Paradoxical evolution of rickettsial genomes

Awa Diop, Didier Raoult, Pierre-Edouard Fournier

Paradoxical evolution of rickettsial genomes

Awa Diop[1], Didier Raoult[2] and Pierre-Edouard Fournier[1*]

[1] UMR VITROME, Aix-Marseille University, IRD, Service de Santé des Armées, Assistance Publique-Hôpitaux de Marseille, Institut Hospitalo-Uuniversitaire Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille, France Tel: +33 413 732 401, Fax: +33 413 732 402.

[2] UMR MEPHI, Aix-Marseille University, IRD, Assistance Publique-Hôpitaux de Marseille, Institut Hospitalo-Uuniversitaire Méditerranée Infection, Marseille, France

[*]Corresponding author: Pr Pierre-Edouard Fournier

Email: pierre-edouard.fournier@univ-amu.fr

Abstract:

Rickettsia species are strictly intracellular bacteria that evolved approximately 150 million years ago from a presumably free-living common ancestor from the order *Rickettsiales* that followed a transition to an obligate intracellular lifestyle. Rickettsiae are best known as human pathogens vectored by various arthropods causing a range of mild to severe human diseases. As part of their obligate intracellular lifestyle, rickettsial genomes have undergone a convergent evolution that includes a strong genomic reduction resulting from progressive gene degradation, genomic rearrangements as well as a paradoxical expansion of various genetic elements, notably small RNAs and short palindromic elements whose role remains unknown. This reductive evolutionary process is not unique to members of the *Rickettsia* genus but is common to several human pathogenic bacteria. Gene loss, gene duplication, DNA repeat duplication and horizontal gene transfer all have shaped rickettsial genome evolution. Gene loss mostly involved amino-acid, ATP, LPS and cell wall component biosynthesis and transcriptional regulators, but with a high preservation of toxin-antitoxin (TA) modules, recombination and DNA repair proteins. Surprisingly the most virulent *Rickettsia* species were shown to have the most drastically reduced and degraded genomes compared to closely related species of milder pathogenesis. In contrast, the less pathogenic species harbored the greatest number of mobile genetic elements. Thus, this distinct evolutionary process observed in *Rickettsia* species may be correlated with the differences in virulence and pathogenicity observed in these obligate intracellular bacteria. However, future investigations are needed to provide novel insights into the evolution of genome sizes and content, for that a better understanding of the balance between proliferation and elimination of genetic material in these intracellular bacteria is required.

Keywords: *Rickettsia*, genomics, evolution, virulence, genome rearrangement, non-coding DNA, gene loss, DNA repeats

1 Introduction

The genus *Rickettsia* (order *Rickettsiales*, family *Rickettsiaceae*) comprises strictly intracellular α-proteobacteria mostly associated with diverse arthropod vectors around the world (Raoult and Roux, 1997; Stothard et al., 1994). *Rickettsia* species evolved approximately 150 million years ago from a common ancestor of *Rickettsiales* that was presumably free-living, and progressively followed a transition to an obligate intracellular lifestyle that occurred 775–525 million years ago and then to primarily infecting arthropod lineages approximately 525–425 million years ago (El Karkouri et al., 2016; Merhej and Raoult, 2011; Lucy A Weinert et al., 2009). These bacteria are also well known to infect mammalian hosts, mostly through arthropod bites or arthropod feces infecting scratching lesions. On the basis of their phenotypic properties, vector hosts and phylogenetic organization, *Rickettsia* species were split into three to four groups by different authors (Figure 1): i) the spotted fever group (SFG, Figure 1) contains many spotted fever-causing species as well as numerous species of as-yet unknown pathogenicity. SFG rickettsiae are mostly associated with ticks, but also fleas and mites (Diop et al., 2017); ii) the second phylogenetic group, the typhus group (TG, Figure 1) is only made of *R. prowazekii* and *R. typhi* that cause epidemic and murine typhus, and are associated with human body lice and rat fleas, respectively (Diop et al., 2017); iii) the ancestral group includes *R. bellii* and *R. canadensis*. These species diverged early from SFG and TG rickettsiae, are associated with ticks but do not cause human disease (Figure 1) (Diop et al., 2017); iv) a fourth group, named transitional group, was proposed by Gillespie *et al.* to include SFG species phylogenetically close to *R. felis* (Gillespie et al., 2007). However, as these species do not exhibit significant differences with other SFG species except their phylogenetic position, several authors discussed the validity of this latter group (Shpynov et al., 2018).

Rickettsia species cause a range of illnesses, from mild and self-limiting to severe and life-threatening diseases (Diop et al., 2017). Currently, the most common rickettsioses are African tick-bite fever caused by *R. africae*, scalp eschar and neck lymphadenopathy (SENLAT) caused by *R. slovaca*, Mediterranean spotted fever (MSF) caused by *R. conorii*, Rocky Mountain spotted fever (RMSF) caused by *R. rickettsii* and murine typhus caused by *R. typhi*. (El Karkouri et al., 2017; Parola et al., 2013; Sahni et al., 2013). *Rickettsia prowazekii*, the historical agent of epidemic typhus, is only rarely encountered currently but has a strong epidemic potential (Parola et al., 2013). Furthermore, recent studies have reported the association of other *Rickettsia* lineages with other reservoirs including protozoa, algae, leeches, plants or insects (Merhej and Raoult, 2011; Murray et al., 2016; Weinert et al., 2009).

In 1998, the first complete *Rickettsia* genome, that of *R. prowazekii* strain Madrid E, was sequenced (Andersson et al., 1998). It was the seventh bacterial genome to be sequenced. Subsequently, the genomes of many *Rickettsia* species have been fully sequenced, allowing a better knowledge of the molecular mechanisms involved in their pathogenicity (Balraj et al., 2009). Genome sequencing also appeared as a potential tool to revolutionize the phylogenetic and evolutionary investigations of prokaryotes, especially endosymbiotic bacteria. Hence, deciphering rickettsial genomes appeared as an efficient tool to understand the evolution of these obligate intracellular bacteria.

2 General features of rickettsia genomes

Rickettsia species have small genome sizes and low G+C contents. SFG and TG rickettsiae exhibit genome sizes of 1.25 to 2.3 Mb, and 1.11 Mb, respectively. They also exhibit G+C contents ranging from 32.2 to 33.0% and 28.9 to 29.0%, respectively. *Rickettsia* species have numbers of predicted protein-coding genes varying between 817 and 2,479 and most of them

maintain a near perfect chromosomal synteny (Diop et al., 2017), which enabled the identification of an ongoing and progressive genome degradation (Ogata, 2001). Rickettsial genomes contain many functional or unfunctional pseudogenes and possess a high percentage of non-coding DNA (Blanc et al., 2007; McLeod et al., 2004) (Fig. 2). This percentage of non-coding DNA ranges from 16.2% for *R. felis* to 31% for *R. massiliae*. *Rickettsia prowazekii*, the most reduced rickettsial genome contains 24% of non-coding sequence. By comparison, *Chlamydia trachomatis*, another strictly intracellular bacterium, possesses only 10% non-coding DNA (Andersson et al., 1998; Holste et al., 2000; Rogozin et al., 2002).This pseudogenization progressively leads to a genome downsizing and results from a switch from a free-living to an obligate intracellular lifestyle. This progressive reductive evolution has allowed rickettsiae to purge unnecessary and redundant genes mainly involved in metabolisms supplied by eukaryotic host cells (Georgiades and Raoult, 2011; Merhej et al., 2009). Paradoxically to this ongoing genomic reduction, rickettsial genomes exhibit another marker of convergent evolution, *i. e.*, the expansion of genetic elements including small RNAs, tandem repeats, short palindromic elements named rickettsia palindromic elements (RPEs) (Ogata et al., 2002), ankyrin and tetratricopeptide repeats and gene family duplication mainly ADP-ATP translocases, toxin-antitoxin modules and type IV secretion system (T4SS). Another unexpected property of rickettsial genomes is the presence of plasmids, the first described in obligate intracellular bacteria. The first plasmid was identified in *R. felis* (Ogata et al., 2005a). To date, at least 20 rickettsial plasmids have been described in 11 species. Their number varies from 1 to 4 per species/strain (Baldridge et al., 2007; G. Blanc et al., 2007; El Karkouri et al., 2016). These findings suggest possible exchanges of genetic material by conjugation, a mechanism that was thought to be absent in obligate intracellular and allopatric bacteria (Georgiades and Raoult, 2011; Merhej et al., 2009; Ogata et al., 2005a).

3 Rickettsia genome in an ongoing convergent evolution

3.1 Ongoing reductive evolution of Rickettsial genomes

Following their adaptation from a free-living to an obligate intracellular lifestyle in eukaryotic cells, rickettsiae underwent genomic changes to fit their specific bottleneck ecosystem, resulting not only in a reducing genome size but also in a specific genomic architecture (Keeling et al., 1994; Sicheritz-Pontén and Andersson, 1997). Comparative genomics revealed that rickettsiae, by taking advantage of host cell metabolites, underwent a genome reductive evolution (Georgiades and Raoult, 2011; Merhej et al., 2009) that occurred through a progressive pseudogenization (Fig. 2) and gene loss of selected biosynthetic pathway components (Andersson et al., 1998; Audia and Winkler, 2006; Fournier et al., 2009; Ogata, 2001; Sakharkar, 2004; Walker, 2005; Wolf and Koonin, 2013). In addition, genomic degradation was detrimental for the G+C content, as it led to an enrichment in A+T, in particular in the high proportion of non-coding DNA (Sakharkar, 2004). However, a great variation in chromosome size, ranging from 1.1 to 2.3 Mb, is observed in rickettsiae (Diop et al., 2017), indicating that some species are at a more advanced stage of reductive genomic evolution (TG rickettsiae) than others (SFG rickettsiae) (Ogata, 2001). In ehrlichiae, a similar genomic reduction is observed, but the G+C content may remain as high as 49.8% in *Anaplasma* species (Dunning Hotopp et al., 2006), suggesting that the reductive process in these bacteria had a lesser impact on the G+C content degradation. Rickettsial genomes are characterized by a high rate of accumulation of slightly harmful deletions, mutations and insertions (Brynnel et al., 1998). Alternatively, gene loss can also result from the accumulations of small mutations. The formation of internal stop codons within intact genes can occur through the creation of a frameshit by single base mutation, insertion or deletion (Ogata, 2001). This induces the genome degradation resulting from fragmented gene accumulation or gene remnants. An unexpected finding of rickettsial genomics was that the

most virulent species had the most reduced genomes (Fournier et al., 2009). Such a finding is not an isolated phenomenon as in *Mycobacterium*, *Streptococcus* spp., *Corynebacterium* spp. and other genera, the highest degree of gene loss is observed in the most virulent species when compared to closely related and milder or nonpathogenic species (Blanc et al., 2007; Merhej et al., 2013; Ogata, 2001). Many of the genes required by free-living bacteria are absent in *Rickettsia* (Bechah et al., 2010) and degraded genes include mostly those involved in the biosynthesis of nutrients (Blanc, 2005; Ogata, 2001; Renesto et al., 2005). For example, *Rickettsia* exhibits few genes for de novo nucleotide synthesis, *i. e.,* only those for conversion of nucleoside monophosphates into all other nucleotides, implying that they take up nucleoside monophosphates from the host (Wixon, 2001). Analysis of *R. conorii* and *R. prowazekii* genomes (Dunning Hotopp et al., 2006; Ogata, 2001) revealed that genes coding glycolytic enzymes and those required for nucleotide or cofactor biosynthesis are totally absent in *R. conorii* and *R. prowazekii* when compared to most genera in the order *Rickettsiales* that have complete glycolytic pathways. Nevertheless, rickettsiae must obtain glycerol-3-phosphate from the host via a glycerol-3-phosphate transporter (Dunning Hotopp et al., 2006). This ATP production profile is similar for *Rickettsia* and mitochondria, as they possess a high number of ATP/ADP translocases, suggesting that they have both evolved from a common ancestor (Andersson et al., 1998; Renesto et al., 2005). In addition, the genome sequencing of *R. prowazekii* revealed a lack of amino acid metabolism such as those for glutamate metabolism (Andersson et al., 1998; Fuxelius et al., 2007). The enzymes involved in the aspartate and alanine metabolism pathways, and those playing a role in the biosynthesis of leucine, valine, isoleucine and aromatic amino acids (tryptophan, tyrosine, phenylalanine) are similarly missing in *Rickettsia* species (Renesto et al., 2005), suggesting the use of host-derived amino acids for their growth, survival and replication. Additionally, all *Rickettsia* species except *R. belli* have a reduced set of folate biosynthesis genes (Fuxelius et

al., 2007). In TG rickettsiae all five genes required for the de novo folate biosynthesis are lacking (Hunter et al., 2015). Furthermore, a limited set of genes for LPS and cell wall component biosynthesis, including lipid-A and peptidoglycan, respectively, were identified in *Rickettsia* species (Fuxelius et al., 2007). The rickettsial surface protein-coding genes *rickA* and *sca*2 are another example of genes that were degraded or eliminated by *Rickettsia* species during their specialization. The RickA protein participates in actin polymerization through the activation of Arp2/3 similar to that found in *Listeria monocytogenes* and *Shigella* spp. (Balraj et al., 2008b; Gouin et al., 2004, 1999). While lacking in the TG, *rickA* is present in all AG and SFG rickettsial genomes avalaible (Baldridge et al., 2005; Balraj et al., 2008a, 2008b; Heinzen et al., 1993; Jeng et al., 2004; McLeod et al., 2004; Ogata, 2001; Ogata et al., 2006, 2005a). The absence of *rickA* in *R. prowazekii* is not surprising if we consider its lack of actin motility. In contrast, *R. typhi* exhibits a unique and erratic actin-based motility despite having a nonfunctional RickA protein (McLeod et al., 2004; Reed et al., 2014). In addition, *R. canadensis* expresses RickA but does not exhibit actin-based motility (Heinzen et al., 1993). These data suggest the possible involvement of other actin polymerization mechanisms and that RickA alone may not be sufficient or required for actin-based rickettsial motility. Nevertheless, it was proposed that RickA originated early in rickettsial evolution and may have been lost during the divergence of the TG. Recent research suggests that *Rickettsia* spp. use also Sca2 for actin-based motility with a distinct mechanism compared to RickA. Sca2 was found to be intact in *R. conorii*, absent in *R. prowazekii* and pseudogenized in *R. typhi* (McLeod et al., 2004). In *R. typhi*, Sca2 lacks the FH1 (formin homology 1) domain and contains only a proline-rich tract and a series of five WH2 domains (β-domains) in different locations with a divergence in sequences (Sears et al., 2012). The evolutionary process of genome degradation in rickettsiae led to loss of transcriptional regulator genes with a decreased translational capacity as observed in *R. prowazekii* (Andersson and Kurland, 1998),

despite conserved gene sets coding for toxins, toxin-antitoxin (TA) modules and recombination and DNA repair proteins most likely needed for protection against host immune response (Moran, 2002).

