Supporting Information

Arambula et al. 10.1073/pnas.1301366110

SI Materials and Methods

Bacterial Strains, Growth, and Mutant Construction. Legionella pneumophila Corby was a kind gift from Natalia Kozak (Centers for Disease Control, Bethesda). L. pneumophila Corby and derivatives were routinely maintained in culture in yeast extract (PYG) broth or on buffered charcoal-yeast extract (BCYE) media as previously described (1). In-frame deletions and substitution mutations were constructed using allelic exchange with the levansucrase (sucB) negative selection marker on BCYE agar containing 7.5% (wt/vol) sucrose (2). L. pneumophila Corby gene loci targeted for mutational analysis included twin-arginine translocation B (tatB), legionella determinant target A (ldtA) (LpC_1853), accessory variability determinant (avd) (LpC_1854), reverse transcriptase (RT) (LpC_1855), and intergenic regions between avd. LdtA containing stem/loops variable repeat (VR) or template repeat (TR), respectively. The broad host vector pMMB208 was used for complementation of mutants and protein overexpression, and cells were grown in media supplemented with 5 μg/mL chloramphenicol (Cm) (3).

Plasmid Construction and Mutagenesis. Plasmid vector pATR (and derivative donor plasmids) for diversity-generating retroelement (DGR) homing assays was constructed by cloning sequences extending from 75 bp upstream of LpC_1854 to the stop codon of LpC_1855 into the broad host range vector pMMB208 and inserting a 20-bp GC tag into position 98 of TR; all donor plasmids are derivatives of this initial pATR construct. The catalytically inactive RT derivative (Fig. 2) contains a mutation replacing the essential RT amino acid motif YVDD with SMAA (4). The initiation of mutagenic homing (IMH+) deletion derivative is missing sequences from TR position 108–140. The avd mutant carries a deletion that removes all sequences between the avd start and stop codons. For protein expression studies, macropage infectivity potentiator (MIP), defective organelle trafficking A (dotA), intracellular multiplexation X (icmX), recombinase A (recA), or ldtA were fused at their 3′ ends to sequences encoding hemagglutinin (HA) epitope tags, cloned into pMMB208 under the control of the tac promoter, and induced for protein expression by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM.

PCR-Based DGR Homing Assays. Homing assays were performed as previously described with minor modifications (4, 5). In short, L. pneumophila Corby cells harboring pATR, or mutant derivatives, were subcultured in PYG broth supplemented with Cm to an OD590 of 0.2. Cells were grown for 4 h and induced for DGR component expression by the addition of IPTG to a final concentration of 1 mM for 4 h. Cells were harvested and DNA extracted by a commercial kit (Qiagen). PCR was used to detect transfer of the invariant tag from donor plasmid TR or chromosomal TR sequences to chromosomal VR sequences using the following primer pairs: P1-GCGGCAATGGACGGATGAGC, P2-ACAGGAACACAAACGCAGAC, P3-GTCTGCGTACGGGAAAGATTTC, and P5-CATGATTCTGGCTTGCCTA. Amplified products were cloned (TOPO;Invitrogen) and sequenced to verify transfer of the tag from TR to VR and to detect adenine mutagenesis.

PCR Screen for DGR Genes. To screen L. pneumophila strains for putative DGR genes, a series of degenerative primers were designed using previously described methods (6). The first iteration of primers was used as follows: 16SFor-AGCATKGTCTAGC-TTGCTAG, 16SRev-TCCTCCCACTGGAAGATG, avdFor-TGTTTGGATGTAACGAAAGATTTC, avdRev-C CGGTAC-CTGCTTGCCA, RTFor-AAATCATCGACGTAACGACCATA, and RTRev-CTTCTGTAGCCGTGTGGTCG.

Sequencing of L. pneumophila Strains. The genomes of L. pneumophila strains D5549, D5572, and D5591 were assembled and sequenced on an Illumina platform as previously described (7). Reads were assembled into contigs using Assembler and viewed with Artemis.

