Surface display of a massively variable lipoprotein by a *Legionella* diversity-generating retroelement

Diego Arambula[a], Wenge Wong[a], Bob A. Medhekar[b], Huatao Guo[a], Mari Gingery[c], Elizabeth Czornyj[a], Minghsun Liu[a], Sanghamitra Dey[d], Partho Ghosh[d], and Jeff F. Miller[a,e,f,1]

Departments of [a]Microbiology, Immunology, and Molecular Genetics and [c]Chemistry and Biochemistry, [e]Molecular Biology Institute, and [f]California NanoSystems Institute, University of California, Los Angeles, CA 90095; [b]Department of Internal Medicine, Section of Cardiovascular Medicine, Yale University School of Medicine, New Haven, CT 06511; and [d]Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093

Edited[†] by John J. Mekalanos, Harvard Medical School, Boston, MA, and approved April 1, 2013 (received for review January 22, 2013)

Diversity-generating retroelements (DGRs) are a unique family of retroelements that confer selective advantages to their hosts by facilitating localized DNA sequence evolution through a specialized error-prone reverse transcription process. We characterized a DGR in *Legionella pneumophila*, an opportunistic human pathogen that causes Legionnaires disease. The *L. pneumophila* DGR is found within a horizontally acquired genomic island, and it can theoretically generate 10^{26} unique nucleotide sequences in its target gene, legionella determinent target A (*ldtA*), creating a repertoire of 10^{19} distinct proteins. Expression of the *L. pneumophila* DGR resulted in transfer of DNA sequence information from a template repeat to a variable repeat (VR) accompanied by adenine-specific mutagenesis of progeny VRs at the 3′end of *ldtA*. *ldtA* encodes a twin-arginine translocated lipoprotein that is anchored in the outer leaflet of the outer membrane, with its C-terminal variable region surface exposed. Related DGRs were identified in *L. pneumophila* clinical isolates that encode unique target proteins with homologous VRs, demonstrating the adaptability of DGR components. This work characterizes a DGR that diversifies a bacterial protein and confirms the hypothesis that DGR-mediated mutagenic homing occurs through a conserved mechanism. Comparative bioinformatics predicts that surface display of massively variable proteins is a defining feature of a subset of bacterial DGRs.

bacterial surface display | TAT | accelerated evolution

Diversity-generating retroelements (DGRs) benefit their hosts by accelerating the evolution of target proteins (1–5). DGRs were first discovered in a *Bordetella* phage, BPP-1, which uses site-specific, error-prone reverse transcription to generate diversity in a gene that encodes tail fibers responsible for host ligand recognition (2). This process requires a DGR-encoded reverse transcriptase (RT), an accessory variability determinant [(Avd) or an equivalent protein), and a template repeat-derived RNA intermediate. The template repeat (TR)-RNA provides a template for reverse transcription, during which TR adenine residues are copied into any of the four nucleotides. The diversified cDNA displaces a variable repeat (VR) at the 3′ end of the target gene (6, 7). This unidirectional, targeted, mutagenic retrotransposition process is called mutagenic homing. Target recognition requires two *cis*-acting sequences at the 3′ end of VR; the initiation of mutagenic homing (IMH) element and a DNA hairpin/cruciform structure (5–7). Mutagenic homing operates through a copy-and-replace mechanism in which *cis*- and *trans*-acting elements required for further rounds of diversification are preserved, allowing iterative optimization of target protein (TP) function. Structural studies with BPP-1 show that TR adenines are precisely positioned to correspond to residues in the ligand binding pocket of a C-type lectin (CLec) domain at the C terminus of the tail fiber protein, reflecting coevolution between the genetic mechanism that generates diversity and the protein scaffold that displays it (5, 7). A second DGR TP, *Treponema denticola* treponema variable protein A (TvpA) (TDE2269), has been shown to share this disposition of TR adenines in a CLec-fold protein (8).

The *Bordetella* BPP-1 phage DGR has provided the sole paradigm for mechanistic studies of this family of retroelements. Bioinformatic analysis, however, shows that the majority of DGRs are chromosomal elements in bacterial genomes, with representatives in every phylum with significant sequence coverage. Here we present a functional analysis of a non–phage-associated bacterial DGR, using *Legionella pneumophila* as a model system. *L. pneumophila* is a Gram-negative intracellular pathogen that has evolved strategies to evade predation in the environment, leading to accidental virulence in humans (9).

A genomic retroelement was identified in *L. pneumophila* strain Corby, which encodes all components that are characteristically shared among DGRs. The element is capable of directing adenine-specific mutagenesis of a target gene, legionella determinant target A (*ldtA*), and our analyses revealed DNA, RNA, and protein requirements for mutagenic homing in the *L. pneumophila* and BPP-1 DGRs are analogous, implying mechanistic conservation. The diversified TP, LdtA, contains twin-arginine translocation (TAT) and lipoprotein processing (LPP) motifs at its N terminus, along with a C-terminal VR. The N-terminal secretion signals position LdtA in the outer face of the outer membrane, allowing surface display of the variable C-terminal domain, which is predicted to have a CLec fold. To determine the distribution of DGRs in *L. pneumophila*, we screened a library of strains and identified elements that are highly homologous, but have coopted different carrier sequences to display related variable domains; this demonstrates the modularity of these systems. Examination of TPs in diverse Gram-negative bacteria suggests that lipoprotein anchoring and surface display of DGR-diversified protein repertoires is a common theme in many Gram-negative bacteria.

