Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries

Andrew L Goodman[1,2], Meng Wu[1] & Jeffrey I Gordon[1]

[1]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri, USA. [2]Present address: Section of Microbial Pathogenesis and Microbial Diversity Institute, Yale School of Medicine, New Haven, Connecticut, USA (A.L.G.). Correspondence should be addressed to A.L.G. (andrew.goodman@yale.edu) or J.I.G. (jgordon@wustl.edu).

Published online 17 November 2011; doi:10.1038/nprot.2011.417

Insertion sequencing (INSeq) is a method for determining the insertion site and relative abundance of large numbers of transposon mutants in a mixed population of isogenic mutants of a sequenced microbial species. INSeq is based on a modified mariner transposon containing MmeI sites at its ends, allowing cleavage at chromosomal sites 16–17 bp from the inserted transposon. Genomic regions adjacent to the transposons are amplified by linear PCR with a biotinylated primer. Products are bound to magnetic beads, digested with MmeI and barcoded with sample-specific linkers appended to each restriction fragment. After limited PCR amplification, fragments are sequenced using a high-throughput instrument. The sequence of each read can be used to map the location of a transposon in the genome. Read count measures the relative abundance of that mutant in the population. Solid-phase library preparation makes this protocol rapid (18 h), easy to scale up, amenable to automation and useful for a variety of samples. A protocol for characterizing libraries of transposon mutant strains clonally arrayed in a multiwell format is provided.

INTRODUCTION

Development of the protocol

By exposing populations of transposon mutants to selective conditions *in vitro* or *in vivo*, it is possible to identify changes in the representation of mutants, thereby allowing identification of genes and pathways that are key to fitness under the conditions being examined. In principle, this approach could be applied to any microbe whose genome has been sequenced and that is amenable to basic genetic manipulation (insertion of foreign DNA, antibiotic selection). INSeq combines existing methods for signature-tagged mutagenesis[1] with new techniques for high-throughput DNA sequencing such that the precise location and relative abundance of tens of thousands of mutants can be determined in parallel. This approach is based on the modification of mariner, the broad host-range, randomly integrating, transposable element, to allow short fragments of genomic DNA adjacent to the site of transposon insertion to be captured, sequenced and quantified[2]. Mutants that decrease in relative abundance in a selective condition are likely to be in genes that are important for fitness under that condition; mutants that increase in abundance under the selection highlight genes that could be deleterious. The protocol described here includes substantial improvements from our earlier publication[2]; because library preparation can now be performed in 96-well, bead-based format, reagent costs and sample losses are minimized, and large numbers of samples can be barcoded and sequenced in a single sequencing run.

Applications of the method

As INSeq library preparation can be completed in multiplex format, experiments can involve material obtained from high-resolution time-series studies of the effect of *in vitro* or *in vivo* selections, and from multiple biological and technical replicates. Any microbe whose genome has been sequenced and that is amenable to mariner transposon mutagenesis could potentially be investigated with INSeq. This method can also be used to efficiently determine the insertion site of transposon mutant strains that have been clonally arrayed in a multiwell format (**Box 1**). This latter technique relies on a combinatorial pooling approach: a small number of pools are defined, and a liquid-handling robot is used to place each mutant strain into a subset of these pools in a unique pattern. The pools are then characterized by INSeq, and the presence/absence pattern of a given insertion site across the pools reflects the unique pattern associated with the corresponding bacterial strain in the multiwell collection. Alternative pooling patterns that use larger numbers of pools to further reduce the likelihood of mistaking one strain for another can also be considered[3–5]. These alternatives would be appropriate if many mutants in the population were likely to carry insertions at the same location in the genome, but they require greater effort in preparing libraries because of the increased number of pools. Furthermore, these alternate techniques were not developed specifically for mapping transposon insertions and may require modification in order to accomplish this goal.

Limitations and alternative techniques

In bacterial genomes, a high percentage (>90%) of insertions can be precisely localized using INSeq. With more complex organisms, the short (16–17 bp) fragments of adjacent chromosomal sequence captured and quantified by MmeI digestion may be insufficient to uniquely identify transposon locations. As with any negative selection technique, stochastic bottlenecks in the selective process need to be considered when determining the optimal number of mutants to be included in an experiment. Because each insertion is mapped with nucleotide-level resolution, comparing the behavior of multiple mutants in the same gene highlights these selective bottlenecks.

Other techniques for mapping and quantifying transposon insertions by high-throughput sequencing are also available[6–10]. They include two protocols named Tn-Seq: one version[8,10] also uses an

Box 1 | Combinatorial pooling and mapping of clonally arrayed transposon mutant libraries by INSeq ● TIMING ~1.6 h per 96-well tray

Timing estimate assumes the use of an EpMotion liquid-handling robot. Additional time is required for preparing the arrayed mutant collection by manually or robotically picking transposon-mutagenized colonies into 96-well trays and conducting INSeq library preparation on the pooled strains.

1. Prepare EpMotion programs. Two sets of pooling patterns are provided in the analysis package INSeq_analysis/Arrayed_library directory (**Supplementary Data 1**): '16384_strings.txt' is the list of 24-bit strings with a minimum Hamming distance of 6; and '13000_strings.txt' is the subset of these that are less likely to be mistaken for each other if multiple archived strains happen to carry transposons at the identical genome coordinate. The larger set can be used to map up to 170 96-well trays, whereas the smaller set is preferable for arrayed libraries of under 135 trays. A script called `write_dws.pl` translates these pooling patterns into programs for the EpMotion robot. Each program will pool 5 trays (20 programs are needed for a 100-tray mutant library, for example).
To create EpMotion .dws files, go to the data analysis package INSeq_analysis/Arrayed_library directory (**Supplementary Data 1**) and type:

```
perl write_dws.pl <pooling pattern> <number of programs>
```

This will create an output directory called 'Write_dws_output_X_programs'. Create a new application on the EpMotion, and then import the .dws files in this directory into the new application and confirm that they are functioning properly. (Note that the provided script is designed for Midwest Scientific 96-well culture trays (cat. no. TP92696) and the EpMotion Thermorack for 24 cryotubes (cat. no. 960002491). Alternatives may require changes to the program). The `write_dws.pl` script will also create a map file that will be used in the data analysis step. Although the software provided is designed for the EpMotion instrument, analogous programs could be written for other liquid-handling robots if they can be run from a text-format file.