Protease Sensitivity Assays and Immunofluorescence Microscopy. L. pneumophila cells expressing genes of interest for localization studies were subcultured in PYG to an OD590 of 0.2, grown for 4 h, and induced for protein expression by the addition of IPTG to a final concentration of 1 mM for 4 h. Cells were harvested by centrifugation and washed with PBS (pH 7.3) supplemented with 5 mM MgCl2. Cells were normalized by OD and treated with 0, 50, 100, or 200 mg/mL of proteinase K (Sigma) for 1 h at room temperature. Cells were harvested by centrifugation and washed with PBS + 5 mM MgCl2. Cell pellets were solubilized; proteins were separated by SDS/PAGE; and proteins of interest were detected by Western blot using anti-mycotylocytosmatosis (Myx) (Covance), anti-HA (Abcam), or anti-L. pneumophila (Abcam) antibodies.

For whole-cell immunofluorescence microscopy, aliquots of untreated or proteinase K-treated cells washed with PBS + 5 mM MgCl2 were dropped onto gelatinized glass slides and incubated at room temperature for 30 min. Unbound cells were removed, and attached cells were fixed with 100 μL PBS + 4% paraformaldehyde for 30 min at room temperature. Fixed cells were washed twice with PBS and then blocked by addition of 100 μL PBS + 5% BSA for 1 h at room temperature. Once blocked, cells were incubated with anti-Myc, anti-HA, or anti-L. pneumophila (1:250) antisera in 100 μL PBS + 5% BSA for 1 h at room temperature. Cells were washed three times in PBS and then incubated in 100 μL PBS + 5% BSA with species-appropriate fluorochrome-conjugated secondary antibody for 1 h at room temperature. Cells were washed three times with PBS, and coverslips then mounted by addition of VectaShield (Vector Laboratories) and sealed with nail polish. Microscopy was performed using an AXI10 Imager Z-10 (Carl Zeiss), and analysis performed using AxioVision.

Isolation of Membrane Fractions and Acrlybiotin-Exchange Chemistry. Isolation of L. pneumophila Corby membrane proteins was performed as previously described (8) using 100-mL cultures and a discontinuous sucrose cushion from 30% to 60% sucrose density. Membrane fractions of L. pneumophila Corby cells expressing wild-type or mutant LdtA were isolated as above, and posttranslational modification of the lipobox cysteine determined through acrybiotin-exchange chemistry as previously described with minor modifications (9). Briefly, membrane proteins were suspended in PBS + protease inhibitor/5 mM EDTA/1% Triton X-100 buffer (buffer 1), which was supplemented with N-ethylmaleimide to a final concentration of 50 mM for 30 min at 4 °C. Proteins were precipitated with methanol/chloroform on ice for 5 min. Precipitated proteins were washed and suspended in 100 μL of PBS + 5 mM EDTA supplemented with hydroxylamine, when indicated, and biotin–1-biotinamido-4-[4-(maleimidomethyl)cyclohexane-carboxamido] (biotin-BMCC; Pierce) then incubated at 4 °C for 1 h. Proteins were precipitated with...
methanol/chloroform on ice for 5 min and then suspended in buffer 1. Biotin-labeled proteins were purified by column affinity purification using streptavidin beads, washed, eluted by boiling in protein running buffer, and analyzed by Western blotting.

**Protein Structure Prediction.** LdtA, LdtB, and LdtC were subjected to 3D structure prediction using the algorithm in Phyre 2 (10). In all cases, no predictions were made for the N-terminal regions of these proteins, but predictions with high confidence were made for the C-terminal ~150 residues. These predictions indicated that the C-terminal domains of LdtA, LdtB, and LdtC have C-type lectin (CLec) folds. The highest-scoring templates in each case were a *Bacteroides ovatus* hypothetical protein (BACOVA_04982; PDB ID code 4EPS) and *Escherichia coli* intimin (PDB ID code 1E5U) (11). The latter has been characterized as CLec-fold proteins; inspection of the former indicates that it has a CLec fold as well. For LdtA, the confidence level to BACOVA_04982 and intimin was 99.9% and 94.0%, respectively; for LdtB, it was 99.7% and 92.7%, respectively; and for LdtC, it was 99.5% and 86.6%, respectively. The template for modeling the structures of the C-terminal domains of LdtA, LdtB, and LdtC was BACOVA_04982 in each case. The LdtA model has a 99% confidence level (except at eight residues), the LdtB model a 99% confidence level (except at eight residues), and the LdtC model a 99% confidence level (except at 10 residues).

Fig. S1. Predicted structures of the C-terminal domains of LdtA, LdtB, and LdtC in ribbon representation. The α-helices (red), β-strands (blue), loops (gray), and the locations of VR residues are indicated. The core secondary structure elements (the paired β1β5 strands, the connecting α1 and α2 helices, and the β2β3β4 sheet) of the CLec fold are labeled. Other secondary structures may form the inserts often found in CLec folds.

