



# Genomic Features of *Rickettsia heilongjiangensis* Revealed by Intraspecies Comparison and Detailed Comparison With *Rickettsia japonica*

Kentarō Kasama<sup>1</sup>, Hiromi Fujita<sup>2</sup>, Seigo Yamamoto<sup>3</sup>, Tadasuke Ooka<sup>4</sup>, Yasuhiro Gotoh<sup>1</sup>, Yoshitoshi Ogura<sup>1</sup>, Shuji Ando<sup>5</sup> and Tetsuya Hayashi<sup>1\*</sup>

<sup>1</sup>Department of Bacteriology, Faculty of Medical Sciences, Kyushu University, Fukuoka, Japan, <sup>2</sup>Mahara Institute of Medical Acarology, Tokushima, Japan, <sup>3</sup>Frontier Science Research Center, University of Miyazaki, Miyazaki, Japan, <sup>4</sup>Department of Microbiology, Graduate School of Medical and Dental Sciences, Kagoshima University, Kagoshima, Japan, <sup>5</sup>Department of Virology-I, National Institute of Infectious Diseases, Tokyo, Japan

## OPEN ACCESS

### Edited by:

Haiwei Luo,  
The Chinese University of  
Hong Kong, China

### Reviewed by:

Ai Takano,  
Yamaguchi University, Japan  
Zuzana Sekeyová,  
Slovak Academy of Sciences,  
Slovakia

### \*Correspondence:

Tetsuya Hayashi  
thayash@bact.med.kyushu-u.ac.jp

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 26 August 2019

**Accepted:** 15 November 2019

**Published:** 06 December 2019

### Citation:

Kasama K, Fujita H, Yamamoto S,  
Ooka T, Gotoh Y, Ogura Y, Ando S  
and Hayashi T (2019) Genomic  
Features of *Rickettsia*  
*heilongjiangensis* Revealed by  
Intraspecies Comparison and  
Detailed Comparison With  
*Rickettsia japonica*.  
*Front. Microbiol.* 10:2787.  
doi: 10.3389/fmicb.2019.02787

*Rickettsia heilongjiangensis* is the causative agent of Far-Eastern spotted fever (FESF). In Japan, a human case of FESF was identified in Sendai in Miyagi Prefecture in 2008, and *R. heilongjiangensis* bacteria were isolated from *Haemaphysalis concinna* ticks collected in the suspected geographical area of infection. Although the intraspecies genome diversity of *Rickettsia* has been poorly investigated, our recent analysis revealed extremely low genomic diversity of *R. japonica*, the agent of Japanese spotted fever, which is a close relative of *R. heilongjiangensis*. In this study, to investigate the genomic diversity of *R. heilongjiangensis* and understand the genetic relationship between Japanese and Chinese isolates, we sequenced three isolates from *H. concinna* ticks collected in Sendai and one isolate from a *H. concinna* tick collected in Inner Mongolia, China, and performed genomic comparisons between these isolates and strain 054, the type strain isolated from a *Dermacentor silvarum* tick in Heilongjiang Province, China. Although the three Japanese strains were isolated in 2008, 2009, and 2012, their genome sequences were identical, indicating that *H. concinna* ticks carrying a single *R. heilongjiangensis* clone have been distributed in Sendai, Japan. Among the five *R. heilongjiangensis* isolates, only 81 SNPs and 13 insertion/deletion sites were identified, despite the significant differences in these isolates both geographically and temporally. A significant portion of the 81 SNPs (16/81) were found to be recombinogenic. These results indicate low genomic diversity of *R. heilongjiangensis*, as observed in *R. japonica*. We further performed a detailed genomic comparison of *R. heilongjiangensis* and *R. japonica* to accurately define conserved and species-specific genes. This analysis revealed that although notable variations were found in the genomic loci encoding RelA/SpoT family proteins and tandem repeats in major surface proteins, there was only a small difference in the gene repertoire between the two species, suggesting that SNPs and small InDels are responsible for the functional or physiological differences between

the two species, if present. Through this analysis, several species-specific genomic regions that can serve as ideal PCR targets for distinguishing *R. heilongjiangensis* and *R. japonica* were also identified.

**Keywords:** *Rickettsia heilongjiangensis*, *Rickettsia japonica*, spotted fever group rickettsia, genome sequence, intraspecies genomic diversity

## INTRODUCTION

Rickettsia are Gram-negative, obligate intracellular bacteria, some of which are known agents of arthropod-borne infectious diseases. These bacteria are divided into the spotted fever group (SFG), the typhus group (TG), and the ancestral group (AG) according to their phylogeny (Parola et al., 2013; Diop et al., 2019). The AG Rickettsia are sometimes divided into two subgroups, the *R. belli* and *R. canadensis* groups. SFG Rickettsia include more than 25 validated species, with additional new species or subspecies being proposed (Parola et al., 2013; Diop et al., 2019). The designation of a “transitional group” has also been proposed for a sublineage of SFG that includes *R. felis*, *R. akari*, and their close relatives (Gillespie et al., 2007).

Among SFG Rickettsia, at least 24 species/subspecies are known or suspected to be pathogenic to humans and cause tick-borne diseases referred to as “spotted fevers” (Parola et al., 2013). The life cycles of SFG Rickettsia are closely associated with ticks. Ticks act as both vectors and reservoirs for most SFG Rickettsia; the bacteria are maintained in ticks through transstadial and transovarial transmission (Socolovschi et al., 2009). Importantly, most SFG members are thought to have adapted to a specific tick species (Parola et al., 2013). Therefore, the distribution of most species is restricted to a relatively limited geographic region, where specific tick species are distributed, and each causes a region-specific “spotted fever.”

*Rickettsia heilongjiangensis* is the causative agent of Far-Eastern spotted fever (FESF) for humans. The endemic areas of FESF have been considered to be northeastern China (Liu et al., 2016), Siberia, and far-eastern Russia (Igolkina et al., 2018). Strain 054, the type strain of this species, was first isolated in Heilongjiang Province, China, in 1982 (Zhang et al., 2000). Determination of its complete genome sequence and subsequent phylogenetic analysis revealed that *R. heilongjiangensis* is closely related to *R. japonica*, the agent of Japanese spotted fever (JSF) (Duan et al., 2011). In Japan, a human case of FESF was identified in Miyagi Prefecture in 2008 (Ando et al., 2010). The diagnosis of this patient was performed by PCR and sequencing analysis of the *rompA*, *gltA*, and 17-kDa antigen genes because bacterial isolation was not succeeded. However, *R. heilongjiangensis* was isolated from a *Haemaphysalis concinna* tick collected in a suspected geographic area of infection (Ando et al., 2010). Multiple *R. heilongjiangensis* isolates were further obtained from *H. concinna* ticks collected in this area from 2008 to 2012 (Fujita, unpublished data), suggesting that *H. concinna* ticks carrying *R. heilongjiangensis* inhabit this region.

In Japan, JSF is the main rickettsiosis; thus, it is important to discriminate *R. japonica* and *R. heilongjiangensis*. However, the two species are very closely related and no PCR or serological tests suitable for this purpose have been developed. This is because, while the genetic diversity of *R. japonica* has been recently analyzed using high-quality whole-genome sequences (WGSs) of 31 *R. japonica* isolates obtained in various regions of Japan (Akter et al., 2017) that of *R. heilongjiangensis* has not yet been investigated.

