

# Complete Genomic DNA Sequence of the East Asian Spotted Fever Disease Agent *Rickettsia japonica*

Minenosuke Matsutani<sup>1</sup>, Motohiko Ogawa<sup>2</sup>, Naohisa Takaoka<sup>1</sup>, Nozomu Hanaoka<sup>2</sup>, Hidehiro Toh<sup>3</sup>, Atsushi Yamashita<sup>3</sup>, Kenshiro Oshima<sup>3</sup>, Hideki Hirakawa<sup>4</sup>, Satoru Kuhara<sup>4</sup>, Harumi Suzuki<sup>1</sup>, Masahira Hattori<sup>3</sup>, Toshio Kishimoto<sup>2</sup>, Shuji Ando<sup>2</sup>, Yoshinao Azuma<sup>1,5\*</sup>, Mutsunori Shirai<sup>1</sup>

**1** Department of Microbiology and Immunology, Yamaguchi University School of Medicine, Ube, Yamaguchi, Japan, **2** Department of Virology 1, National Institute of Infectious Diseases, Tokyo, Japan, **3** Kitasato Institute for Life Sciences, Kitasato University, Sagami-hara, Kanagawa, Japan, **4** Genetic Resources Technology, Faculty of Agriculture, Kyushu University, Fukuoka, Japan, **5** Department of Science and Technology on Food Safety, Kinki University BOST, Kinokawa, Wakayama, Japan

## Abstract

*Rickettsia japonica* is an obligate intracellular alphaproteobacteria that causes tick-borne Japanese spotted fever, which has spread throughout East Asia. We determined the complete genomic DNA sequence of *R. japonica* type strain YH (VR-1363), which consists of 1,283,087 base pairs (bp) and 971 protein-coding genes. Comparison of the genomic DNA sequence of *R. japonica* with other rickettsiae in the public databases showed that 2 regions (4,323 and 216 bp) were conserved in a very narrow range of *Rickettsia* species, and the shorter one was inserted in, and disrupted, a preexisting open reading frame (ORF). While it is unknown how the DNA sequences were acquired in *R. japonica* genomes, it may be a useful signature for the diagnosis of *Rickettsia* species. Instead of the species-specific inserted DNA sequences, rickettsial genomes contain *Rickettsia*-specific palindromic elements (RPEs), which are also capable of locating in preexisting ORFs. Precise alignments of protein and DNA sequences involving RPEs showed that when a gene contains an inserted DNA sequence, each rickettsial ortholog carried an inserted DNA sequence at the same locus. The sequence, ATGAC, was shown to be highly frequent and thus characteristic in certain RPEs (RPE-4, RPE-6, and RPE-7). This finding implies that RPE-4, RPE-6, and RPE-7 were derived from a common inserted DNA sequence.

**Citation:** Matsutani M, Ogawa M, Takaoka N, Hanaoka N, Toh H, et al. (2013) Complete Genomic DNA Sequence of the East Asian Spotted Fever Disease Agent *Rickettsia japonica*. PLoS ONE 8(9): e71861. doi:10.1371/journal.pone.0071861

**Editor:** Richard Cordaux, University of Poitiers, France

**Received:** May 12, 2013; **Accepted:** July 5, 2013; **Published:** September 9, 2013

**Copyright:** © 2013 Matsutani et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported by the Japan Society for the Promotion of Science, Research for the Future Program (JSPS-RETF 00L01411), a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (KAKENHI: 14013043, 15019069, 16012245), and the Venture Business Laboratory, Yamaguchi University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: azuma@waka.kindai.ac.jp

## Introduction

*Rickettsia japonica* is an obligate intracellular bacterium isolated from ticks, wild animals, and human patients [1]. *R. japonica* was first identified as a causative agent of Japanese spotted fever [2] and has been isolated from patients in South Korea, the Philippines, and Thailand only in the past decade [3–5]. For rapid and accurate diagnosis of rickettsial infections, a new diagnostic assay of *R. japonica* using a species specific DNA sequence was proposed on the basis of a comparative genome analysis of rickettsiae isolated from Asian countries [6].

*Rickettsia* species have been divided into at least 3 groups on the basis of phylogenetic relatedness, i.e., spotted fever group (SFG), typhus group (TG), and ancestral group (AG) [7]. *R. japonica* is categorized into SFG as well as *R. rickettsii* and *R. conorii*. The TG includes epidemic and endemic typhus *Rickettsia*, *R. prowazekii* and *R. typhi*, respectively. *R. canadensis* and *R. bellii* belong to the AG [8–11]. This simplistic taxonomical assignment is not always suitable for classifications based on ecological and clinical traits and thus remains controversial to date [12].

Ogata and coworkers showed the existence of highly frequent repeat sequences of 100–150 bp in length in *Rickettsia* genomes called *Rickettsia* palindromic elements (RPEs) [13]. RPEs have been

identified in all *Rickettsia* and *Wolbachia* species sequenced so far [14]. While other bacterial palindromic repeats are mostly located within non-coding regions, certain RPEs have been found in preexisting open reading frames (ORFs) without destroying the frames. In particular, RPE-1, RPE-2, and RPE-3 are frequently observed in ORFs [13,15,16].

In the present study, to develop diagnostic tools and investigate rickettsial genome evolution, we completed the genomic DNA sequence of *R. japonica* type strain YH (VR-1363), and showed common features of inserted DNA sequences involving RPE-4, RPE-6, and RPE-7.

