Comparing Insertion Libraries in Two *Pseudomonas aeruginosa* Strains to Assess Gene Essentiality

Nicole T. Liberati, Jonathan M. Urbach, Tara K. Thurber, Gang Wu, and Frederick M. Ausubel

Summary

Putative essential genes can be identified by comparing orthologs not disrupted in multiple near-saturated transposon insertion mutation libraries in related strains of the same bacterial species. Methods for identifying all orthologs between two bacterial strains and putative essential orthologs are described. In addition, protocols detailing near-saturation transposon insertion mutagenesis of bacteria are presented, including (1) conjugation-mediated mutagenesis, (2) automated colony picking and liquid handling of mutant cultures, and (3) arbitrary polymerase chain reaction amplification and sequencing of genomic DNA adjacent to transposon insertion sites.

Key Words: essential genes; *MAR2xT7*; *mariner*; PA14, *Pseudomonas aeruginosa*.

1. Introduction

The availability of multiple, nearly saturated mutant libraries in related strains of a single bacterial species offers the opportunity of identifying orthologous genes that are nondisrupted in more than one library. The set of nondisrupted genes are putative "essential" genes. In this chapter, methods are described for creating a nearly saturated bacterial transposon insertion library including conjugation-mediated mutagenesis, arraying transposants into plates using a colony-picking robot, and aliquoting mutant cultures using a liquid-handling robot. Specific guidelines, based on quality control testing, are described for automated handling of bacterial cultures that minimize cross-contamination. Methods are also described for identifying transposon insertion sites using two-step polymerase chain reaction (PCR) amplification of the DNA adjacent to the transposon insertion site using arbitrary primers and subsequent sequencing of the PCR products.

This chapter also describes the use of a custom-designed database for automated DNA sequence analysis. DNA sequences adjacent to transposon insertions are entered into the database where they are aligned with the genomic sequence of the mutagenized
strain using the BLAST algorithm. The approximate genomic locus of each insertion site is determined by the alignment with the best BLAST score. The precise location of each insertion is determined using a modified Smith-Waterman algorithm that aligns sequences obtained from each mutant with the 3' end of the transposon sequence. Once all insertion sites have been located, genes that have been disrupted as well as those that have not can be identified. Finally, protocols are described that identify orthologs in two bacterial strains that can be used to detect essential genes based on the absence of insertion mutants in the orthologs in more than one transposon mutant library.

The protocols described in this chapter are based on experiments carried out in our laboratory that were involved in the use of the *mariner*-based transposon *MAR2xT7* to generate a mutation library in *Pseudomonas aeruginosa* strain PA14 (1). Nevertheless, many if not all of the protocols can be readily adapted for generating transposon mutation libraries in most Gram-negative bacterial species.

2. Materials

2.1. Bacterial Strains

1. Recipient strain: the strain to be mutagenized must be λ-*pir*.
2. Donor strain: *pir*<sup>+</sup> *Escherichia coli* strain carrying plasmid containing a transposase, a compatible transposon that confers antibiotic resistance, additional antibiotic resistance marker outside of the transposon, and a bacteriophage λ-*pir*-dependent origin of replication. In the case of the *P. aeruginosa* strain PA14 library (1), the vast majority of mutants were created with *MAR2xT7*, a gentamicin-resistant derivative of the *Himar1* transposon (2, 3). All protocols in this chapter are based on *MAR2xT7* insertion.
3. Helper strain: If the donor plasmid carrying the transposon is mobilizable but not self-transmissible (*mob*<sup>+</sup> *tra*<sup>-</sup>), an *E. coli* strain carrying a broad-host range helper plasmid should be included in the mating to facilitate conjugation. For example, pRK2-derived *IncP* broad-host range plasmids encode the *tra* genes necessary for conjugal transfer in *trans* (4). If using a helper plasmid, the donor plasmid carrying the transposon requires a proper mobilization sequence known as the origin of conjugal replication (5). To mobilize the donor plasmid encoding *MAR2xT7*, we used a helper strain, *E. coli* HB101 carrying the pRK2 derivative pRK2013, in all matings.

2.2. Relational Database

1. A relational database is needed to (a) track mutant location, (b) track processing status information, and (c) analyze sequencing data for each mutant. The database must contain the genomic sequence of the strain being mutagenized and the coordinates of all predicted open reading frames (ORFs). The PA14 Transposon Insertion Mutant Database (PATIMDB) that was developed and used in our laboratory is implemented using the MySQL RDBMS hosted on a multiprocessor Intel system running RedHat Linux (1). The data-entry application is in Java and runs on Windows 2000. PATIMDB is compatible with different genome sequences and is adaptable to library construction applications in other organisms. A “generic” version of PATIMDB that is designed for use with any bacterial genome will be downloadable in the future at http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/downloads.cgi.
2.3. Equipment

1. Q-fill media dispenser (Genetix, Boston, MA).
2. QBot colony-picking robot (Genetix).
4. HiGro block shaker/incubator (Genomic Solutions, Ann Arbor, MI).
5. Thermocyclers with capacity to run multiple 96-well plates in parallel (e.g., ThermoHybaid, Ashford, Middlesex, UK).
6. ABI 3700 PRISM automated sequencer.
7. 12-channel pipettes (Costar or Finnpipette) with compatible tips.
8. Tabletop centrifuges with block/plate attachments (Beckman Coulter Allegra X-22, Fullerton, CA).
9. Laminar flow hood.
10. −80°C and −20°C freezers including racks designed to hold 96-well plates.
11. Aluminum bases to cool 96-well plates.

