1 **Rapid ordering of barcoded transposon insertion libraries of anaerobic bacteria**

2

3 Anthony L. Shiver[1], Rebecca Culver[2], Adam M. Deutschbauer[3,4], Kerwyn Casey

4 Huang[1,5,6,*]

5

6 [1]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

7 [2]Department of Genetics, Stanford University, Stanford, CA 94305, USA

8 [3]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National

9 Laboratory, Berkeley, CA 94720, USA

10 [4]Department of Plant and Microbial Biology, University of California, Berkeley, CA

11 94720, USA

12 [5]Department of Microbiology and Immunology, Stanford University School of Medicine,

13 Stanford, CA 94305, USA

14 [6]Chan-Zuckerberg Biohub, San Francisco, CA 94158

15

16 [*]Correspondence: kchuang@stanford.edu

Abstract

Commensal bacteria from the human intestinal microbiota play important roles in health and disease. Research into the mechanisms by which these bacteria exert their effects is hampered by the complexity of the microbiota and by the strict growth requirements of the individual species. The assembly of ordered transposon insertion libraries, in which nearly all nonessential genes have been disrupted and the strains stored as independent monocultures, would be a transformative resource for research into many microbiota members. However, assembly of these libraries must be fast and inexpensive in order to empower investigation of the large number of species that typically compose gut communities. The methods used to generate ordered libraries must also be adapted to the anaerobic growth requirements of most intestinal bacteria. We have developed a protocol to assemble ordered libraries of transposon insertion mutants that is fast, cheap, and effective for even strict anaerobes. The protocol differs from currently available methods by making use of cell sorting to order the library and barcoded transposons to facilitate the localization of ordered mutations in the library. By tracking transposon insertions using barcode sequencing, our approach increases the accuracy and reduces the time and effort required to locate mutants in the library. Ordered libraries can be sorted and characterized over the course of two weeks using this approach. We expect this protocol will lower the barrier to generating comprehensive, ordered mutant libraries for many species in the human microbiota, allowing for new investigations into genotype-phenotype relationships within this important microbial ecosystem.

Keywords: anaerobic bacteria, human gut microbiota, ordered mutant library, transposon insertion, barcode, Bar-seq, Tn-seq, RB-TnSeq, high-throughput genetics, systems biology

INTRODUCTION

There is a growing appreciation of the role that commensal bacteria play in human health[1]. However, the complexity of the microbiota, the fastidious growth requirements of its members[2], and a lack of genetic tools and resources to study them have complicated efforts to unravel the mechanistic links between commensal bacteria and host health and disease. Ordered mutant libraries in which all non-essential genes have been comprehensively disrupted and individual mutant strains stored in monoculture have been invaluable resources for connecting genotype to phenotype, but such libraries exist for only a limited number of bacterial species[3-11]. This type of resource would be transformative for investigating the mechanistic basis for the effects of the gut microbiota on health; however, the generation of ordered mutant libraries must first be made rapid, inexpensive, and amenable to the anaerobic growth requirements of gut microbiota members.

Transposon mutagenesis is a highly successful mutagenesis approach for non-model organisms. A small number of well-characterized transposon systems (e.g., *mariner*[12] and Tn*5*[13]) are functional in diverse bacteria, including the dominant phyla of the human microbiota. However, transposon insertion libraries are typically generated and assayed as a random pool. Many aspects of a bacterium's interaction with a host cannot be readily studied in the context of a mutant pool, including phenotypes that can be complemented *in trans* such as the production of metabolites[14]. Thus, while transposon mutagenesis is a fruitful first approach for genetic analysis of under-studied gut microbiota members, a major challenge remains in ordering the random pool of transposon mutants in order to analyze individual strains.

Recent technological advances have vastly lowered the cost and effort required to order random transposon insertion libraries[15,16]. Pooling strategies that combine cultures prior to sequencing and then use computational methods to identify the positions of individual strains can reduce the number of samples that need to be processed by two orders of magnitude[5,7,10,16,17]. The strategy behind various pooling protocols can be separated into two parts. The first is pool design: how individual mutants are combined into larger pools to reduce the number of samples that must be processed, while maintaining the ability to deconvolve the locations of each mutant in the library. The second is how each pool is sequenced to identify the mutations contained within. Current sequencing approaches for transposon insertion libraries use a flavor of transposon insertion sequencing (Tn-seq/INSeq/ST-PCR). Because these sequencing approaches map transposon insertions for every individual pool, they remain laborious, often requiring multiple processing steps for each individual pool[7,10,15,16].

We have incorporated two further advances to facilitate sorting of obligate anaerobic bacteria and further lower the cost and effort of locating mutations within an ordered library. First, we use single cell sorting with a fluorescence-activated cell sorting (FACS) machine to order individual strains. Second, we make use of barcoded transposons and barcode sequencing to locate individual mutations within the ordered library. The first advance makes use of the fast speed at which single cells can be isolated using a cell

90   sorter (>5,000 mutants per hour) to order strains in high throughput while keeping
91   aerobic exposure time below 30 minutes. The second advance utilizes the reproducibility
92   and simplicity of barcode sequencing to reduce the effort required to locate mutations in
93   the ordered library.
94
95   We recently developed the use of barcoded transposons and barcode sequencing (RB-
96   TnSeq) as an alternative sequencing approach for high-throughput functional genomics[18].
97   With RB-TnSeq, the entire transposon insertion library is initially sequenced as one
98   sample, allowing association of random DNA barcodes with transposon insertions across
99   the entire pool. Subsequently, the abundance of transposon insertions can be followed
100 with a single PCR step to amplify barcodes followed by next-generation sequencing (Bar-
101 seq). This approach allowed us to meaure genome-wide mutant fitness data for 32 diverse
102 bacteria across dozens of growth conditions[19]. The same principles that make Bar-seq
103 useful for chemical genomics also make it an ideal sequencing approach for locating
104 mutations within an ordered library.
105
106 While Bar-seq simplifies the process of locating mutants within an ordered library, it
107 requires a diverse progenitor pool with a large number of unique barcodes. Therefore, the
108 main limitation of our protocol is that it is restricted to bacterial strains with a relatively
109 high transformation efficiency such that diverse pools of barcoded transposon insertions
110 can be generated. If a high efficiency transformation protocol can be identified for a
111 strain of interest, the use of Bar-seq promises to lower the barrier to generating an
112 ordered library from a random pool of barcoded transposon insertions.
113
114 We have incorporated flow sorting and barcode sequencing into a protocol (Figure 1) that
115 allows us to rapidly generate an ordered collection of >10,000 barcoded transposon
116 insertion mutants in strictly anaerobic bacteria such as those found in the human gut
117 microbiota. To develop this protocol, we first demonstrated that strict anaerobes can
118 survive the temporary exposure to oxygen that occurs during single-cell sorting with a
119 FACS machine, particularly in the presence of specific antioxidants. We capitalized on
120 this finding to develop a protocol built around sorting single cells from a pooled mutant
121 library. We optimized this protocol to reduce the time, cost, and levels of cross-
122 contamination. Finally, we used this protocol to order a barcoded transposon insertion
123 library in the model human gut commensal *Bacteroides thetaiotaomicron* VPI-5482.
124 Analysis of individual mutants from this ordered library allowed us to further
125 characterize chemical sensitivities that were detected from a pooled chemical-genomics
126 screen of *B. thetaiotaomicron* VPI-5482[20].
127
128 Given a diverse transposon insertion pool, both the sorting portion of the protocol and the
129 strain identification portion can be accomplished in two weeks. As transformation
130 methods are developed for more bacterial residents of the human microbiota, we expect
131 that this protocol will enable the rapid and cost-effective investigation of microbe-host
132 interactions, metabolite production, and other genotype-phenotype relationships in an
133 ever-expanding set of bacteria.
134
135 **EXPERIMENTAL DESIGN**

Survival of anaerobic bacteria during sorting

To determine whether we could use cell sorting to generate ordered libraries of anaerobic bacteria, we first sought to quantify the ability of bacteria from the gut microbiota, particularly strict anaerobes, to survive the process of cell sorting. We chose two Bacteroidetes (*Bacteroides finegoldii* and *Parabacteroides johnsonii*), two Firmicutes (*Lactobacillus reuteri* and *Clostridium symbiosum*), and two Actinobacteria (*Bifidobacterium breve* and *Bifidobacterium longum*) to test survival of the sorting process; both *Bifidobacterium* species, *C. symbiosum, P. johnsonii*, and *B. finegoldii* are all obligate anaerobes while *L. reuteri* is a facultative anaerobe[21]. As a control, the facultative anaerobe *Escherichia coli* BW25113 was also sorted. Strains were grown and sorted into commercial media formulations of MRS (*B. breve* and *B. longum*), RCM supplemented with hemin, menadione, tryptophan, and arginine (*C. symbiosum*) and GAM supplemented with hemin, menadione, tryptophan, and arginine (*P. johnsonii, B. finegoldii, L. reuteri*, and *E. coli*).

All plates and cultures spent 1-1.5 h in an aerobic atmosphere while single cells were sorted into five 96-microwell plates for each strain. We counted the number of blank wells after outgrowth in each 96-well plate as a proxy for the number of single cells that did not survive the sorting procedure. We found that even the strict anaerobes survived the aerobic exposure long enough to complete the cell sorting protocol and regrow at high frequency when returned to anaerobic conditions. The fraction of blank wells after sorting varied across species, with *C. symbiosum* having the highest fraction of blank wells (**Figure 2a**); nonetheless, at least 89% of *C. symbiosum* cells grew after sorting.

Each of the commercial media formulations we used contained a significant concentration of cysteine, which we hypothesized was acting as an antioxidant to protect the cells during their transient exposure to oxygen. To test this hypothesis and explore the utility of alternative antioxidants, we sorted *C. symbiosum* and *Clostridium innocuum* in formulations of RCM in which cysteine had been either omitted or replaced with another antioxidant. We tested cysteine, thioglycolic acid (a component of GAM), and ascorbic acid for their ability to protect the two species. In the absence of any added antioxidants, the survival of both species was greatly reduced (Figure 2b), consistent with their sensitivity to oxygen. Both species were protected by the addition of antioxidants. Interestingly, *C. innocuum* preferred ascorbic acid over cysteine while *C. symbiosum* had the opposite preference (Figure 2b). Thioglycolate was the worst performing antioxidant of those tested for both species (Figure 2b). It remains to be determined if the optimal bacterium-antioxidant pair is idiosyncratic or if there is a general, as yet unidentified, protectant that can be applied across species. Nonetheless, it was straightforward to achieve high survival frequency for both of these strict anaerobes.

Multiple factors will be critical for successfully adopting the cell sorting protocol to a species of interest. First, the concentration of bacterium-sized inanimate particles must be low enough that sorting of these false signals does not appreciably reduce the fill rate of wells in the library. Filter sterilization of the growth medium using a 0.22-µm filter is effective for eliminating this source of error; however, particular species may produce

182 these types of particles (e.g., outer membrane vesicles, extracellular polysaccharides)
183 during the course of growth. Second, the majority of cells in the population must be
184 viable. Third, single cells have to survive the time spent in an aerobic environment that is
185 required to sort the library. Fourth, the majority of the cells in the population cannot exist
186 as part of multi-cell aggregates. The first three issues will lower the fill rate of wells in
187 the library, while the fourth problem will lower the fraction of wells in the library that
188 contain a single strain.
189
190 Each of these potential issues is likely to depend on the growth conditions of the library.
191 While extensive optimization of growth conditions for *B. thetaiotaomicron* was not
192 necessary, there are a number of avenues for troubleshooting the sorting step for difficult
193 bacteria. Changing the growth phase of the sorted cells (e.g. log phase versus early
194 stationary phase) is likely to alter the fraction of cells that survive sorting. Testing
195 different growth and recovery media may also be an important aspect of troubleshooting.
196 Finally, while cysteine may need to be avoided as an antioxidant for known hydrogen
197 sulfide producers (see below), there are other antioxidants that can be added as a
198 supplement to the growth and recovery media.
199
200 **Elimination of H_2S production**
201 CAUTION Hydrogen sulfide is a toxic, caustic gas. Build-up of hydrogen sulfide inside
202 the anaerobic chamber will degrade sensitive electronic equipment. Exposure to
203 hydrogen sulfide gas should be minimized. Before working with large volumes of
204 culture, determine the growth conditions that will prevent the production of hydrogen
205 sulfide during growth.
206
207 Enzymes with cysteine desulfhydrase activity degrade cysteine and produce hydrogen
208 sulfide, a toxic gas, as a byproduct. These enzymes are conserved across the eubacteria[22],
209 and *Bacteroides* sp. are known to produce hydrogen sulfide gas[21]. Given the large
210 volumes of culture involved in growing an ordered a library, even moderate production of
211 H_2S has the potential to overcome standard H_2S removal columns and pose a danger to
212 personnel and equipment. Our original recipe for the BHIS growth medium of
213 *Bacteroides thetaiotaomicron* included 8 mM cysteine as an antioxidant, and *B.*
214 *thetaiotaomicron* cultures produced significant levels of hydrogen sulfide in pilot
215 experiments.
216
217 To avoid the potential hazards of H_2S production, we first confirmed that the level of
218 hydrogen sulfide produced by *B. thetaiotaomicron* is proportional to the concentration of
219 added cysteine. We held hydrogen sulfide test strips over 5 mL cultures of *B.*
220 *thetaiotaomicron* with increasing amounts of cysteine. Reaction of hydrogen sulfide with
221 lead acetate in the test strip led to the black lead sulfide product (**Figure 3a**). We then
222 tested growth of *B. thetaiotaomicron* in BHIS as a function of added cysteine and found
223 that cysteine could be left out of the BHIS recipe without adverse effects on anaerobic
224 growth. The growth curve of *B. thetaiotaomicron* grown in BHIS without cysteine was
225 similar to BHIS with cysteine added (**Figure 3b**). Calculation of the maximum growth
226 rate and the maximum OD_{600} (optical density at 600 nm) also showed no significant
227 differences between BHIS with or without added cysteine (**Figure 3c**). From these results

228   and the known oxygen tolerance of *B. thetaiotaomicron*, we decided to proceed without
229   cysteine for sorting the transposon insertion library.

