

# The Genome of the Amoeba Symbiont “*Candidatus Amoebophilus asiaticus*” Reveals Common Mechanisms for Host Cell Interaction among Amoeba-Associated Bacteria<sup>∇†‡</sup>

Stephan Schmitz-Esser,<sup>1\*</sup> Patrick Tischler,<sup>2</sup> Roland Arnold,<sup>2</sup> Jacqueline Montanaro,<sup>1</sup> Michael Wagner,<sup>1</sup> Thomas Rattai,<sup>2</sup> and Matthias Horn<sup>1</sup>

Department of Microbial Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria,<sup>1</sup> and Technische Universität München, Department of Genome Oriented Bioinformatics, Wissenschaftszentrum Weihenstephan, Freising, Germany<sup>2</sup>

Received 21 October 2009/Accepted 7 December 2009

**Protozoa play host for many intracellular bacteria and are important for the adaptation of pathogenic bacteria to eukaryotic cells. We analyzed the genome sequence of “*Candidatus Amoebophilus asiaticus*,” an obligate intracellular amoeba symbiont belonging to the *Bacteroidetes*. The genome has a size of 1.89 Mbp, encodes 1,557 proteins, and shows massive proliferation of IS elements (24% of all genes), although the genome seems to be evolutionarily relatively stable. The genome does not encode pathways for *de novo* biosynthesis of cofactors, nucleotides, and almost all amino acids. “*Ca. Amoebophilus asiaticus*” encodes a variety of proteins with predicted importance for host cell interaction; in particular, an arsenal of proteins with eukaryotic domains, including ankyrin-, TPR/SEL1-, and leucine-rich repeats, which is hitherto unmatched among prokaryotes, is remarkable. Unexpectedly, 26 proteins that can interfere with the host ubiquitin system were identified in the genome. These proteins include F- and U-box domain proteins and two ubiquitin-specific proteases of the CA clan C19 family, representing the first prokaryotic members of this protein family. Consequently, interference with the host ubiquitin system is an important host cell interaction mechanism of “*Ca. Amoebophilus asiaticus*”. More generally, we show that the eukaryotic domains identified in “*Ca. Amoebophilus asiaticus*” are also significantly enriched in the genomes of other amoeba-associated bacteria (including chlamydiae, *Legionella pneumophila*, *Rickettsia bellii*, *Francisella tularensis*, and *Mycobacterium avium*). This indicates that phylogenetically and ecologically diverse bacteria which thrive inside amoebae exploit common mechanisms for interaction with their hosts, and it provides further evidence for the role of amoebae as training grounds for bacterial pathogens of humans.**

Free-living amoebae, such as *Acanthamoeba* spp., are ubiquitous protozoa which can be found in such diverse habitats as soil, marine water, and freshwater and in many engineered environments (62, 100). They are important predators of prokaryotic and eukaryotic microorganisms, thereby having great influence on microbial community composition, soil mineralization, plant growth, and nutrient cycles (14, 100). Interestingly, many well-known pathogens of humans are able to infect, survive, and multiply within amoebae (39, 51). These protozoa can thus serve as reservoirs and vectors for the transmission of pathogenic bacteria to humans, as demonstrated for *L. pneumophila* and *Mycobacterium avium* (2, 115). It is also increasingly being recognized that protozoa are important for the adaptation of (pathogenic) bacteria to higher eukaryotic cells as a niche for growth (2, 24, 42, 78, 89).

In addition to the many recognized transient associations between amoeba and pathogens, stable and obligate relation-

ships between bacteria and amoebae also were described for members of the *Alphaproteobacteria* (11, 34, 48), the *Betaproteobacteria* (49), the *Bacteroidetes* (50), and the *Chlamydiae* (4, 12, 35, 52). These obligate amoeba symbionts show a worldwide distribution, since phylogenetically highly similar strains were found in amoeba isolates from geographically distant sources (51, 107). The phylogenetic diversity and the different lifestyles of these obligate intracellular bacteria—some are located directly in the host cell cytoplasm (11, 34, 48–50, 52), while others are enclosed in host-derived vacuoles (4, 35, 44)—suggest fundamentally different mechanisms of host cell interaction. However, with the exception of chlamydia-related amoeba symbionts (37, 46, 47), our knowledge of the biology of obligate intracellular symbionts of amoebae is still scarce.

Comparative genomics has been extremely helpful for the analysis of intracellular bacteria. Numerous genome sequences from the *Alpha*- and *Gammaproteobacteria* and the *Chlamydiae* are available and have contributed significantly to our understanding of genome evolution, the biology of intracellular bacteria, and the interactions with their host cells (24, 26, 46, 79, 82). In this study, we determined and analyzed the complete genome sequence of “*Candidatus Amoebophilus asiaticus*” strain 5a2 in order to gain novel insights into its biology. “*Ca. Amoebophilus asiaticus*” is a Gram-negative, obligate intracellular amoeba symbiont belonging to the *Bacteroidetes* which

\* Corresponding author. Mailing address: Department of Microbial Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. Phone: 43 1 4277 57311. Fax: 43 1 4277 54389. E-mail: schmitz-esser@microbial-ecology.net.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

‡ The authors have paid a fee to allow immediate free access to this article.

∇ Published ahead of print on 18 December 2009.

has been discovered within an amoeba isolated from lake sediment (107). “*Ca. Amoebophilus asiaticus*” shows highest 16S rRNA similarity to “*Candidatus Cardinium hertigii*,” an obligate intracellular parasite of arthropods able to manipulate the reproduction of its hosts (131). According to 16S rRNA trees, both organisms are members of a monophyletic group within the phylogenetically diverse phylum *Bacteroidetes*, consisting only of symbionts and sequences which were directly retrieved from corals (113). Among members of the *Bacteroidetes*, the genome sequences of only three symbionts, which are only distantly related (75 to 80% 16S rRNA sequence similarity) to “*Ca. Amoebophilus asiaticus*,” have been determined to date: two strains of “*Candidatus Sulcia muelleri*,” a symbiont of sharpshooters, and “*Azobacteroides pseudotriconymphae*,” a symbiont of an anaerobic termite gut ciliate (45, 72, 74, 127).

The genome of “*Ca. Amoebophilus asiaticus*” is only moderately reduced in size compared to those of many other obligate intracellular bacteria (75, 123), but nevertheless, its biosynthetic capabilities are extremely limited. A large fraction of the genome consists of IS elements and an unparalleled high number of proteins with eukaryotic domains, such as ankyrin repeats, TPR/SEL1 repeats, leucine-rich repeats, and domains from the eukaryotic ubiquitin system, all of them most likely important for host cell interaction. Feature enrichment analysis across a nonredundant data set of all bacterial genomes showed that these domains are enriched in the genomes of bacteria (including several pathogens of humans) known to be able to infect amoebae, providing further evidence for an important role of amoebae in the evolution of mechanisms for host cell interaction in intracellular bacteria.

#### MATERIALS AND METHODS

**Cultivation of “*Ca. Amoebophilus asiaticus*”.** The amoeba containing “*Ca. Amoebophilus asiaticus*” 5a2 (ATCC number PRA-228) was isolated from lake sediment in Austria (107) and cultivated using peptone-yeast-glucose-medium (PYG) [containing 20 g/liter proteose peptone, 18 g/liter glucose, 2 g/liter yeast extract, 3.4 mM sodium citrate-dihydrate, 4 mM MgSO<sub>4</sub> · 7H<sub>2</sub>O, 2 mM Na<sub>2</sub>HPO<sub>4</sub> · 2H<sub>2</sub>O, 1.7 mM KH<sub>2</sub>PO<sub>4</sub>, 0.05 mM Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub> · 6H<sub>2</sub>O] as described previously (107). Cultures were incubated at 30°C; fresh medium was applied every 5 to 10 days.

**Purification of “*Ca. Amoebophilus asiaticus*” from its *Acanthamoeba* host and DNA isolation for genome sequencing.** Amoebae (from 3.6-liters culture) were collected by centrifugation (5,000 × g, 10 min), resuspended in buffer A (35 mM Tris-HCl, 250 mM Sucrose, 25 mM KCl, 10 mM MgCl<sub>2</sub>) containing 250 mM EDTA, and subsequently homogenized using a Dounce tissue grinder (Wheaton). The homogenized amoebae were centrifuged (300 × g, 2 min) to pellet nuclei and unlysed amoebae. The supernatant was removed and centrifuged (4,100 × g, 5 min) to pellet the bacteria, which were subsequently resuspended in buffer A (containing 250 mM EDTA) and centrifuged consecutively two times for 30 min at 40,650 × g in an SM24 rotor (Sorvall). The resulting pellet was resuspended in buffer A (without EDTA) and incubated with 15 U DNase I (Sigma) for 1 h on ice in order to degrade remaining amoebal extracellular DNA. The reaction was stopped by addition of 1/10 volume of 0.5 M EDTA, and the bacteria were washed once with buffer A (containing 250 mM EDTA) and centrifuged (4,100 × g, 5 min). The pellet was resuspended in 250 μl 1 × TE buffer and subsequently used for DNA isolation using a modified protocol from the work of Zhou et al. (133). To the resuspended bacteria, 675 μl DNA extraction buffer (100 mM Tris-HCl, 100 mM EDTA, 100 mM sodium phosphate, 1.5 M NaCl, 1% [wt/vol] cetyltrimethylammonium bromide [CTAB], 200 μg/ml proteinase K, pH 8.0) was added, followed by incubation for 30 min at 37°C. After addition of 75 μl 20% (wt/vol) SDS, the samples were incubated at 65°C for 1 h. The samples were mixed with an equal volume of chloroform-isoamylalcohol (24:1 [vol/vol]), the aqueous phase was recovered after centrifugation (11,200 × g, 10 min), and nucleic acids were precipitated with 0.6 volume

isopropanol at room temperature for 1 h. After centrifugation (16,000 × g, 20 min), the pellet was washed with 70% ethanol, centrifuged again (16,000 × g, 20 min), and resuspended in 1 × Tris-EDTA (TE) buffer containing 100 μg/ml RNase, incubated for 20 min at 37°C, and stored at –20°C until use. The amount of remaining nuclear host DNA was estimated using a semiquantitative PCR assay targeting a 294-bp fragment of the *Acanthamoeba* host 18S rRNA gene. For this, the isolated DNA was amplified with the forward primer Ac18SF1 (5′-GGGGAACTTACCAGGTCCG-3′) and reverse primer Ac18SR1 (5′-CCCTCTAAGAAGCACGGACG-3′) using a two-step PCR protocol with 35 cycles of a 15-s denaturation step at 95°C and a 50-s annealing/elongation step at 70°C. For comparison, PCRs with plasmids (of known concentration) containing the *Acanthamoeba* sp. 5a2 18S rRNA gene were used. PCR products were subjected to agarose gel electrophoresis followed by SyBr green staining; the amount of host DNA was estimated by densitometric quantification of gel band intensity using ImageJ (<http://rsb.info.nih.gov/ij/>). The amount of nuclear host DNA in the DNA preparation used for genome sequencing was estimated to be lower than 0.02%.

**Genome sequencing.** Sequencing (using a combination of 8-kb and fosmid (40-kb) libraries and 454 pyrosequencing (using a Roche 454 GS20 instrument with a coverage of at least 10×) and assembly were performed by the U.S. Department of Energy Joint Genome Institute (JGI) under the Community Sequencing Program. According to the newly defined standards for classification of genome sequences (19), the “*Ca. Amoebophilus asiaticus*” genome sequence belongs to the category “finished”.

