Table S1. List of pseudogenes and split genes determined within the REIS genome. A total of 148 pseudogenes (colored black) are present as one ORF and truncated relative to larger genes encoded within other *Rickettsia* genomes. A total of 84 split genes (colored blue) are comprised of two or more ORFs comprising 217 total fragments (avg. 2.6 fragments per split gene). For split genes, asterisks denote genes splintered by insertion sequences. All CDS are listed by coordinates for the chromosome and four plasmids.

Chromosome

PG	REIS	start	stop	Annotation
1	0022	2890	3819	pyridine nucleotide-disulfide oxidoreductase family protein
	0023	4011	5072	putative bifunctional glutamate synthase subunit beta/2-polyprenylphenol hydroxylase
2	0026	8729	8100	acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase
3	0031	11101	11316	conserved hypothetical protein, authentic frameshift
4	0055	29465	29812	guanosine polyphosphate pyrophosphohydrolase/synthetase, putative frameshift
5	0057	35356	31214	outer membrane protein OmpB precursor, putative frameshift
6	0060	39564	38977	rickettsial palindromic element domain protein, putative frameshift
7	0066	44180	43089	periplasmic protein
	0067	44529	44284	protein involved in external DNA uptake
8	0069	48011	46347	DNA ligase, NAD-dependent
9	0075	51461	52192	guanosine polyphosphate pyrophosphohydrolase/synthetase, putative frameshift
10	0084	57752	58498	N-acetylmuramoyl-L-alanine amidase AmiD, authentic point mutation
11	0088	60671	60024	integrase
12	0094	66429	65464	methionine adenosyltransferase, degenerate
13	0109	78159	79315	ATPase, putative frameshift
14	0113	82378	81350	ABC transporter ATP-binding protein Uup
	0114	83199	82591	ABC transporter ATP-binding protein Uup
15	0143	108251	107347	microcin C7 resistance protein, authentic frameshift
16	0145	111742	111608	conserved hypothetical protein
17	0151	115588	116688	transposase, Mutator family, putative frameshift
18	0161	123606	124200	transposase, putative point mutation
19	0173	135968	133863	conserved hypothetical protein, putative point mutation

PG	REIS	start	stop	Annotation
20	0177	139281	138724	bicyclomycin resistance protein, putative frameshift
21	0178	139737	139564	conserved hypothetical protein
22	0179	141219	140050	export ATP-binding/permease protein
	0180	142185	141723	putative export ATP-binding/permease protein
23	0183	142799	143401	queuosine biosynthesis protein QueC, putative frameshift
24	0184	143906	146830	bacterial NAD-glutamate dehydrogenase family protein
25	0189	148260	148622	proline/betaine transporter
26	0190	149376	148880	putative oxidoreductase protein, putative frameshift
27	0202	155519	156271	conserved hypothetical protein
	0203	156277	156591	conserved hypothetical protein
28*	0204	157101	157835	cell surface antigen Sca13, putative frameshift
	0205	158280	157840	transposase, putative point mutation
	0206	158307	158654	cell surface antigen Sca13
29	0207	159075	158716	transcriptional regulator, XRE family, putative point mutation
30	0208	159318	159072	cytotoxic translational repressor of toxin-antitoxin system, putative frameshift
31	0214	162721	163302	nucleotidyltransferase
32	0234	178767	176752	NADH-quinone oxidoreductase subunit G
33	0259	198424	198121	transcriptional regulator, putative point mutation
34	0271	220201	220422	O-linked GlcNAc transferase
35	0273	222637	221759	hydrophobe/amphiphile efflux-1 HAE1 family protein, putative point mutation
36	0309	249324	248421	conserved hypothetical protein, putative frameshift
37*	0326	258655	259311	large extracellular alpha-helical protein
	0327	259939	259499	transposase
	0328	260072	260293	large extracellular alpha-helical protein

PG	REIS	start	stop	Annotation
38	0330	261417	266090	polyketide synthase, type I, putative frameshift
39	0333	267551	267393	conjugal transfer protein TraD
40	0340	271679	271299	arginyl-tRNA synthetase
41	0345	273499	277559	conjugal transfer protein TraA, putative frameshift
42	0347	280524	278912	putative conjugative transfer protein TraD, authentic frameshift
43	0367	292301	291705	putative transcriptional regulator, putative frameshift
44	0386	308182	309139	conserved hypothetical protein, putative frameshift
45	0407	325472	325843	conserved hypothetical protein
	0408	326010	326120	conserved hypothetical protein
	0409	326527	326988	conserved hypothetical protein
	0410	326985	327275	conserved hypothetical protein
	0411	327597	327469	hypothetical protein
	0412	327648	327854	conserved hypothetical protein
	0413	327890	328204	conserved hypothetical protein
	0414	328285	328812	conserved hypothetical protein
46	0415	329038	329571	conserved hypothetical protein
47	0427	340107	339667	transposase, putative point mutation
48	0433	344553	345377	queuine synthase, putative frameshift
49	0439	350818	351078	conserved hypothetical protein
	0440	351109	351315	conserved hypothetical protein
50	0502	401984	398603	putative cell surface antigen Sca2, authentic frameshift
	0503	402833	402675	hypothetical protein
	0504	403279	402830	surface antigen
	0505	403563	403673	hypothetical protein
	0506	403847	403680	surface antigen
51	0520	415006	413500	conserved hypothetical protein, putative frameshift
52	0527	420444	419882	polysaccharide deacetylase, degenerate

PG	REIS	start	stop	Annotation
53	0530	422322	422155	poly-beta-hydroxybutyrate polymerase
54	0539	428178	428568	proline/betaine transporter, putative frameshift
	0541	428673	429584	proline/betaine transporter
55	0543	430841	430482	hypothetical protein
56	0547	433656	433075	DNA-directed RNA polymerase, beta' subunit
	0548	435954	434160	DNA-directed RNA polymerase, beta' subunit, putative frameshift
	0549	437882	437031	DNA-directed RNA polymerase, beta' subunit
57	0556	448589	446513	elongation factor EF-G, putative frameshift
58	0564	455493	454099	periplasmic serine protease, degenerate
59	0567	457518	457778	tRNA pseudouridine synthase B
60	0570	461580	460376	ClpB protein, degenerate
61	0572	462237	463049	conserved hypothetical protein, putative frameshift
62	0599	476532	478948	type IV secretion/conjugal transfer ATPase, VirB4 family, putative frameshift
63	0602	480377	483733	TrbL/VirB6 plasmid conjugal transfer protein, degenerate
	0603	483752	484097	TrbL/VirB6a plasmid conjugal transfer protein
64	0604	484312	485871	TrbL/VirB6 plasmid conjugative transfer protein, putative frameshift
65	0607	493155	495758	TrbL/VirB6 plasmid conjugal transfer protein, authentic point mutation
66	0625	505078	505650	dephospho-CoA kinase, putative point mutation
67	0631	511645	511544	beta-lactamase-like protein
68	0633	512731	512907	actin polymerization protein RickA (tail)
69*	0637	515475	515729	conserved hypothetical protein
	0638	515801	516241	transposase
	0639	516276	516647	conserved hypothetical protein
	0640	516753	517019	conserved hypothetical protein

PG	REIS	start	stop	Annotation
70	0644	521907	521500	conserved hypothetical protein
71	0657	531412	531861	proline/betaine transporter, degenerate
72*	0667	538885	538535	ankyrin repeat-containing domain protein
	0668	538896	539024	hypothetical protein
	0669	539692	539387	ankyrin repeat-containing domain protein
73	0673	541406	542585	conserved hypothetical protein, putative frameshift
	0674	542690	542929	conserved hypothetical protein
74	0692	554476	555479	AmpG, authentic frameshift
75	0694	556731	556850	hypothetical protein
	0695	557930	557056	rare lipoprotein A, putative frameshift
76	0709	568597	569619	conserved hypothetical protein, putative frameshift
77	0712	574125	575249	conserved hypothetical protein, degenerate
78	0715	578346	577825	conserved hypothetical protein, putative frameshift
79	0728	590306	589386	glycerol-3-phosphate dehydrogenase, authentic frameshift
80	0739	600219	600431	cytochrome b561
	0740	600498	600740	cytochrome b561 family protein
81	0779	633814	634113	conserved hypothetical protein
82	0824	669611	666892	putative conjugative transfer protein TraG, degenerate
83	0826	671711	671016	F pilus assembly protein TraF, degenerate
84	0827	673658	671955	conjugative transfer protein TraN, degenerate
85	0832	677994	676042	conjugative transfer protein TraC, authentic frameshift
86	0883	727428	729938	conjugative transfer protein TraC, authentic point mutation

PG	REIS	start	stop	Annotation
87*	0888	732396	733007	conjugative transfer protein TraN, interruption-N
	0889	733457	733017	transposase
	0890	733638	734674	conjugative transfer protein TraN, putative frameshift
88	0891	734825	735585	F pilus assembly protein TraF, putative frameshift
89*	0893	736998	739513	conjugative transfer protein TraG, authentic frameshift
	0894	739958	739518	transposase
	0895	740048	740254	conjugative transfer protein TraG
90	0899	747846	743704	conjugal transfer protein TraA, degenerate
91	0937	781877	781473	conserved hypothetical protein
92	0942	788813	788349	ABC transporter permease protein
93*	0945	790553	790425	ABC transporter substrate-binding protein
	0946	790653	791093	transposase
	0947	791393	791196	ABC transporter substrate-binding protein
94*	0958	801226	800873	putative phage terminase protein
	0959	802252	801314	transposase
	0960	802787	802368	phage uncharacterized protein
	0961	803125	802988	conserved hypothetical protein
95	0989	820268	819960	conserved hypothetical protein
	0990	820720	820232	putative hemagglutinin protein, putative point mutation
	0991	821166	820780	conserved hypothetical protein
	0992	821537	821181	YD repeat protein
96	0993	821647	824139	AsmA family protein, putative point mutation
97*	1006	833733	834480	efflux transporter, RND family, MFP subunit, putative frameshift
	1007	835259	834717	ISPg3, transposase
	1008	835373	835621	efflux transporter, RND family, MFP subunit
98	1011	839731	838397	conserved hypothetical protein, putative point mutation
99	1015	843052	842930	transglycosylase SLT family protein
100	1017	843604	845171	integral membrane protein, putative frameshift

