Original

Positive Selection in F-Box Domain (Ipp0233) Encoded in *Legionella pneumophila* Strains

TAKEHIKO KENZAKA^{1,2*}, MADOKA YASUI², TAKASHI BABA², MASAO NASU^{1,2},
AND KATSUJI TANI^{1,2}

¹Faculty of Pharmacy, Osaka Ohtani University, 3-11-1 Nishikiori-kita, Tondabayashi, Osaka 584-8540, Japan

²Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamada-oka, Suita,
Osaka 565-0871, Japan

Received 5 July, 2017/Accepted 29 September, 2017

Recent studies have shown that the genome of *Legionella pneumophila* is characterized by many foreign genes from a variety of eukaryotes. The eukaryotic like proteins are known to play a role in its multiplication within host cells; however, their evolutionary genetics of *L. pneumophila* in environments is unknown. In this study, we examined the nonsynonymous/synonymous substitution rate of eukaryotic like domain encoding genes among *L. pneumophila* strains. *In silico* analysis revealed that the nonsynonymous/synonymous substitution rate in F-box domain gene (Ipp0233) was higher than those in other eukaryotic like domain and protein encoding genes and housekeeping genes. The F-box domain gene sequences in *L. pneumophila* strains isolated from a natural hot spring were more diversified than those of previously known strains, owing to preferential positive selection.

Key words : *Legionella pneumophila* / F-box / Hot spring / Eukaryotic like protein / Nonsynonymous substitution.

Advances in whole genome sequencing suggest that lateral gene transfer (LGT) has significantly contributed to acquiring new genetic traits, ability exploit new environments, and genetic diversity among numerous microorganisms (Bordenstein and Reznikoff, 2005; Ochman et al., 2000). Recent studies have shown that some microorganisms evolved through LGT across domains (Hotopp et al, 2007)

Housekeeping genes are thought to be relatively recalcitrant to LGT. Such bacterial essential genes appear to be more conserved than non-essential genes over long time scales (Jordan et al., 2002). In contrast to bacterial genes, eukaryote genes have a different evolutionary history. For example, *Saccharomyces cerevisiae* and *Caenorhabditis elegans* have no significant difference in the rate of evolution between essential and non-essential genes (Jordan et al., 2002). Analyses of mouse and rat genes have shown that essential and non-essential genes have comparable rates of evolution (Hurst and

Smith, 1999). Therefore, the concept that essential genes are more conserved than non-essential genes is not always true in eukaryotes.

The foreign genes in *L. pneumophila* probably derived from eukaryotes can modulate host functions (e.g., *Acanthamoeba castellanii*) and play significant roles in growth within eukaryotic host cells in conjunction with secretion systems (Jules and Buchrieser, 2007). On considering the important roles of such eukaryotic like protein or domain encoding genes in growth, they are thought to be conserved. On the other hand, there is an idea that host-parasite interactions will promote rapid evolution of certain proteins, especially surface proteins, which are involved in direct contact between parasite and host (Jiggins et al., 2002). Some eukaryotic proteins secreted from *L. pneumophila* directly interact with host proteins, and thus it is reasonable that they may be highly variable throughout host-parasite coevolution. Although there has been considerable effort in phylogenetic analyses of bacterial protein secretion systems (Ko et al., 2003), our current knowledge regarding the sequence diversity of eukaryotic like protein

*Corresponding author. Tel: +81-721-24-9741, Fax: +81-721-24-9742, E-mail : kenzat(a)osaka-ohtani.ac.jp

or domain encoding genes in *L. pneumophila* strains in natural environments and their relationship to host cell infectivity remains poor. The principal objectives of this study were to assess the evolutionary pattern of eukaryotic like genes in *L. pneumophila* strains.

Eukaryotic like domain encoding genes and eukaryotic like protein encoding genes (*L. pneumophila* genes preferentially found within eukaryotic proteins) were initially selected from the gene lists described in Cazalet et al. (2004). The selected genes, which were found in Legionellaceae but not in other bacterium, were further screened using BLink based on BLASTP from the National Center for Biotechnology Information (Altschul et al., 1990). The nucleotide sequence similarities with *L. pneumophila* strain Paris were calculated using nucleotide BLAST. The genes conserved among three whole-sequenced genomes of *L. pneumophila* strains (Paris [GenBank accession No. NC_006368]; Philadelphia-1

[NC_002942], and Lens [NC_006369]) were subjected to further analysis. Multiple genes, which encode the same eukaryotic like protein and domain, were analyzed item by item, and represented in a Table as shown below.

