Supporting Information

O'Connor et al. 10.1073/pnas.1111678108

SI Materials and Methods

Construction of the *E. coli* Strain DH5α λpir F′. *E. coli* strain DH5α λpir (1) was grown in liquid LB at 37 °C overnight and plated for single colonies on LB containing 50 μg/mL streptomycin to isolate a spontaneous streptomycin resistance strain. The F plasmid was mated from *E. coli* strain XL1 Blue to the DH5α λpir Str^R strain by coplating on LB media at 37 °C overnight. Bacteria were then harvested and plated on LB containing streptomycin and tetracycline to select for *E. coli* DH5α λpir (str^R) F^+ (tet^R) exconjugants.

Microarray Construction. Primer pairs for amplifying each protein encoding gene of *L. pneumophila* Philadelphia 1 (2) were designed as described by Sassetti et al. (3) and purchased from Invitrogen. PCR products for 2,864 genes were defined ranging from 62 to 357 bp, with an average length of 240 bp. PCR amplification and microarray printing were performed by the Whitehead Institute Center for Microarray Technology (see below). PCR products were generated for 2,768 reactions (96.6% success rate). The 96 amplicons not generated included genes in the *lvh* locus (*lpg1228–lpg1271*), which is deleted in strain Lp02. Eight of the PCR reactions contained multiple products. Genes not represented on the array are listed in Dataset S2. Of the 78 genes for which primer pairs could not be designed, 70-mer oligonucleotides were designed for those predicted to encode proteins involved in virulence (Dataset S2) using a similar method to define the PCR amplicons. For increased specificity, 70-mer oligonucleotides for *lpg0234*, *lpg1934*, *lpg2157*, *lpg2370*, and *lpg2509* were also designed. These 13 oligonucleotides were spotted on the array in addition to the gene-specific PCR products.

PCR Amplification and Microarray Printing. Forward and reverse primers with 5′ universal extension sequences, GAACCGATAGGCATCTAGAGCAC and GAAATCCACCGCACTAGTCCTC, respectively, were used to amplify each gene product in the following reaction: 10 pmol of forward and reverse primer, 50 ng of genomic DNA, 250 mM dNTP mix, 2.5 mM $MgCl_2$, 1× PCR buffer (Applied Biosystems), 10 μL of 5 M betaine, and 1 unit of AmpliTaq DNA polymerase (Applied Biosystems) in a 50 μL final volume using the following cycling conditions: 95 °C for 2 min and 36 cycles of 92 °C for 25 s, 51 °C for 30 s, 72 °C for 1 min, and then 72 °C for 7 min. PCR product amplification efficiency was analyzed by separating 10 μL of each PCR in a 2% (wt/vol) eGel (Invitrogen): Results indicated that PCR products were generated for 2,768 of 2,864 reactions (96.6% success rate). Next, a second round of PCR amplification was performed using 2 μL of the first-round PCR and the universal primers GAACCGATAGGCATCTAGAGCAC and GAAATCCACCGCACTAGTCCTC, which are 5′ amino-modified with a six-carbon linker. The PCR included 10 pmol of forward and reverse universal primers, 250 mM dNTP mix, 2.5 mM $MgCl_2$, 1× PCR buffer (Applied Biosystems), and 1 unit of AmpliTaq DNA polymerase (Applied Biosystems) in a 100 μL final volume. PCR products were amplified using the following reaction conditions: 94 °C for 3 min and 25 cycles of 94 °C for 30 s, 65 °C for 30 s, 72 °C for 2 min, and then 72 °C for 10 min. PCR product amplification for a subset of reactions was analyzed by gel electrophoresis as described above, revealing production of single products of the expected size for all reactions tested.

