Evidence for Acquisition of *Legionella* Type IV Secretion Substrates via Interdomain Horizontal Gene Transfer

Karim Suwvan de Felipe,^{1,2} Sergey Pampou,^{2,3} Oliver S. Jovanovic,² Christopher D. Pericone,² Senna F. Ye,² Sergey Kalachikov,³ and Howard A. Shuman^{1,2*}

Integrated Program in Cellular, Molecular & Biophysical Studies,¹ Department of Microbiology,² and
Columbia Genome Center,³ Columbia University Medical Center, 701 West 168th Street,
New York, New York 10032

Received 16 May 2005/Accepted 24 August 2005

Intracellular pathogens exploit host cell functions to create a replication niche inside eukaryotic cells. The causative agent of Legionnaires' disease, the γ -proteobacterium *Legionella pneumophila*, resides and replicates within a modified vacuole of protozoan and mammalian cells. *L. pneumophila* translocates effector proteins into host cells through the Icm-Dot complex, a specialized type IVB secretion system that is required for intracellular growth. To find out if some effector proteins may have been acquired through interdomain horizontal gene transfer (HGT), we performed a bioinformatic screen that searched for eukaryotic motifs in all open reading frames of the *L. pneumophila* Philadelphia-1 genome. We found 44 uncharacterized genes with many distinct eukaryotic motifs. Most of these genes contain G+C biases compared to other *L. pneumophila* genes, supporting the theory that they were acquired through HGT. Furthermore, we found that several of them are expressed and up-regulated in stationary phase in an RpoS-dependent manner. In addition, at least seven of these gene products are translocated into host cells via the Icm-Dot complex, confirming their role in the intracellular environment. Reminiscent of the case with most Icm-Dot substrates, most of the strains containing mutations in these genes grew comparably to the parent strain intracellularly. Our findings suggest that in *L. pneumophila*, interdomain HGT may have been a major mechanism for the acquisition of determinants of infection.

The γ -proteobacterium *Legionella pneumophila* is an opportunistic human pathogen that multiplies within alveolar macrophages and causes the nosocomial and community-acquired pneumonia known as Legionnaires' disease (18, 25, 48). Human disease occurs when aerosolized *L. pneumophila* is inhaled from man-made or natural freshwater reservoirs harboring the bacteria. *L. pneumophila* poses a significant worldwide public health problem, particularly for individuals with compromised immune systems (19, 38, 40). Eradication of the pathogen from freshwater, industrial settings has proven difficult, since *L. pneumophila* thrives in environments that exclude antibacterial agents, such as biofilms and the intracellular compartments of protozoa (8, 46, 47).

In order to create a replicative niche inside eukaryotic cells suitable for replication, *L. pneumophila* is believed to modulate host cell functions by the delivery of effector proteins through a type IVB secretion system known as the Icm-Dot complex (52, 59). Effector proteins presumably regulate several pathways in the host, including up-regulation of phagocytosis (23), delay in phagosome-lysosome fusion (24), recruitment of ARF1 to the phagosome (43), acquisition of endoplasmic reticulum-derived vesicles (29), and nonlytic egress from the host cell (11).

Several laboratories found Icm-Dot substrates through genetic screens and bioinformatic approaches (3, 7, 11, 13, 36, 43, 44, 57). Most of the known effector proteins are not individu-

ally required for intracellular multiplication, since knocking out the corresponding genes does not result in growth defects within protozoa or macrophages under laboratory conditions (except if the effector is required for the integrity of the translocation apparatus itself, which is the case for LidA).

Two hypotheses have been proposed to explain the origins of virulence determinants in bacterial pathogens: convergent evolution and horizontal gene transfer (HGT) (58). In the former, the pathogen's own genetic material is modified to evolve a protein that has a function similar to that of a host protein. In the latter, the pathogen acquires foreign genetic material from eukaryotic cells that is incorporated into its genome and retains at least part of its original activity.

Of the known *L. pneumophila* effector proteins, it has been suggested that at least one (RalF) has been acquired through interdomain HGT (43). Structural studies have confirmed that the three-dimensional structure of this protein resembles the well-known eukaryotic Sec7 domain fold (1). However, it is currently unknown if interdomain HGT is a major mechanism through which *L. pneumophila* and other pathogens acquire effector proteins.

We decided to take advantage of the recently completed genome sequence of *L. pneumophila* Philadelphia-1 (12) to identify genes that may have arisen via HGT from eukaryotes to bacteria. By looking for domains of eukaryotic origin in all open reading frames of the genome, we identified, in addition to several previously identified Icm-Dot substrates (see Table 4), 46 uncharacterized genes (see Table 3). Two of these 46 genes were subsequently studied and found to be effector proteins by Roy and colleagues while the manuscript was in preparation (7). In the course of analyzing the novel genes similar

* Corresponding author. Mailing address: Department of Microbiology, Columbia University Medical Center, 701 West 168th Street, New York, New York 10032. Phone: (212) 305-6913. Fax: (212) 305-1468. E-mail: has7@columbia.edu.

to those in eukaryotes (referred to as “eukaryotic-like genes”), two other isolates of *L. pneumophila* (Paris and Lens) were sequenced, annotated, and found to have some of the same eukaryotic-like genes and/or different genes with similar domains (9) (see Table 5). However, Buchrieser and colleagues highlighted in their genome study (9) only 26 of the 46 eukaryotic-like genes described here, showing that genome annotation alone is not sufficient to find all genes with eukaryotic domains in bacterial genomes. Additionally, no experimental evidence was provided to show that the eukaryotic-like genes from the Paris and Lens isolates were expressed or involved in host-pathogen interactions.

Here, we provide evidence supporting the theory that the *Legionella* eukaryotic-like genes (*leg*) were acquired through interdomain HGT. Furthermore, we show that several of these genes are in fact expressed in *L. pneumophila* and that their expression is up-regulated in stationary phase in an RpoS-dependent manner. We also found that seven of the Leg proteins, none of which were described previously or were highlighted in genome studies, are translocated into host cells via the Icm-Dot complex. Finally, we analyzed the abilities of 20 strains containing transposon insertions in the *leg* genes to grow intracellularly. Reminiscent of strains containing knock-outs in other *Legionella* effector proteins, we found that none of the *leg* genes tested here are required for intracellular growth.

MATERIALS AND METHODS

Plasmid constructions. Plasmids and strains used in the study are summarized in Table 1. Plasmids were constructed essentially as described elsewhere (11) or using standard molecular biology techniques. To create CyaA hybrid proteins, candidate *leg* genes were amplified by PCR using Proofstart polymerase (QIAGEN, California) with primers that contained unique restriction sites at the ends. The PCR products were digested with either KpnI/XbaI or BamHI. These fragments were then ligated into the KpnI/XbaI or BamHI sites of the previously described pJC141 vector (11). For complementation studies, the same fragments were ligated into the KpnI/XbaI or BamHI sites of the pMMB207C (11) vector. For allelic exchange, we amplified 2,000 bp upstream of *legLC8* using Proofstart polymerase (QIAGEN). The PCR product was digested with EcoRV/BamHI. This fragment was then ligated into the EcoRV/BamHI sites of pLAW344 to generate pKS80 (61). We then PCR amplified 2,398 bp downstream of the *legLC8* gene and digested the product with BamHI/NotI. This fragment was ligated into the BamHI/NotI sites of pKS80 to generate pKS81. We then amplified the kanamycin resistance cassette of Tn903dIIlacZ and digested the product with BamHI. The kanamycin resistance cassette was ligated into the BamHI site of pKS81 to generate pKS82. Primer sequences are available upon request.

