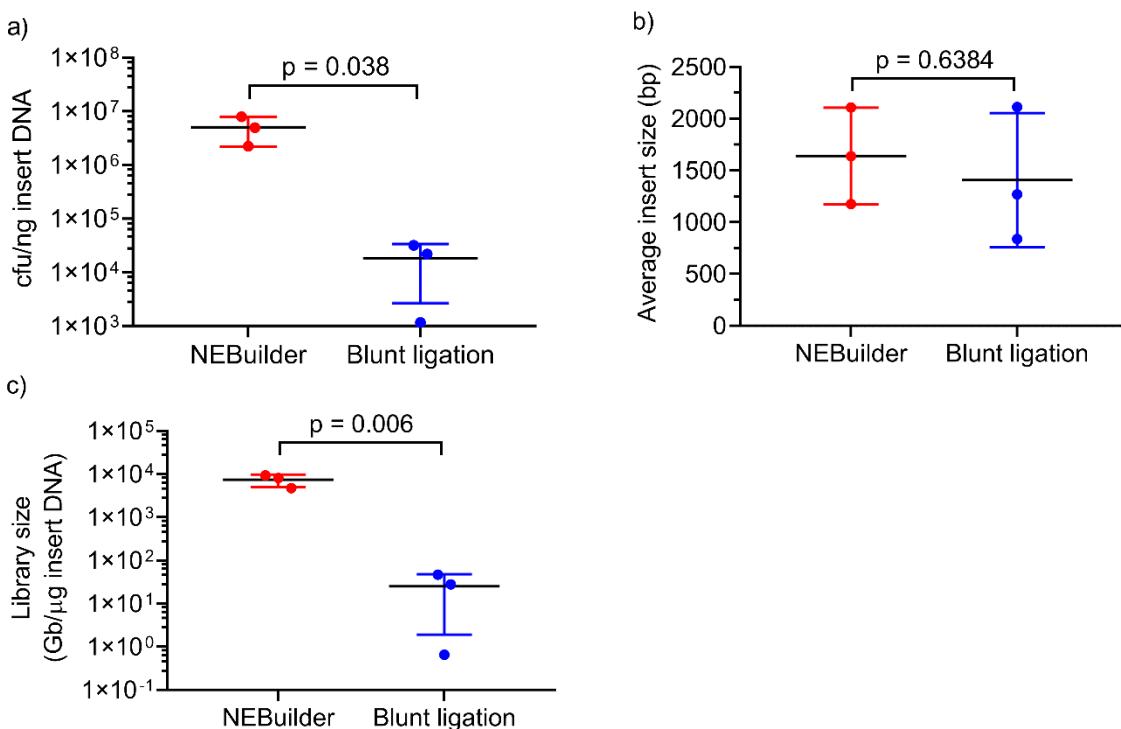


650

651

652 **Figure 4. Comparing metagenomic libraries prepared by assembly or blunt ligation**
653 Functional metagenomic libraries were created using METa assembly via In-Fusion assembly or
654 NEBuilder HiFi assembly and compared to a library constructed using blunt ligation. All
655 libraries used the same input DNA and were each prepared in triplicate. Error bars represent
656 standard deviation of n=3 experiments. Comparisons between blunt ligation and NEBuilder HiFi
657 METa assembly were made using unpaired two tailed t tests. Only a single In-Fusion colony was
658 isolated and therefore we did not perform statistical tests on that method. a) Post-transformation
659 culture titers normalized to the quantity of insert DNA used in the assembly or cloning reaction
660 itself. b) Average insert size for plasmids containing an insert. Colony PCR was performed on
661 the single In-Fusion colony, and 9 colonies were analyzed and averaged for each replicate
662 NEBuilder HiFi assembly or blunt ligation reaction to give three data points per method. c) Final
663 total library size or each replicate measured in gigabase pairs (Gb) of captured metagenomic
664 DNA normalized to the amount of insert (μg) used during cloning or assembly.

665 We found that METa assembly again resulted in significantly higher titers of transformed
666 cells per ng of insert used during library preparation (**Figure 5A**, ~276-fold greater than the blunt
667 ligation libraries, $p = 0.038$) and significantly larger libraries (**Figure 5C**, ~296-fold greater than
668 the blunt ligation libraries, $p = 0.006$). Colonies containing empty vectors occurred much more
669 frequently following blunt ligation reactions (3/7, 4/8, and 5/9 by colony PCR) compared to
670 colonies from METa assembly reactions (0/7, 0/7, and 0/8 by colony PCR) (**Supplemental**
671 **figure 5**). Average insert size (**Figure 5B**) did not appear to differ significantly between
672 methods.



673

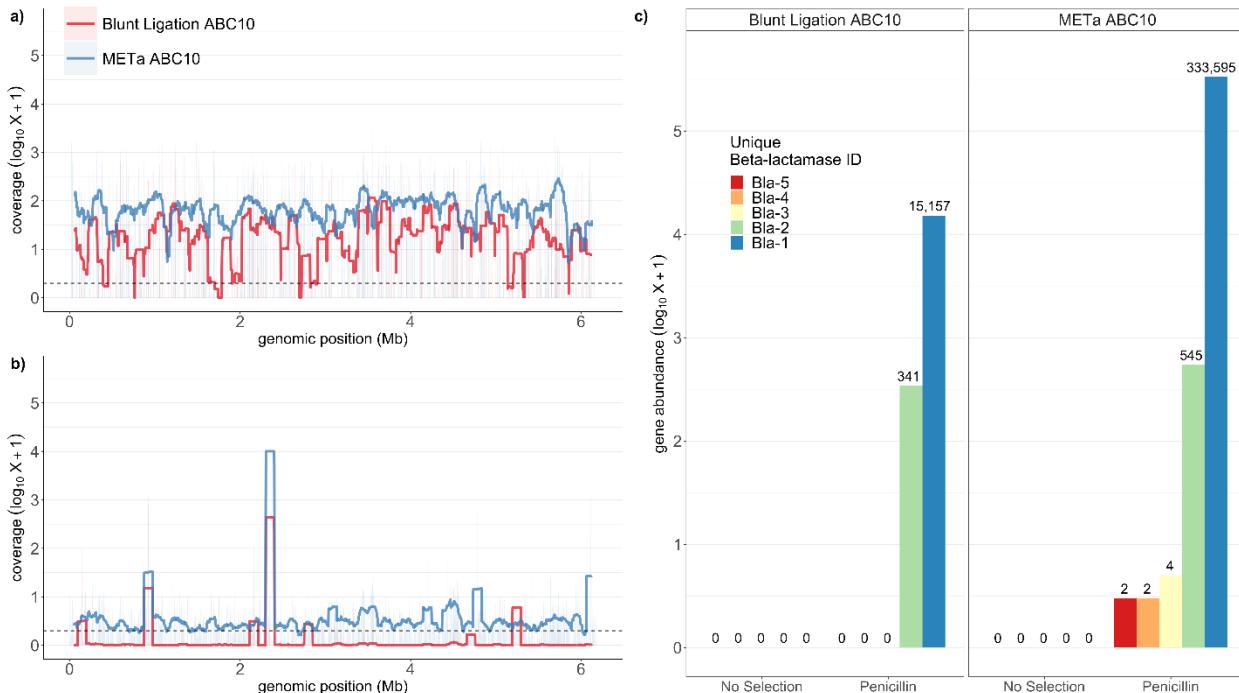
674 **Figure 5. Libraries created by NEBuilder HiFi mediated METa assembly are larger than**
675 **those created by blunt ligation.**

676 Both sets of triplicate metagenomic library assembly/cloning reactions used the same input DNA
677 and were compared using unpaired two tailed t tests. Error bars represent standard deviation of
678 n=3 experiments. a) Culture titers of recovered cells post-transformation normalized to insert
679 DNA mass used in assembly or cloning. b) Vector average insert size determined by colony
680 PCR, excluding colonies containing empty vector constructs. 13 colonies were assayed and insert
681 lengths were averaged for each replicate (n=3) for each construction method. c) Final library size
682 given in Gb/ug of insert DNA.

683

685 After establishing that assembly cloning results in much larger libraries than blunt
686 ligation cloning we next asked if genomic coverage was similar across methods. We sequenced
687 between 355 and 831 random colonies from each of the six libraries. Each library was plated to
688 give approximately 1,000 colonies based on prior titers resulting in an average of approximately
689 600 colonies collected (*ca.* 1562 total colonies from blunt ligation plates, 2008 total colonies
690 from assembly plates). We extracted plasmids from the collected colonies and pooled each set of
691 triplicate libraries into a single pool. These two pools were used as templates in a limited PCR
692 reaction to amplify inserts (**Supplemental figure 12**) which were submitted for long read
693 sequencing on the PacBio Sequel II platform. The resulting reads were mapped back onto
694 published ABC07 and ABC10 genomes (46) and the nucleotide coverage of each library for each
695 genome was calculated. We found qualitatively good agreement between assembly and blunt
696 ligation library coverage of both ABC10 (**Figure 6A**) and ABC07 (**Supplemental figure 6A**).

697 In order to verify the functional aspect of our functional metagenomic libraries we next
698 performed triplicate selections for growth in the presence of 1 mg/ml penicillin. This
699 concentration, about 10-fold higher than what is normally used in microbiology, was chosen
700 based on high level penicillin resistance of strains ABC07 and ABC10. Stocks were plated with
701 the goal of plating enough cells to reach 10-fold coverage of each library, resulting in denser
702 plating for the much larger METa assembly triplicate libraries. As before, colonies were
703 collected for PCR amplification of inserts (**Supplemental figure 12**) and sequenced. Sequencing
704 for both libraries were dominated by reads mapping to one region (ABC07, **Supplemental**
705 **figure 6B**) or two regions (ABC10, **Figure 6B**) of the donor organism genomes. In each case
706 these regions corresponded to predicted β-lactamase genes (**Figure 6C, supplemental figure**
707 **6C**).



708
709 **Figure 6. Assembly and blunt ligation library coverage of ABC010 genome with and**
710 **without penicillin selection**

711 a) Nucleotide depth of coverage for ABC010 genome by functional metagenomic library prepared
712 by assembly (blue) or blunt ligation (red). Coverage is smoothed to a 1 kb resolution. b) Same as
713 in a) but sequenced libraries were first subjected to selection on agar plates containing 1 mg/ml
714 penicillin. c) Gene abundance in post-penicillin selection reads for each of five predicted ABC010
715 β-lactamase genes that had read number >0. From Bla-1 to Bla-5 respective NCBI accession
716 numbers are WP_087722475.1, WP_087721859.1, WP_140413467.1, WP_087721948.1, and
717 WP_087721885.1.
718

719 Preparation of a large soil metagenomic library by METa assembly

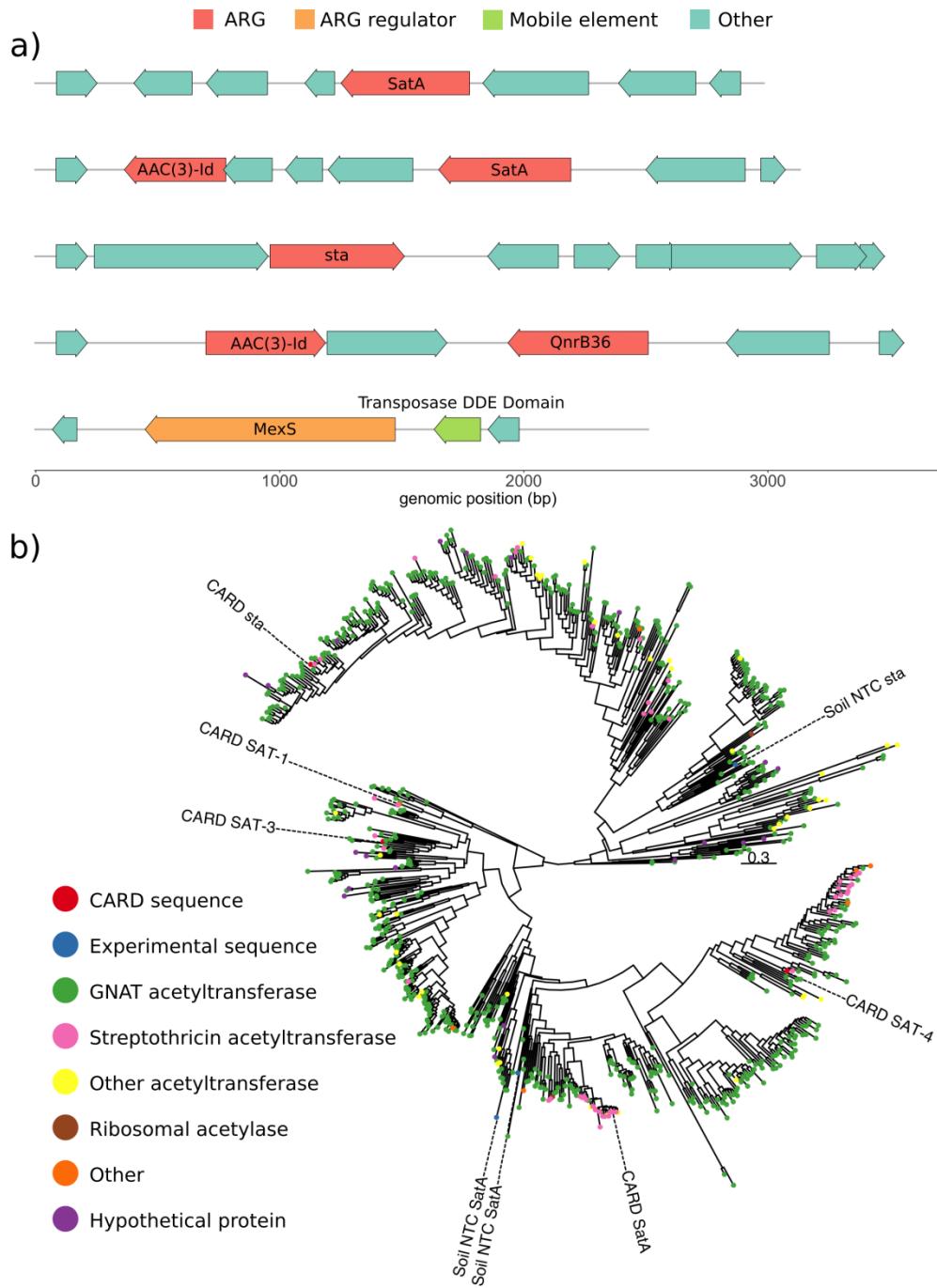
720 We next supplemented our method comparisons (**Figures 4 and 5**) by preparing a

721 functional metagenomic library by METa assembly for direct comparison to published libraries
722 prepared by blunt ligation. A common DNA input quantity for preparation of a single functional
723 metagenomic library by the sonication/blunt ligation workflow is ~5 µg (21, 28, 35, 37, 66, 70).

724 In order to compare METa assembly to the broader literature, we performed a tagmentation
725 reaction on 5 µg of soil metagenomic DNA. Following size selection and polymerase gap filling
726 we retained 1.122 µg of mosaic end containing inserts ready for assembly. Because this quantity

727 of DNA falls well outside the recommended capacity of NEBuilder HiFi reactions we proceeded
728 to first assemble 16.6% (175 ng) of the total in a trial reaction. Transformation of the purified
729 assembly reaction resulted in a 162 Gb library (7.8×10^7 unique clones, average insert size 2.077
730 kb \pm 708.7 bp standard deviation, **Supplemental figure 7**). After scaling up assembly and
731 transformation, the remaining inserts were used to prepare a 529.5 Gb library (2.85×10^8 unique
732 clones, average insert size of 1.858 ± 555.1 bp). Combined these libraries form a 691.5 Gb
733 library from 5 μ g of input DNA.

734 To test the utility of this method, we performed a functional metagenomic selection on
735 the 162 Gb library using the natural product antibiotic nourseothricin. Nourseothricin (also
736 known as cloNAT) is a member of the streptothricin class of aminoglycoside antibiotics first
737 described by Waksman and Woodruff in 1942 (71). Insert amplification (**Supplemental figure**
738 **12**) and sequencing resulted in identification of acetylation to be the dominant mode of resistance
739 in our library (**Supplemental figure 8A**). Among these we identified multiple apparent
740 homologs to known streptothricin acetyltransferases, with some inserts encoding multiple
741 syntenic resistance determinants (**Figure 7A**). Phylogenetic analysis of nourseothricin
742 acetyltransferase proteins places our representative soil-derived resistance genes within this
743 family, with one representative appearing to represent a novel cluster related to the StaT enzyme
744 (26.22% identity, 96.83% coverage) and two representatives clustering near SatA and Sat-4
745 enzymes (50.28%/47.96% identity and 102.17%/95.11% coverage) (**Figure 7B**). This clustering
746 pattern is in agreement with the amino acid sequence percent identities between the soil-derived
747 acetyltransferases and CARD validated streptothricin resistance enzymes.