The reductive evolution of rickettsial genomes is not only the consequence of gene degradation or loss, but it is also linked to a differential expression level of genes (Diop et al., 2017). Some genes under the influence of evolutionary forces are dormant or repressed while others under this effect are overexpressed. Recent research involving two virulent and two milder SFG rickettsiae demonstrated that the two virulent agents *R. conorii* (MSF) and *R. slovaca* (SENLAT) have the most reduced genome and displayed less up-regulated than down-regulated genes than the milder *R. massiliae* and *R. raoultii* causing MSF and SENLAT, respectively (El Karkouri et al., 2017), that have less reduced genomes. Consequently, to adapt to their specific intracellular environment, *Rickettsi*a species were shaped by distinct evolutionary processes. The most pathogenic species are characterized by a strong reductive genomic evolution, with a higher genome degradation rate and accumulation of non-coding DNA than less pathogenic species. These findings suggest that reductive genomic evolution, resulting in protein structural variations, is associated to the emergence of virulence (El Karkouri et al., 2017). It was speculated that the loss of regulator genes, as observed in several intracellular pathogens, is a critical cause of virulence (Darby et al., 2007). This reductive genomic evolution appears to have occurred in several other human pathogens that have no common intracellular ancestor with *Rickettsia* such as *Treponema* spp., *Mycobacterium* spp. or *Yersinia* spp (Merhej et al., 2009; Walker, 2005; Wixon, 2001). Overall, during the course of evolution, rickettsial genomes exhibit a trend toward gene loss rather than acquisition, but strong selective effects co-exist with functional duplication required for survival.

3.2 Gene order, recombination events and ‘‘junk DNA’’ in rickettsial genomes

A comparison of 13 rickettsial genomes (Diop et al., 2017) demonstrated that they exhibit a highly conserved synteny and present few genomic rearrangements, except for *R. bellii* that exhibits little colinearity with other genomes, and *R. felis* that underwent several inversions. In addition, *R. typhi*, underwent a 35-kb inversion close to the replication terminus and a specific 124-kb inversion nearby the origin of replication when compared to *R. prowazekii* and *R. conorii* (McLeod et al., 2004). As in other bacteria, inversions that occured in the origin of replication region are also found in *R. australis*, *R. helvetica* and *R. honei* (X. Dong et al., 2012; Xin Dong et al., 2012; Xin et al., 2012), indicating that this region constitutes a hotspot for genomic rearrangement (Eisen et al., 2000). Homologous intra-chromosomal recombination, the principal mechanism for genomic rearrangement in rickettsiae, occured between repeated sequences or by site-specific recombination. Consequently, duplications, deletions and inversions arose through these structures (Andersson and Kurland, 1998; Krawiec and Riley, 1990). Such events have been observed in Rickettsia spp., in the so-called super-ribosomal protein gene operon. Highly conserved in a broad range of bacteria and archaea, this operon consists of about 40 genes located in seven operons in the same order (Sicheritz-Pontén and Andersson, 1997). Despite their conserved order in many bacteria including *E. coli* and *Bacillus subtilis*, genes in the ribosomal protein gene operon are scattered around the genomes of *Haemophilus influenzae*, *Mycoplasma genitalium* and *R. prowazeki* (Andersson and Kurland, 1998; Fraser et al., 1995). Ribosomal RNA genes in bacterial genomes are normally organized into an operon with a conserved order 16S-23S-5S, and tRNA genes are often found in the spacer between the 16S and the 23S rRNA genes (Krawiec and Riley, 1990). However, an unusual arrangement of rRNA genes has been observed in all avalaible *Rickettsia* genomes, as the 16S rRNA gene is separated from the 23S and 5S rRNA gene cluster (Andersson et al., 1999; Munson et al., 1993). A similar

219 organization is observed in all members of the order *Rickettsiales* (Dunning Hotopp et al.,
220 2006). The upstream spacer of the rearranged 23S rRNA gene in some *Rickettsia* species
221 contains short repetitive sequences that have been eliminated in other related species,
222 suggesting that the rearrangement of rRNA genes occurred by intra-chromosomal
223 recombination prior to speciation in *Rickettsia* spp. Rickettsial genome analysis highlighted a
224 second major genomic rearrangement in rickettsiae, the elongation factor proteins (tuf and
225 fus) being present in more than one copy in *Rickettsia* genomes (Syvänen et al., 1996). These
226 genes can serve as repeat sequences, and initiate a rapid gene loss through intra-chromosomal
227 recombination (Krawiec and Riley, 1990). In addition, the degree and positions of deletions
228 caused by intra-chromosomal recombination in *Rickettsia* is different among the species,
229 which suggests that the homologous recombination is an ongoing process that may result in
230 an ongoing genes loss under weak or no selection pressure.

231 When compared to other bacterial genomes, rickettsial genomes have a high percentage of
232 non-coding DNA sequences which also contains many DNA repeat sequences (Holste et al.,
233 2000; Rogozin et al., 2002). Non-coding DNA in rickettsial genomes is traditionally
234 considered as "junk DNA" resulting from gene degradation. *R. prowazekii* and *R. typhi*, the
235 most reduced rickettsial genomes, harbor high rates of non-coding DNA with 24.6 and 23.7%,
236 respectively. However, *R. bellii* exhibits the lowest rickettsial level of non-coding DNA with
237 14.8% (Diop et al., 2017).

238

239 **3.3 Paradoxical genomic expansions**

240 From a general point of view, rickettsial genomes are typical of those of symbiotic
241 bacteria, in which the reductive trend is the dominant mode of evolution (Andersson and
242 Andersson, 1999; Georgiades and Raoult, 2011; Merhej et al., 2009; Ogata, 2005). However,

despite this reductive evolution, a paradoxical expansion of genetic elements can still occur in rickettsial genomes (Ogata et al., 2002). Genome sequence analysis revealed that rickettsial genome expansion may occur through proliferation of selfish DNA (small non coding RNAs (sRNAs) and rickettsia palindromic elements (RPEs)), gene duplications and horizontal gene transfer (Merhej and Raoult, 2011). Bacterial non-coding RNAs, whose biogenesis is predominantly attributed to either the intergenic regions (trans-acting) or to the antisense strand of an open reading frame (cis-acting) (Schroeder et al., 2015), were well documented in many bacterial taxa including *Enterobacteriaceae*, *Listeria monocytogenes*, *Clostridium perfringens*, *Staphylococcus aureus, Pseudomonas aeruginosa* and *Mycobacterium tuberculosis* (Papenfort and Vanderpool, 2015). sRNAs are classified among the most important post-transcriptional regulators involved in virulence and adaptation depending on the host niche, through transcriptomic regulation (Schroeder et al., 2015). Schroeder et al. (2015) were the first to identify sRNAs in *Rickettsia* species. Twenty to 30% of intergenic regions presumably encode for trans-acting sRNAs (14 to 191 sRNAs, depending on species). These findings may explain the highly conserved intergenic spacers identified by early comparative studies in *Rickettsia* (Ogata, 2001). More than 1,700 trans-acting sRNAs were predicted in 16 genomes of 13 species spanning all rickettsial groups (Schroeder et al., 2015). *Rickettsia prowazekii* was shown to possess stem loop structures after homopolymeric poly(T) stretches in the termination sites where the expression of sRNAs occurs (Woodard and Wood, 2011). Rickettsia palindromic elements (RPEs) were identified in 2002 by Ogata et al. (Ogata et al., 2002). These genetic elements are more abundant in SFG than TG rickettsiae (Figure 2). In the *R. conorii* genome, a total of 656 RPEs, classified into 8 families, were identified (RPE-1 to RPE-8) and represent 3.2% of the entire genome (Ogata et al., 2002). By comparison, only 10 of the 44 RPE-1 copies described in *R. conorii* were found in the *R. prowazekii* genome. Surprisingly, nine of these 10 RPE-1 copies that are present in *R.*

prowazekii are inserted in protein-coding genes, *versus* 19/44 in *R. conorii*. In addition, the RPE-1s inserted into protein-coding genes have a position compatible with the 3-dimentional fold and function of proteins (Ogata et al., 2000). This process of genomic evolution by inserting RPEs within protein-coding genes was initially thought to be unique to *Rickettsia* species but is also encountered in the *Wolbachia* genus (Ogata et al., 2005b; Riegler et al., 2012). Bacteria may use this random strategy to adapt their genetic repertoire in response to selective environmental pressure. The presence of a mobile element inserted in many unrelated genes also suggests the potential role of selfish DNA in rickettsial genome for de novo creation of new protein sequences during the course of evolution, suggesting an implication in the dynamics of genome evolution (Claverie and Ogata, 2003). Moreover, genomic comparison also enabled the identification of several copies of Ankyrin and Tetratricopeptide (TPR)-repeats in rickettsiae. Such repeated elements are frequently found in endosymbionts and assumed to play a role in host-pathogen interaction (Caturegli et al., 2000; Felsheim et al., 2009; Seshadri et al., 2003; Wu et al., 2004). Twenty-two copies of ankyrin- and 11 copies of TPR-repeats were found in *R. felis* (Ogata et al., 2005a). In both species, they were proposed to be linked to pathogenicity. In *Legionella pneumophila*, which exhibits 20 Ankyrin-repeat copies and numerous TPR-repeat copies, these elements are suspected to play a modulatory role in the interactions with the host cytoskeleton and in interferences with the host cell trafficking events, respectively (Cazalet et al., 2004).

In addition to DNA repeat sequences, various gene families are duplicated in rickettsial genomes. Gene duplication was considered as an important source of bacterial adaptation to environmental changes in the host (Hooper, 2003). Following duplication, gene copies can evolve by conserving the same functions or undergoing mutations and becoming non-functional or assuming new functions, thus providing a putative new selective advantage in a new environment (Greub and Raoult, 2003; Walsh, 1995). *Rickettsia prowazekii*, the most

293 reduced and degraded rickettsial genome that lacks the genes encoding the biosynthesis of
294 purines and pyrimidines (Andersson et al., 1998), exhibits five copies of *tlc1* genes. These
295 genes encode ADP/ATP translocases responsible of energy exploitation from host cells
296 (Greub and Raoult, 2003; Renesto et al., 2005). Similar sequences were found in *R. typhi*, *R.*
297 *rickettsii* and *R. montanensis*. Thus, the duplication of the tlc genes in *Rickettsia* is most likely
298 explained by their important role in maintaining an efficient uptake and transport system of
299 host cytoplasmic. ATP Four to 14 copies of *spoT* genes, involved in stringent response and
300 the adaptation to intracellular environment, were also found in rickettsiae (Ogata et al., 2005a;
301 Renesto et al., 2005; Rovery et al., 2005). The *R. conorii* genome has multiple copies of
302 ampG agent encoding β-lactamase, which may explain the resistance of these bacteria to β-
303 lactam antibiotics (Ogata, 2001). The T4SS, a multiple component, membrane-spanning
304 transporter system containing eight distinct classes such as the MPF-T class (P-T4SSs), is
305 largely found in many rickettsial genomes. Rickettsiae possess an incomplete P-T4SS system
306 (related to systems of the IncP group conjugative plasmid) that is characterized by the lack of
307 *virB*5 but the duplication of the *virB*4, *virB*6, *virB*8 and *virB*9 genes (Gillespie et al., 2016).
308 The *R. prowazekii* genome has six Vir components (*virB*4, *virB*8-*virB*11, *virD*4), and the
309 *virB*4 and *virB*9 were duplicated (Gillespie et al., 2009). Seventeen orthologous surface cell
310 antigen-coding genes (*sca*) were identified in rickettsial genomes (Blanc, 2005). SCA proteins
311 autotransporter proteins that were demonstrated to play roles in mammalian cell infection as
312 well as infection of their arthropod host cells, notably by promoting actin-based motility
313 (Sears et al., 2012). The *R. bellii* genome possesses a set of complete conjugation genes, and
314 pilli like-filaments were observed on the bacterial surface (Ogata et al., 2006). Among 13
315 tested Rickettsia collection strains, 11 got positive conjugation gene detection. This suggests
316 that the conjugation elements are widely present among Rickettsia spp (88), and that
317 horizontal gene transfer (HGT) occured at a high rate (Weinert et al., 2009). Within amoebae,

HGTs have given the *Rickettsia* ancestor the access to novel gene pools, with possibility to acquire foreign DNA from other intracellular bacteria, thus, in capability of adaptation environment (Ogata et al., 2006). In addition, a RAGE module, considered as a genetic exchange facilitator, was found in multiple copies in the genome from *Rickettsia* endosymbiont of *Ixodes scapularis* (REIS), the largest rickettsial genome described to date (Gillespie et al., 2014, 2012).