Here, we sequenced three Japanese and one Chinese *R. heilongjiangensis* isolate and performed a WGS-based high-resolution phylogenetic analysis including strain 054 to understand the genomic diversity of this species and the genetic relationships of Japanese and Chinese isolates. A detailed genomic comparison of *R. heilongjiangensis* and *R. japonica* was also performed to better understand the genomic difference between the two species and to identify genetic traits that can be used to discriminate the two species.

## MATERIALS AND METHODS

### Bacterial Isolates

The *R. heilongjiangensis* isolates analyzed in this study are listed in **Table 1**. We sequenced four *R. heilongjiangensis* isolates: one (CH8-1) isolated from a *H. concinna* tick collected in the foothills of the Greater Khingan Range, Inner Mongolia, China, in 1996 (Fujita et al., 2002; Ando et al., 2010), and three (Sendai-29, Sendai-58, and HCN-13) isolated from *H. concinna* ticks collected in Sendai City, Miyagi, Japan, in 2008, 2009, and 2012, respectively (**Table 1**; **Figure 1A**). The sites of tick collection for the Japanese isolates were within 2.5 km of each other. CH8-1 and Sendai-29 were isolated in our previous study (Ando et al., 2010), and Sendai-58 and HCN-13 were isolated in this study by the same procedure. Strain 054, the type strain of *R. heilongjiangensis*, was isolated from a *Dermacentor silvarum* tick collected in Heilongjiang Province, China (Zhang et al., 2000), and was previously sequenced (Duan et al., 2011, 2014). The genome sequence of strain 054 (accession no. NC\_015866.1) was downloaded from the National Center for Biotechnology Information (NCBI) RefSeq database<sup>1</sup>.

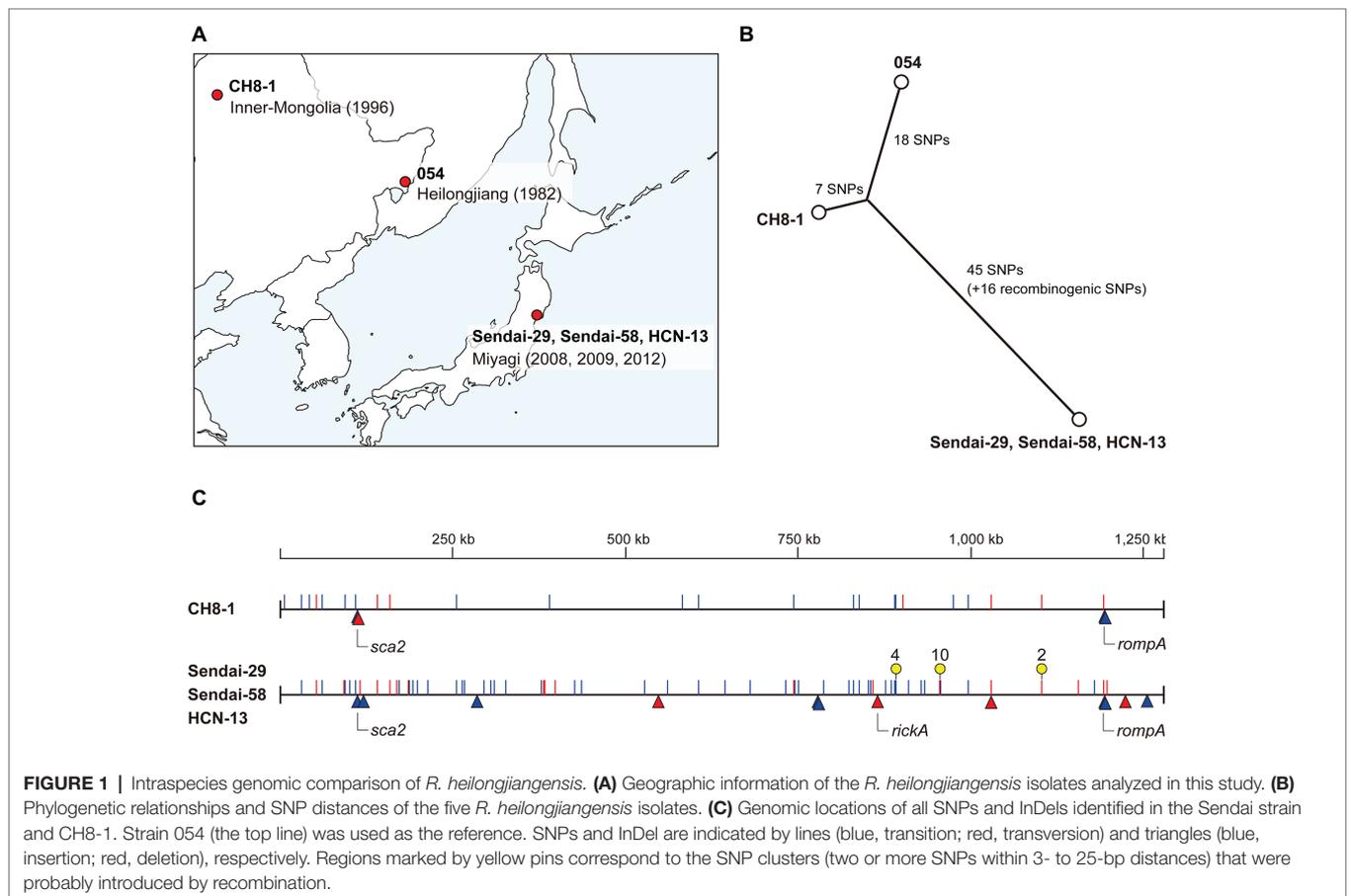
### Genomic DNA Preparation

The bacterial cells were inoculated to L929 cells and incubated in Eagle's minimal essential medium (Nissui, Japan) containing

<sup>1</sup><http://ftp.ncbi.nlm.nih.gov/genomes/refseq/>

**TABLE 1** | The *R. heilongjiangensis* strains analyzed in this study.

Strain/isolate name	Date of isolation (y/m/d)	Location of tick collection	Host tick		Genome information				Reference
			Species	Stage/Sex	Genome size (bp)	CDS (pseudogene)	rRNA/tRNA genes	Accession No.	
054	1982	China, Heilongjiang	<i>D. silvarum</i>	No information	1,278,471	1,485 (275)	3/33	NC_015866.1	Zhang et al. (2000); Duan et al. (2011)
CH8-1	1996/5/27	China, Inner Mongolia	<i>H. concinna</i>	Adult, male	1,279,380	1,485 (276)	3/33	AP019862	Fujita et al. (2002)
Sendai-29	2008/9/23	Japan, Miyagi	<i>H. concinna</i>	Nymph	1,279,369	1,485 (276)	3/33	AP019864	Ando et al. (2010)
Sendai-58	2009/3/29	Japan, Miyagi	<i>H. concinna</i>	Adult, male	1,279,369	1,485 (276)	3/33	AP019865	This study
HCN-13	2012/6/23	Japan, Miyagi	<i>H. concinna</i>	Adult, female	1,279,369	1,485 (276)	3/33	AP019863	This study



**FIGURE 1** | Intrasppecies genomic comparison of *R. heilongjiangensis*. **(A)** Geographic information of the *R. heilongjiangensis* isolates analyzed in this study. **(B)** Phylogenetic relationships and SNP distances of the five *R. heilongjiangensis* isolates. **(C)** Genomic locations of all SNPs and InDels identified in the Sendai strain and CH8-1. Strain 054 (the top line) was used as the reference. SNPs and InDel are indicated by lines (blue, transition; red, transversion) and triangles (blue, insertion; red, deletion), respectively. Regions marked by yellow pins correspond to the SNP clusters (two or more SNPs within 3- to 25-bp distances) that were probably introduced by recombination.