Strain constructions. Transposon-mutagenized strains are listed in Table 2 along with other relevant strains used in this study. The ordered transposon library and pools are described elsewhere (49). By performing PCR in the larger pools (ca. 600 mutants) with a primer annealing to the transposon and one annealing to the upstream region of the gene of interest, we were able to determine, based on the sizes of the PCR products, if there was a transposon insertion in the gene of interest in one of the large pools. If the result was positive, we then performed PCR in the smaller pools representing single microtiter plates (~60 mutants) to identify in what plate our clone of interest was located. Finally, we screened the individual plate by the same PCR technique to identify the mutant of interest. PCRs were performed using *Taq* polymerase (New England Biolabs, Massachusetts) with the following protocol: denaturation at 95°C for 5 min, followed by 35 cycles of denaturation at 95°C for 10 s, annealing at 68°C for 30 s, and amplification at 72°C for 3 min. The PCR products were then sequenced to confirm that the transposon insertion was in fact located in the gene of interest. Primers are available upon request.

We performed allelic exchange as described previously (61) to generate the isogenic *legLC8* strain KS83.

Sequence analysis. PSI-BLAST searches were performed using default parameters in the nonredundant database as of April 2005. Values for overall G+C

TABLE 1. Plasmids used in this study

Plasmid	Relevant characteristics	Reference
pMMB207C	pMMB207 $\Delta mobA$	11
pJC158	pMMB207C <i>lepA-cyaA</i>	11
pJC203	pMMB207C <i>cyaA</i>	11
pKS38	pMMB207C <i>legA3-cyaA</i>	This study
pKS12	pMMB207C <i>legA9-cyaA</i>	This study
pKS10	pMMB207C <i>legK1-cyaA</i>	This study
pKS3	pMMB207C <i>legN-cyaA</i>	This study
pKS39	pMMB207C <i>legC1-cyaA</i>	This study
pKS8	pMMB207C <i>legC2-cyaA</i>	This study
pKS40	pMMB207C <i>legC4-cyaA</i>	This study
pKS41	pMMB207C <i>legC5-cyaA</i>	This study
pKS13	pMMB207C <i>legY-cyaA</i>	This study
pKS15	pMMB207C <i>legLC8-cyaA</i>	This study
pKS9	pMMB207C <i>legLC4-cyaA</i>	This study
pKS14	pMMB207C <i>legL5-cyaA</i>	This study
pKS19	pMMB207C <i>legL7-cyaA</i>	This study
pKS21	pMMB207C <i>legL8-cyaA</i>	This study
pKS4	pMMB207C <i>legD1-cyaA</i>	This study
pKS16	pMMB207C <i>legD2-cyaA</i>	This study
pKS6	pMMB207C <i>legS2-cyaA</i>	This study
pKS11	pMMB207C <i>legP-cyaA</i>	This study
pKS42	pMMB207C <i>legG2-cyaA</i>	This study
pKS43	pMMB207C <i>legT-cyaA</i>	This study
pKS44	pMMB207C <i>legA3</i>	This study
pKS45	pMMB207C <i>legC1</i>	This study
pKS46	pMMB207C <i>legC4</i>	This study
pKS47	pMMB207C <i>legC5</i>	This study
pKS33	pMMB207C <i>legLC8</i>	This study
pKS48	pMMB207C <i>legG2</i>	This study
pKS49	pMMB207C <i>legT</i>	This study
pKS80	pLAW344, 2,000 bp upstream of <i>legLC8</i>	This study
pKS81	pKS80, 2,398 bp downstream of <i>legLC8</i>	This study
pKS82	pKS81 with Kan ^r for allelic exchange	This study

content and G+C content in codon positions 1, 2, and 3 of coding regions were calculated by SeqStat, a C program (28). The program was modified to process a large number of coding regions in FASTA format for analysis of the *Legionella pneumophila* Philadelphia-1 genome.

TABLE 2. Strains used in this study

Strain	Relevant genotype	Reference
JR32	Philadelphia-1, Sm ^r , r ⁻ m ⁺	49
LELA 3118	JR32 <i>dotA::Tn903dIIlacZ</i>	53
LELA 2265	JR32 <i>legA3::Tn903dIIlacZ</i>	This study
LELA 1678	JR32 <i>legA54::Tn903dIIlacZ</i>	This study
LELA 3626	JR32 <i>legA5::Tn903dIIlacZ</i>	This study
LELA 1155	JR32 <i>legA9::Tn903dIIlacZ</i>	This study
LELA 1309	JR32 <i>legA11::Tn903dIIlacZ</i>	This study
LELA 1925	JR32 <i>legA12::Tn903dIIlacZ</i>	This study
LELA 3026	JR32 <i>legK1::Tn903dIIlacZ</i>	This study
LELA 3141	JR32 <i>legC1::Tn903dIIlacZ</i>	This study
LELA 1616	JR32 <i>legC2::Tn903dIIlacZ</i>	This study
LELA 3073	JR32 <i>legC3::Tn903dIIlacZ</i>	This study
LELA 545	JR32 <i>legC4::Tn903dIIlacZ</i>	This study
LELA 4250	JR32 <i>legC5::Tn903dIIlacZ</i>	This study
LELA 1985	JR32 <i>legL1::Tn903dIIlacZ</i>	This study
LELA 3864	JR32 <i>legLC4::Tn903dIIlacZ</i>	This study
LELA 3750	JR32 <i>legLC8::Tn903dIIlacZ</i>	This study
LELA 614	JR32 <i>legD1::Tn903dIIlacZ</i>	This study
LELA 4136	JR32 <i>legS2::Tn903dIIlacZ</i>	This study
LELA 219	JR32 <i>legU2::Tn903dIIlacZ</i>	This study
LELA 583	JR32 <i>legG2::Tn903dIIlacZ</i>	This study
LELA 789	JR32 <i>legT::Tn903dIIlacZ</i>	This study
LM1376	JR32 <i>rpoS4::Tn903dGen^r</i>	20
KS83	JR32 <i>legLC8::Kan^r</i>	This study

FIG. 1. Flow chart of the bioinformatic screen performed to identify *Legionella* eukaryotic-like genes.

RNA extraction, reverse transcription, and real-time PCR. The *L. pneumophila* strain JR32 (a streptomycin-resistant, restriction-defective strain of the Philadelphia-1 isolate) and an isogenic *rpoS* null strain, LM1376, were grown in AYE medium (16) on a shaker at 300 rpm at 37°C. Samples of bacterial culture were removed during logarithmic and stationary phases. Total RNA was isolated using the RNeasy kit (QIAGEN). Reverse transcription was performed using 10 µg of RNA with Superscript II Plus RNase H⁻ reverse transcriptase (RT) (Invitrogen, California) and random primers. After alkaline hydrolysis of source RNA with 1 N NaOH, target cDNA was purified using the QiaQuick PCR purification kit (QIAGEN). Real-time PCR was performed using the SYBR Green PCR master mix (Applied Biosystems, California) on an ABI PRISM 7000 (Applied Biosystems) using 25 pmol of each of the gene-specific primers (available upon request). The following protocol was used: denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s and annealing-extension at 67°C for 30 s. For the amplification reaction we used 100 pg of cDNA from JR32 and LM1376 at log and stationary phases. The amplification levels were determined with the ABI PRISM 7000 SDS software (Applied Biosystems). The following equation was used to calculate fold induction in the RT-PCR experiments: Ratio = $1.7^{\Delta C_T}$, where $\Delta C_T = C_{T \text{ sample1}} - C_{T \text{ sample2}}$.