748
749
750
751
752
753
754
755

Figure 7. Nourseothricin resistance conferring inserts from soil microbiome library selections

a) Genomic context of representative nourseothricin resistance genes including syntenic mobilization or regulatory elements and other antibiotic resistance genes. b) Phylogenetic tree of five CARD streptothrinac acetyltransferase enzymes (red circles, [CARD]), three soil metagenome nourseothricin resistance genes (blue circles, [Soil NTC]), and related enzymes.

756 **METa assembly of functional metagenomic libraries using limited input DNA**

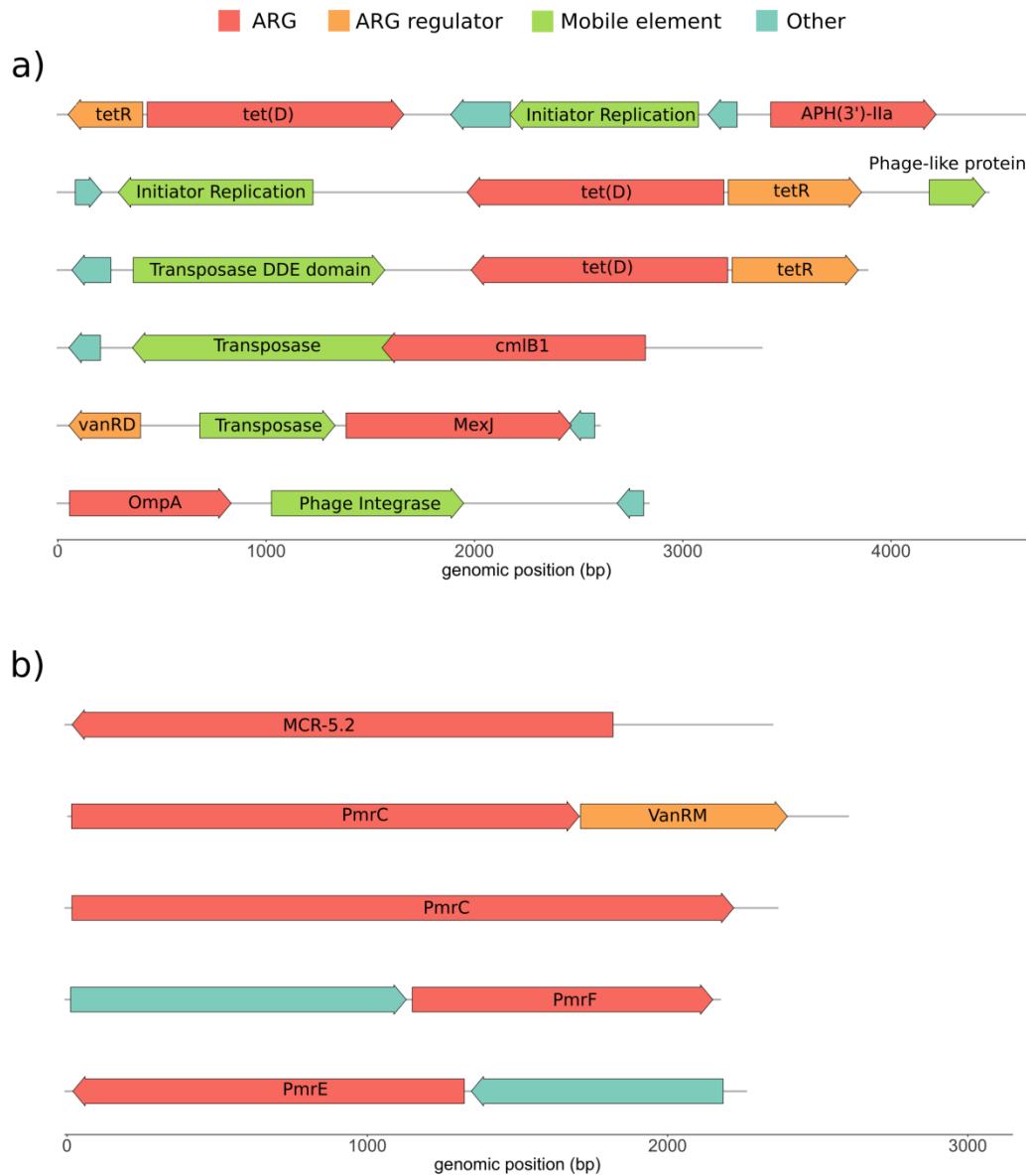
757 Our experiments comparing METa assembly library creation against library creation
758 using blunt ligation suggest that METa assembly, with its potential for low input DNA mass and
759 higher efficiency, may allow for the creation of functional metagenomic libraries using samples
760 with low DNA mass or quality. We first tested this possibility by creating a mock sub-optimal
761 metagenomic DNA sample consisting of a 10 kb λ phage DNA amplicon (in place of high-
762 quality DNA >48 kb suggested for Covaris sonicator). We used 200 ng of the DNA as input into
763 a tagmentation reaction (in place of the suggested 2 μ g to 20 μ g suggested for the sonicator).
764 Following clean-up, size selection, and polymerase repair we obtained 59.4 ng of 1 kb to 5 kb
765 DNA fragments, all of which were used as input for an assembly reaction. This resulted in a
766 ~13.5 Gb library with average insert size of ~1.02 kb \pm 340 bp standard deviation
767 (**Supplemental figure 9**) and a library efficiency of 67.5 Gb/ μ g of input DNA, suggesting that
768 METa assembly can still work well with low input and lower quality (*i.e.* not high molecular
769 weight) DNA.

770 Because it is possible that amplicon DNA behaves differently from metagenomic DNA,
771 we next performed a low input METa assembly library preparation using the same soil
772 metagenomic DNA as in the 5 μ g library. This time we started with a 250 ng DNA input
773 tagmentation reaction and after clean-up, size selection, and polymerase repair we performed an
774 assembly reaction using all 35.2 ng that passed through the processing steps. Electroporation
775 resulted in 3.8×10^6 colonies with an average insert size of 2.029 kb \pm 284 bp standard deviation
776 (**Supplemental figure 10**). We calculated the library size to be 7.71 Gb and the library
777 efficiency to be 30.84 Gb/ μ g, suggesting that METa assembly retains high efficiency with low
778 input metagenomic DNA. Because we had prepared the previous soil library we did not test this

779 library any further.

780 Finally, we prepared one additional low input DNA mass library by METa assembly. We
781 chose to use metagenomic DNA extracted from a Canada goose fecal pellet in order to include
782 DNA sourced from a non-soil microbiome and due to its ubiquitous availability near our research
783 space (**Supplemental figure 11A**). We used 300 ng of input DNA for tagmentation and,
784 following the usual sample processing steps, performed assembly into vector using 60.5 ng of
785 size-selected and repaired inserts. The resulting library was estimated to contain 1.18×10^7 unique
786 clones with colony PCR indicating 95.2% of clones containing an insert, with an average insert
787 length of $2.39 \text{ kb} \pm 651.5 \text{ bp}$ standard deviation (**Supplemental figure 11B**). Together, this
788 suggests a total library size of 26.86 Gb and a library efficiency of 89.67 Gb/ μg .

789 We selected the goose fecal microbiome library against a widely used soil natural product
790 antibiotic, tetracycline, and an antibiotic of last resort, colistin. Metagenomic inserts were
791 amplified from functionally selected plasmids by PCR (**Supplemental figure 12**) and sequenced.
792 The dominant mechanism for tetracycline resistance in the goose gut microbiome appears to be
793 drug efflux (**Supplemental figure 8C**). Many of the genes encoding efflux pumps are syntenic
794 to known regulatory elements (*e.g.* *tetR*) and/or potential mobilization elements (*e.g.* transposase
795 or phage integrase genes) (**Figure 8A**). The dominant mechanisms for colistin resistance in our
796 library consist of modification of lipid A and antibiotic efflux (**Supplemental figure 8B**).
797 Among the predicted lipid A modifying enzymes we identified homologs to the emerging MCR
798 family of mobilized colistin resistance enzymes, including an MCR-5.2 homolog (36.14%
799 identity, 109.89% coverage to CARD representative) (**Figure 8B**).



800
801 **Figure 8. Antibiotic resistance conferring inserts from goose gut microbiome library**
802 **selections**

803 Genomic context of representative resistance genes including syntenic mobilization or regulatory
804 elements and other antibiotic resistance genes following goose microbiome library selection on
805 a) tetracycline or b) colistin.

806

807 DISCUSSION

808 Functional metagenomic libraries have been valuable tools in chemical biology and have

809 been instrumental in the discovery of novel enzymes involved in antibiotic biosynthesis and

810 resistance, pharmaceutical-microbiome interactions, bioremediation of pollutants, and many
811 other activities of value to medicine and industry (15–17). However, the methods used in the
812 preparation of small insert functional metagenomic libraries (**Figure 1**), sonication fragmentation
813 and blunt ligation cloning, currently limit library size, efficiency, and adoption of this method.
814 DNA fragmentation by sonication is widely used because it shows low bias and it is possible to
815 control the resulting fragment size distribution. However, sonication instruments represent a
816 significant capital expense and require a high mass of high quality input DNA (*e.g.* from 2 µg to
817 20 µg of >48 kb length for a Covaris sonicator). This has limited the microbiomes that can be
818 explored by functional metagenomic libraries to those with a large amount of available
819 metagenomic DNA. Blunt ligation, in turn, uses input DNA inefficiently due to the low
820 probabilities involved in the random collisions that bring blunt DNA ends and ligase together to
821 react. Significant input DNA is also wasted by requiring a five-fold molar excess of input DNA
822 to vector DNA to avoid vector self-ligation. Due to this low efficiency, upwards of 20 µg of
823 input DNA may be recommended for library preparation (37). This is illustrated by a recent
824 publication (21) where the authors prepared 22 functional metagenomic libraries with an average
825 size of 18 Gb using approximately 5 µg of metagenomic DNA for each library. As 1 µg of input
826 DNA theoretically represents just under 1 million Gb of DNA, it is self-evident that only a small
827 fraction of the input metagenomic DNA is successfully captured. Two additional considerations
828 also arise from the use of blunt end ligation in current forms of functional metagenomic library
829 preparation. Blunt ligation reactions often require overnight incubation, a significant loss of time,
830 especially for protocols that propose using functional metagenomics as part of a diagnostics
831 pipeline (25). The use of blunt ends also results in library vulnerability to contamination by
832 foreign DNA. Custom agarose gel ladders lacking bacterial DNA and bleaching and washing of

833 electrophoresis apparatus are required to combat this risk (37).

834 We were inspired by the use of transposases in sequencing studies us to see if functional
835 metagenomic library preparation could similarly benefit. We found that transposase-mediated
836 fragmentation can be controlled to yield DNA fragments roughly the size of bacterial ORFs
837 (**Figure 3**) with the following empiric conditions for tagmentation reactions being compatible
838 with our enzyme preparation: 10 mM TAPS buffer pH 8.5, 5 mM MgCl₂, 10% w/v
839 dimethylformamide, input DNA at 10 ng/μl total reaction volume, and assembled transposase at
840 0.5 ng enzyme per ng of input DNA (*i.e.* 5 ng of assembled transposase per μl). The resulting
841 inserts are compatible with classic blunt ligation based cloning (**Figure 2B**) to produce
842 functional metagenomic libraries (**Figures 4 and 5**) but without any evidence of increased
843 efficiency. Tagmentation of metagenomic DNA was successful across several DNA sources
844 including purified bacterial genomic DNA, PCR amplicons, soil metagenomic DNA, and fecal
845 metagenomic DNA.

846 We also realized that installation of mosaic end tags on the 5' ends of fragmented DNA
847 could allow us to replace blunt ligation cloning with higher efficiency homology based assembly
848 methods (**Figures 2C and 2D**). Direct comparisons of functional metagenomic library
849 preparation by assembly cloning demonstrated that METa assembly synergizes dramatically with
850 tagmented DNA fragments and is significantly more efficient (**Figures 4 and 5**) than library
851 preparation *via* blunt ligation. While variation in read counts and the number of input colonies
852 confounded a statistical interpretation, qualitatively it appears that libraries prepared by METa
853 assembly provide equal, if not greater, coverage of input DNA compared to libraries prepared by
854 blunt ligation (**Figure 6A, supplemental figure 6A**).

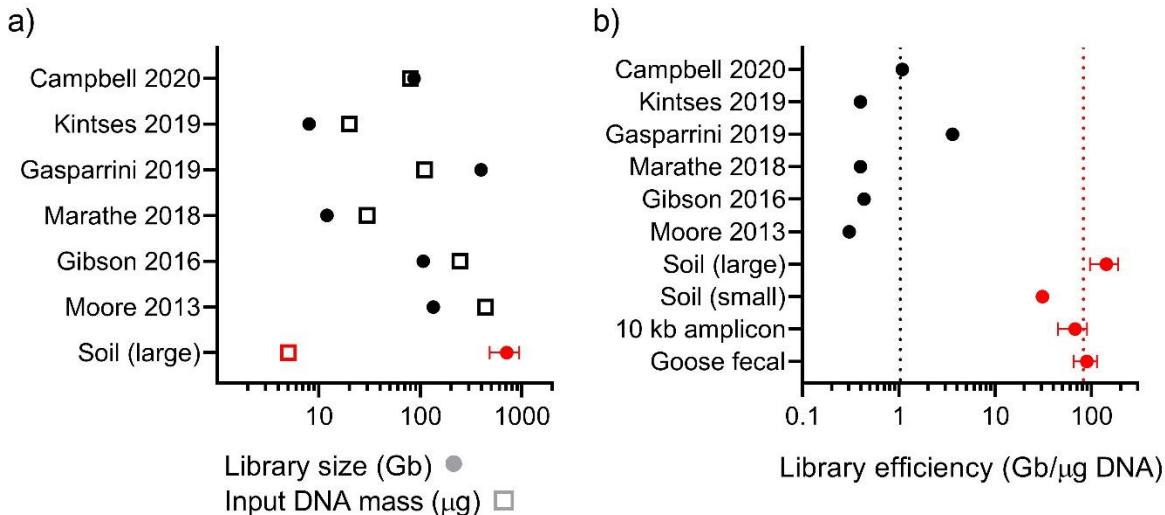


Figure 9. Comparison of METa assembly libraries to literature examples

a) Input DNA mass and functional metagenomic library sizes for six publications (black) compared to the large (5 µg) soil METa assembly library (red). Error bars (not present for literature libraries) calculated based on insert size standard deviation as determined by colony PCR ($n=48$ colonies). Filled shapes correspond to library sizes and empty shapes correspond to input DNA mass. b) Library efficiency (library size in Gb normalized to input DNA mass in µg) calculated for literature examples (black) or METa assembly libraries (red). Vertical dotted lines correspond to literature (black) or METa assembly (red) average efficiency ($n=6$ for literature examples, $n=4$ for METa assembly examples). Standard deviation error bars for METa assembly libraries calculated based on colony PCR as before (from top to bottom colonies tested $n=48$, 28, 14, or 21).

In order to compare METa assembly against functional metagenomic library preparation

by current experts in the field we carried out a literature search of recent articles to find appropriate comparators (**Supplemental figure 13**). Literature protocols with sufficient methods detail (21, 28, 35, 66–68) showed substantial gains in efficiency are made by using METa assembly to prepare functional metagenomic libraries. Literature protocols used initial DNA input masses of between 10 µg and 500 µg to prepare functional metagenomic libraries sized between 10 Gb and 400 Gb. Starting with 5 µg of metagenomic DNA, we used METa assembly to make libraries totaling 712 Gb in size, approximately the same size as the six literature examples combined using a fraction of their total input DNA (**Figure 9A**). Similarly, libraries prepared by METa assembly using only 200 ng to 300 ng of input DNA, more than 10-fold less

878 than the normative 5 µg input for a single library, resulted in libraries of comparable size. When
879 library size is normalized to input metagenomic DNA mass, functional metagenomic libraries
880 prepared by METa assembly significantly outstrip libraries prepared by blunt ligation (**Figure**
881 **9B**).

