Pathogenomic Inference of Virulence-Associated Genes in Leptospira interrogans

Jason S. Lehmann1, Derrick E. Fouts2, Daniel H. Haft2, Anthony P. Cannella1, Jessica N. Ricaldi1, Lauren Brinkac2, Derek Harkins2, Scott Durkin2, Ravi Sanka2, Granger Sutton2, Angelo Moreno1, Michael A. Matthias1

1 Division of Infectious Diseases, Department of Medicine, University of California San Diego School of Medicine, La Jolla, California, United States of America, 2 J. Craig Venter Institute, Rockville, Maryland, United States of America

Introduction

Leptospirosis, caused by spirochete bacteria of the genus Leptospira, is a zoonotic disease of high public health impact [1]. Globally, nearly 900,000 people are infected annually through contact with contaminated water, infected tissue or urine of mammalian reservoir hosts [2]. Phylogenetic analyses have resolved the genus into 3 distinct lineages, which are the focus of a pan-Leptospira genome project supported by the NIAID Genome Sequencing Center: nine pathogenic species; five intermediate species (e.g. L. fainei, L. licerasiae); and six non-infectious saprophytic species (i.e. L. biflexa) (Fig. 1A) [3–6]. The 15 member paralogous gene family of unknown function were identified. This gene and these two Bartonella species came to share this expanded gene family remains an evolutionary mystery. In vivo expression analyses demonstrated up-regulation of 10/11 Leptospira genes identified in the attenuation screen, and profound in vivo, tissue-specific up-regulation by members of the paralogous gene family, suggesting a direct role in virulence and host-pathogen interactions. The pathogenomic experimental design here is generalizable as a functional systems biology approach to studying bacterial pathogenesis and virulence and should encourage similar experimental studies of other pathogens.

Despite its severity and global importance, the molecular pathogenesis of leptospirosis remains poorly understood [8]. Leptospira penetrate mucosal epithelium and damaged integument then hematogenously disseminate to localize within multiple organs, including the liver and kidney, within 72 hours. Leptospiremia may continue for up to two weeks after onset of symptoms with blood bacterial concentrations reaching as high as 105–107 organisms/mL in infected patients [9,10]. The only virulence factor genetically defined to date is the surface lipoprotein Loa22 [11], but mechanisms by which it contributes to disease pathogenesis remain unknown. Other virulence-associated genes include heme oxygenase [12], LPS [13], clpB [14], and flagellar components [15,16]. Although random transposon mutagenesis has been used to identify a few putative leptospiral virulence-related genes [17] [18], further progress has been hindered by the lack of efficient gene-targeted mutagenesis techniques in pathogenic Leptospira [8].

We used a functional systems biology (pathogenomic) approach to identify candidate virulence genes, by genomic comparison of a
**Author Summary**

Leptospirosis is one of the most common diseases transmitted by animals worldwide. It is important because it causes an often lethal febrile illnesses in tropical and subtropical areas associated with poor sanitation and agriculture. Leptospirosis may be epidemic, associated with natural disasters and flooding, or endemic in tropical regions. It is unknown how *Leptospira* cause disease and why different strains cause different severity of illness. In this study we attenuated (weakened) a highly virulent strain of *L. interrogans* by culturing it *in vitro* over several months. Comparison of the whole genome sequence before and after the attenuation process revealed a small set of genes that were mutated, and therefore associated with virulence. We discovered a putative soluble adenylate cyclase with host cell cAMP elevating activity, with implications for immune evasion and a new gene family that is upregulated *in vivo* during acute hamster infection. Interestingly, both *Bartonella bacilliformis* and *Bartonella australis* also have this unique gene family we describe in our pathogenic *Leptospira*. This information aids in our understanding of *Leptospira* evolution and pathogenesis.

culture-attenuated *Leptospira interrogans* serovar Lai strain 56601 (LD<sub>50</sub> = 10<sup>6</sup>) (Fig. 1B) with its virulent, isogenic parent (LD<sub>50</sub> < 100) [19]. *In vivo* relevance of identified candidate genes was determined by quantification of expression of candidate genes on day 4 after hamster infection in blood, liver, and kidney compared to *in vitro* culture.

**Materials and Methods**

**Ethics statement**

This study was carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health in AAALAC-approved facilities. The experimental animal work was approved by the Institutional Animal Care and Use Committee of the University of California San Diego under protocol S03128H.

**Bacterial strain maintenance and attenuation of *L. interrogans* serovar Lai strain 56601**

All strains were maintained *in vitro* in Ellinghausen-McCullough-Johnson -Harris (EMJH) media using standard protocols and are available from BEI Resources. *L. interrogans* serovar Lai strain 56601 was obtained from Dr. David Haake (UCLA). Virulence was selected for by serial passage through hamsters so that the P1 strain used in the present study had an LD<sub>50</sub> of ~10 organisms. *L. interrogans* serovar Lai strain 56601_P1 was attenuated by 18 bi-weekly subcultures *in vitro*. Virulence was assessed every five to ten subcultures using three-week-old male Golden Syrian Hamsters. Following a final subculture, genomic DNA was prepared from this attenuated strain (designated P19) on which next generation sequencing was carried out.