2. Prepare the mutant collection by manually or robotically picking transposon-mutagenized colonies into 96-well trays containing 250 µl of culture medium and appropriate antibiotics. Incubate under standard conditions to allow cultures to grow to turbidity, and note any wells with poor or no growth.

3. For each 96-well culture tray, prepare two 96-well archive trays by adding 30 µl of sterile culture medium plus 40% (vol/vol) glycerol to each well. Transfer 30 µl of the turbid culture to the corresponding well in each of these archive trays, mix well and seal with foil lid before storing archive trays at −80 °C. Turbid cultures may need to be mixed before dispensing into archive trays.
To recover strains from the archive trays, wipe down the foil lid with 70% (vol/vol) ethanol and puncture the well of interest with a sterile pipette tip. Draw a small amount of frozen material onto the tip of the pipette without allowing the entire plate to thaw. Restreak this material onto agar plates, and place a small foil patch over the targeted well in the archive tray before returning it to the −80 °C freezer.

4. Seal culture trays in groups of five with Parafilm and store at 4 °C.

5. On the EpMotion, initiate the Pool_group_1.dws program. Set up the robot deck as shown in the graphical interface with tips, input culture trays nos. 1–5 and a cryovial rack containing 24 2-ml screw-cap cryovials (lids removed and vials labeled with program and pool number). In the cryovial rack, tubes in the top row should be labeled 1–6, the tubes in the second row 7–12, etc. While the program is running, monitor the waste tip bucket to ensure that it does not overflow. When the program is complete, cap the 2-ml output cryovials and store them at −20 °C.

▲ **CRITICAL STEP** Each program takes ~8 h to complete. Consider adding a growth inhibitor if your mutants are likely to grow during this process. If the robot is performing without interventions required, programs can be run overnight. Also, if input trays have been stored at 4 °C, cultures may need to be resuspended manually before being placed in the robot. See **Supplementary Movie 2** for a demonstration of the pooling procedure.

6. Run pooling programs on each subsequent set of five culture trays as described in Step 5.

7. When pooling is complete, thaw all cryovials, vortex well and combine half of the volume of each cryovial with same number from each pooling program to create 24 final pools. Regardless of the number of programs run, there should be 24 pools at this point. Save the remainder of each cryovial at −20 °C.

▲ **CRITICAL STEP** Be sure to combine like pools and not to combine different pools from a single program.

8. Prepare INSeq libraries from each of the 24 pools (Steps 1–78), associating a different barcode with each pool.

9. Run the data analysis pipeline as described in the arrayed library section of the README.txt included in the analysis package.

▲ **CRITICAL STEP** The cutoffs for presence or absence of a strain in a pool may need to be determined empirically by examining the distribution of reads across pools for a given data set. As a starting point, normalize counts per pool for each strain by the overall number of reads in the pool; for each strain, set the presence or absence of a pool to 0 if there are less than three reads associated with that pool and to 1 if there are three or more reads associated with the pool.

MmeI-adapted mariner transposon but does not include a transposon-specific, linear PCR step during library preparation. Compared with INSeq, this simpler approach to library preparation could reduce the effort required for sequencing. The other Tn-Seq protocol[9] does not use an MmeI-adapted mariner transposon; instead, it uses a circularization step to capture the targeted DNA fragments. This could allow the use of transposons that cannot be modified to include MmeI sites in their inverted repeats. Additional approaches, termed 'high-throughput insertion tracking by deep sequencing (HITS)'[6] or 'transposon-directed insertion-site sequencing (TraDIS)'[7] also do not require MmeI sites in the transposon. For these protocols, mutagenized DNA is randomly sheared, adapters are ligated to all fragments, and transposon-chromosomal junctions are enriched by PCR either with (HITS) or without (TraDIS)

an affinity-purification step. In principle, these MmeI-independent methods could allow the capture of longer fragments of genomic sequence adjacent to each transposon, a feature which may be necessary for mapping insertions in more complex genomes.

The INSeq protocol we describe below has several advantages compared with these other methods: (i) introducing an MmeI site in the transposon allows the excised transposon/chromosomal junctions to be of uniform size, thus avoiding size preferences during PCR; (ii) the linear PCR step at the beginning of the protocol enriches the target and increases the efficiency of library preparation; (iii) the solid-phase–based technique separates target sequences from chromosomal DNA background, making it easy to scale up to a 96-well plate format; (iv) robotic automation makes the protocol amenable to monitoring changes in the representation of mutants during studies that require repeated, high-resolution sampling; and (v) descriptions for modifying the INSeq protocol to map arrayed mutant libraries are provided in **Box 1**. If other transposons that cannot be modified with MmeI sites are used, alternative protocols such as the Tn-seq method described by Gallagher *et al.*[9], HITS[6] or TraDIS[7] are preferable.

Experimental design

Overview. An overview of the INSeq protocol is provided in **Figure 1**. Steps 1–16 describe the procedure used for extraction of DNA. In Steps 17–20, DNA is further purified using columns. Steps 21–23 describe a linear PCR procedure that is designed to enrich the transposon/chromosomal junction regions by using a primer complementary to the transposon-specific sequence. Steps 21–23 are necessary for two reasons. First, they add a biotin tag to each target molecule, which facilitates high-throughput, low-volume solid-phase library preparation. Second, the linear PCR enriches for the desired target molecules early in the procedure, minimizing the amount and cost of enzymes needed for subsequent reaction steps. In Steps 24–36, the linear PCR product is bound to magnetic beads, allowing the PCR product to be separated from 'background' chromosomal DNA by affinity capture. Second strand synthesis (and creation of the double-stranded MmeI site) takes place in Steps 37–43; because MmeI requires two sites for efficient DNA cleavage[11], a second MmeI recognition site is provided *in trans* via the double-stranded DNA (dsDNA) fragment M12. Steps 44–52 use MmeI to cut the dsDNA fragment to a uniform length (64–65 bp). The product of this reaction consists of the amplified transposon region generated by the linear PCR, plus 16–17 bp of chromosomal DNA. In Steps 53–63, sequencing adapters are ligated to the transposon/chromosomal junction to introduce the complementary sequence for the sequencing primer used by the Illumina Genome Analyzer/HiSeq instrument. By varying the four-nucleotide barcode sequences in the adapter, multiple samples can be sequenced in a single lane of an Illumina flow cell. With limited cycles of PCR (Steps 64–78), targets are enriched in the linear range. PCR products are then purified by gel extraction to remove extra primers and adapters, and they are then sequenced. Data analysis is described in Steps 79–86; detailed documentation of data analysis tools is available in the README.txt in the INSeq_analysis package (**Supplementary Data 1**). **Box 1** describes an additional protocol for combinatorial pooling and mapping of clonally arrayed transposon mutant libraries by INSeq.