**Genome annotation and analyses.** The PEDANT software system (<http://pedant.gsf.de>) was used for automatic genome sequence analysis and annotation (33). Protein coding genes were predicted using the GeneMarkS software program using default settings (9). Biochemical pathway prediction and reconstruction were performed using the KEGG (60), BRENDA (7), MetaCyc, and BioCyc (16) databases. Proteins containing eukaryotic domains (ankyrin repeats, leucine-rich repeats, TPR/SEL1 repeats, F-Box domains, and U-Box domains) were retrieved from the SIMAP database (96) based on the corresponding domain models from PFAM (32) and SMART (69) as provided by the Interpro database (55). In order to verify the bacterial origins of the proteins with eukaryotic domains and eukaryotic proteins in the “*Ca. Amoebophilus asiaticus*” genome, we performed TBlastN searches against the (yet-incomplete) *Acanthamoeba castellanii* strain Neff genome ([http://www.hgsc.bcm.tmc.edu/projects/microbial/microbial-detail.xsp?project\\_id=163](http://www.hgsc.bcm.tmc.edu/projects/microbial/microbial-detail.xsp?project_id=163)) and *A. castellanii* strain Neff expressed sequence tags (ESTs) available through the Protist EST Program at the TBestDB database (<http://tbestdb.bcm.umontreal.ca/searches/login.php>) (88). *A. castellanii* strain Neff is closely related to the *Acanthamoeba* host of “*Ca. Amoebophilus asiaticus*”. In all cases, the BLAST score (reflecting sequence similarity and alignment quality) of the first BLAST hit was higher when we searched against the GenBank nonredundant (nr) database than with either of the two *Acanthamoeba* data sets. This suggests that the “*Ca. Amoebophilus asiaticus*” eukaryotic proteins and proteins with eukaryotic domains do not represent contaminants from the *Acanthamoeba* host. The enrichment of protein domains was calculated using a self-written Python script for counting the occurrences of domains in the selected genomes and using R (95) (<http://www.R-project.org>) for calculating the statistical significance of enrichments and depletions (Fisher test). Classification of transport proteins into Transport Classification Database (TCDB) Families was done using BLAST (<http://www.tcdb.org/index.php>) (103). Mobile genetic elements were analyzed using the Web-based resource ISfinder (110); repeats were identified using REPuter (67). Pseudogenes were identified using the Ψ-Φ finder software with an E-value cutoff of 10<sup>–15</sup> and at least 30% identity in the initial BLAST step (68). We used all available completed genomes from the phylum *Bacteroidetes* (as available in December 2007) as reference genomes (*Bacteroides fragilis* YCH46, *Bacteroides thetaiotaomicron* VPI-5482, *Cytophaga hutchinsonii* ATCC 33406, “*Gramella forsetii*” KT0803, *Porphyromonas gingivalis* W83, and *Salinibacter ruber* DSM13855). Genes acquired by horizontal gene transfer were predicted by searching for those “*Ca. Amoebophilus asiaticus*” proteins that had their 10 best BLAST hits to proteins from organisms outside the *Bacteroidetes* but within the same phylogenetic lineage (at the phylum or class level). Those proteins that in addition showed bidirectional best BLAST hits using an E-value threshold of 10<sup>–5</sup> were considered to be horizontally acquired and were further subjected to phylogenetic analysis with their homologues. For this, amino acid sequences were aligned with MAFFT (61) and phylogenetic trees were reconstructed with MEGA (66) using the neighbor-joining method and the poisson correction with 1,000× bootstrapping; all positions containing gaps and missing data were eliminated from the data sets. In addition, for all candidates we performed phylogenetic analyses using maximum likelihood (RAxML) with the JTT amino acid substitution model and 1,000× bootstraps (112). Heat maps were created using JColorGrid (57).

TABLE 1. General features of “*Ca. A. asiaticus*” genome and those of other members of the *Bacteroidetes*

Characteristic	Description or value for strain				
	“ <i>Ca. A. asiaticus</i> ” 5a2	“ <i>Ca. Sulcia muelleri</i> ” GWSS	“ <i>Azobacteroides pseudotrichonymphae</i> ” CFP2	<i>Flavobacterium psychrophilum</i> JIP02/86	<i>Bacteroides thetaiotaomicron</i> VPI-5482
Lifestyle	Obligate intracellular, parasitic	Obligate intracellular, mutualistic	Obligate intracellular, mutualistic	Facultative intracellular, parasitic	Extracellular, mutualistic or parasitic
Chromosome size (bp; excluding plasmids)	1,884,364	245,530	1,114,206	2,861,988	6,260,361
G+C content (%)	35.0	22.4	33.0	32.5	42.8
No. of protein-coding genes	1557	228	758	2432	4779
Avg length of protein-coding genes (bp)	990	1006	1047	1003	1170
No. of rRNA operons	1 <sup>a</sup>	1	1 <sup>a</sup>	6	5
No. of tRNA genes	35	31	38	49	71
Coding density (%)	81.8	96.1	71.2	84.5	89.3
No. of putative pseudogenes	222	–	22	20	–
No. of mobile genetic elements	381	–	ND <sup>b</sup>	74	67
% of mobile genetic elements	24%	–	ND	3%	1%
Reference(s)	This study	74, 127	45	27	129

<sup>a</sup> 16S and 23S/5S rRNA genes are unlinked.

<sup>b</sup> ND, not determined.

**Transmission electron microscopy.** *Acanthamoeba* sp. 5a2 was cultured until a 50% monolayer covered the surface of a 25-cm<sup>2</sup> culture flask. Cells were rinsed briefly in 1× Page’s amoebic saline (20 mM NaCl, 0.16 mM MgSO<sub>4</sub> · 7H<sub>2</sub>O, 0.27 mM CaCl<sub>2</sub> · 2H<sub>2</sub>O, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 10 mM KH<sub>2</sub>PO<sub>4</sub>) without dislodging the monolayer and then fixed with 5 ml 2.5% glutaraldehyde in 3 mM cacodylate buffer for 2 h at 20°C. Fixed cells were scraped from the culture flask, pelleted at 6,000 × g for 5 min, and then resuspended in 0.1 M cacodylate buffer. The cells were encapsulated in 12% gelatin (in 0.1 M cacodylate buffer), and the cell pellet was trimmed into 1-mm cubes for further processing. Fixation with 0.5% OsO<sub>4</sub> (in 0.1 M cacodylate buffer) for 2 h was followed by brief distilled-water washes and a dehydration series through 50%, 70%, 95%, and 100% ethanol. Acetone was used as an intermediate step before low-viscosity resin (Agar Scientific) infiltration and embedding. Blocks were polymerized at 60°C overnight. Thin sections (70 nm) were contrasted with uranyl acetate and lead citrate before examination using the transmission electron microscope (Zeiss 902) at 70-kV accelerating voltage.

**Reverse transcriptase PCR.** For transcriptional analysis, amoebae harboring “*Ca. Amoebophilus asiaticus*” were harvested by centrifugation (2,350 × g, 5 min, at 4°C), resuspended in TRIzol (Invitrogen Life Technologies), and immediately homogenized using the Bead-Beater Fast Prep FP120 instrument (BIO 101). Purification of total RNA was performed according to the recommendations of the manufacturer (TRIzol), followed by a DNase treatment using DNase I (Sigma). A control PCR using primers specific for Aasi\_1805 (see below) and *Taq* DNA polymerase instead of reverse transcriptase was performed to ensure that the RNA preparation was free of DNA; 1 µg of RNA was used for cDNA synthesis using the RevertAid First Strand cDNA synthesis kit (MBI-Fermentas). cDNA was used as a template in standard PCRs with a PCR cycling program with 40 cycles and an annealing temperature of 58°C. Negative controls (no DNA added) were included in all PCRs. Primers targeting a 151-bp fragment of Aasi\_0770 (forward primer, 5′-GCAGCTAACCAGAACGAGC-3′; reverse primer, 5′-TGGTGTGGTAGCAGCAAGG-3′) or a 245-bp fragment of Aasi\_1805 (forward primer, 5′-CGCAAGAAGATGCAAGTGAA-3′; reverse primer, 5′-GCTGTCATCTCCGACCATT-3′), respectively, were used. Amplification products were sequenced to ensure that amplification was specific. All experiments were performed in two biological replicates.

**GenBank/EMBL/DBJ accession number.** The genome sequence and annotations for “*Ca. Amoebophilus asiaticus*” have been deposited in the GenBank/EMBL/DBJ databases under accession number CP001102.

## RESULTS AND DISCUSSION

**General features of the “*Ca. Amoebophilus asiaticus*” genome.** The genome of “*Ca. Amoebophilus asiaticus*” consists of a circular chromosome with a size of 1,884,364 bp and a G+C content of 35% (Table 1; see also Fig. S1 in the supplemental material). No plasmids were detected in the assembly or in pulsed-field gel electrophoresis experiments performed to

estimate the genome size prior to genome sequencing (data not shown). In total, 1,557 coding sequences (CDSs) were identified in the “*Ca. Amoebophilus asiaticus*” genome. Approximately 10% of the genes ( $n = 153$ ) did not show protein sequence similarities to other proteins in public databases (unknown proteins). The genome encodes 35 tRNAs (for all amino acids), one unlinked set of rRNA genes, and three further noncoding RNAs (RNaseP, 4.5S RNA, and tmRNA). A high proportion of all predicted genes (24%;  $n = 381$ ) encode putative mobile genetic elements. We identified 222 putative pseudogenes (14% of all CDSs) in the “*Ca. Amoebophilus asiaticus*” genome, most of them ( $n = 197$ ) corresponding to transposases. Consistent with its obligate intracellular lifestyle, the repertoire of regulatory proteins of “*Ca. Amoebophilus asiaticus*” is relatively small: “*Ca. Amoebophilus asiaticus*” encodes two sigma factors (RpoN and RpoD) and 18 putative transcriptional regulators (see the text in the supplemental material). To our surprise, we identified 11 proteins needed for gliding motility, a common feature among the *Bacteroidetes* (71) (see the text in the supplemental material). Taken together, the genome of “*Ca. Amoebophilus asiaticus*” is thus not particularly reduced and streamlined compared to the genomes of obligate mutualistic insect symbionts, which have genome sizes from 0.14 to 0.8 Mbp (73, 79).

**Limited metabolic capabilities of “*Ca. Amoebophilus asiaticus*”.** The percentage of “*Ca. Amoebophilus asiaticus*” proteins affiliating to clusters of orthologous groups (COG) of metabolic categories (C, E, F, G, H, I, M, P, and Q) is remarkably low (17.4%; see Fig. S2 in the supplemental material), suggesting that despite its relatively large genome compared to those of other obligate intracellular bacteria, “*Ca. Amoebophilus asiaticus*” has only very reduced metabolic capabilities. This low percentage in “*Ca. Amoebophilus asiaticus*” is significantly lower than the average percentage of proteins affiliating to metabolism COG categories in the genomes of 40 obligate intracellular bacteria (35%) determined in a recent study (75). While most enzymes of the glycolysis (Embden-Meyerhof-Parnass) pathway are encoded in the genome, phosphofructokinase (PfkA) as the key glycolytic enzyme is absent. In contrast,



fructose biphosphatase (Fbp), the key enzyme of gluconeogenesis, is present, suggesting that the enzymes of the Embden-Meyerhof-Parnass pathway are used for gluconeogenesis rather than for glycolysis. This is further supported by the absence of known specific sugar uptake systems in the “*Ca. Amoebophilus asiaticus*” genome. Furthermore, the tricarboxylic acid cycle is completely missing, as is the oxidative branch of the pentose phosphate pathway. However, interconversions of sugars can be achieved by the nonoxidative branch of the pentose phosphate pathway. Based on the presence of a variety of different proteases and peptidases ( $n = 22$ ) for protein degradation and the presence of amino acid and oligopeptide transporters, we suggest that amino acids or oligopeptides are the main nutrient source of “*Ca. Amoebophilus asiaticus*.”

