IDENTIFICATION OF *ORIENTIA TSUTSUGAMUSHI* USING POLYMERASE CHAIN REACTION TECHNIQUE

WICHITTRA ANUKOOL

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE (IMMUNOLOGY)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2011

Thesis
entitled

IDENTIFICATION OF *ORIENTIA TSUTSUGAMUSHI* USING POLYMERASE CHAIN REACTION TECHNIQUE

……..
Mrs. Wichittra Anukool
Candidate

……..
Assoc.Prof. Pattama Ekpo,
Ph.D.
Major advisor

……..
Prof. Yupin Suputtamongkol,
M.D.
Co-advisor

……..
Lect. Duangjit Kanistanon,
Ph.D., M.D.
Co-advisor

……..
Ms. Patimaporn Wongprompitak,
Ph.D.
Co-advisor

……..
Prof. Banchong Mahaisavariya,
M.D., Dip Thai Board of Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

……..
Prof. Kovit Pattanapunyasat,
Ph.D.
Program Director
Master of Science Program in Immunology
Faculty of Medicine Siriraj Hospital
Mahidol University

Thesis
entitled

IDENTIFICATION OF *ORIENTIA TSUTSUGAMUSHI* USING POLYMERASE CHAIN REACTION TECHNIQUE

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Immunology)
on
February 14, 2011

……..
Mrs. Wichittra Anukool
Candidate

……..
Prof. Roger Frutos,
Ph.D., H.D.R.
Chair

……..	……..
Assoc.Prof. Pattama Ekpo, Ph.D. Member	Prof. Yupin Suputtamongkol, M.D. Member
……..	……..
Lect. Duangjit Kanistanon, Ph.D., M.D. Member	Ms. Patimaporn Wongprompitak, Ph.D. Member
……..	……..
Prof. Banchong Mahaisavariya, M.D., Dip Thai Board of Orthopedics Dean Faculty of Graduate Studies Mahidol University	Prof. Teerawat Kulthanan, M.D. Dean Faculty of Medicine Siriraj Hospital Mahidol University

ACKNOWLEDGEMENTS

First of all, I would like to thank Assoc. Prof. Dr. Pattama Ekpo, my major advisor. It has been an honor to be her student. She has taught me both academic and non-academic knowledge. I also wish to express my appreciation to Prof. Yupin Suputtamongkol. More than funding for my thesis, she provided many precious comments and enlightens the epidemiology of disease to me. Sincere thanks are extended to Dr. Duangjit Kanistanon, on behalf of devoting her precious time and made many valuable suggestions which helped improve especially in PCR parts.

It is an honor for me to acknowledge Dr. Roger Frutos, my thesis chairman, for more understanding in the phylogeny. He patiently answered all my questions and gave me the productive comments for my results.

My eternal gratitude goes to Dr. Patimaporn Wongprompitak, my special mentor. She trained me in both my laboratory and soul. She gave me the precious experienced and also brightened up days and provided some motivation.

Many thanks also to colleagues in the Laboratory of Infectious disease, Siriraj Hospital and also wish to thank my friends in Department of Immunology for helping me get through the difficult times, and for all the emotional support.

I am also indebted to my colleagues at the Veterinary Research and Development Center (Upper northern zone), Department of Livestock Development, my original affiliation, for their helpful while I was in the educational leave.

Finally, I thank my parents, my brothers and my husband for their loving support throughout all my studies at University. Without their encouragement and understanding it would have been impossible for me to finish this work.

During the period of this study, I was supported by the Siriraj Graduate Scholarships for the first year and the Siriraj Graduate Thesis Scholarships for the second year and later.

Wichittra Anukool

IDENTIFICATION OF *ORIENTIA TSUTSUGAMUSHI* USING POLYMERASE CHAIN REACTION TECHNIQUE

WICHITTRA ANUKOOL 5137078 SIIM/M

M.Sc. (IMMUNOLOGY)

THESIS ADVISORY COMMITTEE: PATTAMA EKPO, Ph.D.,
YUPIN SUPUTTAMONGKOL, M.D., DUANGJIT KANISTANON, Ph.D., M.D., PATIMAPORN WONGPROMPITAK, Ph.D.

ABSTRACT

Orientia tsutsugamushi is the causative agent of scrub typhus, a major cause of undifferentiated fever in the Asia-Pacific region, especially Thailand. The genotypic characterization of *O. tsutsugamushi* was identified based on a partial nucleotide sequence of the 56-kDa protein-encoding gene which contained the variable domains (VD) I-IV. There were 106 scrub typhus suspected patients whose DNA were amplified by PCR technique using specific primers for the 56-kDa protein encoding gene. These samples were analyzed by nucleotide sequencing and construction of a dendrogram. Samples were identified into 4 clusters: Karp (44.7%), Kato (3.9%), Gilliam (40.8%), and TA763 (10.5%). Only 2 samples could not be grouped in any cluster. The geographical distribution of the *O. tsutsugamushi* strain in Thailand was studied. However, there were no samples from northern and eastern regions. Karp was the predominant strain in the northeastern region while Gilliam was the dominant strain in the central region as well as in the southern. From this information, the TA763 strain should be included in the panel of antigens used for antibody detection. In addition, strain-specific primers were developed based on the 56-kDa protein encoding gene of the sequences in each cluster to identify *O. tsutsugamushi* strains. Karp- and Gilliam-specific primers generated satisfactory sensitivity and specificity values with approximately 95%. The information obtained from *Orientia* strain identification can be used for monitoring the prevalence of the disease and can also be useful for further vaccine development to prevent *O. tsutsugamushi* infection.

KEY WORDS: *Orientia tsutsugamushi*/ POLYMERASE CHAIN REACTION/ 56-kDa PROTEIN ENCODING GENE

115 pages

การจำแนกเชื้อ *ORIENTIA TSUTSUGAMUSHI* ด้วยเทคนิคปฏิกิริยาลูกโซ่โพลีเมอเรส

IDENTIFICATION OF *ORIENTIA TSUTSUGAMUSHI* USING POLYMERASE CHAIN REACTION TECHNIQUE

วิจิตรา อนุกูล 5137078 SIIM/M

วท.ม. (วิทยาภูมิคุ้มกัน)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์: ปัทมา เอกโพธิ์, Ph.D., ยุพิน ศุพุทธมงคล, M.D., ดวงจิต กนิษฐานนท์, Ph.D., M.D., ปฏิมาพร วงษ์พรหมพิทักษ์, Ph.D.

บทคัดย่อ

Orientia tsutsugamushi เป็นเชื้อก่อโรคสครับ ไทฟัส ซึ่งเป็นสาเหตุหนึ่งของกลุ่มอาการไข้ไม่ทราบสาเหตุในภาคพื้นเอเชียแปซิฟิก โดยเฉพาะในประเทศไทย เชื้อนี้ถูกจำแนกลักษณะพันธุกรรมด้วย variable domain I-IV ของ 56-kDa protein encoding gene ในการศึกษานี้ได้ทำการเพิ่มจำนวนสารพันธุกรรมของ 56-kDa protein encoding gene ตัวอย่างจากผู้ป่วยที่ได้รับการวินิจฉัยว่าเป็นโรคสครับ ไทฟัส จำนวน 106 ตัวอย่าง และนำมาศึกษาลำดับเบสเพื่อหาความสัมพันธ์ทางพันธุกรรมโดยสร้างเป็น dendrogram พบว่าตัวอย่างเหล่านี้สามารถแบ่งได้เป็น 4 กลุ่มตามสายพันธุ์คือ Karp Kato Gilliam และ TA763 คิดเป็น 44.7 3.9 40.8 และ 10.5% มีเพียง 2 ตัวอย่างที่ไม่สามารถจัดกลุ่มได้ จากการศึกษาการกระจายตัวของสายพันธุ์ตามภูมิภาคต่างๆของประเทศไทย ยกเว้นภาคเหนือและตะวันออก พบว่า สายพันธุ์ Karp พบมากที่สุดในภาคตะวันออกเฉียงเหนือ ในขณะที่สายพันธุ์ Gilliam พบมากในภาคกลางและภาคใต้ จากผลการศึกษานี้ เห็นว่าในการทดสอบปฏิกิริยาทางภูมิคุ้มกันเพื่อตรวจหาเชื้อ *O. tsutsugamushi* ควรมีการเพิ่ม antigen ของสายพันธุ์ TA763 ด้วยเนื่องจากพบการระบาดของสายพันธุ์นี้ในประเทศไทย นอกจากนี้ ได้มีการพัฒนาการจำแนกสายพันธุ์ของเชื้อก่อโรคสครับ ไทฟัสด้วยวิธี PCR โดยใช้ primer ที่ออกแบบมาจากลำดับเบสของ 56-kDa protein encoding gene แล้วทำการประเมินการทดสอบโดยหาค่าความไวและความจำเพาะของ primer ดังกล่าว พบว่า primer ที่จำเพาะต่อสายพันธุ์ Karp และ Gilliam มีค่าความไวและความจำเพาะมากกว่า 95% ผลจากการศึกษาและการพัฒนาการจำแนกสายพันธุ์ของเชื้อก่อโรคสครับ ไทฟัส สามารถนำไปพัฒนาและนำไปใช้เพื่อเป็นประโยชน์ในการเฝ้าระวังโรคและศึกษาด้านระบาดวิทยาของเชื้อ รวมทั้งสามารถใช้เป็นข้อมูลเบื้องต้นในการพัฒนาวัคซีนเพื่อป้องกันการติดเชื้อดังกล่าวด้วย

115 หน้า

CONTENTS

	Page
ACKNOWLEDGEMENTS	**iii**
ABSTRACT (ENGLISH)	**iv**
ABSTRACT (THAI)	**v**
LIST OF TABLES	**ix**
LIST OF FIGURES	**x**
LIST OF ABBREVIATIONS	**xi**
CHAPTER I INTRODUCTION	**1**
CHAPTER II LITERATURE REVIEW	**4**
1. *Orientia tsutsugamushi*	4
2. Scrub typhus	6
3. The major antigenic protein of *Orientia tsutsugamushi*	10
3.1 Heat shock protein	12
3.2 Surface proteins	12
3.2.1 The 47-kDa protein	12
3.2.2 The 56-kDa protein	13
3.2.3 The 70-kDa protein	15
3.2.4 The 110-kDa protein	15
4. *Rickettsia* identification based on molecular biology	15
4.1 Phenotypic approaches	16
4.1.1 Serotyping	16
4.1.2 Protein analysis	16
4.2 Genotypic approaches	16
4.2.1 Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP)	16
4.2.2 Pulsed field gel electrophoresis (PFGE)	17

CONTENTS (cont.)

	Page
4.2.3 PCR amplification and direct sequencing	17
4.2.3.1 16S rRNA gene	17
4.2.3.2 Citrate synthase gene or *gltA* gene	18
4.2.3.3 rOmpA and rOmpB protein encoding genes	18
4.2.3.4 56-kDa protein encoding gene	19
CHAPTER III MATERIALS AND METHODS	**21**
1. Patient samples	21
1.1 EDTA blood samples	23
1.2 Chromosomal DNA samples	23
2. Chromosomal DNA preparation	23
3. Quantification of chromosomal DNA	24
4. Amplification of 56-kDa protein encoding gene of *O. tsutsugamushi*	24
4.1 Primer design	24
4.2 Nested polymerase chain reaction	25
5. Detection of PCR products by agarose gel electrophoresis	25
6. DNA sequencing	27
6.1 Preparation of DNA template	27
6.2 DNA purification as a template for automated DNA Sequencing	28
6.3 Automated DNA sequencing using BigDye^{TM} Teminator	28
7. Analysis of nucleotide sequencing data	29
7.1 Editing and assembling of nucleotide sequences	29
7.2 Multiple nucleotide sequences alignment	30
8. Phylogenetic analysis	30

CONTENTS (cont.)

	Page
CHAPTER IV RESULTS	**32**
1. Amplification of 56-kDa protein encoding gene of *O. tsutsugamushi*	32
2. Phylogenetic analysis of partial nucleotide sequences of 56-kDa protein encoding gene of *O. tsutsugamushi*	32
2.1 Multiple sequence alignment and phylogenetic analysis	32
2.2 Geographical distribution	60
3. Strain-specific amplification of *O. tsutsugamushi*	60
CHAPTER V DISCUSSION	**74**
CHAPTER VI CONCLUSION	**80**
REFERENCES	**82**
APPENDICES	**91**
Appendix A Chemical reagents	91
Appendix B Multiple alignment of reference strains of *O. tsutsugamushi* for primer selection	94
Appendix C Nucleotide sequences alignment to select the conserved regions for *O. tsutsugamushi* strain-specific amplification	98
BIOGRAPHY	**113**

LIST OF TABLES

Table		Page
1	The information of patient samples in this study	22
2	Oligonucleotide primers for nested PCR amplification of 56-kDa protein encoding gene	26
3	Description of prototype and reference strains of *O. tsutsugamushi* used in multiple alignments	31
4	The positive results by IFA technique comparing to the amplification of 56-kDa protein encoding gene	34
5	The sequence identity matrix of 78 nucleotide sequence which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database	35
6	The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database	46
7	The geographic distribution of *O. tsutsugamushi* strains in Thailand	62
8	Oligonucleotide primers for nested PCR amplification of 56-kDa protein encoding gene of each strain	65
9	The optimal amount of $MgCl_2$ and optimal annealing temperature of each set of primer for *O. tsutsugamushi* strain-specific amplification	66
10	The evaluation of Karp-specific primers in comparison to strain identification using dendrogram	70
11	The evaluation of Kato-specific primers in comparison to strain identification using dendrogram	71
12	The evaluation of Gilliam-specific primers in comparison to strain identification using dendrogram	72
13	The evaluation of TA763-specific primers in comparison to strain identification using dendrogram	73

LIST OF FIGURES

Figure		Page
1	The epidemiological cycle of scrub typhus	5
2	The tsutsugamushi triangle	8
3	Reported cases of scrub typhus per 100,000 populations in Thailand between 2000-2009	9
4	A map of Thailand demonstrated the reported cases of scrub typhus per 100,000 populations by province in 2009	11
5	The multiple alignment of amino acid sequences of 56-kDa protein from 6 reference strains (Gilliam, Karp, Kato, Kawasaki, Kuroki and Shimokoshi)	14
6	Agarose gel electrophoresis of PCR product amplified using the 56-kDa protein encoding gene primers	33
7	A dendrogram constructed from the alignment of VD I to IV regions of the 56-kDa protein encoding gene based on the N-J algorithm with 1,000 bootstrap values	56
8	Nucleotide comparison of sequence sample no. KP133 using Blastn program	59
9	Nucleotide comparison of sequence sample no. CHUP65 using Blastn program	61
10	Geographical distribution of *O. tsutsugamushi* strains in 4 regions of Thailand	63
11	Agarose gel electrophoresis of PCR product amplification based on nested PCR using strain-specific inner primers for Karp, Kato, Gilliam and TA763	67

LIST OF ABBREVIATIONS

Abbreviation	Term
IH	Incidental host
MH	Maintenance host
MI	Mite island
STG	Scrub typhus group
SFG	Spotted fever group
TG	Typhus group
IFN-γ	Interferon-gamma
LPS	Lipopolysaccharide
MLD	Mice lethal dose
PG	Peptidoglycan
PMN	Polymorphonuclear leukocyte
Th1	T helper 1 cell
DNA	Deoxyribonucleic acid
% G-C content	Percentage of Guanine- Cytosine content
Tm	Melting temperature
MEGA	Molecular Evolutionary Genetics Analysis program
N-J	Neighbor-Joining algorithm
ORF	Open reading frame
VD	Variable domain
CF	Complement fixation
DFA	Direct immunofluorescence
IFA	Indirect immunofluorescence
LAMP-PCR	Loop-mediated isothermal amplification
MIF	Micro-immunofluorescence

LIST OF ABBREVIATIONS (cont.)

Abbreviation	Term
mAb	Monoclonal antibody
NPV	Negative predictive value
PCR	Polymerase chain reaction
PCR-RFLP	Polymerase chain reaction-restriction fragment length polymorphism
PPV	Positive predictive value
PFGE	Pulse field gel electrophoresis
qPCR	Quantitative polymerase chain reaction
SDS-PAGE	Sodiumdodecyl sulfate-polyacrylamide gel electrophoresis
CTAB	Cetyl trimethyl ammonium bromide
dNTP	Deoxyribonucleotide triphosphate
DW	Distilled water
EDTA	Ethylenediaminetetraacetic acid
SDS	Sodiumdodecyl sulfate
TBE	Tris borate EDTA buffer
TE	Tris EDTA buffer
UV	Ultraviolet
bp	Basepair
°C	Degree celcius
kb	Kilobasepair
kDa	Kilodalton
kV	Kilovolt
µg	Microgram
µl	Microliter

LIST OF ABBREVIATIONS (cont.)

Abbreviation	Term
μm	Micrometer
μM	Micromolar
mg/ml	Milligram per milliliter
ml	Milliliter
mM	Millimolar
M	Molar
ng	Nanogram
nm	Nanometer
OD	Optical density
pmole	Picomole
V	Volt
w/v	Weight by volume

CHAPTER I
INTRODUCTION

Orientia tsutsugamushi is a Gram-negative, obligatory intracellular coccobacillus. Its size is approximately 0.5-0.8 μm wide and 1.2-3.0 μm long (1). This bacterium is the causative agent of scrub typhus, a major cause of febrile illness in rural areas of South-East Asia throughout the Asia Pacific rim and Northern Australia. The transmission of this disease involves the larval mites of most rodents, especially *Leptothrombidium* spp. (chigger) or thrombiculid mites (2). The clinical features of scrub typhus are acute fever, severe headache and myalgia. In addition, rash and regional lymphadenopathy are developed in some patients. Eschar, a typical scrub typhus marker, is an ulcer surrounded by a red areola and covered by a dark scab on the skin. It is developed at the site of infection by the bite of infected larval mite. However, the presence of an eschar in patients varies from 7%-97% (3).

Both humoral and cellular immunity play an important role in the pathogen clearance (4). After an infection of with *O. tsutsugamushi*, the macrophage is activated by IFN-γ secreted from T helper 1 (Th1) cells to eliminate the pathogen. Humoral immune response also plays a role in neutralization of the pathogen as well as inhibition of adhesion to and internalization into host cell (5). However, the mechanism of both immune responses depends on the strain of *O. tsutsugamushi* (6). The protective immune response to the homologous strain remains up to 1-3 years while the immune response against the heterologous strains exists only less than 3 months (7-9). Scrub typhus is curable by antibiotic treatment such as doxycycline or chloramphenicol. However, there is a report about the unresponsive to antibiotic treatment such as doxycycline in Thailand. The significant longer time for fever clearance after antibiotic treatment was demonstrated (10).

Scrub typhus in Thailand was first confirmed to be the disease by serological technique in 1953 (11). *O. tsutsugamushi* was first isolated from tissues of Thai rodents (*Rattus* and *Bandicota*) and found that the survival mice from Thai

Orientia infection could resist to the homologous strain inoculation (11). Moreover, eight species of chiggers were collected from different areas in Thailand and five strains of *O. tsutsugamushi* were identified. Karp was the most predominant strain and followed by TA716, TA763, TA686 and Kato (12). Furthermore, the serological identification was used to characterize 77 *Orientia* isolates form 4 different parts of Thailand. The result showed that eight strains; Karp, Kato, Gilliam, TA678, TA686, TA716 (chon strain), TA763 (Fan strain) and TH1817, were identified (13).

Orientia identification was performed on the basis of the phenotypic characterizations including serological assay and protein analysis such as complement fixation (CF), direct fluorescent antibody assay (DFA) and indirect fluorescent antibody assay (IFA) (15, 16). Moreover, the polyclonal antibody was used in the micro-immunofluorescence (MIF) (17). IFA was considered as the gold standard for *Orientia* identification but the great antigenic diversity of antigenic epitopes limited the sensitivity of IFA (18). The genetic characterizations using Polymerase Chain Reaction (PCR) amplification and direct sequencing were accomplished instead of these serological techniques to reduce time-consuming and high-level laboratory requirement. Recently, *Orientia* identification based on 56-kilodalton (kDa) protein encoding gene was performed since this gene contained both conserved and hypervariable domains (19). The hypervariation domains can occur from an unusual degree of insertion or deletion and causes the subsequent antigenic diversity (15). These domains caused this gene to be the most suitable candidate gene for *O. tsutsugamushi* strain identification (15).

The 56-kDa protein is a major outer membrane protein of *O. tsutsugamushi* which played an important role in attachment to or penetration into host cells (20). Consequently, this protein has been considered as type-specific antigen. The analysis of amino acid sequence alignment of 56-kDa gene among six strains of *O. tsutsugamushi* showed four variable domains (VDs; VD I-IV) residing between conserved domains. Nucleotide and amino acid sequences of the conserved domains have very high homology. Hence, these domains are the cause of the cross-reactivity between strains or species-specific domains. The VDs showed the similarities in amino acid sequences between individual strains ranging from 50%-70% which could be responsible for antigenic variation or strain or serotype specificities (19, 21).

The strain identification using PCR and nucleotide sequencing was performed in previous studies. The patients' sample collected from 8 medical centers in Thailand were identified the basis of 56-kDa protein encoding gene and demonstrated that Karp was the major strain followed by Kato (22). However, Gilliam which is the one of prototype strains could not be found. The recent study of the entire open reading frame (ORF) of the gene from 23 isolates of scrub typhus patients demonstrated that Karp was the predominant strain followed by Gilliam, TA716 and TA763 (23).

The information of strain identification in Thailand was rarely published. This information is necessary for the improvement of the antigen panel in IFA technique, the epidemiology study and vaccine consideration for further. Therefore, the objective of this study is to determine the nucleotide sequences of 56-kDa protein encoding gene of *O. tsutsug*amushi Thai samples and construct the dendrogram of nucleotide sequences for cluster analysis. The relationship between *O. tsutsugamushi* detected from Thai patients and that from other geographical regions was analyzed. Finally, the primer sets for *O. tsutsugamushi* strain identification based on 56-kDa protein encoding gene were developed.

CHAPTER II
LITERATURE REVIEW

1. *Orientia tsutsugamushi*

O. tsutsugamushi is a Gram-negative, obligatory intracellular coccobacillus. Its size is approximately 0.5-0.8 micrometers (μm) wide and 1.2-3.0 μm long (1). This bacterium is classified as the *Rickettsiaceae* family and was formerly known as *Rickettsia tsutsugamushi*. *Orientia* was reclassified as a new genus distinct from *Rickettsia* in structure that its outer leaflet of the cell wall is thicker than the inner leaflet while other members of *Rickettsia* such as *R. prowazekii* and *R. rickettsii*, their inner leaflets are thicker than the outer leaflet of cell wall. Moreover, the cell wall of *Orientia* lacks the components of peptidoglycan and lipopolysaccharide (1). *O. tsutsugamushi* is the causative agent of scrub typhus, a major cause of febrile illness in Southeast asia especially in Thailand. The epidemiological cycle of this disease involves mites of most rodents, especially *Leptothrombidium* spp. (chigger) or thrombiculid mites. In the colonies of mites or the 'mite islands', the cycle initiates with the dropping of a parasitic, blood-feeding, hexapod larva from rodents then passing through a nymphochrysalis and followed by eight-legged, non-parasitic nymphal and adult stages (2). The adult mites return to the rodents and finally lay eggs on the rodents (figure 1). The vertical transmission of *O. tsutsugamushi* occurs in these mites when the bacteria in the cytoplasm of various organs including oocytes, is passed to the offspring via transovarial transmission. Then, the bacterium is transferred to rodent by the bite of infected mite vector. Mite is the true reservoir but humans or other effected mammals are the accidental host. Rarely, the bacterium is reverse-transferred from infected animals to mites occurs inefficiently (24). There are many strains of *O. tsutsugamushi* which are composed of both unique and cross-reactive serological epitopes (25). The serological analysis has indicated that three prototype strains which are Karp, Kato and Gilliam are the most

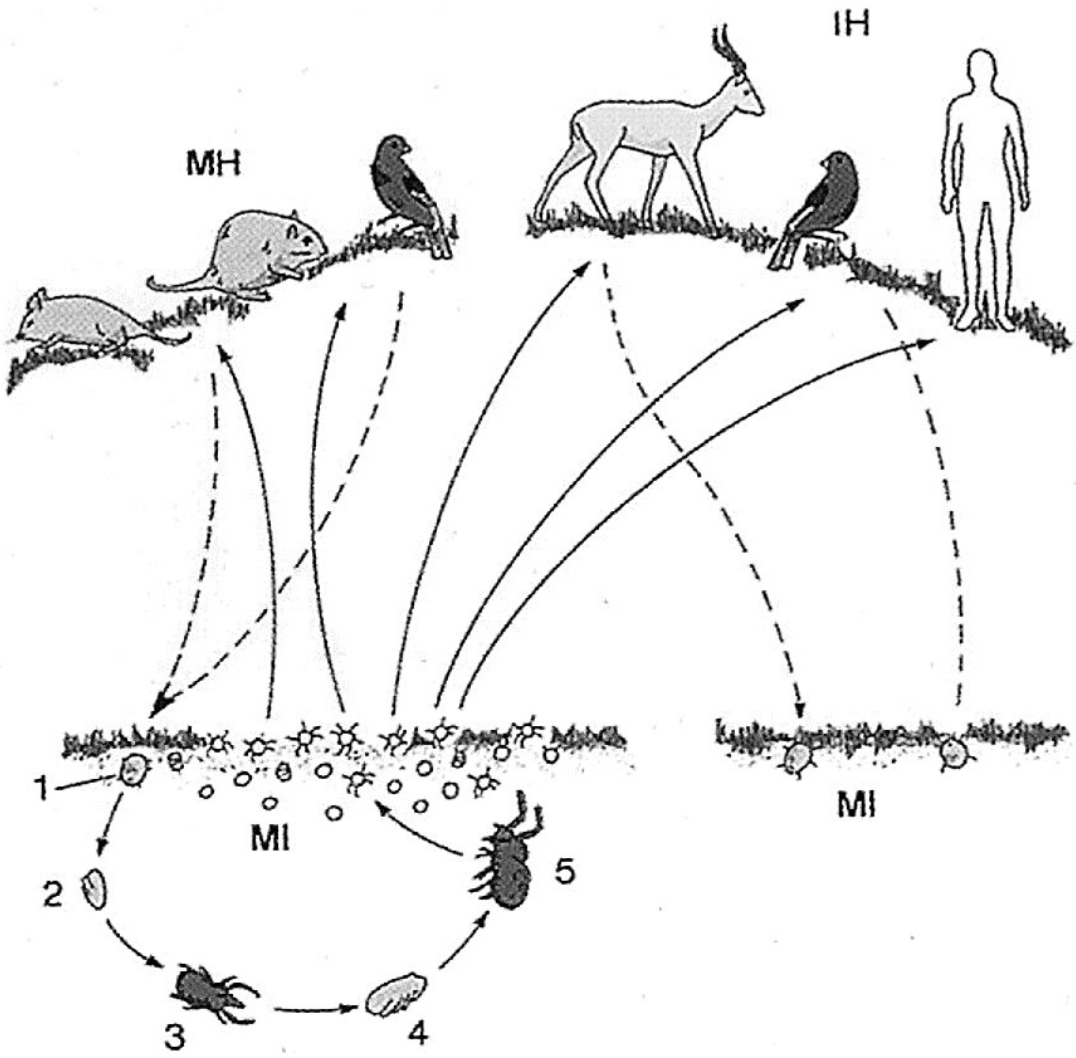


Figure 1. The epidemiological cycle of scrub typhus (MI-mite island, MH-maintenance host, IH-incidental host, 1-a parasitic, blood-feeding, hexapod larva, 2-nymphochrysalis, 3-eight legged, non-parasitic nymphal larva, 4-an imagochrysalis and 5-adult mite) (2).

prominent and represented as the major isolates of *O. tsutsugamushi* (21). In Thailand, the first isolation of *O. tsutsugamushi* from tissue of Thai rodent (Rattus and Bandicota) was reported in 1954 and indicated that mice which survived from Thai *Orientia* infection could resist to the homologous strain inoculation (11). Then, eight species of chiggers were collected form different areas in Thailand and 146 isolates of the bacteria were characterized. Five strains of *O. tsutsugamushi* could be identified. Karp was the most predominant strain (60%) followed by TA716, TA763, TA686 and Kato (12). Furthermore, the serological identification was used to characterize 77 *Orientia* isolates form 4 different regions of Thailand. The result showed that eight strains; Karp, Kato, Gilliam, TA678, TA686, TA716 (chon strain), TA763 (Fan strain) and TH1817, were identified (13).

