Rapid ordering of barcoded transposon insertion libraries of anaerobic bacteria

Anthony L. Shiver1, Rebecca Culver2, Adam M. Deutschbauer3,4 and Kerwyn Casey Huang1,5,6

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, the strict growth requirements of the individual species and a lack of genetic tools and resources. 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 2 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.

Introduction

There is a growing appreciation of the role that commensal bacteria play in human health1. However, the complexity of the microbiota, the fastidious growth requirements of its members2 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 nonessential 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 species3–11. 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 nonmodel organisms. A small number of well-characterized transposon systems (e.g., mariner12 and Tn513) 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 metabolites14. Thus, while transposon mutagenesis is a fruitful first approach for genetic analysis of understudied gut microbiota members, a major challenge remains in ordering the random pool of transposon mutants to analyze individual strains.

Recent technological advances have vastly lowered the cost and effort required to order random transposon insertion libraries15,16. 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. 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 locate 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 one of a range of transposon insertion sequencing approaches, e.g., transposon sequencing (Tn-seq), insertion sequencing (INSeq) and semi-random two-step PCR (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.

Development of the protocol
We have incorporated two further advances to facilitate sorting of obligate anaerobic bacteria and 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 sorter (>5,000 mutants per h) to order strains in high throughput while keeping aerobic exposure time below 30 min. The second advance utilizes the reproducibility and simplicity of barcode sequencing to reduce the effort required to locate mutations in the ordered library.

We previously developed the use of random barcode transposon-site sequencing (RB-TnSeq) as an alternative sequencing approach for high-throughput functional genomics. With RB-TnSeq, the entire transposon insertion library is initially sequenced as one sample, allowing association of random DNA barcodes with transposon insertions across the entire pool. Subsequently, transposon insertion counts can be followed with a single PCR step to amplify barcodes followed by next-generation sequencing, i.e., barcode sequencing (Bar-seq). This approach allowed us to record genome-wide mutant fitness data for 32 diverse bacteria across dozens of growth conditions. The same principles that make Bar-seq useful for chemical genomics, namely the ease of scale and quantitative accuracy, also make it an ideal sequencing approach for locating mutations within an ordered library.

Alternative methods
In this protocol, we used Bar-seq to sequence pools and locate mutant strains within the library. Alternative sequencing methods that have been used in previous studies include INSeq, Tn-seq and ST-PCR. Bar-seq and ST-PCR are 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 inexpensive and simple, it suffers from inherent noise during the first PCR amplification, in which one of the two primers has semirandom specificity. Likely because of this, Bar-seq preserves quantitative information on relative barcode frequency within libraries better than ST-PCR (Supplementary Results, Supplementary Fig. 1), which is important when predicting the positions of repeated barcodes in the library.

Limitations
While Bar-seq simplifies the process of locating mutants within an ordered library, it requires a diverse progenitor pool with a large number of unique barcodes. Therefore, the main limitation of our protocol is that it is restricted to bacterial strains with a relatively high transformation efficiency such that diverse pools of barcoded transposon insertions can be generated. If a high-efficiency transformation protocol can be identified for a strain of interest, the use of Bar-seq promises to lower the barrier to generating an ordered library from a random pool of barcoded transposon insertions.

Sorting using a FACS machine is a fast method for isolating mutants from a large pool, and we have sorted multiple anaerobic species with a high recovery rate when the growth medium is supplemented with an antioxidant (Supplementary Results, Supplementary Fig. 2). However, it may not be the best approach in all situations. Certain strains might be too sensitive to oxygen to survive sorting and will need to remain in an anaerobic environment for the entire protocol. If the number of strains in the original pool is not significantly greater than the number of strains to be sorted, the same mutant strains are likely to be isolated more than once in the ordered library. Finally, a core facility with FACS machines may not be available. In these cases, it may be preferable to directly pick
transformed colonies from a selective plate. Automated colony-picking equipment with a footprint small enough to install in an anaerobic chamber is commercially available\(^1\).

**Optimization and applications of the protocol**

We have incorporated flow sorting and barcode sequencing into a protocol (Fig. 1) that allows us to rapidly generate an ordered collection of >10,000 barcoded transposon insertion mutants in strictly anaerobic bacteria such as those found in the human gut microbiota. To develop this protocol, we first demonstrated that strict anaerobes can survive the temporary exposure to oxygen that occurs during single-cell sorting with a FACS machine, particularly in the presence of specific antioxidants (Supplementary Results, Supplementary Fig. 2). We capitalized on this finding to develop a protocol built around sorting single cells from a pooled mutant library. We optimized this protocol to reduce the time, cost and levels of cross-contamination. Finally, we used this protocol to order a barcoded transposon insertion library in the model human gut commensal *Bacteroides thetaiotaomicron* VPI-5482. Analyses of individual mutants from this ordered library allowed us to further characterize chemical sensitivities that were detected from a pooled chemical genomics screen of *B. thetaiotaomicron* VPI-5482 (ref. \(^2\))

We have demonstrated that a single-cell sorting protocol coupled to barcode sequencing can rapidly and straightforwardly generate a sorted mutant library for anaerobic species. Given a diverse transposon insertion pool, both the sorting portion of the protocol and the strain identification portion can be accomplished in 2 weeks. This protocol significantly lowers the investment required to generate an ordered library of mutants. As transformation methods are developed for more bacterial constituents of the human microbiota, we expect that this protocol will enable the rapid and cost-effective investigation of microbe–host interactions, metabolite production and other genotype–phenotype relationships in an ever-expanding set of bacteria.
Experimental design
Generating a barcoded transposon insertion library
In this approach, 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 nonreplicating 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. The transformation protocol to generate the barcoded library will depend on the strain being studied.

During the creation of a transposon insertion library, an RB-TnSeq experiment as described in Steps 84–94 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.

Optimizing the sorting procedure
The fill rate is the fraction of wells in the 96-well plate in which a culture grows after sorting. The fill rate should ideally be 100% so that fewer 96-well plates will be necessary to isolate the same number of mutants, saving storage space and reagent costs. Multiple factors will be critical for maximizing the fill rate for 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 might produce these types of particles (e.g., outer membrane vesicles, extracellular polysaccharides) during the course of growth. Second, the majority of cells in the population must be viable. The viability of cells will depend on the particular strain and growth phase of the culture. We suggest avoiding cultures that have spent more than a few hours in stationary phase when first determining the optimal sort conditions. Third, single cells have to survive the time spent in an aerobic environment that is required to sort the library.

Given that many species from the gut microbiota are strict anaerobes, we studied the effect of oxygen exposure on different species further. We found that many strict anaerobes isolated from the gut microbiota can survive single-cell sorting with an exposure time of up to ~1 h outside of an anaerobic atmosphere in commercial media (Supplementary Results, Supplementary Fig. 2a). Given that many commercial medium recipes for anaerobes include antioxidants, we were interested in the extent to which antioxidant additives were able to protect sorted cells from death. For two Clostridium sp., the presence of antioxidants was necessary (Supplementary Results, Supplementary Fig. 2b). Furthermore, the addition of different antioxidants had different effects on the two species, indicating that testing a panel of antioxidants will be helpful for protecting sorted cells of different species from oxidative damage.

In addition to maximizing the fill rate, it is important to ensure that every well in the sorted library receives only one cell. Careful consideration of the gate on forward scatter (FSC) signal and side scatter (SSC) signal will prevent the sorting of multicell aggregates. The distribution of trigger pulse length for the sorted cells should be unimodal.