882 We next verified the utility of our libraries as chemical biology tools for bioprospecting
883 by subjecting them to selection on three antibiotics. For our soil library we chose to select with
884 the streptothrin antibiotic nourseothrin. Surveys of antibiotic-producing bacteria have
885 determined that the streptothricins are among the most commonly produced antibiotics in soil
886 ecosystems, with between 10% and ~42% of soil actinomycetes potentially being producers (72,
887 73). Surprisingly for such a widespread antibiotic there are a limited number of described
888 resistance determinants (*sat2-4*, *sttH*, and *natI*), especially compared with other aminoglycoside
889 antibiotics (74–76). Our selection and subsequent analysis suggest that the streptothrin
890 acetyltransferase family is larger and more diverse than currently thought and likely contains
891 unknown major branches. These branches likely incorporate many enzymes which at the
892 moment are solely annotated as Gcn5-related N-acetyltransferases (GNATs) (**Figure 7B**). For
893 selection of our goose gut microbiome library we chose to select on two additional classes of
894 antibiotics (in contrast to aminoglycosides and β-lactams used above): tetracycline due to its
895 widespread historical use in medicine and agriculture, and colistin due to its importance as an
896 antibiotic of last resort. Migratory birds, including species of goose, likely harbor microbiomes
897 richer in antimicrobial resistant bacteria compared to other microbiomes (34, 51). In one study,
898 50% of migratory birds encoded the emerging colistin resistance gene *mcr-1* within their
899 microbiomes, while the most prevalent antimicrobial resistance genes in these microbiomes are
900 against tetracycline (51). The dominant tetracycline resistance mechanism picked up by our

901 experiment is drug efflux. As illustrated in the representative insert gene maps for this selection
902 (**Figure 8A**), many tetracycline efflux pumps from this selection are syntenic to potential
903 mobilization elements, including transposases, and phage- and plasmid-associated genes. In
904 contrast, the dominant colistin resistance mechanism appears to be lipid A modification and our
905 functional metagenomic library selections identified a potentially novel MCR enzyme homology,
906 further confirming the presence of these concerning genes in migratory birds.

907 In summary, our experiments developing and testing METa assembly highlight several
908 advantages of the method. First, the use of transposases to fragment metagenomic DNA has
909 several benefits. It removes the need for costly capital equipment, such as sonicators, while
910 providing the benefits of greater control and experimental flexibility seen with restriction
911 enzyme-mediated fragmentation without the downsides of needing restriction site frequency or
912 methylation dependence (77). Like sonication, fragmentation shows very little to no sequence
913 bias, making it essentially random (40), but unlike sonication it can be applied to low biomass
914 samples. The compatibility with low biomass samples is especially useful, as low biomass or rare
915 microbiomes have forced other research groups being to pool independent samples (21, 35) or
916 use potentially biased DNA amplification techniques (78, 79) to increase input DNA mass. One
917 particularly relevant low biomass sample for METa assembly could be clinical swabs, which
918 have been found on average to yield 371 ng of metagenomic DNA (80) which is well above the
919 inputs we have used here (200 ng to 300 ng).

920 Second, the addition of mosaic end sequences to the DNA inserts allows the use of
921 modern assembly cloning methods. Random fragmentation of metagenomic DNA by sonication
922 or enzymatic digest (by restriction enzyme or fragmentases(67)) results in DNA fragments with
923 little to no information about the DNA sequence at the fragment ends. This limits cloning of

924 these fragments to lower efficiency methods like blunt ligation. More efficient assembly cloning
925 protocols rely on incorporation of insert-matching sequences into the plasmid cloning site to
926 allow insert-vector hybridization to drive ligation specificity and efficiency. As a result, current
927 functional metagenomic library preparation methods have not been able to take advantage of
928 high efficiency cloning techniques that have been available for more than a decade (44, 69).
929 Assembly cloning and transposase fragmentation are therefore synergistic: without assembly
930 cloning, tagmented DNA would be cloned *via* blunt ligation with no gains in efficiency, and
931 without tagmentation, assembly cloning fails for lack of insert-vector hybridization.

932 Third, plasmids used in blunt end cloning, and therefore current functional metagenomic
933 library preparation methods, must undergo phosphatase treatment to prevent self-ligation. This is
934 not a problem in assembly cloning reactions due to the higher temperatures used. We omitted
935 phosphatase treatment from our workflow and while this led to the detection of numerous empty
936 vector colonies in our blunt ligation libraries, only a single METa assembly colony out of 115
937 tested was found to be lacking an insert (**Supplemental figures 3, 5, 7, 9, 10, and 11B**). We
938 used inverse PCR to prepare linearized vectors for both blunt ligation and METa assembly which
939 allowed us to incorporate mosaic end sequences into our vector of choice. This strategy is
940 applicable to any plasmid that is readily amplified by PCR meaning that METa assembly can be
941 easily incorporated into existing functional metagenomic library preparation workflows.

942 Finally, the METa assembly protocol presents significant time savings. The process of
943 fragmenting DNA by tagmentation takes only 7 minutes followed by a 5-minute quench. In
944 contrast, fragmentation by sonication with a *e.g.* a Covaris E220 instrument requires a 60 minute
945 degas time on top of 20 minutes or more of fragmentation time. End repair with DNA
946 polymerase to fill in 3' overhangs following tagmentation takes 15 minutes (though it could

947 likely be accomplished in less time) while end repair following physical fragmentation of DNA
948 with an End-It kit requires 45 minute reactions and 10-minute heat inactivation. Most notably,
949 assembly based cloning takes 15 minutes while it is generally recommended that blunt ended
950 ligation reactions be allowed to react overnight for optimal efficiency. It has been suggested that
951 functional metagenomic library preparation could be used in a rapid workflow for clinical
952 detection of resistance genes (25). The time savings found in METa assembly, most notable in
953 the cloning step, could be invaluable in such a workflow.

954 In conclusion, the synergistic combination of fragmentation by transposase and cloning
955 by assembly allows METa assembly to prepare larger, less DNA greedy, and more robust
956 functional metagenomic libraries. The advantages of METa assembly of functional metagenomic
957 libraries could allow these valuable chemical biology tools to be prepared from sources
958 previously out of reach including those of low biomass (such as from the built environment),
959 requiring fast turnaround (such as in the clinic), or of limited availability (such as exotic or
960 historical samples).

961

962 **AVAILABILITY**

963 Sequencing data and scripts for data analysis and statistics are available at the Hartmann lab
964 github: <https://github.com/hartmann-lab>

965

966 **AUTHOR CONTRIBUTIONS**

967 **TSC:** Conceptualization, formal analysis, investigation, methodology, project administration,
968 resources, supervision, validation, visualization, writing – original draft, writing – review and
969 editing. **AGM:** Data curation, formal analysis, investigation, methodology, software, validation,

970 visualization, writing – original draft, writing – review and editing. **EMH:** Funding acquisition,
971 project administration, resources, supervision, writing – review and editing.

972

973 **ACKNOWLEDGEMENT**

974 We would like to thank Dr. Rickard Sandberg for their deposition of the pTXB1-Tn5 plasmid in
975 the Addgene depository. We wish to acknowledge and thank Dr. Neil L. Kelleher for sharing
976 laboratory space and research support with TSC. We also wish to thank Dr. Gautam Dantas for
977 critical reading of the manuscript.

978

979 **FUNDING**

980 EMH and AGM are supported by The Searle Leadership Fund.

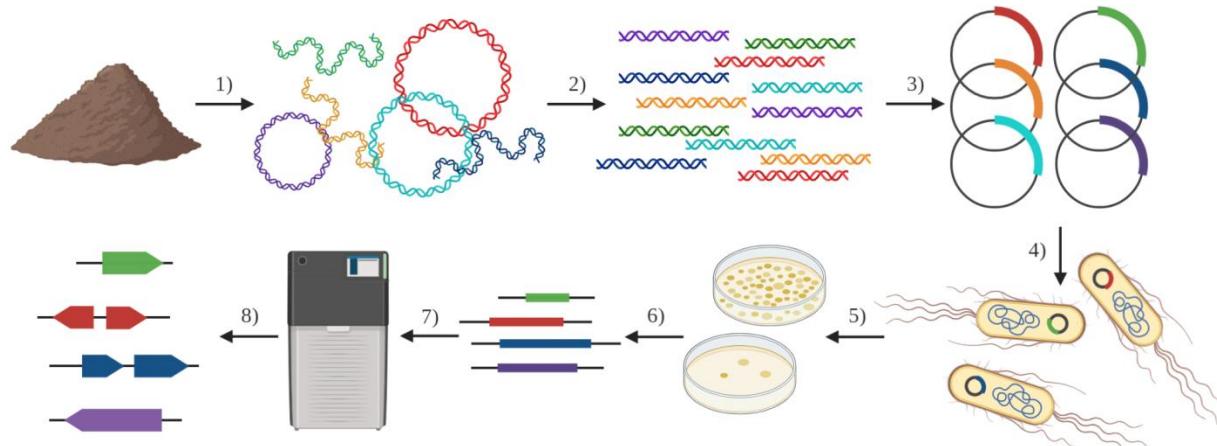
981

982 **COMPETING INTERESTS**

983 Northwestern University, with the authors, has submitted a provisional patent application based
984 on the METa assembly technique.

985

986 TABLES AND FIGURES



987
988

Figure 1. Functional metagenomic library pipeline

989 The general pipeline for the creation and use of functional metagenomic libraries to capture and
990 discover genes from metagenomes.

991 1) Extraction of metagenomic DNA from a microbiome (e.g. soil or fecal samples). 2)
992 Fragmentation of metagenomic DNA to desired size range (e.g. by sonication, restriction enzyme
993 digestion, or tagmentation). 3) Cloning of fragments into expression vectors following size
994 selection (e.g. by blunt ligation or homology-based assembly. 4) Transformation *en masse* of
995 vectors into an expression host (e.g. *E. coli* DH10B) to create functional metagenomic library. 5)
996 Functional selection or screen of library (e.g. on antibiotics to select for resistance). 6)
997 Amplification of selected inserts using vector-specific primers. 7) High-throughput sequencing
998 of selected metagenomic amplicons (e.g. by Illumina or PacBio technologies). 8) Annotation of
999 sequenced amplicons to link novel genes with selected/screened function (e.g. discovery of novel
1000 aminoglycoside acetyltransferases). Figure created in BioRender.com.
1001

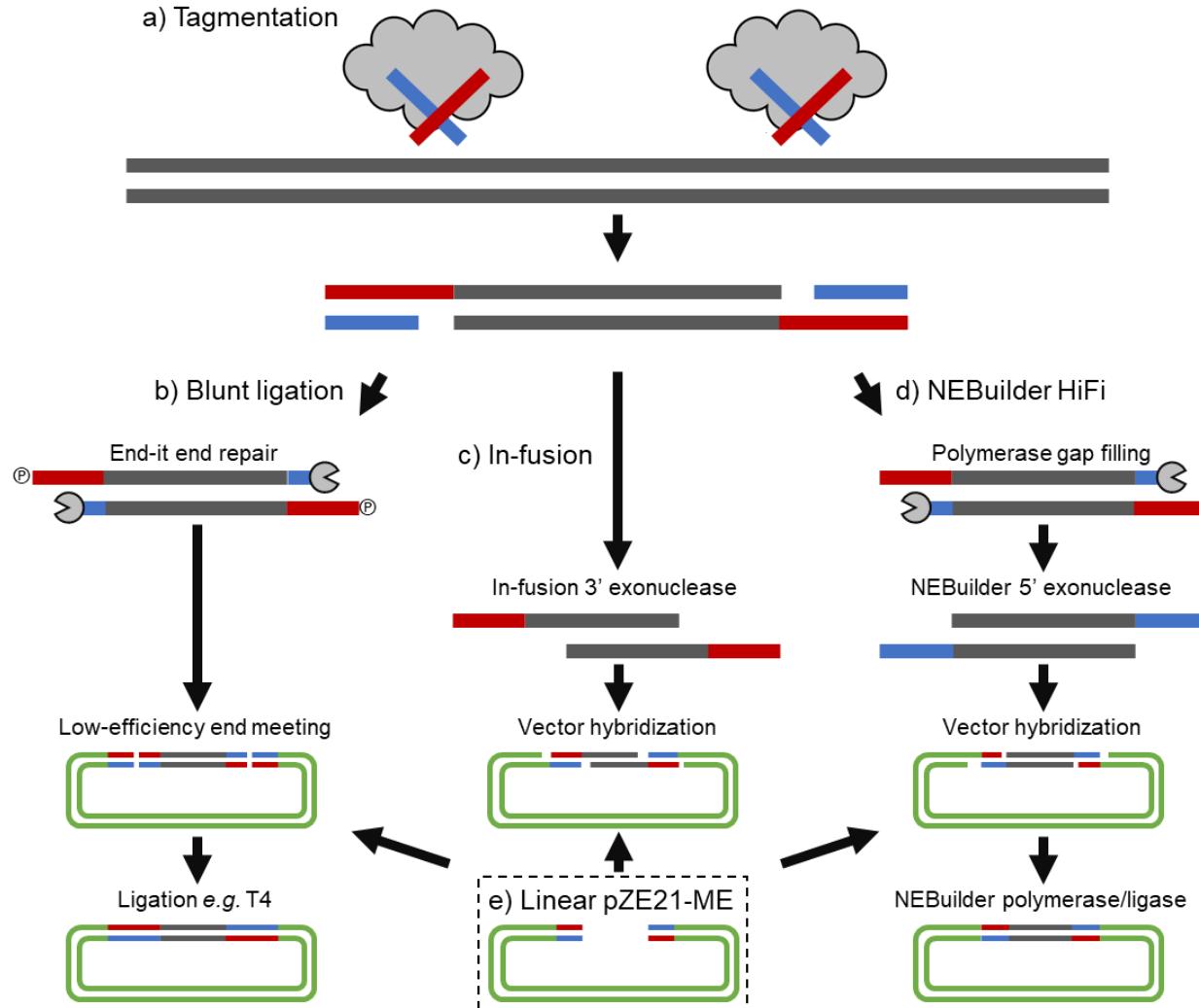
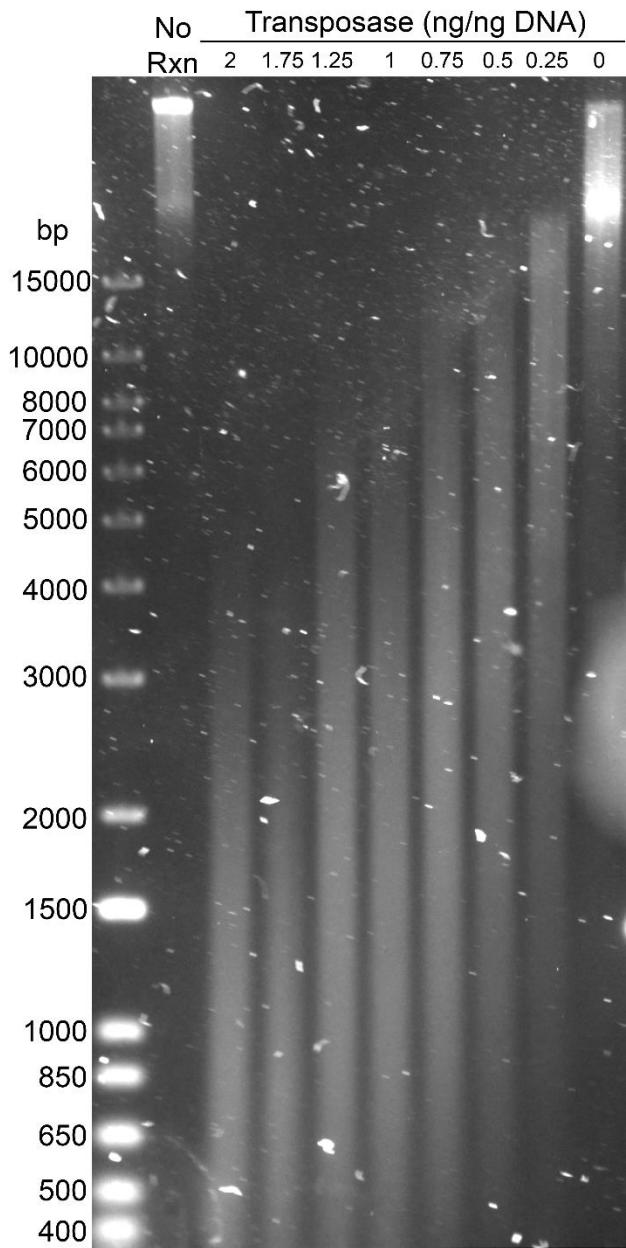
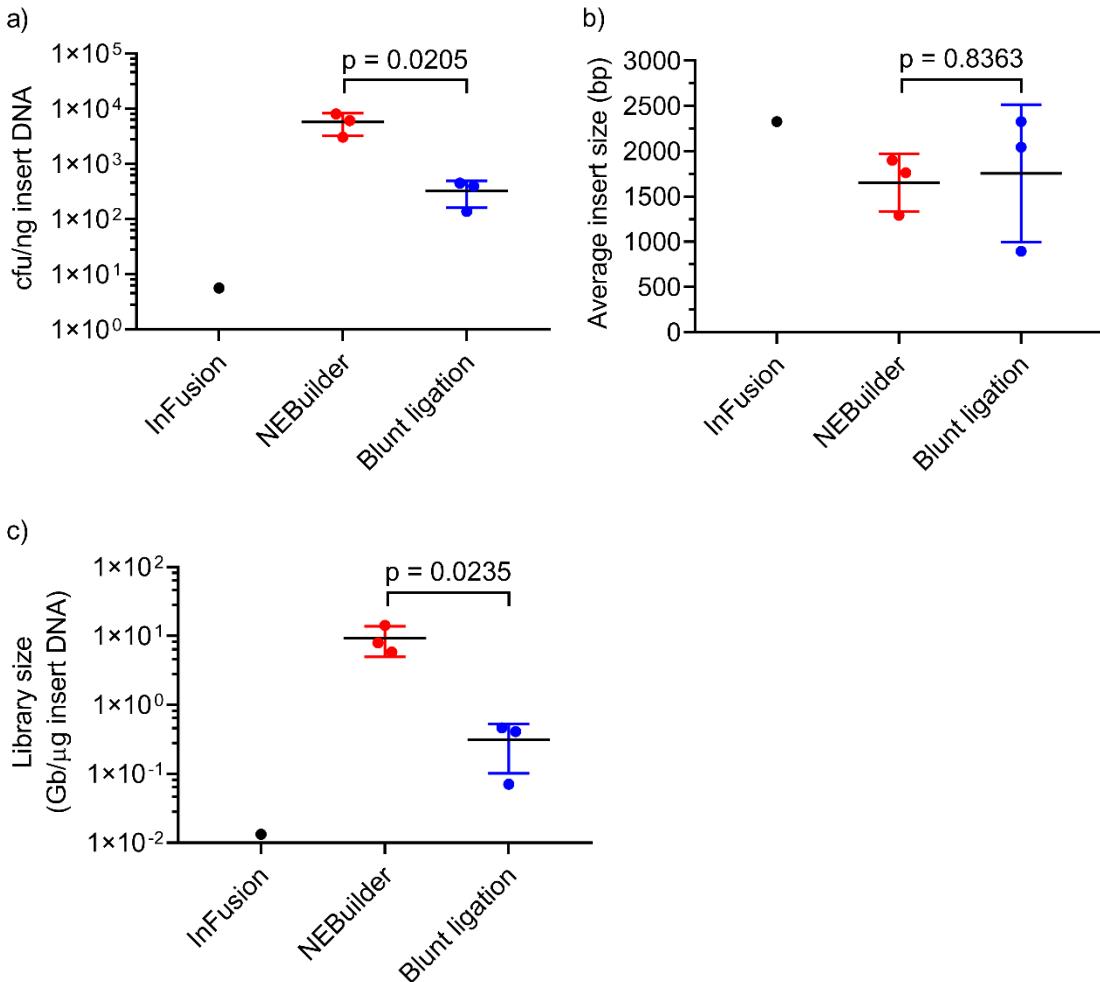