Results

Anatomy of a *Legionella* DGR. The DGR in Fig. 1*A* is located within a 10-kb genomic island in the chromosome of *L. pneumophila* strain Corby. Flanking genes are predicted to encode the heavy metal transport system, cadmium translocating P-type ATPase (*cadA–C*), and a type IV secretion system with associated regulators, legionella vir region (*lvrA–C*, 10). The island shows signs of recent horizontal acquisition, as indicated by

Author contributions: D.A., B.A.M., H.G., P.G., and J.F.M. designed research; D.A., W.W., H.G., S.D., and P.G. performed research; M.G., E.C., and M.L. contributed new reagents/analytic tools; D.A., B.A.M., H.G., M.G., E.C., M.L., S.D., P.G., and J.F.M. analyzed data; and D.A., H.G., M.L., P.G., and J.F.M. wrote the paper.

Conflict of interest statement: J.F.M. is cofounder and chair of, and H.G. is a consultant for, the scientific advisory board of AvidBiotics Inc., a biotherapeutics company in San Francisco.

[†]This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: jfmiller@ucla.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1301366110/-/DCSupplemental.

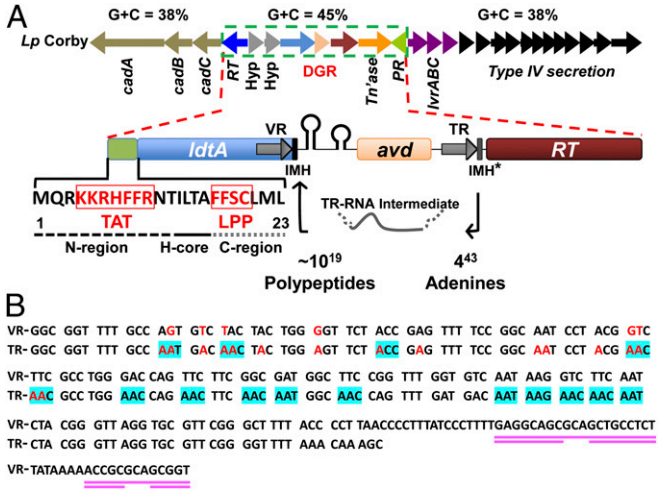

Fig. 1. A *Legionella pneumophila* DGR. (*A*) A DGR in *L. pneumophila* Corby is found within a 10-kb genomic island (green dashed box) adjacent to heavy metal transport gene cluster (*cadA–C*) and type IV secretion system regulatory genes (*lvrA–C*). The DGR-containing genomic island has a higher G + C content (percentile score) than flanking regions, a non-DGR RT (dark blue), genes annotated as hypothetical proteins (gray), putative transposase (orange), and a gene annotated as a phage repressor (light green). DGR locus is expanded with the TP gene (*ldtA*); predicted secretion (TAT) and localization sequences (LPP); accessory protein gene (*avd*); and DGR-encoded reverse transcriptase (*RT*) identified. Additional DGR elements: DNA stem/loops, IMH, and IMH* are indicated. Black arrows represent mutagenic homing as transfer of nucleotide sequence information from TR to VR via an RNA intermediate and the predicted amino acid diversity. The LdtA signal peptide contains a polar N-region (residues 1–11), followed by a hydrophobic core (residues 12–15) and a nonpolar C-region (residues 17–23). A predicted noncanonical TAT motif, KKRHFFR, differs from the *E. coli* consensus [(S/T) RRxFLK]. The LdtA signal peptide contains a lipobox motif, FFSC, which is predicted to be lipid-modified (14). (*B*) Alignment of Corby DGR VR/TR shown in the *ldtA* reading frame. TR adenines (red) usually occupy the first two positions of AAC or AAT codons (shaded) and correspond to substitutions in VR. Nucleotide sequences downstream of the VR stop codon representing DNA stem/loop (underline) structures are shown.

a G + C content (45%) that differs from the rest of the genome (38%), and the presence of a transposase and an unrelated *RT* (11). Interestingly, the 10-kb DGR-containing island is located within a larger (64-kb) integrative and conjugative element (ICE) that has recently been shown to be capable of horizontal transfer (10). The retroelement itself encodes a DGR-type RT (2), an Avd homolog, cognate TR, and VR sequences that differ at sites corresponding to adenines in TR, and tandem stem-loop structures downstream of VR (6, 7). The 148-bp TR contains 43 adenines that most often occupy the first two positions of AAC or AAT codons, allowing maximal amino acid diversity while excluding the possibility of nonsense mutations generated by adenine mutagenesis (Fig. 1*B*). Following mutagenic homing, the *L. pneumophila* TR can theoretically generate 4^{43} ($\sim 10^{26}$) unique DNA sequences capable of encoding $\sim 10^{19}$ different polypeptides—a repertoire of massive proportions. VR-encoded sequences are located at the C terminus of LdtA, within a domain predicted to adopt a CLec fold similar to the distal end of the BPP-1 phage tail fiber (Fig. S1). Two protein-trafficking signals are predicted at the N terminus of LdtA: the first is a TAT motif that differs from the consensus in *Escherichia coli* but is characteristic of known and putative TAT substrates in *L. pneumophila* (12, 13); the second is a lipobox that is also noncanonical. The TAT pathway is an alternative secretion system found in plants and bacteria that can translocate folded proteins or complexes across membranes (12), and lipobox motifs mediate signal peptide cleavage, lipid modification, and anchoring to the inner or outer membrane (14). Although the ability of TAT and lipobox secretion motifs to function in concert has not been thoroughly characterized (15, 16), we hypothesized that the N terminus of LdtA mediates secretion, membrane localization, and, potentially, surface exposure.

***L. pneumophila* DGR Is Functional.** As a first step in characterizing the *L. pneumophila* DGR, we determined if it is capable of mutagenic homing. A PCR assay was used to detect RT-dependent transfer of an invariant sequence tag from TR to VR, and adenine mutagenesis was evaluated by sequencing amplified retrohoming products (Fig. 2). The *avd*, *RT*, and a modified TR containing a 20-bp PCR tag consisting of G + C residues (TR-GC) were expressed *in trans* to the wild-type DGR on a plasmid vector (pATR; Fig. 2*B*). Negative controls included constructs expressing wild-type TR (no tag) or a catalytically inactive RT. Following induction, DNA was extracted and homing products were amplified using primer sets that annealed to the GC tag and sequences upstream (P1/P3) or downstream (P2/P4) of VR. Homing products were readily detected and identified by transfer of the tag from TR-GC to VR (Fig. 2*C*), and sequences of GC-tagged VRs revealed adenine-specific mutagenesis (Fig. S2*A*). These results show that the *L. pneumophila* Corby DGR encodes functional components that are capable of catalyzing adenine-specific mutagenic homing to the VR region of *ldtA*.