**Genome assembly of virulent P1 and attenuated P19 *L. interrogans* serovar Lai strain 56601 and non-synonymous SNV (nsSNV) detection**

We generated 4,379,515 and 5,340,095 unpaired shotgun reads from *L. interrogans* serovar Lai 56601_P1 and 56601_P19, respectively using next generation sequencing technology. All reads were 36 bases long. Both genomes were assembled using the comparative assembler AMOScmp. The AMOScmp-shortReads-alignmentTrimmed pipeline that runs within AMOScmp, was used to look for exact matches of each read to the published *L. interrogans* serovar Lai 56601 genome of at least 20 bp, permitting a maximum consensus error rate of 0.06% (i.e. at most two mismatches in any read). This script runs a reference-based trimming of the 3'-end of the reads prior to assembly. We found that trimming of at most 4 bases from the 3'-end of the reads based on their matches to the reference produced better assemblies than un-trimmed reads. The P1 assembly used 3,919,609 reads, leaving 439,906 unassembled singletons, while the P19 assembly used 4,915,295 leaving 424,800 singleton reads. The 56601_P1 genome was assembled into 167 contigs with an average length of 28,124 kb and an N50 length of 105,604 kb and the P19 genome into 97 contigs, average length 48,417 and N50 of 190,406. We checked the quality of both assemblies using the amosvalidate pipeline, which runs within AMOScmp. This pipeline identifies misassembly features such as increased read depth and correlated SNVs (i.e. one or more reads with the same SNV, which is unlikely to be due to sequencing error), both indicative of collapsed repeats. We found that both assemblies were high quality with at most 5 potential misassembly features in longer contigs. These potential misassemblies were inspected manually using the Hwkeye viewer and reassembled if necessary using minimus, which employs a stricter assembly algorithm. The unfinished 56601_P1a and 56601_P19 genomes were aligned and SNVs identified using the MUMmer v3.22 software package.

**RT-qPCR in vivo gene expression analysis**

Three wk old Golden Syrian Hamsters were infected via intraperitoneal injection with 10<sup>7</sup> low passage *L. interrogans* serovar Lai strain 56601. 96 hours post infection total RNA was collected using TRIzol (Invitrogen) from blood, liver, and kidney tissue, as well as from a 96-hour EMJH culture of *L. lai* strain 56601. 96 hours post infection total RNA was collected using TRIzol (Invitrogen) from blood, liver, and kidney tissue, as well as from a 96-hour EMJH culture of *L. interrogans* grown at 30°C. Total RNA was reverse transcribed using a Quantitect reverse transcription kit (Qiagen). cDNA was amplified using a CFX96 thermal cycler (Bio-Rad) using PerfeCTa SYBR Green FastMix (Quanta Biosciences). PCR was carried out at 95°C for 3 min, a touchdown gradient of 14 cycles of (94°C 10 s, 80°C 45 s) decreasing 1°C/cycle, followed by 40 cycles of (94°C 30 s, 65°C 45 s). Ct values were normalized to the leptospiral 16S rRNA gene and expression fold change calculated using the Pfaffl method [20]. Primer sequences are listed in Table S3 in Text S1.

**Domain architecture analysis of LA_4008 and other related AGC proteins**

Domain architecture comparison of LA_4008 with orthologs of *Mycobacterium xanthus*, *Corallioroscor coralloides*, *Stigmatella aurantiaca*, and *Myobacterium tuberculosis* using NCBI CD Search, SMART, and TPRPred. Protein homology analysis was carried out using BLAST using the following reference sequences: LA_4008 (NP_714188.1), MXAN_4543 (YP_632713.1), COCOR_04748 (YP_005370712.1), STAU_4866 (YP_003954471.1), Rv0386 (NP_714188.1). The coverage for the query sequence, statistical significance (E-value), and maximum amino acid identify ("Max Identity") are indicated at right for each predicted primary sequence. Identified domains were then graphically represented using the DOG 1.0 program (http://dog.biocuckoo.org)