The C_T value represents the cycle at time (T) in which amplification occurs. **Translocation assays.** Translocation assays were performed essentially as described elsewhere (11), except that *Dictyostelium discoideum* was used as a host. Briefly, *Legionella pneumophila* strains carrying the *leg-cyaA* hybrids and controls were grown overnight to stationary phase with the appropriate antibiotics. The bacteria were then subcultured into fresh AYE medium (16) to a final optical density (OD) of 0.08 and grown with shaking at 37°C until they reached an OD of 0.2. Isopropyl-beta-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 1 mM to induce expression of the hybrid proteins, which are under the control of the *Ptac* promoter. The cultures were then incubated at 37°C for an additional 3 h. *Dictyostelium discoideum* DH1 cells (4×10^5) in MB medium {20 mM MES [2-(*N*-morpholino)ethanesulfonic acid] (pH 6.9), 0.7% yeast extract, 1.4% BBL thistone E peptone} containing IPTG to a final concentration of 1 mM were seeded into the wells of 96-well plates and incubated at 25°C for several hours. After IPTG induction, the bacterial cultures were spun

and resuspended in MB medium containing 1 mM IPTG to a final OD of 0.2. The *Dictyostelium discoideum* cells were then infected with 2×10^7 *L. pneumophila* cells (multiplicity of infection = 50) containing the different hybrid proteins or appropriate controls. The microtiter plate was then spun at 2,000 rpm for 15 min at 25°C to synchronize the infection. The plate was then incubated for 15 h at 25°C. The supernatant was removed, and the whole contents of the well were lysed with 100 µl of ice-cold 2.5% perchloric acid. A 0.4 M concentration of Trizma base was then added in appropriate amounts to neutralize the contents of the well. cAMP was then measured by using the cAMP ELISA Biotrak kit (Amersham, New Jersey) following the instructions of the manufacturer.

Intracellular Growth Assays in Eukaryotic Hosts. Intracellular growth curves in *Acanthamoeba castellanii* and terminally differentiated U937 cells were performed essentially as described elsewhere (11) with a multiplicity of infection of 0.001 (4×10^6 host cells and 4×10^3 bacterial cells) in 24-well tissue culture plates. Briefly, infections were synchronized by spinning the bacteria at 2,000 rpm for 10 min onto the adhered host cells. After 1 h of incubation, the wells were washed once to remove host cells that were not adhered and bacterial cells that were not internalized. Fifty-microliter aliquots were removed every 24 h, serially diluted in sterile, distilled water, and plated in charcoal-yeast extract plates (16).

URLs. PSI-BLAST is available at <http://www.ncbi.nlm.nih.gov/BLAST>, Pfam at <http://pfam.wustl.edu/>, SMART at <http://smart.embl-heidelberg.de/>, COILS at http://www.ch.embnet.org/software/COILS_form.html, PESTFIND at <http://bio-web.pasteur.fr/seqanal/interfaces/pestfind.html>, and the Conserved Domains Database at <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>.

RESULTS

Genomewide screen for *L. pneumophila* eukaryotic-like genes. To identify *L. pneumophila* genes containing "eukaryotic domains," we performed computational screens using the web-based programs PSI-BLAST, Pfam, SMART, and COILS (Fig. 1). We scanned every open reading frame (ORF) of the *L. pneumophila* Philadelphia-1 genome (12) for those that

contain specific sequences at the amino acid level that can be classified as “eukaryotic domains.” We consider “eukaryotic domains” to be protein motifs that are widespread in eukaryotic species and significantly unrepresented in archaeal and prokaryotic species (45) and that have known cellular functions associated with eukaryotes. We also used PSI-BLAST to confirm that the full-length ORFs are not widespread in other bacterial species. The *Legionella* eukaryotic-like genes identified by this screen are described in Table 3. Confirming the validity of the approach as a tool to find putative effector proteins, we identified several previously characterized Icm-Dot substrates that also have eukaryotic-like domains (Table 4). The genes identified in this bioinformatics screen include approximately 60% of the previously confirmed Icm-Dot substrates. The expect values listed in Table 4 pertain to the alignment of the domain regions only against the current consensus sequences available in the conserved domains database. These expect values for the domain regions ranged from 5×10^{-4} to 6×10^{-50} . When selecting hits containing coiled-coil domains, we took into consideration only ORFs that had regions of predicted coiled coils for stretches larger than 50 residues and that did not resemble known bacterial coiled-coil proteins based on sequence alignments. Since the current number of complete eukaryotic genomic sequences available is relatively small, it is difficult to predict the flow of horizontal gene transfer by analyzing the gene of the organism to which a given *Legionella* eukaryotic-like gene aligns best. Also, we found that while the regions pertaining to the predicted eukaryotic domains in the *leg* genes align significantly with the consensus sequences of the domains and with the sequences of several eukaryotic genes containing the same types of domains, the remaining parts of most of the genes do not have any clear orthologs in the database (data not shown). This may suggest that a certain degree of amelioration may have occurred since the gene was first acquired or that closer orthologs will be identified only as many more eukaryotic genomes are sequenced. During this work, two other *L. pneumophila* isolates were fully sequenced and annotated (9, 12). This genome study highlighted only 26 out of the 46 genes described here (Table 5). This argues strongly that gene annotation alone greatly underestimates the number of eukaryotic-like genes in bacterial genomes.

leg gene domains. A variety of domains were found in the *leg* genes. Most of these correspond to protein binding motifs commonly found in eukaryotes. Several are predicted to encode products with enzymatic activities related to eukaryotic cell functions. A few, however, are not fully characterized even in eukaryotes. Following is a brief description of the domains and their prevalence in bacterial pathogens.

Ankyrin repeats. The gene products LegA1 to -15 are predicted to contain one or more ankyrin repeats, which are interaction domains usually found in eukaryotic proteins that play a role in the attachment of cytoskeletal components (41). Other bacterial pathogens are also known to carry genes with this type of domain (10, 39), including the closely related intracellular bacterial pathogen *Coxiella burnetii* (the causative agent of Q fever), which contains 13 genes with this type of domain in its genome (55), one of which is conserved in *L. pneumophila* (*legA3* and CBU1292; $E = 1 \times 10^{-93}$). Also, the parasitic endosymbiont *Wolbachia pipientis* contains several

proteins with ankyrin repeats speculated to play a role in its intracellular lifestyle (26, 62).

Leucine-rich repeats. Leucine-rich repeat domains, which usually form protein-binding interfaces (32), were identified in eight of the *leg* genes (*legL1* to *LC8*). Eukaryotic proteins with these domains participate in a wide range of functions. Leucine-rich repeat proteins have also been identified in other bacterial pathogens (30), such as *Salmonella enterica* serovar Typhimurium, *Yersinia pestis*, *Ralstonia solanacearum*, *Shigella flexneri*, and the obligate amoebal symbiont *Parachlamydia*. However, leucine-rich repeat-containing proteins are rarely found in nonpathogenic bacteria but are present in nearly all sequenced eukaryotic species (45).

Coiled coils. Extended regions of coiled coils usually form protein interaction domains that can participate in multiple cell functions, including the formation of fusogenic complexes through SNARE (soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor) proteins (6). Several of the currently known Icm-Dot substrates possess eukaryotic-like coiled-coil regions (11, 36). *legC1* to *-C8*, *legLC4*, and *legLC8* encode products with coiled-coil regions.

Ser/Thr kinases. The *L. pneumophila* genome contains three ORFs predicted to encode Ser/Thr kinases (*legK1* to *-K3*). The presence of kinase effector proteins in bacterial pathogens whose substrates are host cell proteins is not unprecedented (58). LegK1 and *-K3* contain the conserved glycine loop motif and a catalytic aspartate preceded by an arginine—typical features of Ser/Thr kinases. Besides its enzymatic active site, LegK1 has another interesting property: its C terminus is similar to a region in the *Salmonella* effector protein SseG previously shown to contain a golgi localization domain (50).

Signaling lipid-related domains. The *legS2* gene contains an active site predicted to modify the eukaryotic signaling lipid sphingosine 1-phosphate. Sphingosine 1-phosphate can cause the mobilization of Ca^{2+} ions from the endoplasmic reticulum, which in turn activates pathways important for phagocytosis and proper phagosome maturation (37). LegS1 has a predicted lipid phosphoesterase active site that aligns weakly with sphingomyelinases. To our knowledge, similar active sites have not been identified in other prokaryotes.