230 **MATERIALS**
231
232 REAGENTS
233 Media
234 - Bacto^{TM} Brain Heart Infusion (BHI) (Becton Dickinson, #237500)
235 - Sodium bicarbonate (Fisher Bioreagents, #BP328-500)
236 - Hemin (porcine) (Alfa Aesar, A11165)
237 - Erythromycin (Sigma, #E5389-5G)
238 - Glycerol (Fisher Bioreagents, #BP229-1)
239
240 Chemicals
241 - Sodium hydroxide (Fisher Bioreagents, #S318-500)
242 - Hydrochloric acid (Fisher Chemical, #A144-500)
243
244 Plasticware and Consumables
245 - Thermo Scientific^{TM} Nalgene^{TM} Rapid-Flow^{TM} sterile disposable bottle top filters
246   with a PES membrane, 45-mm neck size (Thermo Scientific^{TM}, #5974520)
247   **CRITICAL** Polyether sulfone (PES) is superior to cellulose acetate (CA),
248   surfactant-free cellulose acetate (SCFA), and polyvinylidene fluoride (PVDF) for
249   filtering complex media like BHI without clogging.
250 - PlateOne 2-mL 96-deepwell polypropylene plates (USA Scientific, #1896-2000)
251 - Sterile 96-microwell V-bottom polystyrene plates (Grenier bio-one, #651161)
252 - Sterile microwell plate lids (Grenier Bio-One, #656101)
253 - Standard profile polypropylene reservoirs (Mettler-Toledo, # LR-R2-PB-5)
254 - Breathe-Easier gas permeable sealing membranes (Diversified Biotech, #BERM-
255   2000)
256 - 5 mL polypropylene round bottom tubes (Falcon, #352063)
257 - Vinyl tape (3M, #471)
258 - Nunc Aluminum Seal Tape for 96-well plates (Thermo Scientific^{TM}, #232698)
259 - Alumaseal 384 films (Excel Scientific, #F-384-100)
260 - 1-mL LTS filter tips, for use with BenchSmart (Mettler Toledo, #30296782)
261 - 200-µL LTS filter tips, for use with BenchSmart (Mettler Toledo, #17010646)
262 - 200-µL filter tips, for use with multichannel pipette (Rainin #30389239)
263 - 8.5" x 8.75" OD PAKVF4 MylarFoil pouches, opening on the ZipSeal end
264   (Impak Corporation, #085MFSOZE0875)
265 - 500-cc oxygen absorbing packets (Impak Corporation, #SF500CS1500)
266 - Microtube tough-tags for laser printers, 0.91x0.32 in (Diversified Biotech, #9186-
267   1000)
268 - Microtube tough-tags for laser printers, 1.28x0.5 in (Diversified Biotech,
269   #TTLW-2016)
270
271 Glassware
272 - 2-L Pyrex bottles, 45-mm neck size (Pyrex, #1395-2L)
273 - 1-L Pyrex bottles, 45-mm neck size (Pyrex, #1395-1L)
274 - 250-mL Erlenmeyer flasks (Pyrex, #4980-250)

275

276 PCR and DNA isolation

- QIAamp 96 DNA QIAcube HT Kit (QIAgen #51331)
- 5x Q5 reaction buffer and high GC enhancer (New England Biolabs #B9027S)
- Deoxynucleotide (dNTP) Solution Mix (New England Biolabs #N0447L)
- Q5® High-Fidelity DNA Polymerase (New England Biolabs #M0491L)
- 96-well PCR plate (Eppendorf #0030133382)
- Zymo DNA clean and concentrator kit (Zymo Research #D4013)
- JumpStart DNA polymerase (Sigma Aldrich # D9307-50U)
- Agencourt AMPure XP beads (Beckman Coulter #A63880)
- Indexed primers for Bar-seq and RB-TnSeq (Integrated DNA Technologies; Table S1)
- NEBNext DNA Library Prep Master Mix Set for Illumina (New England Biolabs #E6040S)
- Covaris microTUBE AFA Fiber Pre-Slit Snap-Cap (Covaris #520045)

291 Oligos

- Order indexed primers (Supplementary Data 1) for amplifying the barcodes.
- Order oligos (Supplementary Data 2) for performing RB-TnSeq.

295 Strains

The sorting protocol requires a barcoded transposon insertion library that is compatible with Bar-seq. Briefly, a barcoded transposon insertion library is created using a traditional transposon mutagenesis approach (e.g. electroporation of *in vitro*-assembled transposomes or conjugation of the transposase and transposon on a non-replicating plasmid), with the addition on the transposon construct of a small sequence adjacent to one of the inverted repeats that acts as a barcode. In the approach from Ref. [18], the barcode is a stretch of 20 random nucleotides surrounded by two primer binding sites for amplification of the barcode. During the creation of a transposon insertion library, an RB-TnSeq experiment as described in steps 86-95 is performed to determine the size and quality of the library. The lookup table of insertion locations and the associated DNA barcodes generated by this RB-TnSeq reaction can be used to locate mutants within the ordered library using Bar-seq, or another RB-TnSeq reaction can be run using samples from the ordered library alone (as discussed below).

311 EQUIPMENT

- Benchsmart 96 semi-automated pipetting system (installed within an anaerobic chamber), 1000 µL (Mettler Toledo, #BST-96-1000)
- Stylus Pen (Liberrway, #B01IHBVGOM)
- BenchSmart 96 pipette head, 200 µL (Mettler Toledo, #30296709)
- Forced air incubator for vinyl chambers (Coy Laboratory Products, #2000)
- Teflon-coated 6" hot jaw sealer, portable, constant heat (Impak Corporation, #IPKHS-606T)
- BD GasPak EZ large incubation container (Becton Dickinson, #260672)

320   •   BD Influx Cell Sorter (Becton Dickinson, #646500) or equivalent
321   •   Genesys[TM] 20 visible spectrophotometer (Thermo scientific, #4001-000) or
322   equivalent
323   •   QIAcube HT system (QIAgen, #9001793) or equivalent
324   •   PCR machine (Eppendorf Mastercycler X50a # 6313000018) or equivalent
325   •   Pipet-Lite multichannel pipette L12-200XLS+ (Rainin # 17013810)
326   •   MiSeq system (Illumina # SY-410-1003) or equivalent.
327   •   HiSeq 4000 system (Illumina #SY-401-4001) or equivalent
328   •   Covaris S220 Focused Ultrasonicator (Covaris #500217) or equivalent
329
330 SOFTWARE
331 The software used for random barcode transposon-site sequencing is available from
332 https://bitbucket.org/berkeleylab/feba/src. The scripts for resolving mutant positions
333 within a ordered library are available from
334 https://bitbucket.org/kchuanglab/resolve_barcode_position/src. For installation and
335 configuration, follow the instructions in the respective repository.
336
337 REAGENT SETUP
338 **Bar-seq indexed primer plate**
339 Combine Barseq_P1 with each unique, indexed reverse primer (BarSeq_P2) at a final
340 concentration of 4 µM for each primer and store in a 96-well plate format. Store the oligo
341 mixes at -20 ˚C, where they will be stable for years.
342
343 **RB-TnSeq Y-adapter**
344 Mix the oligos MOD2_TruSeq and Mod2_TS_Univ in a 50 µL volume in a PCR tube at
345 a concentration of 10 µM for each oligo. Use a PCR machine to anneal the oligos using
346 the "Anneal Y-adapter" program. Store the annealed oligos at -20 ˚C, where they will be
347 stable for years.
348
349 **Bar-seq PCR master mix (32.5 µL mix for a 50 µL PCR reaction)**
350 Make the master mix fresh and store on ice until ready to use.

Deionized water	11 µL
5X Q5 reaction buffer	10 µL
5X Q5 GC enhancer	10 µL
10 mM dNTPs	1 µL
Q5 polymerase	0.5 µL

351
352 **RB-TnSeq PCR master mix (50 µL mix for a 100 µL PCR reaction)**
353 Make the master mix fresh and store on ice until ready to use.

Deionized water	21.5 µL
JumpStart Taq buffer	10 µL
10 mM dNTPs	2 µL
100 mM MgCl$_2$	1.5 µL
10 µM Nspacer_barseq_universal	5 µL
10 µM P7_i6	5 µL

DMSO	3 µL
JumpStart Taq polymerase	2 µL

354

Media components

355

356 • Sodium bicarbonate (10% w/v, pH 7.4): To make 1 L, add 700 mL of deionized
357 water to a 1-L beaker. Add 100 g of sodium bicarbonate and stir gently. Insert a
358 calibrated pH meter and slowly add dilute HCl acid (0.1 M) to lower the pH to 7.4.
359 When the target pH has been reached, add deionized water to a final volume of 1
360 L. Filter sterilize the solution with a 0.22-µm PES filter. The solution is stable for
361 at least one month at room temperature with a tightly closed lid.
362 CRITICAL STEP Avoid heat, vigorous stirring, and concentrated acid solutions
363 to prevent driving CO_2 out of solution.

364

365 • Hemin (0.5 mg/mL in 0.01 M NaOH): To make 500 mL, add 250 mg of hemin to
366 5 mL of 1 M NaOH. Once the hemin has dissolved, bring the final volume to 500
367 mL with deionized water. Cover with foil and store at 4 ˚C until use. The solution
368 is stable at 4 ˚C for several months.

369

370 • Antibiotic stock (e.g., 10 mg/mL erythromycin): Make an antibiotic stock at the
371 desired concentration. For the *B. thetaiotaomicron* library, we dissolved 200 mg
372 of erythromycin in 20 mL of absolute ethanol for a final concentration of 10
373 mg/mL, allowing us to make 20 L of BHIS supplemented with 10 µg/mL
374 erythromycin. The stability of antibiotic stocks depends on the antibiotic and
375 solvent. We stored the erythromycin stock at -20 ˚C and used the stock within a
376 week.

377

Sterilization and pre-reduction of media and plasticware

378

379 The volume of media and number of plates have been chosen for a 40-plate library; adapt
380 accordingly for larger libraries.

381

382 Make the sort recovery medium in 5 L batches: for BHIS, add 4 L of deionized water to a
383 5-L beaker. Add 185 g of commercial BHI mix and stir until the solution clears.
384 Microwave 500 mL of deionized water to bring to a boil and add to the solution, continue
385 stirring. Add 50 mL of the hemin solution and 100 mL of the 10% sodium bicarbonate
386 solution. Bring the final volume to 5-L using a graduated cylinder. Filter sterilize into 2-L
387 flasks using a 0.22-µm PES filter. Repeat this process until all 15 L of medium have been
388 made.
389 CRITICAL STEP It is critical to filter sterilize all media in which cells are grown prior to
390 sorting in order to reduce the frequency of empty wells that result from sorting particulate
391 matter. At a minimum, the 1 L of medium used to grow the library before sorting needs
392 to be filter sterilized. We also filter sterilize the recovery medium to avoid wait time and
393 batch effects due to the effects of autoclaving on the medium.

394

395 To prepare 3 L of 50% (w/v) glycerol, transfer 1.5 kg of glycerol to a 5-L beaker. Add
396 water to 3 L final volume and transfer to 3 1-L bottles. Autoclave to sterilize.

397

398 Prepare and sterilize plastic and glassware: wrap 45 2-mL 96-deepwell plates in foil in
399 stacks of 5. Autoclave to sterilize. Wrap 10 reservoirs in foil individually. Autoclave to
400 sterilize. Cover the mouth of 4 250-mL Erlenmeyer flasks with foil. Autoclave to sterilize.
401
402 Move media and plasticware into the anaerobic chamber. Store the bottles of media and
403 glycerol with bottle caps unscrewed to allow gas exchange. This includes:
404 • 7 2-L bottles of sort recovery medium
405 • 1 1-L bottle of sort recovery medium
406 • 2 sterilized reservoirs
407 • 45 sterilized 2 mL 96-deepwell plates
408 • 10 1-mL LTS filter tips
409 • 4 250-mL Erlenmeyer flasks
410 • 1 opened bag of 5 mL polypropylene round bottom tubes
411
412 Finally, let the media sit for 48-72 h in the anaerobic chamber.
413 CRITICAL STEP It is good practice to allow all materials that need to be anaerobic to sit
414 in the anaerobic chamber for more than 48 h. If changes are made in the protocol to adapt
415 for specific needs (e.g., duration of growth in step 32), then corresponding changes must
416 also be made to when materials for the next set are brought into the anaerobic chamber to
417 remove oxygen.
418
419 **EQUIPMENT SETUP**
420
421 BenchSmart Programs
422
423 *Medium transfer*

Option	Setting
Pipetting mode	Advanced
Volume	650 µL
Aspirate/Dispense speed	Asp: 9/ Disp: 9
OPTION: Fixed volume	OFF
OPTION: Volume sequencing	OFF
OPTION: Mix	OFF
OPTION: Cycle counter	OFF
OPTION: Blowout	OFF

424
425 *Dispense glycerol*

Option	Setting
Pipetting mode	Advanced
Volume	557 µL
Aspirate/Dispense speed	Asp: 5 / Disp: 5 / Mix: 5
OPTION: Fixed volume	OFF
OPTION: Volume sequencing	OFF
OPTION: Mix	**ON**
OPTION: Cycle counter	OFF

OPTION: Blowout	OFF
MIX OPTION: Mix volume	750 µL
MIX OPTION: Mix cycles	9

426

427 *Repeat dispense*

Option	Setting
Pipetting mode	MULTI-DISP
Aliquot volume	100 µL
Number of aliquots	1-10
Aspirate/Dispense speed	Asp: 5 / Disp: 3
OPTION: Fixed volume	OFF
OPTION: Volume sequencing	OFF
OPTION: Cycle counter	OFF
OPTION: Dispense blowout	OFF
OPTION: Auto pace	OFF

428

429 *Mix Cultures*

Option	Setting
Pipetting mode	Advanced
Volume	200 µL
Aspirate/Dispense speed	Asp: 5 / Disp: 5 / Mix: 5
OPTION: Fixed volume	OFF
OPTION: Volume sequencing	OFF
OPTION: Mix	**ON**
OPTION: Cycle counter	OFF
OPTION: Blowout	OFF
MIX OPTION: Mix volume	200 µL
MIX OPTION: Mix cycles	9

430

431 *Dispense pools*

Option	Setting
Pipetting mode	MULTI-DISP
Aliquot volume	15 µL
Number of aliquots	2
Aspirate/Dispense speed	Asp: 8 / Disp: 1
OPTION: Fixed volume	OFF
OPTION: Volume sequencing	OFF
OPTION: Cycle counter	OFF
OPTION: Dispense blowout	OFF
OPTION: Auto pace	OFF

432

433 *Combine pools*

Option	Setting
Pipetting mode	Advanced
Volume	100 µL

Aspirate/Dispense speed	Asp: 5 / Disp: 5
OPTION: Fixed volume	OFF
OPTION: Volume sequencing	OFF
OPTION: Mix	OFF
OPTION: Cycle counter	OFF
OPTION: Blowout	OFF

434

435   PCR programs

436

437   *Bar-seq*

Temperature	Time	Cycles
98 ˚C	5 min	
98 ˚C	30 s	
55 ˚C	30 s	25 cycles
72 ˚C	30 s	
72 ˚C	5 min	
12 ˚C	hold	