Starting material. Steps 1–16 describe the procedure used for extraction of DNA from the transposon library (the 'input mutant population') and from biological specimens containing mutants that survive a given selection (the 'output mutant population'). These steps for DNA isolation were originally developed for mouse feces and for samples obtained from the more-proximal regions of the mouse gastrointestinal tract such as the cecum, and can also be used for *in vitro* mutant populations. Transposon mutant libraries can be introduced into a wide variety of biological systems (e.g., various animal or environmental habitats) or characterized in many different *in vitro* formats (e.g., chemostats, tissue culture). In these cases, users should isolate crude DNA using protocols designed for the specific system chosen for study. Whatever the system, experiments need to be designed so that the amount of input and output material is sufficient for library preparation. For example, if the mutagenized species constitutes 100% of a microbial community, then the initial purification steps as described in the PROCEDURE ideally should yield >500 ng of DNA. If the mutagenized species represents only 50% of the population, then the amount of starting material should be doubled.

Oligonucleotide sequence design. The linear PCR primer BioSamA is 5′-biotinylated and encodes a TEG spacer to reduce steric hindrance during the solid-phase enzymatic steps. The primer also encodes a transposon-specific region with a 26-bp sequence complementary to sequences found on both sides of the transposon (underlined in **Supplementary Table 1**). Adapters encode 4-bp barcodes (marked in bold in **Supplementary Table 1**) that allow

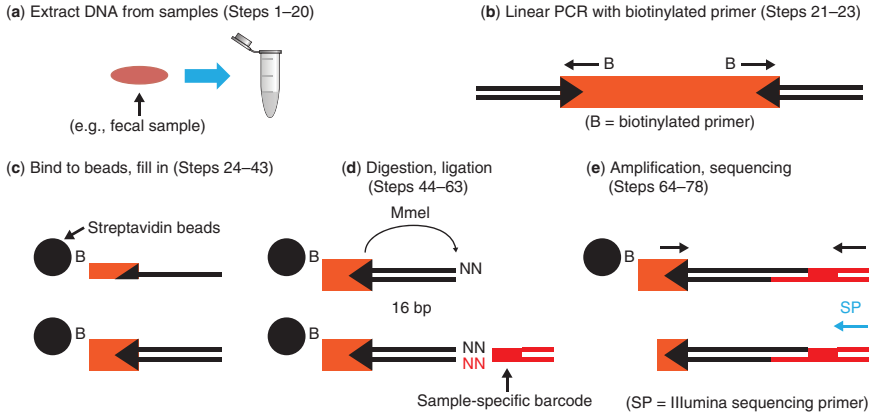

Figure 1 | Overview of insertion sequencing protocol. (**a**) Genomic DNA is extracted and purified from a sample containing the transposon mutant library (e.g., input transposon mutant libraries or output samples, such as fecal samples obtained from gnotobiotic mice after colonization). (**b**) Linear PCR is performed to amplify the transposon with adjacent chromosomal DNA. (**c**) The PCR product is purified by binding to solid-phase beads; second strand DNA is synthesized. (**d**) DNA fragments are digested by MmeI and then ligated to sequencing adapters with a sample-specific barcode. (**e**) Products from stage **d** are amplified in a limited PCR step and sequenced. The inverted repeat region is shown as a black triangle. For convenience, stages **c**–**e** illustrate only one side of the transposon; in reality both sides are amplified and sequenced.

sample multiplexing and produce a 2-bp NN overhang (complementary to the overhang produced by MmeI digestion) for ligation. For each adapter, two complementary oligonucleotide sequences are provided in **Supplementary Table 1**: LIB_AdaptT_x and LIB_AdaptB_x, where x is the unique barcode specified by the adapter). These pairs of oligonucleotide sequences are annealed to each other during the PROCEDURE. The LIB-PCR primers encode the necessary sequences for Illumina library preparation and sequencing.

Replicates and controls. Biological and technical replicates should be performed. The number of biological replicates required depends on the experimental system under study and should be determined empirically. Technical replicates can include identical samples associated with different sample-specific barcodes. The number of technical replicates should be determined by the user; two per biological sample is likely to be sufficient. No-DNA negative controls will help identify cross-contamination between samples; these controls should be run in parallel with DNA-containing samples. Cross-contamination can occur if beads transfer from tube lids to the users' gloves to other samples: this can be avoided with careful technique.