PG	REIS	start	stop	Annotation
101	1024	849009	850034	putative DNA processing protein DprA, putative frameshift
102	1027	853926	853810	proline/betaine transporter
103	1034	858337	857894	conserved hypothetical protein
	1035	858645	858412	conserved hypothetical protei
104	1036	859092	858946	protein CyaY
105*	1037	859803	859228	aminodeoxychorismate lyase
	1038	860273	859833	transposase
	1039	860856	860296	aminodeoxychorismate lyase
106	1043	863384	863193	conserved hypothetical protein
	1044	863583	863461	conserved hypothetical protein
	1045	863833	863576	conserved hypothetical protein
107	1046	864135	863959	ChlI
	1047	865301	864105	competence protein ComM, putative frameshift
108	1070	885202	883424	penicillin-binding protein, putative frameshift
109	1074	888963	888841	proline/betaine transporter
110	1082	902823	902262	guanosine polyphosphate pyrophosphohydrolase/synthetase, putative frameshift
111	1097	920925	924963	conjugal transfer protein TraA, putative frameshift
112	1104	931137	929890	rickA
113*	1109	939572	937573	transcription-repair-coupling factor, putative frameshift
	1110	940372	939932	transposase
	1111	941258	940395	transcription-repair-coupling factor
114	1125	949651	950001	conserved hypothetical protein, putative frameshift
	1126	950254	950985	conserved hypothetical protein
	1127	951435	950995	transposase
	1128	951655	952047	conserved hypothetical protein
115	1138	971566	969993	succinate-semialdehyde dehydrogenase, degenerate

PG	REIS	start	stop	Annotation
116	1144	975684	975088	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase
117	1151	981192	980086	nacht domain protein, authentic point mutation
118	1152	982500	981238	transposase, putative frameshift
119	1153	984420	984022	delta-aminolevulinic acid dehydratase
	1155	985377	984525	delta-aminolevulinic acid dehydratase
120*	1157	986931	987197	conserved hypothetical protein
	1158	987250	987669	putative transposase
	1159	988781	989014	hypothetical protein
	1161	989102	989530	conserved hypothetical protein
121	1167	996987	996527	amino acid permease, degenerate
122	1174	1004738	1006015	transposase, Mutator family, putative frameshift
123	1183	1010230	1010431	conserved hypothetical protein, putative frameshift
124*	1185	1011929	1011276	conserved hypothetical protein
	1186	1011979	1012392	transposase
	1187	1014026	1013417	conserved hypothetical protein
125	1205	1030075	1029989	integration host factor beta subunit
126	1216	1040216	1039905	conserved hypothetical protein
	1217	1040475	1040206	conserved hypothetical protein
127	1218	1040730	1040611	conserved hypothetical protein
	1219	1041092	1040727	conserved hypothetical protein
128	1220	1041685	1041107	conserved hypothetical protein, putative frameshift
129*	1232	1047332	1047604	biotin-protein ligase
	1231	1047887	1047597	transposase subunit
	1233	1048200	1047838	transposase
	1234	1048664	1048248	transposase
	1235	1048728	1049297	putative biotin-protein ligase
130	1245	1059921	1060046	transglycosylase SLT family protein

PG	REIS	start	stop	Annotation
131	1246	1061542	1060132	guanosine pentaphosphate phosphohydrolase, putative frameshift
132	1247	1062655	1061552	VirD4 protein
	2294	1063366	1062858	Vir D4 protein
133	2295	1064706	1063702	VirB11 protein
134	1248	1065049	1064921	VirB9 protein precursor
135	1251	1066438	1067136	VirB8 protein
136	1253	1068102	1068428	multisubunit Na+/H+ antiporter, MnhC subunit
137	1257	1072200	1072583	toxin of toxin-antitoxin system VapC, putative frameshift
138	1267	1085096	1085275	AmpG
139*	1296	1111569	1111186	conserved hypothetical protein
	1298	1112180	1111740	transposase
	1299	1112370	1112203	conserved domain protein
	1300	1112645	1113061	transposase
	1301	1113109	1113471	transposase
	1303	1113476	1113775	transposase
	1302	1114134	1113712	putative transposase
	1304	1114463	1114131	conserved hypothetical protein
140	1307	1115592	1116230	exported protein
	1308	1116445	1116663	exported protein
141	1318	1126763	1126233	conserved hypothetical protein
	1319	1127026	1126838	conserved hypothetical protein
	1320	1127135	1127028	hypothetical protein
142	1351	1153601	1154401	conserved hypothetical protein, putative frameshift
143	1409	1202892	1203548	putative polyketide synthase PksJ
144	1414	1210248	1206187	conjugal transfer protein TraA, putative frameshift
145	1416	1211606	1212139	transposase subunit, putative frameshift

PG	REIS	start	stop	Annotation
146	1432	1223441	1222682	nucleotidyltransferase, putative frameshift
147	1439	1229785	1229663	AmpG
148	1440	1230215	1229778	DNA gyrase subunit B
	1441	1232764	1230920	DNA gyrase subunit B, putative frameshift
149	1446	1239404	1238721	F pilus assembly protein TraH, putative point mutation
150	1459	1247586	1248314	tRNA delta(2)-isopentenylpyrophosphate transferase
	1460	1248325	1248642	tRNA delta(2)-isopentenylpyrophosphate transferase
151	1464	1253618	1253265	conserved hypothetical protein
	1465	1253871	1253764	conserved hypothetical protein
152	1467	1256345	1256190	chitin binding domain protein
153	1470	1257465	1258464	thioredoxin reductase, putative frameshift
154*	1476	1263049	1263495	conserved hypothetical protein
	1477	1263710	1263522	hypothetical protein
	1478	1263776	1264726	conserved hypothetical protein
155	1484	1270703	1269393	proline/betaine transporter, putative point mutation
156	1487	1271672	1275763	conjugal transfer protein TraA, putative frameshift
157	1510	1304170	1301971	conjugative transfer protein TraG, putative frameshift
158	1519	1314437	1312947	F pilus assembly protein TraB, putative point mutation
159	1541	1329944	1329567	conserved hypothetical protein
160	1549	1337470	1338003	conserved hypothetical protein
	1550	1338050	1338982	conserved hypothetical protein
	1551	1339016	1339171	conserved hypothetical protein
161*	1552	1339481	1339173	site-specific DNA adenine methylase
	1553	1339675	1339523	hypothetical protein
	1554	1339856	1340296	transposase
	1555	1340546	1340301	site-specific DNA adenine methylase

PG	REIS	start	stop	Annotation
162	1565	1355222	1354479	protein TraD
	1566	1356183	1355212	conjugative transfer protein
163	1587	1375610	1375729	DNA polymerase III subunit epsilon
	1588	1375732	1376487	DNA polymerase III subunit epsilon
164	1598	1381995	1382717	guanosine polyphosphate pyrophosphohydrolase/synthetase, putative point mutation
165	1638	1421523	1422830	citrate synthase
166	1639	1422940	1423932	putative GTP-binding protein, putative frameshift
167	1661	1441229	1441930	putative permease
168	1664	1443392	1442667	succinyl-diaminopimelate desuccinylase
	1665	1444123	1443898	putative succinyl-diaminopimelate desuccinylase
169	1666	1444603	1444124	type I restriction-modification enzyme, S subunit
170	1688	1459554	1459378	actin polymerization protein RickA
171	1690	1460640	1460741	beta-lactamase-like protein
172	1701	1470905	1468119	2-oxoglutarate dehydrogenase E1 component, putative frameshift
173*	1702	1471738	1471250	thermostable carboxypeptidase 1
	1703	1472756	1471818	transposase
	1704	1473694	1473182	CtaQ
174	1717	1489404	1488271	proline/betaine transporter, putative point mutation
175	1735	1504656	1504195	bicyclomycin resistance protein
	1736	1505986	1505723	bicyclomycin resistance protein, putative frameshift
176	1767	1527428	1526568	dipeptidyl peptidase family member 6
	1768	1528374	1528237	dipeptidyl aminopeptidase/acylaminoacyl-peptidase
177	1777	1536769	1537140	ankyrin repeat-containing domain protein
	1778	1537273	1537494	conserved hypothetical protein

PG	REIS	start	stop	Annotation
178*	1781	1541450	1541617	conserved domain protein
	1782	1542162	1541722	transposase
	1783	1542824	1542270	transposase, IS4 family
	1784	1543665	1543817	conserved hypothetical protein
179	1796	1555129	1556106	signal transduction histidine kinase, degenerate
180	1803	1560216	1560365	hypothetical protein
	1805	1560588	1562657	tetratricopeptide repeat-containing protein
181	1813	1567638	1569128	F pilus assembly protein TraB, authentic point mutation
182	1820	1579002	1578667	conserved hypothetical protein
	1821	1580019	1579135	conserved hypothetical protein
183	1887	1625991	1626968	glycosyltransferase, putative frameshift
184	1890	1631863	1630605	transposase, Mutator family, putative frameshift
185	1893	1632197	1632385	conserved hypothetical protein
186	1903	1640821	1640474	acetate kinase
187	1927	1659654	1659259	MFS type sugar transporter, putative frameshift
188*	1955	1682366	1682482	putative ankyrin repeat protein
	1956	1682554	1682994	transposase
	1957	1683053	1683181	ankyrin repeat-containing domain protein
189	1981	1706324	1708571	tetratricopeptide repeat-containing protein, putative frameshift
190	2005	1725489	1725716	cytochrome c oxidase Cu
	2006	1725721	1725957	cytochrome c oxidase Cu
191	2025	1739817	1740623	valyl-tRNA synthetase
	2026	1740942	1742429	valyl-tRNA synthetase
192	2037	1752788	1752468	phospholipase D
193	2038	1754326	1753116	poly-beta-hydroxybutyrate polymerase, putative frameshift

