For the article “An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants,” by Nicole T. Liberati, Jonathan M. Urbach, Sachiko Miyata, Daniel G. Lee, Eliana Drenkard, Gang Wu, Jacinto Villanueva, Tao Wei, and Frederick M. Ausubel, which appeared in issue 8, February 21, 2006, of *Proc Natl Acad Sci USA* (103:2833–2838; first published February 13, 2006; 10.1073/pnas.0511100103), the authors note that on page 2837, left column, beginning on line 10, “Although gacA has been implicated in biofilm formation in PAO1 (37), we observed a very mild PVC-attachment defect for both the PA14NR Set gacA mutant and for a strain carrying a nonpolar deletion of the gacA gene (19),” should instead read: “Although gacA has been implicated in biofilm formation in PAO1 (37), we observed a very mild PVC-attachment defect for both the PA14NR Set gacA mutant and for a strain carrying a transposon insertion in the gacA gene (19, 20, 38).” The related references appear below. This error does not affect the conclusions of the article.


For the article “Structural basis for mRNA and tRNA positioning on the ribosome,” by Veysel Berk, Wen Zhang, Raj D. Pai, and Jamie H. Doudna Cate, which appeared in issue 43, October 24, 2006, of *Proc Natl Acad Sci USA* (103:15830–15834; first published October 12, 2006; 10.1073/pnas.0607541103), the author name Jamie H. Doudna Cate should have appeared as Jamie H. D. Cate. The online version has been corrected. The corrected author line appears below.

Veysel Berk, Wen Zhang, Raj D. Pai, and Jamie H. D. Cate

www.pnas.org/cgi/doi/10.1073/pnas.0610232103
An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants

Nicole T. Liberati*, Jonathan M. Urbach*, Sachiko Miyata, Daniel G. Lee, Eliana Drenkard, Gang Wu, Jacinto Villanueva†, Tao Wei‡, and Frederick M. Ausubel§

Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114

Contributed by Frederick M. Ausubel, December 22, 2005

Random transposon insertion libraries have proven invaluable in studying bacterial genomes. Libraries that approach saturation must be large, with multiple insertions per gene, making comprehensive genome-wide scanning difficult. To facilitate genome-scale study of the opportunistic human pathogen *Pseudomonas aeruginosa* strain PA14, we constructed a nonredundant library of PA14 transposon mutants (the PA14NR Set) in which nonessential PA14 genes are represented by a single transposon insertion chosen from a comprehensive library of insertion mutants. The parental library of PA14 transposon insertion mutants was generated using *MAR2xT7*, a transposon compatible with transposon-site hybridization and based on *mariner*. The transposon-site hybridization genetic footprinting feature broadens the utility of the library by allowing pooled *MAR2xT7* mutants to be individually tracked under different experimental conditions. A public, internet-accessible database (the PA14 Transposon Insertion Mutant Database, http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/home.cgi) was developed to facilitate construction, distribution, and use of the PA14NR Set. The usefulness of the PA14NR Set in genome-wide scanning for phenotypic mutants was validated in a screen for attachment to abiotic surfaces. Comparison of the genes disrupted in the PA14 transposon insertion library with an independently constructed insertion library in *P. aeruginosa* strain PAO1 provides an estimate of the number of *P. aeruginosa* essential genes.

It is estimated that even in the most-studied organisms, the functions of 30–50% of the genes remain unknown (1). Various strategies have been used to define gene function on a genomic scale, including the assemblage of genome-wide deletion (2–4) or insertion (5–12) mutant libraries or RNA interference libraries (13–15). These approaches allow immediate correlation of a mutant phenotype with a specific gene.

Sequencing the insertion sites of random transposon insertions has been used to assess gene function in a variety of bacterial species (6–12). To ensure saturation, transposon insertion libraries typically consist of multiple insertion alleles in each gene. To streamline genome-wide scanning of the *Pseudomonas aeruginosa* chromosome, we created a nonredundant library of *P. aeruginosa* strain PA14 transposon insertion mutants in which a single mutant has been preselected to represent a particular nonessential gene. Previously, a different *P. aeruginosa* strain (PAO1) was subjected to transposon mutagenesis and a saturating number of insertions sites was sequenced (11). There are several reasons why a genome-wide set of insertion mutations in strain PA14 is valuable in addition to the set in PAO1. First, in contrast to PAO1, PA14 is a primary clinical isolate that has not been passaged in the laboratory. Second, PA14 is a multithagen pathogen that is virulent in a variety of mammalian and nonvertebrate hosts (16–20), and PA14 genes that are absent in PAO1 are known to contribute to its enhanced pathogenicity (21, 22). Third, the PAO1 library was generated using a derivative of the bacterial transposon Tn5, whereas the PA14 library was constructed with a derivative of a eukaryotic *mariner* transposon. Using different transposons to create the two libraries minimizes untargeted gaps due to insertion-site specificities and allows an accurate estimate of the number of essential *P. aeruginosa* genes. Finally, assembling a publicly available nonredundant subset of the PA14 library makes it particularly well suited for carrying out a variety of genome-wide phenotypic screens.

*MAR2xT7*, a derivative of the *mariner* family transposon Himar1 (23, 24), which transposes in both prokaryotic and eukaryotic genomes and exhibits minimal insertion-site specificity, was used to generate most of the PA14 mutants. The 38,976 mutants in the PA14 collection (see Tables 3 and 4, which are published as supporting information on the PNAS web site), containing multiple insertions in most nonessential genes, are described in a public, internet-accessible database [the PA14 Transposon Insertion Mutant Database (PATIMDB), http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/home.cgi]. Version 1.0 of the PA14 nonredundant (PA14NR) set consists of 5,459 mutants selected from the 38,976 member parental library. PA14NR Set mutants were subjected to rigorous quality control procedures, including manual single-colony purification. Using a high-throughput screen for attachment to polyvinylchloride (PVC), we show here that the PA14NR Set is as effective a tool to quickly scan the *P. aeruginosa* chromosome and assign gene function.