438

439   *RB-TnSeq*

Temperature	Time	Cycles
94 ˚C	2 min	
94 ˚C	30 s	
65 ˚C	20 s	25 cycles
72 ˚C	30 s	
72 ˚C	10 min	
12 ˚C	hold	

440

441   *Anneal Y-adapter*

Temperature	Time	Cycles
95 ˚C	2 min	
95 ˚C (-1 ˚C /cycle)	38 s	71 cycles
12 ˚C	hold	

442

443

PROCEDURE

Day 1: Prepare for the sort TIMING 0.5 d

1. (Optional) Remove shelving in the incubator to allow space for 2-mL 96-deepwell plates.
2. Add the selective antibiotic to the final concentration (e.g., for 10 µg/mL erythromycin: add 2 mL of 10 mg/mL erythromycin solution to every 2-L bottle of BHIS).
3. Load sterile 1-mL filter tips onto the BenchSmart. Use the tip box lid to cover a sterile empty reservoir.
4. Uncover the reservoir.
5. Add sort recovery medium to the reservoir up to 250 mL, if necessary.
6. Do two transfers of 650 µL (1.3 mL total) to a 2-mL 96-deepwell plate.
7. Cover the reservoir.
8. Seal the plate with an aluminum seal.
9. Repeat steps 4-8 until every 96-deepwell plate has been filled.
10. When the plates have been filled, unload the 1-mL tips.
    CRITICAL STEP Exercise caution when reusing the tips. Any contamination transferred to the tips through contact with a dirty surface will propagate throughout all subsequent plates. If you touch the tips unintentionally, eject them and load a fresh set before revisiting the medium in the reservoir.
11. Stack the 2-mL 96-deepwell plates inside the anaerobic 37 ˚C incubator.
12. Add the antibiotic to the final concentration to the 1-L bottle of sort recovery medium (e.g., add 1 mL of 10 mg/mL erythromycin to the 1-L bottle of BHIS).
13. Prewarm 4 250-mL Erlenmeyer flasks containing 100 mL sort recovery media with antibiotic.
14. One hour later, thaw a 2-mL aliquot of the mutant library and add the entire contents to one 250-mL flask (the other flasks will be used for maintaining the culture in log phase on the next day).
15. Grow the culture overnight at 37 ˚C.

Day 2: Sort the library TIMING 1 d

16. Determine the optical density of the overnight culture using a UV-Vis spectrophotometer.
    CRITICAL STEP Optical density measurements vary between plate reader machines, and every bacterial strain has an idiosyncratic growth curve in different media. Thus, it will be useful to run a growth curve using the same UV-Vis spectrometer used for this protocol to calibrate OD_{600} readings to the growth phase of your bacterium.
17. Dilute the overnight culture to an OD_{600}=0.05 into one of the pre-warmed Erlenmeyer flasks of sort recovery media.
18. Periodically monitor the OD_{600}, keeping the culture in log phase by diluting it into one of the pre-warmed Erlenmeyer flasks at a 1:4 dilution.
19. Pre-warm four 3-mL aliquots of sort recovery medium in FACS tubes.
20. Before sorting, measure the OD_{600} of the culture and dilute into the FACS tubes to cover a range of OD_{600} values (typically 0.01-0.05).

489   21. Tape the caps to the sides of the FACS tubes using vinyl tape to slow the entrance of
490   oxygen.
491   22. Transfer the FACS tubes to a GasPak EZ container and bring them out of the
492   anaerobic chamber along with the first set of 2-mL 96-deepwell plates.
493   23. Transport the material to the FACS machine, being careful not to agitate the plates.
494   24. Calibrate the cell sorter to ensure droplets are properly sorted into each well.
495   Open the GasPak EZ container, vortex the first tube with the vinyl tape still on.
496   Remove the tape and load the tube on the sample injection port.
497   CRITICAL STEP After unsealing the tube, agitate the sample as little as possible.
498   Without agitation the oxidation front will remain near the top of culture for longer,
499   delaying the onset of oxidation for the cells entering the sample injection port at the
500   bottom of the tube.
501   25. Calibrate the FACS machine to sort single cells: adjust the threshold and gain for
502   forward scatter (FSC) and side scatter (SSC) so that the population of single cells has
503   a clear distribution of signal. Define a restrictive two-dimensional gate on FSC and
504   SSC that selects for the average signal from single cells and excludes larger signals.
505   Use a secondary gate on pulse trigger width to further select for single cells.
506   ?Troubleshooting
507   CRITICAL STEP The gating strategy used to identify single cells is critical for
508   successfully sorting the library and will depend heavily on the strain, growth medium,
509   and physiological state of the cells. Test the exact conditions used to grow the library
510   beforehand to determine the expected signal during the library sort.
511   26. Unseal a 96-deepwell plate and cover with a sterile lid; check for liquid on seal.
512   ?Troubleshooting
513   27. Sort single cells into the plate using the FACS machine.
514   28. Reseal the plate with a Breathe-Easier gas permeable membrane. Use a roller to
515   ensure that a tight seal has formed, especially around the edges of the plate.
516   CRITICAL STEP Ensuring a good seal is important to prevent evaporation from the
517   outer rows and columns of the 96-deepwell plate during the recovery period.
518   CRITICAL STEP It is important to avoid jostling or agitating the plates from this
519   point onwards to minimize cross contamination between the wells. If a plate is jostled
520   while applying the seal or moving to the incubator, check for wetting of the
521   membrane and discard the plate if media has splashed onto the seal.
522   29. When the batch of plates has been sorted, return the plates to the anaerobic chamber
523   and place them in the 37 ˚C incubator.
524   CRITICAL STEP Cycle the airlock with a low-pressure program to avoid peeling the
525   seal from the plates (e.g., 9 cycles of 5 in Hg vacuum pressure and fill with N_2,
526   followed by 9 cycles of 5 in Hg vacuum pressure and fill with the chamber
527   atmosphere).
528   30. Take the next batch of plates out of the 37 ˚C incubator for sorting.
529   CRITICAL STEP Bring the plates out of the anaerobic chamber in small batches.
530   Optimize the size of each batch so that the plates spend as little time in the aerobic
531   environment as possible.
532   31. Repeat steps 26-30 until cells have been sorted into all 96-deepwell plates (Figure 4).
533   32. Once all plates have been sorted, incubate them at 37 ˚C to allow recovery and
534   growth of the sorted cells.

535   ?Troubleshooting

536

537 **Day 3: Label and pre-reduce microwell plates TIMING 1 d**

538 33. Unpack the 96-microwell plates.

539 34. Add a label to the front and to one side.

540 35. Add a sterile lid with a label, store in large sterile plastic bags.

541 36. Move the bags into the anaerobic chamber to drive oxygen off of the plastic. For 6

542   copies of a 40-plate library, process 240 96-microwell plates.

543 37. Move the appropriate number of MylarFoil bags into the anaerobic chamber, equal to

544   the number of processed 96-microwell plates.

545

546 **Day 4/5: Aliquot the library TIMING 1 d**

547 38. Remove the Breathe-Easier membrane from a culture plate, cover with a sterile tip-

548   box lid.

549   CRITICAL STEP Take a moment to examine the growth pattern within the 96-

550   deepwell plate. If the pattern is troubling (no growth, uniform growth in all wells)

551   then discard the plate before it is assigned a number in the following step.

552 39. Add a label to the plate to assign a plate number.

553   CRITICAL STEP This step will be the first point at which the culture plates are

554   identified with a unique label. From this point onwards, it is important to check that

555   the labels match when mixing, aliquoting, and storing the glycerol stock.

556 40. Transfer the culture plate to the BenchSmart, match with the correspondingly labeled

557   96-microwell plates.

558 41. Load 1-mL filter tips.

559 42. Switch the BenchSmart program to "Dispense glycerol".

560 43. Transfer 557 µL of 50% glycerol to the cultures and mix.

561   CRITICAL STEP Track the height of the culture when resuspending to keep the tips

562   of the pipette just below the surface. Submerging the pipette tips when the wells are

563   almost full will push the cultures out of the wells, while allowing the tips to dispense

564   above the culture surface will cause splattering.

565 44. Switch the BenchSmart program to "Repeat dispense".

566 45. Dispense 100 µL to each 96-microwell plate.

567 46. Unload tips.

568 47. Seal the 96-microwell plate glycerol stocks with an AlumaSeal. Use a roller to ensure

569   that a good seal is formed.

570 48. Place the sealed microwell plates in MylarFoil heat seal bags along with a 500-cc

571   oxygen absorber, zip-seal the bags.

572 49. Seal the original culture with an aluminum seal.

573 50. Once enough bagged plates have accumulated, move them outside the anaerobic

574   chamber and use the heat sealer to seal the bags.

575 51. Move the bagged plates to the -80˚C freezer for long-term storage.

576 52. When finished, or as needed to free up space, move the sealed original cultures out of

577   the anaerobic chamber.

578 53. Repeat steps 38-52 until all 96-deepwell plates have been copied as cryostocks

579   (Figure 5).

580 CRITICAL STEP Aliquoting the library will require a team of people to work
581 efficiently. We have designed this protocol around an anaerobic chamber with two
582 sets of gloves to include a team of three researchers. One person will unseal, label and
583 reseal the plates. One person will operate the BenchSmart to mix and aliquot the
584 glycerol stocks. One person will work outside the anaerobic chamber to heat-seal and
585 store the glycerol stocks at -80 ˚C (Figure 5).
586 54. At the end, transfer 200 µL of each of the original culture plates to a flat bottom 96-
587 microwell plate and measure the OD_{600} using a plate reader. This measurement of the
588 final OD_{600} of every well in the library will allow for an estimate of the fraction of
589 wells in which growth occurred. Store the remainder of the culture at 4 ˚C or -20 ˚C.
590 PAUSE POINT The remainder of the cultures in 96-deepwell plates will be used in
591 subsequent steps to pool the library and sequence. The 96-deepwell plates can be
592 stored overnight at 4˚C or for a few days at -20˚C.
593 55. Analyze the final OD_{600} values of the ordered library. Use a cutoff to determine the
594 fraction of wells in which growth was detected. Examine the distribution of
595 growth/no growth and final OD_{600} values across plates. Plates with abnormal growth
596 can be triaged.
597 ?Troubleshooting
598
599 **Pool the library and extract genomic DNA TIMING 1d**
600 56. Thaw the original 96-deepwell plates from step 54.
601 57. Centrifuge each 96-deepwell plate for 5 min at 4,500 x *g* at room temperature.
602 CRITICAL STEP Prevention of cross-contamination during pooling is essential to
603 accurately resolve the position of barcodes within the library. Centrifugation pulls
604 down any culture that may be stuck to the seal and prevents droplets from spraying
605 between wells when the seal is pulled off.
606 58. Allow the 96-deepwell plates to thaw at room temperature.
607 59. Open two sterile 2-mL 96-deepwell plates and cover each one with a sterile tip-box
608 lid. The first 96-deepwell plate will store a mixture of each well from all plates (all-
609 sample plate). The second 96-deepwell plate will store the final pools (final-pool
610 plate).
611 CRITICAL STEP The following steps are designed to work with a 40-plate library so
612 that individual pools have a volume of less than 1.5 mL. To adapt for a larger library,
613 lower the volumes accordingly.
614 CRITICAL STEP The following steps are designed to work with a pool design and
615 library size that need less than 96 pools. If more than 96-pools are required, more
616 than one final-pool plate will be necessary.
617 CRITICAL STEP Mislabeling or cross-contaminating wells in the final-pool plate
618 will lead to significant errors in locating barcodes. Work carefully to ensure that no
619 pipetting errors are made when filling the plate.
620 60. Load sterile 200-µL filter tips onto the BenchSmart.
621 61. Set the BenchSmart-96 to the "Mix cultures" program.
622 62. Aspirate from the centrifuged 96-deepwell plate and dispense into the same wells to
623 mix the cultures.
624 CRITICAL STEP If the cultures are prone to settling, it is important to dispense
625 aliquots into the pools quickly to avoid pipetting errors.

626   63. Set the BenchSmart-96 to the "Dispense pools" program.
627   64. Remove the seal from the centrifuged 96-deepwell plate and cover with a sterile tip
628   box lid.
629   65. Open a sterile, clean, single-well reservoir.
630   66. Multi-dispense 15 µL into the single-well reservoir, creating spots on the bottom of
631   the reservoir.
632   CRITICAL STEP Keep the tips elevated from the bottom of the reservoir to prevent
633   backup of the dispensing liquid, which can lead to spattering of the liquid and cross
634   contamination between tips.
635   67. Multi-dispense 15 µL into the all-sample plate.
636   68. Collect the culture from the single-well reservoir and store in a well of the final-pool
637   plate.
638   69. Repeat steps 64-68 for every plate in the library.
639   70. Pool the library according to rows and columns using the all-sample plate.
640   71. For every row in the all-sample plate, use a 12-channel 200-µL pipette with 12 filter
641   tips loaded. Transfer 120 µL per well to a single-well reservoir. Collect the culture
642   and store in a well of the final-pool plate.
643   72. For every column in the all-sample plate, use a 12-channel 200-µL pipette with 8
644   filter tips loaded. Transfer 180 µL per well to a single-well reservoir. Collect the
645   culture and store in a well of the final-pool plate.
646   OPTIONAL: Steps 73-76 are included to generate DNA for running RB-TnSeq on the
647   ordered library.
648   73. Load 200-µL filter tips onto the BenchSmart.
649   74. Set the BenchSmart-96 to the "Combine pools" program.
650   75. Use the BenchSmart-96 to transfer 100 µL from the all-sample pool plate to a single-
651   well reservoir.
652   76. Transfer the combined sample into a 15-mL conical tube then distribute the sample in
653   1.5 mL aliquots across multiple wells of the final-pool plate.
654   77. Extract genomic DNA from the final-pool plate following the manufacturer's
655   protocol for the QIAamp 96 DNA QIAcube HT kit.
656

657   **Perform Bar-seq to identify the pool inclusion patterns of barcodes TIMING 1d**
658   78. Make enough of the PCR master mix to amplify each pool, add 32.5 µL per reaction
659   to the wells of a 96-well PCR plate.
660   79. Add 5 µL of the 4 µM indexed primer mixes to each well.
661   80. Add 12.5 µL of the extracted genomic DNA to each well.
662   81. Seal the PCR plate with a foil seal and run the Bar-seq PCR program.
663   82. Run 10 µL of each PCR product on a 1% (w/v) agarose gel to check for the correct
664   band (~200 bp).
665   83. Combine 5 µL of each Bar-seq product into a single pool. Store the remainder.
666   84. PCR purify a 50 µL aliquot of this total-library pool by following the manufacturer's
667   protocol for a Zymo DNA clean and concentrator kit. Elute in 20 µL of sterile
668   deionized water.
669   85. Submit the purified PCR product for sequencing using a HiSeq 4000, 1x50 bp reads.
670

OPTIONAL: Perform RB-TnSeq to associate barcodes with insertion sites TIMING 1d

Performing RB-TnSeq on the ordered transposon library in addition to the complete insertion pool is not necessary but may be helpful in certain cases. If the same barcode is associated with more than one insertion site, then the true nature of this insertion is ambiguous: two independent strains, with one insertion per genome each, could serendipitously share the same barcode; alternatively, if the library was generated using conjugation, then repeated barcodes may arise from multiple transposon insertions in the same genome. If the initial insertion pool is complex and has a high fraction of ambiguous barcodes, then performing RB-TnSeq on the ordered library will likely reduce the fraction of ambiguous barcodes since it is improbable that two strains carrying the same barcode will both be sorted.