PG	REIS	start	stop	Annotation
194	2041	1758865	1758008	outer membrane protein OmpA
	2042	1762359	1759074	outer membrane protein A
195	2059	1779094	1779198	AmpG
196	2064	1781464	1780742	conserved hypothetical protein
	2065	1781687	1781550	conserved hypothetical protein
197	2085	1795881	1797076	multidrug resistance protein B, putative frameshift
198	2095	1806235	1806366	conserved hypothetical protein
	2096	1806417	1806653	conserved hypothetical protein
199	2099	1807618	1807782	guanosine polyphosphate pyrophosphohydrolases/synthetase-like protein
	2100	1807882	1808349	guanosine polyphosphate pyrophosphohydrolase/synthetase
200	2101	1808600	1809017	conserved hypothetical protein, putative point mutation
201	0007	1812555	1813086	conserved hypothetical protein, putative frameshift

Plasmid pREIS1

PG	REIS	start	stop	Annotation
202	2108	5937	3436	type I restriction-modification system methyltransferase subunit, putative frameshift
203	2127	23021	24868	conjugative transfer protein TraN, putative point mutation
204	2129	25864	27042	F pilus assembly protein TraH, putative frameshift
205	2130	27182	29740	conjugative transfer protein TraG, putative frameshift
206	2141	35574	36044	transposase
	2142	36097	36633	transposase
207	2152	43106	43579	transposase
	2153	43632	44168	transposase
208	2156	46338	45757	DNA polymerase IV, putative frameshift

Plasmid pREIS2

PG	**REIS**	**start**	**stop**	**Annotation**
209	2174	9330	8803	transposase, Mutator family, putative frameshift
210	2185	17796	17969	ATP-dependent protease La, bacterial type
211	2202	30154	29981	ATP-dependent protease La, bacterial type
212	2238	62525	62022	helicase RecD/TraA
	2239	62838	62569	conjugative transfer protein TraA_Ti

Plasmid pREIS3

PG	**REIS**	**start**	**stop**	**Annotation**
213	2261	17713	18840	tetratricopeptide TPR_2
	2262	18824	19279	O-linked GlcNAc transferase
214	2286	45217	44561	transposase
	2287	45728	45318	transposase

Plasmid pREIS4

PG	**REIS**	**start**	**stop**	**Annotation**
215	1	157	381	hypothetical protein
	2	466	594	hypothetical protein
216	3	608	997	transposase
	4	1093	1557	transposase
217	5	1584	1943	hypothetical protein
	6	1963	2112	conserved hypothetical protein
218	8	3573	2842	proline/betaine transporter
	9	4203	3607	proline/betaine transporter
219	17	7945	7751	putative transposase
	18	8273	8031	putative transposase
220	19	8590	8432	putative transposase
	20	9053	8826	putative transposase

PG	REIS	start	stop	Annotation
221	21	9967	9506	conserved hypothetical protein
222	25	11939	12118	conserved hypothetical protein
223	27	14981	15208	putative transposase
	28	15444	15602	putative transposase
224	29	15761	16003	putative transposase
	30	16089	16283	putative transposase
225	41	22072	21923	hypothetical protein
	42	22451	22092	hypothetical protein
226	43	22942	22478	transposase
	44	23427	23038	transposase
227	45	23569	23441	hypothetical protein
	46	23878	23654	hypothetical protein
228	47	24082	23966	hypothetical protein
	48	24593	24075	putative resolvase
229	52	26122	26316	putative resolvase
	53	26414	26707	putative resolvase
230	54	27212	27550	hypothetical protein
	55	27577	27930	hypothetical protein
231	56	28673	28008	transposase, Mutator family
	57	29266	28742	transposase, Mutator family
232	64	33474	33698	hypothetical protein
	65	33783	33911	hypothetical protein

Table S2. Possible locations of 18 REIS ORFs and two tRNA genes missing from the core Rickettsiaceae (12) and *Rickettsia* (6) genomes.

ORF[A,B]	**Coordinates**	**Description**[C,D,E]
HP (DUF2532)[#]	121012-121180	Flanked by HP (REIS_0158) and TNP (REIS_0159). Likely recombination with breakpoint at 237491-238024. Detected as a probable frameshift in tblastn analysis.
ClpX*, RimJ[#]	143402-143905	Within G18. Gap flanked by ExsB (REIS_0183, frameshift) and duplicated region (REIS_0184-REIS_0189). Reciprocal duplicated region encoded by REIS_0126-REIS_0131, coordinates 93946-101518. ClpX, a chaperone of the Hsp100 family, primarily functions in stress resistance, protein turnover, gene expression, and replication. RimJ is involved in ribosomal protein acetylation and ribosome assembly. Neither ORF was detected in tblastn analysis.
HP (COG3671)[#]	178768-179292	Flanked by NuoG (REIS_0234) and AcnA (REIS_0235). Detected as a full-length ORF in tblastn analysis.
SecE*, tRNA$^{\text{Ala-GGC}}$	446009-446512	Within G30. Gap flanked by NusG (REIS_0555) and EF-G (REIS_0556, frameshift). SecE is an integral component of the general protein secretion pathway. Detected as a full-length ORF in tblastn analysis.
HtrA*	454099-455493	REIS_0564 (degenerate) flanked by sulfur-transferase RHOD (REIS_0563) and HflC (REIS_0565). HtrA is a serine protease and chaperone, primarily involved in protein secretion and quality control. Detected as a probable frameshift in tblastn analysis.
RlpA*	556578- 557930	Split ORF (REIS_0694 and REIS_0695) truncated by G37. Flanked by DacF (REIS_0693) and lipoprotein OsmY (REIS_0696). Rare lipoprotein A has a putative role in lipid anchoring to the cell envelope. Detected as a probable frameshift in tblastn analysis.
GpsA*	589386-590306	Frameshift (REIS_0728). Flanked by TNP (REIS_0727) and metal-dependent hydrolase

		RMMBL (REIS_0729). Likely recombination with breakpoint at 975685-976190. GpsA is a glycerol-3-phosphate dehydrogenase functioning in phospholipid synthesis. Detected as a probable frameshift in tblastn analysis. NOTE: this gene is truncated (86 aa) in *R. candensis*.
UvrC$^{\#}$, (DUF461)*	891000-896068	Within G41. Gap flanked by two singletons: octaprenyl-diphosphate synthase (REIS_1078) and a truncated SpoT ORF (REIS_1079). UvrC functions in DNA lesion repair and was detected as an N-terminal truncation (~145 aa) in tblastn analysis. DUF461 was undetected in tblastn analysis.
HipB*, Rbn*	959187-964144	Within G56. Flanked by GlnA (REIS_1131) and TNP (REIS_1133); typically flanked by GlnA and Sca4 in other *Rickettsia* genomes. Sca4 in REIS is present in three copies in unrelated regions: REIS_1189, REIS_2113 (pREIS1), and REIS_2264 (pREIS3). HipB is a HTH-DNA-binding protein involved in transcription regulation. Detected as a full-length ORF in tblastn analysis. Rbn, a ribonuclease with both exo- and endonuclease function, was undetected in tblastn analysis.
HemB*	985378-986559	Within G60. Flanked by COG1293 (split ORF, REIS_1153 and REIS1155) and TNP fragment (REIS_1156). HemB is a delta-aminolevulinic acid dehydratase involved in heme (porphyrin) metabolism. Detected as an N-terminal truncation (~77 aa) in tblastn analysis.
n2B*	1171500-1172082	Flanked by HP (REIS_1371) and TNP (REIS_1372). Likely recombination with breakpoint at 1386755-1390279. P-loop ATPase. Undetected in tblastn analysis.
Fni$^{\#}$, tRNA$^{\text{Ala-GGC}}$	1333147-1333298	Region between TNP (REIS_1544) and ClpP protease (REIS_1545). Fni is an isopentenyl-diphosphate delta-isomerase involved in isoprene biosynthesis. Undetected in tblastn analysis.
HP (DUF2674)$^{\#}$	1795604-1795880	Flanked by TNP (REIS_2084) and EmrB (REIS_2085, frameshift). Detected as an N-terminal truncation (~26 aa) in tblastn analysis.
GatB*	1800802-1802455	Within G106. Flanked by GatA (REIS_2090)

		and PBPb (REIS_2092). A small singleton (REIS_2092) may be a fragment of GatB. GatB is subunit B of the Asn/Gln amidotransferase. Detected as an N-terminal truncation (~34 aa) in tblastn analysis.
DapB*	1803800-1805862	Flanked by flavin reductase (REIS_2093) and a fragment of BAX inhibitor (BI)-1 like protein family, or YccA (REIS_2094). YccA is recombined in two locales: REIS_2146 (pREIS1) and REIS_2231 (pREIS2). Both plasmid copies are adjacent to the ORFs encoding the small heat shock proteins typically present on rickettsial plasmids. DapB is a dihydrodipicolinate reductase functioning in amino acid, diaminopimelate and lysine biosynthesis. Detected as a probable frameshift in tblastn analysis.

[A] * = missing from Rickettsiaceae core genome, # = missing from *Rickettsia* core genome.

[B] Reference accession nos. for *R. rickettsii* str. Iowa: HP (DUF2532) = YP_001650150, ClpX = YP_001650429, RimJ = YP_001650430, HP (COG3671) = YP_001650579, SecE = YP_001649535, HtrA = YP_001649526, RlpA = YP_001649912, GpsA = YP_001649998, UvrC = YP_001650238, HP (DUF461) = YP_001650239, HipB = YP_001650047, Rbn = YP_001650048, HemB = YP_001650174, n2B = YP_001649646, Fni = YP_001650118, HP (DUF2674) = YP_001649563, GatB = YP_001649554, DapB = YP_001649551. The flanking genes for all SFG rickettsiae subjects were evaluated to determine the conservation of gene order.