2.4. Reagents

1. Arbitrary PCR and sequencing primers (see Methods and Fig. 1).
2. Deoxynucleotide triphosphates (dNTPs), PCR grade (Roche, Indianapolis, IN).
3. Taq DNA polymerase and 10× buffer (no. 1147633; Roche).

---

**Fig. 1.** Transposon insertion library production and analysis workflow.
4. Dimethyl sulfoxide (DMSO) (no. D-8418; Sigma-Aldrich, St. Louis, MO).
5. ExoSAPIT (no. 78205; USB Cleveland, OH).
7. King's B media (2% w/v peptone, 6.57 mM K₂HPO₄, 6.98 mM MgSO₄, 1% w/v glycerol).
8. Luria Broth (1% w/v peptone, 0.5% w/v yeast extract, 0.5% w/v NaCl).
9. Sterile 60% glycerol.
10. Antibiotics to select for the transposon and to select against other mating strains.

2.5. Consumables
1. 20 cm × 20 cm low-profile bioassay dish (no. 240845; Nunc, Rochester, NY).
2. Glass balls 3 mm in diameter (no. 26396-508; VWR, West Chester, PA).
3. Biomek AP96 P250 tips or equivalent (no. 717251; Beckman Coulter, Fullerton, CA).
4. Culture plates/blocks: Greiner 96-well flat-bottom plates with lids (Greiner no. 655185; or alternatively Culture Blocks: 2.0mL 96-well V-bottom polypropylene blocks (no. 3961; Costar/Corning, Acton, MA).
5. Sealing mats for 2.0-mL 96-well blocks (no. 3083; Costar/Corning).
6. Copy plates: low-profile 96-well Serowell V-bottom plates (Bibby Sterilin Serowell no. 611V96 or Abgene no. MP-2000).
7. Lids with stacking rim for 96-well Serotec plates (no. AB-0752; Abgene, Rochester, NY).
8. PCR template plates: 96-well skirted thermo-fast reaction plates (no. AB-0800/150; Abgene).
9. PCR reaction plates: 96-well thin-walled, skirted polycarbonate PCR plates (no. 6511; Costar/Corning).
10. Sterile ARB reaction mix reservoirs (no. 13681501; Fisher Scientific, Pittsburgh, PA).
12. AluminaSeal temperature-resistant seals (no. ALUM-1000; Diversified Biotech Boston, MA).
13. Temperature-resistant cryo-tags (plate labels) (no. SIDE-1000; Diversified Biotech).

3. Methods

3.1. Mutagenesis and Arraying Mutants into 96-Well Plates

If the donor plasmid carrying the transposon is self-transmissible, the strain carrying the donor plasmid is directly mated with the recipient strain. If, however, the donor plasmid is not self-transmissible but is mobilizable, triparental mating with donor, recipient, and helper strains should be performed.

3.1.1. Mating (Note 1)
1. Grow separate saturated cultures of transposon donor and recipient strains (and if necessary, helper strain) with appropriate antibiotics in appropriate media (e.g., Luria Broth; LB).
2. Mix 200 μL recipient strain culture with 400 μL donor strain culture (and, if necessary, 400 μL helper strain culture). Gently pellet cells. Generally a 2:1 ratio of donor to recipient culture volume is recommended but should be tailored to individual mating combinations. It may be necessary to set up multiple mating mixes depending on the frequency of transposition and mating efficiency.
3. Rinse pellet in 1mL mating media. For PA14, King’s B media was used (6). However, other media, including LB, may be appropriate for other mating combinations. Gently pellet cells. Resuspend pellet in 250μL King’s B media.

4. Spot 25-μL aliquots of the resuspended mixture on King’s B media 1.5% agar plates, keeping the area of the drops on the plates as small as possible. Let plates dry before moving to 37°C incubator. Incubate plates for shortest time necessary for transposition (Note 2).

5. Using a sterile pipette tip, scrape one or more mating spots into a tube containing 48mL 0.1M MgSO₄. Resuspend thoroughly by vortexing vigorously. The number of spots that you need to resuspend depends on the frequency of transposition and mating efficiency (Note 3).

6. Using approximately 30 sterile glass balls, spread 1.5mL of the suspension on Luria broth agar (1.5% w/v) in separate 20cm x 20cm bioassay dishes containing appropriate antibiotics. To select for transposants, include the antibiotic that the transposon confers resistance to. To select against the donor strain, include an antibiotic to which the recipient strain is resistant to but the donor strain is not (Notes 4 and 5). Be sure to spread the culture evenly across the plates. For compatibility with the Qbot colony picking robot, the 20cm x 20cm dishes should contain exactly 200mL media.

7. Dry dishes for approximately 45 min in a laminar flow hood until all fluid has dried. Incubate dishes at 37°C for 12 to 15 h.

8. Store dishes at 4°C prior to colony picking.

3.1.2. Label Plates

Create “virtual” plates in the database representing culture plates. The database program should automatically create a set of unique identifiers for those plates. The database should also create “virtual” mutants with their own unique identifiers that are linked to culture plate well positions so that sequencing data for individual mutants can be properly entered and stored. Unique identifiers are numbers or alphanumeric keys that unambiguously specify a particular database entity including plates, mutants, and so forth. Finally, the database should also create human-readable numeric plate labels and, if desired, bar-code labels encoding unique identifiers for each culture plate. These labels should be generated as text files that can be printed (see below).

