Year : 2023

Rhabdochlamydia: a new tick-borne pathogen?

Marquis Bastian

Marquis Bastian, 2023, Rhabdochlamydia: a new tick-borne pathogen?

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_7451AF7A6D264

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Institut de Microbiologie

Rhabdochlamydia, a new tick-borne pathogen?

Thèse de doctorat en médecine et ès sciences (MD-PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Bastian MARQUIS

Médecin diplômé de la Confédération Helvétique

Jury

Prof. Monika Hegi, Présidente et répondante MD-PhD
Prof. Gilbert Greub, Directeur de thèse
Prof. Pierre Cosson, Expert
Prof. Thierry Roger, Expert
Prof. Frédéric Lamoth, Expert

Lausanne 2023

UNIL | Université de Lausanne
Ecole doctorale

Ecole Doctorale
Doctorat MD-PhD

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Madame	Prof. Monika	**Hegi**
Directeur·trice de thèse	Monsieur	Prof. Gilbert	**Greub**
Répondant·e	Madame	Prof. Monika	**Hegi**
Expert·e·s	Monsieur	Prof. Thierry	**Roger**
	Monsieur	Prof. Frederic	**Lamoth**
	Monsieur	Prof. Pierre	**Cosson**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Bastian MARQUIS

Médecin diplômé de la Confédération Helvétique

intitulée

***Rhabdochlamydia*, a new tick-borne pathogen?**

Lausanne, le 21 avril 2023

pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Monika Hegi

Remerciements

Je dois remercier beaucoup de personnes qui m'ont permis d'arriver au bout de ce projet. Commençons par ceux qui ont été impliqués le plus directement : Séb, qui m'a formé aux techniques de labo lorsque je commençais la thèse en sachant tout juste par quel bout tenir une pipette. Silvia, qui restera dans mes souvenirs comme un mélange de Pai Mei et de Yoda. Merci pour tes relectures, ta disponibilité, tes feedbacks et pour les occasionnels échanges de coups de gueule. Trestan, programmeur de profession mais biologiste de cœur, que ma monomanie a bien dû ennuyer quand les discussions revenaient invariablement à Rhabdo et à zdb. Claire, pour avoir cru au potentiel d'un projet de programmation initialement destiné à me simplifier la vie, pour m'avoir donné l'aide nécessaire à en faire zdb et pour m'avoir accordé ta confiance pour la supervision de travaux de first step et de master en biologie. J'espère avoir été à la hauteur. Merci au pothos du bureau 312 pour son éclatante verdeur et sa croissance rapide contrastant avec celle de cette satanée bactérie.

Hors du labo, mais non moins importants, les membres de famille qui m'ont accompagné tout au long de la thèse : Stéphane, Francine, Quentin, Elodie et tout particulièrement Olivia, pour m'avoir supporté au quotidien et avoir toujours su me rappeler l'existence d'un monde extérieur quand ça devenait nécessaire. Capucine, petite chips dont l'arrivée inattendue en milieu de thèse a significativement creusé ma dette de sommeil (p=0.01, Marquis non-parametric pifomètre test for pseudoimaginary retrospective sleep data), mais qui m'a rappelé qu'une couche propre, des nuits de 15 h et un estomac plein devraient en principe suffire à notre bonheur.

Last but not least, merci à Gilbert de m'avoir permis de réaliser cette thèse, de m'avoir laissé la liberté d'explorer le sujet et pour ton enthousiasme jamais pris en défaut pour mes idées les plus farfelues.

Abstract

The *Chlamydiales* is an order of obligate intracellular bacteria that share the same biphasic development cycle. It includes well-known human pathogens, like *Chlamydia trachomatis*, but also species causing diseases in other animals. This order has been continuously expanded in the past decades and it now appears that chlamydiae can infect a wide variety of hosts and can be found in diverse environments. The *Rhabdochlamydiaceae* is a recent addition to the order. It was initially discovered in cockroaches and woodlouse and could eventually be isolated in a culture system from the latter. After its initial discovery, two additional species were described, in ticks and spiders, respectively. *Rhabdochlamydiaceae* were also detected in respiratory samples of hospitalized patients, raising the question of its pathogenic role in lung infections. In this work, we tested the permissivity of various cell lines and determined that the host range of *R. porcellionis* is limited by its thermal sensitivity. In particular, this bacterium cannot grow at 33 °C or 37 °C and cannot withstand exposures as short as 6 h to 37 °C. Mammalian cells were permissive to *R. porcellionis* only when their incubation temperature was lowered to 28 °C, excluding a role of this bacterium in human infectious diseases. In a separate project, we confirmed the presence of *Rhabdochlamydiaceae* in questing ticks at a prevalence of 1.1%. The low copy numbers observed in this study contrasts with previous reports and suggest that ticks heavily infected with *Rhabdochlamydiaceae* may be less fit and less likely to be feeding on a human host. Finally, we sequenced the genome of 5 new rhabdochlamydiae species from different arthropods orders. Interestingly, the analysis of those genomes revealed that rhabdochlamydiae species likely have wide, overlapping host ranges. Ticks and woodlouse having similar thermal preferences and the genomes of arthropod-borne rhabdochlamydiae being closely related, it is likely that the evolutionary pressure that drove the adaptation of *R. porcellionis* towards lower temperatures was also at play for tick-borne species. Tick-borne *Rhabdochlamydiaceae* are thus unlikely to cause infections in humans.

Résumé

L'ordre des *Chlamydiales* est constitué de bactéries intracellulaires obligatoires et inclut des pathogènes d'importance médicale, tels que *Chlamydia trachomatis*, mais aussi d'autres espèces importantes en médecine vétérinaire. Les dernières décennies ont été témoins d'une importante expansion des *Chlamydiales* et il est aujourd'hui évident que ces bactéries se trouvent dans de nombreux hôtes. La famille des *Rhabdochlamydiaceae* est une addition récente à cet ordre initialement découverte chez les cafards et les cloportes, et dont seulement une espèce, *R. porcellionis*, a pu être cultivée. Deux nouvelles espèces, détectées chez les tiques et les araignées, sont venues agrandir cette famille, alors que sa présence dans des échantillons respiratoires de patients hospitalisés a soulevé la question de sa pathogénicité. Dans ce travail, nous avons étudié la permissivité de plusieurs lignées cellulaires à *R. porcellionis* et avons notamment démontré que cette bactérie ne peut pas se répliquer à 37 °C ou à 33 °C. Elle est de plus incapable de supporter des expositions de plus de 6 h à 37 °C. Les cellules de mammifères sont permissives à *R. porcellionis* seulement si elles sont incubées à 28 °C, excluant un rôle pathogène de cette bactérie. Nous avons également confirmé la présence de *Rhabdochlamydiaceae* dans des tiques prélevées sur des humains, à une prévalence de 1.1%. Toutefois, la charge bactérienne des tiques était moins élevée que dans de précédentes études, suggérant que les tiques très infectées sont moins susceptibles d'être trouvées sur des humains. Nous avons finalement séquencé le génome de cinq nouvelles espèces de *Rhabdochlamydiaceae* présentes dans différents ordres d'arthropodes. L'analyse de ces génomes suggère que ces espèces ont des hôtes communs et une capacité à infecter une large variété d'arthropodes. Les Ixodida partageant les mêmes préférences thermiques que le cloporte et les génomes des *Rhabdochlamydiaceae* étant très similaires, il est probable que la pression evolutive ayant mené à une adaptation de *R. porcellionis* à la température de ses hôtes ait aussi agi sur les autres espèces. Il est donc peu vraisemblable que les *Rhabdochlamydiaceae* de tique aient un rôle pathogène.

Résumé pour le grand public

Les chlamydiae forment un groupe hétérogène de bactéries dont la plus connue est certainement *Chlamydia trachomatis*, une bactérie causant des infections sexuellement transmissibles et des infections oculaires pouvant mener à la cécité. Ce groupe inclut de nombreuses autres espèces ayant plusieurs points communs, notamment la nécessité d'envahir une cellule hôte pour pouvoir se répliquer, à la manière d'un virus. Ces bactéries ont souvent des préférences marquées pour les cellules d'un hôte précis. Telle bactérie préfèrera infecter des mammifères, alors que telle autre préfèrera les amibes. Les rhabdochlamydiae font partie du phylum des *Chlamydiae* et ont été découvertes dans les arthropodes, notamment dans les cloportes et les cafards, mais également dans les tiques. La détection de ces bactéries chez des patients souffrant d'infection respiratoire, ainsi que leur présence dans des tiques ont naturellement amené à la question de leur rôle dans les maladies humaines et du risque de transmission par piqûres de tiques. Pour répondre à cette question, nous avons essayé d'infecter différents types de cellules avec la seule rhabdochlamydia isolée en laboratoire pour déterminer si les humains peuvent lui servir d'hôte. Il est apparu que cette bactérie s'est adaptée à la température corporelle plus froide des arthropodes et a perdu ou n'a jamais acquis la capacité de supporter les températures rencontrées dans le corps humain. Cette bactérie ne risque donc pas d'infecter des mammifères. Une autre partie du projet s'intéressait à la diversité des rhabdochlamydiae. Nous avons notamment découvert plusieurs nouvelles espèces infectant divers arthropodes comme les papillons et les coléoptères. Bien qu'il ne soit pas clair si ces nouvelles espèces sont elles aussi sensibles à la chaleur, leur ressemblance génétique à celle testée en laboratoire et les températures similaires de leur hôte favorisent cette hypothèse. Les rhabdochlamydiae ne seraient donc pas pathogènes pour les humains. Étant donnée leur grande diversité, il est toutefois possible que des rhabdochlamydiae infectant d'autres organismes que les arthropodes puissent poser des risques pour la santé humaine.

Glossary

Aa23T *Aedes albopictus* cell line.

ANI Average nucleotide identity.

CLO *Chlamydia*-like organism.

CNRT National centre for tick-borne diseases.

COG Clusters of orthologous groups.

EB Elementary body.

IFU Inclusion forming unit.

KEGG Kyoto encyclopedia of genes and genomes.

NCBI National center for biotechnology information.

NGS Next-generation sequencing.

PAMP Pathogen-associated molecular pattern.

PCR Polymerase chain reaction.

Pfam Protein family database.

PVC Planctomycetota, Verrucomicrobiota, and Chlamydiota superphylum.

qPCR Quantitive polymerase chain reaction.

RB Reticulate body.

S2 *Drosophila melanogaster* cell line.

Sf9 *Spodoptera frugiperda* ovarian cells.

SRA Sequencing read archive.

T3SS Type 3 secretion system.

Contents

List of Figures

List of Tables

1 Introduction

The *Rhabdochlamydiaceae* family is part of the *Chlamydiales*, an order of obligate intracellular bacteria that all share the same biphasic developmental cycle. This order notably includes the *Chlamydiaceae* family, whose members are found in vertebrates and count pathogens relevant for both human and veterinary medicine, such as *Chlamydia trachomatis* and *Chlamydia abortus*. Far from being limited to vertebrates, members of the *Chlamydiales* have been discovered in a wide variety of hosts and environments and some of them are considered as emerging pathogens. This is the case for the *Rhabdochlamydiaceae* family, whose members were initially discovered in arthropods but appear to be present in soil and freshwater environments and whose presence has been detected in samples taken from patients suffering of respiratory infections. Their presence in ticks moreover raised the concern of the risk of tick-borne transmission. This introduction first gives a general description of the *Chlamydiales* order before reviewing the literature on the *Rhabdochlamydiaceae* family. It finally reviews the evidence for the pathogenicity of chlamydiae outside of the *Chlamydiaceae* family.

1.1 The *Chlamydiales* order

The *Chlamydiales* order is part of the *Planctomycetota-Verrucomicrobiota-Chlamydiae* superphylum and is composed of intracellular obligate bacteria that share the same biphasic development cycle. Perhaps the most well-known member of this order is *Chlamydia trachomatis*, a member of the *Chlamydiaceae* family, the causative agent of trachoma

and of a common sexually-transmitted disease [1]. Other members of the *Chlamydiales* are known to causes diseases in humans and other animals, most of them also part of the *Chlamydiaceae* family. For instance, *Chlamydia pneumoniae* is reported to be the cause of about 10% of community-acquired pneumonia [2], while other species affect wild animals [3], cattle [4] or can cause human zoonotic diseases [5]. Far from being limited to those well-known pathogens, the *Chlamydiales* order has been expanding during the past two decades as new species were discovered, growing from a unique family restricted to vertebrate hosts to what now appears as an ubiquitous order whose members are found in a wide range of different hosts and environments [6, 7]. Indeed, in addition to mammals, *Chlamydiales* were identified in reptiles [8, 9], birds [10, 11], arthropods [12–14], amoebae [15–17], fishes [18] and found in unexpected environments, such as marine sediments [19], although no host could be ascertained in the last case. Metagenomics studies moreover revealed a staggering diversity in this order, in terms of taxonomy, with more than 200 predicted family-level taxa [20], but also in terms of gene content, with some chlamydiae having genes necessary for light-driven ATP synthesis [21] or genes encoding a flagellar apparatus and a chemotactism system coherent with motility [22]. Those organisms are collectively referred to as either *Chlamydia*-like organisms, due to a developmental cycle similar to that observed in members of the *Chlamydia* genus, as *Chlamydia*-related bacteria, since they are biologically and phylogenetically related, or as environmental chlamydiae. In contrast to the disease-causing chlamydiae of the *Chlamydiaceae* family, some CLOs confer survival benefits to their host: *Parachlamydia acanthamoebae* appears to protect amoebae against giant viruses [23] and *Neochlamydia* were similarly showed to prevent co-infection by *Legionella* [24, 25]. Overall, the *Chlamydiales* appear as very diverse bacteria ranging from pathogens causing life-threatening pneumonia to symbionts reminiscent of the protective endosymbionts observed in insects.

In spite of their diversity, all *Chlamydiales* share a common developmental cycle composed of an extracellular and infectious form, the elementary body and a replicative intracellular form, the reticulate body [26]. The infectious cycle of the *Chlamydiales*

spans the attachment and internalization of an extracellular elementary body into an eukaryotic host cell, its differentiation into a reticulate body followed by several rounds of replication, the asynchronous differentiation of its progeny into elementary bodies and their exit from the host cell to start a new cycle, either by cell lysis or by extrusion of the whole inclusion [27, 28]. A third form, the aberrant body, can also be observed when reticulate bodies are exposed to stresses such as iron deprivation [29, 30], antibiotics [30] or IFN-γ [31]. Aberrant bodies do not replicate but are able to revert to reticulate bodies and to resume the normal cycle when the stress disappears. This stage has consequently been hypothesized to be a persistence state that could explain the failure of antibiotic treatment of chlamydial infections in some patients [32]. The exact mechanisms allowing the transition between the developmental stages are still unknown and is an area of active research.

There are several key differences between EBs and RBs. EBs (0.1 µm to 0.5 µm) are smaller than RBs (1 µm) and exhibit more diversity in terms of shape compared to the spherical RBs [28]. Indeed, in addition to the coccoid elementary bodies of the *Chlamydiaceae* family, crescent-shaped [33], star-shaped [34, 35] or rod-shaped [13, 14] EBs have also been described. The outer membrane of EBs is enriched in disulfide bridges linking cystein-rich membrane proteins together in a structure termed the outer-membrane complex [27], which have been hypothesized to palliate the absence of peptidoglycan in the envelope of the *Chlamydiales* [36]. Those disulfide bridges are subsequently reduced during the transition from EB to RB [27]. Interestingly, no homolog of those cystein-rich proteins could be found in the genomes of the *Simkaniaceae* [37] and the *Rhabdochlamydiaceae* [38], two families reported to have distinctive rod-shaped EBs [13, 14, 39]. EBs and RBs also differ in terms of metabolism. EBs have historically been likened to spores and were thought to be metabolically inert, but this view has been challenged by recent studies that demonstrated the ability of EBs to metabolize several carbon sources [40–43] and the necessity of an active metabolism for the maintenance of infectivity [40, 41]. In particular, EBs appear to be enriched in proteins involved in central carbon and glu-

cose metabolism [44], which is coherent with the observation that chlamydiae accumulate glycogen during their cycle [45, 46]. The stored glycogen would then serve as an energy source in the extracellular environment, after EBs left their host cell [28]. In contrast, RBs appear as energy parasites scavenging ATP from their host cell to fuel the replication process [28]. Finally, EBs and RBs differ by their DNA compaction level. Histone-like proteins are indeed produced in the later stages of the infectious cycle and induce the condensation of DNA in the future elementary bodies [27].

Once internalized in their host cell, chlamydiae replicate in a host-derived vacuole called the inclusion. The inclusion has been hypothesized to serve as insulation between the bacteria and the innate immune system and to offer an ideal environment for the replication of the bacteria [47]. This however has the downside of isolating the chlamydiae from the nutrient-rich cytoplasm. Chlamydiae indeed have limited biosynthetic and metabolic capacities and are notably auxotroph for most amino-acid and nucleotides [28, 48]. They must therefore import those molecules from their host cell, through the inclusion. The inclusion membrane has however been demonstrated to prevent the diffusion of molecules larger than 520 Da [49, 50], although the possibility of diffusion of smaller metabolites such as sugars or amino-acids into the inclusion has never been investigated [28]. Chlamydiae solve this conundrum with transporters inserted into the inclusion membrane [46, 51] and by hijacking the intracellular traffic of their host cell to acquire the content of the rerouted vesicles [47]. Translocating bacterial transporters to the inclusion and interacting with the host cell are both made possible by the secretion of effectors via several secretion systems. Those systems, particularly the type III secretion system, are well conserved in the *Chlamydiales* [21, 37] and appear to have been acquired early in chlamydial evolution [52]. Indeed, the T3SS appear to be essential to the chlamydial life cyle as its effectors mediate the internalization of the bacteria [53], modulates the host cell immunity [54] and secretes Incs, a family of inclusion membrane proteins that interact with the host cell's intracellular trafficking machinery [55]. Interestingly, while the T3SS is part of the chlamydial core genome, its effectors are much less conserved

and may be a determinant of the host tropism [37]. Chlamydiae thus appear as master manipulators of eukaryotic cells that evolved intricate strategies to acquire the nutrients necessary for their replication, all the while preventing apoptosis of their host cell and mediating the immune response of their host.

1.2 The *Rhabdochlamydiaceae* family

The first members of the *Rhabdochlamydiaceae* family were discovered independently in a laboratory colony of cockroaches and in wild-caught woodlice when sick animals were investigated. Animals either displayed abdominal swelling [56] or white spots in the hepatopancreas [57], while the microscope examination of infected tissues revealed the presence of different forms of intracellular bacteria grouped together in vacuoles. Spherical electron-lucent bacteria occasionally shown to undergo binary fission were indeed observed in the same vacuoles as rod-shaped electron-dense bacteria (Fig. 1.1C). Based on the tissue distribution of bacteria in organisms at different stages of the disease, the infection appeared to start in the epithelium of the digestive tract and to spread to other tissues via the hemolymph. Vacuoles filled with bacteria were indeed observed in the adipose tissue, gonads and hemocytes, although it is not clear if the bacteria are able to complete their developmental cycle in those tissues, as only one of the developmental form could be observed outside of the digestive tract [58]. The affiliation of those bacteria to the *Chlamydiales* order was suspected on the basis of their biphasic developmental cycle and confirmed by the similarity of their 16S rDNA sequences with other members of the order [13, 14]. Both sequences were however distant from those of existing members and clustered together in a separate clade in phylogenetic trees based on the 16S rDNA gene, justifying their inclusion in a new family, the *Rhabdochlamydiaceae*. The name *Rhabdochlamydia* was chosen (the prefix *Rhabdo-* means "rod" in ancient Greek) because of the shape of the elementary bodies observed in electron microscopy (Fig. 1.1A, B and D). The two bacteria were assigned to different species: *Rhabdochlamydia crassificans*, from

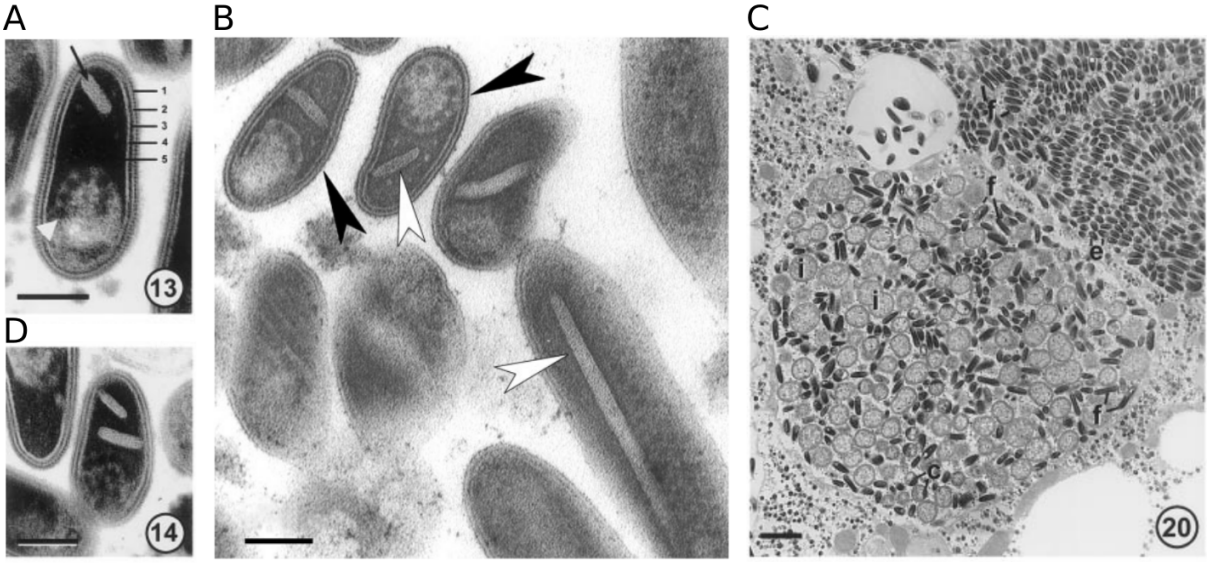

Figure 1.1: Electron microscope images of *R. porcellionis* and *R. crassificans*, adapted from [56] and [14]. (A), (B) and (D) show the typical rod-shaped elementary bodies of *R. crassificans* (A and D) and *R. porcellionis* (B) with their pentalaminar plasma membrane. The black arrow (A) and white arrow (B) point to lamellar bodies. The black arrow in (B) point to the pentalaminar plasma membrane. (C) shows a vacuole with both elementary bodies and reticulate bodies of *R. crassificans*. Black bars: 0.1 µm (A, D and B) or 1 µm (C).

cockroaches [13], and *Rhabdochlamydia porcellionis*, from woodlice [14], although 16S rDNA sequences alone are not sufficient to distinguish members of the *Chlamydiales* at the genus or species level [59] and both bacteria could be part of the same species. *Rhabdochlamydia porcellionis* could later be isolated from infected woodlouse and propagated in Sf9 [60] and is currently the only cultured representative of the *Rhabdochlamydiaceae* family. Interestingly, intracellular bacteria with a similar rod-shaped developmental form had been reported earlier in other arthropods such as scorpions [61], midges [62, 63] and spiders [64]. It is however unclear if those bacteria are part of the same family, as *Simkaniaceae* also have rod-shaped elementary bodies [39] and are also known to infect arthropods [12].

The full host range of the *Rhabdochlamydiaceae* is still unknown. Following the description of *R. porcellionis* and *R. crassificans*, screening by PCR demonstrated the presence of *Rhabdochlamydiaceae* in *Ixodes* ticks [65–68], while less targeted 16S metabarcoding projects also established their presence in spiders [69] and *Dermacentor* ticks [70].

Remarkably, the quantitative PCR used for the screening in [66] showed that samples positive for *Rhabdochlamydiaceae* had important variations in terms of genome copies/µL, with some samples having as many as 8×10^6 copies/µL and others as few as 10 copies/µL, with an average of 20 000 copies/µL. In contrast, samples negative for *Rhabdochlamydiaceae* but positive for *Parachlamydiaceae* had a maximum of 300 copies/µL. Consistently with the high copy numbers observed in qPCR, 16S metabarcoding projects in dwarf spiders and *Dermacentor marginatus* ticks also showed a high relative abundance of *Rhabdochlamydiaceae* in arthropods where this family was detected [69, 70]. Although no microscope evidence is available to confirm this hypothesis, the high copy number observed in the case of *Rhabdochlamydiaceae* is likely due to an infection, while the presence of *Parachlamydiaceae* at low copy numbers can conversely be explained by environmental contamination, in line with previous evidence suggesting that *Parachlamydiaceae* cannot infect insect cells due to their inability to inhibit apoptosis [60, 71, 72]. Although both spiders and ticks belong to the Arachnida class, the detection of *Rhabdochlamydiaceae* in the digestive tract and in the feces of insectivore bats [73, 74] and birds [75], as well as in semi-aquatic insects [74] implies that the host range of this family spans different classes of the Arthropoda phylum. This is coherent with the ability of *R. porcellionis* to grow in Sf9 (*Spodoptera frugiperda*), S2 (*Drosophila melanogaster*) and Aa23T (*Aedes albopictus*) insect cell lines [60].

In addition to their presence in arthropods, members of the *Rhabdochlamydiaceae* were also detected in different water sources and reportedly grown in amoeba co-culture [15, 76, 77]. While the presence of *Rhabdochlamydiaceae* in freshwater environment is coherent with recent publications [21, 75], the evidence for their growth in amoebae is limited and doubtful. Indeed, the only *in-vitro* study of *R. porcellionis* revealed that unlike the other CLOs, *R. porcellionis* is unable to replicate in any of *A. castellanii*, *Acanthamoeba polyphaga* or *Hartmannella vermiformis* [60]. In one study, *A. castellanii* were infected with *R. crassificans* and the progeny used to immunize mice, but without any check to ensure that the bacteria did grow and that any progeny was produced [76].

Western blots loaded with lysed whole bacteria and incubated with immunized mice sera did display bands, but no comparison was made with pre-immune sera. In the absence of evidence for the growth of *R. crassificans* in amoebae, it is questionable whether any bacteria were loaded onto the gel, raising doubts on the nature of the observed bands. In the other two studies [15, 77], the authors attempted to isolate CLOs from water samples, both in *A. castellanii* co-culture and by trying to grow symbiont-containing native amoebae in mixed culture. In one study, *Rhabdochlamydiaceae* could be detected by PCR directly from the samples and in culture with their native, unidentified host, but not in *A. castellanii* co-culture [15]. In the other study, the presence of *Rhabdochlamydiaceae* was demonstrated by PCR in *A. castellanii* co-culture, but as the samples were not filtered before being incubated with amoebae and in the absence of any quantitative evidence of growth, it possible that the PCR amplified the DNA of dead bacteria or that the detected *Rhabdochlamydiaceae* were growing in another host [77]. As *Chlamydiales* species appear to have different tropisms for amoeabe [78], it is possible that *Rhabdochlamydiaceae* only grow in yet untested amoebae or in other protists. Related bacteria were indeed isolated in *Dictyostelium discoideum* [79] and in a ciliate [80]. Isolating these bacteria in culture would therefore require to include a wider diversity of putative host cells in the screens.

Given the difficulties of isolating those organisms, culture-free techniques, such as NGS appear as good alternatives to study their ecology [7, 81] and were successfully used to obtain the genomes of uncultured chlamydiae [18]. Several studies screened public datasets for chlamydial 16S rDNA sequences and performed a meta-analysis of the sequences and their metadata [20, 21, 75]. Such approaches predicted *Rhabdochlamydiaceae* to be the most diverse family of the *Chlamydiales* order, based on the diversity of 16S rDNA sequences assigned to this taxon [20]. More generally, this approach also highlighted the high diversity of environmental chlamydiae, as the full length 16S rDNA sequences clustered in 17 different families, half of which are unknown [20]. In those studies, *Rhabdochlamydiaceae* appeared to be particularly prevalent in soil and freshwater environments [21]. The metadata was however not complete enough to precisely

determine their host organism. Interestingly, the 16S rDNA sequences of *Rhabdochlamydiaceae* obtained in arthropods form a small cluster [75], suggesting that the different clades in this family likely evolved to colonize specific ecological niches and hosts.

Host restriction was further suggested by comparing the genomes of the three sequenced species of the *Rhabdochlamydiaceae* family. In addition to the genome of *R. porcellionis*, the genomes of rhabdochlamydiae infecting ticks [38] (*R. helvetica*) and spiders [75] (*R. oedothoracis*) could also be assembled by shotgun metagenomics of infected hosts. Comparative genomics showed that the genome of *R. porcellionis* was small and relatively transposase-free compared to the larger and transposase-riddled genomes of *R. helvetica* and *R. oedothoracis*. Those observations are compatible with the genome of the two latter species being in the initial steps of the reductive evolution, with a proliferation of transposases and with the genome of the former having an already reduced and stable chromosome. Symbiotic bacteria, either parasitic or beneficial for their host, are indeed known to witness a proliferation of mobile elements before undergoing a rapid genome reduction as they restrict their host range [82–84]. Interestingly, the genome of *R. helvetica* also appears to be shaped by horizontal gene transfers [38]. For instance, some transposases observed in the genome of *R. helvetica* are closely related to transposases found in the genomes of non-chlamydial arthropod symbionts and have likely been transferred horizontally [38]. Horizontal gene transfers could therefore also explain the differences between the genomes of *R. porcellionis* and the other two species; as suggested by Halter et al [75], woodlice might harbor less bacteria than either ticks or spiders and offer less opportunities for gene acquisition by lateral transfer. The genome of *R. porcellionis* might therefore be smaller due the inability of this species to acquire new genes, combined with the known deletion bias in bacteria. However, those kind of analysis are limited by the small number of genomes available for this family.

1.3 *Chlamydiales* as human pathogens

The *Chlamydiales* order includes important human pathogens such as *Chlamydia trachomatis*, the most frequently transmitted sexually-transmitted disease and the leading cause of preventable blindness in the world [85]. *C. trachamotis* was identified in patients suffering of trachoma in the 1900s and could be isolated in a culture system in chicken eggs fifty years later [86]. While initially restricted to *C. trachomatis*, the *Chlamydiales* order has been expanded in the course of the 20$^{\text{th}}$ century and contained 4 species in the 90s, when the first CLOs were discovered[1]: *Chlamydia trachomatis*, *Chlamydia pneumoniae*, *Chlamydia psittaci* and *Chlamydia pecorum*, all grouped in the same family and genus [89]. Both *C. pneumoniae* and *C. psittaci* cause respiratory infections in humans, while *C. pecorum* causes various diseases in livestock and wildlife. In this context, with the *Chlamydiales* order solely composed of pathogens, the discovery of related bacteria sharing the same development cycle naturally led to the suspicion that they too would be pathogenic.

Members of the *Parachlamydiaceae* were among the first CLOs to be investigated for their putative role as human pathogens. *Parachlamydia acanthamoebae* had initially been isolated from the nasal mucosa of a healthy volunteer [90], but its ability to grow in amoebae and its presence in a humidifier associated to an oubreak of fever led to the hypothesis that it could be transmitted and cause lung infections like *Legionella pneumophila* [91–93]. In line with this hypothesis, several studies demonstrated the sero-conversion of patients presenting a pneumonia [91, 94], while other studies based on PCR detected the presence of *Parachlamydia acanthamoebae* in clinical samples from patients suffering of respiratory infections [95–100]. However, the majority of those studies did not include proper controls [94–97, 100] and the studies that did failed to find a statistically significant association with respiratory infections [98, 99], making it impossible to differentiate an infection from a colonization by symbiont-containing amoebae or from an environmental contamination.

[1] the *Chlamydiaceae* now contains 2 genera [11] and 18 species, isolated from mammals [87], birds [11, 88] or reptiles [8, 9]: this family thus appears to be able to infect vertebrates at large

The results from infection models in animals are also equivocal. The instillation of 10^{10} IFU of *Parachlamydia acanthamoebae* in calves by bronchoscopy only induced a mild disease [101]. As a comparison, an infectious dose of 10^9 IFU of *Chlamydia psittaci* killed all the calves in a similar infectious model [102]. In mice, a LD50 of 2.5×10^8 bacteria was determined for a direct injection of the inoculum in the trachea [103], while similar inocula only induced a transient weight loss when injected in the nasal flares [104]. Although such an elevated LD50 could lead to the conclusion of a lack of virulence of *Parachlamydia acanthamoebae*, it is worthwile to remember that mice can be surprisingly resistant to pathogenic bacteria and that similarly high LD50 have been reported for proven humans pathogens [105, 106]. It is however unclear in the case of *Parachlamydia acanthamoebae*, as evidence for its ability to grow in mammalian cells is conflicting. On one hand, *Parachlamydia acanthamoebae* has been reported to replicate in macrophages [107], human pneumocytes and lung fibroblasts [108], but the growth was very limited compared to what was observed in amoebae [109]. On the other hand, a growing body of evidence suggests that *Parachlamydiaceae* are unable to grow in mammalian cells, either due to their inability to inhibit apoptosis [71, 72, 107] or due to temperature sensitivity [110, 111]. Regardless of the growth of *Parachlamydia acanthamoebae* in mammalian cells, the effect of such large inocula in animals can also be explained by the immune response to the presence of pathogen-associated molecular patterns in the lungs [102]. Unfortunately, the controls with heat-inactivated bacteria performed in those studies [103, 104] do not exclude this explanation, as heat could have denatured the PAMP and dampened the immune response [112, 113]. A control with a less destructive inactivation method, as was done for infection models of *Chlamydia pneumoniae* and *Chlamydia psittaci* in mice [114] and calves [102], respectively, would have been preferable to exclude this effect and confirm the pathogenic role of this bacterium. There is however evidence for a higher immunogenicity of heat-inactivated *Parachlamydia acanthamoebae* compared to live bacteria [113] that would mitigate this issue. Altogether, the available data are still insufficient to conclude to the pathogenicity of the *Parachlamydiaceae* for humans.

The evidence for the pathogenicity of *Waddlia chondrophila* is more convincing. This bacterium was first isolated from the homogenate of the lungs and liver of an aborted bovine fetus and later proved to be part of the *Chlamydiales* order [115, 116]. It was isolated independently a second time from the heart of another aborted bovine fetus co-infected with *Neospora caninum*, a known abortifacient agent [117]. Unlike *Parachlamydia acanthamoebae*, *Waddlia chondrophila* does not appear to have a host range restricted to amoebae as it grows well in various mammalian cells, fish cells and insect cells [118–121]. Its discovery in aborted bovine fetuses raised the suspicion that *Waddlia chondrophila* might be an abortifacient agent, particularly since other chlamydia, like *Chlamydia abortus*, were already known to be a major cause of abortions in livestock [122]. A preliminary study in cows demonstrated an association between an history of abortion and the presence of anti-*Waddlia* antibodies [123]. A similar association between miscarriages and a positive *Waddlia* serology was also found in humans [124] and confirmed in subsequent studies that included patients from other countries [125, 126]. Interestingly, the association was statistically significant after adjusting for age and the presence of anti-*Chlamydia trachomatis* antibodies, suggesting that *Waddlia chondrophila* infection might be an independent risk factor for miscarriage [124]. But although the association between miscarriages and a positive *Waddlia* serology was consistently found in different studies, the causality remains to be demonstrated. In particular, no abortions could be observed in murine [127] or bovine [128] models of infections, even if the bacteria could be re-isolated from the placenta and detected in the fetal tissues in one of the challenged animals [128]. Additionally, *Waddlia chondrophila* did not appear to be more prevalent in the placenta or vagina of patients with a positive serology that suffered a miscarriage than in patients that had an uneventful pregnancy [125]. The abortigenic potential of *Waddlia chondrophila* thus appears to be quite reduced compared to that of *Chlamydia abortus* [122]. Moreover, case-control studies are sensitive to confounders; conditions such as the anti-phospholipid syndrome, known to cause both abortions and false positives in syphilis nontreponemal serology [129] could lead to the observed association in the ab-

sence of any effect of *Waddlia chondrophila* on pregnancy. Conversely, the miscarriages in humans could also be caused by the immunological reaction to *Waddlia chondrophila* [127]. Altogether, the evidence for the pathogenicity of *Waddlia chondrophila* is stronger than for any other CLOs, but additional work is necessary to exclude spurious associations and to demonstrate the causal link between the infection and the miscarriages, let alone to justify antibiotic treatments as prophylaxis for miscarriage.

Another family, the *Simkaniaceae*, was also thought to be pathogenic after a high sero-prevalence was reported in the general population [130] following its discovery in 1993 as a cell culture contaminant [131]. It was also associated to bronchiolitis in a case-control study that used a combination of PCR and culture to detect the organism [132]. The conclusions were however weakened by a statistical analysis that did not take into account the concomitant presence of other pathogens known to cause bronchiolitis. Moreover, despite its frequent detection in clinical samples in studies usually lacking controls, no case-control study could confirm the association with respiratory infections (reviewed in [133]). The initial sero-prevalence studies have also been criticized because of a low specificity of the assay used, and could not be replicated [134]. Overall, it seems doubtful that this bacterium has any pathogenic role in humans [134].

As a recent addition to the *Chlamydiales* order, the *Rhabdochlamydiaceae* have had less focus than other families. Unlike the *Parachlamydiaceae*, the *Rhabdochlamydiaceae* appear to have the ability to inhibit apoptosis [60], suggesting that they could indeed infect mammalian cells [71]. Moreover, the presence of *Rhabdochlamydiaceae* in ticks and in the skin of patients with a history of tick bites [67] indicated that they might be vector-borne. Several PCR-based screening studies, with a focus on respiratory infections, detected the presence of *Rhabdochlamydiaceae* in clinical samples (Table 1.1). The wide range of reported prevalences (from 0.4% to 37%) may be due to differences in the sensitivity of the assays, as studies that used *Rhabdochlamydiaceae*-specific real-time PCR systematically detected more *Rhabdochlamydiaceae* than studies that used broad range pan-*Chlamydiales* PCR, although the comparison between studies is complicated by dif-

Table 1.1: Clinical evidence for the pathogenicity of the *Rhabdochlamydiaceae* family. The table shows the proportion of patients tested positive for *Rhabdochlamydiaceae*. The second study [97] included all neonates hospitalized in a neonatology ward, regardless of their diagnosis. The other studies only included patients with specific conditions.

Country	**Disease**	**Patients**	**Controls**	**p-value**	**Reference**
Switzerland[a]	respiratory infection	11/29	0/6	0.06[b]	[98]
Switzerland[a]	-	4/29	-	-	[97]
Switzerland[c]	pneumonia	1/265	2/157	0.56[d]	[99]
Finland[a]	respiratory infection	16/136	-	-	[135]
Germany[e]	pneumonia	2/387	-	-	[95]
Finland[c]	granuloma annulare	2/73	0/39	0.54[d]	[136]

[a] used the *Rhabdochlamydiaceae*-specific real-time PCR developed in [97]

[b] the p-value reported in the paper, obtained by comparing the prevalence of *Rhabdochlamydiaceae*, *Parachlamydia acanthamoebae* and *Protochlamydia naegleriophila* in the controls and patients groups. The p-value if only *Rhabdochlamydiaceae* are taken into account is 0.15, based on the Fisher exact test

[c] used the pan-*Chlamydiales* real-time PCR [99] and amplicon sequencing to identity the family

[d] not reported in the paper, the prevalence in patients and controls were compared with the Fisher exact test

[e] used a nested PCR first amplifying a 1500bp sequence and a 290bp sequence as a second step

ferences in terms of patients population, sampling methods and pathology. However, the absence of controls in half of the studies make it impossible to exclude the possibility of a colonization. A colonization is further suggested by the detection of *Rhabdochlamydiaceae* in nasopharyngeal swabs of healthy volunteers [137]. In addition, the studies that did include controls failed to identify an association with either respiratory infections or skin conditions. The hypothesis of the pathogenicity of *Rhabdochlamydiaceae* is therefore poorly supported by clinical evidence. Interestingly, *Rhabdochlamydiaceae* could be detected in the biofilm of extracorporeal membrane oxygenation catheters in a 16S metabarcoding study [138]. The implication is however unclear, as *Rhabdochlamydiaceae* were only present at a low abundance, with less than 1% of the reads assigned to the family. Altogether, the current evidence is not sufficient to conclude to a pathogenic role of the *Rhabdochlamydiaceae*, as their presence in clinical samples might reflect their abundance in the environment [20, 21, 75].

1.4 Aims

The aims of this project were to:

1. Improve the understanding of the biology of the *Rhabdochlamydiaceae*. In particular, as the host range of this family remains largely unknown, study the permissivity of various cell lines to *R. porcellionis*, the only cultured representative of the *Rhabdochlamydiaceae*. The focus will be on mammalian cell lines, in an attempt to define the pathogenic role of *Rhabdochlamydia* sp. for humans.