**Leptospira Concentrated Culture Supernatant (CCS)**

*L. interrogans* Lai 56601 or *L. licerasiae* Varillal were grown in EMJH media +10% heat inactivated rabbit serum at 37°C on a rotating shaker for 96 hr. Culture was centrifuged for 30 min at
Figure 1. Pathogenomic analysis of *Leptospira interrogans* serovar Lai strain 556021 to identify virulence related genes. (A) Schematic of phylogenetic relatedness of “Pathogenic” (P), “Intermediate” (I) and “Saprophytic” (S) members of the genus *Leptospira*. (B) Workflow to identify putative virulence-associated genes. Asterisk denotes a hypothetical position in which a SNV has been identified. (C) Genomic Locations of SNPs and PF07598 paralogs in the reference genome of *L. interrogans* serovar Lai strain 56601. Each concentric circle represents genomic data and is numbered from the outermost to the innermost circle. The outermost circles represent the predicted CDS on the + and − strands, respectively, colored by functional role categories (see key). The following circle descriptions apply to chromosome I. The third circle notes the location of predicted prophage regions (olive) and the LPS region (slate). The fourth circle indicates those CDS found to have non-synonymous amino acid substitutions (black) as well as the location of CDS annotated as “transposase” in Genbank (salmon). The fifth circle represents the location of the 12 PF07598 family members (blue). The innermost circle denotes atypical regions (x² value). For chromosome II, the outermost and innermost circles are the same as for chromosome I; however, the third circle notes the location of transposases (salmon), while the fourth circle indicates the location of the CDS found to have non-synonymous amino acid substitutions (black).

doi:10.1371/journal.pntd.0002468.g001
10,000 × g. Supernatant was decanted, filtered through a 0.22 μm syringe filter unit (Millipore), and concentrated 10 × in an Amicon Ultra 10K MWCO centrifugal filter unit (Millipore). No leptospires were observed in the CCS after concentration using darkfield microscopy.

**CCS cAMP elevating activity**

CCS was incubated with monolayers of THP-1, a human monocye/macrophage cell line. At 4, 6, and 20 hours cells were rinsed 3 × in PBS and analyzed for cAMP (Direct cAMP EIA kit, Enzo Life Sciences). Secondly, CCS from *L. interrogans* Lai and *L. licerasiae* were incubated with THP-1 monolayers for 4 hours, and assayed for cAMP.

**CCS immunodepletion studies**

Rabbits were used to generate anti-peptide antiseran against LA_4008 using a protein specific, sixteen amino acid fragment (SVEEPDTREIDPRKQK) conjugated to keyhole limpet hemocyanin as a carrier protein (Pacific Immunology, Ramona, CA). The IgG fractions from pre-immunization and production bleed were purified using a Melon Gel IgG Purification kit (Thermo Scientific) and coavalyently linked to magnetic beads using a NanoLink BeadLink Kit (Sohulink). Antibody linked beads were incubated with CCS overnight at 4°C on a rotating shaker. Beads were separated on a QuadroMACS separation unit (Miltenyi Biotec). Depleted CCS was applied to THP-1 monolayers and incubated for 4 hours. Cells were rinsed 3 × in PBS and analyzed for total cAMP using the Direct cAMP EIA kit.

**Phylogenetic analysis of PF07598 paralogous protein family**

*L. interrogans*, *L. burgdorferii* and *B. bacilliformis* full-length sequences were downloaded from the Uniprot database (http://www.uniprot.org) and aligned using MAFFT v7 (http://mafft.cbrc.jp/alignment/software) with default parameters. The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan and Goldman frequency model [21]. Statistical support of the tree topology was obtained from 500 bootstrap replicates. A discrete Gamma distribution was used to model evolutionary rate differences among sites. The rate variation model allowed for some sites to be evolutionarily invariable. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 271 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [22].

**Pan-genomic analysis of attenuated genes and PF07598 orthologs**

The genomic data analyzed here are publically available and are from newly generated, unpublished *Leptospira* whole genome sequence data produced by the JCVI as part of the white paper “Leptospira Genomics and Human Health,” sponsored by the NIAID-funded Genome Sequencing Centers. PanOCT [23] was run using default settings with the exception that a dynamically determined pairwise cutoff was implemented, not available in the current release, but available upon request. The following genomes, representing all 20 *Leptospira* spp. were used: *L. alexandreni* sv. Manha 3 str. L 601 (Genbank:AHMT00000000), *L. altomi* sv. Pingchang str. 80-412 (Genbank:AOHD00000000), *L. biflexa* sv. Patoc str. Patoc I Paris (Genbank:CPO000786), *L. burgdorferii* sv. Javanica str. UI 09931 (Genbank:AHNP00000000), *L. braunii* sv. undetermined str. 33991 (Genbank:AHMO00000000), *L. fainei* sv. Hurbridge str. BUT 6 (Genbank:AKWZ00000000), *L. inalai* sv. Lynne str. 101 (Genbank:AHMM00000000), *L. interrogans* sv. Copenhageni str. Fiocruz L1-130 (Genbank:AE016923), *L. interrogans* sv. Copenhageni str. M20 (Genbank:AOGV00000000), *L. interrogans* sv. Lai str. 56601 (Genbank:AE013090), *L. kirschneri* sv. Cynopteri str. 3522 C (Genbank:AHMN00000000), *L. lowei* sv. undetermined str. Bejo-1soy (Genbank:AHMP00000000), *L. licerasiae* sv. Varillal str. VAR 010 (Genbank:AHOO00000000), *L. meyeri* sv. Hardjo str. West 3 (Genbank:AKXE00000000), *L. noguchii* sv. Panamá str. CZ 214 (Genbank:AKWV00000000), *L. santarosai* sv. Shermani str. 1342K (AOHB00000000), *L. tepliae* sv. Hvaul str. LT 11-33 (Genbank:AOGW00000000), *L. sandhii* sv. Holland str. WaZ, Holland (Genbank:AOGY00000000), *L. weeti* sv. Ranarum str. ICF (Genbank:AOHC00000000), *L. wolbachii* sv. Codice str. CDC (Genbank:AOGZ00000000), *L. wolffii* sv. undetermined str. Khorat-H1 (Genbank:AKWX00000000), *L. yanagatae* sv. Saopaulo str. Sao Paulo (Genbank:AOGX00000000).