[C] G*n* = assembly gaps, where *n* = 1-109.

[D] HP = hypothetical protein, TNP = transposable elements and related derivatives.

[E] Query sequences from *R. rickettsii* str. Iowa were used to perform tblastn searches against the REIS WGS reads.

Table S3. Location of 109 regions of low coverage and uncertainty in the REIS genome assembly.

gap	start	end	length	length of prior contig	Suspected missing genes^{A}
1	5073	6072	1000	5072	NA
2	7177	7376	200	1104	NA
3	8732	8831	100	1355	NA
4	11885	12084	200	3053	NA
5	15440	15939	500	3355	NA
6	16949	17948	1000	1009	NA
7	19543	21542	2000	1594	NA
8	35431	36630	1200	13888	NA
9	44835	45834	1000	8204	NA
10	60740	61239	500	14905	NA
11	63462	63961	500	2222	NA
12	75166	75665	500	11204	NA
13	92197	92296	100	16531	NA
14	124204	125703	1500	31907	NA
15	127041	128040	1000	1337	NA
16	138522	138721	200	10481	NA
17	141221	141720	500	2499	NA
18	143405	143904	500	1684	ClpX*, RimJ$^{\#}$
19	393678	393877	200	249773	NA
20	401987	402286	300	8109	NA
21	405126	405625	500	2839	NA
22	421345	421844	500	15719	NA
23	424869	425368	500	3024	NA
24	426578	426677	100	1209	NA
25	428572	428671	100	1894	NA
26	431043	432042	1000	2371	NA
27	433659	434158	500	1616	NA
28	436029	437028	1000	1870	NA
29	438130	438329	200	1101	NA
30	446011	446510	500	7681	SecE*, tRNA$^{\text{Ala-GGC}}$
31	447444	448443	1000	933	NA
32	479645	479944	300	31201	NA
33	484101	484300	200	4156	NA
34	487601	488600	1000	3300	NA
35	491953	492952	1000	3352	NA
36	542589	542688	100	49636	NA
37	556614	556713	100	13925	RlpA*
38	834616	834715	100	277902	NA
39	853588	853687	100	18872	NA
40	871926	872025	100	18238	NA
41	891006	896005	5000	18980	UvrC$^{\#}$, DUF461*
42	896980	901979	5000	974	NA
43	905357	905556	200	3377	NA
44	909856	910055	200	4299	NA

45	912922	914921	2000	2866	NA
46	917180	918679	1500	2258	NA
47	921146	922645	1500	2466	NA
48	925835	926334	500	3189	NA
49	932585	933084	500	6250	NA
50	934360	934859	500	1275	NA
51	943110	943609	500	8250	NA
52	945353	945452	100	1743	NA
53	946992	947091	100	1539	NA
54	952148	952247	100	5056	NA
55	956094	958093	2000	3846	NA
56	960145	964144	4000	2051	HipB*, Rbn*
57	975686	976185	500	11541	NA
58	978869	978968	100	2683	NA
59	984421	984520	100	5452	NA
60	986241	986340	100	1720	HemB*
61	987741	988740	1000	1400	NA
62	990391	991390	1000	1650	NA
63	1012415	1013414	1000	21024	NA
64	1014712	1016711	2000	1297	NA
65	1020107	1020706	600	3395	NA
66	1057152	1057451	300	36445	NA
67	1062658	1062857	200	5206	NA
68	1064720	1064919	200	1862	NA
69	1078289	1080788	2500	13369	NA
70	1084486	1085085	600	3697	NA
71	1211094	1211593	500	126008	NA
72	1221233	1222232	1000	9639	NA
73	1230216	1230715	500	7983	NA
74	1233098	1235097	2000	2382	NA
75	1236696	1238195	1500	1598	NA
76	1243165	1243364	200	4969	NA
77	1250762	1250961	200	7397	NA
78	1260435	1261934	1500	9473	NA
79	1276914	1277013	100	14979	NA
80	1301771	1301970	200	24757	NA
81	1395675	1397174	1500	93704	NA
82	1414264	1414363	100	17089	NA
83	1427376	1428375	1000	13012	NA
84	1430770	1431269	500	2394	NA
85	1436864	1437363	500	5594	NA
86	1440729	1441228	500	3365	NA
87	1443395	1443894	500	2166	NA
88	1446257	1446856	600	2362	NA
89	1461972	1462371	400	15115	NA
90	1624772	1625271	500	162400	NA
91	1647976	1648175	200	22704	NA
92	1697324	1698023	700	49148	NA
93	1726820	1727319	500	28796	NA
94	1729596	1729795	200	2276	NA
95	1738190	1738289	100	8394	NA

96	1740641	1740940	300	2351	NA
97	1747214	1748713	1500	6273	NA
98	1750796	1750895	100	2082	NA
99	1751952	1752151	200	1056	NA
100	1758872	1759071	200	6720	NA
101	1768315	1769914	1600	9243	NA
102	1771182	1775181	4000	1267	NA
103	1777631	1778630	1000	2449	NA
104	1781784	1782783	1000	3153	NA
105	1794355	1794454	100	11571	NA
106	1800854	1802453	1600	6399	GatB*
107	1803800	1805799	2000	1346	NA
108	1809020	1810019	1000	3220	NA
109	1817934	1818133	200	7914	NA

[A] * = missing from Rickettsiaceae core genome, # = missing from *Rickettsia* core genome. See Table S2 for more information.

Table S4. NCBI accession numbers for proteins used to estimate the phylogeny of the biotin metabolic pathway.

Taxon	**Biotin synthesis genes** [A]					
	BioA (OAT)	**BioD (DTBS)**	**BioC (UbiE)**	**BioH**	**BioF (KBL)**	**BioB (radSAM)**
REIS	ZP_04698248	ZP_04698249	ZP_04698250	ZP_04698251	ZP_04698252	ZP_04698253
Lawsonia intracellularis str. PHE/MN1-00	YP_594802	YP_594801	YP_594800	YP_594799	YP_594798	YP_594797
Neorickettsia risticii str. Illinois	YP_003081804	YP_003081803	YP_003081802	YP_003081801	YP_003081800	YP_003081798
Neorickettsia sennetsu str. Miyayama	YP_506498	YP_506497	YP_506495	YP_506494	YP_506493	YP_506492
Magnetospirillum magneticum str. AMB-1	YP_422126	YP_422125	YP_422124	YP_422123	YP_422122	YP_422120
Achromobacter piechaudii str. ATCC 43553	ZP_06685978	ZP_06685981	ZP_06685981	ZP_06685980	ZP_06685979	ZP_06689510
Bordetella pertussis str. Tohama I	NP_880815	NP_880812	NP_880812	NP_880813	NP_880814	NP_881326
Bordetella avium str. 197N	YP_786961	YP_786964	YP_786964	YP_786963	YP_786962	YP_785360
Bordetella parapertussis str. 12822	NP_884047	NP_884044	NP_884044	NP_884045	NP_884046	NP_883439
Gluconacetobacter diazotrophicus str. PAI 5	YP_001602164	YP_001602167	YP_001602167	YP_001602166	YP_001602165	YP_001603186
Crocosphaera watsonii str. WH 8501	ZP_00514990	ZP_00515520	ZP_00515866	ZP_00515865	ZP_00515864	ZP_00514954
Gluconacetobacter hansenii str. ATCC 23769	ZP_06834781	ZP_06834784	ZP_06834784	ZP_06834783	ZP_06834782	ZP_06833902

Taxon	Biotin synthesis genes [A]					
	BioA (OAT)	**BioD (DTBS)**	**BioC (UbiE)**	**BioH**	**BioF (KBL)**	**BioB (radSAM)**
Bordetella bronchiseptica str. RB50	NP_889867	NP_889870	NP_889870	NP_889869	NP_889868	NP_887883
Nitrococcus mobilis str. Nb-231	ZP_01125956	ZP_01125955	ZP_01127180	ZP_01128606	ZP_01128607	ZP_01127181
Cyanothece sp. ATCC 51142	YP_001804508	YP_001804930	YP_001805563	YP_001805564	YP_001805565	YP_001805764
Zymomonas mobilis subsp. mobilis ZM4	YP_163653	YP_163650	YP_163650	YP_163652	YP_163651	YP_161829
Legionella pneumophila str. Paris	YP_123751	YP_123755	YP_124590	YP_123754	YP_123753	YP_123752
Legionella pneumophila str. Corby	YP_001250207	YP_001250211	YP_001251077	YP_001250210	YP_001250209	YP_001250208
Legionella longbeachae str. D-4968	ZP_06188812	ZP_06188816	ZP_06186482	ZP_06188815	ZP_06188814	ZP_06188813
Parachlamydia acanthamoebae str. Hall's coccus	ZP_06299038	ZP_06299039	ZP_06299041	ZP_06300545	ZP_06299042	ZP_06300151
Chlamydophila pneumoniae str. CWL029	NP_225235	NP_225236	-----	-----	NP_225237	NP_225238
Escherichia coli str. K-12 substr. MG1655	NP_415295	NP_415299	NP_415298	NP_417871	NP_415297	NP_415296
Kurthia sp. 538-KA26	BAB39453	BAB39454	BAB39463	BAB39462	BAB39461	BAB39458
Neisseria meningitidis str. MC58	NP_273774	NP_273775	NP_273521	NP_273326	NP_273519	NP_274174
Pseudomonas aeruginosa str. PAO1	NP_249111	NP_249195	NP_249194	NP_249193	NP_249192	NP_249191
Vibrio cholerae str. B33	-----	ZP_01976490	ZP_01976489	ZP_01973913	ZP_01976491	ZP_01976488