Finally, a large number of mobile genetic elements (MGEs) referred to as mobilome are found in rickettsiae despite their reduced genome size. This mobilome, mostly consisting of plasmids, may ensure DNA movement within and between genomes. To date, at least 20 known rickettsial plasmids have been described in 11 species despite their allopatric lifestyle (Diop et al., 2017). Recent phylogenomic analysis revealed that rickettsial plasmids are undergoing reductive evolutionary events similar to those affecting their co-residing chromosomes (El Karkouri et al., 2016). Rickettsial plasmids were thus shaped by a biphasic model of convergent evolution including a strong reductive evolution as well as an increased complexity via horizontal gene transfer and gene duplication and genesis (El Karkouri et al., 2016). The most reduced and virulent rickettsial genomes have probablely lost plasmid(s) during their evolution when compared to the related milder or non pathogenic species (Darby et al., 2007; El Karkouri et al., 2017; Ogata et al., 2005a).

4 Conclusions and Perspectives

Rickettsia species are strictly intracellular bacteria that are likely to have evolved from a presumably free-living ancestor and followed a transition to an obligate intracellular lifestyle. To adapt to such a bottleneck lifestyle associated with genetic drift, *Rickettsia* species have been shaped by distinct evolutionary processes resulting not only in differences in genome size, but also in genomic architecture. Generally, rickettsial genomes are small and contain a

high ratio of non-coding DNA, which suggests that the reductive trend is their dominant mode of evolution. Comparative sequence analysis has provided important clues on the mechanisms driving the genome-reduction process of *Rickettsia* spp. This phenomenon is marked by a selected loss of genes such as those associated with amino-acid, ATP, LPS and cell wall component biosynthesis with a loss of regulatory genes and a high preservation of toxin-associated proteins and toxin-antitoxin modules. Homologous intra-chromosomal recombination, principal mechanism for genomic rearrangement structures seems play a role in rapid gene loss. Consequently, rickettsiae have evolved under a distinct process including a strong reductive evolution as well as a paradoxical expansion of genetic elements acquired by horizontal gene transfer and gene duplication and genesis. Thus, during the course of evolution, rickettsial genomes had a trend of gene loss rather than gene acquisition or duplication, but these strong selective effects co-exist with functional duplications required for survival. In order to understand the evolution of genome size and content, it is necessary to understand the balance between proliferation and elimination of genetic material in these intracellular bacteria.

5 References

Andersson, J.O., Andersson, S.G., 1999. Genome degradation is an ongoing process in *Rickettsia*. Mol. Biol. Evol. 16, 1178–1191. https://doi.org/10.1093/oxfordjournals.molbev.a026208

Andersson, S.G.., Kurland, C.G., 1998. Reductive evolution of resident genomes. Trends Microbiol. 6, 263–268. https://doi.org/10.1016/S0966-842X(98)01312-2

Andersson, S.G., Stothard, D.R., Fuerst, P., Kurland, C.G., 1999. Molecular phylogeny and rearrangement of rRNA genes in *Rickettsia* species. Mol. Biol. Evol. 16, 987–995. https://doi.org/10.1093/oxfordjournals.molbev.a026188

Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C.M., Podowski, R.M., Näslund, A.K., Eriksson, A.-S., Winkler, H.H., Kurland, C.G., 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396, 133–140.

Audia, J.P., Winkler, H.H., 2006. Study of the five *Rickettsia prowazekii* proteins annotated as ATP/ADP translocases (Tlc): only Tlc1 transports ATP/ADP, while Tlc4 and Tlc5 transport other ribonucleotides. J. Bacteriol. 188, 6261–6268. https://doi.org/10.1128/JB.00371-06

Baldridge, G.D., Burkhardt, N., Herron, M.J., Kurtti, T.J., Munderloh, U.G., 2005. Analysis of fluorescent protein expression in transformants of *Rickettsia monacensis*, an obligate intracellular tick symbiont. Appl. Environ. Microbiol. 71, 2095–2105. https://doi.org/10.1128/AEM.71.4.2095-2105.2005

Baldridge, G.D., Burkhardt, N.Y., Felsheim, R.F., Kurtti, T.J., Munderloh, U.G., 2007. Transposon insertion reveals pRM, a plasmid of *Rickettsia monacensis*. Appl. Environ. Microbiol. 73, 4984–4995. https://doi.org/10.1128/AEM.00988-07

Balraj, P., Karkouri, K.E., Vestris, G., Espinosa, L., Raoult, D., Renesto, P., 2008a. RickA Expression is not sufficient to promote actin-based motility of *Rickettsia raoultii*. PLoS ONE 3, e2582. https://doi.org/10.1371/journal.pone.0002582

Balraj, P., Nappez, C., Raoult, D., Renesto, P., 2008b. Western-blot detection of RickA within spotted fever group rickettsiae using a specific monoclonal antibody. FEMS Microbiol. Lett. 286, 257–262. https://doi.org/10.1111/j.1574-6968.2008.01283.x

Balraj, P., Renesto, P., Raoult, D., 2009. Advances in *Rickettsia* pathogenicity. Ann. N. Y. Acad. Sci. 1166, 94–105. https://doi.org/10.1111/j.1749-6632.2009.04517.x

Bechah, Y., El Karkouri, K., Mediannikov, O., Leroy, Q., Pelletier, N., Robert, C., Medigue, C., Mege, J.L., Raoult, D., 2010. Genomic, proteomic, and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities. Genome Res. 20, 655–663. https://doi.org/10.1101/gr.103564.109

Blanc, G., 2005. Molecular evolution of *Rickettsia* surface antigens: evidence of positive selection. Mol. Biol. Evol. 22, 2073–2083. https://doi.org/10.1093/molbev/msi199

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

Blanc, G., Ogata, H., Robert, C., Audic, S., Suhre, K., Vestris, G., Claverie, J.-M., Raoult, D., 2007. Reductive genome evolution from the mother of *Rickettsia*. PLoS Genet 3, e14.

Brynnel, E.U., Kurland, C.G., Moran, N.A., Andersson, S.G., 1998. Evolutionary rates for tuf genes in endosymbionts of aphids. Mol. Biol. Evol. 15, 574–582.

Caturegli, P., Asanovich, K.M., Walls, J.J., Bakken, J.S., Madigan, J.E., Popov, V.L., Dumler, J.S., 2000. *ankA*: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. Infect. Immun. 68, 5277–5283.

Cazalet, C., Rusniok, C., Brüggemann, H., Zidane, N., Magnier, A., Ma, L., Tichit, M., Jarraud, S., Bouchier, C., Vandenesch, F., Kunst, F., Etienne, J., Glaser, P., Buchrieser, C., 2004. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. Nat. Genet. 36, 1165–1173. https://doi.org/10.1038/ng1447

Claverie, J.-M., Ogata, H., 2003. The insertion of palindromic repeats in the evolution of proteins. Trends Biochem. Sci. 28, 75–80. https://doi.org/10.1016/S0968-0004(02)00036-1

Darby, A.C., Cho, N.-H., Fuxelius, H.-H., Westberg, J., Andersson, S.G.E., 2007. Intracellular pathogens go extreme: genome evolution in the *Rickettsiales*. Trends Genet. 23, 511–520. https://doi.org/10.1016/j.tig.2007.08.002

Diop, A., Raoult, D., Fournier, P.-E., 2017. Rickettsial genomics and the paradigm of genome reduction associated with increased virulence. Microbes Infect. https://doi.org/10.1016/j.micinf.2017.11.009

Dong, X., El Karkouri, K., Robert, C., Gavory, F., Raoult, D., Fournier, P.-E., 2012. Genomic comparison of *Rickettsia helvetica* and other *Rickettsia* species. J. Bacteriol. 194, 2751–2751. https://doi.org/10.1128/JB.00299-12

Dong, X., El Karkouri, K., Robert, C., Raoult, D., Fournier, P.-E., 2012. Genome Sequence of *Rickettsia australis*, the agent of queensland tick typhus. J. Bacteriol. 194, 5129. https://doi.org/10.1128/JB.01117-12

Dunning Hotopp, J.C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S.V., Eisen, J., Seshadri, R., Ren, Q., Wu, M., Utterback, T.R., Smith, S., Lewis, M., Khouri, H., Zhang, C., Niu, H., Lin, Q., Ohashi, N., Zhi, N., Nelson, W., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sundaram, J., Daugherty, S.C., Davidsen, T., Durkin, A.S., Gwinn, M., Haft, D.H., Selengut, J.D., Sullivan, S.A., Zafar, N., Zhou, L., Benahmed, F., Forberger, H., Halpin, R., Mulligan, S., Robinson, J., White, O., Rikihisa, Y., Tettelin, H., 2006. Comparative genomics of emerging human ehrlichiosis agents. PLoS Genet. 2, e21. https://doi.org/10.1371/journal.pgen.0020021

Eisen, J.A., Heidelberg, J.F., White, O., Salzberg, S.L., 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol. 1, research0011–1.

El Karkouri, K., Kowalczewska, M., Armstrong, N., Azza, S., Fournier, P.-E., Raoult, D., 2017. Multi-omics analysis sheds light on the evolution and the intracellular lifestyle strategies of spotted fever group *Rickettsia* spp. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.01363

El Karkouri, K., Mediannikov, O., Robert, C., Raoult, D., Fournier, P.-E., 2016. Genome sequence of the tick-borne pathogen *Rickettsia raoultii*. Genome Announc. 4, e00157–16. https://doi.org/10.1128/genomeA.00157-16

Felsheim, R.F., Kurtti, T.J., Munderloh, U.G., 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

Fournier, P.-E., El Karkouri, K., Leroy, Q., Robert, C., Giumelli, B., Renesto, P., Socolovschi, C., Parola, P., Audic, S., Raoult, D., 2009. Analysis of the *Rickettsia africae* genome reveals that virulence acquisition in *Rickettsia* species may be explained by genome reduction. BMC Genomics 10, 166. https://doi.org/10.1186/1471-2164-10-166

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, R.D., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A., Venter, J.C., 1995. The minimal gene complement of *Mycoplasma genitalium*. Science 270, 397–403.