The genome encodes the pyruvate dehydrogenase complex and several other dehydrogenases. “*Ca. Amoebophilus asiaticus*” can thus oxidize pyruvate to  $\text{CO}_2$  and acetyl coenzyme A, generating NADH. A proton-translocating NADH dehydrogenase I (nuo) complex and an  $\text{F}_0\text{-F}_1$  type ATP synthase are present; however, a complete electron transport chain (cytochrome oxidase and quinone biosynthesis) is missing, indicating that “*Ca. Amoebophilus asiaticus*” does not gain energy from oxidative phosphorylation and that, rather, the ATP synthase is responsible for generating a proton gradient across the inner membrane.

In general, the biosynthetic capabilities of “*Ca. Amoebophilus asiaticus*” are extremely reduced. For example, “*Ca. Amoebophilus asiaticus*” has no pathways for cofactor biosynthesis and misses the genetic repertoire for the generation of purine and pyrimidine nucleotides, a feature shared with rickettsiae and chlamydiae (36, 46). Furthermore, except for *meso*-diaminopimelate, all *de novo* amino acid biosynthesis pathways are lacking; in addition, “*Ca. Amoebophilus asiaticus*” cannot decarboxylate *meso*-diaminopimelate to lysine. In contrast, pathways for lipopolysaccharide and cell wall biosynthesis are present, although reduced in comparison to those of free-living bacteria.

In summary, the absence of glycolysis, the tricarboxylic acid cycle, and the lack of a functional respiratory chain suggest that “*Ca. Amoebophilus asiaticus*” is able to generate energy only via substrate-level phosphorylation from degradation of oligopeptides, amino acids, and other substrates taken up from its amoeba host cell. The very limited anabolic capabilities of “*Ca. Amoebophilus asiaticus*” seem to rule out that any essential substrates are provided to the amoeba host, which is consistent with the observation that the amoeba host can be grown without “*Ca. Amoebophilus asiaticus*” as a symbiont (data not shown). On the other hand, the strict metabolic dependency of “*Ca. Amoebophilus asiaticus*” from its host is characteristic of many other intracellular bacteria (123) and indicates a high level of adaptation to an obligate intracellular lifestyle.

**Proteins for exploitation of host cell metabolism.** In order to compensate for its reduced biosynthetic capabilities, “*Ca. Amoebophilus asiaticus*” has to take up essential nutrients from its host cell. This might be achieved by 82 proteins predicted to be involved in transport of different metabolites (see Table S1 in the supplemental material). This set of transporters offers several redundant possibilities for the uptake of oligopeptides and amino acids, because two copies of Opp

ABCDF-like transporters, two copies of DtpT-like di/tripeptide transporters, and a putative oligopeptide/ $\text{H}^+$  symporter (GltP) are encoded in the genome. In addition, dicarboxylates, such as malate, fumarate, or succinate, can be imported from the host by a DcuAB-like  $\text{C}_4$ -dicarboxylate transporter, partially compensating for the lack of a tricarboxylic acid cycle. The “*Ca. Amoebophilus asiaticus*” genome encodes a  $\text{Na}^+/\text{H}^+$  (CvrA) antiporter and a  $\text{K}^+/\text{H}^+$  (TrkAH) antiporter most likely modulating  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{H}^+$  gradients across the cytoplasmic membrane. The  $\text{Na}^+$ -motive force can then be used to drive other cellular transport activities. Like *Rickettsia felis* and *Orientia tsutsugamushi* (20, 90), the “*Ca. Amoebophilus asiaticus*” genome encodes a large number ( $n = 19$ ) of putative sodium/solute symporters with unknown substrate specificities, which might facilitate the acquisition of nutrients such as proline, galactose, glucose, pantothenate, or nucleosides (59). There are only two specific uptake systems for cofactors from the host: “*Ca. Amoebophilus asiaticus*” lacks an *S*-adenosylmethionine synthetase (MetK) but encodes a putative *S*-adenosylmethionine transporter (Aasi\_1859) (119), and biotin can be taken up by the transporter BioY (Aasi\_1496) (43). In addition, “*Ca. Amoebophilus asiaticus*” exploits its host cell by parasitizing the host ATP pool using an ATP/ADP translocase (Aasi\_0069), a well-known antiporter from chlamydiae and rickettsiae (106, 116, 117, 125). Employing this transporter, “*Ca. Amoebophilus asiaticus*” might be able to supplement energy gained by substrate-level phosphorylation. In summary, many of the truncated metabolic pathways found in “*Ca. Amoebophilus asiaticus*” can be explained by the presence of specific transport proteins that import essential host-derived metabolites.

**Modulation of host cell functions.** After entry into host cells, intracellular bacteria either reside within phagosome-derived vacuoles (e.g., *Chlamydiae*, *Legionella*, and *Ehrlichia*) or they escape the vacuole and subsequently live directly within the cytosol of their host cells (e.g., rickettsiae [41]). Consistent with the latter lifestyle, “*Ca. Amoebophilus asiaticus*” resides within the host cytoplasm (Fig. 1) and thus has direct access to nutrients in this intracellular niche. However, the mechanisms by which “*Ca. Amoebophilus asiaticus*” enters, survives, and replicates within its amoeba host were yet unknown. Analysis of the “*Ca. Amoebophilus asiaticus*” genome now allows the first insights into how these bacteria interact with their host cells.

In order to reach its intracellular niche, “*Ca. Amoebophilus asiaticus*” needs to escape from the host phagosome. For this, “*Ca. Amoebophilus asiaticus*” might employ a mechanism similar to that of rickettsiae, since its genome encodes two putative phospholipases D (Aasi\_0130 and Aasi\_0227) with a high similarity (approximately 40% amino acid sequence identity) to the functionally characterized phospholipase D from *Rickettsia* spp., which is considered the major effector for vacuole exit (56, 98).

Bacteria living in eukaryotes generally use specialized secretion systems to deliver bacterial proteins into the host cytosol (15, 23, 118) in order to modulate host cell functions and to shape the intracellular environment according to their needs. “*Ca. Amoebophilus asiaticus*” encodes a putative type VI secretion system which is distantly related to described type VI secretion systems (Aasi\_1072 to Aasi\_1806). This type of protein secretion system has only recently been discovered and is

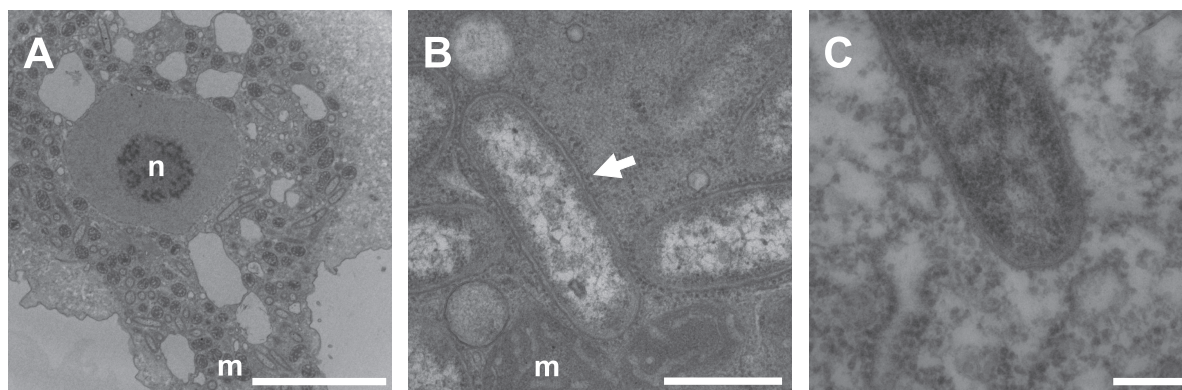


FIG. 1. Ultrastructure of “*Ca. Amoebophilus asiaticus*” 5a2 within its *Acanthamoeba* host cell. (A) Transmission electron micrograph showing the distribution of “*Ca. Amoebophilus asiaticus*” in its *Acanthamoeba* host cell. (B and C) An association of “*Ca. Amoebophilus asiaticus*” with host cell ribosomes (a selected ribosome is marked by an arrow) is clearly visible. No indications for the presence of a membrane surrounding “*Ca. Amoebophilus asiaticus*” are evident. Host cell nucleus (“n”) and selected mitochondria (“m”) are labeled. Bar represents 5  $\mu$ m (A), 500 nm (B), or 200 nm (C).

used primarily by proteobacterial pathogens and symbionts, which generally also encode additional protein secretion systems (94). A Sec-dependent pathway for export of proteins into the periplasm is present in “*Ca. Amoebophilus asiaticus*”, but components of the type II secretion system enabling further transport from the periplasm across the outer membrane are missing. Remarkably, there is also no evidence for the presence of any of the other well-known protein secretion systems (118), yet a large number of “*Ca. Amoebophilus asiaticus*” proteins contain eukaryotic domains and are thus most likely targeted to the host cytosol, where they might act as effector proteins as demonstrated for other intracellular bacteria.

**Unique abundance of proteins with eukaryotic domains.** “*Ca. Amoebophilus asiaticus*” encodes 129 proteins (8% of all CDSs) with domains predominantly found in eukaryotic proteins, including ankyrin repeats, TPR/SEL1 repeats, leucine-rich repeats, or F- and U-box domains (see Tables S2 to S6 in the supplemental material). These domains can mediate protein-protein interactions and are particularly important for the interaction of various intracellular bacterial pathogens with their eukaryotic host cells (3, 17, 18, 40, 65, 84, 91, 120). Compared to all other bacteria for which genome sequences are available, the number and percentage of proteins with eukaryotic domains in the “*Ca. Amoebophilus asiaticus*” genome is unmatched (Fig. 2; see also Table S7 in the supplemental material).

We found 20 proteins containing a leucine-rich repeat (LRR) motif in the “*Ca. Amoebophilus asiaticus*” genome (see Table S2 in the supplemental material). LRR proteins have a characteristic repetitive sequence pattern rich in leucines (8). They are widespread and have been suggested or demonstrated to be involved in protein-protein interactions (8). Many pathogenic bacteria use LRR proteins for the interaction with their host cells (e.g., see reference 10). Interestingly, a high number of LRR proteins have also been identified previously in the phylogenetically distant amoeba symbiont *Protochlamydia amoebophila* (29, 47), although their function is still unknown.

We identified 59 proteins containing tetratricopeptide re-

peats (TPR) or SEL1 repeats (a TPR subfamily; see Table S3 in the supplemental material) in “*Ca. Amoebophilus asiaticus*”. Particularly, SEL1 repeat proteins mediate interaction between bacteria and eukaryotic cells (77). Several “*Ca. Amoebophilus asiaticus*” SEL1 repeat proteins show high similarity (up to 52% amino acid identity) to *L. pneumophila* LidL, which has been shown to be important in early signaling events regulating intracellular trafficking of *L. pneumophila* in macrophages (21, 22, 84).