While extensive optimization of growth conditions for B. thetaiotaomicron was not necessary, there are a number of avenues for troubleshooting the sorting step for difficult bacteria. Changing the growth phase of the sorted cells (e.g., log phase versus early stationary phase) is likely to alter the fraction of cells that survive sorting as well as the extent of aggregation within the culture. Testing different growth and recovery media may also be an important aspect of troubleshooting. Finally, while cysteine may need to be avoided as an antioxidant for known H2S producers (see below), there are other antioxidants that can be added as a supplement to the growth and recovery media.

Eliminating H2S production
H2S is a toxic, caustic gas. Build-up of H2S inside the anaerobic chamber will degrade sensitive electronic equipment. Exposure to H2S gas should be minimized. Before working with large volumes of culture, the growth conditions that will prevent the production of H2S during growth should be determined.
Enzymes with cysteine desulfhydrase activity degrade cysteine and produce \( \text{H}_2\text{S} \) as a byproduct. These enzymes are conserved across the eubacteria\(^2\), and \textit{Bacteroides} sp. are known to produce \( \text{H}_2\text{S} \) gas\(^2\). Given the large volumes of culture involved in growing an ordered library, even moderate production of \( \text{H}_2\text{S} \) has the potential to overcome standard \( \text{H}_2\text{S} \) removal columns and pose a danger to personnel and equipment. Our original recipe for the Brain Heart Infusion supplemented (BHIS) growth medium of \textit{B. thetaiotaomicron} included \(8 \text{ mM} \) cysteine as an antioxidant, and \textit{B. thetaiotaomicron} cultures produced significant levels of \( \text{H}_2\text{S} \) in pilot experiments.

To eliminate \( \text{H}_2\text{S} \) production during the outgrowth step after sorting cells, we first confirmed that the level of \( \text{H}_2\text{S} \) produced by \textit{B. thetaiotaomicron} was proportional to the concentration of added cysteine (Supplementary Results, Supplementary Fig. 3a). Luckily, the growth of \textit{B. thetaiotaomicron} was not dependent on added cysteine (Supplementary Results, Supplementary Fig. 3b,c). From these results and the known oxygen tolerance of \textit{B. thetaiotaomicron}, we decided to proceed without adding cysteine to the medium for sorting the transposon insertion library.

The fill rate was 86% when sorting the barcoded transposon insertion library into 40 96-well plates without adding cysteine to either the FACS tube or the 96-well plates. Further analysis revealed that this 14% of nonsurviving cells was likely unrelated to the physiological effects of the transposon insertion mutant, implicating the transient exposure to oxygen as the likely cause of death (Supplementary Results, Supplementary Fig. 4). For strains that are not aerotolerant like \textit{B. thetaiotaomicron} but produce \( \text{H}_2\text{S} \) from cysteine, it may be more important to replace cysteine with an alternative antioxidant. Alternatively, inclusion of cysteine in only the FACS tube used for sorting might offer partial protection from oxygen during the sort without the risk of producing dangerous amounts of \( \text{H}_2\text{S} \) afterwards.

**Determining the pool design**

Our strategy was to pool according to plates, rows and columns (Fig. 1). This pool design scales linearly with the number of 96-well plates in the library, where \( N_{\text{pools}} = N_{\text{plates}} + 20 \). For larger libraries, previous approaches have used an alternative pool design (plate-row, plate-column, row, column)\(^7\)\(^8\) that scales with the square root of plate number \( N_{\text{pools}} \approx \sqrt{N_{\text{plates}}} + 20 \). Our plate-row-column pool design is less sensitive to ambiguity from multiple occurrences of the same strain in the library, making it easier to probabilistically locate these strains. For example, a strain that is isolated twice in the library will have four potential solutions in a plate-row-column pool design compared with eight potential solutions in the alternative four-dimensional pool design. Furthermore, the ease of Bar-seq reduces the relative cost savings of a pool design that scales with the square root of plate number compared with linearly. Nonetheless, ordered libraries with a larger number of plates might benefit from this alternative pool design. Similarly, ordered libraries with a high degree of repetition (same mutant, multiple locations) would benefit from a DNA Sudoku pool design\(^17\). Regardless of the pool design used, we expect that Bar-seq will improve the quality and reduce the cost and effort to sequence pools. Small modifications to the tables \textit{lookuptable} and \textit{inclusiontable} that are produced at the beginning of our analysis pipeline (Fig. 2) should be sufficient for adapting Bar-seq to analysis of alternative pool designs.

**Preparing scripts and input files for analysis of mutant location**

Two code repositories will have to be downloaded before the mutants can be located within the ordered library. Follow the links in the Software subsection of the Materials section to download the code repositories, and carefully follow the installation instructions to correctly install and configure the two repositories. In subsequent steps, the code available from https://bitbucket.org/berkeleylab/feba/src will be referred to as the FEBA repository.

A genome sequence for the strain of interest will be required to map the barcoded transposon insertions. If the genome sequence is publicly available on NCBI, it can be downloaded from https://www.ncbi.nlm.nih.gov/genome/. Download the genome in a GenBank format so that both the raw sequence and gene annotation files can be generated using scripts from the FEBA repository. If the genome sequence of your strain is not publicly available, you will have to generate the genome sequence and corresponding gene annotations before proceeding to map transposon insertions in the genome.
Materials

Reagents
Media
- Bacto Brain Heart Infusion (BHI; Becton Dickinson, cat. no. 237500)
- Sodium bicarbonate (Fisher Bioreagents, cat. no. BP328-500)
- Hemin, porcine (Alfa Aesar, cat. no. A11165)
- Erythromycin (Sigma, cat. no. E5389-5G)
- Glycerol (Fisher Bioreagents, cat. no. BP229-1)

Chemicals
- Sodium hydroxide (Fisher Bioreagents, cat. no. S318-500)
- Hydrochloric acid (Fisher Chemical, cat. no. A144-500)

PCR and DNA isolation
- QiAamp 96 DNA QiAcube HT Kit (Qiagen cat. no. 51331)
- 5× Q5 reaction buffer and high GC enhancer (New England Biolabs, cat. no. B9027S)
- Deoxynucleotide (dNTP) solution mix (New England Biolabs, cat. no. N0447L)
- Q5 high-fidelity DNA polymerase (New England Biolabs, cat. no. M0491L)

---

Fig. 2 | 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 and splits it into a simple fasta-formatted genome sequence (genome) 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 genome 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).
96-well PCR plate (Eppendorf, cat. no. 0030133382)
Zymo DNA clean and concentrator kit (Zymo Research, cat. no. D4013)
JumpStart DNA polymerase (Sigma Aldrich, cat. no. D9307-50U)
Agencourt AMPure XP beads (Beckman Coulter, cat. no. A63880)
NEBNext DNA Library Prep Master Mix Set for Illumina (New England Biolabs, cat. no. E6040S)
microTUBE AFA Fiber Pre-Slit Snap-Cap (Covaris, cat. no. 520045)
Ultrapure agarose (Invitrogen, cat. no. 16500-500)
SYBR Safe DNA gel stain (Invitrogen, cat. no. S33102)
Gel loading dye, blue (New England Biolabs, cat. no. B7021S)
Quick-Load 100 bp DNA ladder (New England Biolabs, cat. no. N0467S)