GDP-GTP exchange. Two of the *leg* genes (*legG1* and *legG2*) are believed to be involved in GDP-to-GTP exchange in GTP-binding proteins. The gene *legG1* has an RCC1 domain which can act as a guanine-nucleotide dissociation stimulator for the Ran GTP-binding protein (54), while *legG2* has a Ras GEF domain, which exchanges GDP for GTP in Ras. Other pathogens also contain proteins that modify eukaryotic nucleotide-binding proteins (58); however, such proteins are rare in other bacterial species (45).

Ubiquitin-related domains. Two gene products, LegAU13 and LegU1, contain an FBOX domain, which can link target proteins to a ubiquitin ligase in eukaryotes (31). LegAU13 contains both ankyrin repeats and an FBOX domain. LegU2 contains a UBOX domain, which, in eukaryotes, may possess E3 ubiquitin ligase activity (22). To our knowledge, such domains are exclusive to eukaryotes.

Others. The other five predicted enzymes among the *leg* gene products include one SET lysine methyltransferase (*legAS4*), one amylase (*legY*), two putative phytanoyl-coenzyme A (CoA) dioxygenases (*legD1* and *-D2*), and one astacin zinc protease

TABLE 3. Summary of *Legionella* eukaryotic-like genes identified in this study

Gene	ORF	Domain(s)	Best E value	Reference
<i>legA1</i>	lpg2416	Ankyrin repeats	4×10^{-7}	This study; 9
<i>legA2</i>	lpg2215	Ankyrin repeats	5×10^{-4}	This study; 9
<i>legA3</i>	lpg2300	Ankyrin repeats	8×10^{-16}	This study; 9
<i>legAS4</i>	lpg1718	Ankyrin repeats and SET	2×10^{-7}	This study; 9
<i>legA5</i>	lpg2322	Ankyrin repeats	1×10^{-4}	This study; 9
<i>legA6</i>	lpg2131	Ankyrin repeats	3×10^{-6}	This study; 9
<i>legA7</i>	lpg0403	Ankyrin repeats	1×10^{-15}	This study; 9
<i>legA8</i>	lpg0695	Ankyrin repeats	1×10^{-16}	This study; 9
<i>legA9</i>	lpg0402	Ankyrin repeats	5×10^{-20}	This study; 9
<i>legA10</i>	lpg0038	Ankyrin repeats	2×10^{-11}	This study; 9
<i>legA11</i>	lpg0436	Ankyrin repeats	4×10^{-6}	This study; 9
<i>legA12</i>	lpg0483	Ankyrin repeats	2×10^{-4}	This study; 9
<i>legAU13</i>	lpg2144	Ankyrin repeats and FBOX	1×10^{-5}	This study; 9
<i>legA14</i>	lpg2452	Ankyrin repeats	2×10^{-10}	This study; 9
<i>legA15</i>	lpg2456	Ankyrin repeats	6×10^{-9}	This study; 9
<i>legL1</i>	lpg0945	Leucine-rich repeats	2×10^{-11}	This study
<i>legL2</i>	lpg1602	Leucine-rich repeats	3×10^{-6}	This study
<i>legL3</i>	lpg1660	Leucine-rich repeats	9×10^{-6}	This study
<i>legLC4</i>	lpg1948	Leucine-rich repeats and coiled coils	6×10^{-8}	This study
<i>legL5</i>	lpg1958	Leucine-rich repeats	2×10^{-15}	This study
<i>legL6</i>	lpg2392	Leucine-rich repeats	8×10^{-9}	This study
<i>legL7</i>	lpg2400	Leucine-rich repeats	3×10^{-9}	This study
<i>legLC8</i>	lpg1890	Leucine-rich repeats and coiled coils	4×10^{-8}	This study
<i>legC1</i>	lpg1312	Coiled coils	NA ^d	This study
<i>legC2/ylfB</i>	lpg1884	Coiled coils	NA	This study; 7
<i>legC3</i>	lpg1701	Coiled coils	NA	This study
<i>legC4</i>	lpg1953	Coiled coils	NA	This study
<i>legC5</i>	lpg1488	Coiled coils	NA	This study
<i>legC6</i>	lpg1588	Coiled coils	NA	This study
<i>legC7/ylfA</i>	lpg2298	Coiled coils	NA	This study; 7
<i>legC8</i>	lpg2862	Coiled coils	NA	This study
<i>legLC4</i>	lpg1948	Leucine-rich repeats and coiled coils	NA	This study
<i>legLC8</i>	lpg1890	Leucine-rich repeats and coiled coils	NA	This study
<i>legK1</i>	lpg1483	Ser/Thr kinase	2×10^{-22}	This study; 9
<i>legK2</i>	lpg2137	Ser/Thr kinase	2×10^{-5}	This study; 9
<i>legK3</i>	lpg2556	Ser/Thr kinase	9×10^{-22}	This study; 9
<i>legS1</i>	lpg2588	Lipid phosphoesterase ^a	2×10^{-4}	This study
<i>legS2</i>	lpg2176	Sphingosine 1-P lyase ^a	6×10^{-50}	This study; 9
<i>legG1</i>	lpg1976	RCC-1 ^b	7×10^{-9}	This study; 9
<i>legG2</i>	lpg0276	Ras GEF ^b	2×10^{-8}	This study
<i>legAU13</i>	lpg2144	Ankyrin repeats and FBOX ^c	4×10^{-6}	This study; 9
<i>legU1</i>	lpg0171	FBOX ^c	2×10^{-7}	This study; 9
<i>legU2</i>	lpg2830	UBOX ^c	7×10^{-9}	This study; 9
<i>legAS4</i>	lpg1718	Ankyrin repeats and SET	1×10^{-15}	This study; 9
<i>legN</i>	lpg2720	cNMP binding domain	2×10^{-14}	This study
<i>legY</i>	lpg0422	Amylase	2×10^{-52}	This study; 9
<i>legD1</i>	lpg2694	Phytanoyl-CoA dioxygenase	2×10^{-29}	This study; 9
<i>legD2</i>	lpg0515	Phytanoyl-CoA dioxygenase	5×10^{-17}	This study; 9
<i>legP</i>	lpg2999	Astacin protease	1×10^{-52}	This study; 9
<i>legT</i>	lpg1328	Thaumatococcus domain	1×10^{-9}	This study

^a Signaling lipid-related domain.^b GDP-GTP exchange.^c Ubiquitination-related domain.^d NA, not available.

(*legP*). The gene product of *legN* encodes a unique motif within this collection of genes. It contains two predicted cNMP binding domains in the middle portion of the ORF. The LegT protein contains a thaumatococcus domain, which makes part of the more generalized class of domains found in plant pathogene-

sis-related proteins, generally thought to be membrane-permeabilizing factors (4).

Evidence for interdomain horizontal gene transfer. One theory to explain the presence of eukaryotic domains in the *leg* genes is that they may have been acquired by interdomain

TABLE 4. Summary of previously characterized genes that were also identified in this study's bioinformatic screen

Gene	ORF	Domain(s)	Best E value	Reference
<i>ralF</i>	lpg1950	Sec7 (ARF GEF)	3×10^{-41}	43
<i>lepA</i>	lpg2793	Coiled coils	NA ^a	11
<i>lepB</i>	lpg2490	Coiled coils	NA	11
<i>sidA</i>	lpg0621	Coiled coils	NA	36
<i>sidB</i>	lpg1642	Coiled coils	NA	36
<i>sidC</i>	lpg2511	Coiled coils	NA	36
<i>sidA</i>	lpg2510	Coiled coils	NA	36
<i>sidE</i>	lpg0234	Coiled coils	NA	36
<i>sdeA</i>	lpg2157	Coiled coils	NA	36
<i>sdeB</i>	lpg2156	Coiled coils	NA	36
<i>sdeC</i>	lpg2153	Coiled coils	NA	36
<i>sidF</i>	lpg2584	Coiled coils	NA	36
<i>sidG</i>	lpg1355	Coiled coils	NA	36
<i>sidH</i>	lpg2829	Coiled coils	NA	36
<i>sdhA</i>	lpg0376	Coiled coils	NA	36
<i>sdhB</i>	lpg0135	Coiled coils	NA	36

^a NA, not available.

