



Genomic Insight into Symbiosis-Induced Insect Color Change by a Facultative Bacterial Endosymbiont, “*Candidatus Rickettsiella viridis*”

Naruo Nikoh,^a Tsutomu Tsuchida,^b Taro Maeda,^c Katsushi Yamaguchi,^c Shuji Shigenobu,^c Ryuichi Koga,^d
 Takema Fukatsu^{d,e,f}

^aDepartment of Liberal Arts, the Open University of Japan, Chiba, Japan

^bGraduate School of Science and Engineering, University of Toyama, Toyama, Japan

^cNIBB Core Research Facilities, National Institute for Basic Biology, Okazaki, Japan

^dNational Institute of Advanced Industrial Science and Technology, Tsukuba, Japan

^eDepartment of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan

^fGraduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

ABSTRACT Members of the genus *Rickettsiella* are bacterial pathogens of insects and other arthropods. Recently, a novel facultative endosymbiont, “*Candidatus Rickettsiella viridis*,” was described in the pea aphid *Acyrtosiphon pisum*, whose infection causes a striking host phenotype: red and green genetic color morphs exist in aphid populations, and upon infection with the symbiont, red aphids become green due to increased production of green polycyclic quinone pigments. Here we determined the complete genome sequence of the symbiont. The 1.6-Mb circular genome, harboring some 1,400 protein-coding genes, was similar to the genome of entomopathogenic *Rickettsiella grylli* (1.6 Mb) but was smaller than the genomes of phylogenetically allied human pathogens *Coxiella burnetii* (2.0 Mb) and *Legionella pneumophila* (3.4 Mb). The symbiont’s metabolic pathways exhibited little complementarity to those of the coexisting primary symbiont *Buchnera aphidicola*, reflecting the facultative nature of the symbiont. The symbiont genome harbored neither polyketide synthase genes nor the evolutionarily allied fatty acid synthase genes that are suspected to catalyze the polycyclic quinone synthesis, indicating that the green pigments are produced not by the symbiont but by the host aphid. The symbiont genome retained many type IV secretion system genes and presumable effector protein genes, whose homologues in *L. pneumophila* were reported to modulate a variety of the host’s cellular processes for facilitating infection and virulence. These results suggest the possibility that the symbiont is involved in the green pigment production by affecting the host’s metabolism using the secretion machineries for delivering the effector molecules into the host cells.

IMPORTANCE Insect body color is relevant to a variety of biological aspects such as species recognition, sexual selection, mimicry, aposematism, and crypsis. Hence, the bacterial endosymbiont “*Candidatus Rickettsiella viridis*,” which alters aphid body color from red to green, is of ecological interest, given that different predators preferentially exploit either red- or green-colored aphids. Here we determined the complete 1.6-Mb genome of the symbiont and uncovered that, although the red-green color transition was ascribed to upregulated production of green polycyclic quinone pigments, the symbiont genome harbored few genes involved in the polycyclic quinone biosynthesis. Meanwhile, the symbiont genome contained type IV secretion system genes and presumable effector protein genes, whose homologues modulate eukaryotic cellular processes for facilitating infection and virulence in the pathogen *Legionella pneumophila*. We propose the hypothesis that the symbiont may upregu-

Received 11 May 2018 **Accepted** 15 May 2018 **Published** 12 June 2018

Citation Nikoh N, Tsuchida T, Maeda T, Yamaguchi K, Shigenobu S, Koga R, Fukatsu T. 2018. Genomic insight into symbiosis-induced insect color change by a facultative bacterial endosymbiont, “*Candidatus Rickettsiella viridis*.” *mBio* 9:e00890-18. <https://doi.org/10.1128/mBio.00890-18>.

Editor Edward G. Ruby, University of Hawaii at Manoa

Copyright © 2018 Nikoh et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Takema Fukatsu, t-fukatsu@aist.go.jp.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. Solicited external reviewers: Gordon Bennett, Univ. California Merced; Yuichi Hongoh, Tokyo Institute of Technology.

late the host's production of polycyclic quinone pigments via cooption of secretion machineries and effector molecules for pathogenicity.

KEYWORDS *Acyrtosiphon pisum*, *Coxiella*, *Legionella*, "*Candidatus Rickettsiella viridis*", aphid, facultative symbiont, genome, insect body color, polycyclic quinone pigments, polyketide synthase, type IV secretion system

Insects represent a substantial portion of the terrestrial biodiversity (1), and their prevalence and prosperity are ascribed to symbiotic microorganisms associated with them (2, 3). Some symbionts are obligatory partners for their hosts due to, for example, provisioning of essential nutrients deficient in their host's diet (4, 5), whereas other symbionts are facultative associates not necessarily needed for their host's survival (6, 7). Such facultative symbionts often affect a variety of phenotypes of their hosts either positively or negatively, depending on environmental conditions, including manipulating the host's reproduction (8), conferring the host's resistance to natural enemies (9), enhancing the host's tolerance to environmental stresses (10), and modifying the host's food plant range (11).

Aphids (Hemiptera: Aphididae) consist of some 5,000 described species, live solely on plant phloem sap, and include a number of agricultural pest species (12). Among them, the pea aphid *Acyrtosiphon pisum* is the best-studied model species, and the diversity of the symbiotic bacteria and their biological functions have been investigated in detail (4, 6). Almost all aphid species, including *A. pisum*, are associated with the obligate bacterial symbiont *Buchnera aphidicola* in specialized cells called bacteriocytes. *B. aphidicola* exhibits 100% infection frequencies in host populations and complements the host's nutritionally unbalanced plant sap diet by synthesizing essential amino acids (5, 13). In addition to *B. aphidicola*, *A. pisum* frequently harbors facultative bacterial symbionts such as *Serratia symbiotica*, *Regiella insecticola*, *Hamiltonella defensa*, *Rickettsia* sp., *Spiroplasma* sp., and others (14–18), which can have a variety of effects on the host's phenotypes with ecological consequences, including resistance to parasitoid wasps (19), protection against pathogenic fungi (20), tolerance to elevated temperature (10), influence on food plant range (11), and skewing of sex ratios (21).

The genus *Rickettsiella* constitutes the gammaproteobacterial order *Legionellales* together with the genera *Legionella* and *Coxiella* (22). All members of the *Legionellales* are specialized for endoparasitic/symbiotic lifestyle within eukaryotic cells, and some are known as human pathogens. *Coxiella* bacteria are mostly associated with ticks, and *C. burnetii* is known to be the causative agent of Q fever (23, 24). Recently, it has become evident that many, if not all, *Coxiella* bacteria are either facultative or obligate endosymbiotic associates stably maintained through generations of their host ticks (25–27), and some of them exhibit conspicuous reductive genome evolution (28, 29). *Legionella* species are endocellularly associated with aquatic protozoans such as amoebas and ciliates, and *Legionella pneumophila* is notorious as the causative agent of Legionnaires' disease (30, 31). *Rickettsiella* species have been reported to be pathogenic to insects, arachnids, and terrestrial crustaceans (32, 33). Members of the genus *Rickettsiella* from diverse insects and other arthropods, including *R. papillae* from beetle grubs, *R. grylli* from crickets, *R. chironomi* from midges, "*Candidatus Rickettsiella isopodorum*" (here referred to as "*Ca. Rickettsiella isopodorum*") from woodlice, and others, have been described, whereas the designations for a number of "*Candidatus Rickettsiella*" species thus far described have been phylogenetically regarded as synonyms of other species representing pathotypes of different host specificity (34–40). As for genomic information on *Rickettsiella*, draft genome sequences of *R. grylli* (NZ_AAQJ00000000 and MCRF00000000) and "*Ca. Rickettsiella isopodorum*" (LUKY00000000), which consist of 2, 430, and 33 contigs, respectively, have been previously deposited in the GenBank and investigated (41–43).

Recently, a novel *Rickettsiella* lineage was identified in European and American populations of *A. pisum* and described as "*Candidatus Rickettsiella viridis*" (44). "*Ca. Rickettsiella viridis*" is not pathogenic but is regarded as a facultative bacterial symbiont

of the aphid host: its infection shows stable vertical transmission through host generations (45), its infection frequency ranges from 0% to 40% in natural host populations (18, 44–46), and its infection minimally affects fitness parameters of the aphid host (44, 45). Strikingly, a novel phenotypic effect, modification of the host's body color, was discovered for "*Ca. Rickettsiella viridis*." In natural populations of *A. pisum*, red and green morphs commonly coexist and are genetically controlled, with red dominant over green (47), and when experimentally infected with "*Ca. Rickettsiella viridis*," genetically red aphids become green in color (44, 45). Considering the previous ecological studies demonstrating that ladybird beetles preferentially attack red aphids whereas parasitoid wasps tend to oviposit into green aphids (48, 49), the symbiont-induced body color change is potentially of ecologic relevance via affecting interactions with natural enemies (44, 45, 50). Additionally, it was reported that "*Ca. Rickettsiella viridis*" is protective to fungal pathogens via reducing aphid mortality and fungal sporulation (51).

The body color of *A. pisum* mainly consists of two groups of pigment molecules, namely, yellow-red color due to carotenoids and green-blue color due to polycyclic quinones and their glycosides, the so-called "aphins" (47, 52–54). Since the "*Ca. Rickettsiella viridis*" infection scarcely affects carotenoids but significantly increases green pigments in *A. pisum*, it was suggested that the symbiont may be involved in either production or induction of the green pigments (45), although it is unknown what mechanisms may underlie the symbiont-induced aphid color change. In this study, we determined and analyzed the complete genome of "*Ca. Rickettsiella viridis*," thereby gaining insight into the molecular mechanisms involved in the symbiont-induced insect color change.

RESULTS AND DISCUSSION

Sequencing of "*Ca. Rickettsiella viridis*" genome. Quality evaluation using quantitative PCR resulted in estimations indicating that the template DNA contained 3.65 μg of "*Ca. Rickettsiella viridis*" DNA (16.4%), 13.9 μg of *B. aphidicola* DNA (62.5%), and 4.68 μg of *A. pisum* DNA (21.1%). Of 4.5 Gb of Illumina raw sequence reads, 0.89 Gb of the reads were assembled into a circular chromosome of "*Ca. Rickettsiella viridis*." Hence, about 20.0% of the reads represented the "*Ca. Rickettsiella viridis*" genome, which roughly agreed with the quantitative PCR estimate.