Oligos
- Indexed primers (Supplementary Table 1) for amplifying the barcodes (Integrated DNA Technologies, custom order)
- Oligos (Supplementary Table 2) for performing RB-TnSeq (Integrated DNA Technologies, custom order)

Strains
- Barcoded transposon insertion library that is compatible with Bar-seq (see 'Experimental design')

Equipment
- Benchsmart 96 semi-automated pipetting system (installed within an anaerobic chamber), 1,000 µL (Mettler Toledo, cat. no. BST-96-1000)
- Benchsmart 96 semi-automated pipetting system (accessible outside of any anaerobic chamber), 200 µL (Mettler Toledo, cat. no. BST-96-200)
- Stylus pen (Liberrway, cat. no. B01IHBVGOM)
- Forced air incubator for vinyl chambers (Coy Laboratory Products, cat. no. 2000)
- Teflon-coated 6″ hot jaw sealer, portable, constant heat (Impak Corporation, cat. no. IPKHS-606T)
- BD GasPak EZ large incubation container (Becton Dickinson, cat. no. 260672)
- BD Influx cell sorter (Becton Dickinson, cat. no. 646500) or equivalent
- Genesys 20 visible spectrophotometer (Thermo Fisher Scientific, cat. no. 4001-000) or equivalent
- QIAcube HT system (QIAGen, cat. no. 9001793) or equivalent
- PCR machine (Eppendorf Mastercycler X50a, cat. no. 6313000018) or equivalent
- MiSeq system (Illumina, cat. no. SY-410-1003) or equivalent
- HiSeq 4000 system (Illumina, cat. no. SY-401-4001) or equivalent
- S220 focused ultrasonicator (Covaris, cat. no. 500217) or equivalent

Plasticware and consumables
- Nalgene Rapid-Flow sterile disposable bottle top filters with a polyether sulfone (PES) membrane, 45 mm neck size (Thermo Fisher Scientific, cat. no. 5974520) ▲CRITICAL PES is superior to cellulose acetate, surfactant-free cellulose acetate and PVDF for filtering complex media such as BHI without clogging.
- PlateOne 2 mL 96-deepwell polypropylene plates (USA Scientific, cat. no. 1896-2000)
- Sterile 96-microwell V-bottom polystyrene plates (Grenier Bio-One, cat. no. 651161)
- Sterile microwell plate lids (Grenier Bio-One, cat. no. 656101)
- Standard profile polypropylene reservoirs (Mettler Toledo, cat. no. LR-R2-PB-5)
- Breathe-Easier gas-permeable sealing membranes (Diversified Biotech, cat. no. BERM-2000)
- Polypropylene round bottom tubes, 5 mL (Falcon, cat. no. 352063)
- Vinyl tape (3M, cat. no. 471)
- Nunc aluminum seal tape for 96-well plates (Thermo Fisher Scientific, cat. no. 232698)
- Alumaseal 384 films (Excel Scientific, cat. no. F-384-100)
- LTS filter tips, 1 mL, for use with BenchSmart (Mettler Toledo, cat. no. 30296782)
- LTS filter tips, 200 µL, for use with BenchSmart (Mettler Toledo, cat. no. 17010646)
- Filter tips, 200 µL, for use with multichannel pipette (Rainin, cat. no. 30389239)
- OD PAKVF4 MylarFoil pouches, 8.5″ × 8.75″, opening on the ZipSeal end (Impak Corporation, cat. no. 085MFS0ZE0875)
• 500cc oxygen-absorbing packets (Impak Corporation, cat. no. SF500CS1500)
• Microtube Tough-Tags for laser printers, 0.91 × 0.32 in (Diversified Biotech, cat. no. 9186-1000)
• Microtube Tough-Tags for laser printers, 1.28 × 0.5 in (Diversified Biotech, cat. no. TTLW-2016)

Glassware
• Pyrex bottles, 2 L, 45 mm neck size (Pyrex, cat. no. 1395-2L)
• Pyrex bottles, 1 L, 45 mm neck size (Pyrex, cat. no. 1395-1L)
• Erlenmeyer flasks, 250 mL (Pyrex, cat. no. 4980-250)

Software
• Software for RB-TnSeq is available from https://bitbucket.org/berkeleylab/feba/src.
• Scripts for resolving mutant positions within an ordered library are available from https://bitbucket.org/kchuanglab/resolve_barcode_position/src. For installation and configuration, follow the instructions in the respective repository.

Reagent setup

▲ CRITICAL  The volume of media and number of plates have been chosen for a 40-plate library; adapt accordingly for larger libraries. ▲ CRITICAL  It is critical to filter sterilize all media in which cells are grown prior to sorting in order to reduce the frequency of empty wells that result from sorting particulate matter. At a minimum, the 1 L of medium used to grow the library before sorting needs to be filter sterilized. We also filter sterilize the recovery medium to avoid wait time and batch effects due to the effects of autoclaving on the medium.

Bar-seq indexed primer plate
Combine Barseq_P1 with each of the 96 unique, indexed reverse primers (BarSeq_P2; Supplementary Table 1) at a final concentration of 4 µM for each primer, and store in a 96-well plate format. Store the oligo mixes at −20 °C, where they will be stable for years.

RB-TnSeq Y-adapter
Mix the oligos MOD2_TrueSeq and Mod2_TS_Univ (Supplementary Table 2) in a 50 µL volume in a PCR tube at a concentration of 10 µM for each oligo. Use a PCR machine to anneal the oligos using the ‘Anneal Y-adapter’ program (see PCR programs in ‘Equipment setup’). Store the annealed oligos at −20 °C, where they will be stable for years.

Bar-seq PCR master mix (32.5 µL mix for a 50 µL PCR reaction)
Make the master mix fresh, and store on ice until ready to use.

| Component                  | Volume   |
|----------------------------|----------|
| Deionized water            | 11 µL    |
| 5× Q5 reaction buffer      | 10 µL    |
| 5× Q5 GC enhancer          | 10 µL    |
| 10 mM dNTPs                | 1 µL     |
| Q5 polymerase              | 0.5 µL   |

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

| Component                                    | Volume   |
|----------------------------------------------|----------|
| Deionized water                             | 21.5 µL  |
| JumpStart Taq buffer                        | 10 µL    |
| 10 mM dNTPs                                 | 2 µL     |
| 100 mM MgCl₂                                | 1.5 µL   |
| 10 µM Nspacer_barseq_universal              | 5 µL     |
| 10 µM P7_i6 (Supplementary Table 2)         | 5 µL     |
| DMSO                                         | 3 µL     |
| JumpStart Taq polymerase                    | 2 µL     |
Medium components

- Sodium bicarbonate (10% wt/vol, pH 7.4): to make 1 L, add 700 mL of deionized water to a 1 L beaker. Add 100 g of sodium bicarbonate and stir gently. Insert a calibrated pH meter, and slowly add dilute HCl acid (0.1 M) to lower the pH to 7.4. When the target pH has been reached, add deionized water to a final volume of 1 L. Filter sterilize the solution with a 0.22 µm PES filter. The solution is stable for at least 1 month at room temperature (RT; 20–22 °C) with a tightly closed lid. **CRITICAL** Avoid heat, vigorous stirring and concentrated acid solutions to prevent driving CO₂ out of solution.