2. Scrub typhus

The disease caused by *O. tsutsugamushi* has an incubation period of 5-20 days. The major features of scrub typhus are fever, severe headache and myalgia. In addition, rash, regional lymphadenopathy and eschar are developed in some patients. The severity of systemic infection can be ranged from symptomless to fatal. The lesion of scrub typhus initiates from a papule that develops at the site of the bite of infected larvae mite to the formation of an eschar, a typical scrub typhus marker. Eschar is an ulcer surrounded by a red areola and covered by a dark scab. The presence of an eschar in patients varies from 7% to 97% (3).

O. tsutsugamushi can infect endothelial cells, macrophages and polymorphonuclear leukocytes (PMNs) and invade into the cells by induction of phagocytosis. The bacteria attach to the host cells using surface proteins. Lee *et al* reported that the 56-kDa protein which is the outer membrane protein of *O. tsutsugamushi* can bind to host fibronectin and enhance the internalization process (26). Moreover, *O. tsutsugamushi* evades the immune response by phagolysosomal fusion inhibition after the bacterium is entrapped in the phagosome. The bacteria can replicate in cytoplasm of host cell by binary fission and infect other cells via protrusion or budding from the infected cells to neighboring cells which engulf the host membrane-bound or free bacteria (4).

The basic histopathologic lesions of scrub typhus are disseminated perivasculitis and focal interstitial mononuclear infiltrations associated with edema (7). The major symptom is interstitial pneumonia with alveolar edema, hemorrhages, interlobular septal edema and meningoencephalitis. Multiorgan vasculitis causes the diversity of clinical manifestations. Consequently, the failure of respiratory system and peripheral vascular collapse can be the causes of death in humans.

Immunity to *O. tsutsugamushi* depends on both humoral and cellular immunity which plays an important role in the pathogen clearance. In case of cellular immunity, specific T cells, especially T helper 1 (Th1) cells, have a role in protective immunity against *O. tsutsugamushi* by producing IFN-γ which activates the macrophage to eliminate the pathogen. The effect of IFN-γ, however, depends on the strain of *O. tsutsugamushi* and the infected cell type (6). Humoral immunity also plays a role in neutralization of the pathogen by inhibition of the adhesion and internalization process (5). This inhibition efficiency depends on the strain of *O. tsutsugamushi*.

The re-infection can occur frequently among patients living in endemic areas. The immunity to *O. tsutsugamushi* is strain-specific as a result that the protective immune response to the homologous strain last for up to 1-3 years while the immune response against the heterologous strains exists only less than 3 months and absent after a year (7-9). Although vaccines are not available for humans, scrub typhus is responsive to antibiotic treatment such as doxycycline or chloramphenicol. However, there was a report about the clinical response to doxycycline treatment with patient in chiangrai, northern of Thailand. There were 5 out of 12 patients showed the significant longer time for fever clearance after antibiotic treatment comparing to patients in Mae Sod, Tak province (10).

The epidemiology of scrub typhus is confined to a definite geographic region called ‘tsutsugamushi triangle’ extending from northern Japan in the north to northern Australia in the south and to Pakistan and Afghanistan in the west (figure 2) (4). The first case of scrub typhus in Thailand was confirmed to by serologically technique in 1953 (11). Recently, the infected cases of scrub typhus are occurred annually. The reported cases in 2009 presented 5,921 effected cases (morbidity ratio 9.33 per 100,000 populations) and there were 8 fatality cases (figure 3). Most cases

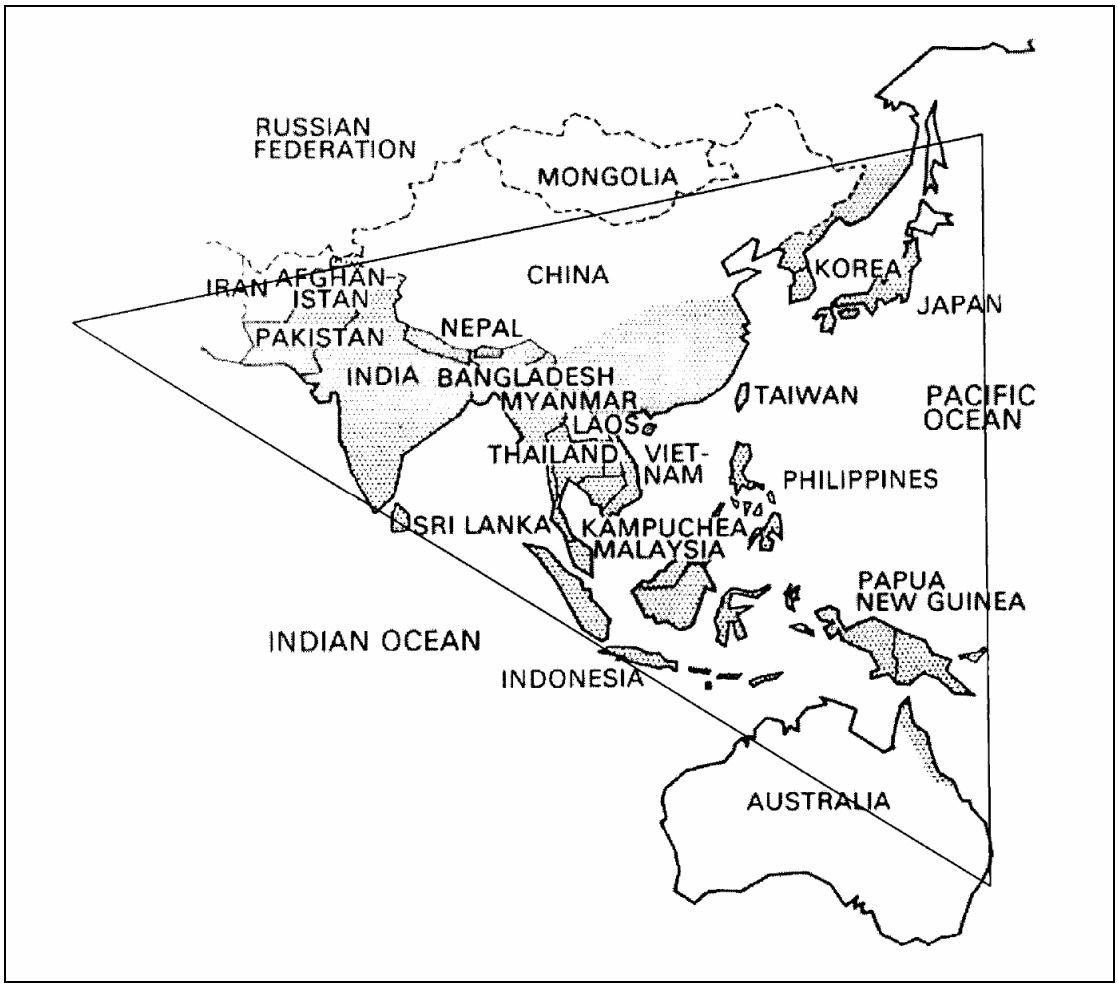


Figure 2. The tsutsugamushi triangle (4).

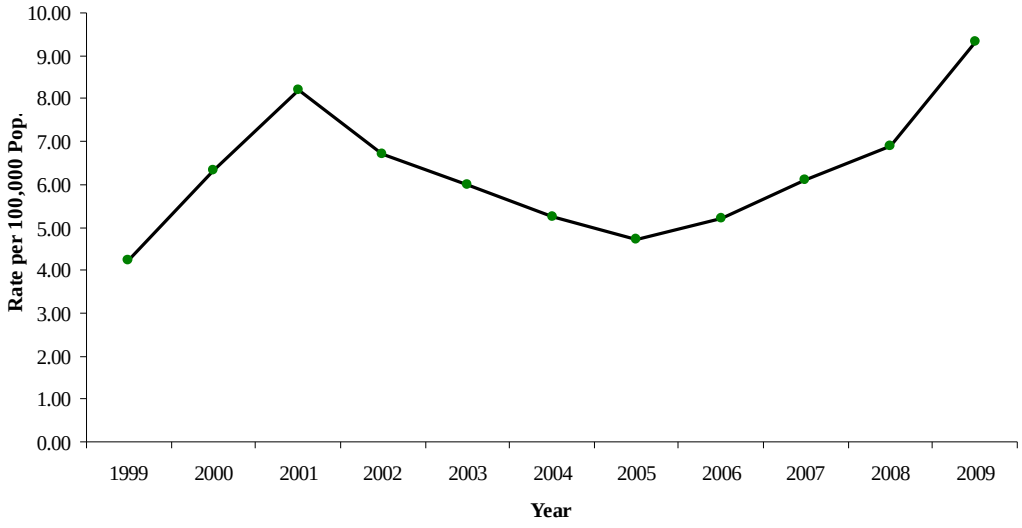


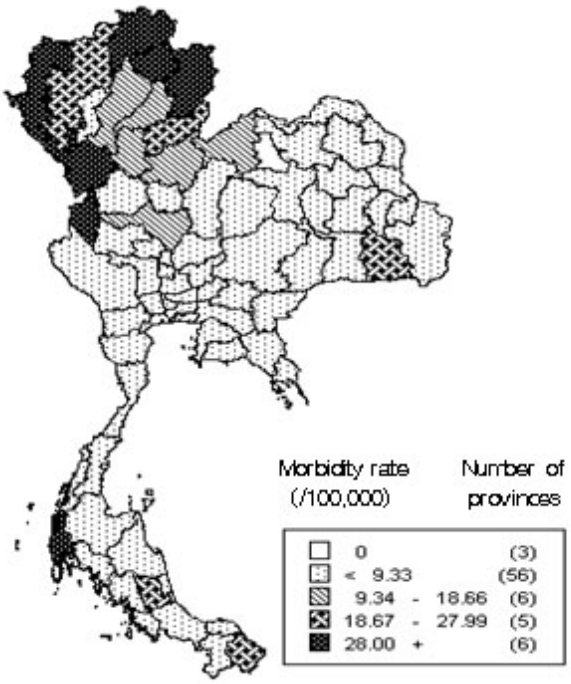
Figure 3. Reported cases of scrub typhus per 100,000 populations in Thailand between 2000-2009 (27).

were found in rainy season during May to October. The northern part of Thailand has the highest infected cases followed by the southern part, the northeastern and the central region, respectively (figure 4) (27).

3. The major antigenic protein of *Orientia tsutsugamushi*

O. tsutsugamushi belongs to alpha-proteobacteria in order *Rickettsiales*. The members of this order, *O. tsutsugamushi* contains the largest genome with 2,127,051 base pair (bp) and 2,008,987 bp for the whole genome of *O. tsutsugamushi* strain Boryong and Ikeda, respectively (24, 28, 29). Its genome is in a single circular chromosome with the GC content approximately 30.5% (29). The genome structure of *O. tsutsugamushi* includes the high degree of repeated sequences in range 37 – 46% (24, 29). Comparing to other bacteria, *Mycoplasma mycoides* and phytoplasmas contain 19.2 and 12.9% repetitive sequences, respectively (30). Even in the *Rickettsiales* order, *Wolbachia pipientis* strain wMel and *Anaplasma phagocytophilum* present 14.2 and 12.7% repetitive sequences, respectively (31). High rate of repetitive sequence causes the unique of genome of *O. tsutsugamushi*. The genome-based metabolic analysis revealed that the set of genes involving in peptidoglycan (PG) biosynthesis were identified. However, the genes involving degradation of PG was also found while those for lipopolysaccharide (LPS) biosynthesis were completely disappeared (24). As a result that PG and LPS were not presented on the cell wall of *O. tsutsugamushi*. Furthermore, comparison of whole genome of *O. tsutsugamushi* strain Ikeda and Boryong found that 540 genes presented as a backbone or conserved in both strains. Interestingly, there were the notable sequence variations in several genes especially surface protein (24).

The expressed proteins of *O. tsutsugamushi* were characterized by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) method. There were at least 30 major proteins including 165-, 150-, 138-, 110-, 72-, 70-, 63-, 60-, 58-, 57-, 56-, 54-, 50-, 49-, 47-, 46-, 43- and 20-kDa proteins (25, 32, 33). Most of these proteins are the surface protein except the 60-kDa protein which exhibits homology to the GroEL or heat shock protein family. This protein is expressed inside



Top Ten Leading Rate

1	Mae HongSon	171.88
2	Nan	122.32
3	Tak	88.47
4	Chiang Rai	64.07
5	Phangnga	62.60
6	Phayao	29.76
7	Chiang Mai	25.31
8	Phatthalung	24.48
9	SiSa Ket	20.99
10	Uttaradit	20.71

Figure 4. A map of Thailand demonstrated the reported cases of scrub typhus per 100,000 populations by province in 2009 (27).

the rickettsial cells (21). In addition, the 47-, 56-, 70-, and 110-kDa proteins were considered as the major antigens.

3.1 Heat shock protein

The 60-kDa protein of *O. tsutsugamushi* is the homolog of the heat shock protein GroEL expressed by *groEL* gene (21). Although it located in the cytoplasm of the organism, the immunoblotting method revealed that patient sera can react with the 60-kDa protein (34). The expression of GroEL protein is upregulated during the infection especially in the early period (35). The 60-kDa protein is highly conserved housekeeping protein which is essential for the survival of these bacteria. However, the nucleotide sequences of 60-kDa protein encoding gene are able to distinguished or differentiated the genus *Orientia* from *Rickettsia* (36).

3.2 Surface proteins

3.2.1 The 47-kDa protein

The 47-kDa protein belongs to the high temperature requirement (HtrA) family of stress response protein and is expressed on the outer membrane of *O. tsutsugamushi*. Using immunoblotting assay, the 47-kDa revealed the cross-reactivity with heterologous antiserum (32). The SDS-PAGE showed that the band of 47-kDa protein moved in different distance between different strains (32). This protein contains both scrub typhus group reactive or species-specific and strain-specific epitopes to B-cells (37). However, the 47-kDa protein encoding gene is highly conserved approximately 97% identity in 25 different strains of *O. tsutsugamushi* (38). Then, the 47-kDa protein and also protein encoding gene is selected as the target to distinguish *O. tsutsugamushi* from other *Rickettsia*. The diagnosis of *O. tsutsugamushi* infection based on 47-kDa protein encoding gene detected 3-21 copies of *Orientia* sequence per microlitre (μl) of blood (39). In addition, the protection of re-infection was accomplished in challenging to the mouse model with the 47-kDa protein. It provided good homologous protection and cross protection against heterologous strains (Ching WM, unpublished data). Unfortunately, the immune response of this protein can cross-react against a human protein, human

serine protease HtrA1 (hHtrA1) due to very high degree of amino acid sequence homology and may cause the autoimmune response (38).

3.2.2 The 56-kDa protein

The 56-kDa protein is the major surface protein of *O. tsutsugamushi* which expressed approximately 10-15% of its total protein (40). The function of this protein involves in the adhesion factor of bacteria to entry into the host cell (41). The strongly reaction by immunoblotting was occurred between the 56-kDa protein to the homologous antiserum whereas the weakly reaction was appeared in the interaction to heterologous antiserum (32). Moreover, mice immunized with the 56-kDa protein could generate neutralizing antibody and resisted to the infection caused by homologous strain of *O. tsutsugamushi* (20). The purified 56-kDa protein can also react with strain-specific monoclonal antibodies by a dot blot immunoassay (42). Therefore, this protein appears to be an immunodominant antigen and exhibits strain-specific characteristics (21). The 56-kDa protein composed of 521-532 amino acids depended on strains of *O. tsutsugamushi* corresponded with the molecular weight in range 54-58 kDa in different strains (19). Recently, the examination of amino acid sequences in databases revealed that the 56-kDa protein comprise of amino acid in range 516-541 amino acids. An unusual degree of nucleotide insertion and deletion may cause of amino acid lengths variation (15).

Ohashi N *et al* (1992) aligned the amino acid sequences of 56-kDa protein of 6 strains (Gilliam, Karp, Kato, Kawasaki, Kuroki and Shimokoshi) and revealed that there were 4 regions of VD I-IV showed the significant different amino acid sequences (figure 5). The amino acid identity between strains was varied from more than 75% to less than 50%. The amino acid length is different in each VD, VDI comprises of 22-28 amino acids where as there are 16-26, 24-35 and 34-40 amino acids in VD II, III and IV, respectively (19). The VD IV is the longest domain but VD II showed the greatest variation of amino acid residues. Comparison of the entired ORF, *O. tsutsugamushi* strains demonstrated high homologies in range of 71-89% (nucleotide sequence) and 56-82% (amino acid sequence). These VDs are considered to responsive for the antigenic variation of each strain of *O. tsutsugamushi*. Furthermore, the homology in each ORF generates the cross-reactivity in immune response.

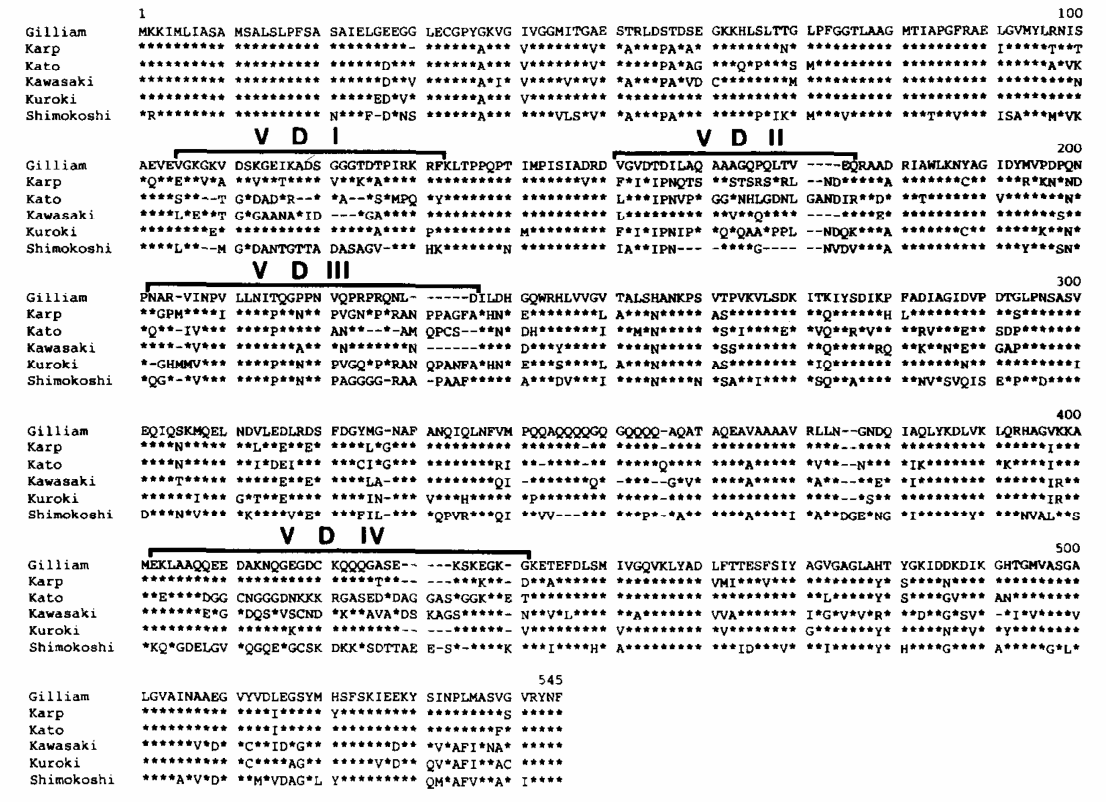


Figure 5. The multiple alignment of amino acid sequences of 56-kDa protein from 6 reference strains (Gilliam, Karp, Kato, Kawasaki, Kuroki and Shimokoshi). The asterisk (*) represented the identical amino acid compared to Gilliam strain. Dash (-) represented the gap. VD I, II, III and IV represented four variable domains (19).

The 56-kDa protein is group-specific protein or unique in *O. tsutsugamushi* which is not expressed in other bacteria (15). This protein is encoded from the 56-kDa protein encoding gene which presented as a single copy containing about 1,550 nucleotides (43). The immunodominant property of this protein demonstrated that it is the highest expressed protein and strain-specific reactivity convinces this protein and gene as an excellent target for the development of vaccine and disease diagnosis. Moreover, the nucleotide sequence analysis of entire ORFs and variable domains can be used as the major target for *O. tsutsugamushi* characterization or identification (15).

3.2.3 The 70-kDa protein

The 70-kDa protein was expressed on the surface of *O. tsutsugamushi* and was presented to react with the heterologous antisera in immunoblotting method implying that these polypeptides contained group or species-specific antigenicity (32).

3.2.4 The 110-kDa protein

The 110-kDa protein is the one of important surface protein of *O. tsutsugamushi*. Based on the affinity-purified antibodies against 110-kDa protein, the SDS-PAGE showed the unique mobility of bands in each strains of Karp, Kato and Gilliam protein. It was possible that the 110-kDa protein contains both cross-reactive and strain-specific epitopes (25).

4. *Rickettsia* identification based on molecular biology

There are 3 families in the order *Rickettsiales* which are *Rickettsiaceae*, *Bartonellaceae* and *Anaplasmataceae*. The family *Rickettsiaceae* composes of 3 genera which are *Coxiella*, *Rochalimaea*, *Rickettsiae*. More than 20 members in the genus *Rickettsiae* are classified into 3 groups: spotted fever group (SFG), typhus group (TG) and scrub typhus group (STG).

Rickettsial isolates can be identified using 2 approaches which are the phenotypic and genotypic approaches.

4.1 Phenotypic approaches

4.1.1 Serotyping

The phenotyping criteria for rickettsial isolate identification are based on serotyping with or without the protein analysis. The earliest identification initiated with the cross-reaction of sera with the somatic antigen of different type of *Proteus spp.* such as OX-19 antigen for *Rickettsia* in the TG, OX-2 and OX-K antigens for the SFG and STG, respectively. This technique called the Weil-Felix test. However, this technique was not used anymore because of the low sensitivity and specificity (18). Then, the MIF based on the serotyping technique was used with mouse polyclonal antisera especially in *Rickettsia spp.* for the SFG differentiation. This technique was developed to detect the specific epitopes of surface protein antigens of *Rickettsia* such as rOmpA, rOmpB and 120-kDa protein (44). In addition, the monoclonal antibody (mAb) was also used in MIF method but there was some limitation since the mAb lacked the standardized cut off to all known isolates (45). Because there are more limitations, this technique requires the comparison between the new isolate to all previous approved species to interpret the result (45).

4.1.2 Protein analysis

The protein analysis such as sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used as another method in the phenotypic criteria (46). Two outer membrane proteins of Rickettsia, rOmpA and rOmpB, were selected to identify the group by SDS-PAGE since these two proteins are the major or dominant antigens expressed in *Rickettsia* (47). However, the *Rickettsia* expressed few proteins then these 2 techniques were not appropriate to identify the intracellular rickettia (45). Furthermore, the highly variation in the molecular weight of rOmpA and rOmpB protein is the limitation of protein analysis method to differentiate the species of *Rickettsia* (45).

4.2 Genotypic approaches

4.2.1 Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP)

Recently, the genotypic criteria were used for rickettsial identification. Initiate with the PCR-RFLP, this technique was used to differentiate

between group of *Rickettsia* based on the amplification of protein-encoding gene of rOmpA and rOmpB. The PCR-RFLP generates the high level of reproducibility to identify genotype of rickettsial isolate by comparison with the reference patterns (48). For example, 9 SFG rickettsiae were differentiated by this technique based on OmpA protein encoding gene (48). Moreover, 2 new rickettsial isolates from China were identified as the new strains of SFG rickettsiae (49). The gene encoding OmpB protein was used to differentiate 36 strains of SFG rickettsiae except *R. africae* and *R. parkeri* (50). The limitation of PCR-RFLP is the appropriate endonuclease enzyme for analysis has been precisely chosen (51).

4.2.2 Pulsed field gel electrophoresis (PFGE)

PFGE was also used to identify the SFG rickettsiae (52) such as 10 isolates of *R. conorii* exhibited the same pattern distinctive from the other species. However, it is necessary to include all rickettsial species in the gel to compare all of the profiles.

4.2.3 PCR amplification and direct sequencing

4.2.3.1 16S rRNA gene

About 5 genes were studied using the direct sequence determination for the identification of rickettsiae. The 16S rRNA gene is universal in bacteria and commonly used for bacterial identification and taxonomy (53). The 16S rRNA (or DNA, in fact) is the small subunit of ribosomal molecule associated with the large unit (23S rRNA). The nucleotide sequence of this gene is approximately 1,550 bp in length containing both variable and conserved regions. The conserved regions are in both ends of sequence whereas the variable region is in between and used for the comparative taxonomy (53). Among the *Rickettsia* group, *Rickettsia spp.* showed the homology of 16S rRNA sequence at least 97.5% but *O. tsutsugamushi* showed only 90% homology when these 2 genera were compared. However, the nucleotide sequence of 16S rRNA gene was not variation enough to differentiate the species of SFG rickettsiae. Then, the comparison of the 16S rRNA gene sequence can accomplish to differentiate the bacteria only at the genus level but not the species or strain levels (53).

