



HAL
open science

Orientia tsutsugamushi, agent of scrub typhus, displays a single metapopulation with maintenance of ancestral haplotypes throughout continental South East Asia

Patimaporn Wongprompitak, Veasna Duong, Wichitra Anukool, Lay Sreyrath, Trinh Thi Xuan Mai, Laurent Gavotte, Catherine Moulia, Emmanuel Cornillot, Pattama Ekpo, Yupin Suputtamongkol, et al.

► To cite this version:

Patimaporn Wongprompitak, Veasna Duong, Wichitra Anukool, Lay Sreyrath, Trinh Thi Xuan Mai, et al.. *Orientia tsutsugamushi, agent of scrub typhus, displays a single metapopulation with maintenance of ancestral haplotypes throughout continental South East Asia*. *Infection, Genetics and Evolution*, 2015, 31, pp.1-8. 10.1016/j.meegid.2015.01.005 . hal-03048806

HAL Id: hal-03048806

<https://hal.science/hal-03048806v1>

Submitted on 18 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Orientia tsutsugamushi, agent of scrub typhus, displays a single metapopulation with maintenance of ancestral haplotypes throughout continental South East Asia

Patimaporn Wongprompitak^{a,b,1}, Veasna Duong^{c,1}, Wichitra Anukool^{a,d}, Lay Sreyrath^c, Trinh Thi Xuan Mai^e, Laurent Gavotte^b, Catherine Moulia^b, Emmanuel Cornillot^f, Pattama Ekpo^a, Yupin Suputtamongkol^{a,g}, Philippe Buchy^c, Roger Frutos^{h,i,*}

^a Department of Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

^b UM2, ISEM, UMR 5554, CNRS-UM2-IRD, Université Montpellier 2 CC065, Place E. Bataillon, 34095 Montpellier Cedex 5, France

^c Virology Unit, Institut Pasteur in Cambodia, 5 Monivong Blvd, PO Box 983, Phnom Penh, Cambodia

^d Graduate Program in Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

^e Institut Pasteur of Nha Trang, 10 Tran Phu Street, Nha Trang, Viet Nam

^f UM1, CPBS, UMR 5236, CNRS-UM1-UM2, 1919 route de Mende, 34293 Montpellier Cedex 5, France

^g Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

^h UM2, CPBS, UMR 5236, CNRS-UM1-UM2, 1919 route de Mende, 34293 Montpellier Cedex 5, France

ⁱ Cirad, UMR 17, Cirad-Ird, TA-A17/G, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France

Orientia tsutsugamushi is the causative agent of scrub typhus, a major cause of febrile illness in rural areas of Asia-Pacific region. A multi-locus sequence typing (MLST) analysis was performed on strains isolated from human patients from 3 countries in Southeast Asia: Cambodia, Vietnam and Thailand. The phylogeny of the 56-kDa protein encoding gene was analyzed on the same strains and showed a structured topology with genetically distinct clusters. MLST analysis did not lead to the same conclusion. DNA polymorphism and phylogeny of individual gene loci indicated a significant level of recombination and genetic diversity whereas the ST distribution indicated the presence of isolated patches. No correlation was found with the geographic origin. This work suggests that weak divergence in core genome and ancestral haplotypes are maintained by permanent recombination in mites while the 56-kDa protein gene is diverging in higher speed due to selection by the mammalian immune system.

1. Introduction

Orientia tsutsugamushi is an obligate intracellular bacterium closely related to the genus *Rickettsia* (α -proteobacteria) (Paris et al., 2013). It is the causative agent of scrub typhus and is transmitted to humans by the bite of trombiculid mites of the genus *Leptotrombidium* (Watt and Parola, 2003). The endemic areas of scrub typhus include rural areas of South-East Asia throughout the Asia Pacific rim and Northern Australia. More than a billion people are believed to be potentially exposed to scrub typhus and about one million cases occur annually (Paris et al., 2013). The clinical manifestations of scrub typhus are very variable, ranging from asymptomatic

infection to deadly acute fever. Late diagnosis is the main problem and subsequent late treatment can lead to severe organ failure associated with a mortality rate of up to 50% (Chattopadhyay and Richards, 2007). This frequent disease is a cause of febrile illness that is largely under-recognized. *O. tsutsugamushi* is characterized by a wide antigenic diversity, and isolates are conventionally classified by reactivity with Karp, Kato, and Gilliam antisera.

Rodents and other small mammals are suspected to be natural hosts (Coleman et al., 2003) where recombination occurs through multiple infections (Duong et al., 2013a). Trombiculid mite larvae usually acquire the pathogen vertically from the mother and after hatching, they feed on only one host without moving from host-to-host. Horizontal transmission was demonstrated on naïve chigger feeding on previously infected rodents (Traub et al., 1975). It was also reported to occur between mites co-feeding on infected rodents (Frances et al., 2000). Chigger mites thus infect their host

* Corresponding author at: University Montpellier 2, CPBS, UMR 5554, CBRS-UM1-UM2, 1919 route de Mende, 34293 Montpellier Cedex 5, France.

E-mail address: roger.frutos@univ-montp2.fr (R. Frutos).

¹ Authors contributed equally.

with the strain they bear and could also become infected with *O. tsutsugamushi* strains infecting previously this same host (Takahashi et al., 1990). Humans are accidental hosts infected when bitten by trombiculid larvae and evolutionary dead-end for the pathogen since they do not allow to finalize the parasitic life-cycle (Duong et al., 2013a).

The genome of *O. tsutsugamushi* displays unique traits including an unusually high amount of repeats, transposons and conjugative elements (Cho et al., 2007; Nakayama et al., 2008, 2010). *O. tsutsugamushi* is characterized by an extremely high number of duplicated gene sequences accounting for almost half of the genome (Nakayama et al., 2008). Horizontal transfer of gene sequences with other species was also demonstrated (Nakayama et al., 2010). Another key aspect of the genomic plasticity is the frequent occurrence of recombination and rearrangement (Duong et al., 2013a; Sonthayanon et al., 2010). Altogether, these plasticity mechanisms participate to host-driven selection, adaptation and evasion of host-defenses (Nakayama et al., 2008, 2010).

The 56-kDa protein, located on the outer membrane surface, is unique to *O. tsutsugamushi*. This protein is dedicated to neutralizing host antibody response and is also involved in cell invasion through specific binding to fibronectin. The 56-kDa TSA is therefore interacting directly with the mammal host, i.e. rodent, and is subjected to host-driven adaptive selection (Enatsu et al., 1999).