MATERIALS

REAGENTS

- Phenol:chloroform:isoamyl alcohol (25:24:1), pH 7.9 (Applied Biosystems, cat. no. AM9712) ! **CAUTION** This mixture is hazardous; handle properly during experiments.
- Sodium acetate (3 M), pH 5.5 (Applied Biosystems, cat. no. AM9740)
- Isopropanol (100%, Fisher Scientific, cat. no. A416-4) ! **CAUTION** Isopropanol is flammable; handle properly during experiments.
- Ethanol (100%, Pharmco-AAPER, cat. no. 111ACS200) ! **CAUTION** Ethanol is flammable; handle properly during experiments.
- TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7; Ambion, cat. no. AM9861)
- NaCl (NaCl powder, Fisher Scientific, cat. no. S271-10)
- Tris base (Trizma base, Sigma-Aldrich, cat. no. T1503-1KG)
- EDTA (EDTA powder, Sigma-Aldrich, cat. no. s657-500)
- RNase A (Qiagen, cat. no. 19101)
- QIAquick PCR purification kit (Qiagen, cat. no. 28106)
- Platinum Pfx DNA polymerase (Invitrogen, cat. no. 11708-021)
- Deoxyribonucleotide triphosphates (dNTPs; 10 mM, Invitrogen, cat. no. 18427-013)
- Dynabeads M-280 streptavidin (Invitrogen, cat. no. 112-05D)
- Hexanucleotide mix (Roche, cat. no. 11277081001)
- Klenow fragment (3′–5′ exo-; 5 U μl^{-1}; New England Biolabs, cat. no. M0212S)
- MmeI (New England Biolabs, cat. no. R0637S)
- T4 DNA ligase (2,000,000 U ml^{-1}; New England Biolabs, cat. no. M0202T)
- Agarose (Roche, cat. no. 11388991-001)
- Electrophoresis buffers (TBE, borate buffer[12] or equivalent)
- NaOH (250 mM, Fisher Scientific, cat. no. S318 3)
- Boric acid (Sigma-Aldrich, cat. no. B7901-500)
- DNA ladder (10 bp, Invitrogen, cat. no. 10821-015)
- GelGreen nucleic acid stain (10,000× stock prepared in water; Biotium, cat. no. 41005)
- Ficoll 400 (20% (wt/vol), Sigma-Aldrich, cat. no. F4375)
- Disodium EDTA (0.1 M, Sigma-Aldrich, cat. no. BP120-500)
- Xylene cyanol (0.25% (wt/vol), Sigma-Aldrich, cat. no. X4126-10G)
- Bromophenol blue; (0.25% (wt/vol), Sigma-Aldrich, cat. no. 114391-5G)
- QIAquick gel extraction kit (Qiagen, cat. no. 28706)
- Primer sequences are provided in **Supplementary Table 1** (Integrated DNA Technologies)
- Magnesium chloride ($MgCl_2$)

For the arrayed library mapping protocol given in Box 1 only

- Glycerol (J.T. Baker, cat. no. 2136-01)

EQUIPMENT

- Zirconium beads (0.1 mm diameter, BioSpec Products, cat. no. 11079101z)
- Screw-top cryovials (2 ml, Axygen Products, cat. no. SCT-200-SS-C-S)
- Phase Lock gel, light, 2 ml (5 Prime, cat. no. 2302820)
- BeadBeater homogenizer (BioSpec)
- Refrigerated microcentrifuge
- Magnetic particle concentrator (MPC) for 1.5-ml tubes (Invitrogen, cat. no. 123-21D)
- MPC for PCR tubes
- Electronic 200-µl eight-channel pipettor with variable speed control
- Heated oven or water bath set to 50 °C
- Vacuum evaporator
- Thermocycler capable of a ramp rate of 0.1 °C s^{-1}
- Electrophoresis equipment for agarose gels
- Non-UV gel illuminator
- Invitrogen QuBit DNA spectrophotometer (or similar, for DNA concentrations ≤ 1 ng μl^{-1})
- DNA sequencing facility with Illumina Genome Analyzer/HiSeq instrument

For the arrayed library mapping protocol given in Box 1 only

- EpMotion 5075PC liquid-handling robot with TS50 single-channel dispensing tool and rack for 24 cryotubes ▲ **CRITICAL** Eppendorf, cat. no. 960002491 was used, but temperature control is not necessary. Note that with the EpMotion, each set of five plates in the arrayed library takes ~8 h to pool. Higher-throughput liquid-handling robots are likely to be faster.
- EpMotion tips (50-µl tips, nonfiltered, 1 × 96 tips per culture tray (Eppendorf, cat. no. 960050200)
- Culture trays, 96 well (Midwest Scientific, cat. no. TP92696)
- Parafilm (Pechiney)

Data analysis resources

- fasta format sequence file and .ptt format annotation file(s) for the genome of interest
- Bowtie short read aligner (Release 0.12.7 (ref. 13))
- INSeq data analysis package (INSeq_analysis; **Supplementary Data 1**)
- INSeq sample data package (INSeq_demo, downloadable from http://gordonlab.wustl.edu/SuppData.html) ▲ **CRITICAL** This package can be run on a computational cluster or a desktop/laptop computer under a UNIX system with 4 GB memory, perl and Bowtie (http://bowtie-bio.sourceforge.net/index.shtml) installed. Basic understanding of Unix and perl are required for data analysis.
- Tab-delimited operon file in the following format (optional; where $P_{A,B}$ is a value between 0 and 1 (inclusive) reflecting the probability that gene A and gene B belong to the same operon[14])

Gene A	$P_{A,B}$
Gene B	$P_{B,C}$
Gene C	$P_{C,D}$

REAGENT SETUP

Buffer A (2×) Buffer A is prepared using 200 mM NaCl, 200 mM Tris, and 20 mM EDTA; pH is adjusted to 8 with concentrated HCl. ▲ **CRITICAL** This and all other buffers can be stored for months at room temperature (20–25 °C).

Bind and wash (B&W) buffer (2×) B&W buffer is prepared using 2 M NaCl, 10 mM Tris, and 1 mM EDTA; pH is adjusted to 7.5 with concentrated HCl.

B&W buffer (1×) Dilute 2× B&W buffer to 1:1 with dH_2O.

LoTE buffer Combine 3 mM Tris and 0.2 mM EDTA; adjust pH to 7.5 with concentrated HCl.

Borate buffer Borate buffer is prepared using 250 mM NaOH in dH_2O; pH is adjusted to 8–8.5 with boric acid.

Xylene cyanol DNA dye (10×) Dye is prepared using 20% (wt/vol) Ficoll 400, 0.1 M disodium EDTA and 0.25% (wt/vol) xylene cyanol.

Bromophenol blue DNA dye (10×) Dye is prepared using 20% (wt/vol) Ficoll 400, 0.1 M disodium EDTA and 0.25% (wt/vol) bromophenol blue.

