



OPEN

Genomic evolution and adaptation of arthropod-associated *Rickettsia*

Khalid El Karkouri^{1,2}, Eric Ghigo², Didier Raoult^{2,3} & Pierre-Edouard Fournier^{1,2}

Rickettsia species are endosymbionts hosted by arthropods and are known to cause mild to fatal diseases in humans. Here, we analyse the evolution and diversity of 34 *Rickettsia* species using a pangenomic meta-analysis (80 genomes/41 plasmids). Phylogenomic trees showed that *Rickettsia* spp. diverged into two Spotted Fever groups, a Typhus group, a Canadensis group and a Bellii group, and may have inherited their plasmids from an ancestral plasmid that persisted in some strains or may have been lost by others. The results suggested that the ancestors of *Rickettsia* spp. might have infected Acari and/or Insecta and probably diverged by persisting inside and/or switching hosts. Pangenomic analysis revealed that the *Rickettsia* genus evolved through a strong interplay between genome degradation/reduction and/or expansion leading to possible distinct adaptive trajectories. The genus mainly shared evolutionary relationships with α -proteobacteria, and also with $\gamma/\beta/\delta$ -proteobacteria, cytophagia, actinobacteria, cyanobacteria, chlamydia and viruses, suggesting lateral exchanges of several critical genes. These evolutionary processes have probably been orchestrated by an abundance of mobile genetic elements, especially in the Spotted Fever and Bellii groups. In this study, we provided a global evolutionary genomic view of the intracellular *Rickettsia* that may help our understanding of their diversity, adaptation and fitness.

Rickettsia species (spp.) (Order Rickettsiales, Family Rickettsiaceae) are obligate intracellular α -proteobacteria of both ecological and medical interest. These bacteria live in close association with a diverse range of hosts including arthropods, mostly, and vertebrates, plants, algae, annelids, amoebae, ciliate and medusae (as primary hosts and vectors)^{1–6}. *Rickettsia* spp. are transmitted to humans or animals, mostly through arthropod bites, and are responsible for mild to fatal diseases such as epidemic typhus and Rocky Mountain spotted fever^{3,7}.

Obligate intracellular bacterial endosymbionts and pathogens evolve in diverse host environments. These biological and ecological lifestyles make their genetic manipulations and understanding the mechanisms of their survival and pathogenicity^{8,9} difficult. To overcome this barrier, evolutionary biology using comparative genomics is a powerful tool to decipher genomic signatures that have shaped bacterial genomes and to understand their diversification, adaptation and fitness. Earlier investigations of *Rickettsia* spp. using comparative genomics have revealed several genetic and evolutionary features within these small (1.1–2.3 Mbp) genomic bacteria, such as (i) a high degree of inter-species genomic synteny; (ii) reductive evolution possibly in relation to their strict intracellular lifestyle; (iii) an enrichment in A + T content, polyA/T homopolymers; (iv) a lack of some metabolic pathways for which host cells provide the missing metabolites; (v) conjugative machinery and mechanisms of adhesion to host and motility; and (vi) a variable distribution of plasmids^{10–28}.

Although initially thought to be devoid of plasmids, several *Rickettsia* spp. were demonstrated to harbour such mobile genetic elements (MGEs)²⁷. While no obvious association between pathogenicity and acquisition of virulence factors was found in *Rickettsia* spp.²⁰, other studies associated the increased virulence with genome reduction in this genus²⁹. However, due to the absence of several genome sequences of the genus *Rickettsia* in public databases, all these comparative genomic investigations were carried out from subsets of species, sometimes drawing conclusions for a species by using a single strain.

In this study, we investigated the genomic evolution and diversity of the obligate intracellular *Rickettsia* spp. using phylogenomic and pangenomic meta-analyses from 34 arthropod-associated species which are represented by 80 genome sequences that we here reannotated. We investigated both the chromosomes and neglected plasmids, as both replicons are part of the historical life of this genus.

¹Aix-Marseille University, IRD, AP-HM, SSA, VITROME, Marseille, France. ²IHU Méditerranée Infection, Marseille, France. ³Aix-Marseille University, IRD, AP-HM, MEPHI, Marseille, France. ✉email: khalid.elkarkouri@univ-amu.fr; pierre-edouard.fournier@univ-amu.fr

Results

In this study, we examined the diversity, evolution and adaptation of the genus *Rickettsia* from 34 species with validated names. The 80 studied strains had been isolated from diverse hosts between 1938 and 2012 worldwide (Table S1). In order to obtain a robust comparative genome analysis, we performed standard re-annotations of 61 complete genomes and 19 draft genomes sequences. Their general genomic features are summarised in Supplementary Table S1.

Phylogenomics and diversity. We first inferred a Maximum Likelihood phylogenomic tree from 461 concatenated core chromosomal proteins of the 80 studied strains (Fig. 1). Overall, our robust evolutionary tree showed that the *Rickettsia* genus has diverged into five distinct and well-supported major groups (BP = 100%): the Spotted Fever groups (SFGI and SFGII), the Typhus group (TG), the Canadensis group (CG) and the Bellii group (BG).

The SFGI rickettsiae group included several closely related species among which the core genome is highly conserved as reflected by short branches in the tree. This lineage is clearly split into two distinct sister sub-clades of rickettsiae (BP node supports = 100%), SFGIa and SFGIb, containing 20 and four species (e.g., *R. conorii* and *R. endosymbiont* of *I. scapularis*), respectively. The next well-resolved group is designated here as the SFGII rickettsiae group (also referred to as the transitional group, TRG), which includes five spotted fever rickettsial species (e.g., *R. felis* and *R. australis*). The core genome of this group appears to be more divergent compared with the SFGI group.

Another member of the SFG species, *R. helvetica*, was placed as an individual clade basal to the TG group but not the SFG group (Fig. 1). Examination of the multiple sequence alignments of *Rickettsia* core genes revealed that *R. helvetica* shares several non-synonymous mutations and large insertions/deletions with the SFGII, TG, CG and/or BG rickettsiae (Supplementary Fig. S1).

Remarkably, the TG group is clearly monophyletic and displays the longest and most diversified branch in the tree as compared with its close relatives (Fig. 1). This group accumulated several mutations in its core genome and then recently diverged into the *R. prowazekii* and *R. typhi* species. Last, the CG and BG groups were placed on two external branches of our phylogenomic tree. The former group includes *R. canadensis* and is clearly basal to the SFG and TG groups, whereas the latter contains *R. bellii* and is the earliest divergent outgroup among studied taxa.

Plasmids. Our genomic annotation detected 41 rickettsial plasmids, of which 11 were newly identified (Fig. 1, Table S1). These plasmids are present in 56% and 33% of examined species and strains, respectively. Thus, 44% and 67% of species and strains, respectively, had no plasmids. By mapping the current large-scale plasmid data set over the five rickettsial phylogroups (SFGI, SFGII, TG, CG and BG), we found that rickettsial plasmids were variably distributed between and within these lineages. They were present in only three phylogroups (SFGI, SFGII and BG), with between one and four plasmid(s) per strain. A large number of rickettsial plasmids (three or four) were only observed in three SFG species, *R. raoultii*, *R. amblyommatidis* and *R. buchneri*.

Interestingly, we detected typical plasmidic genes in the chromosomes from the 11 plasmidless *R. prowazekii* strains and the two *R. raoultii* strains (one exhibiting three plasmids and the other being plasmidless). Indeed, each of the plasmidless *R. prowazekii* strains showed a cluster containing five pseudogenes (RprME0791 to RprME0807) that exhibited high homologies (ids = 67–92%) with rickettsial plasmid genes. Similarly, both *R. raoultii* strains harboured a cluster of six genes (Rra_909 to Rra_0916) that best matched (ids = 82–97%) six genes present exclusively in the *R. peacockii* plasmid (pRpe).

The phylogenetic trees computed from 5 selected genes common to several plasmids (Supplementary Fig. S2) and identified using the pan-plasmidome analysis revealed that each displayed two major ancestral nodes, one gathering all plasmids and the other all chromosomes of *Rickettsia/Orientia*. These results clearly show that these genes originated from a *Rickettsia/Orientia* ancestor harbouring one ancestral chromosome and one ancestral plasmid. Moreover, the supertree constructed from these five individual gene trees confirmed the distinction of both phylogenetic clades and their nodes, which were obtained in each of the 5 phylogenetic trees (Fig. 2).

Arthropod hosts. We compared the diversity of arthropods (as the primary hosts) associated with *Rickettsia* spp. across the five rickettsial phylogroups (Fig. 1). We found that most SFGI *Rickettsia* spp. were mainly associated with ticks (Acari), similarly to CG and BG rickettsiae. For example, the SFGI group was associated with seven genera and 27 species of ticks (*Amblyomma*, *Argas*, *Dermacentor*, *Haemaphysalis*, *Hyalomma*, *Ixodes* and *Rhipicephalus*). Moreover, some SFGI, CG and BG *Rickettsia* spp. were found to be associated with the same tick genus and/or species. However, SFGII rickettsiae were found in ticks, mites (Acari), fleas, and lice (Insecta). Furthermore, some rare strains of the four groups (SFGI, SFGII, CG and BG), as well as all those belonging to TG rickettsiae were identified in vertebrates such as humans or rats.

Pan-chromosome and pan-plasmidome. Using predicted genes from both chromosomes and plasmids, we separately inferred the *Rickettsia* pan-chromosome of 34 species (80 strains) and pan-plasmidome of 19 species (41 plasmids). The pan-chromosome of *Rickettsia* spp. included 4073 cRiGs (orthologous clusters of chromosomal rickettsial genes) clustered from 93,691 protein-coding genes (117,821 CDSs) (see an example of the pan-chromosome patterns across ten genomes in Supplementary Fig. S3), while the pan-plasmidome of *Rickettsia* spp. contained 457 pRiGs (orthologous clusters of plasmidic rickettsial genes) grouped from 1502 protein-coding genes (1893 CDSs).