Figure 2. Blunt cloning protocol compared to METa assembly with NEBuilder HiFi or In-Fusion

1002
1003
1004

1005 a) Transposase enzyme fragments DNA with 5' mosaic end oligos. Inserts can be used as input
1006 for all three methods. All three protocols are compatible with linear pZE21-ME vector prepared
1007 by inverse PCR.
1008 b) Blunt cloning *via* end-repair and ligase. 5' overhangs must be resolved by gap filling and
1009 phosphorylation using end-repair enzyme mixes. Blunt ended inserts can be ligated to blunt
1010 ended vector.
1011 c) METa assembly *via* In-Fusion enzyme mix. In-Fusion 3' exonuclease activity is directly
1012 compatible with transposase fragments. Single stranded DNA overhangs on inserts and vector
1013 hybridize into a stable complex that can be transformed without filling gaps or covalently sealing
1014 nicks.
1015 d) METa assembly *via* NEBuilder HiFi enzyme mix. 5' overhangs must be resolved by DNA
1016 polymerase gap filling. NEBuilder HiFi enzyme mix includes 5' exonuclease to create 3'
1017 overhangs which hybridize with target pZE21-ME. DNA polymerase fills in gaps and ligase
1018 seals nicks.
1019 e) pZE21-ME is prepared and linearized by inverse PCR and is compatible with all three DNA
1020 pipelines.





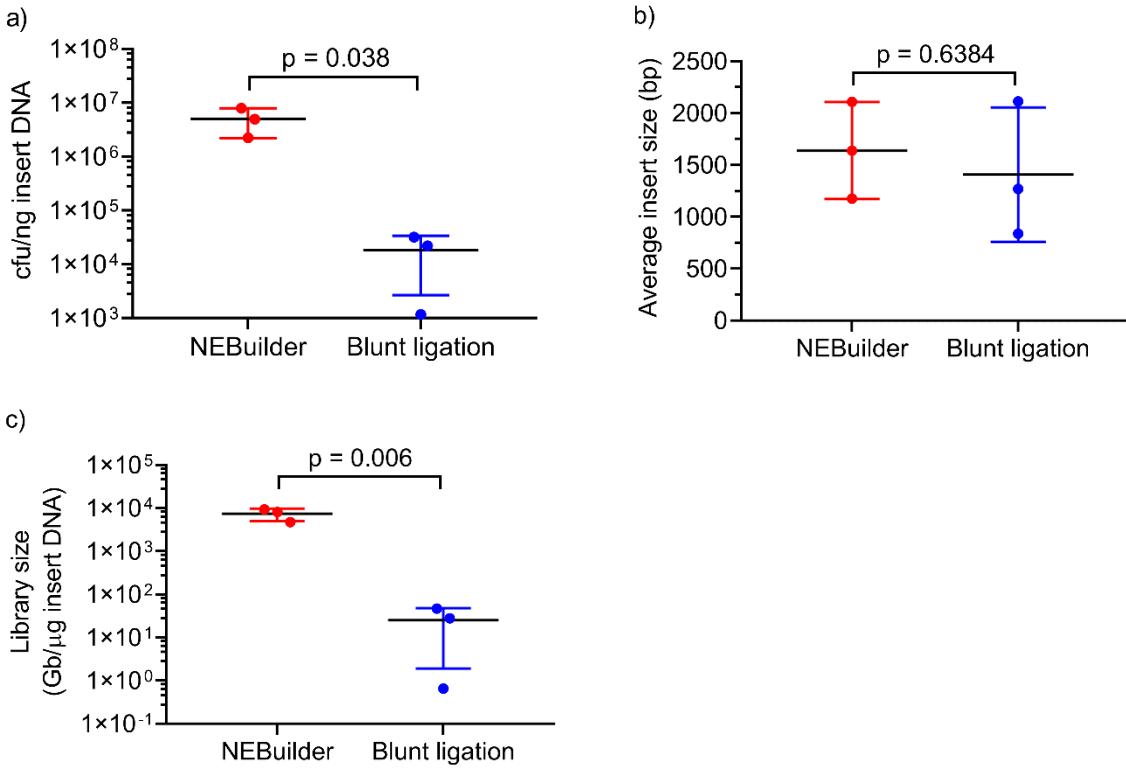
1029

1030

1031 **Figure 4. Comparing metagenomic libraries prepared by assembly or blunt ligation**
1032 Functional metagenomic libraries were created using METa assembly *via* In-Fusion assembly or
1033 NEBuilder HiFi assembly and compared to a library constructed using blunt ligation. All
1034 libraries used the same input DNA and were each prepared in triplicate. Error bars represent
1035 standard deviation of n=3 experiments. Comparisons between blunt ligation and NEBuilder HiFi
1036 METa assembly were made using unpaired two tailed t tests. Only a single In-Fusion colony was
1037 isolated and therefore we did not perform statistical tests on that method. a) Post-transformation
1038 culture titers normalized to the quantity of insert DNA used in the assembly or cloning reaction
1039 itself.

1040 b) Average insert size for plasmids containing an insert. Colony PCR was performed on the
1041 single In-Fusion colony, and 9 colonies were analyzed and averaged for each replicate
1042 NEBuilder HiFi assembly or blunt ligation reaction to give three data points per method.
1043 c) Final total library size or each replicate measured in gigabase pairs (Gb) of captured
1044 metagenomic DNA normalized to the amount of insert (μg) used during cloning or assembly.

1045



1048 **Figure 5. Libraries created by NEBuilder HiFi mediated METa assembly are larger than**
1049 **those created by blunt ligation.**

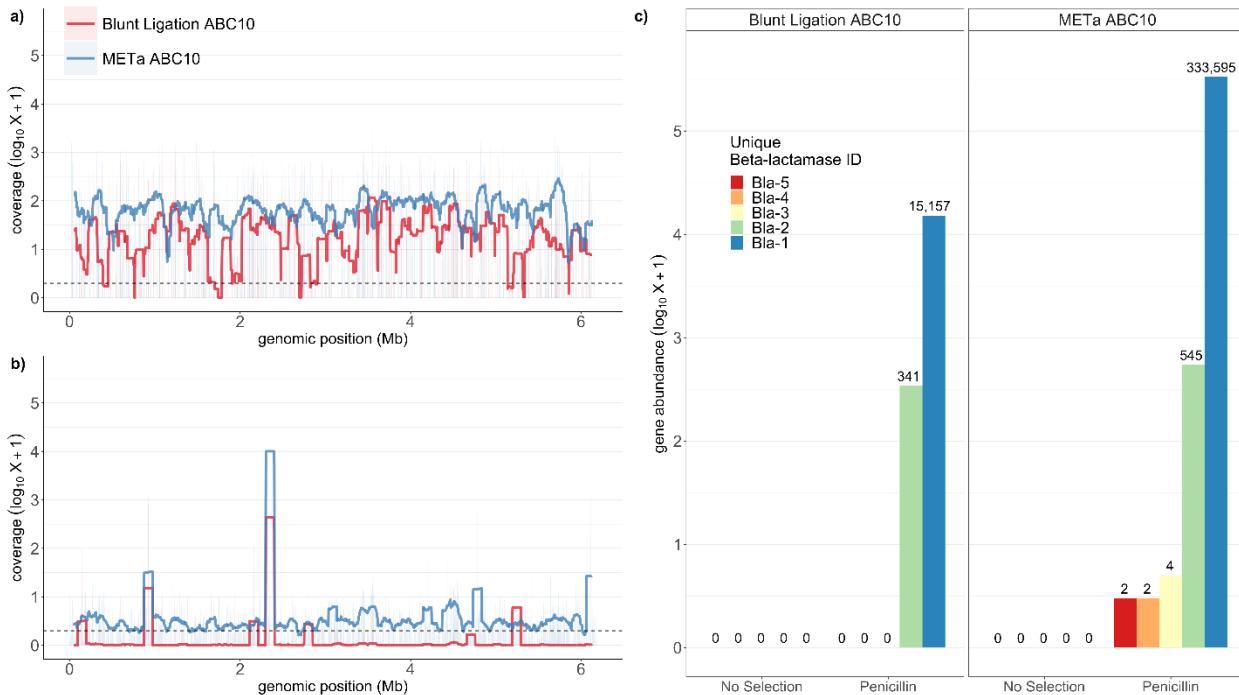
1050 Both sets of triplicate metagenomic library assembly/cloning reactions used the same input DNA
1051 and were compared using unpaired two tailed t tests. Error bars represent standard deviation of
1052 n=3 experiments. a) Culture titers of recovered cells post-transformation normalized to insert
1053 DNA mass used in assembly or cloning.

1054 b) Vector average insert size determined by colony PCR, excluding colonies containing empty
1055 vector constructs. 13 colonies were assayed and insert lengths were averaged for each replicate
1056 (n=3) for each construction method.

1057 c) Final library size given in Gb/μg of insert DNA.

1058

1059



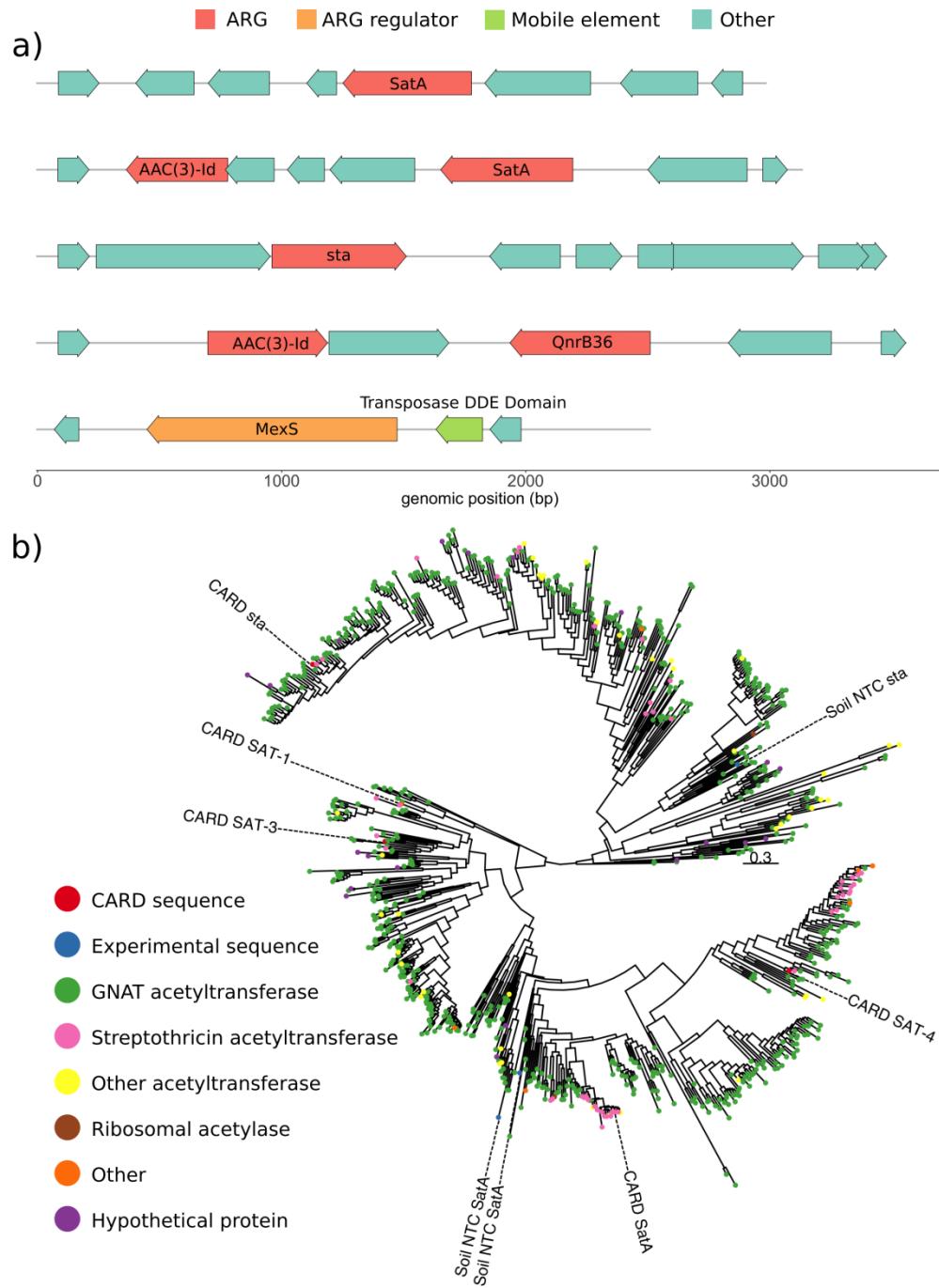
1060
1061 **Figure 6. Assembly and blunt ligation library coverage of ABC010 genome with and**
1062 **without penicillin selection**

1063 a) Nucleotide depth of coverage for ABC010 genome by functional metagenomic library prepared
1064 by assembly (blue) or blunt ligation (red). Coverage is smoothed to a 1 kb resolution.

1065 b) Same as in a) but sequenced libraries were first subjected to selection on agar plates
1066 containing 1 mg/ml penicillin.

1067 c) Gene abundance in post-penicillin selection reads for each of five predicted ABC010 β -
1068 lactamase genes that had read number >0. From Bla-1 to Bla-5 respective NCBI accession
1069 numbers are WP_087722475.1, WP_087721859.1, WP_140413467.1, WP_087721948.1, and
1070 WP_087721885.1.