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

- Antibiotic stock (e.g., 10 mg/mL erythromycin): make an antibiotic stock at the desired concentration. For the *B. thetaiotaomicron* library, we dissolved 200 mg of erythromycin in 20 mL of absolute ethanol for a final concentration of 10 mg/mL, allowing us to make 20 L of BHIS supplemented with 10 µg/mL erythromycin. The stability of antibiotic stocks depends on the antibiotic and solvent. We stored the erythromycin stock at −20 °C and used the stock within 1 week.

Sort recovery medium

Make the sort recovery medium in 5 L batches: for BHIS without cysteine, add 4 L of deionized water to a 5 L beaker. Add 185 g of commercial BHI mix, and stir until the solution clears. Microwave 500 mL of deionized water to bring to a boil, add to the solution and continue stirring. Add 50 mL of the hemin solution and 100 mL of the 10% sodium bicarbonate solution. Bring the final volume to 5 L using a graduated cylinder. Filter sterilize into 2 L flasks using a 0.22 µm PES filter. Repeat this process until all 15 L of medium have been made.

50% (wt/vol) Glycerol

To prepare 3 L of 50% (wt/vol) glycerol, transfer 1.5 kg of glycerol to a 5 L beaker. Add water to 3 L final volume and transfer to three 1 L bottles. Autoclave to sterilize. This solution can be stored for weeks at RT before being brought into the anaerobic chamber to prereduce.

Equipment setup

Prepare and sterilize plastic and glassware

Wrap 45 2 mL 96-deepwell plates in foil in stacks of five. Autoclave to sterilize. Wrap the reservoirs in foil individually. Autoclave to sterilize. Cover the mouth of four 250 mL Erlenmeyer flasks with foil. Autoclave to sterilize.

Prepare anaerobic chamber

Move medium and plasticware into the anaerobic chamber. Store the bottles of medium and glycerol with bottle caps unscrewed to allow gas exchange. This includes

- Seven 2 L bottles of sort recovery medium
- One 1 L bottle of sort recovery medium
- Two sterilized reservoirs
- 45 sterilized 2 mL 96-deepwell plates
- Ten 1 mL LTS filter-tip boxes
- Four 250 mL Erlenmeyer flasks
- One opened bag of 5 mL polypropylene round-bottom tubes

Finally, let the medium sit for 48–72 h in the anaerobic chamber. **CRITICAL** It is good practice to allow all materials that need to be anaerobic to sit in the anaerobic chamber for more than 48 h. If changes are made in the protocol to adapt for specific needs (e.g., duration of growth in Step 34), then corresponding changes must also be made to when materials for the next set are brought into the anaerobic chamber to remove oxygen.

Label 96-microwell plates

Unpack a 96-microwell plate and pair with a sterile lid. Add a label to the lid, the front of the plate and to one side of the plate. Repeat this process for as many plates and copies as desired. For six copies of a 40-plate library, process 240 96-microwell plates. After labeling the plates, store in stacks in large plastic bags until it is time to bring them into the anaerobic chamber.
BenchSmart programs
Transfer media (Step 3)

| 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              |

Dispense glycerol (Step 40)

| Option                          | Setting          |
|---------------------------------|------------------|
| Pipetting mode                  | MULTI-DISP       |
| Volume                          | 557 µL           |
| Number of aliquots              | 1                |
| Aspirate/dispense speed         | Asp: 5/Disp: 5   |
| OPTION: fixed volume            | OFF              |
| OPTION: volume sequencing       | OFF              |
| OPTION: cycle counter           | OFF              |
| OPTION: dispense blowout        | OFF              |
| OPTION: auto pace               | OFF              |

Repeat dispense (Step 42)

| 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              |

Mix cultures (Step 60)

| Option                          | Setting          |
|---------------------------------|------------------|
| Pipetting mode                  | MULTI-DISP       |
| Aliquot volume                  | 150 µL           |
| Number of aliquots              | 1                |
| Aspirate/dispense speed         | Asp: 5/Disp: 5   |
| OPTION: fixed volume            | OFF              |
| OPTION: volume sequencing       | OFF              |
| OPTION: cycle counter           | OFF              |
| OPTION: dispense blowout        | OFF              |
| OPTION: auto pace               | OFF              |
Dispense pools (Step 62)

| 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              |

Combine pools (Step 72)

| 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              |

PCR programs

Bar-seq (Step 79)

| Temperature | Time | Cycles |
|-------------|------|--------|
| 98 °C       | 5 min|        |
| 98 °C       | 30 s | 25 cycles |
| 55 °C       | 30 s | 25 cycles |
| 72 °C       | 30 s | 25 cycles |
| 72 °C       | 5 min|        |
| 12 °C       | Hold |        |

RB-TnSeq (Step 90)

| Temperature | Time | Cycles |
|-------------|------|--------|
| 94 °C       | 2 min| 25 cycles |
| 94 °C       | 30 s | 25 cycles |
| 65 °C       | 20 s | 25 cycles |
| 72 °C       | 30 s | 25 cycles |
| 72 °C       | 10 min|        |
| 12 °C       | Hold |        |

Anneal Y-adapter (see 'Reagent setup')

| Temperature | Time | Cycles |
|-------------|------|--------|
| 95 °C       | 2 min|        |
| 95 °C (−1 °C/cycle) | 38 s| 71 cycles |
| 12 °C       | Hold |        |
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 each 2 L bottle of sort recovery medium 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. Set the BenchSmart program to ‘Transfer Media’ (see ‘Equipment setup’). 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 the level of medium falls below the one-fourth mark in the reservoir during the experiment, add more.
6. Do two transfers of 650 µL (1.3 mL total) from the reservoir to a 2 mL 96-deepwell plate.
7. Cover the reservoir.
8. Seal the 96-deepwell 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 and discard.
   ▲ **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 four 250 mL Erlenmeyer flasks containing 100 mL sort recovery medium with antibiotic from Step 12 to 37 °C by placing in the 37 °C incubator.
14. After 1 h, thaw a 2 mL aliquot of the mutant library and add the entire contents to one 250 mL flask from Step 13 (the other flasks will be used for maintaining the culture in log phase on the next day in Steps 17 and 18).
15. Grow the culture overnight in the anaerobic 37 °C incubator.

Day 2: sort the library ● **Timing 1 d**
▲ **CRITICAL** Sorting the transposon library will require a team of people to work together. One person will travel between the anaerobic chamber and the FACS machine to deliver fresh deepwell plates and return the sorted plates. One person will work at a station adjacent to the FACS machine to unseal new plates, seal sorted plates and assist in loading fresh plates into the FACS machine. One person will operate the FACS machine. It is important that the person running the FACS machine has expertise on the details of operating the machine (Fig. 3).
16. Determine the optical density of the overnight culture using a UV-Vis spectrophotometer.
   ▲ **CRITICAL STEP** Optical density measurements vary between spectrophotometers, and every bacterial strain has an idiosyncratic growth curve in different media. Thus, it will be useful to run a growth curve with the same UV-Vis spectrometer used for this protocol to calibrate OD600 readings to the growth phase of your bacterium.
17. Dilute the overnight culture to an OD600 = 0.05 into another prewarmed Erlenmeyer flask of sort recovery medium from Step 13.
18. Periodically monitor the OD600, keeping the culture in log phase by diluting it into another of the prewarmed Erlenmeyer flasks from Step 13 at a 1:4 dilution.
19. Prewarm between 1 and 12 3 mL aliquots of sort recovery medium in FACS tubes by placing in the anaerobic 37 °C incubator.
20. Before sorting, measure the OD600 of the culture and dilute into the first FACS tube to an OD600 of 0.01–0.05. As an alternative, to save time and effort, estimate the dilution by eye. There should be barely enough culture to see slight turbidity in the FACS tube. If desired, do multiple dilutions into FACS tubes to cover a range of cell concentrations.
21. Vortex the FACS tube(s) containing diluted culture inside the anaerobic chamber, and then tape the caps to the sides of the FACS tubes using vinyl tape to slow the entrance of oxygen.
   ▲ **CRITICAL STEP** Vortexing the tubes helps to break up multicell aggregates and resuspend settled cells, but the tubes should not be agitated outside of the anaerobic chamber.
22 Transfer the FACS tube(s) to a GasPak EZ container, and bring them out of the anaerobic chamber along with the first set of 2 mL 96-deepwell plates.