### Results and Discussion

#### Library Production

We generated a library of random transposon insertion mutations in *P. aeruginosa* strain PA14 as described in *Methods*; in Supporting Methods, which is published as supporting information on the PNAS web site; and at http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/home.cgi. The majority of PA14 library mutants were created by using *MAR2xT7*, a derivative of the *mariner* family transposon Himar1 (Fig. L4) (23, 24). The Himar1 transposase is located outside the *MAR2xT7* coding sequence on the suicide vector used to deliver *MAR2xT7*, preventing genomic integration of the transposase and subsequent *MAR2xT7* transposition to secondary sites. *MAR2xT7* was constructed to allow genetic footprinting of the mutants with transposon-site hybridization (TraSH) analysis (25–27), a method that facilitates the tracking of mutants in a pool under different experimental conditions. *TraSH* promoters placed at both ends of *MAR2xT7* allow each *MAR2xT7* mutant to be uniquely identified by means of a PCR-amplified genomic sequence adjacent to the transposon. The *TraSH* promoters are then used to direct transcription from these PCR products, and the resulting RNA is used to create probes for hybridization to a *P. aeruginosa* DNA microarray to identify the amplified genomic fragments (and therefore the individual mutants) present in the...
Table 1. Summary of PA14/MAR2xT7 library construction results

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrayed</td>
<td>34,176</td>
</tr>
<tr>
<td>PCR-processed and sequenced</td>
<td>30,336</td>
</tr>
<tr>
<td>Processed with high-quality sequence</td>
<td>25,035</td>
</tr>
<tr>
<td>With sequences having BLAST hits in PA14</td>
<td>24,089</td>
</tr>
<tr>
<td>With insertions in PA14 genes</td>
<td>18,977</td>
</tr>
<tr>
<td>With insertions between PA14 genes</td>
<td>5,111</td>
</tr>
<tr>
<td>Insertion locations</td>
<td></td>
</tr>
<tr>
<td>Mapped</td>
<td>24,089</td>
</tr>
<tr>
<td>Unique</td>
<td>20,530</td>
</tr>
<tr>
<td>PA14 genes</td>
<td></td>
</tr>
<tr>
<td>Total predicted</td>
<td>5,962</td>
</tr>
<tr>
<td>Hit internally</td>
<td>4,469</td>
</tr>
<tr>
<td>Not hit</td>
<td>1,493</td>
</tr>
<tr>
<td>Average hits per gene</td>
<td>4.3</td>
</tr>
</tbody>
</table>

As expected, large genes tended to have a higher frequency of insertion than relatively short genes (Fig. 4). The fact that a few genes have a very large number of hits probably reflects a combi-
nation of insertion-site bias and stochastic variability. The distribution of insertion sites within individual genes was relatively random, with the exception of genes that were hit only once (Fig. 5A and Table 6). For these genes, there was a preponderance of insertions at the 3' ends, consistent with the possibility that many of these genes encode essential gene products and that the insertions near the ends of the genes did not disrupt gene function.

Unexpectedly, there was also an enrichment of insertion sites at the 5' ends of genes that were only hit once. The locations of some insertions may have been miscalculated because of poor sequence data, or the translational start sites of some genes may have been miscalculated (such that some insertions occur outside of the ORF). However, it is also possible that transcriptional fusions of the MAR2xT7 sequence with PA14 coding sequences could lead to expression of functional proteins if alternative in-frame start codons are available for translation. In support of the latter explanation, among the insertions near the 5' ends of genes, but not the 3' ends, there is an enrichment of mutants in which Strand A of MAR2xT7 is oriented in the same direction as the coding sequence of the disrupted gene (Figs. 1 and 5B and Table 6). Based on this observation, additional mutants that correspond to genes that are currently represented by this class of mutants need to be included in future releases of the PA14NR Set. Regardless of the location of the insertion within a particular gene, it is possible that a MAR2xT7 insertion in which Strand A is oriented with the coding sequence may not be polar on downstream genes. Conversely, insertions in which Strand B is oriented with the coding sequence are likely to exhibit polar effects on downstream genes.

**Candidate Essential Gene Analysis.** Genome sequence analysis of PA14 carried out in our laboratory shows that, although PA14 and PAO1 share >95% identity, PA14 has a slightly larger chromosome (6,53 megabases versus 6,26 megabases; http://ausubellab.mgh.harvard.edu/pa14sequencing) that encodes 5,962 predicted genes, 392 more than the 5,570 predicted for PAO1. Comparison of the PA14 insertions described here with PAO1 genes targeted in ref. 11 allows a relatively precise estimation of the number of essential *P. aeruginosa* genes. Between the two insertion libraries, >60,000 *P. aeruginosa* mutants have been generated with defined transposon insertion sites. If we consider only PA14 genes that have homologs in PAO1, which we refer to as “PA14/PAO1” orthologs [PAO1 Genome Annotation Project (29) and http://ausubellab.mgh.harvard.edu/pa14sequencing], 608 PA14/PAO1 orthologs were not disrupted in the PAO1 library, and 1,148 PA14/PAO1 orthologs were not disrupted in the PA14 library (Table 2). Table 5 lists 335 *P. aeruginosa* candidate essential genes not disrupted in either library.