Fuxelius, H.-H., Darby, A., Min, C.-K., Cho, N.-H., Andersson, S.G.E., 2007. The genomic and metabolic diversity of *Rickettsia*. Res. Microbiol. 158, 745–753. https://doi.org/10.1016/j.resmic.2007.09.008

Georgiades, K., Raoult, D., 2011. Genomes of the most dangerous epidemic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. PLoS ONE 6, e17962. https://doi.org/10.1371/journal.pone.0017962

Gillespie, J.J., Ammerman, N.C., Dreher-Lesnick, S.M., Rahman, M.S., Worley, M.J., Setubal, J.C., Sobral, B.S., Azad, A.F., 2009. An Anomalous type IV secretion system in *Rickettsia* Is evolutionarily conserved. PLoS ONE 4, e4833. https://doi.org/10.1371/journal.pone.0004833

Gillespie, J.J., Beier, M.S., Rahman, M.S., Ammerman, N.C., Shallom, J.M., Purkayastha, A., Sobral, B.S., Azad, A.F., 2007. Plasmids and rickettsial evolution: insight from *Rickettsia felis*. PLoS ONE 2, e266. https://doi.org/10.1371/journal.pone.0000266

460 Gillespie, J.J., Joardar, V., Williams, K.P., Driscoll, T., Hostetler, J.B., Nordberg, E., Shukla, M., Walenz,
461 B., Hill, C.A., Nene, V.M., Azad, A.F., Sobral, B.W., Caler, E., 2012. A *Rickettsia* genome
462 overrun by mobile genetic elements provides insight into the acquisition of genes
463 characteristic of an obligate intracellular lifestyle. J. Bacteriol. 194, 376–394.
464 https://doi.org/10.1128/JB.06244-11
465 Gillespie, J.J., Kaur, S.J., Rahman, M.S., Rennoll-Bankert, K., Sears, K.T., Beier-Sexton, M., Azad, A.F.,
466 2014. Secretome of obligate intracellular *Rickettsia*. FEMS Microbiol. Rev. n/a–n/a.
467 https://doi.org/10.1111/1574-6976.12084
468 Gillespie, J.J., Phan, I.Q.H., Driscoll, T.P., Guillotte, M.L., Lehman, S.S., Rennoll-Bankert, K.E.,
469 Subramanian, S., Beier-Sexton, M., Myler, P.J., Rahman, M.S., Azad, A.F., 2016. The *Rickettsia*
470 type IV secretion system: unrealized complexity mired by gene family expansion. Pathog. Dis.
471 74, ftw058. https://doi.org/10.1093/femspd/ftw058
472 Gouin, E., Egile, C., Dehoux, P., Villiers, V., Adams, J., Gertler, F., Li, R., Cossart, P., 2004. The RickA
473 protein of *Rickettsia conorii* activates the Arp2/3 complex. Nature 427, 457.
474 Gouin, E., Gantelet, H., Egile, C., Lasa, I., Ohayon, H., Villiers, V., Gounon, P., Sansonetti, P.J., Cossart,
475 P., 1999. A comparative study of the actin-based motilities of the pathogenic bacteria *Listeria*
476 *monocytogenes, Shigella flexneri* and *Rickettsia conorii*. J. Cell Sci. 112, 1697–1708.
477 Greub, G., Raoult, D., 2003. History of the ADP/ATP-translocase-encoding gene, a parasitism gene
478 transferred from a *Chlamydiales* ancestor to plants 1 billion years ago. Appl. Environ.
479 Microbiol. 69, 5530–5535. https://doi.org/10.1128/AEM.69.9.5530-5535.2003
480 Heinzen, R.A., Hayes, S.F., Peacock, M.G., Hackstadt, T., 1993. Directional actin polymerization
481 associated with spotted fever group *Rickettsia* infection of vero cells. Infect. Immun. 61,
482 1926–1935.
483 Holste, D., Weiss, O., Grosse, I., Herzel, H., 2000. Are noncoding sequences of *Rickettsia prowazekii*
484 remnants of ``neutralized'' genes? J. Mol. Evol. 51, 353–362.
485 https://doi.org/10.1007/s002390010097
486 Hooper, S.D., 2003. On the nature of gene innovation: duplication patterns in microbial genomes.
487 Mol. Biol. Evol. 20, 945–954. https://doi.org/10.1093/molbev/msg101
488 Hunter, D.J., Torkelson, J.L., Bodnar, J., Mortazavi, B., Laurent, T., Deason, J., Thephavongsa, K.,
489 Zhong, J., 2015. The *Rickettsia* endosymbiont of *Ixodes pacificus* contains all the genes of de
490 novo folate biosynthesis. PloS One 10, e0144552.
491 Jeng, R.L., Goley, E.D., D'Alessio, J.A., Chaga, O.Y., Svitkina, T.M., Borisy, G.G., Heinzen, R.A., Welch,
492 M.D., 2004. A *Rickettsia* WASP-like protein activates the Arp2/3 complex and mediates actin-
493 based motility: *Rickettsia* RickA activates the Arp2/3 complex. Cell. Microbiol. 6, 761–769.
494 https://doi.org/10.1111/j.1462-5822.2004.00402.x
495 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software Version 7:
496 Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
497 https://doi.org/10.1093/molbev/mst010
498 Keeling, P.J., Charlebois, R.L., Ford Doolittle, W., 1994. Archaebacterial genomes: eubacterial form
499 and eukaryotic content. Curr. Opin. Genet. Dev. 4, 816–822. https://doi.org/10.1016/0959-
500 437X(94)90065-5
501 Krawiec, S., Riley, M., 1990. Organization of the bacterial chromosome. Microbiol. Rev. 54, 502–539.
502 Lechner, M., Findei\s s, S., Steiner, L., Marz, M., Stadler, P.F., Prohaska, S.J., 2011. Proteinortho:
503 detection of (co-) orthologs in large-scale analysis. BMC Bioinformatics 12, 124.
504 McLeod, M.P., Qin, X., Karpathy, S.E., Gioia, J., Highlander, S.K., Fox, G.E., McNeill, T.Z., Jiang, H.,
505 Muzny, D., Jacob, L.S., Hawes, A.C., Sodergren, E., Gill, R., Hume, J., Morgan, M., Fan, G.,
506 Amin, A.G., Gibbs, R.A., Hong, C., Yu, X. -j., Walker, D.H., Weinstock, G.M., 2004. Complete
507 Genome sequence of *Rickettsia typhi* and comparison with sequences of other rickettsiae. J.
508 Bacteriol. 186, 5842–5855. https://doi.org/10.1128/JB.186.17.5842-5855.2004
509 Merhej, V., Georgiades, K., Raoult, D., 2013. Postgenomic analysis of bacterial pathogens repertoire
510 reveals genome reduction rather than virulence factors. Brief. Funct. Genomics 12, 291–304.
511 https://doi.org/10.1093/bfgp/elt015

512 Merhej, V., Raoult, D., 2011. Rickettsial evolution in the light of comparative genomics. Biol. Rev. 86,
513 379–405. https://doi.org/10.1111/j.1469-185X.2010.00151.x
514 Merhej, V., Royer-Carenzi, M., Pontarotti, P., Raoult, D., 2009. Massive comparative genomic analysis
515 reveals convergent evolution of specialized bacteria. Biol. Direct 4, 13.
516 https://doi.org/10.1186/1745-6150-4-13
517 Moran, N.A., 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108, 583–
518 586.
519 Munson, M.A., Baumann, L., Baumann, P., 1993. Buchnera aphidicola (a prokaryotic endosymbiont of
520 aphids) contains a putative 16S rRNA operon unlinked to the 23S rRNA-encoding gene:
521 sequence determination, and promoter and terminator analysis. Gene 137, 171–178.
522 https://doi.org/10.1016/0378-1119(93)90003-L
523 Murray, G.G.R., Weinert, L.A., Rhule, E.L., Welch, J.J., 2016. The phylogeny of *Rickettsia* using
524 different evolutionary signatures: how tree-like is bacterial evolution? Syst. Biol. 65, 265–
525 279. https://doi.org/10.1093/sysbio/syv084
526 Ogata, H., 2005. *Rickettsia felis*, from culture to genome sequencing. Ann. N. Y. Acad. Sci. 1063, 26–
527 34. https://doi.org/10.1196/annals.1355.004
528 Ogata, H., 2001. Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science 293, 2093–
529 2098. https://doi.org/10.1126/science.1061471
530 Ogata, H., Audic, S., Abergel, C., Fournier, P.-E., Claverie, J.-M., 2002. Protein coding palindromes are
531 a unique but recurrent feature in Rickettsia. Genome Res. 12, 808–816.
532 Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P.-E., Raoult, D., M Claverie, J., 2000. Selfish
533 DNA in protein-coding genes of *Rickettsia*.
534 Ogata, H., La Scola, B., Audic, S., Renesto, P., Blanc, G., Robert, C., Fournier, P.-E., Claverie, J.-M.,
535 Raoult, D., 2006. Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in
536 gene exchanges between intracellular pathogens. PLoS Genet. 2, e76.
537 https://doi.org/10.1371/journal.pgen.0020076
538 Ogata, H., Renesto, P., Audic, S., Robert, C., Blanc, G., Fournier, P.-E., Parinello, H., Claverie, J.-M.,
539 Raoult, D., 2005a. The genome sequence of *Rickettsia felis* identifies the first putative
540 conjugative plasmid in an obligate intracellular parasite. PLoS Biol. 3, e248.
541 https://doi.org/10.1371/journal.pbio.0030248
542 Ogata, H., Suhre, K., Claverie, J.-M., 2005b. Discovery of protein-coding palindromic repeats in
543 *Wolbachia*. Trends Microbiol. 13, 253–5. https://doi.org/10.1016/j.tim.2005.03.013
544 Papenfort, K., Vanderpool, C.K., 2015. Target activation by regulatory RNAs in bacteria. FEMS
545 Microbiol. Rev. 39, 362–378. https://doi.org/10.1093/femsre/fuv016
546 Parola, P., Paddock, C.D., Socolovschi, C., Labruna, M.B., Mediannikov, O., Kernif, T., Abdad, M.Y.,
547 Stenos, J., Bitam, I., Fournier, P.-E., Raoult, D., 2013. Update on tick-borne rickettsioses
548 around the World: a geographic approach. Clin. Microbiol. Rev. 26, 657–702.
549 https://doi.org/10.1128/CMR.00032-13
550 Raoult, D., Roux, V., 1997. Rickettsioses as paradigms of new or emerging infectious diseases. Clin.
551 Microbiol. Rev. 10, 694–719.
552 Reed, S.C.O., Lamason, R.L., Risca, V.I., Abernathy, E., Welch, M.D., 2014. Rickettsia actin-based
553 motility occurs in distinct phases mediated by different actin nucleators. Curr. Biol. 24, 98–
554 103. https://doi.org/10.1016/j.cub.2013.11.025
555 Renesto, P., Ogata, H., Audic, S., Claverie, J.-M., Raoult, D., 2005. Some lessons from *Rickettsia*
556 genomics. FEMS Microbiol. Rev. 29, 99–117. https://doi.org/10.1016/j.femsre.2004.09.002
557 Riegler, M., Iturbe-Ormaetxe, I., Woolfit, M., Miller, W.J., O'Neill, S.L., 2012. Tandem repeat markers
558 as novel diagnostic tools for high resolution fingerprinting of *Wolbachia*. BMC Microbiol. 12,
559 S12.
560 Rogozin, I.B., Makarova, K.S., Natale, D.A., Spiridonov, A.N., Tatusov, R.L., Wolf, Y.I., Yin, J., Koonin,
561 E.V., 2002. Congruent evolution of different classes of non-coding DNA in prokaryotic
562 genomes. Nucleic Acids Res. 30, 4264–4271.

563 Rovery, C., Renesto, P., Crapoulet, N., Matsumoto, K., Parola, P., Ogata, H., Raoult, D., 2005.
564 Transcriptional response of *Rickettsia conorii* exposed to temperature variation and stress
565 starvation. Res. Microbiol. 156, 211–218. https://doi.org/10.1016/j.resmic.2004.09.002
566 Sahni, S.K., Narra, H.P., Sahni, A., Walker, D.H., 2013. Recent molecular insights into rickettsial
567 pathogenesis and immunity. Future Microbiol. 8, 1265–1288.
568 https://doi.org/10.2217/fmb.13.102
569 Sakharkar, K.R., 2004. Genome reduction in prokaryotic obligatory intracellular parasites of humans:
570 a comparative analysis. Int. J. Syst. Evol. Microbiol. 54, 1937–1941.
571 https://doi.org/10.1099/ijs.0.63090-0
572 Schroeder, C.L.C., Narra, H.P., Rojas, M., Sahni, A., Patel, J., Khanipov, K., Wood, T.G., Fofanov, Y.,
573 Sahni, S.K., 2015. Bacterial small RNAs in the genus *Rickettsia*. BMC Genomics 16.
574 https://doi.org/10.1186/s12864-015-2293-7
575 Sears, K.T., Ceraul, S.M., Gillespie, J.J., Allen, E.D., Popov, V.L., Ammerman, N.C., Rahman, M.S., Azad,
576 A.F., 2012. Surface proteome analysis and characterization of surface cell antigen (Sca) or
577 autotransporter family of *Rickettsia typhi*. PLoS Pathog. 8, e1002856.
578 https://doi.org/10.1371/journal.ppat.1002856
579 Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.
580 https://doi.org/10.1093/bioinformatics/btu153
581 Seshadri, R., Paulsen, I.T., Eisen, J.A., Read, T.D., Nelson, K.E., Nelson, W.C., Ward, N.L., Tettelin, H.,
582 Davidsen, T.M., Beanan, M.J., others, 2003. Complete genome sequence of the Q-fever
583 pathogen *Coxiella burnetii*. Proc. Natl. Acad. Sci. 100, 5455–5460.
584 Sicheritz-Pontén, T., Andersson, S.G., 1997. GRS: a graphic tool for genome retrieval and segment
585 analysis. Microb. Comp. Genomics 2, 123–139.
586 Shpynov, S.N., Fournier, P.E., Pozdnichenko, N.N., Gumenuk, A.S., Skiba, A.A., 2018. New approaches
587 in the systematics of rickettsiae. New Microbes New Infect. 23, 93-102.
588 https://doi.org10.1016/j.nmni.2018.02.012.
589 Stothard, D.R., Clark, J.B., Fuerst, P.A., 1994. Ancestral divergence of *Rickettsia bellii* from the spotted
590 fever and typhus groups of *Rickettsia* and antiquity of the genus *Rickettsia*. Int. J. Syst. Evol.
591 Microbiol. 44, 798–804.
592 Syvänen, A.-C., Amiri, H., Jamal, A., Andersson, S.G., Kurland, C.G., 1996. A chimeric disposition of the
593 elongation factor genes in *Rickettsia prowazekii*. J. Bacteriol. 178, 6192–6199.
594 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular evolutionary
595 genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
596 https://doi.org/10.1093/molbev/mst197
597 Walker, D.H., 2005. Progress in rickettsial genome analysis from pioneering of *Rickettsia prowazekii*
598 to the recent *Rickettsia typhi*. Ann. N. Y. Acad. Sci. 1063, 13–25.
599 https://doi.org/10.1196/annals.1355.003
600 Walsh, J.B., 1995. How often do duplicated genes evolve new functions? Genetics 139, 421–428.
601 Weinert, L.A., Welch, J.J., Jiggins, F.M., 2009. Conjugation genes are common throughout the genus
602 *Rickettsia* and are transmitted horizontally. Proc. R. Soc. B Biol. Sci. 276, 3619–3627.
603 https://doi.org/10.1098/rspb.2009.0875
604 Weinert, L.A., Werren, J.H., Aebi, A., Stone, G.N., Jiggins, F.M., 2009. Evolution and diversity of
605 *Rickettsia* bacteria. BMC Biol. 7, 6. https://doi.org/10.1186/1741-7007-7-6
606 Wixon, J., 2001. Featured organism: reductive evolution in bacteria: *Buchnera* sp., *Rickettsia*
607 *prowazekii* and *Mycobacterium leprae*. Comp. Funct. Genomics 2, 44–48.
608 Wolf, Y.I., Koonin, E.V., 2013. Genome reduction as the dominant mode of evolution: prospects &
609 overviews. BioEssays 35, 829–837. https://doi.org/10.1002/bies.201300037
610 Woodard, A., Wood, D.O., 2011. Analysis of convergent gene transcripts in the obligate intracellular
611 bacterium *Rickettsia prowazekii*. PLoS ONE 6, e16537.
612 https://doi.org/10.1371/journal.pone.0016537
613 Wu, M., Sun, L.V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J.C., McGraw, E.A., Martin, W.,
614 Esser, C., Ahmadinejad, N., Wiegand, C., Madupu, R., Beanan, M.J., Brinkac, L.M., Daugherty,