▲ CRITICAL STEP Bring the plates out of the anaerobic chamber in small batches. Optimize the size of each batch so that the plates spend as little time in the aerobic environment as possible. We were able to sort into 16 96-deepwell plates in the time it took to travel back and forth between the FACS machine and the anaerobic chamber, so our batch size was 16 plates.

23 Transport the material to the FACS machine, being careful not to agitate the plates.

24 Calibrate the cell sorter to ensure droplets are properly sorted into each well.

25 Open the GasPak EZ container, remove the tape from the first FACS tube and load it onto the sample injection port.

▲ CRITICAL STEP After unsealing the tube, agitate the sample as little as possible. Without agitation, the oxidation front will remain near the top of culture for longer, delaying the onset of oxidation for the cells entering the sample injection port at the bottom of the tube.

26 Calibrate the FACS machine to sort single cells: adjust the threshold and gain for FSC and SSC so that the population of single cells has a clear distribution of signal. Define a restrictive two-dimensional gate on FSC and SSC that selects for the average signal from single cells and excludes larger signals. Use a secondary gate on pulse trigger width to further select for single cells.

▲ CRITICAL STEP The gating strategy used to identify single cells is critical for successfully sorting the library and will depend heavily on the strain, growth medium and physiological state of the cells. Test the exact conditions used to grow the library beforehand to determine the expected signal during the library sort.

? TROUBLESHOOTING

27 Unseal a 96-deepwell plate and check for liquid on seal. If there is liquid on the seal, discard the plate.

? TROUBLESHOOTING

28 Sort single cells into the plate using the FACS machine and conditions defined in Step 26.
Reseal the plate with a Breathe-Easier gas-permeable membrane. Use a roller to ensure that a tight seal has formed, especially around the edges of the plate.

**CRITICAL STEP** Ensuring a good seal is important to prevent evaporation from the outer rows and columns of the 96-deepwell plate during the recovery period.

**CRITICAL STEP** It is important to avoid jostling or agitating the plates from this point onwards to minimize cross-contamination between the wells. If a plate is jostled while applying the seal or moving to the incubator, check for wetting of the membrane and discard the plate if medium has splashed onto the seal.

CRITICAL STEP Ensuring a good seal is important to prevent evaporation from the outer rows and columns of the 96-deepwell plate during the recovery period. It is important to avoid jostling or agitating the plates from this point onwards to minimize cross-contamination between the wells. If a plate is jostled while applying the seal or moving to the incubator, check for wetting of the membrane and discard the plate if medium has splashed onto the seal.

Return all plates containing sorted cells to the anaerobic chamber using a low-pressure cycling program, and place them in the 37 °C incubator. These sorted plates will remain in the 37 °C incubator until the cultures have grown.

**CRITICAL STEP** Cycle the airlock with a low-pressure program to avoid peeling the seal from the plates (e.g., nine cycles of 5 in Hg vacuum pressure and fill with N₂, followed by nine cycles of 5 in Hg vacuum pressure and fill with the chamber atmosphere).

Every hour, dilute cells from the culture in Step 18 into a new prewarmed FACS tube from Step 19 by following Steps 20 and 21.

Take the next batch of plates out of the 37 °C incubator for sorting, and bring them to the FACS machine. If a new FACS tube has been prepared in Step 31, bring it out of the anaerobic chamber in an EZ GasPak container and replace the previous tube at the FACS machine following Step 25.

Repeat Steps 27–32 until cells have been sorted into all 96-deepwell plates (Fig. 3).

Once all plates have been sorted, incubate them at 37 °C to allow recovery and growth of the sorted cells.

**CRITICAL STEP** The length of this incubation will depend on the strain being sorted. An initial sorting experiment would be helpful in determining how long to incubate the sorted plates so that the cultures saturate before continuing with the protocol.

**TROUBLESHOOTING**

Day 4/5: make copies of glycerol stocks of the library ● **Timing 1 d**

**CRITICAL** Making copies of glycerol stocks of the library will require a team of people to work efficiently. We have designed this protocol around an anaerobic chamber with two sets of gloves to include a team of three researchers. One person will unseal, label and reseal the plates; one person will operate the BenchSmart to mix and aliquot the glycerol stocks; one person will work outside the anaerobic chamber to heat-seal and store the glycerol stocks at −80 °C (Fig. 4).

Remove the Breathe-Easier membrane from a 96-deepwell culture plate from Step 34, and cover with a sterile tip-box lid.

**CRITICAL STEP** Take a moment to examine the growth pattern within the 96-deepwell plate. If the pattern is troubling (sparse growth), discard the plate before it is assigned a number in the following step.

Add a label to the 96-deepwell culture plate to assign a plate number.

**CRITICAL STEP** This step will be the first point at which the culture plates are identified with a unique label. From this point onwards, it is important to check that the labels match when mixing, aliquoting and storing the glycerol stock.

Transfer the culture plate to the BenchSmart, and match with the correspondingly labeled 96-microwell plates.

Load 1 mL filter tips.

Switch the BenchSmart program to ‘Dispense glycerol’ (see ‘Equipment setup’).

Transfer 557 µL per well of 50% glycerol to the culture plate, and mix using the ‘Dispense glycerol’ program to aspirate and dispense into the same wells.

**CRITICAL STEP** Track the height of the culture when resuspending to keep the tips of the pipette just below the surface. Submerging the pipette tips when the wells are almost full will push the cultures out of the wells, while allowing the tips to dispense above the culture surface will cause splattering.

Switch the BenchSmart program to ‘Repeat dispense’ (see ‘Equipment setup’).

Dispense 100 µL of culture to each well of the 96-microwell plates.

Unload tips and discard.
45 Seal each of the 96-microwell plate glycerol stocks with an AlumaSeal adhesive seal. Use a roller to ensure that a good seal is formed.

46 Place the sealed 96-microwell plates in MylarFoil heat seal bags along with a 500cc oxygen absorber, and zip-seal the bags.

47 Seal the original culture with an aluminum seal.

48 Before enough bagged plates have accumulated that they impede movement inside the anaerobic chamber, move them outside the anaerobic chamber and use the heat sealer to seal the bags.

49 Move the bagged plates to the −80 °C freezer for long-term storage.

50 When finished, or as needed to free up space, move the sealed original cultures out of the anaerobic chamber and into a 4 °C refrigerator for temporary storage until all glycerol stocks have been made.