The “*Ca. Amoebophilus asiaticus*” genome also encodes 54 proteins containing the ankyrin repeat (ANK) motif (see Table S4 in the supplemental material). The number of ANK repeats per protein varies and is unusually high (up to 53 ANK repeats per protein) in some “*Ca. Amoebophilus asiaticus*” proteins (13, 70, 81) (see Tables S4 and S8 in the supplemental material). ANK repeats are among the most common protein-protein interaction motifs found mainly in eukaryotes and are involved in a variety of cellular processes, such as cytoskeleton formation and cell cycle and transcriptional regulation (70, 81). Some other bacteria also encode ANK proteins but generally at much lower numbers (below 10 ANK proteins per genome) (30, 90, 128). Interestingly, those bacterial genomes with the highest numbers of ANK proteins are from intracellular bacteria (Fig. 2; see also Table S7 in the supplemental material), suggesting that this group of proteins plays a role in the interaction with eukaryotic cells. Recently two ANK proteins from *L. pneumophila* were demonstrated to be important for intracellular replication in amoebae (3, 40). In addition, translocation into the host cell and an essential role for infection of eukaryotic cells (interference with vesicular transport) have recently been shown for ANK proteins from *L. pneumophila* and *Coxiella burnetii* (91, 120).

**Interference with the host ubiquitination system.** A striking feature of the “*Ca. Amoebophilus asiaticus*” genome is the presence of a to date-unmatched number of proteins most likely interfering with the host ubiquitination system (5, 101). In total, 15 proteins with F-box domains and 9 proteins with a U-box domain have been identified (see Tables S5 and S6 in the supplemental material). Both domains are components of the ubiquitination system common to all eukaryotes. Ubiquitin-

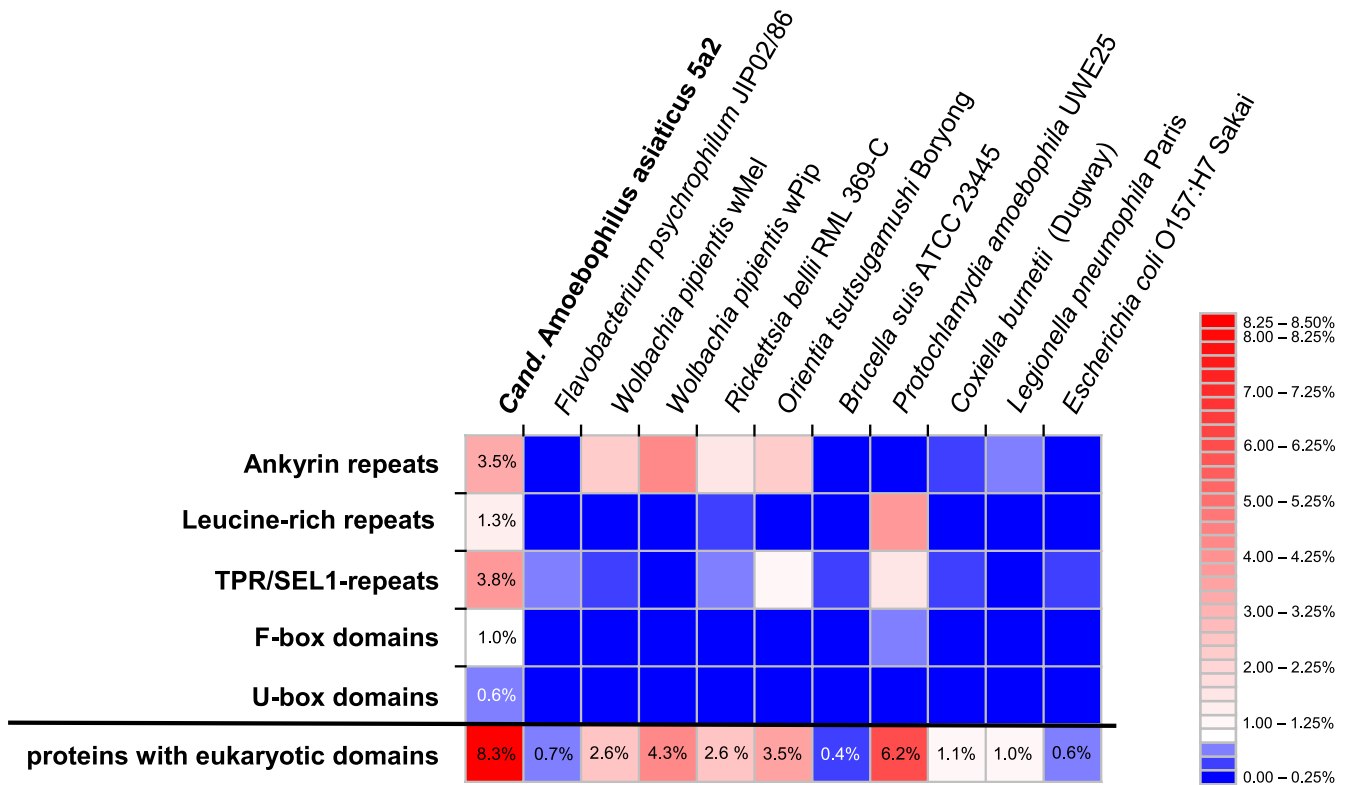


FIG. 2. Abundance of proteins with eukaryotic domains in proteomes of “*Ca. Amoebophilus asiaticus*” and other bacteria. Proteins with eukaryotic domains are most abundant in intracellular bacteria. The percentage of proteins with eukaryotic domains in complete proteomes is shown for selected intracellular and extracellular bacteria, which demonstrates an unprecedented abundance of these proteins in “*Ca. Amoebophilus asiaticus*”. Note that several proteins contain more than one type of eukaryotic domain. For further details, see Table S7 in the supplemental material.

ation (i.e., the covalent attachment of ubiquitin monomers to target proteins) controls the degradation of proteins by the proteasome and many other important cellular functions, such as cell cycle progression, transcriptional regulation, signal transduction, or resistance to pathogens (122).

Interference with ubiquitination using F-box proteins has been described for some plant pathogens, and a recent screen for F-box motifs in public databases revealed in total 31 F-box proteins in all bacterial genomes (5). The highest number of F-box proteins was found in the amoeba symbiont *P. amoebophila* ( $n = 11$ ), which also encodes one of the two known bacterial U-box proteins. The other bacterial U-box protein, lpp2887 from *L. pneumophila*, was recently shown to be secreted into the host cell and characterized as ubiquitin ligase (65). Taken together, the number and percentage of F- and U-box proteins encoded in the “*Ca. Amoebophilus asiaticus*” genome is by far the highest found among prokaryotic genomes (Fig. 2; see Table S7 in the supplemental material).

In addition to these proteins, “*Ca. Amoebophilus asiaticus*” also encodes two proteins with highest amino acid sequence identity (26 and 30%, respectively) to the catalytic domain of eukaryotic cysteine proteases of the ubiquitin-specific protease family (USP), which is the largest and most common group of deubiquitinating enzymes in eukaryotes (85). An alignment of Aasi\_0770 and Aasi\_1805 with known USP proteins revealed

clear conservation of important amino acid residues forming the catalytic triad, as well as other residues which have been shown to be essential for USP function (53, 97) (see Fig. S3 in the supplemental material). Both proteins are significantly shorter (339 and 343 amino acids [aa], respectively) than their eukaryotic homologues, but USP proteases in general show little sequence conservation except for the catalytic core region and are highly variable in size and domain structure (85, 97).

To our knowledge, only five bacterial proteases with deubiquitinating activity have been described so far (76, 102, 130, 132), and all of them belong to the CE clan of cysteine proteases, which are largely unrelated to the USPs of “*Ca. Amoebophilus asiaticus*”. Instead, the two putative “*Ca. Amoebophilus asiaticus*” ubiquitin proteases belong to the CA clan, family C19 (101), and represent the first prokaryotic members of this clan. In order to check whether the two putative USP genes are transcribed during intracellular multiplication of “*Ca. Amoebophilus asiaticus*” in amoebae, we performed reverse transcriptase PCR. We were able to demonstrate the transcription of both USP genes, suggesting that these eukaryotic genes are indeed used by “*Ca. Amoebophilus asiaticus*” (see Fig. S4 in the supplemental material). “*Ca. Amoebophilus asiaticus*” is thus apparently able to interfere with the host ubiquitination system in two ways, by transferring ubiquitin to target proteins (using the F- and U-box



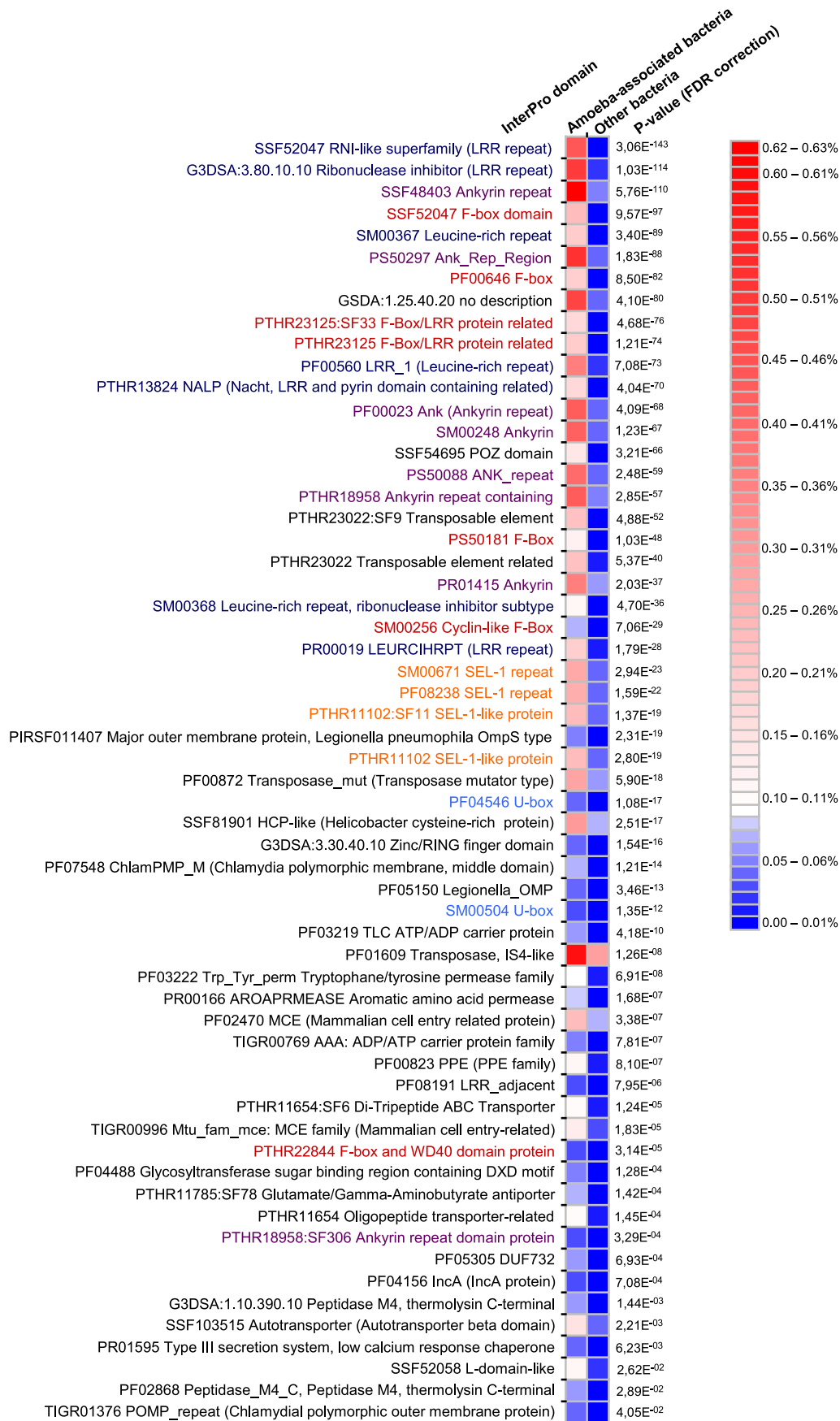
proteins) and by removing ubiquitin from target proteins (using the USPs), a unique combination of features among prokaryotes (101). Interference with the host ubiquitin system is thus presumably vital for the intracellular lifestyle of “*Ca. Amoebophilus asiaticus*.”