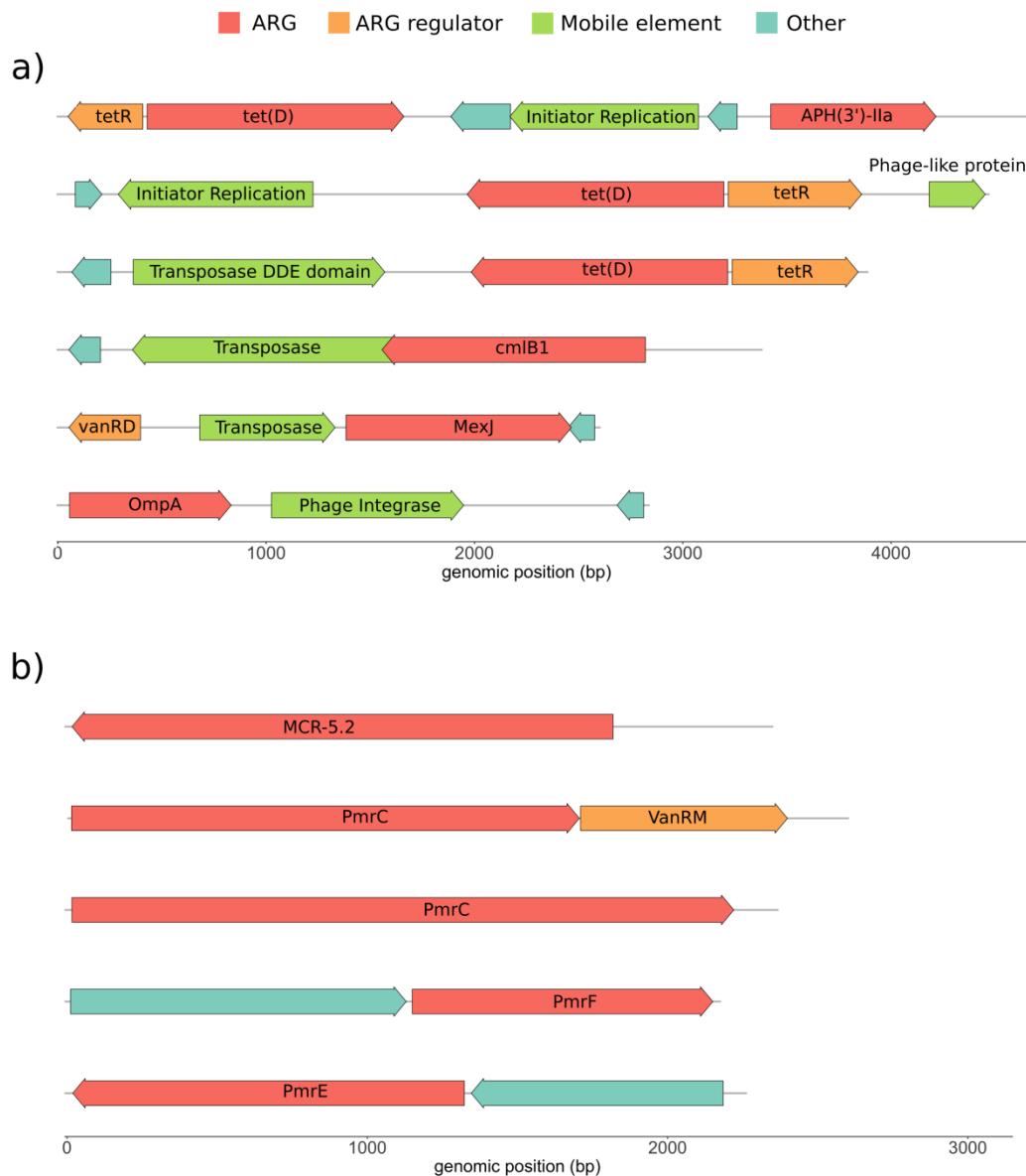
1071
1072
1073



1074
1075 **Figure 7. Nourseothricin resistance conferring inserts from soil microbiome library**
1076 **selections**

1077 a) Genomic context of representative nourseothricin resistance genes including syntenic
1078 mobilization or regulatory elements and other antibiotic resistance genes.
1079 b) Phylogenetic tree of five CARD streptothrinac acetyltransferase enzymes (red circles,
1080 [CARD]), three soil metagenome nourseothricin resistance genes (blue circles, [Soil NTC]), and
1081 related enzymes.

1082



1083

Figure 8. Antibiotic resistance conferring inserts from goose gut microbiome library selections

Genomic context of representative resistance genes including syntenic mobilization or regulatory elements and other antibiotic resistance genes following goose microbiome library selection on a) tetracycline or b) colistin.

1089 .

1090

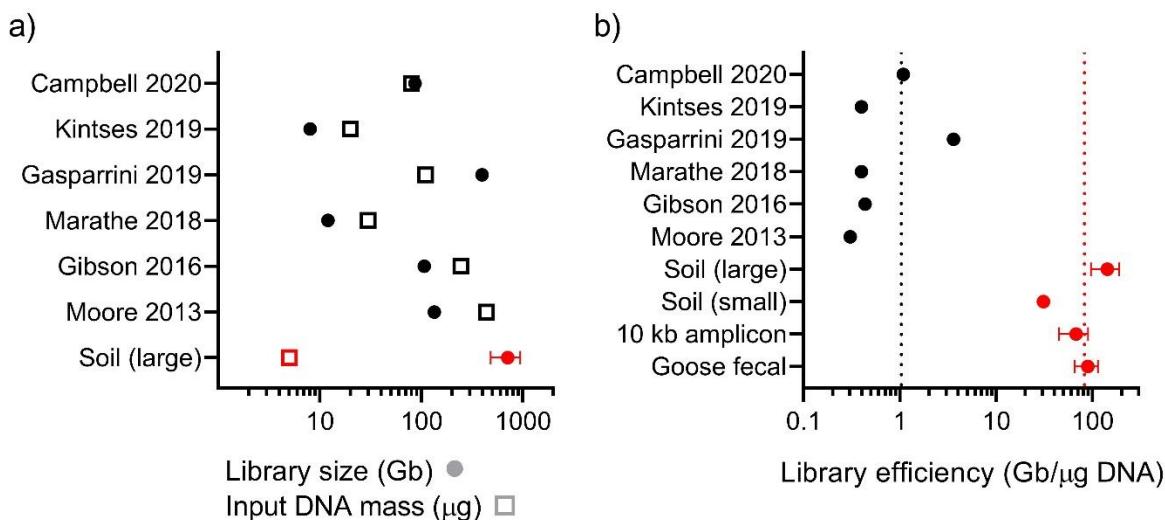
1091

1092

1093

1094

1095



1096
1097 **Figure 9. Comparison of METa assembly libraries to literature examples**

1098 a) Input DNA mass and functional metagenomic library sizes for six publications (black)
1099 compared to the large (5 μg) soil METa assembly library (red). Error bars (not present for
1100 literature libraries) calculated based on insert size standard deviation as determined by colony
1101 PCR (n=48 colonies). Filled shapes correspond to library sizes and empty shapes correspond to
1102 input DNA mass.

1103 b) Library efficiency (library size in Gb normalized to input DNA mass in μg) calculated for
1104 literature examples (black) or METa assembly libraries (red). Vertical dotted lines correspond to
1105 literature (black) or METa assembly (red) average efficiency (n=6 for literature examples, n=4
1106 for METa assembly examples). Standard deviation error bars for METa assembly libraries
1107 calculated based on colony PCR as before (from top to bottom colonies tested n=48, 28, 14, or
1108 21).

1109

1110

1111

1112

1113

1114

1115

1116

1117

1118

1119

1120

1121

1122

1123

1124

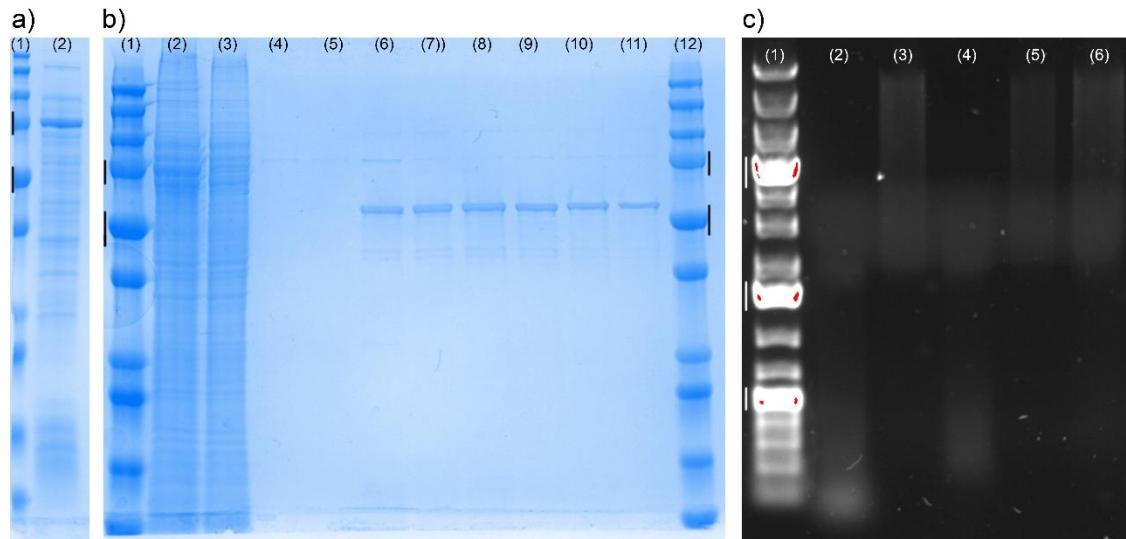
1125 **SUPPLEMENTAL TABLES AND FIGURES**

1126

1127 **Supplemental table 1. Primers and oligonucleotides used**

Oligo name	Oligo sequence	Purpose in study
5Phos_METagA1	/5Phos/CTGTCTCTTATACACATCT	Transposase oligo
METagA2	AGATGTGTATAAGAGACAG	Transposase oligo
6469TSC	CTGTCTCTTATACACATCTCCTC GAGGTCGACGGTATCGATAAGC	Amplification of linear pZE21-ME
6470TSC	CTGTCTCTTATACACATCTCCTT TCTCCTCTTAATGAATT CGGTC AGTGCG	Amplification of linear pZE21-ME
6463TSC	CCAGTTTACTTGCAAGGGCTTCC CAACC	Amplification of pZE21-ME inserts; colony PCR
6464TSC	GCCTTGAGTGAGCTGATACCG CTCG	Amplification of pZE21-ME inserts; colony PCR

1128



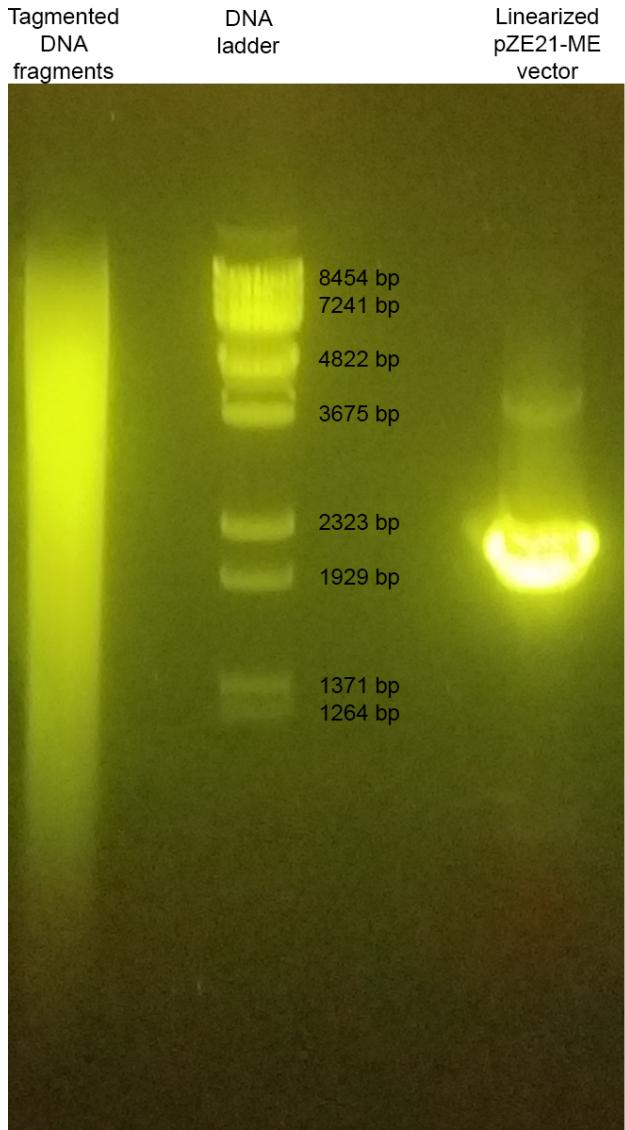
1129

1130 **Supplemental figure 1. Expression, purification, and testing of transposase activity**

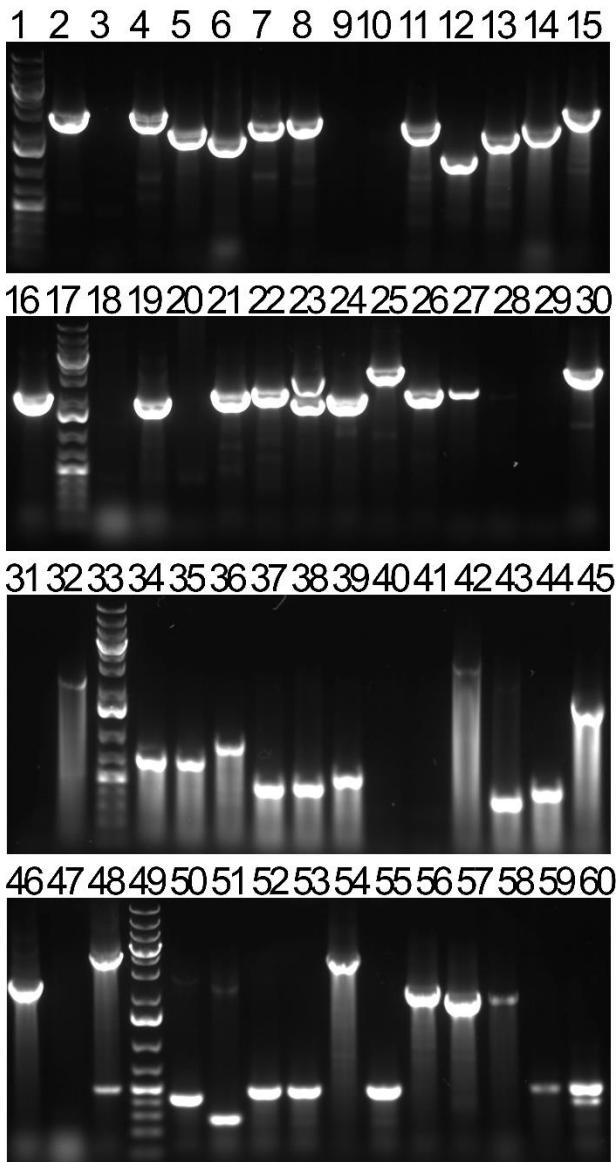
1131 a) Testing auto-induction expression of transposase. (1) Protein ladder with black bars next to 75
1132 kDa and 50 kDa markers, (2) lysate from test expression culture.

1133 b) Transposase purification. (1) Protein ladder with 75 kDa and 50 kDa standards marked, (2)
1134 clarified lysate, (3) flow-through, (4) wash 1, (5) wash 2, (6) 24 hr test elution, (7) 48 hr elution,
1135 (8)-(11) elution washes 1 through 4, (12) ladder. Tn5-CDB fusion protein expected mass is ~75
1136 kDa, Tn5 post-cleavage expected mass is ~50 kDa.

1137 c) Testing transposase activity. (1) DNA ladder with 5,000 bp, 1,500 bp, and 500 bp bands
1138 marked, (2) DNA treated with full transposase reaction, (3) DNA with transposase reaction
1139 minus enzyme, (4)/(5) same as (2)/(3) with column-based kit clean-up, (6) input metagenomic
1140 DNA without treatment.

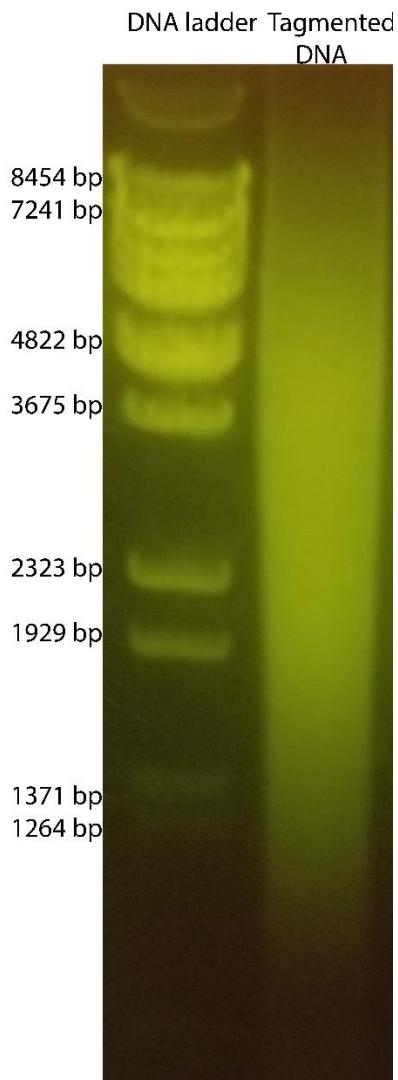


1141
1142 **Supplemental figure 2. Size selection and purification of soil metagenomic tagmented DNA**
1143 **and inverse PCR prepared linear pZE21-ME vector**
1144 Tagmented soil metagenomic DNA (left of ladder) and inverse PCR amplified pZE21-ME (right
1145 of ladder) were purified from agarose gel. DNA fragments were excised from ~1264 bp to ~8454
1146 bp and the major pZE21-ME band at ~2280 bp was excised.