Consequently, five major serotypes were defined, namely Boryong, Gilliam, Karp, Kato and Kawazaki (Tamura et al., 1995) along with five serologically unique strains from Thailand: TA678, TA686, TA716, TA763 and TH1817 (Elisberg et al., 1967). More recently, molecular methods have been developed to target the 56-kDa type specific antigen (TSA) (Ohashi et al., 1990; Stover et al., 1990a,b). Kelly et al. (2009) reported the existence of at least 9 significant clusters: Karp-related, Saitama, Kuroki, TA763, Gilliam, Kawasaki, JG-related (Gilliam type in Japan), Kato and a group of various strains which could be the result of sequencing errors.

Previous works have focused either on MLST analysis or 56-kDa TSA molecular epidemiology in a specific country and no report addressed altogether polymorphism, evolution and phylogeny of *O. tsutsugamushi* at the regional scale. We thus undertook a global analysis on previously characterized strains and newly isolated samples using a comprehensive approach combining MLST markers, DNA polymorphism analysis and 56-kDa TSA phylogenesis to analyze with the same set of tools *O. tsutsugamushi* strains isolated from hospitalized patients diagnosed with scrub typhus in Thailand, Cambodia and Vietnam. We report that the same haplotypes are found throughout South East Asia. We report that DNA polymorphism and phylogeny of individual MLST loci indicated a significant level of recombination and genetic diversity with the presence of the same alleles throughout the same region and no significant clustering. We also report that the 56-kDa phylogenetic analysis showed in contrary the presence of strongly clustered and genetically isolated populations. The conclusion from this work is that extensive recombination process specific to *O. tsutsugamushi* most likely occurs in mites and pathogens populations are further selected depending on the compatibility of the 56-kDa protein with the rodent host. The mites therefore contribute to maintain a genomic stasis and similar haplotypes throughout the region regardless of the 56-kDa host specificity.

2. Materials and methods

2.1. Collection of patient samples

Samples were obtained from patients admitted in hospital with symptoms of febrile illness associated with headache during the year 2000 to 2010 in Cambodia, Vietnam and Thailand. Cambodia

samples (13 samples) were eschar biopsies and/or whole blood samples anticoagulated with EDTA collected from suspect patients admitted to Calmette hospital and Kuntha Bopha hospital between 2008 and 2010. Vietnam samples (7 samples) consist in blood samples obtained from dengue suspect patients who tested negative for dengue infection and from patients suspect for scrub typhus infection. Thailand samples (19 samples) were collected from 266 patients with laboratory confirmed scrub typhus in 6 hospitals (3 in north, 2 in center and 1 in south) between January 2000 and August 2009. Location of sample origin are shown in Supplementary Fig. 2 and Supplementary Table 3. The clinical study in Cambodia and Vietnam was approved by the Cambodian National Ethics Committee for Health Research, the Ethical Committee board of the Institut Pasteur of Nha Trang, and the Institutional Review Board of the Institut Pasteur, Paris, France. Samples from Thailand were a part of the clinical study approved by the Ethical Review Subcommittee of the Public Health Ministry of Thailand, and Siriraj Institutional Review Board Faculty of Medicine Siriraj Hospital, Mahidol University. Patients provided informed consent before sample collection.

2.2. Reference strains and sequences

Five reference strains were used in this work along with those isolated in Southeast Asia. Two of these strains are complete genomes: *O. Tsutsugamushi* strains Boryong (NC 009488) isolated in Korea (Cho et al., 2007) and *O. Tsutsugamushi* strain Ikeda (NC 010793) from Japan (Nakayama et al., 2010). The three other strains are named Karp (AY956315), Kato (M63382) and Gilliam (DQ485289). Information of the time and place of isolation of the reference strains was collected from Kelly et al. (2009). All sequences have been deposited in Genbank 8 separated sets corresponding to each gene addressed in the work. Accession numbers for each set of sequences are: KJ742066 to KJ742104, KJ742105 to KJ742143, KJ742144 to KJ742182, KJ742183 to KJ742221, KJ742222 to KJ742260, KJ742261 to KJ742299, KJ742300 to KJ742338 and KJ742339 to KJ742371 for *adk*, *lepb*, *lipa*, *lipb*, *secY*, *sodB*, *suca* and the 56-kDa protein gene, respectively. All sequences are provided in a Fasta format in Supplementary Table 1.

2.3. Detection of *O. tsutsugamushi* in clinical samples

In Cambodia and Vietnam DNA was extracted from blood clot, buffy coat or eschar biopsies using QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions and analyzed by real time PCR targeting the chaperonin gene *groEL* (Paris et al., 2009). For Thai samples, genomic DNA was extracted from buffy coat using the minipreparation of bacterial genomic DNA method (Ausubel et al., 2002). EDTA blood sample was centrifuged and then buffy coat was collected. The contamination of red blood cells was eliminated by lysis with distilled water. Chromosomal DNA of *O. tsutsugamushi* was extracted from buffy coat by mixing the cell suspension with TE buffer (pH 8.0). Specific extraction and PCR detection procedures were previously described (Duong et al., 2013a,b; Wongprompitak et al., 2013).

2.4. Selection of MLST target loci and PCR reactions

The 7 MLST loci, i.e. *adk*, *lepb*, *lipA*, *lipB*, *secY*, *sodB* and *sucA* (Supplementary Table 2) were selected on the basis of nucleotide analysis of whole genome sequences of 2 reference strains, Boryong and Ikeda. Primer design and PCR amplification of MLST targets were conducted as previously described (Duong et al., 2013a). The 56-kDa gene was amplified as previously described (Wongprompitak et al., 2013).

2.5. DNA sequencing

Amplicons from Cambodian and Vietnamese samples were sequenced in an ABI 3730XL Analyzer (96 capillary type) using the ABI Prism BigDye™ Terminator Cycle Sequencing (Macrogen, Seoul, South Korea) whereas amplicons from Thailand samples were analyzed by gel electrophoresis and submitted to direct DNA sequencing using the BigDye® Terminator v3.1 cycle sequencing (1st BASE company, Malaysia).

2.6. Sequence alignment and phylogenetic analysis

For individual genes phylogenies, alignments were conducted with Muscle, in SeaView 4.2.6 (Gouy et al., 2010). JModelTest (Posada, 2008) was used to select the optimal evolution model by evaluation of the selected parameters using the Akaike Information Criterion (AIC). This approach suggested the model TPM1uf+G for *adk* and *sucA*; HKY+I for *lepB* and *sodB*; TPM1uf+I for *lipA*; Trn+G for *lipB*; and GTR+G for *secY*. Under the selected models, ML analyses were performed with phyML (version 2.4.4) (Guindon and Gascuel, 2003). The robustness of nodes was assessed with 500 bootstrap replicates.