**Enrichment of proteins with eukaryotic domains among amoeba-associated bacteria.** The remarkably high number of proteins with eukaryotic domains in “*Ca. Amoebophilus asiaticus*” prompted us to further investigate the distribution of these protein domains in other bacteria. To achieve an unbiased overview, we performed an enrichment analysis comparing the fraction of all functional protein domains (from the InterPro database) among 514 bacterial proteomes. When we compared the proteomes of bacteria for which the replication in amoeba has been demonstrated (see Table S9 in the supplemental material) with those of 480 other bacteria (representing the complete domain *Bacteria* but including only a single, representative strain if multiple genomes per species were available), we found 59 functional domains to be enriched in amoeba-associated bacteria with high statistical support ( $P$  values  $< 5e^{-2}$ , calculated with the false discovery rate [FDR] correction) (Fig. 3; see also Tables S10 to S12 in the supplemental material). Among these enriched domains, 29 are predominantly found in eukaryotic proteins. Interestingly, the domains potentially involved in host cell interaction described above, such as ANK repeats, LRR, SEL1 repeats, and F- and U-box domains, are among the most highly enriched domains in proteomes of amoeba-associated bacteria. This enrichment was still statistically significant when the enrichment analysis was performed without “*Ca. Amoebophilus asiaticus*” (because of its high number of eukaryotic-like proteins) and even when we included all proteomes from bacteria for which replication in amoebae has been shown only for closely related strains (see Table S12 in the supplemental material). Bacteria that can exploit amoebae as hosts thus share a set of eukaryotic domains important for host cell interaction despite their different lifestyles (within vacuoles versus directly in the cytoplasm; obligate versus facultative intracellular) and their large phylogenetic diversity, spanning five bacterial phyla. This suggests that bacteria thriving within amoeba use similar mechanisms for host cell interaction to facilitate survival in the host cell. Due to the phylogenetic diversity of these bacteria, it is most parsimonious that these traits were acquired independently during evolutionary early interaction with ancient protozoa. The observation that 50% of the amoeba-associated bacteria included in our analysis are also well-known pathogens of humans suggests that the development of these mechanisms not only was fundamental for the survival in amoebae but also primed the evolution of bacteria able to infect higher eukaryotes. Our data thus lend novel support for the suggested role of amoebae as a major training ground for the evolution of bacterial pathogens of humans (2, 24, 42, 78, 89).

**Other candidate proteins for host cell interaction.** In addition to a large number of proteins with eukaryotic domains, “*Ca. Amoebophilus asiaticus*” also encodes several other known bacterial virulence factors. For example, “*Ca. Amoebophilus asiaticus*” contains five putative patatin-like proteins (see Fig. S5 in the supplemental material), glycoproteins with lipolytic activity involved in different aspects of host cell interaction in *L. pneumophila* and *Pseudomonas aeruginosa*, such as

host cell lysis and membrane trafficking (1, 6, 93, 104, 108). Surprisingly, we also found two large proteins similar to insecticidal toxins (toxin complexes [tc]) from various *Photorhabdus* spp. and other members of the *Gammaproteobacteria* (Aasi\_1414 and Aasi\_1417) (see Fig. S6). These highly efficient toxins consist of up to three subunits, one of which is the toxin itself and one or two potentiators (31)—a structural organization conserved in the two “*Ca. Amoebophilus asiaticus*” homologues. Interestingly, “*Ca. Amoebophilus asiaticus*” also encodes a gene cluster for synthesis of lasso peptides (Aasi\_0899 to Aasi\_0902) (see Fig. S7 in the supplemental material), small ribosomally synthesized peptides with diverse functions, such as antimicrobial or antimitochondrial activity or receptor antagonists, known for only few bacterial species (28, 64, 86).

**Stable genome with 24% mobile genetic elements.** The abundance of unusual proteins with eukaryotic domains which are partially shared among amoeba-associated intracellular bacteria and “*Ca. Amoebophilus asiaticus*” indicates that horizontal gene transfer might have had a major impact on the evolutionary history of this bacterium. The high number of mobile genetic elements ( $n = 381$  [24% of all CDSs]) present in “*Ca. Amoebophilus asiaticus*” might represent remnants of past acquisition events of foreign genes (see Fig. S1 in the supplemental material). In total, three phage genes, one truncated reverse transcriptase, which probably represents a degraded form of a group II intron (111), and remnants of a conjugative transposon (CTn) (see the text of the supplemental material) were identified, but by far, most of the mobile genetic elements in the “*Ca. Amoebophilus asiaticus*” genome are transposases representing IS elements (consisting of a transposase gene flanked by inverted and/or direct repeats;  $n = 377$ ). Approximately 156 (41%) of these transposases seem to form intact IS elements, while the majority represent truncated and probably nonfunctional copies. This high number of IS elements is remarkable, since the percentage of IS elements in bacterial chromosomes is usually below 3% (109). A similar abundance of IS elements has been observed in only two other (obligate intracellular) bacterial species: two strains of the alphaproteobacterium *Orientia tsutsugamushi* (19% and 31%) (20, 83) and the  $\gamma$ 1 symbiont of the marine oligochaete *Olavius algarvensis* (20%) (126). An increase in the frequency of mobile genetic elements has been suggested to occur immediately after a bacterium adapts to an intracellular or pathogenic lifestyle (80, 92, 123). The spreading of IS elements has been reported to result in proliferation of pseudogenes, genome rearrangements, and ultimately genome reduction (80, 109). Such extensive genome rearrangements in recent evolutionary time are expected to lead to irregular GC skew patterns. However, unlike the *O. tsutsugamushi* genome and other IS element-rich bacterial genomes, which show a highly irregular GC skew pattern (20, 63, 83, 124, 128), the “*Ca. Amoebophilus asiaticus*” genome displays a GC skew with only two major shifts near the origin and terminus of replication, which is typical for most bacterial genomes (99) (see Fig. S1 in the supplemental material). The overall regular GC skew pattern of the “*Ca. Amoebophilus asiaticus*” genome shows only a few major local deviations (up to 52 kbp in length). Some of these regions also have a slightly different GC content and are associated with tRNA genes. They might thus represent genomic





islands (58). This indicates that despite the abundance of IS elements, which constitute the largest part of the repeats in the “*Ca. Amoebophilus asiaticus*” genome (see the text of the supplemental material), this genome has apparently not been extensively reshuffled recently but rather has remained stable for an extended evolutionary time period. Consequently, it seems unlikely that “*Ca. Amoebophilus asiaticus*” has recently adapted to an intracellular lifestyle.

**Horizontal gene transfer and the evolution of the intracellular lifestyle.** In the “*Ca. Amoebophilus asiaticus*” genome, several genes and gene clusters have clearly been acquired by lateral gene transfer. For example, the gene cluster for lasso peptide synthesis is flanked by two identical IS6-family IS elements (see Fig. S7 in the supplemental material), a genomic organization resembling composite transposons, which can mobilize genomic fragments of up to 17 kb (114, 121). Another unusual protein encoded by “*Ca. Amoebophilus asiaticus*” is the ATP/ADP translocase (106, 117, 125). Horizontal gene transfer has been suggested to be responsible for the occurrence of this protein in largely unrelated intracellular bacteria and plant plastids (38, 54, 105), but no potential mechanism for its spread has been reported. The ATP/ADP translocase gene of “*Ca. Amoebophilus asiaticus*” is flanked by two almost identical copies of a (degraded) IS5 family IS element (98% nucleic acid sequence identity) (see Fig. S8), together probably representing a remnant composite transposon, suggesting that the transporter, which showed highest sequence similarity to chlamydial ATP/ADP translocases (40% amino acid sequence identity), was acquired by transposase-mediated transfer. The ATP/ADP translocase could supply “*Ca. Amoebophilus asiaticus*” with host-derived ATP, which complements the limited energy production and missing nucleotide synthesis capabilities of “*Ca. Amoebophilus asiaticus*” and thus contributed to the adaptation to an intracellular lifestyle.

To get a systematic and more comprehensive overview of indications for horizontally acquired genes in the “*Ca. Amoebophilus asiaticus*” genome, we screened the complete “*Ca. Amoebophilus asiaticus*” proteome for proteins for which all 10 closest-homologue best BLAST hits belonged to organisms from a single phylum or class different from the *Bacteroidetes*. From this list, those proteins that in addition showed bidirectional best BLAST hits with their respective best BLAST hit were considered candidate proteins most likely affected by horizontal gene transfer. Using this conservative approach, we identified in total 54 candidate proteins (see Table S13 in the supplemental material). For 37 of those, phylogenetic analysis with extended data sets resulted in stable, well-supported phylogenetic trees, many of which can be best explained by the origin of these proteins with foreign organisms and acquisition by horizontal gene transfer. In most cases, however, the “*Ca. Amoebophilus asiaticus*” proteins are highly divergent from

their nearest neighbors, suggesting evolutionary ancient gene transfer events whose direction cannot be unambiguously determined (Fig. 4; see also Fig. S9).

Interestingly, five “*Ca. Amoebophilus asiaticus*” proteins might have been transferred from eukaryotes into the “*Ca. Amoebophilus asiaticus*” genome. These proteins show a typical eukaryotic domain structure and include the two putative ubiquitin-specific proteases (see above), a eukaryotic-like protein kinase, a putative flavin-containing monooxygenase, and a putative phosphatidylinositol synthase (Fig. 4; see also Table S13 and Fig. S9 in the supplemental material). Although it is notable that in the two cases where phylogenetic analysis was possible, the “*Ca. Amoebophilus asiaticus*” protein clusters with homologous proteins from the social amoeba *Dictyostelium discoideum*, the exact identity of the eukaryotic donor in these ancient horizontal gene transfer events remains blurred due to low bootstrap support values and a limited data set of eukaryotic homologues. Similar to that of “*Ca. Amoebophilus asiaticus*,” the genome of *Legionella pneumophila* encodes 30 proteins with highest sequence similarity to eukaryotic proteins (18). Recent evidence suggests that one of these proteins, the sphingosine-1-phosphate lyase, which is important for *Legionella* pathogenesis, has been acquired from a protozoan host (25, 87).

The majority (72%) of the “*Ca. Amoebophilus asiaticus*” proteins that were affected by horizontal gene transfer are shared among “*Ca. Amoebophilus asiaticus*” and other amoeba-associated bacteria, particularly rickettsiae. Horizontal gene transfer among amoeba-associated bacteria has been suggested previously as an important mechanism for the evolution of the intracellular lifestyle (89). Our analysis provides further evidence for this hypothesis and suggests that protozoa served as an ancient training ground for intracellular bacteria, not only by facilitating adaptation to the eukaryotic host but also by allowing gene exchange among intracellular bacteria.