## Materials and Methods

### Genomic DNA sequencing

Genome DNA sequencing of *R. japonica* was performed using the same methods as those previously described [17]. Briefly, *R. japonica* YH (VR-1363) was originally isolated from a specimen of a Japanese patient with spotted fever [18]. *R. japonica* was cultured with Vero cells, and the bacterial cells for genomic DNA preparation were purified from the infected host cells using Dounce homogenization, differential centrifugation, and Percoll

density gradient centrifugation. DNA sequencing was performed using Sanger method with a 3730xl sequencer (Life Technologies) through all sequencing processes, i.e., random sequencing and gap closing stages. Whole obtained sequence data (17,664 reads) after removing sequences of vector and sequencing adaptor used here were analyzed by BLASTN against human genome database to remove contamination sequences derived from Vero cells. As a result, shotgun data were obtained with 15,502 reads, which gave 10.0×coverage. The sequences were assembled into 256 contigs using the Phred/Phrap/Consed package software [19]. Additional 1,162 sequence reads were determined by direct sequencing of PCR products to close gaps and to reconfirm sequences of regions with low qualities.

### ORF prediction and annotation

Protein-coding genes were first predicted using a pool of open reading frame (ORF) candidates indicated by 3 programs GenomeGambler ver. 1.51 [20], GeneHacker plus [21], and Glimmer 2.0 [22]. If multiple translation start sites were suggested for an ORF, manual inspection was performed to select the most probable start site on the basis of the package with the nearest upstream ORF, similarities to homologous genes, and the predicted Shine–Dalgarno sequence. Annotation for each ORF was performed using a program BLASTP [23] against genes in *R. conorii* genome [24], the NCBI non-redundant protein database and COGs [25]. Lastly, protein-coding genes were manually identified on the basis of the combination of the results of ORF prediction and annotation. All new sequence data of *R. japonica* YH VR-1363 have been deposited in DDBJ/EMBL/GenBank, the accession number: AP011533.

### RPE analyses

For identification of RPEs in *R. japonica* genome DNA sequences, we used a sequence similarity search program, HMMER3 [26], based on hidden Markov models, and the previously identified RPE sequences [24] (<http://igs-server.cnrs-mrs.fr/mgdb/Rickettsia/>) as a training data set. Rickettsial orthologs of the *R. japonica* genes with RPE-4, RPE-6, and RPE-7 were identified by reciprocal best hit using BLASTP [23] (filtering expectation values: a cut off E-value above  $10^{-10}$  and a sequence overlap cut off below 70%) with considering synteny around the genes. Herein, we showed data using 6 *Rickettsia* species, *R. conorii* [13], *R. africae* [27], *R. felis* [28], *R. prowazekii* [29], *R. typhi* [30], and *R. japonica*. DNA sequences of RPE-4, RPE-6, and RPE-7, and the boundary regions of the *Rickettsia* orthologous genes were aligned with corresponding regions of non-rickettsial orthologous genes using the ClustalW program [31]. Highly frequent sequences were identified from the RPEs by MEME version 4.0.0, which is part of the Meta-MEME package [32]. Motif sequences were illustrated using a sequence LOGO format (<http://weblogo.berkeley.edu>) [33].

The location of amino acid fragments encoded by inserted DNA sequences were analyzed on the basis of homology structure modeling of rickettsial proteins using *Escherichia coli* DNA polymerase III alpha chain, *Bacillus subtilis* tRNA ribosyltransferase-isomerase, and *Thermus thermophilus* phenylalanyl-tRNA synthetase subunit beta (for which the Protein Data Bank Identifiers (PDB IDs) are 2hnhA, 1yy3A, and 1b70B, respectively) [34].

### Unique region analyses

Comparison of the genomic DNA sequence of *R. japonica* YH VR-1363 with other 27 rickettsiae in the public databases (listed in Table S1) were carried out using BLASTN (a cut-off E-value  $>e^{-5}$ ) to figure out *R. japonica* unique regions. Low similarity

regions shown from the first screening were re-analyzed using BLASTN (a cut-off E-value  $>e^{-5}$ ) to assign the unique regions.