The number of non-synonymous differences divided by the number of non-synonymous sites (Ka), and the number of synonymous differences divided by the number of synonymous sites (Ks), were calculated using DnaSP 5.10 software (Librado and Rozas, 2009). The Ka/Ks ratio is an indicator of the selective pressure on a protein-coding gene. Homologous genes with high Ka/Ks ratios are generally considered to have evolved under positive selection, whereas low Ka/Ks ratio means non-positive selection. The Ka/Ks ratios for all strain pairs were calculated using Microsoft Excel XP software (Microsoft, Redmond, WA, USA). A Ka/Ks annotated evolutionary tree was constructed using an online server of Bergen Center for Computational Science Ka/Ks

TABLE 1. Nucleotide similarity of *L. pneumophila* protein genes preferentially found within eukaryotic proteins.

Gene in strain Paris	Sequence similarity ^a		Ka/Ks ^b
	Philadelphia	Lens	
F-box domain lpp0233 (567 bp)	96%	93%	0.40
STPK domain lpp2626 (1404 bp)	95%	94%	0.29
ankyrin repeat lpp0126 (3222 bp)	96%	98%	0.29
ralF lpp1932 (1209 bp)	95%	96%	0.28
ankyrin repeat/F-box domain lpp2082 (465 bp)	99%	99%	0.24
ankyrin repeat protein lpp0037 (1518 bp)	94%	95%	0.22
ankyrin repeat protein lpp2065 (1527 bp)	81%	99%	0.20
ankyrin repeat protein lpp0503 (812 bp)	93%	93%	0.18
ankyrin repeat lpp2517 (2845 bp)	98%	98%	0.17
enhC paralog lpp1310 (1141 bp)	97%	97%	0.15
ankyrin repeat/SET domain lpp1683 (1686 bp)	97%	95%	0.14
STPK domain lpp0267 (2898 bp)	96%	96%	0.13
ankyrin repeat protein lpp0750 (2838 bp)	96%	96%	0.13
STPK domain lpp1439 (1590 bp)	97%	98%	0.12
ankyrin repeat protein lpp0469 (1525 bp)	97%	96%	0.11
ankyrin repeat lpp0547 (1501 bp)	98%	98%	0.083
ankyrin repeat lpp2522 (1306 bp)	93%	96%	0.076
enhC paralog lpp2692 (3603 bp)	98%	98%	0.073
enhC paralog lpp1174 (1473 bp)	99%	98%	0.071
enhC paralog lpp2174 (1122 bp)	99%	99%	0.067
ankyrin repeat protein lpp2270 (1926 bp)	98%	98%	0.052
ankyrin repeat protein lpp2248 (1404 bp)	99%	99%	ND ^c

^aGenomic BLAST ^bRatio of average of Ka and Ks in all pairs.

Ka = Nonsynonymous substitution per nonsynonymous site.

Ks = Synonymous substitution per synonymous site.

^cKs was 0, and thus Ka couldn't be divided.

calculation tool (<http://services.cbu.uib.no/tools/kaks/>).

Their nucleotide sequence similarities against the Paris strain are shown in Table 1. Overall, nucleotide sequence similarities ranged from 93% to 99%, except for *lpp2065*, which exhibited 81% similarity to Paris and Philadelphia-1. The average ($\pm$ standard deviation) Ka/Ks ratio in the eukaryotic domain encoding genes was 0.16 ± 0.092 (ranging from 0 to 0.40). *Lpp0233* (F-box domain) had the highest Ka/Ks value (0.40). The eukaryotic like protein encoding genes among the three *L. pneumophila* strains were also retrieved. The average ($\pm$ standard deviation) Ka/Ks ratio in the eukaryotic like protein encoding genes was 0.10 ± 0.059 (ranging from 0.026 to 0.27). The all Ka/Ks ratios were lower than it in *lpp0233*.