**Conclusion.** The genome sequence of “*Ca. Amoebophilus asiaticus*” revealed a hitherto-unseen arsenal of proteins with eukaryotic domains most likely involved in host cell interaction. Its extremely reduced metabolic capabilities and the presence of a high number of proteins for host cell exploitation suggest that “*Ca. Amoebophilus asiaticus*” is not a mutualistic symbiont but a parasite. “*Ca. Amoebophilus asiaticus*” has not recently adapted to an intracellular lifestyle, but its ancestor might have been associated with eukaryotes for a long evolutionary time frame, possibly even before the divergence of the “*Ca. Amoebophilus asiaticus*” clade and the “*Candidatus Cardinium hertigii*” lineage comprising intracellular insect parasites (131). The available “*Ca. Amoebophilus asiaticus*” genome sequence opens up the possibility of future postgenomic analyses of the mechanisms employed by “*Ca. Amoebophilus*

FIG. 3. Enrichment of protein domains among proteomes of amoeba-associated bacteria. The percentage of InterPro domains in proteomes of bacteria which have been demonstrated to be able to replicate in amoebae was compared against those of 480 other bacterial proteomes. For the 59 listed domains, *P* values (Fisher’s test with FDR correction) indicate a highly significant enrichment in amoeba-associated bacteria (see Table S9 in the supplemental material). Most of these domains are proposed to be involved in host cell interaction. InterPro domains affiliating with the same eukaryotic domain are shown in the same color: Leucine-rich repeats, dark blue; TPR/SEL1 repeats, orange; Ankyrin repeats, purple; F box, red; U box, light blue. For further details, see Table S10.

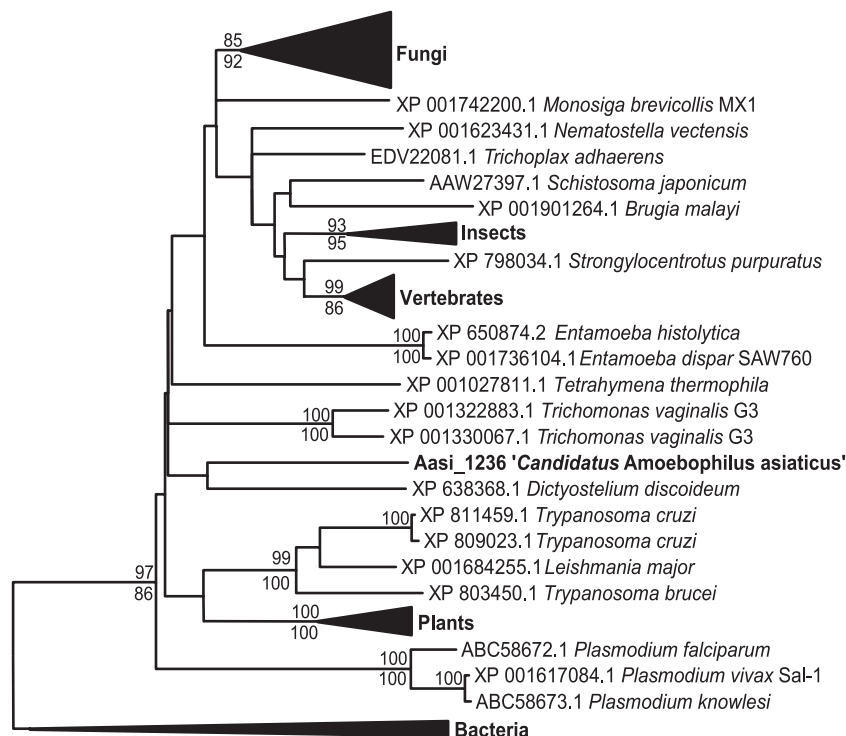


FIG. 4. Phylogenetic relationships of Aasi\_1236 and other CDP-diacylglycerol-inositol 3-phosphatidyltransferases. The phylogenetic tree is based on CDP-diacylglycerol-inositol 3-phosphatidyltransferase (EC number 2.7.8.11) amino acid sequences which were aligned using MAFFT; the tree was calculated with MEGA4 using the neighbor-joining algorithm and 1,000× bootstrapping. GenBank accession numbers are indicated. Neighbor-joining bootstrap values higher than 80% are indicated above the respective nodes. In addition, a maximum-likelihood tree was calculated using RAxML with the JTT amino acid substitution model and 1,000× bootstrapping. RAxML likelihood values higher than 80% are shown below the respective nodes. The bar represents 10% estimated evolutionary distance. Bacterial CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase (EC number 2.7.8.5) sequences were used as an outgroup.

asiaticus” to exploit its host cell, particularly its unique repertoire of proteins able to interfere with the host ubiquitin system.

#### ACKNOWLEDGMENTS

We gratefully acknowledge Waltraud Klepal and the team of the Ultrastructure Laboratory (University of Vienna) for advice and assistance with electron microscopy. We are grateful to Christian Baranyi for excellent technical assistance. Anja Spang is gratefully acknowledged for help with the analysis of IS elements. We are grateful to Mohamed Marahiel for help with the identification of the lasso peptide synthesis gene cluster.

Stephan Schmitz-Esser is supported by FWF grant P19252-B17. Work in the laboratory of Matthias Horn and Michael Wagner was funded by grants from the Austrian Science Fund FWF (Y277-B03) and the University of Vienna (Research Focus Project FS573001). Work in the laboratory of Thomas Rattei was funded by the German Research Foundation (DFG RA 1719/1-1).

#### REFERENCES

- Abd, H., B. Wretling, A. Saeed, E. Idsund, K. Hultenby, and G. Sandstrom. 2008. *Pseudomonas aeruginosa* utilizes its type III secretion system to kill the free-living amoeba *Acanthamoeba castellanii*. *J. Eukaryot. Microbiol.* **55**:235–243.
- Albert-Weissenberger, C., C. Cazalet, and C. Buchrieser. 2007. *Legionella pneumophila*—a human pathogen that co-evolved with fresh water protozoa. *Cell. Mol. Life Sci.* **64**:432–448.
- Al-Khodor, S., C. T. Price, F. Habyarimana, A. Kalia, and Y. Abu Kwaik. 2008. A Dot/Icm-translocated ankyrin protein of *Legionella pneumophila* is required for intracellular proliferation within human macrophages and protozoa. *Mol. Microbiol.* **70**:908–923.
- Amann, R., N. Springer, W. Schonhuber, W. Ludwig, E. N. Schmid, K. D. Muller, and R. Michel. 1997. Obligate intracellular bacterial parasites of acanthamoebae related to *Chlamydia* spp. *Appl. Environ. Microbiol.* **63**:115–121.
- Angot, A., A. Vergunst, S. Genin, and N. Peeters. 2007. Exploitation of eukaryotic ubiquitin signaling pathways by effectors translocated by bacterial type III and type IV secretion systems. *PLoS Pathog.* **3**:e3.
- Banerji, S., P. Aurass, and A. Flieger. 2008. The manifold phospholipases A of *Legionella pneumophila*—identification, export, regulation, and their link to bacterial virulence. *Int. J. Med. Microbiol.* **298**:169–181.
- Barthelmes, J., C. Ebeling, A. Chang, I. Schomburg, and D. Schomburg. 2007. BRENDA, AMENDA and FRENDA: the enzyme information system in 2007. *Nucleic Acids Res.* **35**:D511–D514.
- Bella, J., K. L. Hindle, P. A. McEwan, and S. C. Lovell. 2008. The leucine-rich repeat structure. *Cell. Mol. Life Sci.* **65**:2307–2333.
- Besemer, J., A. Lomsadze, and M. Borodovsky. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**:2607–2618.
- Bierne, H., C. Sabet, N. Personnic, and P. Cossart. 2007. Internalins: a complex family of leucine-rich repeat-containing proteins in *Listeria monocytogenes*. *Microbes Infect.* **9**:1156–1166.
- Birtles, R. J., T. J. Rowbotham, R. Michel, D. G. Pitcher, B. Lascola, S. Alexiou-Daniel, and D. Raoult. 2000. ‘*Candidatus Odysella thessalonicensis*’ gen. nov., sp. nov., an obligate intracellular parasite of *Acanthamoeba* species. *Int. J. Syst. Evol. Microbiol.* **50**:63–72.
- Birtles, R. J., T. J. Rowbotham, C. Storey, T. J. Marrie, and D. Raoult. 1997. Chlamydia-like obligate parasite of free-living amoebae. *Lancet* **349**:925–926.
- Bjorklund, A. K., D. Ekman, and A. Elofsson. 2006. Expansion of protein domain repeats. *PLoS Comput. Biol.* **2**:e114.
- Bonkowski, M. 2004. Protozoa and plant growth: the microbial loop in soil revisited. *New Phytol.* **162**:617–631.
- Cascales, E., and P. J. Christie. 2003. The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* **1**:137–149.