4.2.3.2 Citrate synthase gene or *gltA* gene

The citrate synthase gene or *gltA* containing approximately 1,234 bp of species-specific sequence was studied based on the amplification using the consensus primers complemented to the conserved region at the border of this gene. The citrate synthase is an enzyme of the citric acid cycle and presented in nearly all living cells (54). The multiple alignments of these amino acid sequences demonstrated the homology within the domains *Eukarya* and *Proteobacteria* in range 46-92 and 56-75%, respectively, whereas the lower homology was observed between domains in range 20-26% (55). The citrate synthase gene has the informative sequence for the identification of closely related bacteria (54). Furthermore, this gene was expressed in the chromosomes of all rickettsiae. The *gltA* gene of 28 bacteria belonging to the genus *Rickettsia* were analyzed by nucleotide sequencing and found that the obtained phylogenies were similar to that from 16S rRNA sequencing analysis (54). However, the *gltA* sequence analysis generated more reliable evolutionary relationship with the significant bootstrap values for all nodes (54).

4.2.3.3 rOmpA and rOmpB protein encoding genes

The two proteins, rOmpA and rOmpB, are expressed on the membrane of *Rickettsia* with the high molecular weight as 145 and 120 kDa, respectively. The function of rOmpA involves in the intracellular actin-based motility and adhesion to host cells whereas rOmpB, a protective coat in S-layer, acts as the molecular sieve and ion trap. They were associated with the outer leaflet of outer membrane (56, 57). The *ompA* gene could differentiate the SFG rickettsiae based on the restriction endonuclease analysis of PCR product since *ompA* gene has the variability in nucleotide sequence higher than in either 16S rRNA or *gltA* gene (58). There are the epitopes in rOmpB which are commonly found in both SFG and TG rickettsiae. They demonstrated the cross-reactivity between 2 strains in western blotting analysis (59). *OmpB* gene which is approximately 4,776 bp was analyzed and generated the phylogenetic result similar to the analysis of *gltA* and *ompA* gene except *R. rickettsii* and *R. honei*. However, the higher bootstrap was estimated in the phylogeny of *ompB* but not in either of *gltA* or *ompA* (59, 60).

4.2.3.4 56-kDa protein encoding gene

As previously described, the 56-kDa protein was expressed only in *O. tsutsugamushi*. The 56-kDa protein encoding gene contains the interstrain hypervariable regions (VD I-IV). Then, the gene sequence analysis (either from the hypervariable regions or whole gene) of 56-kDa protein encoding gene can differentiate *O. tsutsugamushi* into the different strain. The homology of 56-kDa protein encoding gene was evaluated to differentiate *O. tsutsugamushi* strain.

In 1999, 31 strains of *O. tsutsugamushi* were classified based on the nucleotide sequence homology of 56-kDa protein encoding gene (61). Approximately 91% of whole gene encompassed all 4 hypervariable regions were sequenced. The strain of *O. tsutsugamushi* were clustered in phylogenetic tree and presented that the strains distributed in northeastern and southeastern Asia are different types (61). Similarly, 14 strains of *O. tsutsugamushi* in Taiwan were characterized based on the nucleotide sequencing of whole gene which is the same region of Enatsu T *et al*, 1999 (61). The result showed that the homology of gene was greater than 99% in the same type or strain where as the percent homology was less than 95.84% in different strains (62). The VD II and III of the isolates from Malaysia were determined and the percent divergence between all pairs of sequence was generated. The local strains of Malaysia were more closely related to Karp, Gilliam and TA strains from South-East Asia (63). The nested PCR was used to identify *O. tsutsugamushi* Thai isolates based on the 56-kDa protein encoding gene followed by the nucleotide sequencing of PCR product. The result showed that the nucleotide sequences of 240 patients from 8 Regional Medical Sciences Centers had the high homology (at 96%) to Karp strain in majority (22). Only VD IV region was used to characterize 12 isolates of *O. tsutsugamushi* from scrub typhus patients. These isolates showed the relationship on a dendrogram based on the percent of disagreement (distances) among the reference isolates. In addition, each isolate could be classified based on the homology related to TA716 or Kato (89% homology), Japanese strain LA1 with 93% homology (64). Moreover, 2 hypervariable regions, VD II and III, were also used to characterize 10 specimens from scrub typhus patients in Thailand. The nucleotide sequences of 2 isolated specimens were classified based on the percent of similarity clearly belongs to Karp cluster. However, some isolates

that were not related to any reference strains were considered as new or undescribed genotypes (14). There was a report that the analysis of entire ORF of 56-kDa protein encoding gene to construct the phylogenetic tree by nucleotide sequencing generated more information than the analysis of few regions. The 23 strains of Thai isolated were characterized and clustered in the majority of Karp (23). This result corresponded to the previous study which only approximately 200-400 bp were analyzed by nucleotide sequencing (22).

CHAPTER III
MATERIALS AND METHODS

1. Patient samples

There were 430 samples from patients who exhibited the symptoms of febrile illness and headache in 4 regions of Thailand during the year 2000 to 2010. The numbers of samples were showed in table 1. All samples were diagnosed using IFA technique as the standard method for scrub typhus at Division of Infectious Diseases and Tropical Medicine, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University. Suspected patient's sera were collected on acute and/or convalescent period. The IFA assay was performed as describe briefly. The sera samples were interacted with pooled antigens of *O. tsutsugamushi* strain Karp, Kato and Gilliam provided by The National Research Institute of Health (NIH), Ministry of Public Health, Thailand. Then, fluorescein isothiocyanate (FITC)-conjugated anti-human IgG or FITC-conjugated anti-human IgM was added to bind with the presented IgG or IgM in suspected patient, respectively. The bound antibody was observed by a fluorescent microscope. A positive result of scrub typhus infection was $\geq$ 1:400 for a single IgM or IgG titer against *O. tsutsugamushi* or $\geq$ 4-fold rising for IgG titer in paired sera. This study was a part of the clinical study (65) 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.

There were 2 types of the patient samples in this study which were Ethylenediaminetetraacetic acid (EDTA) blood and chromosomal DNA samples.

Table 1. The information of patient samples in this study.

Regions of Thailand	Types of samples		Total
	EDTA blood	DNA	
Northeastern part	86	204	290
Northern part	4	NA[a]	4
Southern part	10	20	30
Central part	84	22	106
Total	184	246	430

[a] NA: no sample available

1.1 EDTA blood samples

One hundred and eighty four of blood samples with EDTA blood samples were prepared to collect the buffy coat layer. Five milliliters of EDTA blood samples were centrifuged at 3,000*xg* for 10 minutes at room temperature and then buffy coat was collected. Red blood cell contaminants were lysed with distilled water (DW) (66). The procedures of red blood cell lysis initiated by 2 milliliters (ml) of DW was added into red blood cell contaminated buffy coat, gently mixed and left it at room temperature for 1 minute. Subsequently, 0.9% of NaCl or normal saline solution was added up to 15 ml and centrifuged at 3,000*xg* for 5 minutes. The supernatant was discarded and the pellet was repeated with red blood cell lysis until it contained only white blood cell. The pellet of white blood cell was used for DNA extraction or stored at -20°C until used.

1.2 Chromosomal DNA samples

Two hundred and forty six chromosomal DNA samples which were obtained from the previous study were kept in -20°C until used.

2. Chromosomal DNA preparation

Chromosomal DNA of *O. tsutsugamushi* was extracted from white blood cell pellets of EDTA blood samples based on the minipreparation of bacterial genomic DNA method (67). The pellet was resuspended with TE buffer pH 8.0 (see appendix I) up to final volume of 600 µl. The 3 µl of 20 mg/ml proteinase K (see appendix I) in 30 µl of 10% of SDS (see appendix I) was added and incubated at 37°C for 1 hour. Then, 100 µl of the mixture of 10% of CTAB/ 0.7 molar (M) NaCl (see appendix I) and 80 µl of 5 M NaCl (see appendix I) was added before incubated at 65°C for 20 minutes. DNA was purified by adding 1 volume of the mixture of chloroform-isoamyl alcohol (24:1) (see appendix I), mixed thoroughly by vortexing for 1 minute and centrifuged at 12,000*xg* for 5 minutes. The top layer of the solution was transferred to

a new 1.5 ml-microcentrifuge tube. One volume of the mixture of phenol-chloroform-isoamyl (25:24:1) (see appendix I) was added and mixed thoroughly before centrifuged at 12,000*xg* for 5 minutes. The top layer of solution was transferred to a new 1.5 ml-microcentrifuge tube. DNA was precipitated after 0.6 volume of isopropanol was added and mixed gently. The mixture was left at 4°C for 15 minutes and then centrifuged at 12,000*xg* for 10 minutes. The supernatant was discarded and the DNA pellet was washed with 1 ml of 70% ethanol (see appendix I) by centrifugation at 12,000*xg* for 10 minutes. The 70% ethanol was discarded. DNA pellet was air-dried for 10 minutes then resuspended with 50 µl TE buffer (pH 8.0) and kept at -20°C until used.

3. Quantification of chromosomal DNA

DNA concentration was determined by spectrophotometric measurement at the absorbance of both 260 and 280 nm. TE buffer was used as blank. The conversion factor for DNA of 1 OD_{260} was 50 mg/ml. Purity of DNA was determined by the ratio of the absorbance at 260 and 280 nm. The ratio of 1.7-1.9 was represented as purified DNA.

4. Amplification of 56-kDa protein encoding gene of *O. tsutsugamushi*

4.1 Primer design

The 56-kDa protein encoding gene of *O. tsutsugamushi* was amplified by nested PCR technique using two primer sets of external and internal set. These primer sets were designed by alignment of the nucleotide sequences of prototype strains; Karp (accession number M33004), Kato (M63382) and Gilliam isolate UT196 (EF213079) which was derived from GenBank. These nucleotide sequences were aligned using BioEdit software version 7.0.9.0 (68). The conserved regions were selected as the

priming site of both sets of primers (see appendix II and table 2). PCR amplification product from these primers encompassed the variable domain (VD) I-IV (19).

Each primer length was designed in the range of 20-23 bases with the approximately 60% of G-C content. Primer verification was performed *in silico* by OligoCalc software (69) to avoid the secondary structure formation. The melting temperature (Tm) values which ranged from 50-65°C were determined by Primer-BLAST program.

4.2 Nested polymerase chain reaction (nested PCR)

For the first round of PCR amplification, 0.4 micromolar (μM) of sense and antisense of the external primer set were added in to the reaction mixture containing 5 μl of 10x PCR buffer, 200 μM of each of dNTP, 1 unit of a thermostable DNA polymerase (TopTaq DNA polymerase, Qiagen, Germany) and approximately 1 microgram (μg) of DNA samples. DNase/RNase free sterile water (Ultra PURE^{TM}, Gibco, invitrogen) was added to bring the total volume up to 50 μl. The PCR condition was initiated at 95°C for 2 minutes, then 35 cycles of DNA denaturation at 95°C for 1 minute, annealing at 50°C for 1 minute and extension at 72°C for 2 minutes followed by the final extension at 72°C for 5 minutes in a GeneAmp PCR System 9700 (PE-Applied Biosystem Inc., CA, USA). The new PCR reaction was prepared, for the nested round. The internal primers were mixed in the reaction mixture containing the same reagent as the first round. Then, 3 μl of the first round PCR product was added. The amplification cycle was also the same as the first round of amplification.

5. Detection of PCR products by agarose gel electrophoresis

Agarose gel (SeaKem® LE Agarose, Cambrex) was prepared at a concentration of 0.7% (w/v) in 1x TBE buffer (see appendix I). The suspension was melt in microwave oven until completely dissolved then poured into the gel casting

Table 2. Oligonucleotide primers for nested PCR amplification of 56-kDa protein encoding gene.

Primers	Nucleotide sequences (5'→3')	Positions[a]	Polarity
E1	GCTAAAGTTGGAGTTGTTGGAGG	661-683	sense
E2	CCACATACACACCTTCAGCAGC	2056-2035	antisense
I1	CCATTTGGTGGAACGTTGGCTGC	766-788	sense
I2	GTCAGCATAGAGTTTAACTTGGC	1896-1874	antisense

[a] The positions were given according to the sequence of 56-kDa protein encoding gene of *O. tsutsugamushi* with accession number M33004.

tray with an appropriate comb. Gel was left for polymerization at room temperature for at least 30 minutes. Comb was removed and the gel was placed in an electrophoresis tank (PowerPac^{TM}, BioRad, USA). TBE (1x) buffer was added in the tank until submerged the gel.

Ten microlitres of PCR product was mixed with 2 μl of 6x DNA loading buffer (see appendix I) and loaded into the gel slots. Two microliters of 1 kilobasepair (kb) DNA ladder (GeneRuler^{TM}, Fermentas, CA, USA) was also loaded in the first gel slot as a DNA marker. Gel was electrophoresed at constant voltage at 80 V for 1 hour. The gel was stained in the solution of 10 mg/ml of ethidium bromide (see appendix I) in sterile water for 30 minutes at room temperature. Then, the gel was destained in sterile water for 15 minutes at room temperature to eliminate the excess ethidium bromide. The ethidium bromide-stained gel was examined by UV-transilluminator and taken a photograph using GeneGenius^{TM} (SYNGENE, MD, USA) gel documentation system. The expected band of nested PCR product was 1,131 bp.

Samples which exhibited the expected band were considered as the positive samples. These positive samples were selected for further analysis using DNA sequencing method and phylogenetic analysis.

6. DNA sequencing

6.1 Preparation of DNA template

Whole PCR product was run on 0.7% (w/v) agarose gel and stained with ethidium bromide as previously describe. The gel was then visualized by UV-transilluminator (Ultra-Lum, Inc., CA, USA). A clean, sharp scalpel was used to excise the expected DNA band (1,131 bp) in the gel. The excess gel which was not including DNA was trim out for the minimum size of gel. The excised gel containing DNA was transferred to a new 1.5 ml-microcentrifuge tube for purification.

6.2 DNA purification as a template for automated DNA sequencing

DNA in the excised gel was purified as in the protocol of QIAquick Gel Extraction kit (Qiagen, Germany). Briefly, the excised gel was weighed and considered as 1 volume of gel. Three gel volumes of QG buffer provided in the kit was added into the tube then incubated at 50°C for 10 minutes or until the gel completely dissolved. After incubation, 0.6 gel volume of isopropanol was added and mix thoroughly. Then, the mixture was applied to a QIAquick spin column. In this protocol, all centrifugation steps carried out at 13,000*xg* for 1 minute. The flow-through solution was discarded after centrifugation step and 500 µl of QG buffer was added to column, then centrifuged and the flow-through solution was discarded. The column was washed with 750 µl of PE buffer, after that the column was left at room temperature for 5 minutes before centrifugation once to eliminate the excess salt which affected in direct sequencing. The flow-through solution was discarded and the column was centrifuged for an additional 1 minute to eliminate the excess buffer. After that, the column was placed into a new 1.5 ml-microcentrifuge tube for elution step. DNA elution was performed by addition of 30 µl of EB buffer onto the center of membrane of the column and centrifuged for 1 minute. DNA was kept at -20°C until used as a template for automated DNA sequencing. The purity and the amount of DNA were evaluated by agarose gel electrophoresis.

6.3 Automated DNA sequencing using BigDye^{TM} Terminator

The nucleotide sequence of purified PCR product of 56-kDa protein encoding gene was analyzed using ABI BigDye^{TM} Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, CA, USA). The principle of this kit is based on Sanger's dideoxy sequencing protocol. Each sequencing reaction contained 8 µl of Terminator Ready Reaction Mix with 3.2 picomoles (pmoles) of primers. In this study, 2 primers, I1 and I2 (table 2), were used as sequencing primer for each DNA template to generated the sense and anti-sense stranded of DNA, respectively. The quantity of template was in range 10-40 nanograms (ng) when the PCR product was approximately 1,000-2,000 bp. Deionized water was added to adjust the total volume to 20 µl. The reaction cycles were initiated with 96°C for 1 minute followed by 25

cycles of 96°C for 10 second, 50°C for 5 second and 60°C for 4 minutes. The excess or unincorporated dye terminators were removed using ethanol precipitation method to reduce the obscure results. Each sequencing reaction was mixed with 16 μl of deionized water and 64 μl of 95% ethanol and incubated at room temperature for 15 minutes to precipitate the products followed by centrifugation at 10,000*xg* for 15 minutes at 4°C. The supernatant was completely discarded by inverting the tube onto a paper towel. The tube was left until the pellet was dried approximately 10 minutes. Each pellet was resuspended in 25 μl of template suppression reagent and mixed thoroughly. The reaction was denatured at 95°C for 2 minutes after that chilled it on ice and mixed thoroughly. Subsequently, the reaction was loaded on the ABI 3730xI 96-capillary automatic DNA Analyzer (Applied Biosystems, CA, USA). The DNA was automatically injected into the capillary at 2.5 kV and subjected for 2 hours to the POP-7^{TM} polymer for capillary electrophoresis of long read sequencing at 12.2 kV, 50°C. The ABI Sequencing Analysis Software v5.2 (Applied Biosystems, CA, USA) was used to display instrument status and record the raw data in real time situation. The chromatogram was displayed and analyzed using Sequence Scanner Software v1.0 (Applied Biosystems, CA, USA).

7. Analysis of nucleotide sequencing data

7.1 Editing and assembling of nucleotide sequences

The chromatogram and nucleotide sequence data were editing using ChromasPro software v1.5 (Technelysium Pty. Ltd., Australia). The input sequences were in Applied Biosystems format (.ab1). The chromatogram and the text file of nucleotide sequences could be displayed. The undefined nucleotides (shown as ‘N’) which were found at 5’-end from the excess dye peaks and at 3’-end from unreadable nucleotides were observed by the software. These nucleotides were removed for further analysis. However, it was necessary to check manually for the whole sequences. The anti-sense stranded DNA was reverse-complemented before

assembled to the sense stranded DNA. Eventually, the consensus sequence was generated for further analysis.

7.2 Multiple nucleotide sequences alignment

The nucleotide sequences contained hypervariable (VD) I-IV regions of each samples were aligned with the nucleotide sequences of reference and prototype strains obtained from GenBank database (table 3). The ClustalW (70) options in BioEdit program version 7.0.9.0 (68) was used for multiple alignment of the nucleotide sequences. A sequence alignment file (.aln) was generated.

8. Phylogenetic analysis

The Molecular Evolutionary Genetics Analysis (MEGA) program version 4 (71) was used to analyze and cluster the obtained nucleotide sequences. The .aln file from ClustalW was imported to this program then converted to MEGA file format (.meg) before analysis. The nucleotide sequence alignment file (.meg) was activated then the parameters were set up for dendrogram construction. Neighbor-joining algorithm based on a distance algorithm was chosen and the distance matrix was calculated to generate a dendrogram. A reliability of constructed dendrogram was estimated by bootstrapping with 1,000 replications. In each bootstrapping, the new dendrogram was generated and compared to the original one. If one clade of new dendengram is presented in the original clade, a score will be 1. A score will be 0 when there is not the original clade in a new dendrogram. A thousand of comparisons were estimated in bootstrapping method. The number that showed on clade was the times of a pattern of clade occurred in the bootstrap replicates. In addition, pairwise distance values were computed by the MEGA v.4 program to establish the different sites between two sequences in a multiple alignment.

Table 3. Description of prototype and reference strains of *O. tsutsugamushi* used in multiple alignments.

Accession number	Geographical Origin	Strain	Isolate	References
M33004	New Guinea	Karp	Karp	21
EF213078	Thailand	Karp	UT96	23
EF213098	Thailand	Karp	FPW2031	23
M63382	Japan	Kato	Kato	19
EF213085	Thailand	Gilliam	FPW2016	23
EF213079	Thailand	Gilliam	UT196	23
U80635	Thailand	TA686	TA686	Unpublished
U80636	Thailand	TA763	TA763	Unpublished
U19904	Thailand	TA678	TA678	Unpublished
U19905	Thailand	TA716	TA716	Unpublished
L04956	South Korea	Kuroki	Boryong	72
AF173033	Japan	JG	Ikeda	61
M63380	Japan	Kuroki	Kuroki	19
M63383	Japan	Kawasaki	Kawasaki	19
M63381	Japan	Shimokoshi	Shimokoshi	19

CHAPTER IV
RESULTS

1. Amplification of 56-kDa protein encoding gene of *O. tsutsugamushi*

The nucleotide sequences of 56-kDa protein encoding gene of *O. tsutsugamushi* prototype strains obtained from GenBank database were aligned for PCR primer design. Two sets of oligonucleotide primers, external and internal pairs, were designed for nested PCR amplification. The product of nested PCR was approximately 1,131 bp (figure 6). Then, the amplification products were purified and served as the sequencing template. The direct DNA sequencing was performed using the automated DNA sequencer. There were 106 of 430 samples from scrub typhus suspected patients that could be positively amplified by these primers. However, only 78 of 106 amplification products generated the satisfactory results of sequencing.

The detection of scrub typhus based on IFA technique showed 304 positive samples (70.7%) including 147 with single high IgM or IgG titer and 157 with 4-fold rising titer where as 106 positive results (24.7%) were demonstrated from the amplification of 56-kDa protein encoding gene (table 4).

2. Phylogenetic analysis of partial nucleotide sequences of 56-kDa protein encoding gene of *O. tsutsugamushi*

2.1 Multiple sequence alignment and phylogenetic analysis

There were readable and assembled 78 nucleotide sequences of the nested PCR amplified products. The obtained sequences from samples were compared with 15 reference sequences from GenBank database which shown in table 3. All sequences were aligned using the Clustal W (70) and the alignment result was saved in the fasta file format (.fas). The sequence identity matrix was calculated using BioEdit program to demonstrate the similarity of each sequence to the others (table 5). Sixty-

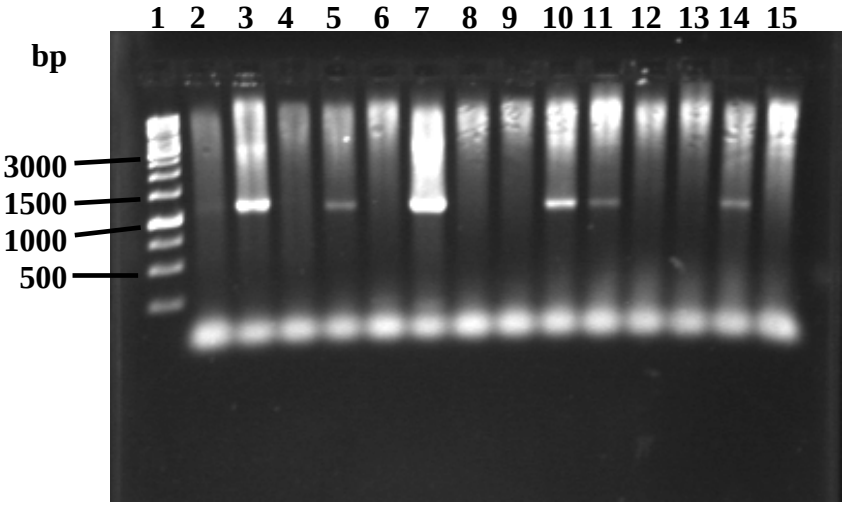


Figure 6. Agarose gel electrophoresis of PCR product amplified using the 56-kDa protein encoding gene primers. The size of PCR product is 1,131 bp. Lane 2 -13 were the clinical samples which the 56-kDa protein encoding gene was amplified in lane 3, 5, 7, 10 and 11. Lane 14 was the positive DNA and lane 15 was the reagent negative control. The 1 kb DNA marker was shown in lane 1.

Table 4. The positive results by IFA technique comparing to the amplification of 56-kDa protein encoding gene.

Regions (n of samples)	Positive results based on	
	IFA, n (%)[a]	56-kDa protein encoding gene amplification, n (%)[a]
Northeastern (290)	216 (74.5)	62 (21.4)
Central (106)	59 (55.7)	31 (29.2)
Southern (30)	25 (83.3)	11 (36.7)
Northern (4)	4 (100)	2 (50)
Total (430)	304 (70.7)	106 (24.7)

[a] The percentage was calculated from the positive result in each region and all regions.

Table 5. The sequence identity matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The identity values were calculated on BioEdit program (page 1 of 10).

Patients / References	KP111	BP288	KP112	KP116	KP68	KP159	KP171	KP133	KP67	KP174
Karp M33004	0.742	**0.963**	0.736	0.919	0.924	**0.954**	**0.988**	0.756	**0.989**	**0.989**
Karp UT76	0.742	0.907	0.736	**0.997**	0.899	0.905	0.920	0.739	0.921	0.921
Karp FPW2031	0.741	0.974	0.737	0.915	**0.936**	0.947	0.973	0.761	0.974	0.974
Kato M63382	0.714	0.721	0.708	0.705	0.715	0.727	0.720	0.686	0.721	0.721
Gilliam FPW2016	0.947	0.727	0.94	0.743	0.720	0.731	0.734	0.711	0.733	0.733
Gilliam UT196	**0.996**	0.737	**0.991**	0.743	0.731	0.743	0.745	0.721	0.745	0.745
TA686 U80635	0.723	0.749	0.715	0.730	0.749	0.746	0.745	0.696	0.746	0.746
TA763 U80636	0.731	0.854	0.724	0.844	0.835	0.862	0.858	**0.776**	0.859	0.859
TA678 U19904	0.692	0.746	0.686	0.739	0.743	0.757	0.743	0.674	0.744	0.744
TA716 U19905	0.689	0.711	0.684	0.703	0.713	0.717	0.715	0.679	0.716	0.716
Kuroki_ Boryong isolate L04956	0.738	0.915	0.734	0.875	0.910	0.906	0.909	0.767	0.910	0.910
JG_Ikeda isolate AF173033	0.946	0.727	0.940	0.731	0.723	0.732	0.732	0.702	0.732	0.732
Kuroki_ Kuroki isolate M63380	0.739	0.919	0.735	0.878	0.913	0.909	0.912	0.768	0.913	0.913
Kawasaki M63383	0.886	0.707	0.879	0.715	0.700	0.710	0.712	0.697	0.712	0.712
Shimokoshi M63381	0.690	0.671	0.684	0.664	0.665	0.669	0.673	0.658	0.673	0.673

The bold number represented the highest values in each samples**.**

Table 5. The sequence identity matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The identity values were calculated on BioEdit program (cont.) (page 2 of 10).