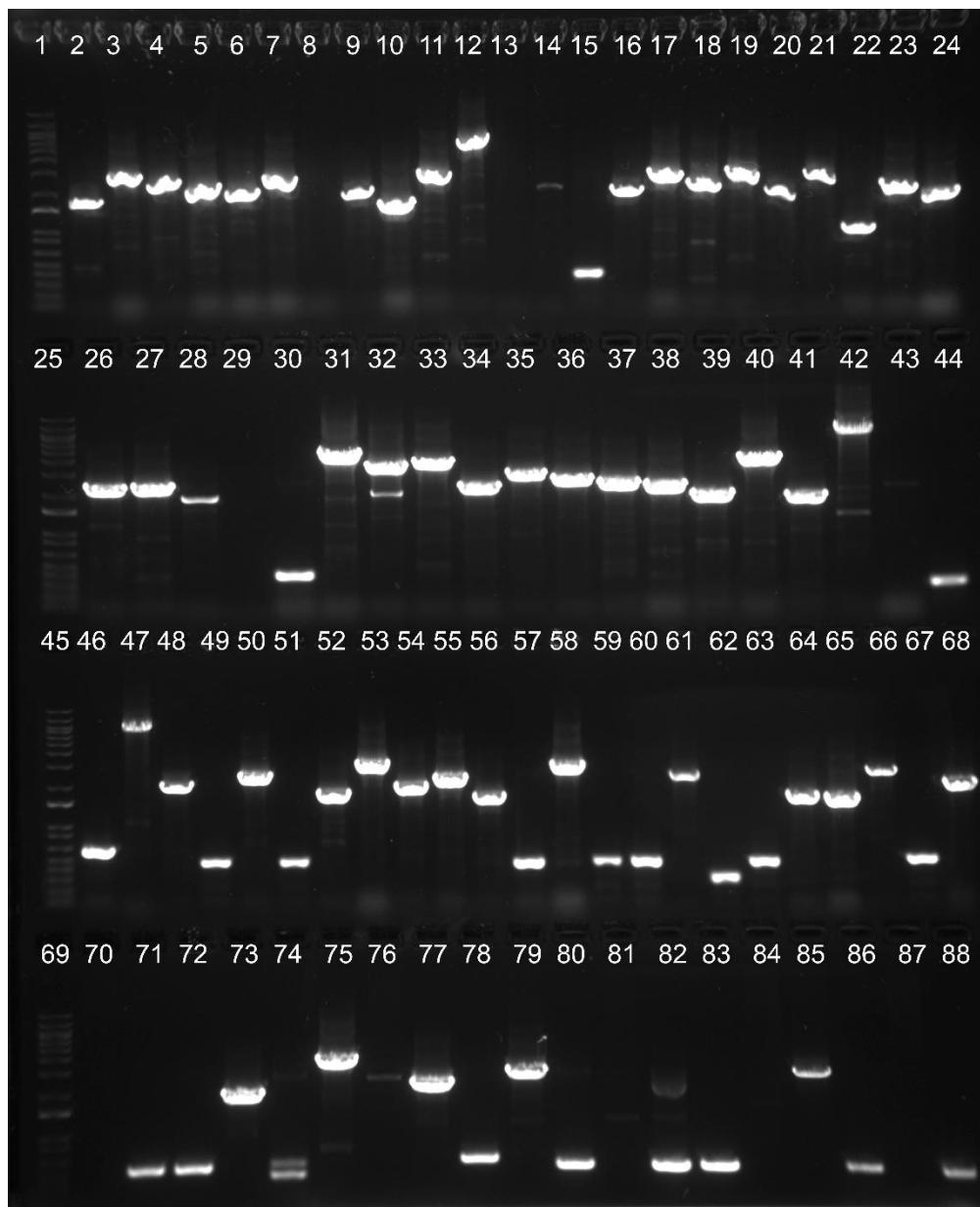
1147
1148 **Supplemental figure 3. Colony PCR to determine library size and average insert length of**
1149 **soil metagenomic DNA prepared by METa assembly (In-Fusion and NEBuilder HiFi) or**
1150 **blunt cloning**

1151 Lanes 1, 17, 33, and 49: DNA ladders with brightest bands corresponding to 5 kb, 1.5 kb, and
1152 500 bp. Lane 2: Single clone resulting from In-Fusion mediated METa assembly. Lanes 3-11,
1153 12-21, and 22-31: Colonies from three replicate METa assemblies using NEBuilder HiFi. Lanes
1154 32-41, 42-51, and 52-60: Colonies from triplicate blunt ligation cloning reactions. Lanes with
1155 bands at 500 bp correspond to amplification of vector backbone only (no inserts) while lanes
1156 with no band indicate colony PCR reaction failure.
1157
1158
1159
1160



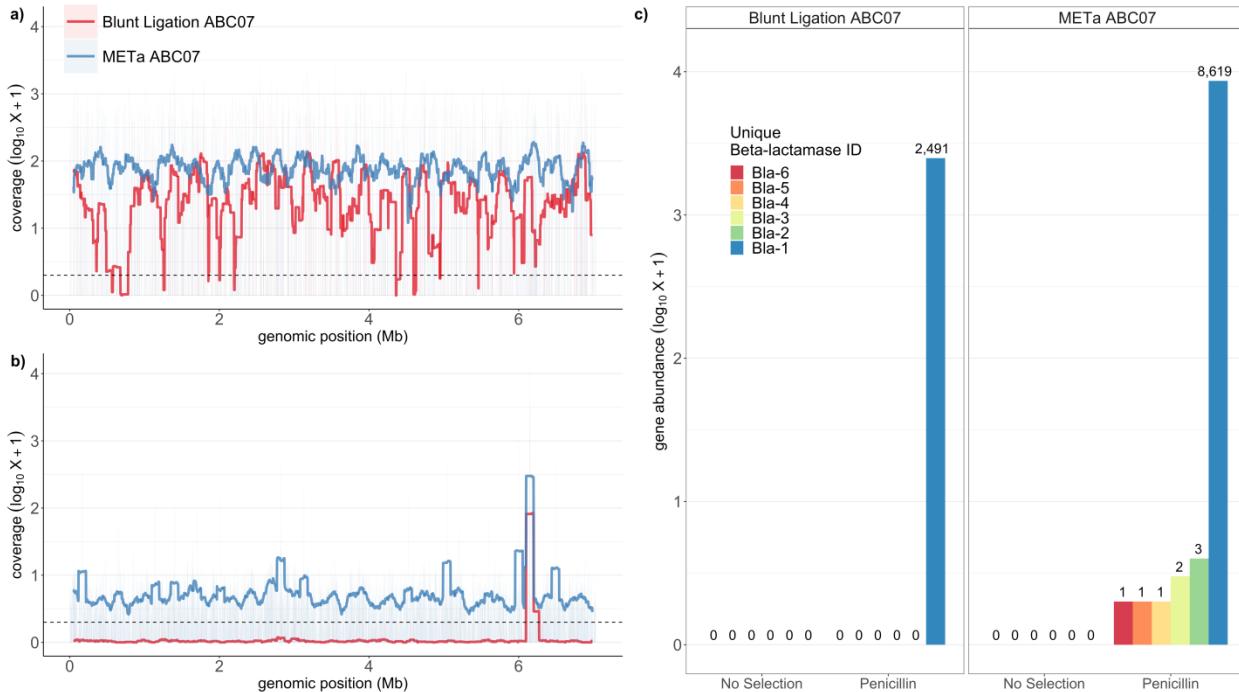
Supplemental figure 4. Size selection of mixed genomic tagmented DNA

Tagmented mixed ABC07, ABC10 genomic DNA was purified from agarose gel. DNA fragments were excised from ~1264 bp to ~8454 bp



1166
1167 **Supplemental figure 5. Colony PCR to determine library size and average insert length of**
1168 **mixed genome library prepared by METa assembly or blunt cloning**

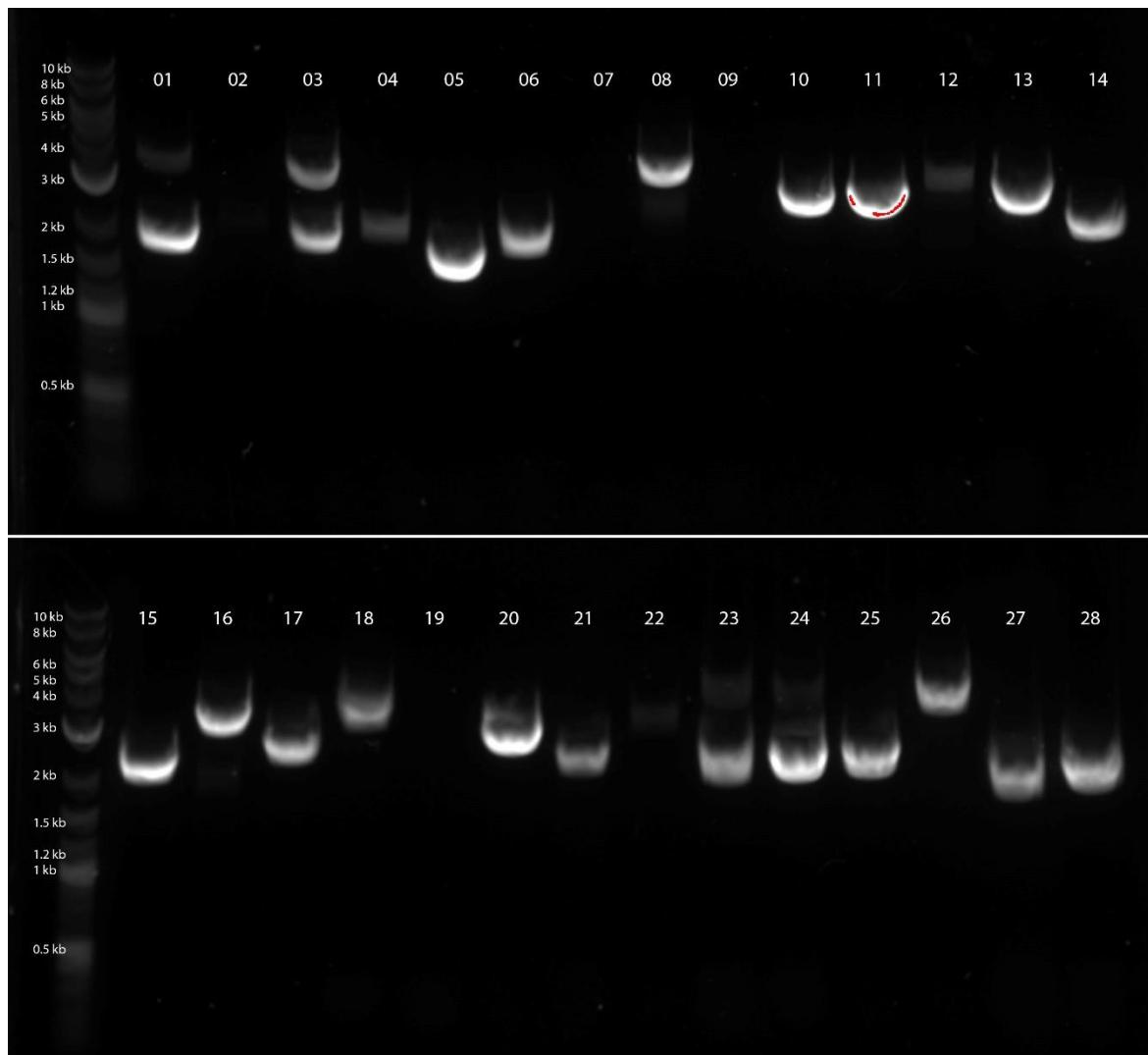
1169 Lanes 1, 25, 45, and 69: DNA ladders with top and brightest bands corresponding to 20 kb and
1170 1.5 kb, respectively. Lanes 2-24, 26-44: Colonies from three replicate METa assemblies using
1171 NEBuilder HiFi (13 colonies per replicate, lanes 15, 30, and 44 from negative control sham
1172 colonies). Lanes 46-68, 70-88: Colonies from triplicate blunt ligation cloning reactions (13
1173 colonies per replicate, lanes 59, 74, and 88 from negative control sham colonies). Reactions
1174 resulting in 500 bp amplicons indicate carriage of a no inserts vector. Reactions resulting in no
1175 visible band indicate technical failure of the colony PCR reaction.
1176



1177
1178 **Supplemental figure 6. Assembly and blunt ligation library coverage of ABC07 genome**
1179 **with and without penicillin selection**

1180 a) Nucleotide depth of coverage for ABC07 genome by functional metagenomic library prepared
1181 by assembly (blue) or blunt ligation (red). Coverage is smoothed to a 1 kb resolution.
1182 b) Same as in a) but sequenced libraries were first subjected to selection on agar plates
1183 containing 1 mg/ml penicillin.
1184 c) Gene abundance in post-penicillin selection reads for each of six predicted ABC07 β -
1185 lactamase genes that had read number >0. From Bla-1 to Bla-6 respective NCBI accession
1186 numbers are WP_087694996.1, WP_003211977.1, WP_003216184.1, WP_087694154.1,
1187 WP_008436310.1, WP_003207471.1.