In the experiment in Fig. 2*C*, efficient transfer of the GC tag was detected following expression of *avd*, TR-GC, and *RT in trans* to the chromosomal DGR. To measure activity of the native element, we inserted the 20-bp GC tag into the chromosomal TR by allelic exchange, and assayed mutagenic homing by PCR. Transfer of the GC tag and adenine mutagenesis was both observed (Fig. S2 *B* and *C*), but detection required an increase in the number of PCR amplification cycles compared with the experiment in Fig. 2*B* (25 vs. 35 cycles). The low level of activity of the native chromosomal element was not surprising. DGR-mediated mutagenesis is stochastic, and the vast majority of diversified target genes are likely to encode inactive products. Mutagenic homing is expected to be tightly controlled to prevent a loss of fitness due to overdiversification (*Discussion*).

***Cis*- and *Trans*-Acting Factors Required for Mutagenic Homing.** Our understanding of mutagenic homing derives almost exclusively from studies of the phage BPP-1 DGR (2, 7). We were curious to determine if observations with BPP-1 are applicable to other DGRs, especially those encoded on bacterial chromosomes. To explore requirements for mutagenic homing by the *L. pneumophila* DGR, *avd*, TR, and *RT* were deleted *en bloc* from the *L. pneumophila* genome, and pATR (Fig. 2*B*) or derivatives with mutations in *trans*-acting factors were tested in our PCR-based assay. The ability to detect homing products required the TR tag and expression of an active RT, and homing was also dependent on expression of *avd* (Fig. 2*D*). The *avd* gene is predicted to encode a small, basic protein (14.1 kDa, pI 9.3) with homologs in other DGRs, including BPP-1, where it serves an essential function and has been shown to form a positively charged pentameric barrel that interacts with RT (2–4). IMH*, located at the 3′ terminus of TR, is a 32-bp sequence with two mismatches to IMH. In BPP-1, IMH and IMH* are essential components that are predicted to facilitate assembly of a complex between VR, TR-RNA, and RT during priming and reverse transcription (6, 7). Deletion of IMH* on the pATR donor plasmid, or IMH on the recipient chromosome, abrogated activity of the *L. pneumophila* DGR (Fig. 2*D*).

In BPP-1, target-site recognition is also dependent on the presence of an inverted repeat that flanks IMH and forms a hairpin/cruciform structure in supercoiled DNA (7). These elements are highly conserved in phage DGRs, where they consist of 7- to 10-bp GC-rich stems and 4-nt loops. A divergent yet potentially analogous element, composed of two tandem

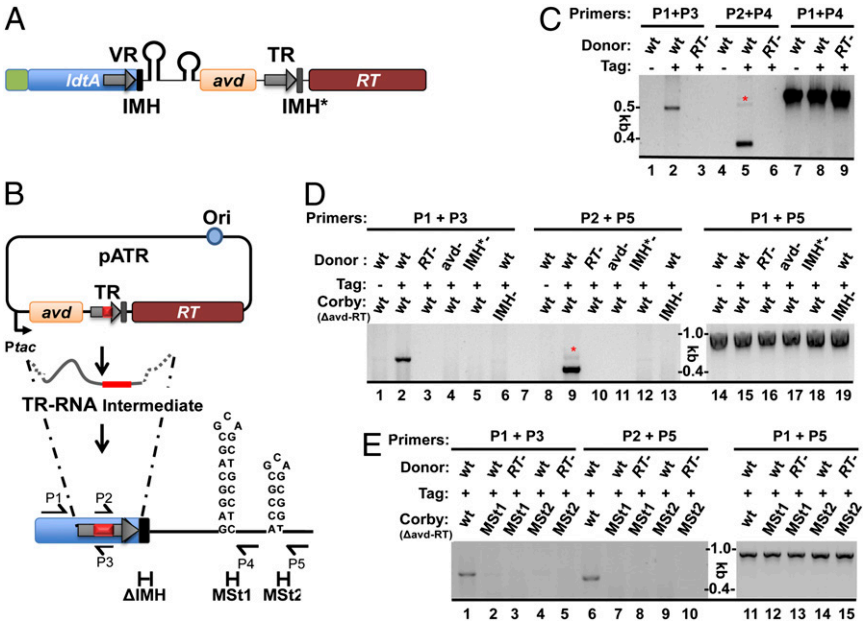

Fig. 2. The *L. pneumophila* DGR is capable of mutagenic homing. (*A*) *L. pneumophila* Corby DGR cassette with components indicated as in Fig. 1*A*. (*B*) Experimental design. pATR encodes *avd*, TR with 20-bp GC PCR tag (red box), and *RT* expressed from P*tac*. DGR-mediated transfer incorporates the GC tag into the chromosomal VR. Primer binding sites (P1–P5) for PCR-based mutagenic homing assays are shown, and deletions removing IMH, or disruption of the first (MSt1) or second (MSt2) stem/loop structure are indicated (black brackets). (*C*) DGR homing assays. Donor plasmids (pATR and derivatives) were transformed into wild-type *L. pneumophila* Corby, P*tac* expression induced, and genomic DNA extracted and used for PCR with primers shown in *B*. Donor pATR derivatives encoded wild-type (*wt*) or mutant (*RT*–) RT alleles, with (+) or without (–) TR tags. Equivalent amounts of template DNA were used for PCR assays (P1 + P4). Nonspecific PCR amplification of donor plasmid indicated (red asterisk). (*D*) pATR constructs used in *C* and additional derivatives with deletions removing *avd* (*avd*–) or IMH (IMH–) were transformed into *L. pneumophila* Corby deleted for sequences from *avd* to *RT* (Δ*avd-RT*), with IMH present (*wt*) or deleted (*IMH*–). DGR homing assays were conducted as in *C* with equal amounts of template DNA (P1 + P5). Nonspecific PCR amplification of donor plasmid indicated (red asterisk). (*E*) pATR donor plasmids with wild-type (*wt*) or mutant RT alleles (*RT*–) were transformed into *L. pneumophila* Corby Δ*avd-RT* with *wt* or disrupted stem/loop structures (MSt1, MSt2). DGR homing assays were conducted as in *C* with equal amounts of template DNA (P1 + P5).