Patients / References	KP20	KP137	KP115	RAP117	KP66	BP269	KP155	CHUP50	BP5	BP41
Karp M33004	0.785	**0.987**	0.919	0.787	**0.956**	0.917	**0.989**	0.709	0.786	0.919
Karp UT76	0.776	0.919	**0.997**	0.778	0.907	**0.995**	0.921	0.709	0.778	**0.997**
Karp FPW2031	0.804	0.972	0.915	0.805	0.949	0.913	0.974	0.707	0.805	0.915
Kato M63382	0.682	0.719	0.705	0.683	0.726	0.704	0.721	0.699	0.683	0.705
Gilliam FPW2016	0.679	0.731	0.743	0.68	0.730	0.743	0.733	0.876	0.680	0.743
Gilliam UT196	0.687	0.743	0.743	0.689	0.742	0.742	0.745	**0.913**	0.688	0.743
TA686 U80635	0.751	0.745	0.730	0.753	0.746	0.728	0.746	0.713	0.752	0.730
TA763 U80636	**0.861**	0.857	0.844	**0.868**	0.864	0.842	0.859	0.693	**0.865**	0.844
TA678 U19904	0.717	0.742	0.739	0.718	0.757	0.737	0.744	0.680	0.718	0.739
TA716 U19905	0.714	0.714	0.703	0.715	0.718	0.702	0.716	0.675	0.715	0.703
Kuroki_ Boryong isolate L04956	0.802	0.908	0.875	0.804	0.908	0.873	0.910	0.702	0.803	0.875
JG_Ikeda isolate AF173033	0.686	0.730	0.731	0.687	0.731	0.730	0.732	0.888	0.687	0.731
Kuroki_ Kuroki isolate M63380	0.804	0.911	0.878	0.806	0.911	0.876	0.913	0.703	0.805	0.878
Kawasaki M63383	0.677	0.712	0.715	0.678	0.709	0.714	0.712	0.871	0.678	0.715
Shimokoshi M63381	0.651	0.672	0.664	0.655	0.669	0.665	0.673	0.675	0.653	0.664

The bold number represented the highest values in each samples**.**

Table 5. The sequence identity matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The identity values were calculated on BioEdit program (cont.) (page 3 of 10).

Patients / References	BP46	BP64	BP100	BP107	BP143	BP151	BP259	BP263	RAP8	RAP39
Karp M33004	0.743	**0.953**	0.919	0.781	0.786	**0.956**	0.742	0.742	0.741	0.919
Karp UT76	0.743	0.906	**0.997**	0.772	0.778	0.907	0.742	0.742	0.741	**0.997**
Karp FPW2031	0.742	0.946	0.915	0.800	0.805	0.949	0.741	0.741	0.740	0.915
Kato M63382	0.714	0.727	0.705	0.679	0.683	0.726	0.714	0.714	0.712	0.705
Gilliam FPW2016	0.948	0.729	0.743	0.676	0.680	0.730	0.947	0.947	0.945	0.743
Gilliam UT196	**0.995**	0.740	0.743	0.684	0.688	0.742	**0.996**	**0.996**	**0.994**	0.743
TA686 U80635	0.724	0.746	0.730	0.748	0.752	0.746	0.723	0.723	0.721	0.730
TA763 U80636	0.731	0.862	0.844	**0.860**	**0.865**	0.864	0.731	0.731	0.728	0.844
TA678 U19904	0.692	0.758	0.739	0.714	0.718	0.757	0.692	0.692	0.690	0.739
TA716 U19905	0.689	0.718	0.703	0.711	0.715	0.718	0.689	0.689	0.687	0.703
Kuroki_ Boryong isolate L04956	0.739	0.907	0.875	0.798	0.803	0.908	0.738	0.738	0.737	0.875
JG_Ikeda isolate AF173033	0.945	0.729	0.731	0.683	0.687	0.731	0.946	0.946	0.944	0.731
Kuroki_ Kuroki isolate M63380	0.740	0.910	0.878	0.800	0.805	0.911	0.739	0.739	0.738	0.878
Kawasaki M63383	0.885	0.707	0.715	0.676	0.678	0.709	0.886	0.886	0.884	0.715
Shimokoshi M63381	0.690	0.667	0.664	0.649	0.653	0.669	0.690	0.690	0.688	0.664

The bold number represented the highest values in each samples**.**

Table 5. The sequence identity matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The identity values were calculated on BioEdit program (cont.) (page 4 of 10).

Patients / References	RAP64	CHUP20	CHUP30	CHUP40	CHUP68	KL36	KL53	KL102	BL8
Karp M33004	0.708	0.838	0.742	0.720	0.833	0.742	**0.988**	0.919	0.739
Karp UT76	0.701	0.823	0.742	0.720	0.823	0.742	0.920	**0.997**	0.739
Karp FPW2031	0.709	0.848	0.741	0.718	0.847	0.741	0.973	0.915	0.738
Kato M63382	**0.961**	0.703	0.714	0.704	0.702	0.714	0.720	0.705	0.712
Gilliam FPW2016	0.699	0.718	0.947	0.919	0.711	0.947	0.732	0.743	0.944
Gilliam UT196	0.713	0.731	**0.996**	**0.952**	0.723	**0.996**	0.744	0.743	**0.993**
TA686 U80635	0.720	0.764	0.723	0.710	0.765	0.723	0.745	0.730	0.721
TA763 U80636	0.704	**0.967**	0.731	0.714	**0.925**	0.731	0.858	0.844	0.727
TA678 U19904	0.797	0.736	0.692	0.681	0.735	0.692	0.743	0.739	0.690
TA716 U19905	0.838	0.729	0.689	0.682	0.732	0.689	0.715	0.703	0.687
Kuroki_ Boryong isolate L04956	0.708	0.853	0.738	0.717	0.845	0.738	0.909	0.875	0.736
JG_Ikeda isolate AF173033	0.702	0.727	0.946	0.909	0.721	0.946	0.731	0.731	0.943
Kuroki_ Kuroki isolate M63380	0.709	0.855	0.739	0.718	0.847	0.739	0.912	0.878	0.737
Kawasaki M63383	0.693	0.706	0.886	0.867	0.698	0.886	0.711	0.715	0.883
Shimokoshi M63381	0.653	0.664	0.690	0.672	0.666	0.690	0.672	0.664	0.688

The bold number represented the highest values in each samples**.**

Table 5. The sequence identity matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The identity values were calculated on BioEdit program (cont.) (page 5 of 10).

Patients / References	BL9	BL133	BP13	RAP25	BP120	RAP22	Si001	Si002	Si005	Si006
Karp M33004	0.919	0.989	0.709	**0.986**	**0.989**	0.742	0.919	0.787	0.740	0.742
Karp UT76	**0.997**	0.921	0.702	0.920	0.921	0.742	**0.997**	0.778	0.740	0.742
Karp FPW2031	0.915	**0.974**	0.710	0.973	0.974	0.741	0.915	0.805	0.739	0.741
Kato M63382	0.705	0.721	**0.965**	0.722	0.721	0.714	0.705	0.683	0.711	0.714
Gilliam FPW2016	0.743	0.733	0.701	0.732	0.733	0.947	0.743	0.680	0.944	0.947
Gilliam UT196	0.743	0.745	0.715	0.744	0.745	**0.996**	0.743	0.689	**0.993**	**0.996**
TA686 U80635	0.730	0.746	0.722	0.745	0.746	0.723	0.730	0.753	0.719	0.723
TA763 U80636	0.844	0.859	0.705	0.856	0.859	0.731	0.844	**0.868**	0.728	0.731
TA678 U19904	0.739	0.744	0.798	0.745	0.744	0.692	0.739	0.718	0.690	0.692
TA716 U19905	0.703	0.716	0.839	0.717	0.716	0.689	0.703	0.715	0.687	0.689
Kuroki_ Boryong isolate L04956	0.875	0.910	0.709	0.909	0.910	0.738	0.875	0.804	0.736	0.738
JG_Ikeda isolate AF173033	0.731	0.732	0.705	0.731	0.732	0.946	0.731	0.687	0.942	0.946
Kuroki_ Kuroki isolate M63380	0.878	0.913	0.710	0.912	0.913	0.739	0.878	0.806	0.737	0.739
Kawasaki M63383	0.715	0.712	0.694	0.711	0.712	0.886	0.715	0.678	0.884	0.886
Shimokoshi M63381	0.664	0.673	0.653	0.67	0.673	0.690	0.664	0.655	0.685	0.690

The bold number represented the highest values in each samples**.**

Table 5. The sequence identity matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The identity values were calculated on BioEdit program (cont.) (page 6 of 10).

Patients / References	Si007	Si008	Si009	Si010	Si011	Si012	Si013	Si015	Si016	Si018
Karp M33004	0.710	0.742	0.739	0.945	0.742	0.742	0.742	0.965	**0.924**	0.741
Karp UT76	0.703	0.742	0.739	0.913	0.742	0.742	0.742	0.909	0.885	0.74
Karp FPW2031	0.710	0.741	0.740	**0.952**	0.741	0.741	0.741	**0.976**	0.916	0.739
Kato M63382	**0.963**	0.714	0.710	0.726	0.714	0.714	0.714	0.723	0.713	0.712
Gilliam FPW2016	0.700	0.947	0.943	0.734	0.947	0.947	0.947	0.729	0.711	0.945
Gilliam UT196	0.715	**0.996**	**0.994**	0.744	**0.996**	**0.996**	**0.996**	0.739	0.722	**0.994**
TA686 U80635	0.721	0.723	0.718	0.762	0.723	0.723	0.723	0.75	0.736	0.721
TA763 U80636	0.705	0.731	0.727	0.862	0.731	0.731	0.731	0.856	0.841	0.73
TA678 U19904	0.798	0.692	0.688	0.756	0.692	0.692	0.692	0.748	0.744	0.69
TA716 U19905	0.839	0.689	0.686	0.722	0.689	0.689	0.689	0.713	0.701	0.687
Kuroki_ Boryong isolate L04956	0.709	0.738	0.737	0.921	0.738	0.738	0.738	0.917	0.88	0.736
JG_Ikeda isolate AF173033	0.704	0.946	0.943	0.733	0.946	0.946	0.946	0.729	0.711	0.944
Kuroki_ Kuroki isolate M63380	0.710	0.739	0.738	0.924	0.739	0.739	0.739	0.921	0.883	0.737
Kawasaki M63383	0.695	0.886	0.883	0.715	0.886	0.886	0.886	0.709	0.690	0.884
Shimokoshi M63381	0.654	0.690	0.686	0.676	0.690	0.690	0.690	0.673	0.655	0.689

The bold number represented the highest values in each samples**.**

Table 5. The sequence identity matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The identity values were calculated on BioEdit program (cont.) (page 7 of 10).

Patients / References	RAP13	BP32	BP119	Si019	CHUP140	RAP107	KP128	CHUP138	RAP36
Karp M33004	0.939	0.742	0.741	0.937	0.742	0.741	0.735	0.743	0.735
Karp UT76	0.916	0.742	0.741	0.905	0.742	0.741	0.735	0.743	0.735
Karp FPW2031	**0.948**	0.741	0.74	**0.943**	0.741	0.74	0.736	0.742	0.734
Kato M63382	0.723	0.714	0.713	0.724	0.714	0.712	0.706	0.714	0.708
Gilliam FPW2016	0.731	0.947	0.946	0.733	0.947	0.943	0.937	0.947	0.938
Gilliam UT196	0.741	**0.996**	**0.995**	0.743	**0.996**	**0.992**	**0.987**	**0.996**	**0.984**
TA686 U80635	0.755	0.723	0.722	0.755	0.723	0.718	0.712	0.722	0.717
TA763 U80636	0.856	0.731	0.730	0.850	0.731	0.728	0.723	0.732	0.724
TA678 U19904	0.75	0.692	0.691	0.749	0.692	0.691	0.684	0.693	0.686
TA716 U19905	0.717	0.689	0.688	0.718	0.689	0.688	0.682	0.690	0.684
Kuroki_ Boryong isolate L04956	0.917	0.738	0.737	0.912	0.738	0.737	0.733	0.739	0.731
JG_Ikeda isolate AF173033	0.73	0.946	0.945	0.732	0.946	0.941	0.937	0.945	0.938
Kuroki_ Kuroki isolate M63380	0.92	0.739	0.738	0.915	0.739	0.738	0.734	0.740	0.732
Kawasaki M63383	0.710	0.886	0.885	0.713	0.886	0.883	0.878	0.887	0.878
Shimokoshi M63381	0.672	0.690	0.689	0.668	0.690	0.686	0.680	0.688	0.684

The bold number represented the highest values in each samples**.**

Table 5. The sequence identity matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The identity values were calculated on BioEdit program (cont.) (page 8 of 10).

Patients / References	CHUP65	Si020	Si021	Si022	Si023	Si024	Si025	Si026	Si027	Si028
Karp M33004	0.713	0.919	0.735	0.737	**0.989**	**0.989**	0.919	0.919	0.736	0.742
Karp UT76	0.712	**0.997**	0.735	0.737	0.921	0.921	**0.997**	**0.997**	0.736	0.742
Karp FPW2031	0.721	0.915	0.736	0.736	0.974	0.974	0.915	0.915	0.737	0.741
Kato M63382	0.797	0.705	0.707	0.709	0.721	0.721	0.705	0.705	0.708	0.714
Gilliam FPW2016	0.666	0.743	0.941	0.943	0.733	0.733	0.743	0.743	0.942	0.947
Gilliam UT196	0.682	0.743	**0.992**	**0.992**	0.745	0.745	0.743	0.743	**0.993**	**0.996**
TA686 U80635	0.779	0.73	0.716	0.718	0.746	0.746	0.730	0.730	0.717	0.723
TA763 U80636	0.740	0.844	0.723	0.725	0.859	0.859	0.844	0.844	0.724	0.731
TA678 U19904	0.782	0.739	0.684	0.686	0.744	0.744	0.739	0.739	0.685	0.692
TA716 U19905	**0.850**	0.703	0.682	0.684	0.716	0.716	0.703	0.703	0.683	0.689
Kuroki_ Boryong isolate L04956	0.730	0.875	0.733	0.733	0.910	0.910	0.875	0.875	0.734	0.738
JG_Ikeda isolate AF173033	0.676	0.731	0.941	0.941	0.732	0.732	0.731	0.731	0.942	0.946
Kuroki_ Kuroki isolate M63380	0.732	0.878	0.734	0.734	0.913	0.913	0.878	0.878	0.735	0.739
Kawasaki M63383	0.665	0.715	0.878	0.881	0.712	0.712	0.715	0.715	0.879	0.886
Shimokoshi M63381	0.642	0.664	0.682	0.684	0.673	0.673	0.664	0.664	0.685	0.690

The bold number represented the highest values in each samples**.**

Table 5. The sequence identity matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The identity values were calculated on BioEdit program (cont.) (page 9 of 10).

Patients / References	Karp M33004	Karp UT76	Karp FPW2031	Kato M63382	Gilliam FPW2016	Gilliam UT196	TA686 U80635	TA763 U80636
Karp M33004	**ID**	0.918	0.964	0.713	0.731	0.742	0.741	0.850
Karp UT76	0.918	**ID**	0.913	0.703	0.742	0.742	0.728	0.842
Karp FPW2031	0.964	0.913	**ID**	0.714	0.731	0.743	0.747	0.847
Kato M63382	0.713	0.703	0.714	**ID**	0.700	0.714	0.727	0.711
Gilliam FPW2016	0.731	0.742	0.731	0.700	**ID**	0.948	0.709	0.720
Gilliam UT196	0.742	0.742	0.743	0.714	0.948	**ID**	0.723	0.731
TA686 U80635	0.741	0.728	0.747	0.727	0.709	0.723	**ID**	0.763
TA763 U80636	0.850	0.842	0.847	0.711	0.720	0.731	0.763	**ID**
TA678 U19904	0.738	0.737	0.741	0.801	0.680	0.692	0.748	0.743
TA716 U19905	0.710	0.701	0.706	0.850	0.675	0.689	0.792	0.739
Kuroki_ Boryong isolate L04956	0.901	0.873	0.925	0.714	0.727	0.740	0.762	0.838
JG_Ikeda isolate AF173033	0.729	0.730	0.732	0.705	0.906	0.949	0.713	0.726
Kuroki_ Kuroki isolate M63380	0.904	0.876	0.928	0.715	0.728	0.741	0.764	0.84
Kawasaki M63383	0.709	0.714	0.708	0.697	0.850	0.886	0.695	0.707
Shimokoshi M63381	0.667	0.662	0.669	0.653	0.672	0.689	0.673	0.662

The bold number represented the highest values in each samples**.**

Table 5. The sequence identity matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The identity values were calculated on BioEdit program (cont.) (page 10 of 10).

Patients / References	TA678 U19904	TA716 U19905	Kuroki_ Boryong isolate L04956	JG_Ikeda isolate AF173033	Kuroki_ Kuroki isolate M63380	Kawasaki M63383	Shimokoshi M63381
Karp M33004	0.738	0.710	0.901	0.729	0.904	0.709	0.667
Karp UT76	0.737	0.701	0.873	0.730	0.876	0.714	0.662
Karp FPW2031	0.741	0.706	0.925	0.732	0.928	0.708	0.669
Kato M63382	0.801	0.850	0.714	0.705	0.715	0.697	0.653
Gilliam FPW2016	0.680	0.675	0.727	0.906	0.728	0.850	0.672
Gilliam UT196	0.692	0.689	0.740	0.949	0.741	0.886	0.689
TA686 U80635	0.748	0.792	0.762	0.713	0.764	0.695	0.673
TA763 U80636	0.743	0.739	0.838	0.726	0.84	0.707	0.662
TA678 U19904	**ID**	0.822	0.742	0.685	0.745	0.675	0.655
TA716 U19905	0.822	**ID**	0.710	0.687	0.712	0.669	0.645
Kuroki_ Boryong isolate L04956	0.742	0.710	**ID**	0.727	0.994	0.702	0.664
JG_Ikeda isolate AF173033	0.685	0.687	0.727	**ID**	0.728	0.880	0.676
Kuroki_ Kuroki isolate M63380	0.745	0.712	0.994	0.728	**ID**	0.703	0.665
Kawasaki M63383	0.675	0.669	0.702	0.880	0.703	**ID**	0.676
Shimokoshi M63381	0.655	0.645	0.664	0.676	0.665	0.676	**ID**

The bold number represented the highest values in each samples**.**

eight nucleotide sequences showed the highest identity values when compared with the prototype nucleotide sequences whereas 10 nucleotide sequences generated the highest identity scores when compared to original Thai strains, TA763 (9 nucleotide sequences) and TA716 (1 nucleotide sequence). Among 68 nucleotide sequences, thirty-four nucleotide sequences were highly similar to Karp strain. Thirty-one nucleotide sequences showed the high identity values to Gilliam while only 3 nucleotide sequences were similar to Kato strain. The ranges of high identity scores were shown in the comparison to prototype strains as 0.924-0.997, 0.961-0.965 and 0.913-0.996 for Karp, Kato and Gilliam, respectively. On the other hand, the highest scores in the lower ranges were shown in comparison of TA763 and TA716 as 0.776-0.967 and 0.850, respectively. However, the highest scores were not shown when compared to other stains, Kuroki, Ikeda, Boryong, Kawasaki and Shimokoshi which were originally identified in Japan or Korea.

The identity scores of nucleotide sequence of samples compared to the references sequences were corresponded to the distance matrix evaluated by the MEGA4 program (71) (table 6). The lowest value in each sample was grouped in the same strains of identitying evaluation. There were 34 nucleotide sequences displayed the lowest distance matrix in comparison to Karp strain with the range of 0.000-0.042. Moreover, 31 and 3 nucleotide sequences showed the lowest distance value to Gilliam and Kato with the range of 0.004-0.068 and 0.039-0.043, respectively. For Thai strains, 9 nucleotide sequences showed the lowest scores to TA763 (0.030-0.207) whereas 1 nucleotide sequence displayed the lowest score to TA716 (0.159) strain.

The dendrogram was constructed from the distance values (table 6) based on the Neighbor-Joining (N-J) algorithms. The reliability of branches was tested by 1,000 bootstraps then the 1,000 topologies of tree were generated. However, the slightly different was found in dendrogram generated by the MEGA4 program (71) as shown in figure 7. The most frequently generated topology was selected to presented on a dendrogram and the percentage of reconstructed times was noted on the branch. At least 4 clusters were divided and demonstrated on a dendrogram. They were Karp, Gilliam, TA763 and Kato.

Karp cluster contained 34 nucleotide sequences corresponded with the group based on the highest and lowest of identity and distance scores, respectively.

Table 6. The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The distance values were calculated based on the maximum likelihood algorithm available on the MEGA4 program (page 1 of 10).

Patients / References	KP111	BP288	KP112	KP116	KP68	KP159	KP171	KP133	KP67	KP174
Karp M33004	0.219	0.017	0.230	0.044	0.051	**0.018**	**0.004**	0.225	**0.004**	**0.004**
Karp UT76	0.225	0.046	0.236	**0.000**	0.050	0.044	0.042	0.230	0.042	0.042
Karp FPW2031	0.229	**0.014**	0.237	0.051	**0.042**	0.028	0.019	0.224	0.019	0.019
Kato M63382	0.222	0.215	0.233	0.227	0.220	0.208	0.217	0.255	0.217	0.217
Gilliam FPW2016	0.009	0.217	0.018	0.218	0.227	0.208	0.212	0.238	0.212	0.212
Gilliam UT196	**0.004**	0.224	**0.010**	0.225	0.231	0.213	0.217	0.238	0.217	0.217
TA686 U80635	0.227	0.226	0.242	0.240	0.224	0.226	0.231	0.282	0.231	0.231
TA763 U80636	0.241	0.120	0.252	0.120	0.135	0.113	0.114	**0.207**	0.114	0.114
TA678 U19904	0.261	0.188	0.270	0.195	0.190	0.182	0.192	0.279	0.192	0.192
TA716 U19905	0.267	0.239	0.277	0.250	0.232	0.231	0.238	0.277	0.238	0.238
Kuroki_ Boryong isolate L04956	0.235	0.069	0.242	0.087	0.077	0.080	0.073	0.218	0.073	0.073
JG_Ikeda isolate AF173033	0.051	0.232	0.059	0.235	0.234	0.221	0.229	0.262	0.229	0.229
Kuroki_ Kuroki isolate M63380	0.233	0.064	0.240	0.083	0.073	0.076	0.069	0.216	0.069	0.069
Kawasaki M63383	0.094	0.240	0.103	0.240	0.250	0.234	0.236	0.249	0.236	0.236
Shimokoshi M63381	0.278	0.294	0.290	0.313	0.305	0.298	0.298	0.324	0.298	0.298

The bold number represented the lowest distance values in each samples.

Table 6. The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The distance values were calculated based on the maximum likelihood algorithm available on the MEGA4 program (cont.) (page 2 of 10).

Patients / References	KP20	KP137	KP115	RAP117	KP66	BP269	KP155	CHUP50	BP5	BP41
Karp M33004	0.150	**0.005**	0.044	0.150	**0.017**	0.044	**0.004**	0.254	0.150	0.044
Karp UT76	0.160	0.043	**0.000**	0.160	0.043	**0.000**	0.042	0.260	0.160	**0.000**
Karp FPW2031	0.136	0.020	0.051	0.136	0.027	0.051	0.019	0.266	0.136	0.051
Kato M63382	0.253	0.219	0.227	0.253	0.210	0.227	0.217	0.249	0.253	0.227
Gilliam FPW2016	0.279	0.213	0.218	0.279	0.210	0.218	0.212	**0.068**	0.279	0.218
Gilliam UT196	0.283	0.219	0.225	0.283	0.215	0.225	0.217	0.070	0.283	0.225
TA686 U80635	0.175	0.233	0.240	0.175	0.226	0.240	0.231	0.242	0.175	0.240
TA763 U80636	**0.083**	0.116	0.120	**0.083**	0.111	0.120	0.114	0.286	**0.083**	0.120
TA678 U19904	0.204	0.194	0.195	0.204	0.182	0.195	0.192	0.282	0.204	0.195
TA716 U19905	0.210	0.240	0.250	0.210	0.231	0.250	0.238	0.293	0.210	0.250
Kuroki_ Boryong isolate L04956	0.145	0.074	0.087	0.145	0.079	0.087	0.073	0.281	0.145	0.087
JG_Ikeda isolate AF173033	0.275	0.230	0.235	0.275	0.223	0.235	0.229	0.093	0.275	0.235
Kuroki_ Kuroki isolate M63380	0.142	0.070	0.083	0.142	0.074	0.083	0.069	0.279	0.142	0.083
Kawasaki M63383	0.280	0.235	0.240	0.280	0.236	0.240	0.326	0.118	0.280	0.240
Shimokoshi M63381	0.318	0.298	0.313	0.318	0.300	0.313	0.298	0.301	0.318	0.313

The bold number represented the lowest distance values in each samples.

Table 6. The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The distance values were calculated based on the maximum likelihood algorithm available on the MEGA4 program (cont.) (page 3 of 10).

Patients / References	BP46	BP64	BP100	BP107	BP143	BP151	BP259	BP263	RAP8	RAP39
Karp M33004	0.219	**0.018**	0.044	0.156	0.150	**0.017**	0.219	0.219	0.221	0.044
Karp UT76	0.225	0.044	**0.000**	0.166	0.160	0.043	0.225	0.225	0.227	**0.000**
Karp FPW2031	0.229	0.028	0.051	0.142	0.136	0.027	0.229	0.229	0.231	0.051
Kato M63382	0.222	0.210	0.227	0.260	0.253	0.210	0.222	0.222	0.226	0.227
Gilliam FPW2016	0.009	0.212	0.218	0.287	0.279	0.210	0.009	0.009	0.011	0.218
Gilliam UT196	**0.004**	0.217	0.225	0.291	0.283	0.215	**0.004**	**0.004**	**0.006**	0.225
TA686 U80635	0.227	0.228	0.240	0.182	0.175	0.226	0.227	0.227	0.231	0.240
TA763 U80636	0.241	0.113	0.120	**0.089**	**0.083**	0.111	0.241	0.241	0.245	0.120
TA678 U19904	0.261	0.182	0.195	0.211	0.204	0.182	0.261	0.261	0.265	0.195
TA716 U19905	0.267	0.231	0.250	0.217	0.210	0.231	0.267	0.267	0.271	0.250
Kuroki_ Boryong isolate L04956	0.235	0.080	0.087	0.152	0.145	0.079	0.235	0.235	0.237	0.087
JG_Ikeda isolate AF173033	0.051	0.225	0.235	0.282	0.275	0.223	0.051	0.051	0.054	0.235
Kuroki_ Kuroki isolate M63380	0.233	0.076	0.083	0.148	0.142	0.074	0.233	0.233	0.235	0.083
Kawasaki M63383	0.094	0.238	0.240	0.284	0.280	0.236	0.094	0.094	0.097	0.240
Shimokoshi M63381	0.278	0.302	0.313	0.326	0.318	0.300	0.278	0.278	0.282	0.313

The bold number represented the lowest distance values in each samples.