S.C., Durkin, A.S., Kolonay, J.F., Nelson, W.C., Mohamoud, Y., Lee, P., Berry, K., Young, M.B., Utterback, T., Weidman, J., Nierman, W.C., Paulsen, I.T., Nelson, K.E., Tettelin, H., O'Neill, S.L., Eisen, J.A., 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: A streamlined genome overrun by mobile genetic elements. PLoS Biol. 2, e69. https://doi.org/10.1371/journal.pbio.0020069

Xin, D., El Karkouri, K., Robert, C., Raoult, D., Fournier, P.-E., 2012. Genomic comparison of *Rickettsia honei s*train RBT and other *Rickettsia* Species. J. Bacteriol. 194, 4145. https://doi.org/10.1128/JB.00802-12

Figure 1: Phylogenomic tree of 29 *Rickettsia* species based on whole-genome sequence analysis using the Maximum Likehood method within the FastTree software. Genomes were aligned using Mugsy software. Values at the nodes are percentages. Numbers at the nodes represent the percentages of bootstrap values obtained by repeating the analysis 1000 times to generate a majority consensus tree. Only values greater than 90 % were reported. AG = Ancestral group; TG = Typhus group; TRG = Transitional group; SFG = Spotted fever group.

Figure 2: Phylogenomic tree based on 591 core proteins and pathogenic and genomic features, of *Rickettsia* species exhibiting various degrees of pathogenesis. For each genome (downloaded from GenBank), gene prediction was obtained using the Prokka software (Seemann, 2014). The core genome was identified using the ProteinOrtho software (Lechner et al., 2011). Then, the amino acid sequences of 591 proteins (Supplementary Table) conserved in all studied genomes were concatenated for each species and multiple alignment was performed using the Mafft software (Katoh and Standley, 2013, p. 2). Gapped positions were removed. The phylogenetic inferences were obtained using the Maximum Likelihood method and the MEGA software version 6 (Tamura et al., 2013). Branching support was evaluated using the bootstrap method with 1000 replications. Bootstrap values greater than 90% are shown at the nodes. Properties of each species were extracted from the following references (Andersson et al., 1998; G. Blanc et al., 2007; Guillaume Blanc et al., 2007b; El Karkouri et al., 2017, 2016; Fournier et al., 2009; McLeod et al., 2004; Ogata, 2001b; Ogata et al., 2006, 2005a). NA = data not available; RPEs = Rickettsia palindromic elements.

Supplementary Table: List of protein sequences used for inferring the phylogenomic relationshiops of ten *Rickettsia* species exhibiting various degrees of pathogenesis and presented in Figure 2.

Clusters ID	Protein function	Genes
cRIG00001	rnpA Ribonuclease P	rnpA
cRIG00002	rplT 50S ribosomal protein L	rpl
cRIG00003	nrdG Organic radical activating enzymes	nrd
cRIG00004	rplY 50S ribosomal protein L	rpl
cRIG00005	ychF GTP-binding protein YchF	ychF
cRIG00006	murE UDP-N-acetylmuramoylalanyl-D-glutamate-- , 2,6-diaminopimelate ligase	mur
cRIG00007	murF UDP-N-acetylmuramoylalanyl-D-glutamyl- , 2,6-diaminopimelate--D-alanyl-D-alanyl ligase	mur
cRIG00008	mraY1 Phospho-N-acetylmuramoyl-pentapeptide- transferase	mraY
cRIG00009	recG ATP-dependent DNA helicase RecG	rec
cRIG00010	traX F pilin acetylation protein TraX	tra
cRIG00011	mviN Integral membrane protein MviN	mviN
cRIG00012	ppa Inorganic pyrophosphatase	ppa
cRIG00013	ccmE Cytochrome c-type biogenesis protein ccmE	ccm
cRIG00014	Sco2 protein precursor	Sco2
cRIG00015	secD Protein-export membrane protein secD	sec
cRIG00016	yajC Preprotein translocase YajC subunit	yajC
cRIG00017	gyrB DNA gyrase subunit B	gyr
cRIG00018	murA UDP-N-acetylglucosamine -1-carboxyvinyltransferase	mur
cRIG00019	surA Parvulin-like peptidyl-prolyl isomerase	surA
cRIG00020	secA Preprotein translocase secA subunit	sec
cRIG00021	Unknown	-
cRIG00022	uvrC Excinuclease ABC subunit C	uvr
cRIG00023	cmk Cytidylate kinase	cmk
cRIG00024	rpsA 30S ribosomal protein S1	rps
cRIG00025	clpP ATP-dependent Clp protease proteolytic subunit	clp
cRIG00026	glmU UDP-N-acetylglucosamine pyrophosphorylase	glmU
cRIG00027	Unknown	-
cRIG00028	Unknown	-
cRIG00029	lpdA2 Dihydrolipoamide dehydrogenase	lpdA
cRIG00030	5 -Formyltetrahydrofolate cyclo-ligase	-
cRIG00031	rnd Ribonuclease D	rnd
cRIG00032	gcvT Glycine cleavage T-protein	gcvT
cRIG00033	putP Na+/proline symporter and signal transduction histidine kinase	putP
cRIG00034	hemC Porphobilinogen deaminase	hemC
cRIG00035	trpS Tryptophanyl-tRNA synthetase	trpS
cRIG00036	plsC 1-acyl-sn-glycerol--phosphate acyltransferase	plsC
cRIG00037	ampG1 AmpG	ampG
cRIG00038	tlc3 ATP/ADP translocase	tlc
cRIG00039	Unknown	-
cRIG00040	ispB Octaprenyl-diphosphate synthase	isp
cRIG00041	potE Putrescine-ornithine antiporter	potE
cRIG00042	iscA2 Iron-sulfur cluster assembly accessory protein	isc

cRIG00043	iscU FeS cluster assembly scaffold IscU	isc
cRIG00044	iscS Cysteine desulfurase IscS	isc
cRIG00045	spl1 NifS-like protein	spl1
cRIG00046	Unknown	-
cRIG00047	ntrY Nitrogen regulation protein NtrY	ntr
cRIG00048	rpsU 30S ribosomal protein S	rps
cRIG00049	Unknown	-
cRIG00050	ileS Isoleucyl-tRNA synthetase	ileS
cRIG00051	accC Acetyl-CoA carboxylase, biotin carboxylase	accC
cRIG00052	pccB Propionyl-CoA carboxylase beta chain precursor	pccB
cRIG00053	aas2 -acylglycerophosphoethanolamine acyltransferase	aas
cRIG00054	znuB Zinc/manganese ABC transporter permease protein	znuB
cRIG00055	ubiG Ubiquinone biosynthesis O-methyltransferase	ubi
cRIG00056	gltX Glutamyl-tRNA synthetase	gltX
cRIG00057	groEL 60 kD chaperonin	groEL
cRIG00058	groES 10 kD chaperonin	groES
cRIG00059	rph Ribonuclease PH	rph
cRIG00060	grpE GrpE protein	grpE
cRIG00061	perM Permease PerM-like protein	perM
cRIG00062	DnaA-like protein	DnaA
cRIG00063	rplQ 50S ribosomal protein L17	rpl
cRIG00064	rpoA DNA-directed RNA polymerase alpha chain	rpo
cRIG00065	rpsK 30S ribosomal protein S11	rps
cRIG00066	rpsM 30S ribosomal protein S13	rps
cRIG00067	adk Adenylate kinase	adk
cRIG00068	secY Preprotein translocase secY subunit	sec
cRIG00069	rplO 50S ribosomal protein L15	rpl
cRIG00070	rpmD 50S ribosomal protein L30	rpm
cRIG00071	rpsE 30S ribosomal protein S5	rps
cRIG00072	rplR 50S riboSomal protein L18	rpl
cRIG00073	rplF 50S riboSomal protein L6	rpl
cRIG00074	rp30SH 30S riboSomal protein S8	rps
cRIG00075	rp30SN 30S riboSomal protein S14	rps
cRIG00076	rplE 50S riboSomal protein L5	rpl
cRIG00077	rplX 50S riboSomal protein L24	rpl
cRIG00078	rplN 50S ribosomal protein L14	rpl
cRIG00079	rpsQ 30S ribosomal protein S17	rps
cRIG00080	rpmC 50S ribosomal protein L29	rpm
cRIG00081	rplP 50S ribosomal protein L16	rpl
cRIG00082	rpsC 30S ribosomal protein S3	rps
cRIG00083	rplV 50S ribosomal protein L22	rpl
cRIG00084	rpsS 30S ribosomal protein S19	rps
cRIG00085	rplB 50S ribosomal protein L2	rpl
cRIG00086	rplW 50S ribosomal protein L23	rpl
cRIG00087	rplD 50S ribosomal protein L4	rpl
cRIG00088	rpsJ 30S ribosomal protein S10	rps
cRIG00089	tuf Elongation factor EF-Tu	tuf
cRIG00090	fumC Fumarate hydratase	fumC

cRIG00091	ftsZ Cell division protein ftsZ	fts
cRIG00092	NifU-like protein	NifU
cRIG00093	ampG2 AmpG	ampG
cRIG00094	rhlE ATP-dependent RNA helicase RhlE	rhlE
cRIG00095	cspA Cold shock-like protein	cspA
cRIG00096	ksgA Dimethyladenosine transferase	ksgA
cRIG00097	Unknown	-
cRIG00098	ostA Organic solvent tolerance protein-like protein	ostA
cRIG00099	xseA Exodeoxyribonuclease VII, large subunit	xse
cRIG00100	xth2 Exodeoxyribonuclease III	xth
cRIG00101	GTP-binding protein	-
cRIG00102	ubiB 2-polyprenylphenol -hydroxylase	ubi
cRIG00103	ubiE Ubiquinone/menaquinone biosynthesis methlytransferase UbiE	ubi
cRIG00104	Unknown	-
cRIG00105	tatD Putative deoxyribonuclease TatD	tat
cRIG00106	metG Methionyl-tRNA synthetase	metG
cRIG00107	tmk1Thymidylate kinase	tmk1
cRIG00108	proP4 Proline/betaine transporter	proP4
cRIG00109	ubiA 4-hydroxybenzoate octaprenyltransferase	ubi
cRIG00110	valS Valyl-tRNA synthetase	valS
cRIG00111	RmuC family protein	RmuC
cRIG00112	exsB Trans-regulatory protein ExsB	exsB
cRIG00113	msbA2 Multidrug resistance protein	msbA
cRIG00114	Unknown	-
cRIG00115	bcr2 MFS-type bicyclomycin resistance protein	bcr2
cRIG00116	Lipoprotein releasing system, transmembrane protein, LolC/E family protein	Lol
cRIG00117	lolD Lipoprotein releasing system ATP-binding protein LolD	Lol
cRIG00118	Unknown	-
cRIG00119	Hemolysin-like protein	-
cRIG00120	ccmF Cytochrome c-type biogenesis protein ccmF	ccm
cRIG00121	ompB, sca5 Outer membrane protein rOmpB	sca
cRIG00122	Beta-glucosidase	-
cRIG00123	lpxK Tetraacyldisaccharide 4'-kinase	lpx
cRIG00124	ligA DNA ligase, NAD-dependent	lig
cRIG00125	tgt Queuine tRNA-ribosyltransferase	tgt
cRIG00126	ABC transporter substrate binding protein	-
cRIG00127	NADHubiquinone oxidoreductase 17,2 kD subunit	-
cRIG00128	rnhA Ribonuclease H	rnh
cRIG00129	Unknown	-
cRIG00130	Unknown	-
cRIG00131	coaE Dephospho-CoA kinase	coaE
cRIG00132	dnaQ DNA polymerase III epsilon chain	dna
cRIG00133	surf1 Surfeit locus protein	surf1
cRIG00134	ATP-dependent helicase	-
cRIG00135	fabD Malonyl CoA-acyl carrier protein transacylase	fab
cRIG00136	tlc5 ATP/ADP translocase	tlc
cRIG00137	tlyC Hemolysin C	tlyC