General features of "*Ca. Rickettsiella viridis*" genome. The facultative bacterial symbiont "*Ca. Rickettsiella viridis*" exhibited a moderately reduced genome consisting of a circular 1,579,735-bp chromosome harboring 1,378 protein-coding open reading frames (ORFs) with an average size of 977 bp, which covered 86% of the whole genome. Of these ORFs, 963 were assigned to putative biological functions, 257 matched hypothetical proteins of unknown function, and 158 were unique to "*Ca. Rickettsiella viridis*". Over half of the unique hypothetical proteins (92/158) were less than 100 amino acid residues in size, and it was unclear whether they represent true gene products or not. Two ribosomal operons, three small RNA genes, and 42 tRNA genes, which include at least one corresponding tRNA for each of the 20 amino acids, were identified (Fig. 1 and Table 1; see also Table S1). Besides these genes, 57 ORFs were truncated and/or interrupted by stop codons and were regarded as presumable pseudogenes (Table 1; see also Table S2). One prophage region and five insertion sequences encoding transposases were also identified on the "*Ca. Rickettsiella viridis*" genome (Table 1).

Genomic comparison of "*Ca. Rickettsiella viridis*" with allied bacterial pathogens and symbionts. The general features of the "*Ca. Rickettsiella viridis*" genome, namely, genome size, AT content, number of ORFs, number of pseudogenes, paucity of mobile genetic elements, etc., were strikingly similar to those of *Rickettsiella grylli*, an endocellular pathogenic bacterium identified from crickets and grasshoppers, in contrast to those of *Coxiella burnetii* and *Legionella pneumophila*, related endocellular human pathogens belonging to the same gammaproteobacterial order, *Legionellales*, and also in contrast to those of other gammaproteobacterial endocellular aphid symbionts such as *Hamiltonella defensa* and *Buchnera aphidicola* (Table 1). The syntenic

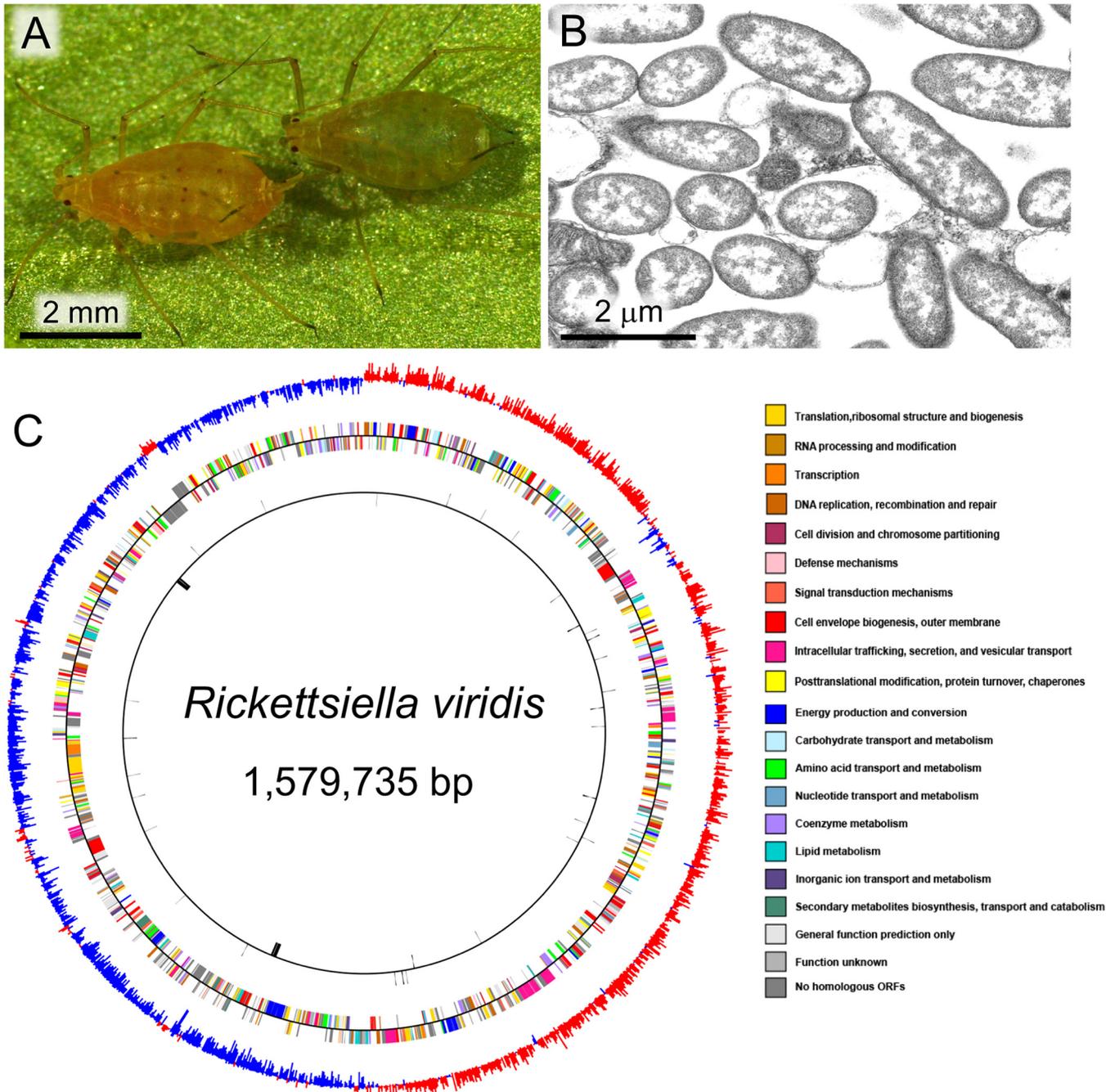


FIG 1 The pea aphid *A. pisum* and its facultative endosymbiont "*Ca. Rickettsiella viridis*." (A) An uninfected red aphid (left) and an infected green aphid (right). (B) Transmission electron microscopy of "*Ca. Rickettsiella viridis*" cells. (C) A circular view of the "*Ca. Rickettsiella viridis*" genome. On the GC skew circle, red and blue indicate high (>0) and low (<0) (G-C)/(G+C) values (1,000 nucleotide window), respectively. On the CDS circle, colors are as indicated in the key.

relationship of orthologous genes was well conserved between the genomes of "*Ca. Rickettsiella viridis*" and *R. grylli* (Fig. 2). Molecular phylogenetic analysis based on 53 ribosomal protein sequences showed close phylogenetic relationships among "*Ca. Rickettsiella viridis*," *R. grylli*, and "*Ca. Rickettsiella isopodorum*" (Fig. 3). All these results consistently support the phylogenetic relationship of "*Ca. Rickettsiella viridis*" to *R. grylli* and "*Ca. Rickettsiella isopodorum*" belonging to the same genus, *Rickettsiella*.

Gene content analysis unveiled notable differences among the genomes of "*Ca. Rickettsiella viridis*," *R. grylli*, and "*Ca. Rickettsiella isopodorum*": 371 of 1,378 protein-coding genes in the "*Ca. Rickettsiella viridis*" genome have no orthologs in the *R. grylli*

TABLE 1 General genomic features of “*Ca. Rickettsiella viridis*” and allied gammaproteobacteria

Bacterium	Host and symbiotic status	Symbiotic niche(s)	Phenotypic feature(s)	Accession no.	Chromosome size (bp)	No. of plasmids	AT content (%)	Coding content (%)	No. of predicted proteins	No. of ribosomal RNAs	No. of transfer RNAs	No. of small RNA genes	No. of pseudogenes	No. of plasmid island(s)	No. of elements
“ <i>Ca. Rickettsiella viridis</i> ”	Aphid facultative symbiont	Endo-/extracellular, systemic	Aphid's body color change	AP018005	1,579,735	0	61	86	1,378	6	42	3	57	1	5
<i>Rickettsiella grylli</i>	Cricketer pathogen	Endo-/extracellular, systemic	Cricket's pathology	AAQJ000000000 ^a	1,581,239	0	62	87	1,374	6	40	3	74	0	19
“ <i>Ca. Rickettsiella isopodorum</i> ” RCFS	Terrestrial isopod pathogen	Endo-/extracellular, systemic	Terrestrial isopod's pathology	LUKY000000000 ^b	1,493,003	0	63	87	1,254	2	40	4	39	1	2
<i>Coxiella burnetii</i> RSA 493	Tick symbiont/human pathogen	Endocellular, systemic	Human pathology (Q fever)	NC_002971	1,995,281	1	57	78	1,811	3	42	4	143	0	28
<i>Legionella pneumophila</i> Philadelphia 1	Amoeba symbiont/human pathogen	Endocellular	Human pathology (legionellosis)	NC_002942	3,397,754	1	62	89	2,922	9	43	1	13	0	24
<i>Hamiltonella defensa</i> SAT	Aphid facultative symbiont	Endo-/extracellular, systemic	Aphid's resistance to parasitic wasp	NC_012751	2,110,331	1	60	81	2,094	9	43	2	188	5	58
<i>Buchnera aphidicola</i> APS	Aphid obligatory symbiont	Endocellular, localized bacteriocyte	Supply of essential nutrients to aphid	NC_002528	640,681	2	74	88	562	3	32	4	14	0	0
<i>Escherichia coli</i> K-12	Human gut bacterium	Extracellular, gut cavity	Free-living gut bacterium	NC_000913	4,639,675	0	49	88	4,243	22	89	65	179	10	42

^aWhole-genome shotgun sequence data, which were assembled into 2 contigs.

^bWhole-genome shotgun sequence data, which were assembled into 33 contigs.