Some genes in Table 5 may actually be nonessential but were not targeted in either the PAO1 or PA14 libraries because they are small or are located in transposition cold spots or in an operon upstream of an essential gene. In contrast, the list may be missing essential genes. As discussed above, genes disrupted very few times with insertion sites only at the extreme ends may actually be essential (Table 6). Furthermore, genes with redundant essential gene functions would be missed by this analysis. A probabilistic calculation of 60,000 random insertions over 6.5 megabases (approximate number of mutations in both the PA14 and PAO1 libraries and approximate size of the *P. aeruginosa* genome) showed that a gene 327 bp in length has a 95% chance of being disrupted (Fig. 8A, which is published as supporting information on the PNAS web site). Fig. 8B shows that, of the putative 335 essential genes, 22% are shorter than 327 bp, whereas only 9% of all genes in the genome are shorter than 327 bp. Assuming that short genes do not constitute a disproportionately large fraction of all essential genes, these data suggest that approximately half of candidate essential genes shorter than 327 bp were probably not disrupted in the PAO1 or PA14 libraries simply because of their small size rather than because they are essential genes.

A different way of assessing the number of genes that should have been disrupted assuming a random insertion distribution is illustrated in Fig. 6, which shows a skewing of the observed length of gaps between MAR2xT7 insertions in the PA14 genome from a model of gap sizes based on a random Monte Carlo simulation of 26,534 insertions (the approximate number of MAR2xT7 insertion sites sequenced). This simulation showed that for a random library, gaps larger than 2.3 kb would be rare. In fact, there are >160 MAR2xT7 gaps larger than 2.3 kb (one as large as 14 kb). Table 7 shows that many PA14/PAO1 orthologs are found in gaps in both the PAO1 and PA14 insertion libraries, suggesting that they are essential. In contrast, other PA14/PAO1 orthologs located in PA14 gaps are targeted in the PAO1 library, suggesting that these gaps likely reflect MAR2xT7 cold spots.

**PA14NR Set Production.** The PA14NR Set, a subset of the parental PA14 MAR2xT7 transposon insertion library, was created to expedite genome-scale screening. The PA14NR Set is a collection of mutants in which each PA14 gene targeted by MAR2xT7 is represented by a single insertion mutant (or in some cases two mutants, see Supporting Methods). A total of 5,459 mutants were included in the PA14NR Set, which correspond to 4,596 predicted PA14 genes (77% of all predicted PA14 genes). An automated prioritization scheme was used to choose the PA14NR Set. MAR2xT7 mutants in the wild-type background were selected in preference to TnphoA

---

**Fig. 3.** MAR2xT7 insertion-site distribution. The number of insertion sites in every 10 kb of PA14 genomic sequence. The upper set of data points represents insertions in which MAR2xT7 Strand B runs 5' to 3' with the top strand of the PA14 chromosome. The lower set of data points denotes MAR2xT7 insertions oriented in the opposite direction. Mb, megabase.

**Fig. 4.** Frequency of insertions within genes. The left y axis shows the number of genes disrupted once, twice, etc. (solid trace). The right y axis shows the average length of the genes at each insertion frequency (dashed trace).
mutants or mutants in *exoU* or *exoUspcU* backgrounds. Mutants with more 5′/H11032 insertions were chosen over mutants with insertion sites further downstream in the same gene. For a complete description of the selection process, see *Methods* and Tables 3 and 4.

Selected mutants were colony-purified to ensure that the PA14NR Set is free of cross-contaminants and to keep phenotypic variant subpopulation sizes to a minimum. The enrichment of phenotypic small colony variants (SCVs) during the construction of the library was of particular concern because we observed that PA14 SCVs, which have properties characteristic of phenotypic variants in other bacterial species (30), grow considerably faster than the wild-type strain under static (microaerobic) conditions. Culture conditions were optimized during subsequent steps to minimize the growth of SCVs. In addition, multiple precautions were taken to ensure contamination-free transfer of cultures to storage plates and to plates designated for dissemination of the library to other laboratories (see Supporting Methods). To assess the integrity of the PA14NR Set library, the insertion sites of *MAR2xT7* in 109 random PA14NR Set mutants was determined by arbitrary PCR sequencing. This analysis showed that 106 of the 109 clones contained the expected insertion. Reporting mislabeled mutant clones will allow us to revise the PA14NR Set catalog to reflect these discrepancies.

Details concerning the construction and features of the PA14NR Set can be found at http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/home.cgi. A transposon insertion map of all PA14 transposon insertions can be found at http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/tmap.cgi. The map allows users to search for specific mutant sequence and BLAST alignments and visually scan predicted PA14 genes for insertion locations.

### Screen for PVC-Attachment Mutants

To assess the utility of the PA14NR Set for high-throughput functional screens, the entire

---

**Table 2. Summary of the PA14 and PAO1 mutant libraries**

<table>
<thead>
<tr>
<th>Strain, study</th>
<th>Predicted PA14/PAO1 orthologs hit</th>
<th>PA14/PAO1 orthologs not hit</th>
<th>PA14/PAO1 orthologs not hit in either library</th>
<th>Unique insertion locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA14*, this study</td>
<td>5,102</td>
<td>3,954</td>
<td>1,148</td>
<td>22,881</td>
</tr>
<tr>
<td>PAO1, ref. 11</td>
<td>5,102</td>
<td>4,494</td>
<td>608</td>
<td>30,100</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>335</td>
</tr>
</tbody>
</table>