To explore the significance of Ka/Ks ratio in eukaryotic like domain encoding genes, we calculated the variance of Ka/Ks ratios in prokaryotic or universal genes. We selected a set of 31 protein-encoding phylogenetic marker genes as described by Wu and Eisen (Table 2) (Wu, and Eisen, 2008). The genes are universally distributed in bacteria, existing as single copy genes within each genome. They are housekeeping genes that are involved in genetic information processing (replication, transcription, and translation) or central metabolism and are thought to be relatively recalcitrant to lateral gene transfer. The average ($\pm$ standard deviation) Ka/Ks ratio in the housekeeping genes was 0.034 ± 0.060 (ranging from 0 to 0.29). The Ka/Ks ratios in the housekeeping genes were significantly lower than those in the eukaryotic like domain encoding genes ($P < 0.01$).

The Ka/Ks ratios of some conserved genes of *Legionella* spp. which were often used in multi-locus sequence typing (MLST) (Gaia et al., 2003), were calculated in Table 2. The average ($\pm$ standard deviation) Ka/Ks ratio in the MLST genes was 0.061 ± 0.043 (ranging from 0.025 to 0.13), and were significantly lower than those in the eukaryotic like domain encoding genes ($P < 0.05$). Our *in silico* analyses of completely sequenced genomes of *L. pneumophila* strains revealed that the F-box domain gene was one of the most diversified eukaryotic like genes within these genomes. On considering evolution under positive selection, we investigated the sequence diversity of the F-box domain gene (*lpp0233*) in *L. pneumophila* isolates.

Natural hot-spring water samples were taken from sites K-5 and K-8 of a natural hot spring in the Kirishima area of southern Japan. Hot-spring water is naturally produced at site K-5 (hot-spring source, $31^{\circ}49'43.0''\text{N}$, $130^{\circ}44'53.1''\text{E}$), and hot-spring water flows into a stream at site K-8 ($31^{\circ}49'55.0''\text{N}$, $130^{\circ}44'32.6''\text{E}$). No specific permission was required for sampling these sites, and no endangered or protected species were involved in this study. These sites are surrounded by

TABLE 2. Ka/Ks ratios in MLST marker and house keeping genes among three whole-sequenced genomes of *L. pneumophila* strains Paris, Philadelphia, and Lens.

Category	Gene	Ka/Ks ^a
House keeping	<i>dnaG</i>	0.15 (0.022/0.15)
	<i>frr</i>	0.018 (0.002/0.082)
	<i>infC</i>	0.29 (0.002/0.008)
	<i>nusA</i>	0.037 (0.002/0.047)
	<i>pgK</i>	0.062 (0.010/0.16)
	<i>pyrG</i>	0.023 (0.002/0.092)
	<i>rplA</i>	0 (0/0.088)
	<i>rplB</i>	0.001 (0/0.057)
	<i>rplC</i>	0.027 (0.002/0.061)
	<i>rplD</i>	0.051 (0.003/0.050)
	<i>rplE</i>	0 (0/0.044)
	<i>rplF</i>	0 (0/0.042)
	<i>rplK</i>	0 (0/0.079)
	<i>rplL</i>	0 (0/0.052)
	<i>rplM</i>	ND ^b (0.002/0)
	<i>rplN</i>	0 (0/0.030)
	<i>rplP</i>	0 (0/0.029)
	<i>rplS</i>	0 (0/0.055)
	<i>rplT</i>	0 (0/0.007)
	<i>rpmA</i>	ND (0/0)
	<i>rpoB</i>	0.019 (0.002/0.088)
	<i>rpsB</i>	0.040 (0.002/0.056)
	<i>rpsC</i>	0.00 (0.003/0.066)
	<i>rpsE</i>	0 (0/0.049)
	<i>rpsI</i>	0 (0/0.063)
	<i>rpsJ</i>	0 (0/0.036)
	<i>rpsK</i>	0 (0/0.034)
	<i>rpsM</i>	0 (0/0.030)
	<i>rpsS</i>	ND (0/0)
	<i>smpB</i>	0.034 (0.002/0.052)
	<i>tsf</i>	0.043 (0.005/0.12)
MLST marker	<i>asd</i>	0.028 (0.002/0.062)
	<i>dotA</i>	0.12 (0.039/0.32)
	<i>flaA</i>	0.040 (0.007/0.18)
	<i>mip</i>	0.027 (0.001/0.044)
	<i>mompS</i>	0.13 (0.014/0.099)
	<i>neuA</i>	0.042 (0.002/0.055)
	<i>pilE</i>	0.066 (0.004/0.059)
	<i>proA</i>	0.025 (0.002/0.079)

^aRatio of average of Ka and Ks in all pairs.