The PCR products were purified using Multiscreen GV purification plates (Millipore) following the manufacturer's instructions and then dried in a hood at room temperature overnight and resuspended in 50 μL of water. The resuspended DNA was transferred to flat-bottomed polystyrene 384-well plates (Corning), dried, and resuspended in 15 μL of printing buffer (3× SSC, 1.5 M betaine). The DNA was printed on UltraGAPS-coated slides (Corning) using an Omnigrid-100 arraying robot (Gene Machines) and Stealth type 3 pins (Telechem). After arraying, the printed slides were rehydrated with a handheld steamer (Conair). DNA was fixed by UV cross-linking with a Stratalinker (Stratagene) set at 60 mJ. The slides were then stored under vacuum at room temperature.

Construction of *L. pneumophila* Mariner *himar*-1–Based Transposon. The *mariner himar*-1–based transposon plasmid was constructed as follows: fxTn3 and fxTn4 (all oligonucleotide sequences are provided in Table S1) were annealed to generate a fragment consisting of a 5′-T_7 promoter element (T_7)-*himar*-1 recognition sequence (*himar*1)-TA dinucleotide (TA)-3′ flanked by SmaI and BamHI cohesive ends at the 5′ and 3′ ends, respectively, and cloned into SmaI-BamHI digested pUC19 (New England Biolabs). Next, the R6K origin of replication and kanamycin resistance cassette fragment from pSC189 (4) were amplified by PCR using primers kanR6Kfor and kanR6Krev and blunt-end ligated into the SmaI site, such that a SmaI restriction site was regenerated at the 5′ end. The resulting plasmid was digested with SmaI and XmnI and ligated to similarly digested pSDfx (pSDfx is a derivative of pSD12 (Eric Rubin, Harvard University, Boston, MA), a pUC-based vector containing a repaired 5′-T_7-*himar*1-TA-3′ SmaI-XmnI fragment). The resulting plasmid, pfxTn2, contained the *himar*-1–based transposon: 5′-TA-*himar*1-T_7-R6K origin of replication-kanamycin resistance cassette-T_7-*himar*1-TA-3′. Next, *sacB* from pSR47s (5) was amplified by PCR using primers sacB1 and sacB2 and cloned as a SacI fragment into the SacI site of pfxTn2. The resulting plasmid was digested with BamHI and ligated to the BamHI fragment of pSC189, which contains an ampicillin resistance cassette and the hyperactive C9 transposase (6) under the control of the constitutive *lac* promoter to generate pTO100, the *Legionella himar*-1 transposon plasmid.

TraSH Screen. ***Preparation of labeled TraSH in vitro probes and control genomic DNA for microarray analysis.*** In vitro probes for array analysis were generated and fluorescently labeled using an adapted protocol from Sassetti et al. (7). To generate labeled control genomic probe, genomic DNA was isolated from *L. pneumophila* Philadelphia 1 strain Lp02 using a Qiagen DNeasy kit. Three micrograms of genomic DNA was combined with 3 μg of Random Primer 6 (New England Biolabs), heated at 99 °C for 10 min, cooled on ice, and then amplified using Klenow (4.5 units; New England Biolabs) in the presence of 0.1–0.15 mM Cy3-dCTP (GE Healthcare) and 0.06 mM dATP/dGTP/dTTP in 1× Buffer 2 (New England Biolabs). Reactions were incubated at 37 °C for 3 h, and DNA was then purified using a Qiagen PCR purification kit. The amount of labeled in vitro and genomic probe was quantified as described by Sassetti et al. (7).

Microarray preparation and hybridization. Microarray slides were incubated in prehybridization solution [25% formamide, 5× SSC, 0.1% SDS, and 0.004% fraction of V BSA (Sigma)] at 42 °C for 45 min, rinsed with ddH_2O, and then dried by centrifugation at 346 × *g* for 2 min. Five picomoles of Cy-labeled genomic and in vitro probe (based on picomoles of Cy dye incorporated) was resuspended in hybridization solution [25% formamide, 5× SSC, 0.1% SDS, 0.8 μg/mL yeast tRNA (Invitrogen), and 2 mg/mL acetylated BSA (Invitrogen)], heated to 95 °C for 5 min, and

then applied to the microarray. Hybridization was performed under a coverslip in a Corning hybridization chamber immersed in a 42 °C water bath for 14–16 h. Microarrays were then washed in 2× SSC, 0.1% SDS at 42 °C for 5 min, 0.1× SSC, 0.1% SDS at room temperature for 10 min, and three times in 0.1× SSC for 1 min and then dried by centrifugation.