86. Combine genomic DNA aliquots from the whole-library pool of step 76 and dilute so that there is 1.5-3 µg DNA per aliquot.
87. Fragment the purified genomic DNA to an average size of 300 bp using a Covaris S2 Ultrasonicator using the manufacturer's protocol.
88. Further size-select the fragmented DNA for an average size of 300 bp with a tighter distribution using AMPure XP beads using the manufacturer's protocol.
89. At the adapter ligation step, ligate the custom TruSeq Y*-adapters (Reagent Setup) to the DNA fragments at a final concentration of 150 nM. In the clean-up step after adapter ligation, elute into 50 µL of elution buffer to use in the next PCR step.
90. Make enough of the PCR Master mix for a 100 µL reaction (See Reagent Setup).
91. Add 50 µL of adapter-ligated DNA.
92. Run the RB-TnSeq PCR program.
93. Run 5 µL of PCR product on a 1% (w/v) agarose gel to check for the correct band (200-300 bp).
94. Purify the PCR product with AMPure XP beads.
95. Submit the purified PCR product for sequencing using a MiSeq, 1x150 bp reads.

Locate the transposon insertion mutants in the library TIMING 1d
96. Create a new directory and assemble the necessary input files inside.
97. Download the appropriate genbank file with gene annotations.
98. Convert the genbank file into *fasta* and *genes* files for further analysis (Figure 6) using the command "*SetupOrg.pl -gbk genbankfile -out organism*".

SetupOrg.pl generates a *fasta* file for the genome and a *genes* tab-delimited file that are used as inputs in subsequent analysis.

99. Move the RB-TnSeq fastq file(s) into a directory named *tnseq*.
100. Move the Bar-seq fastq files into a directory named *barseq*.
101. Create or move a model file into a directory named *model*.

The model file is a model sequence for what should be expected before the transposon-genome junction. Run "*MapTnSeq.pl -h*" for further explanation of

716   how the model file is constructed and see *feba/primers/model_** in the FEBA
717   repository for examples.

718

719   102.   Map reads onto the genome (Figure 6) by running the command *MapTnSeq.pl -*
720   *genome organism/genome.fna -model model/modelfile -first tnseq/fastq > tncount*
721   103.   Generate the lookup table (Figure 6) by running the command *python3*
722   *path_to_repository/scripts/DesignSortedLibrary.py3 --input tncount --genes*
723   *organism/genes.tab --Nmin 10 --output lookuptable*
724   104.   Generate a tab-delimited key file (*barseqkey,* Figure 6) with the columns
725   *path_to_fastQ file*, *index name*, and *pool name*.

726

727   The FEBA codebase uses a restricted set of indexes to identify the barcodes in a
728   Bar-seq read. The index name in the key file indicates the index that should be
729   used for a demultiplexed fastq file using the FEBA set of primers. When using an
730   alternative set of primers for Bar-seq, the scripts will have to be modified
731   accordingly. This key file allows for the association of each sequencing file with
732   the appropriate pool label. The pool names must be P# for plate number, C# for
733   column, and R# for row, e.g. P1, C3, or R5.

734

735   105.   Generate the pool inclusion table (Figure 6) by running the command *python3*
736   *path_to_repository/scripts/SortedPoolCounts.py3 –k barseqkey –o inclusiontable –q*
737   *10 –f 0.01*
738   ?Troubleshooting
739   CRITICAL STEP Unlike the other sequencing approaches, which map reads onto the
740   genome, there is no internal metric that indicates whether a given barcode sequence is
741   real or spurious. Prudently choosing the minimum number of reads to consider a
742   barcode real is important to avoid introducing extra barcode sequences into the
743   analysis pipeline. *SortedPoolCounts.py3* uses the *-f* parameter, the minimum fraction
744   of reads relative to expectation, to control the number of barcode sequences included
745   in the analysis.
746   106.   Locate transposon insertion mutants in the ordered library (Figure 6) by running
747   the command *matlab -nodesktop -nosplash -r *
748   *"addpath(path_to_repository/scripts/); *
749   *resolve_barcode_position(inclusiontable,lookuptable,organism/genes.tab,posi*
750   *tiontable); *
751   *exit;"*
752   107.   Examine the *positiontable* file; this is the final map of strain locations within the
753   library.
754   ?Troubleshooting

755

756   **TROUBLESHOOTING**

757

758   Table 1: Troubleshooting advice for common problems, with the associated step and
759   possible explanation(s).

Step	Problem	Possible reason	Solution

25	Signal from FSC and SSC unclear, hard to identify individual cells.	Media not sufficiently filtered; particulate matter is acting as a signal	Always filter the media
26	Media in plates contaminated	Original bottles not sterilized	Autoclave the 2-L bottles before storage, use sterile technique to filter media
		Contamination during the transfer to plates	Use sterile technique, eject tips if they physically touch anything other than media, lower aspiration and dispensation speeds, cover the reservoir and plates except when pipetting
26	Media on the seal	Too much agitation during transfer	Carry plates by hand
		Cell morphology is aberrant	Re-examine growth conditions, use microscopy to determine cell morphology during growth to detect chaining, clumping, or filamentation. Change culture conditions to obtain more regular cell shapes
32	Wicking of media onto gas permeable membrane, leakage of media from the plates	Too much agitation during transfer	Lower the media volume. Carry plates by hand
32	Production of H_2S during incubation	Culture conditions that promote the production of H_2S gas	Change the culture conditions to eliminate H_2S production. For *B. thetaiotaomicron*, we removed additional cysteine from the medium recipe. Before working with a new strain, check Bergey's Manual of Determinative Bacteriology[21] for hydrogen sulfide production and test small cultures before scaling up

55	Very few empty wells per plate	Cross contamination between wells	Check for signs that the medium has wicked onto the membrane
55	Too many empty wells per plate	Sorting of particulate matter	Filter the media
		Sorting of dead cells	Adjust growth conditions to improve viability. Try a different medium, and/or an earlier point in the growth curve
		Harsh conditions for growth of single cells after the sort	Adjust the concentrations of antioxidants (e.g. cysteine, ascorbic acid). Re-examine culture conditions
	Spatial bias in empty wells	Mis-calibrated sorting	Calibrate the FACS machine
105	Too many or too few barcodes per pool	Nonideal cutoff as the minimum number of reads for a barcode to be analyzed	Adjust the -f parameter when running *SortedPoolCounts.py3*
107	High rate of more than one strain per well	Contamination in the library.	Re-sort library, taking care to reduce possible sources of cross-contamination, both during the sorting step and the handling of plates after sorting
		Contamination during the pooling steps.	Re-pool the ordered library, taking care to pipette carefully
	High rate of wells with no assignment	Poor coverage from pool sequences	Check sequence depth of sequencing run, resequence pools

760

TIMING

Steps 1-11, Fill 96-deepwell plates with medium: 3-4 h for 45 plates

Steps 12-15, Inoculate an overnight culture with a library aliquot: 2 h hands on / 16 h total

Steps 16-18, Grow the library to log phase: 2-5 h (will vary by strain and media)

Steps 19-20, Prepare for the sorting: 0.5 h

Steps 21-31, Sort the library: 2 h for 45 plates

Step 32, Growth of the ordered library: 2-3 d (will vary by strain and media)

Steps 33-35, Labeling of microwell plates: 1 d

Steps 36-37, Transfer of material to anaerobic chamber: 0.5 h

Steps 38-53, Aliquot and store glycerol stocks: 1 d

Step 54, Measure OD_{600} of the library cryostocks: 2 h

Step 55, Analyze cryostock OD_{600} for quality control: 1 h

Steps 56-76, Pool the ordered library for Bar-seq and RB-TnSeq: 1 d

Step 77, Extract genomic DNA from samples: 4 h

Step 78-84, Prepare the Bar-seq library: 1 d

Step 85, Sequence the Bar-seq library: 1 d

Steps 86-94, Prepare the RB-TnSeq library: 1 d

Step 95, Sequence the RB-TnSeq library: 1 d

Steps 96-106, Resolve the positions of barcodes in the library: 3 h

ANTICIPATED RESULTS

As a test of our approach, we sorted single cells from a barcoded transposon insertion library of *B. thetaiotaomicron* VPI-5482 into 40 96-well plates. We then located barcodes within the ordered library using Bar-seq. We predicted a position for 3517 barcodes within the library, of which 235 were found in more than one position and had to be resolved probabilistically[8]. The final map of transposon insertions in the library identified 3059 wells with a single strain, 350 wells with more than one strain, and 431 wells without a barcode assignment (Figure 7a, top). Less than a third of the wells without a barcode overlapped with wells in which the OD_{600} was less than 0.25, indicating that low cell density is not the main cause of our inability to detect a barcode (Figure 7a, middle). There were no more than three barcodes per well in the library (Figure 7a, bottom). We spot-checked 15 strains (7 single insertions and 8 multiple insertions) using PCR and confirmed that all 15 were correctly located (Supplementary Figure 1, Supplementary Data 3).

Sorting clarifies otherwise ambiguous transposon insertion events

In the transposon insertion pool used in this study, 85% of barcodes were associated with a single, unambiguous insertion site (Figure 7b, top). As discussed above, the remaining barcodes were ambiguous because they could have arisen either from multiple transposon insertions in the same cell or from separate strains that serendipitously received the same barcode.

It is extremely unlikely that distinct strains with the same barcode would be sorted into the same well, implying that barcodes with a definite location in the library are associated with a single strain. By performing RB-TnSeq on the ordered library as well as on the original collection, we were able to disambiguate multiple barcodes, as barcodes with a definite location that are associated with multiple insertions are likely to be single strains with multiple insertions. Using this logic, we reduced the fraction of ambiguous barcodes in the ordered library to 4% (Figure 7b, bottom).

Notably, the fraction of strains with more than one insertion was greater when predicted from RB-TnSeq of the ordered library compared with the larger pool (Figure 7b). It is possible that decreasing the diversity of the library by more than an order of magnitude allows for greater power in detecting additional insertion sites. Thus, if a substantial fraction of insertions in the pooled library are ambiguous, performing an additional RB-TnSeq reaction on the ordered library may disambiguate many of them.

Barcode sequencing improves probabilistic solutions to library positions

In this protocol, we used barcode sequencing (Bar-seq) to sequence pools and locate mutant strains within the library. Alternative sequencing methods that have been used in previous studies include insertion sequencing (INSeq)[5], transposon sequencing (Tn-seq)[7], and semi-random two-step PCR (ST-PCR)[8]. Bar-seq and ST-PCR are significantly simpler and more cost-effective than INSeq and Tn-seq; these PCR-based sequencing methods avoid fragmentation, restriction enzyme digestion, size selection, ligation, and multiple purification steps that are required for INSeq and Tn-seq. While ST-PCR is

828 inexpensive and simple, it suffers from inherent noise during the first PCR amplification,
829 in which one of the two primers has semi-random specificity. Because of this issue, we
830 expect that Bar-seq will be more accurate than ST-PCR in quantifying the relative counts
831 of barcodes in the pools, which is important when predicting the probabilistic positions of
832 repeated barcodes in the library.
833
834 The pool designs used here and elsewhere[7,8] are optimized for the time and effort
835 required to pool the libraries. If a barcode occurs in exactly one each of the plate, row,
836 and column pools, then for our pool design it has a definite solution for a single location
837 in the library. If a strain occurs more than once in the library, then its barcode will appear
838 in more than one of each of the different pool types and, with our pool design, there will
839 be multiple possible solutions for its locations in the library. To predict the correct
840 locations of these repeated transposon insertion strains, a previous study[8] developed a
841 probabilistic solution that we implemented here.
842
843 This approach uses the relative read counts of a barcode across pools to predict the most
844 likely solution out of all alternatives. Briefly, the ratio of reads of the same barcode in
845 two pool types (plate, row, or column) should depend only on the number of strains
846 added to each pool. For example, the ratio of reads for a barcode in its row and column
847 pools should be 8/12. When choosing between alternative combinations of pools, if the
848 ratio of reads for one combination is closer to this ideal ratio than the alternatives, then it
849 is the more likely solution. In reality, the ratio of reads is a distribution centered
850 approximately around this ideal ratio due to both systematic and random noise introduced
851 during the amplification process. A tighter distribution will be more predictive when
852 choosing alternatives.
853
854 Because our pool design (plate, row, column) and the pool design from a previous study
855 using ST-PCR (plate-row, plate-column, row, column)[8] share the row and column pool
856 types, we were able to directly compare Bar-seq to ST-PCR according to their ability to
857 distinguish between alternative solutions for indefinite strain locations. We first
858 examined the distribution of row versus column read ratios for definite solutions in the
859 two libraries. The $\log_2$(row/column read ratio) should be a Voigt distribution centered
860 around -0.585[8]. We found that the Bar-seq method (this study) resulted in a tighter
861 distribution than ST-PCR (reanalyzed data from previous work[8]) (Figure 7c, top),
862 indicating that Bar-seq better preserved read count information contained in the genomic
863 DNA pools.
864
865 We next examined the subset of barcodes (this study, $n=169$) or transposon insertion
866 locations (reanalyzed data from previous work[8], $n=1451$) that were detected in exactly
867 two row and two column pools. There are two unique solutions that could give rise to this
868 pool inclusion pattern. The log ratio of the probabilities of the more (p_1) and less (p_2)
869 likely solutions is a measure of the ability of read ratio information to predict a correct
870 solution for the mutants. The Bar-seq method outperformed ST-PCR (Wilcoxon rank sum
871 test: $W=2\times10^5$, $p=7\times10^{-59}$) (Figure 7c, bottom).
872