HGT. Koonin and colleagues have done extensive sequence analysis to identify the origin of genes with eukaryotic domains in bacteria (45). The presence of such genes in both the bacterial and eukaryotic lineages and not the archaeal lineage contradicts the notion of common ancestry, since if genes with these domains had arisen first in bacteria, they should have been transmitted vertically to both eukaryotes and archaea after the division of the domains. This suggests either that such genes were lost very early on in the archaeal lineage or that they were horizontally transferred between eukaryotes and bacteria. Even though interdomain horizontal gene transfer is not a well-characterized phenomenon, there are several reports suggesting that genetic material can be exchanged between prokaryotes and eukaryotes (5, 14, 21, 33, 51).

Two lines of evidence support the theory that at least some of the *leg* genes may have been acquired through interdomain HGT. First, the domains found in the *leg* genes are present in eukaryotes and, in a few cases, in a small number of other bacterial species (mostly pathogenic or endosymbiotic ones),

but not in archaea (45). The presence of the eukaryotic domains found in the *leg* genes in few other bacterial species raises the possibility that some of the *leg* genes were acquired prior to the modern-day speciation of *L. pneumophila* and/or that interbacterial HGT occurred after the gene was originally acquired from a eukaryote. The second line of evidence is that the *leg* genes differ from the average *L. pneumophila* coding region values in both overall G+C content and G+C content by codon position. Differences in G+C content and distribution can often be seen in genes acquired from another species by HGT (35). For comparison, we selected 34 *L. pneumophila* "informational" genes, such as those involved in transcription and translation. These genes are less likely to be subject to HGT (27). We then calculated the overall G+C content and the G+C content in codon positions 1, 2, and 3 for all 46 *leg* genes (Table 3), the 34 "informational" genes, and all 3,005 predicted open reading frames in the *L. pneumophila* Philadelphia-1 genome. The *leg* genes differ in their overall G+C content and G+C content in codon positions 1 and 2 from *L. pneumophila* genomic averages and from the "informational" genes (two-tailed *P* values obtained from *t* tests are provided in Fig. 2). We also found that the previously characterized genes for Icm-Dot substrates identified in our bioinformatics screen (Table 4) have G+C content biases similar to those of the *leg* genes (data not shown). These differences are consistent with the hypothesis that the *leg* genes and other effectors may have been acquired through HGT from a species with sequence characteristics distinct from those of *L. pneumophila*. The fact that the G+C content in codon position 3, the "wobble" codon, did not differ significantly suggests that there has been sufficient time for amelioration to the genomic average in codon position 3. The "wobble" codon allows for the greatest proportion of synonymous mutations and can ameliorate more quickly than codon positions 1 or 2 (34).

RpoS-dependent expression. Several studies have suggested that in *L. pneumophila* the sigma factor RpoS regulates genes involved in intracellular replication (2, 20, 36, 63). We decided to find out if the *leg* genes are actually expressed and if they are regulated by RpoS. We chose the genes *legA3*, *-AS4*, *-K1*, *-C1*, *-C3*, *-C4*, *-C5*, *-L5*, *-L7*, *-LC8*, *-G2*, and *-T* because they rep-

TABLE 5. Domain and gene conservation in different *Legionella pneumophila* isolates

Domain	No. of genes for isolate			Reference(s)
	Philadelphia-1	Paris	Lens	
Ankyrin	15 (1 unique)	20 (4 unique)	20 (2 unique)	This study; 9
FBOX	2	2 (1 unique)	2	This study; 9
UBOX	1	1	0	This study; 9
Sphingosine 1-P lyase	1	1	1	This study; 9
Astacin protease	1	1	1	This study; 9
Phytanoyl CoA dioxygenase	2	2	2	This study; 9
RCC-1 (Ran GEF)	1	1	1	This study; 9
S/T kinases	3	3	3	This study; 9
cNMP-binding	1	1	1	This study
Amylase	1	1	1	This study; 9
Lipid phosphoesterase	1	1	1	This study
Leucine-rich repeats	8 (1 unique)	6	6	This study
Thaumatococcus	1	1	1	This study
Coiled coils	26 (2 unique)	26	26	This study; 7, 11, 36
Sec7	1	1	1	This study; 43
Ras GEF	1	1	1	This study

FIG. 2. Dot plot analysis showing overall and codon position G+C content differences in the *leg* genes. The plot shows overall G+C content (A) and content of codon position 1 (B), codon position 2 (C), and codon position 3 (D) in the coding regions. The solid horizontal lines represent the average values for all 3,005 predicted *L. pneumophila* genes. The closed squares represent the values for each of the 46 *leg* genes (the genes are organized on the x axis in order of their appearance in the chromosome). The open triangles represent the values for each of the 34 “informational” genes thought less likely to be the subject of HGT (the genes are organized on the x axis in order of their appearance in the chromosome). In general, the *leg* genes differ in their overall nucleotide content and nucleotide content in codon positions 1 and 2. Two-tailed *P* values obtained from *t* tests are provided.

resent distinct types of domains. We used 16S rRNA and Mip (15) as negative controls and the gene *lpg0981* as a positive control based on our preliminary studies showing that Mip and 16S rRNA are not up-regulated in stationary phase, while *lpg0981* is highly up-regulated in stationary phase in an RpoS-dependent manner (S. Pampou, S. Kalachikov, and H. A. Shuman, unpublished data). We isolated RNA from cells in logarithmic phase and in stationary phase both in an *rpoS*⁺ strain (JR32) and in an isogenic *rpoS* mutant (LM1376) and performed RT-PCR for the 12 genes listed above and the controls. We calculated the induction ratio for each gene based on the equation described in Methods. All the *leg* genes selected are expressed and up-regulated in stationary phase in an RpoS-dependent manner, albeit to various degrees (Fig. 3).

Translocation of Leg proteins into a eukaryotic host. *L. pneumophila* encodes a type IVB secretion system called the Icm-Dot complex that translocates proteins from the pathogen into host cells (11, 36, 42, 43). The use of gene fusions to the catalytic domain of the calmodulin-dependent adenylate cyclase protein, CyaA of *Bordetella pertussis*, provides a convenient and reproducible method for measuring hybrid protein translocation from bacteria to eukaryotic cells (11, 42). In order to find out if the Leg proteins are translocated into host cells, we constructed hybrid proteins with CyaA, expressed the proteins in *L. pneumophila*, and then measured the capacity of these fusion proteins to be translocated into *Dictyostelium discoideum*. Increases in the level of cAMP in the host cell are indicative of translocation.

We amplified 20 candidate *leg* genes and created C-terminal

translational gene fusions to the catalytic domain coding region of *cyaA*. We measured cAMP production following infection by both JR32 and *dotA* (LELA 3118) *L. pneumophila* expressing these genes. The *dotA* strain of *L. pneumophila* is unable to translocate effector proteins into the host because of a mutation in one of the essential components of the type IVB secretion system (11, 42). We considered the Leg-CyaA proteins to be translocated in an Icm-Dot-dependent manner if there were increases in the levels of host cAMP when using *dot*⁺ *L. pneumophila* and no increases when using the *dotA* mutant of *L. pneumophila* expressing the *leg-cyaA* hybrids. We were able to detect translocation for 8 out of 20 *leg* gene products tested (Fig. 4). Recent work by Roy and colleagues confirms the translocation of YlfB/LegC2 and shows that YlfA/LegC7, which was not tested in this study, is also a translocated eukaryotic-like protein (7). A lack of signal does not mean that the gene product is not translocated, since there are several reasons that could cause the assay to be unsuccessful for any particular hybrid protein. For instance, it has been shown that for the RalF effector, C-terminal CyaA fusions prevent translocation of the hybrid into host cells (42), while for the LepA and LepB effectors, C-terminal CyaA fusions do not block translocation (11).

