


RESEARCH ARTICLE

Open Access



# Comparative Analysis of Genome of *Ehrlichia* sp. HF, a Model Bacterium to Study Fatal Human Ehrlichiosis

Mingqun Lin<sup>1\*</sup> , Qingming Xiong<sup>1</sup>, Matthew Chung<sup>2</sup>, Sean C. Daugherty<sup>2</sup>, Sushma Nagaraj<sup>2</sup>, Naomi Sengamalay<sup>2</sup>, Sandra Ott<sup>2</sup>, Al Godinez<sup>2</sup>, Luke J. Tallon<sup>2</sup>, Lisa Sadzewicz<sup>2</sup>, Claire Fraser<sup>2,3</sup>, Julie C. Dunning Hotopp<sup>2,4,5</sup> and Yasuko Rikihisa<sup>1\*</sup>

## Abstract

**Background:** The genus *Ehrlichia* consists of tick-borne obligatory intracellular bacteria that can cause deadly diseases of medical and agricultural importance. *Ehrlichia* sp. HF, isolated from *Ixodes ovatus* ticks in Japan [also referred to as *I. ovatus Ehrlichia* (IOE) agent], causes acute fatal infection in laboratory mice that resembles acute fatal human monocytic ehrlichiosis caused by *Ehrlichia chaffeensis*. As there is no small laboratory animal model to study fatal human ehrlichiosis, *Ehrlichia* sp. HF provides a needed disease model. However, the inability to culture *Ehrlichia* sp. HF and the lack of genomic information have been a barrier to advance this animal model. In addition, *Ehrlichia* sp. HF has several designations in the literature as it lacks a taxonomically recognized name.

**Results:** We stably cultured *Ehrlichia* sp. HF in canine histiocytic leukemia DH82 cells from the HF strain-infected mice, and determined its complete genome sequence. *Ehrlichia* sp. HF has a single double-stranded circular chromosome of 1,148,904 bp, which encodes 866 proteins with a similar metabolic potential as *E. chaffeensis*. *Ehrlichia* sp. HF encodes homologs of all virulence factors identified in *E. chaffeensis*, including 23 paralogs of P28/OMP-1 family outer membrane proteins, type IV secretion system apparatus and effector proteins, two-component systems, ankyrin-repeat proteins, and tandem repeat proteins. *Ehrlichia* sp. HF is a novel species in the genus *Ehrlichia*, as demonstrated through whole genome comparisons with six representative *Ehrlichia* species, subspecies, and strains, using average nucleotide identity, digital DNA-DNA hybridization, and core genome alignment sequence identity.

**Conclusions:** The genome of *Ehrlichia* sp. HF encodes all known virulence factors found in *E. chaffeensis*, substantiating it as a model *Ehrlichia* species to study fatal human ehrlichiosis. Comparisons between *Ehrlichia* sp. HF and *E. chaffeensis* will enable identification of *in vivo* virulence factors that are related to host specificity, disease severity, and host inflammatory responses. We propose to name *Ehrlichia* sp. HF as *Ehrlichia japonica* sp. nov. (type strain HF), to denote the geographic region where this bacterium was initially isolated.

**Keywords:** *Ehrlichia* sp. HF, Monocytic Ehrlichiosis, Mouse model, Comparative genomic analysis, Core genome alignment, Virulence factors

\* Correspondence: lin.427@osu.edu; rikihisa.1@osu.edu

<sup>1</sup>Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Road, Columbus, OH 43210, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

The incidence of tick-borne diseases has risen dramatically in the past two decades, and continues to rise [1–3]. The 2011 Institute of Medicine report “Critical Needs and Gaps in...Lyme and Other Tick-Borne Diseases” revealed the urgent need for research into tick-borne diseases [4]. *Ehrlichia* species are tick-borne obligate intracellular bacteria, which are maintained via the natural transmission and infection cycle between particular species of ticks and mammals (Table 1). The genus *Ehrlichia* belongs to the family Anaplasmataceae in the order Rickettsiales. According to International Code of Nomenclature of Prokaryotes and International Journal of Systematic and Evolutionary Microbiology [46], and following the reorganization of genera in the family Anaplasmataceae based on molecular phylogenetic analysis [47], the genus *Ehrlichia* currently consists of six taxonomically classified species with validly published names, including *E. chaffeensis*, *E. ewingii*, *E. canis*, *E. muris*, *E. ruminantium*, and a recently culture-isolated *E. minasensis* that is closely related to *E. canis* (Table 1) [19, 37].

Accidental transmission and infection of domestic animals and humans can cause potentially severe to fatal diseases, and four species (*E. chaffeensis*, *E. ewingii*, *E. canis*, and *E. muris*) are known to infect humans and cause emerging tick-borne zoonoses [11, 19–21, 34, 48, 49]. In the US, the most common human ehrlichiosis is human monocytic ehrlichiosis (HME) caused by *E. chaffeensis*, which was discovered in 1986 [12], followed by human Ewingii ehrlichiosis discovered in 1998 [34]. The most recently discovered human ehrlichiosis is caused by *E. muris* subsp. *eaucalirensis* [originally referred to as *E. muris*-like agent (EMLA)] [19, 20]. Human infection with *E. canis* has been reported in South and Central America [21, 22, 49]. Regardless of the *Ehrlichia* species, clinical signs of human ehrlichiosis include fever, headache, myalgia, thrombocytopenia, leukopenia, and elevated serum liver enzyme levels [20, 21, 34, 48–50].

HME is a significant, emerging tick-borne disease with serious health impacts with the highest incidence in people over 60 years of age and immunocompromised individuals [48]. Life-threatening complications such as

**Table 1** Biological characteristics of representative *Ehrlichia* species

| Species <sup>1</sup> (Type strain)                      | Diseases   | Mammalian Host                             | Tick Vector/Host  | Geographic Distribution                   | References |
|---|--|--|---|---|------------|
| <i>Ehrlichia</i> sp. HF (HF565)                         | Acute fatal infection of mice (experimental)                             | Unknown                                    | <i>Ixodes ovatus</i> , <i>I. ricinus</i> , and <i>I. apronophorus</i> ticks | Japan, France, Serbia, Romania            | [5–10]     |
| <i>E. chaffeensis</i> (Arkansas) <sup>2</sup>           | Human monocytic ehrlichiosis (HME)                                       | Deer, Human, Dog, Coyote, Fox <sup>3</sup> | <i>Amblyomma americanum</i> (Lone star tick)                                | USA, Africa, South America, Europe, Japan | [11–16]    |
| <i>E. muris</i> subsp. <i>muris</i> (AS145)             | Murine monocytic ehrlichiosis (chronic systemic infection of mice)       | Mouse, Vole                                | Ticks ( <i>Haemaphysalis flava</i> or <i>Ixodes persulcatus</i> )           | Japan, Russia <sup>4</sup>                | [17, 18]   |
| <i>E. muris</i> subsp. <i>eaucalirensis</i> (Wisconsin) | Human or murine monocytic ehrlichiosis (fatal infection of mice)         | Human, Mouse                               | <i>Ixodes scapularis</i> (black-legged tick)                                | Wisconsin and Minnesota, USA              | [19, 20]   |
| <i>E. canis</i> (Oklahoma)                              | Canine tropical pancytopenia, Venezuelan Human Ehrlichiosis <sup>5</sup> | Dog, Human                                 | <i>Rhipicephalus sanguineus</i> (brown dog tick)                            | Global                                    | [21–25]    |
| <i>E. ruminantium</i> (Welgevonden)                     | Heartwater   | Ruminants (Cattle, Sheep, Goats, Antelope) | Various <i>Amblyomma</i> species of ticks                                   | Africa, Caribbean <sup>6</sup>            | [26–33]    |
| <i>E. ewingii</i> (Stillwater)                          | Canine granulocytic ehrlichiosis, Human ewingii ehrlichiosis             | Deer, Dog, Human                           | <i>Amblyomma americanum</i>   | USA, Japan                                | [34–36]    |
| <i>E. minasensis</i> (UFMG-EV)                          | Ehrlichiosis   | Cattle, Deer, Dog <sup>7</sup>             | <i>Rhipicephalus microplus</i> tick   | Brazil, Global                            | [37–45]    |

<sup>1</sup>Based on International Code of Nomenclature of Prokaryotes, and published in International Journal of Systematic and Evolutionary Microbiology, which lists officially approved list of bacterial classification and nomenclature, the genus *Ehrlichia* currently consists of six validly published species with correct names (<https://psn.dsmz.de/genus/ehrlichia>)

<sup>2</sup>*Ehrlichia* sp. HF, or *Ixodes ovatus Ehrlichia* (IOE) agent, is a field tick isolate of *Ehrlichia* species in Fukushima Prefecture, Japan from 1993 to 1994. *Ehrlichia* sp. HF DNA was also detected in *I. ricinus* tick from Brittany, France and Serbia, and *I. apronophorus* tick in Romania

<sup>3</sup>*E. chaffeensis* DNA was detected in 71% of free-ranging coyotes in Oklahoma and experimentally infected red foxes

<sup>4</sup>*E. muris* DNA was found in *I. persulcatus* ticks and small mammals in Russia

<sup>5</sup>Human infection with *E. canis* with clinical signs was reported in Venezuela, and *E. canis* was culture isolated from a VHE patient. In addition, *E. canis* DNA was detected in human blood bank donors in Costa Rica

<sup>6</sup>Heartwater in Caribbean islands of Guadeloupe was caused *E. ruminantium* Gardel, which is transmitted by *Amblyomma variegatum* (Tropical bont tick) and exceptionally virulent in Dutch goats. More heartwater cases in wild and domestic ruminants have been reported in five Caribbean islands, posing an increasing threat to domestic and wild ruminants in the continental US

<sup>7</sup>*E. minasensis* strain UFMG-EVT was isolated from the haemolymph of engorged *Rhipicephalus microplus* female ticks in Brazil, whereas strain Cuiaba was isolated from the whole blood of a naturally infected cattle. *E. minasensis* DNAs have also been reported in ticks, cervids, and dogs from France, Pakistan, Ethiopia, and Israel

renal failure, adult respiratory distress syndrome, meningoencephalitis, multi-system organ failure, and toxic shock occur in a substantial portion of the patients who are hospitalized and resulting in a case fatality rate of 3% [48]. However, there is no vaccine available for HME [51], and the only drug of choice is doxycycline, which is only effective with early diagnosis and treatment, and is not suitable for all patient groups [48]. In addition, pathogenesis and immunologic studies on human ehrlichiosis have been hampered due to the lack of an appropriate small animal disease model, as *E. chaffeensis* only transiently infects immunocompetent laboratory mice [52, 53]. *E. chaffeensis* naturally infects dogs and deer with mild to no clinical signs [53–55]. However, use of these animals is difficult and cost-prohibitive, while not being suitable for pathogenesis studies.

In an attempt to determine the pathogens harbored by *Ixodes ovatus* ticks prevalent in Japan, Fujita and Watanabe inoculated tick homogenates into the intraperitoneal cavity of laboratory mice, followed by serial passage through naïve mice using homogenized spleens from infected mice [5]. From 1983 to 1994, twelve “HF strains” were isolated from *I. ovatus* ticks in this manner, with the strain named after the scientist Hiromi Fujita who first discovered and isolated this bacterium [5]. Electron micrographs of HF326 showed the typical ultrastructure of *Ehrlichia* in the mouse liver [5]. A few years later, analysis of the 16S rRNA gene of the HF strains showed that four isolates (HF565, HF568-1, HF568-2, and HF639-2) from Fukushima, and two isolates (HF642 and HF652) from Aomori, northern Japan, were identical and closely related to *Ehrlichia* spp. [6]. The phylogenetic comparison of 16S rRNA and GroEL protein sequences of HF565 with those of members of the family *Anaplasmataceae*, and electron micrographs of HF565 verified that the HF strain belongs to the genus *Ehrlichia* [6]. Recent studies indicated that DNA sequences of *Ehrlichia* sp. HF have been detected not only in *I. ovatus* ticks throughout Japan, but also in *Ixodes ricinus* ticks in France [7] and Serbia [8], and *Ixodes apronophorus* ticks in Romania [9].

Unlike *E. muris*, HF565 does not induce splenomegaly but is highly virulent in mice, as intraperitoneal inoculation kills immunocompetent laboratory mice in 6–10 days [5, 6, 10, 56]. HF565 (the HF strain described here) was requested by and distributed to several US laboratories, where the strain was dubbed as *I. ovatus Ehrlichia* (IOE) agent. Using the HF strain-infected mouse spleen homogenate as the source of HF bacterium, pathogenesis studies in inoculated mice revealed that these bacteria induce a toxic shock-like cytokine storm, involving cytotoxic T-cells, NKT cells, and neutrophils similar to those reported in fatal HME [57–68]. Therefore, *Ehrlichia* sp. HF has been increasingly serving as a needed immunocompetent mouse model for studying fatal ehrlichiosis.

The major barriers for advancing research on *Ehrlichia* sp. HF, however, have been the inability to stably culture it in a mammalian macrophage cell line and lack of genome sequence and analysis data. Previously, it was cultured in monkey endothelial RF/6A cells and *Ixodes scapularis* tick embryo ISE6 cells [69]. To facilitate studies using *Ehrlichia* sp. HF, we stably cultured the HF strain in a canine histiocytic leukemia cell line DH82, and obtained the complete whole genome sequence (GenBank accession NZ\_CP007474). Despite many studies being conducted with *Ehrlichia* sp. HF, this bacterium has not been classified into any species, causing confusion in the literature with several different names (IOE agent, *Ehrlichia* sp. HF, the HF strain). Comparative core genome alignment and phylogenetic analysis reveal that *Ehrlichia* sp. HF is a new species that is most closely related to *E. muris* and *E. chaffeensis*, justifying the formal nomenclature of this species. The genome sequencing and analysis, including comparative virulence factor analysis of *Ehrlichia* sp. HF, provides important insights, resources, and validation for advancing the research on emerging human ehrlichioses.

## Results and Discussion

### Culture Isolation of *Ehrlichia* sp. HF and purification of *Ehrlichia* genomic DNA

To obtain sufficient amounts of bacterial DNA free from host cell DNA, we stably cultured *Ehrlichia* sp. HF in DH82 cells. Spleen and blood samples were collected from *Ehrlichia* sp. HF-infected mice euthanized at an acute stage of illness (8 d post inoculation) (Fig. S1A). Diff-Quik staining showed that the bacteria were present in blood monocytes (Fig. S1B). After 2–3 weeks co-culturing with infected spleen homogenates, large vacuoles (inclusions) containing numerous bacteria (known as morulae) were observed in the cytoplasm of DH82 (Fig. S1C) and RF/6A cells (Fig. S1D). *Ehrlichia* sp. HF could also be successfully passaged from DH82 cells to ISE6 cells (Fig. S1E). Morulae of *Ehrlichia* sp. HF in cell cultures were like those seen in the tissue sections of the thymus and the lungs of infected mice [6], and in the endothelial cells of most organs of infected mice [10]. *Ehrlichia* sp. HF cultured in DH82 cells infects and kills mice at 7–10 days post intraperitoneal inoculation, similar to those inoculated with the infected mouse spleen homogenate, demonstrating that *Ehrlichia* sp. HF culture isolate maintains mouse virulence [56]. The mouse LD<sub>50</sub> of *Ehrlichia* sp. HF cultured in DH82 cells is approximately 100 bacteria [56].