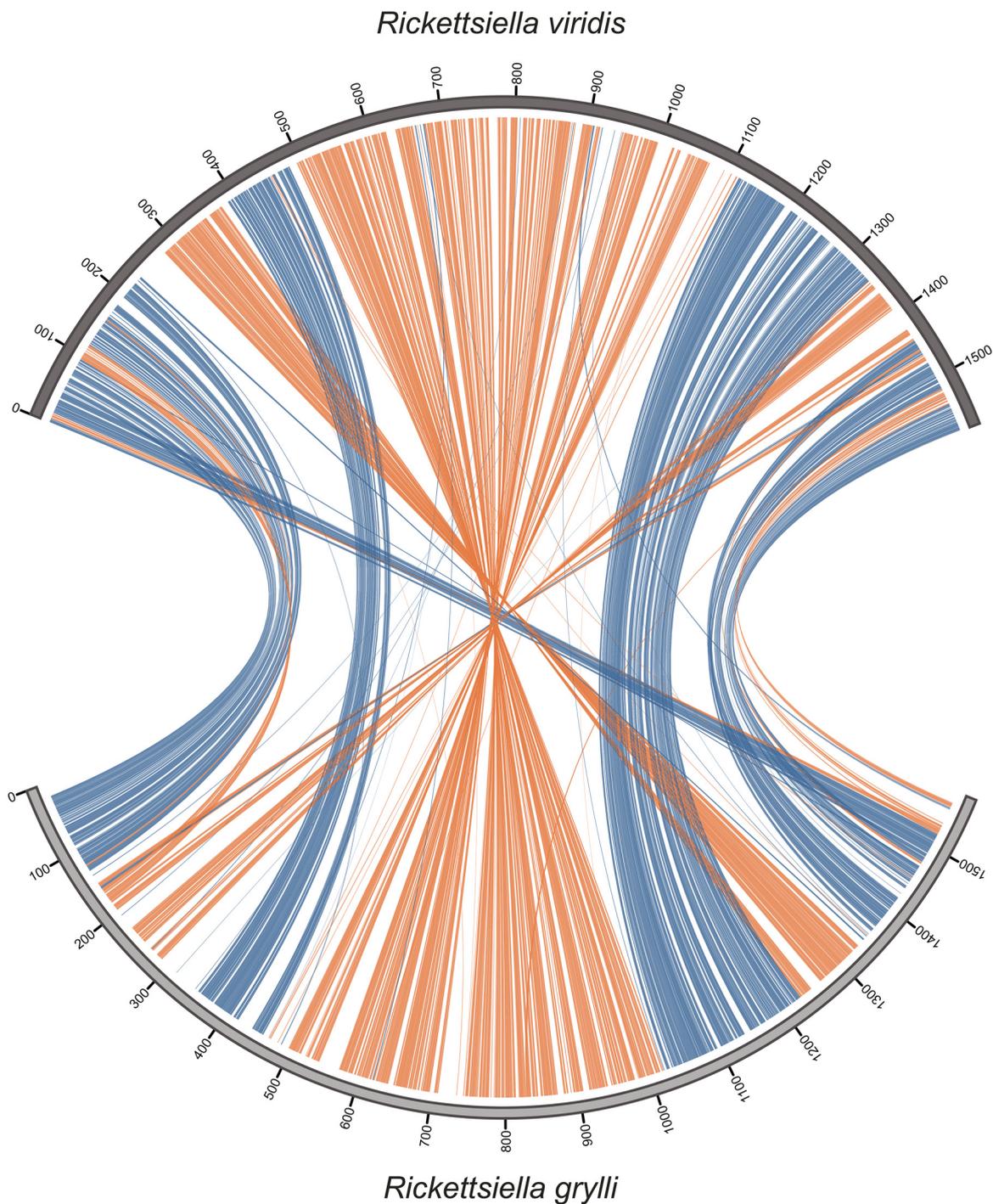


FIG 2 Plot of synteny between the genomes of “*Ca. Rickettsiella viridis*” and *R. grylli*. Orthologous genes are connected between the genomes by lines. Blue and red indicate the orthologous genes orientated in the same direction and in the reverse direction, respectively.

and “*Ca. Rickettsiella isopodorum*” genomes, whereas 257 of 1,374 protein-coding genes in the *R. grylli* genome have no orthologs in the “*Ca. Rickettsiella viridis*” genome (Fig. 4A). Furthermore, the majority of these lineage-specific genes, namely, 344 genes in “*Ca. Rickettsiella viridis*” and 241 genes in *R. grylli*, have no orthologs in the genomes of *C. burnetii* and *L. pneumophila* (Fig. 4B). These patterns suggest that many lineage-specific genes evolved after the divergence of the genus *Rickettsiella*, presumably through repeated gene losses and acquisitions.

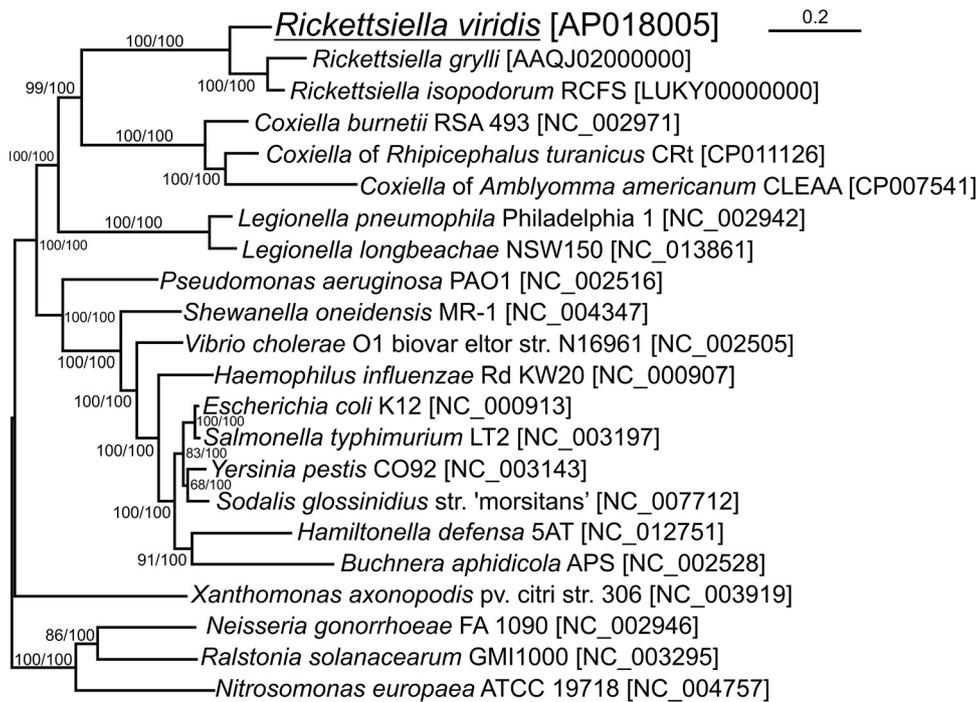


FIG 3 Phylogenetic placement of “*Ca. Rickettsiella viridis*” in the gammaproteobacteria. The maximum likelihood phylogeny is inferred from 53 concatenated ribosomal protein sequences (6,687 aligned amino acid sites). Statistical support values for each clade are shown at each node in the order of maximum likelihood/Bayesian analyses.

Reductive genome evolution in “*Ca. Rickettsiella viridis*” and allied bacterial pathogens and symbionts. The genomes of “*Ca. Rickettsiella viridis*” (1.6 Mb; an insect symbiont) and *R. grylli* (1.6 Mb; an insect pathogen) were smaller than the genome of *C. burnetii* (2.0 Mb; a tick symbiont/human pathogen) and were remarkably smaller than the genome of *L. pneumophila* (3.4 Mb; a protozoan symbiont/human pathogen) (Table 1). Relative-rate tests revealed that the molecular evolutionary rate of “*Ca. Rickettsiella viridis*” was comparable to that of *R. grylli*, slightly but significantly higher than that of *C. burnetii*, and significantly much higher than that of *L. pneumophila*

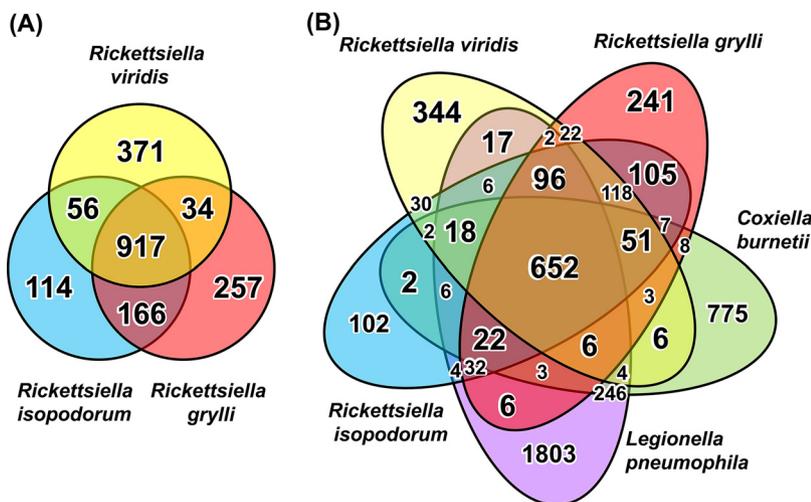


FIG 4 Venn diagrams comparing gene contents among “*Ca. Rickettsiella viridis*,” *R. grylli*, “*Ca. Rickettsiella isopodorum*,” *C. burnetii*, and *L. pneumophila*. (A) “*Ca. Rickettsiella viridis*” versus *R. grylli* versus “*Ca. Rickettsiella isopodorum*.” (B) “*Ca. Rickettsiella viridis*” versus *R. grylli* versus “*Ca. Rickettsiella isopodorum*” versus *C. burnetii* versus *L. pneumophila*.

(Table S3). These patterns suggest the possibility that reductive genome evolution has been more prominent in “*Ca. Rickettsiella viridis*,” *R. grylli*, and *C. burnetii* than in *L. pneumophila*, which may be relevant to their ecology and lifestyle. In their life cycle, *Rickettsiella* and *Coxiella* species are mainly associated with terrestrial arthropods, including insects, ticks, and isopods (23, 25, 33), whereas *Legionella* species are mainly associated with aquatic protozoans such as amoebas and ciliates (30, 55, 56). Considering that terrestrial extrahost conditions entailing desiccation, UV irradiation, and other environmental stresses must be extremely harsh for endocellular bacteria in contrast to aquatic extrahost conditions, horizontal transmissions across their host lineages may be relatively limited for *Rickettsiella* and *Coxiella* species in comparison with *Legionella* species. Actually, stable vertical transmission through host generations has been observed in several arthropod-associated *Rickettsiella* and *Coxiella* strains (25, 45). Although speculative, such differences in their lifestyle and transmission mode may be relevant to the different levels of their reductive genome evolution (57, 58). Notably, recent studies have unveiled that some *Coxiella* strains from ticks exhibit further genome reduction to 1.7 Mb for *Coxiella* sp. strain CRT (symbiont of *Rhipicephalus turanicus*) (28) and even to 0.66 Mb for *Coxiella* sp. strain CLEAA (symbiont of *Amblyomma americanum*) (29), which are both considered vitamin-provisioning nutritional mutualists for their host ticks (28, 29). Presumably, *Coxiella* sp. strain CLEAA may represent a more advanced stage of the host-symbiont coevolution than *Coxiella* sp. strain CRT and “*Ca. Rickettsiella viridis*.”

