AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere

Host-Microbial Interactions | Research Article

Orientia tsutsugamushi: comprehensive analysis of the mobilome of a highly fragmented and repetitive genome reveals the capacity for ongoing lateral gene transfer in an obligate intracellular bacterium

Suparat Giengkam,[1] Chitrasak Kullapanich,[1] Jantana Wongsantichon,[1] Haley E. Adcox,[2] Joseph J. Gillespie,[3] Jeanne Salje[1,4,5,6]

AUTHOR AFFILIATIONS See affiliation list on p. 27.

ABSTRACT The rickettsial human pathogen *Orientia tsutsugamushi* (Ot) is an obligate intracellular Gram-negative bacterium with one of the most highly fragmented and repetitive genomes of any organism. Around 50% of its ~2.3-Mb genome is composed of repetitive DNA that is derived from the highly proliferated Rickettsiales amplified genetic element (RAGE). RAGE is an integrative and conjugative element (ICE) that is present in a single Ot genome in up to 93 copies, most of which are partially or heavily degraded. In this report, we analyzed RAGEs in eight fully sequenced Ot genomes and manually curated and re-annotated all RAGE-associated genes, including those encoding DNA mobilization proteins, P-type (*vir*), and F-type (*tra*) type IV secretion system (T4SS) components, ankyrin repeat- and tetratricopeptide repeat-containing effectors, and other piggybacking cargo. Originally, the heavily degraded Ot RAGEs have led to speculation that they are remnants of historical ICEs that are no longer active. Our analysis, however, identified two Ot genomes harboring one or more intact RAGEs with complete *F-T4SS* genes essential for mediating ICE DNA transfer. As similar ICEs have been identified in unrelated rickettsial species, we assert that RAGEs may play an ongoing role in lateral gene transfer within the Rickettsiales. We also identified a conserved set of gene transfer agent genes in all Ot genomes. Together these findings indicate that, despite their obligate intracellular lifestyle and host range restricted to mites, rodents, and humans, Ot genomes are highly dynamic and shaped through ongoing invasions by mobile genetic elements and virus-like elements.

IMPORTANCE Obligate intracellular bacteria, or those only capable of growth inside other living cells, have limited opportunities for horizontal gene transfer with other microbes due to their isolated replicative niche. The human pathogen Ot, an obligate intracellular bacterium causing scrub typhus, encodes an unusually high copy number of a ~40 gene mobile genetic element that typically facilitates genetic transfer across microbes. This proliferated element is heavily degraded in Ot and previously assumed to be inactive. Here, we conducted a detailed analysis of this element in eight Ot strains and discovered two strains with at least one intact copy. This implies that the element is still capable of moving across Ot populations and suggests that the genome of this bacterium may be even more dynamic than previously appreciated. Our work raises questions about intracellular microbial evolution and sounds an alarm for gene-based efforts focused on diagnosing and combatting scrub typhus.

KEYWORDS *Orientia tsutsugamushi*, Rickettsiales, obligate intracellular bacteria, intracellular pathogens, mobile genetic elements, comparative genomics, gene transfer agents, integrative and conjugative elements, lateral gene transfer

Editor Alfredo G. Torres, The University of Texas Medical Branch at Galveston, Galveston, Texas, USA

Address correspondence to Jeanne Salje, jss53@cam.ac.uk.

The authors declare no conflict of interest.

See the funding table on p. 28.

Received 15 May 2023
Accepted 22 August 2023
Published 18 October 2023

Copyright © 2023 Giengkam et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

Orientia tsutsugamushi (Ot) is an obligate intracellular Gram-negative bacterium that is a symbiont of trombiculid mites and causes the vector-borne human disease scrub typhus. Ot is a member of the alphaproteobacterial order Rickettsiales, which contains three well-studied families: Anaplasmataceae, Rickettsiaceae, and Midichloriaceae (1, 2), as well as four lesser-known families that have recently been described (Deianiraeaceae, Mitibacteraceae, Gamibacteraceae, and Athabascaceae) (3–5). As a lineage within Rickettsiaceae, genus *Orientia* also includes *Candidatus Orientia chiloensis*, which has recently been identified as an endemic species in Chile (6), and *Candidatus Orientia chuto*, which was isolated from a patient in Dubai (7). There is extensive strain diversity within the Ot species, which can be found in rodents, mites, and human patients across Southeast Asia. While strain diversity corresponds to differences in virulence in patients and in animal infection models, the molecular basis of these differences in virulence is not well understood. Ot strains are often classified according to serotype groupings, which are organized based on the human serological response to the highly antigenic surface protein TSA56. Major serotype groups are named after type strains and include Karp, Kato, Gilliam, Japanese-Gilliam, TA763, Saitama, Kuroki, Kawasaki, and Shimokoshi.

At around 2.0–2.5 Mb, the genome of Ot is almost double the size of most Rickettsiales genomes and is one of the most fragmented and repetitive bacterial genomes reported to date (8–10). With almost 50% of the genome composed of repetitive DNA sequences, many experimental approaches are challenging, i.e., primer design, gene and genome sequencing, gene prediction and annotation, and comparative genomics. Complete genome sequences of two strains, Boryong and Ikeda, were published in 2008 using short-read sequencing and bacterial artificial chromosome cloning (8, 9). Six additional strains (Karp, Kato, Gilliam, TA686, UT76, and UT176) were fully sequenced in 2018 using long-read PacBio technology (11). A comparison of these eight genomes enabled the identification of 657 core genes and an open pangenome that is heavily characterized by gene duplication and pseudogenization rather than the import of novel genes (11). Of these eight strains for which complete genome sequences are available, all except TA686 are pathogenic and were isolated from human patients (11). TA686 was isolated from a tree shrew (*Tupaia glis*).

The Ot genome is dominated by an integrative and conjugative element (ICE) called Rickettsiales amplified genetic element (RAGE) (8, 9, 12) that has proliferated rampantly throughout the genome and is present in over 70 copies. RAGEs encode numerous repeated and pseudogenized genes, as well as single-copy cargo genes that appear to be important for bacterial growth and pathogenesis. While the high number of RAGE copies in the Ot genome remains unmatched, similar RAGEs have been described in several other *Rickettsia* species: *Rickettsia bellii* (RBE) (single intact copy) (13), *Rickettsia buchneri* (seven complete or near-complete genomic copies and two plasmid encoded copies) (12), *Rickettsia massiliensis* (single intact copy (14), *Rickettsia parkeri* str. Atlantic Rainforest (single intact copy) (15), *Rickettsia felis* str. LSU-Lb (one plasmid encoded copy) (16), and *Rickettsia peacockii* (one partially degraded copy) (17). In Ot, the ICE has not been controlled by the bacterial host, and the RAGE has replicated to high levels in the genome (8, 9, 11). The reasons for the differential fate of the RAGEs in Rickettsiaceae are unknown, but the small effective population size as well as the presence of population bottlenecks in obligate intracellular bacteria likely explain why it has had the ability to proliferate without strong negative selection in at least some rickettsial species.

Here, we present a thorough Ot phylogenomics analysis and re-annotation of the RAGEs and their cargo genes in eight strains: Gilliam, Boryong, UT76, UT176, Karp, Kato, Ikeda, and TA686 (Table 1). We delineate the start and stop sites of all the intact and degraded RAGEs, allowing us to identify inter-RAGE (IR) regions with conserved clusters of genes. We further describe complete RAGEs in two Ot genomes (Kato and Gilliam). Finally, we annotate and classify the genes associated with DNA mobilizations and divergent type IV secretion systems (T4SSs), as well as the numerous multicopy cargo genes within the RAGEs, including those encoding ankyrin-repeat containing

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

TABLE 1 Accession numbers of genomes used in this study

Strains	Genome accession numbers	Links
Boryong	AM494475.1	https://www.ncbi.nlm.nih.gov/nuccore/AM494475.1
	NC009488.1	https://www.ncbi.nlm.nih.gov/nuccore/NC_009488.1
UT76	LS398552.1	https://www.ncbi.nlm.nih.gov/nuccore/LS398552
UT176	LS398547.1	https://www.ncbi.nlm.nih.gov/nuccore/LS398547.1
Karp	LS398548.1	https://www.ncbi.nlm.nih.gov/nuccore/LS398548.1
Kato	LS398550.1	https://www.ncbi.nlm.nih.gov/nuccore/LS398550.1
Ikeda	AP008981.1	https://www.ncbi.nlm.nih.gov/nuccore/AP008981.1
	NC_010793.1	https://www.ncbi.nlm.nih.gov/nuccore/NC_010793.1
TA686	LS398549.1	https://www.ncbi.nlm.nih.gov/nuccore/LS398549.1
Gilliam	LS398551.1	https://www.ncbi.nlm.nih.gov/nuccore/LS398551.1

proteins (Anks) and tetratricopeptide repeat containing proteins (TPRs), which are putative secreted effectors. This detailed disentangling of the superfluous RAGE-dominated mobilome from the core and accessory Ot genome is expected to enlighten research on Ot biology and overall genome evolution in obligate intracellular bacteria.

RESULTS AND DISCUSSION

RAGE and IR regions

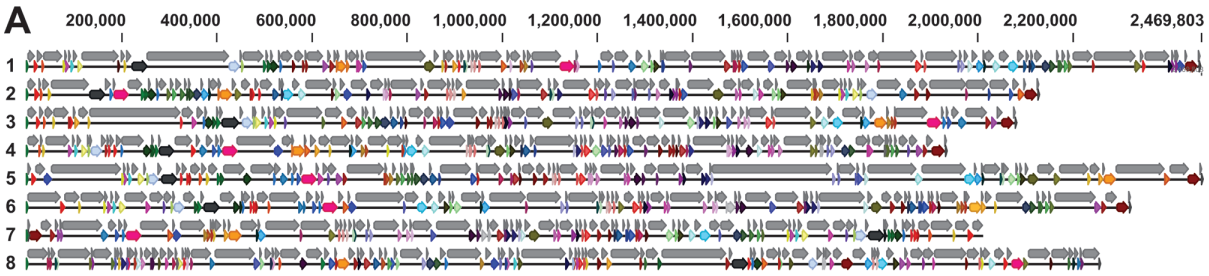
The RAGE is an ICE that is present in Ot and certain other Rickettsiales genomes. The degree of amplification and degradation of RAGE in the Ot genome is so extensive—making up around 50% of the Ot genome in 71–93 distinct genomic regions—that the beginning and end sites of RAGEs cannot be easily identified by visual inspection. Here, we established objective criteria for the classification of genes into Ot RAGEs and manually delineated each RAGE in the eight Ot strains in our study.

First, one or more copies of each of the following mobilization genes must be present: integrase (*int*), transposase, and F-type T4SS genes/relaxosome (*tra/trb*). Second, genes encoding one or more previously defined cargo or regulation proteins (9) must be present: membrane proteins (reclassified here as Ot_RAGE_membrane protein, see below), DNA adenine methyltransferase (*dam*), DNA helicase, ATP-binding proteins (*mrp*), histidine kinases (HKs), SpoT-related proteins (synthetase and/or hydrolase domains), HNH endonuclease, peroxiredoxin, Anks, and TPRs. Third, the RAGE region begins with the first RAGE-associated gene and continues until a previously defined core gene (11) or another RAGE is reached. A run of core genes is classified as an IR element. The identification of boundaries between RAGEs is described in Materials and Methods. Fourth, given the abundance of genes encoding hypothetical proteins (HPs) within RAGEs, those located between mobilization/cargo RAGE genes are classified as being part of that RAGE. However, HP-encoding genes located between RAGE mobilization/cargo genes and Ot core genes that cannot be resolved as being within either RAGE or IR regions are classified as isolated HP-encoding genes. Fifth and final, single or multiple mobilization or cargo RAGE genes are classified as isolated mobile genes or cargo genes, respectively. Some new cargo genes were identified by virtue of residing within RAGE regions in most or all genomes, and these are discussed below.

In this way, the entire genome of each Ot strain was classified into the following regions: RAGE region, IR region, isolated HP-encoding gene, isolated mobilization gene, and isolated cargo gene (data set S1). Using these criteria, we identified 71–93 RAGEs in the eight analyzed genomes (Fig. 1A and B). The patterns of RAGE fragmentation and pseudogenization varied extensively between strains, and it was not possible to map RAGEs between strains (Fig. 1A). This implies that RAGEs entered Ot strains one or more times as intact elements and subsequently underwent replication, pseudogenization, and recombination in independent trajectories.

Most RAGEs in Ot are degraded, both in terms of either completely lacking RAGE-associated genes or retaining genes that have been truncated or fragmented into

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.



B

	Ot strain	Total RAGE	Complete RAGE w/o truncated/ split genes	Complete RAGE w/ truncated/ split genes	Incomplete RAGE	No. IR regions	No. isolated mobilisation genes	No. isolated cargo genes	No. isolated HP-encoding genes
1	Gilliam	91	3	7	81	95	52	10	49
2	Boryong	84	0	1	83	90	25	17	65
3	UT76	71	0	2	69	86	31	11	52
4	UT176	72	0	3	69	96	42	9	40
5	Karp	77	0	13	64	84	25	13	39
6	Kato	81	1	10	70	92	33	12	54
7	Ikeda	76	0	0	76	91	36	16	57
8	TA686	93	0	7	86	103	33	12	46

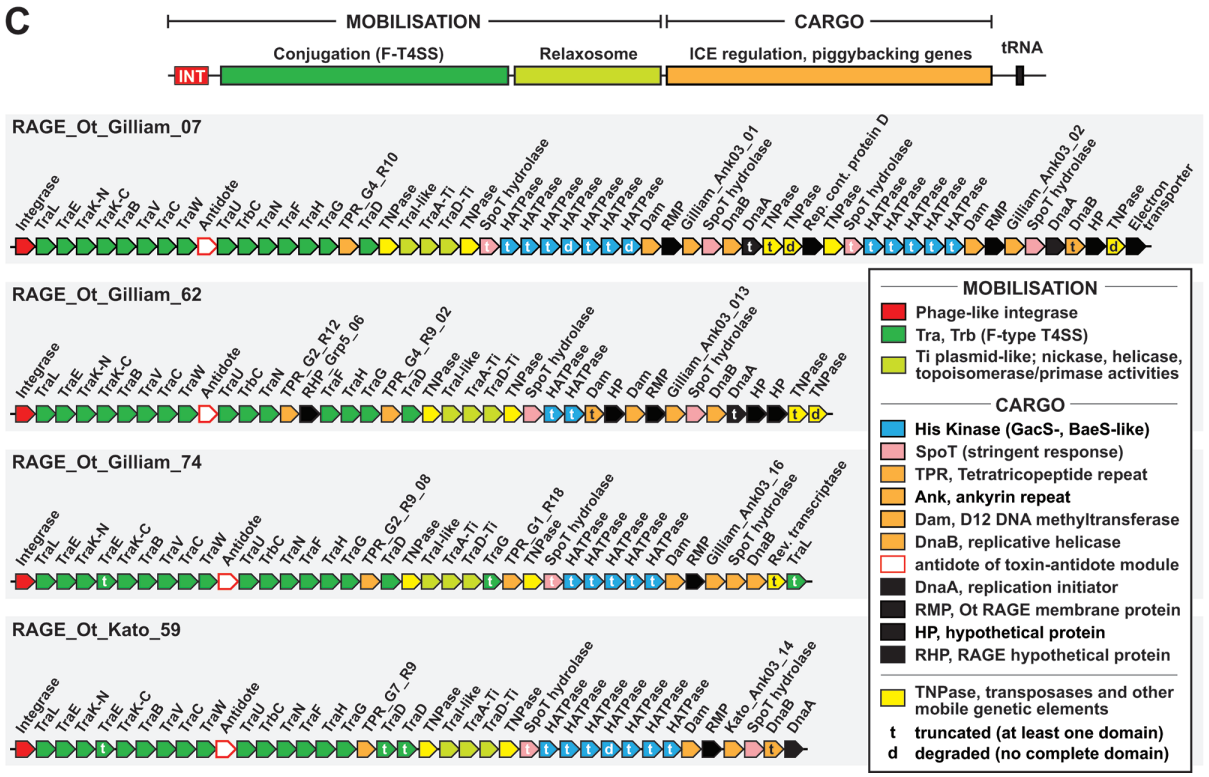


FIG 1 Ot RAGE and IR elements. (A) An overview of the genomes of eight Ot strains with genes classified into RAGE and IR regions. Numbers at the left refer to Ot strains listed in panel B. Gray arrows indicate RAGE regions; colored arrows denote IR regions. The colors correspond to conserved IR regions between strains and demonstrate the lack of synteny between Ot genomes. (B) Table summarizing RAGEs, IR regions and isolated mobile genes, cargo genes, and hypothetical genes that could not be classified into RAGE or IR elements. Ot strains are listed accordingly to a previously estimated phylogeny (11), with numbers corresponding to full genome maps in panel A. (C) Organization of genes in the four complete RAGEs found in our analysis. Detailed analysis of the re-annotation and classification of all genes in the eight genomes are given in data set S1. d, degraded (no identifiable domains present); t, truncated (at least one identifiable domain present).

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

predicted pseudogenes. A previous study found that the Ot strain Ikeda genome lacked any complete RAGES (9). We assessed whether any of the strains in our analysis encoded complete RAGEs, defined as containing a full set of mobilization genes and additional cargo genes as outlined in previous analyses (9). All the strains in our analysis, with the notable single exception of Ikeda, encoded one or more complete sets of RAGE genes (Fig. 1B). However, most of those RAGEs contained one or more mobilization genes that were truncated. Accordingly, we carried out sequence alignments to define each RAGE gene as being full length, truncated (containing one or more identifiable domains) or degraded (containing no identifiable complete domains). We then assessed whether any strains contained complete RAGEs with intact, full-length genes (Fig. 1B). We found that two strains, Gilliam and Kato, encoded complete RAGEs with full-length mobilization genes (Fig. 1C). This suggests that these strains may have obtained these elements recently and that they may be capable of continued mobilization. ICEs normally have a preferred integration site, often within tRNA genes (12), as observed for *Rickettsia* species (18). Despite our discovery of complete RAGEs in these Ot genomes, no identifiable integration sites could be determined.

We previously used RNA sequencing analysis and comparative genomics to show that, despite the lack of synteny between Ot strains driven by the prolific RAGEs, small groups of proximate genes were transcribed at similar levels and maintained synteny across strains (19). This demonstrates selection for gene order at the local level despite it being absent at a global level across Ot genomes. To further identify gene groups evolving under strong selective constraints relative to superfluous RAGEs, we analyzed the IR regions, which harbor most core Ot genes (Fig. 1A). This revealed 84–103 IR regions, ranging in length from 2 to 27 genes (Fig. 1B). The difference in number of IRs between genomes is accounted for by some IR gene blocks being split up in some genomes and not others, rather than the presence of novel genes in some genomes. Identification of these conserved IR gene groups illuminates highly conserved microsynteny that may encompass functionally linked genes sharing expression and/or regulatory programs.

Single-copy cargo genes

The delineation of Ot genomes into RAGE and IR regions enabled us to better characterize RAGE cargo genes (Fig. 2A and B). In addition to the group of highly replicated multicopy cargo genes already described as RAGE components (discussed below), we identified numerous single-copy genes previously overlooked for their occurrence within RAGEs (Fig. 2A). These include genes involved in fundamental processes of bacterial physiology and metabolism, e.g., tyrosine tRNA ligase, RNA polymerase subunit omega, and the ClpP protease, as well as genes encoding predicted secretory effectors likely involved in interactions with host cells, including phospholipase D and autotransporter proteins ScaA and ScaC in all genomes, and ScaB (Boryong), ScaF (TA686) and ScaG (TA686) in individual strains. As many of these single-copy genes have orthologs in other bacterial species that lack the RAGEs, it is likely that they were not introduced by mobile genetic elements. Rather, their current presence within RAGEs indicates they were probably incorporated into RAGE via recombination. However, a case-by-case analysis may reveal certain conserved genes shuttling between Ot genomes via RAGE mobilization. For instance, despite their conservation in all *Rickettsia* genomes, genes encoding secreted effectors and metabolite transporters were previously found piggybacking on RAGEs in the *R. buchneri* genome, illustrating the ability of RAGE to shuttle rickettsial genes important for the obligate intracellular lifestyle (12).