Optional: In the generation of the PA14 transposon library, PATIMDB-created virtual plates for culture, PCR template, PCR reaction, and copy plates, thereby allowing the status of each plate (e.g., whether the ARB 1 reaction has been performed) to be entered and monitored. This tracking feature is not essential but may be useful with different applications. Files containing bar codes encoding unique identifiers and human-readable text labels for each culture plate and the PCR template plates, PCR reaction plates, and copy plates derived from each culture plate are automatically generated by the database once virtual culture plates have been created.

1. Print and apply labels. Files containing bar codes and/or corresponding human-readable labels can be created with Sagian Print and Apply Software (part no. 148640; Sagian Core Systems, Indianapolis, IN). The labels are then printed and transferred to each plate via a print-labeling machine. If this software/hardware is not available, labels for each plate generated by the database can be printed on special temperature-resistant labels (Materials) with general word-processing software and a standard laser printer and applied...
to each plate by hand. Labels should be applied to one of the two short sides of the plates because these sides face out in the racks used to store plates at -80°C.

3.1.3. Robotic Colony Picking and Inoculation of Culture Plates

The QBot colony picking robot comes with a computer workstation and custom software (QSoft) to run the robot (Note 6).

1. Load QBot Platform. Place 20 cm x 20 cm dishes (up to four at a time) containing transposant colonies on the QBot platform. Fill ethanol bath with 70% ethanol.
2. Start QBot software.
3. Camera alignment: Make a hole with a pipette tip in a colony-free area of the agar media in dish no. 1 (the dish located in the upper-left-hand holder on the QBot platform). A dish containing no colonies can also be used. Choose “Align Camera,” select “Assay Tray,” and then click “Yes.” Change “Zoom Focus” to 3 and, using the cursors, set the bull’s-eye on the hole. This procedure sets X-Y coordinates for the picking head.
4. Set picking height. Mark “Stop Short on Z Axis,” and click the center dot on the screen. Select “Pin A1 Down” over dish no. 1. Click the down arrow, carefully directing the pin to move down until the pin just touches (but does not pierce) the agar. Enter “OKAY” to exit. Repeat setting picking height for each of the four dishes on the platform. Return to home position.
5. Test imaging colonies. Each plate is divided into 40 sectors. Select sector 18 of dish no. 1 (the sector located in the middle of dish no. 1) by highlighting it. The camera will move to sector 18 of dish no. 1. Take image by selecting “Picture.” A picture of the selected sector will appear with well-defined colonies in green, unclear colonies in yellow, and poor colonies in red. Select “Tools” and “Threshold” to adjust the light settings to maximize the green-to-red colony ratio (Note 7). Click “Reprocess.” Check other sectors in the same tray to ensure that the green-to-red colony ratio is high. If it is not, return to sector 18 and readjust light threshold as necessary. Check other trays and determine if the set light threshold can be used across all plates on the platform. The light threshold can only be adjusted in sector 18 for each plate. Select “Done” when all trays have been imaged.
6. Set picking run. The software will ask if the steps outlined above have been completed. Select “OKAY” for each of these questions. Specify whether all sectors in all trays are to be picked (full run) or only specific sectors in specific trays (partial run). If a partial run is selected, highlight which sectors should be picked. Once the imaging is complete, a message will appear: “Script is Complete.” View image result in the last picked sector screen, which will show the total colonies identified in the run (green, yellow, and red) and the total colonies that the QBot would pick (green only) based on the set parameters in all sectors in every dish.
7. Select “OKAY” to view the Destination Plate Guide. Based on the number of colonies that will be picked as described above, the plate guide will indicate the number of destination plates and where to load them in the two hotels on the QBot platform.
8. Prepare destination plates. Use the QFill media dispenser to fill labeled 96-well flat-bottom plates with 280 µL LB containing appropriate antibiotics to select for transposants. Replace lids and load into the QBot’s hotels with the cut corner side of the plate positioned outward. If using different plates or media volumes, be sure the media level in the wells is high enough so that the QBot head pins actually touch the media. Select “Done” to exit.
9. Begin picking. A message will appear: “Are you ready to begin?” Select “YES.” The pins are rinsed in 70% ethanol and dried for 10 s between each plate. If more destination plates
are needed, the QBot head will stop in front of the last picked plate in the hotel, and the software will indicate how many more plates to reload. After the picking run, the message “Picking is Complete” will appear. Select “OKAY.” If you want to have the QBot pick from additional trays, select “YES” when the “Do you want to save?” message appears. The QBot will resume picking, inoculating the remaining uninoculated wells, if any, in the last destination plate used (Note 8).

10. Carefully seal destination plates with Aeroseal covers, taking care not to disrupt the media.

3.1.4. Culture Plate Incubation

Incubate the inoculated and sealed Culture plates at 37°C without shaking long enough to ensure that slow-growing mutants produce cultures (Note 9).

3.1.5. Aliquoting Mutant Strains Using a Biomek FX Liquid-Handling Robot (Fig. 1)

The Biomek FX transfers strains from culture plates to several destination plates: a PCR template plate and three copy plates. This can be done by hand using multichannel pipettes but would require extensive labor. The following steps are used to set a Biomek FX program. Once all plates are properly loaded onto the Biomek platform with lids removed, run the program.