2. Describe new species isolated from ticks. Previous sequencing projects were indeed successful in assembling genomes by a shotgun sequencing approach on infected hosts [38, 75]. The aim is to apply the same method to study the diversity and biology of the *Rhabdochlamydiaceae* family.

3. Determine whether ticks could serve as a vector for *Rhabdochlamydiaceae* infections in humans.

The first aim is addressed in Chapter 5, where we present our results on how temperature sensitivity limits the host range of *R. porcellionis*. The second aim is addressed in Chapters 3 and 4. We assembled the genomes of several *Rhabdochlamydiaceae* species sequenced from ticks and other arthropods. To compare those genomes, we developed a tool, zDB, to ease the exploration and comparison of bacterial genomes. We present this tool in Chapter 3, while the preliminary results of genomes analysis are presented in Chapter 4. Aim 3 was only partially addressed: we present our results on the prevalence of *Chlamydiales* and their co-occurence with tick-borne pathogens in ticks in Chapter 2.

2 Presence of *Chlamydiales* in Swiss ticks and co-infection rate with tick-borne pathogens

Authors: Bastian Marquis, Werner Tischhauser, Rahel Ackermann-Gäumann, Reto Lienhard, Marie-Lise Tritten, Virginie Martin, Gilbert Greub

Status: submitted to New Microbes and New Infection

Contribution: BM performed the *Rhabdochlamydiaceae* qPCR, the statistical analysis and wrote the manuscript.

Abstract

Ticks are well known vectors of multiple human pathogens. Previous studies have already established the presence of *Chlamydiae* in ticks, they however did not search for co-infections with human pathogens. In this study, 561 ticks at various developmental stages were collected from humans and screened for the presence of *Chlamydiales*, *Borrelia*, the tick-borne encephalitis virus, *Rickettsia* spp., *Anaplasma phagocytophilum*, *Babesia* spp., *Francisella tularensis* and *Candidatus* Neoehrlichia mikurensis. We confirmed the presence of two families of the *Chlamydiales* order, the *Rhabdochlamydiaceae* (1.1%) and the *Parachlamydiaceae* (3.6%) in ticks, with an overall prevalence of 4.5%. Our data also suggests an association between the presence of *Chlamydiales* and tick-borne pathogens (OR=2.05), although this result is not statistically significant.

2.1 Introduction

The *Chlamydiae* phylum is composed of intracellular obligate bacteria that share the same biphasic development cycle [6] composed of an extracellular and infectious form, the elementary body (EB), and an intracellular and replicative form, the reticulate body (RB) [139]. After having been internalized in its host cell, the EB differentiates into an RB and initiates its replication cycle. After several rounds of replication, RBs differentiate back into EBs and exit the cell, either by lysis or by exocytosis to start a new infection cycle [140]. While initially restricted to the *Chlamydiaceae* family, the *Chlamydiales* order has seen a rapid expansion in the past decades as new families were discovered in various organisms and environments, such as fish [18, 141], arthropods [13, 14, 69, 75] and marine sediments [19, 22]. Overall, members of the *Chlamydiales* order appear to be widespread in the environment [6, 7, 21]. Interestingly, the presence of chlamydiae in ticks has been documented in different regions of the world [65–68, 142].

Intracellular bacteria, such as *Rickettsia* or *Francisella*, are known to be part of the physiological tick microbiota and are necessary for their survival as they synthesize nutrients absent from blood meals [143]. As for chlamydiae, evidences would rather suggest a pathogenic role. Indeed, infection of arthropod cell lines by chlamydiae was demonstrated to have a detrimental effect [58, 60]. In addition, *Rhabdochlamydia* infection in *Porcellio scaber* and *Blatta orientalis* was shown to induce abdominal swelling and to ultimately cause the death of the infected organisms [58]. Finally, the analysis of a chlamydiae genome sequenced from a tick sample suggested a parasitic rather than a symbiotic lifestyle [38]. Several studies hint at interplays between the tick microbiota and the colonization by tick-borne pathogens [144–146]. As chlamydiae were shown to infect the digestive tract of arthropods [13, 56, 58], their presence could have an influence on the colonization of ticks by tick-borne pathogens. In this study, we aimed to determine the prevalence of chlamydiae in questing ticks and to assess how frequent the co-occurrence of chlamydiae and tick-borne pathogens is.

2.2 Methods

Tick collection

Questing ticks were collected in Switzerland during the 02/2018 to 03/2019 period by users of a smartphone tick application ("Zecke – Tick Prevention"). Users sent the collected ticks to the Swiss National Reference Center for Tick-transmitted Diseases (CNRT) for analysis. The application allowed the collection of additional data, such the geographical location where the tick was picked.

The tick developmental stage and engorgement were determined by microscopic examination at the CNRT. To prepare for the screenings, tick samples were homogenized in 600 µL of pre-cooled PBS using the TissueLyser system (Qiagen, Hilden, Germany). After a short centrifugation step, 400 µL of the supernatant were transferred to a Deep-well plate (Eppendorf, Hamburg, Germany), 60 µL of glycerin were added per well and the were plates stored at −80 °C for further use. 100 µL of the supernatant were used for nucleic acid extraction.

Nucleic acid extraction

100 µL of tick homogenate supernatant were lysed in 400 µL of AVL buffer supplemented with InhibitEX Tablets (Qiagen, Hilden, Germany) in a 96-well MagNA Pure processing cartridge (Roche, Penzberg, Germany). NA extraction was performed with the MagNA Pure 96 instrument and the MagNA Pure 96 DNA and Viral NA Large Volume kit, using the Pathogen Universal LV 2.0 protocol, a sample volume of 500 µL and an elution volume of 100 µL.

PCR screening

Individual ticks were screened by qPCR at the CNRT for *Borrelia* spp., the tick-borne encephalitis virus, *Rickettsia* spp., *Anaplasma phagocytophilum*, *Babesia* spp., *Francisella*

tularensis and *Candidatus* Neoehrlichia mikurensis, as described in [147]. As part of the present project, ticks were also screened for *Chlamydiales* and *Rhabdochlamydiaceae* using a pan-*Chlamydiales* qPCR and a *Rhabdochlamydia* qPCR, respectively. Both assays amplify a fragment of the 16S rRNA encoding gene. The pan-*Chlamydiales* qPCR was performed as described in [99], whereas the *Rhabdochlamydia* qPCR was performed as described in [97]. Negative controls with DNA-free water were included in all plates. Amplicons from samples with a CT value lower than 35 with the pan-*Chlamydiales* qPCR were sent for sequencing (Microsynth AG, Balgach). The sequences were used for the identification at the family level by BLAST [148] searches on the NCBI database. The statistical analysis was done with R (v3.6.1). The 95% confidence intervals were calculated using the z-test for proportions and categorical variables were compared using the Fisher exact test.

2.3 Results

A total of 561 ticks were collected (55 adults, 458 nymphs, 43 larvae, the stage could not be determined in 5 cases). Based on macroscopic appearance and previous epidemiological data [65, 66], all ticks likely belonged to the *Ixodes ricinus* species. Of the 561 ticks, 23 were engorged (4.7%, 95% CI: 3.08-7.09%). However, 73 ticks were too damaged for their engorgement to be ascertained.

The pan-*Chlamydiales* qPCR was positive in 20/561 (3.6%, 95% CI: 2.31-5.44%) samples, the *Rhabdochlamydia* qPCR in 6/561 (1.1%, 95% CI: 0.49-2.31%) samples, for an overall prevalence of 25/561 (4.5%, 95% CI: 3.04%-6.50%), with only one sample positive in both assays. All positive samples had a low number of genome copies (median of 710 copies/mL or 178 copies per tick). Three pan-*Chlamydiales* amplicons could be sequenced and all belonged to the *Parachlamydiaceae* family (Supplementary Table 2.3). Nine out of 25 chlamydiae infected ticks were adults (36%, 95% CI: 18.71-57.38%), 15 were nymphs (60%, 95% CI: 38.89-78.19%) and only one was a larva (4%, 95%

Table 2.1: Distribution of tick developmental stages. The percentages indicate the proportion of developmental stages for each given column. The developmental of 5 ticks could not be determined.

Tick stage	**Chlamydiae positive**[a]			**Chlamydiae**
(N=556)	Total	*Rhabdochlamydia*	pan-*Chlamydiales*	**negative**
Larva	1 (4%)	0 (0%)	1 (5%)	42 (7%)
Nymph	15 (60%)[b]	4 (67%)	12 (60%)	443 (83%)
Adult	9 (36%)	2 (33%)	7 (35%)	46 (9%)
Total	25	6	20	531

[a] Ticks were considered as chlamydiae-positive if at least one of the *Rhabdochlamydia* or the pan-*Chlamydiales* qPCR were positive.
[b] One nymph was positive in both qPCR.

CI: 0.21-22.32%). The distribution of tick developmental stages in chlamydia-positive and chlamydia-negative ticks is shown in Table 2.1. Interestingly, the distribution of ticks developmental stages was different in chlamydia-positive and chlamydia-negative ticks (p=0.0008, Fisher exact test). No engorged tick was positive in either the pan-*Chlamydiales* or the *Rhabdochlamydia* qPCRs and there was no association between tick engorgement and the presence of chlamydiae (p-value = 0.61, Fisher exact test).

Tick-borne pathogens were present in 12 of the 25 chlamydia-positive ticks (48.00%, 95% CI 30.03-66.50%) and in 166 of the 536 chlamydia-negative ticks (30.97%, 95% CI 27.20-35.00%). The association between the presence of tick-borne pathogens and the presence of chlamydia was not statistically significant (OR 2.05, 95% CI: 0.84-5.00). Chlamydia-positive ticks carried *Rickettsia* spp. (N=5), *Borrelia* spp (N=6), *Candidatus* Neoehrlichia mikurensis (N=1) and *Anaplasma phagocytophilum* (N=1), with a tick being carrier of both *Rickettsia* spp. and *Borrelia* spp. None of those pathogens was associated with the presence of chlamydia (Table 2.2). The prevalence of tick-borne pathogens was identical in ticks positive with the rhabdochlamydia-specific qPCR (n=3/6, 50%, 95% IC: 18.76%-81.23%) and ticks positive with the pan-*Chlamydiales* qPCR (n=10/20, 50%, 95% IC: 29.93%-70.07%).

Table 2.2: Distribution of pathogens in chlamydiae-positive and chlamydiae-negative ticks. The last columns shows the odds ratio of the presence of tick-borne pathogens in chlamydiae-positive and chlamydiae-negative ticks.

Pathogen	**Chlamydiae positive (N=25)**	**Chlamydiae negative (N=536)**	**Odds ratio (95% CI)**
Rickettsia spp.	5 (20%)	63 (12%)	1.87 (0.53-5.39)
Borrelia spp.	6 (24%)	82 (15%)	1.74 (0.55-4.73)
Anaplasma phagocytophilum	1 (4%)	7 (1%)	3.14 (0.07-26.10)
Candidatus Neoehrichia mikurensis	1 (4%)	30 (5%)	0.70 (0.02-4.63)
Babesia spp.	0 (0%)	12 (2%)	0.00 (0.00-8.04)
Francisella tularensis	0 (0%)	1 (0%)	0.00 (0.00-825.39)
All pathogens	**12 (48%)**	**166 (31%)**	**2.05 (0.84-5.00)**

2.4 Discussion

The prevalence of 4.5% in our samples is in the range of those reported in previous Swiss studies [65, 66]. It is however much smaller than those reported in Algeria (45%), Finland (40%), Australia (26.9%-46.8%) and Italy (28%) [65, 67, 68, 142]. The chlamydiae families found in this study match those reported in previous publications from Finland, Algeria and Switzerland, with a predominance of *Rhabdochlamydiaceae* and *Parachlamydiaceae*. The families are however different from those found in ticks from Australia and Italy. However, it must be noted that the collected ticks were mostly of the *Ixodes* genus in previous studies from Switzerland [65, 66], Finland [67] and Australia [68], whereas *Rhipicephalus* and *Hyalomma* were the most frequent genus in Italy and Algeria [65, 142]. Aside from geographical differences, differences in host range could thus be an explanation for the observed variations in the prevalence of chlamydiae families.

The comparison of the tick developmental stages in chlamydiae-infected and chlamydiae-free ticks suggests a late acquisition in the development cycle, as opposed to the transovarial transmission reported for some tick-borne pathogens and tick symbionts [149]. The acquisition via a blood meal on an unknown reservoir is a possible hypothesis. Interestingly, the important variations in terms of number of genomes copies previously

reported, particularly for *Rhabdochlamydia* [66], was not observed in this study with all samples having less than 2000 copies/mL. As evidence suggests a pathogenic role of *Rhabdochlamydiaceae* in arthropods [56, 58, 60], it is possible that due to a lower fitness, highly infected ticks were less likely to be questing and consequently less likely to be collected on a human host. In contrast, previous studies that relied on flagging [66–68] would likely be less biased towards healthy ticks. Conversely, the low number of genomes copies observed for *Parachlamydiaceae* is consistent with previous evidence [66] and likely reflects the presence of amoebae containing *Parachlamydiaceae* symbionts on the surface of the ticks. Indeed, only *Rhabdochlamydiaceae* and *Simkaniaceae* could be detected when ticks were subjected to a washing step before before DNA extraction [68]; a step demonstrated to be important to avoid surface contaminants [150] that was lacking in most studies [65–67]. Moreover, *Parachlamydiaceae* cannot replicate in either mammalian [71] or insect cell lines [72, 121] as they are unable to inhibit apoptosis [71, 72], unlike *Rhabdochlamydiaceae* and *Simkaniaceae*, both able to grow in insect cell lines [60, 151] and both observed in the tissues of infected arthropods [12, 13]. The low copy numbers of this study might therefore be a consequence of a bias of the collection method in the case of *Rhabdochlamydiaceae* and a consequence of environmental contamination in the case of *Parachlamydiaceae*.

The *Rhabdochlamydiaceae* qPCR was more sensitive than the pan-*Chlamydiales* qPCR: only one sample positive for *Rhabdochlamydiaceae* was also positive with the pan-*Chlamydiales* qPCR. This decreased sensitivity is expected as the pan-*Chlamydiales* qPCR amplifies a longer fragment than the *Rhabdochlamydiaceae* qPCR. In addition, the primer of the pan-*Chlamydiales* qPCR is known to have a mismatch in its target sequence in the *Rhabdochlamydiaceae*, which might further decrease the sensivity of the PCR for this particular family [99, 152]. However, this implies that the true prevalence of *Chlamydiales* in ticks may have been underestimated. Even if no mismatches are reported in the *Parachlamydiaceae*, a real-time PCR amplifying a shorter sequence could have detected the presence of chlamydiae in samples where no amplification was detected with the pan-

Chlamydiales qPCR. A closer estimation could be obtained by using a set of genus or family-specific qPCR amplifying shorter regions.

The difference of prevalence of co-infections in chlamydiae-infected ticks and chlamydiae-free ticks is also worth mentioning. While not statistically significant in this work, the two-fold difference is puzzling. Previous studies established the modulatory role of tick microbiota towards the colonization by pathogens [144–146]. As previously reported [70], the diversity in microbiota seems to be lower in chlamydiae-infected ticks. Chlamydial infection may thus be responsible for the high prevalence of pathogens in ticks by altering their microbiota. Conversely, it may just be an opportunistic colonizer of ticks in dysbiosis and be a marker of poor tick health. Another explanation for the increased prevalence of tick-borne pathogens among chlamydiae-infected ticks would be a similar acquisition mode, i.e. following a blood meal on infected rodents, which are well-known reservoir of several zoonotic pathogens.

In conclusion, this study confirms the presence of members of the *Chlamydiales* order in Swiss ticks, in particular, of the *Parachlamydiaceae* and of the *Rhabdochlamydiaceae* families. It also suggests an association between chlamydial infection and the presence of human tick-borne pathogens. Finally, this work suggests that ticks heavily by *Rhabdochlamydiaceae* may have a reduced fitness and be less prone to bite humans.

2.5 Supplementary materials

Table 2.3: Fragments of the 16S rDNA obtained by amplicon sequencing and the taxonomy of their best blast hit in the NCBI non-redundant nucleotides database. Although 16S rDNA sequences do not allow the classification at the species or genus level, they can be used to make an identification at the family level.

Sequence	**Best hit taxonomy**	**E-value**
CTTTACGACCCTAAGGCCTTCTTCGCTCACA CGGCGTCGCATCGTCAGGCTTTCGCCCATTG CGAATGATTCTCGACTGCAGCCTCCCGTAGG AGTCTGGGCAGTTCTCAGTCCCAGTG	*Protochlamydia naegleriophila*	2×10^{-53}
GCGAAAGAGCTTTACRACCCNAAGGCCTTCA TCGCTCACACGGCGTCGCATCGTCAGGCTTT CGCCCATTGCGAATGATTCTCGACTGCAGCC TCCCGTAGGAGTCTGGGC	*Parachlamydia acanthamoebae*	7×10^{-47}
TTCTCTTGTTCCAGGCGAAGAGCTTTACGAC CCTAAGGCCTTCATCGCTCACACGGCGTCGC ATCGTCAGGCTTTCGCCCATTGCGAATGATT CTCGACTGCAGCCTCCCGTAGGAGTCTGGGC AGTTCTCAGTCCCAGTG	*Parachlamydia acanthamoebae*	3×10^{-61}

2.6 Bibliography

6. Horn M. *Chlamydiae* as symbionts in eukaryotes. Annual review of microbiology 2008;62:113–31.

7. Collingro A, Köstlbacher S, and Horn M. *Chlamydiae* in the environment. Trends in Microbiology 2020;28:877–88.

12. Everett KD, Thao M, Horn M, Dyszynski GE, and Baumann P. Novel chlamydiae in whiteflies and scale insects: endosymbionts '*Candidatus* Fritschea bemisiae' strain Falk and '*Candidatus* Fritschea eriococci' strain Elm. International Journal of Systematic and Evolutionary Microbiology 2005;55:1581–7.

13. Corsaro D, Thomas V, Goy G, Venditti D, Radek R, and Greub G. '*Candidatus* Rhabdochlamydia crassificans', an intracellular bacterial pathogen of the cockroach *Blatta orientalis* (Insecta: Blattodea). Systematic and Applied Microbiology 2007;30:221–8.

14. Kostanjšek R, Štrus J, Drobne D, and Avguštin G. '*Candidatus* Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod *Porcellio scaber* (Crustacea: Isopoda). International journal of systematic and evolutionary microbiology 2004;54:543–9.

18. Taylor-Brown A, Pillonel T, Greub G, Vaughan L, Nowak B, and Polkinghorne A. Metagenomic analysis of fish-associated *Ca.* Parilichlamydiaceae reveals striking metabolic similarities to the terrestrial *Chlamydiaceae*. Genome biology and evolution 2018;10:2587–95.

19. Dharamshi JE, Tamarit D, Eme L, et al. Marine sediments illuminate chlamydiae diversity and evolution. Current Biology 2020;30:1032–1048.e7.

21. Köstlbacher S, Collingro A, Halter T, Schulz F, Jungbluth SP, and Horn M. Pangenomics reveals alternative environmental lifestyles among chlamydiae. Nature Communications 2021;12:1–15.

22. Collingro A, Köstlbacher S, Mussmann M, Stepanauskas R, Hallam SJ, and Horn M. Unexpected genomic features in widespread intracellular bacteria: evidence for motility of marine chlamydiae. The ISME journal 2017;11:2334–44.

38. Pillonel T, Bertelli C, Aeby S, et al. Sequencing the obligate intracellular *Rhabdochlamydia helvetica* within its tick host *Ixodes ricinus* to investigate their symbiotic relationship. Genome biology and evolution 2019;11:1334–44.

56. Radek R. Light and electron microscopic study of *Rickettsiella* species from the cockroach *Blatta orientalis*. Journal of invertebrate pathology 2000;76:249–56.

58. Kostanjšek R and Marolt TP. Pathogenesis, tissue distribution and host response to *Rhabdochlamydia porcellionis* infection in rough woodlouse *Porcellio scaber*. Journal of invertebrate pathology 2015;125:56–67.

60. Sixt BS, Kostanjšek R, Mustedanagic A, Toenshoff ER, and Horn M. Developmental cycle and host interaction of *Rhabdochlamydia porcellionis*, an intracellular parasite of terrestrial isopods. Environmental Microbiology 2013;15:2980–93.

65. Croxatto A, Rieille N, Kernif T, et al. Presence of *Chlamydiales* DNA in ticks and fleas suggests that ticks are carriers of *Chlamydiae*. Ticks and tick-borne diseases 2014;5:359–65.

66. Pilloux L, Aeby S, Gaümann R, Burri C, Beuret C, and Greub G. The high prevalence and diversity of *Chlamydiales* DNA within *Ixodes ricinus* ticks suggest a role for ticks as reservoirs and vectors of *Chlamydia*-related bacteria. Applied and Environmental Microbiology 2015;81:8177–82.

67. Hokynar K, Sormunen JJ, Vesterinen EJ, et al. *Chlamydia*-like organisms (CLOs) in finnish *Ixodes ricinus* ticks and human skin. Microorganisms 2016;4:28.

68. Burnard D, Weaver H, Gillett A, Loader J, Flanagan C, and Polkinghorne A. Novel *Chlamydiales* genotypes identified in ticks from Australian wildlife. Parasites & Vectors 2017;10:46.

69. Vanthournout B and Hendrickx F. Endosymbiont dominated bacterial communities in a dwarf spider. PLoS One 2015;10:e0117297.

70. Zhang YK, Yu ZJ, Wang D, Bronislava V, Branislav P, and Liu JZ. The bacterial microbiome of field-collected *Dermacentor marginatus* and *Dermacentor reticulatus* from Slovakia. Parasites & vectors 2019;12:1–10.

71. Brokatzky D, Kretz O, and Häcker G. Apoptosis functions in defense against infection of mammalian cells with environmental chlamydiae. Infection and immunity 2020;88:e00851–19.

72. Sixt BS, Hiess B, König L, and Horn M. Lack of effective anti-apoptotic activities restricts growth of *Parachlamydiaceae* in insect cells. PloS One 2012;7:e29565.

75. Halter T, Koestlbacher S, Collingro A, et al. Ecology and evolution of chlamydial symbionts of arthropods. ISME Communications 2022;2:1–11.

97. Lamoth F, Aeby S, Schneider A, Jaton-Ogay K, Vaudaux B, and Greub G. *Parachlamydia* and *Rhabdochlamydia* in premature neonates. Emerging Infectious Diseases 2009;15:2072–5.

99. Lienard J, Croxatto A, Aeby S, et al. Development of a new *Chlamydiales*-specific real-time PCR and its application to respiratory clinical samples. Journal of Clinical Microbiology 2011;49:2637–42.

121. Kebbi-Beghdadi C, Fatton M, and Greub G. Permissivity of insect cells to *Waddlia chondrophila*, *Estrella lausannensis* and *Parachlamydia acanthamoebae*. Microbes and Infection. Special issue on intracellular bacteria 2015;17:749–54.

139. Abdelrahman YM and Belland RJ. The chlamydial developmental cycle. FEMS microbiology reviews 2005;29:949–59.

140. Chiarelli TJ, Grieshaber NA, Omsland A, Remien CH, and Grieshaber SS. Single-inclusion kinetics of *Chlamydia trachomatis* development. mSystems 2020;5:e00689–20.

141. Steigen A, Nylund A, Plarre H, Watanabe K, Karlsbakk E, and Brevik Ø. Presence of selected pathogens on the gills of five wrasse species in western Norway. Diseases of Aquatic Organisms 2018;128:21–35.

142. Chisu V, Foxi C, Tanda A, and Masala G. Molecular evidence of *Chlamydiales* in ticks from wild and domestic hosts in Sardinia, Italy. Parasitology Research 2018;117:981–7.

143. Bonnet SI, Binetruy F, Hernández-Jarguín AM, and Duron O. The tick microbiome: why non-pathogenic microorganisms matter in tick biology and pathogen transmission. Frontiers in cellular and infection microbiology 2017;7:236.

144. Abraham NM, Liu L, Jutras BL, et al. Pathogen-mediated manipulation of arthropod microbiota to promote infection. Proceedings of the National Academy of Sciences 2017;114:E781–E790.

145. Narasimhan S, Rajeevan N, Liu L, et al. Gut microbiota of the tick vector *Ixodes scapularis* modulate colonization of the Lyme disease spirochete. Cell host & microbe 2014;15:58–71.

146. Adegoke A, Kumar D, Bobo C, et al. Tick-borne pathogens shape the native microbiome within tick vectors. Microorganisms 2020;8:E1299.

147. Oechslin CP, Heutschi D, Lenz N, et al. Prevalence of tick-borne pathogens in questing *Ixodes ricinus* ticks in urban and suburban areas of Switzerland. Parasites & Vectors 2017;10:558.

148. Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. Basic local alignment search tool. Journal of molecular biology 1990;215:403–10.

149. Hodosi R, Kazimirova M, and Soltys K. What do we know about the microbiome of *I. ricinus*? Frontiers in cellular and infection microbiology 2022;12:990889.

150. Binetruy F, Dupraz M, Buysse M, and Duron O. Surface sterilization methods impact measures of internal microbial diversity in ticks. Parasites & vectors 2019;12:1–10.

151. Vouga M, Baud D, and Greub G. *Simkania negevensis* may produce long-lasting infections in human pneumocytes and endometrial cells. Pathogens and Disease 2017;75:ftw115.

152. Stadhouders R, Pas SD, Anber J, Voermans J, Mes TH, and Schutten M. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. The Journal of Molecular Diagnostics 2010;12:109–17.

3 zDB: bacterial comparative genomics made easy

Authors: Bastian Marquis, Trestan Pillonel, Alessia Carrara and Claire Bertelli

Status: In review, mSystems

Contributions: BM designed the database, made most of the programming work and wrote the draft

Abstract

The analysis and comparison of genomes relies on different tools for tasks such as annotation, orthology prediction and phylogenetic inference. Most tools are specialized for a single task and additional efforts are thus necessary to integrate and visualize the results. To fill this gap, we developed zDB, an application that integrates an analysis pipeline and a visualization platform. Starting from annotated Genbank files, zDB identifies orthologs and infers a phylogeny for each orthogroup. A species phylogeny is also constructed from shared single-copy orthologs. The results can be enriched with Pfam protein domain prediction, COG and KEGG annotations and Swissprot homologs. The web application allows searching for specific genes or annotations, running Blast queries and comparing genomic regions and whole genomes. The metabolic capacities of organisms can be compared at either the module or pathway levels. Finally, users can run queries to examine the conservation of specific genes or annotations across a chosen subset of genomes and display the results as a list of genes, Venn diagram or heatmaps. Those features will make zDB useful for both bioinformaticians and researchers more accustomed to laboratory research. zDB is perfectly suited to process datasets with tens to hundred of genomes on a desktop machine.

3.1 Introduction

Since the publication of the first complete genome in 1995 [153], the number of available sequences has kept on growing, with now 450'000 different species available in the Genbank database [154]. As recent sequencing technologies make it possible to sequence an organism in a matter of hours at a cost affordable even for small research laboratories, this trend is unlikely to abate in the foreseeable future. These technological improvements transferred the burden from sequencing to the actual analysis of the sequences. While a plethora of different tools already exist for this purpose, they are often specialized for specific tasks such as gene calling, orthology prediction or phylogenetic inference. Moreover, these tools are often standalone programs that do not readily integrate each other's results. As the results are often not produced in a format that easily allows their exploration, additional visualization efforts are also necessary.

The need for tools designed to aggregate results from different sources has been illustrated by the success of programs like Prokka [155], which merges the results of different annotation tools in files ready for submission and visualization in genome browsers. The idea is further extended by pipelines like Bactopia [156], TORMES [157] and ASA3P [158] that automate all steps from reads quality control to antibiotic resistance gene prediction and generate simple HTML reports allowing the visualization of the main results. As these pipelines were developed with a focus on clinical microbiology, they are limited in terms of comparative genomics analysis. In contrast, websites dedicated to the comparative genomics of specific group of organisms [159, 160] have been developed and implement powerful interfaces allowing users to make custom queries and to generate complex plots. However, these websites do not allow users to analyze their own sets of genomes. Some web-based comparative genomics platforms, like EDGAR [161], PhyloCloud [162], CoGe [163] or MicroScope [164] implement similar interfaces while allowing users to upload their own dataset. Those platforms are however closed-source and as the analysis are performed on the platform's respective servers, users are required to register

and upload their dataset. The ideal comparative genomics platform would therefore be open source, could run on any infrastructure, be as flexible and scalable as Bactopia, and similarly to MicroScope and EDGAR, offer an extensive interface to visualize the results.

To fill this gap, we developed zDB, an open-source comparative genomics analysis pipeline and visualization platform. The analysis pipeline performs functional annotations, orthology and phylogenetic inference, while the visualization platform offers an interactive modern web-based interface to explore the results. Altogether, the ease of installing and executing the tool and the ability to easily visualize the results will benefit both bioinformaticians and researchers more accustomed to lab work.

3.2 Results

The visualization platform can be started by a single command as soon as the analyses are complete. The command starts a web server that will make the results available via a web browser, either only locally or also possibly extended to the whole Internet depending on the setup. The platform can also be used to visualize archived results imported from another computer.

3.2.1 Visualization toolkit

The visualization platforms implements a set of plots and queries to explore the results of orthology prediction and phylogeny inference. In addition, zDB comes with several features of more general interest like the possibility to run Blast queries, to search for specific annotation or gene using a search bar and to draw interactive Circos plots or genomic regions.

A side bar is present on all pages of the web application (Fig. 3.1A) and allow a quick access to all available analyses. The content of the "Annotations" tab varies in function of which optional analyses were performed. Similarly, the "Metabolism" tab will only be present when the genomes were annotated with KEGG orthologs. Summaries of the

main characteristics of the genomes of interest, including the results from CheckM, can be visualized either as lists or directly annotated in the species phylogeny (Supp. Fig. 3.5) via the "Genomes" and "Phylogeny" tabs, respectively. Finally, the web interface also includes a search bar (Fig. 3.1A) that allows users to look for genes, gene products, bacteria or specific annotations based on their name. The search bar accepts wildcards and logical operators to combine different search terms.

The "Orthology" tab links to pages allowing to explore gene conservation in the dataset. In particular, users can visualize gene conservation across a chosen set of genomes as either heatmaps (Fig. 3.1G), Venn diagrams (Fig. 3.1E) or lists. zDB can also draw the commonly used core and pan-genome plots as well as a plot of the number of orthogroups in function of the number of genomes where they occur (Fig. 3.1H). The latter plot allows to quickly assess the number of singletons, the size of the core genome and to detect group of genes occurring in a subset of the genomes. Finally, zDB implements an interface that allows users to search for genes conserved in a chosen set of genomes but absent in another one.

As searching for specific sequences in organisms of interest is a frequent task, the visualization platform implements an interface that allows users to run their own Blast searches on either the whole dataset or on a specific genome. Several types of blast searches can be performed (blastp, blastn and tblastn), either with a single query or with multiple query in FASTA format. The results are displayed interactively using the BlasterJS library [165]. Moreover, if the search was run on the whole dataset, zDB can display the results as a heatmap of the best blast hits identity linked to the species phylogenetic tree (Fig. 3.1B). This allows to quickly detect patterns in the distribution of Blast hits in function of the phylogenetic distance.

Finally, zDB can draw plots to compare genomic regions sharing orthologous genes (Fig. 3.1C) and Circos plots to compare a set of genomes to a specified reference (Fig. 3.1D). The minimal setup also includes summary pages for every gene and orthogroup. The gene summary page allows to easily access the nucleotide and amino-acid sequences

(for protein coding genes), displays the genomic region of the gene of interest, as well as the list of orthologous genes (Fig. 3.2A). The orthogroup page allows users to examine the gene phylogeny and the distribution of the orthogroup in the genomes of the dataset (Fig. 3.1F and Fig. 3.2C). Both pages are enriched with the results of the optional analyses if they were performed (Fig. 3.2A and Fig. 3.1F).

3.2.2 Pfam, COG and KEGG functional analyses

The conservation of Pfam, COG and KEGG annotations across genomes can be compared in a similar way to orthogroups. In particular, Venn diagrams, heatmaps, pan- and core-genome plots can be drawn for those annotations, while an interface to search for annotations present in a set of chosen genomes but absent in another is also available. Since COG and KEGG orthologs are assigned to high-level functional categories, users can visualize the distribution of annotated genes in those categories across one or several genomes, either as barcharts (Fig. 3.3A and B) or as heatmaps (Fig. 3.3C). This allows users to quickly visualize differences of functional capabilities between organisms.

To further characterize metabolic capacities, zDB implements a parser for the KEGG module definition language, which allows to assess the completeness of a metabolic module based on the KEGG orthologs present in a genome. Module completeness can be compared at the scope of a single KEGG module (Fig. 3.3C), or at the scope of categories or sub-categories (Fig. 3.3E). The results are directly linked to the species phylogeny, making it easy to notice patterns of metabolic capacities linked to specific clades.

The Swissprot homologs are listed both in the orthogroup home page and in the gene homepage (Fig. 3.2B).

3.2.3 Benchmarking

The results of the initial benchmark evaluating the duration of the optional analysis for the 179 genomes dataset is shown in Fig. 3.4c. Despite the use of Diamond instead of

blastp, searching for homologs in the RefSeq database took about 4 times longer than the other analyses. Similarly, searching for homologs in the Swissprot database took as long as performing the KEGG, COG and Pfam annotations together.

Generating a database with all the optional analysis (except the RefSeq homologs search) took 1.9h, 3.9h, 8.6h, 21.0h and 55.6h for the datasets with 10, 20, 40, 90 and 179 genomes respectively. The CPU time spent in the optional analyses seems to increase linearly with the number of genomes This was however not the case for the core analysis. In particular, the cost of orthology prediction increased faster than the other analyses and will likely be the limiting factor for larger datasets. This is expected due to the $O(n^2)$ complexity of the all-against-all genomes comparison performed by Orthofinder [166].

Finally, running the analysis pipeline on the 90 genomes dataset on a desktop machine took 71 hours to complete, showing that datasets with a hundred genomes can be processed on a mainstream computer.

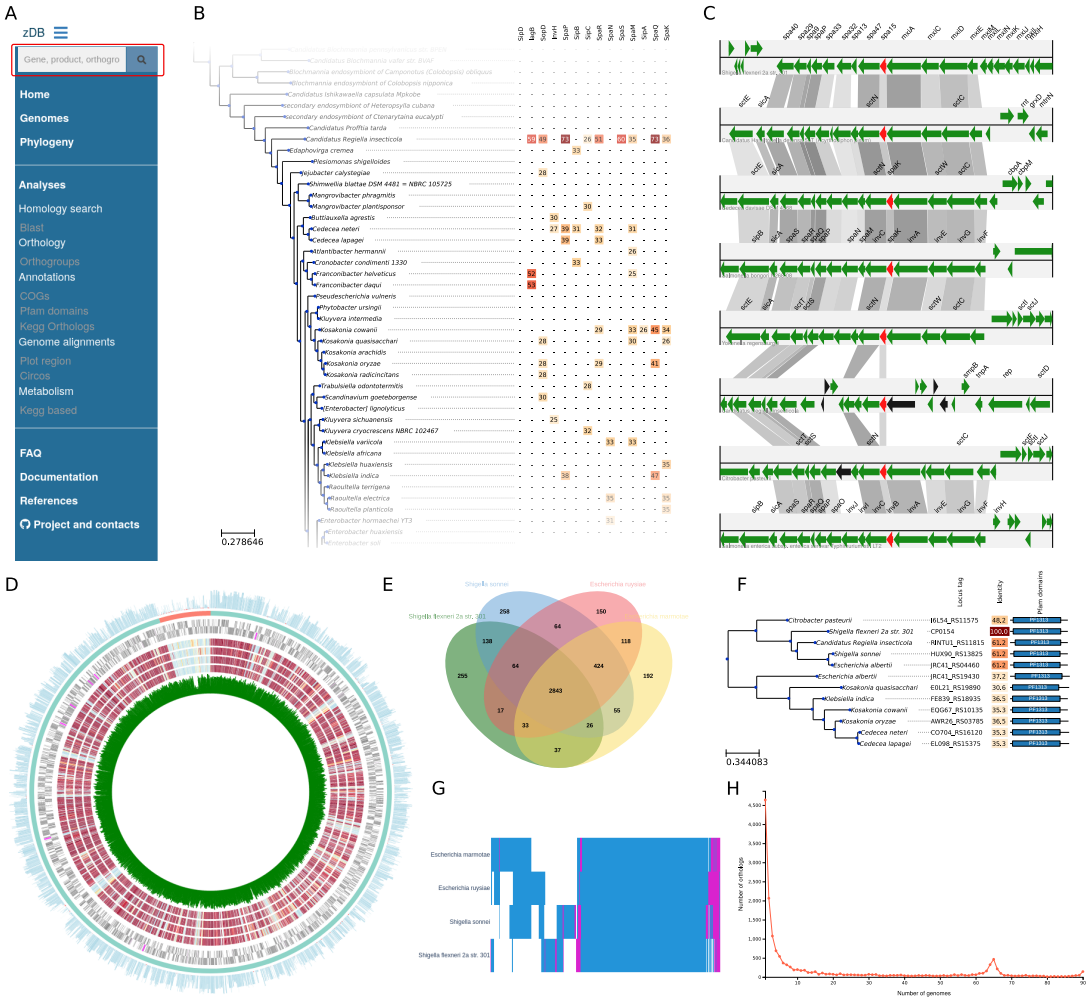

Figure 3.1: **(A)** zDB side panel listing all available analyses. The red box highlights the search bar. **(B)** Example of Blast search result. The whole dataset was searched for several proteins of the type III secretion system of *Salmonella* Typhimurium. The Blast hits are displayed as a heatmap of identities. **(C)** zDB can draw the genomic region of a specified gene of interest and compare it to the regions of its orthologs in other genomes (in red). Black arrow represent pseudogenes. The gray bars link orthologous genes together. Variations in color reflects amino-acid identity. **(D)** Visualization of the conservation of proteins of a reference genome (*Shigella flexneri*) in a set of chosen genomes (4 genomes of the *Escherichia* genus). The inner circle represent the GC content of each open reading frame in the reference genome. The next four circles represent the absence (in blue) or presence (in red) of homologs of proteins from the reference genome in the chosen genomes, with a color scale representing protein identity. The next two circles represent the localisation of the open reading frames in the forward and reverse strands of the reference genome. The last two circles represent the contigs in the reference assembly and a histogram of the number of occurrences of each protein of the reference genome in the chosen genomes. **(E)** Venn diagram illustrating the distribution of the orthologous groups in 4 genomes. **(F)** Example of a gene phylogeny, the identity column shows the identity relative to the CP0154 locus, as it was accessed through the page dedicated to this locus. The rightmost columns displays the domain architecture of the homologous proteins. **(G)** Heatmap of gene conservation in the same 4 genomes as in **(E)**; the pink bar represent genes present in more than one copy in a genome, the blue bars represents genes present in single copy. **(H)** Distribution of the orthologous groups in function of the number of genomes where they occur.

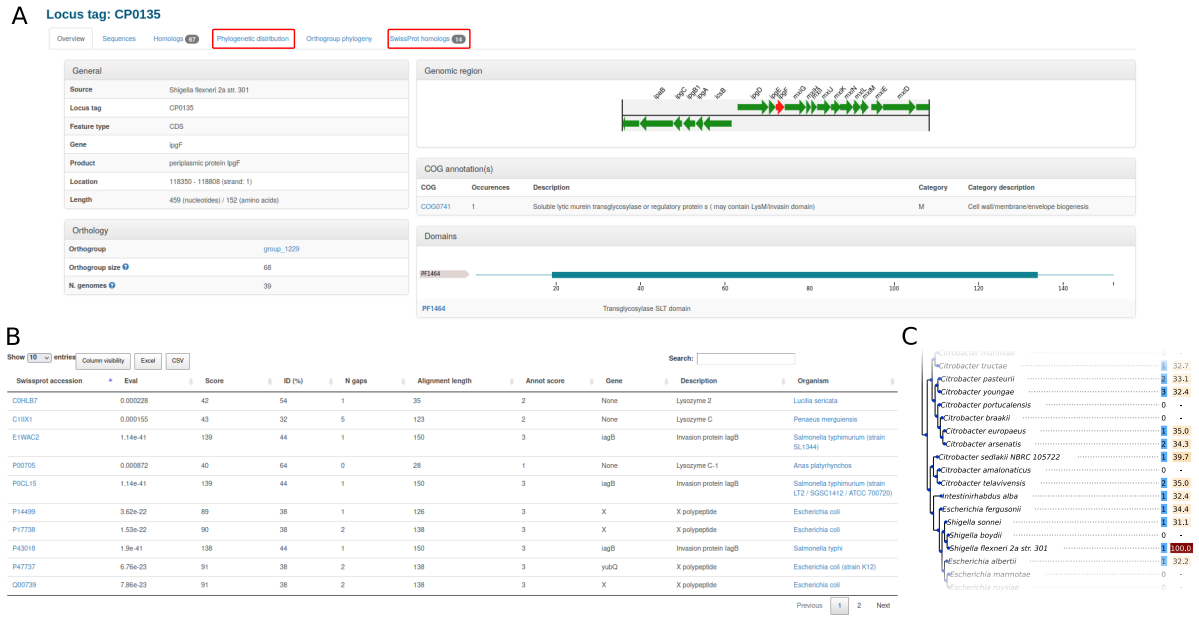

Figure 3.2: **(A)** Example of a gene summary page, with its genomic region, Pfam domains and COG annotation. The phylogenetic distribution and list of Swissprot homologs shown in **(B)** and **(C)** can be accessed from the highlighted tab. **(B)** The list of Swissprot homologs. **(C)** Part of the tree showing the phylogenetic conservation. The first column shows the number of homologs of the gene of interest in a given genome. The second column shows the amino-acid identity between the gene of interest and its closest homolog in a given genome.