1% fetal bovine serum and 2 mM L-glutamine at 34°C for 7–10 days as previously described (Fujita, 2008). After centrifugation at 800 rpm for 5 min to remove L929 cell debris, the supernatants were centrifuged at 14,000 rpm for 10 min to recover bacterial cells. The obtained pellets were suspended in phosphate-buffered saline (PBS, pH 7.2). The cell suspensions were treated with DNase I (Takara Bio, Japan; 2,000 U/ml at a final concentration) for 1 h at 37°C to digest cell-free DNA. DNase I was inactivated by incubation at 65°C for 5 min, and EDTA (Nacalai Tesque, Japan) was added to the solutions at a final concentration of 2 mM.

Bacterial genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Japan) according to the manufacturer's instructions.

## Genome Sequencing, Assembly, and Annotation

Sendai-29 was sequenced using Roche 454 GS-FLX Titanium to generate 354,422 single-end reads (average read length: 316 bp). A total of 64,428 paired-end (PE) reads were also generated from a mate-pair library (insert size: 8 kb). The read sequences were assembled with GS Assembler ver. 2.3

to obtain seven scaffolds. Among these scaffolds, six were found to be derived from mouse DNA by BLASTn analysis. Nine gaps in the remaining scaffold were filled by sequencing PCR products that spanned each gap using an ABI 3730 capillary sequencer to obtain a nearly complete sequence, which contained a single gap in a tandem repeat-containing region of the *rompA* gene. The genome was also sequenced using the Illumina MiSeq platform (PE reads; 150 bp ×2) to verify and correct the sequence.

Genomic DNA libraries of CH8-1, Sendai-58, and HCN-13 were prepared using the QIAseq FX DNA Library Kit (Qiagen) and sequenced on the Illumina MiSeq platform to obtain PE sequences (300 bp ×2). After adapter and low-quality sequences were trimmed using Platanus\_trim ver. 1.07<sup>2</sup>, the reads were mapped to the mouse genome sequence (accession nos. NC\_000067-NC000087, NC\_005089) using Burrows-Wheeler-Aligner (BWA) ver. 0.7.15 (Li and Durbin, 2010) to filter mouse sequences. Unmapped reads were extracted using SAMtools ver. 1.3.1 (Li et al., 2009) and Picard tools<sup>3</sup> and assembled using Platanus Assembler ver. 1.1.0 (Kajitani et al., 2014) to generate a single scaffold that contained two to six gaps in each genome. These gaps, except for that in the tandem repeat-containing region of the *rompA* gene, were closed by capillary sequencing of PCR products as described for Sendai-29. Sequence statuses of each genome, including total numbers of reads obtained, are summarized in **Supplementary Table 1**.

The sequences of the tandem repeat-containing regions in the *rompA* gene were determined using a MinION sequencer (Oxford Nanopore Technologies, UK). The 3.8-kb regions containing the tandem repeats were amplified with the primers RH\_ompA\_F (5'-aagccgctttattaccacc-3') and RH\_ompA\_R (5'-cgcactagtgtgatctac-3') using the PrimeSTAR GXL polymerase kit (Takara Bio Inc.). The PCR products were purified using the Wizard SV Gel and PCR Clean-Up system (Promega Corporation, USA). Sequencing libraries were prepared using the Rapid Barcoding Kit (Oxford Nanopore Technologies; SQK-RBK004) and sequenced using a MinION R9.4.1 flow cell. Sequence reads were base-called using Albacore ver. 2.3.1 and quality-filtered (score > 10, read length ≥ 2,000 bp) using Nanofilt ver. 2.1.1<sup>4</sup>. The filtered reads were assembled using Canu ver. 1.7 (Koren et al., 2017). The assembled sequences were verified by mapping the Illumina reads of each isolate using BWA and SAMtools.

The genome sequences determined in this study were annotated according to the RefSeq annotation for strain 054 (NC\_015866.1). The complete genome sequences determined in this study have been deposited in DDBJ/EMBL/NCBI under accession numbers AP019864 (Sendai-29), AP019865 (Sendai-58), AP019863 (HCN-13), and AP019862 (CH8-1). The raw Illumina reads have also been deposited in the DRA/SRA/ERA database under accession number DRA008226.

<sup>2</sup>[http://platanus.bio.titech.ac.jp/pltanus\\_trim](http://platanus.bio.titech.ac.jp/pltanus_trim)

<sup>3</sup><http://broadinstitute.github.io/picard/>

<sup>4</sup><https://github.com/wdecoster/nanofilt>

## Single Nucleotide Polymorphism and Insertion and Deletion Detection in the *R. heilongjiangensis* Genomes

Single nucleotide polymorphisms (SNPs) and insertion and deletions (InDels) were analyzed between 054, CH8-1, and Sendai-29 because the genomes of the three Japanese isolates were identical. The sequence of strain 054 was used as a reference for SNP and InDel calling, which was performed based on the alignments of the complete sequences constructed using NUCmer ver. 3.1 (Delcher et al., 2002) and MUMmer (Delcher et al., 1999), respectively. Annotation of SNPs was conducted using SnpEff ver. 4.2 (Cingolani et al., 2012).

## Genomic Comparison of *R. heilongjiangensis* and *R. japonica*

The genome sequences of *R. heilongjiangensis* Sendai-29 and *R. japonica* YH\_M (Akter et al., 2017) were used for comparison. The average nucleotide identity according to the BLAST algorithm (ANiB) was calculated with JSpecies ver. 1.2.1 (Richter and Rossello-Mora, 2009). The numbers of SNPs and InDels were determined by constructing WGS alignments using NUCmer and MUMmer, respectively.

To compare the gene repertoires of the two strains using the dataset obtained from the same annotation platform, protein-coding sequences (CDSs) were identified using Prokka ver. 1.12 (Seemann, 2014). Ten and seven CDSs whose counterparts were apparently conserved but were not assigned by Prokka were manually assigned in *R. heilongjiangensis* and *R. japonica*, respectively. CDSs that were conserved in both species were identified by all-to-all blastp analysis (cutoff of >70% identity and > 30% length coverage). All identified CDSs are listed in **Supplementary Table 2** with the “conserved/nonconserved” information for each CDS. In addition, we manually inspected the alignment of the two genomes to determine whether a CDS defined as “nonconserved” was truly specific to one strain or if a homologous nucleotide sequence was present in the other strain, but a corresponding CDS could not be assigned due to the presence of SNPs or small InDels. The functions of CDSs specific to one strain and those that were degraded in one strain (split into two or more CDSs or truncated) were predicted based on the results of homology searches using the NCBI blastp pipeline (see **Supplementary Table 2** for the list of these CDSs and their predicted functions). The genome alignment of the two species was illustrated using GenomeMatcher ver. 2.205 (Ohtsubo et al., 2008).