3.1.5.1. FX Method 1: Library Replication for Storage and PCR Processing (Notes 10, 11, and 12)

1. Label and QFill copy plates with 100 µL LB containing 15% glycerol and antibiotics (Note 13).
2. Set up deck: Two boxes of tips, one set in the home position; one reservoir filled with sterile 60% glycerol; a culture plate containing fully grown cultures; a bar-coded/labeled PCR template plate; and three bar-coded/labeled and QFilled 96-well copy plates.
3. Pick up tip set no. 1.
4. Aspirate 70 µL of each culture from culture plates at 5 mm below the liquid surface, touch tips to side of wells, and transfer to labeled PCR template plates (Notes 10, 11, and 12a). Touch tips to side of wells (Note 12b). PCR template plate cultures are subsequently used as templates for arbitrary PCR.
5. Discard tips.
6. Move tip set no. 2 into the home position.
7. Pick up tip set no. 2.
8. Aspirate 70 µL 60% sterile glycerol from the reservoir 2 mm from the bottom of the reservoir at 70% speed and transfer to the culture plate. Dispense glycerol 5 mm below the liquid surface at 70% speed, with tips following the liquid level as it rises (Notes 10 and 11).
9. Mix at least three times 100 µL at 70% speed, following the liquid level as it rises and falls with each aspiration and dispense step with no blowout. Final glycerol concentration after mixing is 15% glycerol. Other types of bacteria may require different glycerol concentrations or other freezing agents such as DMSO. Adjust as necessary.
10. Using the same tips (or a fresh set of tips if desired) aspirate 25 µL culture/15% glycerol mix from 5 mm below the liquid surface, following liquid level, and touch tips to wells. Transfer culture/glycerol mix to the first labeled 96-well copy plate and dispense 3 mm
from the copy plate bottom at 70% speed. Mix 1 μL once at 100% speed (to remove hanging culture drop) with no blowout. Include a tip touch.

11. Repeat step 10 and dispense in the second copy plate.
12. Repeat step 10 and dispense in the third copy plate.
13. Seal all plates with AluminaSeals.
14. Store culture and copy plates at −80°C. Store PCR template plates at −20°C until PCR processing.

3.2. Insertion Site Identification

3.2.1. Arbitrary PCR

Transposon insertion sites were identified using a two-round arbitrary PCR protocol (Fig. 1) (9).

3.2.1.1. The First Round of Arbitrary PCR (ARB1 Reaction)

1. Thaw PCR template plates containing 70 μL aliquots of the statically grown transformant cultures (see above) and incubate at 95°C for 10 min to lyse the cells.
2. Pellet debris by centrifuging the plate at 3000 rpm for 5 min. The cleared lysate is used as template for the first round of arbitrary PCR (ARB1).
3. Combine reagents for the ARB1 reaction mix in a sterile tube on ice: 1× Taq buffer (Roche), 10% DMSO, 2.5 μM dNTPs (Note 14), 1.25 U Taq DNA polymerase (Roche), 1.0 ng/μL of the transposon specific primer Tn1 (Fig. 1), and an arbitrary primer (Table 1) (Note 15). Taq is added after all other reagents are mixed thoroughly. Once Taq is added, mix by inverting the tube. For P. aeruginosa PA14 MAR2ΔT7 mutants, the transposon-specific primer, PMFL.GM.GB-3a, (5'-TACAGTTTACGAAACGACAGGC-3') was used. Transfer ARB1 reaction mix to a reservoir on ice that will accommodate a 12-channel pipettor.
4. Using a 12-channel pipette, transfer 25 μL ARB1 reaction mix to the wells of a thin-walled 96-well PCR reaction plate sitting on ice in an aluminum plate cooler.
5. Using a 12-channel pipette, transfer 3 μL of the cleared lysates to the reaction mix, pipetting up and down three times to mix.
6. Seal plates with adhesive foil seals.
7. Begin the ARB1 reaction program on the thermocycler:
   (a) 95°C for 5 min
   (b) 95°C for 30 s
   (c) 47°C for 45 s
   (d) 72°C for 1 min
   (e) Repeat steps (b), (c), and (d) for 30 cycles.
   (f) 72°C for 5 min
8. Once the thermocycler reaches 95°C for initial denaturation, transfer the ARB1 reaction plate to the thermocycler.
9. Update the status of plates processed for the ARB1 reaction in the database if desired.

3.2.1.2. The ARB2 Reaction

1. Combine reagents for the ARB2 reaction mix in a sterile tube on ice: 1× Taq buffer (Roche), 10% DMSO, 2.5 μM dNTPs, 1.25 U Taq polymerase (Roche), 1.0 ng/μL of the transposon specific primer Tn2 (Fig. 1), and an arbitrary primer (either ARB2 or ARB2A; Table 1). As in the ARB1 reaction, Taq is added after all other reagents are mixed
### Table 1
**Arbitrary Primers**

<table>
<thead>
<tr>
<th>Primer name</th>
<th>First round Sequence</th>
<th>Primer name</th>
<th>Second round Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARB1</td>
<td>GGCCACCGCGTCGACTAGTACNNNNNNNNNGATAT</td>
<td>ARB2</td>
<td>GGCCACCGCGTCGACTAGTACNNNNNNNNNGATAT</td>
</tr>
<tr>
<td>ARB1A</td>
<td>GGCCACCGCGTCGACTAGTACNNNNNNNNNGTATA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARB1B</td>
<td>GGCCACCGCGTCGACTAGTACNNNNNNNNACNG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARB1C</td>
<td>GGCCACCGCGTCGACTAGTACNNNNNNNGTAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARB1D</td>
<td>GGCCAGGCCCTGCAGATGATGNNNNNNNNGTAT</td>
<td>ARB2A</td>
<td>GGCCAGGCCCTGCAGATGATGNNNNNNNNGTAT</td>
</tr>
<tr>
<td>ARB1E</td>
<td>GGCCAGGCCCTGCAGATGATGNNNNNNNNGTANG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
thoroughly. Once Taq is added, mix by inverting the tube. For MAR2xT7 mutants, the transposon-specific primer, PMFLGM.GB-2a, (5'-TGTTCAACTGGGGTCGCTGGGCTCC-3') was used. Transfer ARB2 reaction mix to a reservoir on ice that will accommodate a 12-channel pipettor.