Ka = Nonsynonymous substitution per nonsynonymous site.

Ks = Synonymous substitution per synonymous site.

^bKs was 0, and thus Ka couldn't be divided.

FIG. 1. Phylogenetic analysis of F-box sequences of *L. pneumophila* strains isolated from a natural hot spring and three whole-sequenced genomes of *L. pneumophila* strains (Paris [lpp0233], Philadelphia [lpg0171], and Lens [lpl0234]).

(A) Phylogenetic tree of F-box amino acid sequences of *L. pneumophila* isolates.

(B) Ka/Ks annotated evolutionary tree of F-box sequences of *L. pneumophila* isolates.

Value of outside node (shown in bold) indicates Ka/Ks of the branch.

Value of inside node indicates Ka (upper) and Ks (lower) of the branch.

Black line and gray line indicate positive selection and negative (or neutral) selection, respectively.

forests. Water samples of >500 ml were collected in sterile bottles in April 2006 and September 2007 and brought to our laboratory on ice.

Water aliquots of 300 ml were filtered through polycarbonate white filters (pore size = 0.2 μ m, diameter = 47 mm; Advantec Toyo Kaisha, Ltd.). The filters were placed in a 15-ml sterile plastic tube, and cells on the filter were suspended in 3-ml sterile deionized water by vortexing. The suspensions were heated at 50°C for 30 min to suppress the growth of non-*Legionella* bacteria. A portion of the concentrated sample (0.1 ml) was spread directly onto the surface of WYO α agar medium, after which plates were incubated at 37°C for 7 days.

Presumptive *Legionella* sp. colonies were counted and subcultured on BCYE and BCYE without cysteine at 37°C for 2 days. All isolates that grew only on BCYE were considered putative *Legionella* strains, and maintained at -80°C in 5% (w/v) yeast extract with 15% (v/v) glycerol. The *Legionella* isolates were serogrouped

using commercial kits containing antisera against *L. pneumophila* Serotype group (SG) 1, 2, 3, 4, 5, 6, and 7-15; *L. bozemanii* 1 and 2; *L. dumoffii*, *L. gormanii*, and *L. micdadei* (*Legionella* antisera "Seiken," Denka Seiken Co. Ltd., Tokyo, Japan). SG 1 was not dominant in our samples. SG7-15 was found in all our samples.

Genomic DNA was extracted from *L. pneumophila* cultures. *L. pneumophila* isolates that grew on WYO α agar were resuspended in the TE buffer and lysed at 37°C for 1 h with 1% SDS and proteinase K (50 μ g per ml). The suspensions were incubated at 90°C for 10 min, frozen in liquid nitrogen, and then successively thawed three times at room temperature. After phenol and phenol-chloroform-isoamyl alcohol extractions, DNA was precipitated in ethanol, dried, and resuspended in 10 mM Tris-HCl buffer (pH 8.5). By targeting *mip* using real-time PCR (Wellinghausen et al., 2001), all isolates were confirmed to be *L. pneumophila*.

The F-box primer set targeted for the conserved

region in the F-box gene (lpg0171-f, 5'-ATGAAAG CAAAATACGAC-3'; lpg0171-r, 5'-CTACAATGGCTCAC AT-3') was designed in this study. PCR was conducted with a standard reaction mixture of 25 μ l. A PCR mixture containing 1 $\times$ PCR buffer II, 2.5 mM MgCl₂, 0.2 mM of each deoxynucleoside triphosphate, 0.4 μ M of an F-box primer set, and 2.5 U of AmpliTaq Gold was made using DNA-free water. This mixture was subjected to 30 amplification cycles at 94°C for 1 min, 55°C for 0.5 min, and 72°C for 2.5 min.