**Statistics**

Data were analyzed using GraphPad Prism 5.0. Significance was assessed using one-way ANOVA followed by Tukey’s HSD post hoc testing. P-values are reported as *** = p<0.001, ** = p<0.01, * = p<0.05.

**Results**

**Pathogenomic identification of protein coding genes in *Leptospira interrogans* serovar Lai and patterns of tissue-specific up-regulation in vivo**

Comparison of the wild type and attenuated *L. interrogans* Lai 55601 genomes identified 41 non-synonymous single nucleotide variants (nsSNVs) in a total of 35 protein-coding genes (CDS; Table S1 in in Text S1). P19 sequence analysis revealed that all SNVs were homogeneous within the culture population; minority populations were not detected at the limit of detection of the Illumina sequencing platform (<4%). For the purposes of this study, therefore, the bacterial populations were considered clonal.

Filtering to include CDS restricted to pathogenic *Leptospira* species identified 11 genes (Fig. 2K). These CDS are highly conserved among pathogenic *Leptospira* species (Fig. 2K). In vivo transcriptional analysis identified that of these 11 pathogen-specific genes, 10 were up-regulated in vivo during acute hamster infection (Fig. 2, normalized to the 16S rDNA gene, Fig. S1 in Text S1). Transcriptional up-regulation of CDS was as high as several thousand-fold, with a much higher dynamic range than found with in vitro conditions used in previously reported systems biology analyses (summarized in Table S2 in Text S1).

**Identification of a putative leptospiral protein with host cAMP elevating activity**

Of particular interest is LA_4008, a putative adenylate/ guanylate cyclase (AGC) that lacks transmembrane helices typical of integral membrane cyclases involved in signal transduction, suggesting that this protein may be soluble. While another adenylate/guanylate cyclase was found in our screen in pathogens and intermediates (Table S1 in Text S1; LA_0027), this protein is predicted to be a housekeeping gene, a membrane-bound and intracellular, and not likely to be found in the extracellular milieu. Orthologs of LA_4008 are found only in pathogenic *Leptospira* and the immediately pathogenic strain *L. fainei*, Fig. 2K. Other bacterial adenylate cyclases lacking transmembrane domains include the soluble cyclase class of toxins of the pathogens
**Figure 2. In vivo transcriptional analysis of putative virulence-associated genes.** In vivo relevance of the identified virulence-related genes, mRNA transcript levels of the genes identified by the pathogenomics approach was assessed by real time, reverse transcriptase quantitative PCR of blood, liver and kidney 4 d after hamster infection, compared to log phase *in vitro* cultured *Leptospira*. Leptospiral gene expression levels in infected tissue vs. EMJH were expressed logarithmically as the log2 of the fold change between the two conditions (A–J). 16S rRNA transcript levels (previously validated [61]) were used to normalize gene expression in tissues and under the different conditions (Fig. S1 in Text S1). Expression of 10/11 identified genes was detectable *in vivo* in all three tissues assayed; the exception was the hypothetical protein LA_0979. The remaining 10 genes were detected in all three tissues assayed. Expression varied between groups of animals, and interestingly, the highest levels of up-regulation were found in leptospires isolated from the blood of infected animals, with transcript levels also being up in bacteria from the liver. Virulence-associated genes were variably up-regulated in kidney. The data represented are the mean ± SEM of 3 independent experiments (n = 7 animals). (K) *Leptospira* species distribution of the 11 virulence-associated genes.
Identification of a paralogous gene family shared by pathogenic *Leptospira*, *Bartonella bacilliformis*, and *Bartonella australis* with profound, tissue-specific up-regulation in vivo in an acute leptospirosis infection model in hamsters.