2. Using a 12-channel pipette, transfer 20µL ARB2 reaction mix to a thin-walled PCR reaction plate sitting on ice in an aluminum plate cooler.

3. Using a 12-channel pipette, transfer 5µL of each ARB1 product to the ARB2 reaction mix in the PCR plate, pipetting up and down three times to mix.

4. Seal plates with adhesive foil seals.

5. Begin the ARB2 reaction program on the thermocycler:
   (a) 95°C for 30s
   (b) 45°C for 45s
   (c) 72°C for 1 min
   (d) Repeat steps (a), (b), and (c) for 40 cycles
   (e) 72°C for 5 min

6. Once the thermocycler reaches 95°C for initial denaturation, transfer the ARB2 reaction plate to the thermocycler.

7. Update the status of plates processed for the ARB2 reaction in the database if desired.

3.2.2. PCR Cleanup and Sequencing

1. Use a 12-channel pipette to transfer 5µL of each ARB2 reaction to a new PCR reaction plate on ice.

2. Mix 2µL EXOSAP-IT enzyme mix into the ARB2 reaction mix by pipetting up and down three times.

3. Seal plates with adhesive foil.

4. Move plate from ice to the thermocycler preheated to 37°C.

5. Incubate plates at 37°C for 15 min.

6. Incubate plates at 80°C for 15 min.

7. Update the status of plates subjected to PCR cleanup in the database if desired.

8. Add 13µL freshly diluted sequencing primer at a concentration of 7.69ng/µL to each sample for a final concentration of 5ng/µL. For MAR2xT7, the Tn3 sequencing primer (Fig. 1) PMFLGM.GB-4a (5'-GACCGAGATAGGTTGAGT-3') was used. Store samples at 4°C prior to sequencing.

9. Subject samples to fluorescently labeled dideoxynucleotide chain termination sequencing according to the kit manufacturer’s instructions.

3.2.3. Uploading Sequences into the Relational Database and Sequence Analysis

Data-uploading methods are dependent on the database and software being used. Here we describe a general scheme that should be tailored accordingly.

1. Assign plate and sample names to sequencing data: Assemble ABI sequencing files into sets of 96, each set of 96 in a folder titled with a human-readable plate name that is recognizable by the database and linked to the unique identifier for the culture plate used as template for the sequences (Note 16). Each ABI file in the folder should have a human-readable and computer-parsable name that includes the culture plate name and the well position of the mutant from which the sequence was derived. ABI file names should be recognized by the database so that each ABI file is correctly associated with the unique
identifier for the proper mutant. For the PA14 library, each folder was given a three-digit numeric name that was linked to the unique culture plate database identifier. In the event that multiple sequencing attempts might be necessary to obtain data on all the mutants in a plate, the three-digit plate names were given a version number, each with its own unique database identifier. For example, folder 242v6 represents culture plate number 242, version 6. Sequences from individual wells in this plate are titled 242v6_A01, 242v6_A02, 242v6_A03, and so on.

3.2.4. Insertion Site Identification

Software used for data analysis should process sequence information in the following way to identify the insertion site of each mutant.

1. Assign a quality score to each base in the original sequence. Using the PHRED software application (www.phred.org), PATIMDB assigns a quality score for each base pair in each uploaded sequence or raw sequence.

2. Trim low-quality sequence. Bases with a quality score of less than 20 are trimmed off the raw sequence to produce a processed sequence (Notes 17 and 18).

3. Perform BLAST alignment with genomic sequences. PATIMDB aligns each processed sequence with both the PA14 and PAO1 genome sequences using the BLAST algorithm (10). The assignment of a location of the transposon insertion site in the PA14 genome for a given processed sequence isolated from a particular mutant is based on the region of the genome with the best BLAST score (Note 19).

4. Identify the transposon sequence immediately adjacent to the genomic sequence junction point. The precise location of the insertion site in the region of the BLAST hit is determined automatically using a Smith-Waterman algorithm built into PATIMDB that searches the first 120 bases of the raw sequence for alignment with a 26-base sequence at the end of the MAR2xT7 transposon, allowing up to seven mismatches or gaps (Fig. 2). Parameters

---

![Diagram](image.png)