873 We conclude that Bar-seq is more robust than ST-PCR and more cost-effective than
874 INSeq and Tn-seq, making it the preferred sequencing method for locating mutants
875 within an ordered library.
876
877 **Limited growth in certain wells is likely due to transient oxygen stress during**
878 **sorting**
879 After aliquoting and storing the library, we measured the OD_{600} after ~60 h of growth. In
880 14% of the wells, the OD_{600} fell below a threshold of 0.25, but in many cases was
881 significantly higher than the expected OD_{600} of a blank well (Figure 8a). Microscopic
882 examination of the cultures in these low OD_{600} wells revealed that the majority still had
883 detectable culture densities (Figure 8a,b). Furthermore, the cultures in these low OD_{600}
884 wells could be recovered by outgrowth (Figure 8c), with a longer lag time as the main
885 feature that distinguished them from wells that exhibited high OD_{600} values during initial
886 outgrowth (Figure 8d). To test if the low OD_{600} was connected to a biological process
887 disrupted by the transposon insertions, we looked for a correlation between low OD_{600}
888 and the disrupted genes of the transposon insertion strains. Of the 61 genes that had more
889 than one insertion strain isolated as a pure culture in the ordered library, we found that
890 one strain having a low OD_{600} did not make it more likely that the second also had a low
891 OD_{600}. In fact, there was zero overlap between pairwise comparisons (Fisher's exact test:
892 $p=0.6$, odds ratio=0). We conclude that low OD_{600} wells in the *B. thetaiotaomicron*
893 ordered library are not due to a fitness defect of the mutant that was sorted. Instead, we
894 hypothesize that the transient exposure to oxygen during sorting temporarily slows
895 growth for a fraction of the sorted wells. The presence of these low OD_{600} wells could be
896 a useful indication that post-sort recovery conditions of a strain can be further optimized.
897
898 **Adaptations to this protocol**
899 This protocol introduces two new advances in creating ordered mutant libraries: flow
900 sorting to isolate single cells of anaerobic species, and barcode sequencing to locate
901 mutant strains within the ordered library.
902
903 Sorting using a FACS machine is a fast method for isolating mutants from a large pool
904 that is able to sort multiple anaerobic species when used in combination with an
905 antioxidant (Figure 2). However, it may not be the best approach in all situations. Certain
906 strains may be too sensitive to oxygen to survive sorting and will need to remain in an
907 anaerobic environment for the entire protocol. If the number of strains in the original pool
908 is not significantly greater than the number of strains to be sorted, the same mutant strains
909 are likely to be isolated more than once in the ordered library. Finally, a core facility with
910 FACS machines may not be available. In these cases, it may be preferable to directly pick
911 transformed colonies from a selective plate. Automated colony picking equipment with a
912 footprint small enough to install in an anaerobic chamber is commercially available[23].
913 Regardless of the approach used to order the library, barcode sequencing still promises to
914 greatly decrease the cost and effort of locating mutations in the ordered library.
915
916 Our strategy was to pool according to plates, rows, and columns (Figure 1). This pool
917 design scales linearly with the number of plates in the library, where $N_{pools} = N_{plates} +$
918 20. For larger libraries, previous approaches have used an alternative pool design (plate-

919   row, plate-column, row, column)[7,8] that scales with the square root of plate number

920   ($N_{\text{pools}} \cong 2\sqrt{N_{\text{plates}}} + 20$). Our plate-row-column pool design is less sensitive to

921   ambiguity from multiple occurrences of the same strain in the library, making it easier to

922   probabilistically locate these strains. For example, a strain that is isolated twice in the

923   library will have 4 potential solutions in a plate-row-column pool design compared with 8

924   potential solutions in the alternative four-dimensional pool design. Furthermore, the ease

925   of Bar-seq reduces the relative cost savings of a pool design that scales with the square

926   root of plate number compared to linearly. Nonetheless, ordered libraries with a larger

927   number of plates may benefit from this alternative pool design. Similarly, ordered

928   libraries with a high degree of repetition (same mutant, multiple locations) would benefit

929   from a DNA Sudoku pool design[17]. Regardless of the pool design used, we expect that

930   Bar-seq will improve the quality (Figure 7c) and reduce the cost and effort to sequence

931   pools. Small modifications to the tables *lookuptable* and *inclusiontable* that are produced

932   at the beginning of our analysis pipeline should be sufficient for adapting Bar-seq to

933   analysis of alternative pooling designs.

934

935   In conclusion, we have demonstrated a single-cell sorting protocol coupled to barcode

936   sequencing that can rapidly and straightforwardly generate a sorted mutant library for

937   anaerobic species. By lowering the investment required to generate an ordered library of

938   mutants, we expect that this protocol will be a powerful advance for researchers

939   investigating genotype-phenotype relationships in the human microbiota, catalyzing new

940   approaches to study the effect of the microbiota on health and disease.

941

Data availability

The sequencing data used in this study are publicly available on NCBI as part of BioProject PRJNA573294. The sequence reads from the Bar-seq experiment on pools of the ordered library are available as part of BioSample SAMN12807646. The sequence reads from the RB-TnSeq experiment on the ordered library are available as part of BioSample SAMN12809978. Plate reader output files, microscopy images, and extracted growth parameters that support the findings of this study (Fig. 2, 3, 6) are available from the corresponding author (K.C.H.) upon request.

Code availability

The code required for locating insertion mutants in ordered libraries is available on BitBucket (https://bitbucket.org/kchuanglab/resolve_barcode_position/src). These scripts rely on previously published code available on BitBucket (https://bitbucket.org/berkeleylab/feba/src).

Author Contributions

A.L.S., A.M.D., and K.C.H. conceived the study. A.L.S. designed experiments and collected data. A.L.S. and R.C. analyzed the data. A.L.S., R.C., A.M.D., and K.C.H. interpreted the data and wrote the manuscript.

Acknowledgements

Figures

Figure 1: Schematic of the entire process of sorting the transposon-insertion library and mapping the barcodes. The protocol includes steps to order a random transposon library and steps to locate individual mutant strains in the library. (Reagent Setup) The media and consumables for a high-throughput library are made and allowed to sit in an anaerobic chamber long enough to drive off oxygen. (Steps 21-31) Single cells from a transposon mutant pool are sorted into individual wells of 96-deepwell plates and allowed to grow to saturation. (Steps 38-53) The cultures derived from sorted single cells are mixed with glycerol to make a cryostock, aliquoted into multiple copies, and stored at -80 ˚C. (Steps 56-76) Left-over cultures are pooled into subsets for sequencing. (Steps 78-85) Barcodes from the pools are PCR-amplified and sequenced. (Steps 96-106) Information on the inclusion of barcodes in each pool is used to locate barcodes within the library. Association of each barcode with a transposon insertion locates mutant strains within the library.

992
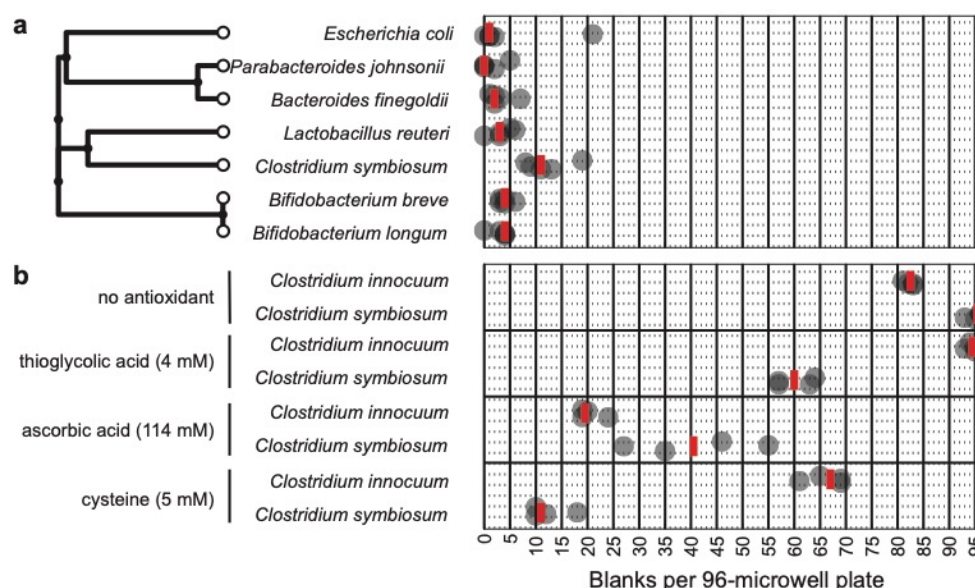
Figure 2: Strict anaerobes isolated from the human gut microbiota survive sorting in the presence of antioxidants.

a) Six human gut isolates and a lab strain of *E. coli* survived sorting and grow in commercial media. Each strain was grown in a commercial formulation of media with added supplements. *Escherichia coli* BW25113, *Parabacteroides johnsonii* CL02T12C29, *Bacteroides finegoldii* CL09T03C10, and *Lactobacillus reuteri* CF48-3A were grown in Gifu Anaerobic Medium (GAM) supplemented with 1.5 µM hemin, 30 µM menadione, 6 mM arginine, and 1 mM tryptophan. *Clostridium symbiosum* WAL-14163 was grown in Reinforced Clostridial Medium (RCM) supplemented with 1.5 µM hemin, 30 µM menadione, 6 mM arginine, and 1 mM tryptophan. *Bifidobacterium breve* DSM20213 and *Bifidobacterium longum* NCC2705 were grown in de Mann, Rogosa, Sharp (MRS) medium. Each strain was sorted into five 96-microwell plates. Dark circles are the number of blank wells in one 96-microwell plate after outgrowth. Red lines are the median value for the 5 replicate 96-microwell plates. All strains except *E. coli* and *L. reuteri* are obligate anaerobes[21]. Phylogenetic distances were extracted from the TimeTree database[24].

b) Two highly oxygen-sensitive strains, *Clostridium symbiosum* WAL-14163 and *Clostridium innocuum* 6_1_30, were sorted and grown in media with an antioxidant. For both strains, we used an alternative formulation of RCM with no cysteine and no agar. We then added either no antioxidant, 4 mM thioglycolic acid, 114 mM ascorbic acid, or 5 mM cysteine. Dark circles are the number of blank wells in one 96-microwell plate after outgrowth. Red lines are the median value for the 4 replicate 96-microwell plates. It is possible to achieve high levels of viability for both of the strains.

1018
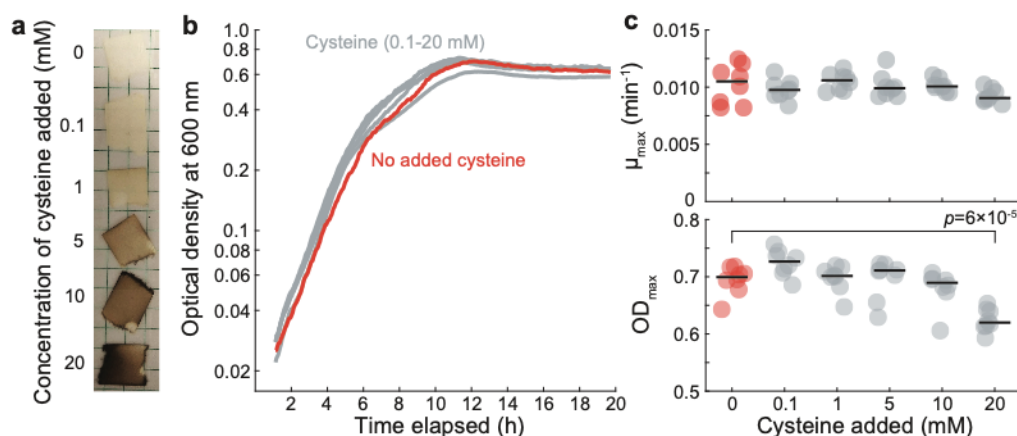
**Figure 3: Elimination of H$_2$S production may be necessary for large volumes of culture**.

a) *B. thetaiotaomicron* produces hydrogen sulfide proportional to the concentration of cysteine added to the medium. The original recipe for BHIS included supplementation with 8 mM cysteine. BHIS media with a range of cysteine concentrations were prepared and inoculated with *B. thetaiotaomicron*. After 24 h of growth, a hydrogen sulfide test strip was held over the culture for 5 s. Reaction of hydrogen sulfide with lead acetate in the strips leads to black discoloration.

b) Anaerobic growth of *B. thetaiotaomicron* is not strongly impacted by the addition of cysteine. A 96-well microwell plate was filled with BHIS and increasing levels of cysteine (0-20 mM). The media were inoculated with *B. thetaiotaomicron* VPI-5482 and growth was monitored by reading optical density at 600 nm. The growth curves show the median value for (n=8) independent measurements.