During our analysis of attenuation mutations we identified two members (LA\_3490, LA\_3388) of a newly discovered paralogous gene family that is shared between pathogenic *Leptospira* but conspicuously absent in the intermediate and saprophytic species. All full-length members of this family (PF07598/UF1561) are predicted to have secretory signal peptides, although degenerate forms do occur. Past the signal peptide, Cys residues are invariant at twelve positions, and occur nowhere else, suggesting a conserved pattern of disulfide bond formation and implying extracellular function (Fig. S2 in Text S1). In a given genome, the most closely related paralogs are often tandem. Otherwise, gene neighborhood analysis provided no clue to protein function. Paralog counts in pathogenic *Leptospira* range from two in the leptospire *L. santarosai* to 12 in *L. kirschneri* serovar Cynopteri and *L. interrogans* (Fig. 5A). Interestingly the PF07598 gene family has also recently described in the unrelated $\varepsilon$-proteobacteria species *Bartonella bacilliformis* and *Bartonella australis*. *B. bacilliformis* has 15 paralogs in its genome with *B. australis* having nearly the same (Fig. 5B) [30]. In addition single gene copies were found in three animal-infecting $\varepsilon$-proteobacteria, *Helicobacter hepaticus*, *H. mustelae*, and *H. celerum*.

There are great phylogenetic distances separating the genera that contain this gene family, but paralogs are restricted to select animal-infecting species within each lineage; suggesting that these proteins may be uniquely related to host adaptation. All 12 members of the leptospiral PF07598 gene family were analyzed for *in vivo* expression in hamsters acutely infected with virulent, wild type *L. interrogans* Lai 55601. All members of this gene family were up-regulated in blood and liver to varying degrees, with LA\_3490 and LA\_3388, both containing secretory signal peptide sequences, being most highly up-regulated (more than $\sim$1000-fold); all members of this gene family were up-regulated in the circulation and liver to varying degrees. In contrast, up-regulation of other members this gene family significantly varied among experimental animals in kidney (Fig. 5C–E).

Other pathogenomically-identified putative virulence genes in *Leptospira* spp.

Other pathogenomically-associated virulence genes include the following:

**LA\_1056**: This gene has two predicted transmembrane helices and shares a conserved PHA00965 domain with tail proteins found in Gram-positive bacteriophages. This protein shows similarity to phage tape measure proteins after repeated rounds PSI-BLAST. Recent studies involving the phage-encoded pblA in *Streptococcus mitis* have identified a sequence weakly reminiscent of a tape measure motif protein by PSI-BLAST as an adhesin-type molecule used for bacterial attachment to platelets [31–34].

**LA\_1765**: This protein has similarity to spvB, a protein from a group of plasmid-encoded virulence genes that mediate lethal infection in nontyphoid *Salmonella* strains [35].

**LA\_1533**: A flavin-dependent thymidylate synthase. This unusual and newly described class of enzyme is expressed by many clinically relevant pathogens, including *Bacillus anthracis*, *Borrelia burgdorferi*, *Campylobacter jejuni*, *Clostridium difficile*, *Helicobacter pylori*, *Mycobacterium tuberculosis*, and *Treponema pallidum* during infection as part of an alternative thymidine synthesis pathway [36–38].

**LA\_0202**: A gene of unknown function previously reported to be transcriptionally up-regulated in virulent *L. interrogans* Lai 55601 when compared to another avirulent strain [39].

**LA\_1568**: A putative lipoprotein with $\beta$-propeller repeats that has not been previously studied. Lipoproteins are important mediators of spirochete virulence, with the *L. interrogans* genome...
encoding over one hundred lipoproteins [40], the function and localization of many remain unclear.

**Discussion**

Here we describe the use of pathogenomics to identify novel potential virulence genes in the pathogenic spirochete *Leptospira interrogans*. Previous work to identify mechanisms of pathogenesis by gene knockouts and transposon mutagenesis has not yet yielded detailed mechanistic insights into the role of individual genes play in the pathogenesis of leptospirosis. It has long been known in the leptospirosis field that serial *in vitro* passage of pathogenic *Leptospira* yielded attenuated organisms; the converse, serial passage of liver homogenates of infected animals selects for virulence. A previous study explored the genomic and proteomic differences between a pathogenic *L. interrogans* serovar Lai strain 56601, and an avirulent strain IPAV [41]. These data must be carefully considered because the analyzed strains are not isogenic (the IPAV strain is of unknown provenance since details of its original isolation are unavailable) nor do they provide *in vivo* relevance for identified genes, focusing instead on proteomic differences between strains during *in vitro* EMJH culture. Our current study, which employed whole genome sequence comparison of an attenuated strain with its isogenic pathogenic parent, yielded a small set of protein coding genes (CDS) with point mutations. While most of the 11 specific mutations found here cannot be quantitatively attributed to specific aspects of virulence or pathogenicity, our pathogenomic approach yielded the identification of a novel leptospiral AGC with cAMP elevating activity in host cells and a hitherto unstudied large gene family that is broadly up-regulated, in a tissue-specific manner, *in vivo* during an animal model of acute leptospirosis.