Within the pan-chromosome, only 15% (599) of cRiGs represented the core chromosome (present in all strains), whereas the remaining 85% (3474) of cRiGs corresponded to the flexible chromosome (present in a

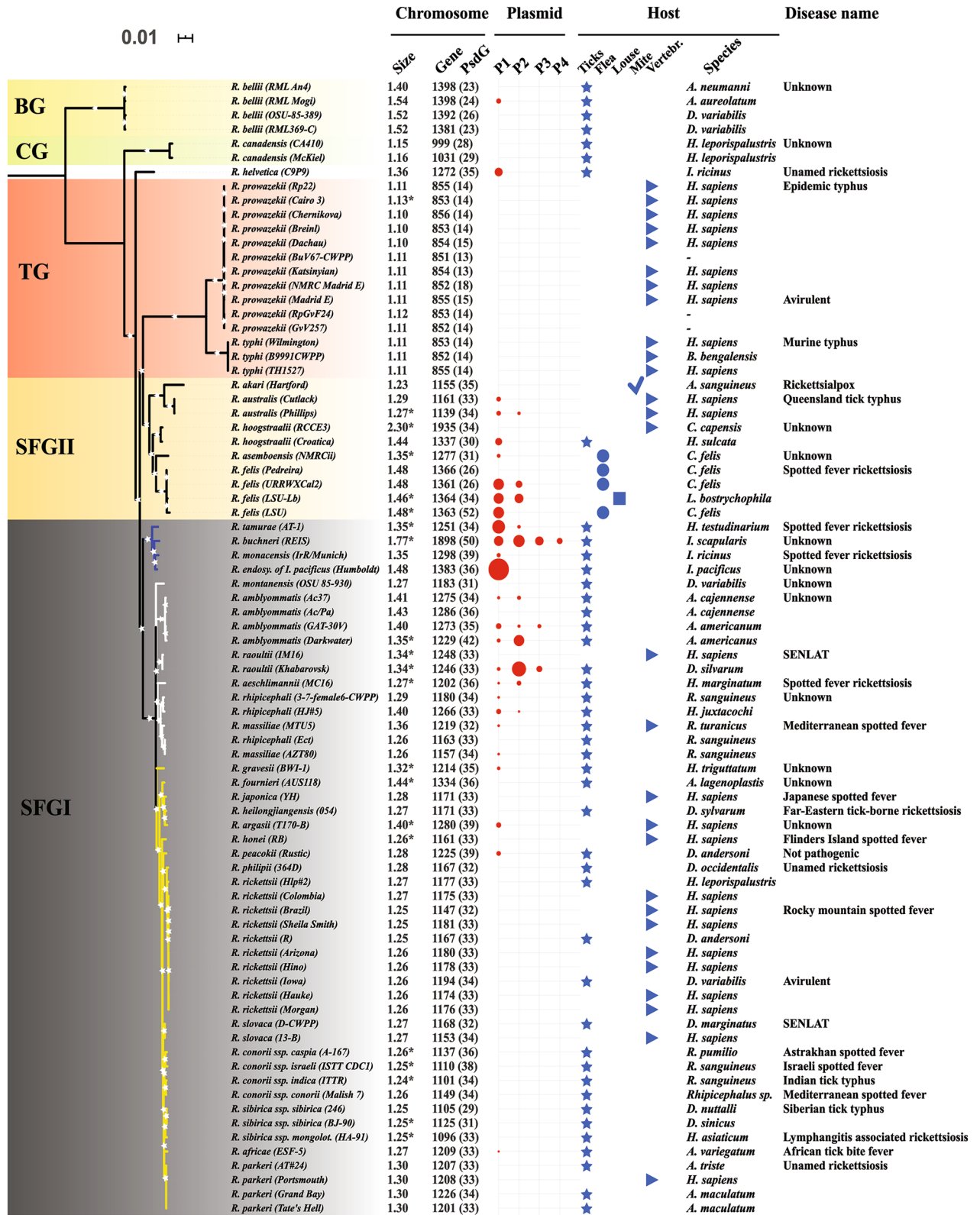


Figure 1. Maximum likelihood phylogenomic tree of arthropod-associated *Rickettsia* spp. estimated from 461 core chromosomal genes containing 36,880 core genes of 34 species and 80 strains. For each strain, the columns show its chromosome size, gene content (pseudogenes in %), number of plasmids and their sizes (presented by the diameter of circles) and host (either familiar or taxonomic name), as well as the name of the disease that it causes in humans. A: *Amblyomma*, H: *Homo sapiens*, D: *Dermacentor*, R: *Rhipicephalus*, A: *lagenoplastis*; *Argas lagenoplastis*, *H. leporispalustris/juxtacochi/sulcata*: *Haemaphysalis leporispalustris/juxtacochi/sulcata*, *H. asiaticum/marginatum/testudinarium/triguttatum*: *Hyalomma asiaticum/marginatum/testudinarium/triguttatum*, I: *Ixodes*, A: *sanguineus*: *Allodermanyssus sanguineus*, C: *Ctenocephalides*, L: *Liposcelis*, C: *capensis*: *Corvus capensis* and B: *Bandicota*.

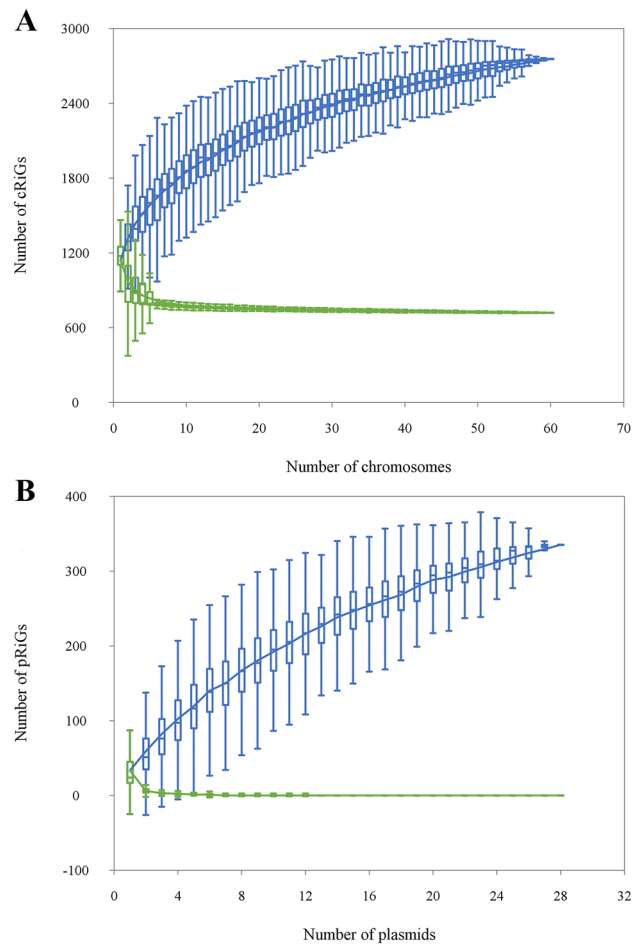


Figure 3. Estimation of the core chromosome ($y = 497 e^{-0.312x} + 737$; $R^2 = 0.95$) and pan-chromosome ($y = 1029 x^{0.23} + 87$; $R^2 = 0.99$) established through the analysis of 60 complete chromosomes of 28 *Rickettsia* spp. (A). Estimation of the core plasmidome ($y = 144.2 e^{-1.48x} + 0.36$; $R^2 = 0.99$) and pan-plasmidome ($y = 100.7 x^{0.42} - 74.3$; $R^2 = 0.99$) simulated from 27 complete plasmids representing 15 *Rickettsia* spp. (B). Box plots show the median and quartiles.

riddled by stop codons which may result from a neutral/adaptive event and/or an incomplete sequence assembly of highly repetitive chromosomes (see below).

Moreover, the analysis of the chromosome expansions across the five phylogroups revealed that the total numbers of chromosomal gene duplications were remarkably variable and higher in the SFGI (from 52 to 837) and SFGII (from 74 to 805) groups (Fig. 4). They decreased with low variations in the BG group (from 163 to 194) and much more in the CG (from 35 to 53) and TG (from 4 to 12) groups. Similarly, the frequency of chromosomal gene amplifications increased in the SFGI (from two to 186 copies) and SFGII (from 2 to 104 copies) groups and decreased in the BG group (from two to 27 copies) and much more in the CG (from two to nine copies) and TG (two copies) groups. Two strains of the SFGI and SFGII groups, *R. buchneri* REIS and *R. hoogstraalii* RCCE3 were a remarkable illustration of a massive genome expansion in the genus.

We then analysed the possible contribution of prophages in the chromosome sizes of some species representatives of each phylogroup that exhibited large variations in their chromosome sizes (Table S2). We detected no prophage in the smallest TG *R. prowazekii* Rp22 (1.1 Mbp) and CG *R. canadensis* (1.15 Mbp) chromosomes nor in the large SFGI *R. asembonensis* NMRCii chromosome (1.3 Mb). Moreover, all detected prophages were defective (either incomplete or questionable) and exhibited sizes ranging from 5.7 to 31.8 kb. When summing the sizes of these prophages for each species, we found that they represent less than 0.004% of the chromosome sizes in all phylogroups and species examined. As examples, the cumulated sizes of prophages represent only 0.0015%, 0.0004% and 0.0007% of the largest chromosomes, i.e., *R. bellii* RML369-C (1.5 Mb, BG), *R. buchneri* REIS (1.7 Mb, SFGI) and *R. hoogstraalii* RCCE3 (2.3 Mb, SFGII), respectively.

Plasmids. Plasmid sizes also varied between and within the SFGI (from 12.3 to 121.4 kb), SFGII (from 19.4 to 63.8 kb) and BG (28.7 kb) phylogroups (Fig. 1, Table S1). The pRaf, pRae1 and pRma plasmids were the smallest plasmids in the *Rickettsia* genus (about 12–15 kb in sizes). Moreover, the numbers of plasmids were also variable and greater in the SFGI (from zero to four plasmids) and SFGII (from zero to three) groups compared to the BG group (zero to one).