### General features of the *Ehrlichia* sp. HF genome

The complete genome of *Ehrlichia* sp. HF was sequenced using both Illumina and PacBio platforms, and the reads from both platforms were combined at multiple levels in order to obtain a reliable assembly. The

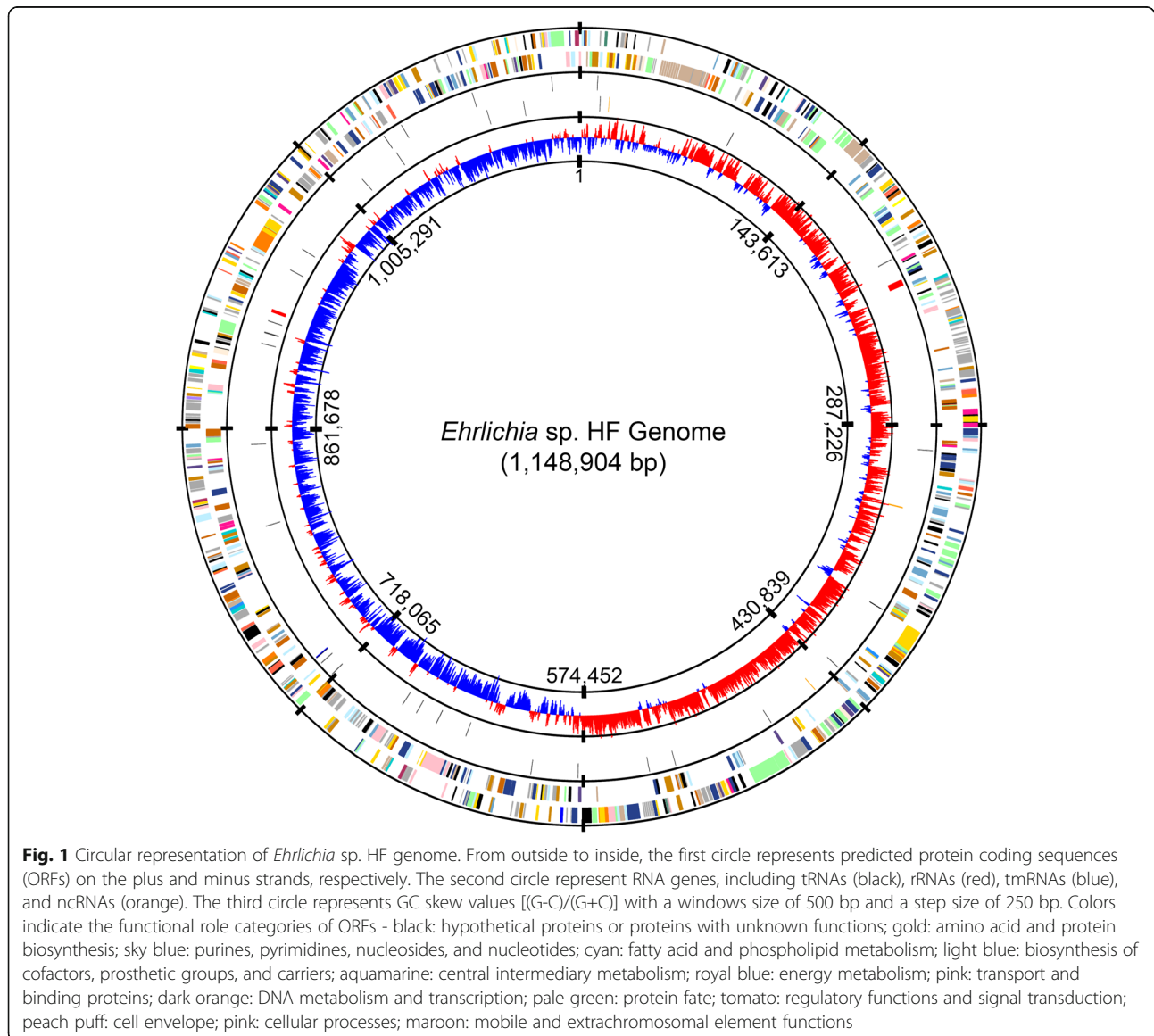
genome was rotated to the replication origin of *Ehrlichia* sp. HF (Fig. 1), which was predicted to be the region between *hemE* (uroporphyrinogen decarboxylase, EHF\_0001) and *tlyC* (hemolysin or related HlyC/CorC family transporter, EHF\_0999) as described for other members in the family Anaplasmataceae [70]. Annotation of the finalized genome assembly was generated using the IGS prokaryotic annotation pipeline [71]. The completed genome of *Ehrlichia* sp. HF is a single double-stranded circular chromosome of 1,148,904 bp with an overall G+C content of ~30%, which is similar to those of *E. chaffeensis* Arkansas [72], *E. muris* subsp. *eauclairensis* Wisconsin [19], and *E. muris* AS145<sup>T</sup> [73] (Table 2).

The *Ehrlichia* sp. HF genome encodes one copy each of the 5S, 16S, and 23S rRNA genes, which are separated

in 2 locations with the 5S and 23S rRNA being adjacent (Fig. 1, red bars in the middle circle) as in other sequenced members in the family Anaplasmataceae [72, 74]. Thirty-six tRNA genes are identified with cognates for all 20 amino acids (AA) (Table 2 and Fig. 1, black bars in the middle circle), similar to other *Ehrlichia* spp. (36 – 37 genes, Table 2).

#### Comparative genomic analysis of *Ehrlichia* sp. HF with other *Ehrlichia* species

Previous studies have shown that some *Anaplasma* spp. and *Ehrlichia* spp. have a single large-scale symmetrical inversion (X-alignment) near the replication origin, which may have resulted from recombination between duplicated, but not identical *rho* termination factors [72, 75, 76]. All genomes of the sequenced *Ehrlichia* spp. encode



**Table 2** Genome properties of representative *Ehrlichia* species

| <i>Ehrlichia</i> Species <sup>1</sup> | EHF         | ECH        | EMU        | EmCRT <sup>2</sup> | ECA        | ERW        |
|---------------------------------------|-------------|------------|------------|--------------------|------------|------------|
| NCBI RefSeq                           | NZ_CP007474 | NC_007799  | NC_023063  | NZ_LANU01000001    | NC_007354  | NC_005295  |
| Size (bp)                             | 1,148,904   | 1,176,248  | 1,196,717  | 1,148,958          | 1,315,030  | 1,516,355  |
| GC (%)                                | 29.6        | 30.1       | 29.7       | 29.8               | 29.0       | 27.5       |
| Protein                               | <b>866</b>  | <b>892</b> | <b>874</b> | <b>866</b>         | <b>933</b> | <b>934</b> |
| tRNA                                  | 36          | 37         | 37         | 36                 | 36         | 36         |
| rRNA                                  | 3           | 3          | 3          | 3                  | 3          | 3          |
| Other RNA                             | 4           | 3          | 3          | 4                  | 3          | 4          |
| Pseudogene                            | 11          | 17         | 24         | 15                 | 10         | 18         |
| Total Gene                            | <b>920</b>  | <b>952</b> | <b>941</b> | <b>924</b>         | <b>985</b> | <b>995</b> |

<sup>1</sup>Abbreviations: EHF *Ehrlichia* sp. HF (HF565), EMU *E. muris* subsp. *muris* AS145, EmCRT *E. muris* subsp. *eaucalarensis* Wisconsin, ECH *E. chaffeensis* Arkansas, ECA *E. canis* Jake, ERW *E. ruminantium* Welgevonden

<sup>2</sup>The genome of *E. muris* subsp. *eaucalarensis* Wisconsin is incomplete, consisting of 3 contigs, NZ\_LANU01000001, NZ\_LANU01000002, and NZ\_LANU01000003

duplicated *rho* genes. Whole genome alignments demonstrate that the *Ehrlichia* sp. HF genome exhibits almost complete synteny with other *Ehrlichia* spp., including *E. muris*, *E. canis*, and *E. ruminantium*, without any significant genomic rearrangements or inversions despite these genomes being oriented in the opposite directions (Fig. 2). However, *Ehrlichia* sp. HF has a single large-scale symmetrical inversion relative to *E. chaffeensis* at the duplicated *rho* genes (Fig. 2b). Large scale inversion was also reported in other bacteria such as *Yersinia* and *Legionella* species when genomes of closely related species are compared [77]. However, the biological meaning and evolutionary implications of such process, if any, are largely unknown.

In order to compare the protein ortholog groups among four closely-related *Ehrlichia* spp., including *Ehrlichia* sp. HF, *E. muris* subsp. *eaucalarensis*, *E. muris* AS145, and *E. chaffeensis* Arkansas, 4-way comparisons were performed using reciprocal BLASTP algorithm with E-value < 1e<sup>-10</sup> (Fig. 3). The four-way comparison showed that the core proteome, defined as the set of proteins present in all four genomes, consists of 823 proteins representing 94.9% of the total 867 protein-coding ORFs in *Ehrlichia* sp. HF (Fig. 3 and Table 3). Among these conserved proteins, the majority are associated with housekeeping functions and are likely essential for *Ehrlichia* survival (Table 3).

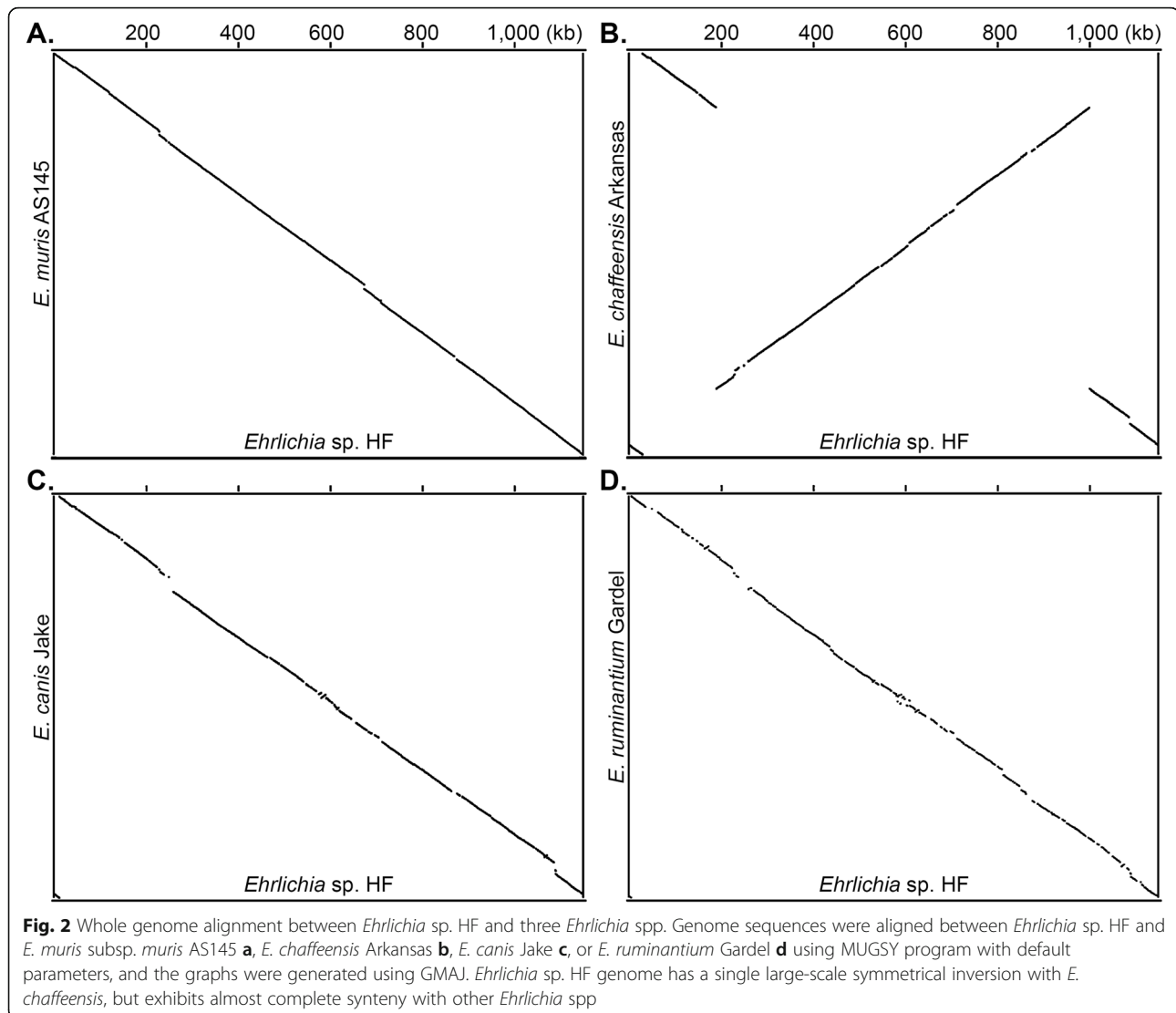
By 4-way comparison, a hypothetical protein (EHF\_RS02845 or MR76\_RS01735) is found only in *Ehrlichia* sp. HF and *E. muris* subsp. *muris*, the two strains that do not infect humans, but not in *E. chaffeensis* and *E. muris* subsp. *eaucalarensis*, which both infect humans [11, 12, 78] (Table S1). On the other hand, the human-infecting strains of *E. chaffeensis* and *E. muris* subsp. *eaucalarensis* have genes encoding a bifunctional DNA-formamidopyrimidine glycosylase/DNA-(apurinic or apyrimidinic site) lyase protein, MutM (ECH\_RS02515 or EMUCRT\_RS01070) (Table

S1). In addition, transposon mutagenesis studies have identified intragenic insertions of genes encoding DNA mismatch repair proteins MutS and MutL in *Ehrlichia* sp. HF [56]. Biological relationship between MutM and the human infectivity remains to be investigated.

*E. muris* subsp. *muris*, *E. muris* subsp. *eaucalarensis*, and *Ehrlichia* sp. HF cause persistent or lethal infection in mice, whereas immunocompetent mice clear *E. chaffeensis* infection within 10 – 16 days [79–81]. A metallophosphoesterase (ECH\_RS03950/ECH\_0964), which may function as a phosphodiesterase or serine/threonine phosphoprotein phosphatase, was found only in *E. chaffeensis* but not in the other three *Ehrlichia* spp. (Table S2).

Except for 28 *E. chaffeensis*-specific proteins, there are less than 10 species-specific proteins present in *Ehrlichia* sp. HF, *E. muris* subsp. *muris* AS145, or *E. muris* subsp. *eaucalarensis* (Table S2), all of which are hypothetical proteins without any known functions or domains. Potentially, these proteins may be involved in differential pathogenesis of these *Ehrlichia* species.

Two-way comparisons identified further proteins that are unique to *Ehrlichia* sp. HF, but absent in other *Ehrlichia* spp. (Table S3). Several of these proteins are involved in DNA metabolism, mutation repairs, or regulatory functions that were only found in *Ehrlichia* sp. HF (Table S3). For example, compared to *Ehrlichia* sp. HF proteomes, *E. chaffeensis* lacks a patatin-like phospholipase family protein (ECH\_RS03820, a pseudogene with internal frameshift at AA<sup>180</sup>), which has phospholipase A<sub>2</sub> activity catalyzing the nonspecific hydrolysis of phospholipids, glycolipids, and other lipid acyl hydrolase activities [82–84]. *E. muris* subsp. *muris* lacks CckA protein, a histidine kinase that can phosphorylate response regulator CtrA and regulate the DNA segregation and cell division of *E. chaffeensis* [85, 86]. However, the absence of these proteins needs to be further validated since sequencing errors and

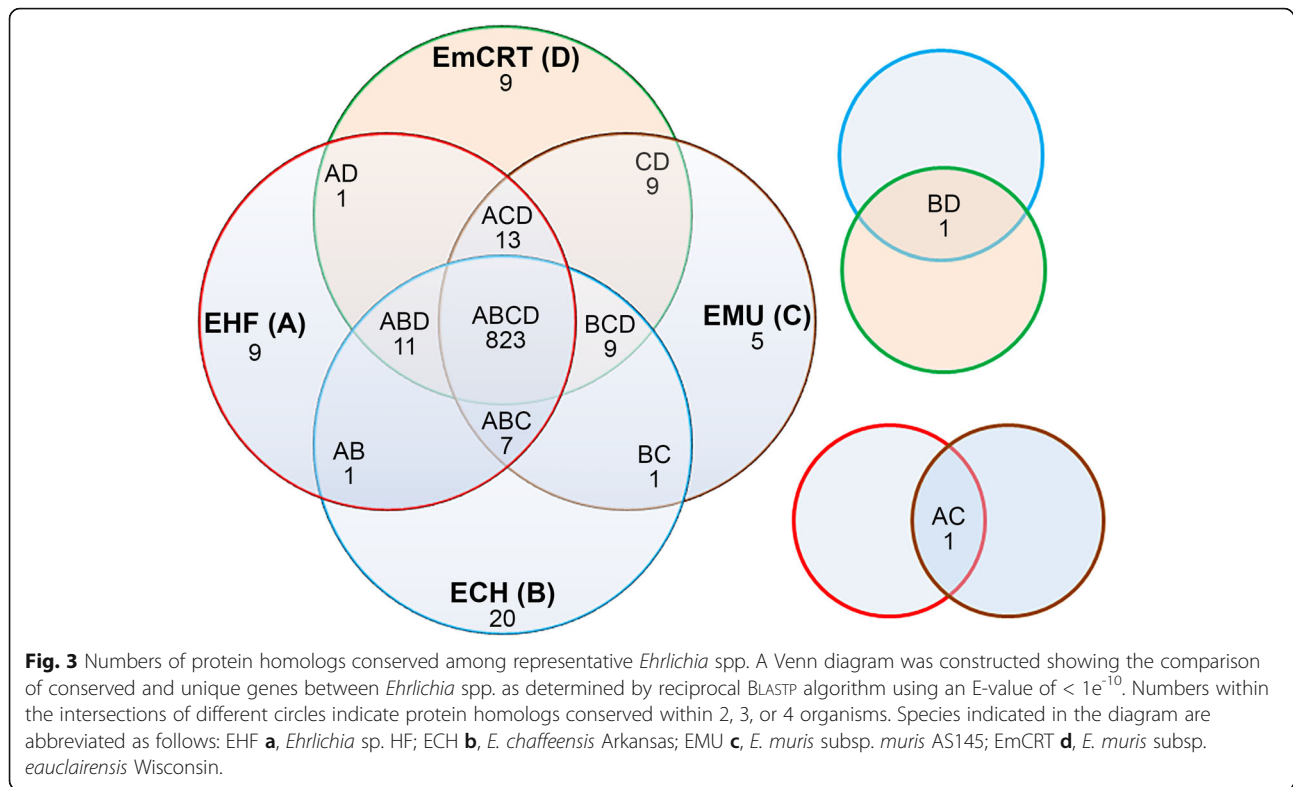


mis-annotations can frequently confound such analyses. For example, although the homolog to *E. chaffeensis* TRP120 was not identified in *E. muris* subsp. *eauclairensis*, TBLASTN searches indicated that this ORF is split into two pseudogenes (EMUCRT\_0995 and EMUCRT\_0731) in two separate contigs of the draft genome sequences. In addition, RpoB/C were misannotated in *E. muris* subsp. *eauclairensis* genome as a concatenated pseudogene EMUCRT\_RS04655, whereas several genes encoding GyrA, PolI, AtpG, and CckA of *E. muris* AS145 were annotated as pseudogenes due to frameshifts in homopolymeric tracts (Table S3).