The identification of a non-transmembrane bound AGC in pathogenic *Leptospira* is particularly important for two reasons. First, the primary structure implies a non-housekeeping function since the protein is not predicted to be membrane-associated, unlike the housekeeping AGC. Second, the host cell cAMP elevating activity of LA_4008 reported in this study is the first demonstrated evidence of a possible biological mechanism that could contribute to virulence for *Leptospira*. Although long established and accepted as a virulence mechanism in other pathogens, the evidence of elevation of host cAMP levels by *L. interrogans* suggests a previously unknown mechanism of pathogenesis and immune evasion for this bacterium, especially given recent evidence that pathogenic *Leptospira* may reside within macrophages *in vivo* [42,43]. Manipulation of intracellular cAMP levels in immune cells may be an important means of attenuating host responses to infection [44], an enticing hypothesis given the up-regulation of this gene upon leptospiral entry into the bloodstream observed in this study. Many human pathogens exploit host cell cAMP signaling during infection, for example, the pore-forming toxin CyaA of the respiratory pathogen *Bordetella pertussis* penetrates host cells where it catalyzes the unregulated conversion of cellular ATP to cAMP, thereby impairing superoxide production, chemotaxis, cytokine production, and phagocytosis [45–47]. Similar effects are caused by the edema factor (EF) of *Bacillus anthracis*, the ExoY toxin of *Pseudomonas aeruginosa*, and the AGC encoding over one hundred lipoproteins [40], the function and localization of many remain uncertain.

Figure 3. Ortholog sequence analysis of pathogenic *Leptospira* adenylate/guanylate cyclase compared to predatory environmental bacteria and the pathogen, *Mycobacterium tuberculosis*. Domain architecture comparison of LA_4008 with orthologs of *Myxococcus xanthus*, *Corallococcus coralloides*, *Stigmatella aurantiaca*, and *Mycobacterium tuberculosis* using NCBI CD Search, SMART, and TPRPred. Protein homology analysis was carried out using BLAST using the following reference sequences: LA_4008 (NP_714188.1), MXAN_4545 (YP_632713.1), COCOR_04748 (YP_005370712.1), STAUR_4866 (YP_003954471.1), Rv0386 (CCP43116). The coverage for the query sequence, statistical significance (E-value), and maximum amino acid identity (“Max Ident”) are indicated at right for each predicted primary sequence.

doi:10.1371/journal.pntd.0002468.g003
Figure 4. Confirmation of cAMP induction in target mammalian cells by LA_4008 activity in leptospiral culture supernatant. (A) THP-1 cell monolayers were treated with leptospire-free concentrated culture supernatant (CCS) from *L. interrogans* Lai or EMJH negative control. (B) THP-1 monolayers were treated with CCS from *L. interrogans* Lai or *L. licerasiae* Varillal, NT = not treated. (C) THP-1 cell monolayers were treated with CCS, CCS that was immunoprecipitated (IP) with specific anti-peptide antibody raised in rabbits and non-specific anti-LA 4008 antibody, and CCS that was digested with proteinase K. Values in all experiments are represented as the mean (n = 3) ± SD.

doi:10.1371/journal.pntd.0002468.g004

toxin of *Yersinia pestis* [24,48–50]. Due to an unexpected loss of the cryogenically preserved stock cultures, we were unable to assess the cAMP elevating activity of the attenuated P19 strain. However, we would hypothesize that the attenuated SNV-containing variant could have either absolute elimination or quantitative reduction in cAMP elevating activity; this possibility will be addressed directly in ongoing experiments by quantifying the effect of recombinantly producing wild type and mutant LA_4008 on THP-1 and other target cells. We also believe that any observed reduction in activity would have been a quantitative not qualitative difference. Regardless, our findings demonstrate that LA_4008 contributes to a transitory increase in cAMP levels in host cells, and that further experiments are certainly needed to assess the functional consequences of cAMP intoxication in host immune cells during leptospirosis. To formally determine the role of LA_4008 in *Leptospira* pathogenesis is the subject of ongoing experiments, including determining whether this protein modulates mechanisms of evading host defenses.