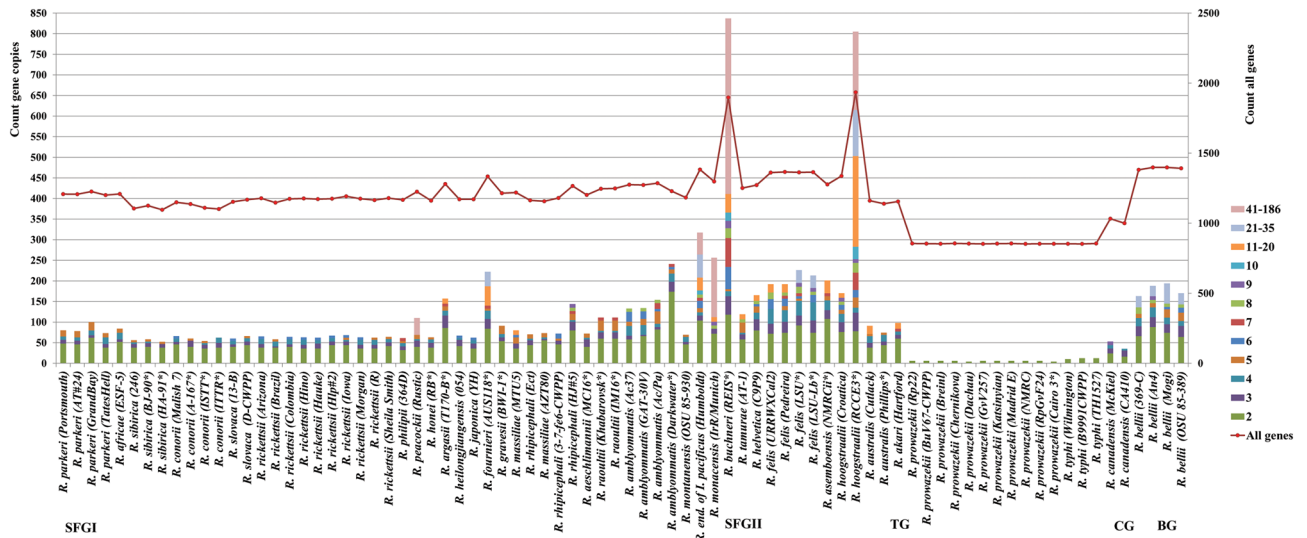


Figure 4. Distributions of gene proliferations identified in the pan-chromosome across 34 *Rickettsia* spp. (80 strains or chromosomes). The x-axis represents the 80 *Rickettsia* chromosomes examined. The left y-axis shows the total number of gene copies present in each chromosome. For each chromosome, colours indicate the numbers of genes according to their copy counts (from 2 to 186). The right y-axis displays the total gene count per chromosome.

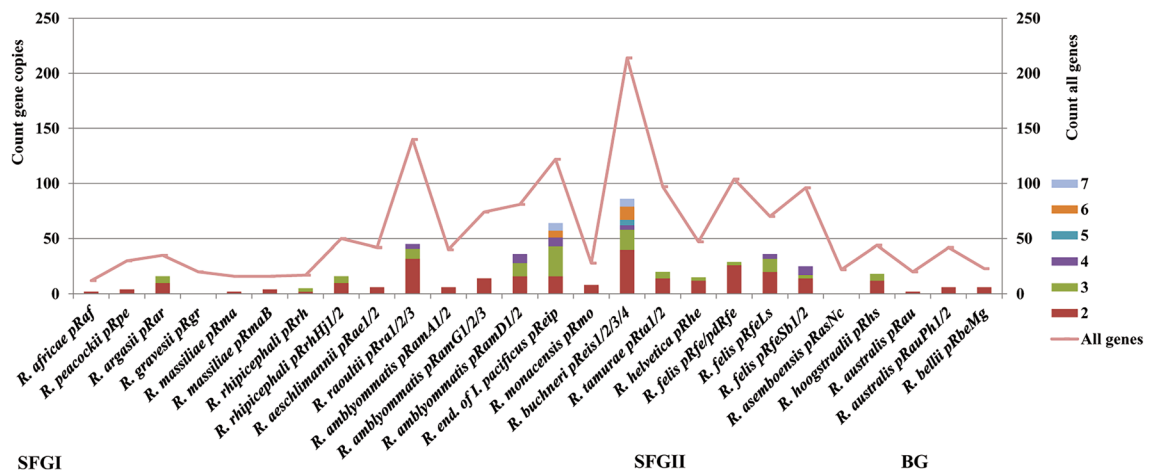


Figure 5. Distributions of gene proliferations identified in the pan-plasmidome across 19 *Rickettsia* spp. (26 strains, 41 plasmids). The x-axis represents the 26 *Rickettsia* strains, each harbouring from 1 to 4 plasmids. The left y-axis shows the total number of plasmidic gene copies present in each strain. For each plasmid, colours indicate the numbers of genes according to their copy counts (from 2 to 7). The right y-axis displays the total plasmidic gene count per strain.

Examination of the plasmid degradation revealed that the proportion of pseudogenes was highly variable in the SFGI (from 0 to 64.7%) and SFGII (from 17.9 to 58.6%) phylogroups (Fig. 1, Table S1). The largest proportions (> 50) of pseudogenes were found in several SFG plasmids, either complete or draft (e.g., pRamA2, pReis4 and pRfELs).

The distribution of the total numbers of gene duplications in plasmids varied within the SFGI (from two to 64) and SFGII (from four to 36) phylogroups (Fig. 5). However, the frequency of gene proliferations in plasmids was low (from two to seven copies).

Evolutionary networks. To infer the evolutionary relationships of the *Rickettsia* genus, we constructed an initial homology network of the pan-chromosome (4073 cRiGs, including 599 core and 3474 flexible genes) by identifying the first best hit of each cRiG in the pan-plasmidome and non-*Rickettsia* lineages (Fig. 6). This network excluded several plasmidic genes that did not have any sequence match with the pan-chromosome. Then, a network of the pan-plasmidome (457 flexible pRiGs) was constructed by detecting the first and second best hits of each pRiG in the pan-chromosome and non-*Rickettsia* lineages (Fig. 7). This latter analysis distinguished three gene sets: a first gene set in which the pan-plasmidome matched the pan-chromosome; a second gene set in

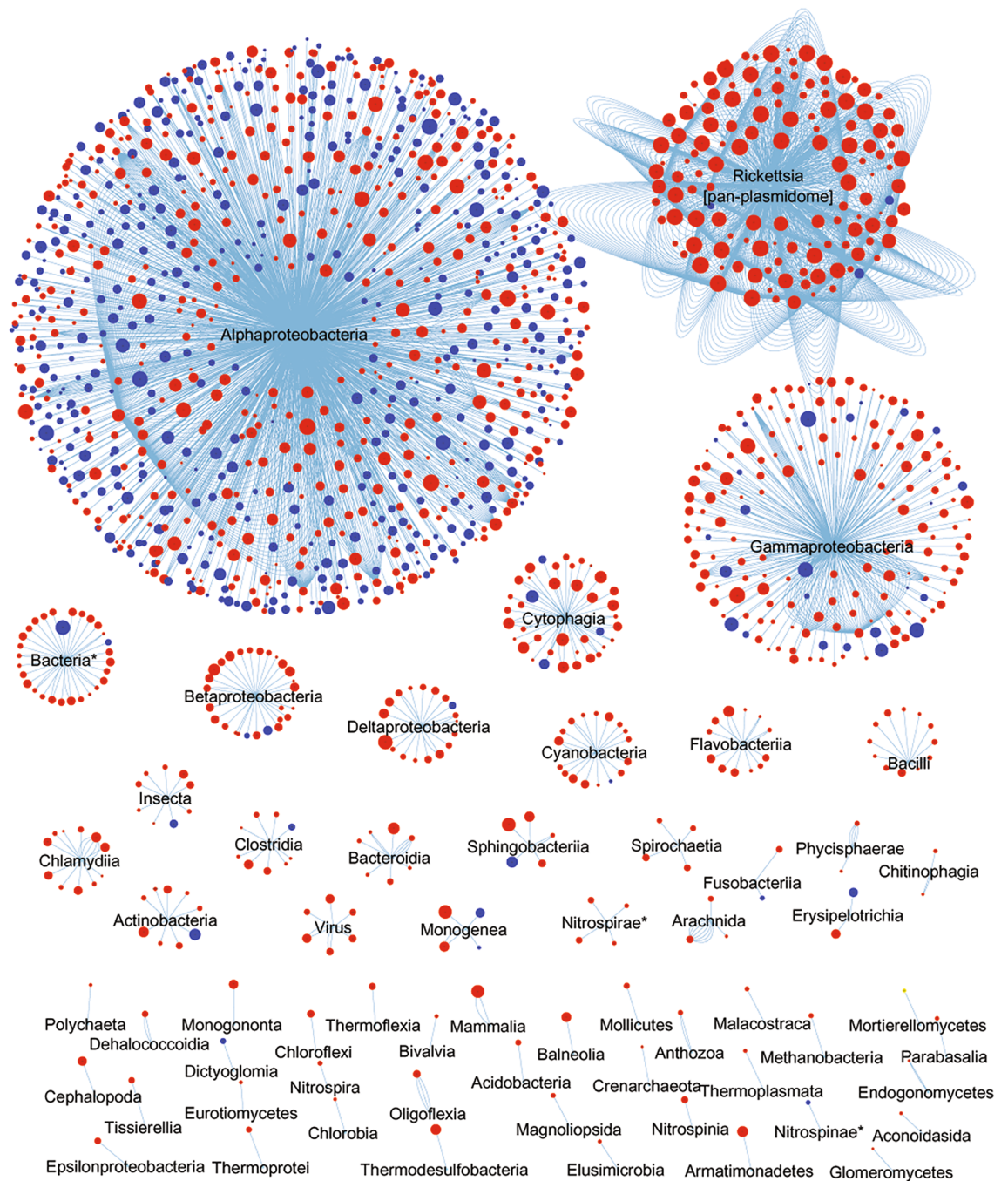


Figure 6. Evolutionary network of the *Rickettsia* pan-chromosome displaying best homologues from the *Rickettsia* pan-plasmidome and customised databases containing non-*Rickettsia* protein sequences of bacteria, archaea, eukarya and viruses. Nodes correspond to the pan-chromosome cRiGs. Node colours correspond to the core (in blue) or flexible (in red) cRiGs. Node sizes refer to sequence identities (from 30 to 100%) between a cRiG and its first best homologue, which are linked by blue edges. * unclassified bacteria.

which the pan-plasmidome matched other lineages; and a third gene set in which the pan-plasmidome matched both the pan-chromosome and other lineages.