Highly abundant multi-copy cargo genes

Analysis of RAGE-associated cargo genes revealed 16 genes or gene groups present in numerous copies in all eight Ot genomes (Fig. 2B). Gene groups included those encoding membrane proteins, Dam DNA methyltransferases, DNA helicases, multi-drug

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

A

○ single copy in RAGE ● single copy in RAGE and inter-RAGE ○ Inter-RAGE
● multi-copy in RAGE ● multi-copy in RAGE and inter-RAGE x absent

RAGE-associated genes	Description	Gilliam	Boryong	UT76	UT176	Karp	Kato	Ikeda	TA686
18 kDa heat shock protein	ACD_sHsps-like: alpha-crystallin-type small heat shock protein	x	x	x	x		x	x	x
MidA family	SAM-dependent methyltransferase	2	2	2	2	2	2	2	2
TimA	Tim44/TimA family of putative adaptor protein (transport)	1	1	1	1	1	1	1	1
glpE	Thiosulfate sulfurtransferase GlpE	x	x	x	x		x	x	x
cspC	Cold shock protein, CspA family								
gcvT	Aminomethyltransferase (glycine degradation)								
MdlB	ABC transporter ATP-binding protein (intracellular bacteria hallmark)	2	2	2,2	2	2	2,2	2	2
lipB	Octanoyltransferase (octanoyl-ACP to lipoate-dependent enzymes)			2	x				
DDE_Tnp_1_3	Transposase DDE domain or inactive derivative	x			x				5
uvrA	Excinuclease ABC subunit A								
pgsA	CDP-diacylglycerol-glycerol-3-P 3-phosphatidyltransferase								
tyrS	Tyrosine--tRNA ligase								
lacA	Ribose-5-phosphate isomerase								
dapD	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase								
ScaA	ScaA autotransporter protein								
MdtL	Multidrug resistance protein MdtL	x	x	x	x		x	x	x
pld	Phospholipase D family protein								
uvrD	DNA helicase II	1	1	1	1	1	1	1	1
rpoZ	DNA-directed RNA polymerase subunit omega								
bamB	Outer membrane protein assembly factor BamB								
ssb	Single-stranded DNA-binding protein								
PPDK	Pyruvate, phosphate dikinase								
map	Type I methionyl aminopeptidase								
ctaA	Heme A synthase (Cytochrome oxidase assembly protein)								
clpP	ATP-dependent Clp protease proteolytic subunit								
HipB	Transcriptional regulator, contains XRE-family HTH domain	2,3	4	4	4	4	4	4	4
ScaC	ScaC autotransporter (T5SS)							x	
bcr/cflA	Bcr/CflA family drug resistance efflux transporter								2
Sco1	Cytochrome oxidase Cu insertion factor								
rne/rng	Ribonuclease E/G								
trpS	Tryptophan--tRNA ligase								
SLC5-6-like_sbd (NT): HK (CT)	Sodium solute symporter fused to sensor HK/response regulator								
rppH	RNA pyrophosphohydrolase								
ndk	Nucleoside-diphosphate kinase								
cyaY	Iron donor protein CyaY								
Dam	Site-specific DNA-adenine methylase			x	x	x	x	x	x
rSAM_XyeB	Radical SAM/SPASM domain peptide maturase, XyeB family			x	x	x	x	x	x
YifB	Predicted ATPase with chaperone activity	3	3	4	4	4	4	4	3
prmA	Ribosomal protein L11 methyltransferase	x	x		x	x		x	
Abhydrolase_6	Alpha/beta hydrolase family (MhpC-like)	x	x	x	x	x			x
prfB|prfB	Peptide chain release factor RF-2	x	x	x	x	x		x	x
gpgP	Glucosyl-3-phosphoglycerate phosphatase	x	x	x	x	x		x	x
HPD	4-hydroxyphenylpyruvate dioxygenase (tyrosine degradation)	x	x	x	x	x	x	x	
sfp	4'-phosphopantetheinyl transferase sfp	x	x	x	x	x	x	x	
glyA	Serine hydroxymethyltransferase	1	1	1	1	1	1	1	
acoA	Acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit alpha	x	x	x	x	x	x	x	
HNHc	Endonuclease signature (viral, prokaryotic, and eukaryotic proteins)	x	x	x	x	x	x	x	
fol_rel_CADD	Putative folate metabolism protein, CADD family	x	x	x	x	x	x	x	
xseB	Exodeoxyribonuclease VII small subunit	1	1	1	1	1	1	1	1
ProP	Proline/glycine betaine transporter	6	6	6	6	6	6	6	5
ltrA	Group II intron-encoded protein LtrA	x	x	x	x	x	x	x	
lysS	lysine--tRNA ligase	1	1	1	1	1	1	1	1
FolP	Dihydropteroate synthase	x	x	x	x	x	x	x	
aroK	Shikimate kinase 1	x	x	x	x	x	x	x	
menG	2-phytyl-1,4-naphtoquinone methyltransferase	x	x	x	x	x	x	x	
dnaQ	DNA polymerase III subunit epsilon	1	1	1	1	1	1	1	1
gpmA	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	x	x	x		x	x	x	
pthXo1	pthXo1 CDS, TAL effector protein PthXo1	x	x	x		x	x	x	x
secA-like	preprotein translocase SecA subunit-like protein	5		3		8	3	3	8
SenC	Required for optimal cytochrome c oxidase activity	x	x	x		x	x	x	x
mta	HTH-type transcriptional activator mta	x	x	x		x	x	x	x
prfB	peptide chain release factor RF-2	x	x	x	x	x			x

B

MOBILISATION — CARGO

		Gilliam	Boryong	UT76	UT176	Karp	Kato	Ikeda	TA686
MOBILISATION	Phage-like integrases (RAGE insertion)	95	77	99	58	89	102	83	98
	Conjugation and relaxosome proteins (*tra* genes)	647	664	607	421	686	621	512	572
CARGO (high copy) — highly abundant	Membrane proteins unique to RAGE	32	56	24	22	41	30	26	27
	Dam, D12 DNA methyltransferases	33	34	29	18	30	26	18	22
	DNA replicative helicases	40	54	46	24	52	50	38	53
	Multidrug Resistance Proteins/Histidine Kinases	103 (5)	57 (10)	103 (7)	50 (7)	119 (7)	95 (7)	82 (8)	129 (7)
	SpoT stringent response regulators	64	54	64	36	78	62	62	49
	Hypothetical Protein-encoding genes	462	547	419	324	297	464	417	473
CARGO (high copy) — put. effectors	Tetratricopeptide epeat containing proteins	48	29	34	31	39	45	21	31
	Ankyrin repeat containing proteins	62	66	63	53	75	53	47	56
CARGO (high copy) — less abundant	Peroxiredoxin	2	2	2	3	5	7	4	5
	XthA2 (exodeoxyribonuclease III)	2	3 (1)	4 (1)	2 (1)	4 (1)	2 (1)	4 (1)	2
	Repeat-containing protein D	5 (1)	9	3	4 (1)	3 (1)	2 (1)	6 (1)	2 (1)
Mobile Genetic Elements	Transposase and inactive derivatives (numerous types)	634	373	275	531	361	279	350	480
	Reverse transcriptase	23	64	12	7	20	38	36	38
	HNH endonucleases	1	2	2	0	22	9	1	18

FIG 2 Single and multi-copy cargo genes encoded on Ot RAGEs. (A) Single or low-copy cargo genes encoded on Ot RAGE. Summary statistics show whether genes are present in single or multiple copies on RAGEs in different strains, and also in single or multiple copies in IRs. The exact number of copies is given for each gene. Blue text indicates the number of copies in IR; red text denotes the number of copies in RAGE. (B) Frequency and distribution of high copy cargo genes (both full length and truncated/degraded) within RAGEs in eight strains of Ot. Numbers in brackets denote additional copies in IRs.

resistance proteins (MRPs) and histidine kinases, SpoT hydrolase and synthetases, hypothetical/uncharacterized proteins, mobile genetic elements (i.e., insertion sequences (ISs), transposases, integrases, and reverse transcriptases), Anks, TPRs, and Vir- and

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

Tra-type T4SS proteins. All of these, except *vir*-type T4SS genes, have been identified as RAGE cargo genes in previous studies (8, 9, 18, 20). For each gene group, we assessed (i) whether all the genes annotated as belonging to this category were paralogs of the same gene or whether multiple distinct genes were present within one group, and (ii) whether some or all genes within a group were truncated and not able to form a full-length protein and, where a functional domain was known, whether this domain was present or not. Genes involved in DNA mobilization, effector proteins, and T4SS genes are discussed in dedicated sections below, while other multi-copy cargo genes are discussed here.

Membrane proteins

The eight Ot genomes encode 21–41 RAGE-associated genes annotated as membrane proteins (Fig. 3A, data set S2). Analyses revealed that each Ot strain encodes exactly one copy of three genes encoding proteins with analogy to characterized membrane proteins: the YccA modulator of protease FtsH, vitamin transporter Vut1, and a gene similar to the rhamnose transporter RhaT. The remaining 18–38 genes encode paralogs of a gene we call Ot_RAGE_membrane protein, ranging in length from 90 to 663 bp. This protein lacks homology to any non-Ot genes, and no known domain could be identified. Thus, the function of this gene in Ot is unknown.

DNA methyltransferases

Ot genomes encode 18–34 genes with similarity to *dam*, all of which are located within RAGEs (Fig. 3B). Sequence alignments demonstrate that 8–26 encode full-length proteins, defined as being equal in length to *Escherichia coli* (ECO) Dam and encoding all seven known domains. The Ot genomes encode an additional 2–26 truncated *dam* genes where some domains are preserved and fewer degraded copies with no identifiable domains. It is not known if these genes are functional, although previous studies (19, 21) showed that they were not detected by proteomics analysis. They may have a specific role in methylation of RAGE during mobilization and/or integration to protect from deleterious effects of single-stranded DNase activity.

DNA helicases

DNA helicases unwind double-stranded DNA and function in DNA and RNA metabolism, with general roles in DNA replication, repair, and recombination. We found that all strains of Ot encode exactly two full-length copies of the DNA helicase *uvrD*, which is involved in DNA repair (Fig. 3C). One copy is located within a RAGE, while the second is located within IR82. The *dnaB* family of helicases, by contrast, is more numerous and degraded in Ot genomes (Fig. 3C). This helicase is involved in DNA replication and is present in 36–52 copies, with 1–23 being full length and mostly located within RAGEs. Each genome encodes one full-length copy located at the interface of IR46 and IR47, which is likely the ancestral non-RAGE paralog involved in genome replication. Other copies are undoubtedly associated with RAGE mobilization. Our previous proteomics analysis (19, 21) detected expression of a *dnaB* gene, but due to sequence similarities between numerous paralogs, it was not possible to determine which specific gene was expressed. Given its role in DNA replication, however, it is expected that at least one paralog would be expressed and functional.

MRPs and histidine kinases

We identified around 100 genes in the Ot genomes that were annotated as MRPs or histidine kinases (Fig. 2B and 3D; Fig. S1; data set S3) or were found to contain histidine kinase domains (e.g., the sodium/proline symporter *putP*). MRPs are members of the ATP-binding cassette (ABC) transporter protein family and were annotated as MRPs based on the presence of a histidine kinase ATPase domain (HATPase). Our analysis determined that genes encoding two "MRP" proteins were distinct from all the other

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

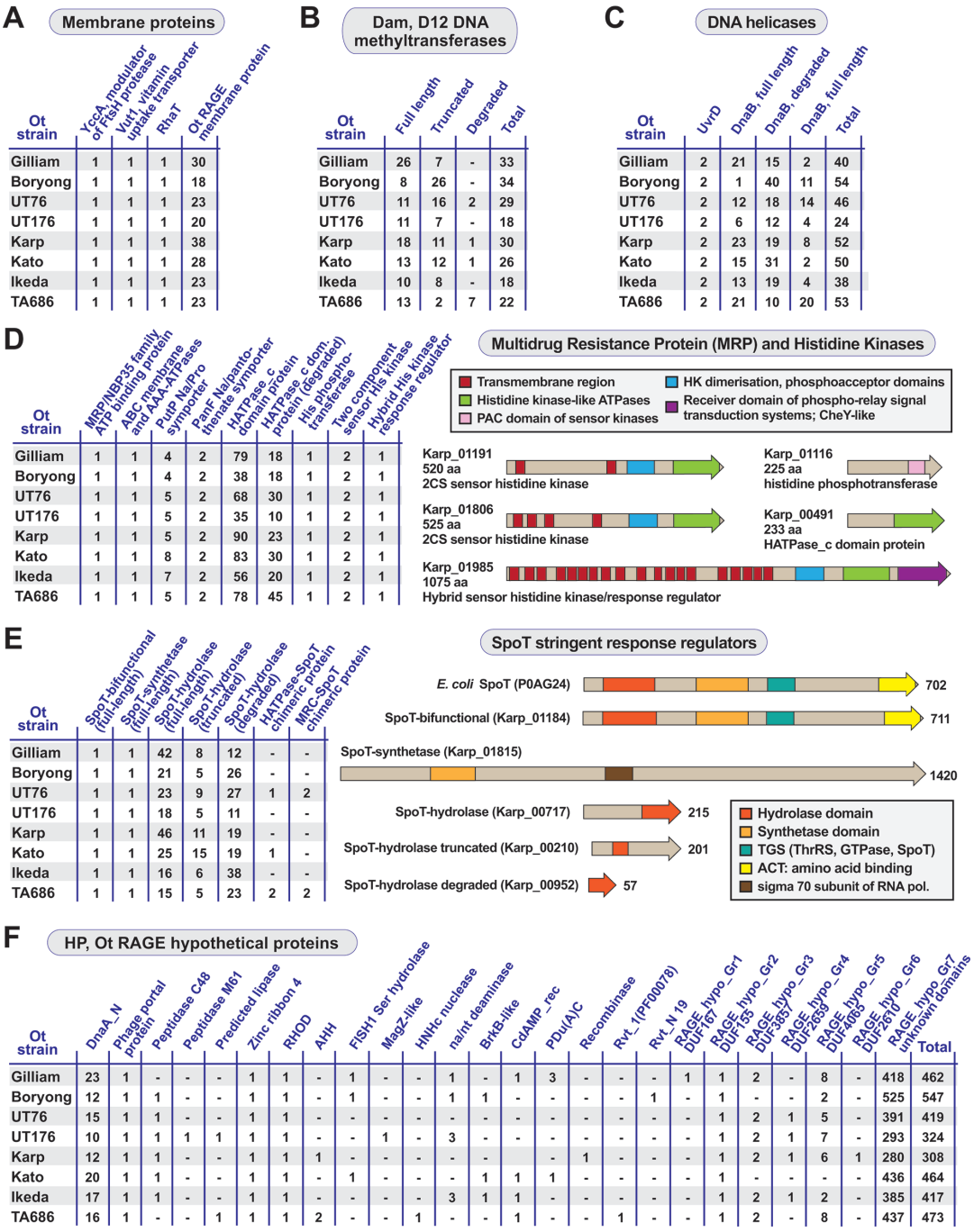


FIG 3 Analysis of high copy cargo genes on RAGE elements in Ot. (A–C) Frequency and distribution of RAGE cargo genes annotated as (A) membrane proteins, (B) Dam DNA methyltransferases, and (C) DNA helicases. DnaB is a replicative DNA helicase, and UvrB is a repair DNA helicase. (D) Frequency and distribution of RAGE cargo genes encoding MRP/histidine kinases, with examples of His kinase divergent architectures. (E) Frequency and distribution of RAGE cargo genes encoding SpoT stringent response regulators, with examples of divergent architectures. The bifunctional SpoT protein is compared to the canonical SpoT protein of *Escherichia coli*. (F) Frequency and distribution of RAGE cargo genes encoding HPs. AHH, adenosyl homocysteine hydrolase; BrkB-like, YihY/virulence factor BrkB family protein; CdAMP_rec, cyclic diAMP receptor proteins; DUF, domain of unknown function; DnaA_N, N-terminal domain of DnaA; MagZ, nucleoside triphosphate pyrophospho-hydrolase; na/nt, nucleic acid/nucleotide deaminase; PDu(A)C, copper chaperone; RHOD, rhodanese homology domain; Rvt_1 (PF00078), reverse transcriptase Pfam PF00078; Rvt_N 19, domain of reverse transcriptase Rvt_N.

HATPase domain containing proteins in Ot: an MRP/NBP35 family ATP-binding protein and an ABC-membrane and AAA ATPase protein, both single copy in each genome. The former was located within an IR region (IR9) and the latter within RAGEs. Analysis of the

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

remaining genes annotated as MRP/histidine kinases led us to identify several full-length orthologs of two-component system (2CS) histidine kinase genes. These include one histidine phosphotransferase gene (which does not contain an HATPase domain), one large hybrid sensor histidine kinase/response regulator gene, and two 2CS histidine kinase genes (Fig. 3D). All were present in the same copy number and locations in each genome. Histidine phosphotransferase and the two 2CS sensor histidine kinase genes were consistently found in IR regions IR43, IR47, and IR63, respectively, while the hybrid sensor histidine kinase/response regulator gene was located within a RAGE. In our UT76 proteomics data set (21), both the histidine phosphotransferase and the hybrid sensor histidine kinase/response regulator genes were detected, while the two sensor histidine kinase proteins were not (data set S1). We also identified two copies of a histidine kinase domain containing sodium/pantothenate symporter, *panF*, present in IR11 and IR55 regions in each genome, as well as a sodium/proline symporter, *putP*, present in four to eight copies and distributed into both RAGEs and IRs. Analysis of our previous proteomics data set showed that *panF* was expressed in strain UT76 as was one copy of *putP* located in IR49 (21) (data set S1). The remaining MRP/histidine kinase genes were paralogs of one another, containing an HATPase domain and being present in 45–113 copies per genome with various degrees of truncation (Fig. S1). We classified these as degraded HATPase domain containing proteins when an intact HATPase domain could no longer be detected due to the short length of the gene (Fig. 3D; Fig. S1; data sets S1 and S3).

SpoT stringent response regulators

SpoT is a bifunctional synthetase/hydrolase that is essential for inducing and regulating the stringent response in *E. coli* and other bacteria through mediating inter-cellular levels of (p)ppGpp, also called alarmone (22, 23). Ot genomes encode 36–78 genes with homology to *spoT*. We identified exactly one full-length *spoT* gene, present in an IR region (IR46), encoding both synthetase and hydrolase domains in all Ot genomes (Fig. 3E). This gene was the only *spoT* homolog found to be expressed in our previous proteomics analysis and was shown to be upregulated in extracellular Ot, consistent with a role in transitioning between different bacterial states (21). We also identified exactly one gene in each genome that encodes only the *spoT* synthetase domain in addition to a long C-terminal domain of unknown function. We then identified 15–46 *spoT* genes that lacked the synthetase domain yet encoded the intact hydrolase domain as well as a further 16–44 *spoT* genes that were truncated or degraded such that a functional hydrolase domain was no longer present. Finally, we identified one to four genes in some genomes in which hydrolase domains are fused to other genes. Most rickettsial genomes harbor 6–12 *spoT* genes, with some of the abovementioned architectures present (data not shown). Curiously, bifunctional (complete hydrolase and synthetase domains) genes are typical of most other Rickettsiales species, though not common in *Rickettsia* species and absent from notable human pathogens (e.g., *Rickettsia prowazekii*, *Rickettsia typhi*, *Rickettsia rickettsii*, and *Rickettsia conorii*) (12). Still, the tendency for all Rickettsiales genomes to retain numerous single-domain *spoT* genes, even when RAGEs are absent, implies their function in some aspect of the stringent response. The presence of such drastic numbers and diverse architectures of *spoT* genes in Ot genomes relative to other rickettsial species is intriguing and deserving of future investigation.