16. Caspi, R., H. Foerster, C. A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S. Y. Rhee, A. G. Shearer, C. Tissier, T. C. Walk, P. Zhang, and P. D. Karp. 2008. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **36**:D623–D631.
17. Caturegli, P., K. M. Asanovich, J. J. Walls, J. S. Bakken, J. E. Madigan, V. L. Popov, and J. S. Dumler. 2000. ankA: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect. Immun.* **68**:5277–5283.
18. 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.
19. Chain, P. S., D. V. Grafham, R. S. Fulton, M. G. Fitzgerald, J. Hostetler, D. Muzny, J. Ali, B. Birren, D. C. Bruce, C. Buhay, J. R. Cole, Y. Ding, S. Dugan, D. Field, G. M. Garrity, R. Gibbs, T. Graves, C. S. Han, S. H. Harrison, S. Highlander, P. Hugenholtz, H. M. Khouri, C. D. Kodira, E. Kolker, N. C. Kyrpides, D. Lang, A. Lapidus, S. A. Malfatti, V. Markowitz, T. Metha, K. E. Nelson, J. Parkhill, S. Pitluck, X. Qin, T. D. Read, J. Schmutz, S. Sozhamannan, P. Sterk, R. L. Strausberg, G. Sutton, N. R. Thomson, J. M. Tiedje, G. Weinstock, A. Wollam, and J. C. Detter. 2009. Genomics Genome project standards in a new era of sequencing. *Science* **326**:236–237.
20. Cho, N. H., H. R. Kim, J. H. Lee, S. Y. Kim, J. Kim, S. Cha, S. Y. Kim, A. C. Darby, H. H. Fuxelius, J. Yin, J. H. Kim, J. Kim, S. J. Lee, Y. S. Koh, W. J. Jang, K. H. Park, S. G. Andersson, M. S. Choi, and I. S. Kim. 2007. The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. *Proc. Natl. Acad. Sci. U. S. A.* **104**:7981–7986.
21. Cirillo, S. L., J. Lum, and J. D. Cirillo. 2000. Identification of novel loci involved in entry by *Legionella pneumophila*. *Microbiology* **146**(Part 6): 1345–1359.
22. 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.
23. Cornelis, G. R. 2006. The type III secretion injectisome. *Nat. Rev. Microbiol.* **4**:811–825.
24. Darby, A. C., N.-H. Cho, H.-H. Fuxelius, J. Westberg, and S. G. E. Andersson. 2007. Intracellular pathogens go extreme: genome evolution in the *Rickettsiales*. *Trends Genet.* **23**:511–520.
25. Degtyar, E., T. Zusman, M. Ehrlich, and G. Segal. 2009. A *Legionella* effector acquired from protozoa is involved in sphingolipids metabolism and is targeted to the host cell mitochondria. *Cell. Microbiol.* **11**:1219–1235.
26. Dubilier, N., C. Bergin, and C. Lott. 2008. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* **6**:725–740.
27. Duchaud, E., M. Boussaha, V. Loux, J. F. Bernardet, C. Michel, B. Kerouault, S. Mondot, P. Nicolas, R. Bossy, C. Caron, P. Bessieres, J. F. Gibrat, S. Claverol, F. Dumetz, M. Le Henaff, and A. Benmansour. 2007. Complete genome sequence of the fish pathogen *Flavobacterium psychrophilum*. *Nat. Biotechnol.* **25**:763–769.
28. Duquesne, S., D. Destoumieux-Garzon, J. Peduzzi, and S. Rebuffat. 2007. Microcins, gene-encoded antibacterial peptides from enterobacteria. *Nat. Prod. Rep.* **24**:708–734.
29. Eugster, M., C. A. Roten, and G. Greub. 2007. Analyses of six homologous proteins of *Protochlamydia amoebophila* UWE25 encoded by large GC-rich genes (lgr): a model of evolution and concatenation of leucine-rich repeats. *BMC Evol. Biol.* **7**:231.
30. Fenn, K., and M. Blaxter. 2006. *Wolbachia* genomes: revealing the biology of parasitism and mutualism. *Trends Parasitol.* **22**:60–65.
31. Ffrench-Constant, R., and N. Waterfield. 2006. An ABC guide to the bacterial toxin complexes. *Adv. Appl. Microbiol.* **58**:169–183.
32. Finn, R. D., J. Tate, J. Mistry, P. C. Coghill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer, and A. Bateman. 2008. The Pfam protein families database. *Nucleic Acids Res.* **36**:D281–D288.
33. Frishman, D., K. Albermann, J. Hani, K. Heumann, A. Metanomski, A. Zollner, and H. W. Mewes. 2001. Functional and structural genomics using PEDANT. *Bioinformatics* **17**:44–57.
34. Fritsche, T. R., M. Horn, S. Seyedirashti, R. K. Gautom, K. H. Schleifer, and M. Wagner. 1999. In situ detection of novel bacterial endosymbionts of *Acanthamoeba* spp. phylogenetically related to members of the order *Rickettsiales*. *Appl. Environ. Microbiol.* **65**:206–212.
35. Fritsche, T. R., M. Horn, M. Wagner, R. P. Herwig, K. H. Schleifer, and R. K. Gautom. 2000. Phylogenetic diversity among geographically dispersed *Chlamydiales* endosymbionts recovered from clinical and environmental isolates of *Acanthamoeba* spp. *Appl. Environ. Microbiol.* **66**:2613–2619.
36. Fuxelius, H. H., A. Darby, C. K. Min, N. H. Cho, and S. G. Andersson. 2007. The genomic and metabolic diversity of *Rickettsia*. *Res. Microbiol.* **158**: 745–753.
37. Greub, G. 2009. *Parachlamydia acanthamoebae*, an emerging agent of pneumonia. *Clin. Microbiol. Infect.* **15**:18–28.
38. Greub, G., and D. Raoult. 2003. History of the ADP/ATP-translocase-encoding gene, a parasitism gene transferred from a *Chlamydiales* ancestor to plants 1 billion years ago. *Appl. Environ. Microbiol.* **69**:5530–5535.
39. Greub, G., and D. Raoult. 2004. Microorganisms resistant to free-living amoebae. *Clin. Microbiol. Rev.* **17**:413–433.
40. Habyarimana, F., S. Al-Khodori, A. Kalia, J. E. Graham, C. T. Price, M. T. Garcia, and Y. A. Kwaik. 2008. Role for the Ankyrin eukaryotic-like genes of *Legionella pneumophila* in parasitism of protozoan hosts and human macrophages. *Environ. Microbiol.* **10**:1460–1474.
41. Hackstadt, T. 1998. The diverse habitats of obligate intracellular parasites. *Curr. Opin. Microbiol.* **1**:82–87.
42. Harb, O. S., L.-Y. Gao, and Y. A. Kwaik. 2000. From protozoa to mammalian cells: a new paradigm in the life cycle of intracellular bacterial pathogens. *Environ. Microbiol.* **2**:251–265.
43. Hebbeln, P., D. A. Rodionov, A. Alfandega, and T. Eitinger. 2007. Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module. *Proc. Natl. Acad. Sci. U. S. A.* **104**:2909–2914.
44. Heinz, E., I. Kolarov, C. Kastner, E. R. Toenshoff, M. Wagner, and M. Horn. 2007. An *Acanthamoeba* sp. containing two phylogenetically different bacterial endosymbionts. *Environ. Microbiol.* **9**:1604–1609.
45. Hongoh, Y., V. K. Sharma, T. Prakash, S. Noda, H. Toh, T. D. Taylor, T. Kudo, Y. Sakaki, A. Toyoda, M. Hattori, and M. Ohkuma. 2008. Genome of an endosymbiont coupling N<sub>2</sub> fixation to cellulolysis within protist cells in termite gut. *Science* **322**:1108–1109.
46. Horn, M. 2008. Chlamydiae as symbionts in eukaryotes. *Annu. Rev. Microbiol.* **62**:113–131.
47. Horn, M., A. Collingro, S. Schmitz-Esser, C. L. Beier, U. Purkhold, B. Fartmann, P. Brandt, G. J. Nyakatura, M. Droege, D. Frishman, T. Rattei, H. W. Mewes, and M. Wagner. 2004. Illuminating the evolutionary history of chlamydiae. *Science* **304**:728–730.
48. Horn, M., T. R. Fritsche, R. K. Gautom, K. H. Schleifer, and M. Wagner. 1999. Novel bacterial endosymbionts of *Acanthamoeba* spp. related to the *Paramecium caudatum* symbiont *Caedibacter caryophilus*. *Environ. Microbiol.* **1**:357–367.
49. Horn, M., T. R. Fritsche, T. Linner, R. K. Gautom, M. D. Harzenetter, and M. Wagner. 2002. Obligate bacterial endosymbionts of *Acanthamoeba* spp. related to the *beta-Proteobacteria*: proposal of 'Candidatus Procabacter acanthamoebae' gen. nov., sp. nov. *Int. J. Syst. Evol. Microbiol.* **52**:599–605.
50. Horn, M., M. D. Harzenetter, T. Linner, E. N. Schmid, K. D. Muller, R. Michel, and M. Wagner. 2001. Members of the *Cytophaga-Flavobacterium-Bacteroides* phylum as intracellular bacteria of acanthamoebae: proposal of 'Candidatus Amoebophilus asiaticus.' *Environ. Microbiol.* **3**:440–449.
51. Horn, M., and M. Wagner. 2004. Bacterial endosymbionts of free-living amoebae. *J. Eukaryot. Microbiol.* **51**:509–514.
52. Horn, M., M. Wagner, K. D. Muller, E. N. Schmid, T. R. Fritsche, K. H. Schleifer, and R. Michel. 2000. *Neochlamydia hartmannellae* gen. nov., sp. nov. (*Parachlamydiaceae*), an endoparasite of the amoeba *Hartmannella vermiformis*. *Microbiology* **146**:1231–1239.
53. Hu, M., P. Li, M. Li, W. Li, T. Yao, J. W. Wu, W. Gu, R. E. Cohen, and Y. Shi. 2002. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell* **111**:1041–1054.
54. Huang, J., and J. P. Gogarten. 2007. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* **8**:R99.
55. Hunter, S., R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A. F. Quinn, J. D. Selengut, C. J. Sigrist, M. Thimma, P. D. Thomas, F. Valentin, D. Wilson, C. H. Wu, and C. Yeats. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37**:D211–D215.
56. Hybiske, K., and R. S. Stephens. 2008. Exit strategies of intracellular pathogens. *Nat. Rev. Microbiol.* **6**:99–110.
57. Joachimiak, M. P., J. L. Weisman, and B. May. 2006. JColorGrid: software for the visualization of biological measurements. *BMC Bioinformatics* **7**:225.
58. Juhas, M., J. R. van der Meer, M. Gaillard, R. M. Harding, D. W. Hood, and D. W. Crook. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* **33**:376–393.
59. Jung, H. 2002. The sodium/substrate symporter family: structural and functional features. *FEBS Lett.* **529**:73–77.
60. Kanehisa, M., M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamaniishi. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**:D480–D484.
61. Katoh, K., and H. Toh. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**:286–298.
62. Khan, N. A. 2006. *Acanthamoeba*: biology and increasing importance in human health. *FEMS Microbiol. Rev.* **30**:564–595.
63. Klasson, L., T. Walker, M. Sebahia, M. J. Sanders, M. A. Quail, A. Lord,