repeats with GC-rich stems of different lengths and identical 3-nt loops, is present downstream of the *L. pneumophila ldtA* IMH (Fig. 2*B*). To examine their role in homing, we generated mutants in which sequences in the 3′ halves of either stem were replaced with complementary nucleotides to disrupt base pairing. As shown in Fig. 2*E*, disruption of stem-loop 1 (SL1) or stem-loop 2 (SL2) eliminated homing. Our results show that the *L. pneumophila* and BPP-1 DGRs operate in a fundamentally similar manner, using conserved *cis*- and *trans*-acting components for adenine-specific mutagenic homing.

LdtA Is Surface-Exposed. *L. pneumophila* Corby cells expressing LdtA or control proteins with C-terminal HA epitope tags were lysed and separated into soluble and membrane fractions. LdtA partitioned to the membrane fraction along with macrophage infectivity potentiator (MIP) [outer membrane (OM) protein control] (17) and defective organelle trafficking A (DotA) [inner membrane (IM) protein control] (18), whereas recombinase A (RecA) appeared in the soluble fraction (Fig. 3*A*) (19). On further separation by isopycnic gradient ultracentrifugation, DotA was enriched in fractions containing IM proteins, whereas MIP and LdtA preferentially partitioned to the OM (Fig. 3*B*). To determine the orientation of LdtA, intact bacterial cells were treated with proteases under conditions that preferentially degrade surface-exposed proteins (Fig. 3*C*). Cells were induced to express HA-tagged LdtA, MIP, DotA, RecA, or intracellular multiplication X (IcmX) as a periplasmic control (20), and incubated with increasing concentrations of proteinase K (21). MIP, an integral OM protein, showed moderate protease sensitivity, whereas periplasmic (IM) and cytoplasmic control proteins were relatively unaffected. In contrast, LdtA was highly sensitive to protease treatment, indicative of surface localization.

Indirect immunofluorescence was used as an independent approach to test surface exposure. Cells expressing full-length LdtA-HA were recalcitrant to visualization attempts, and we hypothesized this was due to sequestration of the epitope tag within a folded structure. Based on structural modeling predictions, we constructed a variant that expressed the first 370 amino acids of LdtA fused to a triple-HA C-terminal epitope tag (LdtA-370-3HA). Surface immunofluorescence was readily detected using intact cells, whereas visualization of the IcmX or DotA negative controls required OM permeabilization (Fig. 3*E*). In sum, membrane fractionation, protease sensitivity, and surface immunofluorescence support the conclusion that LdtA is an OM protein with a surface-exposed C terminus.

LdtA Is a TAT-Secreted Lipoprotein. Translocation through the TAT pathway requires cytoplasmic recognition by the TatABC complex of signal sequences with the consensus motif SRRxFLK, which is conserved in plants and other bacteria but more variable in *Legionella* (12, 13). The predicted LdtA secretion motif retains requisite recognition components, including polar N-terminal residues, a hydrophobic core with the arginine required for TAT transport, and a nonhydrophobic C-terminal residue (Fig. 1*A*). To determine if TAT is required for LdtA secretion, we generated an in-frame deletion in *L. pneumophila* Corby *tatB* (22, 23). LdtA-HA was protease resistant in the Δ*tatB* strain, and protease sensitivity was restored by complementation with *tatB* (Fig. 3*D*). Confirmatory results were obtained by immunofluorescence (Fig. 3*E*).

Having shown that LdtA is secreted through the TAT pathway, we were curious to determine how it becomes localized in the OM. TAT motifs often contain a consensus sequence at their C terminus, Ala-Xaa-Ala, which mediates cleavage by signal peptidase I (12). In contrast, we identified a potential lipobox motif, FFSC, at the analogous position in LdtA (Fig. 1*A*); this suggested a hybrid signal that combines TAT translocation with signal peptide cleavage, lipid modification of the conserved cysteine (Cys-20), and trafficking to the OM by the localization of lipoprotein (LOL) system. To test this hypothesis, we first determined if Cys-20 is required for surface exposure. As shown in Fig. 3*D*, alanine (C20A) or serine (C20S) substitutions at this site rendered LdtA-HA protease resistant, with a corresponding loss of surface immunofluorescence (Fig. 3*F*). To determine the step at which trafficking is blocked, we compared the effects of OM permeabilization on phenotypes of Cys-20 substitutions in *wt* vs. Δ*tatB* mutants. Wild-type *L. pneumophila* expressing LdtA-C20A showed sensitivity to proteinase K under permeabilizing conditions but not in intact cells, as observed with the periplasmic control protein IcmX (Fig. 3 *C*–*F*). In contrast, LdtA-