Microarray data analysis. Arrays were scanned using a Scanarray Microarray scanner (GSI Lumonics) and quantified using Imagene 7.0 software (BioDiscovery, Inc.). Data from individual arrays were first normalized by total intensity in each channel (insertion to genomic probe). Insertion probe intensities for each spot were then normalized to the ratio of the median value of all data points within 1 SD of the mean of the insertion probe dataset relative to the genomic probe dataset. The insertion/genomic probe intensity ratio for each gene was then calculated. Data from individual arrays were normalized, such that the mean population centered around 1; the ratios for each gene from 10 individual experiments were then averaged, and the SD was calculated. All genes with a ratio <3.8 based on the skewing of the histogram plot of the $\log_{10}$ (insertion/genomic probe intensity) (Fig. 1) from a normal distribution and whose ratio was statistically different from the mean population (all values within 0.5 SD of the mean) based on a two-tailed Student *t* test *P* value <0.001 were defined as essential for growth in rich media. All data analysis was performed using Microsoft Excel and GraphPad Prism 4 software.

Genome Sequencing. DNA samples were prepared as follows: Genomic DNA (5 μg) was sheared under pressure using a nebulizer (Invitrogen) and purified using a PCR purification kit (Qiagen). DNA fragments were blunt-ended using an End-It DNA repair kit (Epicentre) and purified as above. DNA was then A-tailed using Klenow Exo-minus (New England Biolabs) and purified using a Qiaquick gel extraction kit (Qiagen). Strain-specific adapters were generated by annealing ggat_1 and ggat_2 and agct_1 and agct_2, respectively, and then ligated to A-tailed DNA using a Fast Link kit (Epicentre). Samples were separated in 2% (wt/vol) agarose, and 450- to 550-bp fragments were gel-extracted using a Qiaquick gel extraction kit. Adapter-ligated fragments were amplified by PCR using PfuUltra II Fusion HS DNA polymerase (Stratagene) and universal primers D139 and D140 and then purified as above. Coverage of the genome was verified by TOPO-cloning (Invitrogen) an aliquot of the library and sequencing 20 individual isolates. Genomic sequencing was performed using an Illumina Genome Analyzer II at the Tufts Core Facility (Tufts University School of Medicine). Sequence analysis was performed using CLC genomics workbench software (CLC bio). Genome sequences of the hextuple 2q and 3a mutants were deposited in GenBank.

1. Kolter R, Inuzuka M, Helinski DR (1978) Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. *Cell* 15:1199–1208.
2. Chien M, et al. (2004) The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* 305:1966–1968.
3. Sassetti CM, Boyd DH, Rubin EJ (2001) Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci USA* 98:12712–12717.
4. Chiang SL, Rubin EJ (2002) Construction of a mariner-based transposon for epitope-tagging and genomic targeting. *Gene* 296:179–185.
5. Merriam JJ, Mathur R, Maxfield-Boumil R, Isberg RR (1997) Analysis of the *Legionella pneumophila flil* gene: Intracellular growth of a defined mutant defective for flagellum biosynthesis. *Infect Immun* 65:2497–2501.
6. Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM (1999) Hyperactive transposase mutants of the Himar1 mariner transposon. *Proc Natl Acad Sci USA* 96: 11428–11433.
7. Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48:77–84.