cRIG00138	Putative metal-dependent hydrolase	-
cRIG00139	lipA Lipoic acid synthetase	lip
cRIG00140	glyA Glycine/serine hydroxymethyltransferase	gly
cRIG00141	Serine esterase	-
cRIG00142	nth Endonuclease III	nth
cRIG00143	Putative methyltransferase	-
cRIG00144	ABC-type transport systems periplasmic component	-
cRIG00145	tatA Twin-arginine translocation protein TatA	tat
cRIG00146	pgpA Phosphatidylglycerophosphatase A	pgpA
cRIG00147	rplU 50S ribosomal protein L21	rpl
cRIG00148	rpmA 50S ribosomal protein L27	rpm
cRIG00149	Unknown	-
cRIG00150	proP5 Proline/betaine transporter	proP
cRIG00151	NAD-specific glutamate dehydrogenase	NAD
cRIG00152	trmE tRNA modification GTPase TrmE	trm
cRIG00153	recA RecA	rec
cRIG00154	fabG 3-oxoacyl reductase	fab
cRIG00155	acpP Acyl carrier protein	acpP
cRIG00156	fabF 3-oxoacyl-	fab
cRIG00157	mreC Rod shape-determining protein MreC	mre
cRIG00158	mreB Rod shape-determining protein MreB	mre
cRIG00159	Putative permeases	-
cRIG00160	pal Peptidoglycan-associated lipoprotein precursor	pal
cRIG00161	rpmF 50S ribosomal protein L32	rpm
cRIG00162	smpA tmRNA-binding protein	smpA
cRIG00163	ftsY Signal recognition particle-docking protein FtsY	fts
cRIG00164	polA DNA polymerase I	pol
cRIG00165	dnaE DNA polymerase III alpha chain	dna
cRIG00166	udg UDP-glucose 6-dehydrogenase	udg
cRIG00167	Unknown	-
cRIG00168	ampG3 AmpG	ampG
cRIG00169	tatC Sec-independent protein translocase protein TatC	tat
cRIG00170	serS Seryl-tRNA synthetase	serS
cRIG00171	virB4-2 VirB4	virB
cRIG00172	Unknown	-
cRIG00173	terC Tellurium resistance protein TerC	terC
cRIG00174	nuoL1 NADH dehydrogenase I chain L	nuo
cRIG00175	nuoM NADH dehydrogenase I chain M	nuo
cRIG00176	ccmA Heme exporter protein A	ccm
cRIG00177	nuoI NADH dehydrogenase I chain I	nuo
cRIG00178	nuoH NADH dehydrogenase I chain H	nuo
cRIG00179	nuoG NADH dehydrogenase I chain G	nuo
cRIG00180	atpG ATP synthase gamma chain	atp
cRIG00181	atpA ATP synthase alpha chain	atp
cRIG00182	atpH ATP synthase delta chain	atp
cRIG00183	lpdA1 Dihydrolipoamide dehydrogenase	lpdA
cRIG00184	Unknown	-
cRIG00185	kefB Glutathione-regulated potassium-efflux system protein KefB	kefB

cRIG00186	Iojap-related protein	lojap
cRIG00187	bolA2 BolA-like protein	bolA
cRIG00188	infA Translation initiation factor IF-	inf
cRIG00189	maf Nucleotide-binding protein implicated in inhibition of septum formation	maf
cRIG00190	dksA DnaK suppressor-like protein	dksA
cRIG00191	xerC Tyrosine recombinase XerC	xerC
cRIG00192	hypothetical protein	-
cRIG00193	ftsK Cell division protein FtsK	fts
cRIG00194	mraY2 Undecaprenyl-phosphate alpha-N-acetylglucosaminyltransferase	mraY
cRIG00195	Unknown	-
cRIG00196	Unknown	-
cRIG00197	Putative outer surface protein	sca
cRIG00198	fdxA Ferredoxin	fdxA
cRIG00199	ccmC Heme exporter protein C	ccm
cRIG00200	Cation diffusion facilitator family transporter	-
cRIG00201	omp 17 kD surface antigen precursor	Sca
cRIG00202	znuC Zinc ABC transporter ATP-binding protein	znu
cRIG00203	uvrA Excinuclease ABC subunit A	uvr
cRIG00204	ssb Single-stranded DNA-binding protein	ssb
cRIG00205	DAP dipeptidyl aminopeptidase/acylaminoacyl-peptidase-like protein	dap
cRIG00206	htpG Heat shock protein htpG	htpG
cRIG00207	hemA 5-aminolevulinic acid synthase	hemA
cRIG00208	tig Trigger factor	tig
cRIG00209	obg GTP-binding protein	obg
cRIG00210	gltA Citrate synthase I	gltA
cRIG00211	Uracil-DNA glycosylase, family	Uracil
cRIG00212	Ribosomal large subunit pseudouridine synthases RluD subfamily protein	RluD
cRIG00213	hemK Methylase of polypeptide chain release factors	hemK
cRIG00214	Sua5/YciO/YrdC/YwlC family putative translation factor protein	Sua5
cRIG00215	glyS Glycyl-tRNA synthetase beta chain	gly
cRIG00216	glyQ Glycyl-tRNA synthetase alpha chain	gly
cRIG00217	Unknown	-
cRIG00218	Unknown	-
cRIG00219	dnaG DNA primase	dna
cRIG00220	sec59 Dolichol kinase	sec
cRIG00221	greA Transcription elongation factor GreA	greA
cRIG00222	pntA2 NAD(P) transhydrogenase subunit alpha	pnt
cRIG00223	pntA1 NAD(P) transhydrogenase subunit alpha	pnt
cRIG00224	lolA Outer membrane lipoprotein-sorting protein LolA	Lol
cRIG00225	Unknown	-
cRIG00226	Putative aspartyl protease	-
cRIG00227	glnQ Glutamine ABC transporter ATP-binding protein	gln
cRIG00228	phnP Metal-dependent hydrolases of the beta-lactamase superfamily I	phnP
cRIG00229	Unknown	-
cRIG00230	Putative permeases	-
cRIG00231	holC DNA polymerase III chi subunit HolC	holC
cRIG00232	dapE Succinyl-diaminopimelate desuccinylase	dap
cRIG00233	Unknown	-

cRIG00234	lipB Putative lipoate-protein ligase B	lip
cRIG00235	rpsP 30S ribosomal protein S16	rps
cRIG00236	mutL DNA mismatch repair protein MutL	mut
cRIG00237	proP7 Proline/betaine transporter	proP
cRIG00238	hemF Coproporphyrinogen III oxidase precursor	hemF
cRIG00239	Putative membrane protein	-
cRIG00240	hemH Putative ferrochelatase	hemH
cRIG00241	hemE Uroporphyrinogen decarboxylase	hemE
cRIG00242	Unknown	-
cRIG00243	trxA Thioredoxin	trx
cRIG00244	rfbE O-antigen export system ATP-binding protein RfbE	rfb
cRIG00245	rfbA O-antigen export system permease protein RfbA	rfb
cRIG00246	gltD NADPH-dependent glutamate synthase beta chain and related oxidoreductases	gltA
cRIG00247	lpxA Acyl-	lpx
cRIG00248	fabZ (3R)-hydroxymyristoyl-	fab
cRIG00249	lpxD UDP-3-O-	lpx
cRIG00250	Putative P-loop hydrolase	-
cRIG00251	znuA Zinc/manganese ABC transporter substrate binding protein	znu
cRIG00252	pcnB Poly(A) polymerase	pcnB
cRIG00253	atpF ATP synthase B chain	atp
cRIG00254	atpX ATP synthase B chain	atp
cRIG00255	atpE ATP synthase C chain	atp
cRIG00256	atpB ATP synthase A chain	atp
cRIG00257	Unknown	-
cRIG00258	dsbG Protein-disulfide isomerase	dsbG
cRIG00259	Transcriptional regulator	-
cRIG00260	Unknown	-
cRIG00261	cvpA Putative colicin V production membrane protein	cvpA
cRIG00262	clpB ClpB	clp
cRIG00263	hypothetical protein	-
cRIG00264	rpsF 30S ribosomal protein S6	rps
cRIG00265	rpsR30 S ribosomal protein S18	rps
cRIG00266	rplI 50S ribosomal protein L9	rpl
cRIG00267	tilS, mesJ tRNA(Ile)-lysidine synthetase	tilS
cRIG00268	ftsH ATP-dependent metalloprotease FtsH	fts
cRIG00269	sdhB Succinate dehydrogenase iron-sulfur protein	sdh
cRIG00270	Unknown	-
cRIG00271	lgt Prolipoprotein diacylglyceryl transferase	lgt
cRIG00272	Putative membrane protein	-
cRIG00273	yidC Preprotein translocase subunit YidC	yidC
cRIG00274	pgsA CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	pgsA
cRIG00275	Unknown	-
cRIG00276	tlc1 ATP/ADP translocase	tlc
cRIG00277	uhpC Sugar phosphate permease	uhpC
cRIG00278	ndk Nucleoside diphosphate kinase	ndk
cRIG00279	gidA Glucose-inhibited division protein A	gid
cRIG00280	soj ATPase involved in chromosome partitioning	soj
cRIG00281	ParB-like partition proteins	ParB

cRIG00282	abcTABC transporter ATP-binding protein	abcT
cRIG00283	Unknown	-
cRIG00284	kdsA -deoxy--phosphooctulonate synthase	kds
cRIG00285	iscAIron-sulfur cluster assembly accessory protein	isc
cRIG00286	dgt Deoxyguanosinetriphosphate triphosphohydrolase	dgt
cRIG00287	argS Arginyl-tRNA synthetase	arg
cRIG00288	Unknown	-
cRIG00289	parC Topoisomerase IV subunit A	Par
cRIG00290	Unknown	-
cRIG00291	dcd Deoxycytidine triphosphate deaminase	dcd
cRIG00292	secB Protein-export protein secB	sec
cRIG00293	czcR Transcriptional activator protein CzcR	czcR
cRIG00294	GTP cyclohydrolase I	GTP
cRIG00295	Unknown	-
cRIG00296	pntB NAD(p) transhydrogenase subunit beta	pnt
cRIG00297	ompW OmpW family outer-membrane protein	Sca
cRIG00298	sam S-adenosylmethionine transporter	sam
cRIG00299	proP1 Proline/betaine transporter	proP
cRIG00300	cysS Cysteinyl-tRNA synthetase	cysS
cRIG00301	rpsB 30S ribosomal protein S2	rps
cRIG00302	tsf Elongation factor EF-Ts	tsf
cRIG00303	kdtA 3-deoxy-D-manno-octulosonic-acid transferase	kdtA
cRIG00304	Unknown	-
cRIG00305	aatA Aspartate aminotransferase A	aatA
cRIG00306	Unknown	-
cRIG00307	vacJ VacJ lipoprotein precursor	vacJ
cRIG00308	ABC-type transporter related to toluene tolerance	-
cRIG00309	alr Alanine racemase	alr
cRIG00310	ABC transporter permease protein	-
cRIG00311	mkl Ribonucleotide ABC transporter ATP-binding protein	mkl
cRIG00312	rpmB 50S ribosomal protein L28	rpm
cRIG00313	rpmE 50S ribosomal protein L31	rpm
cRIG00314	Hypothetical GTP-binding protein	-
cRIG00315	virB3 VirB3	virB
cRIG00316	virB4-1 VirB4	virB
cRIG00317	virB6-3 VirB6	virB
cRIG00318	virB6-4 VirB6	virB
cRIG00319	virB6-5 VirB6	virB
cRIG00320	trmD tRNA (guanine-n1)-methyltransferase	trm
cRIG00321	rplS 50S ribosomal protein L19	rpl
cRIG00322	Unknown	-
cRIG00323	secF Protein-export membrane protein secF	sec
cRIG00324	nuoF NADH dehydrogenase I chain F	nuo
cRIG00325	lepB Signal peptidase I	lepB
cRIG00326	era GTP-binding protein Era	era
cRIG00327	ruvC Crossover junction endodeoxyribonuclease RuvC	ruvC
cRIG00328	Putative nucleoside-diphosphate-sugar epimerase	-
cRIG00329	mrp Mrp	mrp