*All backgrounds and transposons (see Methods and Table 3).*

---

Fig. 5. Insertion site distribution relative to *MAR2xT7* insertion position within each gene. (A) The fraction of all mutants (black bar) and the fraction of mutants carrying an insertion in a gene that was disrupted only once (hatched bar) are shown for each insertion site position as a percentage of gene length in base pairs. Fractions are based on either the sum of all mutants or the sum of all mutants with insertions in genes disrupted only one time. The fraction of total or single mutants containing an insertion in the first 5% of the gene length is represented by the 5% category. (B) The fraction of mutants carrying an insertion in a gene that was disrupted only once in the library at each gene position (black bars), the fraction of mutants with single gene insertions in which *MAR2xT7* Strand A is oriented in the same direction as the coding sequence (hatched bars), and the fraction of mutants with single gene insertions with *MAR2xT7* Strand B oriented in the same direction as the coding sequence (gray bars) are shown. Fractions are based on the sum of all mutants, the sum of all mutants with Strand A insertions in genes disrupted only once, or the sum of all mutants with Strand B insertions in genes disrupted only once.
PA14NR Set (5,459 mutants) was screened for attachment to PVC plastic, a phenotype shown to require several *P. aeruginosa* genes (31). Because attachment to PVC plastic is correlated with the ability to form biofilm (31, 32), mutants with altered PVC attachment profiles may form biofilm improperly. A total of 416 PA14NR Set mutants with a PVC attachment phenotype were identified in the primary screen, including insertions in *pilC, rpoN, algR, clpP, crc, fleR, filP, sadB, sadA*, and *sadR*, which had previously been shown to be required for PVC attachment and biofilm formation (Fig. 6) (31–36). Although *gacA* has been implicated in biofilm formation in PAO1 (37), we observed a very mild PVC-attachment defect for both the PA14NR Set *gacA* mutant and for a strain carrying a nonpolar deletion of the *gacA* gene (19). The discrepancy between the PAO1 and PA14-phenotypes may be due to differences in assay conditions or strain backgrounds. In addition to the genes discussed here, the PA14NR Set screen for attachment to PVC also identified many genes without previously reported attachment phenotypes (data not shown). Together, these findings validate screening the PA14NR Set as an efficient way to scan the genome for genes without previously reported attachment phenotypes.

The screen of the PA14NR Set failed to identify *pilB* and *flgK*, genes previously shown to contribute to attachment to PVC. In contrast, three other *pilB* mutants and three other *flgK* mutants from the parental MAR2xT7 library showed clear PVC attachment deficiencies. Sequencing the PA14NR Set *flgK* mutant revealed that it carried a transposon insertion in a different gene. As discussed above, we expect that 2.8% of the mutants in the PA14NR Set are mislabeled. In the case of the *pilB* nonredundant set mutant, the location of this insertion was predicted with an insertion-site default above, we expect that 2.8% of the mutants in the PA14NR Set are..

The PA14 transposon insertion library (and its accompanying database, PATIMDB) is a powerful complement to the previously described PAO1 transposon library. The availability of both libraries allows verification of phenotypes by examining strains with the orthologous (or comparable) gene disrupted in both collections. An alternate transposon (*the mariner-based MAR2xT7* in PA14 rather than the Tn5-based *IshpoA/A* and *IsacZ/Z* in PAO1) was chosen to increase the chances of generating insertions in loci not represented in the PAO1 library due to transpositional cold spots. Indeed, approximately half (273 of 608) of the PA14/PAO1 orthologs not represented in the PAO1 library were disrupted in the PA14 collection. By combining information from both PAO1 and PA14 insertion libraries, we have arrived at a list of 355 putative essential genes in *P. aeruginosa*.

**Methods**

**Bacterial Strains.** Transposon insertion mutants were generated in wild-type *P. aeruginosa* strain PA14 (19) and in two PA14 derivatives, *ΔexoU* and *ΔexoUspcU*. PA14 *ΔexoU* contains a 2-kb deletion of *exoU*, but the deletion is not in frame, and a newly generated stop codon may alter expression of the downstream gene *spcU* (28). PA14 *ΔexoUspcU* contains a 2.41-kb in-frame deletion encompassing the adjacent PA14 genes *exoU* and *spcU*, which are in the same operon.

**MAR2xT7.** The majority of the PA14 mutants were created with the TraSH-compatible transposon *MAR2xT7*, an engineered derivative of the Himar1 transposon (23, 24) carried on the suicide plasmid pMAR2xT7 that was propagated in the *pir* Escherichia coli strain MC4100.

**Transposon Mutagenesis, Colony Selection, and Work Flow.** *MAR2xT7* insertions were generated by introducing pMAR2xT7 into PA14 or *exoU* PA14 derivatives from *E. coli* MC4100 in 10 separate tripartite matings, selecting for transposants on 20-× 20-cm LB agar plates containing 15 μg/ml gentamicin and 1 μg/ml Irgasan, and robotically picking putative transposants into 250 μl of LB containing 15 μg/ml gentamicin in 96-well microtiter plates. *TphoA* insertions were generated by mating PA14 and *E. coli* SM10pir carrying the suicide vector pRT731 (38), selecting transposants on LB agar containing 200 μg/ml neomycin and 100 μg/ml Irgasan, and picking transposants into 250 μl of LB containing 200 μg/ml kanamycin and 50 μg/ml Irgasan. Aliquots of the putative *MAR2xT7* and *TphoA* transformant cultures were transferred robotically to 96-well microtiter plates for arbitrary PCR, sequencing, and storage in 15% glycerol. Fig. 9, which is published as supporting information on the PNAS web site, shows a flow diagram indicating how putative insertion mutant clones were.
cataloged and partitioned into various microtiter plates during the process of library construction.

Transposon Insertion Site Identification. Transposon insertion sites were identified using a two-round PCR protocol (39) that involved lysing cells at 95°C, amplifying sequence adjacent to a transposon insertion with a transposon-specific primer and an arbitrary primer, followed by a second amplification using a nested transposon-specific primer and a primer corresponding to a nonrandom portion of the arbitrary primer used in the first PCR. A third nested transposon-specific primer was used for sequencing reactions.

PATIMDB. PATIMDB, which carried out process tracking and automated sequence analysis, was implemented using the MUSL relational database managing system hosted on a multiprocessor Intel system running Red Hat (Raleigh, NC) LINUX. PATIMDB can be accessed online at http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/home.cgi. A map of all identified transposon insertions in the PA14 chromosome is available at http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/tmap.cgi.

Statistical Analysis of Transposon Distribution. We created 26,534 simulated theoretical transposon insertions by using a random number generator. Simulated gap sizes, grouped in bins of 200 bp, were measured by counting the number of bases between insertion locations. The simulation was repeated 200 times, and bin totals were averaged, giving the theoretical distribution of gap sizes.