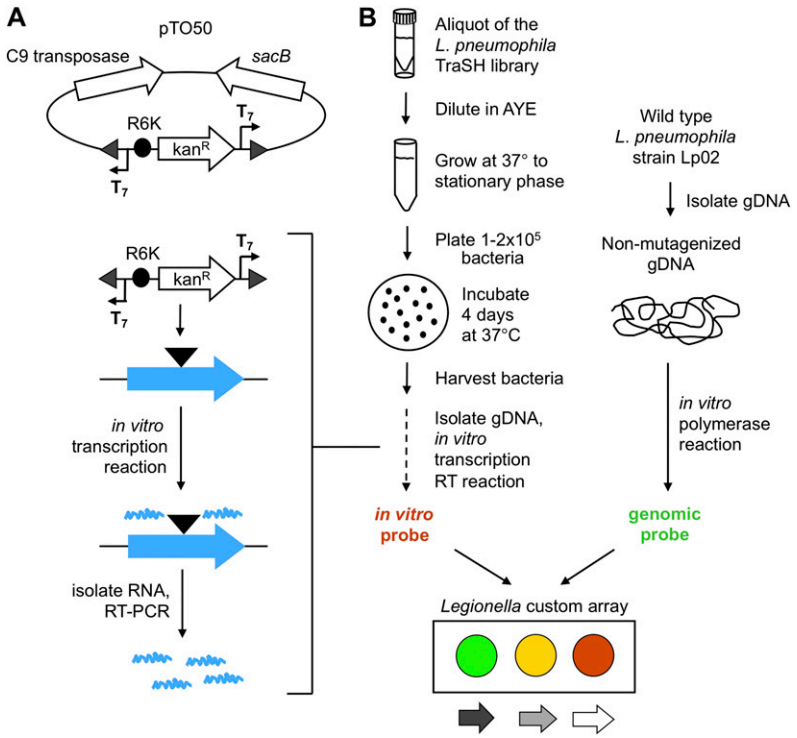

Fig. S1. Genetic screen to identify genes essential for growth of *L. pneumophila* in nutrient-rich bacteriological media. (*A*) Map of the *Legionella* TraSH transposon plasmid, pTO100. (*B*) Genetic screen to identify genes essential for survival and growth of *L. pneumophila* in vitro using TraSH. gDNA, genomic DNA.

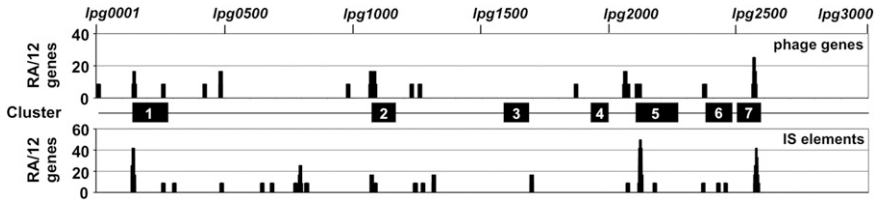

Fig. S2. Mapping of IS elements and phage-related genes relative to the seven nonessential gene clusters. Positional mapping of phage-related genes (*Upper*) and IS elements (*Lower*) along the *L. pneumophila* chromosome relative to the seven nonessential genomic islands (black boxes labeled 1–7), plotted as the running average of the percentage of phage-related genes or IS elements per 12 genes. IS, insertion sequence; RA, running average.