Fig. 2. Insertion site prediction methodology. PATIMDB (1) scans the raw sequence for the sequence aligning with the transposon sequence using the transposon sequence identification tool and (2) performs BLAST alignments with the genomic sequence. The raw sequence coordinates of the last base to align with the transposon sequence (A) and the first base to align with the genomic sequence (B) are used to determine the transposon insertion site in the genomic sequence. Insertion site (in genomic sequence coordinates) = C−(B−A), where C is the genomic sequence coordinate of the first aligning base. This rule applies regardless of which strand aligns with the raw sequence. In cases when the transposon sequence cannot be identified, A is given a fixed value based on where within the raw sequence the last base of a particular transposon is most frequently observed.
such as the maximum number of transposon sequence mismatches tolerated, how far 3′ the algorithm searches for the transposon junction, and how much of the transposon sequence is used in the search can be optimized for individual needs. Where the sequencing primer anneals with respect to the transposon junction, the sequence quality at the transposon junction and the accuracy of the zero positions of individual reads affect the accuracy of transposon sequence identification in the raw sequence. Setting these parameters requires trade-offs, however. For example, longer transposon junction search sequences require that a sequencing primer that anneals further upstream (in the 5′ direction) from the transposon junction be used, requiring a greater allowance for mismatches but allowing a higher degree of confidence that the determined transposon location is correct. For sequences in which the transposon sequence cannot be identified, a default insertion site is selected based on the observation that in most mutant sequences, the transposon junction point lies a set number of bases into the raw sequence (on average, 63 bp for MAR2xT7). Caution must be used, however. In several instances where no alignment with MAR2xT7 was identified, the transposon sequence was found manually beyond the first 120 bases of the raw sequence, suggesting that the search window was set too narrowly. Using the default location of 63 bases in these cases puts the insertion site more than 63 bases away from the actual insertion site.

3.3. Library Mega-Analysis

3.3.1. Insert Distribution Across the Genome

To determine insertion coverage and to detect the presence of hot spots, the genomic coordinates of all transposon insertions are mapped into 1- or 10-kb bins. The number of mapped insertions in each bin is quantified. This analysis should be carried out routinely during library production to assess saturation of the genome.

3.3.2. Insert Distribution Within Predicted ORFs

Combining insertion site coordinates with the start and stop sites of every ORF in the genome gives the number of times each gene has been hit. This analysis also makes apparent which genes have not been hit while library production is in progress. If the fraction of undisturbed genes that are known to be nonessential in other organisms is high, library production should continue. Once near-saturation has been established, genes that were hit only once should be analyzed further to determine if there are more hits at the extreme 5′ and 3′ ends of these genes than would be predicted if insertions were completely random. An enrichment of hits at the extreme 5′ end of genes hit only once indicates possible transcriptional fusions with transposon-derived sequences. An enrichment of hits in the extreme 3′ end of genes hit only once suggests that the insertion did not disrupt gene function. Genes falling into either category may be essential.

3.3.3. Defining Essential Genes: Comparing Transposon Insertion Mutants in Two Different Strains

In the case of P. aeruginosa, transposon mutation libraries using different transposons have been constructed independently in strains PAO1 ([11] and Chapter 9) and PA14 (1). Between the two libraries, the insertions sites of more than 60,000 transposon mutations have been mapped. By comparing the two libraries, we were able to identify
orthologous genes that were not hit in either library and thereby determine a set of putative essential genes. First, we defined orthologs between the two strains (which we call PA14/PAO1 orthologs), and then the orthologs not hit in either library were compared. Genes not hit in either library were considered putative essential genes. As described in Section 3.3.2, for genes hit only once, the hit distribution was skewed toward the extreme 5' and 3' ends of these genes, indicating that some PA14/PAO1 orthologs hit just once at either end of the coding sequence may be also putative essential genes (Table 2).

1. Identify orthologous genes in two different strains. Download and run the “findOrthologs.pl” program (http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/downloads.cgi). This program generates a list of orthologs based on criteria such as the percentage identity and percentage difference of query length to alignment length. After the program is run, some manual curation of the list may be necessary. As a general rule, reciprocal best hits are selected as orthologs, while attempting to maintain synteny along the genome. The program requires:
   (a) FASTA format file or files containing the two single-contig genome sequences named “1” and “2.”
   (b) FASTA format file or files containing predicted ORFs in both genomes. Each gene must occur only once and have a unique gene identifier (GeneID) in the title line.
   (c) Tab-separated values format file containing the fields “GeneID,” “GenomeID,” and “Start” indicating, respectively, the unique gene identifier, the genome identifier (1 or 2), and the start position of each gene.
   (d) Configuration file that contains cutoffs for percent sequence identity and the maximum difference in the length of individual BLAST query sequences.

2. Compare orthologous genes not hit in either library.
   (a) Generate a list of genes in each library that were not mutated (or not “hit”) by identifying those genes from the total gene set in each genome. With a relational database, this requires a so-called left outer join in database parlance (Note 20).
   (b) Select one of the two mutated genomes as the reference genome. Perform a join of nonmutated genes from the nonreference genome library with the orthologs table (Section 3.3.3, No. 1) to generate a list of reference genome orthologs for these nonmutated genes.
   (c) Join the list of nonmutated genes in the reference genome with the list of reference genome orthologs of the nonmutated genes from the nonreference library (above). The

| Table 2 |
| Comparison of PA14 and PAO1 Orthologs Disrupted in Two Mutant Collections |

<table>
<thead>
<tr>
<th>P. aeruginosa strain PA14 (1)</th>
<th>P. aeruginosa strain PAO1 (11)</th>
<th>PA14/PAO1 orthologs in either library</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted PA14/PAO1 orthologs</td>
<td>Predicted PA14/PAO1 orthologs</td>
<td>Unique insertion locations</td>
</tr>
<tr>
<td>5,102</td>
<td>3,954</td>
<td>5,102</td>
</tr>
<tr>
<td>1,148</td>
<td>335</td>
<td>22,881</td>
</tr>
<tr>
<td>30,100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
intersection of these two sets gives a set of putative essential genes. This set excludes genes that are strain-specific or that lack identified orthologs.