2.7. MLST analysis

Identical DNA sequences at a given locus between strains were arbitrarily coded with the same allele number (i.e. each allele or haplotype was given a unique identifier), following MLST conventions (Urwin and Maiden, 2003). Each unique allelic profile was assigned a Sequence Type or ST (a unique series of allele or haplotype, identifiers). The reference isolates for sequence type (ST) determination in this work was the Cambodian isolate, S005 (Duong et al., 2013b). MLST data were used to construct dendrogram by cluster analysis using the unweighted pair group method with arithmetic mean (UPGMA) in the ClonalFrame software package (Didelot and Falush, 2007). Consensus MLST tree was built from combined data from three independent runs with 75% majority rule consensus for inference relatedness (program was run with 50,000 burn-in followed by 50,000 subsequent iterations). Index of association (IA) (Maynard Smith et al., 1993) was calculated using the LIAN software program, v3.5, significant values were determined using the Monte-Carlo method with 1000 resamplings.

2.8. Genetic and DNA polymorphism analyses

Genetic similarity and difference matrices were constructed using Muscle in SeaView 4.2.6 (Gouy et al., 2010). DNA sequence polymorphism, subsequent tests and statistical tests were performed using functions from the DnaSP 5.10.01 package (Librado and Rozas, 2009). The numbers of haplotype (alleles) were analyzed. The DNA polymorphism was investigated by calculation of several measures. Nucleotide diversity, π (π), the average number of nucleotide differences per site between two sequences was calculated, using the Jukes and Cantor correction. Theta (Watterson's mutation parameter) was calculated for the whole sequence from S . η is the total number of mutations, and S is the number of polymorphic sites. K_a (the number of non-synonymous substitutions per non-synonymous site), and K_s (the number of synonymous substitutions per synonymous site) for any pair of sequences were calculated. Tajima's D test was used for statistical testing the hypothesis that all mutations are selectively neutral. Other test, the statistical tests D^* and F^* were used for testing the hypothesis that all mutations are selectively neutral.

2.9. Most recent common ancestor

The analysis was conducted separately on the concatenated MLST loci and on the 56-kDa protein gene sequences using BEAST v 1.6 (Drummond and Rambaut, 2007). Several models were used for each set of sequences (Table 3 and Table 4) with different substitution rate for each codon position (Shapiro et al., 2006) and both strict and relaxed (uncorrelated lognormal) molecular clocks. All Markov Chain Monte Carlo chains were run for sufficient time at 100 million steps to ensure statistical convergence, with 10% removed as burn-in. The degree of uncertainty in each parameter estimate is provided by the 95% highest posterior density (HPD) values, while posterior probability values provide an assessment of the degree of support for each node on the tree.

3. Results

3.1. Identification and analyze of sequence types (ST)

A total of 44 strains of *O. tsutsugamushi* (13 from Cambodia, 7 from Vietnam, 19 from Thailand and 5 reference isolates from Genbank) were identified (Table 1). Each locus was characterized by a

Table 1
Sequence types (ST) of *O. tsutsugamushi* isolated from patient samples.

Strains	MLST loci						
	<i>adk</i>	<i>lepB</i>	<i>lipA</i>	<i>lipB</i>	<i>secY</i>	<i>sodB</i>	<i>sucA</i>
S005	1	1	1	1	1	1	1
S009	2	2	2	2	2	2	2
S072	3	3	2	3	2	3	3
S210	4	4	3	4	3	4	4
S257	5	5	4	5	2	5	5
S358	6	6	2	6	2	2	4
New	7	7	2	7	3	6	4
U0720315	13	6	18	7	17	12	4
U0809161	6	18	2	2	2	2	5
U0820321	6	21	2	2	2	2	5
U0902310	12	19	19	15	16	22	14
U1012321	16	20	17	4	2	24	4
U1014302	6	6	2	2	2	23	16
42BVKH	13	24	11	7	17	28	4
47BVKH	16	6	21	4	2	18	17
47QNg	18	2	2	4	18	26	4
49QNg	17	23	10	16	2	25	4
54BVKH	13	7	11	7	3	12	18
91BVKH	13	7	11	7	3	31	19
92BVKH	18	22	20	7	19	27	4
TH2011	6	15	22	10	20	12	4
TH2033	6	12	6	10	8	12	4
TH2037	11	13	7	7	3	12	10
TH2043	11	14	8	11	9	13	11
TH2059	19	13	6	17	5	29	20
TH2066	12	14	10	18	11	29	4
TH2082	11	15	9	11	10	13	11
TH2085	13	25	23	19	3	25	10
TH2190	12	14	10	12	11	14	5
TH2191	11	7	11	7	3	12	12
TH2208	11	12	12	11	10	15	11
TH2221	11	15	9	11	9	15	11
TH3007	11	13	7	7	2	16	12
TH3029	13	16	13	4	12	17	13
TH4005	14	14	11	13	13	18	14
TH4006	15	17	14	1	14	19	13
TH4007	13	14	15	4	3	20	15
TH4016	12	26	24	4	21	30	21
TH4024	12	6	16	14	15	21	12
Otb	9	8	5	1	4	8	7
Oti	8	9	6	8	5	7	6
Otgm	10	10	6	4	7	9	8
Otkp	8	11	6	9	6	11	4
Otkt	8	9	6	8	5	10	9
N* of Haplotypes	19	26	24	19	21	31	21

Table 2
DNA polymorphism and neutrality tests of MLST loci of *O. tsutsugamushi* isolated from patient samples.

Locus	Hp	S	η	$\eta(S)$	Pa	θ	π	Na	Ns	K_a	K_s	K_a/K_s	Tajima's <i>D</i>	D^*	F^*
<i>Adk</i>	19	28	31	18	10	7.12644	0.00956	17	10	0.00606	0.02414	0.251	-1.59931 ^{NS}	-3.41374**	-3.30358**
<i>lepB</i>	26	37	41	13	24	9.42529	0.02139	11	30	0.00363	0.09110	0.040	-0.32835 ^{NS}	-1.24701 ^{NS}	-1.09754 ^{NS}
<i>lipA</i>	24	37	41	26	11	9.42529	0.00895	11	23	0.00463	0.02333	0.199	-2.12057*	-3.79773**	-3.81262**
<i>lipB</i>	19	22	22	9	13	5.05747	0.00897	11	11	0.00537	0.02153	0.250	-1.33524 ^{NS}	-1.26122 ^{NS}	-1.52422 ^{NS}
<i>secY</i>	21	27	29	12	15	6.66667	0.01074	7	19	0.00175	0.03959	0.044	-1.02707 ^{NS}	-1.88993 ^{NS}	-1.88800 ^{NS}
<i>sodB</i>	31	38	40	15	23	9.19541	0.01723	12	23	0.00757	0.05727	0.132	-1.04306 ^{NS}	-1.52156 ^{NS}	-1.60871 ^{NS}
<i>sucA</i>	21	25	26	16	9	5.97701	0.00574	14	12	0.00233	0.01887	0.124	-2.07465*	-3.13552*	-3.28528**

Hp: number of haplotypes; S: polymorphic sites; η : total number of mutations; $\eta(s)$: number of singletons; Pa: parsimony informative sites; θ : Watterson's mutation parameter (per sequence calculated from); π : nucleotide diversity; Na: number of non-synonymous substitutions; Ns: number of synonymous substitutions; k_a : rate of non-synonymous substitutions; k_s : rate of synonymous substitutions; *Significant at $p < 0.05$; **Significant at $p < 0.02$; NS: not significant. Tajima's *D*: Tajima's *D* test for neutrality; D^* and F^* , Fu and Li's D^* and F^* test for neutrality.