Table 6. The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The distance values were calculated based on the maximum likelihood algorithm available on the MEGA4 program (cont.) (page 4 of 10).

Patients / References	RAP64	CHUP20	CHUP30	CHUP40	CHUP68	KL36	KL53	KL102	BL8
Karp M33004	0.231	0.131	0.219	0.227	0.131	0.219	**0.005**	0.044	0.223
Karp UT76	0.230	0.143	0.225	0.232	0.139	0.225	0.043	**0.000**	0.229
Karp FPW2031	0.232	0.123	0.229	0.236	0.119	0.229	0.020	0.051	0.233
Kato M63382	**0.043**	0.238	0.222	0.228	0.242	0.222	0.219	0.227	0.226
Gilliam FPW2016	0.223	0.242	0.009	**0.013**	0.256	0.009	0.214	0.218	0.011
Gilliam UT196	0.225	0.244	**0.004**	0.020	0.260	**0.004**	0.219	0.225	**0.006**
TA686 U80635	0.250	0.171	0.227	0.235	0.169	0.227	0.233	0.240	0.231
TA763 U80636	0.243	**0.030**	0.241	0.245	**0.063**	0.241	0.116	0.120	0.245
TA678 U19904	0.163	0.195	0.261	0.257	0.196	0.261	0.194	0.195	0.265
TA716 U19905	0.147	0.204	0.267	0.269	0.198	0.267	0.240	0.250	0.271
Kuroki_ Boryong isolate L04956	0.246	0.119	0.235	0.246	0.125	0.235	0.075	0.087	0.239
JG_Ikeda isolate AF173033	0.234	0.242	0.051	0.062	0.254	0.051	0.230	0.235	0.054
Kuroki_ Kuroki isolate M63380	0.244	0.115	0.233	0.243	0.122	0.233	0.070	0.083	0.237
Kawasaki M63383	0.241	0.252	0.094	0.101	0.273	0.094	0.238	0.240	0.097
Shimokoshi M63381	0.300	0.310	0.278	0.288	0.308	0.278	0.300	0.313	0.282

The bold number represented the lowest distance values in each samples.

Table 6. The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The distance values were calculated based on the maximum likelihood algorithm available on the MEGA4 program (cont.) (page 5 of 10).

Patients / References	BL9	BL133	BP13	RAP25	BP120	RAP22	Si001	Si002	Si005	Si006
Karp M33004	0.044	**0.004**	0.229	**0.005**	**0.004**	0.219	0.044	0.150	0.222	0.219
Karp UT76	**0.000**	0.042	0.228	0.043	0.042	0.225	**0.000**	0.160	0.229	0.225
Karp FPW2031	0.051	0.019	0.230	0.020	0.019	0.229	0.051	0.136	0.233	0.229
Kato M63382	0.227	0.217	**0.039**	0.215	0.217	0.222	0.227	0.253	0.227	0.222
Gilliam FPW2016	0.218	0.212	0.220	0.213	0.212	0.009	0.218	0.279	0.013	0.009
Gilliam UT196	0.225	0.217	0.222	0.219	0.217	**0.004**	0.225	0.283	**0.008**	**0.004**
TA686 U80635	0.240	0.231	0.246	0.233	0.231	0.227	0.240	0.175	0.234	0.227
TA763 U80636	0.120	0.114	0.241	0.116	0.114	0.241	0.120	**0.083**	0.244	0.241
TA678 U19904	0.195	0.192	0.161	0.190	0.192	0.261	0.195	0.204	0.264	0.261
TA716 U19905	0.250	0.238	0.145	0.236	0.238	0.267	0.250	0.210	0.271	0.267
Kuroki_ Boryong isolate L04956	0.087	0.073	0.244	0.075	0.073	0.235	0.087	0.145	0.238	0.235
JG_Ikeda isolate AF173033	0.235	0.229	0.230	0.230	0.229	0.051	0.235	0.275	0.056	0.051
Kuroki_ Kuroki isolate M63380	0.083	0.069	0.242	0.070	0.069	0.233	0.083	0.142	0.236	0.233
Kawasaki M63383	0.240	0.236	0.239	0.238	0.236	0.094	0.240	0.280	0.097	0.094
Shimokoshi M63381	0.313	0.298	0.300	0.300	0.298	0.278	0.313	0.318	0.288	0.278

The bold number represented the lowest distance values in each samples.

Table 6. The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The distance values were calculated based on the maximum likelihood algorithm available on the MEGA4 program (cont.) (page 6 of 10).

Patients / References	Si007	Si008	Si009	Si010	Si011	Si012	Si013	Si015	Si016	Si018
Karp M33004	0.227	0.219	0.224	0.034	0.219	0.219	0.219	0.014	**0.028**	0.219
Karp UT76	0.227	0.225	0.230	0.041	0.225	0.225	0.225	0.043	0.040	0.225
Karp FPW2031	0.228	0.229	0.231	**0.031**	0.229	0.229	0.229	**0.011**	0.036	0.229
Kato M63382	**0.040**	0.222	0.229	0.211	0.222	0.222	0.222	0.211	0.217	0.222
Gilliam FPW2016	0.220	0.009	0.014	0.211	0.009	0.009	0.009	0.213	0.219	0.009
Gilliam UT196	0.222	**0.004**	**0.006**	0.218	**0.004**	**0.004**	**0.004**	0.220	0.224	**0.004**
TA686 U80635	0.246	0.227	0.236	0.212	0.227	0.227	0.227	0.224	0.237	0.227
TA763 U80636	0.240	0.241	0.246	0.111	0.241	0.241	0.241	0.117	0.116	0.241
TA678 U19904	0.160	0.261	0.266	0.179	0.261	0.261	0.261	0.185	0.189	0.261
TA716 U19905	0.143	0.267	0.272	0.228	0.267	0.267	0.267	0.236	0.240	0.267
Kuroki_ Boryong isolate L04956	0.242	0.235	0.237	0.069	0.235	0.235	0.235	0.066	0.084	0.235
JG_Ikeda isolate AF173033	0.230	0.051	0.055	0.226	0.051	0.051	0.051	0.228	0.232	0.051
Kuroki_ Kuroki isolate M63380	0.240	0.233	0.235	0.065	0.233	0.233	0.233	0.062	0.080	0.233
Kawasaki M63383	0.237	0.094	0.098	0.230	0.094	0.094	0.094	0.236	0.242	0.094
Shimokoshi M63381	0.296	0.278	0.286	0.291	0.278	0.278	0.278	0.290	0.314	0.278

The bold number represented the lowest distance values in each samples.

Table 6. The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The distance values were calculated based on the maximum likelihood algorithm available on the MEGA4 program (cont.) (page 7 of 10).

Patients / References	RAP13	BP32	BP119	Si019	CHUP140	RAP107	KP128	CHUP138	RAP36
Karp M33004	0.035	0.219	0.221	0.036	0.219	0.220	0.233	0.217	0.232
Karp UT76	0.035	0.225	0.227	0.044	0.225	0.227	0.240	0.223	0.238
Karp FPW2031	**0.032**	0.229	0.231	**0.034**	0.229	0.231	0.240	0.227	0.242
Kato M63382	0.213	0.222	0.224	0.209	0.222	0.225	0.238	0.222	0.233
Gilliam FPW2016	0.213	0.009	0.010	0.210	0.009	0.014	0.020	0.009	0.019
Gilliam UT196	0.220	**0.004**	**0.005**	0.216	**0.004**	**0.009**	**0.013**	**0.004**	**0.017**
TA686 U80635	0.221	0.227	0.229	0.218	0.227	0.236	0.246	0.229	0.236
TA763 U80636	0.119	0.241	0.243	0.121	0.241	0.244	0.256	0.239	0.252
TA678 U19904	0.185	0.261	0.263	0.184	0.261	0.262	0.276	0.259	0.271
TA716 U19905	0.234	0.267	0.269	0.228	0.267	0.269	0.282	0.265	0.277
Kuroki_ Boryong isolate L04956	0.072	0.235	0.237	0.073	0.235	0.236	0.246	0.233	0.248
JG_Ikeda isolate AF173033	0.228	0.051	0.052	0.225	0.051	0.058	0.062	0.052	0.060
Kuroki_ Kuroki isolate M63380	0.067	0.233	0.235	0.069	0.233	0.234	0.244	0.231	0.246
Kawasaki M63383	0.240	0.094	0.095	0.236	0.094	0.098	0.106	0.093	0.104
Shimokoshi M63381	0.302	0.278	0.280	0.303	0.278	0.286	0.296	0.282	0.288

The bold number represented the lowest distance values in each samples.

Table 6. The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The distance values were calculated based on the maximum likelihood algorithm available on the MEGA4 program (cont.) (page 8 of 10).

Patients / References	CHUP65	Si020	Si021	Si022	Si023	Si024	Si025	Si026	Si027	Si028
Karp M33004	0.233	0.044	0.231	0.228	**0.004**	**0.004**	0.044	0.044	0.230	0.219
Karp UT76	0.227	**0.000**	0.238	0.234	0.042	0.042	**0.000**	**0.000**	0.236	0.225
Karp FPW2031	0.226	0.051	0.239	0.238	0.019	0.019	0.051	0.051	0.237	0.229
Kato M63382	0.195	0.227	0.235	0.231	0.217	0.217	0.227	0.227	0.233	0.222
Gilliam FPW2016	0.271	0.218	0.017	0.014	0.212	0.212	0.218	0.218	0.015	0.009
Gilliam UT196	0.273	0.225	**0.009**	**0.009**	0.217	0.217	0.225	0.225	**0.008**	**0.004**
TA686 U80635	0.195	0.240	0.240	0.236	0.231	0.231	0.240	0.240	0.238	0.227
TA763 U80636	0.186	0.120	0.254	0.250	0.114	0.114	0.120	0.120	0.252	0.241
TA678 U19904	0.195	0.195	0.274	0.270	0.192	0.192	0.195	0.195	0.272	0.261
TA716 U19905	**0.159**	0.250	0.281	0.277	0.238	0.238	0.250	0.250	0.279	0.267
Kuroki_ Boryong isolate L04956	0.216	0.087	0.244	0.244	0.073	0.073	0.087	0.087	0.243	0.235
JG_Ikeda isolate AF173033	0.272	0.235	0.058	0.058	0.229	0.229	0.235	0.235	0.056	0.051
Kuroki_ Kuroki isolate M63380	0.212	0.083	0.242	0.242	0.069	0.069	0.083	0.083	0.241	0.233
Kawasaki M63383	0.283	0.240	0.104	0.101	0.236	0.236	0.240	0.240	0.103	0.094
Shimokoshi M63381	0.335	0.313	0.294	0.290	0.298	0.298	0.313	0.313	0.288	0.278

The bold number represented the lowest distance values in each samples.

Table 6. The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The distance values were calculated based on the maximum likelihood algorithm available on the MEGA4 program (cont.) (page 9 of 10).

Patients / References	Karp M33004	Karp UT76	Karp FPW2031	Kato M63382	Gilliam FPW2016	Gilliam UT196	TA686 U80635	TA763 U80636
Karp M33004								
Karp UT76	0.044							
Karp FPW2031	0.023	0.051						
Kato M63382	0.223	0.227	0.223					
Gilliam FPW2016	0.214	0.218	0.222	0.220				
Gilliam UT196	0.219	0.225	0.226	0.222	0.008			
TA686 U80635	0.235	0.240	0.237	0.238	0.224	0.227		
TA763 U80636	0.119	0.120	0.125	0.230	0.234	0.241	0.174	
TA678 U19904	0.197	0.195	0.192	0.155	0.253	0.261	0.231	0.189
TA716 U19905	0.244	0.250	0.248	0.127	0.265	0.267	0.178	0.193
Kuroki_ Boryong isolate L04956	0.077	0.087	0.065	0.236	0.231	0.232	0.221	0.134
JG_Ikeda isolate AF173033	0.230	0.235	0.234	0.229	0.048	0.048	0.234	0.239
Kuroki_ Kuroki isolate M63380	0.073	0.083	0.060	0.234	0.229	0.229	0.217	0.130
Kawasaki M63383	0.238	0.240	0.249	0.231	0.091	0.094	0.247	0.247
Shimokoshi M63381	0.304	0.313	0.306	0.299	0.277	0.280	0.303	0.312

The bold number represented the lowest distance values in each samples.

Table 6. The pairwise distance matrix of 78 nucleotide sequences which encompassed 4 variable regions (VD I-IV) compared to 15 reference nucleotide sequences obtained from GenBank database. The distance values were calculated based on the maximum likelihood algorithm available on the MEGA4 program (cont.) (page 10 of 10).

References \ Patients	TA678 U19904	TA716 U19905	Kuroki_ Boryong isolate L04956	JG_Ikeda isolate AF173033	Kuroki_ Kuroki isolate M63380	Kawasaki M63383	Shimokoshi M63381
Karp M33004							
Karp UT76							
Karp FPW2031							
Kato M63382							
Gilliam FPW2016							
Gilliam UT196							
TA686 U80635							
TA763 U80636							
TA678 U19904							
TA716 U19905	0.144						
Kuroki_ Boryong isolate L04956	0.200	0.248					
JG_Ikeda isolate AF173033	0.262	0.261	0.243				
Kuroki_ Kuroki isolate M63380	0.195	0.244	0.006	0.241			
Kawasaki M63383	0.270	0.285	0.264	0.090	0.262		
Shimokoshi M63381	0.310	0.334	0.316	0.294	0.313	0.289	

The bold number represented the lowest distance values in each samples.

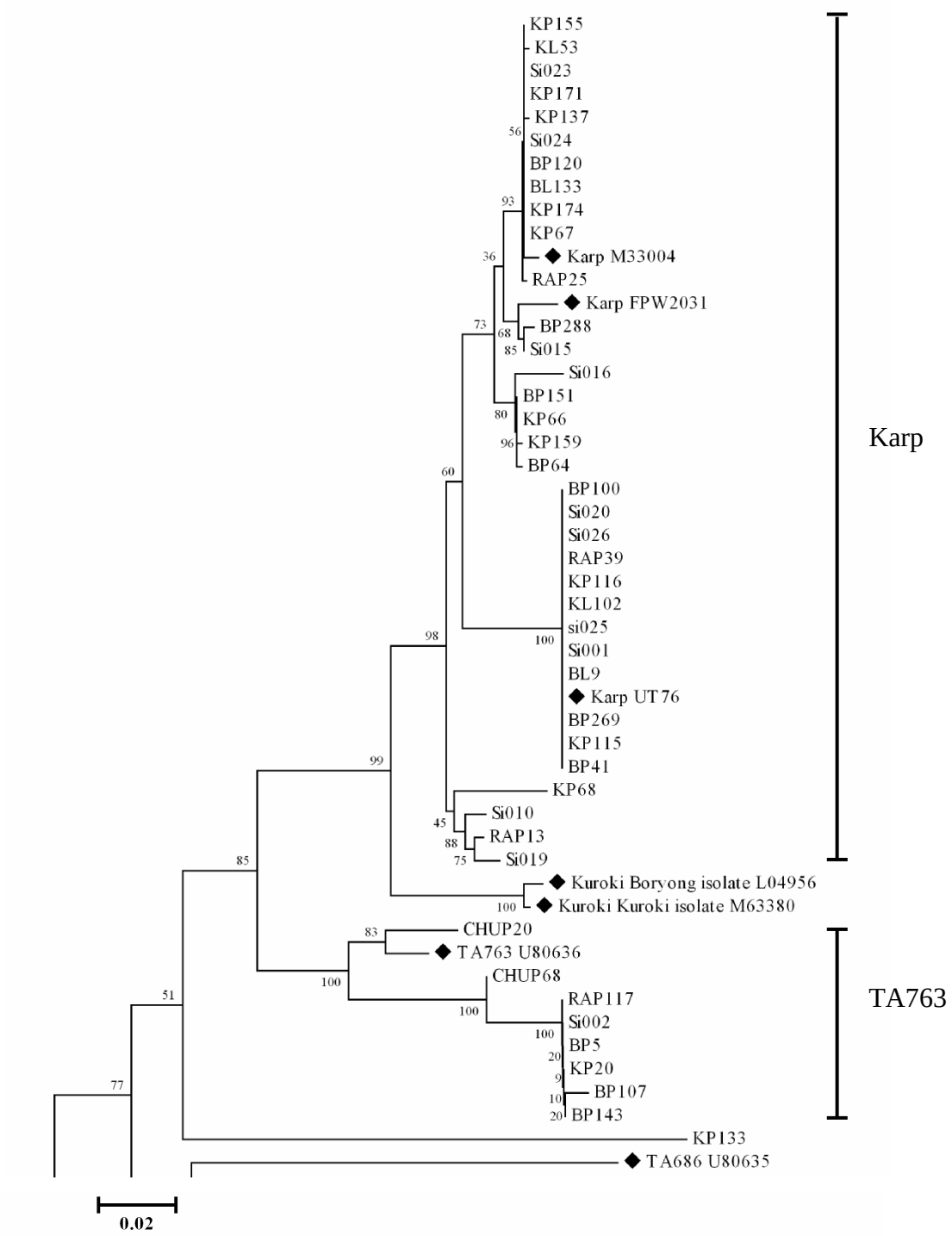


Figure 7. A dendrogram constructed from the alignment of VD I to IV regions of the 56-kDa protein encoding gene based on the N-J algorithm with 1,000 bootstrap values. The *O. tsutsugamushi* with filled diamond represented the reference *O. tsutsugamushi* strains and the strains reported from Asia-Pacific geographical regions obtained from GenBank Database. The numbers at nodes demonstrated the bootstrap values (page 1 of 2).

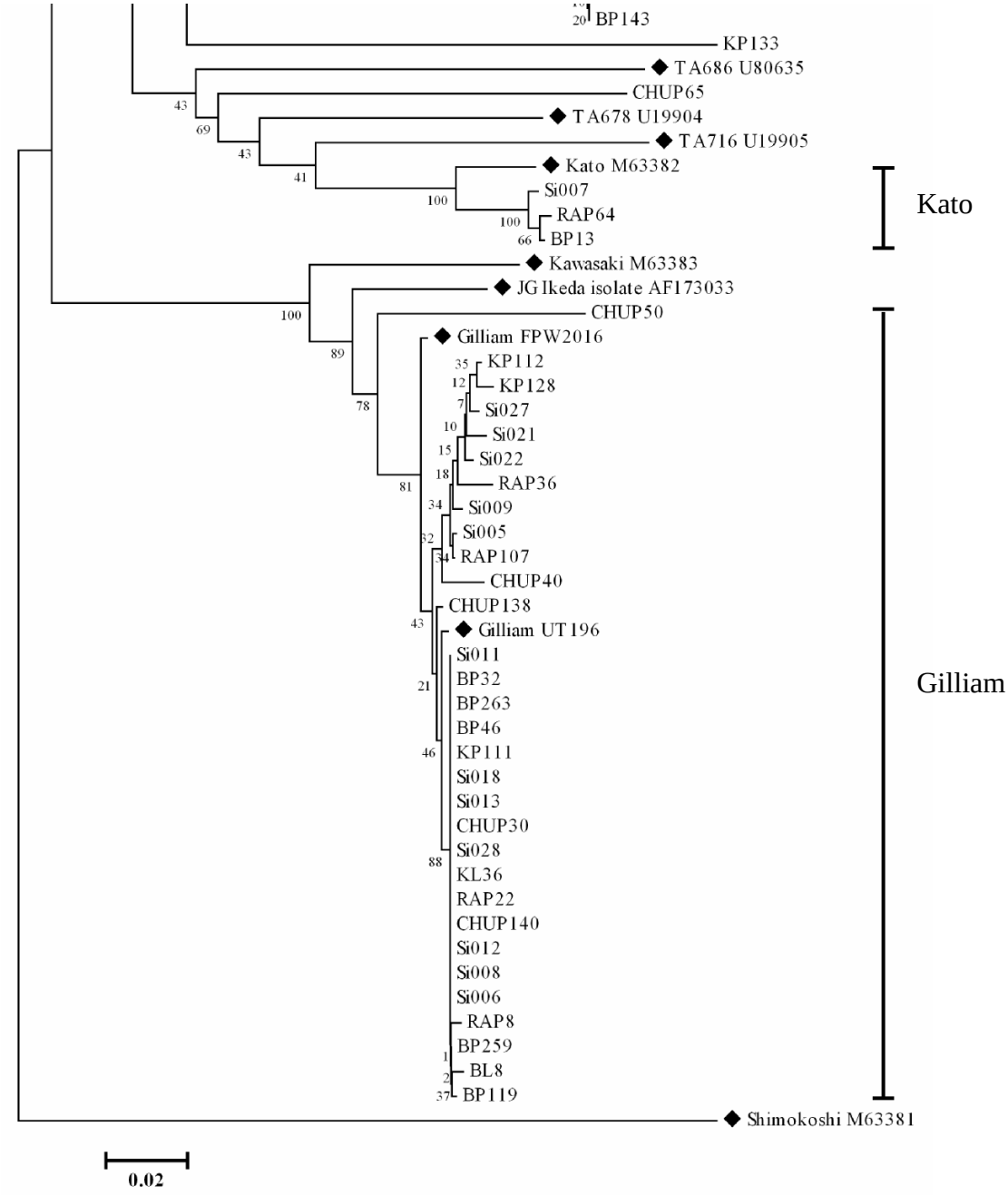


Figure 7. A dendrogram constructed from the alignment of VD I to IV regions of the 56-kDa protein encoding gene based on the N-J algorithm with 1,000 bootstrap values. The *O. tsutsugamushi* with filled diamond represented the reference *O. tsutsugamushi* strains and the strains reported from Asia-Pacific geographical regions obtained from GenBank Database. The numbers at nodes demonstrated the bootstrap values (cont.) (page 2 of 2).

This cluster was bifurcated from two reference nucleotide sequences (Boryong and Kuroki isolate of Kuroki strain). The bootstrap value showed the high percentage at 99% indicated the significant reliability or confidence pattern of this bifurcation pattern.

There were 31 nucleotide sequences in Gilliam cluster corresponded to the scores of identity and distance matrix. The bootstrap value of 100% at the bifurcation of sequences differentiated from Kawasaki strain indicated the significant pattern. Moreover, 89% at the lower node demonstrated that most Gilliam-characterized sequences were similar to the Thai-Gilliam (FPW and UT isolated from Thailand) than the Japanese-Gilliam Ikeda isolate (JG Ikeda isolated from Japan, AF173033). Interestingly, the nucleotide sequence of CHUP50 sample was separated from Gilliam cluster. However, CHUP50 showed the highest identity matrix at 0.913 then it was classified in the Gilliam cluster.

There were 3 nucleotide sequences which showed the highest identity and lowest distance values to Kato strain. These nucleotide sequences were classified in Kato cluster of a dendrogram. The bootstrap value at the node of TA716 and Kato bifurcation was 41% indicating the less confidence of this pattern of branch.

For TA763, the identity and distance matrix demonstrated that 9 nucleotide sequences presented the high relationship to TA763 strain. However, there were only 8 nucleotide sequences clustered in this group with the confidence as 85% of bootstrapping, except sample KP133. With 51% of bootstrap values, it was the low confidence to consider that KP133 was an outgroup of TA763. Although the identity matrix of KP133 was 0.776, the lowest range in this group. Blastn program was performed to confirm the group of KP133. The result showed that there were 2 submitted nucleotide sequences had the high identity to the query KP133 which were *O. tsutsugamushi* strain TT03-1 (accession no. GU120168) and TT0705a (GQ332753) (figure 8). Both nucleotide sequences were the 56-kDa protein encoding gene which isolated from patient's blood sample in Taiwan.

The last nucleotide sequences, CHUP65, it was considered as high relationship with TA716 strain when the highest identity score (0.850) comparing to other reference nucleotide sequences. Moreover, the distance score at 0.159 was the lowest value comparing to others. However, a dendrogram showed that CHUP65 was

Sequences producing significant alignments:	Score (Bits)	E Value	Ident (%)
gb|GU120168.1| *O. tsutsugamushi* strain TT03-1 56-kDa ty...	1764	0.0	98
gb|GQ332753.1| *O. tsutsugamushi* strain TT0705a 56-kDa t...	1764	0.0	98
gb|GU120147.1| *O. tsutsugamushi* strain KM02 56-kDa type...	900	0.0	85
gb|DQ514319.1| *O. tsutsugamushi* strain Neimeng-65 56 kD...	900	0.0	85
gb|DQ485289.1| *O. tsutsugamushi* isolate Taiwan 56kDa ty...	883	0.0	85
gb|GU120169.1| *O. tsutsugamushi* strain TT03-2 56-kDa ty...	865	0.0	92
gb|EF213095.1| *O. tsutsugamushi* strain UT302 56-kDa typ...	839	0.0	87
gb|DQ323175.1| *O. tsutsugamushi* strain Hualien-11 56-kD...	839	0.0	87
gb|GU120173.1| *O. tsutsugamushi* strain TT06-6 56-kDa ty...	833	0.0	87
gb|GU120172.1| *O. tsutsugamushi* strain TT06-1 56-kDa ty...	833	0.0	87
gb|GU120162.1| *O. tsutsugamushi* strain PH05 56-kDa type...	833	0.0	87
gb|GU120153.1| *O. tsutsugamushi* strain KM08 56-kDa type...	833	0.0	87
gb|GU120160.1| *O. tsutsugamushi* strain PH03 56-kDa type...	828	0.0	86
gb|GQ332752.1| *O. tsutsugamushi* strain NT0707a 56-kDa t...	826	0.0	86
gb|AY787232.1| *O. tsutsugamushi* strain Taitung-4 56 kDa...	817	0.0	86
gb|U80636.1|RTU80636 *R. tsutsugamushi* TSA (tsa763) ge...	787	0.0	86
gb|EF213085.1| *O. tsutsugamushi* strain FPW2016 56-kDa t...	756	0.0	82
gb|GU120165.1| *O. tsutsugamushi* strain TT02-1 56-kDa ty...	754	0.0	82
gb|GU120166.1| *O. tsutsugamushi* strain TT02-2 56-kDa ty...	749	0.0	82
gb|GU120149.1| *O. tsutsugamushi* strain KM04 56-kDa type...	749	0.0	86
gb|AY335819.1| *O. tsutsugamushi* strain Taitung-2 56 kDa...	749	0.0	82
gb|GU120171.1| *O. tsutsugamushi* strain TT05 56-kDa type...	743	0.0	82
gb|GU120157.1| *O. tsutsugamushi* strain MZ02 56-kDa type...	743	0.0	82
gb|GQ332760.1| *O. tsutsugamushi* strain KM0606a 56-kDa t...	743	0.0	82
gb|GQ332757.1| *O. tsutsugamushi* strain KHC0707a 56-kDa ...	743	0.0	82
gb|DQ789360.1| *O. tsutsugamushi* strain Hualien-13 56 kD...	743	0.0	82
gb|DQ314548.1| *O. tsutsugamushi* strain Hualien-10 56-kD...	743	0.0	82
gb|AY834393.1| *O. tsutsugamushi* strain Hualien-7 56-kDa...	743	0.0	82
gb|AY834392.1| *O. tsutsugamushi* strain Taitung-5 56-kDa...	743	0.0	82
gb|EF213099.1| *O. tsutsugamushi* strain UT329 56-kDa typ...	741	0.0	82
gb|GU120140.1| *O. tsutsugamushi* strain HL02-1 56-kDa ty...	739	0.0	82
gb|GQ332743.1| *O. tsutsugamushi* strain TY0610a 56-kDa t...	739	0.0	82
gb|GU120164.1| *O. tsutsugamushi* strain TT01-2 56-kDa ty...	737	0.0	82
gb|GQ332745.1| *O. tsutsugamushi* strain TP0708a 56-kDa t...	737	0.0	82
gb|DQ323174.1| *O. tsutsugamushi* strain Hualien-8 56-kDa...	737	0.0	82

Figure 8. Nucleotide comparison of sequence sample no. KP133 using Blastn program.

not in the same clade of TA716 even the low bootstrap value as 69% was presented. Blastn program was used to compare the nucleotide sequence of CHUP65 with databank. The result showed 7 nucleotide sequences generated the maximum identity as 98% to CHUP65. All sequences were submitted from Taiwan isolate; GU120151, AY222633, AY222629, GQ332763, AY222641, AY714315 and GU120163 (figure 9).