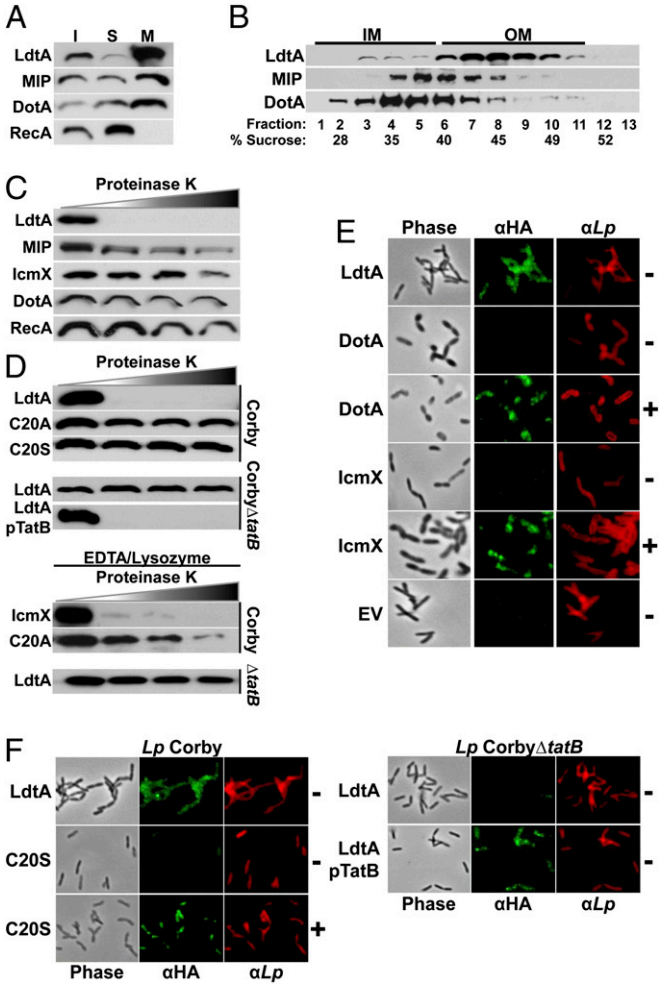

Fig. 3. LdtA is a surface-exposed lipoprotein. Epitope (HA)-tagged LdtA or control proteins were used for subcellular localization experiments. (*A*) Cells were lysed by French press and cellular constituents separated by high-speed centrifugation (I). Proteins with known subcellular locations: MIP (OM), DotA (IM), and RecA (cytoplasmic) were compared with LdtA. RecA was enriched in the soluble fraction(S), whereas LdtA, MIP, and DotA were enriched in membrane fractions(M). Membrane proteins detected in soluble fractions are due to overexpression and increased pools of secretion intermediates. (*B*) High-speed isopycnic sucrose gradient fractionation was used to separate membrane proteins and fractions were probed by Western blotting with anti-HA antibodies. Fractions corresponding to IM or OM are shown. (*C*) *L. pneumophila* Corby cells expressing epitope-tagged proteins were treated with increasing concentrations of proteinase K (0–200 μg/mL) to digest surface-exposed proteins, and whole-cell lysates were probed, as above. HA-tagged IcmX, a periplasmic T4SS component, was included as a control. (*D*) (*Top*) Mutation of the essential Cys in the LdtA LPP processing signal to Ala (C20A) or Ser (C20S) protected LdtA from digestion by proteinase K. (*Middle*) An in-frame deletion in *tatB* (Corby*ΔtatB*) eliminated LdtA digestion, and complementation by plasmid-expressed *tatB* (p*tatB*) restored protease sensitivity. (*Bottom*) *L. pneumophila* Corby cells expressing epitope-tagged IcmX, LdtA C20A, or Corby*ΔtatB* expressing LdtA were permeabilized with EDTA/lysozyme and treated with increasing concentrations of proteinase K. (*E*) Indirect immunofluorescence detection of epitope-tagged surface proteins in *L. pneumophila* Corby. Cells were treated with anti-HA or anti-*L. pneumophila* antibodies under permeabilizing (+) (EDTA and lysozyme) or nonpermeabilizing (–) conditions, as indicated. EV, empty vector control. (*F*) (*Left*) Indirect immunofluorescence detection of epitope-tagged LdtA or LdtA-C20S mutant derivative under permeabilizing (+) or nonpermeabilizing (–) conditions. (*Right*) Indirect immunofluorescence detection of epitope-tagged LdtA in *L. pneumophila* Corby*ΔtatB* in the presence or absence of a complementing plasmid (p*tatB*).

C20 was protease resistant in the Δ*tatB* strain, even after OM permeabilization. Immunofluorescence results paralleled protease sensitivity assays (Fig. 3 *E* and *F*) and showed that mutation of Cys-20 causes missorting to the periplasm, a phenotype that would be predicted to result from the lack of modification of an OM lipoprotein.

Bacteria often modify lipobox cysteines with palmitic acid by linkage to *N*-acyl-*S*-diacylglyceryl-Cys moieties (24, 25). Acylbiotin-exchange chemistry, which provides a sensitive alternative to labeling cells with radioactive fatty acids, was used to detect posttranslational modification of LdtA (Fig. S3 *A* and *B*) (25). Protein lysates from *E. coli* and wild-type or Δ*tatB L. pneumophila* cells expressing LdtA or LdtA-C20S were treated with hydroxylamine to remove acyl-linked lipids, which were then replaced with biotin. Biotinylated proteins were column purified, eluted, and analyzed by Western blotting. Biotin-labeled LdtA was found in wild-type *L. pneumophila* but not in Δ*tatB* mutants, and LdtA-C20S was refractory to biotinylation (Fig. S3*B*). Taken together, our results show that after secretion across the IM by TAT, LdtA is anchored in the outer surface of the OM by an acyl-linked lipid modification.