The PCR products were purified with a MinElute PCR Purification Kit (QIAGEN, Tokyo, Japan). Sequence analysis of the PCR products was conducted using a CEQ8000 genetic analysis system (Beckman Coulter, Fullerton, CA, USA) with lpg0171-f and lpg0171-r primers. Dye terminator cycle-sequencing reactions were performed according to the manufacturer's protocol. Clearly sequenced regions of 474 bp in all strains were subjected to further genetic analysis. Similarity searches for these sequences were performed using BLAST. The GenBank accession numbers of the sequences determined in this study are JX897006 to JX897026.

A phylogenetic tree for the F-box amino acid sequences of the total 21 *L. pneumophila* strains was calculated by the neighbor-joining method with Clustal X (Larkin et al., 2007) and is shown in Fig.1A. Some K-5 strains from the same sampling date formed a cluster I with Paris (lpp0233). K-8 strains and some K-5 strains from a different sampling date formed another cluster II, which was not similar to Paris, Philadelphia-1 (lpg0171), or Lens (lpl0234) strains. Strains K-5-g, K-8-1, and lpg0171 did not belong to either cluster I or II.

The amino acid sequences showed more variability than the nucleotide sequences with similarities ranging from 85% to 100% for all *L. pneumophila* strains. Some strains showed higher sequence diversity than that among completely sequenced genomes of *L. pneumophila* strains, especially K-5-5, K-5-6, and K-8-1 showed low similarity (85%) to Lens strain (lpl0234).

The Ka/Ks ratios of the F-box sequences among all *L. pneumophila* strain pairs were determined (Table 3), which were as high as 2.62. Some Ka/Ks ratios were more than 1 (shown in bold type in the Table). An excess of non-synonymous substitutions (Ka/Ks > 1) is an indicator of positive selection, whereas an excess of synonymous substitutions (Ka/Ks < 1) indicates purifying selection imposed by functional constraints. The Ka/Ks values of more than 1 were not only found in the same samples but also among sampling dates and sites. These values were higher than those among the completely sequenced genomes of the three *L. pneumophila* strains.

The Ka/Ks annotated evolutionary tree is shown in

Fig.1B. A positive selection along branches leading to most strain sequences (shown in black line) was detected. In addition, a positive selection along the branch leading to the K-8-g, K-5-c, K-5-b/K-8-e ancestor, K-5-g/lpg0171 ancestor, K-5-4/K-5-10 ancestor, and K-5-6/K-5-7 ancestor was detected. These results showed that their sequences were more diversified than the sequences previously determined for three known strains.

Phylogenetic analyses showed that the epidemic or sporadic nature of *L. pneumophila* strains could not be distinguished using environmental isolates based on F-box domain gene polymorphisms (Fig.1). That is, there was no relationship between F-box domain gene sequences and the geographical characteristics of Japanese, US (Philadelphia-1), and European (the other) strains (Fig.1). The results on no link between clinical and environmental isolates were consistent with other studies (Aurell et al., 2005). Previous studies suggests that the genomic background with some sets of virulence- and eukaryotic-like genes plays a role in disease causation (Cazalet et al., 2008), and specific *L. pneumophila* strain is a highly stable, worldwide distributed epidemic clone. As well, our results shows that *L. pneumophila* strains with identical F-box domain gene sequences are widely distributed around the world.

The association of *L. pneumophila* with protozoa is a major factor that regulates the occurrence of this bacterium in the environment. Because *H. vermiformis* and *Acanthamoeba* spp. were detected in all samples used in this study (data not shown), eukaryotic proteins, which were secreted from *L. pneumophila* isolates, might closely interact with host proteins. Proteins that have F-box domains in *L. pneumophila* are thought to modulate the eukaryotic ubiquitination machinery. Ubiquitination is a protein modification process that is generally used by cells to tag proteins destined for proteasomal degradation. For these proteins, the F-box domain attaches to ubiquitin and promotes proteasomal degradation (Lomma et al., 2010). The F-box domain gene is involved in the interaction with the host and the host-parasite interactions would promote diversification of even eukaryotic-derived proteins in "arms race" coevolutionary conflicts (Blanc et al., 2005; Kakizawa et al., 2006). The clear positive selections imply an important biological role for F-box domain in host-bacterium interactions, and indicate that the eukaryotic like protein in *L. pneumophila* is variable and thus account for strain differences.