Taxon	**Biotin synthesis genes**[A]					
	BioA (OAT)	**BioD (DTBS)**	**BioC (UbiE)**	**BioH**	**BioF (KBL)**	**BioB (radSAM)**
Vibrio cholerae str. B33	ZP_04400515	ZP_04400519	ZP_04400518	ZP_04399326	ZP_04400517	ZP_04400516
Xylella fastidiosa str. 9a5c	NP_297482	NP_299755	NP_299378	NP_298645	NP_298646	NP_297357
Anaplasma centrale str. Israel	YP_003328623	YP_003328228	YP_003328117	-----	YP_003328822	YP_003328325
Anaplasma phagocytophilum str. HZ	YP_505081	YP_505758	YP_504662	-----	YP_505544	YP_504837
Anaplasma marginale str. St. Maries	YP_153824	YP_154232	YP_154354	-----	YP_153625	YP_154121
Anaplasma marginale str. Virginia	ZP_05279021	ZP_05279432	ZP_05279564	-----	ZP_05278823	ZP_05279325
Anaplasma marginale str. Florida	YP_002563530	YP_002563940	YP_002564061	-----	YP_002563334	YP_002563832
Ehrlichia ruminantium str. Gardel	YP_196321	YP_196818	YP_195935	-----	YP_196094	YP_196599
Ehrlichia ruminantium str. Welgevonden	YP_180252	YP_180714	YP_179888	-----	YP_180041	YP_180514
Ehrlichia ruminantium str. Welgevonden	YP_197275	YP_196818	YP_196885	-----	YP_197049	YP_197558
Ehrlichia chaffeensis str. Sapulpa	ZP_00545089	ZP_00545228	ZP_00544740	-----	ZP_00545310	ZP_00544607
Ehrlichia chaffeensis str. Arkansas	YP_507474	YP_507894	YP_506868	-----	YP_507738	YP_507172
Ehrlichia canis str. Jake	YP_303017	YP_303517	YP_302980	-----	YP_302820	YP_303287

[A] BioA, adenosylmethionine-8-amino-7-oxononanoate aminotransferase; BioD, dethiobiotin synthetase; BioC, carboxyl-O-methyltransferase; BioH, carboxylesterase; BioF, 8-amino-7-oxononanoate synthase; BioB, biotin synthase.

Table S5. Distribution of 1,480 REIS CDS across 15 Rickettsiaceae genomes.

Genome	**Shared OGs** [A]	**No. proteins** [B]	**REIS proteins** [C]
Scrub Typhus Group			
O. tsutsugamushi Ikeda	590	813 (27.4)	689 (14.4)
O. tsutsugamushi Boryong	567	712 (20.4)	700 (19.0)
Ancestral Group			
R. bellii RML369-C	914	933 (2)	1041 (12.2)
R. bellii OSU 85 389	898	906 (0.9)	1029 (12.7)
R. canadensis McKiel	820	827 (0.9)	853 (3.9)
Typhus Group			
R. prowazekii Madrid E	754	755 (0.3)	773 (2.5)
R. typhi Wilmington	756	758 (0.3)	776 (2.6)
Transitional Group			
R. felis URRWXCal2	964	1033 (6.7)	1101 (12.4)
R. akari Hartford	866	900 (3.8)	1087 (20.3)
Spotted fever Group			
R. massiliae MTU5	892	909 (1.9)	1031 (13.5)
R. rickettsii Sheila Smith	874	886 (1.4)	906 (3.5)
R. rickettsii Iowa	928	959 (3.2)	969 (4.2)
R. conorii Malish 7	901	911 (1.1)	941 (4.3)
R. sibirica 246	889	902 (1.4)	925 (3.9)
R. africae ESF-5	858	864 (0.7)	893 (3.9)

[A] Number of orthologous groups (OGs) per genome containing at least one REIS protein.

[B] Total proteins within shared OGs. Parentheses enclose percentage of proteins that are multicopy.

[C] Total REIS proteins within shared OGs. Values in parentheses as described above.

Table S6. Common protein families across nine plasmids from six *Rickettsia* genomes [A].

Protein[B]	Plasmids[C,D,E]								
	pRF (68)	**pREIS1 (53)**	**pREIS2 (81)**	**pREIS3 (43)**	**pREIS4 (52)**	**pRM (17)**	**pRMA (12)**	**pRPR (20)**	**pRAF (11)**
DnaA	396; 19 774; 5	821; 2109	768; 2196	790; 2255	----	820; 16	778; 1	779; 4	740; 1260
TNP_M	250; 30 230; 54	----	221; 2166 224; 2235 224; 2211 224; 2176	----	141; 11 141; 28 221; 44	----	----	----	----
HTH_XRE	220; 29 122; 2	161; 2112	59; 2192	----	----	167; 19 168; 20	161; 12	----	157; 1269
TraA$_{Ti}$	93; 38 907; 39 357; 40	2464; 2159	499; 2223	2076; 2246	----	187; 3	----	----	----
IbpA	196; 52 ----	175; 2145	175; 2232	----	130; 14 130; 25	174; 7	----	175; 12	----
ParA_1	211; 23	210; 2111	----	----	271; 19	196; 18	213; 13	243; 1	223; 1270
Rve	289; 48	178; 2142 178; 2153	----	218; 2286	----	----	----	364; 5 364; 15	----
HSP	163; 51	----	----	----	154; 13 154; 26	158; 6	----	154; 13	----
SR_ResInv	182; 66	----	72; 2167	----	----	----	183; 5	----	194; 1262
TNP_31_1	371; 61	318; 2136	----	----	----	----	347; 6	----	361; 1264
TraD_AP	----	104; 2102	104; 2244 104; 2229	104; 2248	----	----	----	----	99; 1266
ISRpe1	63; 49	156; 2141 157; 2152	----	136; 2287	----	----	----	----	----
TNP_11	109; 1	----	332; 2169 158; 2209 158; 2178	----	----	----	----	----	----
p_TNP	----	312; 2148 307; 2105	----	----	----	296; 22	----	304; 14	----
TPR_1	254; 17 815; 26	----	----	375; 2261	----	----	----	----	----
TraA_AP	137; 65	----	131; 2162	----	----	134; 10	----	----	----
Sca12	476; 25	----	----	852; 2259	----	----	----	----	881; 1268
TNP_rve	----	----	----	----	173; 10 173; 29	130; 11	----	----	----

Protein[B]	**Plasmids[C,D,E]**								
	pRF (68)	**pREIS1 (53)**	**pREIS2 (81)**	**pREIS3 (43)**	**pREIS4 (52)**	**pRM (17)**	**pRMA (12)**	**pRPR (20)**	**pRAF (11)**
SR_ResInv_2	----	----	----	----	111; 12 111; 27	127; 5	----	----	----
HP_1	92; 45	73; 2131	----	----	----	----	----	----	----
TraD_F	309; 44	567; 2132	----	----	----	----	----	----	----
SR_ResInv_3	146; 32	192; 2106	----	----	----	192 23	----	----	----
TraD_Ti	85; 37	74; 2103	74; 2245	----	----	----	----	----	----
TNP_31_2	350; 41	----	----	370; 2251	----	----	----	----	----
Sca4	599; 27	----	----	583; 2264	----	----	----	470; 19	----
DnaG-like	387; 62	----	----	1063; 2288	----	----	----	----	----
HP_2	281; 12	----	----	222; 2263	----	----	----	----	----
TNP_14_rve	102; 64	----	293; 2170	----	----	----	----	----	----
ParA_2	271; 3	----	271; 2193	----	----	----	----	----	----
HP_3	106; 4	----	106; 2194	----	----	----	----	----	----
HP_4	96; 59	----	----	----	----	----	358; 10	----	----
TNP_31_3	----	----	236; 2173	----	----	----	----	----	213; 1263
HP_5	----	----	----	----	74; 11 74; 36	----	----	----	----
ProP	----	----	----	----	323; 31	416; 12	----	----	----
POLIIIe	377; 34	----	75; 2163	----	----	----	----	----	----
HP_6	----	----	36; 2171	----	----	----	90; 7	----	----
RecD/TraA	----	----	722; 2164	----	----	721; 8	----	----	----
MdlB	----	----	590; 2217	----	----	----	----	588; 11	----
TPR_2	419; 18	----	----	----	----	----	103; 8	----	----
HP_7	164; 21	----	----	----	----	----	----	141; 3	----
HP_8	82; 22	----	----	----	----	----	----	82; 2	----
Total	**37**	**17**	**24**	**11**	**15**	**14**	**8**	**11**	**8**
% grouped	**54.4**	**32.1**	**29.6**	**20.8**	**28.8**	**82.4**	**66.7**	**55.0**	**72.7**

[A] Orthologous protein families generated using OrthoMCL (1).