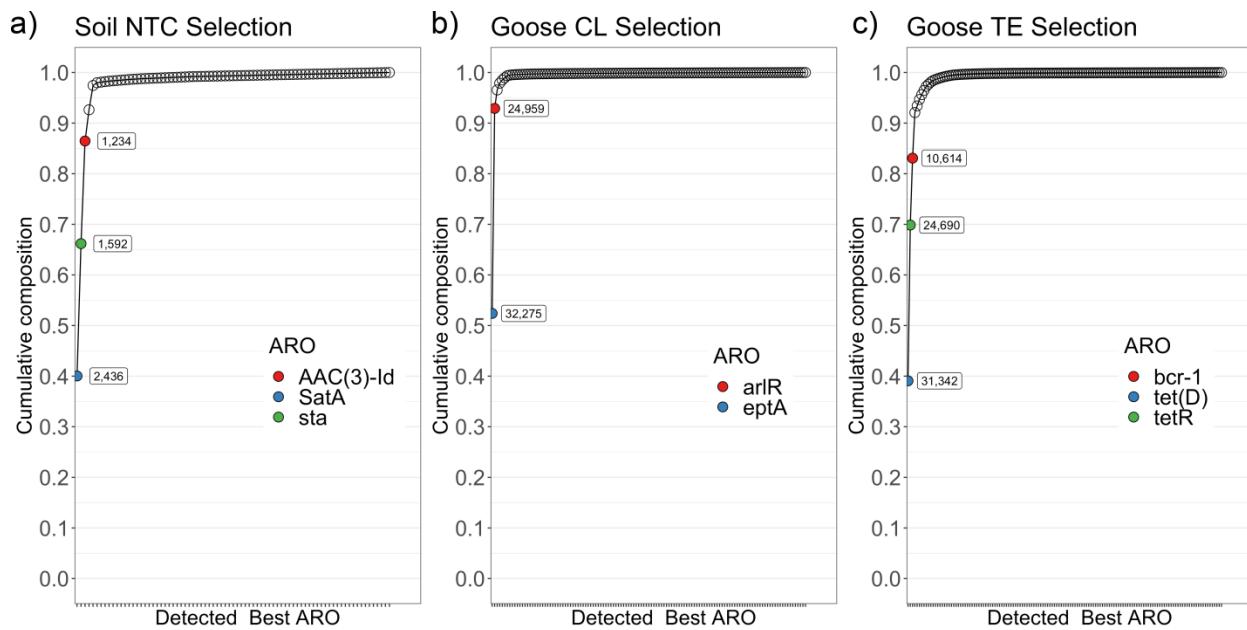
1188



1189
1190
1191
1192
1193
1194
1195
1196

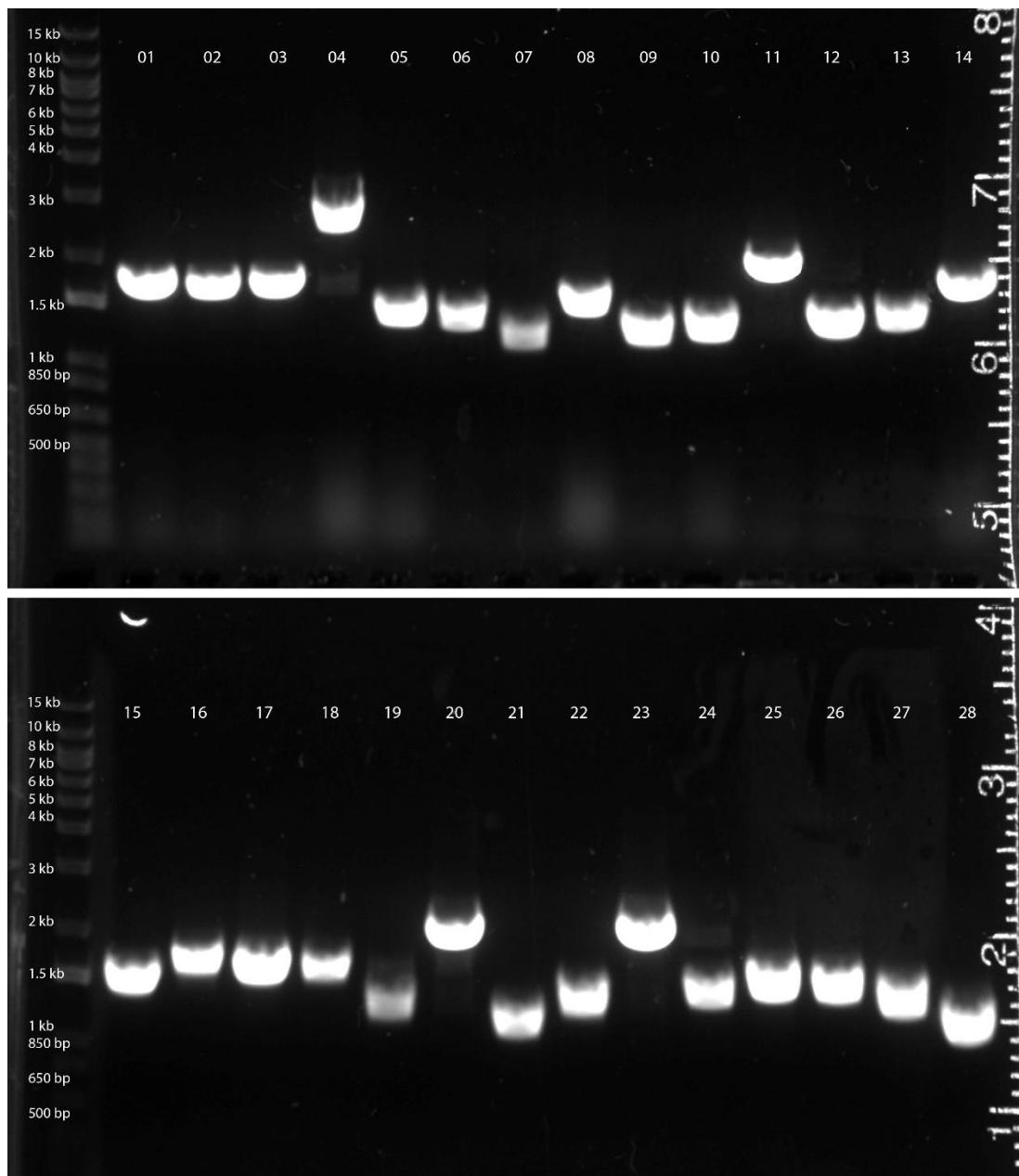
Supplemental figure 7. Representative colony PCR to determine average insert and library size of 5 µg soil library

The first lane of each gel contains ladder, numbered lanes correspond to 28 colony PCR reactions. Reactions for colonies 2, 7, 9, and 19 showed no amplicon indicating reaction failure. Lack of reactions producing 500 bp amplicons indicates all successful reactions contained an insert.



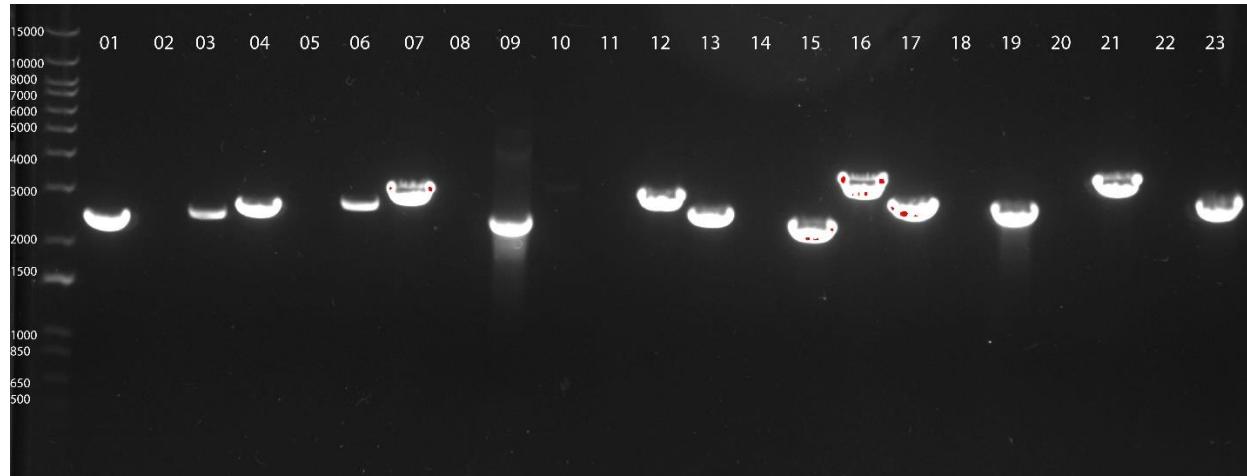
Supplemental figure 8. Antibiotic resistance gene family abundance

Major gene families predicted by CARD Antibiotic Resistance Ontology for a) soil microbiome selected on nourseothricin (NT) or goose gut microbiome selected on b) colistin (CL) or c) tetracycline (TE).



1203
1204 **Supplemental figure 9. Colony PCR to determine average insert size and library size of 10**
1205 **kb amplicon test library**

1206 The first lane of each top and bottom gel contain ladder, numbered lanes correspond to 28 colony
1207 PCR reactions. All tested colonies contained plasmids with inserts as indicated by all amplicons
1208 exceeding 500 bp in size.
1209



1210
1211 **Supplemental figure 10. Colony PCR to determine average insert size and library size of**
1212 **250 ng soil functional metagenomic library**

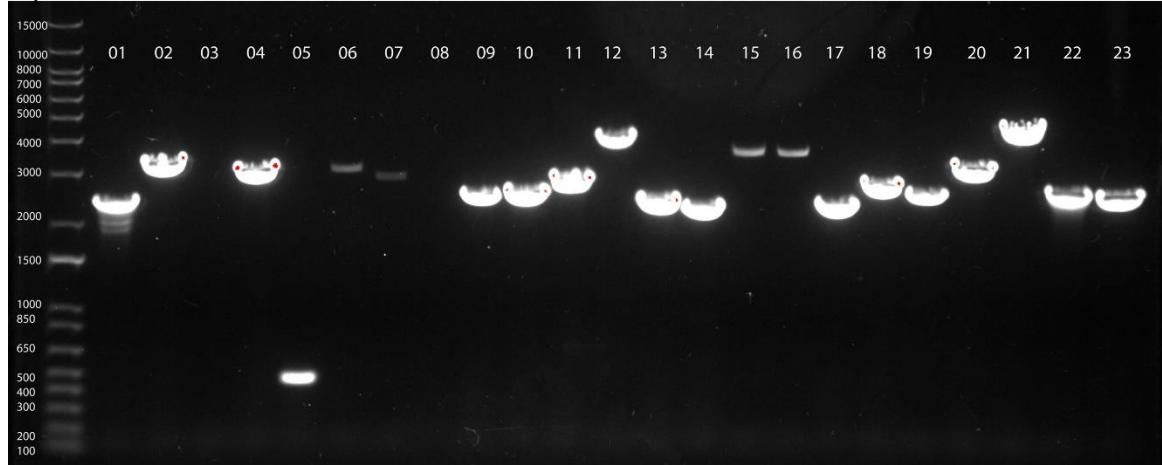
1213 The first lane contains ladder, numbered lanes correspond to 24 colony PCR reactions. Reactions
1214 2, 5, 8, 10, 11, 14, 18, 20, and 22 failed. All successful amplification reactions contained
1215 plasmids with inserts as indicated by all amplicons exceeding 500 bp in size.

1216
1217
1218
1219

a)



b)

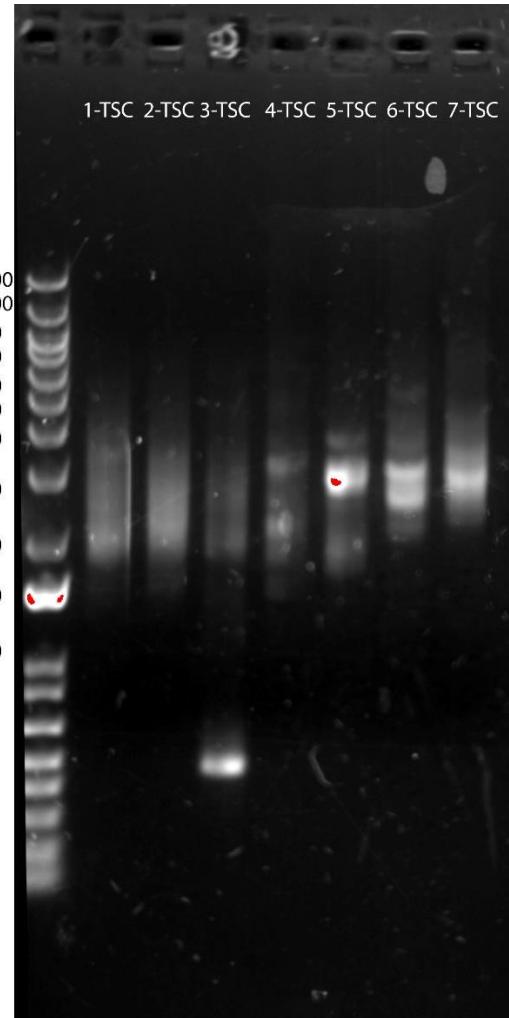


1220

1221 **Supplemental figure 11. Low input goose fecal microbiome functional metagenomic library**

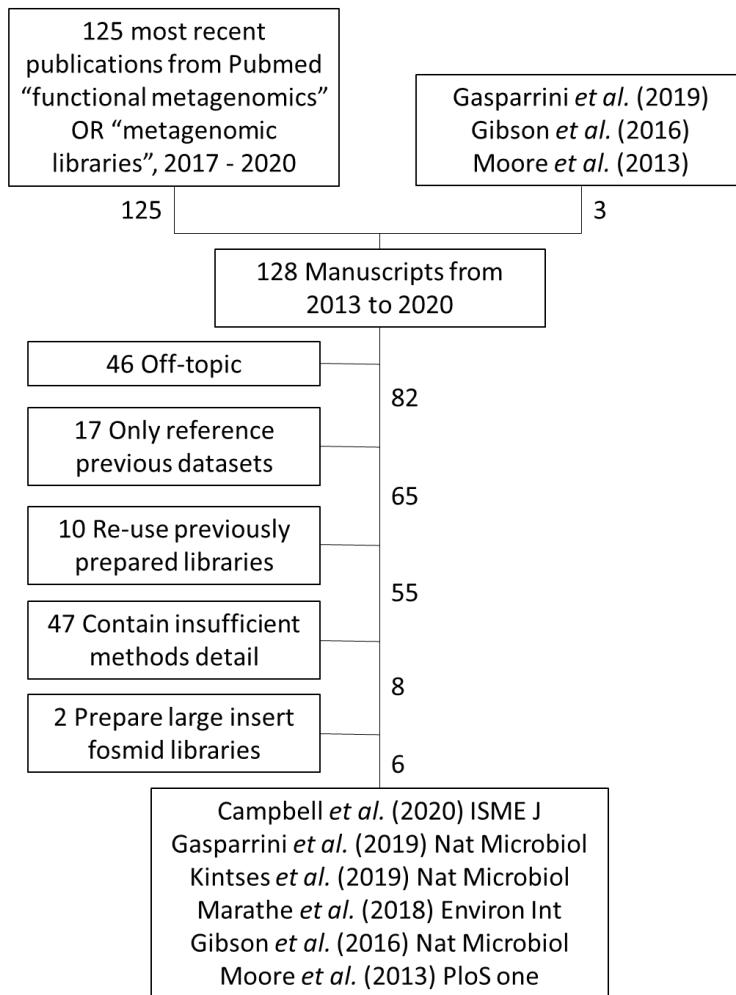
1222 a) The gaggle of geese putatively identified as *Branta canadensis* from which the fecal pellet
1223 was obtained.

1224 b) Colony PCR to identify average insert size and estimate total library size. The first lane
1225 contains ladder and the remaining lanes contain PCR reactions performed on 23 library colonies.
1226 Reactions 3 and 8 failed while the 500 bp amplicon in reaction 5 corresponds to 1/21 colonies
1227 containing no insert.



1228
1229 **Supplemental figure 12. Amplicons from functional metagenomic library minipreps for**
1230 **sequencing**

1231 From left to right: Ladder, (1-TSC) pooled triplicate ABC07/10 library prepared by METa
1232 assembly, no antibiotics, (2-TSC) same as 1-TSC following selection on 1 mg/ml penicillin, (3-
1233 TSC) pooled triplicate ABC07/10 library prepared by blunt ligation, no antibiotics, (4-TSC)
1234 same as 3-TSC following selection on 1 mg/ml penicillin, (5-TSC) 5 µg soil microbiome library
1235 following selection on 64 µg/ml nourseothricin, (6-TSC) 300 ng goose fecal microbiome library
1236 following selection on 8 µg/ml tetracycline, (7-TSC) same as 6-TSC but selection on 4 µg/ml
1237 colistin.
1238



1239

1240 **Supplemental figure 13. Literature search for functional metagenomic library preparation**
1241 **details**

1242 See methods for details.

1243

1244

1245

1246

1247

1248

1249

1250

1251

1252 **REFERENCES**

- 1253 1. Dick G. 2018. Genomic Approaches in Earth and Environmental Sciences. John Wiley &
1254 Sons, Ltd, Chichester, UK.
- 1255 2. Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono
1256 KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J,
1257 Handelsman J, Goodman RM. 2000. Cloning the Soil Metagenome: a Strategy for
1258 Accessing the Genetic and Functional Diversity of Uncultured Microorganisms. Appl
1259 Environ Microbiol 66:2541–2547.
- 1260 3. Milshteyn A, Schneider JS, Brady SF. 2014. Mining the Metabiome: Identifying Novel
1261 Natural Products from Microbial Communities. Chem Biol 21:1211–1223.
- 1262 4. Chittim C, Irwin SM, Balskus EP. 2018. Deciphering human gut microbiota- nutrient
1263 interactions : a role for biochemistry. Biochemistry 57:2567–2577.
- 1264 5. Koppel N, Maini Rekdal V, Balskus EP. 2017. Chemical transformation of xenobiotics by
1265 the human gut microbiota. Science 356.
- 1266 6. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. 1998. Molecular
1267 biological access to the chemistry of unknown soil microbes: a new frontier for natural
1268 products. Chem Biol 5:R245-9.
- 1269 7. Boolchandani M, D’Souza AW, Dantas G. 2019. Sequencing-based methods and
1270 resources to study antimicrobial resistance. Nat Rev Genet 20:356–370.
- 1271 8. Zallot R, Oberg N, Gerlt JA. 2019. The EFI Web Resource for Genomic Enzymology
1272 Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel
1273 Enzymes and Metabolic Pathways. Biochemistry 58:4169–4182.
- 1274 9. Cohen LJ, Kang H-S, Chu J, Huang Y-H, Gordon EA, Reddy BVB, Ternei MA, Craig

- 1275 JW, Brady SF. 2015. Functional metagenomic discovery of bacterial effectors in the
1276 human microbiome and isolation of commendamide, a GPCR G2A/132 agonist. Proc Natl
1277 Acad Sci 201508737.
- 1278 10. Adu-Oppong B, Gasparini AJ, Dantas G. 2016. Genomic and functional techniques to
1279 mine the microbiome for novel antimicrobials and antimicrobial resistance genes. Ann N
1280 Y Acad Sci 1–17.
- 1281 11. Ufarté L, Laville E, Duquesne S, Morgavi D, Robe P, Klopp C, Rizzo A, Pizzut-Serin S,
1282 Potocki-Veronese G. 2017. Discovery of carbamate degrading enzymes by functional
1283 metagenomics. PLoS One 12:e0189201.
- 1284 12. Colin P-Y, Kintses B, Gielen F, Miton CM, Fischer G, Mohamed MF, Hyvönen M,
1285 Morgavi DP, Janssen DB, Hollfelder F. 2015. Ultrahigh-throughput discovery of
1286 promiscuous enzymes by picodroplet functional metagenomics. Nat Commun 6:10008.
- 1287 13. Forsberg KJ, Patel S, Witt E, Wang B, Ellison TD, Dantas G. 2015. Identification of
1288 Genes Conferring Tolerance to Lignocellulose-Derived Inhibitors by Functional
1289 Selections in Soil Metagenomes. Appl Environ Microbiol 82:528–37.
- 1290 14. Sommer MOA, Church GM, Dantas G. 2010. A functional metagenomic approach for
1291 expanding the synthetic biology toolbox for biomass conversion. Mol Syst Biol 6:360.
- 1292 15. Uchiyama T, Miyazaki K. 2009. Functional metagenomics for enzyme discovery:
1293 challenges to efficient screening. Curr Opin Biotechnol 20:616–622.
- 1294 16. Handelsman J. 2004. Metagenomics: Application of Genomics to Uncultured
1295 Microorganisms. Microbiol Mol Biol Rev 68:669–685.
- 1296 17. Helm E Van Der, Genee HJ, Sommer MOA. 2018. The evolving interface between
1297 synthetic biology and functional metagenomics. Nat Chem Biol 1–9.