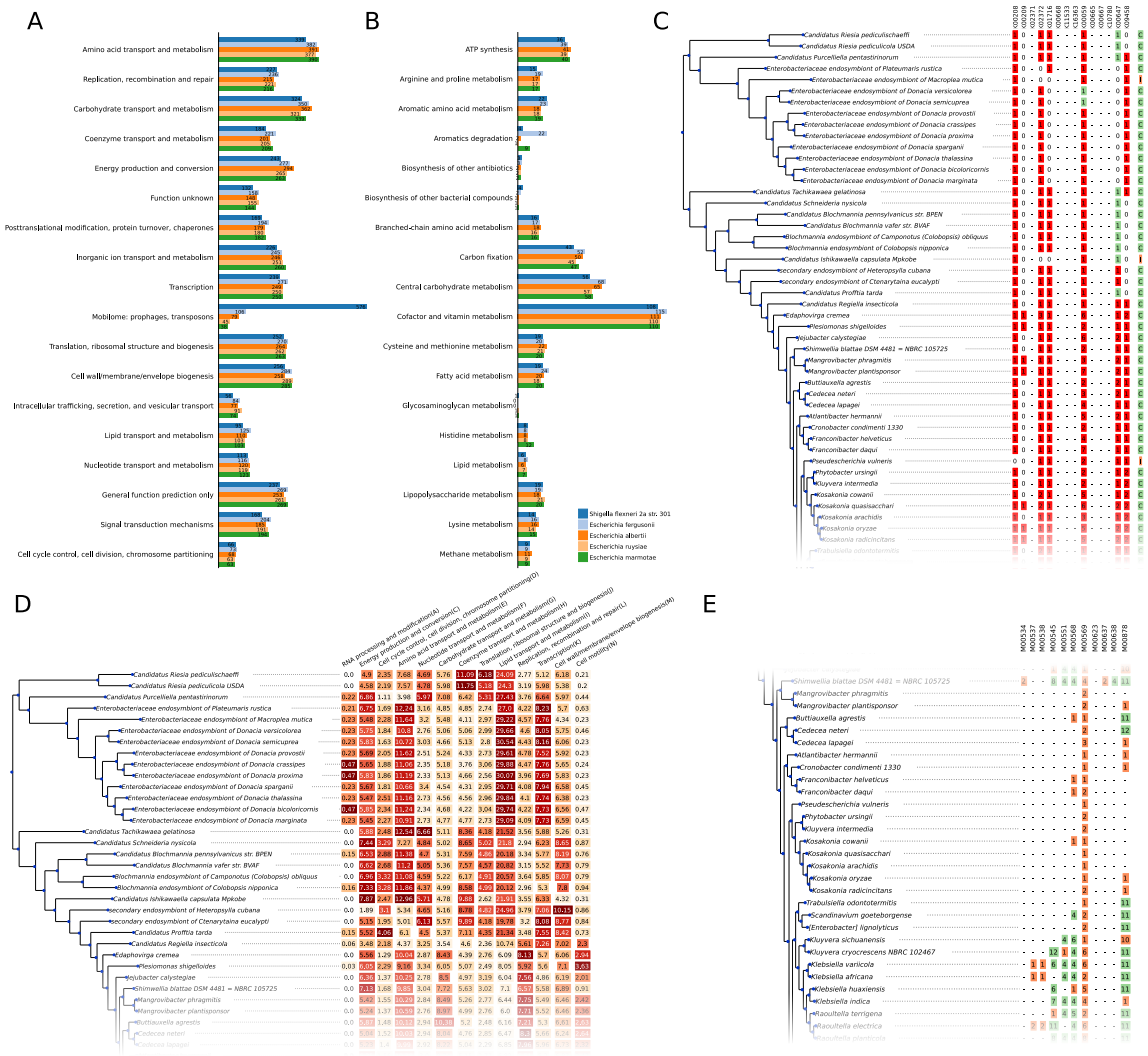

Figure 3.3: **(A)** and **(B)** Distribution of genes annotated with COG and KEGG orthologs in their functional categories for 4 chosen genomes. **(C)** Details of the completeness of KEGG module 83 (fatty acid biosynthesis, elongation). Red squares in the heatmap corresponds to the number of genes annotated as a given KEGG ortholog. Green square corresponds to a gene without KEGG annotation but in the same orthogroup as a gene annotated with a given KEGG ortholog. This may indicate a shared function and in such cases, the corresponding KEGG ortholog is considered as present when estimating the module completeness. The last column indicates module completeness, as determined by the module definition language. **(D)** Proportion of the genes in a genomes assigned to the different COG categories (the columns of some categories were removed for the sake of simplicity). **(E)** Green squares indicate a complete KEGG module, orange squares indicate an incomplete module, while the number indicates the number of genes annotated as KEGG orthologs of a given module.

3.3 Materials and Methods

3.3.1 Design and implementation

zDB is composed of two parts that can be run independently (Fig. 3.4a): an analysis pipeline which performs all the computationally intensive steps and stores the results in a Sqlite3 database, and a visualization platform which renders the results stored in the database in a graphical interface.

The analyses are separated in a set of core analyses focused on orthology prediction and phylogeny inference and a set of independent optional analyses, with a focus on functional annotation. To simplify the installation and make the analyses reproducible and scalable, all steps are run within singularity containers [167] under the control of the Nextflow workflow manager [168]. Nextflow indeed allows the analyses to be easily scaled from high-performance clusters to desktop machines, while containers guarantee the reproducibility and ease of installation by packaging the tools in controlled environments. After the completion of the analysis pipeline, zDB can export the results as a compressed archive for subsequent use. The ability to export the results was developed to facilitate sharing and to accommodate the fact that the analysis may have to be run and exported from a high performance computing (HPC) cluster, where long term storage might not be possible due to disk space constraints.

The visualization platform was implemented as a Django website using a pre-existing project [159] as a scaffold. The Django server can either be instantiated on a desktop computer, for local access, or on an Internet-facing computer, if the website is to be made public (internally within a network or externally). The results are rendered as lists, annotated phylogenetic trees and interactive plots. The phylogenetic trees are drawn as static images with the ete3 toolkit [169], while the interactive plots are generated by a collection of home-made scripts based on the d3.js framework and several libraries such as jvenn.js [170], Circos.js (https://github.com/nicgirault/circosJS)), BlasterJS [165] and

plotly (https://plotly.com)). All the plots drawn by the website can be downloaded as support vector graphics (.svg) images for subsequent use. Finally, users can also retrieve the results directly from the database via a Python [171] interface, if custom analyses are to be performed.

The code is available on GitHub (https://github.com/metagenlab/zDB) and zDB can be installed as a conda package [172].

3.3.2 Minimal analyses: quality control, orthology inference and core genome phylogeny

The minimal set of analyses includes quality control with CheckM [173], the generation of Blast [148] databases, orthology prediction and phylogeny inference. This core analysis should be sufficient for most applications, as it allows core/pangenome analysis, gene annotation based on homology and generates species and gene phylogenies.

zDB takes GenBank files as input and has currently been tested with the output of Prokka [155], PGAP [174] and Bakta [175]. As locally assembled genomes may have duplicated accessions or locus tags, zDB first checks their uniqueness and automatically generates new identifiers if necessary. Amino-acid and nucleotide sequences are then extracted from the GenBank files and used as input for subsequent analyses (Fig. 3.4a). Annotations such as gene names and protein products are also extracted from the GenBank files. Those annotations are indeed particularly valuable when reference genomes are analyzed together with draft genomes, as the annotations of the genes from a reference genome may hint at the function of their homologs in draft genomes.

Orthology is inferred using Orthofinder [166]. The sequences of orthologous proteins are aligned with MAFFT [176] and the alignments are used to infer phylogenetic trees for each orthogroup using FastTree [177]. In addition to gene phylogenies, zDB also generates a species tree with FastTree using the concatenated alignments of the single-copy core orthologs. As some assemblies may be incomplete, the condition that core orthologs must

be present in all genomes can be relaxed to allow missing genes. zDB generates Blast databases with both amino-acids and nucleotides sequences for each individual genomes and for the whole dataset. This allows users to search for sequences in a specific genome without the interference of better hits in other genomes, while still making it possible to perform global searches on the whole dataset.

Table 3.1: Reference databases used by zDB

Database	**Release**	**Size** [1]	**Search tool**
Swissprot [178]	Release 2021_04	86M	Blastp [148]
Refseq nr [2][179]	Release 210	34.9G	Diamond [180]
KEGG hmm profiles [181]	Release 03/2022	1.2G	Kofamscan [182]
CDD [183]	Release 3.19	4.0G	Rpsblast [148]
Pfam-A hmm profiles [184]	Release 35.0	279M	Pfam_scan [184]

[1] Refers to the volume of data to download

[2] As downloading the non-redundant RefSeq database is prone to failure, it is not automatically downloaded by zDB. A script to download and prepare the databases is installed with zDB, but has to be run manually.

3.3.3 Optional analyses: homology search, COG, KEGG and Pfam annotations

To complement the core analysis, zDB can perform optional analyses focused on function prediction. Optional analyses all take the proteins of the non-redundant pan-genome as input and include the assignment to the Cluster of Orthologs Genes (COG), mapping to the Kyoto Encyclopedia of Genes and Genomes (KEGG), prediction of Pfam protein domains and search for homologs in the SwissProt database. COG annotations offer clues regarding protein functions and allow their classification in broad functional categories. The assignment to COG [185] clusters is performed by rps-blast [148] searches using the position-specific score matrices of the NCBI Conserved Domain Database (CDD) [183]. KEGG annotations give insights into the metabolic capacities of the analyzed bacteria. The mapping to KEGG orthologs is performed by Kofamscan [182] using the prokaryotic profiles of the KEGG database [181]. As they can offer functional insights into otherwise unannotated proteins and as domain architecture conservation may be a valuable addition

to a gene phylogeny, Pfam protein annotations were also added in the optional analysis. The annotation is performed with the Pfam_scan [184] tool and the Pfam-A database. Finally, zDB can also perform homology search with blastp [148] against the manually curated entries of the SwissProt [178] database. The reference database used by zDB to perform those analysis are listed in Table 3.1. Of note, the core analyses can be performed without any reference database.

To screen for lateral gene transfers using a well-validated method [186], zDB can search the RefSeq database for homologs of proteins from the non-redundant pangenome. The search is performed by Diamond [180] to reduce the duration of the analysis. The proteins of every orthogroup and their best hits (the best 4 hits of every protein, by default) are then aligned with MAFFT, and the alignment is used by FastTree to infer a phylogenetic tree. As reference genomes downloaded from RefSeq may have been included in zDB input dataset, the best hits from genomes already present in the database are filtered out. If the database was populated with genomes of related bacteria, observing that a protein from a distant taxa clusters more closely than the other proteins from the same orthogroup may indeed indicate a lateral gene transfer.

3.3.4 Benchmarking

Although the analysis pipeline could process more genomes, the visualization platform is designed for datasets ranging from tens to hundreds of genomes. We therefore chose a representative dataset composed of the NCBI's 179 reference genomes of the *Enterobacteriaceae* family to benchmark the analysis pipeline. The genomes were downloaded as Genbank files from the NCBI (the references are shown in Table 3.2). We ran a first benchmark to measure the running time of the different optional analyses on the full dataset. As the search for RefSeq homologs proved to be prohibitively long (Fig. 3.4C), it was not included in the subsequent benchmark. The pipeline was then run on randomly generated subsets of the 179 genomes composed of 10, 20, 40, 90 or all genomes, all with a mean genome size of 3.8Mbp.

The performances of the pipeline were measured using Nextflow –with-report option. All analyses were run on an Ubuntu 18 server (112 Intel Xeon Platinum 8280 2.7GHz CPUs, equipped with 377GB of RAM memory), limiting parallelization to 20 simultaneous processes (with Nextflow cpus option) and total memory usage to 32GB (with Nextflow memory option), to mimic the computing power of a high-end desktop computer. We also tested the 90 genomes dataset on a desktop computer with 6 cores to have a better idea of the performances on a computer with more limited resources.

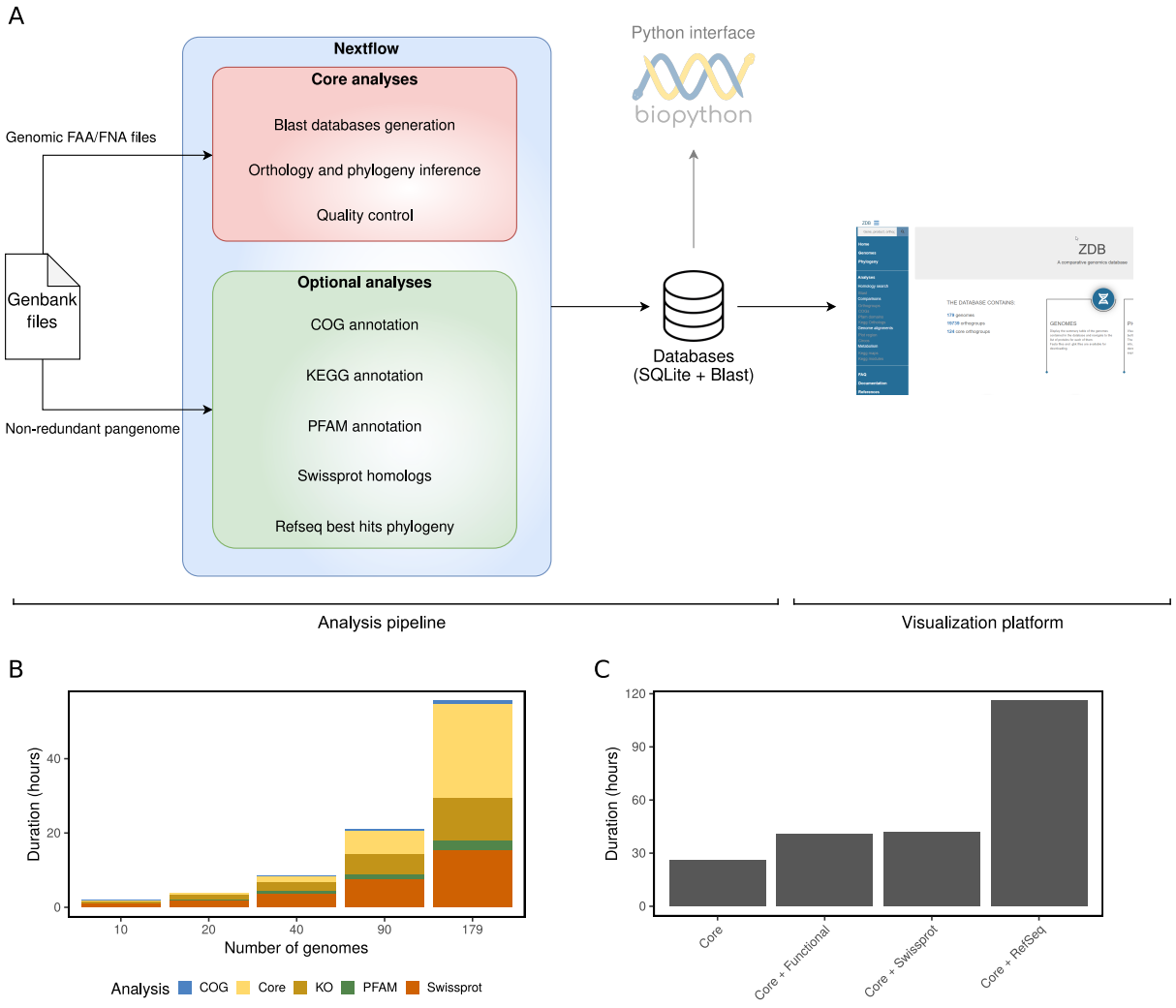

Figure 3.4: **(A)** Representation of the dataflow of zDB. The core analyses are performed for each dataset, while users can choose which optional analyses to perform. It is possible to only perform the core analyses. **(B)** Duration of the analysis split by type (core and optional) in total CPU time according to the number of genomes in a benchmarking. **(C)** Benchmarking of the different analysis types with the 179 genome dataset. Functional analysis include COG and KEGG orthology annotations and Pfam domain prediction. Swissprot and RefSeq represent searching for homologs in Swissprot and Refseq, respectively.

3.4 Discussion

To our knowledge, zDB is the first comparative genomics tool entirely run on user side to provide both a complete analysis pipeline and a visualization platform to explore the results. It was designed to require minimal typing on the command line and requires only three commands from installation to visualization of the results. As shown in the benchmarks, the analysis pipeline can process datasets of one hundred genomes in a matter of days on desktop computer, making dedicated computing infrastructures unnecessary for all but the largest datasets. The possibility to easily run Blast queries, to search for specific genes or to retrieve amino-acids and nucleotides sequences will make zDB useful for researchers more accustomed to lab work, while features such as core and pan-genome analysis and genomic regions comparisons will be useful to seasoned bioinformaticians. Altogether, this makes of zDB a tool easy to use and install for a wide variety of applications such as setting up a public database for an organism of interest, genome browsing or for the characterization of newly sequenced genomes.

Future directions

Large phylogenies are cumbersome to visualize as they may not entirely fit in a computer screen. To alleviate this, we plan to replace the ete3 drawing engine by a custom Javascript library to draw interactive phylogenetic trees allowing the user to collapse and expand branches. As of now, the addition or removal of genomes from an existing database is not possible and requires to repeat all the analysis on the modified dataset. We therefore plan to implement the possibility to add or remove genomes, which will allow users to incrementally improve a database without having to repeat the analyses. Finally, we plan to extend the set of optional analyses with additional annotations such as the prediction of antibiotic resistance genes, protein transmembrane domains and signal peptides. As the project is entirely open source and has been designed to be easily extended, the community is also welcome to implement any new features.

3.5 Supplementary materials

Table 3.2: Genomes included in the test dataset

Scientific Name	**Accession**
Escherichia coli str. K-12 substr. MG1655	GCF_000005845.2
Shigella flexneri 2a str. 301	GCF_000006925.2
Salmonella enterica subsp. *enterica* serovar Typhimurium str. LT2	GCF_000006945.2
Escherichia coli O157:H7 str. Sakai	GCF_000008865.2
Candidatus Blochmannia pennsylvanicus str. BPEN	GCF_000011745.1
Citrobacter koseri ATCC BAA-895	GCF_000018045.1
Candidatus Hamiltonella defensa 5AT (*Acyrthosiphon pisum*)	GCF_000021705.1
Citrobacter rodentium ICC168	GCF_000027085.1
Candidatus Riesia pediculicola USDA	GCF_000093065.1
Candidatus Blochmannia vafer str. BVAF	GCF_000185985.2
Enterobacter soli	GCF_000224675.1
Klebsiella pneumoniae subsp. *pneumoniae* HS11286	GCF_000240185.1
Shimwellia blattae DSM 4481 = NBRC 105725	GCF_000262305.1
Secondary endosymbiont of *Ctenarytaina eucalypti*	GCF_000287335.1
Secondary endosymbiont of *Heteropsylla cubana*	GCF_000287355.1
Enterobacter hormaechei YT3	GCF_000328885.1
Candidatus Blochmannia chromaiodes str. 640	GCF_000331065.1
Candidatus Moranella endobia PCVAL	GCF_000364725.1
Cedecea davisae DSM 4568	GCF_000412335.2
Salmonella bongori N268-08	GCF_000439255.1
Siccibacter colletis	GCF_000696575.1
Trabulsiella guamensis ATCC 49490	GCF_000734965.1
Buttiauxella noackiae	GCF_000737905.1
Citrobacter sedlakii NBRC 105722	GCF_000759835.1
Enterobacter cloacae	GCF_000770155.1
Raoultella planticola	GCF_000783935.2
Enterobacter asburiae	GCF_000799205.1
Candidatus Ishikawaella capsulata Mpkobe	GCF_000828515.1
Candidatus Tachikawaea gelatinosa	GCF_000828815.1
Blochmannia endosymbiont of *Polyrhachis* (*Hedomyrma*) *turneri*	GCF_000973505.1
Blochmannia endosymbiont of *Camponotus* (*Colobopsis*) *obliquus*	GCF_000973545.1
Phytobacter ursingii	GCF_001022135.1
Cronobacter universalis NCTC 9529	GCF_001277175.1
Cronobacter muytjensii ATCC 51329	GCF_001277195.1
Cronobacter malonaticus LMG 23826	GCF_001277215.2
Cronobacter dublinensis subsp. *dublinensis* LMG 23823	GCF_001277235.1
Cronobacter condimenti 1330	GCF_001277255.1
Trabulsiella odontotermitis	GCF_001297845.1
Enterobacter lignolyticus	GCF_001461805.1
Type-D symbiont of *Plautia stali*	GCF_001485335.1
Type-E symbiont of *Plautia stali*	GCF_001485355.1

Table 3.2: Genomes included in the test dataset

Scientific Name	**Accession**
Leclercia adecarboxylata	GCF_001518835.1
Citrobacter amalonaticus	GCF_001558935.2
Kluyvera cryocrescens NBRC 102467	GCF_001571285.1
Buttiauxella gaviniae ATCC 51604	GCF_001654835.1
Buttiauxella ferragutiae ATCC 51602	GCF_001654915.1
Buttiauxella brennerae ATCC 51605	GCF_001654925.1
Mangrovibacter phragmitis	GCF_001655675.1
Kosakonia oryzae	GCF_001658025.2
Kluyvera georgiana	GCF_001682915.1
Kosakonia sacchari	GCF_001683395.1
Enterobacter roggenkampii	GCF_001729805.1
Enterobacter ludwigii	GCF_001750725.1
Shigella boydii	GCF_001905915.1
Shigella dysenteriae	GCF_001932995.2
Enterobacter chengduensis	GCF_001984825.2
Candidatus Riesia pediculischaeffi	GCF_002073895.1
Cedecea neteri	GCF_002393445.1
Superficieibacter electus	GCF_002915575.1
Klebsiella oxytoca	GCF_002984395.1
Pluralibacter gergoviae	GCF_003019925.1
Mangrovibacter plantisponsor	GCF_003182475.1
Franconibacter helveticus	GCF_003207695.1
Klebsiella huaxiensis	GCF_003261575.2
Edaphovirga cremea	GCF_003332275.1
Candidatus Purcelliella pentastirinorum	GCF_003391335.1
Citrobacter gillenii	GCF_003429605.1
Cronobacter sakazakii	GCF_003516125.1
Enterobacter chuandaensis	GCF_003594915.1
Buttiauxella izardii	GCF_003601925.1
Enterobacter mori	GCF_003606205.2
Yokenella regensburgei	GCF_003634235.1
Citrobacter europaeus	GCF_003795375.1
Citrobacter freundii	GCF_003812345.1
Buttiauxella warmboldiae	GCF_003818135.1
Scandinavium goeteborgense	GCF_003935895.2
Enterobacter huaxiensis	GCF_003944645.1
Enterobacter quasiroggenkampii	GCF_003964805.1
Kosakonia cowanii	GCF_004089895.1
Siccibacter turicensis	GCF_004168465.1
Enterobacter wuhouensis	GCF_004331265.1
Enterobacter quasihormaechei	GCF_004331385.1
Kosakonia quasisacchari	GCF_004331415.1
Citrobacter arsenatis	GCF_004353845.1
Citrobacter tructae	GCF_004684345.1
Citrobacter murliniae	GCF_004801125.1
Jejubacter calystegiae	GCF_005671395.1

Table 3.2: Genomes included in the test dataset

Scientific Name	**Accession**
Klebsiella indica	GCF_005860775.1
Raoultella electrica	GCF_006711645.1
Enterobacter asburiae	GCF_007035645.1
Atlantibacter subterranea	GCF_007570865.1
Klebsiella aerogenes	GCF_007632255.1
Atlantibacter hermannii	GCF_008064855.1
Kosakonia radicincitans	GCF_008330085.1
Enterobacter vonholyi	GCF_008364555.1
Citrobacter portucalensis	GCF_008693605.1
Enterobacter sichuanensis	GCF_009036245.1
Enterobacter oligotrophicus	GCF_009176645.1
Kosakonia arachidis	GCF_009363135.1
Citrobacter telavivensis	GCF_009363175.1
Citrobacter braakii	GCF_009648935.1
Intestinirhabdus alba	GCF_009711095.1
Blochmannia endosymbiont of *Camponotus nipponensis*	GCF_009827135.1
Cronobacter turicensis	GCF_011605535.1
Cedecea colo	GCF_011808225.1
Raoultella terrigena	GCF_012029655.1
Plesiomonas shigelloides	GCF_012273295.1
Enterobacteriaceae endosymbiont of *Plateumaris sericea*	GCF_012562605.1
Enterobacteriaceae endosymbiont of *Plateumaris pusilla*	GCF_012562765.1
Enterobacteriaceae endosymbiont of *Plateumaris rustica*	GCF_012562965.1
Enterobacteriaceae endosymbiont of *Plateumaris braccata*	GCF_012563325.1
Enterobacteriaceae endosymbiont of *Donacia fulgens*	GCF_012567545.1
Enterobacteriaceae endosymbiont of *Donacia marginata*	GCF_012567685.1
Enterobacteriaceae endosymbiont of *Donacia bicoloricornis*	GCF_012567955.1
Enterobacteriaceae endosymbiont of *Donacia thalassina*	GCF_012568245.1
Enterobacteriaceae endosymbiont of *Donacia vulgaris*	GCF_012568445.1
Enterobacteriaceae endosymbiont of *Donacia simplex*	GCF_012568645.1
Enterobacteriaceae endosymbiont of *Donacia cincticornis*	GCF_012568845.1
Enterobacteriaceae endosymbiont of *Donacia sparganii*	GCF_012569045.1
Enterobacteriaceae endosymbiont of *Donacia proxima*	GCF_012569285.1
Enterobacteriaceae endosymbiont of *Donacia crassipes*	GCF_012569785.1
Enterobacteriaceae endosymbiont of *Donacia cinerea*	GCF_012569925.1
Enterobacteriaceae endosymbiont of *Donacia provostii*	GCF_012570145.1
Enterobacteriaceae endosymbiont of *Donacia clavipes*	GCF_012570365.1
Enterobacteriaceae endosymbiont of *Donacia semicuprea*	GCF_012570535.1
Enterobacteriaceae endosymbiont of *Donacia dentata*	GCF_012570745.1
Enterobacteriaceae endosymbiont of *Donacia versicolorea*	GCF_012570965.1
Enterobacteriaceae endosymbiont of *Donacia tomentosa*	GCF_012571135.1
Enterobacteriaceae endosymbiont of *Macroplea mutica*	GCF_012571345.1
Enterobacteriaceae endosymbiont of *Macroplea appendiculata*	GCF_012571605.1
Enterobacteriaceae endosymbiont of *Neohaemonia nigricornis*	GCF_012571795.1
Phytobacter diazotrophicus	GCF_012923785.1
Buttiauxella agrestis	GCF_013234275.1

Table 3.2: Genomes included in the test dataset

Scientific Name	**Accession**
Klebsiella variicola	GCF_013305245.1
Candidatus Regiella insecticola	GCF_013373955.1
Shigella sonnei	GCF_013374815.1
Raoultella ornithinolytica	GCF_013457875.1
Escherichia fergusonii	GCF_013892435.1
Kluyvera sichuanensis	GCF_014218705.1
Franconibacter daqui	GCF_014644275.1
Pseudocitrobacter faecalis	GCF_014653055.1
Blochmannia endosymbiont of *Colobopsis nipponica*	GCF_014857065.1
Kluyvera ascorbata	GCF_015099135.1
Enterobacter bugandensis	GCF_015137655.1
Klebsiella michiganensis	GCF_015139575.1
Kosakonia pseudosacchari	GCF_015167415.1
Lelliottia nimipressuralis	GCF_015319205.1
Shimwellia pseudoproteus	GCF_016415625.1
Klebsiella quasipneumoniae	GCF_016415705.1
Citrobacter cronae	GCF_016502155.1
Citrobacter werkmanii	GCF_016505055.1
Lelliottia aquatilis	GCF_016771845.1
Escherichia albertii	GCF_016904755.1
Enterobacter dykesii	GCF_018597265.1
Enterobacter quasimori	GCF_018597345.1
Citrobacter youngae	GCF_018883525.1
Citrobacter pasteurii	GCF_019047765.1
Lelliottia amnigena	GCF_019355955.1
Enterobacter cancerogenus	GCF_019665745.1
Citrobacter farmeri	GCF_019803045.1
Escherichia ruysiae	GCF_019840805.1
Candidatus Schneideria nysicola	GCF_019923545.1
Enterobacter timonensis	GCF_900021175.1
Candidatus Doolittlea endobia	GCF_900039485.1
Candidatus Hoaglandella endobia	GCF_900044015.1
Candidatus Gullanella endobia	GCF_900048035.1
Kosakonia oryziphila	GCF_900094795.1
Kosakonia oryzendophytica	GCF_900094925.1
Cedecea lapagei	GCF_900635955.1
Escherichia marmotae	GCF_900637015.1
Klebsiella quasivariicola	GCF_900978035.1
Klebsiella africana	GCF_900978845.1
Klebsiella spallanzanii	GCF_901563875.1
Klebsiella pasteurii	GCF_902158635.1
Klebsiella grimontii	GCF_902164675.1
Pseudescherichia vulneris	GCF_902164725.1
Metakosakonia massiliensis	GCF_902374985.1
Kluyvera intermedia	GCF_902387965.1
Candidatus Profftia tarda	GCF_904848675.1

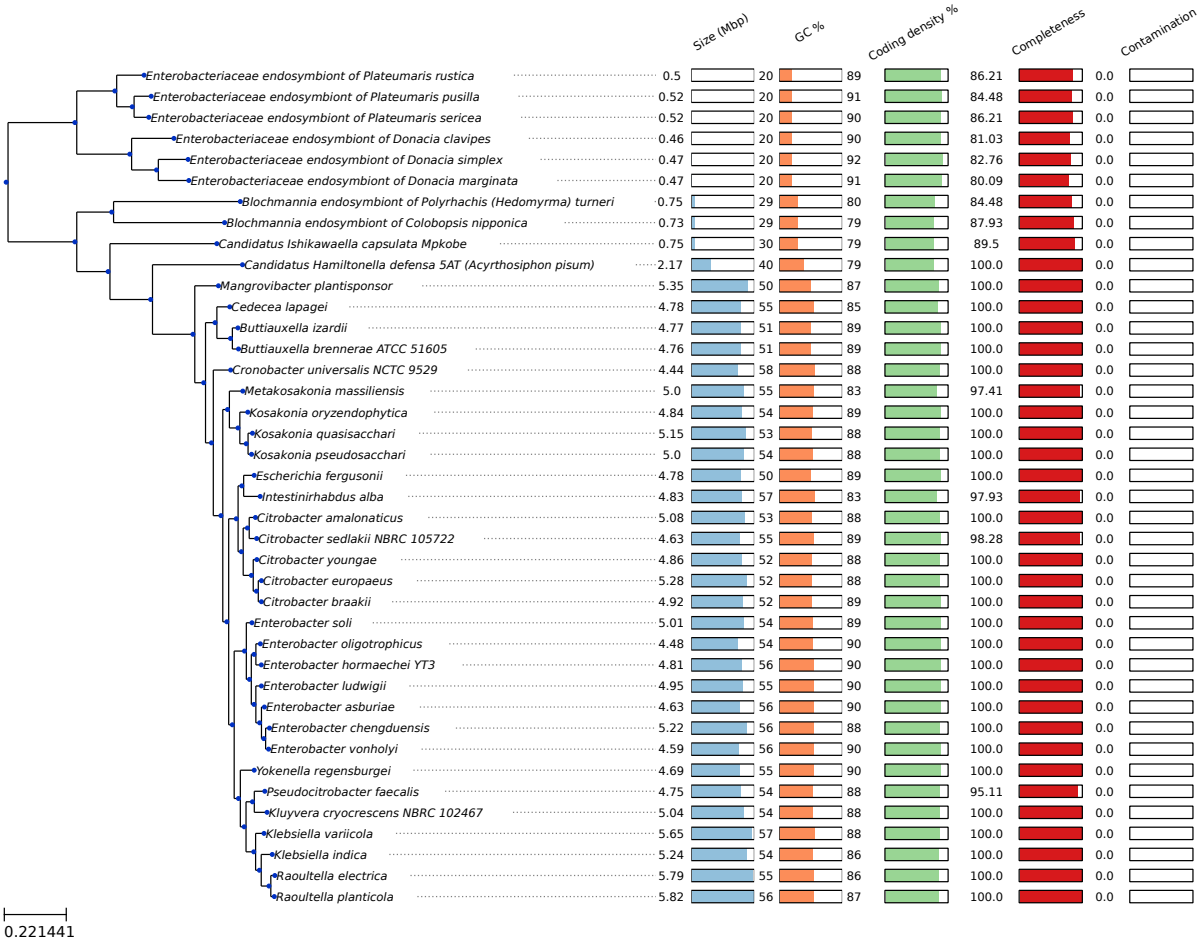

Figure 3.5: Annotated species tree of the 40 genomes dataset, accessed via the "Phylogeny" tab of the web application. The tree was generated by the maximum likelihood method, based on the concatenated alignment of 186 shared single copy orthologs. The tree is annotated with the genome size, the GC content, the coding density (the ratio of the size of the predicted coding regions over the genome size) and estimations of the assemblies completeness and contamination by checkm [173].

Bibliography

148. Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. Basic local alignment search tool. Journal of molecular biology 1990;215:403–10.

153. Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 1995;269:496–512.

154. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, and Karsch-Mizrachi I. GenBank. Nucleic acids research 2019;47:D94–D99.

155. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30: 2068–2069.

156. Petit III RA and Read TD. Bactopia: a flexible pipeline for complete analysis of bacterial genomes. Msystems 2020;5:e00190–20.

157. Quijada NM, Rodríguez-Lázaro D, Eiros JM, and Hernández M. TORMES: an automated pipeline for whole bacterial genome analysis. Bioinformatics 2019;35:4207–12.

158. Schwengers O, Hoek A, Fritzenwanker M, et al. ASA3P: an automatic and scalable pipeline for the assembly, annotation and higher-level analysis of closely related bacterial isolates. PLoS computational biology 2020;16:e1007134.

159. Pillonel T, Tagini F, Bertelli C, and Greub G. ChlamDB: a comparative genomics database of the phylum Chlamydiae and other members of the Planctomycetes-Verrucomicrobiae-Chlamydiae superphylum. Nucleic Acids Research 2020;48:D526–D534.

160. Amos B, Aurrecoechea C, Barba M, et al. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. Nucleic Acids Research 2022;50:D898–D911.

161. Dieckmann MA, Beyvers S, Nkouamedjo-Fankep RC, et al. EDGAR3. 0: comparative genomics and phylogenomics on a scalable infrastructure. Nucleic Acids Research 2021;49:W185–W192.

162. Deng Z, Botas J, Cantalapiedra CP, Hernández-Plaza A, Burguet-Castell J, and Huerta-Cepas J. PhyloCloud: an online platform for making sense of phylogenomic data. Nucleic Acids Research 2022.

163. Grover JW, Bomhoff M, Davey S, Gregory BD, Mosher RA, and Lyons E. CoGe LoadExp+: A web-based suite that integrates next-generation sequencing data analysis workflows and visualization. Plant Direct 2017;1.

164. Vallenet D, Calteau A, Dubois M, et al. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. Nucleic Acids Research 2020;48:D579–D589.

165. Blanco-Míguez A, Fdez-Riverola F, Sánchez B, and Lourenço A. BlasterJS: A novel interactive JavaScript visualisation component for BLAST alignment results. PLoS One 2018;13:e0205286.

166. Emms DM and Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome biology 2019;20:1–14.

167. Kurtzer GM, Sochat V, and Bauer MW. Singularity: Scientific containers for mobility of compute. PloS one 2017;12:e0177459.

168. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, and Notredame C. Nextflow enables reproducible computational workflows. Nature biotechnology 2017;35:316–9.

169. Huerta-Cepas J, Serra F, and Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Molecular biology and evolution 2016;33:1635–8.

170. Bardou P, Mariette J, Escudié F, Djemiel C, and Klopp C. jvenn: an interactive Venn diagram viewer. BMC bioinformatics 2014;15:1–7.

172. Grüning B, Dale R, Sjödin A, et al. Bioconda: sustainable and comprehensive software distribution for the life sciences. Nature methods 2018;15:475–6.

173. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, and Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome research 2015;25:1043–55.

174. Tatusova T, DiCuccio M, Badretdin A, et al. NCBI prokaryotic genome annotation pipeline. Nucleic acids research 2016;44:6614–24.

175. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, and Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microbial genomics 2021;7.

176. Katoh K and Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 2013;30:772–80.

177. Price MN, Dehal PS, and Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one 2010;5:e9490.

178. Bairoch A and Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic acids research 2000;28:45–8.

179. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic acids research 2016;44:D733–D745.

180. Buchfink B, Reuter K, and Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature methods 2021;18:366–8.

181. Kanehisa M, Sato Y, Kawashima M, Furumichi M, and Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic acids research 2016;44:D457–D462.

182. Aramaki T, Blanc-Mathieu R, Endo H, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 2020;36:2251–2.

183. Marchler-Bauer A, Derbyshire MK, Gonzales NR, et al. CDD: NCBI's conserved domain database. Nucleic acids research 2015;43:D222–D226.

184. Mistry J, Chuguransky S, Williams L, et al. Pfam: The protein families database in 2021. Nucleic acids research 2021;49:D412–D419.

185. Tatusov RL, Koonin EV, and Lipman DJ. A genomic perspective on protein families. Science 1997;278:631–7.

186. Marcet-Houben M and Gabaldón T. Acquisition of prokaryotic genes by fungal genomes. Trends in Genetics 2010;26:5–8.

4 Comparative genomics of arthropod-borne chlamydiae

Authors: Bastian Marquis, Trestan Pillonel, Gilbert Greub

Authors: Bastian Marquis assembled the genomes and performed the analyses

Status: Work in progress

4.1 Introduction

Since the publication of the genome of *C. trachomatis* in 1998 [48], numerous other members of the *Chlamydiales* could be sequenced and while this sequencing effort was initially focused on the *Chlamydiaceae* family [187], it quickly extended to CLO [37, 188, 189] and, as the cost of sequencing decreased, became an efficient way to study the biology, diversity and evolution of those elusive intracellular obligate bacteria. Shotgun sequencing of environmental samples indeed highlighted the diversity of the *Chlamydiales* in terms of gene content [22], while the analysis of the genome of *C. trachomatis* led to discoveries such as the presence of a complete peptidoglycan synthesis pathway or the reduced metabolic repertoire of this species [48], contrasting with the metabolic capacities of CLOs [28].

The genome of three rhabdochlamydial species could be sequenced so far: the metagenome of *R. helvetica* was assembled from a pool of *Ixodes ricinus* ticks collected in the wild [38], while the same approach using a different sequencing technology allowed to obtain the complete genome of *R. oedothoracis*, a spider-borne rhabdochlamydia [69, 75]. Finally, the complete genome of *R. porcellionis* has also recently been sequenced [75]. The analysis of those genomes revealed the presence of a large number of transposases in *R. oedothoracis* and *R. helvetica* and their absence from the smaller genome of *R. porcellionis* [38, 75]. The proliferation of transposases in the former two species was interpreted as the initial steps of reductive genome evolution [75], the latest stage of which can be observed in the streamlined genome of *R. porcellionis*. The reductive evolution of those genomes was suggested to be a consequence of the recent adaptation of a protist-borne chlamydiae as an arthropod symbiont [75]. The nature of the symbiontic relationship between those bacteria and their host however remains unknown. In line with the pathogenic role of rhabdochlamydia observed in woodlouse [58], the gene content of *R. helvetica* does not suggest a mutualistic relationship with its tick host, as this species appears to lack the metabolic capacities of nutritional symbionts [38]. However, another type of symbiosis,

such as defensive symbiosis, cannot be excluded.