#### Metabolic and Biosynthetic Potential

The metabolic potential of *Ehrlichia* sp. HF (Table 3) was analyzed by functional role categories using Genome

Properties [87], Kyoto Encyclopedia of Genes and Genomes (KEGG) [88], and Biocyc [89]. In addition, by two and four-way comparisons between *Ehrlichia* sp. HF and *E. chaffeensis* (Fig. 3 and Table 3), results indicated that *Ehrlichia* sp. HF possesses similar metabolic pathways as previously described for *E. chaffeensis* [72]. *Ehrlichia* sp. HF genome encodes pathways for aerobic respiration to produce ATP, including pyruvate metabolism, the tricarboxylic acid (TCA) cycle, and the electron transport chain, but lacks critical enzymes for glycolysis and gluconeogenesis. Similar to *E. chaffeensis*, *Ehrlichia* sp. HF can synthesize fatty acids, nucleotides, and cofactors, but has very limited capabilities for amino acid biosynthesis, and is predicted to make only glycine, glutamine, glutamate, aspartate, arginine, and lysine. *Ehrlichia* sp. HF encodes very few enzymes related to central



**Table 3** Role category breakdown of protein coding genes in *Ehrlichia* species

| Role Category <sup>1</sup>                                 | EHF        | ECH        | EMU        | EmCRT      | Unique in EHF <sup>2</sup> |
|--|------------|------------|------------|------------|----------------------------|
| Amino acid biosynthesis                                    | 22         | 23         | 23         | 22         |                            |
| Biosynthesis of cofactors, prosthetic groups, and carriers | 64         | 60         | 65         | 61         |                            |
| Cell envelope  | 53         | 51         | 51         | 48         | 1                          |
| Cellular processes   | 42         | 41         | 42         | 41         |                            |
| Central intermediary metabolism                            | 3          | 3          | 5          | 3          |                            |
| DNA metabolism   | 41         | 44         | 41         | 42         |                            |
| Energy metabolism  | 84         | 82         | 80         | 83         |                            |
| Fatty acid and phospholipid metabolism                     | 20         | 19         | 21         | 21         |                            |
| Mobile elements  | 4          | 4          | 4          | 4          |                            |
| Protein fate   | 79         | 78         | 77         | 78         |                            |
| Protein synthesis  | 108        | 108        | 107        | 107        |                            |
| Nucleotide biosynthesis                                    | 35         | 35         | 35         | 35         |                            |
| Regulatory functions                                       | 14         | 15         | 13         | 14         |                            |
| Transcription  | 21         | 21         | 19         | 19         |                            |
| Transport and binding proteins                             | 33         | 33         | 32         | 33         |                            |
| Hypothetical proteins or proteins with unknown functions   | 244        | 276        | 268        | 255        | 8                          |
| Total Assigned Functions:                                  | <b>623</b> | <b>617</b> | <b>615</b> | <b>611</b> |                            |
| Total Proteins   | <b>867</b> | <b>893</b> | <b>883</b> | <b>866</b> |                            |

<sup>1</sup>Abbreviations: EHF *Ehrlichia* sp. HF, ECH *E. chaffeensis* Arkansas, EMU *E. muris* subsp. *muris* AS145, EmCRT *E. muris* subsp. *eaucalarensis* Wisconsin.

<sup>2</sup>Proteins specific to *Ehrlichia* sp. HF are based on 4-way comparison analysis among four *Ehrlichia* spp. by Blastp (E < 1e-10)

intermediary metabolism (Table 3) and partially lacks genes for glycerophospholipid biosynthesis, rendering this bacterium dependent on the host for its nutritional needs, like *E. chaffeensis* [90, 91].

*Ehrlichia* species, including the HF strain and *E. chaffeensis*, are deficient in biosynthesis pathways of typical pathogen-associate molecular patterns (PAMPs), including lipopolysaccharide, peptidoglycan, common pili, and flagella. Nevertheless, both *E. chaffeensis* and *Ehrlichia* sp. HF induce acute and/or chronic inflammatory cytokines production in a MyD88-dependent, but Toll-like receptors (TLR)-independent manner [92–94]. Similar to acute severe cases of HME, *Ehrlichia* sp. HF causes an acute toxic shock-like syndrome in mice involving many inflammatory factors and kills mice in 10 days [56, 61, 66, 67], suggesting that *Ehrlichia* species have unique, yet to be identified inflammatory molecules.

### Two-component regulatory systems

A two-component regulatory system (TCS) is a bacterial signal transduction system, generally composed of a sensor histidine kinase and a cognate response regulator, which allows bacteria to sense and respond rapidly to environmental changes [95]. Our previous studies showed that *E. chaffeensis* encodes three pairs of TCSs, including CckA/CtrA, PleC/PleD, and NtrX/NtrY, and that the histidine kinase activities were required for bacterial infection [85, 86]. Analysis showed that all three histidine kinases were identified in four species of *Ehrlichia* including *Ehrlichia* sp. HF (Table 4). However, the response regulator *cckA* gene of *E. muris* subsp. *muris* AS145 was annotated as a pseudogene due to an internal frameshift (Table 4). Since CckA regulates the critical biphasic developmental cycle of *Ehrlichia*, which converts between infectious compact dense-cored cell (DC) and replicative larger reticulate cell (RC) form [85], the mutation of *cckA* in *E. muris* AS145 needs to be further validated to rule out sequencing error in a homopolymeric tract.

### *Ehrlichia* Outer Membrane Proteins (Omps)

*Ehrlichia* spp. encode 14 – 23 tandemly-arrayed paralogous Omp-1/P28 major outer membrane family proteins in a >26 kb genomic region [52, 93, 96–98]. This polymorphic multigene family is located downstream of *tr1*, a putative transcription factor, and upstream of *secA* gene [97]. Compensating for incomplete metabolic pathways, the major outer membrane proteins P28 and Omp-1F of *E. chaffeensis* possess porin activities for nutrient uptake from the host, which allow the passive diffusion of L-glutamine, the monosaccharides arabinose and glucose, the disaccharide sucrose, and even the tetrasaccharide stachyose as determined by a proteoliposome swelling assay [99]. The *Ehrlichia* sp. HF genome

has 23 paralogous *omp-1/p28* family genes, named *omp-1.1* to *omp-1.23* (Fig. 4), and similarly flanked by *tr1* and *secA* genes. Comparing with the *E. chaffeensis* Omp-1/P28 proteins by the best matches from BLASTP search, the HF genome lacks orthologs of *E. chaffeensis* Omp-1Z, C, D, F, and P28-2, but has duplicated Omp-1H and 6 copies of Omp-1E (Fig. 4). Since P28 and Omp-1F of *E. chaffeensis* showed different solute diffusion rates [99], the divergence of *Ehrlichia* sp. HF Omp-1 protein family could affect the effectiveness of nutrient acquisition by these bacteria.

Gram-negative bacteria encode a conserved outer membrane protein Omp85 (or YaeT) for outer membrane protein assembly [100, 101], and a molecular chaperone OmpH that interacts with unfolded proteins as they emerge in the periplasm from the Sec translocation machinery [102, 103]. The outer membrane lipoprotein OmpA of *E. chaffeensis* is highly expressed [104–106], and OmpA family proteins in other gram-negative bacteria are well characterized for their roles in porin functions, bacterial pathogenesis, and immunity [107]. All three outer membrane proteins were identified in *Ehrlichia* sp. HF, and highly conserved in these *Ehrlichia* spp. (Table 4), suggesting their essential roles in bacterial infection and survival.

Our previous studies showed that *E. chaffeensis* uses its outer membrane invasin EtpE to bind host cell receptor DNase X, and regulates signaling pathways required for entry and concomitant blockade of reactive oxygen species production for successful infection of host monocytes [108–111]. Analysis showed that the homologs of EtpE were present in *Ehrlichia* sp. HF as well as other *Ehrlichia* (Table 4), suggesting these bacteria might use similar mechanisms for entry and infection of their host cells.

### Protein secretion systems

*Ehrlichia* sp. HF encodes all major components for the Sec-dependent protein export system to secrete proteins across the membranes. In addition, intracellular bacteria often secrete effector molecules into host cells via Sec-independent pathways, which regulate host cell physiological processes, thus enhancing bacterial survival and/or causing diseases [112]. Analysis of the *Ehrlichia* sp. HF genome identifies the Sec-independent Type I secretion system (T1SS), which can transport target proteins with a C-terminal secretion signal across both inner and outer membranes into the extracellular medium, and twin-arginine dependent translocation (TAT) pathway, which can transport folded proteins across the bacterial cytoplasmic membrane by recognizing N-terminal signal peptides harboring a distinctive twin-arginine motif (Table 4) [113].



**Table 4** Potential pathogenic genes in *Ehrlichia* sp. HF, *E. chaffeensis*, *E. muris* subsp. *muris*, and *E. muris* subsp. *eaucalarensis*

| Organisms <sup>1</sup>                                       | EHF        | ECH        | EMU            | EmCRT          |
|--|------------|------------|----------------|----------------|
| <b>Outer Membrane Proteins:</b>                              |            |            |                |                |
| Omp-1/P28 family proteins                                    | 23         | 22         | 20             | 20             |
| Omp85  | +          | +          | +              | +              |
| OmpH   | +          | +          | +              | +              |
| OmpA family protein  | +          | +          | +              | +              |
| EtpE   | +          | +          | +              | +              |
| <b>Type IV Secretion System:</b>                             |            |            |                |                |
| VirB1/B5   | -          | -          | -              | -              |
| VirB2  | + (5)      | + (4)      | + (4)          | + (5)          |
| VirB3  | +          | +          | +              | +              |
| VirB4  | + (2)      | + (2)      | + (2)          | + (2)          |
| VirB6  | + (4)      | + (4)      | + (4)          | + (4)          |
| VirB7  | +          | +          | +              | +              |
| VirB8  | + (2)      | + (2)      | + (2)          | + (2)          |
| VirB9  | + (2)      | + (2)      | + (2)          | + (2)          |
| VirB10/B11/D4  | +          | +          | +              | +              |
| <b>Putative T4SS Effectors:</b>                              |            |            |                |                |
| Etf-1  | +          | +          | +              | +              |
| Etf-2 <sup>2</sup>   | ±          | +          | ±              | ±              |
| Etf-3  | +          | +          | +              | +              |
| <b>Type I Secretion System<sup>3</sup></b>                   |            |            |                |                |
|  | +          | +          | +              | +              |
| <b>Twin-arginine Translocation (TAT) Pathway<sup>4</sup></b> |            |            |                |                |
|  | +          | +          | +              | +              |
| <b>TRP Proteins</b>  |            |            |                |                |
| TRP32  | + (94 aa)  | + (198 aa) | + (112 aa)     | + (105 aa)     |
| TRP47  | + (255 aa) | + (316 aa) | + (228 aa)     | + (252 aa)     |
| TRP120   | + (584 aa) | + (548 aa) | + (1288 aa)    | + <sup>5</sup> |
| <b>Ankyrin-repeat domain proteins</b>                        |            |            |                |                |
|  | 5          | 5          | 5              | 5              |
| <b>Two-Component Regulatory Systems:</b>                     |            |            |                |                |
| PleC/PleD  | +          | +          | +              | +              |
| NtrY/NtrX  | +          | +          | +              | +              |
| CckA/CtrA  | +          | +          | ± <sup>6</sup> | +              |

<sup>1</sup>Abbreviations: EHF, *Ehrlichia* sp. HF; EMU, *E. muris* subsp. *muris* AS145; ECH, *E. chaffeensis* Arkansas; EmCRT, *E. muris* subsp. *eaucalarensis* Wisconsin. Numbers inside parentheses indicate the copy number of the gene; or else, only a single copy exists. +, genes present; -, homolog of the gene not identified based on Blast searches.

<sup>2</sup>In addition to Etf-2 (ECH\_0261, 264 aa), *E. chaffeensis* encodes six paralogs of Etf-2 with protein sizes range from 190 ~ 350 AA (ECH\_0243, 293 aa; ECH\_0246, 285 aa; ECH\_0247, 316 aa; ECH\_0253, 189 aa; ECH\_0255, 352 aa; and ECH\_0257, 226 aa). However, only low homologies (26 ~ 32% AA sequence identity) to *E. chaffeensis* Etf-2 were identified in other *Ehrlichia* spp. (indicated by ±)

<sup>3</sup>Type I Secretion System is consisting of an outer membrane channel protein TolC, a membrane fusion protein HlyD, and an ATPase HlyB. All are present in these *Ehrlichia* spp

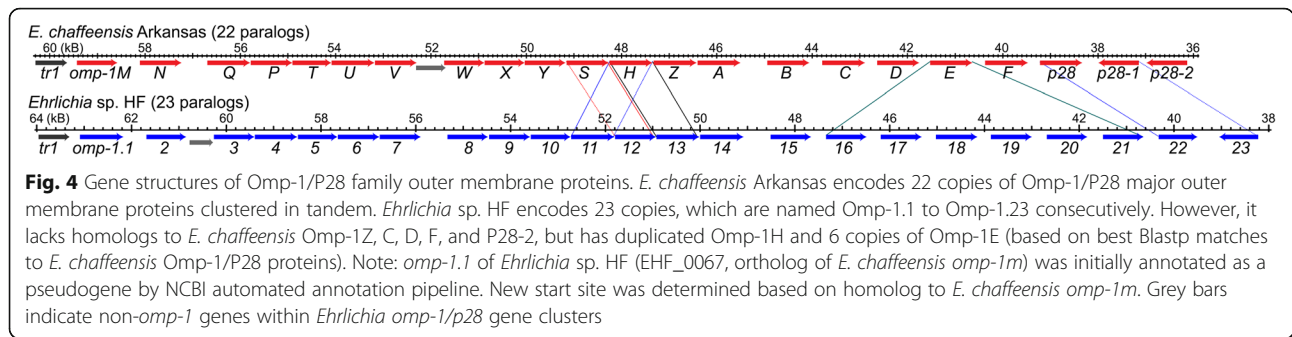
<sup>4</sup>Both twin-arginine translocase subunits TatA and TatC were identified in all *Ehrlichia* spp

<sup>5</sup>Tblastn search indicates that that the homolog of *E. chaffeensis* TRP120 in *E. muris* subsp. *eaucalarensis* Wisconsin is split into two pseudogenes (EMUCRT\_0995 and EMUCRT\_0731) present in two separate contigs (NZ\_LANU01000002 and NZ\_LANU01000003) of the incomplete genome sequences

<sup>6</sup>Gene encoding CtrA protein was identified in *E. muris* subsp. *muris* AS145 genome. However, *cckA* gene is annotated as a pseudogene due to an internal deletion, causing frameshift at 1,123 bp

The Type IV secretion system (T4SS) is a protein secretion system of Gram-negative bacteria that can translocate bacterial effector molecules into host cells and plays a key role in pathogen-host interactions [90, 114]. Except for VirB1 and VirB5, all key components of the

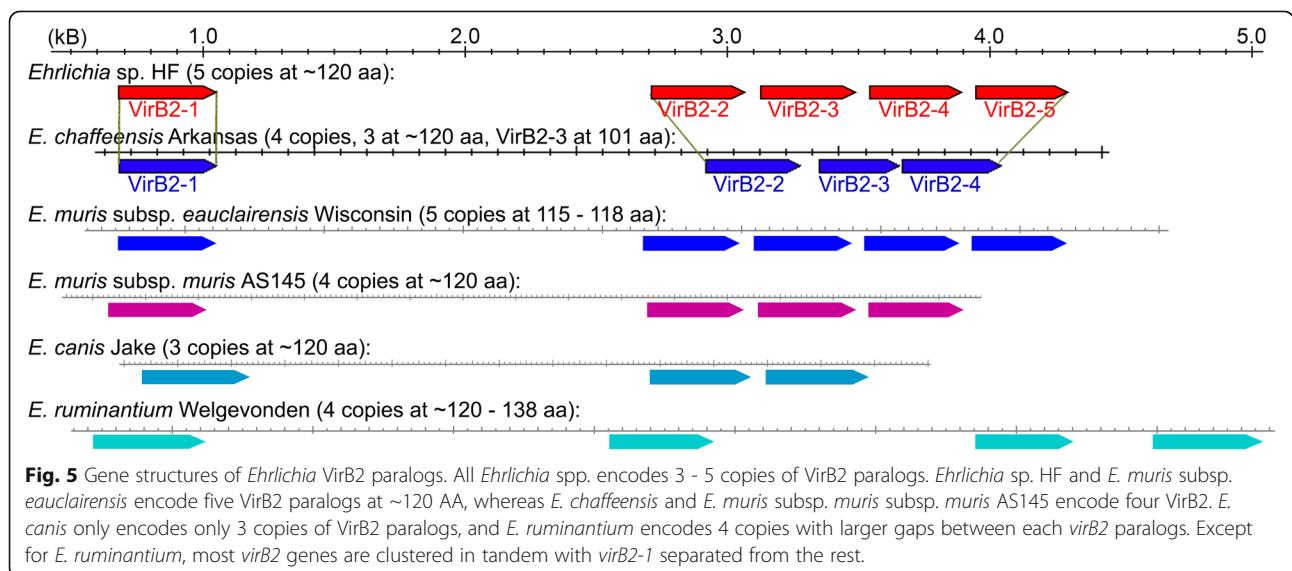
T4SS apparatus were identified in *Ehrlichia* sp. HF, similar to those of *E. chaffeensis* (Table 4). The minor pilus subunit VirB5 is absent in all Rickettsiales [115]. VirB1, which is involved in murein degradation, is not present in *Ehrlichia* spp., likely due to the lack of peptidoglycan.