Distribution of DGRs in *L. pneumophila* Isolates. Though the *Legionella* genus contains over 50 species, more than 84% of cases of legionellosis are caused by *L. pneumophila* serogroup I (Sg1) (26). The DGR in *L. pneumophila* Corby, a Sg1 clinical isolate, is absent in several other sequenced Sg1 strains, and conserved flanking sequences define the endpoints of the DGR genomic island (Figs. 1*A* and 4*A*). To further investigate the distribution of DGRs within *L. pneumophila*, we conducted a PCR screen of 12 additional Sg1 clinical isolates and found three that contained DGR-associated loci (Fig. S4); to determine their composition and architectures, multiplexed genomic libraries of three newly identified DGR-containing isolates were sequenced. Complete sets of conserved DGR components, including *RT*, *avd*, TR, VR, stem loops, IMH, IMH*, and TP loci were identified in strains D5572 and D5591. The third isolate, D5549, contained *RT*, TR, and *avd* genes without an identifiable linked TP. Given the ability of the diversifying machinery to act *in trans* (7), unlinked target proteins may exist in this isolate. In D5591, the DGR is located on a genomic island similar to the one in *L. pneumophila* Corby, with identical regulatory genes to the right (Fig. 4; Fig. S4*D*). A chromosomal rearrangement at the left boundary appears to have substituted different flanking loci and removed genes from the genomic island. For D5572 and D5549, DGRs are encoded on similar genomic islands but flanking sequences could not be identified, even at 50-fold coverage, because the high degree of plasticity in the *L. pneumophila* genomes precluded assembly of long stretches of contiguous sequence (Fig. S4*C*) (27).

Fig. 4*B* shows details of the DGRs in three Sg1 isolates, and their comparison is quite interesting. *RT* and *avd* genes are predicted to encode nearly identical products, and the TR loci share 97% nucleotide identity. VR sequences are also similar, with differences at positions corresponding to adenine residues in their cognate TRs (Fig. S5). Protein threading predicts that the VR domains of LdtA, LdtB, and LdtC adopt similar CLec folds with binding pockets composed of diversified residues (Fig. S1). Though LdtA and LdtB are related throughout their entire length, including secretion and localization signals, sequences upstream of the LdtC VR have significantly diverged. Despite this, the extreme N terminus of LdtC is predicted to encode a Sec-dependent secretion signal followed by a lipobox motif, which could provide an alternative means for surface localization (28). These observations illustrate the modular nature of DGRs and is consistent with the notion that *trans*-acting factors and diversified scaffolds act in a generic manner to evolve ligand binding specificities, which can be adapted to different functions through connection with different N-terminal domains.

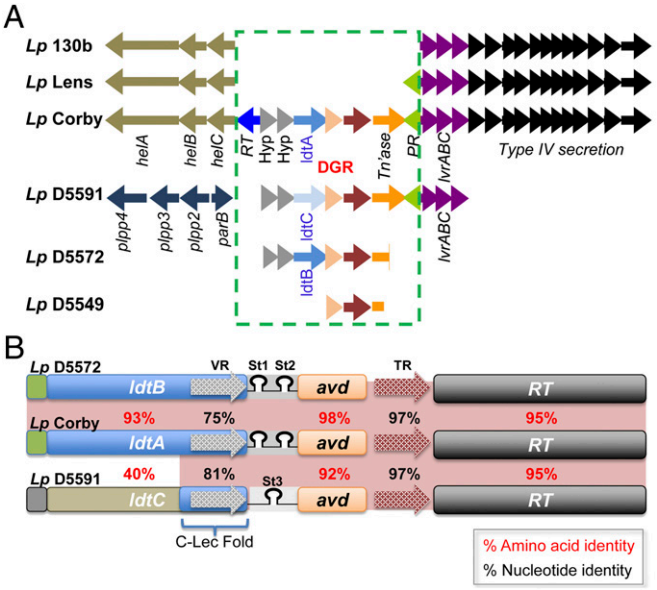

Fig. 4. A subset of *L. pneumophila* strains contain DGRs. (*A*) The genomic island (dashed box) containing the *L. pneumophila* Corby DGR is absent in sequenced and assembled genomes of *L. pneumophila* strains 130b and Lens, and DGR and flanking genomic island sequences are present in Sg1 strains D5549, D5572, and D5591, which were partially sequenced in this study. (*B*) Sequence comparisons between DGRs. TP genes (*IdtA–C*), TAT/LPP secretion signals (green boxes), and Sec/LPP secretion signal (gray box) are shown. Other DGR components are indicated with percent identities between nucleotide (black) and amino acid sequences (red). Predictive programs (*SI Material and Methods*) identified similar stem loops (St1/2) in Corby and D5572, whereas D5591 contains a single, highly structured stem loop (St3). The predicted CLec folds in LdtA, B, and C are modeled in Fig. S1.

LdtA as a Model for Diversified Surface Lipoproteins. The ability to introduce massive amounts of diversity in surface-anchored bacterial proteins seems likely to be an adaptive trait. In support of this idea, a partial survey of sequenced bacterial genomes revealed multiple DGRs that are predicted to diversify TPs with N-terminal lipoprotein sequences. As shown in Fig. 5, predicted DGR-diversified TPs in *Treponema denticola*, *Bacteroides fragilis*, *Bacteroides thetaiotaomicron*, *Vibrio angustum*, and *Shewanella baltica* contain N-terminal LPP motifs. Although LdtA is known, and LdtB is predicted, to be TAT secreted, the other TPs in Fig. 5 have Sec secretion signals. *T. denticola* encodes seven TPs that are particularly interesting; we predict they are diversified by a single DGR, and the majority includes highly predicted LPP processing signals. Spirochetes are known to make extensive use of the LPP secretion pathway to anchor proteins to their outer surface (21). Taken together, our results suggest that lipid modification and surface exposure are conserved features of proteins that are diversified by bacterial chromosomal DGRs.