To obtain a better understanding of the evolution and biology of the *Rhabdochlamydiaceae*, we sequenced and described the genome of 7 additional arthropod-borne species of chlamydiae, 5 of which belong to the *Rhabdochlamydiaceae* family.

4.2 Methods

Sequencing of tick samples

Pools of ticks collected in the study by Pilloux et al [66] were selected for sequencing based on the number of genome copies as measured by the pan-*Chlamydiales* qPCR. Additionally, a culture of *R. porcellionis* maintained in Sf9 cells was also included in the project. For both tick samples and the *R. porcellionis* culture, genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega, Dübendorf, Switzerland). The libraries were prepared with the Nextera XT library kit (Illumina, San Diego, CA) and sequenced on an Illumina MiSeq platform (Illumina).

Reads obtained from a separate sequencing project from Australia were also included in our analysis (courtesy of Dr. Polkinghorne). Briefly, this dataset contains reads sequenced from a pool of ticks collected from marsupials in a previous study [68]. Genomic DNA extracted from the pool underwent host methylated DNA depletion using the NEBNext Microbiome DNA Enrichment kit (New England BioLabs, USA). The enriched DNA was then purified by ethanol precipitation before being subject to multiple displacement amplification using the Qiagen Repli-G mini kit (Qiagen, Germany). The library was sequenced on an Illumina HiSeq platform to produce 125bp paired-end reads at the Australian Genome Research facility.

Sequence read archive query

We ran a SQL query on the SRA archive [190, 191] to identify sequencing projects contaminated with chlamydial reads. The query is shown in Listing 4.5 and takes advantage of the availability of the results of the taxonomic analysis performed by the NCBI on each read archive [192]. The query excludes sequencing projects of *Homo sapiens* and *Mus musculus* as those returned too many results. We further narrowed the results to sequencing projects of the Arthropoda phylum.

Assembly

Illumina reads were trimmed using fastp [193]. Reads from ticks and from the *R. porcellionis* culture were mapped to their host genome (GCA_000973045.2 and GCF_011064685.1, respectively) using bwa [194] and only the unmapped reads were kept for further analysis. The Illumina datasets obtained from the SRA query were also trimmed with fastp, but were not subjected to the filtering based on their host genome. The reads were assembled using spades (`--meta`) [195] and the resulting assembly graphs were visualized with Bandage [196]. A taxonomy was assigned to each contigs by CAT/BAT [197]. All contigs classified as chlamydiae were included in the final assemblies. In addition, for assemblies displaying a well-connect assembly graph, contigs not classified as chlamydiae were also included if they were connected to the chlamydial graph and had a sequencing depth in the range of that of chlamydial contigs.

The *Latrodectus elegans* dataset contained both Nanopore and Illumina reads. To reduce the complexity of the dataset, the Nanopore reads were mapped to the genome of *Latrodectus hesperus* (GCA_000697925.2) with minimap2 [198]. The reads that did not map to the reference genome were assembled using flye (`--meta`) [199]. Chlamydial contig were then identified with CAT/BAT and the assembly graph was visualized with Bandage [196]. The assembly was then corrected with three rounds of pilon (`--fix-all`) using the Illumina reads [200].

Plasmid conformation

R. porcellionis (DSM 27522) was maintained in Sf9 cells grown in Grace (Gibco, Thermo Fisher Scientific, Waltham, USA) medium complemented with 10% FCS. Infected cells were passaged weekly and fresh cells were added when necessary to compensate for the lysis induced by the bacteria. Genomic DNA was extracted from the culture using the Wizard SV Genomic DNA Purification Kit (Promega, Dübendorf, Switzerland) following the manufacturer protocol. To resolve the conformation of the plasmid, PCR were performed in a volumne of 50 µL on 5 µL of genomic DNA with 0.5 µL of Phusion hot start II DNA polymerase (Thermo Fisher Scientific, Waltham, USA), 1 µmol L^{-1} of reverse (conf_R1 or conf_R2 in Table 4.1) and forward primers (conf_F1 or conf_F2), 10 µL 5X Phusion Green HF buffer (Thermo Fisher Scientific, Waltham, USA), 28.5 µL of nuclease-free water and 1 µL of 10 mmol L^{-1} dNTPs (Promega, Dübendorf, Switzerland). The cycling conditions were identical for all PCRs and started with 30 s of denaturation at 98 °C, followed by 35 cycles of 20 s of annealing at 60 °C, 30 s of extension at 72 °C and 10 s of denaturation at 98 °C, with a 10 min final extension step at 72 °C. The amplicons were sent for Sanger sequencing at Microsynth (Microsynth AG, Balgach, Switzerland).

To measure the frequency of the two plasmid conformations during the infectious cycle, we plated 1×10^5 Sf9 cells in a 24-wells plate. *R. porcellionis* were collected from infected Sf9 cells using a freeze-thaw cycle, followed by a filtration through a 5 µm filter. The filtrate was then used to infect the cells at an MOI of 0.1-1. The infected cells were centrifuged at 130 g for 15 minutes, followed by an incubation of 30 minutes at 28 °C. The supernatant was then replaced with fresh medium to remove non-internalized bacteria. Samples were taken every other day for ten days and plasmid DNA was extracted using the NucleoSpin Plasmid kit (Machery-Nagel, Oensingen, Switzerland) following the manufacturer's instructions for low copy number plasmids. The frequency of the two conformations was measured by a SYBR green qPCR. The design of the primers is summarized in Fig. 4.2B. The qPCR were performed in a volume of 25 µL with 5 µL of DNA

Table 4.1: Primers used to resolve the conformation of the plasmid and to measure the frequency of the conformations.

Primer	Sequence
conf_F1	5'-CATCTATCAACACTGTTTAATTCCC-3'
conf_F2	5'-GATGGGAAGAAAAATCCGGTAC-3'
conf_R1	5'-CTCAATCCAATCTCGGCCG-3'
conf_R2/R2[1]	5'-GACTAAATAGCTGTTAGGTCCGG-3'
F	5'-ACGCAAATAAAGGCTGATCCTG-3'
F2	5'-GGAAGCACCTACCTTTCTTGAG-3'
R1	5'-GGCCGATTTTTCTTTACCAACG-3'

[1] used in both the PCR and the qPCR

template, 10 µL of iTaq universal SYBR green supermix (Bio-Rad, Cressier, Switzerland), 300 nmol L^{-1} of reverse and forward primers (R1, R2 and F, F2 from Table 4.1, respectively) and 3.8 µL of nuclease-free water on a QuantStudio3 real-time PCR system (Applied Biosystems, Thermo Fisher Scientific, Waltham, USA). The cycling conditions were identical for all reactions and started with an initial denaturation step of 10 min at 95 °C followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C.

Annotation and comparative genomics

The assemblies were annotated using the NCBI Prokaryotic Genome Annotation Pipeline [174]. zDB (Chapter 3) was then used to compare the assemblies to a selection of reference genomes of the PVC superphylum (Supp. Table 4.5). Pairwise ANI were calculated using pyani [201]. The pairwise comparison of nine phylogenetically informative markers was also computed to classify the assemblies at the family, genus and species levels [59]. The phylogenetic trees were visualized with zdb and iTOL [202].

4.3 Results

4.3.1 Datasets

Following the successful assembly of *R. helvetica* from a tick [38], we decided to assemble a higher diversity of *Rhabdochlamydiaceae* genomes by following the same approach on

other tick samples from the study by Pilloux et al [66]. Nine pools containing a high number of copies of *Rhabdochlamydiaceae* were selected for sequencing (Suppl. Table 4.2). In addition, we also included a dataset sequenced from a pool of ticks retrieved from marsupials in Australia and previously showed to be positive for *Rhabdochlamydiaceae* [68]. Finally, we sequenced the genome of *R. porcellionis* from a culture in Sf9. This species is indeed the only culturable member of the *Rhabdochlamydiaceae* and its genome had not yet been published when we started this project.

In an attempt to assemble genomes of rhabdochlamydiae from a wider diversity of hosts, we queried the SRA for sequencing projects contaminated with chlamydial reads. The rationale behind the query was that scientists performing sequencing projects might inadvertently sequence organisms infected with chlamydiae. The chlamydial reads would therefore contaminate the dataset and could be used to assemble the genomes of yet unknown species. The query therefore lists all the sequencing projects unrelated to chlamydia, but still containing chlamydial reads (Suppl. Listing 4.5). Our query was run on the 8$^{\text{th}}$ of June 2022 and returned 45 100 read archives containing chlamydiae reads, only 384 of which included metadata on the sequenced organism (Supp. Fig. 4.4). We further narrowed our selection to the 17 read archives from arthropods sequencing projects (Supp. Table 4.3). The reads of three of those projects had already been used to assemble the genome of *Rhabdochlamydia oedothoracis* and were not included in our analysis [75], leaving a total of 25 datasets.

4.3.2 Assemblies

We could assemble 20 genomes from those 25 datasets. Four read archives did not contain enough chlamydial reads to obtain complete genomes (SRR15257431, SRR5562871, ERR4790642, SRR4999935). Likewise, the T3186 tick sample contained a majority of tick reads (Suppl. Table 4.2) and even though a complete assembly could be obtained, it was not considered for further analysis due its low coverage (mean coverage of 15) and fragmented assembly graph (Suppl. Fig. 4.5D). Remarkably, it was possible to assemble

a close genome from the *Latrodectus elegans* hybrid dataset (Suppl. Fig. 4.5A). The quality of the other assemblies varied, with some Illumina assemblies having as few as 12 contigs and other up to 226 contigs despite high overall sequencing depths. This likely reflects both differences in sequencing technologies, with some sequencing platforms producing shorter reads and therefore, more fragmented assemblies, and the presence of a high number of transposases in some genomes. Indeed, despite having similar sequencing depths and having been sequenced on the same platform, the *R. porcellionis* assembly is strikingly more contiguous than the tick assemblies (Suppl. Table 4.4). However, the genome of *R. porcellionis*, sequenced by Halter et al [75] was reported to be almost free of transposases, contrasting with the transposase-laden genome of *R. oedothoracis*. Similar differences between the tick genomes and the *R. porcellionis* genomes could explain this discrepancy.

To further analyse the genomes, we created a zdb database with all our selected assemblies and 34 reference genomes, including the 4 existing rhabdochlamydia genomes [38, 75] and an *Akkermansia muciniphila* genome to serve as an outgroup (Supp. Table 4.5). Interestingly, one of the genomes was assembled from a *Bemisia tabaci* sequencing project. As *Fritschea bemisiae* was originally described in this host, but was never sequenced, we wondered if the chlamydia whose genome we assembled could be the same as the organism described by Everett et al [12]. To test this, we retrieved the 16 638 bp sequence (Genbank accession: AY140910.1) containing the rRNA encoding genes obtained in the initial study [12] and searched for homologous sequences with blastn in the whole zdb dataset. A nearly perfect match was obtained in the SRR8142474 assembly, with only 4 mismatches over the whole query. This almost perfect match contrasts with the other genomes, where the percentage identity varied between 77 and 91%, with lower query covers (Supp. Fig. 4.7) and suggests that the genome assembled from the *Bemisia tabaci* dataset was indeed sequenced from *Fritschea bemisiae*.

We performed the taxonomic classification of the newly assembled genomes using nine phylogenetic markers [59] and pairwise ANI comparisons. Interestingly, those two

approaches disagreed at the species-level for the T3538 assembly (Suppl. Fig. 4.6). As no *Rhabdochlamydiaceae* genome was available when the taxogenomics criteria were developed and given the widespread use of the ANI for taxonomic classification [203], we chose to separate the T3538 into its own species. Overall, those genomes classify into 8 species, 5 genera and 3 families (Suppl. Fig. 4.6) and add 7 new species, 2 new genera and 1 family level lineages to the *Chlamydiales* order.

Remarkably, the deep-branching chlamydiae genome sequenced from *Scorpiops tibetanus* is one of the smallest among the CLOs, which contrasts with the well-conserved size of the other arthropod-borne chlamydiae (Fig. 4.1). This could be due to the host specialization of this particular species, as genome reduction is a hallmark of host restriction [82]. Conversely, the uniformly larger genome of the other arthropod-borne chlamydiae suggests a wider host range. This hypothesis is further supported by the absence of apparent clustering in function of the host organism. Tick-borne chlamydiae indeed cluster with beetle-borne chlamydiae whereas butterfly-borne chlamydiae cluster with tick-borne and isopod-borne chlamydiae (Fig. 4.1).

4.3.3 Plasmid conformation

All assemblies sequenced from ticks collected by Pilloux et al [66], as well as the *R. porcellionis* assembly had a similar structure in the assembly graph, shown in Fig. 4.2A. The two larger contigs were identified as of chlamydial origin by CAT/BAT, while the presence of homologs of the parA (pgp5) and the integrase (pgp8) genes confirmed the plasmidic origin of those contigs. The smallest contig had no predicted open-reading frame and had therefore no taxonomic classification. Interestingly, the plasmid had a lower coverage than the chromosome in most of our sequencing project, while it was the opposite in the SRA and *R. porcellionis* datasets (Supp. Table 4.4). This may reflect a bias either in the DNA extraction or library preparation steps, as even low copy plasmids would be expected to have a higher sequencing depth than the chromosome.

We hypothesized that the plasmid existed in a unique conformation and attempted

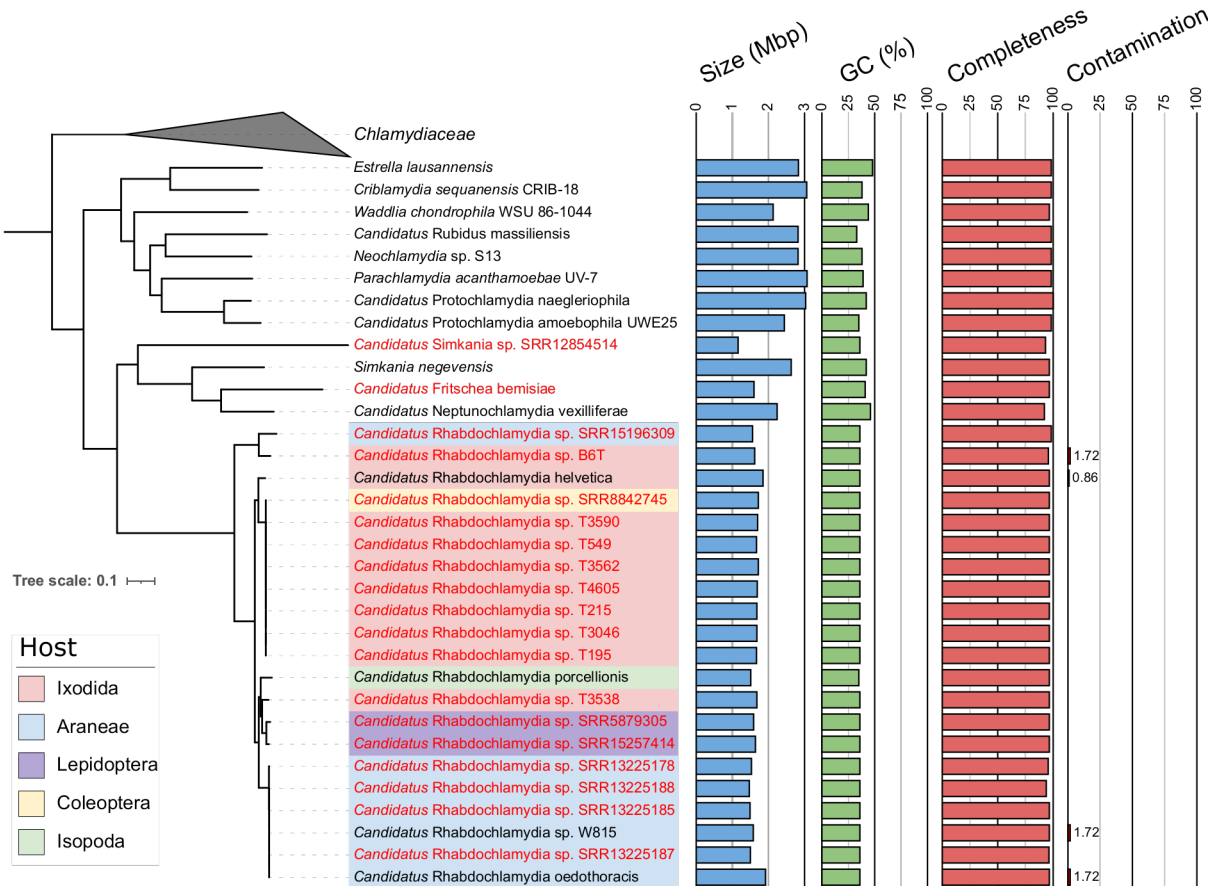

Figure 4.1: Maximum likelihood phylogenetic tree of the genomes sequenced in this study (in red) and of reference genomes of the *Chlamydiales* order (in black). The tree is based on the concatenation of 125 single-copy orthologs and is rooted on *Akkermansia muciniphila*. Genome size (in Mbp), GC content, completeness and contamination (as measured by checkm), are indicated by bars. The hosts of the *Rhabdochlamydiaceae* species included in the tree are indicated by the background color.

to resolve the assembly graph by PCR, on DNA extracted from a *R. porcellionis* culture maintained in Sf9. Surprisingly, the PCR showed that the two conformations co-exist (Fig. 4.2C) and sequencing of the amplicons confirmed that the middle contig is a long inverted repeat that likely prevented Spades from assembling the reads in a complete circular structure. As the F1R2 conformation appear to be more frequent than the F1R1 conformation, we wondered if their frequencies change during the infectious cycle. To test this, we developed two pairs of qPCR, with one qPCR in a pair amplifying the plasmid regardless of its conformation (the FR1 and FR2 PCRs) and the other qPCR amplifying only a specific conformation (F2R1 and F2R2; see Fig. 4.2). This design allows to deduce

the frequency of each conformation by computing the ratio of the results from the two qPCR of a pair. This approach confirmed the impression that the F1R2 conformation is more prevalent than the F1R1 conformation (p-value=0.03, one-sample t-test comparing the averages of the ratio over the timepoints for each replicate to 1). The frequency of the different conformations moreover does not appear to change during the infection cycle (Fig. 4.2D).

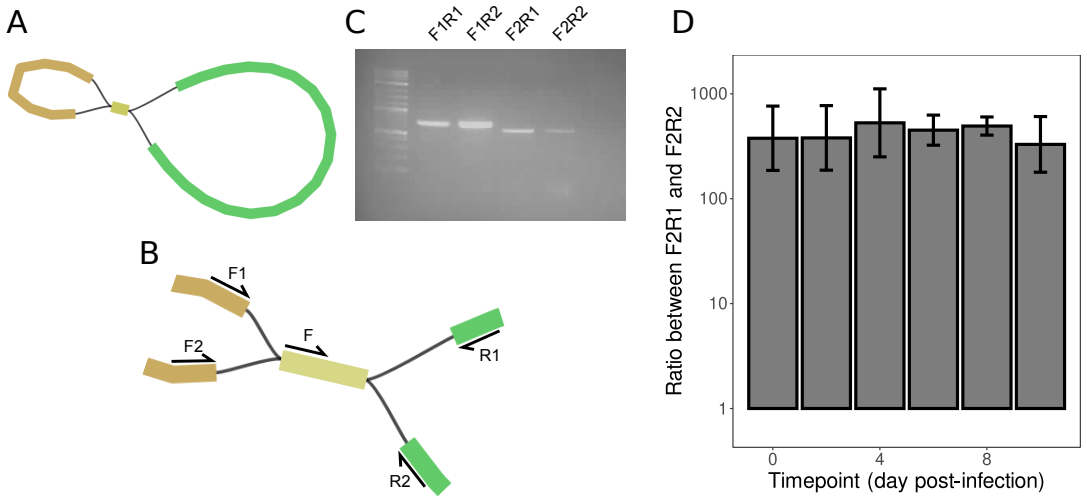

Figure 4.2: The plasmid of *R. porcellionis* has two possible conformations. (A) The assembly graph of the plasmid, composed of three contigs. This separate subgraph was identified as a plasmid due to the presence of proteins typical of chlamydial plasmids. The assembler could not resolve the assembly graph into a single contig as two conformations are possible: the F1R1-F2R2 and the F1R2-F2R1 conformations (B). (B) Primers design for the real-time PCR used to assess the frequency of the different conformations. Two sets of qPCR were used for each conformation: a first qPCR (FR1 or FR2) measured the total copy number of plasmid, while a second qPCR (F2R1 or F2R2, respectively) measured the copy number of plasmid in a given conformation. (C) When trying to resolve the structure of the plasmid by PCR, it appeared that two conformations co-exist. The gel also suggests that the F1R1-F2R2 conformation is less frequent than the F1R2-F2R1 conformation. (D) Evolution of the quantity of F2R1 relative to F2R2 during the infection cycle. The results show the mean and standard deviation of three independent experiments.

4.3.4 Evidence of lateral gene transfer

While assessing the metabolic capacities of the different species using zdb KEGG module completeness prediction, we observed an unusual phylogenetic distribution of the pyridoxal phosphate biosynthesis metabolic pathway (Fig. 4.3A). Such a phylogenetic

distribution could either be explained by the presence of the module in the common ancestor of *Fritschea bemisiae* and the clade of amoebae symbionts and the subsequent loss of the genes in all *Rhabdochlamydiaceae* and most *Simkaniaceae*, or by the acquisition of the module by lateral gene transfer in *Fritschea bemisiae*. The genomic region of the operon do no seem to be conserved between *Fritschea bemisiae* and the amoeba symbionts (Fig. 4.3B). This could be due to genomic rearrangements, but is also compatible with the hypothesis of acquisition by lateral gene transfer. To distinguish between the two possibilities, we searched for homologs of the two genes of the operon in the NCBI non-redundant database (update 2023/01/12) using blastp. The best hit for both genes came from a *Cardinium* endosymbiont (WP_034577597.1 and WP_014934528.1, with 82% identity and a query cover of 100% and 72% identity and 96% query cover, respectively), while running the same query on the orthologous genes of the ameobal symbionts returned hits from chlamydiae. This suggests the acquisition of this metabolic module by lateral gene transfer from a *Cardinium* endosymbiont.

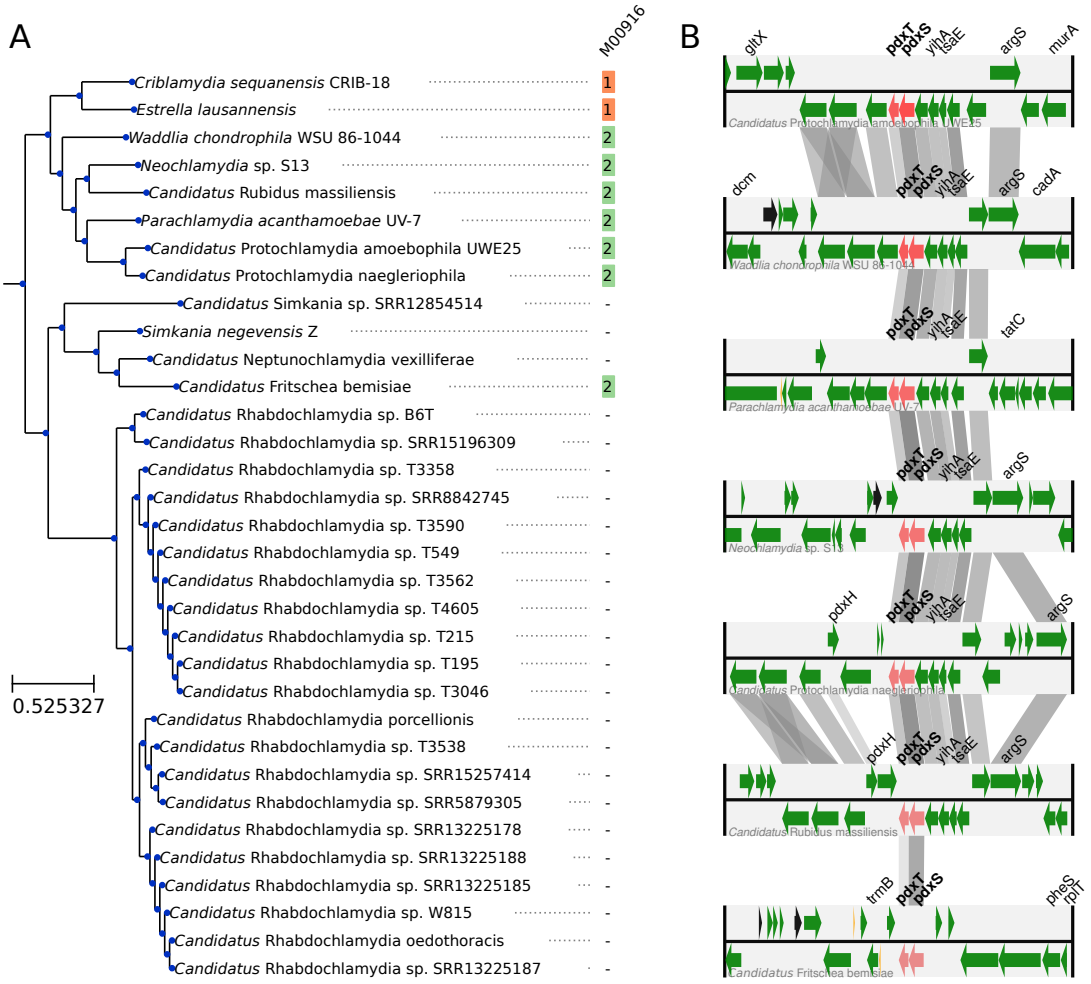

Figure 4.3: The pyridoxal phosphate biosynthesis pathway was likely acquired horizontally by *Fritschea bemisiae*. (A) Phylogenetic distribution of the pyridoxal phosphate biosynthesis KEGG module (module M00916). The numbers indicate the number of genes involved in this module in a genome. Genomes with a complete module are indicated in green. Genomes with an incomplete module are indicated in orange. (B) Genomic neighborhood of the pyridoxal phosphate operon. Orthologous relationships are shown as gray bands. The operon is shown with red arrows. Black arrows indicate pseudo-genes, yellow arrows indicate tRNAs, green arrows indicate protein-coding genes.

4.4 Discussion

In this study, we sequenced and described the genome of 7 new species of arthropod-borne chlamydiae. A majority of those species belong to families already known to infect arthropods, such as the *Rhabdochlamydiaceae* [13, 14, 65, 75] and the *Simkaniaceae* [12], however, one species belong to a new family-level lineage (Supp. Fig. 4.6). Our results thus suggest that the adaptation of chlamydiae to arthropods likely appeared independently at least two times during the evolution of the *Chlamydiae* phylum. Moreover, the comparative analysis of the genomes demonstrated the likely acquisition of a pyridoxal synthesis pathway by lateral gene transfer in *Fritschea bemisiae*, bringing the first evidence of the potential role of a chlamydia as a nutritional symbiont.

One key issue with our results is the possibility that the chlamydial reads were not sequenced from bacteria infecting the arthropods itself but from contaminants present at the surface of the tick. This issue is well documented for tick microbiota studies, for which the necessity of a thorough surface decontamination is well demonstrated [150]. As the SRA metadata do not include the details of sample processing, in particular, whether the surface of the arthropods was sterilized or not, it is impossible to ascertain the absence of such contaminations. The high depth of sequencing of the genomes assembled during this study however make this hypothesis unlikely (Supp. Table 4.4). Surface contaminants are indeed unlikely to represent a significant portion of the samples reads, as suggested in a previous study that screened ticks for the presence of *Chlamydiales* by real-time PCR [66]: samples positive for *Parachlamydiaceae* often had a high CT, in contrast to the low CT in samples positive for *Rhabdochlamydiaceae*. As evidence suggest the inability of *Parachlamydiaceae* to grow in arthropods [72], the samples positive for members of this family were likely due to the presence of infected protists on the surface of the ticks. The high CT of surface contaminants observed in the study by Pilloux *et al.* [66] together with the low quantity of chlamydial reads in samples with high CT observed in this study (Supp. Table 4.4) make it unlikely that entire genomes of contaminants could

be obtained by shotgun sequencing. The taxonomic classification of the assemblies is moreover coherent with the presence of the sequenced bacteria in the arthropod itself, as the assemblies are closely related to *R. porcellionis* (Fig. 4.1), an isolate cultured from woodlouse and thought to be unable to grow in amoebae [60]. It is therefore unlikely that those genomes were sequenced from contaminants.

In conclusion, this study highlighted the extent of the host range of the *Rhabdochlamydiaceae* family, whose members can be found in different orders of the Arthropoda phylum. It also showed that unlike what was suggested in previous studies [75], rhabdochlamydial species do not appear to specialize for particular hosts, as members of the same species could be found in different orders of arthropods.

4.5 Supplementary materials

Table 4.2: Characteristics of the tick pools selected for sequencing. All pools except B6T contained only *Ixodes ricinus* ticks.

Pool	**CT**	**Tick stage**	**N. ticks in pool**	**N. reads**	**Perc. tick reads** [1]
T195	18	Female	5	1 967 732	45%
T215	20	Male	5	4 734 369	43%
T549	22	Nymph	10	1 675 462	47%
T3046	19	Nymph	10	1 886 963	37%
T3186	22	Female	5	2 483 254	80%
T3538	19	Female	5	2 062 190	34%
T3562	19	Female	5	2 230 419	26%
T3590	18	Male	5	1 931 036	18%
T4605	22	Nymph	10	3 991 814	75%
B6T[2]	-	Female	1-5	25 720 123	20%

[1] Based on the mapping of the sequenced on the reference genome used to filter the host DNA.

[2] Sample from Australia, containing an Illumina HiSeq dataset sequenced from an *Ixodes tasmani* pool collected on koalas [68]. The ticks were grouped in pools of 1 to 5 individuals. The exact number of ticks in this pool is not known [68].

Listing 4.1: Query

```
SELECT  meta.acc, meta.sample_acc, meta.biosample,
meta.sra_study, meta.bioproject, meta.sample_name,
meta.libraryselection, meta.instrument, meta.mbytes,
meta.organism, tax.acc,  tax.tax_id, tax.rank,
tax.name, tax.total_count,  tax.self_count,
info.total_spot_count, info.analyzed_spot_count,
info.unaligned_spot_count, info.identified_spot_count,
info.unaligned_only,  info.unaligned_spot_count
FROM ((`nih-sra-datastore.sra_tax_analysis_tool.tax_analysis` AS tax
INNER JOIN `nih-sra-datastore.sra_tax_analysis_tool.tax_analysis_info`
    AS info ON tax.acc = info.acc)
INNER JOIN `nih-sra-datastore.sra.metadata` as meta
    ON tax.acc = meta.acc)
WHERE tax.name IN ("Parachlamydiales", "Chlamydiia", "Chlamydiales")
AND  tax.self_count > 200
AND meta.assay_type="WGS"
AND meta.organism not like "%Chlamydia%"
AND meta.organism not like "Mus␣musculus%"
ORDER BY tax.self_count DESC;
```

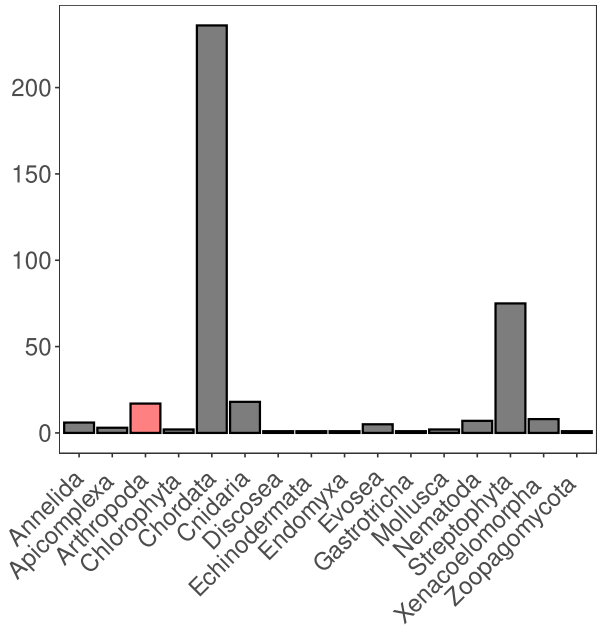

Figure 4.4: Distribution of the sequencing projects retrieved by the query according to the phylum of the sequenced organism. Interestingly, our query also returned plant sequencing projects. This is unexpected as no chlamydiae has ever been isolated from plants, despite several theories postulating that the acquisition of plastids was facilitated by the presence of a chlamydial symbiont [204].

Table 4.3: Arthropod sequencing projects selected for further evaluation

SRA accession	**Organism**	**Class**
ERR4790642	*Crematogaster levior*	Insecta
SRR12854514	*Scorpiops tibetanus*	Arachnida
SRR13225169[1]	*Oedothorax gibbosus*	Arachnida
SRR13225170[1]	*Oedothorax gibbosus*	Arachnida
SRR13225178	*Oedothorax gibbosus*	Arachnida
SRR13225185	*Oedothorax gibbosus*	Arachnida
SRR13225186[1]	*Oedothorax gibbosus*	Arachnida
SRR13225187	*Oedothorax gibbosus*	Arachnida
SRR13225188	*Oedothorax gibbosus*	Arachnida
SRR15196309	*Latrodectus elegans*	Arachnida
SRR15257414	*Melitaea didyma*	Insecta
SRR15257431	*Mimathyma schrenckii*	Insecta
SRR4999935	*Penaeus vannamei*	Malacostraca
SRR5562871	*Aedes aegypti*	Insecta
SRR5879305	*Papilio ambrax*	Insecta
SRR8142474	*Bemisia tabaci*	Insecta
SRR8842745	*Nicrophorus vespilloides*	Insecta

[1] the reads from those archives have already been used to assemble the genomes of *R. oedothoracis* [75].

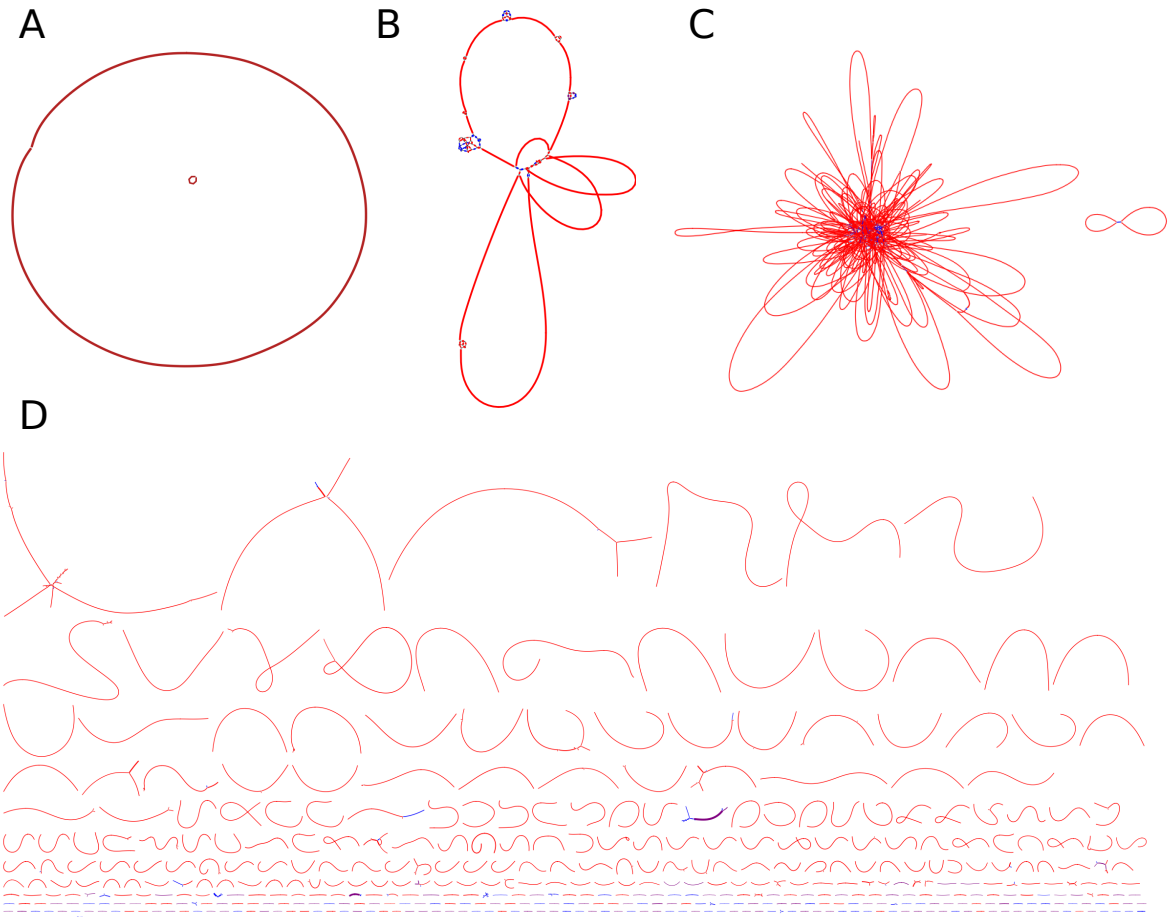

Figure 4.5: Representative assembly graphs, from best (A) to worst (D). This figures shows the assembly graph of (A) the hybrid assembly from the *Latrodectus elegans* dataset, containing a circular chromosome and a plasmid, (B) the SRR12854514 dataset (C) the T549 dataset and (D) the T3186 dataset. Except for the assembly graph in (A), the color of the nodes reflect the taxonomic classification by CAT/BAT. Red: chlamydia. Blue: unclassified. Purple: host organism. Green: other bacteria.

Table 4.4: Assemblies statistics

Assembly	**Cov. chromosome**	**Cov. plasmid**	**N50**	**N. contigs**
SRR12854514	2 685	-[1]	190 517	12
SRR13225178	250	366	10 571	226
SRR13225185	1 548	4 315	9 998	222
SRR13225187	3 518	4 653	10 180	221
SRR13225188	566	1 871	9 299	224
SRR15257414	259	394	82 145	51
SRR15196309	456	957	1 530 314	2
SRR5879305	620	1 151	23 394	108
SRR8142474	440	-[1]	174 556	28
SRR8842745	356	854	19 677	138
T195	151	85	17 485	132
T215	380	207	17 692	132
T3046	177	123	17 692	130
T3538	203	131	19 927	136
T3562	244	135	18 604	136
T3590	238	119	17 175	152
T4605	63	71	18 598	133
T549	123	119	18 598	125
B6T	207	136	16 193	185
R. porcellionis	114	517	172 233	16

[1] no plasmid could be identified in those two assemblies

Table 4.5: List of genomes used as references in the analysis of the assemblies obtained in this study.