These *virB/D* genes encoding T4SS apparatus are split into three major operons as well as single genes in three separate loci that encode VirB7 and duplicated VirB8/9 proteins (Table 4 and Fig. S2). Genes encoding VirB4 are also duplicated, which are clustered with multiple paralogs of *virB2* and *virB6* genes (Table 4 and Fig. S2). *Ehrlichia* sp. HF encodes four tandem functionally uncharacterized VirB6-like paralogs (800 – 1,942 AA), which have increasing masses and are three- to six-fold larger than *Agrobacterium tumefaciens* VirB6 (~300 AA), with extensions found at both N- and C-terminus [116].

In *A. tumefaciens*, VirB2 is the major T-pilus component that forms the main body of this extracellular structure, which is believed to initiate cell-cell contact with plant cells prior to the initiation of T-complex transfer [117, 118]. A yeast two-hybrid screen identified interaction partners in *Arabidopsis thaliana*, suggesting that *Agrobacterium* VirB2 directly contacts the host cell during the substrate translocation process [114, 119, 120]. Compared to *E. chaffeensis* and *E. muris* subsp. *muris* AS145 that encode four VirB2 paralogs, both *Ehrlichia* sp. HF and *E. muris* subsp. *eaucloirensis* encode five VirB-2 paralogs at ~120 AA (Table 4 and Fig. 5).

Most *virB2* genes are clustered in tandem except for *virB2-1*, which is separated from the rest. VirB2 paralogs are quite divergent and only share 26% identities despite their similar sizes and domain architecture among Rickettsiales [115, 121]. Phylogenetic analysis of VirB2 paralogs in representative *Ehrlichia* species showed that VirB2-1 proteins are clustered in a separate branch; whereas the rest of VirB2 paralogs are more divergent (Fig. S3). *A. tumefaciens* VirB2 undergoes a novel head-to-tail cyclization reaction and polymerizes to form the T-pilus [116], and mature VirB2 integrates into the cytoplasmic membrane via two hydrophobic  $\alpha$ -helices [122, 123]. Analysis of *Ehrlichia* sp. HF VirB2-4 showed that it possesses a signal peptide (cleavage site between residues 29 and 30) and two hydrophobic transmembrane  $\alpha$ -helices (Fig. S4A). Alignment of these VirB2 paralogs showed that two hydrophobic  $\alpha$ -helices are completely conserved, although they are more divergent on the N- and C-terminus (Fig. S4B), suggesting that *Ehrlichia* VirB2s could form the secretion channels for mature T4SS pili as in *Agrobacterium* [121]. Our previous study confirmed that VirB2 is expressed on the surface of a closely related bacterium *Neorickettsia risticii* [124].



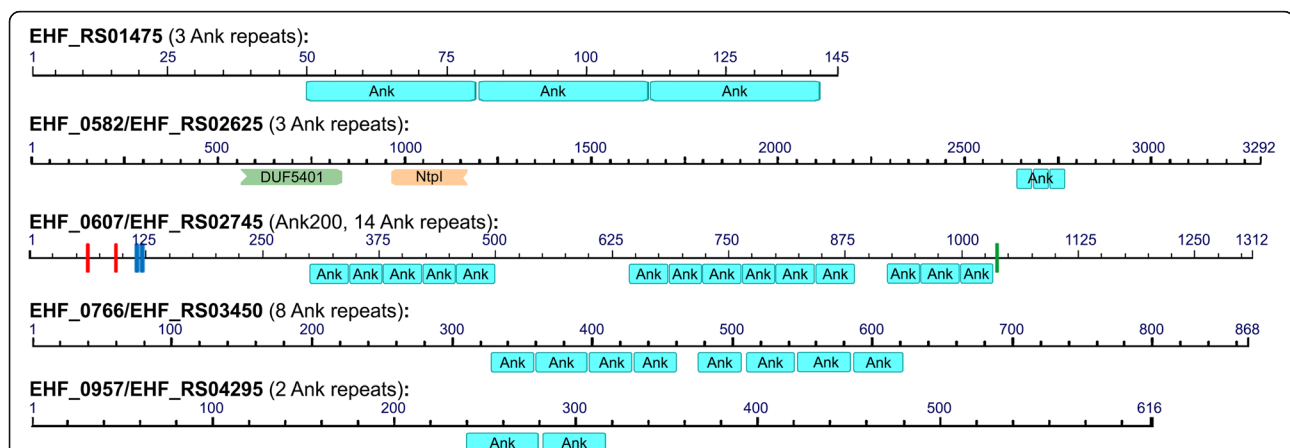
Studies indicated that VirB2 paralogs of *Anaplasma phagocytophilum* are differentially expressed in tick and mammalian cells [125], and an outer membrane vaccine of *Anaplasma marginale* containing VirB2 can protect against the disease and persistent infection [126, 127]. Therefore, the expression of VirB2 paralogs could be specific to the host environment, and their highly divergent C-terminus may offer antigenic variations for protection from host adaptive immunity.

### Putative T4SS Effectors

In contrast to other intracellular pathogens with enormous numbers of effectors (i.e. *Legionella pneumophila*), *E. chaffeensis* encodes much fewer but versatile effectors [128]. Three *E. chaffeensis* T4SS effectors have been experimentally characterized, namely *Ehrlichia* translocated factor (Etf)-1, -2, and -3 [129]. These T4SS effectors are essential for infection of host cells, through inhibition of host apoptosis by Etf-1 [129], acquisition of host nutrients by Etf-1-induced autophagosomal pathways [90], or maintenance of the bacterial replication compartments by Etf-2-mediated inhibition of endosome maturation [130]. Homologs of Etf-1 and Etf-3 were identified in all *Ehrlichia* spp., and they are highly conserved with percent protein identities over 77% and 85%, respectively. Etf-2 proteins are more divergent among *Ehrlichia* spp., and *E. chaffeensis* encodes five paralogs of Etf-2 with protein lengths range from 190 ~ 350 AA; however, only low homologies (26 ~ 32% protein identity) to *E. chaffeensis* Etf-2 were identified in *Ehrlichia* sp. HF and other *Ehrlichia* species (Table 4). Whether these proteins contain a T4SS motif and can be secreted into the host cell cytoplasm remains to be studied.

### Ankyrin-repeat containing proteins

Ankyrin-repeats (Ank) are structural repeating motifs that consist of 33-AA with two anti-parallel  $\alpha$ -helices connected to the next repeat via a loop region [131]. Ank proteins are more common in eukaryotes, which mediate protein–protein interactions involved in a multitude of host processes including cytoskeletal motility, tumor suppression, and transcriptional regulation [131]. AnkA of *A. phagocytophilum* is one of a few known T4SS effectors, which can be translocated into the host cells, tyrosine-phosphorylated, and plays an important role in facilitating intracellular infection by regulating host signaling pathways [132–134]. *A. phagocytophilum* AnkA can also be translocated to the cell nucleus and bind to transcriptional regulatory regions of the *CYBB* locus to suppress host-cell innate immune response [135, 136]. The AnkA homolog in *E. chaffeensis*, Ank200, also contains tyrosine kinase phosphorylation sites and can be tyrosine-phosphorylated in the infected host cells [137, 138]. *E. chaffeensis* Ank200 interacts with Alu-Sx elements to regulate several genes associated with ehrlichial pathobiology [139]. A homolog of *E. chaffeensis* Ank200 was identified in *Ehrlichia* sp. HF (EHF\_0607), which also contains two putative tyrosine kinase phosphorylation sites and SH3 domains in addition to 14 Ank repeats [133] (Fig. 6). Our analysis identified four additional Ank-repeat containing proteins in these representative *Ehrlichia* spp. (Table 4). In *Ehrlichia* sp. HF, these proteins range from ~150 to over 3,000 AA in length and contain 2 - 14 copies of Ank repeats (Fig. 6). It remains to be elucidated if any of the ankyrin repeat-containing proteins in *Ehrlichia* sp. HF can be secreted, and whether these proteins



**Fig. 6** Domain structures of Ankyrin-repeat containing proteins in *Ehrlichia* sp. HF. *Ehrlichia* sp. HF encodes 5 Ank-repeat containing proteins, including *E. chaffeensis* Ank200 homolog (EHF\_0607). Ank-repeat domains were determined by NCBI Conserved Domains Database (CDD, <https://www.ncbi.nlm.nih.gov/Structure/cdd>) [140, 141], and eukaryotic phosphorylation sites were determined by Scansite 4.0 (<https://scansite4.mit.edu/>) [142]. In addition to 14 Ank repeats, Ank200 (EHF\_0607) contains two tyrosine kinase phosphorylation sites (red bars), two SH3 domains (blue), and one Ser/Thr kinase site (green). Domain abbreviations: Ank, Ankyrin repeat; DUF5401, family of unknown function initially found in Chromadorea like *Caenorhabditis elegans*; Ntpl, Archaeal/vacuolar-type  $H^+$ -ATPase subunit I/STV1.

regulate host cell signaling to benefit intracellular ehrlichial infection.

#### Tandem-repeat containing proteins (TRPs)

Using a heterologous *Escherichia coli* T1SS apparatus, studies have identified four potential *E. chaffeensis* T1SS effectors, including ankyrin-repeat containing protein Ank200, and three tandem-repeat containing proteins (TRPs), TRP47, TRP120, and TRP32 [138]. TRP120 protein also contains a motif that is rich in glycine and aspartate and relates to the repeats-in-toxins (RTX) family of exoproteins [93, 138, 143]. Our current analysis identified homologs of *E. chaffeensis* TRP proteins in *Ehrlichia* sp. HF and other representative *Ehrlichia* spp. (Table 4). In *E. chaffeensis*, all three TRP proteins contain various numbers of tandem repeats with repeat lengths ranging from 19 ~ 80 AA. However, bioinformatic analysis of TRP homologs in *Ehrlichia* spp. indicated that these proteins are highly variable, and the length and numbers of repeats are different among all *Ehrlichia* spp. (Fig. 7, Table S4). Unlike *E. chaffeensis* TRP32 and TRP47, no repeats or variable-length PCR target (VLPT) domains were detected in homologs of those in *Ehrlichia* sp. HF and other *Ehrlichia* spp. (Fig. 7a - b). Interestingly, TRP120 homolog of *Ehrlichia* sp. HF has tandem repeats with longer length (100-AA), whereas that of *E. muris* AS145 encodes a very large protein at 1,288 AA with over 12 repeats that are highly enriched in glutamic acid (Fig. 7c, Table S4). TRP120 homolog is also identified in *E. muris* subsp. *eaucclairensis*, which is split into two ORFs in two separate contigs of the incomplete genome sequences, and has a total of ~11 repeats (Fig. 7c, Table S4). Previous studies have indicated that *E. chaffeensis* TRP proteins are highly immunogenic in infected patients and animals [144], and could play important roles in host–pathogen interactions [143, 145–152]. Our recent study using Himar1 transposon mutagenesis of *Ehrlichia* sp. HF recovered a mutant with insertion within *TRP120* gene from DH82 cells, indicating that TRP120 is not essential for survival and infection of *Ehrlichia* sp. HF in DH82 cells [56]. As targeted mutagenesis of *Ehrlichia* is still unavailable, future studies using the cloned TRP120 mutant will benefit functional analysis of TRP120. In addition, it remains to be studied if any of TRPs of *Ehrlichia* sp. HF can be secreted by the T1SS, and whether these proteins regulate host cell signaling to benefit intracellular ehrlichial infection or pathogenicity.

#### *Ehrlichia* sp. HF is a new *Ehrlichia* species based on genome and proteome phylogenetic analysis

To classify *Ehrlichia* sp. HF in the genus *Ehrlichia*, we conducted phylogenetic analyses of *Ehrlichia* sp. HF by using nucleotide-based core genome alignment of

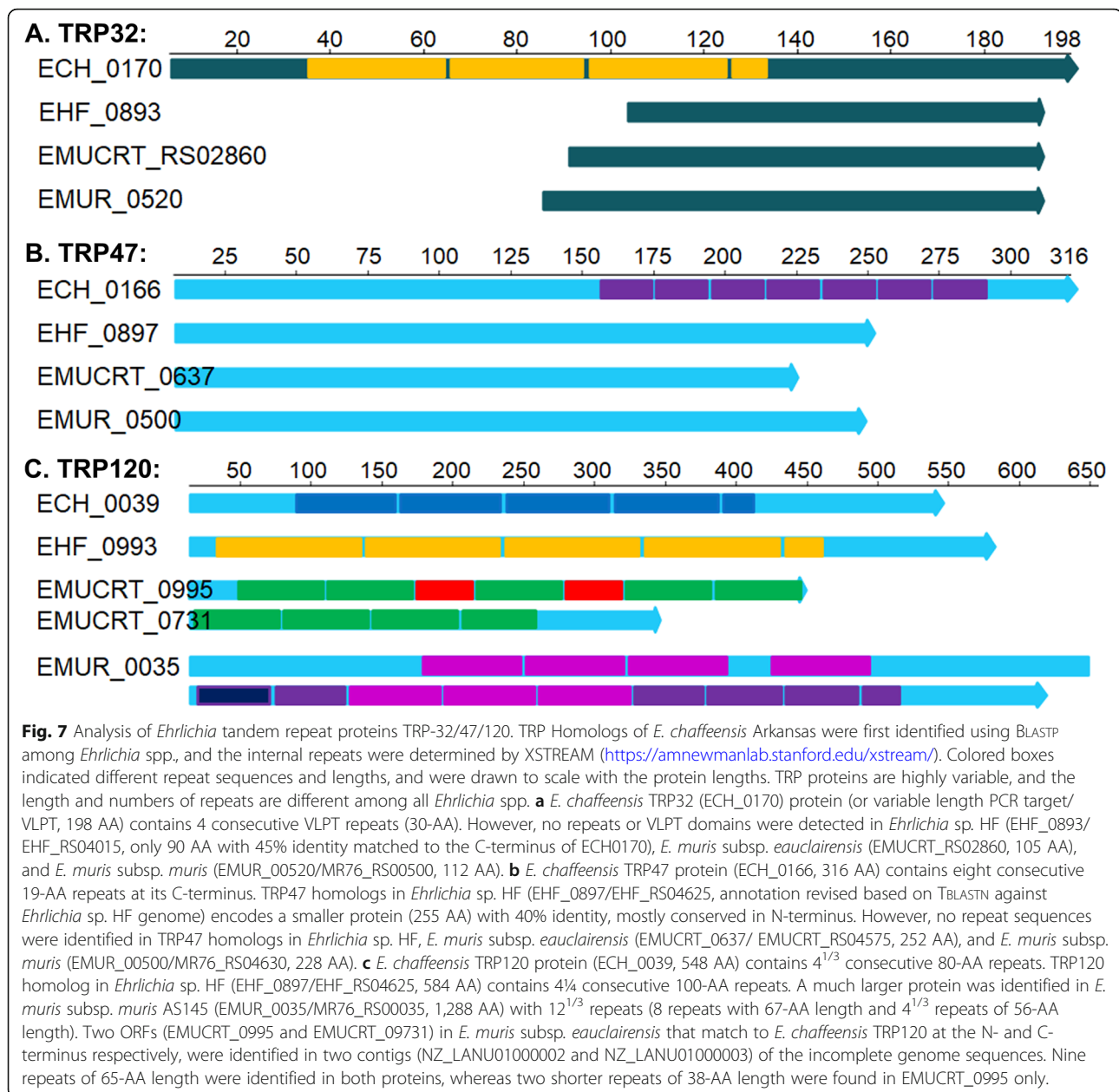
*Ehrlichia* sp. HF and 6 representative *Ehrlichia* species, subspecies, and strains by using three different parameters: (1) average nucleotide identity (ANI) [153], (2) digital DNA–DNA hybridization (dDDH) [154], and (3) core genome alignment sequence identity (CGASI) [155]. ANI values are calculated by first splitting the genome of one organism into 1 kbp fragments, which are then searched against the genome of the other organism. ANI is then calculated by taking the average sequence identity of all matches spanning >70% of their length with >60% sequence identity [153]. dDDH values are calculated by using the sequence similarity of conserved regions between two genomes and taking the sum of all identities found in matches divided by the overall match length [154]. CGASI values between genomes are calculated by generating a core genome alignment, consisting of all positions present in all analyzed genomes, and calculating the sequence identities between them [155].

Using the core genome alignment used to calculate CGASI values, the maximum-likelihood phylogenetic tree of the seven recognized species in the genus *Ehrlichia* showed that *Ehrlichia* sp. HF is a sister taxon to *E. muris* (Fig. 8a), being most closely related to *E. muris* subsp. *AS145*. However, between the two genomes, ANI, dDDH, and CGASI values are 91.8%, 43.2%, and 95.7%, respectively, all below the species cutoffs (95%, 70%, and 96.8%, respectively) [155]. Additionally, the current species designations for these 7 *Ehrlichia* genomes are supported by all three parameters, with the exception of two subspecies in *E. muris* (Fig. 8b).