HPs

The Ot strains harbor 308–547 genes per genome that are annotated as hypothetical or uncharacterized, of which about half are located within RAGEs (Fig. 2B and 3F; data set S4). We determined whether all the RAGE hypothetical/uncharacterized genes were paralogs of a single RAGE gene or if they encoded multiple different genes. Sequence alignments for all the genes annotated as hypothetical or uncharacterized in the Karp genome were performed, which revealed that the genes clustered into 24 groups (Fig. 3F) with 18 of these encoding genes carrying known protein domains. Those without

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

known domains were named RAGE_hypo_Gr1-7, with groups I–VI encoding a domain of unknown function and group VII combining a mixture of all remaining HP genes with no identifiable known domains. Three of these genes with known domains were present in exactly one copy in all genomes and encode a phage portal protein, a zinc ribbon domain, and a rhodanese homology domain. Another single-copy gene found in all eight Ot genomes carries a domain of unknown function (DUF155). Furthermore, hypothetical genes containing a *dnaA* N-terminal domain were identified in 10–23 copies in all Ot genomes. In our prior study, one full-length paralog, located in IR1 was expressed in UT76, while others were not detected (21) (data set S1).

Other hypothetical/uncharacterized genes were distributed sporadically among the genomes. In order to get a sense of the distribution of the remaining RAGE-associated hypothetical genes that were not clustered into conserved groups, we analyzed the remaining hypothetical genes in the RAGEs of the Karp genome only. There were no identifiable domains in any of these and the diversity was such that it was not possible to bin them into homologous groups. We annotated them all as belonging to a large and divergent 25th group (RAGE_hypo_Gr7). While little can be inferred about the function of these hundreds of genes, it is likely that at least some of these play important roles in the biology of Ot.

Putative effectors piggybacking on Ot RAGE

Anks

The ankyrin repeat (AR) is one of the most common protein folds in nature, being widespread in eukaryotes and pervasive in many viruses and host-associated bacteria (24, 25). Ankyrin repeats are used to mediate a myriad of protein-protein interactions, and host-associated prokaryotes and viruses frequently express Anks to hijack or subvert host cell pathways that would be detrimental or beneficial to their survival (26). Previously, several Ot Anks were shown to be secreted via the rickettsial type I secretion system (27). Certain Ank effectors have been functionally characterized in strain Ikeda and shown to play important roles in host cell interactions (27–31). However, a major challenge in comparing the host-pathogen cell biology of different Ot strains has been the difficulty assessing which Anks are most similar to those in other strains. This is important for determining the significance of Anks as species-specific versus strain-specific effectors underlying pathogenesis.

We defined a set of criteria for clustering Anks, with their subsequent characterization within each Ot genome following the well-described Ank repertoire of strain Ikeda (9). We identified several new Ank groups in Ikeda, although some of these lack complete Ank repeats and are likely non-functional (data set S5). Our comparative analysis indicates Ot strains encode 47–66 Anks, with variability (67%–94%) in the number of common versus strain-specific genes per genome (Fig. 4A). Ot Anks often harbor a single F-box domain, which are prominently known components of SCF (Skp1, Cullin1, and F-box) ubiquitin ligase complexes, but recent studies have described their participation in non-SCF protein-protein interactions involved in diverse eukaryotic pathways (32). F-box-resembling pox protein repeats of ankyrin, C-terminal (PRANC) domains and coiled coils were less frequently predicted.

Of the 54 orthologous groups of Ot Anks, all genomes were found to encode at least one copy of seven groups: *ank03, ank08, ank10, ank11, ank12, ank20*, and *ank24* (Fig. 4B). *ank03* is by far the most prominent Ank, being present in 4–32 copies in Ot genomes, with the other six families present in 1–4 copies (Fig. 4B). Curiously, while most Ot Anks are predominantly found within RAGEs, *ank20* is encoded in an IR region (IR84) in all analyzed Ot genomes. Collectively, these seven Anks likely carry out essential functions in Ot biology. However, each strain likely utilizes unique Ank arsenals throughout its life cycle, given that some of the less conserved Anks have characterized functional roles; e.g., Ank01 and Ank06 of Ot str. Ikeda modulate NF-kB transport to the nucleus (30). As such, it is likely that there is functional redundancy between the Ank groups, with some

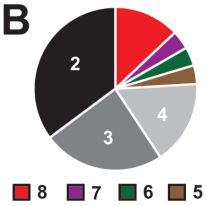
Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

A

Ankyrin repeat containing proteins (*n* = 129)

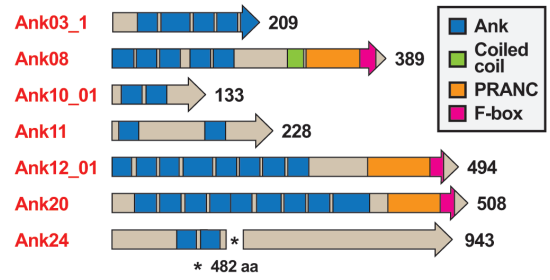
Ot strain	orthologous groups (*n* = 54): No. orthologous groups (OGs)	Tot. proteins (% in OGs)	CC, PRANC, F-box	CC or PRANC, F-box	F-box only	w/o F-box	singletons (*n* = 75): No. singletons	Tot. proteins	CC, PRANC, F-box	CC or PRANC, F-box	F-box only	w/o F-box
Gilliam	25	48, 79%	5	5	24	14	12	15	-	2	5	8
Boryong	22	36, 67%	4	3	7	22	21	28	3	2	2	21
UT76	36	53, 84%	6	8	20	19	8	11	2	2	5	2
UT176	29	51, 94%	2	9	22	18	3	3	-	-	-	3
Karp	25	66, 88%	3	12	35	16	7	9	-	1	2	6
Kato	26	51, 93%	6	5	24	16	4	4	1	-	-	3
Ikeda	22	49, 88%	10	10	20	9	7	7	-	-	-	7
TA686	20	40, 70%	5	5	23	7	14	17	4	2	4	7

B



Ank copy no. range / avg. per genome

Ank03: 4-32 / 17	Ank_06: 1-5 / 1
Ank08: 1-3 / 1	Ank_09: 1-1 / 1
Ank10: 1-4 / 2	Ank_25: 1-1 / 1
Ank11: 1-1 / 1	Ank_38: 1-8 / 3
Ank12: 1-2 / 2	
Ank20: 1-1 / 1	Ank_02: 1-1 / 1
Ank24: 1-1 / 1	Ank_07: 1-4 / 2



C

TPR repeat containing proteins

Ot strain	TPR Group 1	2	3	4	5	6	7	8	9	Total proteins
Gilliam	14	17	2	6	-	1	3	5	-	48
Boryong	7	12	6	1	-	1	1	1	-	29
UT76	7	15	2	3	2	1	-	2	2	34
UT176	9	15	4	1	-	1	-	1	-	31
Karp	16	15	3	4	-	1	-	-	-	39
Kato	11	18	2	5	2	1	1	3	2	45
Ikeda	5	5	3	2	2	-	-	2	2	21
TA686	11	15	-	2	-	1	-	2	-	31
Repeat range	1-18	1-16	1-9	3-11	2	5-11	4-9	1-10	1-2	

D

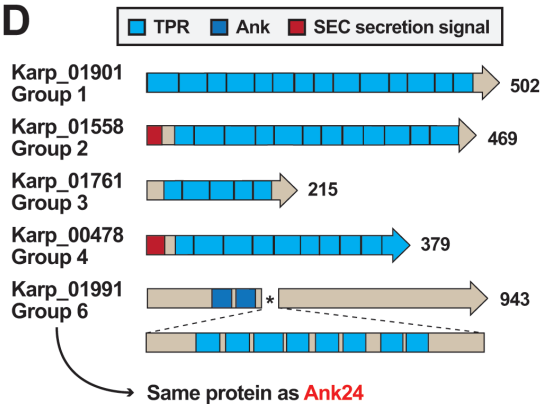


FIG 4 Putative effectors piggybacking on Ot RAGE. (A) Frequency and distribution of Anks in Ot genomes. Anks are broken down into orthologous groups (OGs, present in two or more genomes) or singletons (unique to a genome). CC, coiled coil; PRANC domain found at the C terminus of certain pox virus proteins; F-box, motif of approximately 50 amino acids that functions in protein-protein interactions. (B) (Top) Graphical view of Ank OG strain representation (two to eight genomes). Roughly 25% of Ank OGs are found in five or more strains, with variable levels of conservation in copy number per genome. (Bottom) Architectures for Anks present in all Ot genomes, with proteins from Ot strain Karp. (C) Frequency and distribution of TPRs in Ot genomes. Nine ortholog groups contain all the TPRs across eight Ot genomes. (D) Examples of diverse TPR architectures for six proteins from Ot strain Karp. PRANC, pox protein repeats of ankyrin, C-terminal.

of the 100 other Ank groups not found in Ikeda functioning similarly to *ank01* and *ank06* in genomes lacking these genes.

TPRs

The tetratricopeptide repeat is another protein motif that is commonly used in mediating inter-protein interactions, typically found in subunits of multi-protein complexes (33). TPR repeat containing proteins are widespread in Ot (Fig. 4C), albeit with a lower number of copies per genome than Anks. Ot TPRs have been less characterized than the Ot Anks, with only one report demonstrating a role in inhibition of eukaryotic translation in Ot strain Boryong (34). We compiled the 21–48 TPRs per Ot genome and classified them into nine groups primarily based on within-protein location of tetratricopeptide repeats (Fig. 4C; Fig. S2; data set S5). While the positions were conserved within groups, the number of repeats was variable and indicated expansion and contraction of repeats, as well as processive gene degradation within each group (Fig. 4D). The prediction of SEC signal peptides in certain TPRs indicates at least some of these putative effectors may be secreted to the periplasm with possible translocation across the outer membrane, possibly via TolC as proposed for the RARP-1 effector of *R. typhi*. (35) Still, the

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

lack of N-terminal secretion signals in most TPRs indicates other possible routes for TPR secretion that await characterization.

Mobile genetic elements associated with RAGE

Integrases

ICEs, such as Ot RAGE, encode integrase genes to catalyze genomic integration and conjugative genes (discussed in the next section) to catalyze horizontal gene transfer (36). The Ot genomes analyzed in this study encode 58–102 integrase genes (Fig. 5A), of which only 3–13 per genome remain full length, consistent with progressive degradation of the Ot RAGE. Where present, the integrases are located at the start position of a RAGE (Fig. 1C). However, several integrase genes were located as isolated genes outside RAGE regions, reflecting the high mobility of these genes and the overall high recombination rates in Ot genomes.

Transposable elements

In addition to the OtRAGE, the Ot genome encodes two other types of transposable elements (9): "copy and paste" retrotransposons (group II introns) and "cut and paste" DNA transposons. While these are independent mobile genetic elements, they have been incorporated into the Ot RAGE regions, and it is likely that the different mobilizable elements impact each other's activity. Group II introns are self-splicing retrotransposons that catalyze their integration into genomes via an RNA intermediate using an intron-encoded reverse transcriptase protein (38). The Ot genomes encode 7–64 group II intron reverse transcriptase genes, although only UT76, Karp, and TA686 encode full-length genes, with the others being heavily degraded (Fig. 5A). All full-length reverse transcriptase genes were immediately followed by an HNH endonuclease gene likely required for catalysis.

Ot encodes several families of DNA transposons which have been previously classified in strain Ikeda (9). This class of mobile elements is composed of a transposase gene flanked by inverted repeat (IR) regions on either side, which together make up an insertion sequence (IS) (39). Numerous families of IS have been identified in other bacteria. Given the large number of IS elements in each Ot genome and their highly degrative tendency, we selected two of the many frequently occurring IS genes, ISOt3 and ISOt5, to characterize in detail across all eight genomes (Fig. 5A). These were present in 0–120 (ISOt3) and 0–70 (ISOt5) full-length copies across the genomes. We also analyzed the complete set of IS elements in one strain, Karp, and compared these with those previously predicted for str. Ikeda (Fig. 5B). An example of the analysis of one IS element in Karp, ISOt1, shows the distribution of full-length and degraded copies typical of IS families in all Ot genomes (Fig. 5C). We identified the same set of IS elements that had previously been described in Ikeda (9) (Fig. 5B and C). We followed the classification and nomenclature established in Nakayama et al. (9) in which mISOt1, mISOt2 and mISOt4 denotes "miniature" versions of elements containing the same terminal inverted repeat sequences as ISOt1, ISOt2, and ISOt4, respectively. Within the nine IS classes found in Karp, most were heavily degraded with some, such as IS630 family transposase ISOt3, having no remaining full-length elements. We identified seven additional groups of transposase genes in Karp that were not part of IS elements (Fig. 5B).

Gene transfer agents

In exploring the different machineries that might carry out lateral gene transfer in Ot, we searched for evidence of genes that might have originated from phages. We identified phage-like genes such as that encoding capsid protein in all the Ot genomes in our analysis. Upon closer inspection, these were found to closely resemble the recently described gene transfer agents (GTAs) of *Wolbachia* (40). This is a DNA element that includes several virus-like genes but is distinct from bacteriophages. First described and best characterized in *Rhodobacter capsulatus*, this genetic element packages genomic

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

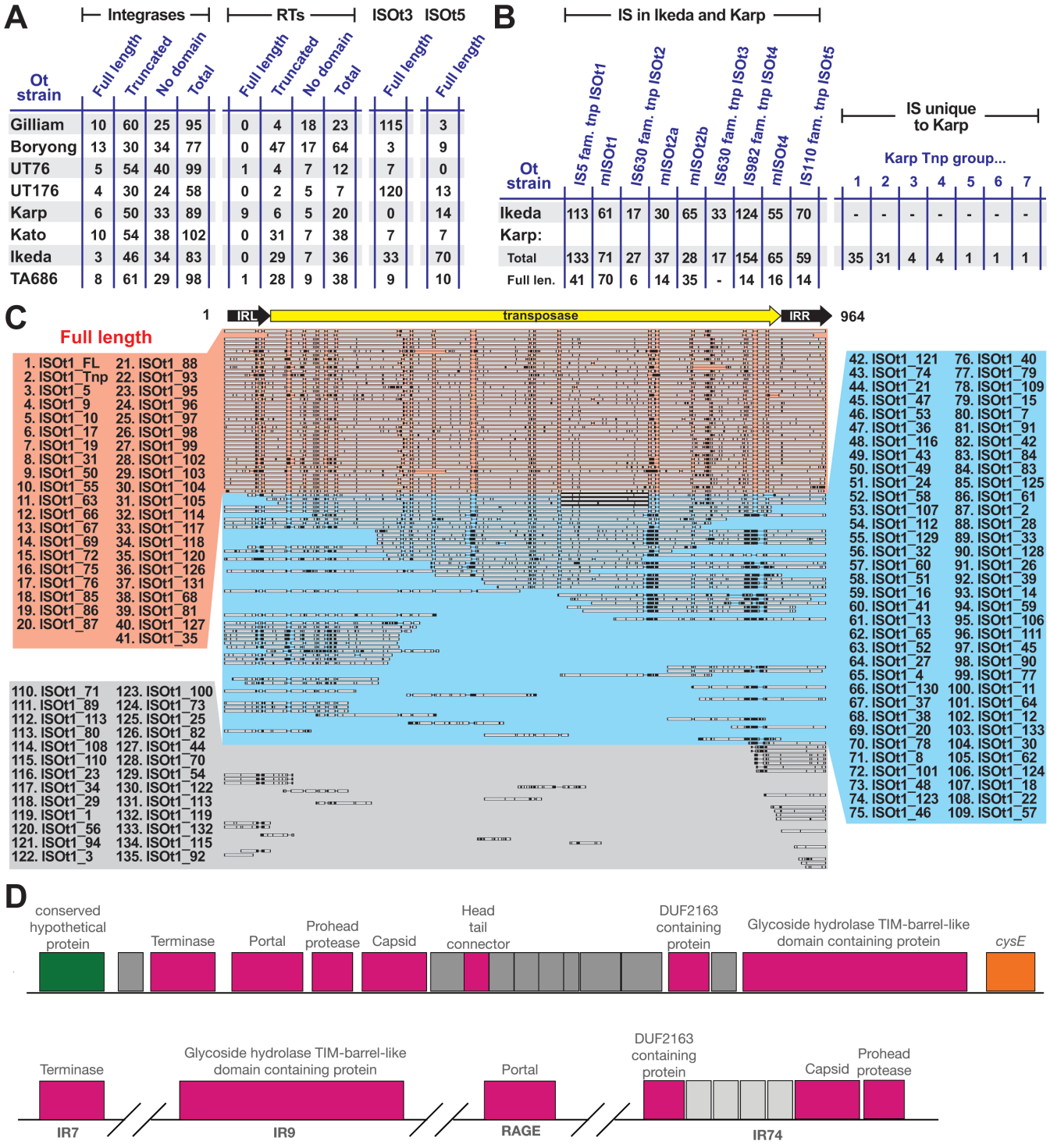


FIG 5 A diversity of mobile genetic elements associates with Ot RAGEs. (A) Frequency and distribution of Ot RAGE-associated genes encoding integrases, group II intron-associated reverse transcriptases, and IS elements ISOt3 and ISOt5. (B) Frequency and distribution of IS elements in Ikeda and Karp strains. (C) Alignment showing classification of ISOt1 elements as full length or degraded. Full-length copies of ISOt1 in Karp are shown by the red dotted box. (D) Overview of GTAs in Ot compared with the canonical GTA cluster in *Rhodobacter capsulatus*. GTA genes conserved between Ot and *R. capsulatus* are shown in magenta. GTA genes in *R. capsulatus* that are not found in Ot are shown in dark gray. Ot genes positioned between GTA genes are shown in light gray. Figure adapted from Lang and Beatty (37). GTA, gene transfer agent.

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

DNA rather than viral DNA and does not cause lysis in recipient cells. GTAs were previously shown to be widespread in alphaproteobacteria, including Rickettsiales (37). We found that all the Ot genomes carry six GTA genes encoding terminase, glycoside hydrolase TIM-barrel-like domain containing protein, portal, DUF2163 containing protein, capsid, and the prohead protease (Fig. 5D; data set S7). All except the portal protein have conserved positions in IRs 7, 9, and 74 in all genomes, while portal is always encoded on a RAGE. The conserved nature of this set of genes in all Ot genomes is highly suggestive of an ongoing role for GTAs, although its activity and potential role in lateral gene transfer remain to be characterized experimentally.