Metabolic capacity of “*Ca. Rickettsiella viridis*” genome. The “*Ca. Rickettsiella viridis*” genome, like other endocellular bacterial genomes, retained many genes responsible for basic cellular processes such as translation, replication, and energy production (Table S4). Gene content analysis revealed that “*Ca. Rickettsiella viridis*” is unable to synthesize most essential amino acids, some nonessential amino acids, and some vitamins and cofactors (Fig. 5 and 6). Probably, these molecules are provided by the host aphid and/or the coexisting primary symbiont *B. aphidicola*. Previous studies showed that, in some sap-sucking insects such as cicadas, spittlebugs, leafhoppers, aphids, and other hemipterans, bacteriocyte-associated cosymbionts often exhibit metabolic complementarity wherein incomplete biosynthetic pathways encoded by each of the two symbiont genomes constitute the complete metabolism as a whole when combined (59–62). However, such relationships were not observed between *B. aphidicola* and “*Ca. Rickettsiella viridis*” in *A. pisum*; for example, B vitamin synthesis genes missing in the “*Ca. Rickettsiella viridis*” genome were often also lacking in the *B. aphidicola* genome (Fig. 7), which should reflect the nonessential and facultative nature of “*Ca. Rickettsiella viridis*” for the host aphid (44, 45), like that of other facultative aphid symbionts, including *S. symbiotica*, *H. defensa*, and *R. insecticola* (63–65).

Absence of biosynthesis genes for polycyclic quinone pigments in “*Ca. Rickettsiella viridis*” genome. The green color of *A. pisum* and other aphids has been attributed to pigment molecules mainly consisting of polycyclic quinones and/or their glucosides, the so-called aphins (45, 52–54), although the biosynthesis pathways and genes responsible for synthesis of aphins have been poorly characterized (54). Polycyclic quinones are synthesized by condensation of acetyl-coenzyme A (acetyl-CoA) and malonyl-CoA via the polyketide pathway, a process related to fatty acid synthesis (66, 67). Several types of polyketide synthase genes have been found in a variety of bacteria, including *Legionella* species (68–70). Polyketide synthases are extremely large multidomain enzymes, usually consisting of eight types of domains (71, 72). Some ORFs in the “*Ca. Rickettsiella viridis*” genome showed sequence similarity to ORFs corresponding to one of the domains of polyketide synthases. For example, *fabF* and *fabD* harbored in the “*Ca. Rickettsiella viridis*” genome exhibited partial sequence similarities to the keto synthetase domain and the acyl transferase domain of polyketide synthase gene *mscE* of *Streptomyces carzinostaticus*. The proteins encoded by these genes are involved in fatty acid biosynthesis. However, no ORFs in the “*Ca. Rickettsiella viridis*” genome

	<i>Rickettsiella viridis</i>	<i>Rickettsiella grylli</i>	<i>Rickettsiella isopodorum</i>	<i>Coxiella burnetii</i>	<i>Legionella pneumophila</i>	<i>Hamiltonella defensa</i>	<i>Buchnera aphidicola</i>	<i>Escherichia coli</i>
Amino acid biosynthesis								
Essential amino acid								
Arginine [10]	3	1	3	2	7	2	10	10
Threonine [4]	1	1	1	1	2	4	4	4
Methionine [4]	1	2	2	2	1	0	1	4
Isoleucine [6]	0	0	0	1	0	0	4	6
Valine [5]	0	0	0	1	0	0	4	5
Leucine [9]	0	0	0	2	0	0	8	9
Lysine [9]	9	9	9	7	9	8	9	9
Histidine [8]	1	1	1	0	8	0	8	8
Phenylalanine [2]	1	1	1	1	2	1	1	2
Tryptophan [5]	0	0	0	4	5	0	5	5
Chorismate [7]	0	0	0	7	7	7	7	7
Non-essential amino acid								
Tyrosine [2]	1	1	1	2	1	1	0	1
Cysteine [2]	1	0	1	1	1	0	2	2
Glycine [1]	1	1	1	1	1	1	1	1
Aspartate [1]	1	1	1	1	1	1	0	1
Glutamate [1]*	1	1	1	1	1	1	2(2)	1
Alanine [1]	0	0	0	1	1	1	1	1
Asparagine [1]	0	1	0	1	0	1	0	1
Glutamine [1]	0	0	0	1	1	0	0	1
Proline [3]	1	0	1	1	3	1	0	3
Serine [3]	0	0	0	3	2	2	1	3
Coenzyme biosynthesis								
Biotin [6]	4	5	5	6	6	6	3	6
Folate [10]	5	5	5	9	8	8	2	9
Coenzyme A [8]	5	5	5	7	6	4	5	8
Flavin [6]	6	6	6	6	6	6	6	6
Pyridoxal 5'-phosphate [7]*	2(2)	2(2)	2(2)	5	5	7	3	7
Thiamine [4]	4	4	4	3	1	1	1	4
NAD [5]	2	2	2	5	5	5	2	5
Ubiquinone [9]	7	7	7	8	9	9	0	9
Glutathione [2]	1	2	2	2	2	2	2	2
Heme [10]	10	10	10	10	10	9	3	10
Lipoate [2]	2	2	2	2	2	2	2	2
Nucleotide biosynthesis								
Purine nucleotide [10]*	9	9	9	10	10	10	9[9]	10
Pyrimidine nucleotide [18]	17	16	16	17	17	17	18	18
Carbohydrate metabolism								
Tricarboxylic acid cycle [14]	14	14	14	14	14	14	3	14
Pentose-phosphate pathway [7]	3	4	3	4	5	7	7	7
Glycolysis [9]	9	9	9	9	9	9	9	9
Beta oxidation of fatty acid [4]	4	1	4	3	4	0	0	4
Energy metabolism								
NADH dehydrogenase [12]	12	12	12	12	12	12	12	12
Cytochrome c oxidase [4]	0	3	2	4	3	4	4	4
Cytochrome bd oxidase [2]	2	2	2	2	2	0	0	2
F ₀ F ₁ ATP synthase [8]	8	8	8	8	8	8	8	8
Cell wall, Membrane								
Fatty acid [11]	11	11	11	10	10	11	5	11
Polyisoprenoid [9]	9	9	9	7	8	9	9	9
Phospholipid [9]	7	7	7	8	9	9	1	9
UDP-N-acetyl-D-glucosamine[3]	3	3	3	3	3	3	3	3
KDO2-Lipid A [13]	11	11	11	11	12	13	0	13
Peptidoglycan [10]	10	10	8	10	10	10	9	10
Flagellar apparatus [26]	0	0	0	0	26	0	20	26

FIG 5 Comparison of metabolic gene repertoires among “*Ca. Rickettsiella viridis*” and allied gamma-proteobacteria. The minimal number of genes for a metabolic pathway is shown in each of the brackets. Colors indicate the ratio of retained genes to the minimal gene set for a metabolic pathway as follows: green for 100%, orange for 99% to 75%, yellow for 74% to 50%, blue for 49% to 25%, and gray for 24% to 0%. Asterisks denote the presence of an alternative pathway for biosynthesis of the final product, and numbers in square brackets show the minimal number of genes for the alternative pathway.

exhibited significant sequence similarities to known polyketide synthase genes in their full lengths. These results suggest that biosynthesis genes for green polycyclic quinone pigments are not harbored in the “*Ca. Rickettsiella viridis*” genome but are likely harbored in the host aphid genome.

Putative biological role of fatty acid catabolism genes in “*Ca. Rickettsiella viridis*” genome. In the “*Ca. Rickettsiella viridis*” genome, we identified genes for

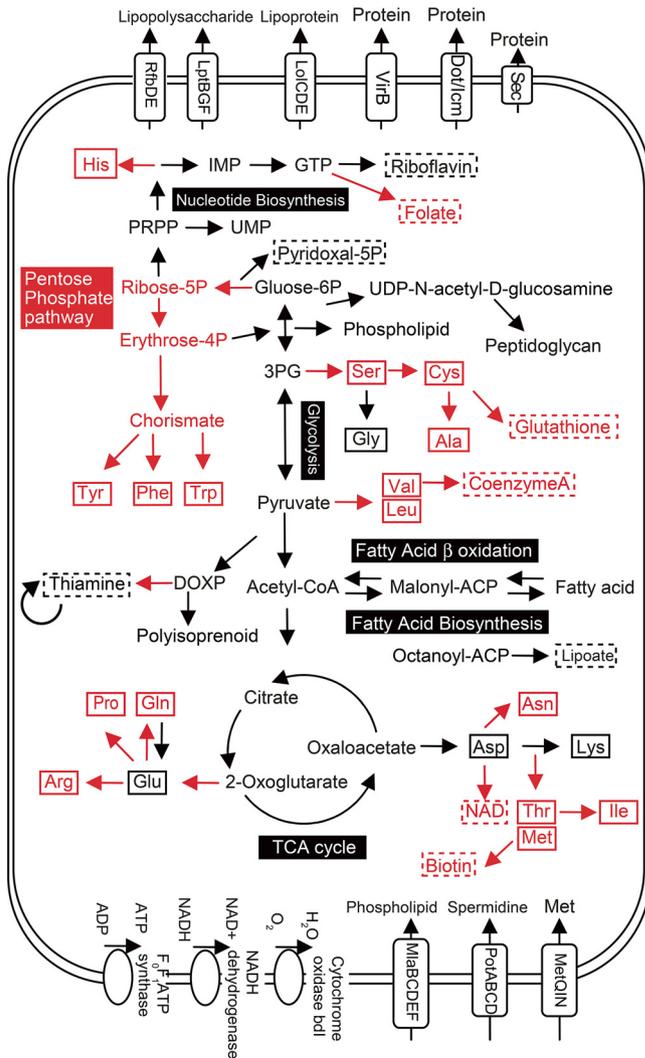


FIG 6 A hypothetical model of the metabolism and transport in “*Ca. Rickettsiella viridis*.” Main elements of metabolic pathways and transporters that are retained and lost in the “*Ca. Rickettsiella viridis*” genome are shown in black and red, respectively. Amino acids are in solid boxes, whereas vitamins and coenzymes are in dashed boxes. ACP, acyl carrier protein; DOXP, 1-deoxy-D-xylulose-5-phosphate pathway.

beta-oxidation of fatty acids, which generate acetyl-CoA and recruit it into the tricarboxylic acid (TCA) cycle for ATP production (Fig. 5 and 6). Here it is notable that other endosymbiotic bacteria of *A. pisum* whose genomes have been determined, namely, *B. aphidicola*, *S. symbiotica*, *H. defensa*, and *R. insecticola*, possess no beta-oxidation pathway (13, 63–65), which may be relevant to the fact that the sole food of the aphid, plant phloem sap, is carbohydrate-rich but devoid of lipids and proteins (5). In this context, the function of the beta-oxidation pathway of “*Ca. Rickettsiella viridis*” may be not only for its own ATP synthesis but also for provisioning acetyl-CoA to the host. Considering that malonyl-CoA or methylmalonyl-CoA is used for synthesizing the carbon backbone of polyketides (73) and that acetyl-CoA carboxylase encoded in the “*Ca. Rickettsiella viridis*” genome catalyzes the conversion of acetyl-CoA into malonyl-CoA (Fig. 6), it seems plausible, although speculative, that the beta-oxidation capability of “*Ca. Rickettsiella viridis*” might be involved in provisioning of the substrates for the host’s synthesis of the green pigments.