cRIG00330	hflK Protease activity modulator HflK	hfl
cRIG00331	hflC2 Membrane protease subunit, stomatin/prohibitin-like protein	hfl
cRIG00332	htrA Periplasmic serine protease	htrA
cRIG00333	Putative sulfurtransferase	-
cRIG00334	sdhC Succinate dehydrogenase cytochrome b-556 subunit	sdh
cRIG00335	yqiY Amino acid ABC transporter permease protein	yqi
cRIG00336	rpsL 30S ribosomal protein S12	rps
cRIG00337	rpsG 30S ribosomal protein S7	rps
cRIG00338	fusA Elongation factor EF-G	fusA
cRIG00339	nusG Transcription antitermination protein NusG	nus
cRIG00340	rplK 50S ribosomal protein L11	rpl
cRIG00341	rplA 50S ribosomal protein L1	rpl
cRIG00342	rplJ 50S ribosomal protein L10	rpl
cRIG00343	rplL 50S ribosomal protein L7/L12	rpl
cRIG00344	rpoB DNA-directed RNA polymerase beta chain	rpo
cRIG00345	rpoC DNA-directed RNA polymerase beta prime chain	rpo
cRIG00346	pepA Aminopeptidase A	pepA
cRIG00347	Chromosome partitioning protein-like protein	-
cRIG00348	aspS Aspartyl-tRNA synthetase	aspS
cRIG00349	Integral membrane protein, interacts with FtsH	fts
cRIG00350	yqiX Amino acid ABC transporter substrate binding protein	yqi
cRIG00351	gatA Glutamyl-tRNA(Gln) amidotransferase subunit A	gat
cRIG00352	gatC Glutamyl-tRNA(Gln) amidotransferase subunit C	gat
cRIG00353	rrf Ribosome recycling factor	rrf
cRIG00354	pyrH Uridylate kinase	pyr
cRIG00355	mnhE Multisubunit Na+/H+ antiporter, MnhE subunit	mnh
cRIG00356	emrB MFS-type multidrug resistance protein B (SPLIT GENE)	emrB
cRIG00357	omp1 Outer membrane protein omp	Sca
cRIG00358	Putative membrane-associated zinc metalloprotease	-
cRIG00359	nusB N utilization substance protein B	nus
cRIG00360	rrmJ Ribosomal RNA large subunit methyltransferase J	rrmJ
cRIG00361	Oligoketide cyclase/lipid transport protein	-
cRIG00362	Unknown	-
cRIG00363	hupA DNA-binding protein HU	hupA
cRIG00364	holB DNA polymerase III delta subunit	hol
cRIG00365	ffh Signal recognition particle protein	ffh
cRIG00366	gltP Na+/H+-dicarboxylate symporters	glt
cRIG00367	Putative 6-pyruvoyl tetrahydropterin synthase	-
cRIG00368	sucB Dihydrolipoamide acetyltransferase component	suc
cRIG00369	sucA2 -oxoglutarate dehydrogenase Ecomponent	suc
cRIG00370	recN DNA repair protein RecN	rec
cRIG00371	comL DNA uptake lipoprotein	comL
cRIG00372	dnaJ DnaJ	dna
cRIG00373	dnaK DnaK	dna
cRIG00374	Heat shock protease	hsl
cRIG00375	Unknown	-
cRIG00376	holA DNA polymerase III, delta subunit	hol
cRIG00377	coq7 Ubiquinone biosynthesis protein coq	coq7

cRIG00378	coxC Cytochrome c oxidase subunit III	cox
cRIG00379	virB2 VirB2-like protein	virB
cRIG00380	RafdapD 2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase	dap
cRIG00381	uspA Universal stress protein UspA and related nucleotide-binding proteins	uspA
cRIG00382	imp TRAP-type uncharacterized transport system, periplasmic component	imp
cRIG00383	Unknown	-
cRIG00384	hscA Heat shock protein hscA	hsc
cRIG00385	rnhB Ribonuclease HII	rnh
cRIG00386	uvrB Excinuclease ABC subunit B	uvr
cRIG00387	grxC1 Glutaredoxin, GrxC family	grxC1
cRIG00388	atm1 Multidrug resistance protein Atm	atm1
cRIG00389	gyrA DNA gyrase subunit A	gyr
cRIG00390	def1 Polypeptide deformylase	def
cRIG00391	fmt Methionyl-tRNA formyltransferase	fmt
cRIG00392	abcT3 Multidrug resistance ABC transporter ATP-binding protein	abcT
cRIG00393	Unknown	-
cRIG00394	cydA Cytochrome d ubiquinol oxidase subunit I	cyd
cRIG00395	thrS Threonyl-tRNA synthetase	thrS
cRIG00396	Unknown	-
cRIG00397	tolC Type I secretion outer membrane protein TolC	tol
cRIG00398	Unknown	-
cRIG00399	Ankyrin repeat	-
cRIG00400	parE DNA topoisomerase IV, B subunit	Par
cRIG00401	ctp Carboxyl-terminal protease	ctp
cRIG00402	barA Histidine kinase sensor protein	barA
cRIG00403	Unknown	-
cRIG00404	Unknown	-
cRIG00405	WD40-like repeat	-
cRIG00406	rplM 50S ribosomal protein L13	rpl
cRIG00407	rpsI 30S ribosomal protein S9	rps
cRIG00408	nudH (Di)nucleoside polyphosphate hydrolase	nudH
cRIG00409	Regulatory components of sensory transduction system	-
cRIG00410	efp Translation elongation factor EF-P	efp
cRIG00411	suhB Extragenic suppressor protein suhB	suhB
cRIG00412	Unknown	-
cRIG00413	psd Phosphatidylserine decarboxylase	psd
cRIG00414	pssA CDP-diacylglycerol--serine O-phosphatidyltransferase	pssA
cRIG00415	hypothetical protein	-
cRIG00416	Unknown	-
cRIG00417	bolA1 BolA-like protein	bolA
cRIG00418	EAL domain containing protein	EAL
cRIG00419	murC UDP-N-acetylmuramate--alanine ligase	mur
cRIG00420	murB UDP-N-acetylenolpyruvoylglucosamine reductase	mur
cRIG00421	ddlB D-alanine--D-alanine ligase	ddlB
cRIG00422	Membrane protein implicated in regulation of membrane protease activity	-
cRIG00423	cycM Cytochrome c	cycM
cRIG00424	lpxC UDP-3-O-	lpx
cRIG00425	xth1 Exodeoxyribonuclease III	xth

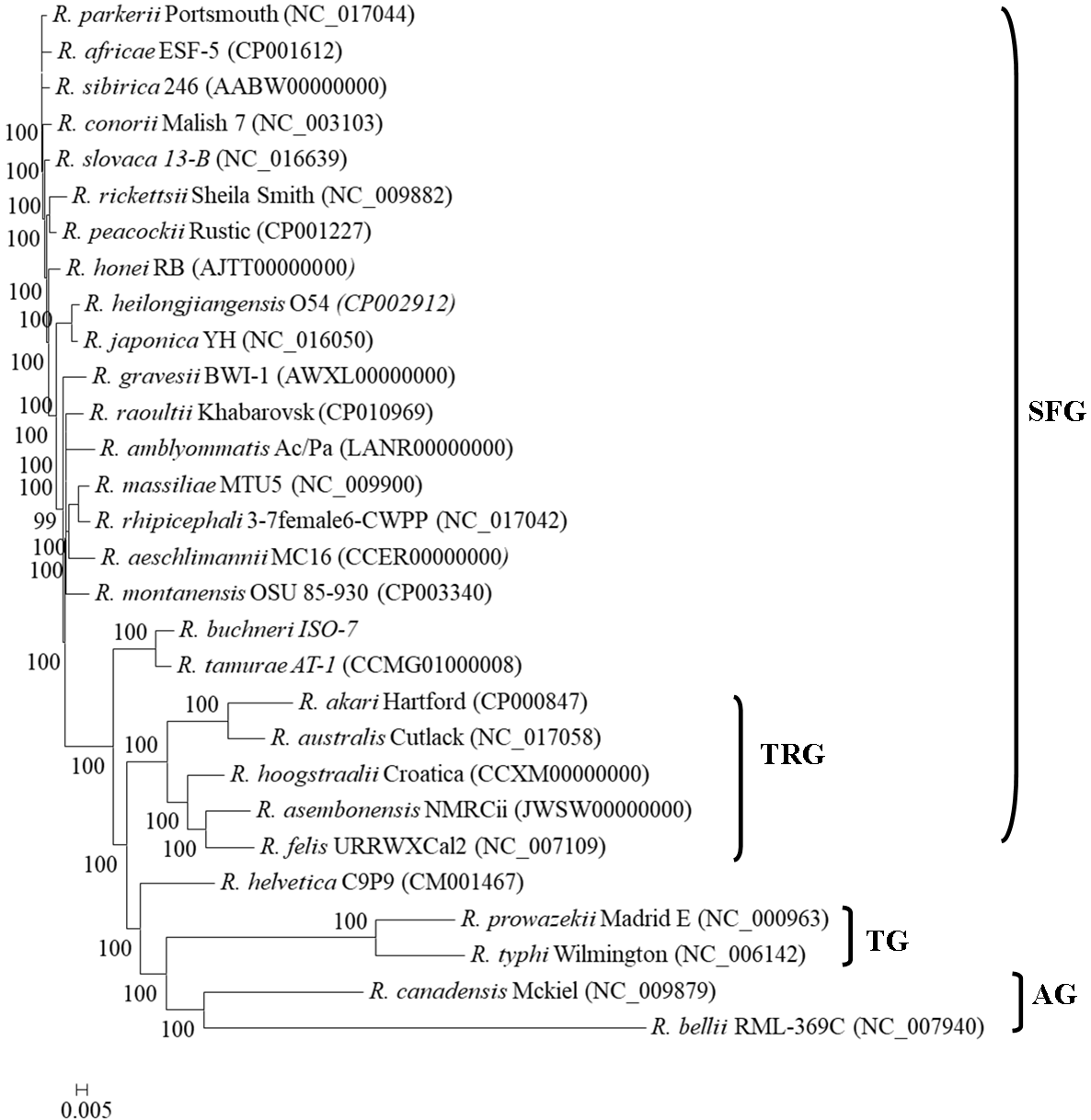
cRIG00426	pdhA Pyruvate dehydrogenase e1 component, alpha subunit precursor	pdh
cRIG00427	pdhB Pyruvate dehydrogenase E1 component, beta subunit precursor	pdh
cRIG00428	typA GTP-binding protein TypA	typ
cRIG00429	hlpA Outer membrane protein	hlp
cRIG00430	icd Isocitrate dehydrogenase, NADP-dependent	icd
cRIG00431	Monovalent cation/proton antiporter, MnhG/PhaG subunit	MnhG
cRIG00432	mnhB Multisubunit Na+/H+ antiporter, MnhB subunit	mnh
cRIG00433	ccmB Heme exporter protein B	ccm
cRIG00434	Unknown	-
cRIG00435	petA Ubiquinol-cytochrome c reductase, iron-sulfur subunit	pet
cRIG00436	petB Cytochrome b	pet
cRIG00437	fbcH Cytochrome c1, heme protein precursor	fbcH
cRIG00438	nuoL2 NADH dehydrogenase I chain L	nuo
cRIG00439	nuoN2 NADHubiquinone oxidoreductase subunit (chain N)	nuo
cRIG00440	mnhC Multisubunit Na+/H+ antiporter, MnhC subunit	mnh
cRIG00441	virB8-1VirB8	virB
cRIG00442	virB8-2 VirB8	virB
cRIG00443	virB9-2 VirB9	virB
cRIG00444	virD4 VirD4	virD
cRIG00445	gppA Guanosine pentaphosphate phosphohydrolase	gppA
cRIG00446	Unknown	-
cRIG00447	Unknown	-
cRIG00448	cysQ 3'(2'),'5-bisphosphate nucleotidase	cysQ
cRIG00449	mutS DNA mismatch repair protein MutS	mut
cRIG00450	lacA Ribose-5-phosphate isomerase	lacA
cRIG00451	nlpD1 Membrane-bound metallopeptidase	nlpD
cRIG00452	thyX Thymidylate synthase, flavin-dependent	thy
cRIG00453	tolB TolB protein precursor	tol
cRIG00454	hypothetical protein	-
cRIG00455	cox11, ctaG Cytochrome c oxidase assembly protein cox 11	cox
cRIG00456	trmU tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	trm
cRIG00457	atrC1 Cationic amino acid transporter-1	atrC
cRIG00458	hisS Histidyl-tRNA synthetase	hisS
cRIG00459	tolQ TolQ	tol
cRIG00460	tolR TolR	tol
cRIG00461	Periplasmic protein TonB, links inner and outer membranes	Ton
cRIG00462	proP10 Proline/betaine transporter	proP
cRIG00463	HlyD family secretion protein	HlyD
cRIG00464	aprD Alkaline protease secretion ATP-binding protein AprD	aprD
cRIG00465	asd Aspartate-semialdehyde dehydrogenase	asd
cRIG00466	Unknown	-
cRIG00467	hslV Heat shock protein HslV	hsl
cRIG00468	hslU Heat shock protein HslVU, ATPase subunit HslU	hsl
cRIG00469	lpxB Lipid-A-disaccharide synthase	lpx
cRIG00470	Aminodeoxychorismate lyase	-
cRIG00471	cyaY CyaY	cyaY
cRIG00472	gltX1 Glutamyl-tRNA synthetase	glt
cRIG00473	topA DNA topoisomerase I	top