Bacteria	**Accession**
Criblamydia sequanensis CRIB-18	GCA_000750955.1
Candidatus Similichlamydia laticola Hat2	GCA_003339615.1
Candidatus Neptunochlamydia vexilliferae K9	GCA_015356785.1
Estrella lausannensis CRIB 30	GCA_900000175.1
Candidatus Protochlamydia amoebophila UWE25	GCA_000011565.2
Candidatus Protochlamydia naegleriophila KNic	GCA_001499655.1
Candidatus Clavichlamydia salmonicola ET	GCA_015356765.1
Candidatus Rhabdochlamydia porcellionis 15C	GCA_015356815.2
Candidatus Rhabdochlamydia helvetica T3358	GCF_901000775.1
Candidatus Rhabdochlamydia sp. W815 W815	GCF_018642185.1
Candidatus Rhabdochlamydia oedothoracis	GCF_019453995.1
Simkania negevensis Z	GCA_000237205.1
Parachlamydia acanthamoebae UV-7	GCA_000253035.1
Waddlia chondrophila WSU 86-1044	GCA_000092785.1
Chlamydiifrater phoenicopteri 14-2711_R47	GCA_902807005.1
Chlamydiifrater volucris 15-2067_O50	GCA_902806995.1
Chlamydia ibidis 10-1398/6	GCA_000454725.1
Candidatus Chlamydia corallus G3/2742-324	GCA_002817655.1
Chlamydia avium 10_881_SC42	GCA_000417735.2
Chlamydia buteonis IDL17-4553	GCA_019056495.1
Chlamydia pecorum DBDeUG_2018	GCA_020459125.1
Chlamydia suis SWA-86	GCA_900169125.1
Chlamydia pneumoniae TW-183	GCA_000007205.1
Chlamydia abortus 15-70d24	GCA_900416725.2
Chlamydia felis Fe/C-56	GCA_000009945.1
Chlamydia caviae GPIC	GCA_000007605.1
Chlamydia muridarum Nigg	GCA_000006685.1
Chlamydia gallinacea 08-1274/3	GCA_000471025.2
Chlamydia psittaci 6BC	GCA_000204255.1
Rubidus massiliensis	GCA_000756735.1
Neochlamydia sp. S13 S13	GCF_000648235.2
Chlamydophila pneumoniae LPCoLN	GCF_000024145.1
Chlamydia trachomatis A/HAR-13	GCF_000012125.1
Akkermansia muciniphila ATCC BAA-835	GCF_000020225.1

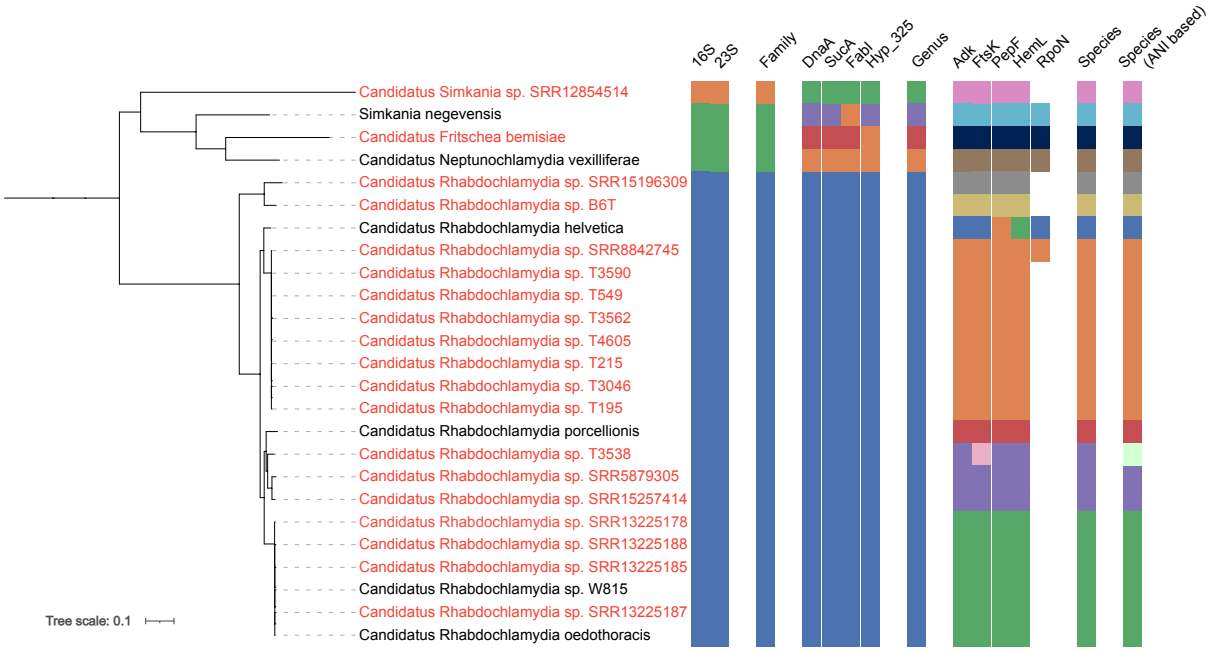

Figure 4.6: Species tree showing the genomes assembled in this study (in red) with existing ones (in black), annotated with clusters grouping the genomes together based on the taxogenomics developed by Pillonel et al [59]. The color are identical for two genomes if the identity for a given taxonomic marker is higher than the proposed threshold allowing their classification in the same taxon. The RpoN marker was absent from most genomes. The last column shows the clustering of the genomes at the species level based on pairwise average nucleotide identity. Genomes were considered as being in the same species if they an ANI higher or equal to 95% and an aligned fraction higher than 65% [203]. Interestingly, the two classification schemes did not agree on all genomes.

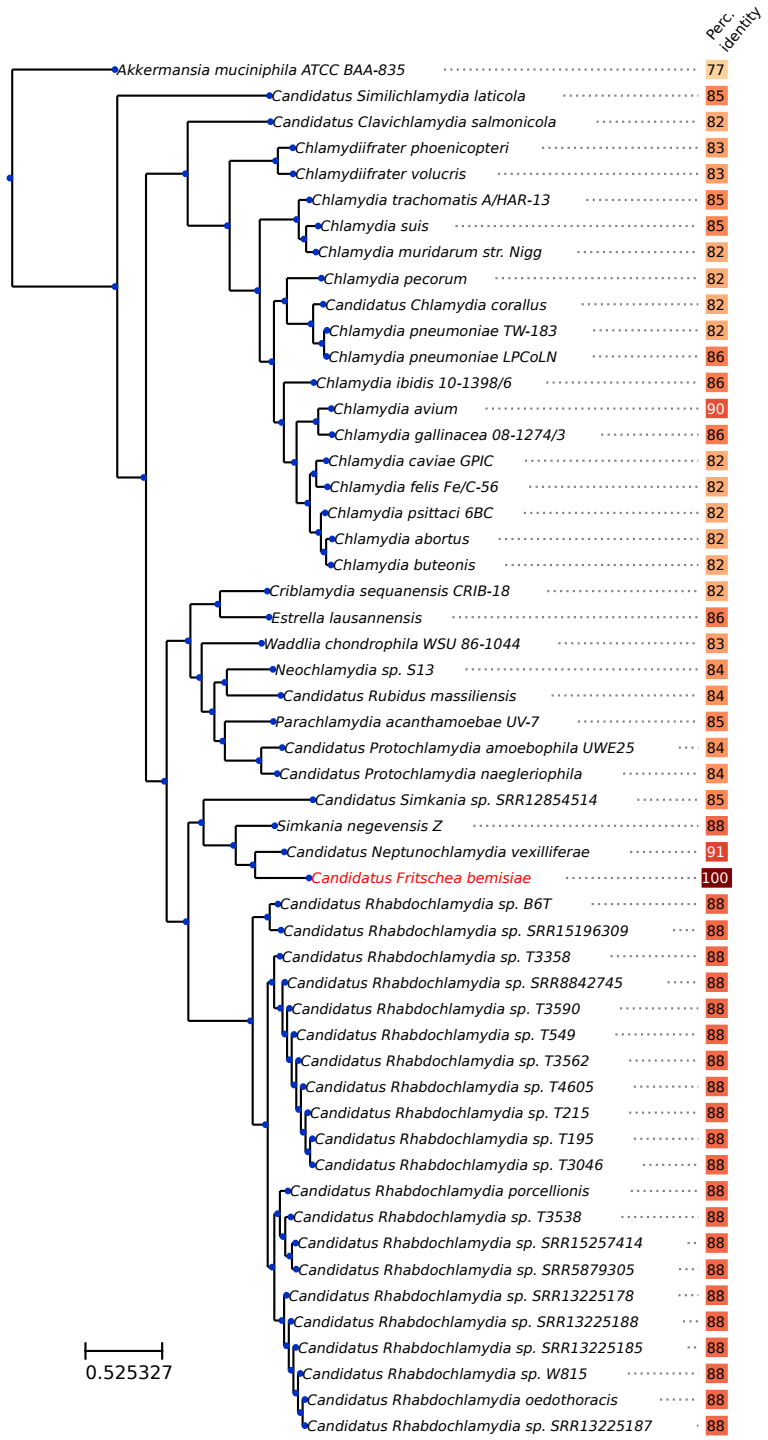

Figure 4.7: Results of searching the genomes of the database for the 16kb sequence of *Fritschea bemisiae* [12] using Blast. The genome assembled from the SRR8142474 dataset has only 4 mismatches over the whole query (16634/16638, query cover of 100%). In the other genomes, the query cover was never higher than 32% and only aligned on the genes of the rRNA operon.

4.6 Bibliography

12. Everett KD, Thao M, Horn M, Dyszynski GE, and Baumann P. Novel chlamydiae in whiteflies and scale insects: endosymbionts '*Candidatus* Fritschea bemisiae' strain Falk and '*Candidatus* Fritschea eriococci' strain Elm. International Journal of Systematic and Evolutionary Microbiology 2005;55:1581–7.

13. Corsaro D, Thomas V, Goy G, Venditti D, Radek R, and Greub G. '*Candidatus* Rhabdochlamydia crassificans', an intracellular bacterial pathogen of the cockroach *Blatta orientalis* (Insecta: Blattodea). Systematic and Applied Microbiology 2007;30:221–8.

14. Kostanjšek R, Štrus J, Drobne D, and Avguštin G. '*Candidatus* Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod *Porcellio scaber* (Crustacea: Isopoda). International journal of systematic and evolutionary microbiology 2004;54:543–9.

28. Omsland A, Sixt BS, Horn M, and Hackstadt T. Chlamydial metabolism revisited: interspecies metabolic variability and developmental stage-specific physiologic activities. FEMS Microbiology Reviews 2014;38:779–801.

37. Collingro A, Tischler P, Weinmaier T, et al. Unity in variety—the pan-genome of the *Chlamydiae*. Molecular biology and evolution 2011;28:3253–70.

38. Pillonel T, Bertelli C, Aeby S, et al. Sequencing the obligate intracellular *Rhabdochlamydia helvetica* within its tick host *Ixodes ricinus* to investigate their symbiotic relationship. Genome biology and evolution 2019;11:1334–44.

40. Sixt BS, Siegl A, Müller C, et al. Metabolic features of *Protochlamydia amoebophila* elementary bodies – A link between activity and infectivity in chlamydiae. PLOS Pathogens 2013;9:e1003553.

48. Stephens RS, Kalman S, Lammel C, et al. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 1998;282:754–9.

58. Kostanjšek R and Marolt TP. Pathogenesis, tissue distribution and host response to *Rhabdochlamydia porcellionis* infection in rough woodlouse *Porcellio scaber*. Journal of invertebrate pathology 2015;125:56–67.

59. Pillonel T, Bertelli C, Salamin N, and Greub G. Taxogenomics of the order *Chlamydiales*. International journal of systematic and evolutionary microbiology 2015;65: 1381–1393.

60. Sixt BS, Kostanjšek R, Mustedanagic A, Toenshoff ER, and Horn M. Developmental cycle and host interaction of *Rhabdochlamydia porcellionis*, an intracellular parasite of terrestrial isopods. Environmental Microbiology 2013;15:2980–93.

65. Croxatto A, Rieille N, Kernif T, et al. Presence of *Chlamydiales* DNA in ticks and fleas suggests that ticks are carriers of *Chlamydiae*. Ticks and tick-borne diseases 2014;5:359–65.

66. Pilloux L, Aeby S, Gaümann R, Burri C, Beuret C, and Greub G. The high prevalence and diversity of *Chlamydiales* DNA within *Ixodes ricinus* ticks suggest a role for ticks as reservoirs and vectors of *Chlamydia*-related bacteria. Applied and Environmental Microbiology 2015;81:8177–82.

68. Burnard D, Weaver H, Gillett A, Loader J, Flanagan C, and Polkinghorne A. Novel *Chlamydiales* genotypes identified in ticks from Australian wildlife. Parasites & Vectors 2017;10:46.

69. Vanthournout B and Hendrickx F. Endosymbiont dominated bacterial communities in a dwarf spider. PLoS One 2015;10:e0117297.

75. Halter T, Koestlbacher S, Collingro A, et al. Ecology and evolution of chlamydial symbionts of arthropods. ISME Communications 2022;2:1–11.

82. McCutcheon JP and Moran NA. Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology 2012;10:13–26.

150. Binetruy F, Dupraz M, Buysse M, and Duron O. Surface sterilization methods impact measures of internal microbial diversity in ticks. Parasites & vectors 2019;12:1–10.

174. Tatusova T, DiCuccio M, Badretdin A, et al. NCBI prokaryotic genome annotation pipeline. Nucleic acids research 2016;44:6614–24.

187. Kalman S, Mitchell W, Marathe R, et al. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nature genetics 1999;21:385–9.

188. Horn M, Collingro A, Schmitz-Esser S, et al. Illuminating the evolutionary history of chlamydiae. Science 2004;304:728–30.

189. Bertelli C, Collyn F, Croxatto A, et al. The *Waddlia* genome: a window into chlamydial biology. PloS one 2010;5:e10890.

190. Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, and O'Sullivan C. The Sequence Read Archive: a decade more of explosive growth. Nucleic acids research 2022;50:D387–D390.

191. Kodama Y, Shumway M, and Leinonen R. The Sequence Read Archive: explosive growth of sequencing data. Nucleic acids research 2012;40:D54–D56.

192. Katz KS, Shutov O, Lapoint R, Kimelman M, Brister JR, and O'Sullivan C. A fast, scalable, MinHash-based k-mer tool to assess Sequence Read Archive next generation sequence submissions. bioRxiv 2021.

193. Chen S, Zhou Y, Chen Y, and Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–i890.

194. Li H and Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics 2009;25:1754–60.

195. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology 2012;19:455–77.

196. Wick RR, Schultz MB, Zobel J, and Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 2015;31:3350–2.

197. Meijenfeldt F von, Arkhipova K, Cambuy DD, Coutinho FH, and Dutilh BE. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome biology 2019;20:1–14.

198. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018;34: 3094–3100.

199. Kolmogorov M, Yuan J, Lin Y, and Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nature biotechnology 2019;37:540–6.

200. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one 2014;9:e112963.

201. Pritchard L, Glover RH, Humphris S, Elphinstone JG, and Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods 2016;8:12–24.

202. Letunic I and Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic acids research 2021;49:W293–W296.

203. Chaumeil PA, Mussig AJ, Hugenholtz P, and Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics (Oxford, England) 2019;36:1925–7.

204. Ball SG, Subtil A, Bhattacharya D, et al. Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? The Plant Cell 2013;25:7–21.

5 Temperature sensitivity affects the host range of *Rhabdochlamydia porcellionis*

Authors: Bastian Marquis, Silvia Ardissone, Gilbert Greub

Status: Published in Applied and Environmental Microbiology

Contributions: BM designed and performed the experiments and the statistical analysis and wrote the manuscript

Abstract

In this work, we tested the permissivity of various mammalian and arthropod cell lines to *R. porcellionis*. No growth could be observed in any mammalian cells, and more surprisingly, in any arthropod cells except in Sf9. As *R. porcellionis* infects organisms with a lower body temperature than mammals, we reasoned that our results might be due to the thermal preferences of this bacterium and repeated the experiment with different incubation temperatures. Incubating Sf9 at 33 °C and 37 °C indeed prevented the replication of the bacteria, while the incubation of mammalian cells at 28 °C allowed its growth. Further demonstrating the sensitivity of *R. porcellionis* to higher temperatures, exposures as short as 6 h at 37 °C were sufficient to irreversibly block its growth. Both developmental stages moreover appear to share similar thermal preferences, as elementary bodies lost their infectivity faster when incubated at 37 °C than when incubated at room temperature.

Those results suggest that *R. porcellionis* either lost or never acquired the ability to grow at the temperatures encountered in mammalian hosts. This species is therefore unlikely to play any pathogenic role in humans or to jump from an arthropod host to a mammal. This work finally highlights the importance of varying the incubation temperatures when attempting to isolate chlamydiae from environmental samples.

AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

ENVIRONMENTAL MICROBIOLOGY

Temperature Affects the Host Range of *Rhabdochlamydia porcellionis*

Bastian Marquis,[a] Silvia Ardissone,[a] Gilbert Greub[a]

[a]Institute of Microbiology of the University Hospital Center and the University of Lausanne, Lausanne, Switzerland

ABSTRACT The *Rhabdochlamydiaceae* family is a recent addition to the *Chlamydiae* phylum. Its members were discovered in cockroaches and woodlice, but recent metagenomics surveys demonstrated the widespread distribution of this family in the environment. It was, moreover, estimated to be the largest family of the *Chlamydiae* phylum based on the diversity of its 16S rRNA encoding gene. Unlike most *Chlamydia*-like organisms, no *Rhabdochlamydiaceae* member could be cultivated in amoebae, and its host range remains unknown. We tested the permissivity of various mammalian and arthropod cell lines to determine the host range of *Rhabdochlamydia porcellionis*, the only cultured representative of this family. While growth could initially be obtained only in the Sf9 cell line, lowering the incubation temperature of the mammalian cells from 37°C to 28°C allowed the growth of *R. porcellionis*. Furthermore, a 6-h exposure to 37°C was sufficient to irreversibly block the replication of *R. porcellionis*, suggesting that this bacterium either lost or never acquired the ability to grow at 37°C. We next sought to determine if temperature would also affect the infectivity of elementary bodies. Although we could not purify enough bacteria to reach a conclusive result for *R. porcellionis,* our experiment showed that the elementary bodies of *Chlamydia trachomatis* and *Waddlia chondrophila* lose their infectivity faster at 37°C than at room temperature. Our results demonstrate that members of the *Chlamydiae* phylum adapt to the temperature of their host organism and that this adaptation can in turn restrict their host range.

IMPORTANCE The *Rhabdochlamydiaceae* family is part of the *Chlamydiae*, a phylum of bacteria that includes obligate intracellular bacteria sharing the same biphasic developmental cycle. This family has been shown to be highly prevalent in the environment, particularly in freshwater and soil, and despite being estimated to be the largest family in the *Chlamydiae* phylum is only poorly studied. Members of the *Rhabdochlamydiaceae* have been detected in various arthropods like ticks, spiders, cockroaches, and woodlice, but the full host range of this family is currently unknown. In this study, we showed that *R. porcellionis*, the only cultured representative of the *Rhabdochlamydiaceae* family, cannot grow at 37°C and is quickly inactivated at this temperature. A similar temperature sensitivity was also observed for elementary bodies of chlamydial species adapted to mammals. Our work demonstrates that chlamydiae adapt to the temperature of their reservoir, making a jump between species with different body temperatures unlikely.

KEYWORDS *Chlamydia*, arthropods, environmental microbiology

The *Chlamydiae* phylum includes obligate intracellular bacteria that share the same biphasic developmental cycle composed of an extracellular infectious stage, the elementary body (EB), and an intracellular replicative form, the reticulate body (1). While historically restricted to the human pathogens of the *Chlamydiaceae* family (2, 3), this order has seen a rapid expansion during the past decades as new species were discovered in the environment (4–12). Estimations based on sequence diversity in the 16S rRNA encoding gene predict hundreds of unknown family-level lineages (5). It is now evident that far

Editor Nicole R. Buan, University of Nebraska-Lincoln

Address correspondence to Gilbert Greub, gilbert.greub@chuv.ch.

The authors declare no conflict of interest.

Received 23 February 2023
Accepted 20 March 2023

from being restricted to mammals, chlamydiae are highly prevalent in the environment and successfully adapted to different ecological niches and host organisms (5, 13–15).

Since they diverged from their common ancestor hundreds of millions of years ago (3, 16), the different families of the *Chlamydiae* phylum specialized for specific hosts, often losing the ability to infect other organisms in the process. For instance, species of the *Chlamydiaceae* family, while highly adapted to vertebrates, are seemingly unable to replicate in amoebae (17–19). Conversely, members of the *Parachlamydiaceae* family grow efficiently in amoebae but poorly, if at all, in mammalian and insect cell lines, likely due to their inability to inhibit apoptosis (20–24). Some other families, like the *Simkaniaceae* and *Waddliaceae* families, seem to have conserved wider host ranges, as cultured representatives can grow in a wide variety of cell types (19, 21, 25–27). It is, however, unclear how the *in vitro* host range of these bacteria translates *in vivo*, as immortalized cell lines grown in axenic medium in the absence of an immune system are far from the conditions expected in most multicellular organisms. The host range of some chlamydiae is thus well determined, while it remains unknown for others.

The picture is clearer for the *Rhabdochlamydiaceae* family, a recent addition to the *Chlamydiae* phylum. Members of this family were initially discovered in cockroaches (28) and woodlice (29) and later detected in ticks and spiders (30–34), while a distant relative was recently identified in the amoeba *Dictyostelium discoideum* (35). Interestingly, *Rhabdochlamydiaceae* were also detected in patients suffering from respiratory infections (36–38) or inflammatory skin disorders (39), suggesting a potential pathogenic role of these bacteria. Despite being hypothesized to be the most diverse clade of the *Chlamydiae* phylum (5), the *Rhabdochlamydiaceae* family is poorly studied, and only one species, isolated from the rough woodlouse, has been cultured so far (40). Similarly to *Chlamydia trachomatis*, *Rhabdochlamydia porcellionis* was shown to inhibit apoptosis and could not be cultured in amoebae (40), hinting at a specialization for multicellular organisms. In line with this hypothesis, several arthropod cell lines were shown to sustain the growth of *R. porcellionis* (40). Unlike the *Chlamydiaceae*, *Rhabdochlamydiaceae* were associated not with animal hosts but with soil and freshwater environments (13), and while various arthropods were demonstrated to be a reservoir for *Rhabdochlamydiaceae*, their full host range is still unknown.

In this study, we sought to study the host range of *R. porcellionis*. We tested the permissivity of mammalian and arthropod cell lines to *R. porcellionis* using immunofluorescence, electron microscopy, and quantitative PCR (qPCR). We could demonstrate that *R. porcellionis* is unable to withstand short exposures to 37°C and cannot grow at 33 and 37°C. Mammalian cells were, however, permissive to these bacteria when incubated at 28°C. *R. porcellionis* thus appears to have adapted to the temperature of its host and to have lost the ability to infect organisms with a higher body temperature in the process.

RESULTS

***R. porcellionis* has a limited host range and a long replication cycle.** To study the host range of *R. porcellionis* in mammalian cells, we tested the permissivity of pneumocytes (A549), endometrial cells (Ishikawa), and fibroblasts (McCoy). The first two cell lines are indeed known to be permissive to different *Chlamydia*-like organisms (25, 27), while McCoy cells are frequently used to propagate members of the *Chlamydiaceae* family. In addition to the Sf9 cell line, already used for the subculture of *R. porcellionis* (40), we tested a tick cell line (IRE/CTVM19) and a mosquito cell line (C6/36), as those were derived from organisms likely closer to the natural reservoir of the *Rhabdochlamydiaceae* family than mammalian cell lines.

Growth could be observed in Sf9 cells (Fig. 1A), with a doubling time of 20.4 h (standard deviation [SD] = 1.9 h), comparable to that of *Simkania negevensis* (21 h) but longer than that of *Waddlia chondrophila* (4 h) in the same cell line (21, 25). Immunofluorescence showed heavily infected Sf9 cells at 6 days postinfection (Fig. 1B); however, reticulate bodies appeared disseminated in the cytoplasm of the host cell and did not seem to be enclosed in an inclusion (Fig. 1B). This observation was confirmed in electron microscopy micrographs, where bacteria appear to replicate in the cytoplasm of the host cell without any

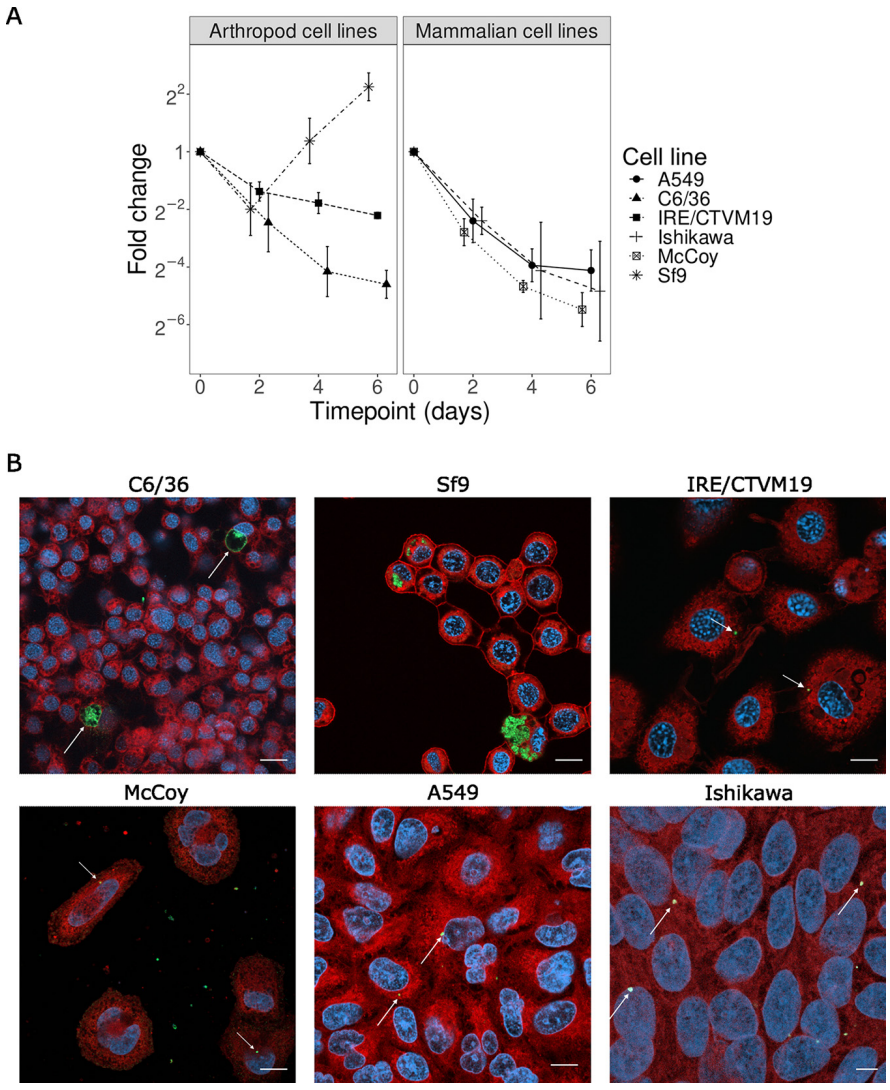

FIG 1 Permissivity of arthropod and mammalian cell lines to *R. porcellionis*. (A) The *y* axis represents the fold change of the number of genome copies per microliter relative to the initial time point. The results are shown as the mean and standard deviation from three biological replicates. (B) *R. porcellionis* in mammalian and arthropod cell lines at 6 days postinfection. Growth could be observed only in Sf9 cells. The reticulate bodies do not appear to be grouped in an inclusion and seem to be replicating directly in the cytoplasm. The enlarged bodies in the C6/36 cell line are likely aberrant bodies. Bacteria appear to have been internalized in all the other cell lines but failed to replicate. White arrows indicate enlarged bacteria in C6/36 cells and internalized EBs in the other cell lines. Cells were stained with concanavalin A (red), DAPI (blue), and anti-*Simkania* antibody (green); bar, 10 μm.

visible inclusion (Fig. 2A to C). Interestingly, unlike *S. negevensis*, *R. porcellionis* could not grow in mammalian cells and, more surprisingly, failed to grow in all arthropod cell lines except Sf9 cells (Fig. 1A). While bacteria were internalized in all cells tested (Fig. 1B), they either failed to initiate replication or formed aberrant bodies measuring more than 5 μm (in C6/36 cells).

***R. porcellionis* is unable to replicate at 37°C.** *Rhabdochlamydiaceae* have been detected in various arthropods such as ticks (34, 41), spiders (42), cockroaches (28), and woodlice (29). As those organisms are poikilothermic and have a lower body temperature than mammals (43, 44), we reasoned that the *Rhabdochlamydiaceae* family might have adapted to the lower temperature of its host organisms and either lost or never acquired the ability to grow at 37°C. To test this hypothesis, we infected Sf9 cells with *R. porcellionis* and incubated them at 20, 28, 33, and 37°C. *R. porcellionis* grew at 20 and 28°C but not at 33 or 37°C (Fig. 3A). The lack of growth at 33 or 37°C could, however, be due to the loss of permissivity of Sf9 cells at those temperatures. As a control for this, we infected Sf9 cells

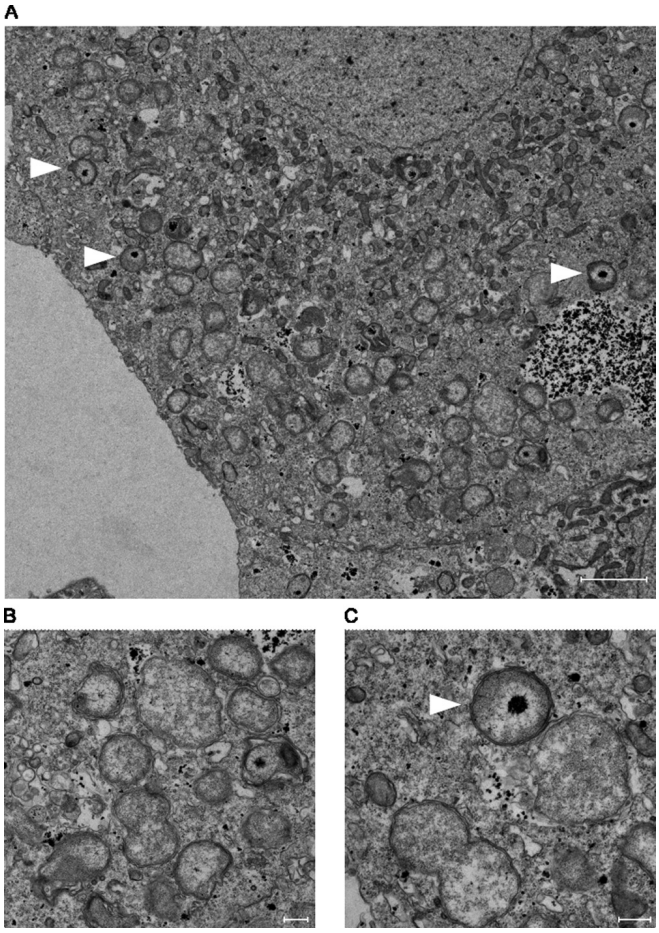

FIG 2 Transmission electron micrographs of *R. porcellionis* in Sf9 cells at 6 days postinfection. (A to C) Infected Sf9 cells harboring numerous bacteria. Reticulate bodies, some of which are undergoing binary fission, can be observed along with intermediate bodies with condensed DNA. The bacteria do not appear to be grouped in an inclusion. Sf9 cells were infected with *R. porcellionis* at an MOI of ~1. White arrowheads, intermediate bodies. Bar, 2 μm (A) or 200 nm (B and C).

with *W. chondrophila*, another *chlamydia*-like organism known to have a wide host range and to grow at both 28 and 37°C (21, 26, 27), and incubated them at 28 and 37°C. Unlike *R. porcellionis*, *W. chondrophila* replicated at both temperatures (see Fig. S1 in the supplemental material), suggesting that Sf9 cells retain their permissivity at 37°C. This, however, does not exclude the possibility of Sf9 cells selectively losing their permissivity to *R. porcellionis* at 37°C.

To test if the absence of growth in mammalian cells was also an effect of temperature, we infected A549, Ishikawa, and McCoy cells with *R. porcellionis* and lowered the incubation temperature to 28°C. This change indeed allowed the growth of *R. porcellionis* (Fig. 3B) in mammalian cells, although the doubling time appeared to be longer and more variable than that in Sf9 cells (Fig. 3C). The long doubling time (Fig. 3C), the absence of intermediate bodies, and the distorted appearance of the bacteria in McCoy cells (Fig. S2) indicate that those cells might be too different from the natural host of *R. porcellionis* to allow efficient growth.

We finally compared the infection efficiencies of *R. porcellionis* in A549, McCoy, Ishikawa, and Sf9 cells by measuring the inclusion-forming unit (IFU) count at 6 days postinfection. As shown in Fig. 3D, the IFU count was significantly lower in Ishikawa and A549 cells than in Sf9 cells. Surprisingly, there was no difference between McCoy and Sf9 cells. This might, however, be due to an underestimation of the IFU count in the latter, as infected Sf9 cells tend to detach from glass coverslips.

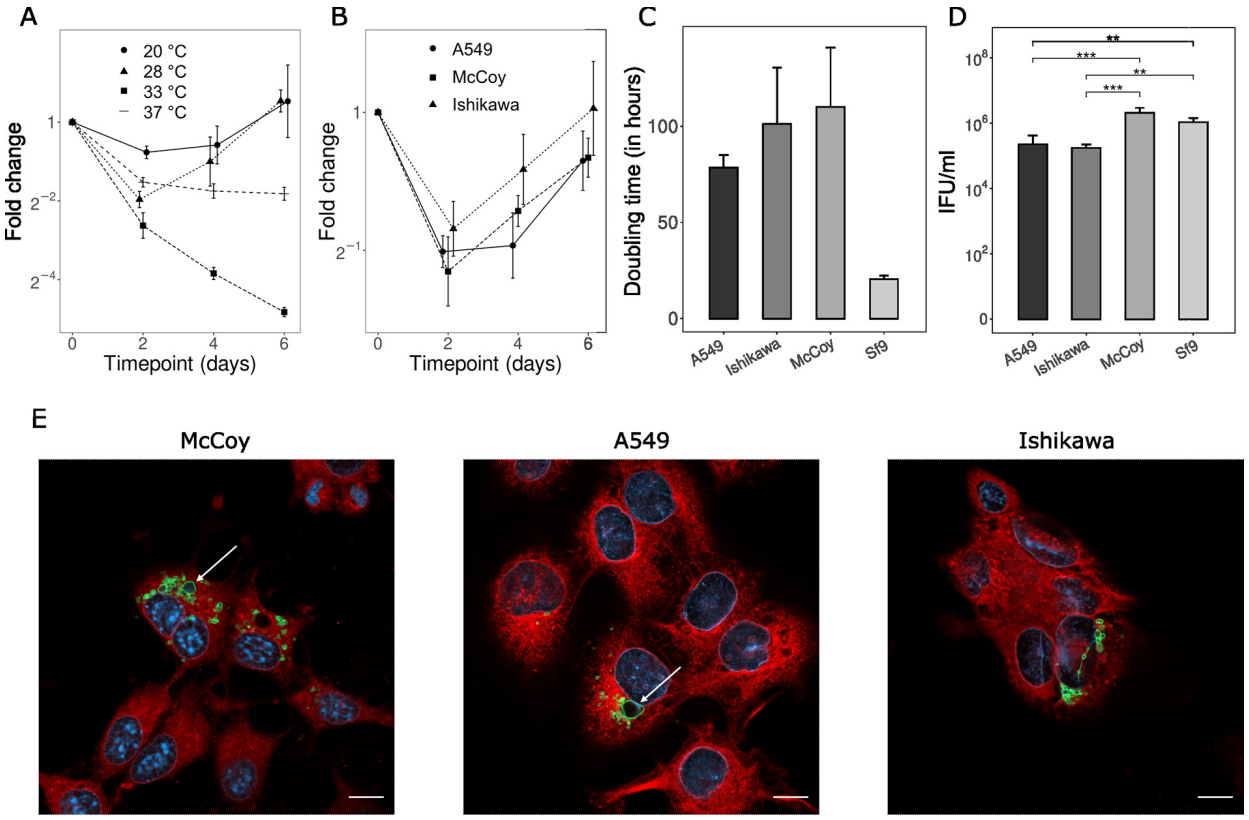

FIG 3 The growth of *R. porcellionis* in mammalian cells depends on temperature. (A) Growth kinetics of *R. porcellionis* in Sf9 cells incubated at 20, 28, 33, and 37°C. (B) Growth kinetics of *R. porcellionis* in mammalian cells incubated at 28°C. In both panel A and panel B, the *y* axis represents the fold change relative to the initial time point. (C) Doubling time of *R. porcellionis* in the different cell lines. Doubling times were estimated by dividing 48 h by the $\log_2$ of the highest fold change observed between two consecutive time points. Despite the marked difference between the doubling time in Sf9 and the other cell lines, the Kruskal-Wallis test was not statistically significant ($P = 0.06$). (D) IFU count of *R. porcellionis* grown in different cell lines at 28°C. The cells were fixed at 6 days postinfection. The tendency of infected Sf9 cells to detach from the glass coverslips could induce an underestimation of the IFU count. A one-way ANOVA revealed that there was a statistically significant difference between at least two cell lines (P value = 0.0002). The plot shows the results of the Tukey honestly significant difference test for the pairwise comparison of the IFU count in the different cell lines (**, <0.01; ***, <0.001). (E) McCoy, A549, and Ishikawa cells infected with *R. porcellionis*, incubated at 28°C, and fixed at 6 days postinfection. The two enlarged bodies in McCoy and A549 cells are likely aberrant bodies (white arrows). Cells were stained with concanavalin A (red), DAPI (blue), and anti-*Simkania* antibodies (green). Bar, 10 μm. The results show the mean and standard deviation from three biological replicates.

Transient exposure to 37°C irremediably blocks the replication of *R. porcellionis*. Members of the *Chlamydiae* are known to enter a third nonreplicative stage, the aberrant body, when exposed to stresses such as antibiotic exposure (45), nutrient deprivation (46), or heat shock (47). Aberrant bodies are typically described as nonreplicating enlarged cells able to resume their regular cycle once the stress disappears, although their morphology was shown to vary as a function of the stresses (48, 49). We thus wondered whether *R. porcellionis* could similarly recover and resume its growth cycle after an exposure to 37°C. To determine this, Sf9 cells at 2 days postinfection were incubated at 37°C for various durations. The infected cells were then further incubated at 28°C for four additional days to check for bacterial growth after stress removal. The growth was assessed by measuring the number of genome copies at the end of the incubation at 37°C and after 4 days of recovery at 28°C (Fig. 4A).

Our experiment shows that a transient exposure to 37°C as short as 6 h irreversibly blocks the replication of *R. porcellionis* (Fig. 4B). The effect of the temperature shift seems to be more deleterious for longer incubation at 37°C, as expected. However, the complete experiment lasted from 6 days, for the samples subjected to the 6-h shift at 37°C, to 8 days, for the samples subjected to the 48-h shift at 37°C. This difference of duration could also explain the more pronounced effect of longer incubations. Indeed, the number of genome copies at the end of the shift to 37°C was not affected by the duration of the exposure (P value = 0.126, one-way analysis of variance [ANOVA]), showing

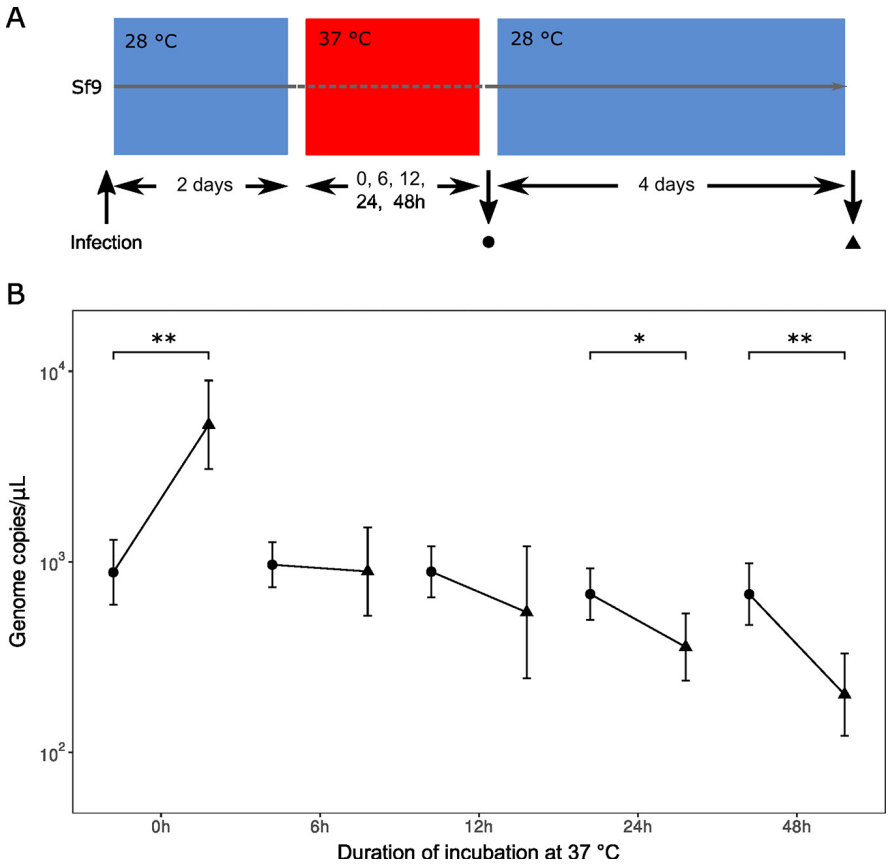

FIG 4 Effect of a transient exposure to 37°C on the replication of *R. porcellionis* in Sf9 cells. (A) Summary of the experimental design. (B) The graph shows the number of bacteria (genome copies/microliter) immediately after the exposure to 37°C (circles) and after 4 days of recovery at 28°C (triangles). The comparison between the two time points was made with a paired *t* test and corrected for multiple testing with the Holm stepdown procedure. The results show the mean and standard deviation from three biological replicates (*, <0.05; **, <0.001).

that it takes at least 48 h for temperature to affect the number of genome copies. It is therefore possible that the apparent lack of difference between 96 h and 0 h for the 6- and 12-h time points is due to the experiment not being long enough for the effect of temperature to manifest in terms of genome copies. The delay might be due to DNA being slowly released—and degraded by cytosolic DNases—from inactivated bacteria.