Table 3
Estimated age of the most recent common ancestor for the concatenated MLST loci.

Model	Age mean value ^a	ESS ^b	95% HPD limits ^c	Clock rate (mean)
Relax, constant, HKY	140.964	364.755	25.529–235.776	1.57×10^{-4}
Strict, constant, HKY	316.741	584.904	34.679–793.026	9.19×10^{-5}
Relax, BSP, GTR	361.57	1189.612	25.722–1098.090	8.01×10^{-5}
Relax, BSP, GTR+g4	306.726	114.373	20.788–1131.545	1.12×10^{-4}
Relax, BSP, HKY+g4	508.695	94.806	30.779–2459.947	9.68×10^{-5}
Strict, constant, GTR+g4	145.853	953.683	24.815–271.296	1.98×10^{-4}
Strict, constant, HKY+g4	360.987	543.354	27.523–493.609	1.74×10^{-4}

^a Value is given in years.

^b ESS = effective sample size.

^c HPD: higher posterior density.

Table 4
Estimated age of the most recent common ancestor for the 56-kDa protein gene.

Model	Age mean value ^a	ESS ^b	95% HPD limits ^c	Clock rate (mean)
Relax, constant, HKY	10698.736	435.787	147.918–33810.385	1.40×10^{-4}
Strict, constant, HKY	24426.885	587.934	618.437–72325.069	3.99×10^{-5}
Relax, constant, GTR	27286.289	287.357	230.835–82040.054	6.88×10^{-5}
Strict, constant, GTR	21380.724	643.776	480.960–75770.973	4.03×10^{-5}
Strict, constant, GTR + g4	24156.207	845.462	778.238–83851.991	4.05×10^{-5}
Strict, constant, HKY + g4	37619.574	505.351	773.025–88060.418	4.14×10^{-5}

^a Value is given in years.

^b ESS = effective sample size.

^c HPD: higher posterior density.

relatively high number of haplotypes. Highest haplotypes number (i.e. 31) was displayed by *sodB* gene whereas *adk* and *lipB* displayed the lowest (i.e. 19). No correlation could be established between the number of haplotypes and the number of mutations (η) or the percentage of diversity ($\eta \times 100/\text{size}$) Percentage of diversity is calculated as the total number of mutations (η) times 100, divided by the number of nucleotides in sequences (sequence size). Among the 44 strains studied, each display a unique profile ST (Fig. 1) and no clonal complex could be identified by e-burst analysis (data not shown). The degree of recombination was determined using the index of association (IA) measuring the level of linkage between alleles at different loci. IA value determined for *O. tsutsugamushi* MLST strains was 0.1199 ($p < 0.001$) suggesting a clear clonal population even if no clonal complex are detected.

3.2. Phylogenetic analysis of the 56-kDa TSA

Analyzing the phylogeny of the 56-kDa allowed to determine the existence of 3 well-structured clusters as well on nucleotides (Fig. 2a) as on amino-acids sequences (Fig. 2b). Cluster 1 includes strains from the three countries and the Kato strain reference. Cluster 2 includes Karp and Boryong strain references and strains from the three countries. In addition, cluster 3 is split in two subclusters,

subcluster 2a including strains from the three countries and Karp reference and subcluster 2b including strains from Vietnam and Cambodia. Cluster 3 is closer to the Ikeda reference strain and include strains from Thailand and Cambodia. Four strains are not included in the 3 clusters and are branched separately, Gilliam (OtgM), 47QNg from Vietnam and TH4024 and TH2085 from Thailand. The latter two branched together.

3.3. Phylogenetic analysis of sequenced MLST loci

Owing to the absence of a relevant outgroup each locus was analyzed for phylogenetic trees construction based on midpoint rooting (Supplementary Fig. 1). The number of isolates analyzed for MLST loci phylogeny (44) was slightly different than that of 56-kDa analysis (38). Each gene displayed a different topology but for all genes isolates are associated together by geographic origin within the same clusters. However, the recurrent feature for these phylogenetic trees was the very low values of bootstrapping involving a weak confidence in topologies. A similar phylogenetic analysis was conducted on the concatenated sequences which yielded a tree with a different topology but with the same characteristics of low bootstrap values and association of isolates from all three countries (Fig. 3).