2.2 Geographical distribution

The samples in this study were collected from 4 regions of Thailand which are northeastern, central, southern and northern regions as shown in table 1. From 430 samples, there were 78 samples which could be sequenced and 76 samples among these were classified as strain in a dendrogram except sample no.KP133 and CHUP65. The distribution of strains in each region could be analyzed as shown in table 7 and figure 10. The percentage was calculated in each region.

Most of samples in this study were collected from northeastern region. Karp was recognized as the major distributed strain with 55.6%. Gilliam, TA763 and Kato were detected at the frequency of 33.3%, 8.9% and 2.2%, respectively. These four strains could also be found in central region which strain distribution was slightly different from the northeastern region. Gilliam were presented as the major strain (47.8%), and then Karp was found at 34.8%. The distribution of Kato and TA763 could be found at an equal level of 8.7%. In southern and northern region, fewer strains were found in comparison to other regions. Gilliam and TA763 were found in the southern region with 71.4% and 28.6%, respectively. Karp was the only one strain found in the northern region.

3. Strain-specific amplification of *O. tsutsugamushi*

The nucleotide sequences of 56-kDa protein encoding gene within and between clusters were compared and used for strain identification by nested PCR

Sequences producing significant alignments:	Score (Bits)	E Value	Ident %
gb|GU120151.1| *O. tsutsugamushi* strain KM06 56-kDa type..	1801	0.0	98
gb|AY222633.1| *O. tsutsugamushi* strain TW44R 56kDa type...	1801	0.0	98
gb|AY222629.1| *O. tsutsugamushi* strain TW62R 56kDa type...	1801	0.0	98
gb|GQ332763.1| *O. tsutsugamushi* strain KHC0606b 56-kDa ...	1796	0.0	98
gb|AY222641.1| *O. tsutsugamushi* strain TWyu11 56kDa typ...	1796	0.0	98
gb|AY714315.1| *O. tsutsugamushi* strain Hualien-4 56-kDa...	1796	0.0	98
gb|GU120163.1| *O. tsutsugamushi* strain TT01-1 56-kDa ty...	1790	0.0	98
gb|U19905.1|RTU19905 *R. tsutsugamushi* TA716 56 kDa ty...	1024	0.0	86
gb|EF213087.1| *O. tsutsugamushi* strain FPW1038 56-kDa t...	966	0.0	84
gb|GU120142.1| *O. tsutsugamushi* strain HL03-1 56-kDa ty...	852	0.0	85
gb|GQ332751.1| *O. tsutsugamushi* strain KHC0704a 56-kDa ...	852	0.0	85
gb|DQ852664.1| *O. tsutsugamushi* strain Hualien-14 56kDa...	852	0.0	85
gb|AY222635.1| *O. tsutsugamushi* strain TW381 56kDa type...	852	0.0	85
gb|GU120139.1| *O. tsutsugamushi* strain HL01 56-kDa type...	846	0.0	85
gb|GQ332752.1| *O. tsutsugamushi* strain NT0707a 56-kDa t...	846	0.0	85
gb|AY222630.1| *O. tsutsugamushi* strain TW521 56kDa type...	846	0.0	85
gb|GU120141.1| *O. tsutsugamushi* strain HL02-2 56-kDa ty...	841	0.0	85
gb|AF173050.1|AF173050 *O. tsutsugamushi* strain LF-1 56 ...	841	0.0	83
gb|GQ332761.1| *O. tsutsugamushi* strain HC0605a 56-kDa t...	835	0.0	82
gb|EF213095.1| *O. tsutsugamushi* strain UT302 56-kDa typ...	835	0.0	86
gb|AY714317.1| *O. tsutsugamushi* strain Hualien-6 56-kDa...	835	0.0	82
gb|AY836148.1| *O. tsutsugamushi* train Taiwan CDC Kato ...	830	0.0	82
gb|AF173040.1|AF173040 *O. tsutsugamushi* strain Omagari ...	830	0.0	82
gb|M63382.1|RIRTST56A *R. tsutsugamushi* 56 kDa type-sp...	830	0.0	82
gb|AF173041.1|AF173041 *O. tsutsugamushi* strain Akita-7 ...	826	0.0	82
gb|U80635.1|RTU80635 *R. tsutsugamushi* TSA (tsa686) ge...	826	0.0	85
gb|GU120149.1| *O. tsutsugamushi* strain KM04 56-kDa type...	824	0.0	85
gb|GU120148.1| *O. tsutsugamushi* strain KM03 56-kDa type...	824	0.0	82
gb|GQ332762.1| *O. tsutsugamushi* strain KM0607b 56-kDa t...	824	0.0	82
gb|AY636101.1| *O. tsutsugamushi* Hualien-3 56 kDa type-s...	824	0.0	82
gb|U80636.1|RTU80636 *O. tsutsugamushi* TSA (tsa763) ge...	824	0.0	85
gb|AY714316.1| *O. tsutsugamushi* strain Hualien-5 56-kDa...	819	0.0	82
gb|GU120160.1| *O. tsutsugamushi* strain PH03 56-kDa type...	802	0.0	85
gb|GU120143.1| *O. tsutsugamushi* strain HL03-2 56-kDa ty...	802	0.0	81
gb|GU120173.1| *O. tsutsugamushi* strain TT06-6 56-kDa ty...	797	0.0	85

Figure 9. Nucleotide comparison of sequence sample no. CHUP65 using Blastn program.

Table 7. The geographic distribution of *O. tsutsugamushi* strains in Thailand.

Regions(n of sequence)	Strain Distribution			
	Karp, n (%)	Kato, n (%)	Gilliam, n (%)	TA763, n (%)
Northeastern (45)	25 (55.6)	1 (2.2)	15 (33.3)	4 (8.9)
Central (23)	8 (34.8)	2 (8.7)	11 (47.8)	2 (8.7)
Southern (7)	NA[a]	NA[a]	5 (71.4)	2 (28.6)
Northern (1)	1 (100)	NA[a]	NA[a]	NA[a]
Total 76 sequences	34 (44.7)	3 (3.9)	31 (40.8)	8 (10.5)

[a] NA: no sample available

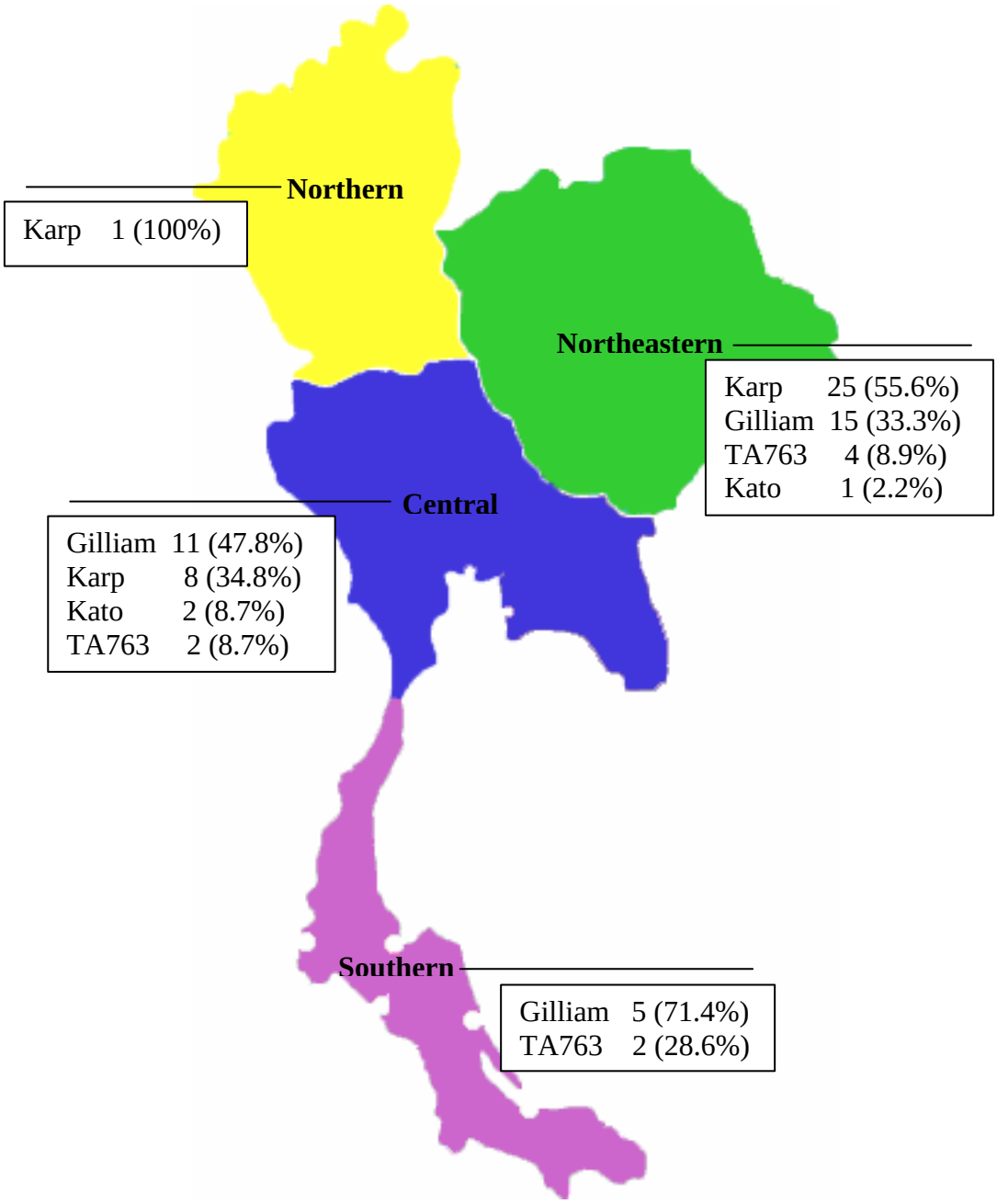


Figure 10. Geographical distribution of *O. tsutsugamushi* strains in 4 regions of Thailand. The percentage in parenthesis was calculated from the total 76 nucleotide sequences.

detection. Primers were designed by manually select the conserved region and the hypervariable which was shown (appendix III). Primers for each stain were evaluated *in silico* by OligoCalc software (69) to avoid the secondary structure formation. Primer-Blast program was used to compare and search for the nucleotide sequences of primers to sequence in databank. The details of primers were shown in table 8.

The nucleotide sequences of the well-characterized strain based on a dendrogram were amplified by all 4 sets of primers by nested PCR to verify the strain-specific primers. Then, the result of PCR amplification was compared with a classification result.

The conditions of nested PCR initiated with the DNA samples. The external primers E1 and E2 (table 2) were used to amplified the 56-kDa protein encoding gene. After that the PCR product was used as the template for the nested PCR amplification using the strain-specific primers. The amount of magnesium chloride ($MgCl_2$) solution and annealing temperature were optimized for each primer set as shown in table 9. However, the other reagents such as 10 µl of 5x Green GoTaq® Flexi Buffer, 10 millimolar (mM) of each of dNTP, 1.25 unit of a thermostable DNA polymerase (GoTaq® DNA polymerase, Promega, USA) and 2 µl of DNA template were used for every reaction. DNase/RNase free sterile water (Ultra PURE™, Gibco, invitrogen) was added up to 50 µl. The PCR condition was initiated at 95°C for 2 minutes, then 35 cycles of DNA denaturation at 95°C for 1 minute, annealing at temperature according to the primer set for 30 second and extension at 72°C for 30 second followed by the final extension at 72°C for 5 minutes in a GeneAmp PCR System 9700 (PE-Applied Biosystem Inc., CA, USA). The PCR product was analysed by 1.5% (w/v) agarose gel electrophoresis. The PCR products were 254, 316, 277 and 194 bp for specific primers of Karp, Kato, Gilliam and TA763, respectively (figure 11).

The efficiency of these specific primers was evaluated by comparing with *O. tsutsugamushi* strains that were identified using dendrogram as a reference method. DNA sample no. KP133 and CHUP65 were excluded because their branches in a dendrogram were not included in any clusters. Eventually, 76 sequences were analyzed for this efficiency testing.

Table 8. Oligonucleotide primers for nested PCR amplification of 56-kDa protein encoding gene of each strain.

Primers	Nucleotide sequences (5'→3')	Positions[a]	Polarity	Product size[b] (bp)
Amplification of Karp strain				
STRAIN_F	CCTCAGCCTACTATAATGCCTATAAG	961-986	sense	254
KARP_R	CCAACAGGATTAGGGTTACCC	1214-1194	antisense	
Amplification of Kato strain				
KATO_F	GCTCTGATGCTGATATTAGATCTGGT	882-908	sense	316
KATO_R	GGCATTAGGCGGACCTTGAGGAA	1197-1176	antisense	
Amplification of Gilliam strain				
STRAIN_F	CCTCAGCCTACTATAATGCCTATAAG	448-473	sense	277
GILLIAM_R	GGTCAAGTATGTTAAGATTTTG	724-703	antisense	
Amplification of TA763 strain				
STRAIN_F	CCTCAGCCTACTATAATGCCTATAAG	409-434	sense	194
763_R	CGCCCAGGATTATTAGGATCCTTC	602-579	antisense	

[a] The positions were given according to the nucleotide sequence of 56-kDa protein encoding gene of *O. tsutsugamushi* in each strain that would be amplified. Primers of Karp strain amplification were compared to Karp reference sequence (M33004). The reference sequences Gilliam FPW2016 (EF213085), TA763 (U80636) and Kato (M63382) were specified the position of priming site of *O. tsutsugamushi* Gilliam, TA763 and Kato strains, respectively.

[b] The product size was evaluated using Primer-Blast program.

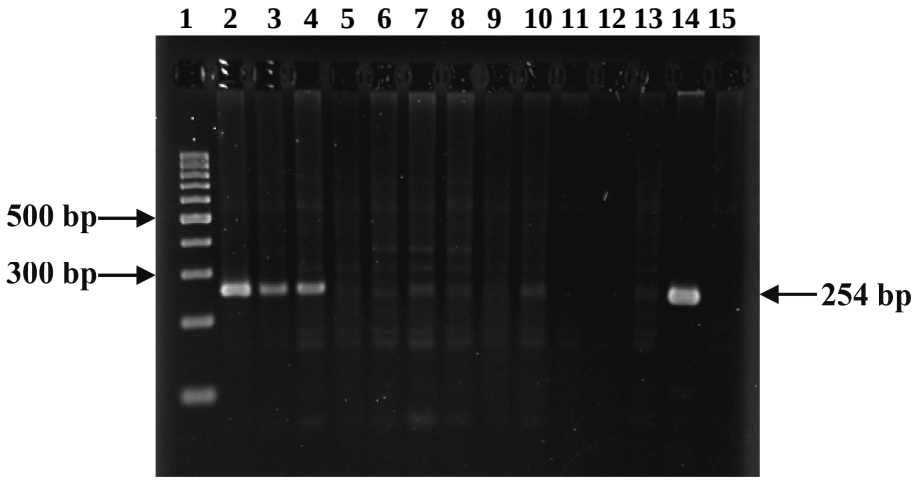
Table 9. The optimal amount of $MgCl_2$ and optimal annealing temperature of each set of primer for *O. tsutsugamushi* strain-specific amplification.

Strain-specific primer set	$MgCl_2$[a] (mM)	Annealing temperature[b] (°C)
Karp	1.0	62
Kato	1.0	63
Gilliam	3.0	62
TA763	1.5	65

[a] 25 mM available in GoTaq® Flexi DNA polymerase, Promega, USA

[b] The reaction was perform on a GeneAmp PCR System 9700 (PE-Applied Biosystem Inc., CA, USA).

A.



B.

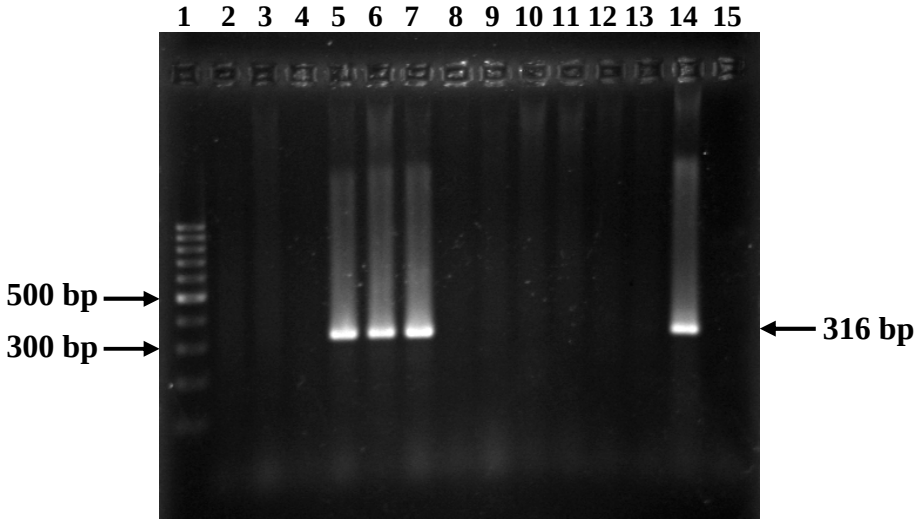
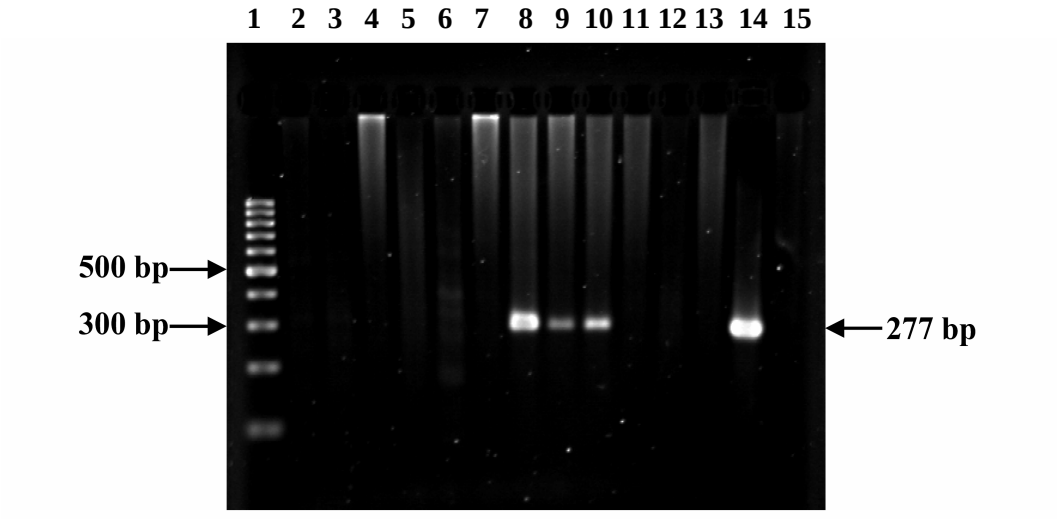


Figure 11. Agarose gel electrophoresis of PCR product amplification based on nested PCR using strain-specific inner primers for Karp (A), Kato (B), Gilliam (C) and TA763 (D). Lane 1, 100 bp DNA marker; lane 2 – 13, DNA samples; lane 14, positive DNA for each specific primers; lane 15, reagent as negative control. The numbers on the right of each panel indicated product size of each pair of primer pair (page 1 of 2).

C.



D.

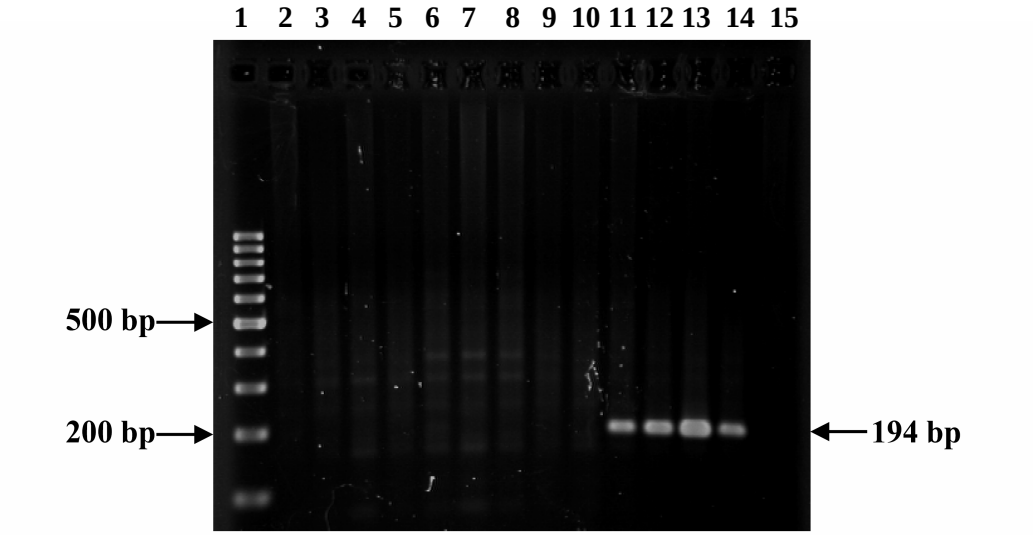


Figure 11. Agarose gel electrophoresis of PCR product amplification based on nested PCR using strain-specific inner primers for Karp (A), Kato (B), Gilliam (C) and TA763 (D). Lane 1, 100 bp DNA marker; lane 2 – 13, DNA samples; lane 14, positive DNA for each specific primers; lane 15, reagent as negative control. The numbers on the right of each panel indicated product size of each pair of primer pair (cont.) (page 2 of 2).

All 34 sequences in Karp cluster were amplified by Karp-specific primers. The primers did not amplify 40 of 42 sequences in non-Karp cluster. The sensitivity and specificity of Karp-specific primers were evaluated as shown in table 10 (A). The sensitivity and specificity was 100 and 95.24%, respectively. These results could be implied that 100% of DNA samples in Karp cluster were amplified by Karp-specific primers whereas approximately 95% of DNA samples in non-Karp cluster were not amplified by these specific primers. The positive predictive (PPV) and negative predictive (NPV) values as shown in table 10 (B) were approximately 94% of probability that the positive result of nested PCR came from DNA samples in Karp cluster whereas 100% of the negative result of nested PCR derived from DNA samples in non-Karp cluster.

The efficiency of Kato-specific primers was also evaluated. The high percentage of sensitivity, specificity and NPV was demonstrated as 100, 95.89 and 100%, respectively (table 11A and B). The low percentage was presented on the PPV as 50% which means there was only 50% of positive result of nested PCR which came from the real Karp cluster.

The high percentage of evaluation was shown in Gilliam-specific primers (table 12A and B). Among 31 DNA samples in Gilliam cluster, 90% of them could be amplified by the primers when 96% of non-Gilliam cluster DNA samples could not be amplified. Moreover, the probability that the positive result of nested PCR derived from Gilliam cluster was the same ratio with the probability that the negative result of nested PCR derived from non-Gilliam cluster at 93%.

The evaluation values of TA763 specific primers were similar to the values of Kato-specific primers. The high percentage of sensitivity, specificity and NPV values could be found at 100, 91 and 100%, respectively (table 13A and B). However, the low percentage was presented in the PPV at only 57%.

Table 10. The evaluation of Karp-specific primers in comparison to strain identification using dendrogram.

A.

	Strain identification by dendrogram		Total
	Karp	Non-Karp	
PCR positive	34	2	36
PCR negative	0	40	40
Total	34	42	76

B.

Evaluation	Calculation	Values (%)
Sensitivity	34/34	100
Specificity	40/42	95.24
Positive predictive value (PPV)	34/36	94.44
Negative predictive value (NPV)	40/40	100

Table 11. The evaluation of Kato-specific primers in comparison to strain identification using dendrogram.

A.

	Strain identification by dendrogram		Total
	Kato	Non-Kato	
PCR positive	3	3	6
PCR negative	0	70	70
Total	3	73	76

B.

Evaluation	Calculation	Values (%)
Sensitivity	3/3	100
Specificity	70/73	95.89
Positive predictive value (PPV)	3/6	50
Negative predictive value (NPV)	70/70	100

Table 12. The evaluation of Gilliam-specific primers in comparison to strain identification using dendrogram.

A.

	Strain identification by dendrogram		Total
	Gilliam	Non-Gilliam	
PCR positive	28	2	30
PCR negative	3	43	46
Total	31	45	76

B.