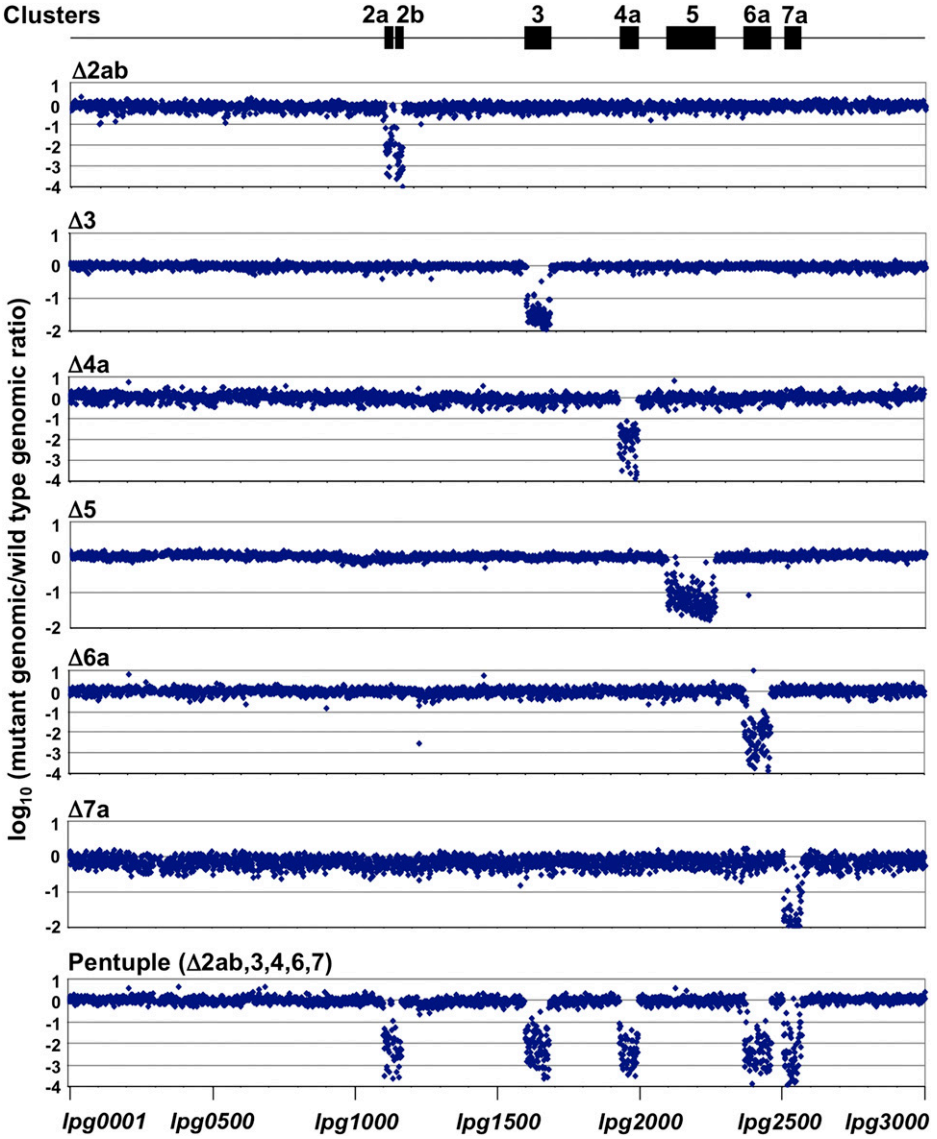

Fig. S3. Microarray analysis to verify genome composition of cluster deletion mutant strains by microarray analysis. Genomic DNA from each cluster deletion mutant and the WT strain was used as a template in individual DNA polymerase reactions using random oligonucleotides and either Cy3-dCTP (WT) or Cy5-dCTP (mutant). For pairwise comparison of each mutant against the WT strain, differentially labeled DNA samples were simultaneously hybridized to an *L. pneumophila* custom microarray. The presence or absence of each gene was determined by comparing the relative intensity of each probe at each spot. The $\log_{10}$ (spot intensity ratio) for each gene for representative mutants is shown. Array hybridization and analyses were performed as described in *SI Materials and Methods*.