Similar results are observed in a phylogenetic analyses based on the 16S rRNA sequences (Fig. S5A) or eight concatenated protein sequences (3,188 AA total) consisting of five conserved housekeeping proteins (TyrB/Mdh/Adk/FumC/GroEL) and three more divergent surface proteins like major outer membrane or T4SS apparatus proteins (P28/VirB2-1/VirB6-1) (Fig. S5B). However, the nodes on the phylogenetic tree generated using the core nucleotide alignment consistently have higher bootstrap support values than those of 16S rRNAs or concatenated proteins (Fig. 8a and S5). Based on these analyses, we proposed the following new classification of *Ehrlichia* sp. HF.

#### Description of *Ehrlichia japonica* sp. nov. (*japonica*, N.L. fem. adj. *japonica* from Japan)

The distances observed between *Ehrlichia* sp. HF and other *Ehrlichia* species by whole genome sequence-based phylogenetic analysis indicate that *Ehrlichia* sp. HF represents a new species in the genus *Ehrlichia*. This species is therefore named as *Ehrlichia japonica* sp. nov. to denote the geographic region where this bacterium was initially isolated. The type strain, HF<sup>T</sup>, was named

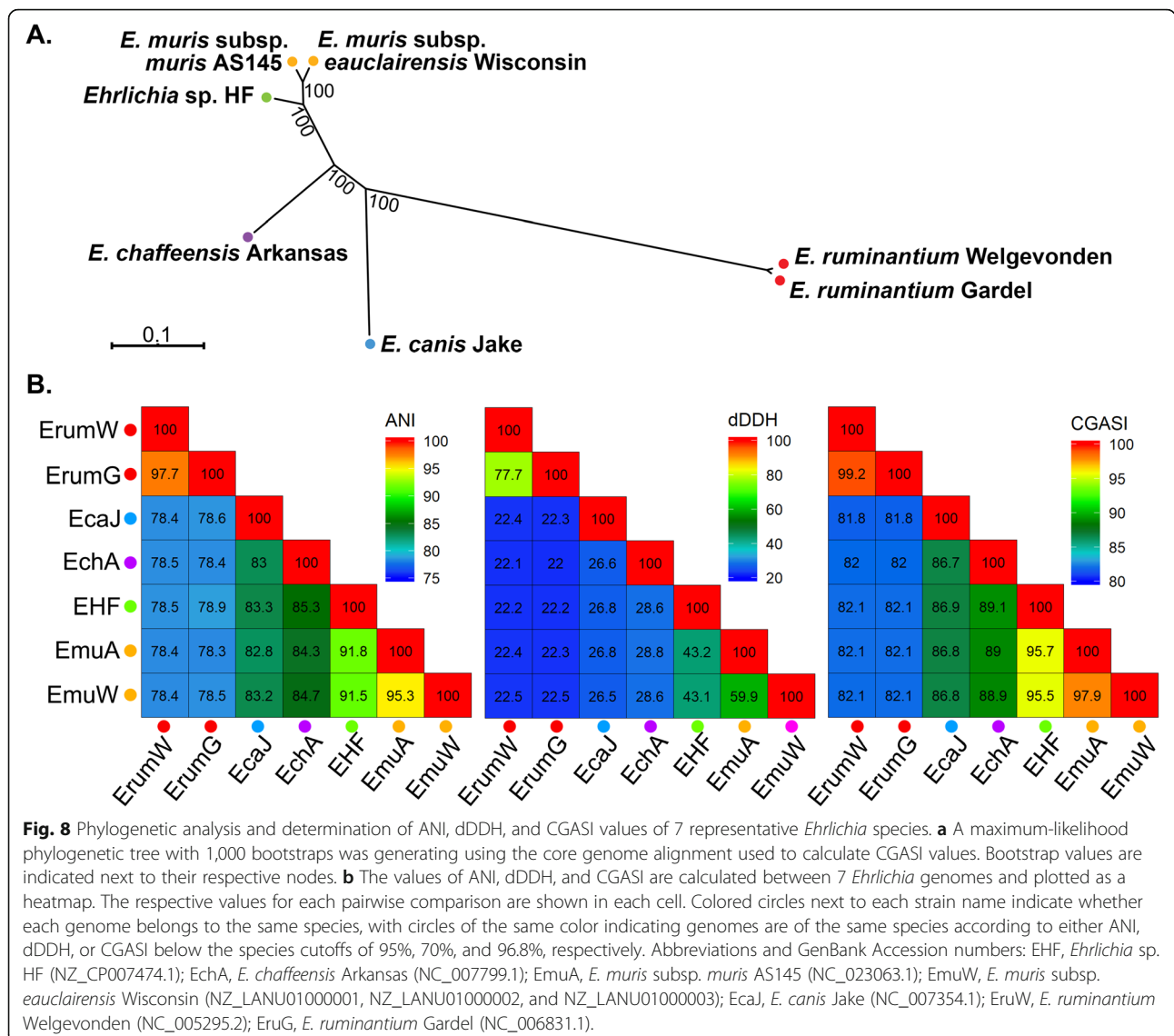


after the scientist Hiromi Fujita who first discovered and isolated this bacterium [5].

To date all *E. japonica* was found in various *Ixodes* species of ticks in Japan, France, Serbia, and Romania. This species is highly pathogenic to mice. *E. japonica* can be distinguished by PCR of 16S RNA using *Ehrlichia* sp. HF-specific primer pair HF51f/HF954r (923 bp target size, Table S5, Fig. S6) from other *Ehrlichia* species [156]. *E. japonica* HF<sup>T</sup> can be stably cultured in DH82 cells, which is available from BEI Resources (Deposit ID# NR-46450, Manassas, VA) and Collection de Souches de l'Unité des Rickettsies (CSUR Q1926, Marseille, France).

## Conclusions

By comparing with closely related *Ehrlichia* spp., this study indicates that the genome of *Ehrlichia* sp. HF encodes all homologs to virulence factors of *E. chaffeensis* required to infect host cells, including outer membrane proteins, protein secretion systems and effectors, supporting that this species can serve as a model bacteria to study in vivo pathogenesis and immune responses for fatal ehrlichiosis. Whole genome alignment and phylogenetic analyses indicate that *Ehrlichia* sp. HF can be classified as a new species in the genus *Ehrlichia*, and we propose to name it as *Ehrlichia japonica* sp. nov. Availability of this bacterial strain in macrophage cultures and complete



**Fig. 8** Phylogenetic analysis and determination of ANI, dDDH, and CGASI values of 7 representative *Ehrlichia* species. **a** A maximum-likelihood phylogenetic tree with 1,000 bootstraps was generating using the core genome alignment used to calculate CGASI values. Bootstrap values are indicated next to their respective nodes. **b** The values of ANI, dDDH, and CGASI are calculated between 7 *Ehrlichia* genomes and plotted as a heatmap. The respective values for each pairwise comparison are shown in each cell. Colored circles next to each strain name indicate whether each genome belongs to the same species, with circles of the same color indicating genomes are of the same species according to either ANI, dDDH, or CGASI below the species cutoffs of 95%, 70%, and 96.8%, respectively. Abbreviations and GenBank Accession numbers: EHF, *Ehrlichia* sp. HF (NZ\_CP007474.1); EchA, *E. chaffeensis* Arkansas (NC\_007799.1); EmuA, *E. muris* subsp. *muris* AS145 (NC\_023063.1); EmuW, *E. muris* subsp. *eaucalarensis* Wisconsin (NZ\_LANU01000001, NZ\_LANU01000002, and NZ\_LANU01000003); EcaJ, *E. canis* Jake (NC\_007354.1); ErumW, *E. ruminantium* Welgevonden (NC\_005295.2); ErumG, *E. ruminantium* Gardel (NC\_006831.1).

whole genome sequence data will greatly advance ehrlichiosis researches, including *in vivo* virulence factors, therapeutic interventions, and vaccine studies.

**Methods**

**Culture isolation of *Ehrlichia* sp. HF**

Two C57BL/6 mice (Envigo, Indianapolis, IN) were intraperitoneally inoculated with mouse spleen homogenates containing *Ehrlichia* sp. HF in RPMI-1640 (Mediatech, Manassas, VA) freezing medium containing 20% fetal bovine serum (FBS; Atlanta Biologicals, Lawrenceville, GA) and 10% DMSO (Millipore Sigma, Burlington, MA), which are stored in liquid nitrogen at approximate 0.35 ml, equivalent to 1/2 of an infected spleen. Clinical signs and body weight were monitored daily. Moribund mice at 8 day post inoculation were euthanized by CO<sub>2</sub> inhalation and cervical dislocation.

Blood samples were collected by cardiac puncture, and buffy coat was separated by centrifugation at 1,000 × g. The presence of *Ehrlichia* sp. HF in monocytes in the blood smear was confirmed by Diff-Quik staining (Thermo Fisher Scientific, Waltham, MA). The spleen was aseptically excised and a single-cell suspension was prepared in 0.7-ml of RPMI-1640 media after lysing red blood cells with ammonium chloride. DH82 cells were cultured in DMEM (Dulbecco minimal essential medium; Mediatech) supplemented with 5% FBS and 2 mM L-glutamine (L-Gln; GIBCO, Waltham, MA) at 37°C under 5% CO<sub>2</sub> in a humidified atmosphere as described previously [157]. RF/6A cells (ATCC) were cultured in advanced minimal essential medium (AMEM, Gibco) supplemented with 5% FBS and 2 mM L-glutamine. The ISE6 cell line, derived from the *Ixodes scapularis* tick embryo, was cultured in L15C300 medium at 34°C as

described previously [158]. Half of buffy coat cells and spleen cell suspension from one mouse were overlaid on DH82 and RF/6A cells in respective culture media, and cultured with the addition of 0.1 µg/mL cycloheximide (Millipore Sigma). To assess the degree of *Ehrlichia* infection in host cells, a drop of infected cells was centrifuged onto a slide in a Shandon Cytospin 4 cyto-centrifuge (Thermo Fisher), and the presence of *Ehrlichia*-containing inclusions was examined in both cell types by Diff-Quik staining every 3 – 4 days. *Ehrlichia* sp. HF was continuously passaged in DH82 cells with the addition of 0.1 µg/mL cycloheximide.

#### Culture and purification of host cell-free *Ehrlichia* sp. HF and bacterial genomic DNA

Twelve T175 flasks of *Ehrlichia* sp. HF-infected DH82 cells (>80% infectivity) at 3 d post infection (pi) were homogenized in 30 ml of 1× SPK buffer (0.2 M sucrose and 0.05 M potassium phosphate, pH 7.4) for 30 times with type A tight-fitting pestle in a dounce homogenizer (Wheaton, Millville, NJ). After centrifugation at 700 × *g* (Sorvall 6000D, Thermo Fisher), the pellet was further homogenized for additional 30 times. Homogenates were combined and step-wise centrifuged at 700, 1,000, and 1,500 × *g* for 10 min without using the break function of the centrifuge to avoid disturbing the loosely-packed pellets, then passed through 5.0- and 2.7-µm filters, and centrifuged at 10,000 × *g* for 10 min (Sorvall RC 5C Plus using SS-34 rotor). The purity of bacteria was determined by Diff-Quik staining (Fig. S6A). Genomic DNA samples were prepared using Qiagen genomic tips (Qiagen, Germantown, MD) according to the manufacturer's instructions, and resuspended in TE buffer. The quantity and quality of genomic DNA were determined by Nanodrop (8.41 µg total DNA; Thermo Fisher) as well as 0.9% agarose gel electrophoresis with BioLine markers (Fig. S6B). The purity of bacterial genomic DNA was confirmed by PCR and agarose gel electrophoresis using specific primers targeting *Ehrlichia* sp. HF 16S rRNA gene (HF51f/HF954r) and canine G3PDH DNA (Table S5 and Fig. S6) [156, 159]. The contamination of host DNA was estimated to be satisfactorily low for shotgun sequencing to obtain complete genome sequence (Fig. S6B).

#### Sequencing and annotation

Indexed Illumina mate pair libraries were prepared following the mate pair library v2 sample preparation guide (Illumina, San Diego, CA), with two modifications. First, the shearing was performed with the Covaris E210 (Covaris, Wobad, MA) using the following conditions: duty cycle, 10; time, 120 sec; intensity 4; and cycles per burst, 200. The DNA was purified between enzymatic reactions and the size selection of the library was performed with

AMPure XT beads (Beckman Coulter Genomics, Danvers, MA).

Paired-end genomic DNA libraries for sequencing using Illumina platform were constructed using the KAPA library preparation kit (Kapa Biosystems, Woburn, MA). DNA was fragmented with the Covaris E210 and the libraries were prepared using a modified version of manufacturer's protocol. The DNA was purified between enzymatic reactions and the size selection of the library was performed with AMPure XT beads (Beckman Coulter Genomics), using 33.3 µl beads for 50 µl purified ligation product. For indexed samples, the PCR amplification step was performed with primers containing a six-nucleotide index sequence.

Concentration and fragment size of libraries were determined using the DNA High Sensitivity Assay on the LabChip GX (Perkin Elmer, Waltham, MA) and qPCR using the KAPA Library Quantification Kit (Complete, Universal) (Kapa Biosystems, Woburn, MA). The mate pair libraries were sequenced on an Illumina HiSeq 2500 (Illumina), producing 23.8 M reads (4.8G bases), while the paired-end libraries were sequenced on an Illumina MiSeq (Illumina), producing 1.6 M reads (826.2M bases).

DNA samples for PacBio sequencing were sheared to 8 kbp using the Covaris gTube (Woburn, MA). Sequencing libraries were constructed and prepared for sequencing using the SMRTbell Express Template Prep Kit 2.0 (3kbp - 10kbp) and the DNA/Polymerase Binding Kit 2.0 (Pacific Biosciences, Menlo Park, CA). Libraries were loaded onto v2 SMRT Cells, and sequenced with the DNA Sequencing Kit 2.0 (Pacific Biosciences), producing 81,741 reads (388.8M bases). All sequence reads were deposited at NCBI Sequence Read Archive (SRA, BioProject accession number PRJNA187357).

Five assemblies were generated with various combinations of the data and assembly algorithms: (1) Celera Assembler v7.0 of only PacBio data, (2) Celera Assembler v7.0 of PacBio data with correction using Illumina paired-end data, (3) HGAP assembly of only PacBio data, (4) MaSuRCA 1.9.2 assembly of Illumina paired-end data subsampled to 50× coverage, and (5) MaSuRCA 1.9.2 assembly of Illumina paired-end data subsampled to 80× coverage. The first assembly was the optimal assembly, namely the one generated with Celera Assembler v7.0 with only the PacBio data. The data set was subsampled to ~22× coverage of the longest reads using an 8 Kbp minimum read length cutoff, with the remainder of the reads used for the error correction step. The resulting single-contig assembly totaled ~89.4 Kbp with 41.68% GC-content. The genome was trimmed to remove overlapping sequences, oriented, circularized, and rotated to the predicted origin of replication.

Annotation for this finalized genome assembly was generated using the IGS prokaryotic annotation pipeline [71] and deposited in GenBank (accession number NZ\_CP007474.1).

#### ANI, dDDH, and CGASI calculations

ANI values were calculated using OrthANIu v1.2 [153] paired with USEARCH v9.2.64 [160] run with default settings. dDDH values calculated using the web service GGDC v2.1 (ggdc.dsmz.de) [161] paired with the BLAST+ alignment tool [162] with default settings. For CGASI calculations, core genome alignments were constructed using Mugsy v1r2.2 [163] and mothur v1.40.4 [164]. CGASI values were calculated from the core genomes using the *PairwiseAlignments* function of the R package Biostrings. A maximum-likelihood phylogenetic tree was generated and plotted using IQTree v1.6.2 [165], with ModelFinder [166] and 100 ultra-fast bootstraps [167], and iTOL v5.5.1 [168], respectively.

#### Bioinformatic Analysis

For phylogenetic analysis, 16S rRNA genes from seven representative *Ehrlichia* spp., including *Ehrlichia* sp. HF, *E. chaffeensis* Arkansas, *E. muris* subsp. *muris* AS145, *E. muris* subsp. *eaucلائrens* Wisconsin, *E. canis* Jake, *E. ruminantium* Welgevonden, and *E. ruminantium* Gardel, were aligned individually using MegAlign program of DNASTar Lasergene 12 (Madison, WI) with MUSCLE algorithm. Alternatively, eight concatenated proteins [169], including 5 conserved housekeeping proteins (aspartate aminotransferase [TyrB], malate dehydrogenase [Mdh], adenylate kinase [Adk], fumarate hydratase [FumC], and 60 kDa chaperon [GroEL]), and 3 divergent outer membranes proteins (P28/VirB2-1/VirB6-1) from these *Ehrlichia* spp. were aligned individually using MegAlign program with CLUSTAL OMEGA. The evolutionary analyses were inferred by using the Maximum Likelihood method and Tamura-Nei model for 16S rRNA [170] or JTT matrix-based model for concatenated proteins [171], and bootstrap values for 1,000 replicates were obtained in MEGA X software [172]. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood approach, and then selecting the topology with superior log likelihood value. Phylogenetic trees were drawn to scale with branch lengths shown under each branch, and the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown above the branches.