The pan-chromosome network. We found that 75% of the pan-chromosome displayed best sequence homologues (identities = 30–100%) with the pan-plasmidome and diverse bacterial, archaeal, eukaryotic and viral taxa (56 classes/105 orders/162 families) as well as other unclassified organisms (Fig. 6). Within this taxonomic pattern, the pan-chromosome shared best sequence similarities (34.6%, including 503 core and 906 flexible genes) with the α -proteobacteria (Fig. 6, see examples of genes in Table S3). Remarkably, a second large fraction of the pan-chromosome (25.3% including only three core and 1026 flexible genes) exhibited best sequence similarity matches with the pan-plasmidome. Interestingly, a third gene set of the pan-chromosome contained

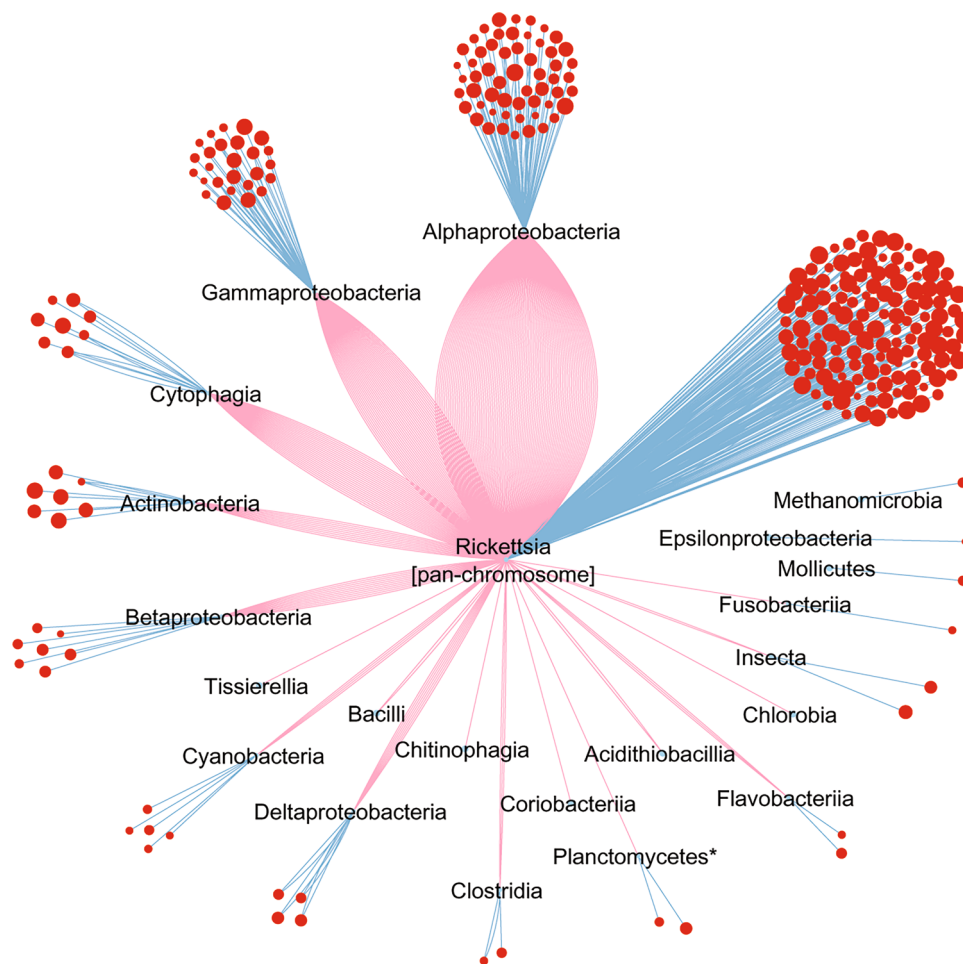


Figure 7. Evolutionary network of the *Rickettsia* pan-plasmidome exhibiting homologues from the *Rickettsia* pan-chromosome and customised databases containing non-*Rickettsia* protein sequences of bacteria, archaea, eukarya and viruses. Nodes represent the flexible pan-plasmidome pRiGs (in red). Node sizes refer to sequence identities (from 30 to 100%) between a pRiG and its first best homologue (blue edges). Some pRiGs also displayed second best homologues (red edges).

small proportions of genes (from 7.1 to 1.2%; including 60 core and 551 flexible genes) that exhibited a best hit with distant bacterial lineages largely belonging to γ -proteobacteria, cytophagia, and β/δ -proteobacteria (Fig. 6, Table S3). Smaller fractions of genes (from 0.9 to 0.02%) best matched 54 more distant taxonomic classes (either identified or unidentified) such as cyanobacteria, chlamydia and viruses (Fig. 6, Table S3). The remaining pan-chromosome (25%) had no homologue with the pan-plasmidome and non-*Rickettsia* lineages available in the databases.

The pan-plasmidome network. Overall, 72% of the pan-plasmidome genes exhibited best similarities (ids = 30–100%) with the pan-chromosome, diverse lineages (15 classes/35 orders/47 families) and/or other unclassified organisms (Fig. 7), whereas the remaining 28% had no homologue with our datasets and databases. Among this taxonomic profile, 10% (as first gene set of the 72%) of the pan-plasmidome displayed best matches with the pan-chromosome only, and 18% (as second gene set of the 72%) exhibited best hits with α -proteobacteria (9%), followed by γ -proteobacteria (3.9%), cytophagia (0.9%), actinobacteria (1.3%), δ -proteobacteria (1.3%), β -proteobacteria (1.1%), and cyanobacteria (0.4%), and then with much smaller proportions with four lineages (0.2%) (Fig. 7, Table S3). The remaining 44% (as third gene set of the 72%) of the pan-plasmidome showed best sequence similarities with the pan-chromosome and several lineages including α -proteobacteria (19.1%), γ -proteobacteria (8.3%), cytophagia (5.4%), β -proteobacteria (2.4%), actinobacteria (2.2%), δ -proteobacteria (1.5%), cyanobacteria (0.7%) and other lineages (from 0.4 to 0.2%) (Fig. 7, Table S3).

Discussion

In this study, we examined the diversity, evolution and adaptation of arthropod-associated *Rickettsia* spp. using genomics, phylogenomics and pangenomics data analysis from 34 species (80 strains).

Overall, using phylogenomic analysis we found that the *Rickettsia* genus has diverged into five distinct major groups: two Spotted Fever groups (SFGI and SFGII), the Typhus group (TG), the Canadensis group (CG) and the Bellii group (BG). Our phylogenomic tree corroborates other rickettsial phylogenies estimated from subsets of species, strains and/or core genes/proteins^{2,21,22,26,30–36}. However, although some of these phylogenies included a novel rickettsial clade (from *Adalia*)^{2,30,36} and four others (from *Hydra* spp., *Torix* spp., *Rhizobius* spp. and meloidae)^{2,4–6,30}, they would not impact our robust five rickettsial phylogroups. Indeed, the *Adalia* clade was located between the CG and BG groups, while the other four clades were all external to the BG group. The rare specimens^{2,4–6,30,36} previously clustered in these new clades were not included in our study because no genomic sequence of these bacteria was available. Therefore, the BG phylogroup represents the outgroup of our data set, as it is the earliest divergent group among our taxa.

The 80 *Rickettsia* strains that we studied were isolated from diverse hosts worldwide, including Acari, Insecta and vertebrates. Examination of the diversity of arthropods of these bacterial species across the five rickettsial phylogroups revealed that the SFGI, CG and BG *Rickettsia* spp. were mainly associated with ticks (Acari). However, some species of these groups were also identified in mosquitoes, fleas and/or sheep keds (SFGI in Insecta)^{37–39}, and in beetles, mosquitoes and/or flies (CG and BG in Insecta)^{2,40}. The examined SFGII rickettsiae were detected in ticks, mites (Acari), fleas or lice (Insecta) and, in the case of *R. felis*, in ticks, mites (Acari) and mosquitoes (Insecta)^{3,39–43}. Other *Rickettsia* spp. of our dataset were found to be associated with vertebrates². For example, the TG species (*R. prowazekii* and *R. typhi*) were identified in humans and rats, respectively. However, these species are known to be associated with fleas and/or lice (Insecta)³, but were also detected in ticks^{44,45}. On the other hand, our study found that 55% of *Rickettsia* strains were associated with ticks.

Previous data indicated that *Rickettsia* species are mainly associated with acarian and/or insect hosts, with a predominance in ticks. This entails two evolutionary scenarios that might have occurred for the *Rickettsia*-host association. First, an ancestor of *Rickettsia* spp. might have been associated with an acarian and/or an insect prior to its diversification. This suggests that *Rickettsia* spp. may have been maintained and have diverged inside diverse arthropod species through vertical transmissions (transovarial and trans-stadial). However, co-speciation between *Rickettsia* and arthropod species appears to be ruled out by the incongruence of phylogenetic trees^{31,46}. Moreover, some *Rickettsia* spp. were found in association with various species of acarians and/or insects, suggesting that a strict co-speciation process is unlikely². Episodic host changes between acarians, insects and/or vertebrates may have occurred by blood feeding or directly from infectious faeces of arthropods during the diversification of *Rickettsia* spp.^{2,33,42,47}. Therefore, the large proportion of *Rickettsia* strains (*i.e.*, 55%) detected in ticks does not necessarily indicate that they are more abundant in ticks than in other arthropods but may be related to the fact that ticks are more sampled than insects for medical but not environmental purposes.

Alternatively, a rickettsial ancestor might have originally been associated with other eukaryotic hosts, as suggested by recent studies that identified two novel rickettsial clades within the genus *Rickettsia*: the torix clade associated with amoeba, leeches, amphipods and arthropods, and the hydra clade associated with protists and unidentified environmental hosts^{2,4–6}. Both clades are ancestral outgroups to our five phylogroups, suggesting other possible host changes over time².

Thus, all these data indicate that *Rickettsia*-host associations, diversity and ecology are complex and may reflect complex adaptations and trajectories. While arthropods are the most diverse animal phylum living in diverse ecological ecosystems^{48,49}, using a limited diversity of both partners may seriously influence and undermine knowledge of the evolutionary history of *Rickettsia* and their hosts. Further studies on the evolution, diversity and ecology of *Rickettsia* and relatives, together with their partners, would improve our understanding not only of *Rickettsia*-host relationships but also of *Rickettsia* phylogeny and classification.

Regarding plasmids, which are mobile genetic elements harboured by prokaryotic cells, they replicate independently of the chromosome and act as a major driving force in prokaryotic diversity, evolution and fitness^{50,51}. In this study, we identified 41 plasmids in members of the *Rickettsia* genus. The presence of these plasmids suggests two possible evolutionary scenarios regarding their origin. First, the ancestor of rickettsiae harboured one plasmid system (or more), that was(were) inherited and vertically transmitted to its progeny. Indeed, the 41 studied plasmids are present in three rickettsial phylogroups (SFGI, SFGII, and BG). In addition to these SFG species, *R. asiatica* and *R. akari* also harbour plasmids^{52–54}, respectively. Moreover, although, it has been reported that there were no instances of plasmid-associated genes within TG rickettsial genomes²², our pangenomic study showed that the plasmidless TG *R. prowazekii* species (11 strains) harboured a gene cluster integrated in its chromosomes that best matched rickettsial plasmidic genes. Similarly, the SFG species *R. raoultii* (two strains, one with three plasmids and the other being plasmidless) exhibited a gene cluster that best matched a *R. peacockii* plasmid in its chromosomes. These results indicated that the TG *R. prowazekii* and the SFG *R. raoultii* species likely had one and a fourth plasmid, respectively, that were disrupted and integrated in their chromosomes during rickettsial evolution²⁸. Overall, we bring evidence that rickettsial plasmids were present in four of the five rickettsial phylogroups as either individual replicons or plasmid genes integrated in the chromosomes. In addition, our supertree gathered plasmids into a single phylogenetic clade which is distinct from that of the chromosomes, suggesting that rickettsial plasmids originated from a common plasmid ancestor. Altogether, these results on 41 plasmids confirmed our previous work performed on 20 plasmids²⁸. A potential ancestral plasmid was also suggested for the genus *Chlamydia*, another strictly intracellular pathogen of humans and animals⁵⁵.