Secretion systems encoded by the “*Ca. Rickettsiella viridis*” genome. Secretion of macromolecules across the cell envelope is important for establishing bacterial

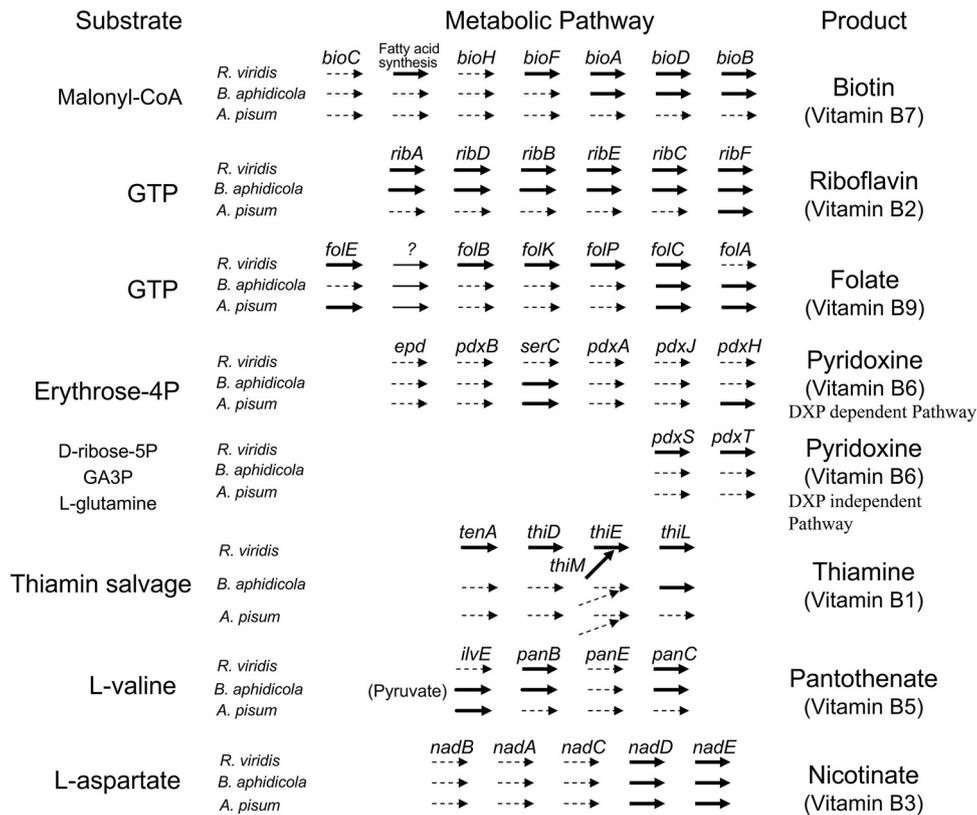


FIG 7 Biosynthetic pathway genes for B vitamins encoded in the genomes of “*Ca. Rickettsiella viridis*,” *B. aphidicola*, and *A. pisum*.

infection and virulence, and a variety of specialized molecular machineries, the so-called secretion systems, are operating in bacteria and have been classified into several evolutionarily and functionally related groups, such as the type I, II, III, IV, and V secretion systems (74). Among these, the secretion machineries ancestrally related to the bacterial conjugation systems are called the type IV secretion systems (75). The type IV secretion systems are membrane-associated transporter complexes that function to deliver DNA and protein substrates across the bacterial cell envelope to other bacterial cells, to eukaryotic host cells, or to extracellular milieu (76). A survey of the “*Ca. Rickettsiella viridis*” genome revealed several secretion systems, one of which was the Dot/Icm type IV secretion system commonly present in the endocellular bacterial members of the *Legionellales*, including *R. grylli*, “*Ca. Rickettsiella isopodorum*,” *C. burnetii*, and *L. pneumophila* (Table S5) (77). Of 25 genes encoding the major components of the secretion system, 21 were found in the “*Ca. Rickettsiella viridis*” genome, whose orthologs were mostly conserved in the genomes of *R. grylli*, “*Ca. Rickettsiella isopodorum*,” *C. burnetii*, and *L. pneumophila* (Table S5). In addition, the “*Ca. Rickettsiella viridis*” genome contained the VirB type IV secretion system. Of 12 genes comprising the major components of the secretion system, 9 intact genes and a pseudogene were present in the “*Ca. Rickettsiella viridis*” genome (Table S5), wherein the genes were found in two different operon-like structures. Each of the operons was located in a mobile genetic element-like region with a transposase gene and an integrase gene (see Fig. S1 in the supplemental material), suggesting the possibility that these operons may potentially be transposable. While *R. grylli* and *C. burnetii* were devoid of the VirB type IV secretion system, *L. pneumophila* possessed 10 genes comprising the secretion system (Table S5).

Effector proteins for Dot/Icm type IV secretion system. Thus far, a number of studies have extensively screened for *L. pneumophila* genes encoding effector proteins secreted by the Dot/Icm type IV secretion system and have identified some 300

confirmed or presumable effector protein-encoding genes in the bacterial genome, although the biological functions are still unknown for the majority of them (78, 79). In the "*Ca. Rickettsiella viridis*" genome, we identified 10 genes, *arp/ankH/sdcA*, *legK1/legK3*, *lepB*, *L. pneumophila* g2359 (*lpg2359*), *lpg2628*, *mavN*, *ravC*, *ravJ*, *rsmE*, and *sidP*, that are orthologous to the effector genes identified in *L. pneumophila* (Table S6). In *L. pneumophila*, *sidP* contains a CX₅R motif, encodes a phosphoinositide-3-phosphatase, and presumably affects the host's lysosomal trafficking by delaying the phosphoinositide cascade (80); *legK1*, one of five *L. pneumophila* genes with eukaryotic protein kinase motifs, is suggested to be involved in modulation of macrophage defense or inflammatory responses via activating the host NF- κ B signaling (81); *ankH* encodes an effector protein with two ankyrin motifs and is needed for intracellular proliferation of *L. pneumophila* (82); and *mavN* encodes a putative transmembrane protein, targets vacuolar membrane, and facilitates intravacuolar iron acquisition by *L. pneumophila* (83, 84). It seems likely, although speculative, that these effector orthologs of "*Ca. Rickettsiella viridis*" may be similarly translocated via the Dot/Icm type IV secretion system and involved in infection processes and cytological modifications in the aphid-*Rickettsiella* association. The list of the effector gene candidates detected by a homology-based survey of the "*Ca. Rickettsiella viridis*" genome (Table S6) is no doubt a conservative estimate, and many more effectors are to be identified in future studies. It should also be noted that there may be additional effector molecules related to the VirB type IV secretion system encoded in the "*Ca. Rickettsiella viridis*" genome (see Table S5).

"*Ca. Rickettsiella viridis*" proteins with eukaryotic motifs. As depicted above (see Table S6), the effector proteins of "*Ca. Rickettsiella viridis*," and also those of *L. pneumophila*, often contain eukaryotic protein motifs and functional domains (85, 86). Hence, in an attempt to identify more candidate effector protein genes in "*Ca. Rickettsiella viridis*," we surveyed the "*Ca. Rickettsiella viridis*" genome using SMART (87), thereby retrieving a total of 31 protein-coding genes with eukaryotic motifs: 16 genes with ankyrin repeats; 6 genes with coiled coils; 3 genes with a serine/threonine kinase motif; and 6 genes with a GMP reductase motif, RCC1 repeats, tetratricopeptide repeats, an AAA motif, a Ras motif, or an orotidine-5'-phosphate decarboxylase motif (Table S7). Although biological aspects of these genes are unknown, it is possible that they represent symbiont-produced effector molecules functioning in the association between the aphid and "*Ca. Rickettsiella viridis*."

Conclusion and perspective. The genome of "*Ca. Rickettsiella viridis*" determined in this study, which is the first completely determined genome of a *Rickettsiella* species, illuminates the nature of the bacterium as a moderately genome-reduced, facultative endosymbiont of the aphid *A. pisum* which is phylogenetically allied to entomopathogenic *Rickettsiella* species. The gene repertoire analyses strongly suggest that "*Ca. Rickettsiella viridis*" is involved in the symbiosis-induced body color change indirectly, in which the symbiont itself cannot produce the green pigments in the absence of relevant synthetic genes but probably activates the host's pathways for synthesis of the green pigments. Future studies should focus on transcriptomic and functional analyses of the genes involved in synthesis of the green pigments encoded in the *A. pisum* genome (88), in which polyketide synthesis genes and fatty acid synthesis genes are likely candidates.

Not only the mechanisms of the green pigment synthesis but also the mechanisms as to how "*Ca. Rickettsiella viridis*" activates the host's biosynthesis pathway for the green pigments are of profound interest. In the "*Ca. Rickettsiella viridis*" genome, we identified a number of type IV secretion system genes and possible effector protein genes. These molecular machineries may function for infection and proliferation of the symbiont in the host insect cells. It seems likely, although speculative, that the molecular machineries may also be involved in manipulation of the host's cellular processes and metabolic pathways, thereby realizing the upregulation of the host's green pigment production. Judging from the comparative genomics of *Coxiella* and

Legionella (and *Rickettsiella* in this study), members of the gammaproteobacterial order *Legionellales* must have ancestrally acquired the molecular mechanisms consisting of the Dot/Icm type IV secretion system (77), which has been followed by the evolution of extremely diverse effector proteins for establishing their endocellular lifestyle within eukaryotic host cells (89). The presence of these molecular machineries may have predisposed the ancestor of “*Ca. Rickettsiella viridis*” to achieve the striking symbiotic phenotype of modifying the host’s body color.