51 Repeat Steps 36–50 until all 96-deepwell plates have been copied as cryostocks (Fig. 4).

**PAUSE POINT** The remainder of the cultures in 96-deepwell plates will be used in Steps 54–75 to pool the library and sequence. The 96-deepwell plates can be stored overnight at 4 °C or for a few days at −20 °C.

52 At the end, transfer 200 µL of each of the original culture plates to a flat-bottom 96-microwell plate and measure the OD\textsubscript{600} using a plate reader. This measurement of the final OD\textsubscript{600} of every well in the library will allow for an estimate of the fraction of wells in which growth occurred. Store the remainder of the culture at 4 °C or −20 °C.

53 Analyze the final OD\textsubscript{600} values of the ordered library from Step 52. Use a heuristic cutoff to determine the fraction of wells in which growth was detected. Examine the distribution of growth/no growth and final OD\textsubscript{600} values across plates. Plates with abnormal growth can be discarded along with the 96-microwell glycerol stocks.

**TROUBLESHOOTING**

Pool the library and extract genomic DNA ● **Timing 2 d**

▲ **CRITICAL** See Fig. 5 for pooling strategy.

54 Thaw the original 96-deepwell plates from Step 50.

55 Centrifuge each 96-deepwell plate for 5 min at 4,000g at RT.

▲ **CRITICAL STEP** Prevention of cross-contamination during pooling is essential to accurately resolve the position of barcodes within the library. Centrifugation pulls down any culture that
may be stuck to the seal and prevents droplets from spraying between wells when the seal is pulled off.

56 Allow the 96-deepwell plates to thaw at RT. Plates will take 1 h to thaw from −20 °C but can be used immediately if stored at 4 °C in Step 50.

57 Open two sterile 2 mL 96-deepwell plates, and cover each one with a sterile tip-box lid. The first 96-deepwell plate will store a mixture of each well from all plates (All-Sample plate). The second 96-deepwell plate will store the final pools (Final-Pool plate).

▲ CRITICAL STEP The following steps are designed to work with a 40-plate library so that individual pools have a volume of less than 1.5 mL. To adapt for a larger library, lower the volumes accordingly.

▲ CRITICAL STEP The following steps are designed to work with a pool design and library size that need less than 96 pools. If more than 96 pools are required, more than one Final-Pool plate will be necessary.

▲ CRITICAL STEP Mislabeling or cross-contaminating wells in the Final-Pool plate will lead to significant errors in locating barcodes. Work carefully to ensure that no pipetting errors are made when filling the plate.
Protocol

58 Load sterile 200 µL filter tips onto the BenchSmart.
59 Remove the seal from the centrifuged 96-deepwell plate, and cover with a sterile tip-box lid.
60 Set the BenchSmart-96 to the 'Mix cultures' program (see 'Equipment setup').
61 Aspirate and dispense into the same wells in the centrifuged 96-deepwell plate from Step 59 to mix the cultures.

**A CRITICAL STEP** If the cultures are prone to settling, it is important to dispense aliquots into the pools quickly after mixing to avoid pipetting errors.

62 Set the BenchSmart-96 to the 'Dispense pools' program (see 'Equipment setup').
63 Position an upside-down tip-box lid from the fresh tip box in Step 58 so that you can pipette into it using the BenchSmart.
64 Dispense 15 µL from each well of the centrifuged 96-deepwell plate from Step 61 into the upside-down tip-box lid from Step 63 using the BenchSmart, creating spots on the bottom.

**A CRITICAL STEP** Keep the tips slightly elevated from the lid to prevent backup of the dispensing liquid, which can lead to spattering of the liquid and cross contamination between tips.

65 Dispense 15 µL from each well of the centrifuged 96-deepwell plate from Step 61 into the All-Sample plate from Step 57 (Fig. 5a).
66 Collect the culture from the tip-box lid in Step 64 by holding the tip-box at a slight angle and using the tip of a 1 mL pipette to disturb the droplets so that they fall to the bottom corner of the lid. Once enough culture has accumulated, use the 1 mL pipette to aspirate from the bottom corner and then dispense across the face of the lid to mix the culture. Mix thoroughly.

67 Store the mixed culture in a well of the Final-Pool plate from Step 57 (Fig. 5b).
68 Repeat Steps 58–67 for every 96-deepwell culture plate in the library, collecting into the same All-Sample and Final-Pool plates.

**PAUSE POINT** The pooled cultures in the All-Sample and Final-Pool plates can be stored for weeks at −80 °C. If the cultures are stored frozen, plan to allow the plates to thaw for at least 1 h before executing the next step.

69 For every row in the All-Sample plate, use a 12-channel 200 µL pipette with 12 filter tips loaded. Transfer 120 µL per well to a single-well reservoir. Collect the culture, and store in a well of the Final-Pool plate (Fig. 5c).
70 For every column in the All-Sample plate, use a 12-channel 200-µL pipette with eight filter tips loaded. Transfer 180 µL per well to a single-well reservoir. Collect the culture, and store in a well of the Final-Pool plate (Fig. 5d).
71 Load 200 µL filter tips onto the BenchSmart.

**A CRITICAL STEP** Steps 71–74 are included as an option to generate DNA for performing RB-TnSeq on the ordered library in Steps 84–93. Running an RB-TnSeq experiment on the ordered library is most important when the lookup table generated with RB-TnSeq on the original pooled library ('Experimental design') is incomplete because of high complexity of the pooled insertion library. With insufficient coverage in the lookup table, a significant fraction of mapped barcodes within the ordered library will have no recorded insertion site. Because the ordered library can be orders of magnitude less complex than the original pooled library, much higher coverage can be achieved by performing RB-TnSeq on it as opposed to the original pool. If the lookup table generated when characterizing the original pool is sufficient to associate the majority of sorted barcodes with an insertion site, this step will not be necessary.

72 Set the BenchSmart-96 to the 'Combine pools' program (see 'Equipment setup').
73 Use the BenchSmart-96 to transfer 100 µL from each well of the All-Sample pool plate from Step 68 to a single-well reservoir.
74 Swirl the reservoir to mix the cultures, and then transfer the combined sample by pouring into a 15 mL conical tube. Distribute the sample by pipetting 1.5 mL aliquots from the conical tube across at least six wells of the Final-Pool plate from Step 68 (Fig. 5e).
75 Extract genomic DNA from each well of the Final-Pool plate following the manufacturer’s protocol for the QIAamp 96 DNA QIAcube HT kit (Fig. 5f). Store the genomic DNA at −20 °C.

**PAUSE POINT** The genomic DNA can be stored at −20 °C for years.

**Perform Bar-seq to identify the pool inclusion patterns**

**Timing 0.5 d**

76 Make enough of the PCR master mix (see 'Reagent setup') to amplify DNA from each well of the Final-Pool plate, and add 32.5 µL per reaction to the wells of a 96-well PCR plate.
77 Add 5 µL of the 4 µM indexed primer mixes (see 'Reagent setup') to each well.
78 Add 12.5 µL of each extracted genomic DNA from Step 75 to each well of the PCR plate.
79 Seal the PCR plate with a foil seal, and run the Bar-seq PCR program (see 'Equipment setup').
80 Run 10 µL of each PCR product on a 1% (wt/vol) agarose gel to check for the correct band (~200 bp). Run the gel at 100 V for 20 min at RT. Add SYBR Safe at a 1× concentration to the gel before casting, and visualize the band with UV or blue light.
81 Combine 5 µL of each Bar-seq product into a single pool, called a total-library pool. The remainder can be stored at −20 °C for years.
82 Purify a 50 µL aliquot of the total-library pool by following the manufacturer’s protocol for a Zymo DNA clean and concentrator kit. Elute in 20 µL of sterile deionized water.
83 Submit the purified PCR product for sequencing using a MiSeq, Reagent Kit v3, 1× 75 bp reads.