Type IV secretion systems

F-T4SS

The RAGE encodes a conjugative T4SS highly similar to the F-T4SS of the archetypal F plasmid of *E. coli* (*tra/trb*) (8, 9). Previous comparisons of the RAGE T4SS with that of the *E. coli* F plasmid showed that it encodes 14 proteins predicted to form the T4SS scaffold, some of which are analogous to components within P-type T4SSs (18, 41) (Fig. 6A and B; data set S8). While syntenic to the *E. coli tra*/*trb* T4SS, the RAGE T4SS lacks genes involved in the regulation of conjugation, as well as other assembly factors and lytic transglycosylases (Fig. 6C). In this way, the RAGE T4SS is a streamlined version of the canonical F-T4SS. The Ot RAGE T4SS is also highly similar in gene order and composition to F-T4SSs characterized in the RAGEs of *R. buchneri* (12), *R. bellii* (13), *R. felis (*16), and *Rickettsia massiliae*. (14) As with these prior reports, we also did not identify a gene encoding a pilin protein (typically TraA in F-T4SSs) in RAGE T4SSs, though it may be that a pilus is synthesized using a different pilin gene since the RAGE-harboring *R. bellii* forms large pili during host infection (13). Experiments are needed to determine if the RAGE T4SS builds a pilus or functions pilus-less, as is noted for the P-T4SS of *Rickettsia* species (41), *Neorickettsia risticii* (42), and likely all Rickettsiales (43). Another common peculiarity of these F-T4SSs is the split gene encoding TraK, the significance of which is unknown.

The relaxosome of the *E. coli tra*/*trb* F-T4SS encodes one multi-functional relaxase, TraI, which excises and binds single-stranded plasmid DNA (45). In contrast, the Ot RAGE carries three genes, *traI*, $traA_{Ti}$, and $traD_{Ti}$, predicted to comprise the relaxosome that mobilizes RAGE (Fig. 6C). *E. coli* TraI harbors four distinct domains required for nicking, binding, and unwinding DNA. By contrast, Ot TraI lacks a domain for nicking DNA and shares very limited similarity to *E. coli* TraI. However, Ot $TraA_{Ti}$ carries a MobA-like domain that cleaves single- and double-stranded DNA at specific sites (46). Curiously, all the domains encompassed by Ot TraI and $TraA_{Ti}$ proteins are found in a single *Rickettsia* RAGE protein named $TraA_{Ti}$-I, which is highly similar to both Ot TraI and $TraA_{Ti}$ but shares limited similarity to *E. coli* TraI. As their annotation indicates, RAGE $TraA_{Ti}$ and $TraD_{Ti}$ are similar to relaxosome proteins of plasmid Ti of *Agrobacterium tumefaciens* (ATU), TraA and TraD, which are required for T-DNA translocation into plant cells via the *vir* T4SS (47). The significance of different relaxosome structures between Ot and *Rickettsia* RAGEs is unclear, although $traA_{Ti}$ and $traD_{Ti}$ genes are common on *Rickettsia* plasmids even when RAGEs are absent (48). It may be that multiple RAGE types exist in the rickettsial mobilome and are defined by their cognate relaxosomes. The presence of transposases flanking relaxosome genes in all complete Ot and *Rickettsia* RAGEs may also signify that RAGEs evolve by recombining different relaxosome cassettes into the conjugation and cargo genes.

Over 50% of the RAGE T4SS and relaxosome genes are present as truncated pseudogenes, and some of these clusters encode only a subset of the 18 possible Ot RAGE mobilization genes (Fig. 6D and E). Given the high degree of pseudogenization, we sought to examine whether any strain encoded any RAGE mobilization gene clusters containing a complete complement of full-length genes. We found that Karp, Kato, Gilliam, and UT76 encoded at least one complete RAGE mobilization gene set, while Ikeda, Boryong, TA686, and UT176 did not (Fig. 6D and E; Fig. S3; data set S7). There was a positive correlation between strains containing complete sets of RAGE mobilization gene

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

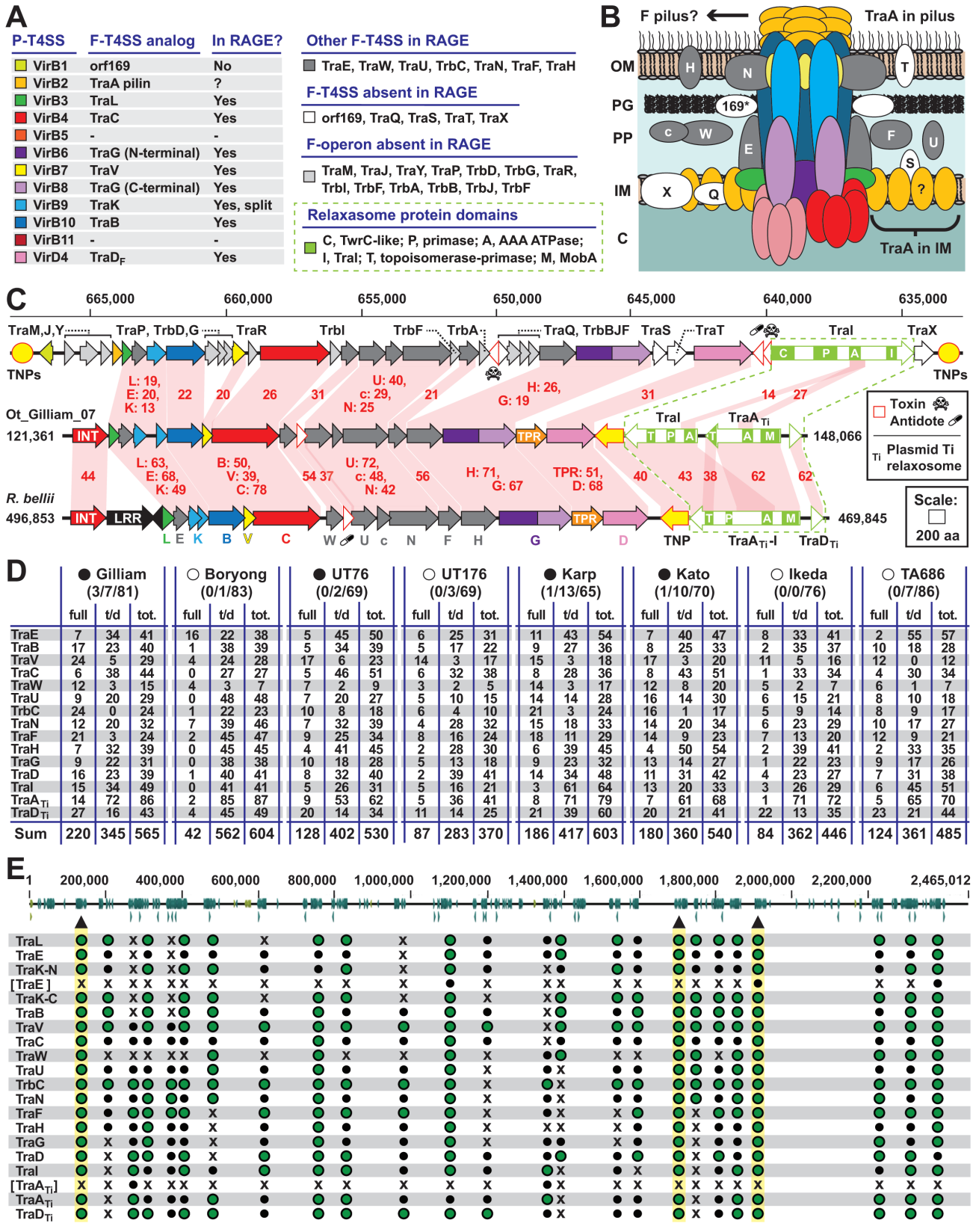


FIG 6 Characteristics of the F-type T4SS and relaxosome proteins encoded on Ot RAGEs. (A) Composition of the Ot RAGE F-T4SS in relation to the *Agrobacterium tumefaciens vir* P-T4SS and the *Escherichia coli tra/trb* F-T4SS from the F operon. Analogs across divergent T4SSs are colored similarly, with other colors as follows: dark gray, RAGE T4SS proteins found in F-T4SSs but not P-T4SSs; white, *E. coli* F-T4SS scaffold genes not present in RAGE T4SSs; light gray, other *E. coli* F operon

(Continued on next page)

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

FIG 6 (Continued)

genes not present in RAGE. For relaxosome proteins (olive green), domains were predicted with SMART (44). (B) Theoretical assembly of the RAGE T4SS in relation to data from other F- and P-type T4SSs. The uncertain synthesis of a pilus is depicted (see text for details). (C) Comparison of the *E. coli* F operon to mobilization genes of complete RAGEs from Ot str. Gilliam and *Rickettsia bellii* str. RML369-C. This *E. coli* strain, K-12 ER3466 (CP010442), has the F operon on a chromosomal segment flanked by transposases (yellow circles). Red shading and numbers indicate % aa identity across pairwise alignments. Dashed lines enclose the relaxosome genes, whose protein domains are described in panel A. (D) Frequency and distribution of of full-length and truncated *tra*/*trb* genes in Ot strains. Complete circles, genomes containing full sets of *tra*/*trb* genes within one or more RAGE; open circles, no complete *tra*/*trb* gene sets. Numbers in parentheses: number of complete RAGEs/number of complete RAGE genes containing truncated genes/incomplete RAGEs. Details of truncated genes and gene fusions are given in data sets S1 and S8. (E) Genomic location of *tra*/*trb* gene clusters in Ot str. Gilliam. Triangles and highlighting depict complete RAGEs. Bracketed TraE and $TraA_{TI}$ are commonly occurring pseudogenized duplications. Green circles, complete gene; small black circles, predicted pseudogene; Xs, gene absent with *tra*/*trb* gene cluster. INT, integrase; LRR, leucine rich repeat protein.

sets and the total number of full-length mobilization genes (Fig. 6D and E). Moreover, the complete RAGE mobilization gene clusters in Gilliam and Kato were located within complete RAGE regions (Fig. 1C and 6E; Fig. S3). While several genomes lack complete RAGE mobilization gene clusters, all except Boryong encode at least one full-length copy of each RAGE mobilization gene, albeit not in a contiguous cluster. Therefore, it is possible that all strains except Boryong could assemble a functional F-type T4SS competent to mediate transfer of RAGEs.

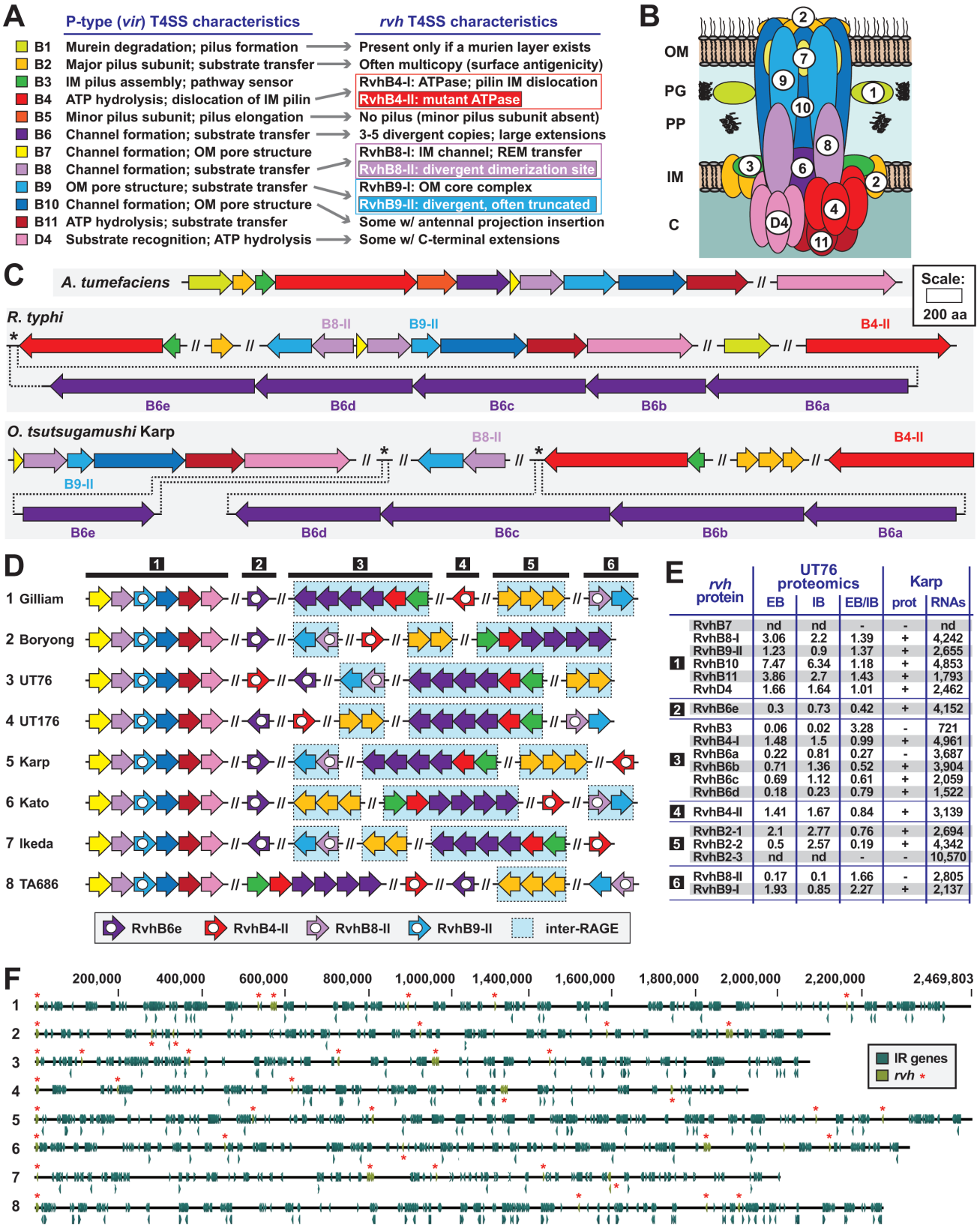
P-T4SS

Like other rickettsial species, Ot encodes a P-type T4SS related to the archetypal *vir* T4SS of the pTi plasmid of *A. tumefacians*. Thus, Ot encodes a second class of T4SS in addition to the RAGE-encoded F-T4SS. Relative to *vir*, this Rickettsiales *vir* homolog (*rvh*) T4SS has distinct features, including the scattered distribution of *rvh* gene clusters; duplication of *rvhB8*, *rvhB9*, and *rvhB4*; three to five copies of *rvhB6*; and no gene encoding an equivalent to the VirB5 minor pilin subunit (43, 49, 50) (Fig. 7A and B). These characteristics are nuanced: (i) RvhB4-II, RvhB8-II, and RvhB9-II carry atypical structural deviations from described VirB4, VirB8, and VirB9 family proteins; (ii) RvhB6 proteins have large insertions flanking the VirB6-like membrane spanning region; and (iii) a lack of a minor pilin subunit precludes formation of a T-pilus (50, 51). There is evidence that structurally different RvhB8-I and RvhB8-II proteins of *R. typhi* cannot dimerize (51), which leads to the hypothesis that divergent duplications may autoregulate effector secretion (50). The recent identification of *rvh* genes in all seven Rickettsiales families implies a highly important function (3, 52). Thus, we assessed the properties of the Ot *rvh* T4SS in the face of its rampant mobile genetic element-induced genome shuffling.

A single set of *rvh* genes is present in all the Ot genomes analyzed here and has features resembling those described in other Rickettsiales species (Fig. 6A through C). Interestingly, Ot shares two key characteristics with the *rvh* T4SS of distantly related Anaplasmataceae species as opposed to that of closely related *Rickettsia* species. First, while Ot lacks genes for a VirB5 protein and therefore cannot form extracellular pili used for cellular attachment, it encodes two to three copies of *rvhB2*, the major pilus subunit. Given that Ot lacks lipopolysaccharide (LPS) on its surface, multiple RvhB2 proteins may act as divergent surface antigens in a similar fashion previously posited for Anaplasmataceae species, which collectively lack LPS biosynthesis genes and have multiple *rvhB2* genes throughout their genomes (43, 53). Second, Ot lacks any identifiable *rvhB1* gene, which is present only in *Rickettsia* spp. This gene encodes a lytic transglycosylase predicted to cleave peptidoglycan (PGN) to allow T4SS scaffold assembly (49). Compared with *Rickettsia* spp., which synthesize a canonical PGN layer (54, 55), the presence of a minimal cell wall in Ot and Anaplasmataceae species is consistent with the absence of *rvhB1* from these genomes (56, 57). These collective differences in *rvh* T4SS architecture present clear convergent evolution in Ot and Anaplasmataceae species in the context of shared cell wall morphology and probable responses to host cell immune pressures (58).

Our analysis shows that the identities of six *rvh* gene clusters are conserved across Ot strains, while the genomic positions of the clusters vary between strains (Fig. 7D).

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.



	rvh protein	UT76 proteomics EB	UT76 proteomics IB	UT76 proteomics EB/IB	Karp prot	Karp RNAs
	RvhB7	nd	nd	-	-	nd
	RvhB8-I	3.06	2.2	1.39	+	4,242
	RvhB9-II	1.23	0.9	1.37	+	2,655
1	RvhB10	7.47	6.34	1.18	+	4,853
	RvhB11	3.86	2.7	1.43	+	1,793
	RvhD4	1.66	1.64	1.01	+	2,462
2	RvhB6e	0.3	0.73	0.42	+	4,152
	RvhB3	0.06	0.02	3.28	-	721
	RvhB4-I	1.48	1.5	0.99	+	4,961
3	RvhB6a	0.22	0.81	0.27	-	3,687
	RvhB6b	0.71	1.36	0.52	+	3,904
	RvhB6c	0.69	1.12	0.61	+	2,059
	RvhB6d	0.18	0.23	0.79	+	1,522
4	RvhB4-II	1.41	1.67	0.84	+	3,139
	RvhB2-1	2.1	2.77	0.76	+	2,694
5	RvhB2-2	0.5	2.57	0.19	+	4,342
	RvhB2-3	nd	nd	-	-	10,570
6	RvhB8-II	0.17	0.1	1.66	-	2,805
	RvhB9-I	1.93	0.85	2.27	+	2,137

FIG 7 Synopsis of Ot P-type (*vir*-like) T4SS genes. (A) Description of the general *rvh* T4SS characteristics, summarized from prior studies (41, 43, 49, 50). (B) Theoretical assembly of the RAGE T4SS in relation to data from other P-type T4SSs. There is no synthesis of a T-pilus (see text for details). (C) Comparison of genes encoding *vir* T4SS in *Agrobacterium tumefaciens*, the archetypal P-T4SS, and those encoding the *rvh* T4SS in *Rickettsia typhi* and Ot. (D) Arrangement of

(Continued on next page)

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

FIG 7 (Continued)

rvh genes in Ot genomes. Red genes are located in RAGE regions, while blue are located in IR regions. (E) Previously published RNAseq and proteomics data showing relative expression levels of *rvh* genes in strains UT76 and Karp. These are taken from Atwal et al. (UT76) (21) and Mika-Gospodorz et al. (19) (Karp). UT76 data show relative peptide counts in IB and EB. Karp data show presence or absence of detectable peptides from proteomics analysis (±) and relative RNA transcripts from RNAseq data (TPM). (F) Distribution of *vir* genes across Ot genomes showing lack of conservation of absolute position, despite similarities in gene groupings as shown in Fig. 7D. EB, extracellular bacteria; IB, intracellular bacteria; TPM, transcripts per million.