Notes
1. Multiple matings over the course of library production also minimize the number of redundant mutants. Ten different matings were used to produce the PA14 library.
2. In control experiments related to the construction of the *P. aeruginosa* PA14 transposon mutation library, *MAR2xT7* transposants were first apparent at 2h incubation, indicating that with the particular donor cells, transposon, and recipient cells used for these experiments, a 2-h mating time was sufficient to obtain a high frequency of transposition events but minimized the amplification of transposants, thereby reducing the frequency of isolating mutants containing the same mutation. It is important to determine the optimal time and donor:recipient culture ratio for matings with the particular strains to be used prior to scaling up.
3. The transposition frequency for a given mating combination must also be determined so that the number of mating spots that must be harvested to obtain the desired number of transposants is known. For PA14 mated with MC4100/pMAR2xT7 and HB101/pRK2013, three pooled mating spots consistently generated more than 4800 transposants. Transposant colonies approximately 2.5 mm in width can be recognized by the QBot.
4. The antibiotic resistance marker on the backbone of the donor plasmid carrying the transposon serves to verify that only the transposon has inserted into the genome, and integration of the entire donor plasmid has not occurred. In initial experiments, putative transposants should be tested for donor plasmid integration.
5. In the case of *P. aeruginosa* strain PA14, we have observed that liquid cultures grown statically under microaerobic conditions contain a high frequency of so-called phenotypic variants that are resistant to high concentrations of multiple antibiotics (7, 8). These antibiotic-resistant variants (called RSCVs for rough small colony variants) exhibit a variety of transient phenotypic changes in addition to antibiotic resistance including high surface hydrophobicity that results in increased biofilm formation and reduced virulence. We do not know whether RSCV formation is a common feature of all *P. aeruginosa* strains or other bacterial species. Because the frequency of RSCVs increases with high levels of antibiotics, we determined the minimal concentrations of gentamicin and Irgasan required to select for PA14/MAR2xT7 transposants. Gentamicin resistance is encoded by *MAR2xT7*, and PA14 is naturally resistant to Irgasan. Therefore, Irgasan is used to select for *P. aeruginosa* and against *E. coli*.
6. If a QBot is not available, colonies can be picked by hand, but depending on the library size, this could be a daunting task.
7. It is essential to set the QBot picking parameters to ensure that the robot does not mistake two overlapping colonies for a single colony. Look at the trays after the picking run to verify that only individual colonies were disrupted by the pins.
8. Trays can be stored at 4°C and repicked if necessary.
9. As previously described, experiments in our lab with *P. aeruginosa* strain PA14 showed that static long-term culture (more than 16h) greatly increases the frequency of RSCVs in PA14 cultures. However, growing PA14 in deep-well Costar/Corning blocks (600 μL media in 2.0 mL 96-well deep-well blocks) with agitation in a HiGro shaker (Materials) or in a standard plate shaker that can hold blocks for 16h or less prevents RSCV formation and makes the cultures easier to transfer (the cultures tend to have a more uniform consistency).
However, because the QBot can only inoculate low-profile plates, inoculation of deep-well blocks had to be done by hand.

10. In the construction of the PA14 library, an individual rack of transfer tips was used three times to transfer the same cultures in steps 4, 10, 11, and 12. Moreover, to keep costs down, the tips used to add glycerol to culture plates by dispensing glycerol from above the plate into the cultures were saved and reused for 24 different culture plates. Subsequently, however, we found by carrying out quality-control experiments that the tips used to dispense glycerol did become contaminated with PA14. We therefore recommend instead that fresh tips be used to add glycerol to each plate.

11. This example illustrates the importance of quality-control testing to ensure that the liquid-handling method used to aliquot cultures be cross-contaminant free. Using a control plate consisting of some wells inoculated with the mutagenized strain interspersed with many uninoculated (sterile) wells is a simple way to determine the level of potential cross-contamination at each step of the protocol. Store source and destination plates at 37°C for several days to confirm the absence of cell growth in uninoculated wells. Control plate tests of our methods defined several critical parameters essential for minimizing cross-contamination when handling PA14 cultures. These parameters are discussed below.

12. The following guidelines arose from thorough testing of the Biomek FX method used to transfer PA14 cultures.

(a) **Glycerol (to final concentration of 15%) must be added and mixed into PA14NR set cultures before transfer.** Transferring PA14 cultures grown in LB or LB + 15 μg/mL gentamicin for various lengths of time, with or without agitation, to either 96- or 384-well plates, resulted in a high frequency of cross-contamination of both adjacent and nonadjacent wells. This occurred whether the transfer was performed by the Biomek FX robot or by hand using a multichannel pipette. We assume this cross-contamination is the result of aerosols from the tips as they are held over destination plates. We found that the addition of glycerol (final concentration of 15%) prior to transfer greatly decreased the frequency of cross-contamination of wells. Whatever the cause of cross-contamination, it is essential that 15% glycerol be added to cultures to be transferred when making copies of the library. Because glycerol inhibits the PCR reaction, transfer of culture to be used as PCR templates from culture plates should be performed before addition of glycerol. This step, therefore, must be thoroughly tested for well-well cross-contamination.