PROCEDURE

Isolation of crude DNA ● TIMING 4 h

1| Add 250 µl of zirconium beads to the sample (e.g., ~10^9 colony-forming units of an *in vitro* mutant population or <600 mg of cecal/fecal contents from mice colonized with a mutant population) in a 2-ml screw-top vial (cryovial). These beads help with cell disruption. Note that this procedure describes the processing of a single sample, but multiple samples can readily be processed in parallel.

2| Add 500 µl of 2× buffer A, 210 µl of 20% (wt/vol) SDS and 500 µl of phenol:chloroform:isoamyl alcohol (25:24:1). Chill the sample on ice.

3| Use the BeadBeater on the 'homogenize' setting for 2 min. Rest the sample on ice for 2 min and repeat bead-beating for 2 min.

4| Centrifuge the tubes in a refrigerated microcentrifuge (4 °C, 6,800*g*, 3 min).

5| Transfer the aqueous phase (~600 µl) to a Phase Lock gel tube (prespun as described in the manufacturer's instructions).

6| Add an equal amount (600 µl) of phenol:chloroform:isoamyl alcohol to the Phase Lock gel tube and mix by inversion.

7| Centrifuge the tube in a microcentrifuge (room temperature, 18,000*g*, 5 min).

8| Transfer the aqueous phase to a microcentifuge tube and discard the organic phase.

9| Add 600 µl of cold (−20 °C) 100% isopropanol to the microcentrifuge tube.

10| Add 60 µl (one-tenth volume) of 3 M sodium acetate (pH 5.5), and mix thoroughly by vortexing.

11| Incubate at −20 °C for at least 1 h.
■ **PAUSE POINT** The sample can be stored overnight or longer at −20 °C at this point.

12| Centrifuge the tube in a refrigerated microcentrifuge (4 °C, 18,000*g*, 20 min). Carefully decant and discard the supernatant.

13| Wash the pellet with 500 µl of 100% ethanol and centrifuge it in a refrigerated microcentrifuge (4 °C, 18,000*g*, 3 min). Carefully decant and discard the supernatant.

14| Remove any excess ethanol by gently tapping the tube upside-down on a laboratory tissue.

15| Evaporate any remaining supernatant in a vacuum evaporator (no heat, check at 5-min intervals until dry).

16| Resuspend the pellet in 200 µl of TE buffer (pH 7). Incubate in an oven or water bath at 50 °C for 30 min, vortexing every ~10 min. The pellet will dissolve much faster once it has been dislodged from the wall of the microcentrifuge tube.

RNase treatment and cleanup of crude DNA using QIAquick columns ● TIMING 1 h

17| Transfer 100 µl of the crude DNA sample to a new tube. The remainder should be stored at −20 °C or −80 °C.

18| Add 0.5 µl of RNase A and incubate the tube at room temperature for 2 min. RNase treatment ensures accurate DNA quantitation in subsequent steps.

19| Clean up the sample using QIAquick PCR purification columns according to the manufacturer's instructions. Make sure that all the ethanol-containing wash buffer is removed from each column before elution by pipetting around the inner rim with a 10-µl tip, if necessary. Elute each sample in 52 µl of buffer EB (part of QIAquick kit).

20| Measure the DNA concentration of each sample using a spectrophotometer (either a UV-based or dye-based system is acceptable).
■ **PAUSE POINT** The sample can be stored for months at −20 °C or −80 °C at this point.

Linear PCR ● **TIMING 2 h**

21| Assemble the linear PCR reactions on ice.

Component	Volume per sample
dH_2O	to 100 µl
Pfx buffer	10 µl
10 mM dNTPs	2 µl
50 mM $MgCl_2$	2 µl
BioSamA (1 pmol $µl^{-1}$)	5 µl
Clean DNA (from Step 20)	0.5–2 µg
Pfx polymerase	1 µl

22| Split the reaction into 2 × 50 µl in PCR tubes and run them on a thermocycler as follows: 94 °C for 2 min, followed by 50 cycles of 94 °C for 15 s and 68 °C for 1 min.

23| Pool the tubes containing the same DNA sample, run them over a QIAquick PCR cleanup column according to the manufacturer's instructions, and elute them in 50 µl of buffer EB.

Bind linear PCR products to beads ● **TIMING 1 h**

24| Resuspend streptavidin-coated beads by shaking.

25| Add beads (32 µl per sample) to a new microcentrifuge tube (1 ml maximum; use multiple tubes if a larger volume is required).

26| Place the tube on the MPC for 1–2 min.

27| Carefully remove the supernatant with a pipette.

28| Remove the tube from MPC and add 1,000 µl of 1× B&W buffer; gently resuspend by pipetting.

29| Repeat Steps 26–28 twice for a total of three washes.

30| Remove the final wash and add 2× B&W buffer (52 µl per sample). Aliquot into PCR strip tubes (one tube per sample, 50 µl per tube).

31| Add the entire volume of one sample from Step 23 to the tube.

32| Incubate at room temperature with gentle mixing for 30 min.

33| Place the tube on the MPC for 2 min.

34| Carefully remove the supernatant with a pipette.
▲ **CRITICAL STEP** To avoid disturbing the beads, set the electronic multichannel pipettor to its slowest setting, place the end of the tip against the opposite side of the tube from the beads, and slowly move the tip downward as the supernatant is removed. See **Supplementary Movie 1** for a demonstration.
? TROUBLESHOOTING

35| Remove the tube from the MPC and add 100 µl of 1× B&W buffer; gently resuspend by pipetting.

36| Repeat Steps 33–35 twice, but resuspend beads in 100 µl of LoTE buffer each time.
■ **PAUSE POINT** The sample can be stored at 4 °C overnight at this point.

Second strand synthesis ● **TIMING 1 h**

37| Denature the sample by heating in a thermocycler: 95 °C for 2 min, then chill quickly to 4 °C.

38| Prepare second strand mix on ice.