Evaluation	Calculation	Values (%)
Sensitivity	28/31	90.32
Specificity	43/45	95.56
Positive predictive value (PPV)	28/30	93.33
Negative predictive value (NPV)	43/46	93.48

Table 13. The evaluation of TA763-specific primers in comparison to strain identification using dendrogram.

A.

	Strain identification by dendrogram		Total
	TA763	Non-TA763	
PCR positive	8	6	14
PCR negative	0	62	62
Total	8	68	76

B.

Evaluation	Calculation	Values (%)
Sensitivity	8/8	100
Specificity	62/68	91.18
Positive predictive value (PPV)	8/14	57.14
Negative predictive value (NPV)	62/62	100

CHAPTER V
DISCUSSION

O. tsutsugamushi is the causative agent of scrub typhus. There was a report in Thailand that 16% of 845 patients with acute undifferentiated fever were diagnosed as scrub typhus (73). Although this disease is curable with antibiotics, late- or mis-diagnosis can cause the severity of disease from non- or delayed-treatment.

Protective vaccines for scrub typhus had been developed using several technologies such as recombinant protein vaccine, DNA vaccine and combined technology (74). However, the protective antigens of *O. tsutsugamushi* are different among strain-specific epitopes. The protective immunity was high and approximately 1-3 years against *O. tsutsugamushi* homologous strain infections while the protective immunity was low and short-lived only about 1-3 months against heterologous strains (6, 9).

The indirect immunofluorescense assay (IFA) was considered as the gold standard assay for scrub typhus diagnostic using antibody detection (18). However, the sensitivity of this technique was low because of the high antigenic diversity of 56-kDa protein of the bacteria. At least 8 strains were identified as common strain in Thailand, Karp, Kato, Gilliam, TA678, TA686, TA716 (Chon), TA763 (Fan) and TH1817 (13). The information of strain prevalence in local area is the important part in orders to select the country specific or local antigenic strains for the diagnostic test and vaccine production.

Among several antigenic genes of *O. tsutsugamushi*, the 56-kDa protein encoding gene was used to identify strains or study in phylogeny because of the diversities of this gene. The analysis of this gene using 31 strains of the bacteria indicated that all isolates originated in Southeast Asia (61). Moreover, 14 samples of *O. tsutsugamushi* isolated from Taiwan were classified into 10 strains based on this gene and indicated that these strains were different from the distribution in other

countries (62). In addition, the 56-kDa protein encoding gene was targeted to study the distribution of antibiotic resistant strain of *O. tsutsugamushi* in Thailand (64).

In this study, the 56-kDa protein encoding gene was used for *O. tsutsugamushi* identification. The sensitivity of PCR amplification of this gene was lower than IFA method with 24.7% and 70.7%, respectively. The results of IFA assay was indicated from serum titer of primary blood collection with high titer and secondary blood collection with 4-fold rising titer. These causes IFA has more chance to show the positive result. However, the serologic identification requires a large panel of specific antisera therefore this approach can be performed only in reference laboratories (18). The sensitivity of PCR amplification of 16S rRNA gene was demonstrated more sensitivity than 56-kDa protein encoding gene (3). The low variation in nucleotide sequence of 16S rRNA limited the bacterial genus identification (53). The 56-kDa protein encoding gene was expressed only in *O. tsutsugamushi* which contained 4 hypervariable regions (15, 19). The PCR amplifying and direct sequencing were the methods of choice to gain the nucleotide sequence information of the gene. The genus specific primers were selected from the site in the conserved regions of the 56-kDa protein encoding gene from reference nucleotide sequences (see appendix II) by encompassing all variable regions. After that, the PCR product was analyzed by nucleotide sequencing and each nucleotide sample was compared using the similarity and the distance values. The dendrogram was generated and demonstrated that four clusters of *O. tsutsugamushi* strains were classified. Karp was the predominant strain followed by Gilliam, TA763 and Kato, respectively. Karp and Gilliam were the predominant strains in many countries of Asia. In Japan, several methods such as monoclonal antibody typing, RFLP, indirect fluorescent antibody assay (IFA), indirect immunoperoxidase assay (IP) and nucleotide sequencing of 56-kDa protein encoding gene were used to identify strains of multi-sourced samples (human, rodent and mite). Karp was the predominant strain in most regions followed by Gilliam and other local strains such as Kawasaki and Kuroki (15, 61, 75). Gilliam showed the high prevalence in Shandong province, China (15). However, TA763 was detected in human samples as the dominant strain in Malaysia but the major strain of mite samples was infected with Karp (15).

Several studies in Thailand reported that Karp was a predominant strain. The nucleotide sequences of prototype strains and patient samples showed very high similarity in range of 93-96%. Moreover, the nucleotide sequences of samples were similar to Karp (M33004) strain with approximately 92-99%, whereas, the similarity among different prototype was 67-75%, 69-86% and 68-80% in comparison to Gilliam, TA763 and Kato, respectively. Another prototype strain, Gilliam, the high similarity to the reference strain (EF213085 and EF213079) was shown in this study in a range of 91-99% in which related to the previous study with 88.9-91.8% of similarity (23). Although there is one sequence sample was separated from the Gilliam cluster (No. CHUP50) but with 91% of similarity value indicated that this sample was closely related to Gilliam than Ikeda group of Japanese isolate. In Kato cluster, three nucleotide sequences showed the similarity of 96-97%. These data were corresponded to the values of previous study at 97% similarity (22).

The previous study, TA763 could be found in 28% of 146 mite samples collected in Thailand which was reported to express the weakly reactive to antibody in the direct fluorescent assay (DFA) (12). However, there were few publications about the identification of TA763 based on molecular techniques. In this study, a dendrogram demonstrated that samples were identified into TA763 with different similarity level both high and low relationships. The sample no. CHUP20 showed the highest similarity to TA763 at 96.7% indicated that this sample closely related to TA763 while 7 samples (CHUP68, RAP117, Si002, BP5, KP20, BP107 and BP143) demonstrated the weak relationship to TA763 (the similarity in a range of 77.6-86.8%). However, the distance values of these 7 samples were smallest in a range of 0.063-0.089 comparing to TA763 whereas higher values were shown when compared to other reference sequences (0.119-0.318). As a result, the taxa of these 7 samples demonstrated the closer positions to TA763 than other reference strains in a dendrogram.

Interestingly, 2 samples, KP133 and CHUP65, did not showed the relation to any reference sequences in a dendrogram with the high distance values 0.207 for KP133 and 0.159 for CHUP65 in comparison to TA763 and TA716, respectively. Even KP133 was the most similar to TA763 with 77.6% whereas CHUP65 showed 85.0% similarity to TA716. These values indicated that the nucleotide samples might

different from the prototype or reference sequences. The results from Blastn program showed that these 2 sequences related to Taiwanese strains. The previous studies also showed that Thai isolates were more related to Taiwanese strain than Japanese strain (23). In addition, 8 of 10 isolates in previous study which were considered as new genotypes showed some association between Thai and Taiwanese strains (14).

The previous research articles demonstrated the geographical *O. tsutsugamushi* strain distribution in Thailand based on the genotyping technique. Although the dominant strain in Thailand of most articles was Karp, the difference could be found among the regions of Thailand. In this study, 76 nucleotide samples were used to analyze the strain distribution except 2 nucleotide samples which could not grouped in any reference clusters. Moreover, the distribution in the northern region could not be analyzed because there was only one sample which was insufficient for statistically analysis. Most samples in this study were from Northeastern region. The distribution of strain in this region was more diverse than previous study (23). In addition to Karp, TA763 and Gilliam, one sample was grouped in Kato cluster. The limitation of the study was the low sample numbers in some clusters and also some regions.

The information of the previous study (22) showed the difference in geographic distribution of *O. tsutsugamushi* strains in Thailand. Only 200-400 bp of 56-kDa protein encoding gene was used for dendrogram construction. Karp was the predominant strain in 4 regions of Thailand (northern, northeastern, central and southern regions), and other strains were rarely found except Kato which was detected in the southern region (22). The length of nucleotide sequences in the study might involve in strain variety because nucleotide sequences including more VDs region can generate higher variation between strains than only 200-400 bp of PCR products which were sequenced in this previous study (22). Even ORFs of 56-kDa protein encoding gene were sequenced, the less amount of samples might generate the low variety of strain distribution (23).

Antigens of only 3 prototype strains, Karp, Kato and Gilliam were commonly used in IFA technique but the antigenic variation of other strains had to be concerned. The local strains of antigens should be added to improve the efficiency of IFA technique (16). Moreover, the molecular techniques were used to identify the

strains of *O. tsutsugamushi*. Several genes were studied to identify *O. tsutsugamushi* strains such as 16S rRNA, *groEL* and 47-kDa protein encoding gene but the nucleotide sequences of these genes were less divergence (less than 1% for 16S rRNA gene and 3.5% for groEL gene) (28, 76). The high divergence was presented in 56-kDa protein encoding gene especially in VDs. Thus, the 56-kDa protein encoding gene was the important gene for strain-specific primer selection. In this study, the specific primers for distributed strains in Thailand were designed. Karp- and Gilliam-specific primers were verified on clustered-clinical samples and provided the high sensitivity and specificity values as well as positive and negative predictive value (PPV and NPV). Kato- and TA763-specific primers also presented the high sensitivity and specificity but the PPV was low in both primers (50% and 57.14%, respectively). It indicated that the high rate of false positive was detected. However, the NPV of these two specific primers was 100%.

There were many studies reported about *Orientia* detection using different genes with several molecular techniques (21). Most of the published papers showed the primers designed from the conserved regions for *Orientia* species detection by PCR amplification. The first report demonstrated the primers derived from the 56-kDa protein encoding gene of 3 prototype strains to amplify mouse blood specimens based on conventional PCR. These primers generated the sensitivity approximately 3-5 MLD_{50}/ml of infected blood and the satisfied specificity value (77). Since the most objective of gene amplification was the detection of *O. tsutsugamushi*, so several genes was selected for many techniques to accomplish more sensitivity and specificity. The real-time quantitative PCR (qPCR) for *O. tsutsugamushi* detection was developed with primers derived from 47-kDa protein encoding gene, generated more sensitivity for *O. tsutsugamushi* detection than mouse inoculation (78). Moreover, duplex PCR was reported the higher sensitivity than conventional PCR using primers designed from *groEL* gene (79). Recently, the loop-mediated isothermal PCR assay (LAMP-PCR) was developed based on *groEL* gene (80). However, these techniques were emphasized only on the genus level but not the strain identification.

The first report of strain identification was demonstrated by Furuya Y *et al* (81). Nested PCR was performed using primers derived from the 56-kDa protein encoding gene of 5 strains (81). These primers were used in several studies (3, 22,

82). Although the specificity of these primers was almost 100%, the sensitivity of these primers was only 29% for blood samples, and was increased as 86% when applied to the sample from eschar (3, 82).

The information presented in this study showed that the 56-kDa protein encoding gene can be used for identification of the *O. tsutsugamushi* strain in blood sample of patient with scrub typhus. Moreover, the 56-kDa protein encoding gene of each strain was able to be the target gene for strain-specific primer development. These strain-specific primers were useful for the further development of strain identification based on PCR and other molecular techniques.

CHAPTER VI
CONCLUSION

In order to identify the *O. tsutsugamushi* genotypes from clinical samples, buffy coat samples were collected from patients who was diagnosed as scrub typhus and extracted for chromosomal DNA. Two pairs of primers were designed to amplify the 56-kDa protein encoding gene. There were 106 of 430 scrub typhus suspected samples showed the positively amplification which were selected for further DNA sequencing. Eventually, the nucleotide sequences of readable samples were assembled. All nucleotide sequence samples including 15 reference nucleotide sequences obtained from GenBank database were aligned using computer software, Clustal W. Then, a dendrogram was generated using MEGA4 software.

There were 4 clusters derived from a dendrogram, Karp, Kato, Gilliam and TA763. Most nucleotide sequences were clustered as Karp (34/76, 44.7%) followed by Gilliam, TA763 and Kato with 31/76 (40.8%), 8/76 (10.5%) and 3/76 (3.9%), respectively. Two nucleotide sequences were undifferentiated in any clusters. However, they showed the highest similarity to *O. tsutsugamushi* Taiwanese strains when Blastn program was corresponded to previous study that reported the relationship between *O. tsutsugamushi* Thai and Taiwanese strains was closely than those with Japanese strains. Geograpical distribution of *O. tsutsugamushi* genotypes was studied in 3 regions of Thailand, Northeastern, Central and Southern. Although mostly sequences were in Karp cluster, it was the predominant strain only in Northeastern. Gilliam and TA763 were commonly found in other 3 regions.

Furthermore, 4 strain-specific primer sets were designed to identify each strain by PCR amplification. The sensitivity and specificity were also evaluated by comparison of the result to the direct nucleotide sequencing and a dendrogram. Karp- and Gilliam-specific primers generated the satisfactory results whereas Kato- and TA763-specific primers should be improved.

The results in this study were not only presented the strain distribution but also advantaged for *O. tsutsugamushi* strain detection and identification based on serological and molecular techniques. In serological diagnosis, the antigen of *O. tsutsugamushi* TA763 strain should be included in pooled antigen for antibody detection. Moreover, the Taiwanese strain should be concerned. The information from this study might be used for developing of *O. tsutsugamushi* strain identification in routine diagnosis. The strain prevalence is also useful for epidemiological study and prevention infection by the bacteria with vaccine development.

REFERENCES

1. Tamura A, Ohashi N, Urakami H, Miyamura S. Classification of *Rickettsia tsutsugamushi* in a new genus, *Orientia* gen. nov., as *Orientia tsutsugamushi* comb. nov. Int J Syst Bacteriol 1995; 45: 589-91.
2. Peters W, Pasvol G. Tropical Medicine and Parasitology. 5th edition. London, Mosby. 2002.
3. Sonthayanon P, Chierakul W, Wuthiekanun V, Blacksell SD, Pimda K, Suputtamongkol Y *et al*. Rapid diagnosis of scrub typhus in rural Thailand using polymerase chain reaction. Am J Trop Med Hyg 2006; 75: 1099-102.
4. Seong SY, Choi MS, Kim IS. *Orientia tsutsugamushi* infection: overview and immune responses. Microbes Infect 2001; 3: 11-21.
5. Hanson BA. Effect of immune serum on infectivity of *Rickettsia tsutsugamushi*. Infect Immun 1983; 42: 341-9.
6. Hanson B. Comparative susceptibility to mouse interferons of *Rickettsia tsutsugamushi* strains with different virulence in mice and of *Rickettsia rickettsii*. Infect Immun 1991; 59: 4134-41.
7. Guerrant RL, Walker DH, Weller PF. Tropical Infectious Disease: Principles, pathogens and practice. vol 2, Philadelphia, Churchill Livingstone. 1999.
8. Rights FL, Smadel JE, Jackson EB. Studies on scrub typhus (tsutsugamushi disease). III. Heterogeneity of strains of *Rickettsia tsutsugamushi* as demonstrated by cross vaccination studies. J Exp Med 1948; 87: 339-51.
9. Smadel JE, Ley HL, Diercks FH, Paterson PY, Wisseman CL, Traub R. Immunization against scrub typhus: duration of immunity in volunteers following combined living vaccine and chemoprophylaxis. Am J Trop Med Hyg 1952; 1: 87-99.
10. Watt G, Chouriyagune C, Ruangweerayud R, Watcharapichat P, Phulsuksombati D, Jongsakul K *et al*. Scrub typhus infections poorly responsive to antibiotics in northern Thailand. Lancet 1996; 348: 86-9.

11. Traub R, Johnson PT, Miesse ML, Elbel RE. Isolation of *Rickettsia tsutsugamushi* from rodents from Thailand. Am J Trop Med Hyg 1954; 3: 356-9.
12. Shirai A, Tanskul PL, Andre RG, Dohany AL, Huxsoll DL. *Rickettsia tsutsugamushi* strains found in chiggers collected in Thailand. Southeast Asian J Trop Med Public Health 1981; 12: 1-6.
13. Elisberg BL, Sangkasuvana V, Campbell JM, Bozeman FM, Bodhidatta P. Physiogeographic distribution of scrub typhus in Thailand. Acta Med Biol (Niigata) 1967; 15: 61-7.
14. Fournier PE, Siritantikorn S, Rolain JM, Suputtamongkol Y, Hoontrakul S, Charoenwat S *et al*. Detection of new genotypes of *Orientia tsutsugamushi* infecting humans in Thailand. Clin Microbiol Infect 2008; 14: 168-73.
15. Kelly DJ, Fuerst PA, Ching WM, Richards AL. Scrub typhus: the geographic distribution of phenotypic and genotypic variants of *Orientia tsutsugamushi*. Clin Infect Dis 2009; 48: S203-30.
16. Koh GC, Maude RJ, Paris DH, Newton PN, Blacksell SD. Diagnosis of scrub typhus. Am J Trop Med Hyg 2010; 82: 368-70.
17. Parola P, Blacksell SD, Phetsouvanh R, Phongmany S, Rolain JM, Day NP *et al*. Genotyping of Orientia tsutsugamushi from humans with scrub typhus, Laos. Emerg Infect Dis 2008; 14: 1483-5.
18. La Scola B, Raoult D. Laboratory diagnosis of rickettsioses: current approaches to diagnosis of old and new rickettsial diseases. J Clin Microbiol 1997; 35: 2715-27.
19. Ohashi N, Nashimoto H, Ikeda H, Tamura A. Diversity of immunodominant 56-kDa type-specific antigen (TSA) of *Rickettsia tsutsugamushi*. Sequence and comparative analyses of the genes encoding TSA homologues from four antigenic variants. J Biol Chem 1992; 267: 12728-35.
20. Seong SY, Huh MS, Jang WJ, Park SG, Kim JG, Woo SG *et al*. 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 1997; 65: 1541-5.
21. Stover CK, Marana DP, Carter JM, Roe BA, Mardis E, Oaks EV. The 56-kilodalton major protein of *Rickettsia tsutsugamushi*: Molecular cloning

and sequence analysis of the *sta*56 gene and precise identification of a strain-specific epitope. Infect Immun 1990; 58: 2076-84.

22. Manosroi J, Chutipongvivate S, Auwanit W, Manosroi A. Determination and geographic distribution of *Orientia tsutsugamushi* serotypes in Thailand by nested polymerase chain reaction. Diagn Microbiol Infect Dis 2006; 55: 185-90.
23. Blacksell SD, Luksameetanasan R, Kalambaheti T, Aukkanit N, Paris DH, McGready R *et al*. Genetic typing of the 56-kDa type-specific antigen gene of contemporary *Orientia tsutsugamushi* isolates causing human scrub typhus at two sites in north-eastern and western Thailand. FEMS Immunol Med Microbiol 2008; 52: 335-42.
24. Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, Fukuhara M *et al*. The Whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. DNA Res 2008; 15: 185-99.
25. Oaks EV, Rice RM, Kelly DJ, Stover CK. Antigenic and genetic relatedness of eight *Rickettsia tsutsugamushi* antigens. Infect Immun 1989; 57: 3116-22.
26. Lee JH, Cho NH, Kim SY, Bang SY, Chu H, Choi MS *et al*. Fibronectin facilitates the invasion of *Orientia tsutsugamushi* into host cells through interaction with a 56-kDa type-specific antigen. J Infect Dis 2008; 198: 250-7.
27. Bureau of Epidemiology, Department of Disease control, Ministry of Public Health. Annual epidemiological surveillance report 2009 available on http://203.157.15.4/Annual/Total_Annual.html
28. Eremeeva ME, Madan A, Shaw CD, Tang K, Dasch GA. New perspectives on rickettsial evolution from new genome sequences of rickettsia, particularly *R. canadensis*, and *Orientia tsutsugamushi*. Ann N Y Acad Sci 2005; 1063: 47-63.
29. Cho NH, Kim HR, Lee JH, Kim SY, Kim J, Cha S *et al*. 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 2007; 104: 7981-6.

30. Bai X, Zhang J, Ewing A, Miller SA, Jancso Radek A, Shevchenko DV *et al*. Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. J Bacteriol 2006; 188: 3682-96.
31. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC *et al*. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2004; 2: E69.
32. Tamura A, Ohashi N, Urakami H, Takahashi K, Oyanagi M. Analysis of polypeptide composition and antigenic components of *Rickettsia tsutsugamushi* by polyacrylamide gel electrophoresis and immunoblotting. Infect Immun 1985; 48: 671-5.
33. Oaks EV, Stover CK, Rice RM. Molecular cloning and expression of *Rickettsia tsutsugamushi* genes for two major protein antigens in *Escherichia coli*. Infect Immun 1987; 55: 1156-62.
34. Ohashi N, Tamura A, Suto T. Immunoblotting analysis of anti-rickettsial antibodies produced in patients of Tsutsugamushi disease. Microbiol Immunol 1988; 32: 1085-92.
35. Karlin S, Brocchieri L. Heat shock protein 60 sequence comparisons: duplications, lateral transfer, and mitochondrial evolution. Proc Natl Acad Sci U S A 2000; 97: 11348–53.
36. Paris DH, Aukkanit N, Jenjaroen K, Blacksell SD, Day NP. A highly sensitive quantitative real-time PCR assay based on the *groEL* gene of contemporary Thai strains of *Orientia tsutsugamushi*. Clin Microbiol Infect 2009; 15: 488-95.
37. Hickman CJ, Stover CK, Joseph SW, Oaks EV. Murine T-cell response to native and recombinant protein antigens of *Rickettsia tsutsugamushi*. Infect Immun 1993; 61: 1674-81.
38. Chen HW, Zhang Z, Huber E, Chao CC, Wang H, Dasch GA *et al*. Identification of cross-reactive epitopes on the conserved 47-kilodalton antigen of *Orientia tsutsugamushi* and human serine protease. Infect Immun 2009; 77: 2311-9.

39. Jiang J, Chan TC, Temenak JJ, Dasch GA, Ching WM, Richards AL. Development of a quantitative real-time polymerase chain reaction assay specific for *Orientia tsutsugamushi*. Am J Trop Med Hyg 2004; 70: 351-6.
40. Hanson B. Identification and partial characterization of *Rickettsia tsutsugamushi* major protein immunogens. Infect Immun 1985; 50: 603-9.
41. Rydkina E, Sahni A, Silverman DJ, Sahni SK. *Rickettsia rickettsii* infection of cultured human endothelial cells induces heme oxygenase 1 expression. Infect Immun 2002; 70: 4045-52.
42. Ohashi N, Tamura A, Ohta M, Hayashi K. Purification and partial characterization of a type-specific antigen of *Rickettsia tsutsugamushi*. Infect Immun 1989; 57: 1427-31.
43. Min CK, Yang JS, Kim S, Choi MS, Kim IS, Cho NH. Genome-based construction of the metabolic pathways of *Orientia tsutsugamushi* and comparative analysis within the Rickettsiales order. Comp Funct Genomics 2008: 623145.
44. Philip RN, Casper EA, Burgdorfer W, Gerloff RK, Hughes LE, Bell EJ. Serologic typing of rickettsiae of the spotted fever group by microimmunofluorescence. J Immunol 1978; 121: 1961-8.
45. Fournier PE, Dumler JS, Greub G, Zhang J, Wu Y, Raoult D. Gene sequence-based criteria for identification of new rickettsia isolates and description of *Rickettsia heilongjiangensis* sp. nov. J Clin Microbiol 2003; 41: 5456-65.
46. Anacker RL, Mann RE, Gonzales C. Reactivity of monoclonal antibodies to *Rickettsia rickettsii* with spotted fever and typhus group rickettsiae. J Clin Microbiol 1987; 25: 167-71.
47. Xu W, Raoult D. Distribution of immunogenic epitopes on the two major immunodominant proteins (rOmpA and rOmpB) of *Rickettsia conorii* among the other rickettsiae of the spotted fever group. Clin Diagn Lab Immunol 1997; 4: 753-63.
48. Regnery RL, Spruill CL, Plikaytis BD. Genotypic identification of rickettsiae and estimation of intraspecies sequence divergence for portions of two rickettsial genes. J Bacteriol 1991; 173: 1576-89.

49. Yu X, Jin Y, Fan M, Xu G, Liu Q, Raoult D. Genotypic and antigenic identification of two new strains of spotted fever group rickettsiae isolated from China. J Clin Microbiol 1993; 31: 83-8.

50. Eremeeva M, Yu X, Raoult D. Differentiation among spotted fever group rickettsiae species by analysis of restriction fragment length polymorphism of PCR-amplified DNA. J Clin Microbiol 1994; 32: 803-10.

51. Roux V, Fournier PE, Raoult D. Differentiation of spotted fever group rickettsiae by sequencing and analysis of restriction fragment length polymorphism of PCR-amplified DNA of the gene encoding the protein rOmpA. J Clin Microbiol 1996; 34: 2058-65.

52. Roux V, Raoult D. Genotypic identification and phylogenetic analysis of the spotted fever group rickettsiae by pulsed-field gel electrophoresis. J Bacteriol 1993; 175: 4895-904.

53. Clarridge JE 3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev 2004; 17: 840-62.

54. Roux V, Rydkina E, Eremeeva M, Raoult D. Citrate synthase gene comparison, a new tool for phylogenetic analysis, and its application for the rickettsiae. Int J Syst Bacteriol 1997; 47: 252-61.

55. Sutherland KJ, Henneke CM, Towner P, Hough DW, Danson MJ. Citrate synthase from the thermophilic archaebacterium *Thermoplasma acidophilium*. Cloning and sequencing of the gene. Eur J Biochem 1990; 194: 839-44.

56. Uchiyama T, Kawano H, Kusuhara Y. The major outer membrane protein rOmpB of spotted fever group rickettsiae functions in the rickettsial adherence to and invasion of Vero cells. Microbes Infect 2006; 8: 801-9.

57. Ching WM, Dasch GA, Carl M, Dobson ME. Structural analyses of the 120-kDa serotype protein antigens of typhus group rickettsiae. Comparison with other S-layer proteins. Ann N Y Acad Sci 1990; 590: 334-51.

58. Fournier PE, Roux V, Raoult D. Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA. Int J Syst Bacteriol 1998; 48: 839-49.

59. Roux V, Raoult D. Phylogenetic analysis of members of the genus *Rickettsia* using the gene encoding the outer-membrane protein rOmpB (*ompB*). Int J Syst Evol Microbiol 2000; 50: 1449-55.

60. Carl M, Dobson ME, Ching WM, Dasch GA. Characterization of the gene encoding the protective paracrystalline-surface-layer protein of *Rickettsia prowazekii*: presence of a truncated identical homolog in *Rickettsia typhi*. Proc Natl Acad Sci U S A 1990; 87: 8237-41.

61. Enatsu T, Urakami H, Tamura A. Phylogenetic analysis of *Orientia tsutsugamushi* strains based on the sequence homologies of 56-kDa type-specific antigen genes. FEMS Microbiol Lett 1999; 180: 163-9.