c) The maximum growth rate (μ_{max}) (top) and maximum OD$_{600}$ (OD$_{max}$) (bottom) were extracted from individual growth curves (n=8). The individual data points (circles) and median (solid line) are plotted for each concentration of cysteine. The omission of cysteine (red circles) did not significantly alter the maximum growth rate or the maximum OD$_{600}$. The only significant comparison between the no-cysteine condition and the cysteine-added conditions was with 20 mM cysteine for the OD$_{max}$ parameter (multiple hypothesis corrected one-way ANOVA, p=6x10^{-5}).
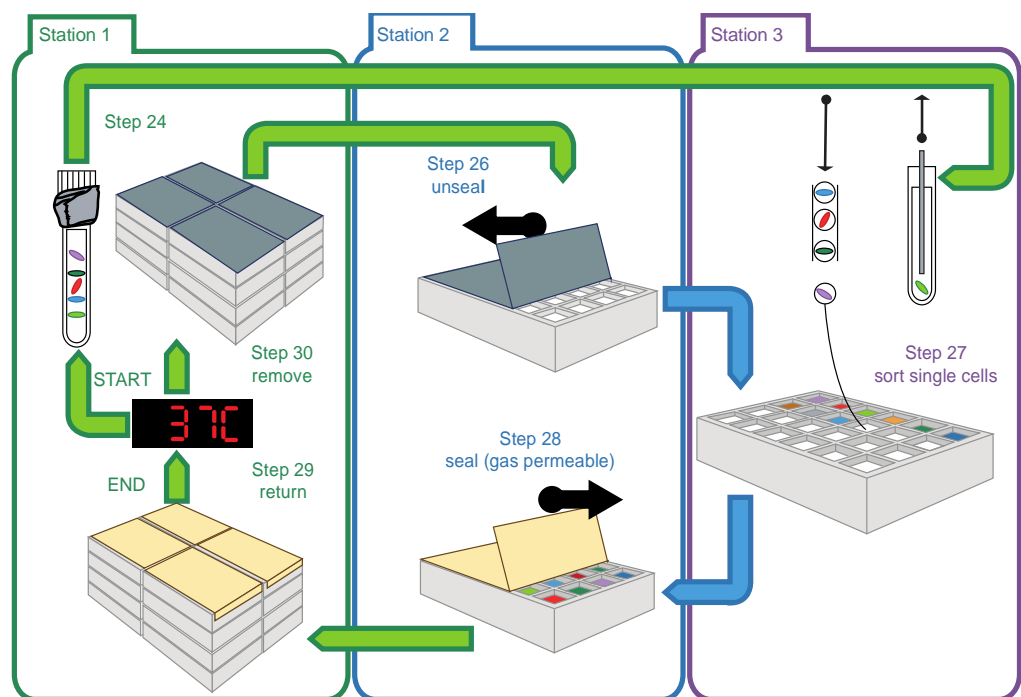
1040
Figure 4: Sorting schematic. Three people work together to rapidly sort single cells. Station 1 is centered at the anaerobic chamber. The person working at station 1 cycles between the anaerobic chamber and the cell sorter, bringing sets of fresh (unsorted) 96-deepwell plates to the cell sorter and taking back the set of sorted plates that have accumulated. Station 2 is next to the cell sorter. This person works to unseal fresh plates, pass them to station 3, take sorted plates, and reseal them with a gas permeable membrane. Station 3 is the cell sorter. This person works to set up the equipment for the sort and run the single-cell sorting protocol. The arrows indicate movement of consumables between stations. The arrows are color coded by the person doing the work.
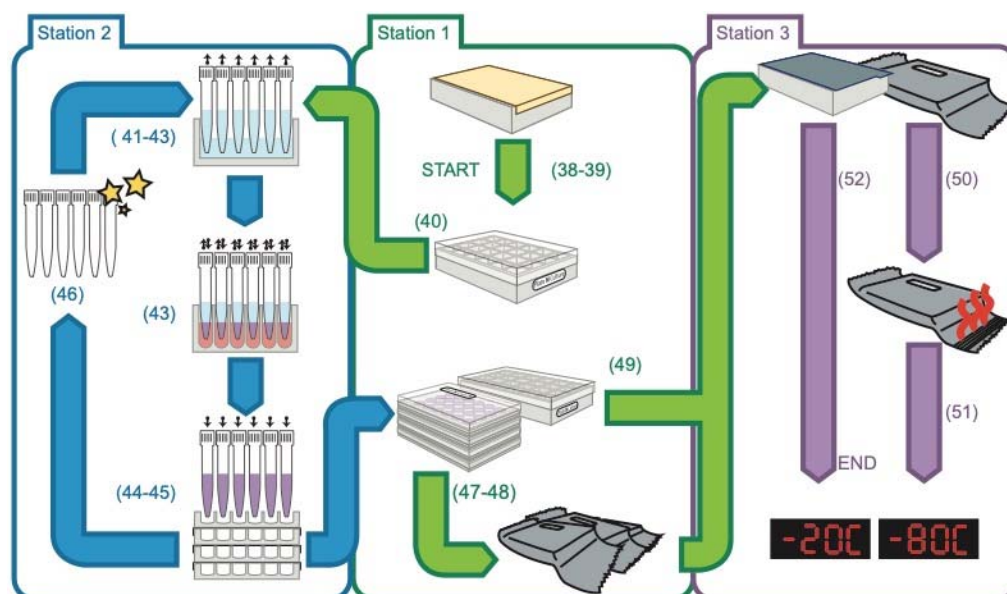
Figure 5: Cryostocking schematic. Three people work together to rapidly make and aliquot cryostocks of the library. Station 1 is at the incubator inside the anaerobic chamber. This person works to unseal the 96-deepwell plates, label them, and pass the plates to Station 2. Station 2 is at the BenchSmart inside the anaerobic chamber. This person works to mix the cultures with glycerol and aliquot the cryostocks into microwell plates. After aliquoting a set of cryostocks, this person passes them back to Station 1. Station 1 then seals the cryostocks with an aluminum seal, places them in a MylarFoil heat seal bag, and passes the cryostocks through the airlock to Station 3. Station 3 works next to the anaerobic chamber airlock. This person takes MylarFoil bags, heat-seals them, and stores the cryostocks at -80 ˚C.

1061
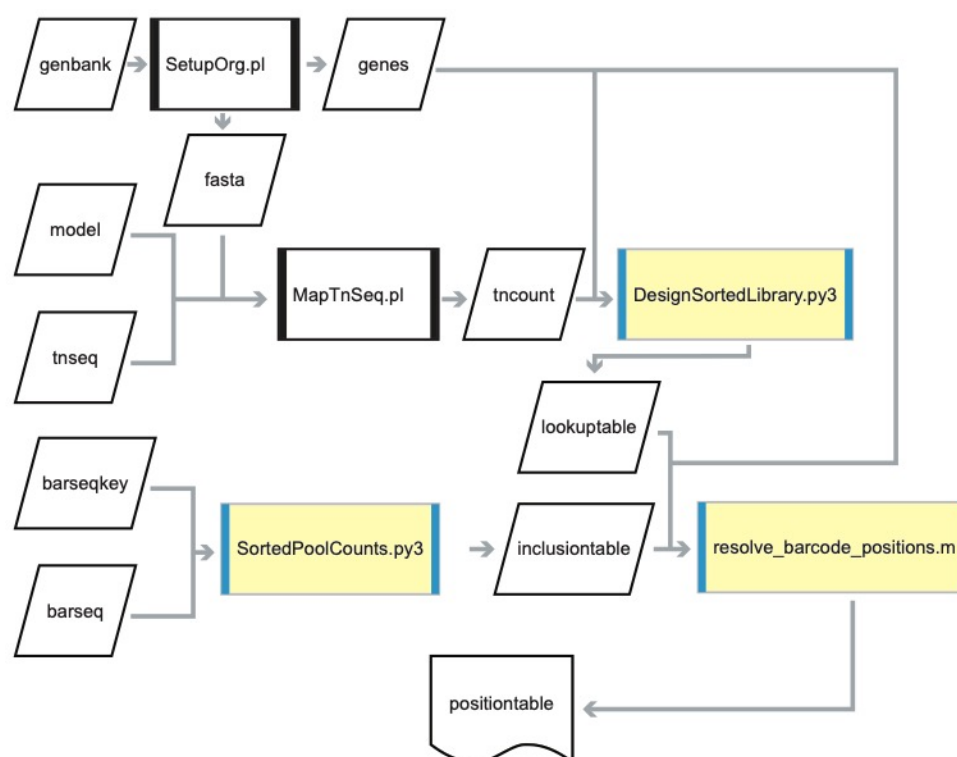
Figure 6: Analysis workflow to locate mutant strains in the ordered library. Files
and data are shown as parallelograms. Scripts are shown as rectangles. Rectangles with
black edges are scripts from the previously published FEBA code repository. Yellow
rectangles with blue edges are scripts written for this study. The language of the scripts is
reflected in the file extensions (Perl scripts end in .pl, Python scripts end in .py3, and
MATLAB scripts end in .m). See the code availability statement for more information.
Input files to the analysis are aligned on the left. The final output, *positiontable*, is at the
bottom. SetupOrg.pl takes a genbank-formatted genome sequence (*genbank*) as input as
splits it into a simple fasta-formatted sequence (*fasta*) and a table of gene locations
(*genes*). MapTnSeq.pl takes the RB-TnSeq data (*tnseq*), a model of the expected
transposon sequence in the RB-TnSeq data (*model*), and *fasta* and outputs a table of read
counts for every unique combination of barcode and insertion site location (*tncount*).
DesignSortedLibrary.py3 takes the information in *tncount* and processes it to create a
lookup table connecting unique barcodes to their likely insertion site(s) in the genome
(*lookuptable*). SortedPoolCounts.py3 takes the Bar-seq data (*barseq*) and a formatting
key (*barseqkey*) and creates a table listing read counts of every barcode in every pool
(*inclusiontable*). The final script, resolve_barcode_positions.m, uses the *inclusiontable*,
lookuptable, and *genes* files to locate transposon insertion strains in the library and record
this information in a table (*positiontable*).

1081
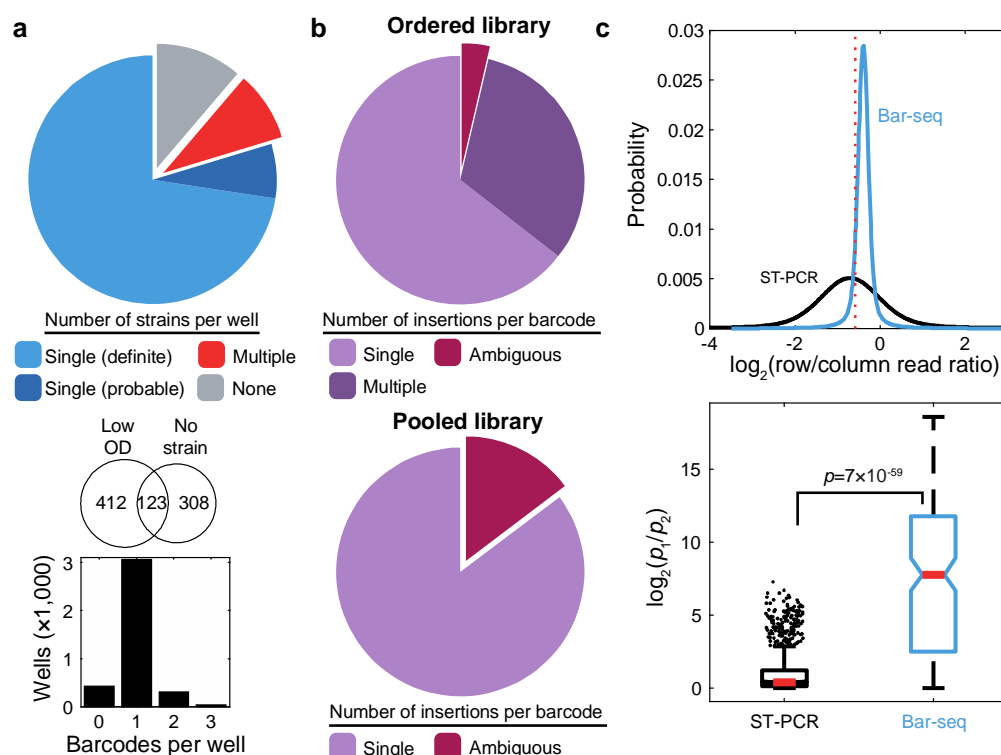
1082   **Figure 7: Statistics of strain location identification in a *B. thetaiotaomicron***

1083   **transposon insertion library sorted into 40 plates.**

1084   a) Top: we identified 2787 wells carrying a single mutant with a definite solution
1085   and 272 wells carrying a single mutant with a probable solution. 350 wells carried
1086   more than one strain and 431 wells had none assigned. Middle: only 123 of the
1087   431 wells with no assigned strain also had an $OD_{600}<0.25$, indicating that low cell
1088   density was not the sole cause of missing barcodes. Bottom: the majority of wells
1089   in the library were assigned a single barcode; no well contained more than 3
1090   unique barcodes.

1091   b) Barcodes associated with more than one insertion have an ambiguous assignment
1092   in pooled libraries. Performing RB-TnSeq on the ordered library shows that the
1093   vast majority of barcodes with more than one insertion are strains with multiple
1094   insertions (top), significantly reducing the fraction of ambiguous assignments
1095   relative to RB-TnSeq of the pooled library (bottom).

1096   c) The predictive power of Bar-seq for probabilistic location of mutants in the
1097   library is greater than ST-PCR. Our pool design shares the same row and column
1098   pools as a previous ST-PCR study, allowing us to compare the performance of the
1099   sequencing methods. Top: the $\log_2$(ratio of row/column reads) should be a Voigt
1100   distribution centered around -0.585[8] (red dotted line).The fit Voigt distributions
1101   for Bar-seq (blue, this study) and ST-PCR (black, our re-analysis of previous
1102   work[8]) were normalized to an area under the curve of 1 and plotted. Bottom: The
1103   fit Voigt distributions (above) were used to predict the likelihood of row and
1104   column solutions for all mutant strains that appeared in exactly two row and two
1105   column pools. There are two unique solutions to this 2x2 pool inclusion pattern.
1106   The log of the ratio of the probabilities for the more (p_1) and less (p_2) likely

1107   solution ($\log_2(p_1/p_2)$) is plotted for Bar-seq (blue, $n=169$) and ST-PCR (black,
1108 $n=1451$). The red dashes are the medians of the distributions, the edges of the
1109   boxes are the upper and lower quartiles, whisker length is 1.5 times the inter-
1110   quartile range, and the outliers are plotted individually. Bar-seq has a significantly
1111   higher log ratio, signifying better predictions.
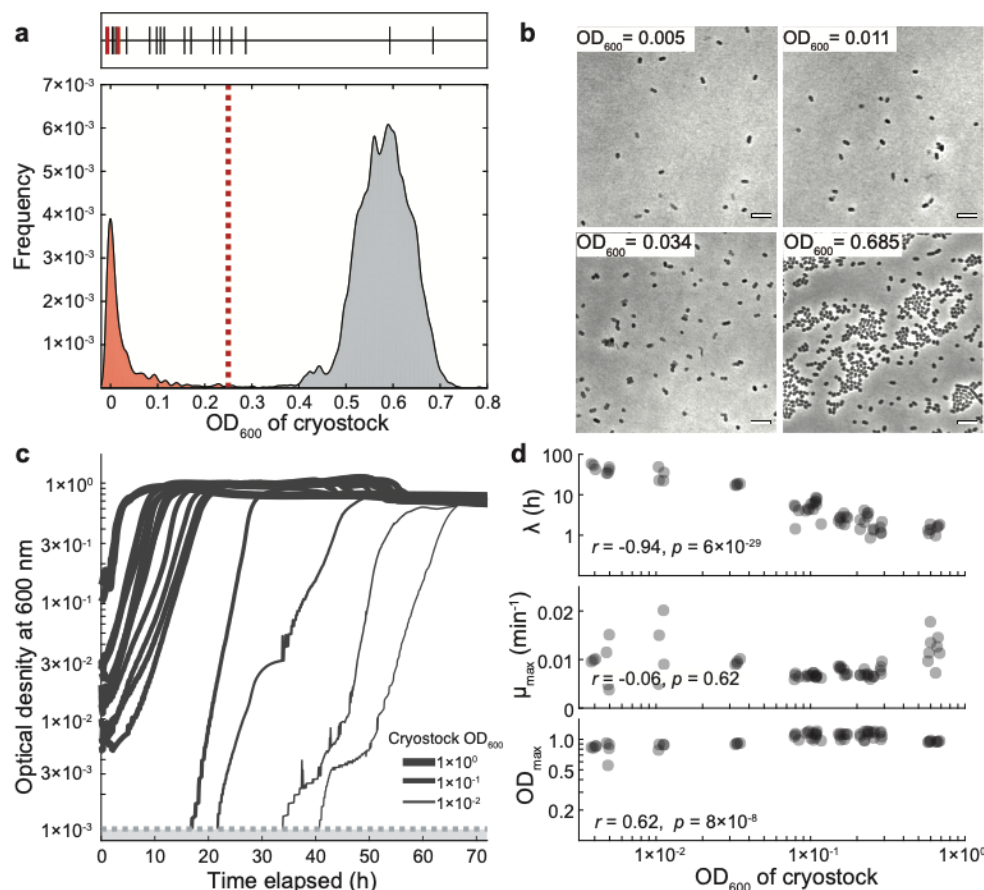
Figure 8: Growth properties of the 40-plate *B. thetaiotaomicron* barcoded transposon insertion library.