Temperature affects the infectivity of elementary bodies. Given the effect of temperature on the replication of *R. porcellionis*, we wondered whether it would also affect the infectivity of elementary bodies. To determine this, we incubated elementary bodies of *R. porcellionis* at 20°C (room temperature [RT]) or 37°C and measured the number of IFU every day for 4 days. To be closer to natural conditions, we did not perform the freeze-and-thaw cycle and directly filtered the supernatant from infected Sf9 cells. In addition, as Sf9 cells tend to detach after several days of infection, we used McCoy cells grown at 28°C for IFU quantification of *R. porcellionis* to avoid any bias due to cell detachment. We used a random intercepts mixed-effects linear model to predict the log-transformed count of IFUs based on an interaction variable between the duration of incubation in days and the temperature. When none of the counted cells was infected, we conservatively assumed an IFU count corresponding to one infected cell in 100.

As shown in Fig. 5A, *R. porcellionis* elementary bodies (EBs) incubated for 24 h at 37°C were less infectious than their counterpart incubated at 20°C, although the incubation temperature did not significantly predict the IFU count (R^2 = 0.77, beta for the interaction term = −0.55, *P* value = 0.06). The interaction term implies that for every day of incubation at 37°C, the IFU count will be reduced by a factor of 3.55 ($10^{0.55}$)

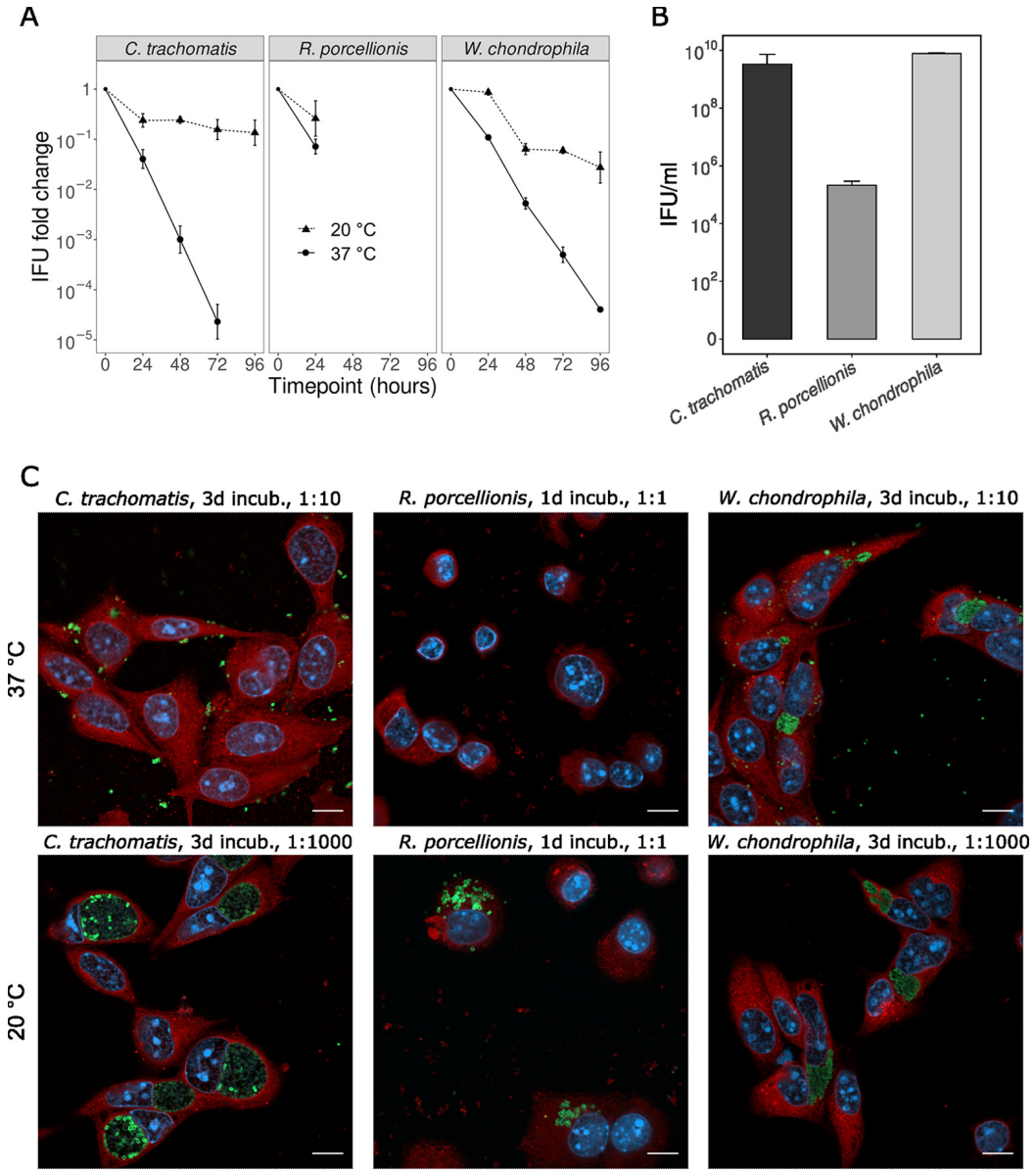

FIG 5 Effect of different incubation temperatures on the infectivity of elementary bodies. (A) Evolution of the number of IFU after incubation at 20°C or 37°C, normalized to the initial IFU count. (B) IFU count at the initial time point. Panels A and B show the results as the mean and standard deviation from three biological replicates. (C) Confocal images of McCoy cells infected with serial dilutions of EBs incubated at 20°C or 37°C. This highlights the deleterious effect of an incubation at 37°C for the EBs of all three species. Interestingly, the inclusions formed by EBs incubated at 37°C also tended to be smaller. The difference between the well-formed inclusions for *C. trachomatis* and *W. chondrophila* and the dissemination of reticulate bodies in the host cytoplasm for *R. porcellionis* is also striking. Cells were fixed at 1 day (for *C. trachomatis* and *W. chondrophila*) or 6 days (for *R. porcellionis*) postinfection and stained with concanavalin A (red), DAPI (blue), and antibodies against the different bacteria (green). Bar, 10 μm. Incub., incubation duration.

compared to the same incubation at room temperature. Due to the low initial quantity of infectious particles (Fig. 5B), no infected cell could be observed after the first time point. Using a freeze-and-thaw cycle would only marginally improve the IFU count (Fig. 3D) and introduce an experimental bias, while resorting to bead beating to increase elementary body yield was shown to be counterproductive for *R. porcellionis* (40). In the absence of alternatives to obtain results for later time points, we did not repeat the experiment with these bacteria. We instead resorted to more tractable chlamydiae from two different families, *W. chondrophila* and *C. trachomatis*, to check whether exposure to different temperatures also affected their infectivity. As both bacteria infect mammals, we expected their EBs to be adapted to the temperature of their host and the number of IFU to decrease faster at 20°C

than at 37°C. This proved to be false, as the IFU count decreased faster at 37°C than at 20°C for both *W. chondrophila* ($R^2 = 0.97$, beta = −0.68, *P* value = 3.27×10^{-11}) and *C. trachomatis* ($R^2 = 0.94$, beta = −1.06, *P* value = 2.08×10^{-10}). Compared to an incubation at room temperature, the IFU count of *C. trachomatis* and *W. chondrophila* is thus predicted to be reduced by a factor of 11.48 and 4.79 for every day of incubation at 37°C, respectively. Altogether, these results suggest that while the reticulate bodies of the *Chlamydiae* have different thermal preferences, the elementary bodies of distant families share the same temperature sensitivity.

DISCUSSION

Several factors are known to influence the host range of the *Chlamydiae*. The ability to inhibit apoptosis has, for instance, been suggested to be a hallmark of chlamydiae infecting multicellular organisms (20, 23), while chlamydiae adapted to multicellular organisms appear to have lost the ability to grow in amoebae (3, 19). The adaptation to specific temperatures has already been demonstrated to be an important determinant for the host range of the members of the *Parachlamydiaceae* family (22, 50, 51). In the present study, we demonstrated that this extends to other families of the *Chlamydiae* phylum by showing that *R. porcellionis* is likely specialized for the temperature ranges encountered in *Porcellio scaber* (43, 52). As a consequence of this specialization, *R. porcellionis* is quickly inactivated if exposed to the body temperature of mammals, making it unlikely for this bacterium to be pathogenic for mammals and restricting its host range to organisms with a temperature lower than 33°C (Fig. 3A). The use of different incubation temperatures should thus be considered in addition to the inclusion of diverse host cells (40) when attempting to isolate environmental chlamydiae. Indeed, independently of its ability to grow in amoebae, a chlamydial symbiont sharing the same temperature sensitivity as *R. porcellionis* would fail to grow in the *Acanthamoeba castellanii* subculture system frequently used for the isolation of chlamydiae from environmental samples (6, 12, 53).

The inability to formally exclude that our observations are due to the effect of temperature on the host cells alone is a key limitation of this study. Our experiments, however, make this hypothesis unlikely, as an effect of the temperature exclusively on the host cells would imply the rapid (Fig. 4B) loss of permissivity of Sf9 cells for *R. porcellionis* at 37°C but not for *W. chondrophila* (Fig. 3A and see Fig. S1 in the supplemental material), as well as the gain of permissivity of mammalian cells at 28°C (Fig. 3B). In addition, a previous work showed that insect cell lines do not lose their permissivity to intracellular bacteria and can notably be used for the culture of *Mycoplasma* species when grown at 37°C (54). Incubations at 37°C also do not seem to affect protein synthesis in Sf9 cells, and heat-tolerant cells growing at 37°C could even be obtained (55). The tolerance of Sf9 cells to high temperatures was hypothesized to be due to *Spodoptera frugiperda* being adapted to warm climates (55). Along with previous evidence showing a similar temperature sensitivity in species of the *Parachlamydiaceae* family (50, 51), an effect of the temperature on the bacteria themselves thus appears to be the a more parsimonious explanation. Conversely, the effect of cold shock on mammalian cells remains poorly understood and has been studied essentially for its use in the production of recombinant proteins (56, 57). The possibility that the growth of *R. porcellionis* in mammalian cells is due in part to an effect of the temperature on the host cell therefore cannot be excluded.

The attempt to assess whether elementary bodies share the temperature sensitivity of reticulate bodies was impaired by the difficulty of purifying enough *R. porcellionis* EBs. The unequivocal results obtained for two chlamydial species from different families and the trend observed for *R. porcellionis* (Fig. 5A), however, suggest that the elementary bodies of this bacterium also lose their infectivity faster at 37°C. Elementary bodies of different chlamydial species may therefore be more similar in terms of thermal preferences than their respective reticulate bodies (50, 51), although this trend would need to be confirmed in other chlamydiae such as *Simkania negevensis* or in species of the *Parachlamydiaceae* family. Of note, our results also suggest that the better extracellular survival of *Chlamydia*-like organisms than of *C. trachomatis* observed in a

previous study (58) may be due to differences in incubation temperatures rather than the metabolic capacity of those organisms.

The ability to grow in both insect and mammalian cells (Fig. 1A and Fig. 3B) that we observed for *R. porcellionis* has also been reported for various *Chlamydiaceae* (59, 60) species and suggests that despite the long evolutionary distances between those organisms, their cells are still similar enough to allow the growth of bacteria from both families. The long doubling times and frequent aberrant bodies observed in mammalian cells may be due either to differences in the host cell physiology or to the repercussion of a lower fitness of the host cell on the bacteria. In contrast to the previous observation that *R. porcellionis* can grow in different insect cells (40), we could not observe growth in either the C6/36 or the IRE/CTVM19 cell line. The lack of growth in the latter cell line might be due to the specialization of *R. porcellionis* for isopods at the cost of the ability to grow in cells originating from the more distantly related arachnids. This explanation, however, fits poorly with the observation that *R. porcellionis* grows in mammalian cells. The permissivity of cell lines might therefore be more related to the cell type than to the host species, as suggested by the restricted tissue distribution of *R. porcellionis* and *Rhabdochlamydia crassificans* in their respective hosts (28, 61).

It is unclear how the thermal preferences of *R. porcellionis* generalize to the other members of the *Rhabdochlamydiaceae*. The high predicted diversity of the *Rhabdochlamydiaceae* family (5, 42) and the demonstration that species of the same chlamydial family can have different thermal preferences (50, 51) indeed imply the possibility that other rhabdochlamydiae could have adapted to various ranges of temperatures. However, the similar thermal preferences exhibited by ixodid ticks (44, 62) and woodlice (43, 52) suggest the possibility that the same evolutionary mechanism that drove the adaptation of *R. porcellionis* toward lower temperatures could have had the same effect in tick-borne rhabdochlamydiae (30, 34, 41). This will, however, remain purely speculative until the thermal preferences of additional rhabdochlamydial species can be assessed. Interestingly, a similar temperature sensitivity has also been reported for *Wolbachia* (63, 64), another arthropod symbiont whose host range broadly overlaps that of the *Rhabdochlamydiaceae* (65). Of note, *Wolbachia* appears to influence the thermal preferences of its host to favor its replication (66). A similar effect might also be found in *R. porcellionis*, as it could prevent the accidental cure of the host of the bacterial infection by a short exposure to temperatures higher than 33°C (Fig. 4B).

This work demonstrates the adaptation of *R. porcellionis* to the range of temperatures encountered in its host, at the cost of the ability to successfully infect other species with higher body temperatures. In particular, the temperature sensitivity of *R. porcellionis* precludes its transmission to mammals and excludes a pathogenic role of this bacterium for humans (36–38, 67). It finally highlights the importance of testing different incubation temperatures when attempting to recover chlamydiae from environmental samples.

MATERIALS AND METHODS

Cell culture. *Spodoptera frugiperda* ovarian epithelial cells (Sf9, ATCC CRL-1711) were cultured at 28°C in Grace insect medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% fetal calf serum (FCS). *Aedes albopictus* cells (C6/36, ATCC CRL-1660) were cultured at 28°C in the presence of 5% CO_2 in Dulbecco modified Eagle medium (DMEM) (PAN-Biotech, Aidenbach, Germany) supplemented with 10% FCS. *Ixodes ricinus* (IRE/CTVM19) cell lines were maintained at 28°C in Leibovitz L-15 medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% tryptose phosphate broth (Gibco, Thermo Fisher Scientific, Waltham, MA, USA), 20% FCS, and 1% L-glutamine (Sigma-Aldrich, Buchs, Switzerland), as described in reference 68. Human pneumocytes (A549, ATCC CCL-185), mouse fibroblasts (McCoy, ATCC CRL-1696), and human endometrial cells (Ishikawa, gift of G. Canny) were cultured in DMEM supplemented with 10% FCS and grown at 37°C in the presence of 5% CO_2. The *Acanthamoeba castellanii* strain (ATCC 30010) was cultured at 25°C in peptone-yeast extract-glucose (PYG) medium.

Bacterial strains. The *R. porcellionis* strain was acquired from the DSMZ collection (DSM 27522) and cultivated in Sf9 cells. The infected cells were passaged once a week, and fresh cells were added approximately every four passages to compensate for host cell death due to the presence of the bacteria. *Waddlia chondrophila* strain WSU 86-1044 (ATCC VR-1470) was cocultivated with *A. castellanii* in PYG broth at 32°C. Suspensions of EBs were collected at 7 days postinfection, diluted 10 times, and used to infect fresh *A. castellanii*. *Chlamydia trachomatis* (ATCC VR-902B) was cultivated in McCoy cells incubated at 37°C in DMEM supplemented with 10% FCS and 1 μg mL^{-1} cycloheximide.

Infection procedure. Cells were seeded in a 24-well plate (Corning) at a density of 1×10^5 or 3×10^5 cells per well 2 h before infection. The infection procedures were performed as described in previous publications (26, 40).

For *R. porcellionis*, suspensions of infected Sf9 cells were subjected to a freeze-thaw cycle to disrupt the cells, followed by a filtration through a 5-μm-pore filter to remove the debris. Plated cells were then infected with the filtrate at a multiplicity of infection (MOI) of ~0.1 to 1 and centrifuged for 15 min at $130 \times g$ at room temperature, followed by an incubation of 30 min at 28°C. The medium was then replaced to remove noninternalized bacteria.

W. chondrophila EBs were collected from the supernatant of *A. castellanii* at 5 days postinfection. The supernatant was then filtered through a 5-μm-pore filter to remove cell debris. Plated cells were infected with the filtrate at an MOI of ~0.1 to 1 and centrifuged at $1{,}790 \times g$ for 10 min at room temperature. After 30 min of incubation at either 28°C (for Sf9 cells) or 37°C (for mammalian cells), the medium was replaced to remove noninternalized bacteria.

C. trachomatis EBs were collected from the supernatant of infected McCoy cells at 3 days postinfection. The supernatant was then filtered with a 5-μm-pore filter. Plated cells were infected with the filtrate and centrifuged at $900 \times g$ for 15 min at room temperature. The infected cells were then incubated for 30 min at 37°C, and the medium was replaced to remove noninternalized bacteria.

After the infection, the cells were incubated at their usual growth temperature, unless specified otherwise. The samples were collected at various time points for quantification by quantitative PCR (qPCR) and immunofluorescence staining. The MOI was estimated by measuring the proportion of infected cells with 10-fold serial dilutions of EB filtrate in immunofluorescence.

Inclusion-forming unit (IFU) quantification. Cells were plated in a 24-well plate at a density of 3×10^5 cells per well 2 h before the infection and were then infected with serial 10-fold dilutions of EB suspensions. After the initial 30 min of incubation, the medium was replaced with fresh medium supplemented with 1 μg mL^{-1} of cycloheximide (69). The cells infected with *W. chondrophila* or *C. trachomatis* were incubated for 24 h at 37°C with 5% CO_2, while the cells infected with *R. porcellionis* were incubated for 6 days at 28°C, with 5% CO_2 for mammalian cells. The cells were then fixed and stained for immunofluorescence, and the proportion of infected cells was determined using an epifluorescence microscope. At least 100 cells were counted for each condition.

Effect of temperature on EB infectivity. *C. trachomatis* and *W. chondrophila* EBs were collected from the supernatant of infected cells as in a standard infection procedure. To be closer to a natural infection, EBs of *R. porcellionis* were collected from the supernatant of infected cells, without any prior lysis step. The supernatants were filtered with a 5-μm-pore filter. The filtrate was then diluted 1:1 in PYG for *W. chondrophila*, 1:1 in DMEM supplemented with 10% FCS for *C. trachomatis*, and 1:1 in Grace medium supplemented with 10% FCS for *R. porcellionis*. The suspensions of elementary bodies were then incubated at either 20°C or 37°C in 24-well plates. In the case of *C. trachomatis*, the plate was incubated in the presence of 5% CO_2. IFUs were quantified using McCoy cells immediately after the dilution in fresh medium or after 1, 2, 3, or 4 days of incubation.

Effect of incubation temperature on growth. Sf9 cells were plated in a 24-well plate at a density of 10^5 cells per well, infected with *R. porcellionis* at an MOI of ~0.1 to 1, and incubated for 48 h at 28°C. The plates were then incubated at 37°C for 6, 12, 24, or 48 h before being switched back to 28°C for four additional days. Samples were taken for bacterial growth quantification by qPCR right after the switch to 28°C and after the subsequent 4 days of incubation.

Quantitative PCR. Genomic DNA was extracted using the Wizard SV genomic DNA purification kit (Promega, Dübendorf, Switzerland) following the manufacturer's protocol. Quantitative PCR for *R. porcellionis* (36) or *W. chondrophila* (70) was performed on 5 μL of genomic DNA with iTaq Supermix (Bio-Rad, Cressier, Switzerland), 200 nM primers (WadF4, 5′-GGCCCTTGGGTCGTAAAGTTCT-3′, and WadR4, 5′-CGGAGTTAGCCGGTGCTTCT-3′, for *W. chondrophila*; RcF, 5′-GACGCTGCGTGAGTGATGA-3′, and RcR, 5′-CCGGTGCTTCTTTACGCAGTA-3′, for *R. porcellionis*) and 100 nM probe (WadS2, 5′-6-carboxyfluorescein [FAM]-CATGGGAACAAGAGAAGGATG-BHQ1-3′, and RcS, 5′-FAM-CTTTCGGGTTGTAAAACTCTTTCGCGCA-BHQ1-3′). The cycling conditions were identical for both qPCRs: 3 min at 95°C and 40 cycles of 15 s at 95°C and 1 min at 60°C. The qPCRs were performed on a QuantStudio3 real-time PCR system (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA).

Immunofluorescence staining. Infected cells grown on glass coverslips were fixed with ice-cold methanol for 5 min at different time points after the infection. Cells were then washed three times with phosphate-buffered saline (PBS) and incubated for at least 2 h in PBS with 0.1% saponin, 0.04% NaN_3, and 10% FCS (blocking solution). The coverslips were then incubated at room temperature for 2 h in blocking solution with rabbit anti-*Simkania negevensis* antibodies (25) (dilution at 1:1,000), rabbit anti-*Waddlia chondrophila* antibodies (71) (dilution at 1:1,000), or goat antibodies targeting the major outer membrane protein of *Chlamydia trachomatis* (dilution at 1:1,000) (LSBio, Seattle, WA, USA). Anti-*Simkania* antibodies were used to detect *R. porcellionis*, as antibodies raised against a chlamydial species often cross-react with related species (40, 72). After the incubation with the primary antibody, the coverslips were washed three times in PBS with 0.1% saponin and incubated for 1 h at room temperature in blocking solution with 1.6 μg mL^{-1} 4′,6-diamidino-2-phenylindole (DAPI) dilactate (Molecular Probes, Thermo Fisher Scientific, Waltham, MA, USA), 100 μg mL^{-1} concanavalin A-Texas Red conjugate (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA), and Alexa 488-conjugated chicken anti-goat or goat anti-rabbit antibodies (1:1,000 dilution) (Life Technologies, Thermo Fisher Scientific, Waltham, MA, USA). The coverslips were then embedded in Mowiol (Sigma-Aldrich, Buchs, Switzerland) and kept in the dark at 4°C until further use. The coverslips were examined with a confocal microscope (Zeiss LSM 900; Zeiss, Oberkochen, Germany).

Electron microscopy imaging. Sf9 and McCoy cells were plated in T25 flasks at a density of 10^6 cells per flask and infected with *R. porcellionis* at an MOI of ~1. After the initial 30 min of incubation at 28°C, the old medium was replaced with fresh medium supplemented with 1 μg mL^{-1} cycloheximide. The cells were then incubated for 6 days at 28°C before collection. The cell suspension was centrifuged at 500 $\times$ *g* for 10 min, and the pellet was resuspended in a solution of 4% paraformaldehyde (Electron Microscopy Sciences [EMS], Hatfield, PA, USA) and 2.5% glutaraldehyde (Fluka, Buchs, Switzerland) in a 0.1 mol L^{-1} phosphate buffer at pH 7.4 (PB buffer) and incubated at 4°C for 4 h. After an additional centrifugation at 500 $\times$ *g* for 10 min, the cells were resuspended in a solution of 1% paraformaldehyde in PB buffer. They were then directly postfixed by a fresh mixture of 1% osmium tetroxide (EMS, Hatfield, PA, USA) with 1.5% potassium ferrocyanide (Sigma, St. Louis, MO, USA) in PB buffer for 1 h at room temperature (RT). The samples were then washed three times in distilled water and spun down in 2% low-melting-temperature agarose in H_2O (Sigma, St. Louis, MO, USA), allowed to solidify on ice, cut in 1-mm^3 cubes, and dehydrated in acetone solution (Sigma, St. Louis, MO, USA) at graded concentrations (30%, 40 min; 50%, 40 min; 70%, 40 min; 100%, twice for 1 h). This was followed by infiltration in Epon (Sigma, St. Louis, MO, USA) at graded concentrations (Epon 1/3 acetone, 2 h; Epon 3/1 acetone, 2 h, Epon 1/1, 4 h; Epon 1/1, 12 h) and finally polymerization for 48 h at 60°C in an oven. Ultrathin sections of 50 nm were cut on a Leica Ultracut microtome (Leica Microsystems GmbH, Vienna, Austria) and picked up on a copper slot grid (2 by 1 mm; EMS, Hatfield, PA, USA) coated with a polyethyleneimine (PEI) film (Sigma, St. Louis, MO, USA). Sections were poststained with 2% uranyl acetate (Sigma, St. Louis, MO, USA) in H_2O for 10 min, rinsed several times with H_2O followed by Reynolds lead citrate in H_2O (Sigma, St. Louis, MO, USA) for 10 min, and rinsed several times with H_2O. Images were taken with a Philips CM100 1201 microscope at the Lausanne University electron microscopy facility.

Statistical analysis. The results of this study are given as means with standard deviations. The linear regression models with random effect were fitted using the lme4 package (73). Doubling times were calculated by dividing 48 h (the interval between time points used in this work) by the $\log_2$ of the highest fold change observed between two consecutive time points. All statistics were performed with R (v4.2.0).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 1 MB.

ACKNOWLEDGMENTS

Bastian Marquis is funded by the Jürg Tschopp MD-PhD scholarship.

Conceptualization, B.M. and G.G.; data curation, B.M.; formal analysis, B.M.; funding acquisition, B.M. and G.G.; investigation, B.M., G.G., and S.A.; methodology, B.M. and S.A.; project administration, G.G.; resources, G.G. and S.A.; software, not applicable; supervision, G.G. and S.A.; validation, G.G. and S.A.; visualization, B.M.; writing – original draft, B.M.; writing – review & editing, G.G. and S.A.

REFERENCES

1. Moulder JW. 1991. Interaction of chlamydiae and host cells in vitro. Microbiol Rev 55:143–190. https://doi.org/10.1128/mr.55.1.143-190.1991.
2. Page LA. 1968. Proposal for the recognition of two species in the genus Chlamydia Jones, Rake, and Stearns, 1945. Int J Syst Evol Microbiol 18:51–66.
3. Horn M. 2008. Chlamydiae as symbionts in eukaryotes. Annu Rev Microbiol 62:113–131. https://doi.org/10.1146/annurev.micro.62.081307.162818.
4. Pillonel T, Bertelli C, Greub G. 2018. Environmental metagenomic assemblies reveal seven new highly divergent chlamydial lineages and hallmarks of a conserved intracellular lifestyle. Front Microbiol 9:79. https://doi.org/10.3389/fmicb.2018.00079.
5. Lagkouvardos I, Weinmaier T, Lauro FM, Cavicchioli R, Rattei T, Horn M. 2014. Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the Chlamydiae. ISME J 8:115–125. https://doi.org/10.1038/ismej.2013.142.
6. Thomas V, Casson N, Greub G. 2006. Criblamydia sequanensis, a new intracellular Chlamydiales isolated from Seine river water using amoebal co-culture. Environ Microbiol 8:2125–2135. https://doi.org/10.1111/j.1462-2920.2006.01094.x.
7. Rurangirwa FR, Dilbeck PM, Crawford TB, McGuire TC, McElwain TF. 1999. Analysis of the 16S rRNA gene of micro-organism WSU 86–1044 from an aborted bovine foetus reveals that it is a member of the order Chlamydiales: proposal of Waddliaceae fam. nov., Waddlia chondrophila gen. nov., sp. nov. Int J Syst Bacteriol 49:577–581. https://doi.org/10.1099/00207713-49-2-577.
8. Kahane S, Metzer E, Friedman MG. 1995. Evidence that the novel microorganism "Z" may belong to a new genus in the family Chlamydiaceae. FEMS Microbiol Lett 126:203–207. https://doi.org/10.1111/j.1574-6968.1995.tb07417.x.
9. Fritsche TR, Gautom RK, Seyedirashti S, Bergeron DL, Lindquist TD. 1993. Occurrence of bacterial endosymbionts in *Acanthamoeba* spp. isolated from corneal and environmental specimens and contact lenses. J Clin Microbiol 31:1122–1126. https://doi.org/10.1128/jcm.31.5.1122-1126.1993.
10. Amann R, Springer N, Schönhuber W, Ludwig W, Schmid EN, Müller KD, Michel R. 1997. Obligate intracellular bacterial parasites of acanthamoebae related to Chlamydia spp. Appl Environ Microbiol 63:115–121. https://doi.org/10.1128/aem.63.1.115-121.1997.
11. Kahane S, Gonen R, Sayada C, Elion J, Friedman MG. 1993. Description and partial characterization of a new Chlamydia-like microorganism. FEMS Microbiol Lett 109:329–333. https://doi.org/10.1111/j.1574-6968.1993.tb06189.x.
12. Lienard J, Croxatto A, Prod'hom G, Greub G. 2011. Estrella lausannensis, a new star in the Chlamydiales order. Microbes Infect 13:1232–1241. https://doi.org/10.1016/j.micinf.2011.07.003.
13. Köstlbacher S, Collingro A, Halter T, Schulz F, Jungbluth SP, Horn M. 2021. Pangenomics reveals alternative environmental lifestyles among chlamydiae. Nat Commun 12:4021. https://doi.org/10.1038/s41467-021-24294-3.
14. Collingro A, Köstlbacher S, Horn M. 2020. Chlamydiae in the environment. Trends Microbiol 28:877–888. https://doi.org/10.1016/j.tim.2020.05.020.
15. Dharamshi JE, Tamarit D, Eme L, Stairs CW, Martijn J, Homa F, Jørgensen SL, Spang A, Ettema TJG. 2020. Marine sediments illuminate chlamydiae diversity and evolution. Curr Biol 30:1032–1048.e7. https://doi.org/10.1016/j.cub.2020.02.016.

16. Greub G, Raoult D. 2003. History of the ADP/ATP-translocase-encoding gene, a parasitism gene transferred from a *Chlamydiales* ancestor to plants 1 billion years ago. Appl Environ Microbiol 69:5530–5535. https://doi.org/10.1128/AEM.69.9.5530-5535.2003.
17. Essig A, Heinemann M, Simnacher U, Marre R. 1997. Infection of *Acanthamoeba castellanii* by *Chlamydia pneumoniae*. Appl Environ Microbiol 63:1396–1399. https://doi.org/10.1128/aem.63.4.1396-1399.1997.
18. Wirz M, Polkinghorne A, Dumrese C, Ziegler U, Greub G, Pospischil A, Vaughan L. 2008. Predator or prey? Chlamydophila abortus infections of a free-living amoebae, Acanthamoeba castellani 9GU. Microbes Infect 10:591–597. https://doi.org/10.1016/j.micinf.2008.01.006.
19. Coulon C, Eterpi M, Greub G, Collignon A, McDonnell G, Thomas V. 2012. Amoebal host range, host-free survival and disinfection susceptibility of environmental Chlamydiae as compared to Chlamydia trachomatis. FEMS Immunol Med Microbiol 64:364–373. https://doi.org/10.1111/j.1574-695X.2011.00919.x.
20. Sixt BS, Hiess B, König L, Horn M. 2012. Lack of effective anti-apoptotic activities restricts growth of Parachlamydiaceae in insect cells. PLoS One 7:e29565. https://doi.org/10.1371/journal.pone.0029565.
21. Kebbi-Beghdadi C, Fatton M, Greub G. 2015. Permissivity of insect cells to Waddlia chondrophila, Estrella lausannensis and Parachlamydia acanthamoebae. Microbes Infect 17:749–754. https://doi.org/10.1016/j.micinf.2015.09.014.
22. Hayashi Y, Nakamura S, Matsuo J, Fukumoto T, Yoshida M, Takahashi K, Mizutani Y, Yao T, Yamaguchi H. 2010. Host range of obligate intracellular bacterium Parachlamydia acanthamoebae. Microbiol Immunol 54:707–713. https://doi.org/10.1111/j.1348-0421.2010.00265.x.
23. Brokatzky D, Kretz O, Häcker G. 2020. Apoptosis functions in defense against infection of mammalian cells with environmental chlamydiae. Infect Immun 88:e00851-19. https://doi.org/10.1128/IAI.00851-19.
24. Greub G, Desnues B, Raoult D, Mege JL. 2005. Lack of microbicidal response in human macrophages infected with Parachlamydia acanthamoebae. Microbes Infect 7:714–719. https://doi.org/10.1016/j.micinf.2005.01.009.
25. Vouga M, Baud D, Greub G. 2017. Simkania negevensis may produce long-lasting infections in human pneumocytes and endometrial cells. Pathog Dis 75:ftw115. https://doi.org/10.1093/femspd/ftw115.
26. Kebbi-Beghdadi C, Batista C, Greub G. 2011. Permissivity of fish cell lines to three Chlamydia-related bacteria: Waddlia chondrophila, Estrella lausannensis and Parachlamydia acanthamoebae. FEMS Immunol Med Microbiol 63:339–345. https://doi.org/10.1111/j.1574-695X.2011.00856.x.
27. Kebbi-Beghdadi C, Cisse O, Greub G. 2011. Permissivity of Vero cells, human pneumocytes and human endometrial cells to Waddlia chondrophila. Microbes Infect 13:566–574. https://doi.org/10.1016/j.micinf.2011.01.020.
28. Corsaro D, Thomas V, Goy G, Venditti D, Radek R, Greub G. 2007. 'Candidatus Rhabdochlamydia crassificans', an intracellular bacterial pathogen of the cockroach Blatta orientalis (Insecta: Blattodea). Syst Appl Microbiol 30:221–228. https://doi.org/10.1016/j.syapm.2006.06.001.
29. Kostanjšek R, Strus J, Drobne D, Avgustin G. 2004. 'Candidatus Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod Porcellio scaber (Crustacea: Isopoda). Int J Syst Evol Microbiol 54:543–549. https://doi.org/10.1099/ijs.0.02802-0.
30. Croxatto A, Rieille N, Kernif T, Bitam I, Aeby S, Péter O, Greub G. 2014. Presence of Chlamydiales DNA in ticks and fleas suggests that ticks are carriers of Chlamydiae. Ticks Tick Borne Dis 5:359–365. https://doi.org/10.1016/j.ttbdis.2013.11.009.
31. Hokynar K, Sormunen JJ, Vesterinen EJ, Partio EK, Lilley T, Timonen V, Panelius J, Ranki A, Puolakkainen M. 2016. Chlamydia-like organisms (CLOs) in Finnish Ixodes ricinus ticks and human skin. Microorganisms 4:28. https://doi.org/10.3390/microorganisms4030028.
32. Vanthournout B, Hendrickx F. 2015. Endosymbiont dominated bacterial communities in a dwarf spider. PLoS One 10:e0117297. https://doi.org/10.1371/journal.pone.0117297.
33. White JA, Styer A, Rosenwald LC, Curry MM, Welch KD, Athey KJ, Chapman EG. 2020. Endosymbiotic bacteria are prevalent and diverse in agricultural spiders. Microb Ecol 79:472–481. https://doi.org/10.1007/s00248-019-01411-w.
34. Pilloux L, Aeby S, Gaümann R, Burri C, Beuret C, Greub G. 2015. The high prevalence and diversity of *Chlamydiales* DNA within *Ixodes ricinus* ticks suggest a role for ticks as reservoirs and vectors of *Chlamydia*-related bacteria. Appl Environ Microbiol 81:8177–8182. https://doi.org/10.1128/AEM.02183-15.
35. Haselkorn TS, Jimenez D, Bashir U, Sallinger E, Queller DC, Strassmann JE, DiSalvo S. 2021. Novel Chlamydiae and Amoebophilus endosymbionts are prevalent in wild isolates of the model social amoeba Dictyostelium discoideum. Environ Microbiol Rep 13:708–719. https://doi.org/10.1111/1758-2229.12985.
36. Lamoth F, Aeby S, Schneider A, Jaton-Ogay K, Vaudaux B, Greub G. 2009. Parachlamydia and Rhabdochlamydia in premature neonates. Emerg Infect Dis 15:2072–2075. https://doi.org/10.3201/eid1512.090267.
37. Lamoth F, Jaton K, Vaudaux B, Greub G. 2011. Parachlamydia and Rhabdochlamydia: emerging agents of community-acquired respiratory infections in children. Clin Infect Dis 53:500–501. https://doi.org/10.1093/cid/cir420.
38. Niemi S, Greub G, Puolakkainen M. 2011. Chlamydia-related bacteria in respiratory samples in Finland. Microbes Infect 13:824–827. https://doi.org/10.1016/j.micinf.2011.04.012.
39. Tolkki L, Hokynar K, Meri S, Panelius J, Puolakkainen M, Ranki A. 2018. Granuloma annulare and morphea: correlation with Borrelia burgdorferi infections and Chlamydia-related bacteria. Acta Derm Venereol 98:355–360. https://doi.org/10.2340/00015555-2831.
40. Sixt BS, Kostanjšek R, Mustedanagic A, Toenshoff ER, Horn M. 2013. Developmental cycle and host interaction of Rhabdochlamydia porcellionis, an intracellular parasite of terrestrial isopods. Environ Microbiol 15:2980–2993. https://doi.org/10.1111/1462-2920.12252.
41. Pillonel T, Bertelli C, Aeby S, de Barsy M, Jacquier N, Kebbi-Beghdadi C, Mueller L, Vouga M, Greub G. 2019. Sequencing the obligate intracellular Rhabdochlamydia helvetica within its tick host Ixodes ricinus to investigate their symbiotic relationship. Genome Biol Evol 11:1334–1344. https://doi.org/10.1093/gbe/evz072.
42. Halter T, Köstlbacher S, Collingro A, Sixt BS, Tönshoff ER, Hendrickx F, Kostanjšek R, Horn M. 2022. Ecology and evolution of chlamydial symbionts of arthropods. ISME Commun 2:45. https://doi.org/10.1038/s43705-022-00124-5.
43. Antoł A, Rojek W, Singh S, Piekarski D, Czarnoleski M. 2019. Hypoxia causes woodlice (Porcellio scaber) to select lower temperatures and impairs their thermal performance and heat tolerance. PLoS One 14:e0220647. https://doi.org/10.1371/journal.pone.0220647.
44. Fieler AM, Rosendale AJ, Farrow DW, Dunlevy MD, Davies B, Oyen K, Xiao Y, Benoit JB. 2021. Larval thermal characteristics of multiple ixodid ticks. Comp Biochem Physiol A Mol Integr Physiol 257:110939. https://doi.org/10.1016/j.cbpa.2021.110939.
45. Matsumoto A, Manire GP. 1970. Electron microscopic observations on the effects of penicillin on the morphology of *Chlamydia psittaci*. J Bacteriol 101:278–285. https://doi.org/10.1128/jb.101.1.278-285.1970.
46. Coles AM, Reynolds DJ, Harper A, Devitt A, Pearce JH. 1993. Low-nutrient induction of abnormal chlamydial development: a novel component of chlamydial pathogenesis? FEMS Microbiol Lett 106:193–200. https://doi.org/10.1111/j.1574-6968.1993.tb05958.x.
47. Kahane S, Friedman MG. 1992. Reversibility of heat shock in Chlamydia trachomatis. FEMS Microbiol Lett 76:25–30.
48. Brockett MR, Liechti GW. 2021. Persistence alters the interaction between *Chlamydia trachomatis* and its host cell. Infect Immun 89:e0068520. https://doi.org/10.1128/IAI.00685-20.
49. Scherler A, Jacquier N, Kebbi-Beghdadi C, Greub G. 2020. Diverse stress-inducing treatments cause distinct aberrant body morphologies in the *Chlamydia*-related bacterium, *Waddlia chondrophila*. Microorganisms 8:89. https://doi.org/10.3390/microorganisms8010089.
50. Sampo A, Matsuo J, Yamane C, Yagita K, Nakamura S, Shouji N, Hayashi Y, Yamazaki T, Yoshida M, Kobayashi M, Ishida K, Yamaguchi H. 2014. High-temperature adapted primitive Protochlamydia found in Acanthamoeba isolated from a hot spring can grow in immortalized human epithelial HEp-2 cells. Environ Microbiol 16:486–497. https://doi.org/10.1111/1462-2920.12266.
51. Yamane C, Yamazaki T, Nakamura S, Matsuo J, Ishida K, Yamazaki S, Oguri S, Shouji N, Hayashi Y, Yoshida M, Yimin Yamaguchi H. 2015. Amoebal endosymbiont Parachlamydia acanthamoebae Bn9 can grow in immortal human epithelial HEp-2 cells at low temperature; an in vitro model system to study chlamydial evolution. PLoS One 10:e0116486. https://doi.org/10.1371/journal.pone.0116486.
52. Antoł A, Berg MP, Verberk WC. 2021. Effects of body size and lung type on desiccation resistance, hypoxia tolerance and thermal preference in two terrestrial isopods species. J Insect Physiol 132:104247. https://doi.org/10.1016/j.jinsphys.2021.104247.
53. Jacquier N, Aeby S, Lienard J, Greub G. 2013. Discovery of new intracellular pathogens by amoebal coculture and amoebal enrichment approaches. J Vis Exp (80):e51055. https://doi.org/10.3791/51055.
54. Volokhov DV, Kong H, George J, Anderson C, Chizhikov VE. 2008. Biological enrichment of *Mycoplasma* agents by cocultivation with permissive cell cultures. Appl Environ Microbiol 74:5383–5391. https://doi.org/10.1128/AEM.00720-08.
55. Gerbal M, Fournier P, Barry P, Mariller M, Odier F, Devauchelle G, Duonor-Cerutti M. 2000. Adaptation of an insect cell line of Spodoptera frugiperda