For a long time, *C. burnetii* and *L. pneumophila* have been recognized as human pathogens that cause Q fever and Legionnaires’ disease, respectively (24, 30). However, recent ecological studies have revealed that these human pathogens are rather exceptional among the diversity of *Coxiella* and *Legionella* species; the majority of them exist in the environment as endocellular bacteria in ticks for *Coxiella* species (25, 26) and as endocellular bacteria in aquatic amoebas and ciliates for *Legionella* species (30, 31). *Rickettsiella* species are known as endocellular pathogens/symbionts of insects and other arthropods without known vertebrate hosts (32, 33). The genomic commonalities and differences observed in the members of *Legionellales* highlight the molecular, cellular, and evolutionary continuum across pathology, virulence, manipulation, and other physiological and ecological consequences of parasitic/symbiotic associations between the endocellular bacteria and their eukaryotic hosts.

MATERIALS AND METHODS

Insect and DNA preparation and histology. We used a laboratory *A. pisum* strain, 4TV^{amp}/RA04^{acg}, which was generated by artificial transfer of “*Ca. Rickettsiella viridis*” from donor aphid strain RA04^{acg} into recipient aphid strain 4 TV^{amp} (45). These aphid strains were originally collected in locations around Rennes, France, from the red clover *Trifolium pratense* and the alfalfa *Medicago sativa*, respectively, and the experimental aphid strains have been maintained in the laboratory on seedlings of the broad bean *Vicia faba* (45). In total, 48 adult female aphids were dissected in phosphate-buffered saline. Debris was removed through the use of 100-, 50-, 11-, and 5- μ m-pore-size filters, and the filtrate was subjected to DNA extraction. Aliquots of the DNA sample were subjected to quantitative PCR to estimate the relative DNA contents of “*Ca. Rickettsiella viridis*,” *B. aphidicola*, and *A. pisum* by targeting the following genes: the *gyrB* gene of “*Ca. Rickettsiella viridis*” using primers RclGyrB-AF1 and RclGyrB-AR1 (45); the *dnaK* gene of *B. aphidicola* using primers BuchDnaK-AF1 and BuchDnaK-AR1 (90); and the elongation factor 1 α gene of *A. pisum* using primers ApisEF-422F and ApisEF-522R (91). The DNA amount of the each target organism was calculated as $M = \Gamma \times N \times M_{bp} \times N_A^{-1}$, where Γ is the genome size (600 Mb for *A. pisum* [88], 0.64 Mb for *B. aphidicola* [13], and, tentatively, 1.5 Mb for “*Ca. Rickettsiella viridis*”), N is the copy number of the target sequence (determined by real-time PCR), M_{bp} is the mean molar mass of a base pair (660 g mol⁻¹), and N_A is Avogadro’s number (6.02 $\times 10^{23}$ mol⁻¹). For transmission electron microscopy, dissected aphid tissues were prefixed with glutaraldehyde, postfixed with osmium tetroxide, embedded in epoxy resin, processed into ultrathin sections, stained with uranyl acetate and lead citrate, and observed under a transmission electron microscope as previously described (45).

Genome analysis. Using the genome DNA extracted as described above, a genomic library was constructed with an S220 Focused-Ultrasonicator (Covaris, Woburn, MA, USA) for fragmentation, Pippin Prep (Sage Science, Beverly, MA, USA) for size selection (target fragment size = 200 to 400 bp), and a TruSeq DNA Sample Prep kit (Illumina, CA, USA) for adapter ligation and library amplification. The library was sequenced (101 bp from each end) on a HiSeq 2000 platform (Illumina). A total of 44,139,360 raw reads (4.5 Gb) was obtained from the library. The obtained reads were assembled after an *in silico* elimination of contaminated sequences derived from the primary symbiont *B. aphidicola*. After filtering low-quality and adapter sequences, we mapped the reads to the genomes of *B. aphidicola* strains A5 (NC_011833) and APS (NC_002528) using bowtie2 (ver. 2.0.0) and used the unmapped 0.89 Gb of sequences for Velvet (ver.1.2.07) assembling. After several rounds of optimization of the Velvet parameters, we obtained the best assembly with k-mer = 91 and exp_cov = 30. The resultant assembly comprised 149 scaffolds (>1,000 bp) containing 1,865 kbp, among which 4 scaffolds (1,554 kbp, in total) were identified as representing the “*Ca. Rickettsiella viridis*” genome on the basis of high similarity to the genome of *R. gryllii* (NZ_AAQJ02000000). The gaps among the scaffolds were closed by Sanger sequencing. Putative protein-coding sequences (CDSs) were predicted using Glimmer3.0 (92). The annotation of CDSs was based on results of BLASTP searches against UniProt and the NCBI nonredundant protein database. Table S1 lists the annotated CDSs. The CDSs that exhibited database matches to a functional gene of other bacteria but were interrupted by frame shifts and/or stop codons were regarded as pseudogenes. To find pseudogenes in spacer regions between CDSs, BLASTX searches were conducted against UniProt using the spacer sequences as the queries. Table S2 lists the identified pseudogenes. tRNA genes were predicted by tRNAscan-SE (93). Other noncoding RNAs were identified in accordance with their similarity to *Escherichia coli* homologues. The bacterial insertion sequences were predicted using IS finder (94).

Molecular phylogenetic and evolutionary analyses. A set of 53 ribosomal protein genes for which orthologs were commonly identified in “*Ca. Rickettsiella viridis*” and other related representatives of the

family *Legionellales* was selected for phylogenetic analyses (Table S1). Each of the ortholog sets was aligned using MAFFT 5.6 (95), and all the alignments were concatenated. Alignment gaps and ambiguously aligned sites were excluded from analysis. The Le and Gascuel (LG) gamma-distributed rate (+G) invariable sites (+I) substitution model for the amino acid invariable sites was selected under the Akaike criterion using ProtTest v3.4.2 (96). Molecular phylogenetic analyses were conducted by two methods: maximum likelihood analysis using RAxML version 8.2.0 (97) and Bayesian analysis using MrBayes 3.1.2 (98). Bootstrap values for maximum likelihood phylogeny were obtained by 1,000 resamplings. Posterior probabilities were estimated for Bayesian phylogeny. A relative-rate test was performed using RRTree (99) on the basis of amino acid distances calculated from the concatenated alignment of the 53 ribosomal protein sequences.

Accession number(s). The annotated genome sequence of “*Ca. Rickettsiella viridis*” has been deposited in the DNA Data Bank of Japan with the accession number [AP018005](https://www.ddb-japan.org/entry/AB018005).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.00890-18>.

FIG S1, PDF file, 0.1 MB.

TABLE S1, PDF file, 0.3 MB.

TABLE S2, PDF file, 0.1 MB.

TABLE S3, PDF file, 0.1 MB.

TABLE S4, PDF file, 0.1 MB.

TABLE S5, PDF file, 0.1 MB.

TABLE S6, PDF file, 0.1 MB.

TABLE S7, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Hiroyo Asao and Sachiko Wakazuki for next-generation sequencing, and Kaoru Nikoh and Junko Makino for technical assistance.

This work was supported by JSPS KAKENHI JP22128007 and JP17H06388 to T.F. and also by NIBB Collaborative Research Program 16-402 to T.F. and 11-704 to T.T.

REFERENCES

- Grimaldi D, Engel MS. 2005. Evolution of the insects. Cambridge University Press, New York, NY.
- Bourtzis K, Miller TA. 2003. Insect symbiosis. CRC Press, Boca Raton, FL.
- Zchori-Fein E, Bourtzis K. 2011. Manipulative tenants: bacteria associated with arthropods. CRC Press, Boca Raton, FL.
- Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42:165–190. <https://doi.org/10.1146/annurev.genet.41.110306.130119>.
- Douglas AE. 2009. The microbial dimension in insect nutritional ecology. *Funct Ecol* 23:38–47. <https://doi.org/10.1111/j.1365-2435.2008.01442.x>.
- Oliver KM, Degnan PH, Burke GR, Moran NA. 2010. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu Rev Entomol* 55:247–266. <https://doi.org/10.1146/annurev-ento-112408-085305>.
- Feldhaar H. 2011. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecol Entomol* 36:533–543. <https://doi.org/10.1111/j.1365-2311.2011.01318.x>.
- Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 6:741–751. <https://doi.org/10.1038/nrmicro1969>.
- Oliver KM, Smith AH, Russell JA. 2014. Defensive symbiosis in the real world—advancing ecological studies of heritable, protective bacteria in aphids and beyond. *Funct Ecol* 28:341–355. <https://doi.org/10.1111/1365-2435.12133>.
- Montllor CB, Maxmen A, Purcell AH. 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrtosiphon pisum* under heat stress. *Ecol Entomol* 27:189–195. <https://doi.org/10.1046/j.1365-2311.2002.00393.x>.
- Tsuchida T, Koga R, Fukatsu T. 2004. Host plant specialization governed by facultative symbiont. *Science* 303:1989. <https://doi.org/10.1126/science.1094611>.
- Favret C, Eades DC. 2009. Introduction to aphid species file. *Redia* 92:115–117. http://favret.aphidnet.org/pubs/Favret_Eades_2009.pdf.
- Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. *APS. Nature* 407:81–86. <https://doi.org/10.1038/35024074>.
- Chen DQ, Campbell BC, Purcell AH. 1996. A new *Rickettsia* from a herbivorous insect, the pea aphid *Acyrtosiphon pisum* (Harris). *Curr Microbiol* 33:123–128. <https://doi.org/10.1007/s002849900086>.
- Fukatsu T, Tsuchida T, Nikoh N, Koga R. 2001. *Spiroplasma* symbiont of the pea aphid *Acyrtosiphon pisum* (Insecta: Homoptera). *Appl Environ Microbiol* 67:1284–1291. <https://doi.org/10.1128/AEM.67.3.1284-1291.2001>.
- Moran NA, Russell JA, Koga R, Fukatsu T. 2005. Evolutionary relationships of three new species of Enterobacteriaceae living as symbionts of aphids and other insects. *Appl Environ Microbiol* 71:3302–3310. <https://doi.org/10.1128/AEM.71.6.3302-3310.2005>.
- Guay JF, Boudreault S, Michaud D, Cloutier C. 2009. Impact of environmental stress on aphid clonal resistance to parasitoids: role of *Hamiltonella defensa* bacterial symbiosis in association with a new facultative symbiont of the pea aphid. *J Insect Physiol* 55:919–926. <https://doi.org/10.1016/j.jinsphys.2009.06.006>.
- Gauthier JP, Outreman Y, Mieuze L, Simon JC. 2015. Bacterial communities associated with host-adapted populations of pea aphids revealed by deep sequencing of 16S ribosomal DNA. *PLoS One* 10:e0120664. <https://doi.org/10.1371/journal.pone.0120664>.
- Oliver KM, Russell JA, Moran NA, Hunter MS. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc Natl Acad Sci U S A* 100:1803–1807. <https://doi.org/10.1073/pnas.0335320100>.
- Scarborough CL, Ferrari J, Godfray HCJ. 2005. Aphid protected from pathogen by endosymbiont. *Science* 310:1781. <https://doi.org/10.1126/science.1120180>.
- Simon JC, Boutin S, Tsuchida T, Koga R, Le Gallie JF, Frantz A, Outreman Y, Fukatsu T. 2011. Facultative symbiont infections affect aphid reproduction. *PLoS One* 6:e21831. <https://doi.org/10.1371/journal.pone.0021831>.
- Garrity GM, Bell JA, Lilburn T. 2005. *Legionellales* ord. nov., p 210. In