The alternative hypothesis that rickettsial plasmids may have been acquired laterally from other organisms^{17,19–21,54} does not appear to be relevant. Indeed, we identified several orthologous gene clusters that were common to most plasmids, and our supertree did not display any evidence of plasmid horizontal transfer, thus supporting only the vertical inheritance hypothesis²⁸.

Rickettsial plasmids were identified in 56% and 33% of studied species and strains, respectively, at a rate of one to four plasmids. In contrast, 44% and 67% of studied species and strains had no plasmids, respectively. This result indicated that, during rickettsial diversification from ancestral lineages and a likely vertical

transmission, plasmids were probably retained, multiplied and/or lost between ancestral and present lineages. The retention, multiplication and loss of plasmids may be due to evolutionary adaptations of rickettsiae to either single or several eukaryotic hosts during their obligate intracellular lifestyle. The rickettsial adaptation by the persistence and multiplication of plasmids (\geq one replicon) in some strains suggests that most plasmids may harbour functionally indispensable genes (either laterally acquired and/or duplicated, see “Discussion” below) conferring upon them beneficial traits that could contribute to their survival and/or fitness in arthropod and/or vertebrate hosts. Moreover, the retention, multiplication and potential key roles of plasmids in several *Rickettsia* spp. make this genus remarkable in comparison with several closely related genera in the order Rickettsiales (families Rickettsiaceae [genera *Orientia*, *Occidentia*, *Phycorickettsia*], Anaplasmataceae [*Anaplasma*, *Ehrlichia*, *Neorickettsia*] and *Candidatus* Midichloriaceae [*Candidatus* Midichloria, *Candidatus* Aquarickettsia]), which are devoid of plasmids^{21,56–59}. Moreover, only a single putative plasmid was identified in the intracellular genus *Wolbachia*, another member of the family Anaplasmataceae⁶⁰. This plasmid is smaller (9.2 Kb) than rickettsial plasmids (> 12.3 kb) and does not match their most common gene, the DnaA-like gene. This suggests a possible distinct origin from that of rickettsial plasmids.

On the other hand, the rickettsial adaptation by the loss of plasmids in some strains may result from the fact that plasmids became non-essential and represented a biological cost for these bacteria. This adaptive loss of plasmids may be total and due to the reductive evolutionary process, which may confer to rickettsiae another strategy of survival and fitness inside their hosts. The adaptive loss of plasmids may also be partial by maintaining, after a reductive disruption, some gene clusters via lateral gene incorporation into the chromosomes, as shown for example, in the plasmidless *R. prowazekii*, and in the *R. raoultii* species²⁸, as well as in a *Coxiella burnetii* strain⁶¹. The loss of plasmids may also be accidental or result from numerous passages of the bacteria in selective cell culture, as observed for *R. felis*⁶². Some species, such as *R. parkeri*, *R. typhi* and *R. canadensis*, which are represented by a small number of strains (\leq four), all plasmidless, require exhaustive population genomic investigation to confirm their status, as has been carried out for *R. japonica*⁶³ and *Orientia tustugamushi*, the closest relative genus to *Rickettsia*⁵⁶. Therefore, more genomic data are needed to improve our knowledge of the diversity and distribution of plasmids in *Rickettsia* spp.

Genome reduction (or shrinkage) and expansion are other evolutionary processes that shape bacterial evolution, adaptation and diversity, leading to either their simplification or complexification^{11,19,64}. Genome reduction is known to have occurred during the transition of bacteria from a free-living to obligate intracellular lifestyle^{11,64}, and can result from both degradation or pseudogenisation and the loss of non-essential genes (either neutral ratchet or adaptive) under relaxed selective pressure and genetic drift^{65–70}. However, genome expansion may take place by gene duplication and amplification (GDA forming paralogues) and horizontal gene transfer (HGT forming xenologues)^{71,72}. Several studies showed that the genus *Rickettsia* undergoes reductive evolution^{10–14,19,20,28,29,31,34}, but genome degradation and expansions remained less investigated in both the chromosomes and plasmids.

This study revealed consistent variations in the genome degradation and expansion between and within the five rickettsial phylogroups. Indeed, the SFG rickettsiae, followed by the BG rickettsiae, displayed the largest genomes in the genus and a high and differential degree of degradation and/or expansion by duplication in both the chromosomes and plasmids and/or an increase in plasmid numbers. It has been suggested that most gene expansions arrive via HGT and not via duplication⁷¹. The genome expansion identified in the SFG and BG species originated more from gene duplications and proliferations of both orthologous and/or xenologous genes. In contrast, the TG and CG rickettsiae have the smallest genomes in the genus and exhibited the lowest degree of genome degradation and expansions. These rickettsiae have the most reduced genomes in the genus^{10–14,19,20}, and the loss of plasmids in these phylogroups may have contributed to this evolutionary process. Overall, these results show that the genus *Rickettsia* clearly fits with a general biphasic model of genomic evolution⁶⁸ including genomic reduction by pseudogenisation and loss of genes and/or plasmids^{13,14,19,20,28,29,31,34}, as well as genomic expansion by gene amplification^{73,74} and/or plasmid multiplication²⁸. The SFG and BG species promote expansion rather than contraction, whereas the TG and CG species undergo a reduction rather than amplification. This suggests that members of the genus *Rickettsia* may have followed two distinct adaptive trajectories during its evolutionary history. Similar genomic evolutions were also described in several intracellular bacteria that underwent a genomic expansion (e.g., *Orientia tsutsugamushi*^{56,58}, *Wolbachia* strain wFol⁷⁵, *Burkholderia malleri*⁶⁷ and *Cardinium* strain⁷⁶) and/or a genomic contraction (e.g., *Candidatus* Fokinia solitaria⁷⁷, *Buchnera aphidicola*, *Mycobacterium leprae*^{67,78}, and *Chlamydia* spp.⁷⁹).

SFG and BG rickettsiae as well as several closer or distant bacteria seemingly defied the reductive force which is characteristic of several symbiotic and/or pathogenic bacteria living in bottleneck intracellular niches. Moreover, SFG rickettsiae showed the highest core genome conservation in the genus as observed in the phylogenomic tree. All these results suggest that SFG and BG rickettsiae may be less dependent upon host cell resources for their survival and be less pathogenic than TG and CG rickettsiae. It may also explain why virulence in SFG rickettsiae may be multifactorial⁸⁰. However, although the increase in virulence was associated with strong reductive evolution in TG rickettsiae²⁹, the high level of core gene mutations observed in this phylogroup suggests that adaptive substitutions may also contribute to the pathogenicity and deadly diseases caused by these bacteria²⁵, which requires further investigation.

Although they are intracellular bacteria exhibiting small genomes (from about 1.1 to 2.3 Mb), *Rickettsia* spp. exhibited a comprehensively open pan-chromosome and a completely open pan-plasmidome. Moreover, both *Rickettsia* “pans” shared genes with each other (25.3% and 39%, respectively) and/or α -proteobacteria (34.6% and 15%, respectively) as well as with distantly related and unrelated lineages largely belonging to $\gamma/\beta/\delta$ -proteobacteria, cytophagia, actinobacteria cyanobacteria, chlamydia, viruses and/or unknown taxa. These results suggest that rickettsial genomes harbour an important genomic diversity, and like several free-living prokaryotes⁸¹, *Rickettsia* spp. exhibited several genes which may have been laterally acquired (and duplicated

or not) between their chromosomes, plasmids, some close α -proteobacteria relatives and/or distant lineages. Moreover, *Rickettsia* spp. contained an important proportion of genes (25–28%) that have no significant non-*Rickettsia* homologue in databases, suggesting the occurrence of more potential xenologues, whose origins remain to be discovered. More computational and experimental investigations are needed to determine the functions of critical genes for rickettsial survival and/or pathogenicity.

In addition, it is likely that arthropod-associated *Rickettsia* spp. have exchanged genetic material with microbiome members of these arthropods or other reservoirs in order to acquire genetic advantages and/or to amplify existing or new functions profitable for their survival, competition and/or fitness. Indeed, several lineages that shared homologous genes with *Rickettsia* spp. were previously identified among microbial communities in these ecological niches^{82–87}. Among these taxa, *Wolbachia* (Rickettsiales), *Candidatus Odysella* (Holosporales), *Maritalea* (Rhizobiales), *Francisella* spp., *Legionella massiliensis*, *Pseudomonas*, *Candidatus Hamiltonella* and/or *Berkiella* (γ -proteobacteria), *Chromobacterium* spp. (β -proteobacteria), *Lawsonia intracellularis* (δ -proteobacteria), *Cardinium* spp. (Cytophagia), *Spirulina major* (Cyanobacteria), *Candidatus Protochlamydia amoebophila* (Chlamydia) and viruses (Caudovirales) can be cited as examples. Moreover, the mobile and integrative genetic elements found in rickettsial genomes, including 41 plasmids, 1036 transposase and integrase genes (23% of the pan-chromosome and 23% of the pan-plasmidome) and 178 conjugative *tra* genes (3.8% of the pan-chromosome and 5% of the pan-plasmidome)^{19,36,58,73,74}, may have greatly orchestrated the strong cross-exchanges (or crosstalk), particularly for the SFG and BG phylogroups, and consequently the adaptation and fitness of *Rickettsia* spp. Thus, understanding the structure and dynamics of microbial communities in arthropods through metagenomic studies together with the mobile genetic elements will provide new insights into the dynamics of gene exchanges and interactions between *Rickettsia* and these micro-organisms.