Discussion

To date, over 300 unique DGRs have been identified in phage, plasmid, or bacterial genomes, and they are associated with an array of diverse ecological niches. Despite their widespread distribution in nature, and capacity to confer selective advantages, only a single, phage-encoded DGR has been studied in mechanistic detail (2–7). Our discovery of a functional DGR in *L. pneumophila* provides a bacterial system for comparative analysis. We demonstrate that targeted, adenine-specific mutagenesis occurs in *L. pneumophila*, providing direct support for the hypothesis that this is a capability common to all DGRs. Both the *cis*- and *trans*-acting requirements for DGR activity in *L. pneumophila* are analogous to those in *Bordetella* phage, highlighting the conserved nature of the mutagenic homing mechanism. Of equal importance, our analysis of the LdtA TP demonstrates that DGRs have been adapted to mediate the display of variable protein repertoires on the surface of bacterial cells.

To our knowledge, the amount of diversity that can be generated by *L. pneumophila* DGRs is greater than for any characterized biological system, including the diversification of Ig scaffolds during mammalian immune responses (29). The comparison between DGR-mediated diversity and the generation of immunity is instructive in several ways. In both cases, the genetic mechanisms responsible for creating diversity have coevolved with protein scaffolds to display it; the Ig fold for antibodies and T-cell receptors and the CLec fold for DGRs. Additionally, the same basic sequence of target-gene diversification, surface display of variable proteins, and selection leading to amplification appears to hold in either case. DGRs, however, operate under a unique constraint. Although mutagenesis is highly directed, it is inherently stochastic, and the vast majority of mutagenic events are likely to be deleterious. For TPs that have function, we predict the frequency of mutagenic homing will be low, or regulated, to balance the loss of fitness resulting from mutagenesis

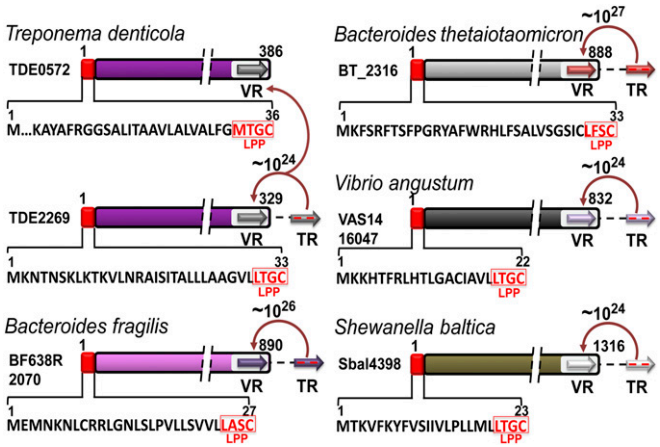

Fig. 5. Putative DGR target proteins in diverse bacteria are predicted lipoproteins. Species and protein ORFs are indicated. Analysis of putative TPs (28) identified N-terminal lipoboxes (red) and C-terminal VR sequences (colored arrows) found within predicted CLec domains (white box). The *T. denticola* 35405 genome encodes seven TPs (two are shown), which are predicted to be diversified *in trans* by a single DGR linked to TDE2269. All other TPs are encoded within their cognate DGRs. Numbers above arrows represent theoretical levels of nucleotide diversity generated by mutagenic homing.

with the advantages conferred by accelerated evolution. This relationship may explain the low level of basal activity observed for the native *L. pneumophila* DGR, and our ability to increase mutagenesis by exogenous expression of *trans*-acting components. In support of this idea, preliminary data suggests that DGR-encoded *trans*-acting factors are up-regulated in *L. pneumophila* during the transition from replicative to transmissive states (30).

We have demonstrated that LdtA traffics across the inner membrane via the TAT translocon, is lipidated, and localized to the external face of the outer membrane, presumably by the LOL processing system. Although the manner in which the TAT translocation and LOL processing systems intersect, and the mechanism through which lipoproteins are resolved to the outer leaflets of outer membranes remain as open questions, our results clearly demonstrate that LdtA adorns the surface of *L. pneumophila*. We show that lipid modification at the N terminus of LdtA correlates with surface exposure, and that the C-terminal CLec domain is available to the extracellular milieu. CLec domains have been reported as general ligand-binding domains found in metazoan, bacterial, and viral proteins, and known binding partners include other proteins, sugars, lipids, and inorganic ligands (8). Though the function of LdtA is presently unknown, a likely possibility is that it facilitates *L. pneumophila* receptor–ligand interactions of importance for survival in the environment and/or interactions with host cells.

The *L. pneumophila* Corby and D5591 DGRs are located in conserved genomic islands that bear the hallmarks of recent horizontal acquisition. In *L. pneumophila* Corby, the DGR island appears to be a recently acquired element within a much larger genomic island corresponding to an ICE, which appears, on the basis of G + C content, to be ancestral within the Legionellaceae (10). In our initial screen for DGR elements in *L. pneumophila* Sg1 clinical isolates, we identified *RT* and *avd* homologs in 25% of the strains tested. It will be important to investigate the distribution of DGRs in larger sample sets that include both clinical and environmental isolates to determine if correlations exist between the presence or absence of DGRs, or the nature of variable proteins, and virulence for humans. Our analysis of DGRs in clinical isolates revealed that all *L. pneumophila* DGRs share nearly identical *avd*, TR, and *RT* loci as well as highly conserved targeting elements (IMH and stem loops). Despite this similarity in the diversification machinery, each VR displays a unique pattern of adenine-specific mutagenesis, which is consistent with the hypothesis that these elements are active in nature, and target protein functions are under selection. The observation that similar VRs have been fused to entirely different N-terminal sequences in *L. pneumophila* D5591 vs. *L. pneumophila* Corby/D5572 provides further illustration of the modular nature of diversified proteins and the versatility of the VR-encoded CLec scaffold. The widespread distribution of DGRs in nature and their adaptation to mediate both phage and bacterial surface display is not surprising given their utility as internally programmed, self-renewable systems that accelerate the evolution of adaptive traits.