Component	Volume per sample (µl)
dH_2O	16
10× hexanucleotide mix	2
10 mM dNTPs	1
Klenow (exo-)	1

39| Collect the beads with the MPC, carefully discard the supernatant, remove the tube from the MPC, and then gently resuspend the sample in 20 µl of second strand mix.

40| Incubate the samples in a thermocycler at 37 °C for 30 min. Mix by gently tapping the tube every 10–15 min.

41| Add 100 µl of LoTE buffer to the sample, collect the beads in the MPC, and then carefully discard the supernatant.

42| Repeat Step 41.

43| Resuspend the beads in 100 µl of LoTE buffer.

■ **PAUSE POINT** The sample can be stored at 4 °C overnight at this point.

MmeI digestion ● TIMING 2.5 h

44| Prepare 50 µM double-stranded M12 oligonucleotide[11] by combining the following in a new PCR tube.

M12_top (100 µM in EB)	15 µl
M12_bot (100 µM in EB)	15 µl
1 M NaCl	1.5 µl

45| Anneal oligonucleotides in a thermocycler: 95 °C for 5 min; cool to 4 °C at a rate of 0.1 °C s^{-1}; store in 5-µl aliquots at −20 °C for future use.

46| Prepare MmeI buffer mix on ice.

Component	Volume per sample (µl)
dH_2O	16.8
10× NEBuffer 4[a]	2
32 mM SAM[a]	0.08
M12 dsDNA	0.2

[a]Part of the MmeI restriction enzyme package from NEB.

47| Collect the beads from Step 43 with the MPC, carefully discard the supernatant, remove the tube from the MPC, and then gently resuspend each sample in 19 µl of MmeI buffer mix.

48| Add 1 µl of MmeI to the sample.

49| Incubate in the thermocycler at 37 °C for 1 h. Gently mix the sample every 10–15 min.

50| Add 100 µl of LoTE buffer to the sample, collect the beads in the MPC, and then carefully discard the supernatant.

51| Repeat Step 50.

52| Resuspend the beads in 100 µl of LoTE buffer.

■ **PAUSE POINT** The sample can be stored at 4 °C overnight at this point.

Linker ligation ● TIMING 2.5 h

53| Prepare a 50 μM stock of barcoded, double-stranded sequencing adapters (one barcode sequence per sample) by combining the following in a new PCR tube.

LIB_AdaptT_(barcode) (100 μM in EB)	15 μl
LIB_AdaptB_(barcode) (100 μM in EB)	15 μl
1 M NaCl	1.5 μl

54| Anneal oligonucleotides in a thermocycler: 95 °C for 5 min; cool to 4 °C at a rate of 0.1 °C s^{-1}.

■ **PAUSE POINT** Adapters can be stored for months in 5-μl single-use aliquots at −20 °C for future use.

55| If they are frozen, thaw dsDNA sequencing adapters on ice (one barcode per sample). Dilute each 50 μM stock of dsDNA sequencing adapter to 5 μM in ice-cold 1× T4 DNA ligase buffer.

▲ **CRITICAL STEP** If the adapters are not thawed on ice, they can dissociate into single-stranded DNAs.

56| Prepare ligation mix on ice.

Component	Volume per sample (μl)
dH_2O	16.4
10× T4 DNA ligase buffer	2

57| Collect the beads from Step 52 with the MPC, carefully discard the supernatant, remove the tube from the MPC and gently resuspend the sample in 18.4 μl of ligation mix.

58| Add 0.6 μl of 5 μM dsDNA sequencing adapter (from Step 55) containing a unique barcode to the sample. Record which barcode is associated with the sample.

59| Add 1 μl of T4 DNA ligase to the sample.

60| Incubate in a thermocycler at 16 °C for 1 h. Gently mix every 10–15 min.

61| Add 100 μl of LoTE buffer to the sample, collect the beads in the MPC and carefully discard the supernatant.

62| Repeat Step 61.

63| Resuspend the beads in 100 μl of LoTE buffer.

■ **PAUSE POINT** The sample can be stored at 4 °C overnight at this point.

PCR and final purification ● TIMING 4 h

64| Assemble the PCR mix on ice.

Component	Volume per sample (μl)
dH_2O	31.5
10× Pfx buffer	10 (used at 2× standard concentration)
10 mM dNTPs	2
50 mM $MgCl_2$	2
5 μM LIB-PCR5	2
5 μM LIB-PCR3	2
Pfx polymerase	0.5

65| Collect the beads from Step 63 with the MPC, carefully discard the supernatant, remove the tube from the MPC and gently resuspend the sample in 50 μl of PCR mix on ice.

66| Run on a thermocycler, as follows, and prepare a 2% (wt/vol) agarose gel (1–2 lanes per sample plus 2 ladder lanes per gel; GelGreen dye at 1:10,000 dilution; wide-tooth comb) while PCR is running. Run at 94 °C for 2 min followed by 18 cycles of: 94 °C for 15 s, 60 °C for 1 min, 68 °C for 2 min and then 68 °C for 4 min.
■ **PAUSE POINT** The sample can be stored at 4 °C overnight at this point.

67| Collect the beads on the MPC and transfer the supernatant to a new PCR tube.

68| Prepare the DNA ladder.

10-bp DNA ladder	4 µl
Bromophenol blue loading dye	1× final concentration
dH_2O	20 µl final volume

69| Add xylene cyanol loading dye to the sample (supernatant from Step 67) to 1× final concentration.
▲ **CRITICAL STEP** Xylene cyanol is used because bromophenol blue can migrate near the same position as the desired 125-bp PCR product, thereby obscuring the band in the gel. The use of bromophenol blue in the ladder lanes will provide an approximate position of the samples in the gel during electrophoresis.

70| Load the DNA ladder onto the first and last lanes of the gel.

71| Load the sample into its own lane of the gel. If the wells are too small for the full volume, split the sample over two lanes.

72| Run the gel for 30 min at 200 V.
▲ **CRITICAL STEP** Time and voltage may need to be adjusted on the basis of gel size and buffer composition; these parameters work well for 7-cm gels run in borate buffer.