(b) **Avoid carryover of culture mix on transfer tips.** Culture aspiration and dispensing is prone to drops of culture hanging from tip ends that can easily cross-contaminate wells as the robotic head moves over the plates laid out on the deck. To avoid this, we programmed the Biomek FX to touch the tips to the side of the wells with each aspiration and dispensing step (see below). In addition, all “blow-out” steps were skipped because, in our hands, this formed bubbles of culture/glycerol mix on the ends of the tips, a potential source of contamination as the tips move over the blocks and plates on the deck.

(c) **Library propagation should be performed in a 96-well format.** Even with glycerol addition to cultures prior to transfer, we were unable to inoculate 384-well plates either by hand or robotically without cross-contaminating adjacent wells. This is presumably due to the necessity of touching the tips to the side of the wells both after aspiration of culture in source plates and after dispensing culture in destination plates. The mostly likely reason why the wells in 384-well plates get cross-contaminated is that the well walls are shared between wells. In contrast, well walls in 96-well plates are not shared between adjacent wells.
(d) *Keep culture times to a minimum.* We generally grew *P. aeruginosa* PA14 cultures for 14 h but not more than 16 h. Although some mutants may take longer to get to saturation, they are almost always at a high enough density after 14 h to be used for transfer to copy plates.

(e) *Grow cultures with agitation in deep-well blocks.* Reduced aeration of PA14 cultures greatly increases the incidence of RSCV formation. We have tested cultures grown in a Genomic Solutions, Inc., HiGro shaker/incubator with supplemental O₂ and cultures grown in a regular shaker with no extra O₂ added. Both conditions prevented the appearance of variants.

(f) *Keep the surface-to-volume ratio high, ensuring proper aeration of the culture.* Growing cultures larger than 750 µL in 2.0-mL 96-well culture blocks resulted in the appearance of variants. When growing large cultures, we grew 600-µL cultures in deep-well blocks (catalog no. 7556–9600, USA Scientific, Ocala FL), resulting in a surface-to-volume ratio of 0.92.

(g) *Avoid extensive dilution.* The larger the degree of dilution, the higher the frequency of variants. We assume this is because the increased amount of cell division required to saturate the culture promotes the appearance of variants in the culture. Preliminary tests have shown that inoculating 600-µL media with an average-size wild-type colony picked with a pipette tip is sufficient to prevent the appearance of variants in the saturated culture. When using thawed liquid stocks to inoculate media, we generally diluted 1:50.

(h) *Keep surface area of tip coated with culture/glycerol mix to a minimum.* Even when source cultures contain 15% glycerol, transfer of cultures will result in well-to-well cross-contamination of source and destination plates if the majority of the surface area of the transfer tip is coated with culture/glycerol mix. To avoid this problem, the Biomek FX is programmed to have transfer tips aspirate culture mix no more than 10 mm below the liquid surface.

(i) *Seal plates on the deck of the robot.* The simple act of moving the plates from the robot to the bench top before sealing has led to cross-contamination in test runs. Seal the plates thoroughly on the deck, using a roller to ensure that each well is firmly sealed.

13. QFilled plates can be stored at 4°C overnight before run.

14. Others have modified our protocol, replacing DMSO with 1.25 M Betaine (catalog no. 14300; Fluka, Sigma Aldrich, St. Louis, MO) with excellent sequencing success rates (D. Ewen Cameron and J. Mekalanos, personal communication).

15. Originally, ARB1 PCR was performed using the ARB1 primer ([Table 1](#)), with a success rate of approximately 95%. Over time, the efficiency of sequencing with this primer dropped. We found that many of the products of sequencing were extremely short—only as long as the transposon sequence—suggesting that under the conditions of the particular PCR reaction employed, the ARB1 primer had an affinity for the end of the transposon. Several additional ARB primers were created ([Table 1](#)), including two (ARB1D and ARB1E) in which the defined sequence was changed to be less likely to hybridize to the transposon sequence. ARB1D and ARB1E gave the greatest sequencing success rates (~75%).

16. When using PATIMDB to analyze sequences, each plate folder containing 96 sequences is in turn placed into a folder entitled with the transposon name (2×T7). The transposon folder is placed in a folder titled “PA14.” PATIMDB is launched and “Sequence Analysis” is selected. Each folder is uploaded into PATIMDB individually.

17. We found that a Phred quality score of 20 adequately filtered out poor sequences; however, different values can be used if desired.
18. Because the sequence quality at the very beginning of reads is often poor, it is advisable to have the sequencing primer positioned a sufficient number of base pairs from the transposon junction so that at least 30 bases of the transposon sequence are routinely present in the resulting sequencing reads beyond the region of poor sequence.

19. In most cases for the PA14 library, alignment with one region in the genome had a high BLAST score, whereas other genomic regions had much lower BLAST scores. The difference between the best BLAST score and the second-best BLAST score, which we called the Bit Score Separation, was large, indicating a high confidence that the region with the highest BLAST score was the site of the insertion. Therefore, the position of the insertion was based on that BLAST hit. If, however, more than one region of the genome aligns well with a raw sequence and the Bit Score Separation is zero (i.e., in cases of gene duplications), the insertion location cannot be unambiguously determined.

20. If a relational database is not available, a spreadsheet can be used to compare lists of non-mutated orthologs in each library. Orthologs that are common between the two lists are putative essential genes.

Acknowledgments

Funding was provided by the NIH Heart Lung and Blood Institute HL66678, NIH R01, AI064332, and a grant from the Cystic Fibrosis Foundation.

References