Probabilistic Calculation of Insertion Likelihood as a Function of Gene Size. Assuming a random distribution of transposon insertions, the probability of getting at least one insertion in a gene of length g given a genome of size N and a library containing n mutants is p (one or more insertions given n mutants) = 1 − (1 − 1/g)^n.

PA14/PAO1 Orthogon Assignment. PA14/PAO1 orthologs were picked by using an automated Perl script that performed reciprocal BLAST alignments of PA14 and PAO1 protein sequences. Orthologs were required to have the same amino acid length within 30% and to have at least 70% of the amino acid sequence length align with a minimum of 70% identity across the aligned sequence. In the case of multiple high-scoring hits, orthologs were assigned to maintain synteny between the PA14 and PAO1 genomes. Cases of redundancy due to gene duplication were resolved manually.

PA14NR Set Selection and Production. Mutants were prioritized as follows for inclusion in the PA14NR Set. First, priority was given to insertions with BLAST scores of >80. Second, priority was given to PA14/MAR2xT7insertions and PA14/ TphoA over exoU/MAR2xT7 or exoUspcU/MAR2xT7 insertions. Third, priority was given to more 5’ insertions. Fourth, if all other criteria were equal, the mutant with the higher BLAST score received priority. In cases for which there were mutants available in the PA14 background but a more 5’ mutant was available in either the exoU or exoUspcU background, both mutants were included in the set. Mutants selected for inclusion in the PA14NR Set were picked manually from “working plates” (see Fig. 9) and were colony-purified by streaking onto LB agar containing 15 μg/ml gentamicin.

Small colony variants were avoided except in cases where nonviables were not available (<20 mutants). PA14NR Set members were grown in deep-well microtiter plates in a HiGro incubator (Genomic Solutions, Ann Arbor, MI) with O2 injection, and a Biomek FX (Beckman Coulter) liquid-handling robot was used to transfer cultures between microtiter plates using a specially designed protocol that minimized cross-well contamination.

PVC Attachment Screen. The PVC plastic attachment assay was carried out essentially as described in ref. 31. Briefly, cultures were grown statically at 37°C in M63 media containing 1% casamino acids/0.3% glucose/0.5 mM MgSO4/0.025% vitamin BL. The plates were stained with 1% crystal violet for 10 min after removing the media and scored by eye. The absorbance of control samples at 550 nm was also recorded.

Detailed Methods. Detailed descriptions of the methods used to construct and analyze the library, assemble the PA14NR Set, and carry out the PVC attachment screen can be found in Supporting Methods and at http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/productionmethods.cgi.

Note Added in Proof. We plan to add ~400 additional mutants to the PA14NR Set to represent genes that are currently represented in the PA14NR Set by mutants that, as previously mentioned, may produce transcriptional fusions and mutants with miss-called insertion sites.

We thank Tara Holmes and Kalyani Gumpta (Automation Core Facility, Department of Molecular Biology, Massachusetts General Hospital) for help with colony picking and culture transfer and G. O’Toole and S. Lory for helpful discussions. This work was supported by National Heart, Lung, and Blood Institute Grant U01 HL66678 and by a grant from the Cystic Fibrosis Foundation.

Supporting Methods

**Bacterial Strains.** Transposon insertion mutants were generated in wild type *Pseudomonas aeruginosa* strain PA14 (1) and in two PA14 derivatives, ΔexoU and ΔexoUspcU. The rationale for using an exoU background was to decrease the cytotoxicity of PA14 (2). PA14 ΔexoU contains a 2-kb deletion of exoU, but the deletion is not in-frame, and a newly generated stop codon may alter the expression pattern of the immediate downstream gene spcU (2). An in-frame deletion encompassing the adjacent PA14 genes exoU and spcU, which are in the same operon, was constructed by replacing 4.29 kb of wild-type sequence with a 1.88-kb PCR-amplified fragment that contained a 2.41-kb deletion. The following oligonucleotide primers, based on *P. aeruginosa* strain PAO1 sequence data, were used to generate the exoUspcU deletion. ExoU111-Sacl: 5'-CTGGAGCTCGAGTATCGACGCTGTAGCTAACG-3'; ExoU22-XbaI: 5'-TAACGCCTGTTTCTAGATTTGATATGCATGTTCGCTC-3'; SpcU3-XbaI: 5'-CATATCCAATCTAGAACCAGCGCGTTAGTGTTCG-3'; SpcU4-HindIII: 5'-GTCAAGCTTATGGTGCAGACCGTCCAGGC-3'. The PCR-amplified fragment containing the deletion was subcloned into the SacI and HindIII sites of pEX18Ap (3) generating plasmid pEX18exoUspcU8, which was subsequently used to introduce the deletion into the wild-type PA14 genome by homologous recombination as described (3).