a) The distribution of OD_{600} measurements in the cryostocks of the ordered library revealed a surprisingly large population of low OD_{600} wells. A heuristic cutoff of 0.25 (red dotted line) was chosen to discriminate between high (grey distribution) and low OD_{600} (red distribution) in the library. Tick marks at the top of the distribution denote the OD_{600} of wells that were chosen for investigation. Black tick marks record wells in which cells were found by microscopy, while the two red tick marks represent the OD_{600} of wells in which no cells could be found in a 1-μL aliquot.

b) Cells were found using microscopy in most of the low OD_{600} wells in the library, even in wells with $OD_{600}<0.01$, indicating that many of the low OD_{600} wells are not due to a problem with sorting. Scale bar: 5 μm. The cryostock OD_{600} in the library is shown for each image/well. Images were acquired with a Nikon Ti-E outfitted with a 100X (NA 1.4) objective and an Andor Zyla 5.5 sCMOS camera. Micromanager was used for image acquisition[25].

c) Growth curves after recovery from the cryo-stock showed that many of the low OD_{600} wells can recover to wild-type densities. The thickness of the line is proportional to the log of the cryostock OD_{600}. All curves represent the mean of ($n=4$) independent measurements. The horizontal line at an OD_{600} of 0.001 represents the detection limit of the plate reader. The two wells in which cells

1134   were not detected did not recover during the 72 h period of monitoring and are not
1135   plotted.
1136   d) Growth parameters extracted from the recovery curves ($n=4$) show that the
1137   transposon insertion strains isolated from low OD_{600} wells have maximum OD_{600}
1138   values (OD_{max}) and maximum growth rates (μ_{max}) similar to wild-type. As
1139   expected, the lag phase (λ) of the cultures was negatively correlated with the final
1140 OD_{600} in the library (Pearson's $r=-0.94$, $p=6\times10^{-29}$).

References (Key reference highlighted in red)

1   Gentile, C. L. & Weir, T. L. The gut microbiota at the intersection of diet and human health. *Science* **362**, 776-780 (2018).

2   Tramontano, M. *et al.* Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nature Microbiology* **3**, 514-522, doi:10.1038/s41564-018-0123-9 (2018).

3   Liberati, N. T. *et al.* An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. *Proceedings of the National Academy of Sciences* **103**, 2833-2838 (2006).

4   De Berardinis, V. *et al.* A complete collection of single‐gene deletion mutants of Acinetobacter baylyi ADP1. *Molecular systems biology* **4**, 174 (2008).

5   Goodman, A. L. *et al.* Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell host & microbe* **6**, 279-289 (2009).

6   Porwollik, S. *et al.* Defined single-gene and multi-gene deletion mutant collections in Salmonella enterica sv Typhimurium. *PloS one* **9**, e99820 (2014).

7   Vandewalle, K. *et al.* Characterization of genome-wide ordered sequence-tagged Mycobacterium mutant libraries by Cartesian Pooling-Coordinate Sequencing. *Nature communications* **6**, 7106 (2015).

8   Baym, M., Shaket, L., Anzai, I. A., Adesina, O. & Barstow, B. Rapid construction of a whole-genome transposon insertion collection for Shewanella oneidensis by Knockout Sudoku. *Nature communications* **7**, 13270 (2016).

9   Koo, B.-M. *et al.* Construction and analysis of two genome-scale deletion libraries for Bacillus subtilis. *Cell systems* **4**, 291-305. e297 (2017).

10   Dale, J. L. *et al.* Comprehensive Functional Analysis of the Enterococcus faecalis Core Genome Using an Ordered, Sequence-Defined Collection of Insertional Mutations in Strain OG1RF. *mSystems* **3**, e00062-00018 (2018).

11   Baba, T. & Mori, H. The construction of systematic in-frame, single-gene knockout mutant collection in Escherichia coli K-12. *Methods Mol Biol* **416**, 171-181, doi:10.1007/978-1-59745-321-9_11 (2008).

12   Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J. & Robertson, H. M. Hyperactive transposase mutants of the Himar1 mariner transposon. *Proceedings of the National Academy of Sciences* **96**, 11428-11433 (1999).

13   Goryshin, I. Y. & Reznikoff, W. S. Tn5 in vitro transposition. *Journal of Biological Chemistry* **273**, 7367-7374 (1998).

14   Dodd, D. *et al.* A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature* **551**, 648 (2017).

15   Goodman, A. L., Wu, M. & Gordon, J. I. Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nature protocols* **6**, 1969 (2011).

16   Anzai, I. A., Shaket, L., Adesina, O., Baym, M. & Barstow, B. Rapid curation of gene disruption collections using Knockout Sudoku. *nature protocols* **12**, 2110 (2017).

17   Erlich, Y. *et al.* DNA Sudoku—harnessing high-throughput sequencing for multiplexed specimen analysis. *Genome research* (2009).

1186   18   Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria
1187   by sequencing randomly bar-coded transposons. *MBio* **6**, e00306-00315 (2015).
1188   19   Price, M. N. *et al.* Mutant phenotypes for thousands of bacterial genes of
1189   unknown function. *Nature* **557**, 503 (2018).
1190   20   Liu, H. *et al.* Large-scale chemical-genetics of the human gut bacterium
1191   Bacteroides thetaiotaomicron. *BioRxiv*, 573055 (2019).
1192   21   Holt J.G., K. N. R., Sneath P.H.A., Staley J.T., Williams S.T. *Bergey's Manual of*
1193   *Determinative Bacteriology*. 9th ed. edn, (Williams and Willkins, 1994).
1194   22   Shatalin, K., Shatalina, E., Mironov, A. & Nudler, E. H2S: a universal defense
1195   against antibiotics in bacteria. *Science* **334**, 986-990 (2011).
1196   23   Hughes, S. R. *et al.* High-throughput screening of cellulase F mutants from
1197   multiplexed plasmid sets using an automated plate assay on a functional
1198   proteomic robotic workcell. *Proteome science* **4**, 10 (2006).
1199   24   Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for
1200   timelines, timetrees, and divergence times. *Molecular Biology and Evolution* **34**,
1201   1812-1819 (2017).
1202   25   Edelstein, A. D. *et al.* Advanced methods of microscope control using μManager
1203   software. *Journal of biological methods* **1** (2014).
1204
1205

1206 **Supplemental Information**

1207

1208 **Supplementary Data 1: Table of oligos used for Bar-seq.**

1209

Oligo name	Oligo sequence	PAGE purified	Index name	Index sequence	Description
Barseq_P1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGTCGACCTGCAGCGTACG	yes			Common primer for "truseq-indexed" Bar-seq
Barseq_P2_IT 001	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	yes	IT001	ATCACG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 002	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	yes	IT002	CGATGT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 003	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	yes	IT003	TTAGGC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 004	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	yes	IT004	TGACCA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 005	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	yes	IT005	ACAGTG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 006	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGTCCACGAGGTCTCT	yes	IT006	GCCAAT	truseq-derived indexed Bar-seq primer

Barseq_P2_IT 007	CAAGCAGAAGAC GGCATACGAGAT GATCTGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT007	CAGATC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 008	CAAGCAGAAGAC GGCATACGAGAT TCAAGTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT008	ACTTGA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 009	CAAGCAGAAGAC GGCATACGAGAT CTGATCGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT009	GATCAG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 010	CAAGCAGAAGAC GGCATACGAGAT AAGCTAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT010	TAGCTT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 011	CAAGCAGAAGAC GGCATACGAGAT GTAGCCGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT011	GGCTAC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 012	CAAGCAGAAGAC GGCATACGAGAT TACAAGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT012	CTTGTA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 013	CAAGCAGAAGAC GGCATACGAGAT TTGACTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT013	AGTCAA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 014	CAAGCAGAAGAC GGCATACGAGAT GGAACTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT014	AGTTCC	truseq-derived indexed Bar-seq primer

Barseq_P2_IT 015	CAAGCAGAAGAC GGCATACGAGAT TGACATGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT015	ATGTCA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 016	CAAGCAGAAGAC GGCATACGAGAT GGACGGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT016	CCGTCC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 017	CAAGCAGAAGAC GGCATACGAGAT CTCTACGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT017	GTAGAG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 018	CAAGCAGAAGAC GGCATACGAGAT GCGGACGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT018	GTCCGC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 019	CAAGCAGAAGAC GGCATACGAGAT TTTCACGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT019	GTGAAA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 020	CAAGCAGAAGAC GGCATACGAGAT GGCCACGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT020	GTGGCC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 021	CAAGCAGAAGAC GGCATACGAGAT CGAAACGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT021	GTTTCG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 022	CAAGCAGAAGAC GGCATACGAGAT CGTACGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT022	CGTACG	truseq-derived indexed Bar-seq primer

Barseq_P2_IT 023	CAAGCAGAAGAC GGCATACGAGAT CCACTCGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT023	GAGTGG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 024	CAAGCAGAAGAC GGCATACGAGAT GCTACCGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT024	GGTAGC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 025	CAAGCAGAAGAC GGCATACGAGAT ATCAGTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT025	ACTGAT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 026	CAAGCAGAAGAC GGCATACGAGAT GCTCATGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT026	ATGAGC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 027	CAAGCAGAAGAC GGCATACGAGAT AGGAATGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT027	ATTCCT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 028	CAAGCAGAAGAC GGCATACGAGAT CTTTTGGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT028	CAAAAG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 029	CAAGCAGAAGAC GGCATACGAGAT TAGTTGGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT029	CAACTA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 030	CAAGCAGAAGAC GGCATACGAGAT CCGGTGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT030	CACCGG	truseq-derived indexed Bar-seq primer

Barseq_P2_IT 031	CAAGCAGAAGAC GGCATACGAGAT ATCGTGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT031	CACGAT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 032	CAAGCAGAAGAC GGCATACGAGAT TGAGTGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT032	CACTCA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 033	CAAGCAGAAGAC GGCATACGAGAT CGCCTGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT033	CAGGCG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 034	CAAGCAGAAGAC GGCATACGAGAT GCCATGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT034	CATGGC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 035	CAAGCAGAAGAC GGCATACGAGAT AAAATGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT035	CATTTT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 036	CAAGCAGAAGAC GGCATACGAGAT TGTTGGGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT036	CCAACA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 037	CAAGCAGAAGAC GGCATACGAGAT ATTCCGGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT037	CGGAAT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 038	CAAGCAGAAGAC GGCATACGAGAT AGCTAGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT038	CTAGCT	truseq-derived indexed Bar-seq primer

Barseq_P2_IT 039	CAAGCAGAAGAC GGCATACGAGAT GTATAGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT039	CTATAC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 040	CAAGCAGAAGAC GGCATACGAGAT TCTGAGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT040	CTCAGA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 041	CAAGCAGAAGAC GGCATACGAGAT GTCGTCGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT041	GACGAC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 042	CAAGCAGAAGAC GGCATACGAGAT CGATTAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT042	TAATCG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 043	CAAGCAGAAGAC GGCATACGAGAT GCTGTAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT043	TACAGC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 044	CAAGCAGAAGAC GGCATACGAGAT ATTATAGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT044	TATAAT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 045	CAAGCAGAAGAC GGCATACGAGAT GAATGAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT045	TCATTC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 046	CAAGCAGAAGAC GGCATACGAGAT TCGGGAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT046	TCCCGA	truseq-derived indexed Bar-seq primer

Barseq_P2_IT 047	CAAGCAGAAGAC GGCATACGAGAT CTTCGAGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT047	TCGAAG	truseq-derived indexed Barseq primer
Barseq_P2_IT 048	CAAGCAGAAGAC GGCATACGAGAT TGCCGAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT048	TCGGCA	truseq-derived indexed Barseq primer
Barseq_P2_IT 049	CAAGCAGAAGAC GGCATACGAGAT ATGTTTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT049	AAACAT	truseq-derived indexed Barseq primer
Barseq_P2_IT 050	CAAGCAGAAGAC GGCATACGAGAT TGCTTTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT050	AAAGCA	truseq-derived indexed Barseq primer
Barseq_P2_IT 051	CAAGCAGAAGAC GGCATACGAGAT GCATTTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT051	AAATGC	truseq-derived indexed Barseq primer
Barseq_P2_IT 052	CAAGCAGAAGAC GGCATACGAGAT TTTGTTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT052	AACAAA	truseq-derived indexed Barseq primer
Barseq_P2_IT 053	CAAGCAGAAGAC GGCATACGAGAT CAAGTTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT053	AACTTG	truseq-derived indexed Barseq primer
Barseq_P2_IT 054	CAAGCAGAAGAC GGCATACGAGAT AGTCTTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT054	AAGACT	truseq-derived indexed Barseq primer

Barseq_P2_I T 055	CAAGCAGAAGAC GGCATACGAGAT TCGCTTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT055	AAGCGA	truseq-derived indexed Bar-seq primer
Barseq_P2_I T 056	CAAGCAGAAGAC GGCATACGAGAT GTCCTTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT056	AAGGAC	truseq-derived indexed Bar-seq primer
Barseq_P2_I T 057	CAAGCAGAAGAC GGCATACGAGAT CCTATTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT057	AATAGG	truseq-derived indexed Bar-seq primer
Barseq_P2_I T 058	CAAGCAGAAGAC GGCATACGAGAT GTTTGTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT058	ACAAAC	truseq-derived indexed Bar-seq primer
Barseq_P2_I T 059	CAAGCAGAAGAC GGCATACGAGAT AGATGTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT059	ACATCT	truseq-derived indexed Bar-seq primer
Barseq_P2_I T 060	CAAGCAGAAGAC GGCATACGAGAT CTGGGTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT060	ACCCAG	truseq-derived indexed Bar-seq primer
Barseq_P2_I T 061	CAAGCAGAAGAC GGCATACGAGAT GCCGGTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT061	ACCGGC	truseq-derived indexed Bar-seq primer
Barseq_P2_I T 062	CAAGCAGAAGAC GGCATACGAGAT TATCGTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT062	ACGATA	truseq-derived indexed Bar-seq primer