## Results and Discussion

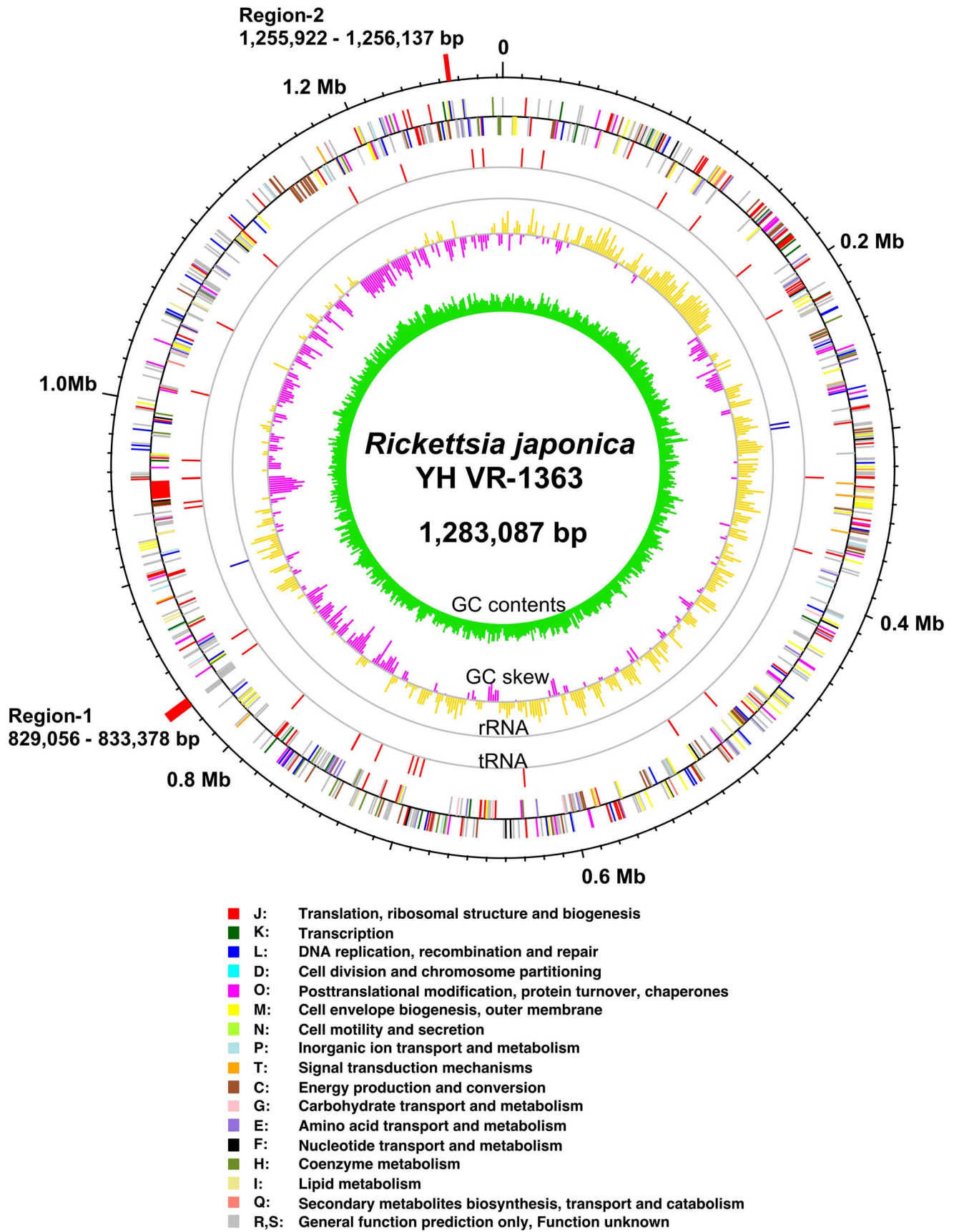
### General features of the *R. japonica* genome

The genomic DNA of *R. japonica* strain YH consists of a 1,283,087-bp circular chromosome with 32.4% GC content, and no plasmids were identified (Fig. 1). A total of 971 protein-coding genes were identified in the genome, and those ORFs covered 80.1% of the chromosome. Putative functions were assigned to 706 genes. The genomes of *Rickettsia* species belonging to the SFG are well known to contain numerous pseudogenes, and many of their orthologs in *R. felis* are still intact [28,35]. We identified 202 pseudogenes in the *R. japonica* genome, and the 144 orthologs in *R. felis* are unbroken. The *R. japonica* genome includes all known virulence factors conserved among SFG rickettsiae, such as rOmpA (which functions as the initial adhesion and is conserved in SFG [36]), rOmpB (which plays an important role for recognition and invasion, and is conserved widely in rickettsiae [37]), phospholipase A<sub>2</sub> (which has been proposed to mediate escape from phagosomes [38]), InvA and RickA (which mediate actin polymerization and motility [39]), and 3 hemolysins (i.e., hemolysin A, hemolysin C, and hemolysin-like protein).

Recently the draft genome sequence of *R. japonica* type strain YH (VR-1336) have been published [40]. There are no differences between DNA sequences of 16S rRNA genes (1,508 bp) of *R. japonica* YH (VR-1336) and *R. japonica* YH (VR-1363), and 23S-5S rRNA gene operons (3,151 bp) as well (a phylogenetic tree was shown in Figure S1A). In contrary, amino acid sequences of the *R. japonica* VR-1336 gene products are relatively diverse from *R. japonica* YH VR-1363, resulted in presentation of a different phylogenetic tree from one based on rRNA genes (Figure S1B). The precise taxonomical classification and nomenclatural scheme of *Rickettsia* species should be clarified in the future.