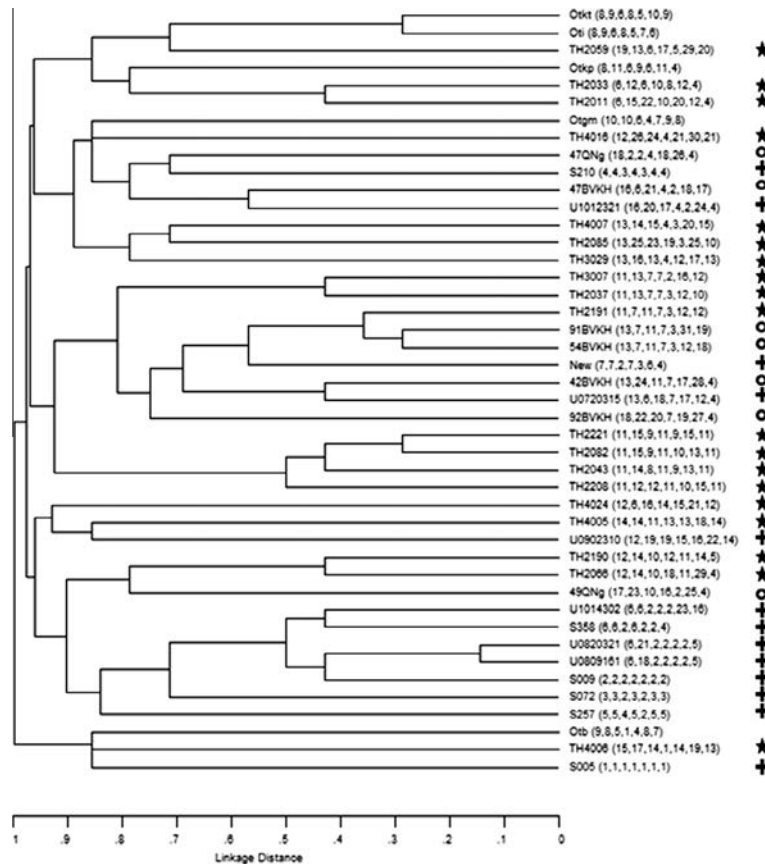


Fig. 1. UPGMA tree of *O. tsutsugamushi* strains ST. The strains showed in this figure compose of the reference strain from GenBank and patient samples isolated from Vietnam (circle), Cambodia (cross) and Thailand (star). Otkb: *O. tsutsugamushi* strain Boryong. Oti: *O. tsutsugamushi* strain Ikeda. Otkm: *O. tsutsugamushi* Gilliam reference type. Otkc: *O. tsutsugamushi* Kato reference type. Otkp: *O. tsutsugamushi* Karp reference type.

3.4. DNA polymorphism and selective pressure

DNA polymorphism data are summarized in Table 2. These data indicate differing trends depending on locus and country of origin. The first noticeable feature is the confirmation of the higher diversity of MLST loci in *O. tsutsugamushi* than in related Rickettsiales, i.e. *Ehrlichia ruminantium* (Adakal et al., 2010). This higher level of diversity previously reported in Cambodia (Duong et al., 2013a) is also observed in Thailand and Vietnam. θ was ranging from 0.265 to 5.487 in *E. ruminantium* depending on locus and population whereas it ranged from 0.816 to 9.345 in *O. tsutsugamushi* depending on locus and country of origin. The mean value of θ was 2.003 for *E. ruminantium* (Adakal et al., 2010) and 4.531 for *O. tsutsugamushi* (this work). This high level of DNA polymorphism in *O. tsutsugamushi* MLST targets is accompanied by low K_a/K_s ratios and mostly non-significant neutrality test, i.e. Tajima's D and F_u and $Li D^*$ and F^* tests indicating that the null hypothesis of selective neutrality could not be rejected. Exceptions however were observed for *adk*, *lipA* and *sucA* for which the null hypothesis could be rejected. However, all tests yielded largely negative values indicative of excess of newly non fixed single mutations and an expanding population. Taken together with the low K_a/K_s ratio and the maintenance of haplotypes this indicates the occurrence of conflicting trends: high mutational rate generating a high polymorphism (and thus different ST) on one hand and negative selective pressure conserving the sequences from a functional standpoint and leading to stasis on another hand. Variations were observed with a higher level of polymorphism in Thai strains for *lipB*, *secY* and *sucA*, in Cambodian strains for *lepB* and *sodB* and a

slightly higher level for Vietnamese strains for *lipA*. For *adk*, Thai and Cambodian strains displayed very close level of polymorphism.

3.5. Age of most recent common ancestor

The age of the most recent common ancestor (MRCA) was estimated for both the concatenated MLST loci (Table 3) and the 56-kDa protein gene (Table 4) using different evolutionary models. MRCA age for MLST loci was estimated between 140.964 years and 508.695 years depending upon the model considered (Table 3) whereas MRCA age for the 56-kDa protein gene was 10698.736 years and 37619.574 years depending upon the evolutionary model (Table 4). The molecular clock rate estimates were in similar ranges for both MLST loci and 56-kDa protein gene. It ranged from 9.19×10^{-5} to 1.98×10^{-4} for MLST loci (Table 3) and from 6.88×10^{-5} to 1.40×10^{-4} for the 56-kDa protein gene (Table 4).

4. Discussion

The present work is the first comparative analysis of MLST genotyping conducted on *O. tsutsugamushi* at the scale of the South East Asian region and compared to a molecular epidemiology analysis based on the 56-kDa protein gene.

The clinical samples collected throughout continental South East Asia displayed an overall pseudo-clonal structure similar to that previously observed at the smaller geographic scale of a single country, i.e. Cambodia (Duong et al., 2013a). Each clinical isolate sequenced is unique and characterized by a unique ST (Duong