Materials and Methods

Detailed experimental procedures are found in *SI Materials and Methods*. *L. pneumophila* Corby was a gift from Natalia Kozak (Centers for Disease Control, Bethesda). Mutants were constructed using allelic exchange as previously described (22). Homing assays (6), protease sensitivity assays (21), Immunofluorescence (21, 22) microscopy, and acylbiotin-exchange chemistry (25) were performed, with minor modification, as previously described. *In silico* analysis and protein threading was performed as described in *SI Materials and Methods*.

ACKNOWLEDGMENTS. This work was supported by National Institutes for Health Grants T32 AI007323, R01 AI069838, and R21 DE021528.

1. Gogvadze E, Buzdin A (2009) Retroelements and their impact on genome evolution and functioning. *Cell Mol Life Sci* 66(23):3727–3742.
2. Doulatov S, et al. (2004) Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. *Nature* 431(7007):476–481.
3. Alayyoubi M, et al. (2013) Structure of the essential diversity-generating retroelement protein bAvd and its functionally important interaction with reverse transcriptase. *Structure* 21(2):266–276.
4. Medhekar B, Miller JF (2007) Diversity-generating retroelements. *Curr Opin Microbiol* 10(4):388–395.
5. Miller JL, et al. (2008) Selective ligand recognition by a diversity-generating retroelement variable protein. *PLoS Biol* 6(6):e131.
6. Guo H, et al. (2008) Diversity-generating retroelement homing regenerates target sequences for repeated rounds of codon rewriting and protein diversification. *Mol Cell* 31(6):813–823.
7. Guo H, et al. (2011) Target site recognition by a diversity-generating retroelement. *PLoS Genet* 7(12):e1002414.
8. Le Coq J, Ghosh P (2011) Conservation of the C-type lectin fold for massive sequence variation in a Treponema diversity-generating retroelement. *Proc Natl Acad Sci USA* 108(35):14649–14653.
9. Isberg RR, O'Connor TJ, Heidtman M (2009) The Legionella pneumophila replication vacuole: Making a cosy niche inside host cells. *Nat Rev Microbiol* 7(1):13–24.
10. Lautner M, Schunder E, Herrmann V, Heuner K (2013) Transfer of genomic islands in Legionella pneumophila. *J Bacteriol* 195(7):1583–1597.
11. Langille MG, Hsiao WW, Brinkman FS (2010) Detecting genomic islands using bioinformatics approaches. *Nat Rev Microbiol* 8(5):373–382.
12. Stanley NR, Palmer T, Berks BC (2000) The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*. *J Biol Chem* 275(16):11591–11596.
13. De Buck E, et al. (2004) A putative twin-arginine translocation pathway in Legionella pneumophila. *Biochem Biophys Res Commun* 317(2):654–661.
14. Narita S, Tokuda H (2010) Sorting of bacterial lipoproteins to the outer membrane by the Lol system. *Methods Mol Biol* 619:117–129.
15. Shruthi H, Babu MM, Sankaran K (2010) TAT-pathway-dependent lipoproteins as a niche-based adaptation in prokaryotes. *J Mol Evol* 70(4):359–370.
16. Gralnick JA, Vali H, Lies DP, Newman DK (2006) Extracellular respiration of dimethyl sulfoxide by Shewanella oneidensis strain MR-1. *Proc Natl Acad Sci USA* 103(12):4669–4674.
17. Helbig JH, et al. (2001) Immunolocalization of the Mip protein of intracellularly and extracellularly grown Legionella pneumophila. *Lett Appl Microbiol* 32(2):83–88.
18. Roy CR, Isberg RR (1997) Topology of Legionella pneumophila DotA: An inner membrane protein required for replication in macrophages. *Infect Immun* 65(2):571–578.
19. Miyamoto H, et al. (1993) Protein profiles of Legionella pneumophila Philadelphia-1 grown in macrophages and characterization of a gene encoding a novel 24 kDa Legionella protein. *Microb Pathog* 15(6):469–484.
20. Matthews M, Roy CR (2000) Identification and subcellular localization of the Legionella pneumophila IcmX protein: A factor essential for establishment of a replicative organelle in eukaryotic host cells. *Infect Immun* 68(7):3971–3982.
21. Schulze RJ, Zückert WR (2006) Borrelia burgdorferi lipoproteins are secreted to the outer surface by default. *Mol Microbiol* 59(5):1473–1484.
22. De Buck E, et al. (2005) Legionella pneumophila Philadelphia-1 tatB and tatC affect intracellular replication and biofilm formation. *Biochem Biophys Res Commun* 331(4):1413–1420.
23. Rossier O, Cianciotto NP (2005) The Legionella pneumophila tatB gene facilitates secretion of phospholipase C, growth under iron-limiting conditions, and intracellular infection. *Infect Immun* 73(4):2020–2032.
24. Kamalakkannan S, Murugan V, Jagannadham MV, Nagaraj R, Sankaran K (2004) Bacterial lipid modification of proteins for novel protein engineering applications. *Protein Eng Des Sel* 17(10):721–729.
25. Linder ME, Deschenes RJ (2007) Palmitoylation: Policing protein stability and traffic. *Nat Rev Mol Cell Biol* 8(1):74–84.
26. Kozak NA, et al. (2009) Distribution of lag-1 alleles and sequence-based types among Legionella pneumophila serogroup 1 clinical and environmental isolates in the United States. *J Clin Microbiol* 47(8):2525–2535.
27. Cazalet C, et al. (2004) Evidence in the Legionella pneumophila genome for exploitation of host cell functions and high genome plasticity. *Nat Genet* 36(11):1165–1173.
28. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786.
29. Sidhu SS, et al. (2004) Phage-displayed antibody libraries of synthetic heavy chain complementarity determining regions. *J Mol Biol* 338(2):299–310.
30. Dalebroux ZD, Svensson SL, Gaynor EC, Swanson MS (2010) ppGpp conjures bacterial virulence. *Microbiol Mol Biol Rev* 74(2):171–199.