## Analyses of Tandem Repeat Structures in the *rompA*, *sca1*, and *sca2* Genes

We analyzed the inter- and intraspecies variation of tandem repeats in the *rompA*, *sca1*, and *sca2* genes of all available *R. heilongjiangensis* and *R. japonica* genomes. The genome sequences of *R. japonica* strains were downloaded from the NCBI RefSeq database (last accessed on 2019-01-07; accession numbers are listed in **Supplementary Table 3**). Tandem repeats in the three genes were identified using Tandem Repeat Finder (Benson, 1999).

## RESULTS AND DISCUSSION

### General Genomic Features of the Four *R. heilongjiangensis* Isolates Sequenced in This Study

We determined the complete genome sequences of the three Japanese and one Chinese *R. heilongjiangensis* isolates. The three Japanese isolates (Sendai-29, Sendai-58, and HCN-13) were isolated from *H. concinna* ticks collected at very close locations (within 2.5 km) in Sendai city, Miyagi Prefecture. The Chinese isolate (CH8-1) was also isolated from a *H. concinna* tick collected in Inner Mongolia, China, in 1996 (Table 1; Figure 1A).

Although the three Japanese isolates were obtained in 2008, 2009, and 2012, their genome sequences were identical, showing no genomic variations. Therefore, they are hereafter collectively referred to as the “Sendai strain.” This result indicates that three isolates represent a single clone and that *H. concinna* ticks carrying this clone have been distributed in this region. This region is the suspected area, where the first Japanese FESF patient identified in 2008 was infected by *R. heilongjiangensis*. Although *R. heilongjiangensis* was not isolated from this patient, the patient was likely infected by this clone via a *H. concinna* tick bite. If so, there is a considerable risk of *R. heilongjiangensis* infections in this region. No FESF patients have been identified in other regions in Japan. *R. heilongjiangensis* carried by *H. concinna* ticks have not yet been isolated in other regions and no systematic investigation of the distribution of *H. concinna* ticks has been conducted. However, it is also possible that there are unidentified (hidden) FESF patients in Japan due to the poor understanding of spotted fever in the northern part of Japan, which is the nonendemic area of JSF. It will be important to systematically and continuously survey FESF patients and *R. heilongjiangensis*-carrying ticks in Japan.

The genome of the Sendai strain was comprised of a circular chromosome of 1,279,369 bp in length (GC content; 32.3%) and contained 1,485 CDSs (including 276 pseudogenes), three rRNA genes (5S, 16S, and 23S), and 33 tRNA genes. The genome size of CH8-1 was 1,279,380 bp, and the genome contained the same numbers of CDSs and rRNA and tRNA genes (Table 1). Plasmids were not found in either isolates, same as strain 054, which was isolated from a *D. silvarum* tick collected in Heilongjiang Province, China (Duan et al., 2011, 2014). Although it was previously reported that strain 054 contained many insertion sequences (ISs) (Duan et al., 2014), we identified no IS elements in the isolates sequenced in this study or in strain 054. Therefore, the *R. heilongjiangensis* genomes are devoid of IS elements. Prophages were also not detected in the *R. heilongjiangensis* genomes.

### Comparison of the Genomes of Japanese and Chinese *R. heilongjiangensis* Isolates

Comparison of the genomes of the Sendai strain and two Chinese isolates (CH8-1 and 054) revealed that the *R. heilongjiangensis* genome sequences were highly conserved, despite the fact that isolation conditions of the three isolates differed significantly geographically, temporally, and in terms of the host tick species.

We identified only 86 SNP and 13 InDel sites at the WGS level (listed in Supplementary Tables 3, 4, respectively). The identified SNPs and InDels revealed the phylogenetic relationships and genetic distances between the three isolates. The Sendai strain differed from strain 054 by 79 SNPs and 12 InDels, while CH8-1 differed from strain 054 by 25 SNPs and four InDels (Figures 1B,C). However, the Sendai strain contained three genomic regions (nucleotide positions in strain 054: 888,963–888,979, 955,989–956,014, and 1,155,544–1,155,547) with multiple accumulated SNPs (two or more SNPs within 3- to 25-bp distances) (Figure 1C; Supplementary Table 4). In addition, many of the SNPs in these regions (10/16) were generated by transversion, sharply contrasting with the rate among other SNPs (18/70). These features indicate that the three SNP clusters are most likely to have been introduced by recombination. Therefore, the actual genetic distances between the Sendai strain and the two Chinese strains should be considered excluding these recombinogenic SNPs (Figure 1C).

Among the 13 InDel sites, three were shared by the Sendai strain and CH8-1, and 9 were found only in the Sendai strain (Figure 1C; Supplementary Table 5). Importantly, the three InDel sites shared by the two strains were associated with tandem repeats in three genes encoding surface proteins: *rickA*, *rompA*, and *sca2*. Thus, the presence of these InDels does not reliably indicate the phylogeny of the strains. One InDel site unique to CH8-1 was also associated with tandem repeats in the *sca2* gene.

The very low level of genomic diversity observed for *R. heilongjiangensis* is comparable to that observed in the analysis of 31 *R. japonica* isolates obtained in various regions in Japan: 32–39 SNPs between Lineage I (the major lineage) and Lineage II (a minor lineage) and 59–66 SNPs between Lineage I and Lineage III (another minor lineage) (Akter et al., 2017). Although the two Chinese isolates were isolated from different tick species, they are more closely related than the Chinese and Japanese isolates are, suggesting a stronger geographic association than the host tick-association. However, such evolutionary relationships have not yet been described for other rickettsia species and other bacteria with similar tick-associated life cycles (transovarial and transstadial transmission and horizontal transmission via blood feeding of infected animals), such as *Coxiella burnetii* and related species (Eldin et al., 2017) and *Borrelia miyamotoi* (Wagemakers et al., 2015). In addition, as *R. heilongjiangensis* was detected in *H. concinna* and *H. longicornis* but not in *D. silvarum* in a recent surveillance conducted in China (Liu et al., 2016), it remains unclear whether *D. silvarum* is a true host of *R. heilongjiangensis*. Therefore, additional isolates obtained in other geographic regions in Far East Asia, including other regions in Japan, need to be analyzed to fully understand the genomic diversity, phylogeography, and host association of *R. heilongjiangensis*.

### Detailed Genomic Comparison of *R. heilongjiangensis* and *R. japonica*

Duan et al. (2014) reported the results of a genomic comparison of *R. heilongjiangensis* strain 054 and *R. japonica* strain YH.