### Unique traits of the *R. japonica* genome

Comparison of the genomic DNA sequence of *R. japonica* YH VR-1363 with the sequences of other 27 rickettsiae in the public databases (Table S1) showed that 2 regions (Region-1 and Region-2) were unique to only a few rickettsiae including *R. japonica* (a cut off E-value above  $e^{-5}$ ) (Fig. 1). The Region-1, which was located at 829,056–833,378 bp in *R. japonica* genome, was conserved only in *R. japonica* and *R. heilongjiangensis* [41], and Region-2 at 1,255,922–1,256,137 bp was conserved in *R. japonica*, *R. heilongjiangensis*, *Candidatus Rickettsia amblyommii* (Accession number: CP003334) and *Rickettsia montanensis* (CP003340). Polymerase chain reaction (PCR) analysis clarified that Region-2 was conserved among all 5 strains of *R. japonica* (i.e., YH, DT-1, FLA-1, HH-8, and HH-9) [6]. Region-1 consists of 12 genes encoding 3 integrases (RJP\_0637-RJP\_0639), 2 protein kinases (RJP\_0642-RJP\_0643), and 7 hypothetical proteins (RJP\_0640, RJP\_0641, RJP\_0644-RJP\_0648). Region-2 carries a hypothetical protein (RJP\_0988). Similarities of all 13 genes to other organisms except rickettsiae were rather low, and the organisms carrying these similar genes varied widely from alphaproteobacteria to plants and animals. For instance, amino acid sequences of the integrases are partially similar to integrases from *R. belli*, *R. felis*, *R. massiliae*, and *Orientia tsutsugamushi*, but not to integrases from *Rickettsia* species closely related in the SFG. Interestingly, the hypothetical protein in Region-2 contains ankyrin repeats and is partially similar to mammalian proteins, such as *Tupaia chinensis* myosin-XVI (ELW47454) with 50% identity out of 32 aa of RJP\_0988. Lastly, since the DNA sequences of Region-1 and



**Figure 1. Circular exhibition of the *Rickettsia japonica* YH genome.** The outermost scale is marked for nucleic acid position in Mbp, and Region-1 and -2 (red). From the outside (track 1), gene positions and directions (clockwise on the outside and anti-clockwise on the inside) of each gene were classified and colored based on COGs [45]. Track 2 and 3: tRNA (red) and rRNA (blue), respectively. Track 4: GC skew, outside yellow and inside purple indicate values >0 and values <0 as calculated by  $(G-C)/(G+C)$  [46]. Track 5: innermost, GC contents. Based on the cumulus of the GC skew values, accompanied by other *Rickettsia* genomes, a hypothetical origin, shown as *ori*, was determined and the base numbers were counted from the origin.

doi:10.1371/journal.pone.0071861.g001

Region-2 are unique to only a few *Rickettsia* species and there are some differences in DNA sequences of the species, it is very useful to recognize and distinguish *Rickettsia* species [6].

The *R. japonica* genome shows genome-wide synteny against SFG rickettsial genomes, while there are 2 genome rearrangements occurred between 668 and 754 kb, and 851 and 939 kb in the *R. japonica* YH genome (Figure S2). These regions are thought to be hotspots for genome rearrangements of *Rickettsia* [30] since frequent rearrangements were observed.

### Common features of RPE-4, RPE-6, and RPE-7

RPEs in the genes of *R. japonica* were identified by hidden Markov models. The data showed that the *R. japonica* genes with RPEs and their inserted positions were conserved in SFG *Rickettsia* genome evolution. In the case of genes with RPE-1, RPE-2, and RPE-3, rickettsial orthologs of the genes contained the same group of RPEs at similar sites [13]. In contrast, in the case of genes carrying RPE-4, RPE-6, and RPE-7, no typical RPEs were observed in some rickettsial orthologs; or they contained different groups of RPEs located near to each other but at different positions (Table 1). It seems that there are at least 2 possibilities. It could be that the insertion sites for RPE-4, RPE-6, and RPE-7 are hotspots for insertions of different origins. The other possibility is that RPE-4, RPE-6, and RPE-7 originated from the same inserted DNA sequence but widely diverged thereafter.