to grow at 37 degrees C: characterization of an endodiploid clone. In Vitro Cell Dev Biol Anim 36:117–124. https://doi.org/10.1290/1071-2690(2000)036%3C0117:AOAICL%3E2.0.CO;2.
56. Al-Fageeh MB, Marchant RJ, Carden MJ, Smales CM. 2006. The cold-shock response in cultured mammalian cells: harnessing the response for the improvement of recombinant protein production. Biotechnol Bioeng 93:829–835. https://doi.org/10.1002/bit.20789.
57. Bartolo-Aguilar Y, Chávez-Cabrera C, Flores-Cotera LB, Badillo-Corona JA, Oliver-Salvador C, Marsch R. 2022. The potential of cold-shock promoters for the expression of recombinant proteins in microbes and mammalian cells. J Genet Eng Biotechnol 20:173. https://doi.org/10.1186/s43141-022-00455-9.
58. Sixt BS, Siegl A, Müller C, Watzka M, Wultsch A, Tziotis D, Montanaro J, Richter A, Schmitt-Kopplin P, Horn M. 2013. Metabolic features of Protochlamydia amoebophila elementary bodies – a link between activity and infectivity in Chlamydiae. PLoS Pathog 9:e1003553. https://doi.org/10.1371/journal.ppat.1003553.
59. Derré I, Pypaert M, Dautry-Varsat A, Agaisse H. 2007. RNAi screen in Drosophila cells reveals the involvement of the Tom complex in Chlamydia infection. PLoS Pathog 3:1446–1458. https://doi.org/10.1371/journal.ppat.0030155.
60. Elwell C, Engel JN. 2005. Drosophila melanogaster S2 cells: a model system to study Chlamydia interaction with host cells. Cell Microbiol 7:725–739. https://doi.org/10.1111/j.1462-5822.2005.00508.x.
61. Kostanjšek R, Pirc Marolt T. 2015. Pathogenesis, tissue distribution and host response to Rhabdochlamydia porcellionis infection in rough woodlouse Porcellio scaber. J Invertebr Pathol 125:56–67. https://doi.org/10.1016/j.jip.2015.01.001.
62. Eisen RJ, Eisen L, Ogden NH, Beard CB. 2016. Linkages of weather and climate with Ixodes scapularis and Ixodes pacificus (Acari: Ixodidae), enzootic transmission of Borrelia burgdorferi, and Lyme disease in North America. J Med Entomol 53:250–261. https://doi.org/10.1093/jme/tjv199.
63. Power NR, Rugman-Jones PF, Stouthamer R, Ganjisaffar F, Perring TM. 2022. High temperature mortality of Wolbachia impacts the sex ratio of the parasitoid Ooencyrtus mirus (Hymenoptera: Encyrtidae). PeerJ 10:e13912. https://doi.org/10.7717/peerj.13912.
64. van Opijnen T, Breeuwer JA. 1999. High temperatures eliminate Wolbachia, a cytoplasmic incompatibility inducing endosymbiont, from the two-spotted spider mite. Exp Appl Acarol 23:871–881. https://doi.org/10.1023/A:1006363604916.
65. Kaur R, Shropshire JD, Cross KL, Leigh B, Mansueto AJ, Stewart V, Bordenstein SR, Bordenstein SR. 2021. Living in the endosymbiotic world of Wolbachia: a centennial review. Cell Host Microbe 29:879–893. https://doi.org/10.1016/j.chom.2021.03.006.
66. Hague MTJ, Caldwell CN, Cooper BS. 2020. Pervasive effects of Wolbachia on host temperature preference. mBio 11:e01768-20. https://doi.org/10.1128/mBio.01768-20.
67. Haider S, Collingro A, Walochnik J, Wagner M, Horn M. 2008. Chlamydia-like bacteria in respiratory samples of community-acquired pneumonia patients. FEMS Microbiol Lett 281:198–202. https://doi.org/10.1111/j.1574-6968.2008.01099.x.
68. Bell-Sakyi L, Zweygarth E, Blouin EF, Gould EA, Jongejan F. 2007. Tick cell lines: tools for tick and tick-borne disease research. Trends Parasitol 23:450–457. https://doi.org/10.1016/j.pt.2007.07.009.
69. Grieshaber S, Grieshaber N, Yang H, Baxter B, Hackstadt T, Omsland A. 2018. Impact of active metabolism on *Chlamydia trachomatis* elementary body transcript profile and infectivity. J Bacteriol 200:e00065-18. https://doi.org/10.1128/JB.00065-18.
70. Goy G, Croxatto A, Posfay-Barbe KM, Gervaix A, Greub G. 2009. Development of a real-time PCR for the specific detection of Waddlia chondrophila in clinical samples. Eur J Clin Microbiol Infect Dis 28:1483–1486. https://doi.org/10.1007/s10096-009-0804-7.
71. Croxatto A, Greub G. 2010. Early intracellular trafficking of Waddlia chondrophila in human macrophages. Microbiology (Reading) 156:340–355. https://doi.org/10.1099/mic.0.034546-0.
72. Casson N, Entenza JM, Greub G. 2007. Serological cross-reactivity between different *Chlamydia*-like organisms. J Clin Microbiol 45:234–236. https://doi.org/10.1128/JCM.01867-06.
73. Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48. https://doi.org/10.18637/jss.v067.i01.

6 Discussion and conclusion

Despite being reported as one of the most diverse family of the *Chlamydiales* order [20], the *Rhabdochlamydiaceae* have received little interest from the research community. Only one isolate of this family could be cultured and it has not been the subject of any publication after its initial description in 2013 [60]. Sequencing those organisms proved to be easier than culturing them, as the genomes of three species of *Rhabdochlamydiaceae* could be assembled [38, 75]. Aside from their presence in the environment, the *Rhabdochlamydiaceae* were also detected in clinical samples [95, 97, 98], questioning whether some members of this family could be human pathogens. We thus aimed to better determine the host range of *R. porcellionis* and to assess whether it is compatible with the hypothesized pathogenic role of this bacterium. We also aimed at describing other rhabdochlamydia species to gain a better understanding of the evolution and biology of this family.

In this work, we determined that *R. porcellionis* fails to grow at 37 °C or 33 °C and cannot recover from short exposure to 37 °C. The bacterium could however grow in mammalians cells incubated at 28 °C, but the long doubling times and the presence of aberrant bodies suggest that those cell lines do not offer ideal conditions for the growth of *R. porcellionis*. Our sequencing efforts led to the assembly of 7 new species of chlamydiae, obtained from arthropods of 6 different orders. Finally, in the course of this project, we developed a tool to ease the analysis of bacterial genomes and used it to gain insight into the evolution of the *Rhabdochlamydiaceae* and *Simkaniaceae* families.

6.1 Pathogenicity of the *Rhabdochlamydiaceae*

One of the main aim of this project was to gain a better understanding of the risk posed by the *Rhabdochlamydiaceae* to human health. Previous studies suggested a potential role in respiratory and skin diseases [95, 97, 98, 135, 136], but the level of evidence was low, with no statistical association between the presence of the bacterium and the diseases of interest. Our results showing that *R. porcellionis* cannot grow at either 33 °C or 37 °C and its apparent inability to withstand short exposures at 37 °C are hardly compatible with the hypothesis of the pathogenicity of this particular species. In particular, those results make unlikely an involvement of this bacterium in lower respiratory tract infections or its blood-borne dissemination after a tick bite. Interestingly, a similar deleterious effect of an incubation at 37 °C was also observed for *Mycobacterium leprae*, the causative agent of leprosy [205]. The temperature sensitivity of the bacterium is responsible for the marked tropism of the disease for the coldest regions of the human body [205]. However, *Mycobacterium leprae* still has a reported optimal growth temperature of 33 °C [205]. The failure of *R. porcellionis* to grow at this temperature makes even skin or upper respiratory tract infections unlikely. As the thermal preferences of *Porcellio scaber* [206] seem to be similar to that of most ixodid ticks [207, 208], it is likely that the tick-borne rhabdochlamydiae share the same temperature sensitivity as *R. porcellionis*. In addition, the short evolutionary distances between the newly sequenced rhabdochlamydiae and *R. porcellionis* suggests overlapping host ranges, and likely, common thermal preferences. However, this reasoning only applies to arthropod-borne rhabdochlamydia; more distantly related species of *Rhabdochlamydiaceae* might have a higher thermal tolerance. Such a variation in temperature sensitivity has indeed already been observed in the *Parachlamydiaceae* family. *Parachlamydia acanthamoebae* was demonstrated to be unable to grow at 37 °C [111], while another *Parachlamydiaceae* isolated from amoebae thriving in hot springs grew well at this temperature [16]. Unfortunately, the heat-resistant *Parachlamydiaceae* was not sequenced. Including heat-resistant and heat-sensitive parachlamydiae in the

comparative genomics analysis would indeed have given hints as to the minimal evolutionary distance necessary to observe differences in thermal sensitivity. Even if the arthropod-borne rhabdochlamydiae likely share the same thermal tolerance as *R. porcellionis*, the *Rhabdochlamydiaceae* family still contains members whose host is currently unknown [21, 75]. It is therefore impossible to infer the non-pathogenicity of the whole family from our findings on *R. porcellionis*.

Even though this work failed to clarify the pathogenic role of the *Rhabdochlamydiaceae* family as a whole, it is worth questioning the premise that it might be pathogenic at all. As mentioned in the introduction, the prior probability of the pathogenicity of a newly discovered *Chlamydiales* was high in the 1990s and early 2000s, when the order was mostly composed of mammalian pathogens. The explosion of new *Chlamydiales* discovered in the environment in the past decade [20, 21] would tend to lower this probability, as the prevalence of proven pathogens in the order decreases. Had the *Rhabdochlamydiaceae* family been discovered in the past years, its discovery would have been a mere addition to an already densely populated order and the question of its pathogenicity would likely never have arisen. The absence of any statistical association with human disease makes a pathogenic role even more unlikely and would rather suggest that humans are frequently exposed to this bacterium without consequences.

In the past, the isolation of new emerging pathogens seems to have been successful when efforts were made to identify a causative agent in patients suffering of a suspected infectious disease [209]. The formal identification of *Tropheryma whipplei* and *Candidatus* Neoehrlichia mikurensis were indeed achieved by eubacterial PCR in patients suffering from Whipple's disease [210] and of a bacteremia of unknown cause [211, 212]. More recently, the SARS-CoV-2 was rapidly identified as the causative agent of a cluster of pneumonia [213]. Similarly, the Bourbon virus, an hypothesized tick-borne virus unknown as of 2014 was identified in a matter of months after its first fatality [214]. As modern diagnostic techniques appear sensitive enough to identify new emerging pathogens, it appears unlikely that the evidence for the pathogenicity of the *Rhabdochlamydiaceae*

would be so weak if this family caused outbreaks or severe diseases. It is however possible that, similarly to *Legionella pneumophila*, some *Rhabdochlamydiaceae* cause self-limited mild diseases such as the Pontiac fever [215].

The hypothesis that ticks could be vectors of chlamydial infection [65, 66] is also questionable. Even though ticks are vectors of multiple human diseases, they also carry numerous non-pathogenic bacteria as part of their physiological microbiota [143]. Indeed, ticks have a complex microbiota that notably include bacteria from genera known to contain human pathogens, such as *Coxiella* [216], *Francisella* [217] and *Rickettsia* [149] and while those tick symbionts are closely related to human pathogens, they do not appear to have any role in human disease [149, 216, 218]. The presence of a bacteria in ticks is therefore not a sufficient condition to deduce that it can cause tick borne diseases, all the more so if its pathogenicity is still unproven. Moreover, the contrast between the high copy numbers observed in ticks collected by flagging [66] and the low copy numbers observed in questing ticks suggests that the presence of *Rhabdochlamydiaceae* might cause a fitness cost on its host and lowers the risk of an infected tick biting a human. The hypothesis that rhabdochlamydiae cause tick-borne diseases thus appear very unlikely, while the hypothesis of its pathogenicity seems to have more to do with the timing of its discovery than with an actual pathogenic role.

6.2 Ecology and evolution

Remarkably, it appears that the *Rhabdochlamydiaceae* and the *Simkaniaceae* independently evolved to infect arthropods. The *Simkaniaceae* includes at least two members able to replicate in protists: *Simkania negevensis* whose natural reservoir is still unknown but that was demonstrated to grow well in amoebae [151] and *Neptunochlamydia vexilliferae*, that was discovered in marine amoebae [17]. The branching order in the *Simkaniaceae* family (Fig. 4.1) suggests that *Fritschea bemisiae* evolved from a protist symbiont and likely underwent a reductive genome evolution after its specialization for

arthropods. As for the *Rhabdochlamydiaceae*, the recent discovery of a closely related bacteria in ciliates [80] and the evidence that the ancestor of extant chlamydia was a protist symbiont [7] also suggest that the common ancestor of the *Simkaniaceae* and the *Rhabdochlamydiaceae* infected protists and that the ability to infect arthropods appeared after the divergence of the two families.

Another surprising finding was the presence of the pyridoxal synthesis operon in *Fritschea bemisiae* that was likely acquired by horizontal gene transfer from a *Cardinum* symbiont and suggests a transition from a parasitic to a beneficial symbiotic relationship. Lateral gene transfers have indeed been reported as drivers of evolution towards a beneficial symbiotic relationship [219]. Moreover, beneficial symbiosis has already been documented for other chlamydiae in the form of a defensive symbiosis protecting against infections by other pathogens [23, 24]. In the case of *Fritschea bemisiae*, the ability to synthesize pyridoxal would rather suggest a role as a metabolic symbiont. Arthropods with a restricted diet indeed need bacterial symbionts to synthesize the nutrients absent from their diet [217]. This is notably the case for sap-feeding insects such as *Bemisia tabaci* [220]. *Fritschea bemisiae* might thus be the first documented chlamydiae playing the role of a metabolic symbiont. This would however need to be demonstrated in-vivo, by comparing the fitness of *Bemisia tabaci* with and without *Fritschea bemisia*. It might indeed be the case that the acquisition of this metabolic module increases the fitness of the bacteria without any effect on the host.

6.3 Future perspectives

Chlamydiae in ticks

The low prevalence of *Chlamydiales* in collected ticks makes it difficult to reach conclusive evidence about their potential association with tick-borne pathogens. To circumvent this issue [65, 66], a case-control approach could be used to assess their potential association with tick-borne pathogens. In particular, frozen samples from the study by Pilloux et al

[66] could be reused to compare the prevalence of specific tick-borne pathogens in ticks already known to be positive or negative for *Rhabdochlamydia*, although the pooling of ticks would complicate the analysis. Moreover, having a prior knowledge of the prevalence of tick-borne pathogens in Swiss ticks [147] would allow the estimation of a sample size necessary to reach a satisfying statistical power and to be more confident in the conclusion.

Ecology and evolution

Our knowledge of the *Rhabdochlamydiaceae* family is currently limited to arthropod-borne species. As this family appears to also be prevalent in freshwater [15, 21], the isolation of other rabdochlamydia species could be attempted in alternative culture systems such ciliates, using an approach similar to what Corsaro et al did [15]. Isolating and sequencing rhabochlamydia growing in protists would indeed be invaluable to better understand the evolution of chlamydia and how they specialize for different hosts. In particular, this would allow the comparison with the *Simkaniaceae* family: as both families appear to have independently evolved to infect arthropods, it would be of great interest to determine if this evolution towards a common host shaped their genomes in similar ways.

To study the different stages of infection and the tissue distribution of the bacteria, it would be possible to attempt an artificial infection of laboratory-reared ticks with *R. porcellionis*, assuming that it is able to multiple in this host. The inability of *R. porcellionis* to grow in the IRE/CTVM19 cell line might be more related to the cell type than to the host species. *R. porcellionis* indeed clusters with other rhabdochlamydia species growing in ticks and butterflies (Supplementary Fig. 4.6), suggesting that those closely related species might have the same host range.

Temperature sensitivity

The results on the temperature sensitivity of *R. porcellionis* raise additional questions that would be interesting to investigate. For instance, the striking effect of the incubation temperature on IFU count could be explained both by starvation due to a faster

metabolism at higher temperatures and by a stronger activation of the T3SS. Indeed, EBs were demonstrated to have glycogen stores [45, 46] that are hypothesized to serve as an energy source in the extracellular environment. A faster metabolic activity could therefore lead to the depletion of metabolic substrates and to a loss of infectivity [40]. To assess this, it would be possible to measure the quantity of glycogen in samples incubated at either 37 °C or room temperature and determine whether glycogen is indeed lost faster at higher incubation temperatures.

The loss of infectivity could also be related to a stronger activation of the T3SS at 37 °C than at a room temperature [221]. As EBs are pre-packaged with T3SS effectors [42, 44], the spurious activation of secretion could lead to an eventual depletion of effectors and a loss of infectivity. Using the same buffers as in Jamison et al [221], the presence of proteins in the supernatant could be compared at both incubation temperature to check whether this correlates with the IFU count. Moreover, if higher incubation temperatures indeed activate the T3SS of EBs, this would allow the identification of yet unknown early T3SS effectors. The secreted effectors could indeed be retrieved directly from the supernatant and identified by mass spectrometry, while avoiding the problem of contamination by host cell proteins [222].

Finally, it would also possible to attempt the generation of a temperature-resistant strain of *R. porcellionis* by gradually increasing the incubation temperature of infected Sf9. The generation of variants was indeed successful in an evolution experiment forcing the vertical and horizontal transmission of *Parachlamydia acanthamoebae* between host cells [223] and in the case of *R. porcellionis*, could lead to the identification of the determinants of temperature sensitivity. If this experiment is successful, it would provide hints at how fast *Chlamydiales* can adapt to environmental changes and might even provide clues as to which incubation temperatures to try for the isolation of sequenced but yet uncultured chlamydiae.

Bibliography

1. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2020. URL: `https://www.cdc.gov/std/statistics/2020/overview.htm#Chlamydia` (visited on 12/29/2022).
2. Burillo A and Bouza E. *Chlamydophila pneumoniae*. Infectious Disease Clinics 2010;24:61–71.
3. Polkinghorne A, Hanger J, and Timms P. Recent advances in understanding the biology, epidemiology and control of chlamydial infections in koalas. Veterinary microbiology 2013;165:214–23.
4. Livingstone M, Wattegedera SR, Palarea-Albaladejo J, et al. Efficacy of two *Chlamydia abortus* subcellular vaccines in a pregnant ewe challenge model for ovine enzootic abortion. Vaccines 2021;9:898.
5. Hogerwerf L, De Gier B, Baan B, and Van Der Hoek W. *Chlamydia psittaci* (psittacosis) as a cause of community-acquired pneumonia: a systematic review and meta-analysis. Epidemiology & Infection 2017;145:3096–105.
6. Horn M. *Chlamydiae* as symbionts in eukaryotes. Annual review of microbiology 2008;62:113–31.
7. Collingro A, Köstlbacher S, and Horn M. *Chlamydiae* in the environment. Trends in Microbiology 2020;28:877–88.
8. Chaiwattanarungruengpaisan S, Thongdee M, Anuntakarun S, et al. A new species of *Chlamydia* isolated from Siamese crocodiles (*Crocodylus siamensis*). Plos one 2021;16:e0252081.
9. Staub E, Marti H, Biondi R, et al. Novel *Chlamydia* species isolated from snakes are temperature-sensitive and exhibit decreased susceptibility to azithromycin. Scientific reports 2018;8:1–14.
10. Stokes HS, Berg ML, and Bennett AT. A review of chlamydial infections in wild birds. Pathogens 2021;10:948.
11. Vorimore F, Hölzer M, Liebler-Tenorio E, et al. Evidence for the existence of a new genus *Chlamydiifrater* gen. nov. inside the family *Chlamydiaceae* with two new species isolated from flamingo (*Phoenicopterus roseus*): *Chlamydiifrater phoenicopteri* sp. nov. and *Chlamydiifrater volucris* sp. nov. Systematic and Applied Microbiology 2021;44:126200.

12. Everett KD, Thao M, Horn M, Dyszynski GE, and Baumann P. Novel chlamydiae in whiteflies and scale insects: endosymbionts '*Candidatus* Fritschea bemisiae' strain Falk and '*Candidatus* Fritschea eriococci' strain Elm. International Journal of Systematic and Evolutionary Microbiology 2005;55:1581–7.

13. Corsaro D, Thomas V, Goy G, Venditti D, Radek R, and Greub G. '*Candidatus* Rhabdochlamydia crassificans', an intracellular bacterial pathogen of the cockroach *Blatta orientalis* (Insecta: Blattodea). Systematic and Applied Microbiology 2007;30:221–8.

14. Kostanjšek R, Štrus J, Drobne D, and Avguštin G. '*Candidatus* Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod *Porcellio scaber* (Crustacea: Isopoda). International journal of systematic and evolutionary microbiology 2004;54:543–9.

15. Corsaro D and Venditti D. Detection of *Chlamydiae* from freshwater environments by PCR, amoeba coculture and mixed coculture. Research in Microbiology 2009;160:547–52.

16. Sampo A, Matsuo J, Yamane C, et al. High-temperature adapted primitive *Protochlamydia* found in *Acanthamoeba* isolated from a hot spring can grow in immortalized human epithelial HEp-2 cells. Environmental Microbiology 2014;16:486–97.

17. Pizzetti I, Schulz F, Tyml T, et al. Chlamydial seasonal dynamics and isolation of '*Candidatus* Neptunochlamydia vexilliferae' from a Tyrrhenian coastal lake. Environmental Microbiology 2016;18:2405–17.

18. Taylor-Brown A, Pillonel T, Greub G, Vaughan L, Nowak B, and Polkinghorne A. Metagenomic analysis of fish-associated *Ca.* Parilichlamydiaceae reveals striking metabolic similarities to the terrestrial *Chlamydiaceae*. Genome biology and evolution 2018;10:2587–95.

19. Dharamshi JE, Tamarit D, Eme L, et al. Marine sediments illuminate chlamydiae diversity and evolution. Current Biology 2020;30:1032–1048.e7.

20. Lagkouvardos I, Weinmaier T, Lauro FM, Cavicchioli R, Rattei T, and Horn M. Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the *Chlamydiae*. The ISME journal 2014;8:115–25.

21. Köstlbacher S, Collingro A, Halter T, Schulz F, Jungbluth SP, and Horn M. Pangenomics reveals alternative environmental lifestyles among chlamydiae. Nature Communications 2021;12:1–15.

22. Collingro A, Köstlbacher S, Mussmann M, Stepanauskas R, Hallam SJ, and Horn M. Unexpected genomic features in widespread intracellular bacteria: evidence for motility of marine chlamydiae. The ISME journal 2017;11:2334–44.

23. Arthofer P, Delafont V, Willemsen A, Panhölzl F, and Horn M. Defensive symbiosis against giant viruses in amoebae. Proceedings of the National Academy of Sciences 2022;119:e2205856119.

24. Maita C, Matsushita M, Miyoshi M, et al. Amoebal endosymbiont *Neochlamydia* protects host amoebae against *Legionella pneumophila* infection by preventing *Legionella* entry. Microbes and Infection 2018;20:236–44.

25. Okubo T, Matsushita M, Nakamura S, Matsuo J, Nagai H, and Yamaguchi H. *Acanthamoeba* S13WT relies on its bacterial endosymbiont to backpack human pathogenic bacteria and resist *Legionella* infection on solid media. Environmental microbiology reports 2018;10:344–54.

26. Moulder JW. Interaction of chlamydiae and host cells in vitro. Microbiological Reviews 1991;55:143–90.

27. Elwell C, Mirrashidi K, and Engel J. Chlamydia cell biology and pathogenesis. Nature Reviews Microbiology 2016;14:385–400.

28. Omsland A, Sixt BS, Horn M, and Hackstadt T. Chlamydial metabolism revisited: interspecies metabolic variability and developmental stage-specific physiologic activities. FEMS Microbiology Reviews 2014;38:779–801.

29. Ardissone S, Scherler A, Pillonel T, Martin V, Kebbi-Beghdadi C, and Greub G. Transcriptional landscape of *Waddlia chondrophila* aberrant bodies induced by iron starvation. Microorganisms 2020;8:1848.

30. Scherler A, Jacquier N, Kebbi-Beghdadi C, and Greub G. Diverse stress-inducing treatments cause distinct aberrant body morphologies in the *Chlamydia*-related bacterium, *Waddlia chondrophila*. Microorganisms 2020;8:E89.

31. Beatty WL, Byrne GI, and Morrison RP. Morphologic and antigenic characterization of interferon gamma-mediated persistent *Chlamydia trachomatis* infection in vitro. Proceedings of the National Academy of Sciences 1993;90:3998–4002.

32. Wyrick PB. *Chlamydia trachomatis* persistence in vitro: an overview. The Journal of Infectious Diseases 2010;201:S88–S95.

33. Greub G and Raoult D. Crescent bodies of *Parachlamydia acanthamoeba* and its life cycle within *Acanthamoeba polyphaga*: an electron micrograph study. Applied and environmental microbiology 2002;68:3076–84.

34. Lienard J, Croxatto A, Prod'hom G, and Greub G. *Estrella lausannensis*, a new star in the *Chlamydiales* order. Microbes and Infection 2011;13:1232–41.

35. Rusconi B, Lienard J, Aeby S, Croxatto A, Bertelli C, and Greub G. Crescent and star shapes of members of the *Chlamydiales* order: impact of fixative methods. Antonie Van Leeuwenhoek 2013;104:521–32.

36. Hatch TP. Disulfide cross-linked envelope proteins: the functional equivalent of peptidoglycan in chlamydiae? Journal of Bacteriology 1996;178:1–5.

37. Collingro A, Tischler P, Weinmaier T, et al. Unity in variety—the pan-genome of the *Chlamydiae*. Molecular biology and evolution 2011;28:3253–70.

38. Pillonel T, Bertelli C, Aeby S, et al. Sequencing the obligate intracellular *Rhabdochlamydia helvetica* within its tick host *Ixodes ricinus* to investigate their symbiotic relationship. Genome biology and evolution 2019;11:1334–44.

39. Pilhofer M, Aistleitner K, Ladinsky MS, König L, Horn M, and Jensen GJ. Architecture and host interface of environmental chlamydiae revealed by electron cryotomography. Environmental microbiology 2014;16:417–29.

40. Sixt BS, Siegl A, Müller C, et al. Metabolic features of *Protochlamydia amoebophila* elementary bodies – A link between activity and infectivity in chlamydiae. PLOS Pathogens 2013;9:e1003553.

41. Grieshaber S, Grieshaber N, Yang H, Baxter B, Hackstadt T, and Omsland A. Impact of active metabolism on *Chlamydia trachomatis* elementary body transcript profile and infectivity. Journal of Bacteriology 2018;200:e00065–18.

42. König L, Siegl A, Penz T, et al. Biphasic metabolism and host interaction of a chlamydial symbiont. mSystems 2017;2:e00202–16.

43. Haider S, Wagner M, Schmid MC, et al. Raman microspectroscopy reveals long-term extracellular activity of chlamydiae. Molecular Microbiology 2010;77:687–700.

44. Saka HA, Thompson JW, Chen YS, et al. Quantitative proteomics reveals metabolic and pathogenic properties of *Chlamydia trachomatis* developmental forms. Molecular microbiology 2011;82:1185–203.

45. Colpaert M, Kadouche D, Ducatez M, et al. Conservation of the glycogen metabolism pathway underlines a pivotal function of storage polysaccharides in *Chlamydiae*. Communications Biology 2021;4:296.

46. Gehre L, Gorgette O, Perrinet S, et al. Sequestration of host metabolism by an intracellular pathogen. Elife 2016;5:e12552.

47. Triboulet S and Subtil A. Make it a sweet home: Responses of *Chlamydia trachomatis* to the challenges of an intravacuolar lifestyle. Microbiology Spectrum 2019;7:7–2.

48. Stephens RS, Kalman S, Lammel C, et al. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 1998;282:754–9.

49. Heinzen RA and Hackstadt T. The *Chlamydia trachomatis* parasitophorous vacuolar membrane is not passively permeable to low-molecular-weight compounds. Infection and immunity 1997;65:1088–94.

50. Grieshaber S, Swanson JA, and Hackstadt T. Determination of the physical environment within the *Chlamydia trachomatis* inclusion using ion-selective ratiometric probes. Cellular microbiology 2002;4:273–83.

51. Fisher DJ, Fernández RE, Adams NE, and Maurelli AT. Uptake of biotin by *Chlamydia* Spp. through the use of a bacterial transporter (BioY) and a host-cell transporter (SMVT). PLoS One 2012;9:e46052.

52. Dharamshi JE, Köstlbacher S, Schön ME, Collingro A, Ettema TJ, and Horn M. Gene gain facilitated endosymbiotic evolution of Chlamydiae. Nature Microbiology 2023;8:40–54.

53. Gitsels A, Sanders N, and Vanrompay D. Chlamydial Infection From Outside to Inside. Frontiers in Microbiology 2019;10.

54. Sixt BS, Bastidas RJ, Finethy R, et al. The *Chlamydia trachomatis* inclusion membrane protein CpoS counteracts STING-mediated cellular surveillance and suicide programs. Cell host & microbe 2017;21:113–21.

55. Faris R, Merling M, Andersen SE, Dooley CA, Hackstadt T, and Weber MM. *Chlamydia trachomatis* CT229 subverts Rab GTPase-dependent CCV trafficking pathways to promote chlamydial infection. Cell reports 2019;26:3380–90.

56. Radek R. Light and electron microscopic study of *Rickettsiella* species from the cockroach *Blatta orientalis*. Journal of invertebrate pathology 2000;76:249–56.

57. Drobne D, Štrus J, Žnidaršič N, and Zidar P. Morphological description of bacterial infection of digestive glands in the terrestrial isopod *Porcellio scaber* (Isopoda, Crustacea). Journal of invertebrate pathology 1999;73:113–9.

58. Kostanjšek R and Marolt TP. Pathogenesis, tissue distribution and host response to *Rhabdochlamydia porcellionis* infection in rough woodlouse *Porcellio scaber*. Journal of invertebrate pathology 2015;125:56–67.

59. Pillonel T, Bertelli C, Salamin N, and Greub G. Taxogenomics of the order *Chlamydiales*. International journal of systematic and evolutionary microbiology 2015;65: 1381–1393.

60. Sixt BS, Kostanjšek R, Mustedanagic A, Toenshoff ER, and Horn M. Developmental cycle and host interaction of *Rhabdochlamydia porcellionis*, an intracellular parasite of terrestrial isopods. Environmental Microbiology 2013;15:2980–93.

61. Morel G. Studies on *Porochlamydia buthi* gn, sp. n., an intracellular pathogen of the scorpion *Buthus occitanus*. Journal of Invertebrate Pathology 1976;28:167–75.

62. Götz P. "*Rickettsiella chironomi*": an unusual bacterial pathogen which reproduces by multiple cell division. Journal of Invertebrate Pathology 1972;20:22–30.

63. Federici BA. Reproduction and morphogenesis of *Rickettsiella chironomi*, an unusual intracellular procaryotic parasite of midge larvae. Journal of bacteriology 1980;143:995–1002.

64. Morel G. Isolement de deux chlamydiales (rickettsies) chez un arachnide: L'araignée *Pisaura mirabilis* Cl. Two chlamydiales (rickettsias) in an arachnida: The spider *Pisaura mirabilis* Cl. Experientia 1978;34:344–6.

65. Croxatto A, Rieille N, Kernif T, et al. Presence of *Chlamydiales* DNA in ticks and fleas suggests that ticks are carriers of *Chlamydiae*. Ticks and tick-borne diseases 2014;5:359–65.

66. Pilloux L, Aeby S, Gaümann R, Burri C, Beuret C, and Greub G. The high prevalence and diversity of *Chlamydiales* DNA within *Ixodes ricinus* ticks suggest a role for ticks as reservoirs and vectors of *Chlamydia*-related bacteria. Applied and Environmental Microbiology 2015;81:8177–82.

67. Hokynar K, Sormunen JJ, Vesterinen EJ, et al. *Chlamydia*-like organisms (CLOs) in finnish *Ixodes ricinus* ticks and human skin. Microorganisms 2016;4:28.

68. Burnard D, Weaver H, Gillett A, Loader J, Flanagan C, and Polkinghorne A. Novel *Chlamydiales* genotypes identified in ticks from Australian wildlife. Parasites & Vectors 2017;10:46.

69. Vanthournout B and Hendrickx F. Endosymbiont dominated bacterial communities in a dwarf spider. PLoS One 2015;10:e0117297.

70. Zhang YK, Yu ZJ, Wang D, Bronislava V, Branislav P, and Liu JZ. The bacterial microbiome of field-collected *Dermacentor marginatus* and *Dermacentor reticulatus* from Slovakia. Parasites & vectors 2019;12:1–10.

71. Brokatzky D, Kretz O, and Häcker G. Apoptosis functions in defense against infection of mammalian cells with environmental chlamydiae. Infection and immunity 2020;88:e00851–19.

72. Sixt BS, Hiess B, König L, and Horn M. Lack of effective anti-apoptotic activities restricts growth of *Parachlamydiaceae* in insect cells. PloS One 2012;7:e29565.

73. Fritschi J, Marti H, Seth-Smith HMB, et al. Prevalence and phylogeny of *Chlamydiae* and hemotropic mycoplasma species in captive and free-living bats. BMC microbiology 2020;20:182.

74. Hokynar K, Vesterinen E, Lilley T, et al. Molecular evidence of chlamydia-like organisms in the feces of *Myotis daubentonii* bats. Applied and Environmental Microbiology 2017;83:e02951–16.

75. Halter T, Koestlbacher S, Collingro A, et al. Ecology and evolution of chlamydial symbionts of arthropods. ISME Communications 2022;2:1–11.

76. Casson N, Entenza JM, and Greub G. Serological cross-reactivity between different *Chlamydia*-like organisms. Journal of clinical microbiology 2007;45:234–6.

77. Corsaro D, Feroldi V, Saucedo G, Ribas F, Loret JF, and Greub G. Novel *Chlamydiales* strains isolated from a water treatment plant. Environmental Microbiology 2009;11:188–200.

78. Coulon C, Eterpi M, Greub G, Collignon A, McDonnell G, and Thomas V. Amoebal host range, host-free survival and disinfection susceptibility of environmental *Chlamydiae* as compared to *Chlamydia trachomatis*. FEMS immunology and medical microbiology 2012;64:364–73.

79. Haselkorn TS, Jimenez D, Bashir U, et al. Novel Chlamydiae and *Amoebophilus* endosymbionts are prevalent in wild isolates of the model social amoeba *Dictyostelium discoideum*. Environmental Microbiology Reports 2021;13:708–19.

80. Wang R, Sun R, Zhang Z, Vannini C, Di Giuseppe G, and Liang A. "*Candidatus* Euplotechlamydia quinta," a novel chlamydia-like bacterium hosted by the ciliate *Euplotes octocarinatus* (Ciliophora, Spirotrichea). Journal of Eukaryotic Microbiology 2022:e12945.

81. Taylor-Brown A, Madden D, and Polkinghorne A. Culture-independent approaches to chlamydial genomics. Microbial genomics 2018;4.

82. McCutcheon JP and Moran NA. Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology 2012;10:13–26.

83. Moran NA and Bennett GM. The tiniest tiny genomes. Annual review of microbiology 2014;68:195–215.

84. Moran NA, McCutcheon JP, and Nakabachi A. Genomics and evolution of heritable bacterial symbionts. Annual Review of Genetics 2008;42:165–90.

85. Burton MJ and Mabey DC. The global burden of trachoma: a review. PLoS neglected tropical diseases 2009;3:e460.

86. Taylor-Robinson D. The discovery of *Chlamydia trachomatis*. Sexually Transmitted Infections 2017;93:10–10.

87. Azuma Y, Hirakawa H, Yamashita A, et al. Genome sequence of the cat pathogen, *Chlamydophila felis*. DNA research 2006;13:15–23.

88. Laroucau K, Vorimore F, Aaziz R, et al. *Chlamydia buteonis*, a new *Chlamydia* species isolated from a red-shouldered hawk. Systematic and applied microbiology 2019;42:125997.

89. Clarke IN. Evolution of *Chlamydia trachomatis*. Annals of the New York Academy of Sciences 2011;1230:E11–E18.

90. Amann R, Springer N, Schönhuber W, et al. Obligate intracellular bacterial parasites of acanthamoebae related to Chlamydia spp. Applied and environmental microbiology 1997;63:115–21.

91. Birtles R, Rowbotham T, Storey C, Marrie T, and Raoult D. *Chlamydia*-like obligate parasite of free-living amoebae. Lancet (British edition) 1997;349:925–6.

92. Greub G and Raoult D. *Parachlamydiaceae*: potential emerging pathogens. Emerging infectious diseases 2002;8:626.

93. Greub G. *Parachlamydia acanthamoebae*, an emerging agent of pneumonia. Clinical Microbiology and Infection 2009;15:18–28.

94. Greub G, Boyadjiev I, Scola BL, Raoult D, and Martin C. Serological hint suggesting that *Parachlamydiaceae* are agents of pneumonia in polytraumatized intensive care patients. Annals of the New York Academy of Sciences 2003;990:311–9.

95. Haider S, Collingro A, Walochnik J, Wagner M, and Horn M. *Chlamydia*-like bacteria in respiratory samples of community-acquired pneumonia patients. FEMS microbiology letters 2008;281:198–202.

96. Hokynar K, Kurkela S, Nieminen T, et al. *Parachlamydia acanthamoebae* detected during a pneumonia outbreak in southeastern Finland, in 2017–2018. Microorganisms 2019;7:141.

97. Lamoth F, Aeby S, Schneider A, Jaton-Ogay K, Vaudaux B, and Greub G. *Parachlamydia* and *Rhabdochlamydia* in premature neonates. Emerging Infectious Diseases 2009;15:2072–5.

98. Lamoth F, Jaton K, Vaudaux B, and Greub G. *Parachlamydia* and *Rhabdochlamydia*: emerging agents of community-acquired respiratory infections in children. Clinical Infectious Diseases 2011;53:500–1.

99. Lienard J, Croxatto A, Aeby S, et al. Development of a new *Chlamydiales*-specific real-time PCR and its application to respiratory clinical samples. Journal of Clinical Microbiology 2011;49:2637–42.

100. Casson N, Posfay-Barbe KM, Gervaix A, and Greub G. New diagnostic real-time PCR for specific detection of *Parachlamydia acanthamoebae* DNA in clinical samples. Journal of Clinical Microbiology 2008;46:1491–3.

101. Lohr M, Prohl A, Ostermann C, Diller R, Greub G, and Reinhold P. Effect of *Parachlamydia acanthamoebae* on pulmonary function parameters in a bovine respiratory model. Veterinary Journal (London, England: 1997) 2016;213:9–15.

102. Reinhold P, Ostermann C, Liebler-Tenorio E, et al. A bovine model of respiratory *Chlamydia psittaci* infection: challenge dose titration. PloS one 2012;7:e30125.

103. Casson N, Entenza JM, Borel N, Pospischil A, and Greub G. Murine model of pneumonia caused by *Parachlamydia acanthamoebae*. Microbial pathogenesis 2008;45:92–7.

104. Pilloux L, Casson N, Sommer K, et al. Severe pneumonia due to *Parachlamydia acanthamoebae* following intranasal inoculation: a mice model. Microbes and Infection. Special issue on intracellular bacteria 2015;17:755–60.