The GC-skew was calculated as  $(C-G)/(C+G)$  in windows of 500 bp with step size of 250 bp along the chromosome. Whole genome alignments between *Ehrlichia* spp. were generated using Mugsy program with

default parameters [163], and the graphs were generated using GMAJ (<http://globin.bx.psu.edu/dist/gmaj/>).

To determine protein orthologs conserved among *Ehrlichia* spp., and *Ehrlichia* species-specific genes compared to other related organisms, orthologous clusters were determined by using the reciprocal Basic Local Alignment Search Tool (BLAST) algorithm BLASTP with an E-value of  $< 1e^{-10}$ .

Metabolic pathways and transporters were compared across genomes using (1) the Protein homologs generated with reciprocal BLASTP, (2) Genome Properties [87], (3) TransportDB [173], (4) Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.kegg.jp>), and (5) Biocyc [174]. Signal peptides and membrane proteins were predicted using the pSort-B algorithm (<http://psort.org/psortb/>) [175], and lipoproteins were predicted by LipoP 1.0 (<http://www.cbs.dtu.dk/services/LipoP>) [176].

#### NCBI GenBank Accession numbers and abbreviations of bacteria

*Ehrlichia* sp. HF (EHF), NZ\_CP007474.1; *E. chaffeensis* Arkansas (ECH), NC\_007799.1; *E. muris* subsp. *muris* AS145 (EMU), NC\_023063.1; *E. muris* subsp. *eaucلائrens* Wisconsin (EmCRT, three contigs), NZ\_LANU01000001, NZ\_LANU01000002, and NZ\_LANU01000003; *E. canis* Jake (ECA), NC\_007354.1; *E. ruminantium* Welgevonden (ERU), NC\_005295.2.

#### Notes

During the review process of the present paper, Wang et al. [152] reported (published online on August 3, 2020) that an Himar1 transposon insertion mutant in *TRP120* gene of *E. chaffeensis* was recovered in DH82 cells. This mutant had an initial lag phase but recovered afterwards in DH82 cell culture; however, it could not infect dogs when mixtures of *E. chaffeensis* transposon mutants were inoculated into dogs. Future availability of *TRP120* mutants of multiple *Ehrlichia* species will help comparative functional analysis of *TRP120*.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-020-07309-z>.

**Additional file 1: Table S1.** *Ehrlichia* proteins shared in two species by 4-way comparison analysis

**Additional file 2: Table S2.** *Ehrlichia* species-specific proteins by 4-way comparison analysis

**Additional file 3: Table S3.** *Ehrlichia* sp. HF-specific proteins by 2-way comparison analysis

**Additional file 4: Table S4.** Internal Repeats of *Ehrlichia* Tandem Repeat Proteins

**Additional file 5: Table S5.** Primers used in this study

**Additional file 6: Figure S1.** Culture Isolation of *Ehrlichia* sp. HF from infected mouse buffy coat and spleen. (A) Body weight of mice



inoculated with mouse spleen homogenates containing *Ehrlichia* sp. HF following days post inoculation. (B) *Ehrlichia* sp. HF (white arrows) in the blood monocytes from buffy coat smear by Diff-Quik staining. (C-D) Large *Ehrlichia*-containing inclusions (white arrows) in DH82 cells at ~ 3 weeks post infection (pi) or RF/6A cells at 2 weeks pi. (E) ISE6 cells were infected with purified host cell-free *Ehrlichia* sp. HF-infected DH82 cells and cultured in L15C300 media at 34°C. Infectivity reached 10% at 3 - 5 d pi with large morulae packed with *Ehrlichia*. Bar, 10 µm.

**Additional file 7: Figure S2.** Gene Structures of *Ehrlichia* sp. HF Type IV Secretion System. *Ehrlichia* sp. HF encodes a Type IV secretion system. These *virB/D* genes are split into three major operons: *virB2/4*, *virB3/4/6virB3/4/6*, *virB8/9/10/11/D4*, and three separate loci: *virB7* and duplicated *virB8-2* and *virB9-2*. *virB2* genes are duplicated into 5 copies, whereas *virB6* into 4 copies. Genes encoding *virB1* and *virB5* are not present in HF genome. Note: Due to the short protein length and low homology, *virB7* was not annotated as an ORF by NCBI automated annotation pipeline. However, by TBLASTN using *A. marginale* VirB7 protein sequence [121] against the entire HF genome sequence, a putative *virB7* gene was identified and manual curated.

**Additional file 8: Figure S3.** Phylogenetic analysis of *Ehrlichia* VirB2 paralogs of representative *Ehrlichia* species. Phylogenetic tree of VirB2 paralogs of representative *Ehrlichia* species, including *Ehrlichia* sp. HF (EHF), *E. chaffeensis* Arkansas (EchArk), *E. muris* subsp. *muris* AS145 (Emuris), and *E. muris* subsp. *eaucalarensis* Wisconsin (EmCRT). The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model, and the tree with the highest log likelihood is shown. The tree is drawn to scale with branch lengths measured in the number of substitutions per site (below branches), and the percentage of trees in which the associated taxa clustered together is shown above the branches. Evolutionary analyses were conducted in MEGA X.

**Additional file 9: Figure S4.** Domain structures and alignment of VirB2 paralogs of representative *Ehrlichia* species (A) Domain structures of *Ehrlichia* sp. HF VirB2-4. Analysis of *Ehrlichia* sp. HF VirB2-4 showed that it possesses a signal peptide (cleavage site between residues 29 and 30) and two putative transmembrane motifs. The signal peptide and transmembrane helices (TM) were predicted by SignalP-5.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) and TMHMM Server 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>), respectively. Hydrophobicity was analyzed by Protean program (DNASTar). (B) Alignment of VirB2 paralogs of representative *Ehrlichia* species showed that although these proteins are more divergent on the N- and C-terminus, they are highly conserved in the central transmembrane motifs or hydrophobic regions (indicated by red boxes). \*, conserved among all *Ehrlichia* VirB2 proteins.

**Additional file 10: Figure S5.** Phylogenetic trees of representative *Ehrlichia* species based on 16S rRNA sequences and concatenated protein sequences. (A) 16S rRNA genes from seven representative *Ehrlichia* spp. were aligned individually using MegAlign (1,514 nucleotides of aligned nucleotides). (B) Eight *Ehrlichia* proteins, including 5 conserved housekeeping proteins (TyrB, Mdh, Adk, FumC, and GroEL) and 3 divergent outer membranes proteins (P28, VirB2-1, and VirB6-1) from these *Ehrlichia* spp. were aligned using MegAlign. The aligned protein sequences were trimmed and concatenated (3,188 AA total). The evolutionary analyses were inferred by using the Maximum Likelihood method and Tamura-Nei model for 16S rRNA, or JTT matrix-based model for concatenated proteins. Bootstrap values for 1,000 replicates were obtained using MEGA X. The trees with the highest log likelihood (-2609.17 for 16S rRNA, and -17382.50 for proteins) were shown, and the percentage of trees in which the associated taxa clustered together in the bootstrap test was shown above each branch. The tree is drawn to scale with branch lengths measured in the number of average nucleotide substitutions per site (shown under each branch). GenBank Accession numbers for seven representative *Ehrlichia* spp.: *Ehrlichia* sp. HF, NZ\_CP007474.1; *E. chaffeensis* Arkansas, NC\_007799.1; *E. muris* subsp. *muris* AS145, NC\_023063.1; *E. muris* subsp. *eaucalarensis* Wisconsin, LANU01000000; *E. canis* Jake, NC\_007354.1; *E. ruminantium* Welgevonden, NC\_005295.2; *E. ruminantium* Gardel, NC\_006831.1.

**Additional file 11: Figure S6.** Purification of host cell-free *Ehrlichia* sp. HF and bacterial genomic DNA. (A) Twelve T175 flasks of *Ehrlichia* sp. HF-

infected DH82 cells (>80% infectivity) at 3d pi were homogenized in 30 ml of 1× SPK for 30 times with type B tight-fitting pestle. Pellet following centrifugation at 700 × g was homogenized for additional 30 times. Both homogenates were step-wise centrifuged at 700, 1,000, and 1,500 × g, passed through 5.0- and 2.7-µm filters, and centrifuged at 10,000 × g for 10 min. Host cell-free *Ehrlichia* sp. HF was purified with very low host nuclear contamination under Diff-Quik staining. Bar, 10 µm. (B) Genome DNAs of *Ehrlichia* sp. HF (EHF) were purified using Qiagen genomic tips and dissolved in TE buffer. DNAs were resolved using 0.9% agarose with BioLine molecular weight (MW) markers with DNA concentrations of each band showing inside parenthesis. Genomic DNA bands above 20 kB were visible, and the concentration was above 15 ng/µl. PCR reactions were carried out with 35 cycles at 98°C for 30-second, 60°C for 30-second, and 68°C for 1-minute. Primers targeting 16S rRNA gene of *Ehrlichia* sp. HF detected specific bands at 1/100 dilutions, but dog G3PDH primers did not amplify any bands under any dilutions (Dilutions: .1, 1/10, .01, 1/100 dilution; Pos.: positive control using DNA isolated from dog DH82 cells; -, negative control without DNA input).

### Abbreviations

AA: Amino acid; ANI: Average nucleotide identity; BLAST: Basic local alignment search tool; bp: Base pair; CGAS: Core genome alignment sequence identity; dDDH: Digital DNA-DNA hybridization; Etf: *Ehrlichia* translocated factor; HME: Human monocytic ehrlichiosis; ORF: Open reading frame; pi: Post infection; Omp: Outer membrane protein; T1SS: Type I secretion system; T4SS: Type IV secretion system; TCS: Two-component regulatory system; TRP: Tandem-repeat containing protein

### Acknowledgements

We thank Drs. Makoto Kawahara and Haibin Huang for providing and maintaining C57BL/6 mice infected with *Ehrlichia* sp. HF.

### Authors' contributions

ML, YR, and JCDH designed research; ML and QX performed research; SD, SN, NS, SQ, and AG contributed to genome sequencing and annotation; LT, LS, CMF, and JCDH oversaw sequencing project management; ML, JCDH, MC, and YR analyzed data; and ML and YR wrote the paper. All authors have read and approved the manuscript for publication.

### Funding

This project has been funded by the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Department of Health and Human Services under contract number HHSN272200900009C, and NIH R01 AI 047885. The funding agency played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Availability of data and materials

The datasets supporting the results of this article are included within the article and supplementary information. The genome sequence has been deposited in GenBank with Accession number NZ\_CP007474, *Ehrlichia japonica* strain HF is available from BEI Resources (Deposit ID# NR-46450) and CSUR (Q1926).

### Ethics approval and consent to participate

All animal experiments were performed in accordance with the Ohio State University Institutional Animal Care and Use Committee guidelines and approved protocol. The university program has full continued accreditation by the Association for Assessment and Accreditation of Laboratory Animal Care International under 000028, dated 9 June 2000, and has Public Health Services assurance renewal A3261-01, dated 6 February 2019 through 28 February 2023. The program is licensed by the USDA, number 31-R-014, and is in full compliance with Animal Welfare Regulations.

### Consent for publication

All authors have read and approved the manuscript for publication.

### Competing interests

No potential conflicts of interest were disclosed.

**Author details**

<sup>1</sup>Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Road, Columbus, OH 43210, USA. <sup>2</sup>Institute for Genome Sciences, University of Maryland School of Medicine, 801 W. Baltimore St, Baltimore, MD 21201, USA. <sup>3</sup>Department of Medicine, University of Maryland School of Medicine, 801 W. Baltimore St, Baltimore, MD 21201, USA. <sup>4</sup>Department of Microbiology and Immunology, University of Maryland School of Medicine, 801 W. Baltimore St, Baltimore, MD 21201, USA. <sup>5</sup>Greenebaum Cancer Center, University of Maryland School of Medicine, 801 W. Baltimore St, Baltimore, MD 21201, USA.

Received: 9 June 2020 Accepted: 7 December 2020

Published online: 06 January 2021

**References**

- Paddock CD, Yabsley MJ. Ecological havoc, the rise of white-tailed deer, and the emergence of *Amblyomma americanum*-associated zoonoses in the United States. *Curr Topics Microbiol Immunol*. 2007;315:289–324.
- Centers for Disease Control and Prevention: Weekly cases of notifiable diseases, United States, U.S. territories, and Non-U.S. Residents weeks ending December 21, 2019 (week 51). In: *MMWR Morb Mortal Wkly Rep*. 2019/12/21 edn; 2019.
- Rosenberg R, Lindsey NP, Fischer M, Gregory CJ, Hinckley AF, Mead PS, Paz-Bailey G, Waterman SH, Drexler NA, Kersh GJ, et al. Vital Signs: Trends in Reported Vectorborne Disease Cases - United States and Territories, 2004–2016. *MMWR Morb Mortal Wkly Rep*. 2018;67(17):496–501.
- Institute of Medicine. The Short-Term and Long-Term Outcomes: Workshop Report. In: *Critical Needs and Gaps in Understanding Prevention, Amelioration, and Resolution of Lyme and Other Tick-Borne Diseases*. Washington; National Academies Press (US); 2011.
- Fujita H, Watanabe Y. Ehrlichial organisms isolated from *Ixodes ovatus* ticks and field rodents in Japan. *Ann Rep Ohara Hosp (In Japanese)*. 1994;37:13–7.
- Shibata S, Kawahara M, Rikihisa Y, Fujita H, Watanabe Y, Suto C, Ito T. New Ehrlichia species closely related to *Ehrlichia chaffeensis* isolated from *Ixodes ovatus* ticks in Japan. *J Clin Microbiol*. 2000;38(4):1331–8.
- Matsumoto K, Joncour G, Lamanda P, Inokuma H, Brouqui P. Detection of *Anaplasma phagocytophilum* and *Ehrlichia* sp. HF strains in *Ixodes ricinus* ticks in Brittany, France. *Clin Microbiol Infect*. 2007;13(3):338–41.
- Li K, Stanojevic M, Stamenkovic G, Ilic B, Paunovic M, Lu M, Pesic B, Duric Maslovara I, Siljic M, Cirkovic V, et al. Insight into diversity of bacteria belonging to the order Rickettsiales in 9 arthropods species collected in Serbia. *Sci Rep*. 2019;9(1):18680.
- Andersson MO, Radbea G, Frangoulidis D, Tomaso H, Rubel F, Nava S, Chitimia-Dobler L. New records and host associations of the tick *Ixodes apronophorus* and the first detection of Ehrlichia sp. HF in Romania. *Parasitol Res*. 2018;117(4):1285–9.
- Okada H, Tajima T, Kawahara M, Rikihisa Y. Ehrlichial proliferation and acute hepatocellular necrosis in immunocompetent mice experimentally infected with the HF strain of *Ehrlichia*, closely related to *Ehrlichia chaffeensis*. *J Comp Pathol*. 2001;124(2–3):165–71.
- Dawson JE, Anderson BE, Fishbein DB, Sanchez JL, Goldsmith CS, Wilson KH, Duntley CW. Isolation and characterization of an *Ehrlichia* sp. from a patient diagnosed with human ehrlichiosis. *J Clin Microbiol*. 1991;29(12):2741–5.
- Maeda K, Markowitz N, Hawley RC, Ristic M, Cox D, McDade JE. Human infection with *Ehrlichia canis*, a leukocytic Rickettsia. *New Engl J Med*. 1987;316(14):853–6.
- Anderson BE, Dawson JE, Jones DC, Wilson KH. *Ehrlichia chaffeensis*, a new species associated with human ehrlichiosis. *J Clin Microbiol*. 1991;29(12):2838–42.
- Kawahara M, Tajima T, Torii H, Yabutani M, Ishii J, Harasawa M, Isogai E, Rikihisa Y. *Ehrlichia chaffeensis* infection of sika deer, Japan. *Emerg Infect Dis*. 2009;15(12):1991–3.
- Davidson WR, Lockhart JM, Stallknecht DE, Howerth EW. Susceptibility of red and gray foxes to infection by *Ehrlichia chaffeensis*. *J wildlife Dis*. 1999;35(4):696–702.
- Kocan AA, Levesque GC, Whitworth LC, Murphy GL, Ewing SA, Barker RW. Naturally occurring *Ehrlichia chaffeensis* infection in coyotes from Oklahoma. *Emerg Infect Dis*. 2000;6(5):477–80.
- Wen B, Rikihisa Y, Mott J, Fuerst PA, Kawahara M, Suto C. *Ehrlichia muris* sp. nov., identified on the basis of 16S rRNA base sequences and serological, morphological, and biological characteristics. *Int J Syst Bacteriol*. 1995;45(2):250–4.
- Rar VA, Livanova NN, Panov WV, Doroschenko EK, Pukhovskaya NM, Vysochina NP, Ivanov LI. Genetic diversity of *Anaplasma* and *Ehrlichia* in the Asian part of Russia. *Ticks Tick Borne Dis*. 2010; 1(1):57–65.
- Pritt BS, Allerdice MEJ, Sloan LM, Paddock CD, Munderloh UG, Rikihisa Y, Tajima T, Paskewitz SM, Neitzel DF, Hoang Johnson DK, et al. Proposal to reclassify *Ehrlichia muris* as *Ehrlichia muris* subsp. *muris* subsp. nov. and description of *Ehrlichia muris* subsp. *eaucclairiensis* subsp. nov., a newly recognized tick-borne pathogen of humans. *Int J Syst Evol Microbiol*. 2017;67(7):2121–6.
- Pritt BS, Sloan LM, Johnson DK, Munderloh UG, Paskewitz SM, McElroy KM, McFadden JD, Binnicker MJ, Neitzel DF, Liu G, et al. Emergence of a new pathogenic *Ehrlichia* species, Wisconsin and Minnesota, 2009. *New Engl J Med*. 2011;365(5):422–9.
- Perez M, Bodor M, Zhang C, Xiong Q, Rikihisa Y. Human infection with *Ehrlichia canis* accompanied by clinical signs in Venezuela. *Ann N Y Acad Sci*. 2006;1078:110–7.
- Bouza-Mora L, Dolz G, Solorzano-Morales A, Romero-Zuniga JJ, Salazar-Sanchez L, Labruna MB, Aguiar DM. Novel genotype of *Ehrlichia canis* detected in samples of human blood bank donors in Costa Rica. *Ticks Tick Borne Dis*. 2017;8(1):36–40.
- Unver A, Perez M, Orellana N, Huang H, Rikihisa Y. Molecular and antigenic comparison of *Ehrlichia canis* isolates from dogs, ticks, and a human in Venezuela. *J Clin Microbiol*. 2001;39(8):2788–93.
- Donatien A, Lestoquard F. Existence en Algérie d'une Rickettsia du chien. *Bull Soc Pathol Exot*. 1935;28:418–9.
- Ristic M, Kreier JP. Genus I Anaplasma Theiler 1970, 7AL. In: *Bergey's Manual of Systematic Bacteriology*. Edited by Krieg NR, Holt JG, vol. 1, 1st ed. Baltimore: Williams and Wilkins; 1984: 720–722.
- Cowdry EV. Studies on the Etiology of Heartwater 1. Observation of a *Rickettsia*, *Rickettsia Ruminantium* (N. Sp.), in the Tissues of Infected Animals. *J Exp Med*. 1925;42(2):231–52.
- Cowdry EV. Studies on the Etiology of Heartwater 2. *Rickettsia Ruminantium* (N. Sp.) in the Tissues of Ticks Transmitting the Disease. *J Exp Med*. 1925;42(2):253–74.
- Du Plessis JL. A method for determining the *Cowdria ruminantium* infection rate of *Amblyomma hebraeum*: effects in mice injected with tick homogenates. *Onderstepoort J Vet Res*. 1985;52(2):55–61.
- Uilenberg G, Camus E, Barre N. A strain of *Cowdria ruminantium* isolated in Guadeloupe (French West Indies). *Rev Elev Med Vet Pays Trop*. 1985;38(1):34–42.
- Muramatsu Y, Ukegawa SY, Rahim A, El Hussein M, Abdel Rahman MB, Abdel Gabbar KM, Chitambo AM, Komiya T, Mwase ET, Morita C, et al. Ehrlichia ruminantium, Sudan. *Emerg Infect Dis*. 2005;11:1792–3.
- Kasari TR, Miller RS, James AM, Freier JE. Recognition of the threat of *Ehrlichia ruminantium* infection in domestic and wild ruminants in the continental United States. *J Am Vet Med Assoc*. 2010;237(5):520–30.
- Zhang J, Kelly P, Guo W, Xu C, Wei L, Jongejan F, Loftis A, Wang C. Development of a generic *Ehrlichia* FRET-qPCR and investigation of ehrlichioses in domestic ruminants on five Caribbean islands. *Parasit Vectors*. 2015;8:506.
- Kelly PJ, Lucas H, Yowell C, Beati L, Dame J, Urdaz-Rodriguez J, Mahan S. *Ehrlichia ruminantium* in *Amblyomma variegatum* and domestic ruminants in the Caribbean. *J Med Entomol*. 2011;48(2):485–8.
- Buller RS, Arens M, Hmiel SP, Paddock CD, Sumner JW, Rikihisa Y, Unver A, Gaudreault-Keener M, Manian FA, Liddell AM, et al. *Ehrlichia ewingii*, a newly recognized agent of human ehrlichiosis. *N Engl J Med*. 1999;341(3):148–55.
- Ewing SA, Roberson WR, Buckner RG, Hayat CS. A new strain of *Ehrlichia canis*. *J Am Vet Med Assoc*. 1971;159(12):1771–4.
- Anderson BE, Greene CE, Jones DC, Dawson JE. *Ehrlichia ewingii* sp. nov., the etiologic agent of canine granulocytic ehrlichiosis. *Int J Syst Bacteriol*. 1992;42(2):299–302.
- Cabezas-Cruz A, Zwegarth E, Vancova M, Broniszewska M, Grubhoffer L, Passos LMF, Ribeiro MFB, Alberdi P, de la Fuente J. Ehrlichia minasensis sp. nov., isolated from the tick *Rhipicephalus microplus*. *Int J Syst Evol Microbiol*. 2016;66(3):1426–30.
- Cabezas-Cruz A, Zwegarth E, Ribeiro MF, da Silveira JA, de la Fuente J, Grubhoffer L, Valdes JJ, Passos LM. New species of *Ehrlichia* isolated from *Rhipicephalus (Boophilus) microplus* shows an ortholog of the *E. canis* major