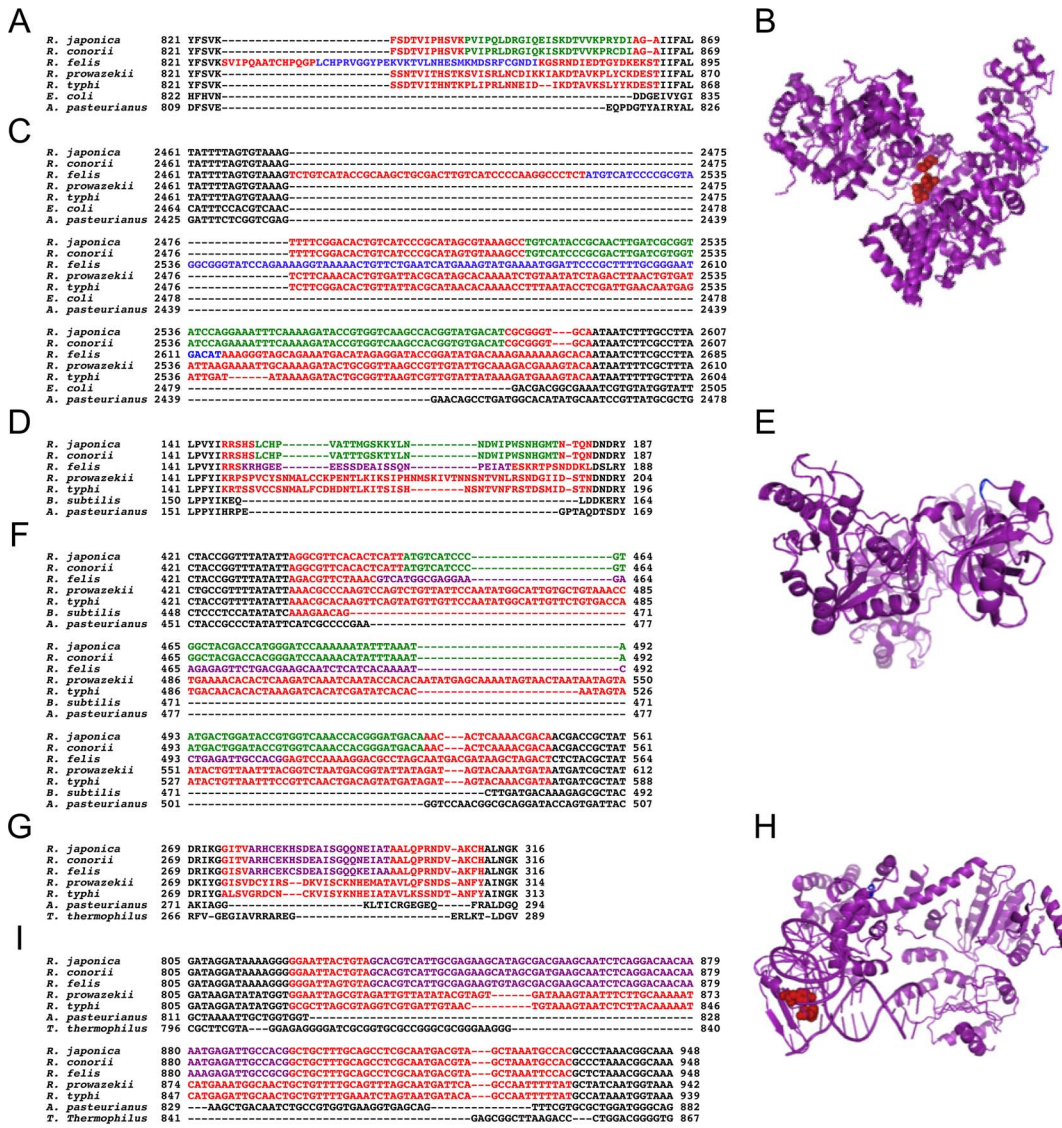
To clarify the possibilities, we precisely aligned 90 RPE loci using amino acid and, thus, DNA sequences from 18 orthologs of 5 rickettsial genomes (3 ortholog examples are shown in Fig. 2) with orthologs from other bacteria. The result obviously indicated that all *Rickettsia* orthologs contained inserted DNA sequences at the same sites, but the insertions were extended from typical RPEs called extended RPE-4 (eRPE-4). eRPE-4 sequences were identified even in the orthologs without typical RPEs (Table 2 and Fig. 2). Location analysis of amino acid fragments encoded by eRPE-4s illustrated that the protein domains, including the extended regions, were located in solvent-exposed areas (Fig. 2), which was similar to RPE-1, RPE-2, and RPE-3 [15]. Analysis using MEME software [42] showed that ATGAC sequences, which were originally found as part of palindromic sequences in RPE-4, RPE-6, and RPE-7, were shown to be highly frequent not only in the typical RPEs but also in the extended region of eRPE-4s (Fig. 2J, Table 2). The frequencies of the ATGAC motif were significantly higher in eRPE-4 than in non-RPE regions (Table 2), and those in RPE-1, RPE-2, and RPE-3 were significantly lower than the frequencies in eRPE-4s and any other genome region. In addition, we found that biotin-(acetyl-CoA carboxylase) ligase genes of *Wolbachia* and *Rickettsia* retained *Wolbachia* palindromic elements (WPEs) [43] and eRPE-4 in the same region of each gene. Additionally, the frequency of the ATGAC motif in WPEs was high (data not shown). However, the possibility that the RPEs

**Table 1.** List of genes with RPE-4, RPE-6 and RPE-7 and insertion lengths in *R. japonica*.

<i>R. japonica</i>			<i>R. conorii</i>	<i>R. africae</i>	<i>R. felis</i>	<i>R. prowazekii</i>	<i>R. typhi</i>	
Gene ID	RPE length	RPE group	RPE group	RPE group	RPE group	RPE group	RPE group	Product
RJP_0284	84	RPE-4	RPE-4	RPE-4	ND	ND	ND	multisubunit Na <sup>+</sup> /H <sup>+</sup> antiporter, mnhB subunit
RJP_0560	142	RPE-4	RPE-4	RPE-4	RPE-4	ND	ND	putative aminomethyltransferase
RJP_0434	98	RPE-4	RPE-4	RPE-4	RPE-6	ND	ND	recB family exonuclease
RJP_0891	114	RPE-4	RPE-4	RPE-4	RPE-6	ND	ND	DNA polymerase III subunit alpha
RJP_0222	115	RPE-4	RPE-4	RPE-4	RPE-7	ND	ND	tRNA ribosyltransferase-isomerase
RJP_0300	90	RPE-6	RPE-6	RPE-6	RPE-7	ND	ND	NADH dehydrogenase subunit N
RJP_0790	192	RPE-4	RPE-4	RPE-4	RPE-7	ND	ND	proline/betaine transporter
RJP_0384	137	RPE-4	RPE-4	RPE-4	RPE-7	ND	RPE-6	putative lipid A core - O-antigen ligase
RJP_0202	150	RPE-6	RPE-6	RPE-6	RPE-6	ND	RPE-4	bifunctional penicillin-binding protein 1C
RJP_0524	96	RPE-6	RPE-6	RPE-6	RPE-4	ND	ND	biotin-(acetyl-CoA carboxylase) ligase
RJP_0634	101	RPE-6	RPE-6	RPE-6	RPE-4	ND	ND	excinuclease ABC subunit C
RJP_0964	108	RPE-6	RPE-6	RPE-6	RPE-7	ND	ND	tRNA (guanine-N(7)-methyltransferase
RJP_0013	68	RPE-7	RPE-7	RPE-7	RPE-7	ND	RPE-7	ABC transporter substrate binding protein
RJP_0722	144	RPE-7	RPE-7	RPE-7	RPE-6	ND	ND	ATP-dependent DNA helicase
RJP_0124	114	RPE-7	RPE-7	RPE-7	RPE-6	ND	ND	putative nucleoside-diphosphate-sugar epimerase
RJP_0449	108	RPE-7	RPE-7	RPE-7	RPE-7	ND	ND	phenylalanyl-tRNA synthetase beta subunit
RJP_0691	144	RPE-7	RPE-7	RPE-7	RPE-7	ND	ND	isoleucyl-tRNA synthetase
RJP_0552	96	RPE-7	RPE-7	RPE-7	RPE-7	RPE-4	ND	soluble lytic murein transglycosylase