(Optional) Perform RB-TnSeq to associate barcodes with insertion sites ● Timing 1 d
84 Combine all wells of purified genomic DNA from Step 75 that correspond to the whole-library pool aliquots from Step 74. Dilute the genomic DNA so that there is 1.5–3 µg DNA per 130 µL.
85 Transfer 130 µL of the whole-library genomic DNA to a Covaris microtube without creating any bubbles. Fragment the purified genomic DNA to an average size of 300 bp using a Covaris S2 ultrasonicator using the manufacturer’s protocol.
86 Further size-select the fragmented DNA for an average size of 300 bp with a tighter distribution using AMPure XP beads and the manufacturer’s protocol.
87 Create a sequencing library from the sheared genomic DNA using a NEBNext kit and following the manufacturer’s instructions, with the following modifications: at the adapter ligation step, ligate the custom TruSeq Y*-adapters (see Supplementary Table 2 and '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.
88 Make enough of the PCR master mix for a 100 µL reaction (see 'Reagent setup' and Supplementary Table 2).
89 Add 50 µL of adapter-ligated DNA from Step 87.
90 Run the RB-TnSeq PCR program (see 'Equipment setup').
91 Run 5 µL of PCR product on a 1% (wt/vol) agarose gel to check for the correct band (200–300 bp).
92 Purify the PCR product with AMPure XP beads, following the manufacturer’s instructions.
93 Submit the purified PCR product for sequencing using a MiSeq, Reagent Kit v3, 1× 150 bp reads.

Locate the transposon insertion mutants in the library ● Timing 1 d
94 Download the GenBank file for your bacterial strain (genbank, Fig. 2) in a new directory.
95 Convert the GenBank file into genome and genes files for further analysis (Fig. 2) using the command

```
SetupOrg.pl -gbk genbankfile -out organism
```

SetupOrg.pl generates a fasta file for the genome and a tab-delimited genes file that are used as inputs in subsequent analysis.
96 Move the RB-TnSeq fastq file(s) from Step 93 into a directory named tnseq.
97 Move the forward read Bar-seq fastq files from Step 83 into a directory named barseq.
98 Create or move a model file (modelfile) into a directory named model.
99 Map reads onto the genome (Fig. 2) using the command

```
MapTnSeq.pl -genome organism/genome -model model/modelfile \ 
-first tnseq/fastq > tncount
```

100 Generate the lookup table (Fig. 2) using the command

```
python3 path_to_repository/scripts/DesignSortedLibrary.py3 \ 
--input tncount --genes organism/genes --Nmin 10 \ 
--output lookuptable
```
101 Generate a tab-delimited key file (barseqkey, Fig. 2) with the columns path to fastQ file, index name and pool name.

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

102 Generate the pool inclusion table (inclusiontable, Fig. 2) by running the command

```bash
python3 path_to_repository/scripts/SortedPoolCounts.py3 \
-k barseqkey -o inclusiontable -q1 0 -f 0.01
```

**CRITICAL STEP** Unlike other sequencing approaches that map reads onto the genome, there is no internal metric for Bar-seq that indicates whether a given barcode sequence is real or spurious. Prudently choosing the minimum number of reads to consider a barcode real is important to avoid introducing extra barcode sequences into the analysis pipeline. `SortedPoolCounts.py3` uses the `-f` parameter, the minimum fraction of reads relative to expectation, to control the number of barcode sequences included in the analysis. If you believe spurious barcodes are being included in the analysis, increase the value of the `-f` parameter when running the script.

? TROUBLESHOOTING

103 Locate transposon insertion mutants in the ordered library (Fig. 2) by running the command

```bash
matlab -nodesktop -nosplash -r \
"addpath(genpath('path_to_repository'));\ 
resolve_barcode_position('inclusiontable','lookuptable',... \ 
'organism/genes','positiontable');\ 
exit;"
```

104 Examine the positiontable file (Fig. 2); this is the final map of strain locations within the library.

? TROUBLESHOOTING

**Troubleshooting**

Troubleshooting advice can be found in Table 1.

### Table 1 | Troubleshooting table

| Step | Problem | Possible reason | Solution |
|------|---------|----------------|----------|
| 26   | Signal from FSC and SSC unclear; hard to identify individual cells | Medium not sufficiently filtered; particulate matter is acting as a signal | Always filter the medium |
|      |         | Cell morphology is aberrant | Reexamine 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 |
| 27   | Medium in plates contaminated | Original bottles not sterilized | Autoclave the 2 L bottles before storage; use sterile technique to filter medium |
|      |         | Contamination during the transfer to plates | Use sterile technique; eject tips if they physically touch anything other than medium; lower aspiration and dispensation speeds; cover the reservoir and plates except when pipetting |
| 34   | Medium on the seal | Too much agitation during transfer | Carry plates by hand |
|      | Wicking of medium onto gas-permeable membrane; leakage of medium from the plates | Too much agitation during transfer | Lower the medium volume; carry plates by hand |

Table continued
Timing

Steps 1–15, day 1: prepare for sort: 0.5 d
Steps 16–35, day 2: sort the library: 1 d
Steps 36–53, day 4/5: make copies of glycerol stocks of the library: 1 d
Steps 54–75, pool the library and extract genomic DNA: 2 d
Steps 76–83, perform Bar-seq to identify the pool inclusion patterns: 0.5 d
Steps 84–93, (Optional) perform RB-TnSeq to associate barcodes with insertion sites: 1 d
Steps 94–104, locate the transposon insertion mutants in the library: 1 d

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 3,517 barcodes within the library, of which 235 were found in more than one position and had to be resolved probabilistically. The final map of transposon insertions in the library identified 3,059 wells with a single barcode, 350 wells with more than one barcode and 431 wells without a barcode assignment (Fig. 6a, left). Less than a third of the wells without a barcode overlapped with wells in which the OD₆₀₀ was less than 0.25, indicating that low cell density is not the main cause of our inability to detect a barcode (Fig. 6a, top right). There were no more than three barcodes per well in the library (Fig. 6a, bottom right). We spot-checked 15 strains (7 single insertions and 8 multiple insertions) using PCR and confirmed that all 15 were correctly located (Supplementary Fig. 5, Supplementary Table 3).

In the transposon insertion pool used in this study, 85% of barcodes were associated with a single, unambiguous insertion site (Fig. 6b, left). The remaining barcodes were ambiguous because they were associated with multiple insertion sites, which could have arisen either from multiple transposon insertions in the same cell or from separate strains that serendipitously received the same barcode.
Given that it is extremely unlikely that distinct strains with the same barcode would be sorted into the same well, transposon insertions that share the same barcode and the same single location in the library are likely to be from multiple insertions in the same strain. By performing RB-TnSeq on the ordered library, we identified multiple instances of distinct transposon insertions that both shared the same barcode and were found only once in the library at the same location. By classifying these instances as a single strain with multiple insertions, we were able to reduce the fraction of ambiguous barcodes in the ordered library to 4% (Fig. 6b, right).