Clusters 1 (*rvhB7*, *rvhB8-I*, *rvhB9-II*, *rvhB10*, *rvhB11*, and *rvhD4*), 2 (*rvhB6e*), and 4 (*rvhB4-II*) are located within RAGEs in all strains, with the other *rvh* genes consistently located in IR regions except for cluster 6 and clusters 3 and 6 in UT176 and TA686, respectively (Fig. 7D and F). Analysis of published data sets of proteomics and RNAseq data in Karp and UT76 (19, 21) show that RvhB2-3 and RvhB7 proteins are not detected under growth conditions used in those analyses, although transcription levels of *rvhB2-3* are high (Fig. 7E). All the other Rvh proteins are detected in UT76, and most are detected in Karp. The UT76 data set compared peptide levels in two different bacteria populations: intracellular bacteria (IB) and extracellular bacteria (EB) (21). The EB:IB ratio of some multi-copy Rvh proteins differs between paralogs, e.g., RvhB2-1, which is present at a ratio of 0.76 compared with the ratio of 0.19 for RvhB2-2. This suggests expression of these proteins may be differentially regulated, potentially reflecting functional differences. Our collective analyses indicate that Ot Rvh genes can form a functional P-T4SS, despite the pervasive mobile element-induced gene shuffling in Ot genomes. While no Ot *rvh* transported effector has been described to date, it is highly likely that Ot utilizes the *rvh* T4SS during host cell infection, as secreted proteins that interact with the T4SS gatekeeper, RvhD4, have been described for *R. typhi* (59–61), *R. rickettsii* (60, 62), *Anaplasma marginale* (63), *Anaplasma phagocytophilum* (64–70) and *Ehrlichia chaffeensis* (66, 71–74).

Ot lacks defense mechanisms against invasive DNA

We show here that the Ot genome is exceptional in its abundance of invasive mobile genetic elements, including ICEs, transposases, group II introns, and GTAs. Bacteria have evolved a range of anti-viral mechanisms to minimize damage caused by mobile genetic elements (75–77). We therefore sought to assess if Ot lacks these protective systems, possibly explaining the invasion and subsequent proliferation of mobile genetic elements. We used DefenseFinder to carry out a systematic search for all known antiphage systems including restriction modification systems, CRISPr/Cas systems, and toxin-antidote defense modules (76, 78). We found that none of the Ot strains in our study had any identifiable defense systems. While it is possible that this is due to sequence divergence, small size (e.g., certain toxin-antidote modules), or systems that have not yet been discovered, the software was able to detect three different restriction modification systems and the newly described Pyscar defense system (79) in the closely related free-living alphaproteobacterial *Caulobacter crescentus* (CCS). In addition to lacking identifiable anti-viral defense systems, Ot also has limited homologous recombination capability, a system that is frequently used in anti-viral defense (76). While Ot encodes RecA and the alternative homologous recombination pathway RecFOR, it lacks the major repair complex RecBCD that can defend against some mobile genetic elements by degrading linear double-stranded DNA (11). Overall, Ot lacks identifiable mobile genetic element defense systems likely explaining the proliferation of mobile DNA in these genomes.

Conclusions

The identification of complete RAGEs in two Ot strains raises the possibility that these ICEs remain actively mobilizable both within and between Ot genomes. Evidence for this hypothesis awaits the sequencing of intact and complete genomes of large numbers

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

of Ot isolates beyond the small number currently available. While only two genomes encode complete RAGEs with full length genes, all encode all the genes required for RAGE mobilization, albeit in dispersed locations across the genome. Future research is needed to determine whether such mosaic RAGEs can be mobilized or not.

The identification of potentially active RAGEs in Ot raises the question of how they can be transferred between Ot organisms during their life cycle. Ot is an obligate intracellular bacterium, and therefore, bacterial cells have limited interactions with other bacteria of the same or different species. It has been shown that simultaneous infection with multiple Ot strains is found during human infection (80), and it is therefore possible that different strains of Ot can infect the exact same host cell in a mite or a rodent during co-infection by two species. Albeit rare, this could occur with sufficient frequency to enable horizontal gene transfer between species. Alternatively, it is possible that the extracellular form of Ot retains sufficient residual metabolic activity to support lateral DNA transfer in the cell-free extracellular state. Mites typically feed in a tight cluster, for example, on the ear of an infested rodent, and the co-feeding pool may provide the environment for close encounters between Ot cells in an intracellular or extracellular state to mediate conjugation.

In conclusion, this study has led to the manual re-annotation of the genomes of eight strains of Ot, enabling the delineation of RAGE and IR regions. Open questions remain. Importantly, while intact RAGEs have been identified in two strains, the dynamics of the Ot RAGE are completely unknown. It is also unknown whether the current set of RAGEs within one genome results from one or multiple invasion events. The Ot RAGE encodes an F-T4SS, but it is not known if these are active, nor what they transport beyond the ICE itself. Progress toward answering these questions will enable further insights into the biology and pathogenicity of this important human pathogen.

MATERIALS AND METHODS

Identification of RAGE and inter-RAGE regions in the genome of Ot

The boundaries of RAGE and IRs were manually delineated in each genome using defined criteria. First, groups of genes whose positions relative to one another were conserved across strains were identified manually by comparing the genomes of eight Ot strains (Ikeda, Boryong, Karp, Kato, Gilliam, TA686, UT76, and UT176). This led to the identification and numbering of IR regions. RAGE regions were subsequently identified using criteria largely drawn from Nakayama et al. (9).

The element was classified as a "complete RAGE" if the element encoded a full-length integrase gene at the left end (N-terminus), a full-length transposase gene, full-length set of conjugative transfer genes (tra genes: *traA*, *traB*, *traC*, *traD*, *traE*, *traF*, *traG*, *traH*, *traI*, *traK*, *traL*, *traN*, *traU*, *traV*, and *traW*), and non-conjugative genes (RAGE-associated cargo genes) including one or all of the following: SpoT-related proteins [ppGpp hydrolase, (p)ppGpp synthetase, SpoT synthase, and SpoT hydrolase], DNA methyltransferase, DNA helicase, histidine kinases, ATP-binding proteins (mrp), HNH endonuclease, membrane proteins, ankyrin repeat proteins, and hypothetical proteins. The RAGE-associated cargo genes in the "complete RAGE gene" can be either full-length or truncated genes.

The element was classified as a "complete RAGE with truncated genes" if the sequence encoded the same gene set as above, but where one or more of the integrase, transposase, or *Tra* conjugative transfer genes were truncated.

The element was classified as an "incomplete RAGE" if the sequence encoded integrases or transposases, and at least one RAGE-associated cargo gene.

Finally, we identified single genes, or groups of genes, that could not be classified as a RAGE or IR region using these criteria. The "isolated mobile gene" was defined as encoding one or more integrase or transposase without RAGE-associated cargo genes. The "isolated cargo gene" was defined as encoding one or more cargo genes without

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

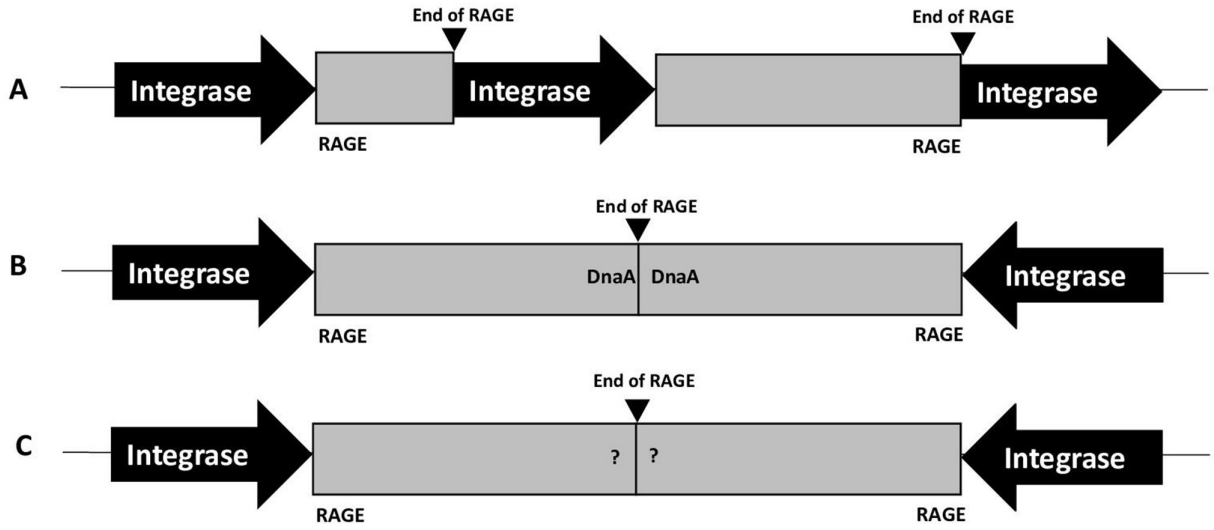


FIG 8 Identification of RAGE termination sites.

transposases, integrases, or *tra* genes. The "isolated hypothetical protein" was defined as encoding one or more hypothetical proteins at the boundary of conserved IRs or RAGEs.

The presence of a *dnaA* gene was used as an indicator gene for defining the end of a RAGE element (Fig. 8). However, the criteria could not be applied for all RAGE elements when hypothetical proteins and transposases were located at the end of RAGE masking the original *dnaA* terminus. In the first case (Fig. 8A), RAGE elements are located next to each other in the same direction. In this case RAGE is terminated when the integrase gene of the next RAGE is found. In the second case (Fig. 8B), RAGE elements are located next to each other in opposite direction, and two *dnaA* genes are located next to each other. In this case, the RAGE is terminated at the *dnaA* gene, which belongs to RAGE on the left (forward direction) and RAGE on the right (reverse direction). In the third case (Fig. 8C), RAGE elements are located next to each other in opposite direction. Two RAGEs were combined into one RAGE if a *dnaA* gene was not present in either RAGE.

Identification of RAGE-associated cargo genes

The list of conserved non-conjugative genes or RAGE-associated cargo genes were extracted based on Nakayama et al. (9). The sequences were then inspected manually by observing the length of genes, conserved motifs/domains, and other elements such as signal peptides and transmembrane regions. A gene was defined as a "full-length gene" if the sequence encoded a complete set of domains, a "truncated gene" if the sequence encoded only a partial set of domains and a "degraded gene" if no domain was identified on the sequence.

Analysis of multi-copy cargo genes not associated with DNA mobilization

Membrane proteins

Membrane proteins were manually extracted from the genome database, and the SMART search engine (44) was used to identify membrane domains and other elements. A large group of genes with homology to each other, but no identifiable known domain, was renamed "OT_RAGE_membrane_protein." The membrane protein encoding gene was assigned a new name if significant domains/main domains were found, such as autotransporter proteins (Sca family), vut1- putative vitamin uptake transporter, RhaT-permease of the drug/metabolite transporter superfamily, and Bax inhibitor-1 (BI-1)/YccA inhibitor of FtsH protease domains.

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

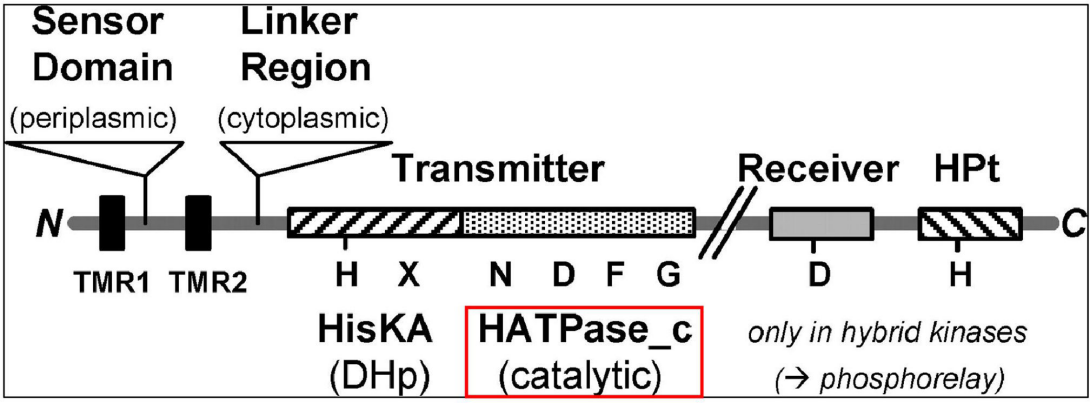


FIG 9 An overview of protein domains histidine protein kinases. Figure is modified from reference (84).

MRP/histidine kinases

All genes previously annotated as HK and multi-drug resistance-associated proteins (MRPs) were extracted from the genome databases and analyzed using SMART (44). Both HK and mrp proteins contain an HATPase_c domain (histidine kinase-type ATPases catalytic domain) (Fig. 9). This domain is found in several ATP-binding proteins, including histidine kinase (81), DNA gyrase B (82), topoisomerases, and heat shock protein HSP90 (83). The HK and mrp genes were renamed based on the presence of an HATPase_c domain. HK or mrp proteins were renamed as "HATPase_c domain containing protein" if the search found an HATPase_c domain. HK or mrp proteins were renamed as "degraded HATpase_c" if no HATPase_c domains were found in the search. Where additional domains were identified, the genes were appropriately renamed, e.g., genes comprising two component sensor systems. Where additional domains were not identified, the HK and mrp proteins in this study were classified as truncated genes because they only contained the catalytic part (HATPase_c) and lacked other major domains such as sensor domain, HisKA (histidine kinase A domain dimerization/His phospotransfer), and receiver Hpt domain (histidine phosphotransfer) (Fig. 9) (84).

In addition, the same HATPase_c domain was also found in symporter proteins. The sodium:proline symporter "PutP" was classified as full length if the sequence contained symporter, HisKA, HATPase_c, and REC domains. PutP was was classified as "truncated PutP" or "degraded PutP" if the sequence lacked the symporter and REC domains. The symporter Sodium:pantothenate symporter "PanF" was classified as full-length if the sequence contained only the symporter region.

(p)ppGpp hydroplase/synthetases

We manually extracted all genes annotated as SpoT, (p)ppGpp, synthetase, and hydrolase from the genome databases. The sequences were then compared to their respective orthologs in ECO and CCS. Literature searches and GenomeNet motif search (pfam) were used to identify motifs and conserved regions in these sequences as shown (Table 2).

We employed a naming system based on the presence of domains/motifs in each CDS. Genes containing all domains (HD, SYNTH, TGS, helical, CC, and ACT) were annotated as SpoT. Genes having only the HD domain were annotated as SpoT hydrolase, while genes encoding only the SYNTH domain were annotated as SpoT synthetase. Genes encoding the HD domain but lacking one or more of the conserved histidines were annotated as truncated hydrolase. Short fragments that could be aligned to hydrolase but lack complete domains were annotated as degraded hydrolase. Genes containing HD domain merging with a part of HATPase were named as HATPase-SpoT

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

TABLE 2 Motifs conserved in SpoT[a]

Domains	Conserved	Motifs	References
HD domain (hydrolase)	Region	AIDYAIHYHGXQTRESGDPYYYHPLHVALIIAQMKXDTVSVITAL LHDTVEDTELTLSDIEREFGKEVAXLVDGVTKLXKLRFQSYHXQQ AXNFRKLLLAISNDIRVLLVKLADRLHNMRTIESIKLLNKRIRIALE TXEIYAPLAERIGA	(85)
	H1	HXXXXR/KXXG/QXXYXXXP/Q/WXX	(86)
	H2	I/VT/IAXLHD/N	(86)
	H3	XLLXKLXXRXHNXXXX	(86)
SYNTH domain (synthetase)	Region	G/ARXKXXYSIXXKMXXKXIXXXQLXDXXAXRXIXXXXXXXXXXC YXXLXXIHXXYXXXPXXXQDFIXXPKXNGYQSXHTXIXGPXXXXI EVQIRTXXMHXXXXXGXAAHWXYK	(85)
	G	GRHK	(86)
	YQS	NGYQSXHT	(87)
TGS domain (Thr-tRNA synthetase, GTPase, and SpoT domain)	Region	CFTPXGKLIALPKGATVVDFAYKXHSELGNKCIGAKISNKVVPLD TQLQNGDQVEIIT	(85)
	H	DFAYXXHXXXG	(88)
Helical domain	Region	TFAVTGKAQSEIRKFIRXQAYKKYIDLGKEILIQTLKKIQVANINV CIAKIAHXLNKKNVEEVFFXIGXELLSKKEIIKIIT	GenomeNet motif search (pfam)
CC domain (conserved cysteine/RIS-ribosome inter-subunit domain	CC	CCYPLPGDLIIGLCT	(85, 89)
ACT domain (aspartokinase, horismate mutase, and TyrA/RRM RNA recognition motif)	Region	RNKIGSLASITTILENNNXNICNIKTTNXTQSTXQIIIDIEISTLEQL NKIXNILQSSXDIISVXR	(85)

[a]Underlined bases indicate important amino acids in the motif.

hydrolase, and genes containing HD domain merging with a part of Mrp were named as Mrp-hydrolase, respectively.

DNA methyltransferases

DNA methyltransferase (MTase) genes were extracted from the genome and analyzed using SMART (44) for protein domain annotation. However, SMART does not provide the details of MTase motifs within the predicted domain. Therefore, multiple sequence alignments of DNA methyltransferase were carried out for identification of motifs using Geneious. We used the conserved amino acid residues in the Dam protein of *E. coli* (acc.no. P0AEE9) as a reference for identification of motifs I–VII and motif X of MTase at the C-terminal region (90). The protein sequence of DNA methyltransferase containing motifs I–VII and motif X was classified as a full-length gene. The protein sequence of DNA methyltransferase with incomplete motifs and unidentified motifs was indicated as truncated gene and degraded gene, respectively.

Replicative DNA helicases

DNA helicase genes were filtered from the genome, and their protein domains were characterized by SMART (44). Multiple sequence alignments of the helicase genes were then carried out in order to identify motifs. We used the conserved amino acid residues in the DnaB protein of *E. coli* K12 (accession no. NC000913.3) as a reference for the identification of motifs I–VII at the C-terminal region (91). The protein sequence of DnaB containing motifs I–VII was indicated as full-length gene. The protein sequence of DnaB with incomplete motif and unidentified motifs was indicated as truncated gene and degraded gene, respectively.

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

Uncharacterized proteins

Between 308 and 464 genes annotated as hypothetical or uncharacterized were found in the eight genomes of *Orientia*. In this study, we only manually analyzed uncharacterized proteins from Karp strain as a model to minimize the analysis time. Uncharacterized proteins from Karp were filtered from the genome, and the protein domain was characterized by SMART (44). Where clear groups of homologous genes within the set of Karp genes were found, these were classified into 25 defined groups. These were renamed according to known domains with which they had homology or named Ot_RAGE_hypo_group 1–7. These 25 groups were then aligned to the other seven genomes in our data set in order to determine the conservation of the groups of genes.

Some uncharacterized proteins were renamed and no longer classified as hypothetical proteins if they aligned to known genes such as DnaA, Phage portal protein, and Lipase3.

Analysis of multi-copy genes involved in DNA mobilization

Insertion sequence transposable elements

The presence of IS elements in Ot was investigated using the online search tool ISfinder (92) to match with attributes and nomenclatures previously submitted for *Orientia*-specific IS (9). Each IS match was manually traced for completeness with flanking IRs and direct repeats along respective genome sequences. Extensive analysis was performed with Karp strain to identify the complete set of IS elements and classified into classes. Only ISOt3 and ISOt5 were systematically analyzed across all eight different Ot genomes.

Integrases

Genes annotated as integrase genes were extracted from the genomes, and protein domains were screened by SMART (44). Integrase in Ot is a phage integrase which is classified into two major families: the tyrosine recombinases and the serine recombinases, based on mode of catalysis (93). Multiple sequence alignment of phage integrase domain was analyzed for identification of motifs using Geneious. We used the conserved amino acid residues in Bacteriophage P2-integrase (accession no. AF063097.1) and Enterobacteria phage P2-integrase (accession no. NC_009488.1) as references for the identification of three domains; arm-type binding motifs at N-terminal region, core-type binding, and catalysis at C-terminal region. The His-X-X-Arg motifs and second conserved arginine on catalytic domain were also included in the alignment (93, 94). The protein sequence of phage integrase containing arm-type binding, core-type binding, and catalysis motifs was used to identify full-length genes. The protein sequence of phage integrase with incomplete motif and unidentified motif were annotated as truncated gene and degraded gene, respectively.

