The Genome of the Amoeba Symbiont “Candidatus Amoebophilus asiaticus” Reveals Common Mechanisms for Host Cell Interaction among Amoeba-Associated Bacteria

Stephan Schmitz-Esser, University of Vienna
Patrick Tischler, Technische Universität München
Roland Arnold, Technische Universität München
Jacqueline Montanaro, University of Vienna
Michael Wagner, University of Vienna, et al.

Available at: https://works.bepress.com/stephan-schmitz-esser/14/
The Genome of the Amoeba Symbiont “Candidatus Amoebophilus asiaticus” Reveals Common Mechanisms for Host Cell Interaction among Amoeba-Associated Bacteria

Stephan Schmitz-Esser, Patrick Tischler, Roland Arnold, Jacqueline Montanaro, Michael Wagner, Thomas Ratter, and Matthias Horn

Department of Microbial Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria, and Technische Universität München, Department of Genome Oriented Bioinformatics, Wissenschaftszentrum Weihenstephan, Freising, Germany

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-
has been discovered within an amoeba isolated from lake sedi-
ment (107). “Ca. Amoebophilus asiaticus” shows highest 16S rRNA similarity to “Candidatus Cardinium hertigi,” an obli-
gate intracellular parasite of arthropods able to manipulate the repro-
duction of its hosts (131). According to 16S rRNA trees, both organisms are members of a monophyletic group within the phylo-
genetically 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 distinctly 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 sharpsnouters, and “Azobacteroides pseudotrichonym-
phae,” a symbiont of an anaerobic termite gut ciliate (45, 72, 74, 127).

The genome of “Ca. Amoebophilus asiaticus” is only moder-
ately reduced in size compared to those of many other ob-
ligate intracellular bacteria (75, 123), but nevertheless, its bio-
synthetic capacities 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 analy-
sis 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 im-
portant 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. Amoebphilus asiaticus” S2a (ATCC number PRA-228) was isolated from lake sedi-
mant 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 MgSO4 · 7H2O, 2 mM Na2HPO4 · 2H2O, 1.7 mM KH2PO4, 0.05 mM Fe(NH4)2(SO4)2 · 6H2O] as de-
scribed 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. Acanthamoeba (from 3.6-liters culture) were

centrifuged (5,000 × g, 10 min), resuspended in buffer A (35 mM Tris-HCl, 250 mM Sucrose, 25 mM KCl, 10 mM MgCl2) containing 250 μM
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 centri-
fuged (4,100 × g, 5 min) to pellet the bacteria, which were subsequently resus-
pended in buffer A (containing 250 mM EDTA and centrifuged consecutively
two times for 30 min at 40,650 × g in a SM24 rotor (Sorvall). The resulting pellet was resuspended in buffer A (without EDTA) and incubated with 15 μM
DNase I (Sigma) for 1 h on ice in order to degrade remaining amoeba extra-
cellular 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 × 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% [vol/vol] cetrimide/trimethylammonium bromide [CTAB], 200 μg/ml protease K, pH 8.0) was added, followed by incubation for 30 min at
37°C. After addition of 75 μl 20% (vol/vol) SDS, the samples were incubated at
65°C for 1 h. The samples were mixed with an equal volume of chloroform-
isooamylic alcohol (24:1 [vol/vol]), the aqueous phase was recovered after centrifu-
gation (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′-GGGGAAACTCTACGGTGCC-3′) and reverse primer Ac18SR1 (5′-CC
CTTACAAAGACCCGAGC-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. S2a 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 annota-
tion (33). Protein coding genes were predicted using the GeneMark software
package, selecting default settings. No biochemical parameters were used. Transla-
tion construction were performed using the KEGG (60), 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 eu-
karyotic domains and eukaryotic proteins in the “Ca. Amoebophilus asiaticus”
genome, we performed TBlastN searches against the (yet-incomplete) Acan-
thamoeba castellanii strain Neff genome (http://www.hgsb.tmc.edu/projects/
/ microbial/microbial-detail.jsp?project_id=163) and A. castellanii strain Neff
expressed sequence tags (ESTs) available through the Prosite 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. Amoe-
bophilus 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 performed using a custom script for the Conserved Domain Database 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.ncbi.nlm.nih.gov/Taxonomy.
php) (103). Mobile genetic elements were analyzed using the Web-based resource
Bfinder (110); repeats were identified using REPutter (67). Pseudogenes were identified
using the I.Φ finder software with an E-value cutoff of 10−15 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,
Cynobacterium hutchinsoni ATCC 33406, “Gramella forsetii” KT0803, Porphyromo-
nas gingivalis W83, and Salmonibacter ruber DSM13855). Genes acquired by hori-
zontal 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 10−5 were considered to be horizontally
acquired and were further subjected to phylogenetic analysis with their homo-
logues. 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 p-distance option with 1,000× bootstrapping; all posi-
tions 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).
The genome sequence and annotation of Acanthamoeba sp. 5a2 was used. The transmission electron microscopy and reverse transcriptase PCR were performed. The results showed that the RNA preparation was free of DNA. 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. The genome of Ca. Amoeobophillus asiaticus is relatively small: 1.88 Mbp, 600 Gb, and 245,530. The genome encodes 11 proteins needed for gliding motility, a common characteristic of obligate intracellular bacteria. The genome has a relatively large genome compared to those of other obligate intracellular bacteria. The genome is relatively small: 1.88 Mbp, 600 Gb, and 245,530. The genome encodes 11 proteins needed for gliding motility, a common characteristic of obligate intracellular bacteria. The genome has a relatively large genome compared to those of other obligate intracellular bacteria.
fructose bisphosphatase (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 CO2 and acetyl coenzyme A, CaCl2. Amoebophilus asiaticus, and several other dehydrogenases. “Ca. Amoebophilus asiaticus” genome encodes the pyruvate dehydrogenase complex (nuo) and an F0-F1 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 mesodiaminopimelate, 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 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
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, leucinerich 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 Protochloramydia amoebaphila (29, 47), although their function is still unknown.

We identified 59 proteins containing tetratrico peptide repeats (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-
tion (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*/*H11005* 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 (USP) family, 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

![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.](http://jb.asm.org/)

---

<table>
<thead>
<tr>
<th>Proteins with eukaryotic domains</th>
<th>Ankyrin repeats</th>
<th>Leucine-rich repeats</th>
<th>TPR/SEL1-repeats</th>
<th>F-box domains</th>
<th>U-box domains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.5%</td>
<td>1.3%</td>
<td>3.8%</td>
<td>1.0%</td>
<td>0.6%</td>
</tr>
<tr>
<td></td>
<td>8.3%</td>
<td>0.7%</td>
<td>2.6%</td>
<td>2.6%</td>
<td>3.5%</td>
</tr>
<tr>
<td></td>
<td>3.5%</td>
<td>0.7%</td>
<td>2.6%</td>
<td>2.6%</td>
<td>3.5%</td>
</tr>
</tbody>
</table>
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 phylogenic 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 phylogenic 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 γ1 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 in 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
3-phosphatidyltransferase (EC number 2.7.8.5) sequences were used as an outgroup. 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.

FIG. 4. Phylogenetic relationships of Aasi_1236 and other CDP-diacylglycerol-inositol 3-phosphatidyltransferases. The phylogenetic tree is 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. 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.

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 Barani 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