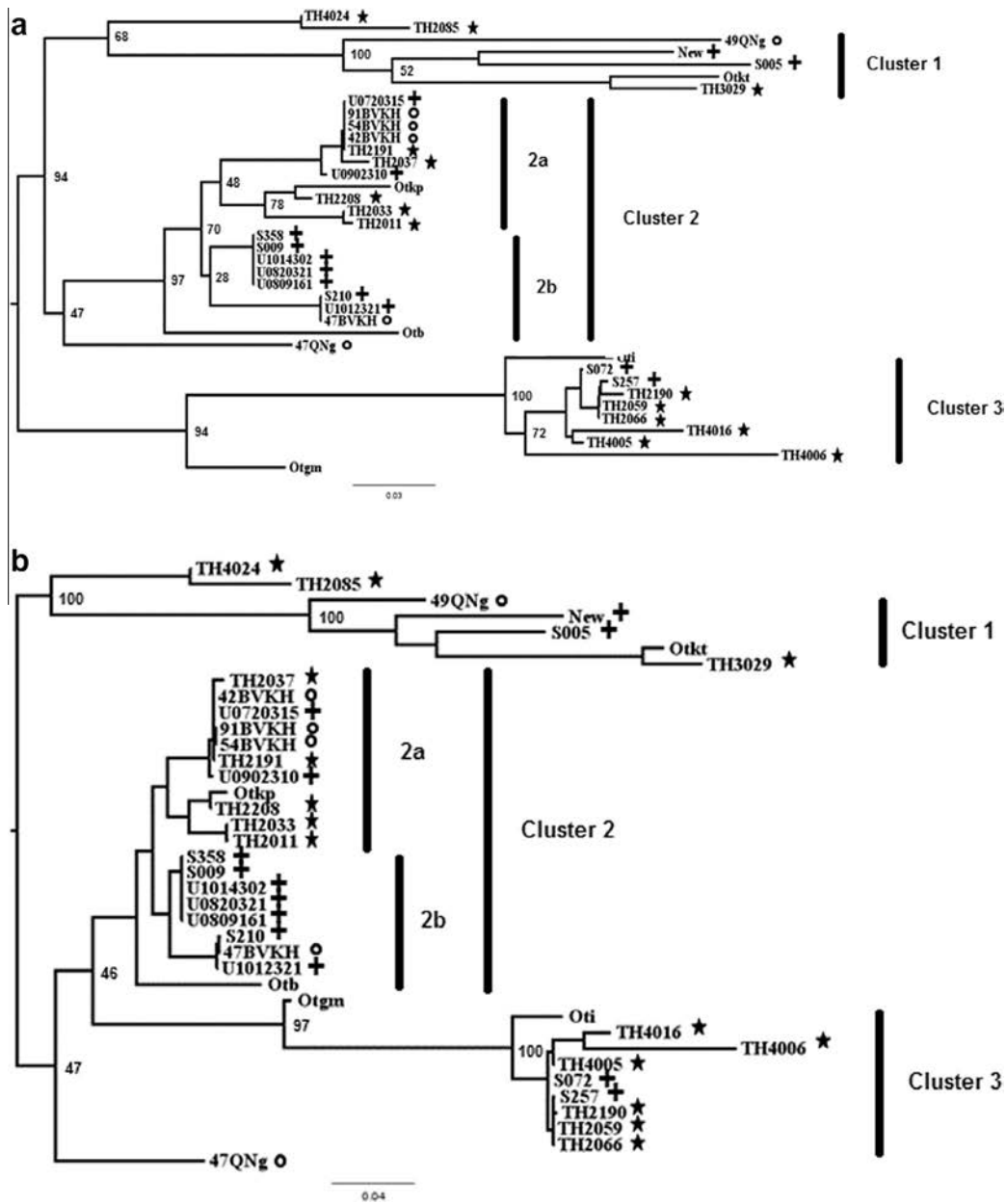


Fig. 2. Phylogenetic trees of 56-kDa protein gene from VD I to VD IV region. The strains showed in this figure compose of the reference strain from GenBank (underlined text) and patient samples isolated from Vietnam (circle), Cambodia (cross) and Thailand (star). (a) Tree from nucleotide sequences (maximum-likelihood, GTR+G Model). (b) Tree from amino acid sequences (distance method). The numbers at nodes demonstrated the bootstrap values. Clusters and subclusters are identified by numbers. Otb: *O. tsutsugamushi* strain Boryong. Oti: *O. tsutsugamushi* strain Ikeda. Otm: *O. tsutsugamushi* Gilliam reference type. Otk: *O. tsutsugamushi* Kato reference type. Otkp: *O. tsutsugamushi* Karp reference type.

et al., 2013a). This apparent clonal organization could be the result of a very high genomic plasticity of *O. tsutsugamushi* and a random infection in humans who are accidental dead-end host (Duong et al., 2013a). Owing to the significant rate of plasticity and polymorphism observed, the combinations of MLST alleles borne by human clinical isolates are only driven by chance and unlikely to be repeated. As a result, clinical isolates act as mountain peaks of the complex *O. tsutsugamushi* genomic landscape and do not reflect the actual *O. tsutsugamushi* population structure. The apparent clonal structure reflected by human clinical isolates is more likely not present in *O. tsutsugamushi* at the scale of the whole species in its true host, the chigger mite, where a more continuous genetic background is expected.

The major discrepancy observed between MLST and the 56-kDa gene indicates indeed divergent selection processes. The 56-kDa protein gene phylogenetic tree is highly structured with clusters associated with high bootstrap values whereas MLST loci phylogenetic trees are characterized by a general lack of structure. An example of this diverging topology is provided the group containing strains, TH2043, TH2221, TH2208 and TH2082. This group is the only relatively structured cluster in the phylogenetic analysis of the concatenated sequences with a bootstrap of 72 (Fig. 3). When considering the well-structured 56-kDa gene phylogenetic tree, strain TH2208 no longer associate with TH2043, TH2221 and TH2082 but instead with the Karp reference strain (Otkp) and strains TH2011 and TH2033.

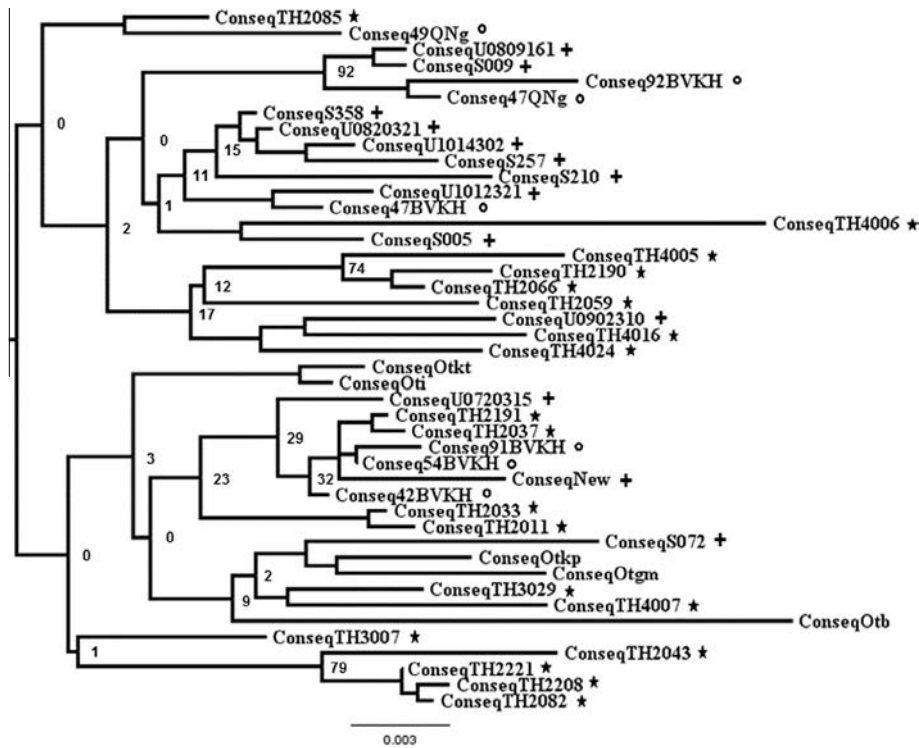


Fig. 3. Maximum-likelihood tree under the GTR+G model of concatenated MLST loci sequences. ML tree was drawn using midpoint rooting. The numbers at nodes demonstrated the bootstrap values. The strains showed in this figure compose of the reference strain from GenBank and patient samples isolated from Vietnam (circle), Cambodia (cross) and Thailand (star). Otkp: Karp reference type of *O. tsutsugamushi*. Otb: *O. tsutsugamushi* strain Boryong. Oti: *O. tsutsugamushi* strain Ikeda. Otgm: Gilliam reference type of *O. tsutsugamushi*. Otkk: Kato reference type of *O. tsutsugamushi*.