The identification of a paralogous protein family shared by pathogenic *Leptospira* spp., and two *α*-proteobacteria *B. bacilliformis* and *B. australis* was particularly intriguing. The observation that this gene family expanded in pathogenic *Leptospira* and the two *Bartonella* spp. suggests that ancestors of these pathogens must have co-existed at some time and place in the past. Phylogenetic analysis suggests a common origin of this gene family, and revealed a greater divergence in the *Bartonella* members, indicated by greater branch length differences (Fig. 5B). Regardless of the source of the primary ortholog, the founding gene was presumably transferred after the branching of pathogenic *Leptospira* from the other clades of *Leptospira*, although it is also possible that gene loss occurred in intermediates or saprophytes evolved from pathogens. Although we cannot speculate on the molecular mechanism of gene transfer, it is interesting to consider the conditions that would have been conducive to such an event. *L. interrogans* is a globally distributed bacterium that can infect many vertebrate hosts as well as live in the environment; it is considered an extracellular parasite, although evidence is mounting that *Leptospira* [42,43,51] are able to persist within macrophages and transverse epithelial cells [52]. *B. bacilliformis* and *B. australis* are facultative intracellular pathogens found only in a specific region of South America [53] and Australia respectively. The PF07598 family shared between pathogenic *Leptospira* might be shared by other *Bartonella* species that have yet to be sequenced or even identified, such as those recently found in Thailand [54,55]. The maintenance of multiple members of this paralogous gene family clearly must confer a selective advantage to these pathogens. We performed a meta-analysis of 6 previous studies [12,56–60] that explored transcriptional responses of *L. interrogans* during exposure to host-like physiological conditions (Table S2 in Text S1), and discovered that the expression levels of several of these genes occurs in response to multiple stimuli. This implies that *L. interrogans* responds to signals from the host milieu that lead to the alteration of expression of these genes in a differential manner during its infection cycle.

The identification of leptospiral AGC and PF07598 gene family orthologs in specific species of evolutionary distant alpha and delta-proteobacteria was an unexpected and exciting discovery. Given the broad host range of *Leptospira* as well as their environmental persistence, the horizontal gene transfer our findings imply emphasize how the soil context within the unique transmission cycle of *Leptospira* has likely shaped the evolution of pathogenic mechanisms for these bacteria.

Our investigation was not without limitations. The attenuation experiment was done only once. While genes of pathogenetic interest were identified here, whether these mutations occurred stochastically or not remains to be determined. Accumulation of mutations during the attenuation process was not assessed so that step-wise accumulation of mutations could not be attributed to a
### A.

| Protein   | Accession     | Function                          | P | P | P | P | P | P | P | P | P | I | I | I | S | S | S | S | S |
|-----------|---------------|-----------------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LA_3388  | NP_713568.1   | hypothetical protein              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| LA_3490  | NP_713670.1   | hypothetical protein              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| LA_0591  | NP_710772.1   | hypothetical protein              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| LA_0620  | NP_710801.1   | hypothetical protein              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| LA_0769  | NP_710950.2   | hypothetical protein              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| LA_0835  | NP_711016.1   | hypothetical protein              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| LA_0934  | NP_711115.1   | hypothetical protein              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| LA_1400  | NP_711581.2   | hypothetical protein              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| LA_2628  | NP_712809.1   | hypothetical protein              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| LA_3271  | NP_713451.1   | hypothetical protein              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

### B.

![Leptospira Pathogenomics Network](image)

**Leptospira**

**Bartonella bacilliformis**

**Helicobacter**

### C.

**Gene Upregulation in Blood**

![Graph showing gene expression levels](image)

**PF07598 Gene**

### D.

**Gene Upregulation in Liver**

![Graph showing gene expression levels](image)

**PF07598 Gene**

### E.

**Gene Upregulation in Kidney**

![Graph showing gene expression levels](image)

**PF07598 Gene**
level of virulence. Proteomic comparisons between ex-creted isolated and EMJH cultured leptospires were not performed, as our study only focused on gene transcriptional levels, which do not necessarily correlate with protein expression levels. It would be interesting to undertake such ex-creted proteomic investigations in Leptospira, especially given the vast transcriptional up-regulation of identified genes upon entry into host tissues. Further investigation remains to define the precise mechanisms of how the identified genes in our study relate to the virulence and pathogenesis of leptospirosis, as a majority of these genes have undiscovered functions.

We show here that a systems biology-pathogenomic approach to infer virulence-related genes in Leptospira interrogans identified a notable set of hitherto unstudied genes with both pathogenic and evolutionary significance, including a putative soluble adenylate/guanulate cyclase (AGC), and a paralogous gene family shared by pathogenic Leptospira and the distantly related pathogens B. bacilliformis, a human-specific pathogen geographically restricted to the Andes mountains of South America, and B. australis, a species currently known to only infect kangaroos. This pathogenomic approach is generalizable beyond prokaryotes and particularly relevant to novel virulence gene identification in any pathogen capable of in vitro attenuation. Given the recalcitrant nature of pathogenic leptospires to genetic manipulation, this approach represents an improved method to identify important virulence genes in pathogens whose pathogenesis remains poorly defined by current research strategies, and highlights the extraordinary insights into bacterial pathogenesis and evolutionary biology that large scale genomic sequencing can produce in the context of simple experimentation. These genes will hopefully spur much needed research into the pathogenesis of this neglected disease, but many may also represent rational choices for new vaccine studies.