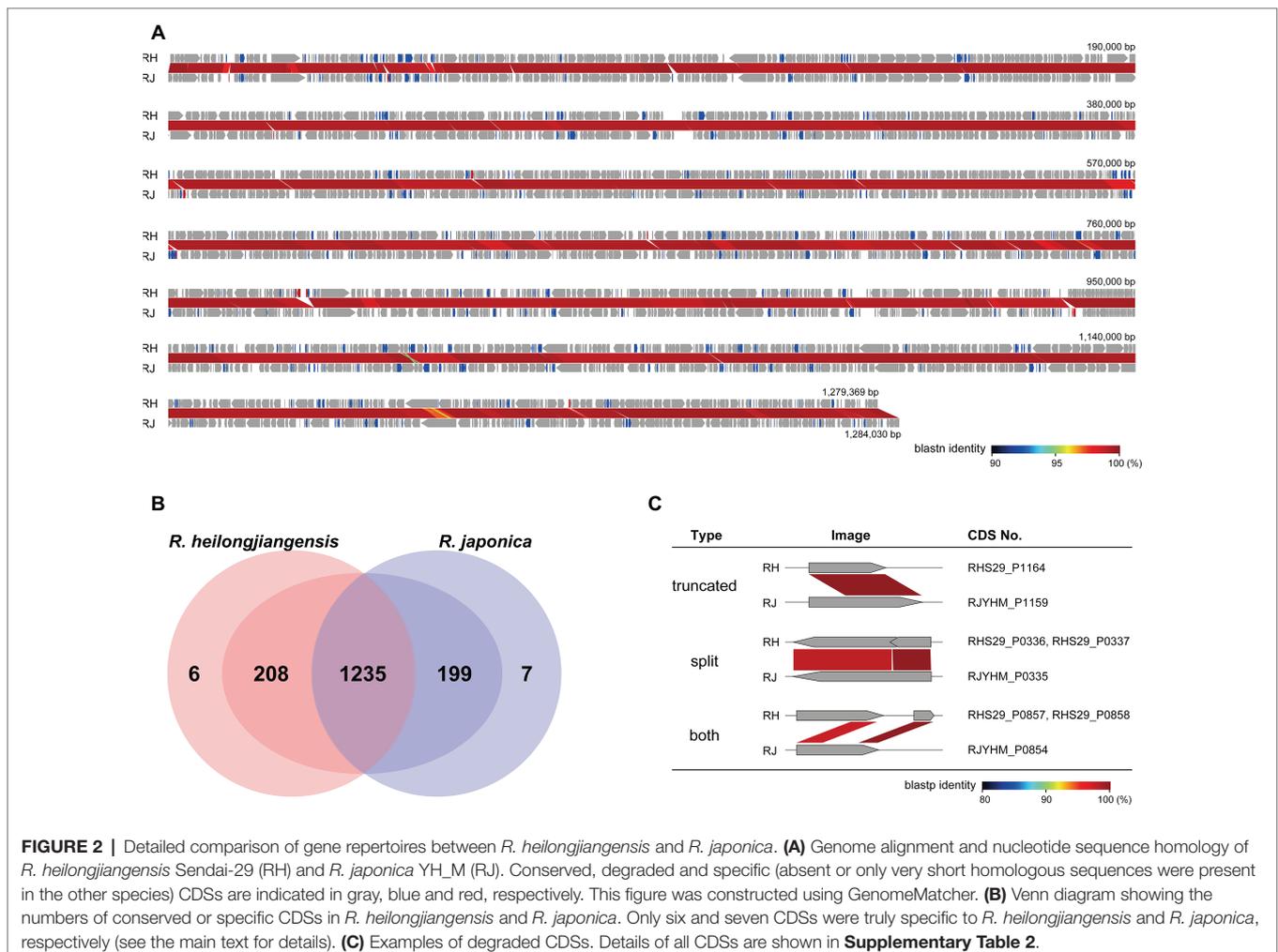
However, the sequence of YH used for comparison has been found to contain significant numbers of sequencing errors (Akter et al., 2017). Therefore, we reanalyzed the genomic differences between the two closely related species using the genome sequences of the *R. heilongjiangensis* Sendai strain and the *R. japonica* YH\_M strain (Akter et al., 2017).

The average nucleotide identity (ANIb) between the two strains was 99.2%. WGS alignment revealed perfect collinearity of the sequences and identified 7,359 SNPs and 157 InDels (>10 bp). The 157 InDels included two relatively large InDels (>1 kb). One was a 2.5-kb *R. heilongjiangensis*-specific region encoding a RelA/SpoT family protein, which was split into two CDSs and two hypothetical proteins. The other was a 1.1-kb region specific to *R. japonica* that corresponded to one of the sequences duplicated in *R. japonica*. A RelA/SpoT family protein is also encoded in this region and split into two CDSs (see below for a more detailed analysis of the genes encoding RelA/SpoT family proteins). The two regions existed in all sequenced strains in each species and were therefore species specific.

To compare the CDS repertoires, CDSs were identified using the same program (Prokka) and by manual inspection of all CDSs. Among the 1,449 and 1,441 CDSs identified in *R.*

*heilongjiangensis* and *R. japonica*, respectively, 1,235 were clearly conserved between the two species (**Figure 2; Supplementary Table 2**). Although 214 and 206 CDSs were specific to *R. heilongjiangensis* and *R. japonica*, respectively, in the initial autoannotation, many of these CDSs were apparently degraded (split, truncated, or both) due to the presence of SNPs and small InDels (**Figure 2; Supplementary Table 2**). Only six CDSs were specific to *R. heilongjiangensis* (homologous sequences were absent or were only very short in *R. japonica*). Similarly, only seven CDSs were specific to *R. japonica*. The functional annotation of these CDSs indicated that most of these CDSs encode hypothetical proteins or apparently truncated proteins (**Supplementary Table 2**).

Analysis of the CDSs that have been degraded in either of the two species (**Supplementary Table 2**) revealed that most of these CDSs also encode small hypothetical proteins or proteins that appear to have already been truncated. Several of the CDSs may encode functional proteins, such as a phosphate acetyltransferase and a CorA-like magnesium transporter (both are split in *R. heilongjiangensis*). However, in the case of the phosphate acetyltransferase gene, the acetate kinase gene (located downstream of the phosphate acetyltransferase gene) was highly



degraded in both species, indicating that the acetyl-CoA biosynthesis pathway has been inactivated in both species. In the case of the CorA-like magnesium transporter gene, another CDS encoding the magnesium transporter MgtE was conserved in both species, indicating that *R. heilongjiangensis* is also able to transport magnesium. Similarly, although a CDS encoding a putative 4 $\alpha$ -hydroxy-tetrahydrobiopterin dehydratase appeared to be truncated only in *R. japonica*, both species and other SFG *Rickettsia* lack the tetrahydrobiopterin biosynthesis pathway because other components of the pathway are missing.

The conservation of the genes encoding RelA/SpoT family proteins, which have been demonstrated in many bacteria to regulate (p)ppGpp production and contribute to the adaptation to environmental stress (Potrykus and Cashel, 2008; Atkinson et al., 2011), was complex. We identified eight loci containing RelA/SpoT family protein-encoding CDSs in either *R. heilongjiangensis* or *R. japonica* (Figure 3). Among these loci, three (loci 2, 5, and 7) were conserved in both genomes, and one (locus 8) was truncated in *R. japonica*. One (locus 3) and three (loci 1, 4, and 6) loci were specifically present in *R. heilongjiangensis* and *R. japonica*, respectively. Loci 3 and 6 correspond to the abovementioned 2.5-kb and 1.1-kb regions specific to *R. heilongjiangensis* and *R. japonica*, respectively. Although it has been reported that mutation of a RelA/SpoT-encoding gene affects the plaque phenotype in *R. rickettsii* (Clark et al., 2011), it is unknown whether or how the difference in the repertoires of RelA/SpoT family proteins between *R. heilongjiangensis* and *R. japonica* affects any phenotypes or resistance to environmental stress in the two species.