Such patterns are expected since MLST genes are selected to be under negative selection (housekeeping genes) whereas the 56-kDa protein acts as the primary immunogen eliciting neutralizing antibodies (Stover et al., 1990a; Seong et al., 1997), interacts with host fibronectin through binding of the antigenic domain III during cell invasion (Lee et al., 2008; Cho et al., 2010). 56-kDa gene clusters are characterized by major deletions indicative of specific structure–function relationship in each cluster and possibly a specific selection in a given host range of mammalian host. This could be thus regarded as a distribution map of the small mammalian hosts capable of maintaining *O. tsutsugamushi* strains displaying human pathogenicity. Indeed, since all strains analyzed in this work were isolated from clinical samples, they were also pathogenic to humans.

In spite of this high genomic plasticity, both MLST markers and 56-kDa gene display a region-wide distribution with the same haplotypes found over the three countries. ST index of association also indicates a common pseudo-clonal population among a large geographic scale. Owing to the distance, presence of mountains and rivers difference and differing environments and climates, between the different countries one would have expected the presence of genetically isolated populations of *O. tsutsugamushi* over South East Asia. The presence of the same alleles over the whole region would thus suggest the occurrence of a very active and massive region-wide movement of hosts, vectors and pathogen capable of maintaining the dispersion of alleles. However, this is very difficult to envision and it is more parsimonious to consider that there might rather be a single metapopulation of *O. tsutsugamushi* within which MLST haplotypes are maintained by negative selective pressure on housekeeping genes in spite of the high recombination rate. The occurrence of the same haplotypes over the whole region is therefore the consequence of this maintenance over time of alleles, i.e.

stasis, rather than that of dynamic diffusion. Incongruities between ages determined for the most recent common ancestor for the concatenated MLST genes (i.e. 140–508 years) and 56-kDa protein gene (i.e. 10698–37619 years) support the hypothesis of a weak divergence in the core genome, i.e. no segregation despite geographic isolation, maintained by permanent recombination in natural hosts (i.e. chigger mites) while the 56-kDa protein gene is diverging at a higher speed owing to selective pressure following regular passages through rodents and small mammalian hosts. The immune response would then select efficient 56-kDa forms for infection regardless of the genomic background. A hypothesis to explain this apparent discrepancy and combine all the results would be the occurrence of recombination in mites for all genes, including the 56-kDa genes, resulting in any possible combination. Because of the lack of congruence between the MLST loci and 56-kDa gene phylogenies, this step is most likely to occur in mites. Indeed, if it occurred in the rodents, the same structure as for the 56-kDa would be expected with most likely a diverging evolution of the MLST loci. This is not observed. Under this hypothesis, following infection of rodents after mite bites, only the strains with the relevant 56-kDa gene would survive in a given group of rodents leading to the maintenance of the associated combination of alleles.

These conclusions represent an expected outcome of adaptation to intracellular parasitism. Following initial genome size reduction and massive loss of genes as part of the adaptation to the intracellular lifestyle (Moran and Mira, 2001; Sallstrom and Andersson, 2005), the genome of cell parasites remain highly stable over long periods of time (Klasson and Andersson, 2004). The maintenance of similar sequences throughout South East Asia, the pseudo clonal dynamic and the high level of recombination make highly difficult the development of reliable genotyping methods. It would thus

sound more appropriate to focus on the improvement of diagnostic tests for early detection in order to administer antibiotic treatment as soon as possible regardless of the strain or genotype encountered.

Acknowledgements

This study was supported in part by a grant from the Institut Pasteur, Paris, France (Actions Concertées Inter-Pasteuriennes). Roger Frutos was supported in part by the Franco-Thai Hubert Curien project PHC 20624VK and by the project ANR Biodiv CERO-Path (ANR 07 BDIV 012). Patimaporn Wongprompitak was supported by ERASMUS MUNDUS Action 2 scholarship through the MAHEVA project.

The authors certify that there is no conflict of interest of any kind associated to this work. There is no source of funding which could potentially affect the outcome and sincerity of the work. There is no personal or institutional conflict with any other institution or bodies which could affect the results and conclusions presented in this article.