ND: typical RPE-4, RPE-6 and RPE-7 were not detected while orthologous genes and insertions exist.

doi:10.1371/journal.pone.0071861.t001



**Figure 2. Multiple alignments of amino acid and DNA sequences, including inserted DNA sequences.** Each panel shows A–C) DNA polymerase III alpha chain; D–F) tRNA ribosyltransferase-isomerase; G–I) phenylalanyl-tRNA synthetase subunit beta; A, D, G) amino acid alignments; B, E, H) structure models; C, F, I) DNA alignments. Letters in alignments indicate green, typical RPE-4s; blue, typical RPE-6s; purple, typical RPE-7s; red, extended RPE-4 regions; black, boundary regions of genes with inserted DNA sequences. Amino acids encoded by inserted DNA sequences were shown by red balls in the protein structures. Gene sequences were extracted from each genome sequence; *Rickettsia japonica* (this work); *R. conorii*, AE006914 [13]; *R. felis*, CP000053 [28]; *R. prowazekii*, AJ235269 [29]; *R. typhi*, AE017197 [30]; *Acetobacter pasteurianus* IFO 3283-01, AP011121 [47]. Protein Data Bank Identifiers (PDB IDs) for amino acid sequences of *Escherichia coli* DNA polymerase III alpha chain, *Bacillus subtilis* tRNA ribosyltransferase-isomerase, and *Thermus thermophilus* phenylalanyl-tRNA synthetase subunit beta are 2hnhA, 1yy3A, and 1b70B, respectively. J) The highly conserved short sequence in e-RPEs were illustrated using a sequence LOGO format [33]. doi:10.1371/journal.pone.0071861.g002

were inserted into certain hotspots is still not negligible. The observations implicated that RPE-4, RPE-6, and RPE-7 (and possibly WPEs) were derived from the same origin, which was

different from the origins of RPE-1, RPE-2, and RPE-3. Thus, the initial insertion occurred before the divergence of Rickettsiaceae after the separation of Rickettsiaceae from alphaproteobacteria.

**Table 2.** Frequency of the ATGAC sequences in extended RPE-4s (eRPE-4s).

Regions	Category of eRPE-4		Length (bp)	Amount of ATGAC	Frequency (per kb)
eRPE-4	eRPE-4 including RPE-4	Total	2,243	35	15.6
		RPE-4	806	27	33.5
		Extended	1,437	8	5.6
eRPE-4 including RPE-6	eRPE-4 including RPE-6	Total	1,805	35	19.4
		RPE-6	1,235	20	16.2
		Extended	570	15	26.3
eRPE-4 including RPE-7	eRPE-4 including RPE-7	Total	2,291	33	14.4
		RPE-7	878	18	20.5
		Extended	1,413	15	10.6
eRPE-4 without RPE-4,6,7	eRPE-4 without RPE-4,6,7	Extended	3,565	17	4.8
RPE-1,-2,-3,-5		24,452	3	0.1	
Other regions		1,248,731	2,311	1.9	
Whole genome of <i>R. japonica</i>		1,283,087	2,417	1.9	

ND: Not detected.

doi:10.1371/journal.pone.0071861.t002

## Conclusions

An accurate diagnosis of rickettsial infections is desirable to distinguish these infections from other similar diseases such as leptospirosis and dengue fever, because rickettsioses is remediable by suitable antimicrobial therapy. In this study, we performed complete sequencing of the genomic DNA of *R. japonica*, which causes tick-borne Japanese spotted fever prevalent in East Asia [44]. Compared to other *Rickettsia* species, *R. japonica* YH (VR-1363) genome contains 2 unique regions, Region-1 and Region-2. Evolutional origins of these regions are not clear, but the DNA sequences will be a useful tool for the diagnosis of *Rickettsia* species [6]. Two rearrangement regions specific to *R. japonica* might be valuable for distinguishing *Rickettsia* species in SFG as well. The precise alignments of DNA sequences of genes involving RPE-4, RPE-6, and RPE-7 showed that insertions of these sequences could occur in the same loci of orthologous genes. Those observations imply that RPE-4, RPE-6, and RPE-7 diverged from a common origin and, thus, may be a valuable signature for the diagnosis of *Rickettsia* species.