- Brenner DJ, Krieg NR, Staley JT, Garrity GM (ed), Bergey's manual of systematic bacteriology, 2nd ed, vol 2 (the proteobacteria), part B (the gammaproteobacteria). Springer, New York, NY.
23. Duron O, Sidi-Boumedine K, Rousset E, Moutailler S, Jourdain E. 2015. The importance of ticks in Q fever transmission: what has (and has not) been demonstrated? *Trends Parasitol* 31:536–552. <https://doi.org/10.1016/j.pt.2015.06.014>.
 24. Eldin C, Mélenotte C, Mediannikov O, Ghigo E, Million M, Edouard S, Mege JL, Maurin M, Raoult D. 2017. From Q fever to *Coxiella burnetii* infection: a paradigm change. *Clin Microbiol Rev* 30:115–190. <https://doi.org/10.1128/CMR.00045-16>.
 25. Duron O, Noël V, McCoy KD, Bonazzi M, Sidi-Boumedine K, Morel O, Vavre F, Zenner L, Jourdain E, Durand P, Arnathau C, Renaud F, Trape JF, Biguezoton AS, Cremaschi J, Dietrich M, Léger E, Appelgren A, Dupraz M, Gómez-Díaz E, Diatta G, Dayo GK, Adakal H, Zoungrana S, Vial L, Chevillon C. 2015. The recent evolution of a maternally inherited endosymbiont of ticks led to the emergence of the Q fever pathogen, *Coxiella burnetii*. *PLoS Pathog* 11:e1004892. <https://doi.org/10.1371/journal.ppat.1004892>.
 26. Machado-Ferreira E, Vizzoni VF, Balsemão-Pires E, Moerbeck L, Gazeta GS, Piesman J, Voloch CM, Soares CA. 2016. *Coxiella* symbionts are widespread into hard ticks. *Parasitol Res* 115:4691–4699. <https://doi.org/10.1007/s00436-016-5230-z>.
 27. Guizzo MG, Parizi LF, Nunes RD, Schama R, Albano RM, Tirloni L, Oldiges DP, Vieira RP, Oliveira WHC, Leite MS, Gonzales SA, Farber M, Martins O, Vaz IDS, Oliveira PL. 2017. A *Coxiella* mutualist symbiont is essential to the development of *Rhipicephalus microplus*. *Sci Rep* 7:17554. <https://doi.org/10.1038/s41598-017-17309-x>.
 28. Smith TA, Driscoll TJ, Gillespie JJ, Raghavan R. 2015. A *Coxiella*-like endosymbiont is a potential vitamin source for the lone star tick. *Genome Biol Evol* 7:831–838. <https://doi.org/10.1093/gbe/evv016>.
 29. Gottlieb Y, Lizarz I, Klasson L. 2015. Distinctive genome reduction rates revealed by genomic analyses of two *Coxiella*-like endosymbionts in ticks. *Genome Biol Evol* 7:1779–1796. <https://doi.org/10.1093/gbe/evv108>.
 30. Fields BS, Benson RF, Besser RE. 2002. *Legionella* and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev* 15:506–526. <https://doi.org/10.1128/CMR.15.3.506-526.2002>.
 31. Taylor M, Ross K, Bentham R. 2009. *Legionella*, protozoa, and biofilms: interactions within complex microbial systems. *Microb Ecol* 58: 538–547. <https://doi.org/10.1007/s00248-009-9514-z>.
 32. Fournier PE, Raoult D. 2005. Genus II. *Rickettsiella* Philip 1956, 267AL, p 241–247. In Brenner DJ, Krieg NR, Staley JT, Garrity GM (ed), Bergey's manual of systematic bacteriology, 2nd ed, vol 2 (the proteobacteria), part B (the gammaproteobacteria). Springer, New York, NY.
 33. Bouchon D, Cordaux R, Greve P. 2011. *Rickettsiella*, intracellular pathogens of arthropods, p 127–148. In Zchori-Fein E, Bourtzis K (ed), Manipulative tenants; bacteria associated with arthropods. CRC Press, Boca Raton, FL.
 34. Kleespies RG, Federici BA, Leclerque A. 2014. Ultrastructural characterization and multilocus sequence analysis (MLSA) of "*Candidatus Rickettsiella isopodorum*", a new lineage of intracellular bacteria infecting woodlice (Crustacea: Isopoda). *Syst Appl Microbiol* 37:351–359. <https://doi.org/10.1016/j.syapm.2014.04.001>.
 35. Leclerque A, Kleespies RG, Ritter CR, Schuster CS, Feiertag SF. 2011. Genetic and electron-microscopic characterization of "*Rickettsiella agriotidis*", a new *Rickettsiella* pathotype associated with wireworm, *Agriotes* sp. (Coleoptera: Elateridae). *Curr Microbiol* 63:158–163. <https://doi.org/10.1007/s00284-011-9958-5>.
 36. Kleespies RG, Marshall SDG, Schuster C, Townsend RJ, Jackson TA, Leclerque A. 2011. Genetic and electron-microscopic characterization of *Rickettsiella* bacteria from the manuka beetle, *Pyronota setosa* (Coleoptera: Scarabaeidae). *J Invertebr Pathol* 107:206–211. <https://doi.org/10.1016/j.jip.2011.05.017>.
 37. Leclerque A, Kleespies RG, Schuster C, Richards NK, Marshall SDG, Jackson TA. 2012. Multilocus sequence analysis (MLSA) of "*Rickettsiella costelytrae*" and "*Rickettsiella pyronotae*", intracellular bacterial entomopathogens from New Zealand. *J Appl Microbiol* 113:1228–1237. <https://doi.org/10.1111/j.1365-2672.2012.05419.x>.
 38. Cordaux R, Paces-Fessy M, Raimond M, Michel-Salzat A, Zimmer M, Bouchon D. 2007. Molecular characterization and evolution of arthropod-pathogenic *Rickettsiella* bacteria. *Appl Environ Microbiol* 73: 5045–5047. <https://doi.org/10.1128/AEM.00378-07>.
 39. Leclerque A, Kleespies RG. 2008. 16S rRNA-, GroEL- and MucZ-based assessment of the taxonomic position of "*Rickettsiella melolonthae*" and its implications for the organization of the genus *Rickettsiella*. *Int J Syst Evol Microbiol* 58:749–755.
 40. Leclerque A, Kleespies RG. 2008. Genetic and electron-microscopic characterization of *Rickettsiella tipulae*, an intracellular bacterial pathogen of the crane fly, *Tipula paludosa*. *J Invertebr Pathol* 98:329–334. <https://doi.org/10.1016/j.jip.2008.02.005>.
 41. Leclerque A. 2008. Whole genome-based assessment of the taxonomic position of the arthropod pathogenic bacterium *Rickettsiella grylli*. *FEMS Microbiol Lett* 283:117–127. <https://doi.org/10.1111/j.1574-6968.2008.01158.x>.
 42. Leclerque A, Kleespies RG. 2008. Type IV secretion system components as phylogenetic markers of entomopathogenic bacteria of the genus *Rickettsiella*. *FEMS Microbiol Lett* 279:167–173. <https://doi.org/10.1111/j.1574-6968.2007.01025.x>.
 43. Wang Y, Chandler C. 2016. Candidate pathogenicity islands in the genome of "*Candidatus Rickettsiella isopodorum*", an intracellular bacterium infecting terrestrial isopod crustaceans. *PeerJ* 4:e2806. <https://doi.org/10.7717/peerj.2806>.
 44. Tsuchida T, Koga R, Fujiwara A, Fukatsu T. 2014. Phenotypic effect of "*Candidatus Rickettsiella viridis*," a facultative symbiont of the pea aphid (*Acyrtosiphon pisum*), and its interaction with a coexisting symbiont. *Appl Environ Microbiol* 80:525–533. <https://doi.org/10.1128/AEM.03049-13>.
 45. Tsuchida T, Koga R, Horikawa M, Tsunoda T, Maoka T, Matsumoto S, Simon JC, Fukatsu T. 2010. Symbiotic bacterium modifies aphid body color. *Science* 330:1102–1104. <https://doi.org/10.1126/science.1195463>.
 46. Russell JA, Weldon S, Smith AH, Kim KL, Hu Y, Łukasik P, Doll S, Anastopoulos I, Novin M, Oliver KM. 2013. Uncovering symbiont-driven genetic diversity across North American pea aphids. *Mol Ecol* 22:2045–2059. <https://doi.org/10.1111/mec.12211>.
 47. Moran NA, Jarvik T. 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* 328:624–627. <https://doi.org/10.1126/science.1187113>.
 48. Losey JE, Harmon J, Ballantyne F, Brown C. 1997. A polymorphism is maintained by opposite patterns of parasitism and predation. *Nature* 388:269–272. <https://doi.org/10.1038/40849>.
 49. Libbrecht R, Gwynn DM, Fellowes MDE. 2007. *Aphidius ervi* preferentially attacks the green morph of the pea aphid, *Acyrtosiphon pisum*. *J Insect Behav* 20:25–32. <https://doi.org/10.1007/s10905-006-9055-y>.
 50. Polin S, Le Gall JC, Simon JC, Tsuchida T, Outreman Y. 2015. Conditional reduction of predation risk associated with a facultative symbiont in an insect. *PLoS One* 10:e0143728. <https://doi.org/10.1371/journal.pone.0143728>.
 51. Łukasik P, van Asch M, Guo H, Ferrari J, Godfray HC. 2013. Unrelated facultative endosymbionts protect aphids against a fungal pathogen. *Ecol Lett* 16:214–218. <https://doi.org/10.1111/ele.12031>.
 52. Bowie JH, Cameron DW, Findlay JA, Quartey JAK. 1966. Haemolymph pigments of aphids. *Nature* 210:395–397. <https://doi.org/10.1038/210395a0>.
 53. Brown KS. 1975. The chemistry of aphids and scale insects. *Chem Soc Rev* 4:263–288. <https://doi.org/10.1039/c9750400263>.
 54. Tsuchida T. 2016. Molecular basis and ecological relevance of aphid body colors. *Curr Opin Insect Sci* 17:74–80. <https://doi.org/10.1016/j.cois.2016.07.005>.
 55. Lau HY, Ashbolt NJ. 2009. The role of biofilms and protozoa in *Legionella* pathogenesis: implications for drinking water. *J Appl Microbiol* 107:368–378. <https://doi.org/10.1111/j.1365-2672.2009.04208.x>.
 56. Richards AM, Von Dwingelo JE, Price CT, Abu Kwaik YA. 2013. Cellular microbiology and molecular ecology of *Legionella*-amoeba interaction. *Virulence* 4:307–314. <https://doi.org/10.4161/viru.24290>.
 57. Bright M, Bulgheresi S. 2010. A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol* 8:218–230. <https://doi.org/10.1038/nrmicro2262>.
 58. McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26. <https://doi.org/10.1038/nrmicro2670>.
 59. Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, Moran NA, Eisen JA. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol* 4:e188. <https://doi.org/10.1371/journal.pbio.0040188>.
 60. McCutcheon JP, McDonald BR, Moran NA. 2009. Convergent evolution