105. Casadevall A. The pathogenic potential of a microbe. Msphere 2017;2:e00015–17.

106. Yoshida SI and Mizuguchi Y. Multiplication of *Legionella pneumophila* Philadelphia-1 in cultured peritoneal macrophages and its correlation to susceptibility of animals. Canadian journal of microbiology 1986;32:438–42.

107. Greub G, Mege JL, and Raoult D. *Parachlamydia acanthamoeba* enters and multiplies within human macrophages and induces their apoptosis. Infection and immunity 2003;71:5979–85.

108. Casson N, Medico N, Bille J, and Greub G. *Parachlamydia acanthamoebae* enters and multiplies within pneumocytes and lung fibroblasts. Microbes and Infection 2006;8:1294–300.

109. Vouga M, Diabi H, Boulos A, Baud D, Raoult D, and Greub G. Antibiotic susceptibility of *Neochlamydia hartmanellae* and *Parachlamydia acanthamoebae* in amoebae. Microbes and infection 2015;17:761–5.

110. Hayashi Y, Nakamura S, Matsuo J, et al. Host range of obligate intracellular bacterium *Parachlamydia acanthamoebae*. Microbiology and Immunology 2010;54:707–13.

111. Yamane C, Yamazaki T, Nakamura S, et al. Amoebal endosymbiont *Parachlamydia acanthamoebae* Bn9 can grow in immortal human epithelial HEp-2 cells at low temperature; an in vitro model system to study chlamydial evolution. PloS One 2015;10:e0116486.

112. Fukumoto T, Matsuo J, Okubo T, et al. *Acanthamoeba* containing endosymbiotic chlamydia isolated from hospital environments and its potential role in inflammatory exacerbation. BMC microbiology 2016;16:1–8.

113. Roger T, Casson N, Croxatto A, et al. Role of MyD88 and Toll-like receptors 2 and 4 in the sensing of *Parachlamydia acanthamoebae*. Infection and immunity 2010;78:5195–201.

114. Sommer K, Njau F, Wittkop U, et al. Identification of high-and low-virulent strains of *Chlamydia pneumoniae* by their characterization in a mouse pneumonia model. FEMS Immunology & Medical Microbiology 2009;55:206–14.

115. Dilbeck P, Evermann J, Crawford T, et al. Isolation of a previously undescribed rickettsia from an aborted bovine fetus. Journal of clinical microbiology 1990;28:814–6.

116. Rurangirwa FR, Dilbeck PM, Crawford TB, McGuire TC, and McElwain TF. Analysis of the 16S rRNA gene of micro-organism WSU 86-1044 from an aborted bovine foetus reveals that it is a member of the order *Chlamydiales*: proposal of *Waddliaceae* fam. nov., *Waddlia chondrophila* gen. nov., sp. nov. International Journal of Systematic Bacteriology 1999;49 Pt 2:577–81.

117. Henning K, Schares G, Granzow H, et al. *Neospora caninum* and *Waddlia chondrophila* strain 2032/99 in a septic stillborn calf. Veterinary microbiology 2002;85: 285–292.

118. Dille S, Kleinschnitz EM, Kontchou CW, Nölke T, and Häcker G. In contrast to *Chlamydia trachomatis*, *Waddlia chondrophila* grows in human cells without inhibiting apoptosis, fragmenting the Golgi apparatus, or diverting post-Golgi sphingomyelin transport. Infection and immunity 2015;83:3268–80.

119. Kebbi-Beghdadi C, Cisse O, and Greub G. Permissivity of Vero cells, human pneumocytes and human endometrial cells to *Waddlia chondrophila*. Microbes and Infection 2011;13:566–74.

120. Goy G, Croxatto A, and Greub G. *Waddlia chondrophila* enters and multiplies within human macrophages. Microbes and infection 2008;10:556–62.

121. Kebbi-Beghdadi C, Fatton M, and Greub G. Permissivity of insect cells to *Waddlia chondrophila*, *Estrella lausannensis* and *Parachlamydia acanthamoebae*. Microbes and Infection. Special issue on intracellular bacteria 2015;17:749–54.

122. Turin L, Surini S, Wheelhouse N, and Rocchi MS. Recent advances and public health implications for environmental exposure to *Chlamydia abortus*: from enzootic to zoonotic disease. Veterinary Research 2022;53:1–17.

123. Dilbeck-Robertson P, McAllister MM, Bradway D, and Evermann JF. Results of a new serologic test suggest an association of *Waddlia chondrophila* with bovine abortion. Journal of Veterinary Diagnostic Investigation 2003;15:568–9.

124. Baud D, Thomas V, Arafa A, Regan L, and Greub G. *Waddlia chondrophila*, a potential agent of human fetal death. Emerging infectious diseases 2007;13:1239.

125. Baud D, Goy G, Osterheld MC, et al. Role of *Waddlia chondrophila* placental infection in miscarriage. Emerging infectious diseases 2014;20:460.

126. Hornung S, Thuong B, Gyger J, et al. Role of *Chlamydia trachomatis* and emerging *Chlamydia*-related bacteria in ectopic pregnancy in Vietnam. Epidemiology & Infection 2015;143:2635–8.

127. Baud D, Ammerdorffer A, Buffe Y, Vouga M, Greub G, and Stojanov M. Impact of *Waddlia chondrophila* infection on pregnancy in the mouse. New Microbes and New Infections 2020;33:100619.

128. Wheelhouse N, Flockhart A, Aitchison K, et al. Experimental challenge of pregnant cattle with the putative abortifacient *Waddlia chondrophila*. Scientific reports 2016;6:1–10.

129. Kasper D, Fauci A, Hauser S, Longo D, Jameson J, and Loscalzo J. Harrison's principles of internal medicine, 19e. Vol. 1. 2. Mcgraw-hill New York, NY, USA: 2015:2134–6.

130. Friedman M, Galil A, Greenberg S, and Kahane S. Seroprevalence of IgG antibodies to the chlamydia-like microorganism '*Simkania* Z'by ELISA. Epidemiology & Infection 1999;122:117–23.

131. Kahane S, Gonen R, Sayada C, Elion J, and Friedman MG. Description and partial characterization of a new *Chlamydia*-like microorganism. FEMS microbiology letters 1993;109:329–33.

132. Kahane S, Greenberg D, Friedman MG, Haikin H, and Dagan R. High prevalence of "*Simkania* Z," a novel *Chlamydia*-like bacterium, in infants with acute bronchiolitis. The Journal of infectious diseases 1998;177:1425–9.

133. Vouga M, Baud D, and Greub G. *Simkania negevensis*, an insight into the biology and clinical importance of a novel member of the *Chlamydiales* order. Critical reviews in microbiology 2017;43:62–80.

134. Vouga M, Kebbi-Beghdadi C, Liénard J, Baskin L, Baud D, and Greub G. What is the true clinical relevance of *Simkania negevensis* and other emerging *Chlamydiales* members? New microbes and new infections 2018;23:1–5.

135. Niemi S, Greub G, and Puolakkainen M. *Chlamydia*-related bacteria in respiratory samples in Finland. Microbes and Infection 2011;13:824–7.

136. Tolkki L, Hokynar K, Meri S, Panelius J, Puolakkainen M, and Ranki A. Granuloma annulare and morphea: correlation with *Borrelia burgdorferi* infections and chlamydia-related bacteria. Acta Dermato-Venereologica 2018;98:355–60.

137. Corsaro D and Venditti D. Detection of novel *Chlamydiae* and *Legionellales* from human nasal samples of healthy volunteers. Folia Microbiologica 2015;60:325–34.

138. Yu Y, Kim YH, Cho WH, Son BS, and Yeo HJ. Biofilm microbiome in extracorporeal membrane oxygenator catheters. PloS one 2021;16:e0257449.

143. Bonnet SI, Binetruy F, Hernández-Jarguín AM, and Duron O. The tick microbiome: why non-pathogenic microorganisms matter in tick biology and pathogen transmission. Frontiers in cellular and infection microbiology 2017;7:236.

147. Oechslin CP, Heutschi D, Lenz N, et al. Prevalence of tick-borne pathogens in questing *Ixodes ricinus* ticks in urban and suburban areas of Switzerland. Parasites & Vectors 2017;10:558.

149. Hodosi R, Kazimirova M, and Soltys K. What do we know about the microbiome of *I. ricinus*? Frontiers in cellular and infection microbiology 2022;12:990889.

151. Vouga M, Baud D, and Greub G. *Simkania negevensis* may produce long-lasting infections in human pneumocytes and endometrial cells. Pathogens and Disease 2017;75:ftw115.

205. Scollard DM, Adams L, Gillis T, Krahenbuhl J, Truman R, and Williams D. The continuing challenges of leprosy. Clinical microbiology reviews 2006;19:338–81.

206. Antoł A, Rojek W, Singh S, Piekarski D, and Czarnoleski M. Hypoxia causes woodlice (*Porcellio scaber*) to select lower temperatures and impairs their thermal performance and heat tolerance. PLoS One 2019;14:e0220647.

207. Fieler AM, Rosendale AJ, Farrow DW, et al. Larval thermal characteristics of multiple ixodid ticks. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 2021;257:110939.

208. Eisen RJ, Eisen L, Ogden NH, and Beard CB. Linkages of weather and climate with *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae), enzootic transmission of *Borrelia burgdorferi*, and Lyme disease in North America. Journal of medical entomology 2016;53:250–61.

209. Vouga M and Greub G. Emerging bacterial pathogens: the past and beyond. Clinical Microbiology and Infection 2016;22:12–21.

210. Relman DA, Schmidt TM, MacDermott RP, and Falkow S. Identification of the uncultured bacillus of Whipple's disease. New England journal of medicine 1992;327:293–301.

211. Grankvist A, Andersson PO, Mattsson M, et al. Infections with the tick-borne bacterium "*Candidatus* Neoehrlichia mikurensis" mimic noninfectious conditions in patients with B cell malignancies or autoimmune diseases. Clinical Infectious Diseases 2014;58:1716–22.

212. Fehr JS, Bloemberg GV, Ritter C, et al. Septicemia caused by tick-borne bacterial pathogen *Candidatus* Neoehrlichia mikurensis. Emerging infectious diseases 2010;16:1127.

213. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. The New England Journal of Medicine 2020;382:727–33.

214. Kosoy OI, Lambert AJ, Hawkinson DJ, et al. Novel Thogotovirus associated with febrile illness and death, United States, 2014. Emerging Infectious Diseases 2015;21:760–764.

215. Phin N, Parry-Ford F, Harrison T, et al. Epidemiology and clinical management of Legionnaires' disease. The Lancet. Infectious Diseases 2014;14:1011–21.

216. Brenner AE, Muñoz-Leal S, Sachan M, Labruna MB, and Raghavan R. *Coxiella burnetii* and related tick endosymbionts evolved from pathogenic ancestors. Genome biology and evolution 2021;13:evab108.

217. Duron O, Morel O, Noël V, et al. Tick-bacteria mutualism depends on B vitamin synthesis pathways. Current Biology 2018;28:1896–902.

218. Buysse M and Duron O. Evidence that microbes identified as tick-borne pathogens are nutritional endosymbionts. Cell 2021;184:2259–60.

219. Drew GC, Stevens EJ, and King KC. Microbial evolution and transitions along the parasite–mutualist continuum. Nature Reviews Microbiology 2021;19:623–38.

220. Douglas AE. The B vitamin nutrition of insects: the contributions of diet, microbiome and horizontally acquired genes. Current Opinion in Insect Science 2017;23:65–9.

221. Jamison WP and Hackstadt T. Induction of type III secretion by cell-free *Chlamydia trachomatis* elementary bodies. Microbial Pathogenesis 2008;45:435–40.

222. Kebbi-Beghdadi C, Pilloux L, Martin V, and Greub G. Eukaryotic cell permeabilisation to identify new putative chlamydial type III secretion system effectors secreted within host cell cytoplasm. Microorganisms 2020;8:361.

223. Herrera P, Schuster L, Wentrup C, et al. Molecular causes of an evolutionary shift along the parasitism–mutualism continuum in a bacterial symbiont. Proceedings of the National Academy of Sciences 2020;117:21658–66.

224. Marquis B, Opota O, Jaton K, and Greub G. Impact of different SARS-CoV-2 assays on laboratory turnaround time. Journal of Medical Microbiology 2021;70:001280.

A Appendix

A.1 Impact of different SARS-CoV-2 assays on laboratory turnaround time

Authors: Bastian Marquis, Onya Opota, Katia Jaton, Gilbert Greub

Status: Published in *Journal of Medical Microbiology* [224]

Contribution: BM participated in the validation of SARS-CoV-2 assays results, performed the statistical analysis and wrote the manuscript.

During the first wave of the COVID-19 pandemic, I volunteered to help for the biomedical validation of the SARS-CoV-2 assays, at a moment when the research laboratories were closed by the lockdown. During this period, I had the occasion to analyze the turnaround time of SARS-CoV-2 assays and to publish an article summarizing the main findings.

JOURNAL OF MEDICAL MICROBIOLOGY

RESEARCH ARTICLE

Marquis *et al.*, *Journal of Medical Microbiology* 2021;70:001280
DOI 10.1099/jmm.0.001280

MICROBIOLOGY SOCIETY

OPEN ACCESS

Impact of different SARS-CoV-2 assays on laboratory turnaround time

Bastian Marquis, Onya Opota, Katia Jaton and Gilbert Greub*

Abstract

Introduction. Clinical microbiology laboratories have had to cope with an increase in the volume of tests due to the emergence of the SARS-CoV-2 virus. Short turnaround times (TATs) are important for case tracing and to help clinicians in patient management. In such a context, high-throughput systems are essential to process the bulk of the tests. Rapid tests are also required to ensure shorter TATs for urgent situations. In our laboratory, SARS-CoV-2 assays were initially implemented on our custom platform using a previously published method. The commercial cobas 6800 (Roche diagnostics) assay and the GeneXpert Xpress (Cepheid) SARS-CoV-2 assay were implemented on 24 March and 8 April 2020, respectively, as soon as available.

Hypothesis/Gap Statement. Despite the abundant literature on SARS-CoV-2 assays, the articles focus mainly on the diagnostic performances. This is to our knowledge the first article that specifically studies the TAT of different assays.

Aim. We aimed to describe the impact of various SARS-CoV-2 assays on the TAT at the beginning of the outbreak.

Methodology. In this study, we retrospectively analysed the TAT of all SARS-CoV-2 assays performed in our centre between 24 February and 9 June, 2020.

Results. We retrieved 33900 analyses, with a median TAT of 6.25 h. TATs were highest (6.9 h) when only our custom platform was used (24 February to 24 March, 2020). They were reduced to 6.1 h when the cobas system was introduced (24 March to 8 April, 2020). The implementation of the GeneXpert further reduced the median TAT to 4.8 h (8 April to 9 June, 2020). The GeneXpert system had the shortest median TAT (1.9 h), followed by the cobas (5.5 h) and by our custom platform (6.9 h).

Conclusion. This work shows that the combination of high-throughput systems and rapid tests allows the efficient processing of a large number of tests with a short TAT. In addition, the use of a custom platform allowed the quick implementation of an in-house test when commercial assays were not yet available.

BACKGROUND

Coronavirus disease (COVID-19) is a disease caused by a novel coronavirus, the SARS-CoV-2, that initially appeared in the Wuhan area, China and was later declared a pandemic [1, 2]. In Switzerland, the first case was documented on 24 February 2020 and the disease spread and reached its peak on 23 March 2020 with 1454 new documented cases. Overall, 30988 cases of COVID-19 were documented on 9 June 2020, with a total of 1633 deaths [3].

The microbiology laboratories were central in the response against COVID-19 as they had to quickly implement SARS-CoV-2 assays, to adapt to a sharp increase in the volume of tests and to maintain short turnaround times (TAT) [4–6]. Short TATs are indeed important to allow a quick tracing of cases, to optimize the use of scarce resources such as negative pressure rooms and to guide clinicians in patient management. However, the increasing amount of scientific literature on SARS-CoV-2 diagnostic assays has focused on assessment of their performance [7–12] and to our knowledge, this is the first publication

Received 29 June 2020; Accepted 23 March 2021; Published 06 May 2021
Author affiliations: [1]Institute of Microbiology, University Hospital Center and University of Lausanne, Lausanne, Switzerland.
***Correspondence:** Gilbert Greub, gilbert.greub@chuv.ch
Keywords: COVID; diagnostic; NAAT; SARS-CoV-2; TAT ; time to results.
Abbreviations: BAL, bronchoalveolar lavage; CHUV, Centre Hospitalier Universitaire VAudois (= Lausanne University hospital); COVID, coronavirus disease; CSF, cerebrospinal fluid; h., hour; Inv., invalid; MOLIS, modern laboratory interface system; PCR, polymerase chain reaction; qPCR, quantitative PCR; RT, retrotranscription; SARS-CoV-2, severe acute respiratory syndrome coronavirus-2; TAT, turnaround time; vs, versus.
One supplementary table is available with the online version of this article.

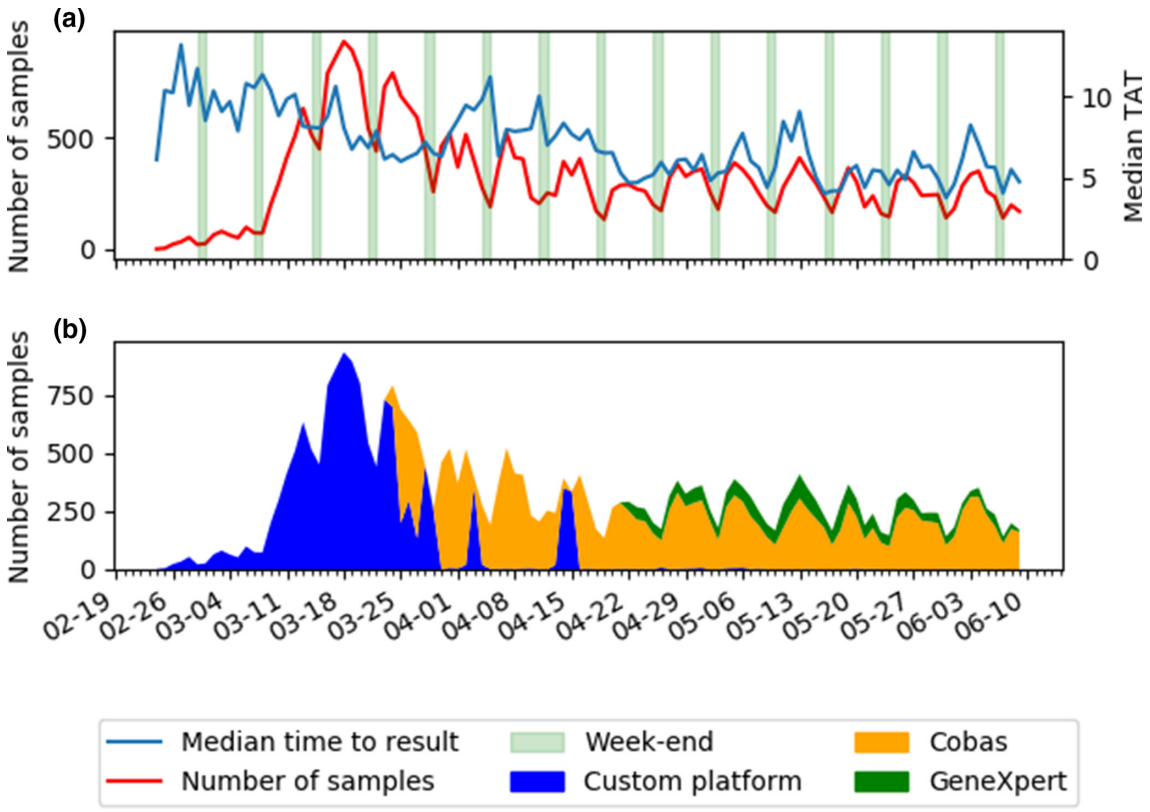

Fig. 1. (a) Median turnaround time in hours and number of samples received per day (b) Samples in function of platform (adapted from [17, 23])

to specifically address reducing the TAT of SARS-CoV-2 assays.

In this paper, we describe how the implementation of the different assays affected the TAT. This work also shows the impact of the sample type, analytical errors and of the result of the analyses on the TAT.

METHODS

We extracted information from all the SARS-CoV-2 analyses performed at CHUV (Lausanne University Hospital) from 24 February to 9 June, 2020. As our laboratory is a reference centre for COVID-19 testing, samples also originated from surrounding hospitals and screening centres. Information on the analyses included the sample type, the type of assays that were used (cobas, GeneXpert and/or our custom platform), their result and the different timestamps (time of reception at the pre-analytic laboratory and the time of the biomedical validation of the result). We excluded all analyses performed for quality control and all the analyses that were cancelled after their registration in our laboratory information system. The data were obtained during a quality enhancement project at our institution. According to the Swiss national law, conducting and publishing the results of such a project is permitted without ethics committee approval.

Assays

The molecular diagnosis laboratory of the Lausanne University Hospital developed a custom platform for automated testing, as described elsewhere [13]. An in-house SARS-CoV-2 assay based on the work of Corman *et al.* [14] was implemented for this custom platform. This assay targets the E and the RdRp genes, but due to the low performances of the RdRp RT-PCR [7, 15], only the E gene RT-PCR was used after the first 893 samples.

The Roche SARS-CoV-2 assay for the cobas 6800 system [16] was implemented on 24 March 2020. The Cepheid GeneXpert Xpress SARS-CoV-2 assay [8] was implemented on 21 April 2020. Both assays showed perfect agreement with our in-house RT-qPCR, with kappa values of 0.98 (as published by Opota *et al.* [17]). Several publications showed similar results [8, 9, 11, 12]. The GeneXpert assay was however reported to have a better sensitivity for samples with a low viral load [8, 9].

Organization

After reception at the pre-analytic laboratory, the samples first had to be registered in our laboratory information system, MOLIS (CompuGroup Medical, AG). The time of reception timestamp corresponds to the time of registration in MOLIS. Some analyses were prioritized over others: we prioritized samples from the emergency department, for patients in

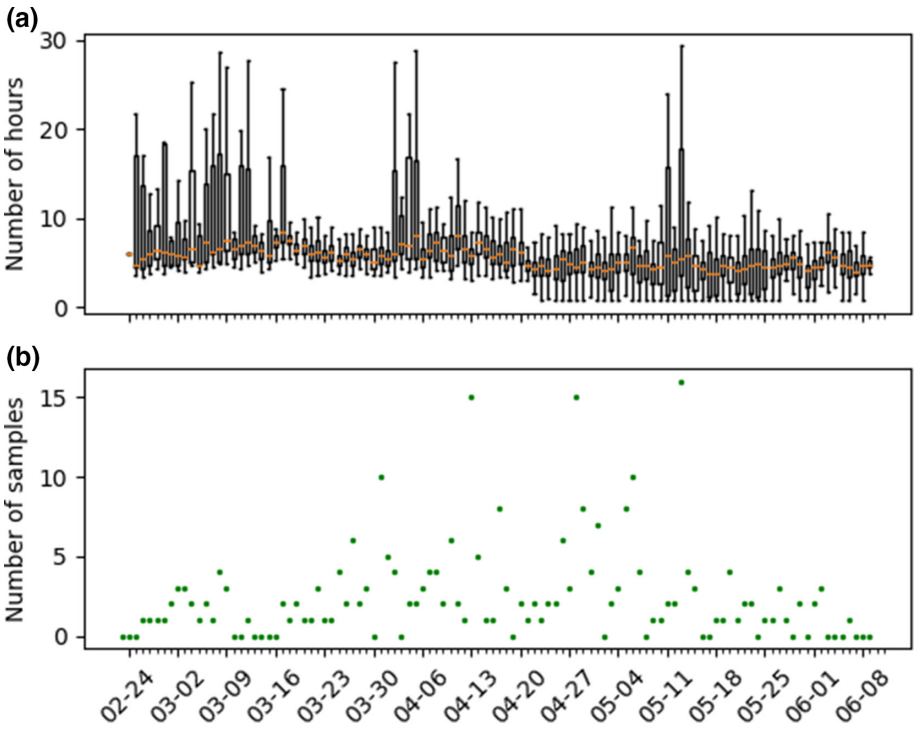

Fig. 2. (a) Boxplot of the turnaround time (outliers not shown) (b) Number of samples with a turnaround time (TAT) >24 h.

need of urgent surgery or samples tagged as urgent by the clinicians. Then came the samples from the different wards of the CHUV, from the external hospitals and finally, from the screening centres.

Initially, all analyses were performed on our custom platform. The analyses were progressively transferred onto the cobas system after its introduction. The GeneXpert system was used for specific cases (samples from the emergency department or in the context of pre-operative assessment for surgeries or organ transplantation).

Before being transmitted to the clinicians, all results had to be validated by a laboratory technician (technical validation) and a clinical microbiologist (biomedical validation). An automated validation by expert systems was progressively introduced to perform the biomedical validation. Its use was initially restricted to the validation of negative results, but starting from 8 April 2020 it was extended to validate all analysis. The results of our custom platform were automatically validated by FastFinder (UgenTec, Hasselt, Belgium). The results of the cobas and the GeneXpert systems were automatically validated by their accompanying expert systems [18].

Statistics

All statistics were done with Python v3.7.3 [19] and the scipy package v1.1.0 [20]. The Mann-Whitney U test was used unless otherwise specified.

RESULTS

We retrieved a total of 33900 tests for analysis, with a median TAT of 6.25 h (range 0.9–678). Of those analyses, 18153 (53.5%) were performed on the cobas system (median TAT=5.5 h, range: 2.8–114.8), 12941 (38.2%) on our custom platform (median TAT=6.9 h, range 3.5–678), 2756 (8.1%) on the GeneXpert system (median TAT=1.9 h, range: 0.8–53.0) and 50 (0.1%) were performed on more than one platform (median TAT=7.4 h, range: 4.1–34.2). The cobas system allowed shorter TATs than our custom platform (5.5 vs 6.9, P <0.001). The GeneXpert system was faster than both the cobas (1.9 vs 5.5, P <0.001) and our custom platform (1.9 vs 6.9, P <0.001). The increase of the number of samples received per day is shown in Fig. 1(a). There was a mean number of 317 samples per day (range 2–933). Fig. 1(b) shows the repartition of the samples on the different diagnostic platforms. The median TAT when only our custom platform was in use was 6.9 h (range 3.5–297.5). It decreased to 6.1 h (P <0.001) after the introduction of the cobas system (range 2.9–678.1) and was further improved to 4.8 h (P <0.001) with the implementation of the GeneXpert system (range 0.8–408.4). The evolution of the TAT is shown in Fig. 2(a). Most analyses were performed on the cobas system after its introduction. Due to technical problems and maintenance (3 April and, 14–15 April 2020 respectively) of the cobas system, testing was temporarily transferred to our custom platform.

As shown in Tables 1 and S1 (available in the online version of this article), there were 267 samples (0.8%) with a TAT

Table 1. TAT according to the platform, the results and the type of sample

Platform	Sample	No. of pos./Neg./Inv. (% pos)*	Median TAT for pos./Neg./Inv./All.	No. of samples with TAT >24 h (%)
Overall	**Total**	**4513/29216/171** (13.31%)	**6.25/5.82/10.47/5.90**	**267** (0.79%)
cobas	**Total**	**1740/16323/90** (9.59%)	**5.65/5.43/10.44/5.47**	**91** (0.50%)
	Nasopharyngeal swabs	1674/15811/73 (9.53%)	5.62/5.43/10.62/5.45	69 (0.34%)
	Respiratory samples†	54/330/9 (13.74%)	7.13/6.09/9.90/6.37	25 (6.36%)
	Other Samples‡	12/182/8 (5.94%)	6.80/5.80/11.07/5.96	6 (2.97%)
Platform	**Total**	**2709/10204/28** (20.93%)	**6.65/6.90/27.88/6.85**	**172** (1.33%)
	Nasopharyngeal swabs	2440/8730/11 (21.82%)	6.58/6.82/19.68/6.77	54 (0.48%)
	Respiratory samples†	260/1280/2 (16.86%)	7.29/7.49/13.37/7.47	13 (0.84%)
	Other samples‡	9/194/15 (4.13%)	8.98/23.20/47.77/23.58	105 (48.17%)
GeneXpert	**Total**	**64/2683/9 (2.32%)**	**1.51/1.28/2.70/1.28**	**3** (0.11%)
	Nasopharyngeal swabs	63/2656/9 (2.31%)	1.52/1.28/2.70/1.30	3 (0.11%)
	Respiratory samples†	0/5/0 (0.00%)	- / 1.72 / - / 1.72	0 (0.00%)
	Other Samples‡	1/22/0 (4.35%)	1.00/1.02/- / 1.02	0 (0.00%)
Multiple platforms	**Total**	0/1/49 (0.00%)	- / **18.68/7.33/7.37**	**1** (2.00%)

*Result of the first analysis. Invalid analyses were repeated (the result of the repeated analysis are not shown).
†includes sputum, oropharyngeal, nasal and mouth swabs and bronchoalveolar lavages (BAL) and mini-BAL.
‡includes blood, urine, stools, anal swabs, bile, obstetrical samples, CSF and biopsies.

longer than 24 h (median: 29.75; range 24.0–678.1), with a mean number of 2.5 such analyses per day (range 0–16). The number of analyses with a TAT longer than 24 h is shown in Fig. 2(b). The majority of those were nasopharyngeal swabs (117 of 267). Analysis with long TAT were however over-represented in unusual samples like bone marrow (1 of 1), ophthalmologic samples (2 of 2), bile aspiration (1 of 1), blood (25 of 48) or CSF (33 of 45). The peaks of analyses with long TAT on 13 April, 28 April and 13 May, 2020 (Fig. 2b) were due to groups of external samples that had to be delayed due to our priority policy.

Some analyses had technical problems during the run and their result is referred to as invalid (due to the inhibition of the reaction [21, 22] or clotting [12]). There were 140 such results on the cobas system (0.8%), 28 on our custom platform (0.2%), and nine on the GeneXpert system (0.004%). Repeating the analysis on leftover material allowed a conclusive result to be reached in most cases (cobas: 134 of 140; custom platform 28 of 28; GeneXpert: 9 of 9). Fifty of initially invalid cobas results were repeated on a different platform (GeneXpert in 38 of 50 and our custom platform in 12 of 50), allowing to reach a conclusive result in all cases.

Surprisingly, the result of the analyses also had an impact on the TAT: for the cobas and the GeneXpert systems, negative results had a shorter TAT than positive results (5.4 vs 5.7, P <0.001 and 1.3 vs 1.5, P <0.05 respectively). For the cobas system, this difference is due to the automatic results validation system (TAT of 5.5 h and negative=5.4 vs positive=6.0, P <0.001). The same comparison cannot be performed for the GeneXpert system, as all its results were automatically validated. The effect was opposite for our custom platform (negative: 6.9 vs positive: 6.7, P <0.001), however no cause could be determined for this effect.

As shown in Fig. 3, most samples were received between 8 am and 9 pm, with three peaks centred at 10 am, 5 pm and 8 pm. The two last peaks correspond to samples from external hospitals that were received in batches. The median TAT was 5.9 h between 6 am and 8 pm, 15.0 h between 9 pm and

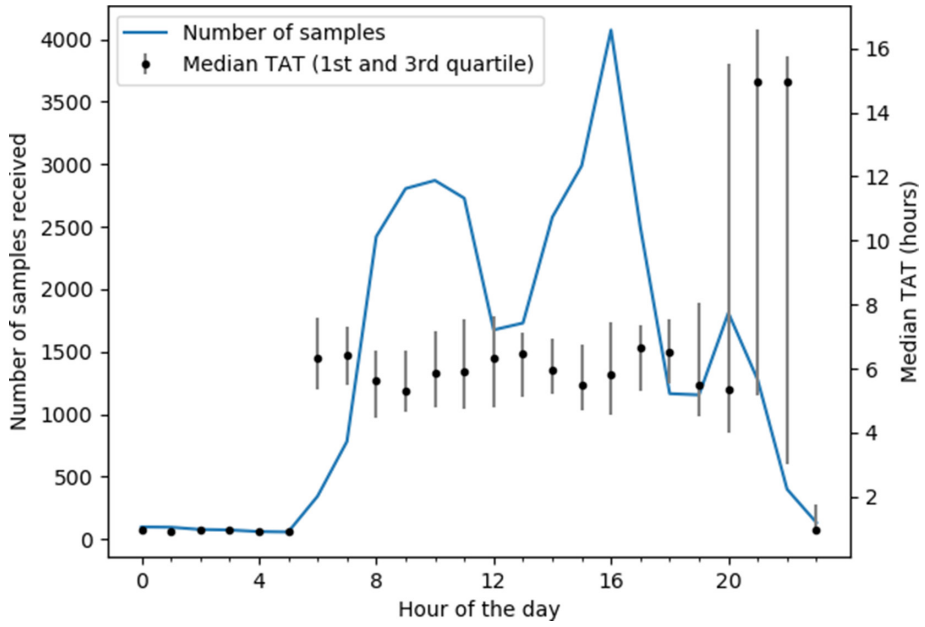

Fig. 3. Number of samples received and turnaround time in function of the hour of the day

10 pm (late arrivals from the wards and the external hospitals that were performed the next day) and 1.0 h between 11 pm and 5 am (samples from the Emergency department performed on the GeneXpert).

DISCUSSION

This work aimed to study the impact of different analysis platforms on the TAT.

These results show that the association of a high-throughput system like the cobas system and a faster system for individual samples like the GeneXpert system allows short turnaround times. As already noted [13], the additional flexibility gained by using a custom platform allowed the quick implementation of a SARS-CoV-2 assay and to cope with a sudden increase in the number of tests, when the cobas SARS-CoV-2 assay was still not approved by the U.S. Food and Drug Administration. However, despite the automation of most steps of the qPCR, our custom platform was predictably slower than the cobas system, that fully automates all steps [16]. The complete automation on the cobas system also reduces the workload of laboratory technicians: as many as six technicians were necessary to operate our custom platform, while only two could achieve a similar throughput of 900 tests per day on the cobas system.

Additionally, in the context of a global shortage of reagents, pipet tips and PCR plates, it is important to maximize testing efficiency. However, this may come at the cost of longer TATs as samples which arrive early may be held until sufficient numbers are reached to maximize use of a PCR plate. Assuming the minimum TAT for the cobas and the custom platform (2.9 and 3.5 h, respectively) is close to the optimal TAT, this shows that a significant proportion of the TAT is spent waiting run completion (median of 5.5 h vs an optimal 2.9 h for the cobas and median of 6.9 h vs optimal 3.5 h for our custom platform). The use of smaller PCR plates may help mitigate the problem, if feasible. Additionally, having both the cobas and the custom platform was useful to alleviate shortage problems. Tips, reagents, processing plates or waste-covers for the cobas system were particularly impacted by such problems and made it necessary to rely on our custom platform despite slightly longer TATs. Having two high-throughput systems was helpful to cope with maintenance or unforeseen downtimes: we were able to transfer the samples from the cobas system to our custom platform without causing delays when the former had to be shutdown.

The introduction of the GeneXpert system allowed a faster track of analysis for urgent samples (such as an assessment for eligibility for an organ transplant), but shortages in reagents limited its widespread use. Additionally, operating the GeneXpert system does not require specialized technicians, which allows analyses for samples collected at nighttime to be run without delay.

While the choice of the analysis platform obviously affects the TAT, some other factors also have an impact: counter-intuitively, using an automated system to validate the analysis delayed the validation of positive results as compared to a manual validation, due to long running times. Inevitably, prioritizing some samples over others may lengthen the TAT of others, and in our case, it caused the three peaks in the number of TAT. The priority policy should be carefully planned. Overall, analysis with TAT longer than 24 h were rare and were over-represented in unusual samples.

CONCLUSION

This work shows the result of using a combination of high-throughput systems for the bulk of the analysis and a faster system for selected individual samples. The former allows a high volume of analyses and the latter allows shorter TAT for urgent samples. With this organization, we achieved a median TAT of 6.25 h. Overall, TATs were shorter than 8 h in 82.1% of the cases, less than 12 h in 89.3% of cases and less than 24 h in 99.2% of cases.

Funding information
Bastian Marquis is funded by the Jürg Tschopp MD-PhD scholarship.

Acknowledgements
We would like to deeply thank all the staff of the Institute of Microbiology of the Lausanne University Hospital. In particular, we thank all the staff of the Laboratory of Molecular Diagnostic of the Institute of Microbiology of the University of Lausanne and all the training FAMH of the COVID-19 post-analytic team of our institute.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
According to the Swiss national law, conducting and publishing the results of such a project is permitted without ethics committee approval.

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F *et al*. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020;395:507–513.
2. Spiteri G, Fielding J, Diercke M, Campese C, Enouf V *et al*. First cases of coronavirus disease 2019 (COVID-19) in the WHO European region, 24 January to 21 February 2020. *Euro Surveill* 2020;25.
3. Maladie coronavirus (COVID-19). Rapport sur La situation épidémiologique en Suisse et dans La Principauté de Liechtenstein 2019.
4. Lippi G, Plebani M. The critical role of laboratory medicine during coronavirus disease 2019 (COVID-19) and other viral outbreaks. *Clin Chem Lab Med* 2020;58:1063–1069.
5. Weemaes M, Martens S, Cuypers L, Van Elslande J, Hoet K *et al*. Laboratory information system requirements to manage the COVID-19 pandemic: a report from the Belgian national reference testing center. *J Am Med Inform Assoc* 2020;27:1293–1299.
6. Hawkins RC. Laboratory turnaround time. *Clin Biochem Rev* 2007;28:179–194.
7. Muenchhoff M, Mairhofer H, Nitschko H, Grzimek-Koschewa N, Hoffmann D *et al*. Multicentre comparison of quantitative PCR-based assays to detect SARS-CoV-2, Germany, March 2020. *Eurosurveillance* 2020;25:2001057.
8. Moran A, Beavis KG, Matushek SM, Ciaglia C, Francois N *et al*. The detection of SARS-CoV-2 using the Cepheid Xpert Xpress SARS-CoV-2 and Roche Cobas SARS-CoV-2 assays. *J Clin Microbiol* 2020.
9. Lieberman JA, Pepper G, Naccache SN, Huang M-L, Jerome KR *et al*. Comparison of commercially available and laboratory-developed assays for in vitro detection of SARS-CoV-2 in clinical laboratories. *J Clin Microbiol* 2020;58
10. Wang W, Xu Y, Gao R, Lu R, Han K *et al*. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* 2020.
11. Pfefferle S, Reucher S, Nörz D, Lütgehetmann M. Evaluation of a quantitative RT-PCR assay for the detection of the emerging coronavirus SARS-CoV-2 using a high throughput system. *Euro Surveill* 2020;25.
12. Poljak M, Korva M, Knap Gašper N, Fujs Komloš K, Sagadin M *et al*. Clinical evaluation of the COBAS SARS-CoV-2 test and a diagnostic platform switch during 48 hours in the midst of the COVID-19 pandemic. *J Clin Microbiol* 2020;58.
13. Greub G, Sahli R, Brouillet R, Jaton K. Ten years of R&D and full automation in molecular diagnosis. *Future Microbiol* 2016;11:403–425.
14. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A *et al*. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 2020;25.
15. Pillonel T, Scherz V, Jaton K, Greub G, Bertelli C. Letter to the editor: SARS-CoV-2 detection by real-time RT-PCR. *Euro Surveill* 2020;25.
16. Cobb B, Simon CO, Stramer SL, Body B, Mitchell PS *et al*. The cobas® 6800/8800 system: a new era of automation in molecular diagnostics. *Expert Rev Mol Diagn* 2017;17:167–180.
17. Opota O, Brouillet R, Greub G, Jaton K. Comparison of SARS-CoV-2 RT-PCR on a high-throughput molecular diagnostic platform and the COBAS SARS-CoV-2 test for the diagnostic of COVID-19 on various clinical samples. *Pathog Dis* 2020;78
18. Mueller L, Scherz V, Greub G, Jaton K, Opota O. Computer-Aided medical microbiology monitoring tool: a strategy to adapt to the SARS-CoV-2 epidemic and that highlights RT-PCR consistency. *MedRxiv* 2020.
19. Van Rossum G, Drake FL. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace; 2009.
20. Jones E, Oliphant T, Peterson P. Others. SciPy: open source scientific tools for python 2001.
21. Opota O, Jaton K, Greub G. Microbial diagnosis of bloodstream infection: towards molecular diagnosis directly from blood. *Clinical Microbiology and Infection* 2015;21:323–331.
22. Tahamtan A, Ardebili A. Real-Time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev Mol Diagn* 2020;20:1–2.
23. Greub G, Opota O, Brouillet R, Jaton K. Diagnostic par RT-PCR de l'infection par le virus SARS-CoV-2. *Pipette - Swiss Lab Med* 2020.