**Transposons.** The vast majority of the PA14 mutants were created with MAR2xT7, an engineered derivative of the Himar1 transposon (4, 5). pMAR2xT7, a plasmid carrying MAR2xT7, was derived from pMFLGM.GB. pMFLGM.GB (obtained from R. Vance and J. Mekalanos, Harvard Medical School, Boston) was constructed by inserting the linker acgcgtgaagttcctatactttctagagaagttcctatctactttctatcctatactttcagagaagttcctatcctactacgtag (which contains FRT sites and SpeI, NheI, NcoI and SacII restriction sites) into the MluI site in pFAC (6). The gentamicin-resistance cassette from pBRR1MCS-5 (6) was excised with NcoI and SacII and cloned into the modified pFAC to create pMFLGM.GB.
To create pMAR2xT7, the primers 5'-
CATCGATCGTCTAGTAACAGGTGCTGATAAGTCCCCGTTCTCTAGACCCTA
TAGTGAGTCTATTACGCCGGCAGCCGCGGGAC-3' and 5'-
GTCCCCGCGGCGCCGCTATACGACTCACAATATAGGGCTAGTCTAGAGACCAGG
ACTTATACGCAAACCTGTACTAGACGATCGATG-3' were annealed, digested with
PvuI, extended with Taq polymerase, digested with SacII, and ligated to replace the
SacII/PvuII fragment in pMFLGM.GB. The primers encode a T7 promoter and an
adjacent inverted repeat (IR) sequence. The sequence encoding a T7 promoter and its
adjacent IR sequence was amplified from pMycoMar (5) by using the following
oligonucleotides: 5'-CATAGCTAGCCGCGGGACCGAGATAGGGTTGAGTG-3' and
5'-TCCCCCAGGTTCTTACTAGACGATCGATG-3'. The NheI-digested PCR fragment
was inserted into the NheI site of the modified pMFLGM.GB to create pMAR2xT7,
which was propagated in the pir+ Escherichia coli strain MC4100. A map and sequence
of pMAR2xT7 are available at http://ausubellab.mgh.harvard.edu/cgi-
bin/pa14/downloads.cgi.

Transposon Mutagenesis, Colony Selection, and Work Flow. Tripartite matings were
performed using saturated cultures of PA14 or PA14 derivatives, E. coli
MC4100/pMAR2xT7, and the helper strain E. coli HB101/pRK2013 (7, 8). 1:2:2 mixtures
of the cultures, respectively, were dropped on Kings’ B media plates (2% wt/vol peptone,
6.57 mM K₂HP₀₄, 6.08 mM MgSO₄, 1% vol/vol glycerol) (8, 9) and incubated at 37°C
for 2 h. Independent matings were spread on a single rectangular 20- × 20-cm LB agar
plate containing 15 µg/ml gentamicin and 1 µg/ml Irgasan at a density of ≈1 cfu/cm².
After incubation at 37°C for 12-15 h, colonies were robotically picked with a Qbot
(Genetix, Boston) and grown statically in 250 µl of LB containing 15 µg/ml gentamicin
in 96-well microtiter plates for ≈40 h at 37°C. For TnphoA matings, saturated cultures of
PA14 and E. coli SM10λpir carrying pRT731 were mixed in a 1:1 ratio, dropped onto
King’s B media plates, and incubated for 6 h at 37°C (10). Independent matings were
collected, resuspended in 100 mM MgSO₄, and plated onto a rectangular LB agar plate as
described above containing 200 µg/ml neomycin and 100 µg/ml Irgasan. Colonies were
picked robotically as above into 250 µl of LB containing 200 µg/ml kanamycin and 50
µg/ml Irgasan and grown at 37°C statically for ≈40 h. Seventy microliters of the putative MAR2xT7 and TnphoA transformant cultures were transferred robotically to 96-well microtiter plates for arbitrary PCR and sequencing (see below). Glycerol was added to the cultures before transfer to long-term storage plates at a final concentration of 15%.

Fig. 9 shows a flow diagram indicating how putative insertion mutant clones were cataloged and partitioned into various microtiter plates during the process of library construction.

**Transposon Insertion Site Identification.** Transposon insertion sites were identified by using a two-round arbitrary PCR protocol (11). Aliquots (70 µl) of the statically grown transformant cultures (see above) were incubated at 95°C for 10 min to lyse the cells and spun at 3,000 rpm for 5 min. Of the cleared lysate, 3 µl was used as template for the first round of arbitrary PCR (ARB 1). ARB 1 PCR mix contained 1× Taq Buffer (Roche), 10% DMSO, 2.5 µM dNTPs, 1.25 units of Taq polymerase (Roche), and 1.0 ng of each primer per microliter of mix. For MAR2xT7 mutants, the transposon-specific primer, PMFLGM.GB-3a, (5'-TACAGTTTACGAACCGAACAGGC-3') was used. For TnphoA mutants, Tn5Ext (5'-GAACGTTACCATGTTAGGAGGTC-3') was used as the transposon-specific primer. Although different ARB primers were used at different times during library production to maximize the efficiency of obtaining useful sequence data, the ARB 1 primer used most frequently was ARB 1D (5'-GGCCAGGCCTGCAGATGATGNNNNNNNGTAT-3'). Plates were sealed with adhesive foil (Alum-1000, Diversified Biotech). After initial denaturation at 95°C for 5 min, ARB 1 plates (Costar catalog no. 6511, Corning Incorporated) were cycled 30 times at 95°C for 30 s, 47°C for 45 s, and 72°C for 1 min. Plates were incubated 5 min at 72°C to allow for extension of PCR products.

For the second round of arbitrary PCR, 5 µl of the ARB 1 reaction was used as template. ARB 2 reaction mix was the same as the ARB 1 reaction except that nested transposon-specific primers and a nested ARB 2 primer were used. The PMFLGM.GB-2a primer (5'-TGCTCAACTGGGTTGTCGCTTCATCCG-3') was used for MAR2xT7 mutants, and the Tn5Int2 primer (5'-GGAGGTCACATGGAAGTCAGATCCTGG-3') was used for
TnphoA mutants, respectively, as transposon-specific primers. The ARB 2 primer, ARB 2A (5’-GGCCAGGCCTGCAGATGATG-3’) was used for both MAR2xT7 and TnphoA mutants. Plates were sealed with adhesive foil. Reactions were cycled 40 times: 95°C for 30 s, 45°C for 30 s, and 72°C for 1 min with a 5-min extension at 72°C.

For PCR cleanup, 5 µl of ARB 2 reaction was mixed with 2 µl of EXOSAP-IT enzyme mix (catalog no. 78205, USB) in a new reaction plate (catalog no. AB-0800, Abgene). Plates were sealed with adhesive foil and incubated according to the manufacturer’s instructions. Sequencing primer was added directly to each PCR cleanup for a final concentration of 5 ng/µl. The PMFLGM.GB-4a primer (5’-GACCGAGATAGGGTTGAGTG-3’) was used as the sequencing primer for MAR2xT7 mutants and the Tn5Int primer (CGGGAAAGGTCTGGTCCAGGACGC-3’) was used for TnphoA mutants.