- of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci U S A* 106:15394–15399. <https://doi.org/10.1073/pnas.0906424106>.
61. McCutcheon JP, Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol* 2:708–718. <https://doi.org/10.1093/gbe/evq055>.
 62. Lamelas A, Gosálbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A. 2011. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genet* 7:e1002357. <https://doi.org/10.1371/journal.pgen.1002357>.
 63. Burke GR, Moran NA. 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol Evol* 3:195–208. <https://doi.org/10.1093/gbe/evr002>.
 64. Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA. 2009. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc Natl Acad Sci U S A* 106:9063–9068. <https://doi.org/10.1073/pnas.0900194106>.
 65. Degnan PH, Leonardo TE, Cass BN, Hurwitz B, Stern D, Gibbs RA, Richards S, Moran NA. 2010. Dynamics of genome evolution in facultative symbionts of aphids. *Environ Microbiol* 12:2060–2069. <https://doi.org/10.1111/j.1462-2920.2009.02085.x>.
 66. Hopwood DA, Sherman DH. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet* 24:37–66. <https://doi.org/10.1146/annurev.ge.24.120190.000345>.
 67. Pankewitz F, Hilker M. 2008. Polyketides in insects: ecological role of these widespread chemicals and evolutionary aspects of their biogenesis. *Biol Rev* 83:209–226. <https://doi.org/10.1111/j.1469-185X.2008.00040.x>.
 68. Luckner M. 1984. Secondary metabolism in microorganisms, plants and animals, p 563. Springer-Verlag, Berlin, Germany.
 69. Katz L, Donadio S. 1993. Polyketide synthesis: prospects for hybrid antibiotics. *Annu Rev Microbiol* 47:875–912. <https://doi.org/10.1146/annurev.mi.47.100193.004303>.
 70. Donadio S, Monciardini P, Sosio M. 2007. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep* 24:1073–1109. <https://doi.org/10.1039/b514050c>.
 71. Hutchinson CR, Fujii I. 1995. Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. *Annu Rev Microbiol* 49:201–238. <https://doi.org/10.1146/annurev.mi.49.100195.001221>.
 72. Staunton J, Weissman KJ. 2001. Polyketide biosynthesis: a millennium review. *Nat Prod Rep* 18:380–416. <https://doi.org/10.1039/a909079g>.
 73. Chan YA, Podevels AM, Kevany BM, Thomas MG. 2009. Biosynthesis of polyketide synthase extender units. *Nat Prod Rep* 26:90–114. <https://doi.org/10.1039/B801658P>.
 74. Thanassi DG, Hultgren SJ. 2000. Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr Opin Cell Biol* 12:420–430. [https://doi.org/10.1016/S0955-0674\(00\)00111-3](https://doi.org/10.1016/S0955-0674(00)00111-3).
 75. Backert S, Meyer TF. 2006. Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol* 9:207–217. <https://doi.org/10.1016/j.mib.2006.02.008>.
 76. Alvarez-Martinez CE, Christie PJ. 2009. Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* 73:775–808. <https://doi.org/10.1128/MMBR.00023-09>.
 77. Segal G, Feldman M, Zusman T. 2005. The Icm/Dot type-IV secretion systems of *Legionella pneumophila* and *Coxiella burnetii*. *FEMS Microbiol Rev* 29:65–81. <https://doi.org/10.1016/j.femsre.2004.07.001>.
 78. Ninio S, Roy CR. 2007. Effector proteins translocated by *Legionella pneumophila*: strength in numbers. *Trends Microbiol* 15:372–380. <https://doi.org/10.1016/j.tim.2007.06.006>.
 79. Zhu W, Banga S, Tan Y, Zheng C, Stephenson R, Gately J, Luo ZQ. 2011. Comprehensive identification of protein substrates of the Dot/Icm type IV transporter of *Legionella pneumophila*. *PLoS One* 6:e17638. <https://doi.org/10.1371/journal.pone.0017638>.
 80. Toulabi L, Wu X, Cheng Y, Mao Y. 2013. Identification and structural characterization of a *Legionella* phosphoinositide phosphatase. *J Biol Chem* 288:24518–24527. <https://doi.org/10.1074/jbc.M113.474239>.
 81. Ge J, Xu H, Li T, Zhou Y, Zhang Z, Li S, Liu L, Shao F. 2009. A *Legionella* type IV effector activates the NF- κ B pathway by phosphorylating the I κ B family of inhibitors. *Proc Natl Acad Sci U S A* 106:13725–13730. <https://doi.org/10.1073/pnas.0907200106>.
 82. Habyarimana F, Price CT, Santic M, Al-Khodori S, Kwaik YA. 2010. Molecular characterization of the Dot/Icm-translocated AnkH and AnkJ eukaryotic-like effectors of *Legionella pneumophila*. *Infect Immun* 78:1123–1134. <https://doi.org/10.1128/IAI.00913-09>.
 83. Isaac DT, Laguna RK, Valtz N, Isberg RR. 2015. MavN is a *Legionella pneumophila* vacuole-associated protein required for efficient iron acquisition during intracellular growth. *Proc Natl Acad Sci U S A* 112:E5208–E5217. <https://doi.org/10.1073/pnas.1511389112>.
 84. Portier E, Zheng H, Sahr T, Burnside DM, Mallama C, Buchrieser C, Cianciotto NP, Héchard Y. 2015. *IroT/mavN*, a new iron-regulated gene involved in *Legionella pneumophila* virulence against amoebae and macrophages. *Environ Microbiol* 17:1338–1350. <https://doi.org/10.1111/1462-2920.12604>.
 85. Brüggemann H, Cazalet C, Buchrieser C. 2006. Adaptation of *Legionella pneumophila* to the host environment: role of protein secretion, effectors and eukaryotic-like proteins. *Curr Opin Microbiol* 9:86–94. <https://doi.org/10.1016/j.mib.2005.12.009>.
 86. Pan X, Lührmann A, Satoh A, Laskowski-Arce MA, Roy CR. 2008. Eukaryotic repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* 320:1651–1654. <https://doi.org/10.1126/science.1158160>.
 87. Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART: a Web-based tool for the study of genetically mobile domains. *Proc Natl Acad Sci U S A* 95:5857–5864. <https://doi.org/10.1073/pnas.95.11.5857>.
 88. International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol* 8:e1000313. <https://doi.org/10.1371/journal.pbio.1000313>.
 89. Burstein D, Amaro F, Zusman T, Lifshitz Z, Cohen O, Gilbert JA, Pupko T, Shuman HA, Segal G. 2016. Genomic analysis of 38 *Legionella* species identifies large and diverse effector repertoires. *Nat Genet* 48:167–175. <https://doi.org/10.1038/ng.3481>.
 90. Koga R, Tsuchida T, Fukatsu T. 2003. Changing partners in an obligate symbiosis: a facultative endosymbiont can compensate for loss of the essential endosymbiont *Buchnera* in an aphid. *Proc Biol Sci* 270:2543–2550. <https://doi.org/10.1098/rspb.2003.2537>.
 91. Sakurai M, Koga R, Tsuchida T, Meng XY, Fukatsu T. 2005. *Rickettsia* symbiont of the pea aphid *Acyrtosiphon pisum*: novel cellular tropism, effect on the host fitness, and interaction with the essential symbiont *Buchnera*. *Appl Environ Microbiol* 71:4069–4075. <https://doi.org/10.1128/AEM.71.7.4069-4075.2005>.
 92. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679. <https://doi.org/10.1093/bioinformatics/btm009>.
 93. Lowe TM, Eddy SR. 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <https://doi.org/10.1093/nar/25.5.0955>.
 94. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34:D32–D36. <https://doi.org/10.1093/nar/gkj014>.
 95. Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33:511–518. <https://doi.org/10.1093/nar/gki198>.
 96. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–1165. <https://doi.org/10.1093/bioinformatics/btr088>.
 97. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>.
 98. Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>.
 99. Robinson-Rechavi M, Huchon D. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* 16:296–297. <https://doi.org/10.1093/bioinformatics/16.3.296>.