## Supporting Information

**Figure S1 Neighbor-joining phylogenetic trees of SFG *Rickettsias*.** A) DNA sequences of 23s-5s rRNA gene operons were aligned with ClustalW [31] and the phylogenetic tree was constructed using Seaview4 [48]. B) Amino acid sequences produced by concatenation of 670 gene products shared with SFG *Rickettsia* were aligned with ClustalW [31] and the phylogenetic tree was constructed using MEGA5.05 [49]. SFG *Rickettsias* used here were *R. africae* ESF-5 (CP001612), *R. parkeri* Portsmouth (CP003341), *R. conorii* Malish 7 (AE006914), *R. slovaca* 13-B (CP002428), *R. peacockii* Rustic (CP001227), *R. philipii* 364D (CP003308), *R. rickettsii* Sheila Smith (CP000848), *R. heilongjiangensis*

## References

- Mahara F (2006) Rickettsioses in Japan and the far East. *Ann N Y Acad Sci* 1078: 60–73.
- Uchida T, Uchiyama T, Kumano K, Walker DH (1992) *Rickettsia japonica* sp. nov., the etiological agent of spotted fever group rickettsiosis in Japan. *Int J Syst Bacteriol* 42: 303–305.
- Chung MH, Lee SH, Kim MJ, Lee JH, Kim ES, et al. (2006) Japanese spotted fever, South Korea. *Emerg Infect Dis* 12: 1122–1124.
- Gaywee J, Sunyakumthorn P, Rodkvamtook W, Ruang-arereate T, Mason CJ, et al. (2007) Human infection with *Rickettsia* sp. related to *R. japonica*, Thailand. *Emerg Infect Dis* 13: 671–673.

054 (CP002912), *R. japonica* YH VR-1363 (AP011533), *R. japonica* YH VR-1336 (AMRT00000000), *R. rhipicephali* 3–7-female6-CWPP (CP003342), *R. massiliae* MTU5 (CP000683), *R. montanensis* OSU 85-930 (CP003340), and *R. felis* URRWXCal2 (CP000053). (TIF)

**Figure S2 Dot-plot analyses of *R. japonica* with *Rickettsia*.** Highly similar regions in genomic DNA sequences of *Rickettsia* species to *R. japonica* YH VR-1363 were bi-directionally shown by dot-plot analysis using GenomeMatcher [50]. Red and green spots indicate similar regions in forward- and complementary-strands of each genome sequence, respectively. Inverted and translocation positions were indicated by 4 vertical lines at 668, 754, 851 and 939 kb of *R. japonica* genome. *Rickettsias* used here were *R. japonica* YH VR-1363 (AP011533), *R. conorii* Malish 7 (AE006914), *R. rickettsii* Sheila Smith (CP000848), *R. heilongjiangensis* 054 (CP002912), *R. massiliae* MTU5 (CP000683), *R. felis* URRWXCal2 (CP000053) and *R. prowazekii* str. Madrid E (AJ235269). (TIF)

**Table S1 *Rickettsia* species for analyses.** (DOC)

## Acknowledgments

We thank S. Yoshioka (Yamaguchi University), and K. Furuya, C. Yoshino, A. Tamura and Y. Hattori (Kitasato University) for technical assistance.

## Author Contributions

Conceived and designed the experiments: MM YA MS. Performed the experiments: MO NT MM YA. Analyzed the data: MM NH HT AY KO HH SK HS YA. Contributed reagents/materials/analysis tools: MO MH TK SA. Wrote the paper: MM YA.