The PA14 Transposon Insertion Mutant Database (PATIMDB). The PATIMDB was implemented using the MYSQL relational database managing system hosted on a multiprocessor Intel system running Red Hat (Raleigh, NC) LINUX. The data-entry application was written in JAVA and runs on WINDOWS 2000. This application implemented process tracking and an automated sequence analysis pipeline. Sequences in the form of ABI files were imported into the application, and base-calling was performed using PHRED. The resulting raw sequence was trimmed to select only high-quality sequence, which was then compared by BLAST (12) to the PA14 genome to identify the insertion site of the transposon and the identity of the disrupted ORF. The data-retrieval system was implemented using a PERL-based common gateway interface (CGI) web application hosted on a multiprocessor Intel system running Red Hat LINUX with APACHE. The functions of the data-retrieval web application include queries and the downloading of data from PATIMDB over the web, such as a list of mutants with identified insertion locations. PATIMDB is compatible with different genome sequences and is adaptable to library construction applications in other organisms. Extensive quality assurance testing was performed on PATIMDB, the data-input application, and the data-retrieval application, to ensure that the database accurately related and processed each
mutant DNA sequence file. PATIMDB can be accessed at http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/home.cgi.

A map of all identified transposon insertions in the PA14 chromosome is also available at http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/tnmap.cgi. The transposon insertion map is a web-based application written with PERL CGI and JAVASCRIPT and is deployed on the same server as the data-retrieval system. This database-driven application accesses PATIMDB via PERL database interface. To give users the best viewing experience, the application dynamically detects user screen resolution to decide the map size that can best fit the screen. Because of the large amount of data that has to be processed to generate the map, when a map window contains data for 50 kb or more, the application performance can be degraded. We used a server-side cache technique to pregenerate all large-window maps, so users can quickly retrieve maps of all sizes. Extensive quality assurance has been done to ensure that all maps display the correct data. A genome navigation bar (at the top of the screen) can locate any genome location with one click. An integrated search tool (below the map) helps users find specific transposon insertions. Two buttons on each side of the page allow users to scan nearby genome locations. With multiple viewing options and data filters, users can select data of interest. Links to information such as the genetic background of the mutant carrying each displayed insertion and gene and mutant reports are also incorporated into the transposon insertion map application.

**Statistical Analysis of Transposon Insertion Distribution.** Gaps (in base pairs) between transposon insertions in the PA14 genome were identified by recording insertion locations along the 6,534,396 base pairs in the PA14 genome. Gap sizes were measured by counting the number of bases between insertion locations. The number of gaps of a given size, or in bins of a given size range, was then tabulated.

A theoretical genome array the same length as the actual genome size was used in a Monte Carlo prediction of random insertion events. We created 26,534 simulated transposon insertions by using a random number generator, and insertion locations were
recorded. Simulated gap sizes were then measured as with the actual data by counting the number of bases between insertion locations. Gap sizes were grouped in bins of 200 bp. The simulation was repeated 200 times and bin totals were averaged giving the theoretical distribution of gap sizes. Gaps larger than 2,500 bp were identified.

**Probabilistic Calculation of Insertion likelihood as a function of Gene Size.**
Assuming a random distribution of transposon insertions, the probability of getting at least 1 insertion in a gene of length $l$ given a genome of size $g$ and a library containing $n$ mutants is:

$$p(\text{one or more insertions given } n \text{ mutants}) = 1 - (1 - (l/g))^n$$

Therefore, a gene 327 bp in length has a 95% likelihood of insertion given a genome size of 6.5 megabases and library size of 60,000 insertions.

**PA14/PAO1 Ortholog Assignment.** PA14/PAO1 orthologs were picked using an automated PERL script that performed reciprocal BLAST alignments of PA14 and PAO1 protein sequences. Orthologs were required to have the same amino acid length within 30% and to have at least 70% of the amino acid sequence length align with a minimum of 70% identity across the aligned sequence. In the case of multiple high-scoring hits, orthologs were assigned to maintain synteny between the PA14 and PAO1 genomes. Cases of redundancy due to gene duplication were resolved manually.

**PA14NR Set Selection and Production.** *Mutant selection criteria.* Selection of mutants for the nonredundant set was automated by creating a hierarchical set of priorities. Mutants were prioritized as follows. (i) The mutants corresponding to a particular gene were divided into two groups: those with BLAST bit scores less than or equal to 80 and those with scores of >80. Priority was given to those with BLAST scores of >80. (ii) Each of these two groups was further subdivided by strain background (PA14; *exoU; exoUspcU*) and transposon (*MAR2xT7; TnphoA*), with priority given to PA14/*MAR2xT7* over PA14/TnphoA and PA14/TnphoA over *exoU/MAR2xT7* or *exoUspcU/MAR2xT7.*
Each of these six groups was ordered by distance of the transposon insertion site from the start of the gene, with priority given to more 5' insertions. A 200-bp handicap was given to sequences that exhibited evidence of contamination based on the sequencing. Finally, the mutants were ordered by BLAST bit score such that if all other criteria were equal the mutant with the higher BLAST score would have priority. In cases where there was a mutant available in the PA14 background but a more 5' mutant was available in either the exoU or exoUspcU background, both mutants were included in the PA14NR Set.