[B] Protein names as follows (from top to bottom of table): 'DnaA', chromosomal replication initiation protein (CTD); 'TNP_M', MULE transposase; 'HTH_XRE', helix-turn-helix-type transcriptional regulator xre; 'TraA_{Ti}', conjugative transfer relaxase; 'IbpA', molecular chaperone (small heat shock protein); 'ParA_1', chromosome partitioning protein A; 'Rve', transposase rve superfamily (integrase core domain); 'HSP', alpha-crystallin domain (ACD) of alpha-crystallin-type small(s) heat shock proteins; 'SR_ResInv', serine recombinase (SR) family, resolvase and invertase subfamily, catalytic domain; ’TNP_31_1’, putative transposase, YhgA-like; pfam04754; ‘TraD_AP’, TraD associated protein; ‘ISRpe1’, N-terminal region of transposase ISRpe1; ‘TNP_11’, transposase DDE domain, pfam01609; ‘p_TNP’, putative transposase T_den_put_tspse; ‘TPR_1’, tetratricopeptide repeat domain-containing protein; ‘TraA_AP’, TraA associated protein; ‘Sca12’, surface antigen 12; ‘TNP_rve’, transposase rve superfamily (integrase core domain); ‘SR_ResInv_2’, serine recombinase (SR) family, resolvase and invertase subfamily, catalytic domain; ‘HP_1’, hypothetical protein known only from REIS and *R. felis*; ‘TraD_F’, TraD F-T4SS coupling protein (VirD4 homolog); ‘SR_ResInv_3’, serine recombinase (SR) family, resolvase and invertase subfamily, catalytic domain; ‘TraD_Ti’, TraD conjugative transfer protein; ‘TNP_31_2’, putative transposase, YhgA-like, with Yae1_N domain (COG5464); ‘Sca4’, surface antigen 4; ‘DnaG-like’, topoisomerase-primase (TORPIM) nucleotidyl transferase/hydrolase domain; ‘HP_2’, hypothetical protein known only from REIS and *R. felis*; ‘TNP_14_rve’, transposase 14 (NTD) with rve superfamily (integrase core domain) (CTD); ‘ParA_2’, chromosome partitioning protein A; ‘HP_3’, hypothetical protein known only from REIS and *R. felis*; ‘HP_4’, length variable hypothetical protein known only from *Rickettsia* plasmids; ’TNP_31_3’, putative transposase, YhgA-like; pfam04754; ‘HP_5’, hypothetical protein known only from pREIS4; ‘ProP’, proline/betaine transporter; ‘POLIIIe’, DNA polymerase III, epsilon subunit-like; ‘HP_6’, hypothetical protein known only from *Rickettsia* plasmids, possibly with TPR repeat domain; ‘RecD/TraA, helicase similar to RecD (exodeoxyribonuclease V alpha chain); ‘MdlB’, ABC-type multidrug efflux pump; ‘TPR_2’, tetratricopeptide repeat domain-containing protein; ‘HP_7’, hypothetical protein known only from *Rickettsia* plasmids, possible transcription regulator; ‘HP_8’, hypothetical protein with limited distribution outside of *Rickettsia* spp.

[C] Total number of ORFs included in clustering in parentheses. Plasmid associations with rickettsial genomes are illustrated in Fig. 3 of the text.

[D] Protein length (aa) before semicolon; plasmid ORF ID number after.

[E] ORFs colored red are truncated or split relative to the largest ORF in each orthologous group.

1. **Li, L., C. J. Stoeckert, Jr., and D. S. Roos.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res **13:**2178-2189.

Table S7. Taxonomic breakdown of the best non-*Rickettsia* blastp hits to 769 REIS proteins.

Taxon	**MGE**		**Other**		**All**	
	No.[B]	**Avg.[C]**	**No.[B]**	**Avg.[C]**	**No.[B]**	**Avg.[C]**
Legionella pneumophila Paris	126	4.20E-04	1	1.00E-11	127	4.16E-04
Geobacter uraniireducens Rf4	103	1.27E-07	0	NA	103	1.27E-07
Microscilla sp. PRE1	100	5.48E-11	0	NA	100	5.48E-11
Orientia tsutsugamushi Boryong	44	6.59E-04	1	2.60E-41	45	6.44E-04
Microcystis aeruginosa PCC 7806	41	2.05E-02	0	NA	41	2.05E-02
Wolbachia (*D. simulans*)	3	2.40E-67	0	NA	3	2.40E-67
Trichomonas vaginalis G3	2	0.0037	6	0.2983	8	0.2247
Lawsonia intracellularis PHE/MN1-00	0	NA	10	2.60E-56	10	2.60E-56
Wolbachia (*D. melanogaster*)	1	4.00E-27	9	1.33E-05	10	1.20E-05
Protochlamydia amoebophila UWE25	8	0.0022	1	1.20E-06	9	0.002
Mycoplasma hyopneumoniae J	1	0.47	1	0.13	2	0.3
Klebsiella pneumoniae MGH 78578	6	4.67E-24	1	0.94	7	1.34E-01
Gammaproteobacteria	50	2.59E-02	24	2.98E-01	74	1.14E-01
Firmicutes	7	5.14E-01	14	0.2154	21	3.06E-01
Cyanobacteria	13	1.33E-02	2	1.80E-18	15	1.15E-02
Bacteroidetes	17	0.0156	2	2.60E-04	19	0.014
Actinobacteria	3	0.3013	10	1.38-01	13	0.1760
Alphaproteobacteria	10	9.00E-02	9	1.44E-01	19	1.16E-01
Betaproteobacteria	13	0.0409	4	0.9025	17	0.2436
Epsilonproteobacteria	9	3.11E-02	1	0.049	10	3.29E-02
other bacteria [A]	4	0.2501	5	0.691	9	0.4951
Eukaryota	43	0.4803	47	0.6673	90	0.578
Archaea	5	2.02E-01	10	0.3049	15	2.71E-01

[A] Aquificae (1), Chloroflexi (3), Deinococcus-Thermus (1), Fusobacteria (1), Spirochaetes (1), and Tenericutes (2)

[B] No. of top blastp hits.

[C] Avg. e-value for top blastp hits.

Table S8. Distribution of 653 REIS transposases and inactive derivatives across 15 other Rickettsiaceae genomes. Roughly half of the REIS transposases (TNPs) are singletons or duplicated singletons. The other half are only found in protein families that are non-conserved across the Rickettsiaceae genomes (Note: no TNPs are present within the core Rickettsiaceae or *Rickettsia* genomes). The majority of these TNPs (305) are found within only 26 protein families, illustrating the extreme proliferation of several TNPs within the REIS genome. The distribution of 317 TNPs across each Rickettsiaceae genome is shown in Fig. S9.

Distribution	**Makeup** [A]	**Breakdown** [B]			
		Size [C]	**w/TNP** [D]	**TNP** [E]	**%** [F]
Rickettsiaceae core	P	459	0	0	--
	I	9	0	0	--
Rickettsia core	P	159	0	0	--
	I	7	0	0	--
Non-conserved	S	338	12	12	3.6
	M	137	26	305	1,173
REIS singletons	T	388	NA	39	10.1
	D	351	NA	263	74.9

[A] Membership across genomes: P, perfect (or one protein per genome); I, perfect with at least one genome contributing more than one protein; S, one protein per 2-15 genomes; M, one genome with more than one protein; T, true singleton; D, dupleton (duplicated singleton). See Fig. S3 for more information pertaining to orthologous group (OG) clustering.

[B] Distribution of 619 REIS TNPs across 15 other Rickettsiaceae genomes. Note: 34 putative TNPs from pREIS4 were not included in OG clustering.

[C] Total OGs for all except singletons and dupletons, which have total proteins.

[D] No. OGs containing TNPs. For singletons and dupletons, total TNPs are listed.

[E] Total no. of TNPs.

[F] Percentage of TNPs per group.

Table S9. Distribution in the REIS genome of putative virulence factors, common phylogenetic markers, and other proteins important in obligate intracellular lifestyle.

ORF [A]	**Description** [B]
GltA	Citrate synthase. Truncated (340 aa, REIS_1638) relative to other rickettsial proteins, which are highly conserved in length (435 aa).
LipA	17kDa surface antigen. Full length (REIS_1933).
Sca0 (RompA)	Surface antigen. Two split ORFs (REIS_2041 and REIS_2042) flanked by TNP (REIS_2043) and interrupted by gap (G100).
Sca1	Surface antigen. Full length (REIS_1976).
Sca2	Surface antigen and actin nucleator. Four fragments (REIS_0502-REIS_0506) flanked by TNP (REIS_0501) and interrupted by gap (G20).
Sca3	Surface antigen. Absent. Possible fragment within gap G55, as Sca3 flanks ATP-dependent protease La (REIS_1130) in other genomes.
Sca4	Surface antigen. Three copies. Chromosome ORF (REIS_1189, 615 aa) is truncated by gap (G64). pREIS1 (REIS_2113, 320 aa) and pREIS3 (REIS_2264, 583 aa) ORFs closely resemble *R. peacockii* pRPR_19, which is a chimeric protein with only ~100 aa at the N-terminus having homology with Sca4.
Sca5 (RompB)	Surface antigen (ubiquitous). Frameshift (REIS_0057) flanked by TNP (REIS_0056) and gap (G8). Likely an assembly problem, as a full length RompB sequence (ABQ57416) has been generated from a different population of REIS.
Sca6	Surface antigen. Absent.
Sca7	Surface antigen. Absent.
Sca8	Surface antigen. Absent.
Sca9	Surface antigen. Two split ORFs (REIS_0202, REIS_0203) flanked by TNP (REIS_0201) and Sca13 region (see below). Combined ORFs (354 aa) are shorter than the largest ortholog (*R. felis*, 557 aa).
Sca10	Surface antigen. Absent.
Sca11	Surface antigen. Absent.
Sca12	Surface antigen. Full length (REIS_2259) and encoded on pREIS3.
Sca13	Two split ORFs (REIS_0204, REIS_0206) interrupted by TNP (REIS_0205). Flanked by Sca9 region (see above) and toxin-anti-toxin module

	(REIS_0207, REIS_0208).
Sca14	Surface antigen. Absent.
Sca15	Surface antigen. Absent.
Sca16	Surface antigen. Absent.
Msp4_1	Surface antigen/adhesin. Full length (REIS_1939).
Msp4_2	Surface antigen/adhesin. Full length (REIS_1938).
Omp1 (YaeT)	Surface antigen. Full length (REIS_2072).
NlpD	Outer membrane antigenic lipoprotein precursor. Full length (REIS_1390).
PrsA	Peptidylprolyl isomerase, binds host histone. Full length (REIS_1435).
Pld	Phospholipase D. Two copies: first chromosome copy (REIS_0873) is between RAGE-bellii and RAGE-F in a region with multiple TNPs; second chromosome copy (REIS_2037) is truncated and flanked by TNP (REIS_2036) and a frameshift (REIS_2038; Poly-beta-hydroxybutyrate polymerase).
Pat	Phospholipase, patatin B1. Full length (REIS_1734).
Pat2	Phospholipase, patatin B2. Two split ORFs (REIS_1222, REIS_1223) flanked by BirA (REIS_1221) and HP (REIS_1224).
PpcE	Post proline cleaving enzyme (prolyl oligopeptidase). Absent.
PtrB	Protease B (prolyl oligopeptidase). Full length (REIS_1259).
Dap2	Dipeptidyl aminopeptidase 2 (prolyl oligopeptidase). Two copies: full length (REIS_0738) and split ORF (REIS_1767, REIS_1768) flanked by TNP (REIS_1766) and NtrY (REIS_1769).
TlyA	Hemolysin A. Full length (REIS_1734).
TlyC	Hemolysin C. Full length (REIS_0496).
RickA	Actin tail polymerization. Split with the majority of the ORF (REIS_1104) flanked by TNP (REIS_1103) and MraY (REIS_1105); C-terminal region is duplicated by recombination event (REIS_0633, REIS_1688). See Fig. S5.
RalF	Putative guanine nucleotide exchange protein. Like other SFG rickettsiae, lacks Sec7 domain and contains only C-terminal “tail” (REIS_0937), which is flanked by TNP (REIS_0936) and malate acetyltransferase (REIS_0938).