Conclusion

During their diversification and adaptation to eukaryotic hosts (mainly arthropods), rickettsial genomes may have been shaped by diverse evolutionary processes. This study is the first large comparative evolutionary genomic investigation of members of the genus *Rickettsia* carried out on fully re-annotated chromosomes (80) and plasmids (41) of 34 identified species. The study shows that rickettsial phylogroups, as well as species or strains, may follow common and/or specific adaptive trajectories for their survival inside arthropods and/or probably their pathogenicity for vertebrate hosts. *Rickettsia* evolved through genomic degradation and reduction, and/or genomic expansion punctuated by probably short, explosive and innovative episodes of complexification. This trade-off evolution in the genus *Rickettsia* corroborates the general model of contraction/expansion of organisms proposed by Wolf and Koonin⁶⁸ and may be due to their long-term intracellular lifestyle and partnerships as well as to the probable switching of arthropod hosts and interaction with their microbiomes.

Using recently updated public databases, our current results provide a robust insight into the evolutionary relationships between the genus *Rickettsia* and diverse distant lineages, thus improving the results of similar previous studies performed on some rickettsial species such as *R. prowazekii*¹³, *R. bellii*¹⁸, *R. felis*^{17,22,31,33,88} and *R. massiliae*⁸⁹, or groups of *Rickettsia* species^{19,90}. Deeper investigations on population genomics, novel *Rickettsia* species and metagenomics of arthropods are needed to improve our knowledge of their diversity, ecology and evolution.

Our work has shed light on the complex genomic history of the obligate intracellular *Rickettsia* spp. It represents a comprehensive overview into post-genomic and experimental studies such as RNAseq and small regulatory RNAs^{91–94}, proteomics^{27,95–98}, genetic transformation⁹ and gene functions^{16,99–101} aiming at better understanding the molecular mechanisms of critical genes driven by evolutionary processes, which have influenced the adaptation, survival and virulence of *Rickettsia* spp.

Material and methods

Genome sequences and reannotation. A total of 34 *Rickettsia* species depicted by 80 strains were investigated in this study. Studied strains were isolated from diverse arthropod hosts, animals or humans (clinical patients) around the world (Table S1). Their publicly available genomic sequences were downloaded from the ftp server (<ftp://ftp.ncbi.nih.gov/Genome/>) of the National Center for Biotechnology Information (NCBI). Of all studied genomes, 61 were complete and 19 were draft sequences. To avoid potential biases across the originally data deposited in NCBI that were obtained by different gene prediction methods and annotation pipelines, we fully re-analysed the raw genomic sequences. First, we distinguished between unidentified contigs of plasmids and chromosomes. This was performed by aligning each sequenced draft genome against the plasmid contigs of the complete genomes using the MAUVE software^{102,103}. Second, all separated chromosomes and plasmid contigs (27 previously and 14 newly identified plasmids) were subjected to a standard CDS (CoDing Sequence > = 40aa) prediction with the AMIgene software¹⁰⁴. Manual curation was then performed to reject false predicted CDSs. Automated functional re-annotation was performed against the RickBase¹⁹ and/or non-redundant protein databases using pipRick (an in-house annotation pipeline written in Perl language) according to the MicroScope Genome Annotation platform criteria¹⁰⁵ and using the BLASTp algorithm¹⁰⁶. Thus, two standardised databases from *Rickettsia* genomes and plasmids were constructed for further investigation. The prediction of the presence or absence of prophages in *Rickettsia* chromosomes was investigated using PHASTER (Phage Search Tool Enhanced release)¹⁰⁷.

Pangenome inference. To construct the *Rickettsia* pangenome, we separately inferred the pan-chromosome and pan-plasmidome in the 34 *Rickettsia* species (80 strains). For each “pan”, we first performed an all-against-all protein search using both OrthoFinder¹⁰⁸ (with default parameters) and Proteinortho¹⁰⁹ (Cutoffs: E -value = 10^{-6} and coverage $\geq 60\%$) software programmes. We then extracted consensus pan-chromosome and

pan-plasmidome, each containing clusters of orthologues of chromosomal and plasmidic rickettsial genes (that we refer to here as cRiGs and pRiGs, respectively) but not ORFs or CDSs as follows:

- (i) for both “pans”, we checked and uncollapsed each orthogroup (from OrthoFinder) with the help of cluster(s) of orthologous/paralogous genes (from Proteinortho) by performing visual/manual/scrutiny inspection and comparison of their accession numbers and/or protein sequence similarities and identifying clear cases of segregating paralogues which may be co-orthologues and not in-paralogues within a single orthogroup³⁰. This strategy greatly helps the automatic and robust inference of pangenomic statistics of clusters of orthologous genes.
- (ii) retrieve pseudogenes defined as either split genes or gene fragments (i.e., genes interrupted by frameshifts or internal stop codons with at least one CDS having a match coverage < 80%) using a reference protein having the longest aa sequences within each cRiG or pRiG, BLAST best matches and annotation^{18,19,27,28}. Indeed, in highly degraded genomes, such as *Rickettsia* genomes, the recognition of orthology/paralogy for pseudogenes remains a difficult task for pangenomic prediction programmes^{19,27}. They can mistakenly place orthologous/paralogous genes into distinct or single orthology groups, consequently leading to an overestimation of the statistics of orthologous/paralogous clusters and the gene contents of taxa in a pan-genome study.
- (iii) discard some clusters displaying short (< 60 aa) and/or chimeric (< 100 aa) genes having no match or no significant hits with the GenBank database¹⁹.

Last, we estimated the pan-chromosome and pan-plasmidome sizes using the distance method implemented in the PanGP software¹¹⁰.

Network analysis. To delineate networks of evolutionary relationships between the *Rickettsia* pan-chromosome or *Rickettsia* pan-plasmidome and microbial and non-microbial organisms, we first searched homologies between the pan-chromosome (4073 reference protein sequences of cRiGs) and the non-redundant protein database (excluding *Rickettsia* species) using BLASTp (E -value = 10^{-6} , coverage > = 60 and identity > = 30%). The same approach was then performed for the pan-plasmidome using the same criteria. All best matches were visually examined. After that, the pan-chromosome was searched for homologies against a customised database (i.e., a database constructed from the best NR and plasmid homologues packaged into single protein FASTA files) using the same BLASTp criteria to identify the first and/or the second best homologues that could be either plasmids and/or other microbial or non-microbial genes. Finally, results of the last homology searches were summarised into evolutionary networks between the pan-chromosome, pan-plasmidome and microbial as well as non-microbial protein sequences using the CYTOSCAPE software¹¹¹. A similar strategy was used for the construction of the evolutionary network of the pan-plasmidome (457 reference protein sequences of each pRiG) as compared to the pan-chromosome, as well as microbial and non-microbial protein sequences (excluding *Rickettsia* species) from the non-redundant protein databases using the same thresholds.

Phylogenomic and supertree analysis. Multiple sequence alignment of the core chromosome (152,764 aligned sites of 461 cRiGs) was carried out using the MAFFT software¹¹². The phylogenomic tree was obtained using the Maximum Likelihood method and the $-m$ PROTGAME parameter using raxmlHPC-PTHREADS, and then plotted with the iTOL¹¹³ and MEGA¹¹⁴ software programmes. Node robustness was estimated using Bootstrap (BP) analysis of 300 replicates.

To perform plasmid phylogenetic analysis and investigate their evolutionary origins, we first selected 5 genes common to several plasmids and identified using the pan-plasmidome analysis. Overall, we searched for complete genes that are present in plasmids. However, we removed genes that are degraded, as *Rickettsia* species evolved by reductive evolution through progressive gene degradation. Degraded genes or short gene fragments cannot be used in the construction of phylogenies as they can blur the signals. The list of the 5 plasmidic genes used in our analysis includes dnaA-like replication initiator protein, patatin-like phospholipase, helix-turn-helix DNA-binding domain, cell surface antigen Sca12 and heat shock protein. Then, we searched for the best homologous genes of these five genes using the Blast tool against our *Rickettsia* pan-chromosome gene and GenBank databases, as previously described²⁸. Sequences of each of the 5 genes were subjected to multiple sequence alignment, and phylogenetic analyses using Neighbour-joining (NJ) or Maximum Likelihood (ML) methods. NJ and ML trees were performed using the JTT amino acid substitution matrix and the WAG model plus the Nearest-Neighbour-Interchange (NNI) under gamma (Γ) distribution, respectively. We then constructed a supertree from Neighbour-joining (NJ) or Maximum Likelihood (ML) newick trees of five selected genes using the Robinson-Foulds algorithm¹¹⁵. Supertree methods enable synthesizing collections of phylogenetic trees with incomplete taxon overlap into comprehensive trees, or supertrees, that combine the information contained in a set of input trees and include all taxa found in the input trees^{115,116}.

Received: 11 August 2021; Accepted: 16 February 2022

Published online: 09 March 2022

References

1. Perlman, S. J., Hunter, M. S. & Zchori-Fein, E. The emerging diversity of *Rickettsia*. *Proc. Biol. Sci.* **273**, 2097–2210 (2006).
2. Weinert, L. A., Werren, J. H., Aebi, A., Stone, G. N. & Jiggins, F. M. Evolution and diversity of *Rickettsia* bacteria. *BMC Biol.* **7**, 6 (2009).