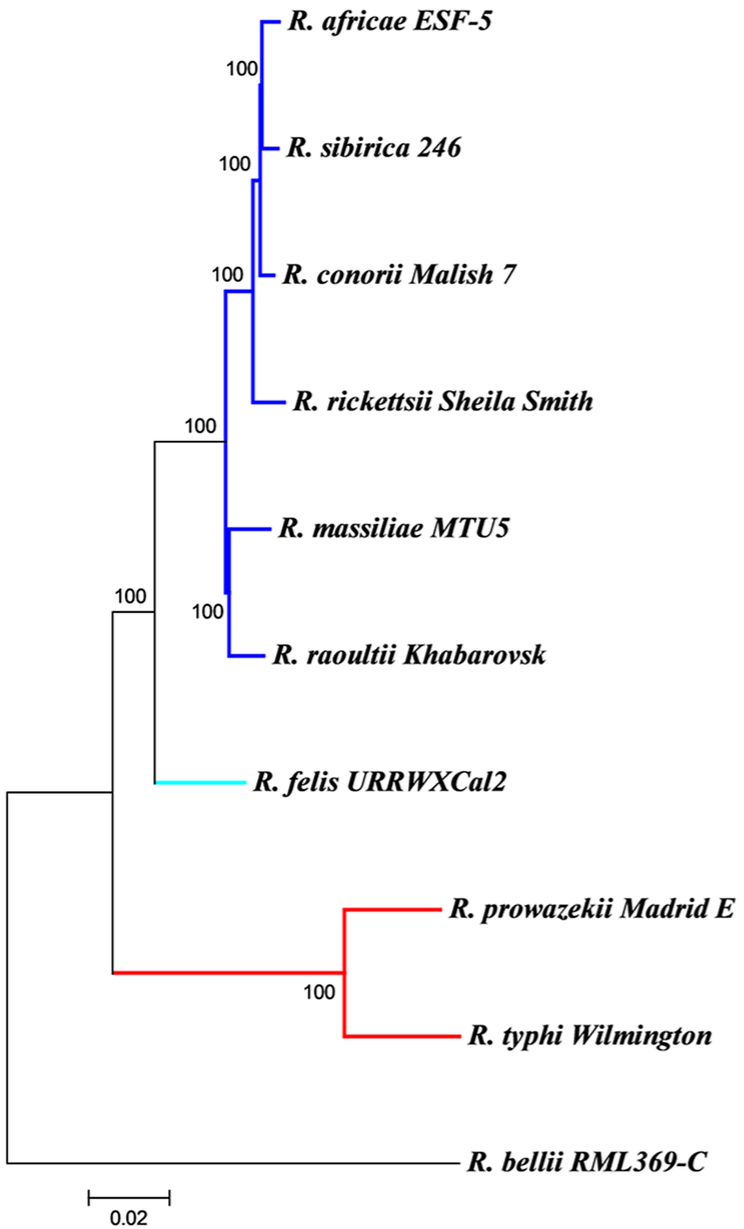
cRIG00474	tdpX1 Thioredoxin peroxidase	tdpX
cRIG00475	hflC1 Membrane protease subunit, stomatin/prohibitin-like protein	hfl
cRIG00476	Putative membrane-associated metal-dependent hydrolase	-
cRIG00477	Efflux transporter, RND family, MFP subunit	rnd
cRIG00478	Putative hydrolase/acyltransferase	-
cRIG00479	Glycosyltransferase	-
cRIG00480	rpsD 30S ribosomal protein S4	rps
cRIG00481	cyoB, ctaB Protoheme IX farnesyltransferase	cyoB
cRIG00482	rimM 16S rRNA processing protein RimM	rimM
cRIG00483	Unknown	-
cRIG00484	xseB Exodeoxyribonuclease VII small subunit	xse
cRIG00485	mpg DNA-3-methyladenine glycosidase	mpg
cRIG00486	nuoE NADH dehydrogenase I chain E	nuo
cRIG00487	nuoD NADH dehydrogenase I chain D	nuo
cRIG00488	nuoC NADH dehydrogenase I chain C	nuo
cRIG00489	nuoA NADH dehydrogenase I chain A	nuo
cRIG00490	cutE Apolipoprotein N-acyltransferase	cut
cRIG00491	lysS Lysyl-tRNA synthetase	lys
cRIG00492	Putative permease	-
cRIG00493	tme Malate oxidoreductase	tme
cRIG00494	proP3 Proline/betaine transporter	proP
cRIG00495	mdh Malate dehydrogenase	mdh
cRIG00496	tlc2 ATP/ADP translocase	tlc
cRIG00497	pyrG CTP synthase	pyr
cRIG00498	kdsB 3-deoxy-manno-octulosonate cytidylyltransferase	kds
cRIG00499	folE GTP cyclohydrolase I	fol
cRIG00500	proS Prolyl-tRNA synthetase	proS
cRIG00501	ruvB Holliday junction DNA helicase RuvB	ruv
cRIG00502	msbA1 Multidrug resistance protein	msbA
cRIG00503	dacF Penicillin-binding protein dacF precursor	dac
cRIG00504	rlpA Rare lipoprotein A precursor	rlpA
cRIG00505	osmY Putative periplasmic or secreted lipoprotein	osmY
cRIG00506	ispZ Intracellular septation protein A	isp
cRIG00507	FTR1 family protein	FTR1
cRIG00508	Unknown	-
cRIG00509	Unknown	-
cRIG00510	Protocatechuate-3,4-dioxygenase, beta subunit	-
cRIG00511	tlpA Thioldisulfide interchange protein tlpA	tlpA
cRIG00512	sppA1 Signal peptide peptidase SppA, 36K type	sppA
cRIG00513	dut Deoxyuridine 5'-triphosphate nucleotidohydrolase	dut
cRIG00514	mltESoluble lytic murein transglycosylase precursor	mltE
cRIG00515	mccF Microcin C7 self-immunity protein	mccF
cRIG00516	RecB family exonuclease	rec
cRIG00517	Putative glutamine amidotransferase	-
cRIG00518	coxA Cytochrome c oxidase polypeptide I	cox
cRIG00519	coxB Cytochrome c oxidase polypeptide II	cox
cRIG00520	nlpD2 Membrane-bound metallopeptidase	nlpD
cRIG00521	ftsW Cell division protein ftsW	fts

cRIG00522	murG UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	mur
cRIG00523	Unknown	-
cRIG00524	Glycosyltransferase	-
cRIG00525	dapF Diaminopimelate epimerase	dap
cRIG00526	MiaB-like tRNA modifying enzyme	Mia
cRIG00527	pheS Phenylalanyl-tRNA synthetase alpha chain	phe
cRIG00528	pheT Phenylalanyl-tRNA synthetase beta chain	phe
cRIG00529	dnaN DNA polymerase III beta chain	dna
cRIG00530	Unknown	-
cRIG00531	rbfA Ribosome-binding factor A	rbf
cRIG00532	Putative membrane protein	-
cRIG00533	RDD family protein	rdd
cRIG00534	recR Recombination protein RecR	rec
cRIG00535	ppnK Putative inorganic polyphosphate/ATP-NAD kinase	ppnk
cRIG00536	Putative hydrolase of the metallo-beta-lactamase superfamily	-
cRIG00537	gpsA Glycerol--phosphate dehydrogenase	pgsA
cRIG00538	Putative permease	-
cRIG00539	Putative hydrolase/acyltransferase	-
cRIG00540	trxB1 Thioredoxin reductase	trx
cRIG00541	lgtD Glycosyl transferase	lgt
cRIG00542	uvrD DNA helicase II	uvr
cRIG00543	Unknown	-
cRIG00544	tdcB Threonine dehydratase	tdcB
cRIG00545	lon1 ATP-dependent protease La	lon1
cRIG00546	yhbH Putative sigma(54) modulation protein	yhbH
cRIG00547	folD Methylenetetrahydrofolate dehydrogenase	fol
cRIG00548	trxB2 Thioredoxin reductase	trx
cRIG00549	nrdA Ribonucleoside-diphosphate reductase alpha chain	nrd
cRIG00550	nrdB Ribonucleoside-diphosphate reductase beta chain	nrd
cRIG00551	Unknown	-
cRIG00552	kpsF KpsF	kpsF
cRIG00553	pnp Polyribonucleotide nucleotidyltransferase	pnp
cRIG00554	rpsO 30S ribosomal protein S15	rps
cRIG00555	truB tRNA pseudouridine synthase B	truB
cRIG00556	tlc4 ATP/ADP translocase	tlc
cRIG00557	sca Cell surface antigen Sca	Sca
cRIG00558	glnA Glutamine synthetase	gln
cRIG00559	Unknown	-
cRIG00560	ppdK Pyruvate,phosphate dikinase precursor	ppdk
cRIG00561	Glutathione S-transferase	-
cRIG00562	folC Folylpolyglutamate synthase	fol
cRIG00563	sodB Superoxide dismutase	sodB
cRIG00564	rssA Putative esterase of the alpha/beta hydrolase superfamily protein	rssA
cRIG00565	birA Biotin-(acetyl-CoA carboxylase) ligase	birA
cRIG00566	rho Transcription termination factor	rho
cRIG00567	mraZ MraZ protein	mra
cRIG00568	mraW S-adenosyl-methyltransferase MraW	mra

cRIG00569	ftsL Cell division protein FtsL	fts
cRIG00570	pbpA2 Penicillin-binding protein	pbpA
cRIG00571	pbpA1 Penicillin-binding protein	pbpA
cRIG00572	Unknown	-
cRIG00573	Unknown	-
cRIG00574	ntrX Nitrogen assimilation regulatory protein NtrX	ntr
cRIG00575	ubiH 2-polyprenyl-6-methoxyphenol 4-hydroxylase	ubi
cRIG00576	nusA N utilization substance protein A, transcription termination factor NusA	nus
cRIG00577	infB Translation initiation factor IF-2	inf
cRIG00578	Putative glycoprotein endopeptidase	-
cRIG00579	Unknown	-
cRIG00580	N6-adenine-specific methylase	N6
cRIG00581	rluA2 Ribosomal large subunit pseudouridine synthase	rluA
cRIG00582	dnaB Replicative DNA helicase	dna
cRIG00583	ubiX 3-octaprenyl-4-hydroxybenzoate carboxy-lyase	ubi
cRIG00584	priA Primosomal protein N'	priA
cRIG00585	hemB Delta-aminolevulinic acid dehydratase	hemB
cRIG00586	Unknown	-
cRIG00587	nuoN1 NADHubiquinone oxidoreductase subunit (chain N)	nuo
cRIG00588	proP6 Proline/betaine transporter	proP
cRIG00589	dus Putative dihydrouridine synthase Dus	dus
cRIG00590	phbC Poly-beta-hydroxybutyrate polymerase	phb
cRIG00591	Unknown	-

[a]cRIG = Rickettsial orthologous cluster


R. parkerii Portsmouth (NC_017044)
R. africae ESF-5 (CP001612)
R. sibirica 246 (AABW00000000)
R. conorii Malish 7 (NC_003103)
R. slovaca 13-B (NC_016639)
R. rickettsii Sheila Smith (NC_009882)
R. peacockii Rustic (CP001227)
R. honei RB (AJTT00000000)
R. heilongjiangensis O54 (CP002912)
R. japonica YH (NC_016050)
R. gravesii BWI-1 (AWXL00000000)
R. raoultii Khabarovsk (CP010969)
R. amblyommatis Ac/Pa (LANR00000000)
R. massiliae MTU5 (NC_009900)
R. rhipicephali 3-7female6-CWPP (NC_017042)
R. aeschlimannii MC16 (CCER00000000)
R. montanensis OSU 85-930 (CP003340)
R. buchneri ISO-7
R. tamurae AT-1 (CCMG01000008)
R. akari Hartford (CP000847)
R. australis Cutlack (NC_017058)
R. hoogstraalii Croatica (CCXM00000000)
R. asembonensis NMRCii (JWSW00000000)
R. felis URRWXCal2 (NC_007109)
R. helvetica C9P9 (CM001467)
R. prowazekii Madrid E (NC_000963)
R. typhi Wilmington (NC_006142)
R. canadensis Mckiel (NC_009879)
R. bellii RML-369C (NC_007940)
100
99
SFG
TRG
TG
AG
0.005



Virulence	Size (bp)	protein-coding genes	% coding sequences	RNAs	pseudogenes	RPEs
Milder	1,278,540 pRaf:12,377	1112	78.26	39	246	460
Virulent	1,250,021	1083	77.76	36	200	NA
Virulent	1,268,755	1374	81.5	39	252	559
Virulent	1,257,710	1345	78.5	36	233	NA
Mild	1,360,898 pRma:15,286	968	69	39	286	562
Mild	1,344,605 pRra1:20,840 pRra2:83,219 pRra3:34,583	1180	71.2	39	339	NA
Mild	1,485,148 pRF:6,282 pRFð:39,268	1444	83.8	39	130	726
Highly virulent	1,111,523	834	76.2	39	181	120
Virulent	1,111,496	838	76.3	39	185	121
Unknown	1,522,076	1429	82.5	40	100	526