- S. Sanders, J. Earl, S. L. O'Neill, N. Thomson, S. P. Sinkins, and J. Parkhill. 2008. Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Mol. Biol. Evol.* **25**:1877–1887.
64. Knappe, T. A., U. Linne, S. Zirah, S. Rebuffat, X. Xie, and M. A. Marahiel. 2008. Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of *Burkholderia thailandensis* E264. *J. Am. Chem. Soc.* **130**:11446–11454.
  65. Kubori, T., A. Hyakutake, and H. Nagai. 2008. *Legionella* translocates an E3 ubiquitin ligase that has multiple U-boxes with distinct functions. *Mol. Microbiol.* **67**:1307–1319.
  66. Kumar, S., M. Nei, J. Dudley, and K. Tamura. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **9**:299–306.
  67. Kurtz, S., and C. Schleiermacher. 1999. REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics* **15**:426–427.
  68. Lerat, E., and H. Ochman. 2004. Psi-Phi: exploring the outer limits of bacterial pseudogenes. *Genome Res.* **14**:2273–2278.
  69. Letunic, I., T. Doerks, and P. Bork. 2009. SMART 6: recent updates and new developments. *Nucleic Acids Res.* **37**:D229–D232.
  70. Li, J., A. Mahajan, and M. D. Tsai. 2006. Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry* **45**:15168–15178.
  71. McBride, M. J. 2001. Bacterial gliding motility: multiple mechanisms for cell movement over surfaces. *Annu. Rev. Microbiol.* **55**:49–75.
  72. McCutcheon, J. P., B. R. McDonald, and N. A. Moran. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl. Acad. Sci. U. S. A.* **106**:15394–15399.
  73. McCutcheon, J. P., B. R. McDonald, and N. A. Moran. 2009. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet.* **5**:e1000565.
  74. McCutcheon, J. P., and N. A. Moran. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* **104**:19392–19397.
  75. Merhej, V., M. Royer-Carenzi, P. Pontarotti, and D. Raoult. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Direct* **4**:13.
  76. Misaghi, S., Z. R. Balsara, A. Catic, E. Spooner, H. L. Ploegh, and M. N. Starnbach. 2006. *Chlamydia trachomatis*-derived deubiquitinating enzymes in mammalian cells during infection. *Mol. Microbiol.* **61**:142–150.
  77. Mittl, P. R., and W. Schneider-Brachert. 2007. Sell-like repeat proteins in signal transduction. *Cell. Signal.* **19**:20–31.
  78. Molmeret, M., M. Horn, M. Wagner, M. Santic, and Y. Abu Kwaik. 2005. Amoebae as training grounds for intracellular bacterial pathogens. *Appl. Environ. Microbiol.* **71**:20–28.
  79. Moran, N. A., J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42**:165–190.
  80. Moran, N. A., and G. R. Plague. 2004. Genomic changes following host restriction in bacteria. *Curr. Opin. Genet. Dev.* **14**:627–633.
  81. 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.
  82. Moya, A., J. Pereto, R. Gil, and A. Latorre. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* **9**:218–229.
  83. Nakayama, K., A. Yamashita, K. Kurokawa, T. Morimoto, M. Ogawa, M. Fukuhara, H. Urakami, M. Ohnishi, I. Uchiyama, Y. Ogura, T. Ooka, K. Oshima, A. Tamura, M. Hattori, and T. Hayashi. 2008. The whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. *DNA Res.* **15**:185–199.
  84. Newton, H. J., F. M. Sansom, J. Dao, A. D. McAlister, J. Sloan, N. P. Cianciotto, and E. L. Hartland. 2007. Sell repeat protein LpnE is a *Legionella pneumophila* virulence determinant that influences vacuolar trafficking. *Infect. Immun.* **75**:5575–5585.
  85. Nijman, S. M., M. P. Luna-Vargas, A. Velds, T. R. Brummelkamp, A. M. Dirac, T. K. Sixma, and R. Bernards. 2005. A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123**:773–786.
  86. Niklison Chirou, M., A. Bellomio, F. Dupuy, B. Arcuri, C. Minahk, and R. Morero. 2008. Microcin J25 induces the opening of the mitochondrial transition pore and cytochrome c release through superoxide generation. *FEBS J.* **275**:4088–4096.
  87. Nora, T., M. Lomma, L. Gomez-Valero, and C. Buchrieser. 2009. Molecular mimicry: an important virulence strategy employed by *Legionella pneumophila* to subvert host functions. *Future Microbiol.* **4**:691–701.
  88. O'Brien, E. A., L. B. Koski, Y. Zhang, L. Yang, E. Wang, M. W. Gray, G. Burger, and B. F. Lang. 2007. TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). *Nucleic Acids Res.* **35**:D445–D451.
  89. Ogata, H., B. La Scola, S. Audic, P. Renesto, G. Blanc, C. Robert, P. E. Fournier, J. M. Claverie, and D. Raoult. 2006. Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.* **2**:e76.
  90. Ogata, H., P. Renesto, S. Audic, C. Robert, G. Blanc, P. E. Fournier, H. Parinello, J. M. Claverie, and D. Raoult. 2005. The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol.* **3**:e248.
  91. Pan, X., A. Luhrmann, A. Satoh, M. A. Laskowski-Arce, and C. R. Roy. 2008. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* **320**:1651–1654.
  92. Parkhill, J., M. Sebahia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabinowitz, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skellton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, and D. J. Maskell. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* **35**:32–40.
  93. Pukatzki, S., R. H. Kessin, and J. J. Mekalanos. 2002. The human pathogen *Pseudomonas aeruginosa* utilizes conserved virulence pathways to infect the social amoeba *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. U. S. A.* **99**:3159–3164.
  94. Pukatzki, S., S. B. McAuley, and S. T. Miyata. 2009. The type VI secretion system: translocation of effectors and effector-domains. *Curr. Opin. Microbiol.* **12**:11–17.
  95. The R Development Core Team. 2005. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
  96. Rattei, T., P. Tischler, R. Arnold, F. Hamberger, J. Krebs, J. Krumsiek, B. Wachinger, V. Stumpfen, and W. Mewes. 2008. SIMAP—structuring the network of protein similarities. *Nucleic Acids Res.* **36**:D289–D292.
  97. Renatus, M., S. G. Parrado, A. D'Arcy, U. Eidhoff, B. Gerhartz, U. Hassleien, B. Pierrat, R. Riedl, D. Vinzenz, S. Worpenberg, and M. Kroemer. 2006. Structural basis of ubiquitin recognition by the deubiquitinating protease USP2. *Structure* **14**:1293–1302.
  98. Renesto, P., P. Dehoux, E. Gouin, L. Touqui, P. Cossart, and D. Raoult. 2003. Identification and characterization of a phospholipase D-superfamily gene in rickettsiae. *J. Infect. Dis.* **188**:1276–1283.
  99. Rocha, E. P. 2008. The organization of the bacterial genome. *Annu. Rev. Genet.* **42**:211–233.
  100. Rodriguez-Zaragoza, S. 1994. Ecology of free living amoebae. *Crit. Rev. Microbiol.* **20**:225–241.
  101. Rytkenon, A., and D. W. Holden. 2007. Bacterial interference of ubiquitination and deubiquitination. *Cell Host Microbe* **1**:13–22.
  102. Rytkenon, A., J. Poh, J. Garmendia, C. Boyle, A. Thompson, M. Liu, P. Freemont, J. C. Hinton, and D. W. Holden. 2007. SseL, a *Salmonella* deubiquitinase required for macrophage killing and virulence. *Proc. Natl. Acad. Sci. U. S. A.* **104**:3502–3507.
  103. Saier, M. H., Jr., M. R. Yen, K. Noto, D. G. Tamang, and C. Elkan. 2009. The Transporter Classification Database: recent advances. *Nucleic Acids Res.* **37**:D274–D278.
  104. Sato, H., and D. W. Frank. 2004. ExoU is a potent intracellular phospholipase. *Mol. Microbiol.* **53**:1279–1290.
  105. Schmitz-Esser, S., I. Haferkamp, S. Knab, T. Penz, M. Ast, C. Kohl, M. Wagner, and M. Horn. 2008. *Lawsonia intracellularis* contains a gene encoding a functional rickettsia-like ATP/ADP translocase for host exploitation. *J. Bacteriol.* **190**:5746–5752.
  106. Schmitz-Esser, S., N. Linka, A. Collingro, C. L. Beier, H. E. Neuhaus, M. Wagner, and M. Horn. 2004. ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to Chlamydiae and Rickettsiae. *J. Bacteriol.* **186**:683–691.
  107. Schmitz-Esser, S., E. R. Toenschhoff, S. Haider, E. Heinz, V. M. Hoenninger, M. Wagner, and M. Horn. 2008. Diversity of bacterial endosymbionts of environmental *Acanthamoeba* isolates. *Appl. Environ. Microbiol.* **74**:5822–5831.
  108. 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. U. S. A.* **102**:4866–4871.
  109. Signier, P., J. Filee, and M. Chandler. 2006. Insertion sequences in prokaryotic genomes. *Curr. Opin. Microbiol.* **9**:526–531.
  110. Signier, P., J. Perochon, L. Lestrade, J. Mahillon, and M. Chandler. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**:D32–D36.
  111. Simon, D. M., S. A. Kelchner, and S. Zimmerly. 2009. A broadscale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. *Mol. Biol. Evol.* **26**:2795–2808.
  112. Stamatakis, A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**:2688–2690.
  113. Sunagawa, S., T. Z. DeSantis, Y. M. Piceno, E. L. Brodie, M. K. DeSalvo, C. R. Voolstra, E. Weil, G. L. Andersen, and M. Medina. 2009. Bacterial diversity and white plague disease-associated community changes in the Caribbean coral *Montastraea faveolata*. *ISME J.* **3**:512–521.

114. **Tan, H. M.** 1999. Bacterial catabolic transposons. *Appl. Microbiol. Biotechnol.* **51**:1–12.
115. **Thomas, V., and G. McDonnell.** 2007. Relationship between mycobacteria and amoebae: ecological and epidemiological concerns. *Lett. Appl. Microbiol.* **45**:349–357.
116. **Tjaden, J., H. H. Winkler, C. Schwoppe, M. Van Der Laan, T. Mohlmann, and H. E. Neuhaus.** 1999. Two nucleotide transport proteins in *Chlamydia trachomatis*, one for net nucleoside triphosphate uptake and the other for transport of energy. *J. Bacteriol.* **181**:1196–1202.
117. **Trentmann, O., M. Horn, A. C. van Scheltinga, H. E. Neuhaus, and I. Haferkamp.** 2007. Enlightening energy parasitism by analysis of an ATP/ADP transporter from chlamydiae. *PLoS Biol.* **5**:e231.
118. **Tseng, T. T., B. M. Tyler, and J. C. Setubal.** 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol.* **9**(Suppl. 1):S2.
119. **Tucker, A. M., H. H. Winkler, L. O. Driskell, and D. O. Wood.** 2003. S-adenosylmethionine transport in *Rickettsia prowazekii*. *J. Bacteriol.* **185**:3031–3035.
120. **Voth, D. E., D. Howe, P. A. Beare, J. P. Vogel, N. Unsworth, J. E. Samuel, and R. A. Heinzen.** 2009. The *Coxiella burnetii* ankyrin repeat domain-containing protein family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. *J. Bacteriol.* **191**:4232–4242.
121. **Wagner, A.** 2006. Cooperation is fleeting in the world of transposable elements. *PLoS Comput. Biol.* **2**:e162.
122. **Welchman, R. L., C. Gordon, and R. J. Mayer.** 2005. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat. Rev. Mol. Cell Biol.* **6**:599–609.
123. **Wernegreen, J. J.** 2005. For better or worse: genomic consequences of intracellular mutualism and parasitism. *Curr. Opin. Genet. Dev.* **15**:572–583.
124. **Westberg, J., A. Persson, A. Holmberg, A. Goesmann, J. Lundeberg, K. E. Johansson, B. Pettersson, and M. Uhlen.** 2004. The genome sequence of *Mycoplasma mycoides* subsp. *mycoides* SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP). *Genome Res.* **14**:221–227.
125. **Winkler, H. H., and H. E. Neuhaus.** 1999. Non-mitochondrial ATP transport. *Trends Biochem. Sci.* **24**:64–68.
126. **Woyke, T., H. Teeling, N. N. Ivanova, M. Huntemann, M. Richter, F. O. Gloeckner, D. Boffelli, I. J. Anderson, K. W. Barry, H. J. Shapiro, E. Szeto, N. C. Kyrpides, M. Mussmann, R. Amann, C. Bergin, C. Ruehland, E. M. Rubin, and N. Dubilier.** 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**:950–955.
127. **Wu, D., S. C. Daugherty, S. E. Van Aken, G. H. Pai, K. L. Watkins, H. Khouri, L. J. Tallon, J. M. Zaborsky, H. E. Dunbar, P. L. Tran, N. A. Moran, and J. A. Eisen.** 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* **4**:e188.
128. **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.
129. **Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper, and J. I. Gordon.** 2003. A genomic view of the human-*Bacteroides thetaotaomicron* symbiosis. *Science* **299**:2074–2076.
130. **Ye, Z., E. O. Petrof, D. Boone, E. C. Claud, and J. Sun.** 2007. *Salmonella* effector AvrA regulation of colonic epithelial cell inflammation by deubiquitination. *Am. J. Pathol.* **171**:882–892.
131. **Zchori-Fein, E., S. J. Perlman, S. E. Kelly, N. Katzir, and M. S. Hunter.** 2004. Characterization of a ‘*Bacteroidetes*’ symbiont in *Encarsia* wasps (Hymenoptera: Aphelinidae): proposal of ‘*Candidatus Cardinium hertigii*.’ *Int. J. Syst. Evol. Microbiol.* **54**:961–968.
132. **Zhou, H., D. M. Monack, N. Kayagaki, I. Wertz, J. Yin, B. Wolf, and V. M. Dixit.** 2005. *Yersinia* virulence factor YopJ acts as a deubiquitinase to inhibit NF-kappa B activation. *J. Exp. Med.* **202**:1327–1332.
133. **Zhou, J., M. A. Bruns, and J. M. Tiedje.** 1996. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**:316–322.