3. Parola, P. *et al.* Update on tick-borne rickettsioses around the world: A geographic approach. *Clin. Microbiol. Rev.* **26**, 657–702 (2013).
4. Galindo, L. J. *et al.* Combined cultivation and single-cell approaches to the phylogenomics of nuclearioid amoebae, close relatives of fungi. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* <https://doi.org/10.1098/rstb.2019.0094> (2019).
5. Pilgrim, J. *et al.* Torix *Rickettsia* are widespread in arthropods and reflect a neglected symbiosis. *Authorea* <https://doi.org/10.22541/au.159534851.19125003> (2020).
6. Park, E. & Poulin, R. Widespread Torix *Rickettsia* in New Zealand amphipods and the use of blocking primers to rescue host COI sequences. *Sci. Rep.* **10**, 16842 (2020).
7. Raoult, D. & Roux, V. Rickettsioses as paradigms of new or emerging infectious diseases. *Clin. Microbiol. Rev.* **10**, 694–719 (1997).
8. Renesto, P., Ogata, H., Audic, S., Claverie, J. M. & Raoult, D. Some lessons from *Rickettsia* genomics. *FEMS Microbiol. Rev.* **29**, 99–117 (2005).
9. Burkhardt, N. Y. *et al.* Development of shuttle vectors for transformation of diverse *Rickettsia* species. *PLoS ONE* **6**, e29511 (2011).
10. Andersson, S. G., Eriksson, A. S., Naslund, A. K., Andersen, M. S. & Kurland, C. G. The *Rickettsia prowazekii* genome: A random sequence analysis. *Microb. Comp. Genom.* **1**, 293–315 (1996).
11. Andersson, S. G. E. & Kurland, C. G. Reductive evolution of resident genomes. *Trends Microbiol.* **6**, 263–268 (1998).
12. Andersson, S. G. E. *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140 (1998).
13. Andersson, J. O. & Andersson, S. G. E. Insights into the evolutionary process of genome degradation. *Curr. Opin. Genet. Dev.* **9**, 664–671 (1999).
14. Ogata, H. *et al.* Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. *Science* **293**, 2093–2098 (2001).
15. McLeod, M. P. *et al.* Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other rickettsiae. *J. Bacteriol.* **186**, 5842–5855 (2004).
16. Gouin, E. *et al.* The RickA protein of *Rickettsia conorii* activates the Arp2/3 complex. *Nature* **427**, 457–461 (2004).
17. Ogata, H. *et al.* The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol.* **3**, 248 (2005).
18. Ogata, H. *et al.* Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.* **2**, 76 (2006).
19. Blanc, G. *et al.* Reductive genome evolution from the mother of *Rickettsia*. *PLoS Genet.* **3**, 14 (2007).
20. Darby, A. C., Cho, N. H., Fuxelius, H. H., Westberg, J. & Andersson, S. G. Intracellular pathogens go extreme: Genome evolution in the Rickettsiales. *Trends Genet.* **23**, 511–520 (2007).
21. Gillespie, J. J. *et al.* Plasmids and rickettsial evolution: Insight from *Rickettsia felis*. *PLoS ONE* **2**, 266 (2007).
22. Gillespie, J. J. *et al.* *Rickettsia* phylogenomics: Unwinding the intricacies of obligate intracellular life. *PLoS ONE* **3**, e2018 (2008).
23. Ellison, D. W. *et al.* Genomic comparison of virulent *Rickettsia rickettsii* Sheila Smith and avirulent *Rickettsia rickettsii* Iowa. *Infect. Immun.* **76**, 542–550 (2008).
24. Felsheim, R. F., Kurtti, T. J. & Munderloh, U. G. Genome sequence of the endosymbiont *Rickettsia peacockii* and comparison with virulent *Rickettsia rickettsii*: identification of virulence factors. *PLoS ONE* **4**, e8361 (2009).
25. Bechah, Y. *et al.* Genomic, proteomic, and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities. *Genome Res.* **20**, 655–663 (2010).
26. Gillespie, J. J. *et al.* Secretome of obligate intracellular *Rickettsia*. *FEMS Microbiol. Rev.* **39**, 80 (2015).
27. El Karkouri, K. *et al.* Multi-omics analysis sheds light on the evolution and the intracellular lifestyle strategies of spotted fever group *Rickettsia* spp. *Front. Microbiol.* **8**, 1363. <https://doi.org/10.3389/fmicb.2017.01363> (2017).
28. El Karkouri, K., Pontarotti, P., Raoult, D. & Fournier, P. E. Origin and evolution of rickettsial plasmids. *PLoS ONE* **11**, e0147492. <https://doi.org/10.1371/journal.pone.0147492> (2016).
29. Fournier, P. E. *et al.* Analysis of the *Rickettsia africae* genome reveals that virulence acquisition in *Rickettsia* species may be explained by genome reduction. *BMC Genom.* **10**, 166 (2009).
30. Murray, G. G., Weinert, L. A., Rhule, E. L. & Welch, J. J. The phylogeny of *rickettsia* using different evolutionary signatures: How tree-like is bacterial evolution?. *Syst. Biol.* **65**, 265–279 (2016).
31. Merhej, V. & Raoult, D. Rickettsial evolution in the light of comparative genomics. *Biol. Rev. Camb. Philos. Soc.* **86**, 379–405 (2011).
32. Vitorino, L., Chelo, I. M., Bacellar, F. & Zé-Zé, L. Rickettsiae phylogeny: A multigenic approach. *Microbiol. Read.* **153**, 160–168 (2007).
33. Gillespie, J. J. *et al.* Genomic diversification in strains of *Rickettsia felis* Isolated from different arthropods. *Genome Biol. Evol.* **7**, 35–56 (2014).
34. Merhej, V., Angelakis, E., Socolovschi, C. & Raoult, D. Genotyping, evolution and epidemiological findings of *Rickettsia* species. *Infect. Genet. Evol.* **25**, 122–137 (2014).
35. Stothard, D. R., Clark, J. B. & Fuerst, P. A. Ancestral divergence of *Rickettsia bellii* from the spotted fever and typhus groups of *Rickettsia* and antiquity of the genus *Rickettsia*. *Int. J. Syst. Bacteriol.* **44**, 798–804 (1994).
36. Weinert, L. A., Welch, J. J. & Jiggins, F. M. Conjugation genes are common throughout the genus *Rickettsia* and are transmitted horizontally. *Proc. Biol. Sci.* **276**, 3619–3627 (2009).
37. Guo, W., Tian, J. & Lin, X. Extensive genetic diversity of Rickettsiales bacteria in multiple mosquito species. *Sci. Rep.* **6**, 38770 (2016).
38. Liu, D. *et al.* First report of *Rickettsia raoultii* and *R. slovacica* in *Melophagus ovinus*, the sheep ked. *Parasit. Vect.* **9**, 600 (2016).
39. Radzijeuskaja, J. *et al.* Prevalence and diversity of *Rickettsia* species in ectoparasites collected from small rodents in Lithuania. *Parasit. Vect.* **11**, 375. <https://doi.org/10.1186/s13071-018-2947-9> (2018).
40. Zhang, J. *et al.* Molecular detection of *Rickettsia felis* and *Rickettsia bellii* in mosquitoes. *Vect. Borne Zoon. Dis.* <https://doi.org/10.1089/vbz.2019.2456> (2019).
41. Oliveira, K. A. *et al.* Molecular identification of *Rickettsia felis* in ticks and fleas from an endemic area for Brazilian spotted fever. *Mem. Inst. Oswaldo Cruz* **103**, 191–194 (2008).
42. Choi, Y. J. *et al.* Molecular detection of various rickettsiae in mites (acari: trombiculidae) in southern Jeolla Province, Korea. *Microbiol. Immunol.* **51**, 307–312 (2007).
43. Brown, L. D. & Macaluso, K. R. *Rickettsia felis*, an emerging flea-borne rickettsiosis. *Curr. Trop. Med. Rep.* **3**, 27–39 (2016).
44. Medina-Sanchez, A. *et al.* Detection of a typhus group *Rickettsia* in *Amblyomma* ticks in the state of Nuevo Leon, Mexico. *Ann. N. Y. Acad. Sci.* **1063**, 327–332 (2005).
45. Sprong, H. *et al.* *Ixodes ricinus* ticks are reservoir hosts for *Rickettsia helvetica* and potentially carry flea-borne *Rickettsia* species. *Parasit. Vect.* **2**, 41 (2009).
46. Modeo, L. *et al.* Candidatus *Trichorickettsia mobilis*, a Rickettsiales bacterium, can be transiently transferred from the unicellular eukaryote Paramecium to the planarian *Dugesia japonica*. *PeerJ* **8**, e8977. <https://doi.org/10.7717/peerj.8977> (2020).
47. Ng-Nguyen, D. *et al.* Domestic dogs are mammalian reservoirs for the emerging zoonosis flea-borne spotted fever, caused by *Rickettsia felis*. *Sci. Rep.* **10**, 4151 (2020).
48. Edgecombe, G. D. & Legg, D. A. Origins and early evolution of arthropods. *Palaeontology* **57**, 457–468 (2014).