Barseq_P2_IT 063	CAAGCAGAAGAC GGCATACGAGAT GAGAGTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT063	ACTCTC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 064	CAAGCAGAAGAC GGCATACGAGAT TCTTCTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT064	AGAAGA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 065	CAAGCAGAAGAC GGCATACGAGAT CTATCTGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT065	AGATAG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 066	CAAGCAGAAGAC GGCATACGAGAT GATGCTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT066	AGCATC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 067	CAAGCAGAAGAC GGCATACGAGAT AGCGCTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT067	AGCGCT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 068	CAAGCAGAAGAC GGCATACGAGAT CGGCCTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT068	AGGCCG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 069	CAAGCAGAAGAC GGCATACGAGAT CCGTATGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT069	ATACGG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 070	CAAGCAGAAGAC GGCATACGAGAT TAGGATGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT070	ATCCTA	truseq-derived indexed Bar-seq primer

Barseq_P2_IT 071	CAAGCAGAAGAC GGCATACGAGAT ATAGATGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT071	ATCTAT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 072	CAAGCAGAAGAC GGCATACGAGAT GCGTGGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT072	CCACGC	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 073	CAAGCAGAAGAC GGCATACGAGAT CATGGGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT073	CCCATG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 074	CAAGCAGAAGAC GGCATACGAGAT TTGCGGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT074	CCGCAA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 075	CAAGCAGAAGAC GGCATACGAGAT CTAAGGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT075	CCTTAG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 076	CAAGCAGAAGAC GGCATACGAGAT TTCTCGGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT076	CGAGAA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 077	CAAGCAGAAGAC GGCATACGAGAT CAGCAGGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT077	CTGCTG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 078	CAAGCAGAAGAC GGCATACGAGAT GGTTTCGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT078	GAAACC	truseq-derived indexed Bar-seq primer

Barseq_P2_IT 079	CAAGCAGAAGAC GGCATACGAGAT TTATTCGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT079	GAATAA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 080	CAAGCAGAAGAC GGCATACGAGAT TCCGTCGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT080	GACGGA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 081	CAAGCAGAAGAC GGCATACGAGAT TATATCGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT081	GATATA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 082	CAAGCAGAAGAC GGCATACGAGAT AGCATCGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT082	GATGCT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 083	CAAGCAGAAGAC GGCATACGAGAT CCTTGCGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT083	GCAAGG	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 084	CAAGCAGAAGAC GGCATACGAGAT AAGTGCGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT084	GCACTT	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 085	CAAGCAGAAGAC GGCATACGAGAT TAAGGCGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT085	GCCTTA	truseq-derived indexed Bar-seq primer
Barseq_P2_IT 086	CAAGCAGAAGAC GGCATACGAGAT TGGAGCGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT086	GCTCCA	truseq-derived indexed Bar-seq primer

Barseq_P2_IT 087	CAAGCAGAAGAC GGCATACGAGAT TGTGCCGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT087	GGCACA	truseq-derived indexed Barseq primer
Barseq_P2_IT 088	CAAGCAGAAGAC GGCATACGAGAT CAGGCCGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT088	GGCCTG	truseq-derived indexed Barseq primer
Barseq_P2_IT 089	CAAGCAGAAGAC GGCATACGAGAT GGTAGAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT089	TCTACC	truseq-derived indexed Barseq primer
Barseq_P2_IT 090	CAAGCAGAAGAC GGCATACGAGAT CATTCAGTGACTG GAGTTCAGACGT GTGCTCTTCCGAT CTGATGTCCACG AGGTCTCT	yes	IT090	TGAATG	truseq-derived indexed Barseq primer
Barseq_P2_IT 091	CAAGCAGAAGAC GGCATACGAGAT ATGGCAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT091	TGCCAT	truseq-derived indexed Barseq primer
Barseq_P2_IT 092	CAAGCAGAAGAC GGCATACGAGAT CCAGCAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT092	TGCTGG	truseq-derived indexed Barseq primer
Barseq_P2_IT 093	CAAGCAGAAGAC GGCATACGAGAT GCGCCAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT093	TGGCGC	truseq-derived indexed Barseq primer
Barseq_P2_IT 094	CAAGCAGAAGAC GGCATACGAGAT TTCGAAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT094	TTCGAA	truseq-derived indexed Barseq primer

Barseq_P2_I T 095	CAAGCAGAAGAC GGCATACGAGAT GGAGAAGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT095	TTCTCC	truseq-derived indexed Bar-seq primer
Barseq_P2_I T 096	CAAGCAGAAGAC GGCATACGAGAT AAACCTGTGACT GGAGTTCAGACG TGTGCTCTTCCGA TCTGATGTCCACG AGGTCTCT	yes	IT096	AGGTTT	truseq-derived indexed Bar-seq primer

1210

1211 **Supplementary Data 2: Table of oligos used for RB-TnSeq.** The oligo MOD2_TruSeq
1212 is modified to carry 5'-phosphate, as denoted by '/5'P/'. The oligo Mod2_TS_Univ is
1213 modified by an internal phosphorothioate bond, as indicated by the '*'. The
1214 MOD2_TruSeq and Mod2_TS_Univ oligos are annealed to create a Y-adapter, as
1215 specified in Reagent Setup.

Oligo name	Oligo sequence	PAGE purified	Index name	Index sequence	Description
MOD2_ TruSeq	/5'P/GATCGGAA GAGCACACGT CTGAACTCCA GTCA	yes			Y-adapter for ligation to genomic DNA fragments (1/2)
Mod2_ TS_Uni v	ACGCTCTTCCG ATC*T	yes			Y-adapter for ligation to genomic DNA fragments (2/2)
Nspacer _barseq _univer sal	ATGATACGGC GACCACCGAG ATCTACACTCT TTCCCTACACG ACGCTCTTCCG ATCTNNNNNN GATGTCCACG AGGTCT	yes			Tn-specific primer for amplification of Tn-seq library
P7_i6	CAAGCAGAAG ACGGCATACG AGATATTGGC GTGACTGGAG TTCAGACGTGT GCTCTTCCGAT CT	yes	i6	GCCAAT	Adapter-specific primer for amplification of RB-TnSeq library, indexed

1216

1217 **Supplementary Data 3: Table of oligos used to spot check strain assignments in the**
1218 **ordered library.** The transposon insertion targeted for the PCR check is annotated with
1219 position-(orientation). Every PCR check used Erm-out as the forward primer and a strain-
1220 specific reverse primer.

Oligo name	Oligo sequence	Description
Erm-out	GCAGGTAATACGACTCACTATAG	outward-oriented primer for erm cassette
BtErm-Check-P06G05	GATAAAATTGCATTGCACATTATCAC	1224-(+) transposon insertion, P6-G5
BtErm-Check-P18G11	GAGACACATCCTGAAATCATG	796028-(-) transposon insertion, P18-G11
BtErm-Check-P20A06	TAATGCATGGGAGCTTCATG	4810775-(+) transposon insertion, P20-A6
BtErm-Check-P22B06	TACGGATAGCGTCGTC	3905852-(-) transposon insertion, P22-B6
BtErm-Check-P23E01	TTGGTTCAGCGGTATGTG	233663-(-) transposon insertion, P23-E1
BtErm-Check-P26A09	CAATACTCTCCATTTTCATCTGAC	4064363-(-) transposon insertion, P26-A9
BtErm-Check-P27E09	CTTAGCTAAATACAGTTACCTGG	4662683-(+) transposon insertion, P27-E9
BtErm-Check-P17B02-A	ATTAAAGGATGCATTACGTATATGTC	3223007-(+) transposon insertion, P17-B2
BtErm-Check-P17B02-B	CGTTCAGCTTATAAGCTTCAG	5788004-(-) transposon insertion, P17-B2
BtErm-Check-P18C02-A	CCCGAAACGTGTTTTCAATG	1864693-(+) transposon insertion, P18-C2
BtErm-Check-P18C02-B	GTTCGGGACAAAGAGATTC	2999897-(-) transposon insertion, P18-C2
BtErm-Check-P26D09-A	CAGCCGTATGATTGTCTATG	3737285-(+) transposon insertion, P26-D9
BtErm-Check-P26D09-B	ATCAGTAGGTTCACCTGTAG	5488616-(+) transposon insertion, P26-D9
BtErm-Check-P31D11-A	GATACAGCAAAGCGAATACG	1408781-(+) transposon insertion, P31-D11
BtErm-Check-P31D11-B	TGCTACGACGATCTGTTC	1994095-(-) transposon insertion, P31-D11
BtErm-Check-P32F07-A	AGGTGGTGTATTGCCTG	18210-(-) transposon insertion, P32-F7
BtErm-Check-P32F07-B	CTTCATGCAAAAGCGATTG	207619-(-) transposon insertion, P32-F7
BtErm-Check-P38E06-A	GTACGACAGAGCAAAAATTAGC	3926701-(+) transposon insertion, P38-E6
BtErm-Check-P38E06-B	CAAATAGCCTGTTTTACCCC	4207683-(+) transposon insertion, P38-E6
BtErm-Check-P39E08-A	TAATGTTGTCCTCCTTGCTG	2248214-(+) transposon insertion, P39-E8
BtErm-Check-P39E08-B	GCGTGGAATATTACCACATATTTC	3630139-(+) transposon insertion, P39-E8
BtErm-Check-P40D11-A	AACAGCGAAGGAGACAAG	2500909-(-) transposon insertion, P40-D11
BtErm-Check-P40D11-B	CGACTTCGTTCTAGAAACC	3061114-(+) transposon insertion, P40-D11

1221

1222 **Supplementary Figure Legends**

1223

1224
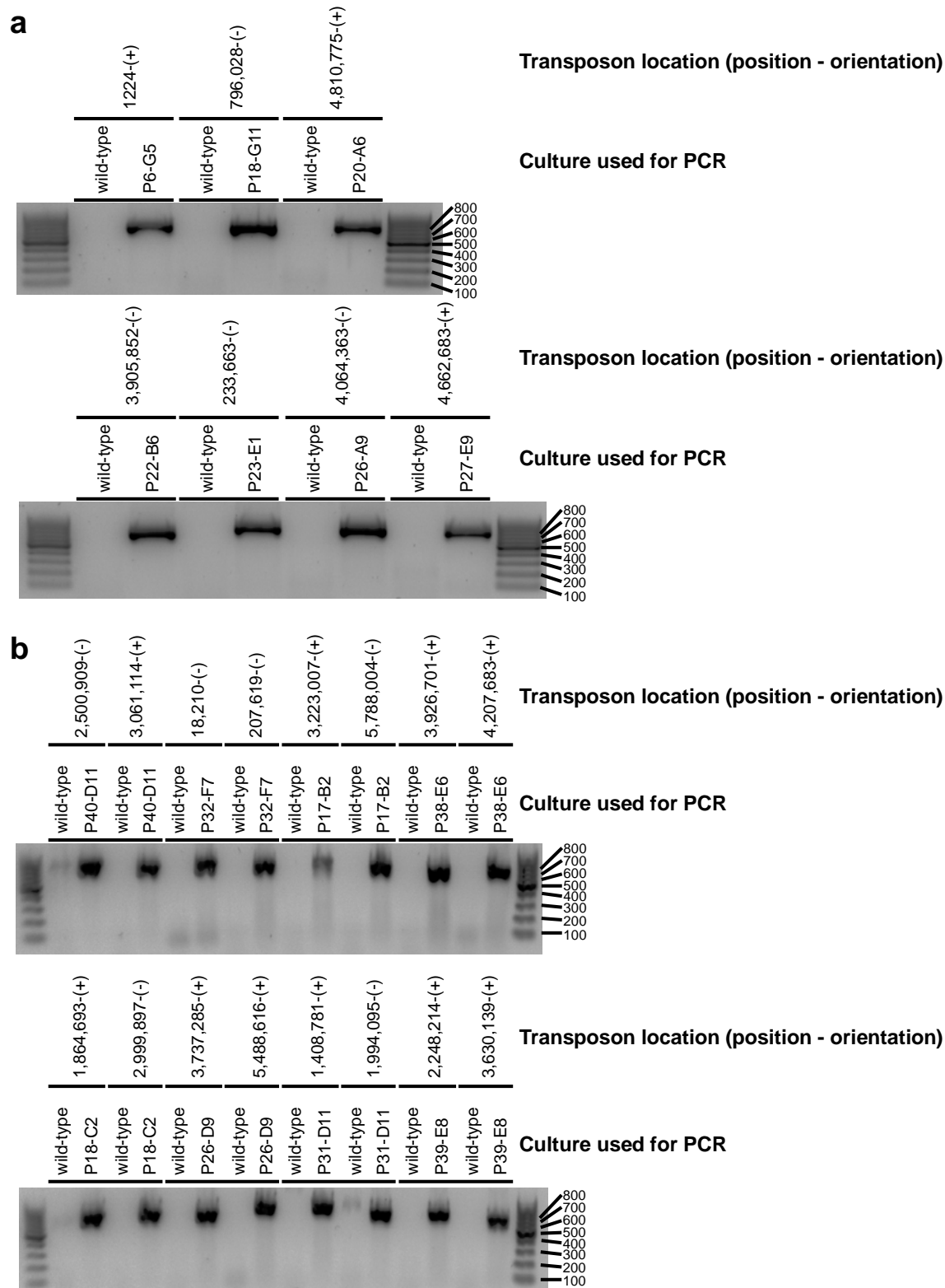
1225 **Supplementary Figure 1: PCR validations of 15 transposon insertion strain**
1226 **assignments within the ordered library.** PCRs were used to confirm that the protocol

1227 correctly assigns transposon insertion strains to locations in the ordered library. PCRs
1228 were performed with a universal outward facing primer that bound to the Erm cassette
1229 and a specific reverse primer that bound to the genome near the insertion point.

1230   a) Seven PCR validations of wells with a single strain with a single transposon
1231   insertion. The top label denotes the transposon insertion location that was targeted
1232   using a position-(orientation) notation. The bottom label records the culture used
1233   as input in the PCR (either an overnight culture of wild-type *B. theta* VPI-5482 or
1234   an aliquot of the cryostock in the ordered library at the position indicated, with P#
1235   abbreviating plate #). All seven wells contained the assigned transposon insertion.

1236   b) Eight PCR validations of wells with a single strain with more than one insertion.
1237   Two insertion sites were checked for each well. The lanes are labeled as in (a).
1238   All eight wells contained both of the tested transposon insertions.