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 (Fig. 6b). It is possible that decreasing the diversity of the library by more than one order of magnitude allows for greater power in detecting additional insertion sites. Thus, performing RB-TnSeq on the ordered library improved the detection of strains with multiple insertions per genome in two ways: it identified new insertion sites because of the greater read depth, and it made it clear that many transposon insertions that share the same barcode do in fact occur in the same strain.

**Reporting Summary**
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**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 (Supplementary Figs. 2–4) are available from the corresponding author upon request.

**Code availability**
The code required for locating insertion mutants in ordered libraries is available on BitBucket (https://bitbucket.org/kchuanglab/resolveBarcode_position/src). These scripts rely on previously published code available on BitBucket (https://bitbucket.org/berkeleylab/feba/src). The code in this protocol has been peer reviewed.
References

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. *Nat. Microbiol.* **3**, 514–522 (2018).

3. Liberati, N. T. et al. An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc. Natl Acad. Sci.* **103**, 2833–2838 (2006).

4. De Berardinis, V. et al. A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* **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. *Nat. Commun.* **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. *Nat. Commun.* **7**, 13270 (2016).

9. Koo, B.-M. et al. Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. *Cell Syst.* **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 (2008).

12. Lampe, D. J., Akery, B. J., Rubin, E. J., Mekalanos, J. J. & Robertson, H. M. Hyperactive transposase mutants of the Himar1 mariner transposon. *Proc. Natl Acad. Sci.* **96**, 11428–11433 (1999).

13. Goryshin, I. Y. & Reznikoff, W. S. Tn5-coded transposons. *mBio* **6**, 1253 (2015).

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

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

16. Erlich, Y. et al. DNA Sudoku—harnessing high-throughput sequencing for multiplexed specimen analysis. *Genome Res.* **19**, 1243–1253 (2009).

17. Wetmore, K. M. et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* **6**, e00306–e00315 (2015).

18. Price, M. N. et al. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* **557**, 503 (2018).

19. Hughes, S. R. et al. High-throughput screening of cellulase F mutants from multiplexed plasmid sets using an automated plate assay on a functional proteomic robotic workcell. *Proteome Sci.* **4**, 10 (2006).

20. Liu, H. et al. Functional genetics of human gut commensal *Bacteroides thetaiotaomicron* reveals metabolic requirements for growth across environments. *Cell Rep.* **34**, 108789 (2021).

21. Shatalin, K., Shatalina, E., Mironov, A. & Nudler, E. H2S: a universal defense against antibiotics in bacteria. *Science* **334**, 986–990 (2011).

22. Holt, J.G., Kreig, N. R., Sneath, P. H. A., Staley J. T. & Williams S. T. *Bergey’s Manual of Determinative Bacteriology* 9th edn (Williams & Willkins, 1994).

Acknowledgements

The authors thank the Huang lab for helpful discussions. This work was supported in part by the Allen Discovery Center at Stanford University on Systems Modeling of Infection (to K.C.H.) and NIH grants GM135102 (to A.M.D. and K.C.H.) and ROI AI147023 (to K.C.H.). K.C.H. is a Chan Zuckerberg Biohub Investigator. The authors also acknowledge the hospitality of the Aspen Center for Physics, which is supported by NSF grant PHY-1607611. The following reagents were obtained through BEI Resources, NIAID, NIH as part of the Human Microbiome Project: *Lactobacillus reuteri*, strain CF48-3A, HM-102; *Clostridium innocuum*, strain 6_1_30, HM-173; *Clostridium symbiosum*, strain WAL-14163, HM-309; *Bacteroides fragilis*, strain CL109T03C10, HM-727; *Parabacteroides johnsonii*, strain CL082T12C29, HM-731.

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.

Competing interests

The authors declare no competing interests.
Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-021-00531-3.
Correspondence and requests for materials should be addressed to K.C.H.
Peer review information Nature Protocols thanks Buz Barstow and Matthew Waldor for their contribution to the peer review of this work.
Reprints and permissions information is available at www.nature.com/reprints.
Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 February 2019; Accepted: 16 February 2021;
Published online: 21 May 2021

Related links
Key references using this protocol
Price, M. N. et al. Nature 557, 503 (2018): https://doi.org/10.1038/s41586-018-0124-0
Liu, H. et al. Cell Rep. 34, 108789 (2021): https://doi.org/10.1016/j.celrep.2021.108789
**Reporting Summary**

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| Item                                                                 | Confirmed |
|----------------------------------------------------------------------|-----------|
| The exact sample size \( (n) \) for each experimental group/condition, given as a discrete number and unit of measurement | n/a       |
| A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly | n/a       |
| The statistical test(s) used AND whether they are one- or two-sided  | n/a       |
| Only common tests should be described solely by name; describe more complex techniques in the Methods section. | n/a       |
| A description of all covariates tested                              | n/a       |
| A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons | n/a       |
| A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) | n/a       |
| For null hypothesis testing, the test statistic (e.g. \( F, t, r \)) with confidence intervals, effect sizes, degrees of freedom and \( P \) value noted | n/a       |
| Give \( P \) values as exact values whenever suitable.              | n/a       |
| For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings | n/a       |
| For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes | n/a       |
| Estimates of effect sizes (e.g. Cohen’s \( d \), Pearson’s \( r \)), indicating how they were calculated | n/a       |

Our web collection on [statistics for biologists](https://www.nature.com/nresearch/research-reporting) contains articles on many of the points above.

### Software and code

**Policy information about availability of computer code**

**Data collection**

Data from plate reader experiments was collected with Gen5 v. 3.04. Microscopy images were collected with MicroManager v. 1.4. Sequencing reads were collected with an Illumina HiSeq 4000 (BioSample SAMN12807646) or an Illumina MiSeq (SAMN12809978).

**Data analysis**

Data from plate reader experiments were analyzed using custom code in MATLAB 2018b. Microscopy images were processed with FIJI v. 2.0.0-rc-69/1.52i. Data from barcode sequencing experiments were analyzed with custom perl scripts (https://bitbucket.org/berkeleylab/feb/ea/src) and by custom code in Python3 and MATLAB 2018b (https://bitbucket.org/kchuanglab/resolve_barcode_position/src).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

### Data

**Policy information about availability of data**

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The sequencing data used in this study are publicly available on NCBI as part of the 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.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | Sample sizes for plate reader experiments were chosen based on the total number of samples, along with relevant controls, that could be fit within a 96-well plate format. |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data exclusions | The differential plot of individual growth curves was fit with a model function to estimate the maximum growth rate. The fit was examined for every sample, and if the fit was clearly poor the sample was excluded from analysis. Because the lag time calculation depends on the fit of maximum growth rate, if the maximum growth rate measurement for a sample was excluded, so was the lag time measurement. 3 samples were excluded from the max growth rate and lag time parameters plotted in Figure 6D. |
| Replication | Independent measurements were taken for every measurement with sample size dictated by the parameters of the experiment. |
| Randomization | Where relevant, data collection was conducted on the same day to prevent batch effects. No further randomization was conducted. |
| Blinding | Investigators were not blinded during data collection or analysis. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a  Involved in the study
☒  Antibodies
☒  Eukaryotic cell lines
☒  Palaeontology
☒  Animals and other organisms
☒  Human research participants
☒  Clinical data

Methods

n/a  Involved in the study
☒  ChIP-seq
☒  Flow cytometry
☒  MRI-based neuroimaging