49. Giribet, G. & Edgecombe, G. D. The phylogeny and evolutionary history of arthropods. *Curr. Biol.* **29**, R592–R602 (2019).
50. Tamminen, M., Virta, M., Fani, R. & Fondi, M. Large-scale analysis of plasmid relationships through gene-sharing networks. *Mol. Biol. Evol.* **29**, 1225–1240 (2012).
51. Hülter, N. F., Wein, T., Effe, J., Garoña, A. & Dagan, T. Intracellular competitions reveal determinants of plasmid evolutionary success. *Front. Microbiol.* **11**, 2062. <https://doi.org/10.3389/fmicb.2020.02062> (2020).
52. Thu, M. J. *et al.* Complete genome sequence of *Rickettsia asiatica* strain Maytaro1284, a member of spotted fever group Rickettsiae isolated from an *Ixodes ovatus* tick in Japan. *Microbiol. Resour. Announc.* **8**, 00886–00919 (2019).
53. Baldrige, G. D., Burkhardt, N. Y., Felsheim, R. F., Kurtti, T. J. & Munderloh, U. G. Plasmids of the pRM/pRF family occur in diverse *Rickettsia* species. *Appl. Environ. Microbiol.* **74**, 645–652 (2008).
54. Baldrige, G. D. *et al.* Wide dispersal and possible multiple origins of low-copy-number plasmids in rickettsia species associated with blood-feeding arthropods. *Appl. Environ. Microbiol.* **76**, 1718–1731 (2010).
55. Szabo, K. V., O'Neill, C. E. & Clarke, I. N. Diversity in chlamydial plasmids. *PLoS ONE* **15**, e0233298 (2020).
56. Fleshman, A. *et al.* Comparative pan-genomic analyses of *Orientia tsutsugamushi* reveal an exceptional model of bacterial evolution driving genomic diversity. *Microb. Genom.* **4**, 000199 (2018).
57. Mediannikov, O. *et al.* High quality draft genome sequence and description of *Occidentia massiliensis* gen. nov., sp. Nov., a new member of the family Rickettsiaceae. *Stand. Genomic. Sci.* **9**, 9 (2014).
58. Cho, N. H. *et al.* The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7981–7986 (2007).
59. Klings, J. G. *et al.* Phylogenetic, genomic, and biogeographic characterization of a novel and ubiquitous marine invertebrate-associated Rickettsiales parasite, *Candidatus Aquarickettsia rohweri*, gen. nov., sp. Nov. *ISME J.* **13**, 2938–2953 (2019).
60. Reveillaud, J. *et al.* The *Wolbachia* mobilome in *Culex pipiens* includes a putative plasmid. *Nat. Commun.* **10**, 1051 (2019).
61. Willems, H., Ritter, M., Jäger, C. & Thiele, D. Plasmid-homologous sequences in the chromosome of plasmidless *Coxiella burnetii* Scurry Q217. *J. Bacteriol.* **179**, 3293–3297 (1997).
62. Fournier, P. E. *et al.* Variations of plasmid content in *Rickettsia felis*. *PLoS ONE* **3**, 2289 (2008).
63. Akter, A. *et al.* Extremely low genomic diversity of *Rickettsia japonica* distributed in Japan. *Genome Biol. Evol.* **9**, 124–133 (2017).
64. Bryant, J., Chewapreecha, C. & Bentley, S. D. Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. *Fut. Microbiol.* <https://doi.org/10.2217/fmb.12.108> (2012).
65. Wixon, J. Featured organism: reductive evolution in bacteria: *Buchnera* sp., *Rickettsia prowazekii* and *Mycobacterium leprae*. *Comparat. Funct. Genom.* **2**, 44–48 (2001).
66. Khachane, A. N., Timmis, K. N. & Santos, V. A. Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. *Mol. Biol. Evol.* **24**, 449–456 (2007).
67. Song, H. *et al.* The early stage of bacterial genome-reductive evolution in the host. *PLoS Pathog.* **6**, e1000922 (2010).
68. Wolf, Y. I. & Koonin, E. V. Genome reduction as the dominant mode of evolution. *BioEssays* **35**, 829–837 (2013).
69. Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. & Scheffler, K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832 (2015).
70. Weinert, L. A. & Welch, J. J. Why might bacterial pathogens have small genomes?. *Trends Ecol. Evol.* **32**, 936–947 (2017).
71. Treangen, T. J. & Rocha, E. P. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* **7**, e1001284. <https://doi.org/10.1371/journal.pgen.1001284> (2011).
72. Elliott, K. T., Cuff, L. E. & Neidle, E. L. Copy number change: Evolving views on gene amplification. *Fut. Microbiol.* **8**, 887–899 (2013).
73. Gillespie, J. J. *et al.* A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. *J. Bacteriol.* **194**, 376–394 (2012).
74. Hagen, R., Verhoeve, V. L., Gillespie, J. J. & Driscoll, T. P. Conjugative Transposons and Their Cargo Genes Vary across Natural Populations of *Rickettsia buchneri* Infecting the Tick *Ixodes scapularis*. *Genome Biol. Evol.* **10**, 3218–3229 (2018).
75. Kampfraath, A. A. *et al.* Genome expansion of an obligate parthenogenesis-associated *Wolbachia* poses an exception to the symbiont reduction model. *BMC Genom.* **20**, 106 (2019).
76. Brown, A. M. V. *et al.* Comparative genomics of *Wolbachia-Cardinium* dual endosymbiosis in a plant-parasitic nematode. *Front. Microbiol.* **9**, 2482. <https://doi.org/10.3389/fmicb.2018.02482> (2018).
77. Floriano, A. M. *et al.* The genome sequence of ‘*Candidatus Fokinia solitaria*’: Insights on reductive evolution in Rickettsiales. *Genome Biol. Evol.* **10**, 1120–1126 (2018).
78. Chong, R. A., Park, H. & Moran, N. A. Genome evolution of the obligate endosymbiont *Buchnera aphidicola*. *Mol. Biol. Evol.* **36**, 1481–1489. <https://doi.org/10.1093/molbev/msz082> (2019).
79. Sigalova, O. M. *et al.* *Chlamydia* pan-genomic analysis reveals balance between host adaptation and selective pressure to genome reduction. *BMC Genom.* **20**, 710. <https://doi.org/10.1186/s12864-019-6059-5> (2019).
80. Clark, T. R. *et al.* Comparative genome sequencing of *Rickettsia rickettsii* strains that differ in virulence. *Infect. Immun.* **83**, 1568–1576 (2015).
81. Galperin, M. Y., Kristensen, D. M., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Microbial genome analysis: The COG approach. *Brief Bioinform.* **20**, 1063–1070. <https://doi.org/10.1093/bib/bbx117> (2019).
82. Greay, T. L. *et al.* Recent insights into the tick microbiome gained through next-generation sequencing. *Parasit. Vect.* **11**, 12 (2018).
83. Ravi, A. *et al.* Metagenomic profiling of ticks: Identification of novel rickettsial genomes and detection of tick-borne canine parvovirus. *PLoS Negl. Trop. Dis.* **13**, e0006805 (2019).
84. Zhuang, L. *et al.* Identification of tick-borne pathogen diversity by metagenomic analysis in *Haemaphysalis longicornis* from Xinyang China. *Infect. Dis. Poverty* **7**, 45 (2018).
85. Tokarz, R. *et al.* Microbiome analysis of *Ixodes scapularis* ticks from New York and Connecticut. *Ticks Tick Borne Dis* **10**, 894–900 (2019).
86. Thoendel, M. Targeted metagenomics offers insights into potential tick-borne pathogens. *J. Clin. Microbiol.* **58**, e01893-e1920 (2020).
87. Duan, D. Y., Liu, G. H. & Cheng, T. Y. Microbiome analysis of the saliva and midgut from partially or fully engorged female adult *Dermacentor silvarum* ticks in China. *Exp. Appl. Acarol.* **80**, 543–558 (2020).
88. Merhej, V., Notredame, C., Royer-Carenzi, M., Pontarotti, P. & Raoult, D. The rhizome of life: the sympatric *Rickettsia felis* paradigm demonstrates the random transfer of DNA sequences. *Mol. Biol. Evol.* **28**, 3213–3223. <https://doi.org/10.1093/molbev/msr239> (2011).
89. Blanc, G. *et al.* 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> (2007).
90. Duron, O. Lateral transfers of insertion sequences between *Wolbachia Cardinium* and *Rickettsia* bacterial endosymbionts. *Heredity* **111**, 330–337 (2013).
91. Schroeder, C. L. C. *et al.* Transcriptional profiling of *Rickettsia prowazekii* coding and non-coding transcripts during in vitro host-pathogen and vector-pathogen interactions. *Ticks Tick Borne Dis* **8**, 827–836 (2017).
92. Narra, H. P. *et al.* Small Regulatory RNAs of *Rickettsia conorii*. *Sci. Rep.* **6**, 36728. <https://doi.org/10.1038/srep36728> (2016).

93. Narra, H. P. *et al.* Comparative transcriptomic analysis of *Rickettsia conorii* during in vitro infection of human and tick host cells. *BMC Genom.* **21**, 665. <https://doi.org/10.1186/s12864-020-07077-w> (2020).
94. Mika-Gospodorz, B. *et al.* Dual RNA-seq of *Orientia tsutsugamushi* informs on host-pathogen interactions for this neglected intracellular human pathogen. *Nat. Commun.* **11**, 3363 (2020).
95. Renesto, P. *et al.* *Rickettsia conorii* and *R. prowazekii* proteome analysis by 2DE-MS: a step toward functional analysis of rickettsial genomes. *Ann. N. Y. Acad. Sci.* **1063**, 90–93 (2005).
96. Ogawa, M., Renesto, P., Azza, S., Moinier, D. & Fourquet, P. Proteome analysis of *Rickettsia felis* highlights the expression profile of intracellular bacteria. *Proteomics* **7**, 1232–1248 (2007).
97. Martins, L. A. *et al.* The intracellular bacterium *Rickettsia rickettsii* exerts an inhibitory effect on the apoptosis of tick cells. *Parasit. Vectors.* **13**, 603 (2020).
98. Csicsay, F. *et al.* Proteomic analysis of *Rickettsia akari* proposes a 44 kDa-OMP as a potential biomarker for Rickettsialpox diagnosis. *BMC Microbiol.* **20**, 200. <https://doi.org/10.1186/s12866-020-01877-6> (2020).
99. Rennoll-Bankert, K. E. *et al.* Which way in? The RalF Arf-GEF orchestrates *Rickettsia* host cell invasion. *PLoS Pathog.* **11**, e1005115. <https://doi.org/10.1371/journal.ppat.1005115> (2015).
100. Engström, P. *et al.* Evasion of autophagy mediated by *Rickettsia* surface protein OmpB is critical for virulence. *Nat. Microbiol.* **4**, 2538–2551 (2019).
101. Voss, O. H. *et al.* Risk1, a phosphatidylinositol 3-kinase effector promotes *Rickettsia typhi* intracellular survival. *MBio* **11**, e00820–e920. <https://doi.org/10.1128/mBio.00820-20> (2020).
102. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
103. Rissman, A. I. *et al.* Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* **25**, 2071–2073 (2009).
104. Bocs, S., Cruveiller, S., Vallet, D., Nuel, G. & Medigue, C. AMIGene: Annotation of microbial genes. *Nucleic Acids Res.* **31**, 3723–3726 (2003).
105. Vallet, D. *et al.* MicroScope: a platform for microbial genome annotation and comparative genomics. *Database* **21**, 2 (2009).
106. Altschul, S. F. *et al.* Gapped BLAST and PSIBLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
107. Arndt, D., Marcu, A., Liang, Y. & Wishart, D. S. PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. *Brief Bioinform.* **19**, 1560–1567. <https://doi.org/10.1093/bib/bbx121> (2019).
108. Emmis, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
109. Lechner, M. *et al.* Proteinortho: Detection of (co-)orthologs in large-scale analysis. *BMC Bioinform.* **12**, 124. <https://doi.org/10.1186/1471-2105-12-124> (2011).
110. Zhao, Y. *et al.* PanGP: A tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics* **30**, 1297–1299 (2014).
111. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
112. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* **20**, 1160–1166. <https://doi.org/10.1093/bib/bbx108> (2019).
113. Letunic, I. & Bork, P. Interactive tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
114. Kumar, S., Stecher, G., Li, M., Nnyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
115. Bansal, M. S., Burleigh, J. G., Eulenstein, O. & Fernández-Baca, D. Robinson-foulds supertrees. *Algorithms Mol. Biol.* **5**, 18. <https://doi.org/10.1186/1748-7188-5-18> (2010).
116. Akanni, W. A., Wilkinson, M., Creevey, C. J., Foster, P. G. & Pisani, D. Implementing and testing Bayesian and maximum-likelihood supertree methods in phylogenetics. *R. Soc. open sci.* **2**, 140436–140436. <https://doi.org/10.1098/rsos.140436> (2015).

Acknowledgements

The study was funded by the Fondation Méditerranée Infection.

Author contributions

K.E. and P.F. conceived the project. K.E. performed data analysis, wrote and edited the manuscript. K.E., E.G., P.F. and D.R. revised the manuscript. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-07725-z>.

Correspondence and requests for materials should be addressed to K.E.K. or P.-E.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022