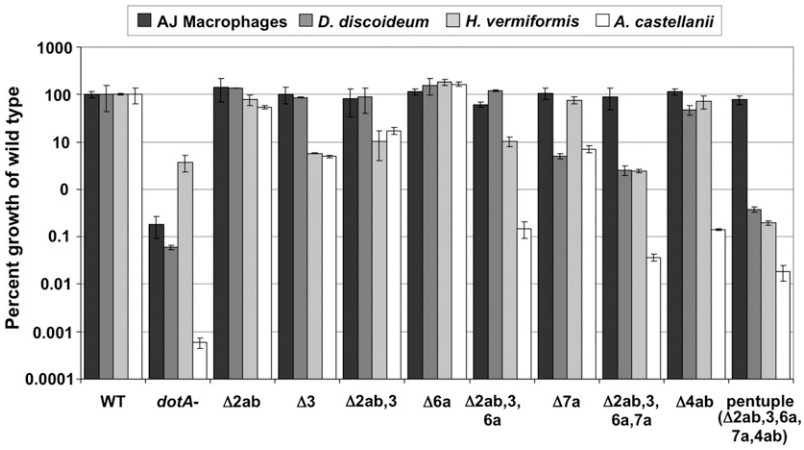

Fig. S4. Intracellular growth of single and combination cluster deletion mutants in natural hosts of *L. pneumophila*. The growth of the WT, the *dotA*$^{-}$ translocation deficient, and various single and combination cluster deletion strains in bone marrow-derived A/J macrophages, *D. discoideum*, *H. vermiformis*, and *A. castellanii* was monitored over 3–4 d and is presented as the percent of growth of each mutant relative to the WT strain based on the total average number of cfus recovered on solid media from lysed cells. Data represent the average of two to three independent experiments of three technical replicates each; error bars indicate ± SD.

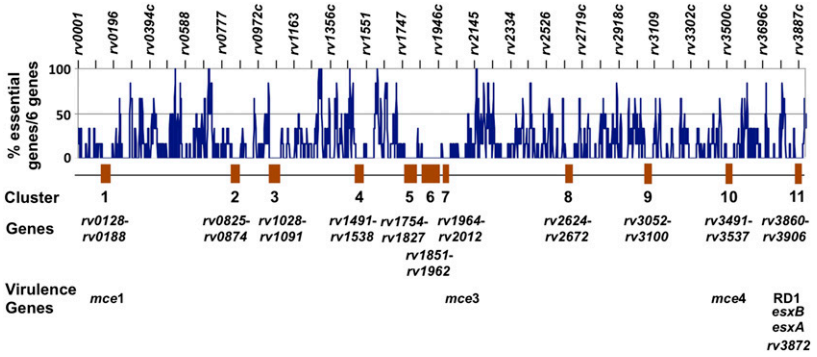

Fig. S5. Nonessential gene clusters in *M. tuberculosis* H37Rv. When the 614 essential genes identified by Rubin and colleagues (1) for growth in rich media were mapped to their position along the chromosome, 11 genomic loci consisting of more than 47 genes were identified dispersed throughout the genome. Based on the size of the genome (4,039 ORFs) and the average number of genes separating consecutive essential genes, the probability of a cluster of 47 genes occurring randomly is less than 1%. Cluster boundaries and virulence-associated *mce* (2, 3), PE/PPE (Proline-Glutamate/Proline-Proline-Glutamate) protein (4) encoding, and RD1 (5) gene loci are indicated.

1. Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48:77–84.
2. Gioffré A, et al. (2005) Mutation in mce operons attenuates Mycobacterium tuberculosis virulence. *Microbes Infect* 7:325–334.
3. Senaratne RH, et al. (2008) *Mycobacterium tuberculosis* strains disrupted in *mce3* and *mce4* operons are attenuated in mice. *J Med Microbiol* 57:164–170.
4. Ramakrishnan L, Federspiel NA, Falkow S (2000) Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. *Science* 288:1436–1439.
5. Brodin P, et al. (2006) Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. *Infect Immun* 74:88–98.

Table S1. Strains, plasmids, and oligonucleotides

Table S1 (DOC)

Dataset S1. *L. pneumophila* genes for which mutations cause a growth disadvantage in nutrient-rich media

Dataset S1 (XLS)

Dataset S2. *L. pneumophila* genes not represented on the custom *Legionella* microarray

Dataset S2 (XLS)