GpsA	Glycerol-3-phosphate dehydrogenase. Frameshift (REIS_0728) flanked by TNP (REIS_0727) and ribonuclease J 1 (REIS_0729). Likely recombination between REIS_0727 and REIS_1146 (TNP) as REIS_1145 (auxin efflux carrier) flanks GpsA in all other *Rickettsia* genomes. REIS_1145 is truncated by gap (G57) supporting the probable recombination event. Conservation of *gpsA* across *Rickettsia* genomes is shown in Fig S6A.
GlpT	Glycerol-3-phosphate transporter. Full length (REIS_0397). A second ORF (REIS_1488) is slightly truncated (312 aa) due to gap (G79) and present within the RAGE-A insert (Fig. RA). Estimated phylogeny of GlpT transporters is shown in Fig. S2.
MetK	S-adenosylmethionine synthase. Pseudogene (REIS_0094) flanked by conserved hypothetical protein (REIS_0093) and HicB (REIS_0095).
EamA	S-adenosylmethionine (SAM) transporter. Full length (REIS_0437) Estimated phylogeny of EamA transporters is shown in Fig. S2.
Tlc	ATP/ADP translocase. Tlc1 full length (REIS_0396), Tlc3 full length (REIS_0804), Tlc4 full length (REIS_1190), and Tlc5 full length (REIS_0531). Tlc2 full length (REIS_0396) with a second full-length copy (REIS_1498) found within the RAGE-A insert (Fig. RA). This latter region is duplicated and recombined, yielding a fragmented Tlc2 (REIS_0285, 188 aa). Estimated phylogeny of Tlc translocases is shown in Fig. S11.
InvA	Dinucleoside polyphosphate hydrolase. Full length (REIS_1328).
SodB	Superoxide dismutase. Full length (REIS_1227).
ProP	Proline/betaine transporter. Seven major groups in *Rickettsia* spp. (Prop1-7). ProP6 absent. ProP1 (REIS_0438), ProP2 (REIS_1056), Prop3 (REIS_0282, with RPE3), and ProP7 (REIS_0013) full length. ProP4 (REIS_1965) full length with two duplicated small fragments (REIS_1027, REIS_1074). ProP5 (REIS_0131) full length with a truncated copy (REIS_0189, 120 aa) in a duplicated region (coordinates 93946-101518); a second full length ProP5 (REIS_1484) is associated with RAGE-A (Fig. 5, Fig. 6). A frameshift (REIS_0539, REIS_0541) is recombined with ProP transporters encoded on two separate regions of pREIS4, all of which are associated with an integrase (REIS_0538), a SpoT hydrolase (REIS_0540), and a SAM-methyltransferase (REIS_0542). A separate pseudogene (REIS_0657) is also associated with an integrase (REIS_0658) and is flanked by TNP (REIS_0656). Estimated phylogeny of ProP transporters is shown in Fig. S13.
ProQ	Proline/betaine transporter regulator. Found only in REIS. Two full length ORFs on pREIS1 (REIS_2104, REIS_2149) and one full length ORF on pREIS3 (REIS_2252).
SpoT/ RelA	Stringent response regulators; (p)ppGpp hydrolase domain-containing proteins of varying length: REIS_0540 (244 aa), REIS_0867 (248 aa, RAGE-Be), REIS_1055 (235 aa), REIS_1079 (103 aa), REIS_1482 (240 aa, RAGE-

	A), REIS_1557 (217 aa), REIS_1799 (151 aa), REIS_2099 (54 aa) + REIS_2100 (155 aa), REIS_2260 (197aa, pREIS3). Two ORFS within a duplicated region of pREIS4, ORF10326 (122 aa) and ORF10359 (245 aa) is recombined with a region on the chromosome (coords. 427259-430984). (p)ppGpp synthetase domain-containing proteins of varying length: REIS_0058 (201 aa), REIS_0864 (215 aa, RAGE-Be), REIS_1891 (139 aa). Compilation of SpoT/RelA ORFs in *Rickettsia* genomes is shown in Fig. S12.
Ndk	Nucleoside diphosphate kinase. Full length (REIS_0398).
GppA	Guanosine pentaphosphate phosphohydrolase. Frameshift (REIS_1246) flanked by RvhB1 (REIS_1245, truncated) and RvhD4 (REIS_1247, frameshift).
AmpG	Peptidoglycan import. AmpG1 full length (REIS_0806), AmpG2 full length (REIS_0045), AmpG3 full length (REIS_0267) with four duplicated fragments (REIS_0692, REIS_1267, REIS_1439, REIS_2059), AmpG4 absent. Estimated phylogeny of AmpG transporters is shown in Fig. S14.
ANK	Ankyrin repeat-domain containing proteins (ANKs) of varying length: REIS_0667 (116 aa) + REIS_0669 (101 aa), REIS_1102 (238 aa), REIS_1344 (584 aa), REIS_1695 (68 aa), REIS_1777 (123 aa) + REIS_1778 (73 aa), REIS_1920 (51 aa), REIS_1955 (38 aa) + REIS_1957 (42 aa), REIS_2011 (189 aa), REIS_2045 (147 aa). NOTE: None of the ANKs are homologous with those present within the RAGE of *O. tsutsugamushi*, rather they are similar to the ANKs that are overrepresented in the *R. bellii* and *R. felis* genomes.
LPS	Lipopolysaccharide biosynthesis; Wzt (RfbB) (REIS_0020), Wzm (RfbA) (REIS_0021), LpxD (REIS_0029), LpxL (REIS_0064), LpxK (REIS_0065), LpxB (REIS_1040), LpxC (REIS_1297), WaaA (REIS_2012) full length. LpxA (REIS_0026) truncated by gap (G30). RfaG (TG-specific WaaG-like sugar transferase, MtfB-like), RfaG (TG-specific WaaG-like sugar transferase,WavL-like), and RfaJ (TG-specific WaaJ-like sugar transferase, A4GalT-like) are absent.
rvh T4SS	Rickettsiales *vir* homolog (*rvh*) P-like T4SS. Note: There are no known homologs to VirB5 in the *rvh* T4SS. RvhB2 (REIS_1676), RvhB3 (REIS_2293), RvhB4a (REIS_2293), RvhB4b (REIS_0264), RvhB6d (REIS_0606), RvhB7 (REIS_1250), RvhB8a (REIS_1251), RvhB8b (REIS_1249), and RvhB9a (REIS_1252) are full length. RvhB11 was not assembled due to gap (G68) but was found complete in the WGS reads. RvhB1 (REIS_2293) is a frameshift. RvhB6a (REIS_0602 + REIS_0603) is a split gene. RvhB6b (REIS_0604) and RvhB6c (REIS_0605) are truncated. RvhB6e (REIS_0607) is a frameshift. RvhB9b (REIS_1248) is truncated. RvhB10 absent but partially detected in *I. scapularis* genome data (gcontig_1108379345628). RvhD4 (REIS_1247 + REIS_2294) is split by gap (G67).
BarA	Histidine kinase sensor protein. Full length (REIS_1341).

PhoR	Histidine kinase phosphatase synthesis sensor protein. Full length (REIS_0144).
CzcR	Transcriptional activator protein. Full length (REIS_0432).
EnvZ	Histidine kinase osmolarity sensor protein. Full length (REIS_0748).
OmpR	Transcriptional activator protein. Full length (REIS_0747).
NtrY	Histidine kinase nitrogen sensor protein. Full length (REIS_1769).
NtrX	Nitrogen assimilation regulatory protein. Full length (REIS_1404).

[A] Red indicates probable pseudogenes.

[B] HP = hypothetical protein, TNP = transposase and inactive derivatives,

Table S10. Genes encoded within the RAGE-A piggyback region that are typical of aminoglycoside antibiotic resistance and biosynthesis gene clusters.