73| Use a non-UV gel illuminator to excise the band at ~125 bp, minimizing surplus agarose in the gel fragment. Place the gel fragment in a microcentrifuge tube (if two lanes were used per sample, the similar fragments can be placed in the same tube as long as the total weight remains under 300 mg).
? TROUBLESHOOTING

74| Clean up the sample using QIAquick gel purification columns according to the manufacturer's instructions (including the isopropanol step for small fragments). Be certain that all of the ethanol-containing wash buffer is removed from the column before elution by pipetting around the inner rim with a 10-µl tip, if necessary. Elute in 32 µl of buffer EB. Gel purification ensures that primer-dimers or other erroneous products do not contribute to the DNA quantification or sequencing in subsequent steps.

75| Quantify the sample on a QuBit or similar spectrophotometer.
▲ **CRITICAL STEP** DNA concentration is typically 0.1–5 ng μl^{-1}, which is below the reliable detection limit of many UV absorbance–based spectrophotometers.

76| All samples should subsequently be normalized to the same concentration. If all samples are ≥0.83 ng μl^{-1}, adjust an aliquot of each sample to 10 nM in EB buffer. On the basis of an expected fragment length of 125 bp, the formula for a 10 nM normalization is volume of DNA (µl) = 10 × final volume (µl) / (DNA concentration (ng μl^{-1}) × (10^6) × (1/649) × (1/125)). If sample concentrations are <0.83 ng μl^{-1}, adjust an aliquot of each sample to 1 nM in EB buffer. Tip: the formula for a 1 nM normalization is the volume of DNA (µl) = 1 × final volume (µl) / (DNA concentration (ng μl^{-1}) × (10^6) × (1/649) × (1/125)).

77| Combine an equal volume of each sample into a single tube.
▲ **CRITICAL STEP** Required volumes may vary by sequencing facility. Be sure to inform the sequencing facility of whether the sample is at 10 or 1 nM.

78| Submit for Illumina sequencing. The optimum loading concentration may vary by sequencing center or by Genome Analyzer model; we currently use 6 pM for GA-IIx and HiSeq machines.
? TROUBLESHOOTING

Data analysis ● TIMING ~3 min per million raw reads
79| Download the data analysis package (INSeq_analysis.zip; **Supplementary Data 1**) and sample data package (INSeq_demo.zip; downloadable from http://gordonlab.wustl.edu/SuppData.html) and save each to your own directory.

80| Bowtie is used to map sequences to the reference genome. Download this mapping software from http://bowtie-bio.sourceforge.net/index.shtml and install in your system. Edit the 'config.txt' file in the data analysis package (INSeq_analysis) by changing the bowtie_dir variable to the path where bowtie was installed. Detailed instructions are provided in the README file in the INSeq_analysis folder (see **Supplementary Data 1**).

81| Go to the uncompressed data analysis package (INSeq_analysis; see **Supplementary Data 1**) and then to the 'indexes' directory. From this location, create a new directory with the organism's name (e.g., BthetaVPI_5482). Put the genome file in .fasta format and the annotation files (in .ptt format) in the new directory. Go into the new directory and construct a bowtie index by typing the command:

```
"path to bowtie directory"/bowtie-build <fasta file> <name of index, same as the new directory>
```

? TROUBLESHOOTING

82| Create an experiment directory where you would like to store your analysis results. For example:

```
mkdir /Users/mwu/Documents/Experiment_1
```

83| Create a mapping file that contains the barcode for each sample in tab-delimited format in the experiment directory:

```
<barcode><Sample Name>
```

84| From the experiment directory, run the analysis pipeline by running the wrapper script INSeq_pipeline.pl with the usage:

```
perl "path for analysis package"/INSeq_pipeline.pl -i <the raw reads file> -m
<Barcodes mapping file> -s <indexed genome name> -d <length_disrupt_percent (max=1)>
```

Required arguments: -i gives the input raw reads file, -m gives the mapping file and -s gives the name of the indexed genome to which the reads should be mapped.

Optional arguments: -d gives the region of the gene in which insertions are expected to disrupt gene function. The default is 1, which means that when the insertion falls anywhere within the gene (100%), the gene's function will be considered to be disrupted. Setting the -d argument to 0.9, for example, would exclude insertions in the distal 10% of the gene when calculating the total number of reads/insertions for that gene.
? TROUBLESHOOTING

85| The INSeq_pipeline.pl will generate several mappingjobsXXX.job under the experiment directory. On a desktop/laptop, run them in series by typing `sh mappingjobsXXX.job` for each job, or (on a cluster) run in parallel. Several output files will be generated:

INSEQ_experiment.scarf_assigned.txt: reads are assigned to different samples on the basis of the sample-specific barcode.

INSEQ_experiment.scarf.log: contains some statistics for the analysis process, including the total number of reads, the percentage of reads being mapped and trimmed (that have the transposon), how many insertions in the sample (coverage) with the number of raw reads (reads after filtering in the normalization step) and the scale factor being used for normalization.

Several output files will be placed in the 'results' folder:

INSEQ_experiment.scarf_Samplename.bowtiemap: a raw mapping output file from bowtie.

INSEQ_experiment.scarf_Samplename.bowtiemap_processed.txt_chromosomename: a text file containing the processed mapping output, with the format:

```
<chromosome name> <insertion position> <reads mapped to the left side of the insertion> <reads mapped to the right side of the insertion> <the total number of reads mapped to that position>
```

INSEQ_experiment.scarf_Samplename.bowtiemap_processed.txt_chromosomename_filter_cpm.txt: a text file with positions having more than three total reads, and normalized to counts per million reads.