Role of the *leg* genes in intracellular multiplication. To evaluate the contribution of the *leg* genes to intracellular growth, we took advantage of an ordered library of transposon insertions in the *L. pneumophila* genome. We were able to find 20 *leg* knockout strains (*legA3*, *-AS4*, *-A5*, *-A9*, *-A11*, *-A12*, *-L1*,

FIG. 3. Stationary-phase up-regulation of *leg* genes in an RpoS-dependent manner. We performed RT-PCR for 12 *leg* genes, 2 negative controls (16S rRNA and Mip), and 1 positive control (lpg0981). We isolated bacterial RNA from cells growing in stationary and logarithmic phases in both the parent wild-type (JR32) and *rpoS* strains using standard RNA isolation procedures. We performed quantitative PCR to estimate the amount of mRNA present in the cell at the time of isolation. The data are plotted as a ratio of induction based on what cycle of the PCR amplification started to occur. The black bars represent the ratio of stationary-phase cells to logarithmic-phase cells in the JR32 background, while the gray bars represent the ratio of JR32 to *rpoS* backgrounds in stationary phase. A ratio greater than 2 is interpreted to be up-regulation of the gene under the assigned condition. All *leg* genes tested are up-regulated in stationary phase in an RpoS-dependent manner. The values shown represent the means for two independently grown bacterial cultures $\pm$ standard deviations.

-LC4, -LC8, -C1, -C2, -C3, -C4, -C5, -K1, -S2, -G2, -U2, -D1, and -T) of the 44 searched for in the transposon library.

To find out if the *leg* gene products are required for intracellular growth, we tested the abilities of the 20 knockout strains to grow within *Acanthamoeba castellanii*. Most of the tested strains grow comparably to the parent strain in this host. However, the strain containing a transposon insertion in gene *legLC8* is unable to grow to the same levels of JR32 (data not shown). Despite the fact that we could complement this growth defect by providing the gene in *trans*, when we constructed a full deletion of the gene by allelic exchange we did not further observe an intracellular growth defect (data not shown). This raises the possibility that an N-terminal fragment of the protein that is expressed has a negative effect in the context of intracellular growth. We conclude that consistent with previous findings from several laboratories (7, 36, 43, 44), most strains lacking the genes for putative effector proteins do not have intracellular growth defects.

DISCUSSION

Currently, there are two nonexclusive theories to explain how bacterial pathogens acquire genes that increase their ability to exploit host cells. One theory suggests that mutations in the bacterial chromosome and selective pressure can generate novel factors that lead to increased infectivity (58). The second theory proposes that these factors are acquired horizontally from host cells or other eukaryotic cells with conserved biological pathways. Additionally, the “food hypothesis” suggests

that certain unicellular organisms that feed on bacteria are exposed to prokaryotic DNA that may be stably incorporated into their chromosomes (14, 17, 33). The “food hypothesis” can be extended to explain the flow of genes in the opposite direction: intracellular bacteria that parasitize on eukaryotic cells are exposed to eukaryotic DNA and may take up the foreign genetic material to enhance their infectivity. In addition, it has been reported that several pathogens, including *L. pneumophila*, are naturally competent (56), offering a plausible mechanism for the acquisition of eukaryotic DNA. It has been hypothesized that bacterial effector proteins have been acquired by both horizontal gene transfer and convergent evolution (58), and yet it is not clear which mechanism is predominant or if the lifestyle of a particular pathogen favors one mechanism over the other. Our findings suggest that in *L. pneumophila*, HGT may have been a major means through which effector proteins were acquired.

Eukaryotic-like genes in bacterial genomes are not always identified by genome annotation alone. Specific bioinformatics screens that search for genes that could have been acquired by eukaryotes can be considerably more sensitive. Here, for instance, we identified 20 genes that were not highlighted before in genome studies, despite the fact that three *L. pneumophila* strains had been fully sequenced and annotated. Importantly, among these 20 genes at least 8 are translocated into eukaryotic hosts. It is interesting that we could detect translocation for only three classes of *leg* genes: those containing coiled coils, leucine-rich repeats, or the one putative effector containing a Ras GEF domain. Since a CyaA translocation assay can fail for

FIG. 4. Measurement of Icm-Dot-dependent translocation of several Leg-CyaA hybrid proteins into *D. discoideum*. JR32 and the isogenic strain *dotA* harboring constructs with Leg-CyaA fusion proteins were used to infect *D. discoideum* in microtiter plates. We obtained the cAMP levels of infected cells for JR32 and *dotA* strains harboring the hybrid proteins. The proteins LegL3, LegL4, LegL5, LegL7, LegL8, LegC2, LegC5, and LegG2 are considered to be translocated, since there is an increase in host cAMP levels following infection using the JR32 strain and no significant increase over those for *D. discoideum* cells alone or when using the Icm-Dot-defective mutant *dotA*. LepA-CyaA served as a positive control, an effector protein previously shown to be translocated (11). The results shown represent the means of cAMP levels/well obtained from triplicate wells $\pm$ standard deviations.

many reasons, it will be interesting to determine if the other *leg* genes are indeed translocated and if they play a role in virulence. Since many paralogs of certain eukaryotic-like genes exist (presumably a product of gene duplication), it is possible that only some are translocated and that the remaining ones may eventually be lost from the genome.

Several theories have been proposed to explain the lack of intracellular multiplication defects of *L. pneumophila* strains containing mutations in the genes that encode effector proteins. These theories include functional redundancy, nonnatural laboratory conditions, and host specificity. We studied mutations in 20 of the *leg* genes (including 8 whose gene products are translocated) and found that none are required for intracellular multiplication. Given that many of the *leg* genes have the same type of eukaryotic domains, it seems that there may be significant functional redundancy among this collection of genes. It seems counterintuitive that *L. pneumophila* would retain so many genes that can fully substitute for one another. It has been argued that retaining functionally redundant genes may prevent single genetic lesions from making the bacteria unable to multiply intracellularly. However, a single mutation in several genes of the Icm-Dot complex can render *L. pneumophila* completely avirulent. The *leg* genes identified here contain domains that seem to have been duplicated and shuffled throughout the chromosome. It is therefore possible that a negative pressure exists on the *leg* genes and that the high level of redundancy may eventually be reduced. It is interesting that a few of the *leg* genes are unique to either the Paris, Lens, or Philadelphia-1 isolates, indicating that deletions and/or duplications of the eukaryotic-like genes are still occurring in isolates of the same species. Alternatively, it is possible that

genes that seem functionally redundant may actually play an essential role under particular environmental conditions or in other hosts that were not tested. This seems like a more plausible hypothesis to explain the lack of intracellular growth defect in strains containing mutations in effector proteins that do not have paralogs within the genome, such as RalF or LegG2.

Many of the eukaryotic-like genes (36 out of the 46 *leg* genes) encode putative protein binding domains. It should be possible to identify host cell interaction partners for these Leg proteins, particularly the ones for which we could detect translocation. Most of the other Leg proteins contain predicted enzymatic active sites. It will be interesting to determine if the three predicted Ser/Thr kinases have host substrates, since there is such a precedent for phosphorylation of host proteins by kinases in other pathogenic bacteria (58, 60). Also, *L. pneumophila* has two proteins with domains predicted to modify eukaryotic signaling lipids that participate in intracellular trafficking. An attractive hypothesis is that *Legionella* modifies both host proteins and signaling lipids in order to subvert organelle trafficking.