**Colony purification.** Mutants that were selected for inclusion in the PA14NR Set were picked manually from frozen 96-well “working plates” (see Fig. 9) on dry ice and streaked onto LB agar containing 15 µg/ml gentamicin. Plates were incubated at 37°C for 14–16 h, cooled to room temperature, and stored at 4°C for no more than 8 days before picking. Colonies were examined under a dissecting scope. A single colony with a morphology representative of the majority of colonies for a given mutant was used to inoculate 600 µl of LB containing 15 µg/ml gentamicin in a 96-well deep-well block (USA Scientific, Ocala, FL). Small colony variants were avoided except in cases where nonvariants were not available (<20 mutants).

**Culture conditions.** Master deep-well blocks were incubated in a HiGro incubator/shaker (Genomic Solutions, Ann Arbor, MI) set to 37°C with O₂ injection for 2 s every 5 min of incubation for 14–16 h. Plates were set at room temperature for 0–6 h before transferring an aliquot of each culture to master blocks (see Fig. 9).

**Biomek transfer to master plates.** A Biomek FX (Beckman Coulter) liquid-handling robot was used to transfer cultures between microtiter plates. We found that unless glycerol (15%) is added to PA14 cultures before transfer, both manual and robotic culture transfer is prone to cross-contamination, possibly because of aerosolization of the cultures. Even with the inclusion of glycerol in transferred cultures, we were unable to define a methodology for transfer to 384-well plates that did not result in cross-contamination of adjacent wells. For this reason, the PA14NR Set has been arranged in a
96-well format. Exhaustive testing of the following protocol demonstrated that during test runs, no cross-well contamination was observed. We transferred 200 µl of sterile 60% glycerol from a reservoir to each deep-well master block and mixed each three times by pipetting 200 µl each time up and down. Tips were touched to the side of the wells to remove liquid clinging to the end of the tips and then discarded. Fresh tips were used to transfer 40 µl of the mixed cultures to each of 10 96-well master plates containing 160 µl LB with 15 µg/ml gentamicin and 15% glycerol. The Biomek FX was programmed to pipet culture 1 mm from the top of the culture for each transfer to minimize the surface area of the transfer tip coated with culture. Transferred cultures were dispensed by dropping the culture slowly into the recipient wells without blowing air out of the tips and without mixing. Tips were dipped into the culture/media mix in the recipient wells, touched to the side of the well and removed. Master blocks and plates were sealed (Alumina Seals, Diversified Biotech, Boston) and stored at –80°C immediately after a set of 10 plates were transferred and sealed.

**Quality Control.** Initially, Southern blotting was used to confirm a single insertion in 10 independent MAR2xT7 mutants. Further confirmation of single MAR2xT7 insertions was obtained from the sequence derived from the ARB PCR product of each mutant. The vast majority of mutants displayed clear, easily readable sequences rather than mixed sequences that would be expected if MAR2xT7 were prone to multiple insertion events.

Extensive quality controls were put in place to prevent cross-contamination of colony-purified cultures and to avoid propagation of variant subpopulations during construction of the PA14NR Set. A full description of the techniques used and a “users manual” for the PA14NR Set are available at http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/home.cgi.

From the PA14NR Set, 120 random clones were resubjected to arbitrary PCR and sequencing to check whether expected mutants were located in the appropriate well positions. Readable sequence was isolated from 109 clones. Of these, all but three
sequences (2.8%) were the same as the sequences assigned by PATIMDB to the corresponding well positions.

**PVC Attachment Screen.** Testing the PA14NR Set for attachment to polyvinylchloride (PVC) plastic was carried out essentially as described in ref. 13. Briefly, a sterilized, metal, 96-pin head was used to inoculate LB with stored, frozen cultures from the PA14NR Set. After 16 h at 37°C with shaking at 250 rpm (Innova4400, New Brunswick Scientific), cultures were diluted 1:10 in M63 media containing 1% casamino acids, 0.3% glucose, 0.5 mM MgSO₄, and 0.025% vitamin B1. These 1:10 dilutions were diluted again in the same medium in PVC plates for a final dilution of 1:100. Diluted cultures (1:10 and 1:100) were grown statically for 8 h at 37°C. Media was removed from the PVC plates by pipeting, and the plates were stained with 1% crystal violet for 10 min. After several destaining rinses in water, stained rings at the air–liquid interface were scored by eye. A subset of rings was dissolved in 95% ethanol, and absorbance at 550 nm was recorded. After 8 h of static growth at 37°C, the optical density of cultures grown from the 1:10 dilutions were read at a wavelength of 600 nm using a Spectra Max Plus plate reader (Molecular Devices, Sunnyvale, CA).


### PA14 Parental Transposon Insertion Library Production Flow Chart

<table>
<thead>
<tr>
<th>-80°C Storage</th>
<th>ARB PCR</th>
<th>Culture (wor) (280 µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-80°C Storage</td>
<td>ARB PCR</td>
<td>Add glycerol Mix and Seal</td>
</tr>
<tr>
<td>Master (mas)</td>
<td>ARB PCR</td>
<td>Supernatant (sup)</td>
</tr>
<tr>
<td>Duplicate (dup)</td>
<td>ARB PCR</td>
<td>Add glycerol Mix and Seal</td>
</tr>
<tr>
<td>Working (wor) Leftover culture/Glycerol Mix</td>
<td>ARB PCR</td>
<td>Add glycerol Mix and Seal</td>
</tr>
</tbody>
</table>

### PA14NR Set Production Flow Chart

1. **PA14 Tn Library plate (wor)**
   - Streak PA14NR Set clones onto LB + Antibiotic
   - Manually pick 5,459 single/(non-SCV) colonies into 600 µL LB + antibiotic
   - Grow in HiGro 16 hours
   - Robotically add glycerol (15%) and mix
   - Transfer 40 µL glycerol/culture mix to 10 storage plates containing 160 µL LB+antibiotic+glycerol (15%)- store-80°C
   - Inoculate media in deep well blocks with thawed culture from storage plates.
   - Grow and transfer to storage plates as described above.

2. **~60 Master blocks**
3. **Storage plates (10 copies)**
4. **PA14NR Set copies for Distribution**