References
1. Bharti AR, Nally JE, Ricaldi JN, Matthias MA, Diaz MM, et al. (2003) Leptospiral virulence. A zoonotic disease of global importance. Lancet Infect Dis 3: 757–771.
2. LERG (2011) Report of the Second Meeting of the Leptospirosis Burden Epidemiology Reference Group. Geneva, Switzerland: World Health Organization.
3. Ricaldi JN, Deuts DE, Selengut JD, Harkins DM, Patra KP, et al. (2012) Whole Genome Analysis of Leptospira licentiae Provides Insight into Leptospiral Evolution and Pathogenicity. PLoS Negl Trop Dis 6 (10): e1853. doi:10.1371/journal.pntd.0001853.
4. Matthias MA, Diaz MM, Campos KJ, Calderon M, Willig MR, et al. (2005) Diversity of Bartonella in the Peruvian Amazon inferred by baysian phylogenetic analysis of 16S ribosomal DNA sequences. Am J Trop Med Hyg 73: 964–974.
5. Brenner DJ, Kaufmann AF, Sulzer KR, Steigerwalt AG, Rogers FC, et al. (2016) Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with proposals for seven new Leptospira species. Int J Syst Bacteriol 66: 839–850.

Supporting Information
Text S1 Supporting information. Includes: Figure S1. Validation of 16S rDNA Gene to Normalize Leptospira In Vivo Gene Expression. Figure S2. Alignment of Bartonella bacilliformis and Leptospira interrogans serovar Lai anonymous paralog families. Table S1. Leptospira Species Distribution of Pathogenically-Discovered Genes. Table S2. Differential Expression of Gene Family Members During Exposure of L. interrogans to Host-like Conditions. Table S3. Primers used for In-vido RT-qPCR Analysis. Table S4. Genome locus tags and GenBank protein sequence accession numbers for Bartonella bacilliformis and Helicobacter spp. PF07598 family homologs used to construct Figure 5A. (DOCX)

Acknowledgments
We thank Dr. Douglas Berg (UC San Diego) for critical review of the manuscript and are grateful to Dr. David Haake (UCLA) for providing an isolate of Leptospira interrogans serovar Lai strain 55601. We are grateful to Professor Michael Minnick of the University of Montana for discussions about Bartonella and for his calling to our attention the recent availability of the complete genome sequence of Bartonella australis in GenBank. We thank Paula Maguina, Staff Research Associate, UC San Diego, for her important and key scientific and logistical contributions to the work reported here, and Jason Tanseco for contributions and early discussions of this project.

Author Contributions
Conceived and designed the experiments: JSL APC JNR JMV MAM. Performed the experiments: JSL APC JNR LR LB DH SD RS GS AM JMV MAM. Contributed reagents/materials/analysis tools: JSL DEF DHH APC JNR LR LB DH SD RS GS AM JMV MAM. Wrote the paper: JSL DEF DHH APC JNR LR LB DH SD RS GS AM JMV MAM.

Additional Supporting Information

6. Yasuda PH, Steigerwalt AG, Sulzer KR, Kaufmann AF, Rogers F, et al. (1987) Desoxyribonucleic acid relatedness between serogroups and serovars in the family Leptospiraceae with proposals for seven new Leptospira species. Int J Syst Bacteriol 37: 407–415.
7. Ko Al, Galvao Reis M, Ribeiro Durado CM, Johnson WD, Jr., Riley LW (1999) Urban epidemic of severe leptospirosis in Brazil. Salvador Leptospirosis Study Group. Lancet 354: 820–825.
8. Ko Al, Gourant C, Picardeau M (2009) Leptospira: the dawn of the molecular genetics era for an emerging zoonotic pathogen. Nat Rev Microbiol 7: 736–747.
9. Lozouathy K, Aivat F, Picardeau M (2009) Use of quantitative real-time PCR for studying the dissemination of Leptospira interogans in the guinea pig infection model of leptospirosis. J Med Microbiol 58: 648–655.
10. Segura E, Ganoza C, Campos K, Ricaldi JN, Torres S, et al. (2003) Clinical spectrum of pulmonary involvement in leptospirosis in an endemic region, with quantification of leptospirotic burden. Clin Infect Dis 40: 346–351.
11. Rintow P, Bouhy F, da Cruz McBride FW, Figueira CP, Hauser M, et al. (2007) The OmpA-like protein Loa22 is essential for leptospiral virulence. PLoS Pathog 3: 497.
Leptospira Pathogenomics

12. Lo M, Bulach DM, Powell DR, Haake DA, Matsunaga J, et al. (2006) Effects of temperature on gene expression patterns in Leptospira interrogans serovar Lai as assessed by whole-genome microarrays. Infect Immun 74: 3846–3859.

13. Murray GL, Nystedt B, Toft C, Zaremba-Niedzwiedzka K, Berglund EC, et al. (2013) Genome-wide transposon mutagenesis in pathogenic Leptospira species. Infect Immun 81: 810–816.

14. Ren SX, Fu G, Jiang Z, Zeng R, Xiao YG, et al. (2003) Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing. Nature 422: 888–893.

15. Piffan MV (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45.

16. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45.