Bacterial essential genes appear to be more conserved than non-essential genes over long time scales (Jorden et al., 2002). In contrast to bacterial genes, eukaryotes have no significant difference in the rate of evolution between essential and non-essential genes (Hurst and

TABLE 3. Ka/Ks^a matrix of F-box sequence among *L. pneumophila* isolates.

	5-1	5-2	5-4	5-5	5-6	5-7	5-8	5-9	5-10	5-b	5-c	5-e	5-f	5-g	8-1	8-4	8-5	8-7	8-c	8-e	8-g	Lpp	Lpl	Lpg
5-1																								
5-2	0.19																							
5-4	ND ^b	0.19																						
5-5	0.58	0.57	0.58																					
5-6	ND	0.38	ND	0.29																				
5-7	ND	0.29	ND	0.43	ND																			
5-8	ND	0.38	ND	0.58	ND	ND																		
5-9	ND	0.19	ND	0.58	ND	ND	ND																	
5-10	0.86	0.21	0.86	0.29	0.29	0.00	1.44	0.86																
5-b	0.69	0.53	0.69	0.66	0.86	0.82	0.77	0.69	0.71															
5-c	0.82	0.55	0.82	0.62	0.73	0.68	0.91	0.82	0.59	2.62														
5-e	0.67	0.51	0.67	0.65	0.86	0.82	0.76	0.67	0.70	0.87	ND													
5-f	0.62	0.49	0.62	0.61	0.78	0.74	0.62	0.62	0.65	0.48	1.15	0.28												
5-g	0.71	0.48	0.71	0.66	1.01	0.93	0.85	0.71	0.74	0.66	0.97	0.62	0.52											
8-1	0.51	0.34	0.51	0.39	0.42	0.38	0.60	0.51	0.45	0.59	0.51	0.60	0.57	0.59										
8-4	0.58	0.46	0.58	0.58	0.74	0.70	0.66	0.58	0.62	0.67	1.44	0.57	0.43	0.60	0.54									
8-5	0.54	0.45	0.54	0.55	0.67	0.64	0.54	0.54	0.58	0.48	0.77	0.47	0.57	0.54	0.51	0.43								
8-7	0.62	0.49	0.62	0.61	0.78	0.74	0.62	0.62	0.65	0.86	1.74	0.86	0.28	0.69	0.57	ND	0.29							
8-c	0.54	0.43	0.54	0.55	0.70	0.66	0.62	0.54	0.58	0.48	1.16	0.28	0.28	0.52	0.51	ND	0.57	ND						
8-e	0.67	0.51	0.67	0.65	0.86	0.82	0.76	0.67	0.70	0.87	ND	ND	0.28	0.62	0.60	0.57	0.47	0.86	0.28					
8-g	0.81	0.61	0.81	0.62	0.82	0.77	0.81	0.81	0.66	2.62	ND	ND	0.57	0.97	0.57	1.44	0.57	1.15	1.15	ND				
Lpp	ND	0.19	ND	0.58	ND	ND	ND	ND	0.86	0.69	0.82	0.67	0.62	0.71	0.51	0.58	0.54	0.62	0.54	0.67	0.81			
Lpl	0.42	0.38	0.42	0.44	0.51	0.49	0.46	0.42	0.53	0.43	0.50	0.40	0.39	0.44	0.49	0.36	0.45	0.39	0.33	0.40	0.50	0.42		
Lpg	0.18	0.18	0.18	0.23	0.28	0.25	0.23	0.18	0.28	0.25	0.30	0.23	0.23	0.16	0.39	0.21	0.23	0.23	0.19	0.23	0.30	0.18	0.36	

^aKa = Nonsynonymous substitution per nonsynonymous site. Ks = Synonymous substitution per synonymous site.

^bKs was 0, and thus Ka couldn't be divided.

Smith, 1999). Despite of the important roles in growth, F-box domain gene was more diversified than those of previously known strains owing to preferential positive selection. Therefore, an idea that host-parasite interactions would promote rapid evolution of certain proteins which are involved in direct contact between parasite and host, would be appropriate to eukaryotic like genes in *L. pneumophila* (Jiggins et al., 2002). The positive selections imply an important biological role for F-box domain in host-bacterium interactions, and indicate that the eukaryotic like domain in *L. pneumophila* accounts for strain differences and the rapid adaptation to environmental settings. These findings will broaden our understanding the survival strategies of *L. pneumophila* in the natural environment.