Reverse transcriptases

Reverse transcriptase (RVT) genes were filtered from the genome, and protein domains were characterized by SMART (44). Multiple sequence alignment of *rvt* was then analyzed for identificatiable motifs. We used the conserved amino acid residues of group II intron reverse transcriptase/maturase (LtrA) in *E. coli* (accession no. WP_096836589.1) and *Lactococcus lactis* (accession no. NZ_CP059048.1) as a reference for the identification of three domains: reverse transcriptases (RVT_N) at the N-terminal site, reverse transcriptases (RVT), and group II intron, maturase-specific domain (GIIM) at the C-terminal site (95–97). The protein sequence of reverse transcriptase containing RVT_N, RVT, and GIIM was used to identify full-length genes. Genes encoding an incomplete reverse transcriptase motif or no identifiable reverse transcriptase motif were classified as truncated and degraded genes, respectively.

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

Transposases

Genes annotated as transposase genes were first extracted from the genome. Then, protein domains and motifs were further characterized by SMART (44) and Geneious, respectively. Transposase genes in Ot belong to restriction endonuclease-like proteins or PD-(D/E)XK nucleases and DD[E/D]- transposase, which generally contain the catalytic domain and transposon-binding domain (98, 99). Some transposases additionally contain C-terminal or N-terminal domains (100). In this study, we used the conserved amino acid residues of in *E. coli* (accession no. NC 002695.2) as a reference for the identification of restriction endonuclease-like motifs (I–IV). Three conserved active sites in motifs II and III, one of which is aspartic acid (D), one is either glutamic (E) or aspartic acid (D) and/or the last one is lysine (K), were identified for characterization of PD-(D/E)XK nucleases and DD[E/D]- transposase motifs (98, 99). The protein sequence of transposase containing motifs (I–IV) and PD-(D/E)XK signature residues was used to identify full-length genes. Genes encoding an incomplete transposase motif or no identifiable motif were classified as truncated and degraded genes, respectively.

GTA genes

GTA genes in each of the Ot genomes were identified by BLASTp search using Wolbachia GTA genes (37, 40).

Analysis of ankyrin repeat containing proteins

To identify AR proteins (Anks) from previously annotated records, all Ank sequences were extracted and analyzed using SMART (44) to predict Ank repeats and other domains including coiled coil, F-box, and PRANC. SMART (44) defines an Ank as a 33-residue motif. SMART predictions were confirmed by manual inspection and additional BLAST searches using Ank repeat domains. In some instances, additional Ank repeat domains were identified that were not detected by SMART. The method used to identify new Ank domains is given in data set S5.

Identification of homologous Ank repeats across Ot strains was carried out manually using Geneious. The criteria for identification of a homologous Ank was based on sequence similarity and repeated units. Individual Ank sequences of the other seven strains were blasted against Ot strain Ikeda. Then, the sequence that presented the

```
O04242/1-30      NGHTALHIAASK-----------------GDEQCVKLLLEHGA------DPNA
CONSENSUS/80%    .t.sslhhsh.t.................tp.phhphllp.t.......pht.
CONSENSUS/65%    pstosLphAstp.................sphphlphLlptss......shsh
CONSENSUS/50%    sGpTsLHhAsps.................sshcllchLlspus......slst
```

Class	Key	Residues
alcohol	o	S,T
aliphatic	l	I,L,V
any	.	A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y
aromatic	a	F,H,W,Y
charged	c	D,E,H,K,R
hydrophobic	h	A,C,F,G,H,I,K,L,M,R,T,V,W,Y
negative	-	D,E
polar	p	C,D,E,H,K,N,Q,R,S,T
positive	+	H,K,R
small	s	A,C,D,G,N,P,S,T,V
tiny	u	A,G,S
turnlike	t	A,C,D,E,G,H,K,N,Q,R,S,T

FIG 10 Consensus terms used by SMART for the identification of Ank domains.

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

TABLE 3 Vir proteins and their motifs

Subunit(s)	Motif(s)	Reference(s)
VirB2-1	TM region 1 (GXXXXXXXXXXXXXXXIXXXXG), TM region 2 (AI**I/VI/V**XXX**A/S**XX)[a]	(101–103)
VirB2-2	TM region 1 (GXXXXXXXXXXXXXXXIXXXXG), TM region 2 (AI**I/VI**XXX**A/S**XX)[b]	
VirB2-3	TM region 1 (GXXXXXXXXXXXXXXXIXXXXG), TM region 2 (AI**I/VI**XXXSXX)	
VirB3	L-TRP-GV motif (LXXXXTRPXXXXGV)	(103, 104)
VirB4-1	Walker A (GPXGXGKT), motif C (FDKDRGXE), Walker B (RIXXXXDGXXXXXXXDE), motif D (LXXXRKXN), motif E (IXATQ)	(103, 105)
VirB4-2	Walker A (GXXXXG**K/R**), Walker B (SLXXXXXXXIXXXDX)	
VirB6-1	Variable TM region (VXAFXXLYXXXXGXXILLX), conserved cytoplasmic loop (PXXXXXXXFXXTXXXXXXW)	(105–107)
VirB6-2	Variable TM region (XXAALXLYXXFFXXXXXXX), conserved cytoplasmic loop (PXXXXXXXFXXTXXXXXXW)	
VirB6-3	Variable TM region (IXAXLXLYXMXXGXXFXLG), conserved cytoplasmic loop (PXXXXXXXFXXTXXXXXXW)	
VirB6-4	Variable TM region (XXXXXXLYXTXXGXXFXLG), conserved cytoplasmic loop (PXXXXXXXFXX**T/I**XXXXXXW)	
VirB6-5	Variable TM region (VXXXLXLXXXFXGXXFLIG), conserved cytoplasmic loop (PXXXXXXXFXXTXXXXXXW)	
VirB7	Conserved cys between positions 15–35, possibility of being a small lipoprotein	(103)
VirB8-1	Homodimerization domain I (YXXXREXY), VirB4 interaction region (XXXY**K**)	(103, 108, 109)
VirB8-2	Homodimerization domain I (YXXXREXY), NPxG motif (NPXG), VirB4 interaction region (XXXY**R**)	
VirB9-1	DXR-YXP motif (DXRXXXXXYXP), Beta 1 (NXXYXX), Beta 2–3-OM extrusion region (PXXXXDXXXXTXXXF - PXXXXX**G/DD**XXXE), VirB7 interaction region (RXGXXXXCXXN)	(103)
VirB9-2	DXR-YXP motif (DXRXXXXXYXP)	
VirB10	OM pore gating (DXLGXXGXXGXV), Beta 6 a (VLXSAX), Beta 7a (XTXXXNQG)	(105, 110)
VirB11	Linker A (IRXXSXXXXXL), Beta 1 (XXEXXXNXPG), Beta 5 (LPXXXRXQXXXPP), ATPase region/Beta 7-Alpha E (GXTXXXKTT), Beta 8 (ERXIXXED), Alpha F-Beta 10 (LXXXXLRXRPDRIXXXE), Beta 11 (GHPGSIXTXH)	(111)
VirD4	DNA-binding motif A (APTXXGKGXGXVIPXXXXXXXSVXXXDXK), DNA-binding motif B (XFLLDEFXXLGKXXX)	(112, 113)

[a]Underlining indicates important amino acids in the motif.
[b]Boldface indicates variability in amino acid sequence.

highest identity (>80%–90%) was chosen to verify the similarity of ankyrin repeats and other domains. The sequences were given the name based on the published Ikeda Anks if the overall sequence's identity to Ikeda Anks was more than 80% and presented a similar set of ankyrin repeats. The sequences were given a new name if the overall sequence identity to Ikeda Anks was less than 80% and/or presented a different set of ankyrin repeats; this included extra ankyrin repeat units or missing units.

Novel Anks or hypothetical proteins were manually identified and analyzed using Geneious. To identify additional Anks in Ikeda and the seven other strains in this study, each repeat unit of an individual published Ikeda Ank was imported into the "Find motifs" tool, and the maximum mismatches were set up to 10. Hits were verified using SMART. Then, the sequence of the novel Ank was BLASTed against the published Anks and within the strain to check whether it was different from previously identified Anks or not. The newly identified Anks were given a name by continued ranking after the published Ikeda Anks, starting with Ank21, Ank22, and Ank23 (Fig. 10).

Analysis of tetratricopeptide repeat containing proteins

To identify TPR proteins from the previous annotated records of eight Ot strains, all TPR sequences were extracted and analyzed using SMART (44). The program predicted the location of TPR motifs and other domains including signal peptide and transmembrane region. All identified TPR proteins were grouped based on the similarity of the location of TPR motifs. Each group consists of one long gene, the master gene, and multiple shorter duplication remnants. Novel TPRs were identified by manual search using Geneious. All TPR proteins were renamed based on the group number followed the number of TPR repeats.

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

TABLE 4 Motifs conserved in Tra and Trb proteins

Subunit	Motif(s)	References
TraE	Anchor region (**LVKY**NKXLLXXTXXL/IAXXXX), predicted conserved region 1 (SXXXXXXXYLXXXA), predicted conserved region 2 (KXXXXXSXFFXXXXXV), predicted conserved region 3 (VXIXGXXXXWXXXXKXXXXX**K/R**XYXLXXK)-GenomeNet Motif search (TraE region)	(114–116)
TraB	Coiled-coil domain (IX**XXXQ/K**XXXX**L/F**XXXXKXXXX), predicted conserved region (GXXSSERAXXR)-GenomeNet Motif search (TrbI-like region), OM pore gating (GXXGXXGXV), Alpha-3 (GXXXGXXX**A/V**XXXLXDXXIK**R/Q**AXXXXP)	(117, 118)
TraV	Conserved cysteine region (C**X**XXXXXXXXXXX**F/L/V**XCXXXXXXC), predicted conserved region (LXXL**F/L**XXXXXX**G/C**E)-**GenomeNet Motif search (TraV region)**[a]	(119, 120)
TraC	Predicted conserved region 1 (YXXYXX**E/K**XXLFXNXXXXGFXLXXXP), predicted conserved region 2 (YXXLXXQXXXXXFXLXXXXD)-GenomeNet Motif search (TraC region), Walker A (GXXGXGKX), Motif C (**V/A**XDXGXXXK), Walker B (RXXXXLXXIDEXW), motif D-E (RXXXGXXXXXTQ)[b]	(118, 121, 122)
TraW	TrbC interaction region (EXXXLXVIMXXLXXXXXXGXXXXXXXXF), predicted conserved region 1 (N**P/S**LXXXXXXXXLXXIXGDDXXQVXWXK), predicted conserved region 2 (FDQXXXLXXXXXIXXXPA)-InterPro Motif search (TraW region)	(123)
TraU	Signaling domain (AXXXCXG), Hydrophobic-2 (CMVXL**G/W**), Hydrophobic-3 (YWLXIXX), Hydrophobic-4 (FXNXXAXXAC-XAD), Hydrophilic-1 (KXXXRXQM), Hydrophilic-2 (**W/L**RKRX**C/Y**)	(114, 124)
TrbC	Signaling domain (MXIRVMXLXXLLXVNN), predicted conserved region 1 (FVSFSXXXXXLK), predicted conserved region 2 (GXXXXXR**G/R**XXNNXXXXT)-GenomeNet Motif search (TrbC region), TraW interaction region (ID**P/S**XLFXXYXXXX**VP/L**XXVX)	(123, 125)
TraN	Conserved cysteine region 1 (SCXEGXX), conserved cysteine region 2 (SXCXXXE), conserved cysteine region 3 (IGXXC), conserved cysteine region 4–5 (CXXXKXXYCXFXS**K/R**LAXXX**Q/H**), conserved cysteine region 6 (CR**G/D**XTVX**E/K**LQXXXF), predicted conserved region-1 (ECX**E**), Predicted conserved region 2 (CXLXXXXC), predicted conserved region 3 (CLXXXXXYX**C**), predicted conserved region 4 (CXKXXXXXX**N/H**CC)-GenomeNet Motif search (TraN region)	(126)
TraF	Predicted conserved region (**G/X**XXWYNX)-GenomeNet motif search (TraF region), C-X-X-C motif (CXXC), Beta 10 (**V**PXXX**L/S**X), Alpha 7 (ISX**D/N/E**XXXXXXL)	(127)
TraH	TrbI interaction region 1 (TXXGXXQX**Q/L**AAGYYXXGXLXXRT), TrbI interaction region (NIXXXAX), predicted conserved region 1 (CXXIXXYLXSFSXIX**G/V/RE**XL), predicted conserved region 2 (FLSSIGXXXXXXXYXXXXXISG), predicted conserved region 3 (LXQXXXEXXXXXR)-GenomeNet motif search (TraH)	(122, 128)
TraG	Membrane spanning region 1 (**M/W**XWXXXXXXIXXXLXXX), membrane spanning region 2 (QSVXXX**L/V**XXXXXXXVFPMXXLXXXXXIXKXWIXXIIWVXSWPVXF), membrane spanning region 3 (XX**A/ST/M**XXXLAXXXPXLSWXVX**K/N**XXXXXXXXLXXXFSXXXV), cleavage site (ASXXGX), predicted conserved region 1 (SXXXXLSXXL), predicted conserved region 2 (KQXXEQXXXXXXYXXQXS)-GenomeNet motif search (TraG N-terminal region)	(129)
TraD	Transmembrane region 1 (MXXQXXXNXXXIGLXXXXXWXXXXYQ), transmembrane region 2 (FLXXSXXXEXXXXFXIX), Walker A (GTXGXGKXX), Walker B (XXWFXXDELP)	(114, 130)
TraI	Helicase region 5 (XHGYAXTXXXX**Q/K**XAXXXXXXVLXXXXXXXXX), predicted conserved region 1 (IXEGXEXXXXLXXXXIXG-XIIXXXX**I/V**XXXXNXX**P/L**XX**G/S**), predicted conserved region 2 (AVXNXVXXXXAXXVX**E/D**XKXXXXXXXKXXFNXVLKXX**GL**)-GenomeNet motif search (Toprim region)	(131)
$TraA_{Ti}$	Nickase region 1 (AIXFXXXXXXXR**S/I**XGXXSCX**K/N**XXYXXXXXXXXXXXXXXXXXXXXXXXXXVXH), Nickase region 2 (NEVE**R/Q**XXX**X**XXNSXXXXXIVI**A/V**L**P/Q**), Nickase region 3 (NXHX**H/N**XXXXXRXXXXG), helicase region 1 (XXGXAGXGKXXXXXX**A/V**), helicase region 2 (XX**V/I**XD**E/K**AGM**V/A**), helicase region 3 (XXXXXLXGDXXQ**L/R**XXXEXGXXFXXXXXXXXXXL), helicase region 5 (XHGYAXTXXKX**Q/H**GAXXXXXX**V/I**LXXXXXXXXX), helicase region 6 (Y**V/T**XMT/IR**H/Y**XXXXXLY)	(131, 132)
$TraD_{Ti}$	Predicted conserved region 1 (RKXXX**R/Q**XXXXX**G/A**XX**V/L**XXAXL), predicted conserved region 2 (IGXXXFXXXXXN)-GenomeNet motif search (TraD region)	(131)

[a]Underlining indicates important amino acids in the motif.
[b]Boldface indicates variability in amino acid sequence.

Analysis of P-type IV secretion systems (*rvh*)

Literature search and BLAST search [National Center for Biotechnology Information (NCBI) and Kyoto Encyclopedia of Genes and Genomes (KEGG)] were performed to identify the presence of each Rvh subunit (RvhB1 to VirB11 and RvhD4) in Ot. The amino acid sequences of each Vir subunit present in Ot Boryong strain were compared to their respective orthologs in RBE and ATU for their percent of amino acids identity, length of

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

amino acid sequences, and presence of motifs. The presence of motifs was used as the major criteria to identify Vir subunits as indicated in Table 3.

VirB1 and VirB5 could not be identified in Ot. However, VirB7 was annotated based on gene positioning, conserved cysteine(s), and its important role of being a small lipoprotein in T4SS of *Rickettsia* species.

These *rvh* genes identified in Boryong were then set as a reference for other Ot strains. Each *vir* gene was then blasted (nucleotide blast using Geneious Prime) to identify the presence of each *rvh* across the eight strains of Ot (Boryong, Ikeda, Karp, Kato, Gilliam, TA686, UT76, and UT176). The amino acid sequences of each Rvh subunit from the eight strains of *Ot* were then aligned (multiple alignment, Clustal Omega) to verify their motifs. Even though some of the gene copies appear to be a truncation or pseudogene due to loss of some motif(s) like that in RvhB4-II, RvhB8-II, and RvhB9-II, they are well characterized in literature. So, these names were kept the same in our annotation.

Analysis of F-type IV secretion systems (RAGE T4SS)

Literature search and blast search (NCBI and KEGG) were performed to identify the presence of each Tra subunit (F-type T4SS: TraA to TraN, TraU to TraW, TrbC, and TrbE; P-type T4SS: $TraA_{Ti}$ and $TraD_{Ti}$) in Ot. The amino acid sequences of each Tra subunit present in Ot Ikeda strain (OTT) were then compared to their respective orthologs in RBE and *Escherichia coli* (ECO) for F-type T4SS or ATU for P-type T4SS. The presence of motifs was used as the major criteria to identify Tra subunits as indicated in Table 4.

Note that $TraA_{Ti}$ found in *Rickettsia* is fragmented into $TraA_{Ti}$ and TraI in *Ot*. The longest *tra* and *trb* genes identified in Ikeda were set as references and were then blasted (nucleotide blast using Geneious Prime) to identify their presence across the eight strains of Ot (Boryong, Ikeda, Karp, Kato, Gilliam, TA686, UT76, and UT176). The amino acid sequences of each Tra subunit from the eight strains of Ot were then aligned (multiple alignment, Clustal Omega) to verify their motifs. Those amino acid sequences with difference greater than 10% from full-length gene in Ikeda or missing motif(s) were classified as pseudogenes.

ACKNOWLEDGMENTS

We thank all the members of the lab in Cambridge and Bangkok for discussions and support. We are also grateful to Liz Batty and Nick Day in MORU for discussions, encouragement, and support. We thank Sarah Bordenstein for taking the time to send helpful comments on the manuscript after reading it on BioRXivs, and the anonymous reviewers for their thorough analysis and helpful comments on this work.

This work was funded by a Royal Society Dorothy Hodgkin Fellowship DH140154 (J.S.); Swiss National Science Foundation IZSTZ0_193925 (J.S. and C.K.); Wellcome Trust Senior Research Fellowship 224277/Z/21/Z (J.S.), grants from the National Institutes of Health, USA, R21 AI156762 and R21 AI166832 (J.J.G.), and the American Heart Association grant 20PRE35210610 (H.E.A.).

Project design and supervision: J.S.; data analysis and figure preparation: S.G., C.K., J.W., H.E.A., J.S., and J.J.G.; original manuscript writing: J.S.; manuscript revisions: S.G., C.K., J.W., J.J.G., and J.S.