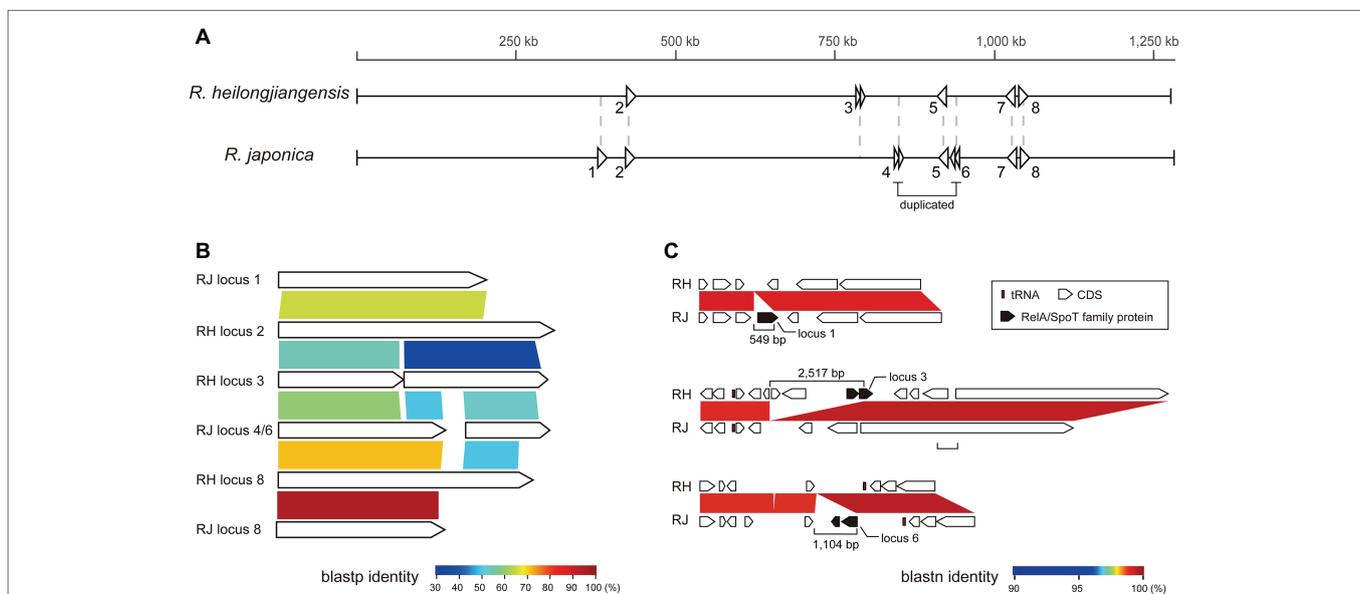
Taken together, the results of our detailed CDS comparison indicate that there is only a small difference in the gene repertoire between the two species. This finding implies that

functional or physiological differences between the two species (at least between the two strains analyzed) may, if present, result from SNPs and small InDels rather than a difference in the gene repertoire. Importantly, however, the genomic regions found to be specific to each species, particularly the 2.5-kb and 1.1-kb regions specific to *R. heilongjiangensis* and *R. japonica*, respectively, can serve as ideal PCR targets for distinguishing the two species.

## Inter- and Intraspecies Variation in Tandem Repeats in the Genes Encoding Major Surface Proteins

As described above, we observed that tandem repeat sequences in the genes encoding major surface proteins were the major sources of InDels. A similar finding was obtained in a genomic comparison of tandem repeat structures in *R. japonica* (Akter et al., 2017). Therefore, we analyzed the inter- and intraspecies variation in tandem repeat structures in the *rompA*, *sca1*, and *sca2* genes using five *R. heilongjiangensis* and 32 *R. japonica* isolates to expand our previous analysis of 31 Japanese *R. japonica* isolates (Akter et al., 2017). The 32 *R. japonica* isolates included one Chinese isolate that was recently deposited in a public database (accession no. GCA\_003454715.1, Supplementary Table 3). Note that only 12 *R. japonica* genomes were used for the analysis of *rompA* because the presence of tandem repeat sequences in the *rompA* genes of other isolates has not been determined.

rOmpA has been associated with adherence to and invasion of human endothelial cells by interacting with integrin  $\alpha$ 2 $\beta$ 1 (Hillman et al., 2013). The tandem repeat structure was first analyzed in the *rompA* gene of *R. rickettsii*



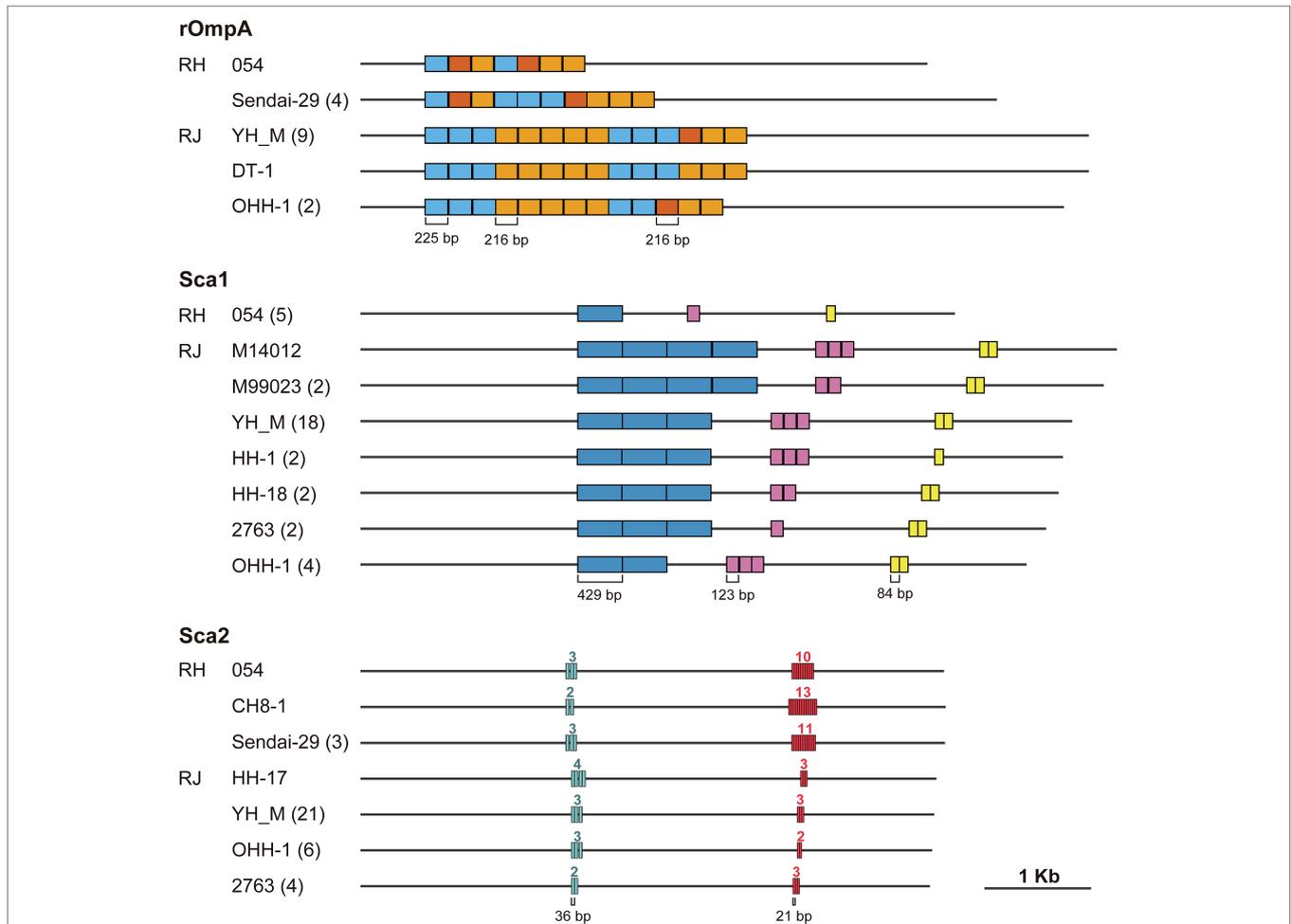
**FIGURE 3** | The genes encoding RelA/SpoT family proteins in *R. heilongjiangensis* and *R. japonica*. **(A)** The genomic locations of eight loci encoding RelA/SpoT family proteins in *R. heilongjiangensis* and *R. japonica*. **(B)** Blastp comparison of RelA/SpoT family proteins. The proteins encoded by loci 5 and 7 are not shown because they presented only low similarity to other proteins. **(C)** Three species-specific loci encoding RelA/SpoT family proteins.