ACKNOWLEDGMENTS

This work was supported by the JSPS Grant-in-Aid for

Scientific Research on Innovative Areas (24117522), Grant-in-Aid for Scientific Research (C) (15K00571), and Grant-in-Aid for Scientific Research on Innovative Areas (15H05946).

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
- Aurell, H., Farge, P., Etienne, J., Jarraud, S., and Bernard, C. (2005) Clinical and environmental isolates of *Legionella pneumophila* serogroup 1 cannot be distinguished by sequence analysis of two surface proteingenes and three housekeeping genes. *App. Environ. Microbiol.* **71**, 282-289.
- Blanc, G., Ngwamidiba, M., Ogata, H., Fournier, P. E., Claverie, J. M., and Raoult, D. (2005) Molecular evolution of rickettsia surface antigens: evidence of positive selection. *Mol. Biol. Evol.* **22**, 2073-2083.
- Bordenstein, S. R., and Reznikoff, W. S. (2005) Mobile DNA

- in obligate intracellular bacteria. *Nat. Rev. Microbiol.* **3**, 688-699.
- Cazalet, C., Rusniok, C., Brüggemann, H., Zidane, N., Magnier, A., Ma, L., Tichit, M., Jarraud, S., Bouchier, C., Vandenesch, F. et al. (2004) Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat. Genet.* **36**, 1165-1173.
- Cazalet, C., Jarraud S., Ghavi-Helm, Y., Kunst, F., Glaser, P., Etienne, J., and Buchrieser, C. (2008) Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species. *Genome Res.* **18**, 431-441.
- Gaia, V., Fry, N. K., Harrison, T. G., and Peduzzi, R. (2003) Sequence-based typing of *Legionella pneumophila* serogroup 1 offers the potential for true portability in legionellosis outbreak investigation. *J. Clin. Microbiol.* **41**, 2932-2939.
- Hotopp, J. C., Clark, M. E., Oliveira, D. C., Foster, J. M., Fischer, P., Torres, M. C., Giebel, J. D., Kumar, N., Ishmael, N., Wang S et al. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**, 1753-1756.
- Hurst, L. D., and Smith, N. G. (1999) Do essential genes evolve slowly? *Curr. Biol.* **9**, 747-750.
- Jiggins, F. M., Hurst, G. D., and Yang, Z. (2002) Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. *Mol. Biol. Evol.* **19**, 1341-1349.
- Jordan, I. K., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2002) Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* **12**, 962-968.
- Jules, M., and Buchrieser, C. (2007) *Legionella pneumophila* adaptation to intracellular life and the host response: Clues from genomics and transcriptomics. *FEBS Let.* **581**, 2829-2838.
- Kakizawa, S., Oshima, K., Jung, H. Y., Suzuki, S., Nishigawa, H., Arashida, R., Miyata, S., Ugaki, M., Kishino, H., and Namba, S. (2006) Positive selection acting on a surface membrane protein of the plant-pathogenic phytoplasmas. *J. Bacteriol.* **188**, 3424-3428.
- Ko, K. S., Lee, H. K., Park, M. Y., and Kook, Y. H. (2003) Mosaic structure of pathogenicity islands in *Legionella pneumophila*. *J. Mol. Evol.* **57**, 63-72.
- Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P. A., McWilliam H., Valentin F., Wallace I. M., Wilm A., Lopez R., Thompson J. D., Gibson T. J., and Higgins D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948.
- Librado, P., and Rozas, J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.
- Lomma, M., Dervins-Ravault, D., Rolando, M., Nora, T., Newton, H. J., Sansom, F. M., Sahr, T., Gomez-Valero, L., Jules, M., Hartland EL, et al. (2010) The *Legionella pneumophila* F-box protein Lpp2082 (AnkB) modulates ubiquitination of the host protein parvin B and promotes intracellular replication. *Cell. Microbiol.* **12**, 1272-1291.
- Ochman, H., Lawrence, J. G., and Groisman EA. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304.
- Wellington, N., Frost, C., and Marre, R. (2001). Detection of *Legionellae* in hospital water samples by quantitative real-time LightCycler PCR. *Appl. Environ. Microbiol.* **67**, 3985-3993.
- Wu, M., and Eisen, J. A. (2008) A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* **9**, R151.