- 1298 18. Gudeta DD, Bortolaia V, Pollini S, Docquier J-D, Rossolini GM, Amos GCA, Wellington
1299 EMH, Guardabassi L. 2016. Expanding the Repertoire of Carbapenem-Hydrolyzing
1300 Metallo- β -Lactamases by Functional Metagenomic Analysis of Soil Microbiota. *Front
1301 Microbiol* 7:1–9.
- 1302 19. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, Dantas G. 2014.
1303 Bacterial phylogeny structures soil resistomes across habitats. *Nature* 509:612–6.
- 1304 20. Pehrsson EC, Tsukayama P, Patel S, Mejía-Bautista M, Sosa-Soto G, Navarrete KM,
1305 Calderon M, Cabrera L, Hoyos-Arango W, Bertoli MT, Berg DE, Gilman RH, Dantas G.
1306 2016. Interconnected microbiomes and resistomes in low-income human habitats. *Nature*
1307 533:212–6.
- 1308 21. Gasparrini AJ, Wang B, Sun X, Kennedy EA, Hernandez-Leyva A, Ndao IM, Tarr PI,
1309 Warner BB, Dantas G. 2019. Persistent metagenomic signatures of early-life
1310 hospitalization and antibiotic treatment in the infant gut microbiota and resistome. *Nat
1311 Microbiol* 4:2285–2297.
- 1312 22. Torres-Cortés G, Millán V, Ramírez-Saad HC, Nisa-Martínez R, Toro N, Martínez-
1313 Abarca F. 2011. Characterization of novel antibiotic resistance genes identified by
1314 functional metagenomics on soil samples. *Environ Microbiol* 13:1101–1114.
- 1315 23. Wichmann F, Udikovic-Kolic N, Andrew S, Handelsman J. 2014. Diverse antibiotic
1316 resistance genes in dairy cow manure. *MBio* 5.
- 1317 24. Riesenfeld CS, Goodman RM, Handelsman J. 2004. Uncultured soil bacteria are a
1318 reservoir of new antibiotic resistance genes. *Environ Microbiol* 6:981–989.
- 1319 25. van der Helm E, Imamovic L, Hashim Ellabaan MM, van Schaik W, Koza A, Sommer
1320 MOA. 2017. Rapid resistome mapping using nanopore sequencing. *Nucleic Acids Res*

- 1321 45:e61.
- 1322 26. Perron GG, Whyte L, Turnbaugh PJ, Goordial J, Hanage WP, Dantas G, Desai MM. 2015.
1323 Functional Characterization of Bacteria Isolated from Ancient Arctic Soil Exposes
1324 Diverse Resistance Mechanisms to Modern Antibiotics. PLoS One 10:e0069533.
- 1325 27. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. 2012. The shared
1326 antibiotic resistome of soil bacteria and human pathogens. Science 337:1107–11.
- 1327 28. Moore AM, Patel S, Forsberg KJ, Wang B, Bentley G, Razia Y, Qin X, Tarr PI, Dantas G.
1328 2013. Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes.
1329 PLoS One 8:e78822.
- 1330 29. Forsberg KJ, Patel S, Wencewicz TA, Dantas G. 2015. The Tetracycline Destructases: A
1331 Novel Family of Tetracycline-Inactivating Enzymes. Chem Biol 1–10.
- 1332 30. Cameron A, Barbieri R, Read R, Church D, Adator EH, Zaheer R, McAllister TA. 2019.
1333 Functional screening for triclosan resistance in a wastewater metagenome and isolates of
1334 Escherichia coli and Enterococcus spp. from a large Canadian healthcare region. PLoS
1335 One 14:e0211144.
- 1336 31. Tao W, Lee MH, Wu J, Kim NH, Kim JC, Chung E, Hwang EC, Lee S-W. 2012.
1337 Inactivation of chloramphenicol and florfenicol by a novel chloramphenicol hydrolase.
1338 Appl Environ Microbiol 78:6295–6301.
- 1339 32. Allen HK, Moe LA, Rodbumrer J, Gaarder A, Handelsman J. 2009. Functional
1340 metagenomics reveals diverse B-lactamases in a remote Alaskan soil. ISME J 3:243–251.
- 1341 33. Sommer MOA, Dantas G, Church GM. 2009. Functional characterization of the antibiotic
1342 resistance reservoir in the human microflora. Science 325:1128–31.
- 1343 34. Martiny AC, Martiny JBH, Weihe C, Field A, Ellis JC. 2011. Functional metagenomics

- 1344 reveals previously unrecognized diversity of antibiotic resistance genes in gulls. *Front*
1345 *Microbiol* 2:1–11.
- 1346 35. Campbell TP, Sun X, Patel VH, Sanz C, Morgan D, Dantas G. 2020. The microbiome and
1347 resistome of chimpanzees, gorillas, and humans across host lifestyle and geography. *ISME*
1348 *J* 14:1584–1599.
- 1349 36. Crofts TS, Gasparrini AJ, Dantas G. 2017. Next-generation approaches to understand and
1350 combat the antibiotic resistome. *Nat Rev Microbiol* 15:422–434.
- 1351 37. Boolchandani M, Patel S, Dantas G. 2017. *Antibiotics*. Springer New York, New York,
1352 NY.
- 1353 38. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie
1354 AP, Caruccio NC, Zhang X, Shendure J. 2010. Rapid, low-input, low-bias construction of
1355 shotgun fragment libraries by high-density in vitro transposition. *Genome Biol* 11.
- 1356 39. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony RK. 2015.
1357 Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One*
1358 10:1–15.
- 1359 40. Marine R, Polson SW, Ravel J, Hatfull G, Russell D, Sullivan M, Syed F, Dumas M,
1360 Wommack KE. 2011. Evaluation of a Transposase Protocol for Rapid Generation of
1361 Shotgun High-Throughput Sequencing Libraries from Nanogram Quantities of DNA.
1362 *Appl Environ Microbiol* 77:8071–8079.
- 1363 41. Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014. Tn5
1364 transposase and tagmentation procedures for massively scaled sequencing projects.
1365 *Genome Res* 24:2033–40.
- 1366 42. Hennig BP, Velten L, Racke I, Tu CS, Thoms M, Rybin V, Besir H, Remans K, Steinmetz

- 1367 LM. 2018. Large-Scale Low-Cost NGS Library Preparation Using a Robust Tn5
- 1368 Purification and Tagmentation Protocol. *G3* (Bethesda) 8:79–89.
- 1369 43. Rinke C, Low S, Woodcroft BJ, Raina J, Skarszewski A, Le XH, Butler MK, Stocker R,
- 1370 Seymour J, Tyson GW, Hugenholtz P. 2016. Validation of picogram- and femtogram-
- 1371 input DNA libraries for microscale metagenomics. *PeerJ* 4:e2486.
- 1372 44. Gibson DG, Young L, Chuang R, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic
- 1373 assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6:343–5.
- 1374 45. Studier FW. 2005. Protein production by auto-induction in high density shaking cultures.
- 1375 *Protein Expr Purif* 41:207–34.
- 1376 46. Crofts TS, Wang B, Spivak A, Gianoulis TA, Forsberg KJ, Gibson MK, Johnsly LA,
- 1377 Broomall SM, Rosenzweig CN, Skowronski EW, Gibbons HS, Sommer MOA, Dantas G.
- 1378 2017. Draft Genome Sequences of Three β -Lactam-Catabolizing Soil Proteobacteria.
- 1379 *Genome Announc* 5:8–10.
- 1380 47. Crofts TS, Wang B, Spivak A, Gianoulis TA, Forsberg KJ, Gibson MK, Johnsly LA,
- 1381 Broomall SM, Rosenzweig CN, Skowronski EW, Gibbons HS, Sommer MOA, Dantas G.
- 1382 2018. Shared strategies for β -lactam catabolism in the soil microbiome. *Nat Chem Biol*
- 1383 14:556–564.
- 1384 48. Dantas G, Sommer MOA, Oluwasegun RD, Church GM. 2008. Bacteria subsisting on
- 1385 antibiotics. *Science* 320:100–3.
- 1386 49. Lutz R, Bujard H. 1997. Independent and tight regulation of transcriptional units in
- 1387 *escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic*
- 1388 *Acids Res* 25:1203–1210.
- 1389 50. Sambrook J, Russell DW. 2006. Transformation of *E. coli* by Electroporation. *Cold*

- 1390 Spring Harb Protoc 2006:pdb.prot3933.
- 1391 51. Cao J, Hu Y, Liu F, Wang Y, Bi Y, Lv N, Li J, Zhu B, Gao GF. 2020. Metagenomic
1392 analysis reveals the microbiome and resistome in migratory birds. *Microbiome* 8:26.
- 1393 52. Lehman SS, Mladinich KM, Boonyakanog A, Mima T, Karkhoff-Schweizer RR,
1394 Schweizer HP. 2016. Versatile nourseothricin and streptomycin/spectinomycin resistance
1395 gene cassettes and their use in chromosome integration vectors. *J Microbiol Methods*
1396 129:8–13.
- 1397 53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
1398 Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
1399 25:2078–2079.
- 1400 54. Zeileis A, Grothendieck G. 2005. Zoo: S3 infrastructure for regular and irregular time
1401 series. *J Stat Softw* 14.
- 1402 55. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W,
1403 Nguyen AL V., Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE,
1404 Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva
1405 A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire
1406 F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar G V., McArthur AG.
1407 2020. CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic
1408 resistance database. *Nucleic Acids Res* 48:D517–D525.
- 1409 56. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*
1410 30:2068–2069.
- 1411 57. Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for
1412 the analysis of massive data sets. *Nat Biotechnol* 35:1026–1028.

- 1413 58. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H,
1414 Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P. 2019. EggNOG 5.0: A
1415 hierarchical, functionally and phylogenetically annotated orthology resource based on
1416 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314.
- 1417 59. UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res*
1418 43:D204-12.
- 1419 60. Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New
1420 York, NY.
- 1421 61. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997.
1422 Gapped BLAST and PSI-BLAST: A new generation of protein database search programs.
1423 *Nucleic Acids Res* 25:3389–3402.
- 1424 62. Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: A web server for clustering
1425 and comparing biological sequences. *Bioinformatics* 26:680–682.
- 1426 63. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
1427 Improvements in performance and usability. *Mol Biol Evol* 30:772–780.
- 1428 64. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood
1429 trees for large alignments. *PLoS One* 5.
- 1430 65. Yu G. 2020. Using ggtree to Visualize Data on Tree-Like Structures. *Curr Protoc*
1431 *Bioinforma* 69:1–18.
- 1432 66. Kintses B, Méhi O, Ari E, Számel M, Györkei Á, Jangir PK, Nagy I, Pál F, Fekete G,
1433 Tengölcs R, Nyerges Á, Likó I, Bálint A, Molnár T, Bálint B, Vásárhelyi BM,
1434 Bustamante M, Papp B, Pál C. 2019. Phylogenetic barriers to horizontal transfer of
1435 antimicrobial peptide resistance genes in the human gut microbiota. *Nat Microbiol* 4:447–

- 1436 458.
- 1437 67. Marathe NP, Janzon A, Kotsakis SD, Flach C-F, Razavi M, Berglund F, Kristiansson E,
1438 Larsson DGJ. 2018. Functional metagenomics reveals a novel carbapenem-hydrolyzing
1439 mobile beta-lactamase from Indian river sediments contaminated with antibiotic
1440 production waste. Environ Int 112:279–286.
- 1441 68. Gibson MK, Wang B, Ahmadi S, Burnham C-AD, Tarr PI, Warner BB, Dantas G. 2016.
1442 Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome.
1443 Nat Microbiol 1:16024.
- 1444 69. Irwin CR, Farmer A, Willer DO, Evans DH. 2012. In-fusion® cloning with vaccinia virus
1445 DNA polymerase. Methods Mol Biol 890:23–35.
- 1446 70. Gasparrini AJ, Markley JL, Kumar H, Wang B, Fang L, Irum S, Symister CT, Wallace M,
1447 Burnham CAD, Andleeb S, Tolia NH, Wencewicz TA, Dantas G. 2020. Tetracycline-
1448 inactivating enzymes from environmental, human commensal, and pathogenic bacteria
1449 cause broad-spectrum tetracycline resistance. Commun Biol 3.
- 1450 71. Waksman SA, Woodruff HB. 1942. Streptothricin, a New Selective Bacteriostatic and
1451 Bactericidal Agent, Particularly Active Against Gram-Negative Bacteria. Proc Soc Exp
1452 Biol Med 49:207–210.
- 1453 72. Baltz RH. 2007. Antimicrobials from actinomycetes: Back to the future. Microbe 2:125–
1454 131.
- 1455 73. Cox G, Sieron A, King AM, De Pascale G, Pawlowski AC, Kotova K, Wright GD. 2017.
1456 A Common Platform for Antibiotic DerePLICATION and Adjuvant Discovery. Cell Chem
1457 Biol 24:98–109.
- 1458 74. Krügel H, Fiedler G, Smith C, Baumberg S. 1993. Sequence and transcriptional analysis

- 1459 of the nourseothricin acetyltransferase-encoding gene *nat1* from *Streptomyces noursei*.
1460 Gene 127:127–131.
1461 75. Van Hoek AHAM, Mevius D, Guerra B, Mullany P, Roberts AP, Aarts HJM. 2011.
1462 Acquired antibiotic resistance genes: An overview. Front Microbiol 2:1–27.
1463 76. Hamano Y, Matsuura N, Kitamura M, Takagi H. 2006. A novel enzyme conferring
1464 streptothricin resistance alters the toxicity of streptothricin D from broad-spectrum to
1465 bacteria-specific. J Biol Chem 281:16842–16848.
1466 77. Gabor EM, De Vries EJ, Janssen DB. 2003. Efficient recovery of environmental DNA for
1467 expression cloning by indirect extraction methods. FEMS Microbiol Ecol 44:153–163.
1468 78. Allemann A, Kraemer JG, Korten I, Ramsey K, Casaulta C, Wüthrich D, Ramette A,
1469 Endimiani A, Latzin P, Hilty M, Barben J, Casaulta C, Jung A, Kieninger E, Korten I,
1470 Latzin P, Moeller A, Mornand A, Hafen G, Müller-Suter D, Regamey N, Rochat I, Singer
1471 F, Spinas R, Trachsel D, Yammine S, Zanolari M. 2019. Nasal resistome development in
1472 infants with cystic fibrosis in the first year of life. Front Microbiol 10.
1473 79. Clemente JC, Pehrsson EC, Blaser MJ, Sandhu K, Gao Z, Wang B, Magris M, Hidalgo G,
1474 Contreras M, Noya-Alarcón Ó, Lander O, McDonald J, Cox M, Walter J, Oh PL, Ruiz JF,
1475 Rodriguez S, Shen N, Song SJ, Metcalf J, Knight R, Dantas G, Dominguez-Bello MG.
1476 2015. The microbiome of uncontacted Amerindians. Sci Adv 1.
1477 80. Biehl LM, Garzetti D, Farowski F, Ring D, Koeppel MB, Rohde H, Schafhausen P,
1478 Stecher B, Vehreschild MJGT. 2019. Usability of rectal swabs for microbiome sampling
1479 in a cohort study of hematological and oncological patients. PLoS One 14:e0215428.
1480