- immunogenic glycoprotein gp36 with a new sequence of tandem repeats. *Parasit Vectors*. 2012;5:291.
39. Cabezas-Cruz A, Zwegarth E, Broniszweska M, Passos LM, Ribeiro MF, Manrique M, Tobes R, de la Fuente J. Complete Genome Sequence of *Ehrlichia mineirensis*, a Novel Organism Closely Related to *Ehrlichia canis* with a New Host Association. *Genome Announc*. 2015;3(1):e01450–14.
  40. Aguiar DM, Ziliani TF, Zhang X, Melo AL, Braga IA, Witter R, Freitas LC, Rondelli AL, Luis MA, Sorte EC, et al. A novel *Ehrlichia* genotype strain distinguished by the TRP36 gene naturally infects cattle in Brazil and causes clinical manifestations associated with ehrlichiosis. *Ticks Tick Borne Dis*. 2014;5(5):537–44.
  41. Aguiar DM, Araujo JP Jr, Nakazato L, Bard E, Cabezas-Cruz A. Complete Genome Sequence of an *Ehrlichia minasensis* Strain Isolated from Cattle. *Microbiol Resour Announc*. 2019;8(15):e00161–19.
  42. Cicculi V, Masse S, Capai L, de Lamballerie X, Charrel R, Falchi A. First detection of *Ehrlichia minasensis* in *Hyalomma marginatum* ticks collected from cattle in Corsica, France. *Vet Med Sci*. 2019;5(2):243–8.
  43. Rehman A, Conraths FJ, Sauter-Louis C, Krucken J, Nijhof AM. Epidemiology of tick-borne pathogens in the semi-arid and the arid agro-ecological zones of Punjab province, Pakistan. *Transbound Emerg Dis*. 2019;66(1):526–36.
  44. Hailemariam Z, Krucken J, Baumann M, Ahmed JS, Clausen PH, Nijhof AM. Molecular detection of tick-borne pathogens in cattle from Southwestern Ethiopia. *PLoS One*. 2017;12(11):e0188248.
  45. Thomson K, Yaaran T, Belshaw A, Curson L, Tisi L, Maurice S, Kiddle G. A new TaqMan method for the reliable diagnosis of *Ehrlichia* spp. in canine whole blood. *Parasit Vectors*. 2018;11(1):350.
  46. Parte AC, Sarda Carbasse J, Meier-Kolthoff JP, Reimer LC, Goker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol*. 2020.
  47. Dumler JS, Barbet AF, Bekker CP, Dasch GA, Palmer GH, Ray SC, Rikihisa Y, Rurangirwa FR. Reorganization of genera in the families *Rickettsiaceae* and *Anaplasmataceae* in the order *Rickettsiales*: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. *Int J Syst Evol Microbiol*. 2001;51(Pt 6):2145–65.
  48. Paddock CD, Childs JE. *Ehrlichia chaffeensis*: a prototypical emerging pathogen. *Clin Microbiol Rev*. 2003;16(1):37–64.
  49. Perez M, Rikihisa Y, Wen B. *Ehrlichia canis*-like agent isolated from a man in Venezuela: antigenic and genetic characterization. *J Clin Microbiol*. 1996;34(9):2133–9.
  50. Martinez MC, Gutierrez CN, Monger F, Ruiz J, Watts A, Mijares VM, Rojas MG, Triana-Alonso FJ. *Ehrlichia chaffeensis* in child, Venezuela. *Emerg Infect Dis*. 2008;14(3):519–20.
  51. Paules CI, Marston HD, Bloom ME, Fauci AS. Tickborne Diseases - Confronting a Growing Threat. *New Engl J Med*. 2018;379(8):701–3.
  52. Ohashi N, Unver A, Zhi N, Rikihisa Y. Cloning and characterization of multigenes encoding the immunodominant 30-kilodalton major outer membrane proteins of *Ehrlichia canis* and application of the recombinant protein for serodiagnosis. *J Clin Microbiol*. 1998;36(9):2671–80.
  53. Davidson WR, Lockhart JM, Stallnecht DE, Howerth EW, Dawson JE, Rechav Y. Persistent *Ehrlichia chaffeensis* infection in white-tailed deer. *J Wildlife Dis*. 2001;37(3):538–46.
  54. Dawson JE, Ewing SA. Susceptibility of dogs to infection with *Ehrlichia chaffeensis*, causative agent of human ehrlichiosis. *Am J Vet Res*. 1992;53(8):1322–7.
  55. Unver A, Rikihisa Y, Stich RW, Ohashi N, Felek S. The omp-1 major outer membrane multigene family of *Ehrlichia chaffeensis* is differentially expressed in canine and tick hosts. *Infect Immun*. 2002;70(8):4701–4.
  56. Bekebrede H, Lin M, Teymournejad O, Rikihisa Y. Discovery of in vivo Virulence Genes of Obligatory Intracellular Bacteria by Random Mutagenesis. *Front Cell Infect Microbiol* 2020, 10(2):2.
  57. Winslow GM, Bitsaktis C, Yager E. Susceptibility and resistance to monocytic ehrlichiosis in the mouse. *Ann N Y Acad Sci*. 2005;1063:395–402.
  58. Bitsaktis C, Huntington J, Winslow G. Production of IFN-gamma by CD4 T cells is essential for resolving ehrlichia infection. *J Immunol*. 2004;172(11):6894–901.
  59. Bitsaktis C, Winslow G. Fatal recall responses mediated by CD8 T cells during intracellular bacterial challenge infection. *J Immunol*. 2006;177(7):4644–51.
  60. Bitsaktis C, Nandi B, Racine R, MacNamara KC, Winslow G. T-Cell-independent humoral immunity is sufficient for protection against fatal intracellular ehrlichia infection. *Infect Immun*. 2007;75(10):4933–41.
  61. Ismail N, Soong L, McBride JW, Valbuena G, Olano JP, Feng HM, Walker DH. Overproduction of TNF-alpha by CD8+ type 1 cells and down-regulation of IFN-gamma production by CD4+ Th1 cells contribute to toxic shock-like syndrome in an animal model of fatal monocytotropic ehrlichiosis. *J Immunol*. 2004;172(3):1786–800.
  62. Stevenson HL, Estes MD, Thirumalapura NR, Walker DH, Ismail N. Natural killer cells promote tissue injury and systemic inflammatory responses during fatal *Ehrlichia*-induced toxic shock-like syndrome. *Am J Pathol*. 2010;177(2):766–76.
  63. Ismail N, Stevenson HL, Walker DH. Role of tumor necrosis factor alpha (TNF-alpha) and interleukin-10 in the pathogenesis of severe murine monocytotropic ehrlichiosis: increased resistance of TNF receptor p55- and p75-deficient mice to fatal ehrlichial infection. *Infect Immun*. 2006;74(3):1846–56.
  64. Ismail N, Walker DH. Balancing protective immunity and immunopathology: a unifying model of monocytotropic ehrlichiosis. *Ann N Y Acad Sci*. 2005;1063:383–94.
  65. Thomas RJ, Dumler JS, Carlyon JA. Current management of human granulocytic anaplasmosis, human monocytic ehrlichiosis and *Ehrlichia ewingii* ehrlichiosis. *Expert Rev Anti Infect Ther*. 2009;7(6):709–22.
  66. Stevenson HL, Crossley EC, Thirumalapura N, Walker DH, Ismail N. Regulatory roles of CD1d-restricted NKT cells in the induction of toxic shock-like syndrome in an animal model of fatal ehrlichiosis. *Infect Immun*. 2008;76(4):1434–44.
  67. Yang Q, Ghose P, Ismail N. Neutrophils mediate immunopathology and negatively regulate protective immune responses during fatal bacterial infection-induced toxic shock. *Infect Immun*. 2013;81(5):1751–63.
  68. Kader M, Alaoui-El-Azher M, Vorhauer J, Kode BB, Wells JZ, Stolz D, Michalopoulos G, Wells A, Scott M, Ismail N. MyD88-dependent inflammasome activation and autophagy inhibition contributes to *Ehrlichia*-induced liver injury and toxic shock. *PLoS Pathog*. 2017;13(10):e1006644.
  69. Munderloh UG, Silverman DJ, MacNamara KC, Ahlstrand GG, Chatterjee M, Winslow GM. *Ixodes ovatus Ehrlichia* exhibits unique ultrastructural characteristics in mammalian endothelial and tick-derived cells. *Ann N Y Acad Sci*. 2009;1166:112–9.
  70. Ioannidis P, Dunning Hotopp JC, Sapountzis P, Siozios S, Tsiamis G, Bordenstein SR, Baldo L, Werren JH, Bourtzis K. New criteria for selecting the origin of DNA replication in *Wolbachia* and closely related bacteria. *BMC Genomics*. 2007;8:182.
  71. Galens K, Orvis J, Daugherty S, Creasy HH, Angiuoli S, White O, Wortman J, Mahurkar A, Giglio MG. The IGS Standard Operating Procedure for Automated Prokaryotic Annotation. *Standards Genomic Sci*. 2011;4(2):244–51.
  72. Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, Eisen JA, Seshadri R, Ren Q, Wu M, Utterback TR, et al. Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet*. 2006;2(2):e21.
  73. Thirumalapura NR, Qin X, Kuriakose JA, Walker DH. Complete Genome Sequence of *Ehrlichia muris* Strain AS145<sup>T</sup>, a Model Monocytotropic Ehrlichia Strain. *Genome Announc*. 2014;2(1):e01234–13.
  74. Massung RF, Lee K, Mauel M, Gusa A. Characterization of the rRNA genes of *Ehrlichia chaffeensis* and *Anaplasma phagocytophila*. *DNA Cell Biol*. 2002;21(8):587–96.
  75. Frutos R, Viari A, Vachieri N, Boyer F, Martinez D. *Ehrlichia ruminantium*: genomic and evolutionary features. *Trends Parasitol*. 2007;23(9):414–9.
  76. Nene V, Kole C. Genome mapping and genomics in animal-associated microbes. Berlin: Springer; 2009.
  77. Darling AE, Miklos I, Ragan MA. Dynamics of genome rearrangement in bacterial populations. *PLoS Genet*. 2008;4(7):e1000128.
  78. Paddock CD, Sumner JW, Shore GM, Bartley DC, Elie RC, McQuade JG, Martin CR, Goldsmith CS, Childs JE. Isolation and characterization of *Ehrlichia chaffeensis* strains from patients with fatal ehrlichiosis. *J Clin Microbiol*. 1997;35(10):2496–502.
  79. Ohashi N, Zhi N, Zhang Y, Rikihisa Y. Immunodominant major outer membrane proteins of *Ehrlichia chaffeensis* are encoded by a polymorphic multigene family. *Infect Immun*. 1998;66(1):132–9.
  80. Chapes SK, Ganta RR. Defining the immune response to *Ehrlichia* species using murine models. *Vet Parasitol*. 2008;158(4):344–59.