INSEQ_experiment.scarf_Samplename.bowtiemap_processed.txt_chromosomename_filter_cpm.txt_mapped: a text file that lists genes mapped by insertions, only when the insertions located at the proximate XX percentage (specify by the -d option) are considered to interrupt the function of genes. The format is:

```
<gene name> <the total number of unique insertions in the gene> <the sum of normalized read counts in that gene> <gene annotation from the ptt file>
```

? TROUBLESHOOTING

86| *Cleanup*: as there are many intermediate files generated along the pipeline, if storage space is a big concern, deleting them is optional. You can run 'clean_up.sh' from the experiment folder by using command `sh clean_up.sh`, which will remove the sorted barcode assigned reads file, the folder 'bcsortedseqs' and the mappingjob files.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
34	Beads are lost during wash steps	Electronic multichannel pipettor is too fast	If alternate pipettors are not available, perform this step manually
73	Bands are faint or inconsistent between samples	Ligation adapters not properly annealed, or have barcodes that prevent proper annealing	Use technical replicates (same input DNA with different sample-specific barcodes) to determine whether the barcode or the input DNA is the problem
78	Illumina sequencing produces fewer reads than expected	Sequence characteristics prevent proper function of Illumina base-calling software	Provide the sequencing facility with information about common and variable bases in your expected sequences on the basis of the barcodes used
81	Error message 'bowtie-build: No such file or directory'	Command did not specify the location of bowtie	Make sure to include the entire path for bowtie, e.g., '/Users/XXX/bowtie-0.12.7/bowtie-build Bt.fna BtVPI 5482'
84	Analysis pipeline cannot recognize the raw reads file	The raw reads file is not in the scarf format	Convert your raw reads file to scarf format
85	Error message 'bowtie: No such file or directory'	The analysis package could not locate bowtie	Check the 'config.txt' in Step 80; make sure the bowtie_dir variable is the path where your bowtie installation is located. For example, bowtie_dir='/Users/XXX/bowtie-0.12.7'
	Error message 'Could not locate a bowtie index corresponding to base name 'XXX''	The analysis package could not find the bowtie index. The analysis package is looking for the index file located in 'INSEQ_analysis/indexes/XXX/XXX' (XXX is the 'indexed genome name' specified in Step 84 by argument -s)	Check the location and the name of your index file. Make sure it is UNDER the 'INSeq_analysis/indexes' directory (it should be in the position exactly inside of the 'indexes' folder), and check that the name of the index, which is the prefix of the .X.ebwt file, is consistent with the name of the directory. If not, remake the ebwt file using bowtie-build and repeat Step 81

(continued)

TABLE 1 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
85	The gene-mapped results show all genes have zero insertions, whereas the insertion results show that there are several insertions along the genome	The analysis package matches the gene annotation file (.ptt) to chromosome (.fna) by names. If the naming is not consistent, the pipeline can not locate the insertions to genes correctly	Remove any spaces in any names, and rename the header names of the .fna files to match the .ptt files, or rename the .ptt files to match the header names of the .fna files

● TIMING

Steps 1–16, Isolation of crude DNA: 4 h
Steps 17–20, Cleanup of crude DNA using QIAquick columns: 1 h
Steps 21–23, Linear PCR: 2 h
Steps 24–36, Bind linear PCR product to beads: 1 h
Steps 37–43, Second strand synthesis: 1 h
Steps 44–52, MmeI digestion: 2.5 h
Steps 53–63, Linker ligation: 2.5 h
Steps 64–78, PCR and final purification: 4 h
Steps 79–86, Data analysis: initial analysis requires ~3 min per million raw reads on a 3.06 GHz Intel Core 2 Duo processor with 4 GB memory
Box 1, arrayed library preparation: ~1.6 h per 96-well tray, plus time for picking colonies into trays and INSeq library preparation

ANTICIPATED RESULTS

The INSeq analysis pipeline produces several output files with data at the level of individual insertion locations, genes and (optionally) operons. Good libraries should have >95% of reads that contain the transposon sequence, consistent read counts generated from both sides of each insertion and consistent data from biological and technical replicates. A real data set from a successful library preparation is provided in the INSeq_demo files.

Note: Supplementary information is available via the HTML version of this article.

ACKNOWLEDGMENTS This work was supported by grants from the National Institutes of Health (DK30292, DK70977, and DK064540 to J.I.G.; F32AI078628 and K01DK089121 to A.L.G.) and by the Crohn's and Colitis Foundation of America. We thank L. Kyro and J. Eberle for their assistance in making the movies that accompany this article.

AUTHOR CONTRIBUTIONS A.L.G., J.I.G. and M.W. designed the experiments and software. A.L.G., M.W. and J.I.G. wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Published online at http://www.natureprotocols.com/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Mazurkiewicz, P., Tang, C.M., Boone, C. & Holden, D.W. Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. *Nat. Rev. Genet.* **7**, 929–939 (2006).
2. 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).
3. Erlich, Y. *et al.* DNA Sudoku—harnessing high-throughput sequencing for multiplexed specimen analysis. *Genome Res.* **19**, 1243–1253 (2009).
4. Prabhu, S. & Pe'er, I. Overlapping pools for high-throughput targeted resequencing. *Genome Res.* **19**, 1254–1261 (2009).
5. Xin, X. *et al.* Shifted Transversal Design smart-pooling for high coverage interactome mapping. *Genome Res.* **19**, 1262–1269 (2009).
6. Gawronski, J.D., Wong, S.M., Giannoukos, G., Ward, D.V. & Akerley, B.J. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc. Natl. Acad. Sci. USA* **106**, 16422–16427 (2009).
7. Langridge, G.C. *et al.* Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res.* **19**, 2308–2316 (2009).
8. van Opijnen, T., Bodi, K.L. & Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* **6**, 767–772 (2009).
9. Gallagher, L.A., Shendure, J. & Manoil, C. Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. *MBio* **2**, 00315-10 (2011).
10. van Opijnen, T. & Camilli, A. Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms. *Curr. Protoc. Microbiol.* **19**, 1E.3.1–1E.3.16 (2010).
11. Morgan, R.D., Bhatia, T.K., Lovasco, L. & Davis, T.B. MmeI: a minimal type II restriction-modification system that only modifies one DNA strand for host protection. *Nucleic Acids Res.* **36**, 6558–6570 (2008).
12. Brody, J.R. & Kern, S.E. Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. *Biotechniques* **36**, 214–216 (2004).
13. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
14. Westover, B.P., Buhler, J.D., Sonnenburg, J.L. & Gordon, J.I. Operon prediction without a training set. *Bioinformatics* **21**, 880–888 (2005).