AUTHOR AFFILIATIONS

[1]Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand
[2]Department of Microbiology and Immunology, Virginia Commonwealth University Medical Center, School of Medicine, Richmond, Virginia, USA
[3]Department of Microbiology and Immunology, School of Medicine, University of Maryland Baltimore, Baltimore, Maryland, USA
[4]Department of Pathology, University of Cambridge, Cambridge, United Kingdom

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

[5]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom
[6]Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom

AUTHOR ORCIDs

Joseph J. Gillespie http://orcid.org/0000-0002-5447-7264
Jeanne Salje http://orcid.org/0000-0001-5802-4580

FUNDING

Funder	Grant(s)	Author(s)
Royal Society (The Royal Society)	DH140154	Jeanne Salje
Wellcome Trust (WT)	224277/Z/21/Z	Jeanne Salje
HHS | National Institutes of Health (NIH)	R21 AI156762	Joseph J. Gillespie
HHS | National Institutes of Health (NIH)	R21 AI166832	Joseph J. Gillespie

AUTHOR CONTRIBUTIONS

Suparat Giengkam, Data curation, Formal analysis, Investigation | Chitrasak Kullapanich, Formal analysis, Investigation | Jantana Wongsantichon, Formal analysis | Haley E. Adcox, Formal analysis | Joseph J. Gillespie, Conceptualization, Formal analysis, Methodology, Writing – review and editing | Jeanne Salje, Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

All data generated by this work are available within the article and supporting information.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Data Set S1 (mSphere00268-23-s0001.xlsx). Genomes.
Data Set S2 (mSphere00268-23-s0002.xlsx). Membrane proteins.
Data Set S3 (mSphere00268-23-s0003.xlsx). MRP and HK.
Data Set S4 (mSphere00268-23-s0004.xlsx). Hypothetical proteins.
Data Set S5 (mSphere00268-23-s0005.xlsx). Ankyrin repeat proteins.
Data Set S6 (mSphere00268-23-s0006.xlsx). Tetratricopeptide repeat-containing proteins.
Data Set S7 (mSphere00268-23-s0007.xlsx). Gene transfer agents.
Data Set S8 (mSphere00268-23-s0008.xlsx). Type 4 secretion systems.
Supplemental Figures (mSphere00268-23-s0009.pdf). Figures S1 to S3.
Captions (mSphere00268-23-s0010.docx). Captions for Data Sets S1 to S8.

REFERENCES

1. Giannotti D, Boscaro V, Husnik F, Vannini C, Keeling PJ. 2022. "The "other" Rickettsiales: an overview of the family "Candidatus midichloriaceae"" Appl Environ Microbiol 88:e0243221. https://doi.org/10.1128/aem.02432-21
2. Salje J. 2021. Cells within cells: Rickettsiales and the obligate intracellular bacterial lifestyle. Nat Rev Microbiol 19:375–390. https://doi.org/10.1038/s41579-020-00507-2
3. Schön ME, Martijn J, Vosseberg J, Köstlbacher S, Ettema TJG. 2022. The evolutionary origin of host association in the Rickettsiales. Nat Microbiol 7:1189–1199. https://doi.org/10.1038/s41564-022-01169-x
4. Castelli M, Sabaneyeva E, Lanzoni O, Lebedeva N, Floriano AM, Gaiarsa S, Benken K, Modeo L, Bandi C, Potekhin A, Sassera D, Petroni G. 2019. Deianiraea, an extracellular bacterium associated with the ciliate paramecium, suggests an alternative scenario for the evolution of Rickettsiales. ISME J 13:2280–2294. https://doi.org/10.1038/s41396-019-0433-9

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

5. Ettema T, Schön ME, JoranM, ulianV. 2022. The evolutionary origin of host association in an ancient bacterial clade. https://doi.org/10.1038/s41564-022-01169-x
6. Weitzel T, Dittrich S, López J, Phuklia W, Martinez-Valdebenito C, Velásquez K, Blacksell SD, Paris DH, Abarca K. 2016. Endemic scrub typhus in South America. N Engl J Med 375:954–961. https://doi.org/10.1056/NEJMoa1603657
7. Izzard L, Fuller A, Blacksell SD, Paris DH, Richards AL, Aukkanit N, Nguyen C, Jiang J, Fenwick S, Day NPJ, Graves S, Stenos J. 2010. Isolation of a novel *Orientia* species (*O. chuto* sp. nov.) from a patient infected in Dubai. J Clin Microbiol 48:4404–4409. https://doi.org/10.1128/JCM.01526-10
8. Cho N-H, Kim H-R, Lee J-H, Kim S-Y, Kim J, Cha S, Kim S-Y, Darby AC, Fuxelius H-H, Yin J, Kim JH, Kim J, Lee SJ, Koh Y-S, Jang W-J, Park K-H, Andersson SGE, Choi M-S, Kim I-S. 2007. The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. Proc Natl Acad Sci U S A 104:7981–7986. https://doi.org/10.1073/pnas.0611553104
9. Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, Fukuhara M, Urakami H, Ohnishi M, Uchiyama I, Ogura Y, Ooka T, Oshima K, Tamura A, Hattori M, Hayashi T. 2008. The whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. DNA Res 15:185–199. https://doi.org/10.1093/dnares/dsn011
10. Fleshman A, Mullins K, Sahl J, Hepp C, Nieto N, Wiggins K, Hornstra H, Kelly D, Chan T-C, Phetsouvanh R, Dittrich S, Panyanivong P, Paris D, Newton P, Richards A, Pearson T. 2018. Comparative pan-genomic analyses of *Orientia tsutsugamushi* reveal an exceptional model of bacterial evolution driving genomic diversity. Microb Genom 4:e000230. https://doi.org/10.1099/mgen.0.000230
11. Batty EM, Chaemchuen S, Blacksell SD, Paris D, Bowden R, Chan C, Lachumanan R, Day N, Donnelly P, Chen SL, Salje J. 2018. Long-read whole genome sequencing and comparative analysis of six strains of the human pathogen *Orientia tsutsugamushi*. Genomics. https://doi.org/10.1101/280958
12. Gillespie JJ, Joardar V, Williams KP, Driscoll T, Hostetler JB, Nordberg E, Shukla M, Walenz B, Hill CA, Nene VM, Azad AF, Sobral BW, Caler E. 2012. A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. J Bacteriol 194:376–394. https://doi.org/10.1128/JB.06244-11
13. Ogata H, La Scola B, Audic S, Renesto P, Blanc G, Robert C, Fournier P-E, Claverie J-M, Raoult D. 2006. Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. PLoS Genet 2:e76. https://doi.org/10.1371/journal.pgen.0020076
14. Blanc G, Ogata H, Robert C, Audic S, Claverie J-M, Raoult D. 2007. Lateral gene transfer between obligate intracellular bacteria: evidence from the *Rickettsia massiliae* genome. Genome Res 17:1657–1664. https://doi.org/10.1101/gr.6742107
15. Londoño AF, Mendell NL, Valbuena GA, Routh AL, Wood TG, Widen SG, Rodas JD, Walker DH, Bouyer DH. 2019. Whole-genome sequence of *Rickettsia parkeri* strain atlantic rainforest, isolated from a Colombian tick. Microbiol Resour Announc 8:e00684-19. https://doi.org/10.1128/MRA.00684-19
16. Gillespie JJ, Driscoll TP, Verhoeve VI, Utsuki T, Husseneder C, Chouljenko VN, Azad AF, Macaluso KR. 2014. Genomic diversification in strains of *Rickettsia felis* isolated from different arthropods. Genome Biol Evol 7:35–56. https://doi.org/10.1093/gbe/evu262
17. Felsheim RF, Kurtti TJ, Munderloh UG. 2009. Genome sequence of the endosymbiont *Rickettsia peacockii* and comparison with virulent *Rickettsia rickettsii*: identification of virulence factors. PLoS One 4:e8361. https://doi.org/10.1371/journal.pone.0008361
18. Hagen R, Verhoeve VI, Gillespie JJ, Driscoll TP, Martinez-Romero E. 2018. Conjugative transposons and their cargo genes vary across natural populations of *Rickettsia buchneri* infecting the tick *Ixodes scapularis*. Genome Biol Evol 10:3218–3229. https://doi.org/10.1093/gbe/evy247
19. Mika-Gospodorz B, Giengkam S, Westermann AJ, Wongsantichon J, Kion-Crosby W, Chuenklin S, Wang LC, Sunyakumthorn P, Sobota RM, Subbian S, Vogel J, Barquist L, Salje J. 2020. Dual RNA-seq of *Orientia tsutsugamushi* informs on host-pathogen interactions for this neglected intracellular human pathogen. Nat Commun 11:3363. https://doi.org/10.1038/s41467-020-17094-8
20. Nakayama K, Kurokawa K, Fukuhara M, Urakami H, Yamamoto S, Yamazaki K, Ogura Y, Ooka T, Hayashi T. 2010. Genome comparison and phylogenetic analysis of *Orientia tsutsugamushi* strains. DNA Res 17:281–291. https://doi.org/10.1093/dnares/dsq018
21. Atwal S, Wongsantichon J, Giengkam S, Saharat K, Pittayasathornthun YJ, Chuenklin S, Wang LC, Chung T, Huh H, Lee S-H, Sobota RM, Salje J. 2022. The obligate intracellular bacterium *Orientia tsutsugamushi* differentiates into a developmentally distinct extracellular state. Nat Commun 13:3603. https://doi.org/10.1038/s41467-022-31176-9
22. Atkinson GC, Tenson T, Hauryliuk V. 2011. The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. PLoS One 6:e23479. https://doi.org/10.1371/journal.pone.0023479
23. Hauryliuk V, Atkinson GC, Murakami KS, Tenson T, Gerdes K. 2015. Recent functional insights into the role of (p)ppGpp in bacterial physiology. Nat Rev Microbiol 13:298–309. https://doi.org/10.1038/nrmicro3448
24. Mosavi LK, Minor DL Jr, Peng Z-Y. 2002. Consensus-derived structural determinants of the ankyrin repeat motif. Proc Natl Acad Sci U S A 99:16029–16034. https://doi.org/10.1073/pnas.252537899
25. Jernigan KK, Bordenstein SR. 2014. Ankyrin domains across the tree of life. PeerJ 2:e264. https://doi.org/10.7717/peerj.264
26. Frank AC. 2019. Molecular host mimicry and manipulation in bacterial symbionts. FEMS Microbiol Lett 366:fnz038. https://doi.org/10.1093/femsle/fnz038
27. VieBrock L, Evans SM, Beyer AR, Larson CL, Beare PA, Ge H, Singh S, Rodino KG, Heinzen RA, Richards AL, Carlyon JA. 2014. *Orientia tsutsugamushi* ankyrin repeat-containing protein family members are type 1 secretion system substrates that traffic to the host cell endoplasmic reticulum. Front Cell Infect Microbiol 4:186. https://doi.org/10.3389/fcimb.2014.00186
28. Beyer AR, VieBrock L, Rodino KG, Miller DP, Tegels BK, Marconi RT, Carlyon JA. 2015. *Orientia tsutsugamushi* strain Ikeda ankyrin repeat-containing proteins recruit SCF1 ubiquitin ligase machinery via poxvirus-like F-box motifs. J Bacteriol 197:3097–3109. https://doi.org/10.1128/JB.00276-15
29. Beyer AR, Rodino KG, VieBrock L, Green RS, Tegels BK, Oliver LD, Marconi RT, Carlyon JA. 2017. *Orientia tsutsugamushi* Ank9 is a multifunctional effector that utilizes a novel GRIP-like Golgi localization domain for Golgi-to-endoplasmic reticulum trafficking and interacts with host COPB2. Cell Microbiol 19. https://doi.org/10.1111/cmi.12727
30. Evans SM, Rodino KG, Adcox HE, Carlyon JA. 2018. *Orientia tsutsugamushi* uses two Ank effectors to modulate NF-κB P65 nuclear transport and inhibit NF-κB transcriptional activation. PLoS Pathog 14:e1007023. https://doi.org/10.1371/journal.ppat.1007023
31. Adcox HE, Hatke AL, Andersen SE, Gupta S, Otto NB, Weber MM, Marconi RT, Carlyon JA. 2021. *Orientia tsutsugamushi* nucleomodulin Ank13 exploits the radar nuclear import pathway to modulate host cell transcription. mBio 12:e0181621. https://doi.org/10.1128/mBio.01816-21
32. Kipreos ET, Pagano M. 2000. The F-box protein family. Genome Biol 1:REVIEWS3002. https://doi.org/10.1186/gb-2000-1-5-reviews3002
33. Perez-Riba A, Itzhaki LS. 2019. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. Curr Opin Struct Biol 54:43–49. https://doi.org/10.1016/j.sbi.2018.12.004
34. Bang S, Min C-K, Ha N-Y, Choi M-S, Kim I-S, Kim Y-S, Cho N-H. 2016. Inhibition of eukaryotic translation by tetratricopeptide-repeat proteins of *Orientia tsutsugamushi*. J Microbiol 54:136–144. https://doi.org/10.1007/s12275-016-5599-5
35. Kaur SJ, Rahman MS, Ammerman NC, Beier-Sexton M, Ceraul SM, Gillespie JJ, Azad AF. 2012. TolC-dependent secretion of an ankyrin repeat-containing protein of *Rickettsia typhi*. J Bacteriol 194:4920–4932. https://doi.org/10.1128/JB.00793-12
36. Johnson CM, Grossman AD. 2015. Integrative and conjugative elements (ICEs): what they do and how they work. Annu Rev Genet 49:577–601. https://doi.org/10.1146/annurev-genet-112414-055018

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

37. Lang AS, Beatty JT. 2007. Importance of widespread gene transfer agent genes in alpha-proteobacteria. Trends Microbiol 15:54–62. https://doi.org/10.1016/j.tim.2006.12.001
38. Belfort M, Lambowitz AM. 2019. Group II intron RnpS and reverse transcriptases: from retroelements to research tools. Cold Spring Harb Perspect Biol 11:a032375. https://doi.org/10.1101/cshperspect.a032375
39. Siguier P, Gourbeyre E, Chandler M. 2014. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev 38:865–891. https://doi.org/10.1111/1574-6976.12067
40. Bordenstein SR, Bordenstein SR. 2022. Widespread phages of endosymbionts: phage WO genomics and the proposed taxonomic classification of Symbioviridae. PLoS Genet 18:e1010227. https://doi.org/10.1371/journal.pgen.1010227
41. Gillespie JJ, Kaur SJ, Rahman MS, Rennoll-Bankert K, Sears KT, Beier-Sexton M, Azad AF. 2015. Secretome of obligate intracellular *Rickettsia*. FEMS Microbiol Rev 39:47–80. https://doi.org/10.1111/1574-6976.12084
42. Lin M, Zhang C, Gibson K, Rikihisa Y. 2009. Analysis of complete genome sequence of *Neorickettsia risticii*: causative agent of potomac horse fever. Nucleic Acids Res 37:6076–6091. https://doi.org/10.1093/nar/gkp642
43. Gillespie JJ, Brayton KA, Williams KP, Diaz MAQ, Brown WC, Azad AF, Sobral BW. 2010. Phylogenomics reveals a diverse Rickettsiales type IV secretion system. Infect Immun 78:1809–1823. https://doi.org/10.1128/IAI.01384-09
44. Letunic I, Khedkar S, Bork P. 2021. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res 49:D458–D460. https://doi.org/10.1093/nar/gkaa937
45. Matson SW, Ragonese H. 2005. The F-plasmid Trai protein contains three functional domains required for conjugative DNA strand transfer. J Bacteriol 187:697–706. https://doi.org/10.1128/JB.187.2.697-706.2005
46. Scherzinger E, Lurz R, Otto S, Dobrinski B. 1992. *In vitro* cleavage of double- and single-stranded DNA by plasmid RSF1010-encoded mobilization proteins. Nucleic Acids Res 20:41–48. https://doi.org/10.1093/nar/20.1.41
47. Escobar MA, Dandekar AM. 2003. Agrobacterium tumefaciens as an agent of disease. Trends Plant Sci 8:380–386. https://doi.org/10.1016/S1360-1385(03)00162-6
48. Weinert LA, Welch JJ, Jiggins FM. 2009. Conjugation genes are common throughout the genus *Rickettsia* and are transmitted horizontally. Proc Biol Sci 276:3619–3627. https://doi.org/10.1098/rspb.2009.0875
49. Gillespie JJ, Ammerman NC, Dreher-Lesnick SM, Rahman MS, Worley MJ, Setubal JC, Sobral BS, Azad AF. 2009. An anomalous type IV secretion system in *Rickettsia* is evolutionarily conserved. PLoS One 4:e4833. https://doi.org/10.1371/journal.pone.0004833
50. Gillespie JJ, Phan IQH, Driscoll TP, Guillotte ML, Lehman SS, Rennoll-Bankert KE, Subramanian S, Beier-Sexton M, Myler PJ, Rahman MS, Azad AF. 2016. The *Rickettsia* type IV secretion system: unrealized complexity mired by gene family expansion. Pathog Dis 74:ftw058. https://doi.org/10.1093/femspd/ftw058
51. Gillespie JJ, Phan IQH, Scheib H, Subramanian S, Edwards TE, Lehman SS, Piitulainen H, Rahman MS, Rennoll-Bankert KE, Staker BL, Taira S, Stacy R, Myler PJ, Azad AF, Pulliainen AT. 2015. Structural insight into how bacteria prevent interference between multiple divergent type IV secretion systems. mBio 6:e01867–15. https://doi.org/10.1128/mBio.01867-15
52. Verhoeve VI, Gillespie JJ. 2022. Origin of rickettsial host dependency unravelled. Nat Microbiol 7:1110–1111. https://doi.org/10.1038/s41564-022-01187-9
53. Sutten EL, Norimine J, Beare PA, Heinzen RA, Lopez JE, Morse K, Brayton KA, Gillespie JJ, Brown WC. 2010. Anaplasma marginale type IV secretion system proteins VirB2, VirB7, VirB11, and VirD4 are immunogenic components of a protective bacterial membrane vaccine. Infect Immun 78:1314–1325. https://doi.org/10.1128/IAI.01207-09
54. Oyler BL, Rennoll-Bankert KE, Verhoeve VI, Rahman MS, Azad AF, Gillespie JJ, Goodlett DR. 2023. *Rickettsia typhi* peptidoglycan mapping with data-dependent tandem mass spectrometry. Microbiology. https://doi.org/10.1101/2022.03.10.483817
55. Figueroa-Cuilan WM, Irazoki O, Feeley M, Smith E, Nguyen T, Cava F, Goley ED. 2023. Quantitative analysis of morphogenesis and growth dynamics in an obligate intracellular bacterium. Mol Biol Cell 34:ar69. https://doi.org/10.1091/mbc.E23-01-0023
56. Atwal S, Giengkam S, Chaemchuen S, Dorling J, Kosaisawe N, VanNieuwenhze M, Sampattavanich S, Schumann P, Salje J. 2017. Evidence for a peptidoglycan-like structure in *Orientia tsutsugamushi*. Mol Microbiol 105:440–452. https://doi.org/10.1111/mmi.13709
57. Atwal S, Chuenklin S, Bonder EM, Flores J, Gillespie JJ, Driscoll TP, Salje J. 2021. Discovery of a diverse set of bacteria that build their cell walls without the canonical peptidoglycan polymerase aPBP. mBio 12:e0134221. https://doi.org/10.1128/mBio.01342-21
58. Gillespie JJ, Salje J. 2023. *Orientia* and *Rickettsia*: different flowers from the same garden. Curr Opin Microbiol 74:102318. https://doi.org/10.1016/j.mib.2023.102318
59. Rennoll-Bankert KE, Rahman MS, Guillotte ML, Lehman SS, Beier-Sexton M, Gillespie JJ, Azad AF. 2016. RalF-mediated activation of Arf6 controls *Rickettsia typhi* invasion by co-opting phosphoinositol metabolism. Infect Immun 84:3496–3506. https://doi.org/10.1128/IAI.00638-16
60. Lehman SS, Noriea NF, Aistleitner K, Clark TR, Dooley CA, Nair V, Kaur SJ, Rahman MS, Gillespie JJ, Azad AF, Hackstadt T. 2018. The rickettsial ankyrin repeat protein 2 is a type IV secreted effector that associates with the endoplasmic reticulum. mBio 9:e00975–00918. https://doi.org/10.1128/mBio.00975-18
61. Voss OH, Gillespie JJ, Lehman SS, Rennoll SA, Beier-Sexton M, Rahman MS, Azad AF. 2020. Risk1, a phosphatidylinositol 3-kinase effector, promotes *Rickettsia typhi* intracellular survival. mBio 11:e00820-20. https://doi.org/10.1128/mBio.00820-20
62. Aistleitner K, Clark T, Dooley C, Hackstadt T. 2020. Selective fragmentation of the trans-Golgi apparatus by *Rickettsia rickettsii*. PLoS Pathog 16:e1008582. https://doi.org/10.1371/journal.ppat.1008582
63. Lockwood S, Voth DE, Brayton KA, Beare PA, Brown WC, Heinzen RA, Broschat SL. 2011. Identification of Anaplasma marginale type IV secretion system effector proteins. PLoS One 6:e27724. https://doi.org/10.1371/journal.pone.0027724
64. Lin M, den Dulk-Ras A, Hooykaas PJJ, Rikihisa Y. 2007. *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. Cell Microbiol 9:2644–2657. https://doi.org/10.1111/j.1462-5822.2007.00985.x
65. Niu H, Kozjak-Pavlovic V, Rudel T, Rikihisa Y. 2010. *Anaplasma phagocytophilum* Ats-1 is imported into host cell mitochondria and interferes with apoptosis induction. PLoS Pathog 6:e1000774. https://doi.org/10.1371/journal.ppat.1000774
66. Rikihisa Y, Lin M. 2010. *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins. Curr Opin Microbiol 13:59–66. https://doi.org/10.1016/j.mib.2009.12.008
67. Rikihisa Y, Lin M, Niu H. 2010. Type IV secretion in the obligatory intracellular bacterium *Anaplasma phagocytophilum*. Cell Microbiol 12:1213–1221. https://doi.org/10.1111/j.1462-5822.2010.01500.x
68. Park JM, Genera BM, Fahy D, Swallow KT, Nelson CM, Oliver JD, Shaw DK, Munderloh UG, Brayton KA. 2023. An *Anaplasma phagocytophilum* T4SS effector, AteA, is essential for tick infection. bioRxiv:2023.02.06.527355. https://doi.org/10.1101/2023.02.06.527355
69. Zhu J, He M, Xu W, Li Y, Huang R, Wu S, Niu H. 2019. Development of TEM-1 beta-lactamase based protein translocation assay for identification of *Anaplasma phagocytophilum* type IV secretion system effector proteins. Sci Rep 9:4235. https://doi.org/10.1038/s41598-019-40682-8
70. Kim Y, Wang J, Clemens EG, Grab DJ, Dumler JS. 2022. *Anaplasma phagocytophilum* Ankyrin A protein (AnkA) enters the nucleus using an importin-beta-, RanGTP-dependent mechanism. Front Cell Infect Microbiol 12:828605. https://doi.org/10.3389/fcimb.2022.828605
71. Liu H, Bao W, Lin M, Niu H, Rikihisa Y. 2012. *Ehrlichia* type IV secretion effector ECH0825 is translocated to mitochondria and curbs ROS and apoptosis by upregulating host MnSOD. Cell Microbiol 14:1037–1050. https://doi.org/10.1111/j.1462-5822.2012.01775.x
72. Rikihisa Y. 2021. "The "biological weapons" of *Ehrlichia chaffeensis*: novel molecules and mechanisms to subjugate host cells". Front Cell Infect Microbiol 11:830180. https://doi.org/10.3389/fcimb.2021.830180
73. Yan Q, Zhang W, Lin M, Teymournejad O, Budachetri K, Lakritz J, Rikihisa Y. 2021. Iron robbery by intracellular pathogen via bacterial effector-induced ferritinophagy. Proc Natl Acad Sci U S A 118:e2026598118. https://doi.org/10.1073/pnas.2026598118