Mutagenic homing by the *L. pneumophila* Corby DGR. (A) Sequence analysis of products from the PCR-based mutagenic homing assays shown in Fig. 2C, lane 2, reveals mutagenesis at VR positions corresponding to adenines in the cognate TR. TR-tagged indicates the sequence of TR containing the invariant tag, and VR indicates the original, nonmutagenized VR sequence. VR1–4 are independent clones of progeny VRs from a single mutagenic homing assay. (B) The invariant tag from Fig. 2B was knocked into the genomic TR in wild-type Corby cells generating *L. pneumophila* Corby TR with G+C residues (TR-GC). Cells were grown in rich media, genomic DNA extracted, and probed by PCR for transfer of the tag from TR to VR. Bands (P1+P3) observed with wild-type samples contain mutagenic homing products as shown in C, and bands appearing in ΔRT samples are a result of template switching (detected due to the increased number of cycles) and do not represent mutagenic homing products (verified by sequencing PCR products). (C) Sequencing of PCR products from Corby TR-GC (P1+P3) reveals mutagenesis of nucleotide positions in VR corresponding to adenines in TR. The two nucleotide mismatches observed in the invariant tag of VR1–4 are an artifact of the reverse primer and do not represent DGR-mediated mutagenesis.
Fig. S3. LdtA trafficking and modification. (A) Posttranslational modification of the conserved lipobox cysteine (C20) in LdtA. *E. coli* cells expressing epitope-tagged LdtA or LdtA-C20S were induced for protein expression. Cells were disrupted and total protein lysate was blocked with N-ethylmaleimide followed by cleavage of posttranslational modifications of cysteine by hydroxylamine, as indicated. Modifications removed by cleavage were replaced with biotin-BMCC (Pierce), and protein lysates run over a streptavidin column to bind biotinylated proteins, eluted, run on a protein gel, and probed by Western blotting with anti-Myc antibodies. Differential labeling of wild-type LdtA and the cysteine mutant C20S identifies posttranslational palmitoylation of the lipobox-conserved cysteine (red arrow ~54 kDa). A nonspecific product is indicated by a black arrow. (B) *L. pneumophila* Corby cells expressing epitope-tagged LdtA, LdtA-C20S, or CorbyΔtatB cells expressing LdtA were induced for protein expression and total membrane proteins isolated as in A. Unmodified cysteines in membrane proteins were blocked with N-ethylmaleimide followed by cleavage of posttranslational modifications by hydroxylamine, as indicated, and replacement of the modification with biotin-BMCC (Pierce). Labeled proteins were run over a streptavidin column, washed, eluted, and run on a protein gel for analysis by Western blot. Lysates were probed with anti-Myc antibodies to detect epitope-tagged LdtA (green) or fluorophore-conjugated streptavidin-detected biotinylated proteins (red). CorbyΔtatB was complemented by expression of tatB from plasmid. Red bands observed in membrane proteins without hydroxylamine (Left) treatment are biotinylated proteins that have nonspecifically bound and been eluted from the streptavidin column.
**Fig. S4.** A subset of *L. pneumophila* strains contain DGRs. (A) PCR screen detects conserved nucleotide motifs in *avd*, *RT*, and *L. pneumophila*-specific 16S rRNA using *L. pneumophila* strain 130b (DGR−) and strain Corby (DGR+) as controls. (B) A library of clinical *L. pneumophila* isolates was screened for putative DGR genes using primers from A. Results from screens for conserved *RT* genes and 16S loci are shown. Primers against *avd* gave similar results as those for *RT*. (C) Genomic libraries of putative DGR-containing strains were generated and sequenced. Reads for D5591 were assembled into contigs and a pairwise assembly (SI Material and Methods) using Corby as a reference genome. (D) The average G + C content was determined using 200-bp segments for a 10-kb region containing the putative DGR island (dashed green box) for strain D5591. The average G + C content for the entire island is 46%, and the average G + C content for the flanking regions is 35% (dashed red lines).

**Fig. S5.** Analysis of *L. pneumophila* TRs and VRs. Alignment of (A) TRs and (B) VRs from *L. pneumophila* Corby, D5572, and D5591. Nucleotide differences from the Corby sequences are shown.