5. Camer GA, Alejandria M, Amor M, Satoh H, Muramatsu Y, et al. (2003) Detection of antibodies against spotted fever group *Rickettsia* (SFGR), typhus group *Rickettsia* (TGR), and *Coxiella burnetii* in human febrile patients in the Philippines. Japanese journal of infectious diseases 56: 26–28.
6. Hanaoka N, Matsutani M, Kawabata H, Yamamoto S, Fujita H, et al. (2009) Diagnostic assay for *Rickettsia japonica*. Emerg Infect Dis 15: 1994–1997.
7. Uchiyama T, Kishi M, Ogawa M (2012) Restriction of the growth of a nonpathogenic spotted fever group rickettsia. FEMS Immunol Med Microbiol 64: 42–47.
8. Fournier PE, Belghazi L, Robert C, Elmkouri K, Richards AL, et al. (2008) Variations of plasmid content in *Rickettsia felis*. PLoS One 3: e2289.
9. Merhej V, Raoult D (2011) Rickettsial evolution in the light of comparative genomics. Biol Rev Camb Philos Soc 86: 379–405.
10. Blanc G, Ogata H, Robert C, Audic S, Suhre K, et al. (2007) Reductive genome evolution from the mother of *Rickettsia*. PLoS Genet 3: e14.
11. Stothard DR, Clark JB, Fuerst PA (1994) Ancestral divergence of *Rickettsia bellii* from the spotted fever and typhus groups of *Rickettsia* and antiquity of the genus *Rickettsia*. Int J Syst Bacteriol 44: 798–804.
12. Gillespie JJ, Joardar V, Williams KP, Driscoll T, Hostetler JB, et al. (2012) A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. J Bacteriol 194: 376–394.
13. Ogata H, Audic S, Barbe V, Artiguenave F, Fournier PE, et al. (2000) Selfish DNA in protein-coding genes of *Rickettsia*. Science 290: 347–350.
14. Ogata H, La Scola B, Audic S, Renesto P, Blanc G, et al. (2006) Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. PLoS Genet 2: e76.
15. Ogata H, Audic S, Abergel C, Fournier PE, Claverie JM (2002) Protein coding palindromes are a unique but recurrent feature in *Rickettsia*. Genome Res 12: 808–816.
16. Blanc G, Ogata H, Robert C, Audic S, Claverie JM, et al. (2007) Lateral gene transfer between obligate intracellular bacteria: evidence from the *Rickettsia massiliae* genome. Genome Res 17: 1657–1664.
17. Azuma Y, Hirakawa H, Yamashita A, Cai Y, Rahman MA, et al. (2006) Genome sequence of the cat pathogen, *Chlamydomphila felis*. DNA research 13: 15–23.
18. Uchida T, Tashiro F, Funato T, Kitamura Y (1986) Isolation of a spotted fever group *Rickettsia* from a patient with febrile exanthematous illness in Shikoku, Japan. Microbiol Immunol 30: 1323–1326.
19. Gordon D, Desmarais C, Green P (2001) Automated finishing with autofinish. Genome Res 11: 614–625.
20. Sakiyama T, Takami H, Ogasawara N, Kuhara S, Kozuki T, et al. (2000) An automated system for genome analysis to support microbial whole-genome shotgun sequencing. Biosci Biotechnol Biochem 64: 670–673.
21. Yada T, Nakao M, Totoki Y, Nakai K (1999) Modeling and predicting transcriptional units of *Escherichia coli* genes using hidden Markov models. Bioinformatics 15: 987–993.
22. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27: 4636–4641.
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
24. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, et al. (2001) Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science 293: 2093–2098.
25. Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278: 631–637.
26. Eddy SR (2011) Accelerated Profile HMM Searches. PLoS Comput Biol 7: e1002195.
27. Fournier PE, El Karkouri K, Leroy Q, Robert C, Giumelli B, et al. (2009) Analysis of the *Rickettsia africae* genome reveals that virulence acquisition in *Rickettsia* species may be explained by genome reduction. BMC genomics 10: 166.
28. Ogata H, Renesto P, Audic S, Robert C, Blanc G, et al. (2005) The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. PLoS biology 3: e248.
29. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, et al. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396: 133–140.
30. McLeod MP, Qin X, Karpathy SE, Gioia J, Highlander SK, et al. (2004) Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other rickettsiae. J Bacteriol 186: 5842–5855.
31. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.
32. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. (2009) MEME-SUITE: tools for motif discovery and searching. Nucleic Acids Res 37: W202–208.
33. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14: 1188–1190.
34. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, et al. (2009) Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc 4: 1–13.
35. Andersson JO, Andersson SG (1999) Genome degradation is an ongoing process in *Rickettsia*. Mol Biol Evol 16: 1178–1191.
36. Li H, Walker DH (1998) rOmpA is a critical protein for the adhesion of *Rickettsia rickettsii* to host cells. Microb Pathog 24: 289–298.
37. Uchiyama T (2003) Adherence to and invasion of Vero cells by recombinant *Escherichia coli* expressing the outer membrane protein rOmpB of *Rickettsia japonica*. Ann N Y Acad Sci 990: 585–590.
38. Winkler HH, Miller ET (1982) Phospholipase A and the interaction of *Rickettsia prowazekii* and mouse fibroblasts (L-929 cells). Infect Immun 38: 109–113.
39. Gouin E, Egile C, Dehoux P, Villiers V, Adams J, et al. (2004) The RickA protein of *Rickettsia conorii* activates the Arp2/3 complex. Nature 427: 457–461.
40. Dong X, El Karkouri K, Robert C, Raoult D, Fournier PE (2012) Genomic analysis of *Rickettsia japonica* strain YHT. J Bacteriol 194: 6992.
41. Duan C, Tong Y, Huang Y, Wang X, Xiong X, et al. (2011) Complete genome sequence of *Rickettsia heilongjiangensis*, an emerging tick-transmitted human pathogen. J Bacteriol 193: 5564–5565.
42. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings/International Conference on Intelligent Systems for Molecular Biology 2: 28–36.
43. Ogata H, Suhre K, Claverie JM (2005) Discovery of protein-coding palindromic repeats in *Wolbachia*. Trends in microbiology 13: 253–255.
44. Walker DH (2007) Rickettsiae and rickettsial infections: the current state of knowledge. Clin Infect Dis 45 Suppl 1: S39–44.
45. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, et al. (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4: 41.
46. Grigoriev A (1998) Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res 26: 2286–2290.
47. Azuma Y, Hosoyama A, Matsutani M, Furuya N, Horikawa H, et al. (2009) Whole-genome analyses reveal genetic instability of *Acetobacter pasteurianus*. Nucleic Acids Res 37: 5768–5783.
48. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27: 221–224.
49. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
50. Ohtsubo Y, Ikeda-Ohtsubo W, Nagata Y, Tsuda M (2008) GenomeMatcher: a graphical user interface for DNA sequence comparison. BMC Bioinformatics 9: 376.