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

74. Yan Q, Lin M, Huang W, Teymournejad O, Johnson JM, Hays FA, Liang Z, Li G, Rikihisa Y. 2018. *Ehrlichia* type IV secretion system effector Etf-2 binds to active RAB5 and delays endosome maturation. Proc Natl Acad Sci U S A 115:E8977–E8986. https://doi.org/10.1073/pnas.1806904115
75. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, Bernheim A. 2022. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun 13:2561. https://doi.org/10.1038/s41467-022-30269-9
76. Rocha EPC, Bikard D. 2022. Microbial defenses against mobile genetic elements and viruses: who defends whom from what? PLoS Biol 20:e3001514. https://doi.org/10.1371/journal.pbio.3001514
77. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359:eaar4120. https://doi.org/10.1126/science.aar4120
78. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. 2014. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. PLoS One 9:e110726. https://doi.org/10.1371/journal.pone.0110726
79. Tal N, Morehouse BR, Millman A, Stokar-Avihail A, Avraham C, Fedorenko T, Yirmiya E, Herbst E, Brandis A, Mehlman T, Oppenheimer-Shaanan Y, Keszei AFA, Shao S, Amitai G, Kranzusch PJ, Sorek R. 2021. Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. Cell 184:5728–5739. https://doi.org/10.1016/j.cell.2021.09.031
80. Sonthayanon P, Peacock SJ, Chierakul W, Wuthiekanun V, Blacksell SD, Holden MTG, Bentley SD, Feil EJ, Day NPJ. 2010. High rates of homologous recombination in the mite endosymbiont and opportunistic human pathogen *Orientia tsutsugamushi*. PLoS Negl Trop Dis 4:e752. https://doi.org/10.1371/journal.pntd.0000752
81. Song Y, Peisach D, Pioszak AA, Xu Z, Ninfa AJ. 2004. Crystal structure of the C-terminal domain of the two-component system transmitter protein nitrogen regulator II (NRII; NtrB), regulator of nitrogen assimilation in *Escherichia coli*. Biochemistry 43:6670–6678. https://doi.org/10.1021/bi049474r
82. Bellon S, Parsons JD, Wei Y, Hayakawa K, Swenson LL, Charifson PS, Lippke JA, Aldape R, Gross CH. 2004. Crystal structures of *Escherichia coli* topoisomerase IV ParE subunit (24 and 43 kilodaltons): a single residue dictates differences in novobiocin potency against topoisomerase IV and DNA gyrase. Antimicrob Agents Chemother 48:1856–1864. https://doi.org/10.1128/AAC.48.5.1856-1864.2004
83. Wright L, Barril X, Dymock B, Sheridan L, Surgenor A, Beswick M, Drysdale M, Collier A, Massey A, Davies N, Fink A, Fromont C, Aherne W, Boxall K, Sharp S, Workman P, Hubbard RE. 2004. Structure-activity relationships in purine-based inhibitor binding to HSP90 Isoforms. Chem Biol 11:775–785. https://doi.org/10.1016/j.chembiol.2004.03.033
84. Mascher T, Helmann JD, Unden G. 2006. Stimulus perception in bacterial signal-transducing histidine kinases. Microbiol Mol Biol Rev 70:910–938. https://doi.org/10.1128/MMBR.00020-06
85. AtkinsonGemma C, TanelT, VasiliH. 2011. The RelA/SpoT Homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. Available from: https://doi.org/10.1371/journal.pone.0023479
86. Prusińska JM, Boniecka J, Dąbrowska GB, Goc A, AnnaGOC. 2019. Identification and characterization of the *Ipomoea nil Rela/spot Homologs (Inrshs)* and potential directions of their transcriptional regulation. Plant Sci 284:161–176. https://doi.org/10.1016/j.plantsci.2019.01.026
87. Gratani FL, Horvatek P, Geiger T, Borisova M, Mayer C, Grin I, Wagner S, Steinchen W, Bange G, Velic A, Maček B, Wolz C. 2018. Regulation of the opposing (P)ppGpp synthetase and Hydrolase activities in a Bifunctional Rela/spot Homologue from *Staphylococcus Aureus*. PLoS Genet 14:e1007514. https://doi.org/10.1371/journal.pgen.1007514
88. Takada H, Roghanian M, Murina V, Dzhygyr I, Murayama R, Akanuma G, Atkinson GC, Garcia-Pino A, Hauryliuk V. 2020. "The C-terminal RRM/ACT domain is crucial for fine-tuning the activation of 'long' Rela/spot Homolog enzymes by Ribosomal complexes". Front Microbiol 11:277. https://doi.org/10.3389/fmicb.2020.00277
89. Jain V, Saleem-Batcha R, China A, Chatterji D. 2007. Molecular dissection of the Mycobacterial stringent response protein Rel. Protein Sci 15:1449–1464. https://doi.org/10.1110/ps.062117006
90. Hagemann M, Gärtner K, Scharnagl M, Bolay P, Lott SC, Fuss J, Huettel B, Reinhardt R, Klähn S, Hess WR. 2018. Identification of the DNA methyltransferases establishing the methylome of the cyanobacterium synechocystis sp. PCC 6803. DNA Res 25:343–352. https://doi.org/10.1093/dnares/dsy006
91. Leipe DD, Aravind L, Grishin NV, Koonin EV. 2000. The bacterial replicative helicase DnaB evolved from a RecA duplication. Genome Res. 10:5–16.
92. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32–6. https://doi.org/10.1093/nar/gkj014
93. Groth AC, Calos MP. 2004. Phage integrases: biology and applications. J Mol Biol 335:667–678. https://doi.org/10.1016/j.jmb.2003.09.082
94. Abremski KE, Hoess RH. 1992. Evidence for a second conserved arginine residue in the integrase family of recombination proteins. Protein Eng Des Sel 5:87–91. https://doi.org/10.1093/protein/5.1.87
95. Gladyshev EA, Arkhipova IR. 2011. A widespread class of reverse transcriptase-related cellular genes. Proc Natl Acad Sci U S A 108:20311–20316. https://doi.org/10.1073/pnas.1100266108
96. Blocker FJH, Mohr G, Conlan LH, Qi L, Belfort M, Lambowitz AM. 2005. Domain structure and three-dimensional model of a group II Intron-encoded reverse transcriptase. RNA 11:14–28. https://doi.org/10.1261/rna.7181105
97. Zhao J, Lambowitz AM. 2005. A bacterial group II intron-encoded reverse transcriptase localizes to cellular poles. Proc Natl Acad Sci U S A 102:16133–16140. https://doi.org/10.1073/pnas.0507057102
98. Nesmelova IV, Hackett PB. 2010. DDE transposases: structural similarity and diversity. Adv Drug Deliv Rev 62:1187–1195. https://doi.org/10.1016/j.addr.2010.06.006
99. Knizewski L, Kinch LN, Grishin NV, Rychlewski L, Ginalski K. 2007. Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. BMC Struct Biol 7:40. https://doi.org/10.1186/1472-6807-7-40
100. Davies DR, Goryshin IY, Reznikoff WS, Rayment I. 2000. Three-dimensional structure of the Tn 5 synaptic complex transposition intermediate. Science 289:77–85. https://doi.org/10.1126/science.289.5476.77
101. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting Transmembrane protein Topology with a hidden Markov model: Application to complete Genomes. Journal of Molecular Biology 305:567–580. https://doi.org/10.1006/jmbi.2000.4315
102. Lai EM, Kado CI. 2000. The T-Pilus of *Agrobacterium Tumefaciens*. Trends Microbiol 8:361–369. https://doi.org/10.1016/s0966-842x(00)01802-3
103. Gillespie JJ, Ammerman NC, Dreher-Lesnick SM, Rahman MS, Worley MJ, Setubal JC, Sobral BS, Azad AF. 2009. An anomalous type IV secretion system in *Rickettsia* is Evolutionarily conserved. PLoS One 4:e4833. https://doi.org/10.1371/journal.pone.0004833
104. Cao TB, Saier MH. 2001. Conjugal type IV macromolecular transfer systems of gram-negative bacteria: Organismal distribution, structural constraints and evolutionary conclusions. Microbiology (Reading) 147:3201–3214. https://doi.org/10.1099/00221287-147-12-3201
105. Gillespie JJ, Phan IQH, Driscoll TP, Guillotte ML, Lehman SS, Rennoll-Bankert KE, Subramanian S, Beier-Sexton M, Myler PJ, Rahman MS, Azad AF. 2016. The Rickettsia type IV secretion system: Unrealized complexity Mired by gene family expansion. Pathog Dis 74:ftw058. https://doi.org/10.1093/femspd/ftw058
106. Judd PK, Mahli D, Das A. 2005. Molecular characterization of the Agrobacterium tumefaciens DNA transfer protein Virb6. Microbiology (Reading) 151:3483–3492. https://doi.org/10.1099/mic.0.28337-0
107. Lawley TD, Klimke WA, Gubbins MJ, Frost LS. 2003. F factor conjugation is a true type IV secretion system. FEMS Microbiol Lett 224:1–15. https://doi.org/10.1016/S0378-1097(03)00430-0
108. Bailey S, Ward D, Middleton R, Grossmann JG, Zambryski PC. 2006. *Agrobacterium tumefaciens* Virb8 structure reveals potential protein–protein interaction sites. Proc Natl Acad Sci U S A 103:2582–2587. https://doi.org/10.1073/pnas.0511216103
109. Terradot L, Bayliss R, Oomen C, Leonard GA, Baron C, Waksman G. 2005. Structures of two core subunits of the bacterial type IV secretion system,Virb8 from *Brucella Suis* and Comb10 from *Helicobacter Pylori*. Proc Natl Acad Sci U S A 102:4596–4601. https://doi.org/10.1073/pnas.0408927102

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.

110. Banta LM, Kerr JE, Cascales E, Giuliano ME, Bailey ME, McKay C, Chandran V, Waksman G, Christie PJ. 2011. An *Agrobacterium* Virb10 Mutation conferring a type IV secretion system gating defect. J Bacteriol 193:2566–2574. https://doi.org/10.1128/JB.00038-11
111. Ripoll-Rozada J, Zunzunegui S, de la Cruz F, Arechaga I, Cabezón E. 2013. Functional interactions of Virb11 traffic Atpases with Virb4 and Vird4 molecular Motors in type IV secretion systems. J Bacteriol 195:4195–4201. https://doi.org/10.1128/JB.00437-13
112. Leloup L, Lai E-M. 2002. Identification of a Chromosomal Tra-like region in Agrobacterium tumefaciens 267:115–123. https://doi.org/10.1007/s00438-002-0646-9
113. Kumar RB, Das A. 2002. Polar location and functional domains of the *Agrobacterium Tumefaciens* DNA transfer protein Vird4. Mol Microbiol 43:1523–1532. https://doi.org/10.1046/j.1365-2958.2002.02829.x
114. Frost LS, Ippen-Ihler K, Skurray RA. 1994. Analysis of the sequence and gene products of the transfer region of the F sex factor. Microbiol Rev 58:162–210. https://doi.org/10.1128/mr.58.2.162-210.1994
115. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053
116. Bragagnolo N, Rodriguez C, Samari-Kermani N, Fours A, Korouzhdehi M, Lysenko R, Audette GF. 2020. Protein Dynamics in F-like bacterial conjugation. Biomedicines 8:362. https://doi.org/10.3390/biomedicines8090362
117. Gilmour MW, Gunton JE, Lawley TD, Taylor DE. 2003. Interaction between the inchi1 plasmid R27 coupling protein and type IV secretion system: TraG associates with the coiled-coil mating pair formation protein TrhB. Mol Microbiol 49:105–116. https://doi.org/10.1046/j.1365-2958.2003.03551.x
118. Gillespie JJ, Phan IQH, Driscoll TP, Guillotte ML, Lehman SS, Rennoll-Bankert KE, Subramanian S, Beier-Sexton M, Myler PJ, Rahman MS, Azad AF. 2016. The *Rickettsia* type IV secretion system: Unrealized complexity Mired by gene family expansion. Pathog Dis 74:ftw058. https://doi.org/10.1093/femspd/ftw058
119. Harris RL, Hombs V, Silverman PM. 2001. Evidence that F-Plasmid proteins Trav, Trak and Trab assemble into an envelope-spanning structure in *Escherichia Coli*. Mol Microbiol 42:757–766. https://doi.org/10.1046/j.1365-2958.2001.02667.x
120. Harris RL, Silverman PM. 2002. Roles of internal Cysteines in the function, localization, and reactivity of the Trav outer membrane lipoprotein encoded by the F Plasmid. J Bacteriol 184:3126–3129. https://doi.org/10.1128/JB.184.11.3126-3129.2002
121. Gillespie JJ, Ammerman NC, Dreher-Lesnick SM, Rahman MS, Worley MJ, Setubal JC, Sobral BS, Azad AF. 2009. An anomalous type IV secretion system in *Rickettsia* is Evolutionarily conserved. PLoS One 4:e4833. https://doi.org/10.1371/journal.pone.0004833
122. Lawley TD, Klimke WA, Gubbins MJ, Frost LS. 2003. F factor conjugation is a true type IV secretion system. FEMS Microbiol Lett 224:1–15. https://doi.org/10.1016/S0378-1097(03)00430-0
123. Shala-Lawrence A, Bragagnolo N, Nowroozi-Dayeni R, Kheyson S, Audette GF. 2018. The interaction of Traw and Trbc is required to facilitate conjugation in F-like Plasmids. Biochem Biophys Res Commun 503:2386–2392. https://doi.org/10.1016/j.bbrc.2018.06.166
124. Moore D, Maneewannakul K, Maneewannakul S, Wu JH, Ippen-Ihler K, Bradley DE. 1990. Characterization of the F-Plasmid Conjugative transfer gene traU. J Bacteriol 172:4263–4270. https://doi.org/10.1128/jb.172.8.4263-4270.1990
125. Maneewannakul S, Maneewannakul K, Ippen-Ihler K. 1991. Characterization of trbC, a new F Plasmid Tra Operon gene that is essential to Conjugative transfer. J Bacteriol 173:3872–3878. https://doi.org/10.1128/jb.173.12.3872-3878.1991
126. Klimke WA, Rypien CD, Klinger B, Kennedy RA, Rodriguez-Maillard JM, Frost LS. 2005. The mating pair stabilization protein, Tran, of the F Plasmid is an outer-membrane protein with two regions that are important for its function in conjugation. Microbiology (Reading) 151:3527–3540. https://doi.org/10.1099/mic.0.28025-0
127. Elton TC, Holland SJ, Frost LS, Hazes B. 2005. F-like type IV secretion systems Encode proteins with Thioredoxin folds that are putative Dsbc Homologues. J Bacteriol 187:8267–8277. https://doi.org/10.1128/JB.187.24.8267-8277.2005
128. Denis A, Barbara A, Jan M. 2010. F Plasmid TRAF and Trah are components of an outer membrane complex involved in conjugation. J Bacteriol. https://doi.org/10.1128/jb.00726-09
129. Firth N, Skurray R. 1992. Characterization of the F Plasmid Bifunctional conjugation gene, traG. Mol Gen Genet 232:145–153. https://doi.org/10.1007/BF00299147
130. Lessl M, Pansegrau W, Lanka E. 1992. Relationship of DNA-transfer-systems: Essential transfer factors of Plasmids Rp4, TI and F share common sequences. Nucleic Acids Res 20:6099–6100. https://doi.org/10.1093/nar/20.22.6099
131. Farrand SK, Hwang I, Cook. DM. 1996. The tra region of the nopaline-type Ti plasmid is a chimera with elements related to the transfer systems of RSF1010, RP4, and F. https://doi.org/10.1128/jb.178.14.4233-4247.1996
132. Alt-Mörbe J, Stryker JL, Fuqua C, Li PL, Farrand SK, Winans SC. 1996. The conjugal transfer system of Agrobacterium tumefaciens Octopine-type TI Plasmids is closely related to the transfer system of an Incp Plasmid and distantly related to TI Plasmid VIR genes. J Bacteriol 178:4248–4257. https://doi.org/10.1128/jb.178.14.4248-4257.1996

Downloaded from https://journals.asm.org/journal/msphere on 03 November 2023 by 195.89.33.213.