81. Saito TB, Walker DH. A Tick Vector Transmission Model of Monocytotropic Ehrlichiosis. *J Infect Dis.* 2015;212(6):968–77.
82. Hirschberg HJ, Simons JW, Dekker N, Egmond MR. Cloning, expression, purification and characterization of patatin, a novel phospholipase A. *Eur J Biochem.* 2001;268(19):5037–44.
83. Mordue DG, Scott-Weathers CF, Tobin CM, Knoll LJ. A patatin-like protein protects *Toxoplasma gondii* from degradation in activated macrophages. *Mol Microbiol.* 2007;63(2):482–96.
84. Tobin CM, Knoll LJ. A patatin-like protein protects *Toxoplasma gondii* from degradation in a nitric oxide-dependent manner. *Infect Immun.* 2012;80(1):55–61.
85. Cheng Z, Kumagai Y, Lin M, Zhang C, Rikihisa Y. Intra-leukocyte expression of two-component systems in *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* and effects of the histidine kinase inhibitor closantel. *Cell Microbiol.* 2006;8(8):1241–52.
86. Kumagai Y, Cheng Z, Lin M, Rikihisa Y. Biochemical activities of three pairs of *Ehrlichia chaffeensis* two-component regulatory system proteins involved in inhibition of lysosomal fusion. *Infect Immun.* 2006;74(9):5014–22.
87. Haft DH, Selengut JD, Brinkac LM, Zafar N, White O. Genome Properties: a system for the investigation of prokaryotic genetic content for microbiology, genome annotation and comparative genomics. *Bioinformatics.* 2005;21(3):293–306.
88. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 1999;27(1):29–34.
89. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform.* 2019;20(4):1085–93.
90. Lin M, Liu H, Xiong Q, Niu H, Cheng Z, Yamamoto A, Rikihisa Y. *Ehrlichia* secretes Etf-1 to induce autophagy and capture nutrients for its growth through RAB5 and class III phosphatidylinositol 3-kinase. *Autophagy.* 2016;12(11):2145–66.
91. Lin M, Grandinetti G, Hartnell LM, Bliss D, Subramaniam S, Rikihisa Y. Host membrane lipids are trafficked to membranes of intravacuolar bacterium *Ehrlichia chaffeensis*. *Proc Natl Acad Sci U S A.* 2020;117(14):8032–43.
92. Koh YS, Koo JE, Biswas A, Kobayashi KS. MyD88-dependent signaling contributes to host defense against ehrlichial infection. *PLoS One.* 2010;5(7):e11758.
93. Rikihisa Y. Molecular Pathogenesis of *Ehrlichia chaffeensis* Infection. *Annu Rev Microbiol.* 2015;69:283–304.
94. Miura K, Matsuo J, Rahman MA, Kumagai Y, Li X, Rikihisa Y. *Ehrlichia chaffeensis* induces monocyte inflammatory responses through MyD88, ERK, and NF-kappaB but not through TRIF, interleukin-1 receptor 1 (IL-1R1)/IL-18R1, or toll-like receptors. *Infect Immun.* 2011;79(12):4947–56.
95. Dorman CJ, McKenna S, Beloin C. Regulation of virulence gene expression in *Shigella flexneri*, a facultative intracellular pathogen. *Int J Med Microbiol.* 2001;291(2):89–96.
96. Cheng C, Paddock CD, Reddy Ganta R. Molecular heterogeneity of *Ehrlichia chaffeensis* isolates determined by sequence analysis of the 28-kilodalton outer membrane protein genes and other regions of the genome. *Infect Immun.* 2003;71(1):187–95.
97. Ohashi N, Rikihisa Y, Unver A. Analysis of transcriptionally active gene clusters of major outer membrane protein multigene family in *Ehrlichia canis* and *E. chaffeensis*. *Infect Immun.* 2001;69(4):2083–91.
98. Singu V, Liu H, Cheng C, Ganta RR. *Ehrlichia chaffeensis* expresses macrophage- and tick cell-specific 28-kilodalton outer membrane proteins. *Infect Immun.* 2005;73(1):79–87.
99. Kumagai Y, Huang H, Rikihisa Y. Expression and porin activity of P28 and OMP-1F during intracellular *Ehrlichia chaffeensis* development. *J Bacteriol.* 2008;190(10):3597–605.
100. Gentle IE, Burri L, Lithgow T. Molecular architecture and function of the Omp85 family of proteins. *Mol Microbiol.* 2005;58(5):1216–25.
101. Voulhoux R, Bos MP, Geurtsen J, Mols M, Tommassen J. Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science.* 2003;299(5604):262–5.
102. Hirvas L, Coleman J, Koski P, Vaara M. Bacterial 'histone-like protein I' (HLP-I) is an outer membrane constituent? *FEBS Lett.* 1990;262(1):123–6.
103. Walton TA, Sousa MC. Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation. *Mol Cell.* 2004;15(3):367–74.
104. Huang H, Lin M, Wang X, Kikuchi T, Mottaz H, Norbeck A, Rikihisa Y. Proteomic analysis of and immune responses to *Ehrlichia chaffeensis* lipoproteins. *Infect Immun.* 2008;76(8):3405–14.
105. Lin M, Kikuchi T, Brewer HM, Norbeck AD, Rikihisa Y. Global proteomic analysis of two tick-borne emerging zoonotic agents: *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. *Front Microbiol.* 2011;2:24.
106. Cheng Z, Miura K, Popov VL, Kumagai Y, Rikihisa Y. Insights into the CtrA regulon in development of stress resistance in obligatory intracellular pathogen *Ehrlichia chaffeensis*. *Mol Microbiol.* 2011;82(5):1217–34.
107. Confer AW, Ayalew S. The OmpA family of proteins: roles in bacterial pathogenesis and immunity. *Vet Microbiol.* 2013;163(3-4):207–22.
108. Mohan Kumar D, Lin M, Xiong Q, Webber MJ, Kural C, Rikihisa Y. EtpE Binding to DNase X Induces Ehrlichial Entry via CD147 and hnRNP-K Recruitment, Followed by Mobilization of N-WASP and Actin. *mBio.* 2015;6(6):e01541–15.
109. Mohan Kumar D, Yamaguchi M, Miura K, Lin M, Los M, Coy JF, Rikihisa Y. *Ehrlichia chaffeensis* uses its surface protein EtpE to bind GPI-anchored protein DNase X and trigger entry into mammalian cells. *PLoS Pathog.* 2013;9(10):e1003666.
110. Teymournejad O, Lin M, Rikihisa Y. *Ehrlichia chaffeensis* and Its Invasin EtpE Block Reactive Oxygen Species Generation by Macrophages in a DNase X-Dependent Manner. *mBio.* 2017;8(6):e01551–17.
111. Teymournejad O, Rikihisa Y. *Ehrlichia chaffeensis* Uses an Invasin To Suppress Reactive Oxygen Species Generation by Macrophages via CD147-Dependent Inhibition of Vav1 To Block Rac1 Activation. *mBio.* 2020;11(2):e00267–20.
112. Papanikou E, Karamanou S, Economou A. Bacterial protein secretion through the translocase nanomachine. *Nat Rev Microbiol.* 2007;5(11):839–51.
113. Lee PA, Tullman-Ercek D, Georgiou G. The bacterial twin-arginine translocation pathway. *Annu Rev Microbiol.* 2006;60:373–95.
114. Cascales E, Christie PJ. The versatile bacterial type IV secretion systems. *Nat Rev Microbiol.* 2003;1(2):137–49.
115. Gillespie JJ, Kaur SJ, Rahman MS, Rennoll-Bankert K, Sears KT, Beier-Sexton M, Azad AF. Secretome of obligate intracellular *Rickettsia*. *FEMS Microbiol Rev.* 2015;39(1):47–80.
116. Alvarez-Martinez CE, Christie PJ. Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev.* 2009;73(4):775–808.
117. Fullner KJ, Lara JC, Nester EW. Pilus assembly by *Agrobacterium* T-DNA transfer genes. *Science.* 1996;273(5278):1107–9.
118. Lai EM, Kado CI. Processed VirB2 is the major subunit of the promiscuous pilus of *Agrobacterium tumefaciens*. *J Bacteriol.* 1998;180(10):2711–7.
119. Hwang HH, Gelvin SB. Plant proteins that interact with VirB2, the *Agrobacterium tumefaciens* pilin protein, mediate plant transformation. *Plant Cell.* 2004;16(11):3148–67.
120. Backert S, Meyer TF. Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol.* 2006;9(2):207–17.
121. Gillespie JJ, Brayton KA, Williams KP, Diaz MA, Brown WC, Azad AF, Sobral BW. Phylogenomics reveals a diverse Rickettsiales type IV secretion system. *Infect Immun.* 2010;78(5):1809–23.
122. Kalkum M, Eisenbrandt R, Lanka E. Protein circllets as sex pilus subunits. *Curr Protein Pept Sci.* 2004;5(5):417–24.
123. Silverman PM. Towards a structural biology of bacterial conjugation. *Mol Microbiol.* 1997;23(3):423–9.
124. Lin M, Zhang C, Gibson K, Rikihisa Y. Analysis of complete genome sequence of *Neorickettsia risticii*: causative agent of Potomac horse fever. *Nucleic Acids Res.* 2009;37(18):6076–91.
125. Nelson CM, Herron MJ, Felsheim RF, Schloeder BR, Grindle SM, Chavez AO, Kurtti TJ, Munderloh UG. Whole genome transcription profiling of *Anaplasma phagocytophilum* in human and tick host cells by tiling array analysis. *BMC Genomics.* 2008;9:364.
126. Suttan EL, Norimine J, Beare PA, Heinzen RA, Lopez JE, Morse K, Brayton KA, Gillespie JJ, Brown WC. *Anaplasma marginale* type IV secretion system proteins VirB2, VirB7, VirB11, and VirD4 are immunogenic components of a protective bacterial membrane vaccine. *Infect Immun.* 2010;78(3):1314–25.
127. Lopez JE, Palmer GH, Brayton KA, Dark MJ, Leach SE, Brown WC. Immunogenicity of *Anaplasma marginale* type IV secretion system proteins in a protective outer membrane vaccine. *Infect Immun.* 2007;75(5):2333–42.
128. Rikihisa Y. Role and Function of the Type IV Secretion System in *Anaplasma* and *Ehrlichia* Species. *Curr Topics Microbiol Immunol.* 2017;413:297–321.

129. Liu H, Bao W, Lin M, Niu H, Rikihisa Y. Ehrlichia type IV secretion effector ECH0825 is translocated to mitochondria and curbs ROS and apoptosis by upregulating host MnSOD. *Cell Microbiol.* 2012;14(7):1037–50.
130. Yan Q, Lin M, Huang W, Teymournejad O, Johnson JM, Hays FA, Liang Z, Li G, Rikihisa Y. Ehrlichia type IV secretion system effector Etf-2 binds to active RAB5 and delays endosome maturation. *Proc Natl Acad Sci U S A.* 2018; 115(38):E8977–86.
131. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* 2004;13(6): 1435–48.
132. Lin M, den Dulk-Ras A, Hooikaas PJ, Rikihisa Y. Anaplasma phagocytophilum Anka secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cell Microbiol.* 2007;9(11):2644–57.
133. Rikihisa Y, Lin M, Niu H, Cheng Z. Type IV secretion system of Anaplasma phagocytophilum and Ehrlichia chaffeensis. *Ann N Y Acad Sci.* 2009;1166: 106–11.
134. Udo JW, Carlson AC, Kennedy EL. Anaplasma phagocytophilum Anka is tyrosine-phosphorylated at EPIYA motifs and recruits SHP-1 during early infection. *Cell Microbiol.* 2007;9(5):1284–96.
135. Caturegli P, Asanovich KM, Walls JJ, Bakken JS, Madigan JE, Popov VL, Dumler JS. Anka: an Ehrlichia phagocytophila group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect Immun.* 2000;68(9):5277–83.
136. Park J, Kim KJ, Choi KS, Grab DJ, Dumler JS. Anaplasma phagocytophilum Anka binds to granulocyte DNA and nuclear proteins. *Cell Microbiol.* 2004; 6(8):743–51.
137. Rikihisa Y, Lin M. Anaplasma phagocytophilum and Ehrlichia chaffeensis type IV secretion and Ank proteins. *Curr Opin Microbiol.* 2010;13(1):59–66.
138. Wakeel A, den Dulk-Ras A, Hooikaas PJ, McBride JW. Ehrlichia chaffeensis tandem repeat proteins and Ank200 are type 1 secretion system substrates related to the repeats-in-toxin exoprotein family. *Front Cell Infect Microbiol.* 2011;1:22.
139. Zhu B, Nethery KA, Kuriakose JA, Wakeel A, Zhang X, McBride JW. Nuclear Translocated Ehrlichia chaffeensis Ankyrin Protein Interacts with the Mid A-Stretch of Host Promoter and Intronic Alu Elements. *Infect Immun.* 2009; 77(10):4243–55.
140. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 2017;45(D1):D200–3.
141. Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 2004;32:W327–331.
142. Obenauer JC, Cantley LC, Yaffe MB. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* 2003;31(13):3635–41.
143. Wakeel A, Zhu B, Yu XJ, McBride JW. New insights into molecular Ehrlichia chaffeensis-host interactions. *Microbes Infect.* 2010;12(5):337–45.
144. Luo T, Zhang X, Nicholson WL, Zhu B, McBride JW. Molecular characterization of antibody epitopes of Ehrlichia chaffeensis ankyrin protein 200 and tandem repeat protein 47 and evaluation of synthetic immunodeterminants for serodiagnosis of human monocytotropic ehrlichiosis. *Clin Vaccine Immunol.* 2010;17(1):87–97.
145. Luo T, Kuriakose JA, Zhu B, Wakeel A, McBride JW. Ehrlichia chaffeensis TRP120 interacts with a diverse array of eukaryotic proteins involved in transcription, signaling, and cytoskeleton organization. *Infect Immun.* 2011; 79(11):4382–91.
146. McBride JW, Walker DH. Molecular and cellular pathobiology of Ehrlichia infection: targets for new therapeutics and immunomodulation strategies. *Expert Rev Mol Med.* 2011;13:e3.
147. Zhu B, Das S, Mitra S, Farris TR, McBride JW. Ehrlichia chaffeensis TRP120 Moonlights as a HECT E3 Ligase Involved in Self- and Host Ubiquitination To Influence Protein Interactions and Stability for Intracellular Survival. *Infect Immun.* 2017;85(9):e00290–17.
148. Zhu B, Kuriakose JA, Luo T, Ballesteros E, Gupta S, Fofanov Y, McBride JW. Ehrlichia chaffeensis TRP120 binds a G+C-rich motif in host cell DNA and exhibits eukaryotic transcriptional activator function. *Infect Immun.* 2011; 79(11):4370–81.
149. Lina TT, Farris T, Luo T, Mitra S, Zhu B, McBride JW. Hacker within! Ehrlichia chaffeensis Effector Driven Phagocyte Reprogramming Strategy. *Front Cell Infect Microbiol.* 2016;6:58.
150. Klema VJ, Sepuru KM, Fullbrunn N, Farris TR, Dunphy PS, McBride JW, Rajarathnam K, Choi KH. Ehrlichia chaffeensis TRP120 nucleomodulin binds DNA with disordered tandem repeat domain. *PLoS One.* 2018;13(4): e0194891.
151. Mitra S, Dunphy PS, Das S, Zhu B, Luo T, McBride JW. Ehrlichia chaffeensis TRP120 Effector Targets and Recruits Host Polycomb Group Proteins for Degradation To Promote Intracellular Infection. *Infect Immun.* 2018;86(4):e00845–17.
152. Wang Y, Nair ADS, Alhassan A, Jaworski DC, Liu H, Trinkl K, Hove P, Ganta CK, Burkhardt N, Munderloh UG, et al. Multiple Ehrlichia chaffeensis Genes Critical for Its Persistent Infection in a Vertebrate Host Are Identified by Random Mutagenesis Coupled with In Vivo Infection Assessment. *Infect Immun.* 2020;88(10):e00316–20.
153. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek.* 2017;110(10):1281–6.
154. Auch AF, von Jan M, Klenk HP, Goker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards Genomic Sci.* 2010;2(1):117–34.
155. Chung M, Munro JB, Tettelin H, Dunning Hotopp JC. Using Core Genome Alignments To Assign Bacterial Species. *mSystems.* 2018;3(6):e00236–18.
156. Kawahara M, Rikihisa Y, Lin Q, Isogai E, Tahara K, Itagaki A, Hiramitsu Y, Tajima T. Novel genetic variants of Anaplasma phagocytophilum, Anaplasma bovis, Anaplasma centrale, and a novel Ehrlichia sp. in wild deer and ticks on two major islands in Japan. *Appl Environ Microbiol.* 2006;72(2):1102–9.
157. Rikihisa Y, Stills H, Zimmerman G. Isolation and continuous culture of Neorickettsia helminthoeca in a macrophage cell line. *J Clin Microbiol.* 1991; 29(9):1928–33.
158. Munderloh UG, Jauron SD, Fingerle L, Leitritz L, Hayes SF, Hautman JM, Nelson CM, Huberty BW, Kurtz TJ, Ahlstrand GG, et al. Invasion and intracellular development of the human granulocytic ehrlichiosis agent in tick cell culture. *J Clin Microbiol.* 1999;37(8):2518–24.
159. Xiong Q, Bao W, Ge Y, Rikihisa Y. Ehrlichia ewingii infection delays spontaneous neutrophil apoptosis through stabilization of mitochondria. *J Infect Dis.* 2008;197(8):1110–8.
160. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26(19):2460–1.
161. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* 2013;14:60.
162. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinform.* 2009;10: 421.
163. Angiuoli SV, Salzberg SL. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics.* 2011;27(3):334–42.
164. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009;75(23):7537–41.
165. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.
166. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587–9.
167. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol.* 2018;35(2): 518–22.
168. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016;44(W1):W242–5.
169. Konstantinidis KT, Ramette A, Tiedje JM. Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Appl Environ Microbiol.* 2006;72(11):7286–93.
170. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 1993;10(3):512–26.
171. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 1992;8(3):275–82.
172. Kumar S, Stecher G, Li M, Niyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol.* 2018;35(6):1547–9.
173. Ren Q, Chen K, Paulsen IT. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.* 2007;35:D274–9.

174. Krieger CJ, Zhang P, Mueller LA, Wang A, Paley S, Arnaud M, Pick J, Rhee SY, Karp PD. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 2004;32:D438–42.
175. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics.* 2010;26(13):1608–15.
176. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 2003;12(8):1652–62.

### **Publisher's Note**

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