The large number of domains encoded by the *leg* genes reinforces the theory that *L. pneumophila* uses a varied arsenal of proteins to modulate intracellular trafficking to its own advantage. This suggests that this pathogen subverts multiple host cell functions during different stages of infection and that the biogenesis of a replicative vacuole probably involves many cellular factors. The large number of effector proteins in *L. pneumophila* may also help explain why this pathogen is capable of multiplying within a variety of phylogenetically unrelated eukaryotic hosts. Understanding the origin and mechanistic

function of the *leg* proteins will certainly help clarify how *L. pneumophila* acquired effector proteins and how they subvert host cell functions. It will also be informative to find out if other pathogens encode such a large number of eukaryotic-like genes and if interdomain HGT is a general mechanism for the acquisition of factors necessary for bacterial pathogenesis.

ACKNOWLEDGMENTS

We are grateful to the reviewers for many insightful suggestions. We thank David Figurski, Paul Planet, Richard Sugang, Jonathan Dworin, Sean Murray, and all members of the Shuman and Kessin laboratories for technical advice and helpful discussions.

This work was supported by NIH grants AI23549 (H.A.S.) and AI64481 (H.A.S.).

REFERENCES

- Amor, J. C., J. Swails, X. Zhu, C. R. Roy, H. Nagai, A. Ingmundson, X. Cheng, and R. A. Kahn. 2005. The structure of RalF, an ADP-ribosylation factor guanine nucleotide exchange factor from *Legionella pneumophila*, reveals the presence of a cap over the active site. *J. Biol. Chem.* **280**:1392–1400.
- Bachman, M. A., and M. S. Swanson. 2001. RpoS co-operates with other factors to induce *Legionella pneumophila* virulence in the stationary phase. *Mol. Microbiol.* **40**:1201–1214.
- Bardill, J. P., J. L. Miller, and J. P. Vogel. 2005. IcmS-dependent translocation of SdeA into macrophages by the *Legionella pneumophila* type IV secretion system. *Mol. Microbiol.* **56**:90–103.
- Batalia, M. A., A. F. Monzingo, S. Ernst, W. Roberts, and J. D. Robertus. 1996. The crystal structure of the antifungal protein zeamatin, a member of the thaumatin-like, PR-5 protein family. *Nat. Struct. Biol.* **3**:19–23.
- Budd, A., S. Blandin, E. A. Levashina, and T. J. Gibson. 2004. Bacterial alpha2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biol.* **5**:R38.
- Burkhard, P., J. Stetefeld, and S. V. Strelkov. 2001. Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* **11**:82–88.
- Campodonico, E. M., L. Chesnel, and C. R. Roy. 2005. A yeast genetic system for the identification and characterization of substrate proteins transferred into host cells by the *Legionella pneumophila* Dot/Icm system. *Mol. Microbiol.* **56**:918–933.
- Cargill, K. L., B. H. Pyle, R. L. Sauer, and G. A. McFeters. 1992. Effects of culture conditions and biofilm formation on the iodine susceptibility of *Legionella pneumophila*. *Can. J. Microbiol.* **38**:423–429.
- Cazalet, C., C. Rusniok, H. Bruggemann, N. Zidane, A. Magnier, L. Ma, M. Tichit, S. Jarraud, C. Bouchier, F. Vandenesch, F. Kunst, J. Etienne, P. Glaser, and C. Buchrieser. 2004. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat. Genet.* **36**:1165–1173.
- Chang, J. H., J. M. Urbach, T. F. Law, L. W. Arnold, A. Hu, S. Gombar, S. R. Grant, F. M. Ausubel, and J. L. Dangel. 2005. A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*. *Proc. Natl. Acad. Sci. USA* **102**:2549–2554.
- Chen, J., K. S. de Felipe, M. Clarke, H. Lu, O. R. Anderson, G. Segal, and H. A. Shuman. 2004. *Legionella* effectors that promote nonlytic release from protozoa. *Science* **303**:1358–1361.
- Chien, M., I. Morozova, S. Shi, H. Sheng, J. Chen, S. M. Gomez, G. Asamani, K. Hill, J. Nuara, M. Feder, J. Rineer, J. J. Greenberg, V. Steshenko, S. H. Park, B. Zhao, E. Teplitskaya, J. R. Edwards, S. Pampou, A. Georgioui, I. C. Chou, W. Iannuccilli, M. E. Ulz, D. H. Kim, A. Geringer-Sameth, C. Goldsberry, P. Morozov, S. G. Fischer, G. Segal, X. Qu, A. Rzhetsky, P. Zhang, E. Cayanis, P. J. De Jong, J. Ju, S. Kalachikov, H. A. Shuman, and J. J. Russo. 2004. The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* **305**:1966–1968.
- Conover, G. M., I. Derre, J. P. Vogel, and R. R. Isberg. 2003. The *Legionella pneumophila* LidA protein: a translocated substrate of the Dot/Icm system associated with maintenance of bacterial integrity. *Mol. Microbiol.* **48**:305–321.
- de Koning, A. P., F. S. Brinkman, S. J. Jones, and P. J. Keeling. 2000. Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Mol. Biol. Evol.* **17**:1769–1773.
- Engleberg, N. C., C. Carter, D. R. Weber, N. P. Cianciotto, and B. I. Eisenstein. 1989. DNA sequence of *mip*, a *Legionella pneumophila* gene associated with macrophage infectivity. *Infect. Immun.* **57**:1263–1270.
- Feeley, J. C., R. J. Gibson, G. W. Gorman, N. C. Langford, J. K. Rasheed, D. C. Mackel, and W. B. Baine. 1979. Charcoal-yeast extract agar: primary isolation medium for *Legionella pneumophila*. *J. Clin. Microbiol.* **10**:437–441.
- Field, J., B. Rosenthal, and J. Samuelson. 2000. Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. *Mol. Microbiol.* **38**:446–455.
- Fields, B. S., J. M. Barbaree, E. B. Shotts, Jr., J. C. Feeley, W. E. Morrill, G. N. Sanden, and M. J. Dykstra. 1986. Comparison of guinea pig and protozoan models for determining virulence of *Legionella* species. *Infect. Immun.* **53**:553–559.
- Gump, D. W., and M. Keegan. 1986. Pulmonary infections due to *Legionella* in immunocompromised patients. *Semin. Respir. Infect.* **1**:151–159.
- Hales, L. M., and H. A. Shuman. 1999. The *Legionella pneumophila* *rpoS* gene is required for growth within *Acanthamoeba castellanii*. *J. Bacteriol.* **181**:4879–4889.
- Hall, C., S. Brachat, and F. S. Dietrich. 2005. Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. *Eukaryot. Cell* **4**:1102–1115.
- Hatakeyama, S., and K. I. Nakayama. 2003. U-box proteins as a new family of ubiquitin ligases. *Biochem. Biophys. Res. Commun.* **302**:635–645.
- Hilbi, H., G. Segal, and H. A. Shuman. 2001. Icm/dot-dependent upregulation of phagocytosis by *Legionella pneumophila*. *Mol. Microbiol.* **42**:603–617.
- Horwitz, M. A., and F. R. Maxfield. 1984. *Legionella pneumophila* inhibits acidification of its phagosome in human monocytes. *J. Cell Biol.* **99**:1936–1943.
- Horwitz, M. A., and S. C. Silverstein. 1980. Legionnaires' disease bacterium (*Legionella pneumophila*) multiples intracellularly in human monocytes. *J. Clin. Invest.* **66**:441–450.
- Iturbe-Ormaetxe, I., G. R. Burke, M. Riegler, and S. L. O'Neill. 2005. Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J. Bacteriol.* **187**:5136–5145.
- Jain, R., M. C. Rivera, and J. A. Lake. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. USA* **96**:3801–3806.
- Jovanovic, O. S. 2002. Computational, genetic and biochemical analysis of broad host-range plasmid RK2. Doctoral thesis. Columbia University, New York, N.Y.
- Kagan, J. C., and C. R. Roy. 2002. *Legionella* phagosomes intercept vesicular traffic from endoplasmic reticulum exit sites. *Nat. Cell Biol.* **4**:945–954.
- Kedzierski, L., J. Montgomery, J. Curtis, and E. Handman. 2004. Leucine-rich repeats in host-pathogen interactions. *Arch. Immunol. Ther. Exp. (Warsz)* **52**:104–112.
- Kipreos, E. T., and M. Pagano. 2000. The F-box protein family. *Genome Biol* **1**:REVIEWS3002.
- Kobe, B., and J. Deisenhofer. 1995. Proteins with leucine-rich repeats. *Curr. Opin. Struct. Biol.* **5**:409–416.
- Kondo, N., N. Nikoh, N. Ijichi, M. Shimada, and T. Fukatsu. 2002. Genome fragment of *Wolbachia endosymbiont* transferred to X chromosome of host insect. *Proc. Natl. Acad. Sci. USA* **99**:14280–14285.
- Lawrence, J. G., and H. Ochman. 1997. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44**:383–397.
- Lawrence, J. G., and H. Ochman. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* **95**:9413–9417.
- Luo, Z. Q., and R. R. Isberg. 2004. Multiple substrates of the *Legionella pneumophila* Dot/Icm system identified by interbacterial protein transfer. *Proc. Natl. Acad. Sci. USA* **101**:841–846.
- Malik, Z. A., C. R. Thompson, S. Hashimi, B. Porter, S. S. Iyer, and D. J. Kusner. 2003. Cutting edge: Mycobacterium tuberculosis blocks Ca²⁺ signaling and phagosome maturation in human macrophages via specific inhibition of sphingosine kinase. *J. Immunol.* **170**:2811–2815.
- Marrie, T. F., and R. S. Martin. 1988. Persistent *Legionella pneumophila* bacteraemia in an immunocompromised host. *J. Infect.* **16**:203–204.
- McGraw, E. A., and S. L. O'Neill. 2004. *Wolbachia pipientis*: intracellular infection and pathogenesis in *Drosophila*. *Curr. Opin. Microbiol.* **7**:67–70.
- Miyara, T., K. Tokashiki, T. Shimoji, K. Tamaki, M. Koide, and A. Saito. 2002. Rapidly expanding lung abscess caused by *Legionella pneumophila* in immunocompromised patients: a report of two cases. *Intern. Med.* **41**:133–137.
- Mosavi, L. K., T. J. Cammett, D. C. Desrosiers, and Z. Y. Peng. 2004. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **13**:1435–1448.
- Nagai, H., E. D. Cambronne, J. C. Kagan, J. C. Amor, R. A. Kahn, and C. R. Roy. 2004. A C-terminal translocation signal required for Dot/Icm-dependent delivery of the *Legionella* RalF protein to host cells. *Proc. Natl. Acad. Sci. USA* **102**:826–831.
- Nagai, H., J. C. Kagan, X. Zhu, R. A. Kahn, and C. R. Roy. 2002. A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. *Science* **295**:679–682.
- Ninio, S., D. M. Zuckman-Cholon, E. D. Cambronne, and C. R. Roy. 2005. The *Legionella* IcmS-IcmW protein complex is important for Dot/Icm-mediated protein translocation. *Mol. Microbiol.* **55**:912–926.
- Ponting, C. P., L. Aravind, J. Schultz, P. Bork, and E. V. Koonin. 1999. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* **289**:729–745.
- Rogers, J., A. B. Dowsett, P. J. Dennis, J. V. Lee, and C. W. Keevil. 1994.