References

- Adakal, H., Gavotte, L., Stachurski, F., Konkobo, M., Henri, H., Zoungrana, S., Huber, K., Vachieri, N., Martinez, D., Morand, S., Frutos, R., 2010. Clonal origin of emerging populations of *Ehrlichia ruminantium* in Burkina-Faso. *Infect. Genet. Evol.* 10, 903–912.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K., 2002. *Short Protocols in Molecular Biology*. John Wiley & Sons, New York.
- Chattopadhyay, S., Richards, A.L., 2007. Scrub typhus vaccines: past history and recent developments. *Hum. Vaccines* 3, 73–80.
- Cho, B.A., Cho, N.H., Seong, S.Y., Cho, M.S., Kim, I.S., 2010. Intracellular invasion by *Orientia tsutsugamushi* is mediated by integrin signaling and actin cytoskeleton rearrangements. *Infect. Immun.* 78, 1915–1923.
- Cho, N.H., Kim, H.R., Lee, J.H., Kim, S.Y., Kim, J., Cha, S., Kim, S.Y., Darby, A.C., Fuxelius, H.H., Yin, J., Kim, J.H., Kim, J., Lee, S.J., Koh, Y.S., Jang, W.J., Park, K.H., Andersson, S.G., Choi, M.S., Kim, I.S., 2007. The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7981–7986.
- Coleman, R.E., Monkanna, T., Linthicum, K.J., Strickman, D.A., Frances, S.P., Tanskul, P., Kollars Jr., T.M., Inlao, I., Watcharapichat, P., Khlaimanee, N., Phulsuksombati, D., Sangiun, N., Lerdtusnee, K., 2003. Occurrence of *Orientia tsutsugamushi* in small mammals from Thailand. *Am. J. Trop. Med. Hyg.* 69, 519–524.
- Didelot, X., Falush, D., 2007. Inference of bacterial microevolution using multilocus sequence data. *Genetics* 175, 1251–1266.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Duong, V., Blassdell, K., May, T.T., Sreyrath, L., Gavotte, L., Morand, S., Frutos, R., Buchy, P., 2013a. Diversity of *Orientia tsutsugamushi* clinical isolates in Cambodia reveals active selection and recombination process. *Infect. Genet. Evol.* 15, 25–34.
- Duong, V., Mai, T.T., Blasdel, K., Lo, L.V., Morvan, C., Lay, S., Anukool, W., Wongprompitak, P., Suputtamongkol, Y., Laurent, D., Richner, B., Ra, C., Chien, B.T., Frutos, R., Buchy, P., 2013b. Molecular epidemiology of *Orientia tsutsugamushi* in Cambodia and Central Vietnam reveals a broad region-wide genetic diversity. *Infect. Genet. Evol.* 15, 35–42.
- Elisberg, B.L., Sangkasuvana, V., Campbell, J.M., Bozeman, F.M., Bodhidatta, P., 1967. Physiogeographic distribution of scrub typhus in Thailand. *Acta Med. Biol.* 15, 61–67.
- Enatsu, T., Urakami, H., Tamura, A., 1999. Phylogenetic analysis of *Orientia tsutsugamushi* strains based on the sequence homologies of 56-kDa type-specific antigen genes. *FEMS Microbiol. Lett.* 180, 163–169.
- Frances, S.P., Watcharapichat, P., Phulsuksombati, D., Tanskul, P., 2000. Transmission of *Orientia tsutsugamushi*, the aetiological agent for scrub typhus, to co-feeding mites. *Parasitology* 120, 601–607.
- 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.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Kelly, D.J., Fuerst, P.A., Ching, W.M., Richards, A.L., 2009. Scrub typhus: the geographic distribution of phenotypic and genotypic variants of *Orientia tsutsugamushi*. *Clin. Infect. Dis.* 48, S203–230.
- Klasson, L., Andersson, S.G.E., 2004. Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol.* 12, 37–43.
- Lee, J.H., Cho, N.H., Kim, S.Y., Bang, S.Y., Chu, H., Choi, M.S., Kim, I.S., 2008. Fibronectin facilitates the invasion of *Orientia tsutsugamushi* into host cells through interaction with a 56-kDa type-specific antigen. *J. Infect. Dis.* 198, 250–257.
- Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
- Maynard Smith, J., Smith, N.H., O'Rourke, M.G.S.B., 1993. How clonal are bacteria? *Proc. Natl. Acad. Sci. U.S.A.* 90, 4384–4388.
- Moran, N.A., Mira, A., 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol.* 54, 1–12.
- Nakayama, K., Kurokawa, K., Fukuhara, M., Urakami, H., Yamamoto, S., Yamazaki, K., Ogura, M., Ooka, T., Hayashi, T., 2010. Genome comparison and phylogenetic analysis of *Orientia tsutsugamushi* strains. *DNA Res.* 17, 281–291.
- Nakayama, K., Yamashita, A., Kurokawa, K., Morimoto, T., Ogawa, M., Fukuhara, M., Urakami, H., Ohnishi, M., Uchiyama, I., Ogura, Y., Ooka, T., Oshima, K., Tamura, A., Hattori, M., Hayashi, T., 2008. The Whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. *DNA Res.* 15, 185–199.
- Ohashi, N., Nashimoto, H., Ikeda, H., Tamura, A., 1990. Cloning and sequencing of the gene (tsg56) encoding a type-specific antigen from *Rickettsia tsutsugamushi*. *Gene* 91, 119–122.
- Paris, D.H., Aukkanit, N., Jenjaroen, K., Blacksell, S.D., Day, N.P., 2009. A highly sensitive quantitative real-time PCR assay based on the *groEL* gene of contemporary Thai strains of *Orientia tsutsugamushi*. *Clin. Microbiol. Infect.* 15, 488–495.
- Paris, D.H., Shelite, T.R., Day, N.P., Walker, D.H., 2013. Unresolved problems related to scrub typhus: a seriously neglected life-threatening disease. *Am. J. Trop. Med. Hyg.* 89, 301–307.
- Posada, D., 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256.
- Sallstrom, B., Andersson, S.G., 2005. Genome reduction in the alpha-proteobacteria. *Curr. Opin. Microbiol.* 8, 579–585.
- Seong, S.Y., Huh, M.S., Jang, W.J., Park, S.G., Kim, J.G., Woo, S.G., Choi, M.S., Kim, I.S., Chang, W.H., 1997. Induction of homologous immune response to *Rickettsia tsutsugamushi* Boryong with a partial 56-kilodalton recombinant antigen fused with the maltose-binding protein MBP-Bor56. *Infect. Immun.* 65, 1541–1545.
- Shapiro, B., Rambaut, A., Drummond, A.J., 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* 23, 7–9.
- Sonthayanon, P., Peacock, S.J., Chierakul, W., Wuthiekanun, V., Blacksell, S.D., Holden, M.T., Bentley, S.D., Feil, E.J., Day, N.P., 2010. High rates of homologous recombination in the mite endosymbiont and opportunistic human pathogen *Orientia tsutsugamushi*. *PLoS Neglected Trop. Dis.* 4, e752.
- Stover, C.K., Marana, D.P., Carter, J.M., Roe, B.A., Mardis, E., Oaks, E.V., 1990a. The 56-kilodalton major protein antigen of *Rickettsia tsutsugamushi*: molecular cloning and sequence analysis of the sta56 gene and precise identification of a strain-specific epitope. *Infect. Immun.* 58, 2076–2084.
- Stover, C.K., Marana, D.P., Dasch, G.A., Oaks, E.V., 1990b. Molecular cloning and sequence analysis of the Sta58 major antigen gene of *Rickettsia tsutsugamushi*: sequence homology and antigenic comparison of Sta58 to the 60-kilodalton family of stress proteins. *Infect. Immun.* 58, 1360–1368.
- Takahashi, M., Murata, M., Hori, E., Tanaka, H., Kawamura Jr., A., 1990. Transmission of *Rickettsia tsutsugamushi* from *Apodemus speciosus*, a wild rodent, to larval trombiculid mites during the feeding process. *Jpn. J. Exp. Med.* 60, 203–208.
- Tamura, A., Ohashi, N., Urakami, H., Miyamura, S., 1995. Classification of *Rickettsia tsutsugamushi* in a new genus, *Orientia* gen. nov., as *Orientia tsutsugamushi* comb. nov. *Int. J. Syst. Bacteriol.* 45, 589–591.
- Traub, R., Wisseman Jr., C.L., Jones, M.R., O'Keefe, J.J., 1975. The acquisition of *Rickettsia tsutsugamushi* by chiggers (trombiculid mites) during the feeding process. *Ann. N. Y. Acad. Sci.* 266, 91–114.
- Urwin, R., Maiden, M.C., 2003. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol.* 11, 479–487.
- Watt, G., Parola, P., 2003. Scrub typhus and tropical rickettsioses. *Curr. Opin. Infect. Dis.* 16, 429–436.
- Wongprompitak, P., Anukool, W., Wongsawat, E., Silpasakorn, S., Duong, V., Buchy, P., Morand, S., Frutos, R., Ekpo, P., Suputtamongkol, Y., 2013. Broad-coverage molecular epidemiology of *Orientia tsutsugamushi* in Thailand. *Infect. Genet. Evol.* 15, 53–58.