REIS ORF	**Top 5 Blastp subjects Taxon**	**Accession no.**	**Phyl**[A]	**Gene**[B]	**Score**	**Expect**
1491	*Streptomyces tenjimariensis*	CAI59977	Act	IstN	131	1e-28
	Micromonospora olivasterospora	CAF31551	Act	ForN	116	2e-24
	Frankia sp. CcI3	YP_482457	Act	SAM-MT	100	1e-19
	Trichodesmium erythraeum	YP_721874	Cya	SAM-MT	82.0	7e-14
	Synechococcus sp. CC9902	YP_377481	Cya	SAM-MT	55.5	6e-06
1492	*Streptomyces tenjimariensis*	CAI59979	Act	IstM	319	5e-85
	Streptomyces lividus	CAG38698	Act	LivM	224	2e-56
	Streptomyces fradiae	BAD95821	Act	NeoD	222	7e-56
	Streptomyces fradiae	CAF33313	Act	NeoM	220	2e-55
	Streptomyces kanamyceticus	BAD20768	Act	GT-1	220	3e-55
1493	*Bacillus circulans*	Q8G8Y2	Fir	BtrR	206	5e-51
	Streptomyces sp. C	ZP_07291433	Act	StsC	201	2e-49
	Streptomyces sp. ACT-1	ZP_06281476	Act	StsC	197	3e-48
	Streptomyces platensis	ADC52823	Act	StsC	197	3e-48
	Streptomyces kanamyceticus	BAD20758	Act	WecE	193	4e-47
1495	*Streptomyces* sp. DSM 40477	CAF33032	Act	AprU	68.9	7e-10
	Bacillus cereus AH1273	ZP_04174987	Fir	PKc_like	59.3	7e-07
	Streptoalloteichus hindustanus	CAI47639	Act	AprU	56.6	4e-06
	Bacillus weihenstephanensis	YP_001645433	Fir	PKc_like	54.3	2e-05
	Clostridium butyricum E4	ZP_04525810	Fir	PKc_like	52.0	1e-04
1496	REIS (REIS_0288)	ZP_04698620	Alp	HemL	899	0.0
	Streptomyces kanamyceticus	BAD20756	Act	HemL	239	7e-61
	Micromonospora olivasterospora	CAF31546	Act	ForB	238	1e-60
	Micromonospora echinospora	CAF31437	Act	GenB4	236	6e-60
	Frankia sp. CcI3	YP_482453	Act	HemL	234	1e-59
1500	*Streptomyces tenjimariensis*	CAH60159	Act	IstC	269	4e-70
	Streptoalloteichus hindustanus	BAE98082	Act	AlloH	263	3e-68
	Acetivibrio cellulolyticus CD2	ZP_07327975	Fir	DHQ_s	263	3e-68
	Streptomyces sp. KCTC 9047	ACR82896	Act	AmgC	257	2e-66
	Streptoalloteichus hindustanus	CAI47651	Act	AprC	255	6e-66
1501	*Micromonospora inyonensis*	ACN38356	Act	Tdh	157	2e-36
	Micromonospora inyonensis	ABX38722	Act	Tdh	156	2e-36
	Micromonospora echinospora	CAF34032	Act	GacH	155	7e-36
	Streptomyces tenebrarius	CAE22477	Act	TacD	154	1e-35
	Streptomyces rimosus subsp. *paromomycinus*	CAF32372	Act	ParE	154	2e-35
1502	*Micromonospora inyonensis*	ACN38340	Act	BetA	439	3e-121
	Streptomyces rimosus subsp. *paromomycinus*	CAF32378	Act	ParQ	424	1e-116
	Streptomyces lividus	CAG38701	Act	LivQ	416	3e-114
	Micromonospora echinospora	CAF34042	Act	GacJ	409	4e-112
	Streptomyces fradiae	BAD95824	Act	NeoG	408	8e-112
1503	*Micromonospora inyonensis*	ACN38339	Act	HemL	248	1e-63
	Bacillus circulans	BAE07066	Fir	BtrB	246	3e-63
	Streptomyces ribosidificus	CAG34709	Act	RbmH	241	1e-61

	Micromonospora echinospora	CAF34043	Act	GacK	238	9e-61
	Streptomyces rimosus subsp. *paromomycinus*	CAF32385	Act	ParB	234	2e-59
1504	*Streptomyces tenebrarius*	CAE22472	Act	TbmB	321	9e-86
	Bacillus circulans	Q8G8Y2	Fir	BtrR	206	5e-51
	Streptomyces sp. DSM 40477	CAF33048	Act	AprS	318	9e-85
	Streptomyces kanamyceticus	BAD20758	Act	WecE	317	3e-84
	Pseudomonas entomophila L48	YP_608081	Gam	WecE	313	2e-83
1505	*Pseudomonas entomophila* L48	YP_608079	Gam	PIG-L	101	8e-20
	Streptomyces rimosus subsp. *paromomycinus*	CAF32383	Act	ParD	100	2e-19
	Streptomyces sp. DSM 40477	CAH18559	Act	TobN	96.7	2e-18
	Streptomyces lividus	CAG38705	Act	LivD	94.7	8e-18
	Streptomyces tenjimariensis	CAH60149	Act	IstD	94.7	8e-18

[A] Major phylogenetic group: Act, Actinobacteria; Cya, Cyanobacteria; Fir, Firmicutes; Alp, *Alphaproteobacteria*; Gam, *Gammaproteobacteria*.

[B] Annotation of top hits (used in Fig. 6): IstN, methyltransferase; IstM, putative NDP-D-glucosaminyltransferase; Btr, L-glutamine:2-deoxy-scyllo-inosose aminotransferase; AprU, putative kinase; HemL, glutamate-1-semialdehyde 2,1-aminomutase 1; IstC, 2-deoxy-scyllo-inosose synthase; Tdh, 2-deoxy-scyllo-inosamine dehydrogenase; BetA (Sis6), putative sugar-alcohol dehydrogenase; HemL, glutamate-1-semialdehyde 2,1-aminomutase 1; TbmB (KanB), L-glutamine:2-deoxy-scyllo-inosose aminotransferase; PIG-L, putative amidase, AGA deacetylase.

Table S11. REIS genes with highly similar homologs in the "*Candidatus* Amoebophilus asiaticus" genome, and their distribution in other *Rickettsia* genomes.

Gene [A]	**Aasi** [B] **ORF**	**Top hit**	**REIS** [C] **ORF**	**Top hit**	**Dist.** [D]
AlkD	0285	*Rickettsia bellii* RML369-C (5e-75)	------		D
AmpD	0795	*Orientia tsutsugamushi* str. Boryong (5e-66)	------		D
AraC_E	1875	*Rickettsia bellii* RML369-C (5e-45)	------		D
RhlE	1916	*Wolbachia* (*Culex quinquefasciatus* Pel) (1e-123)	0046	Candidatus *Amoebophilus asiaticus* 5a2 (3e-99)	R
FabH	0115	*Streptomyces avermitilis* MA-4680 (32e-74)	0100	Candidatus *Puniceispirillum marinum* IMCC1322 (9e-92)	R
PanF	0446	*Orientia tsutsugamushi* str. Ikeda (2e-57)	0144	Candidatus *Amoebophilus asiaticus* 5a2 (6e-121)	D
ISRpe1	0884	*Rickettsia* endosymbiont of *Ixodes scapularis* (5e-169)	0170	Candidatus *Amoebophilus asiaticus* 5a2 (2e-172)	D
EamA	1859	*Rickettsia typhi* str. Wilmington (2e-68)	0437	Candidatus *Amoebophilus asiaticus* 5a2 (1e-65)	R
HP	0297	*Rickettsia akari* str. Hartford (5e-19)	0637-0640	Candidatus *Amoebophilus asiaticus* 5a2 (1e-16)	D
HP	0973	*Rickettsia bellii* RML369-C (3e-68)	0995	Candidatus *Amoebophilus asiaticus* 5a2 (1e-49)	D
SplB	1237	*Rickettsia felis* URRWXCal2 (1e-130)	0998	Candidatus *Amoebophilus asiaticus* 5a2 (3e-55)	D
ProP	1183	*Rickettsia canadensis* str. McKiel (1e-118)	1056	Candidatus *Amoebophilus asiaticus* 5a2 (5e-101)	P
TypA	0730	*Wolbachia* (*Culex quinquefasciatus* Pel) (0.0)	1285	*Magnetospirillum magnetotacticum* MS-1 (0.0)	D
PilT	1214	*Rickettsia bellii* RML369-C (8e-40)	1287	*Thioalkalivibrio* sp. HL-EbGR7 (2e-27)	D
MdlB	0731	*Legionella longbeachae* NSW150 (1e-150)	1370	*Parachlamydia acanthamoebae* str. Hall's coccus (1e-123)	R

Gene [A]	Aasi [B] ORF	Top hit	REIS [C] ORF	Top hit	Dist. [D]
HP	0712	*Parachlamydia acanthamoebae* str. Hall's coccus (3e-17)	1420	*Parachlamydia acanthamoebae* str. Hall's coccus (6e-12)	R
Taq	0861	*Rickettsia akari* str. Hartford (8e-155)	1530	Candidatus *Amoebophilus asiaticus* 5a2 (2e-152)	D
F-box	0025	*Nematostella vectensis* (2e-11)	1559	Candidatus *Amoebophilus asiaticus* 5a2 (2e-59)	D
HtrA	0325	*Rickettsia typhi* str. Wilmington (3e-104)	1710	Candidatus *Amoebophilus asiaticus* 5a2 (2e-98)	D
Antitoxin	1215	*Rickettsia bellii* OSU 85-389 (2e-21)	1775	Candidatus *Amoebophilus asiaticus* 5a2 (6e-05)	D
MdlB	0674	*Parachlamydia acanthamoebae* str. Hall's coccus (7e-126)	2217	*Parachlamydia acanthamoebae* str. Hall's coccus (2e-122)	R

[A] Genes identified as highly homologous across *Rickettsia* genomes and “*Candidatus* Amoebophilus asiaticus”, either here or previously (1).

[B] “*Candidatus* Amoebophilus asiaticus” ORF number and top blastp hit.

[C] Complementary REIS ORF number (if present) and top blastp hit.

[D] Distribution in other *Rickettsia* genomes: R, conserved; D, patchy; P, proliferated.

1. **Schmitz-Esser, S., P. Tischler, R. Arnold, J. Montanaro, M. Wagner, T. Rattei, and M. Horn.** 2010. The genome of the amoeba symbiont "*Candidatus* Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. J Bacteriol **192:**1045-1057.