- Influence of temperature and plumbing material selection on biofilm formation and growth of *Legionella pneumophila* in a model potable water system containing complex microbial flora. *Appl. Environ. Microbiol.* **60**:1585–1592.
47. Rogers, J., and C. W. Keevil. 1992. Immunogold and fluorescein immunolabelling of *Legionella pneumophila* within an aquatic biofilm visualized by using episcopic differential interference contrast microscopy. *Appl. Environ. Microbiol.* **58**:2326–2330.
 48. Rowbotham, T. J. 1980. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J. Clin. Pathol.* **33**:1179–1183.
 49. Sadosky, A. B., L. A. Wiater, and H. A. Shuman. 1993. Identification of *Legionella pneumophila* genes required for growth within and killing of human macrophages. *Infect. Immun.* **61**:5361–5373.
 50. Salcedo, S. P., and D. W. Holden. 2003. SseG, a virulence protein that targets Salmonella to the Golgi network. *EMBO J.* **22**:5003–5014.
 51. Schlieper, D., M. A. Oliva, J. M. Andreu, and J. Lowe. 2005. Structure of bacterial tubulin BtubA/B: evidence for horizontal gene transfer. *Proc. Natl. Acad. Sci. USA* **102**:9170–9175.
 52. Segal, G., M. Purcell, and H. A. Shuman. 1998. Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome. *Proc. Natl. Acad. Sci. USA* **95**:1669–1674.
 53. Segal, G., J. J. Russo, and H. A. Shuman. 1999. Relationships between a new type IV secretion system and the icm/dot virulence system of *Legionella pneumophila*. *Mol. Microbiol.* **34**:799–809.
 54. Seki, T., N. Hayashi, and T. Nishimoto. 1996. RCC1 in the Ran pathway. *J. Biochem. (Tokyo)* **120**:207–214.
 55. Seshadri, R., I. T. Paulsen, J. A. Eisen, T. D. Read, K. E. Nelson, W. C. Nelson, N. L. Ward, H. Tettelin, T. M. Davidsen, M. J. Beanan, R. T. Deboy, S. C. Daugherty, L. M. Brinkac, R. Madupu, R. J. Dodson, H. M. Khouri, K. H. Lee, H. A. Carty, D. Scanlan, R. A. Heinzen, H. A. Thompson, J. E. Samuel, C. M. Fraser, and J. F. Heidelberg. 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc. Natl. Acad. Sci. USA* **100**:5455–5460.
 56. Sexton, J. A., and J. P. Vogel. 2004. Regulation of hypercompetence in *Legionella pneumophila*. *J. Bacteriol.* **186**:3814–3825.
 57. Shohdy, N., J. A. Efe, S. D. Emr, and H. A. Shuman. 2005. Pathogen effector protein screening in yeast identifies *Legionella* factors that interfere with membrane trafficking. *Proc. Natl. Acad. Sci. USA* **102**:4866–4871.
 58. Stebbins, C. E., and J. E. Galan. 2001. Structural mimicry in bacterial virulence. *Nature* **412**:701–705.
 59. Vogel, J. P., H. L. Andrews, S. K. Wong, and R. R. Isberg. 1998. Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* **279**:873–876.
 60. Walburger, A., A. Koul, G. Ferrari, L. Nguyen, C. Prescianotto-Baschong, K. Huygen, B. Klebl, C. Thompson, G. Bacher, and J. Pieters. 2004. Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science* **304**:1800–1804.
 61. Wiater, L. A., A. B. Sadosky, and H. A. Shuman. 1994. Mutagenesis of *Legionella pneumophila* using Tn903 dllaCZ: identification of a growth-phase-regulated pigmentation gene. *Mol. Microbiol.* **11**:641–653.
 62. Wu, M., L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J. C. Brownlie, E. A. McGraw, W. Martin, C. Esser, N. Ahmadinejad, C. Wiegand, R. Madupu, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, J. F. Kolonay, W. C. Nelson, Y. Mohamoud, P. Lee, K. Berry, M. B. Young, T. Utterback, J. Weidman, W. C. Nierman, I. T. Paulsen, K. E. Nelson, H. Tettelin, S. L. O'Neill, and J. A. Eisen. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**:E69.
 63. Zusman, T., O. Gal-Mor, and G. Segal. 2002. Characterization of a *Legionella pneumophila* *relA* insertion mutant and roles of RelA and RpoS in virulence gene expression. *J. Bacteriol.* **184**:67–75.