(Anderson et al., 1990), but the functions of the tandem repeats remain to be elucidated. Our analysis showed that the tandem repeat regions in *rompA* are composed of three very similar sequences of 225, 216, and 216 bp in the two species (**Figure 4**). Although a previous study of *R. rickettsii* (Anderson et al., 1990) defined two types of repeat units (type-I, 225 bp; type-II, 216 bp), we found that the type-II unit can be divided into two subtypes (type-IIa and type-IIb, see **Supplementary Figure 1** for sequence alignment). While a certain level of intraspecies variation was detected in the two species, one striking interspecies difference was observed: the *R. japonica* genes contained more repeat units than the *R. heilongjiangensis* genes.

The Sca1 protein of *R. conorii* expressed in *Escherichia coli* has been demonstrated to mediate adhesion to mammalian epithelial cells (Riley et al., 2010). While three types of repeat sequences showing remarkable intraspecies variation were identified in the *sca1* genes of *R. japonica*, no tandem repeats were identified in the *R. heilongjiangensis* genes.

Functional and structural analysis of the Sca2 proteins of *R. conorii*, *R. rickettsii* and *R. parkeri* revealed that Sca2 functionally mimics the host protein formin and promotes the nucleation and elongation of actin filaments and the actin-based intracellular motility of *Rickettsia* cells (Haglund et al., 2010; Kleba et al., 2010; Cardwell, 2012; Madasu et al., 2013). Our analysis identified two types of tandem repeats in the *sca2* genes of *R. heilongjiangensis* and *R. japonica*. While both types of repeats showed intraspecies variation in each species, the 21-bp repeat has been markedly amplified in *R. heilongjiangensis*.

The surface proteins encoded by these three genes have been demonstrated or are thought to be associated with various functions related to interaction with the host (Gillespie et al., 2015). The functions of the repeat structures in these proteins are currently unknown. However, as all tandem repeat-containing regions in the three proteins are exposed on cell surfaces, analyses of the inter- and intraspecies variation in these tandem



**FIGURE 4 |** Tandem repeat structures in the *rompA*, *sca1*, and *sca2* genes of *R. heilongjiangensis* and *R. japonica* isolates. Tandem repeat sequences were identified using Tandem Repeats Finder. The names of the representative isolates exhibiting each structure are indicated with the number of isolates presenting the same structure in parenthesis. The numbers above each repeat region in the *sca2* gene represent the numbers of repeat units.

repeats in a wider range of SFG *Rickettsia* species and assessment of the structure-function relationships of this variation may provide novel insights into the biology and pathogenicity of SFG *Rickettsia*.

## CONCLUSION

We determined the WGSs of four *R. heilongjiangensis* isolates, including three Japanese isolates and one Chinese isolate. The identical WGSs of the Japanese isolates indicate that *H. concinna* ticks carrying a single *R. heilongjiangensis* clone have been distributed in Sendai, Japan. Genomic comparison of the *R. heilongjiangensis* isolates including a previously sequenced Chinese strain revealed that, while some sign of geographic separation was noted, the *R. heilongjiangensis* isolates showed very low genomic diversity, comparable with that in their close relative *R. japonica*. We further performed a detailed comparison of the CDS repertoires between *R. heilongjiangensis* and *R. japonica* and accurately identified conserved and species-specific CDSs. The results of this analysis revealed that there is only a small difference in gene repertoires between the two species, suggesting that SNPs and small InDels are mainly responsible for the functional or physiological differences between the two species. Several species-specific genomic regions were also identified in this analysis, which can serve as ideal PCR targets for distinguishing *R. heilongjiangensis* and *R. japonica*.

## REFERENCES

- Akter, A., Ooka, T., Gotoh, Y., Yamamoto, S., Fujita, H., Terasoma, F., et al. (2017). Extremely low genomic diversity of *Rickettsia japonica* distributed in Japan. *Genome Biol. Evol.* 9, 124–133. doi: 10.1093/gbe/evw304
- Anderson, B. E., McDonald, G. A., Jones, D. C., and Regnery, R. L. (1990). A protective protein antigen of *Rickettsia rickettsii* has tandemly repeated, near-identical sequences. *Infect. Immun.* 58, 2760–2769. doi: 10.1007/s11856-015-1234-0
- Ando, S., Kurosawa, M., Sakata, A., Fujita, H., Sakai, K., Sekine, M., et al. (2010). Human *Rickettsia heilongjiangensis* infection, Japan. *Emerg. Infect. Dis.* 16, 1306–1308. doi: 10.3201/eid1608.100049
- Atkinson, G. C., Tenson, T., and Haurlyuk, V. (2011). The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppppp synthetases and hydrolases across the tree of life. *PLoS One* 6:e23479. doi: 10.1371/journal.pone.0023479
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC148217/pdf/270573.pdf>
- Cardwell, M. M. (2012). Identification and characterization of the mammalian association and actin nucleating domains in the *Rickettsia conorii* autotransporter protein, Sca2. *Cell. Microbiol.* 14, 1485–1495. doi: 10.1111/j.1462-5822.2012.01815.x
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Clark, T. R., Ellison, D. W., Kleba, B., and Hackstadt, T. (2011). Complementation of *Rickettsia rickettsii* RelA/SpoT restores a nonlytic plaque phenotype. *Infect. Immun.* 79, 1631–1637. doi: 10.1128/iai.00048-11
- Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O., and Salzberg, S. L. (1999). Alignment of whole genomes. *Nucleic Acids Res.* 27, 2369–2376. doi: 10.1093/nar/27.11.2369

## DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the DDBJ/EMBL/NCBI under accession numbers AP019864 (strain Sendai-29), AP019865 (strain Sendai-58), AP019863 (strain HCN-13), AP019862 (strain CH8-1), and DRA008226 (Illumina raw data).

## AUTHOR CONTRIBUTIONS

KK and TO performed sequencing analysis and genome assembly. HF, SY, and SA isolated *R. heilongjiangensis* and prepared genomic DNA. KK analyzed the data with support from YG, YO, SA, and TH. KK and TH wrote the manuscript. All authors gave final approval for the manuscript.

## FUNDING

This work was supported by AMED under Grant Number JP19kfk0108068h0502 to TH, and by JSPS Grant-in-Aid for challenging Exploratory Research Grant Number JP19K22536 to TH.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.02787/full#supplementary-material>

- Delcher, A. L., Phillippy, A. M., Carlton, J., and Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 30, 2478–2483. doi: 10.1093/nar/30.11.2478
- Diop, A., Raoult, D., and Fournier, P. E. (2019). Paradoxical evolution of rickettsial genomes. *Ticks Tick. Borne. Dis.* 10, 462–469. doi: 10.1016/j.ttbdis.2018.11.007
- Duan, C., Tong, Y., Huang, Y., Wang, X., Xiong, X., and Wen, B. (2011). Complete genome sequence of *Rickettsia heilongjiangensis*, an emerging tick-transmitted human pathogen. *J. Bacteriol.* 193, 5564–5565. doi: 10.1128/JB.05852-11
- Duan, C., Xiong, X., Qi, Y., Gong, W., Jiao, J., and Wen, B. (2014). Genomic and comparative genomic analyses of *Rickettsia heilongjiangensis* provide insight into its evolution and pathogenesis. *Infect. Genet. Evol.* 26, 274–282. doi: 10.1016/j.meegid.2014.05.028
- Eldin, C., Mélenotte, C., Mediannikov, O., Ghigo, E., Million, M., Edouard, S., et al. (2017). From Q fever to *Coxiella burnetii* infection: a paradigm change. *Clin. Microbiol. Rev.* 30, 115–190. doi: 10.1128/CMR.00045-16
- Fujita, H. (2008). Cell culture system for isolation of disease agents: 15 years of experience in Ohara research laboratory. *Annu. Rep. Ohara Hosp.* 48, 21–42.
- Fujita, H., Takada, N., and Chaitong, U. (2002). Preliminary report on rickettsial strains of spotted fever group isolated from ticks of China, Nepal and Thailand. *Annu. Rep. Ohara Hosp.* 44, 15–18.
- Gillespie, J. J., Beier, M. S., Rahman, M. S., Ammerman, N. C., Shallom, J. M., Purkayastha, A., et al. (2007). Plasmids and rickettsial evolution: insight from *Rickettsia felis*. *PLoS One* 2:e266. doi: 10.1371/journal.pone.0000266
- Gillespie, J. J., Kaur, S. J., Sayeedur Rahman, M., Rennoll-Bankert, K., Sears, K. T., Beier-Sexton, M., et al. (2015). Secretome of obligate intracellular *Rickettsia*. *FEMS Microbiol. Rev.* 39, 47–80. doi: 10.1111/1574-6976.12084
- Haglund, C. M., Choe, J. E., Skau, C. T., Kovar, D. R., and Welch, M. D. (2010). *Rickettsia* Sca2 is a bacterial formin-like mediator of actin-based motility. *Nat. Cell Biol.* 12, 1057–1063. doi: 10.1038/ncb2109
- Hillman, R. D., Baktash, Y. M., and Martinez, J. J. (2013). OmpA-mediated rickettsial adherence to and invasion of human endothelial cells is dependent upon interaction with  $\alpha 2\beta 1$  integrin Robert. *Cell. Microbiol.* 15, 727–741. doi: 10.1002/cbm.20244.DNA

- Igolkina, Y., Vysochina, N., Ivanov, L., Epikhina, T., Pukhovskaya, N., Rar, V., et al. (2018). Genetic variability of *Rickettsia* spp. in *Dermacentor* and *Haemaphysalis* ticks from the Russian Far East. *Ticks Tick. Borne. Dis.* 9, 1594–1603. doi: 10.1016/j.ttbdis.2018.07.015
- Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 24, 1384–1395. doi: 10.1101/gr.170720.113
- Kleba, B., Clark, T. R., Lutter, E. L., Ellison, D. W., and Hackstadt, T. (2010). Disruption of the *Rickettsia rickettsii* Sca2 autotransporter inhibits actin-based motility. *Infect. Immun.* 78, 2240–2247. doi: 10.1128/IAI.00100-10
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116.Freely
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Liu, H., Li, Q., Zhang, X., Li, Z., Wang, Z., Song, M., et al. (2016). Characterization of rickettsiae in ticks in northeastern China. *Parasit. Vectors* 9, 1–9. doi: 10.1186/s13071-016-1764-2
- Madasu, Y., Kast, D. J., Kovar, D. R., Suarez, C., and Dominguez, R. (2013). *Rickettsia* Sca2 has evolved formin-like activity through a different molecular mechanism. *Proc. Natl. Acad. Sci.* 110, E2677–E2686. doi: 10.1073/pnas.1307235110
- Ohtsubo, Y., Ikeda-Ohtsubo, W., Nagata, Y., and Tsuda, M. (2008). GenomeMatcher: a graphical user interface for DNA sequence comparison. *BMC Bioinformatics* 9, 1–9. doi: 10.1186/1471-2105-9-376
- Parola, P., Paddock, C. D., Socolovschi, C., Labruna, M. B., Mediannikov, O., Kernif, T., et al. (2013). Update on tick-borne rickettsioses around the world: a geographic approach. *Clin. Microbiol. Rev.* 26, 657–702. doi: 10.1128/CMR.00032-13
- Potrykus, K., and Cashel, M. (2008). (p)ppGpp: still magical? *Annu. Rev. Microbiol.* 62, 35–51. doi: 10.1146/annurev.micro.62.081307.162903
- Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Riley, S. P., Goh, K. C., Hermanas, T. M., Cardwell, M. M., Chan, Y. G. Y., and Martinez, J. J. (2010). The *Rickettsia conorii* autotransporter protein sca1 promotes adherence to nonphagocytic mammalian cells. *Infect. Immun.* 78, 1895–1904. doi: 10.1128/IAI.01165-09
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Socolovschi, C., Mediannikov, O., Raoult, D., and Parola, P. (2009). The relationship between spotted fever group rickettsiae and ixodid ticks. *Vet. Res.* 40:34. doi: 10.1051/vetres/2009017
- Wagemakers, A., Staarink, P. J., Sprong, H., and Hovius, J. W. R. (2015). *Borrelia miyamotoi*: a widespread tick-borne relapsing fever spirochete. *Trends Parasitol.* 31, 260–269. doi: 10.1016/j.pt.2015.03.008
- Zhang, J. Z., Fan, M. Y., Wu, Y. M., Fournier, P., Roux, V., and Raoult, D. (2000). Genetic classification of “*Rickettsia heilongjiangii*” and “*Rickettsia hulini*,” two Chinese spotted fever group rickettsiae. *J. Clin. Microbiol.* 38, 3498–3501.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Kasama, Fujita, Yamamoto, Ooka, Gotoh, Ogura, Ando and Hayashi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.