62. Qiang Y, Tamura A, Urakami H, Makisaka Y, Koyama S, Fukuhara M *et al*. Phylogenetic characterization of *Orientia tsutsugamushi* isolated in Taiwan according to the sequence homologies of 56-kDa type-specific antigen genes. Microbiol Immunol 2003; 47: 577-83.

63. Tay ST, Rohani YM, Ho TM, Shamala D. Sequence analysis of the hypervariable regions of the 56 kDa immunodominant protein genes of *Orientia tsutsugamushi* strains in Malaysia. Microbiol Immunol 2005; 49: 67-71.

64. Kollars TM Jr, Bodhidatta D, Phulsuksombati D, Tippayachai B, Coleman RE. Short report: variation in the 56-kD type-specific antigen gene of *Orientia tsutsugamushi* isolated from patients in Thailand. Am J Trop Med Hyg 2003; 68: 299-300.

65. Suputtamongkol Y, Niwattayakul K, Suttinont C, Losuwanaluk K, Limpaiboon R, Chierakul W *et al*. An open, randomized, controlled trial of penicillin, doxycycline, and cefotaxime for patients with severe leptospirosis. Clin Infect Dis 2004; 39: 1417-24.

66. Sermswan RW, Wongratanacheewin W, Anuntagool N, Sirisinha S. Comparison of the polymerase chain reaction and serologic tests for diagnosis of septicemic melioidosis. Am J Trop Med Hyg 2000; 63: 146-9.

67. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA *et al*. Short protocols in molecular biology. Vol 1, fifth ed., John Wiley & Sons, US. 2002.

68. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Window 95/98/NT. Nucleic Acids Symp Ser 1999; 41: 95-8.

69. Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 2007; 35: 43-6.

70. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 11: 4673-80.

71. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 2007; 24, 1596-9.

72. Kim IS, Seong SY, Woo SG, Choi MS, Chang WH. High-level expression of a 56-kilodalton protein gene (*bor56*) of *Rickettsia tsutsugamushi* Boryong and its application to enzyme-linked immunosorbent assays. J Clin Microbiol 1993; 31: 598-605.

73. Suttinont C, Losuwanaluk K, Niwatayakul K, Hoontrakul S, Intaranongpai W, Silpasakorn S *et al*. Causes of acute, undifferentiated, febrile illness in rural Thailand: results of a prospective observational study. Ann Trop Med Parasitol 2006; 100: 363-70.

74. Chattopadhyay S, Richards AL. Scrub typhus vaccines: past history and recent developments. Hum Vaccin 2007; 3: 73-80.

75. Kawamori F, Akiyama M, Sugieda M, Kanda T, Akahane S, Uchikawa K *et al*. Epidemiology of Tsutsugamushi disease in relation to the serotypes of *Rickettsia tsutsugamushi* isolated from patients, field mice, and unfed chiggers on the eastern slope of Mount Fuji, Shizuoka Prefecture, Japan. J Clin Microbiol 1992; 30: 2842-6.

76. Lee JH, Park HS, Jang WJ, Koh SE, Kim JM, Shim SK *et al*. Differentiation of rickettsiae by *groEL* gene analysis. J Clin Microbiol 2003; 41: 2952-60.

77. Kelly DJ, Marana DP, Stover CK, Oaks EV, Carl M. Detection of *Rickettsia tsutsugamushi* by gene amplification using polymerase chain reaction techniques. Ann N Y Acad Sci 1990; 590: 564-71.

78. Singhsilarak T, Leowattana W, Looareesuwan S, Wongchotigul V, Jiang J, Richards AL *et al*. Short report: detection of *Orientia tsutsugamushi* in clinical samples by quantitative real-time polymerase chain reaction. Am J Trop Med Hyg 2005; 72: 640-1.
79. Park HS, Lee JH, Jeong EJ, Kim JE, Hong SJ, Park TK *et al*. Rapid and simple identification of *Orientia tsutsugamushi* from other group rickettsiae by duplex PCR assay using groEL gene. Microbiol Immunol 2005; 49: 545-9.
80. Paris DH, Blacksell SD, Newton PN, Day NP. Simple, rapid and sensitive detection of *Orientia tsutsugamushi* by loop-isothermal DNA amplification. Trans R Soc Trop Med Hyg 2008; 102: 1239-46.
81. Furuya Y, Yoshida Y, Katayama T, Yamamoto S, Kawamura A Jr. Serotype-specific amplification of *Rickettsia tsutsugamushi* DNA by nested polymerase chain reaction. J Clin Microbiol 1993; 31: 1637-40.
82. Kim DM, Kim HL, Park CY, Yang TY, Lee JH, Yang JT et al. Clinical usefulness of eschar polymerase chain reaction for the diagnosis of scrub typhus: a prospective study. Clin Infect Dis 2006; 43: 1296-300.

APPENDICES

APPENDIX A

GENERAL REAGENTS

1.0 M TrisCl pH 8.0

Tris base (Sigma)	12.11	gm

Dissolve in distilled water, adjust to pH 8.0 with concentrated HCl and adjust to 100 ml final volume with distilled water. Sterilized by autoclaving for 15 minutes at 121°C, 15 lb/square inches.

0.5 M EDTA pH 8.0

EDTA ($C_{10}H_{14}N_2O_8Na_2 \cdot 2H_2O$) (Amresco)	18.61	gm

Dissolve in distilled water, adjust to pH 8.0 with NaOH and adjust to 100 ml final volume with distilled water. Sterilized by autoclaving for 15 minutes at 121°C, 15 lb/square inches.

1.0 N NaOH

NaOH (Fluka)	4	gm
Distilled water	100	ml

REAGENTS FOR DNA EXTRACTION

TE pH 8.0 (10 mM TrisCl pH 8.0, 1 mM EDTA pH 8.0)

1.0 M TrisCl pH 8.0	1	ml
0.5 M EDTA pH 8.0	0.20	ml

Dissolve and adjust the volume to 100 ml with distilled water.

Proteinase K (20 mg/ml)

Proteinase K (Promega)	0.20	gm
Distilled water	10	ml

The solution was aliquoted and stored at -20°C.

10% Sodium dodecyl sulfate (SDS) (w/v)

Sodium dodecyl sulfate (Sigma)	1	gm
Distilled water	10	ml

CTAB/NaCl solution (10% CTAB in 0.7 M NaCl)

NaCl (Fluka)	4.10	gm
CTAB (hexadecyltrimethyl ammonium bromide) (Fluka)	10	gm
Distilled water	100	ml

5.0 M NaCl

NaCl (Fluka)	29.22	gm
Distilled water	100	ml

Chloroform:Isoamyl alcohol (24:1)

Chloroform (Merck)	24	parts
Isoamyl alcohol (Merck)	1	part

Mixed them together and stored in a dark bottle at room temperature.

Phenol:Chloroform:Isoamyl alcohol (25:24:1)

Phenol (Merck)	25	parts
Chloroform (Merck)	24	parts
Isoamyl alcohol (Merck)	1	part

Mixed them together and stored in a dark bottle at room temperature.

70% Ethanol

Absolute ethanol (Merck)	70	ml
Distilled water	30	ml

REAGENTS FOR AGAROSE GEL ELECTROPHORESIS

1xTBE buffer

Tris base (Sigma)	10.80	gm
Boric acid (Fluka)	5.50	gm
0.5 M EDTA pH 8.0	4	ml

Dissolve and adjust the volume to 1,000 ml with distilled water.

0.7% Agarose gel (w/v)

Agarose gel (Gibco)	0.7	gm
1X TBE buffer	100	ml

Melt in microwave oven until dissolved completely.

1.5% Agarose gel (w/v)

Agarose gel (Gibco)	1.5	gm
1X TBE buffer	100	ml

Melt in microwave oven until dissolved completely.

6x Gel-loading buffer

Bromophenol blue (Sigma)	25	mg
Xylene cyanol FF (Sigma)	25	mg
Glycerol (Sigma)	3	ml

Dissolve and adjust the volume to 10 ml with distilled water.

Ethidium bromide (10 mg/ml)

Ethidium bromide (Sigma)	1	gm
Distilled water	100	ml

The solution was stored in a dark bottle at 4°C.

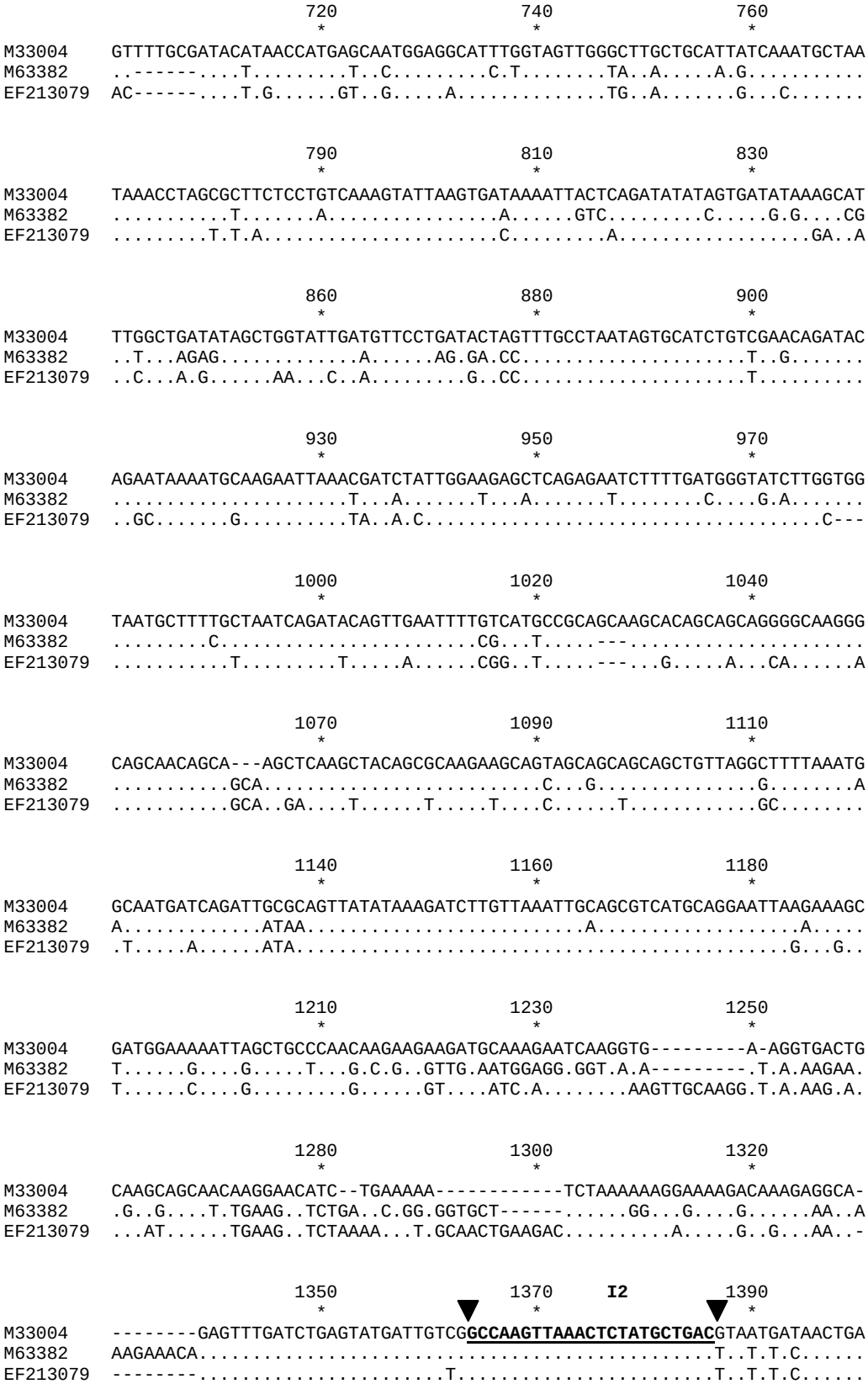
APPENDIX B

MULTIPLE ALIGNMENT OF THE REFERENCE STRAINS OF O. TSUTSUGAMUSHI FOR PRIMER SELECTION

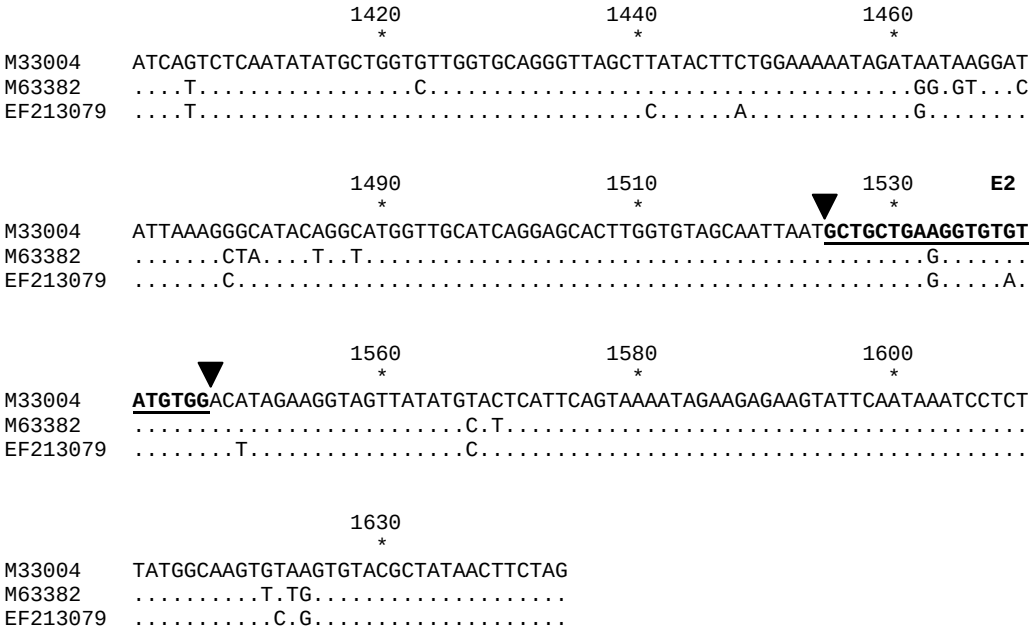
Multiple sequence alignment of 56-kDa protein encoding gene of the prototype strains of *O. tsutsugamushi*, Karp (accession no. M33004), Kato (M63382) and Gilliam (EF213079) which obtained from GenBank database. Two sets of primers were selected from the conserved regions as the priming site to amplify the gene. Dots (.) represented the identical nucleotides to the top sequence whereas the indel situation was replaced with dash (-).

```
                          20                  40                  60
                           *                   *                   *
M33004    ATGAAAAAAATTATGTTAATTGCTAGTGCAATGTCTGCGTTGTCGTTGCCATTTTCAGCTAGTGCAATAG
M63382    .....................................A.....A.....G.............G....
EF213079  ................................................G.............G....

                          90                  110       E1        130
                           *                   *                   *
M33004    AATTGGGGGA---AGAAGGATTAGAGTGTGGTCCTTATGCTAAAGTTGGAGTTGTTGGAGGAATGATTAC
M63382    ..........TGA..G.....................................C................
EF213079  ..........TGA..G.....................................C...........T.....

                          160                 180                 200
                           *                   *                   *
M33004    TGGCGTAGAATCTGCTCGCTTGGATCCAGCTGATGCTGAAGGCAAAAAACACTTGTCATTAACAAATGGG
M63382    ............A.......................GT...........A...C.........CCTC.
EF213079  ............A...................A..A.T.........T.....C.........CCTC.

                    I1    230                 250                 270
                           *                   *                   *
M33004    CTGCCATTTGGTGGAACGTTGGCTGCAGGTATGACAATCGCTCCAGGATTTAGAGCAGAGATAGGTGTTA
M63382    A............T..A..A.....................G.................C....G....
EF213079  A............T..A..A.....G.......T...T..A.................C.........

                          300                 320                 340
                           *                   *                   *
M33004    TGTACCTTACAAATATAACTGCTCAGGTTGAAGAAGGTAAAGTTAAGGCAGATTCTGTAGGTGAGACAAA
M63382    ........G.G...G...AA..AG....G...TC.......AC.-------GC....------------.
EF213079  .........G........GC...G.........T........GC....T.......AG.......GT...

                          370                 390                 410
                           *                   *                   *
M33004    GGCAGATTCTGTAGGTGGGAAAGATGCTCCTATACGTAAGCGGTTTAAACTTACACCTCCTCAGCCTACT
M63382    T..T...AT.AG.TC...TGC....T.......G.C.C......A............A............
EF213079  ...........G.T......T.....................CC............A............

                          440                 460                 480
                           *                   *                   *
M33004    ATAATGCCTATAAGTATAGCTGTACGTGACTTTGGGATTGATATTCCTAACCAGACCTCAGCAG----CA
M63382    ................T..G.AT......C.....G.............GTAC.TCA..G..GAGCT.
EF213079  ......................AT.....TGAG...G......C.GA..TT.TTG.TCAG..T.------

                          510                 530                 550
                           *                   *                   *
M33004    AGCAC--AAGCCGCAGCCTCAGGCTTAATGATGAGCAACGTGCTGCAGCTAGGATCGCTTGGTTAAAGAA
M63382    .T...CTGG.TGAT.A...TG.TGC.......ATT.GG.......AC.A.......A.......G.....
EF213079  -CTG.----TGG...A..A...-....C..T......G..........AG.....T........G.....

                          580                 600                 620
                           *                   *                   *
M33004    TTGTGCTGGTATTGACTATAGGGTAAAAAACCCTAATGATCCTAATGGGCCTATGGTTATAAATCCGATA
M63382    ..A.......G.........T...TCC.G.T......A.....C.G.---...GAA..G.......AG.G
EF213079  ..A.................T...CCC.G.T...C.GA.....C...---...GA...........TG.G

                          650                 670                 690
                           *                   *                   *
M33004    TTGTTAAATATTCCACAGGGTAACCCTAATCCTGTTGGAAATCCACCGCAGCGAGCAAATCCGCCTGCAG
M63382    C.A...........T..A...CCG.....------..C......---TAGA.A...T.TG.AA...TGTA
EF213079  ............A.T..A..GCCA.....------C.T.C.G..---TAGA.---.TCGG.AAGA.CTTA
```

Multiple sequence alignment of 56-kDa protein encoding gene of prototype strains of *O. tsutsugamushi* (page 1 of 3)



Multiple sequence alignment of 56-kDa protein encoding gene of prototype strains of *O. tsutsugamushi* (page 2 of 3)

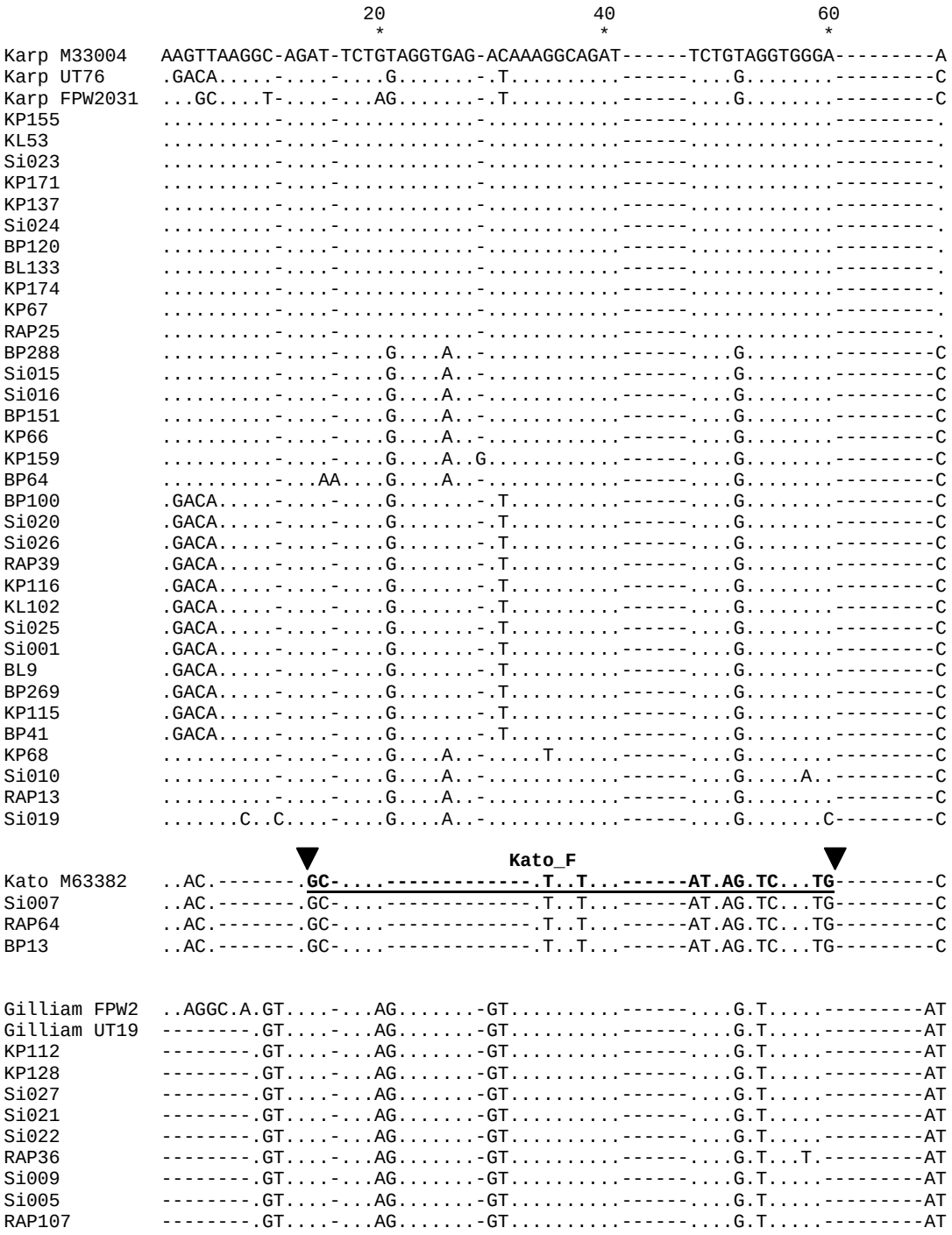


Multiple sequence alignment of 56-kDa protein encoding gene of prototype strains of *O. tsutsugamushi* (page 3 of 3)

APPENDIX C

NUCLEOTIDE SEQUENCE ALIGNMENT TO SELECT THE CONSERVED REGIONS FOR O. TSUTSUGAMUSHI STRAIN-SPECIFIC AMPLIFICATION

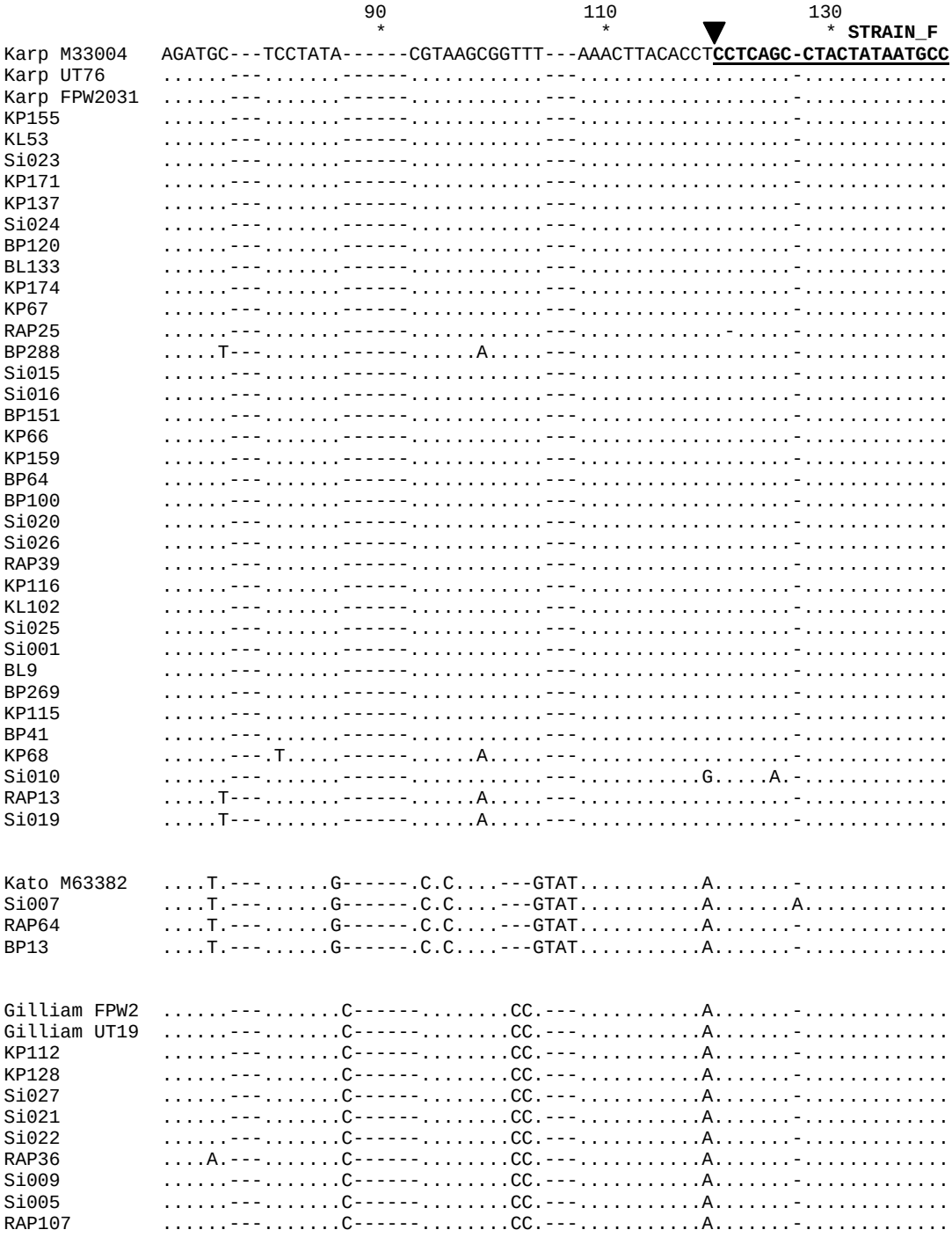
Nucleotide sequence alignment of 56-kDa protein encoding gene of 76 samples (excluded sample no. KP133 and CHUP65 which showed low identity scores) were characterized as strain based on a dendrogram. Four sets of primers were selected from the conserved regions of each strain as the priming site to amplify the gene. Dots (.) represented the identical nucleotides to the top sequence whereas the indel situation was replaced with dash (-).



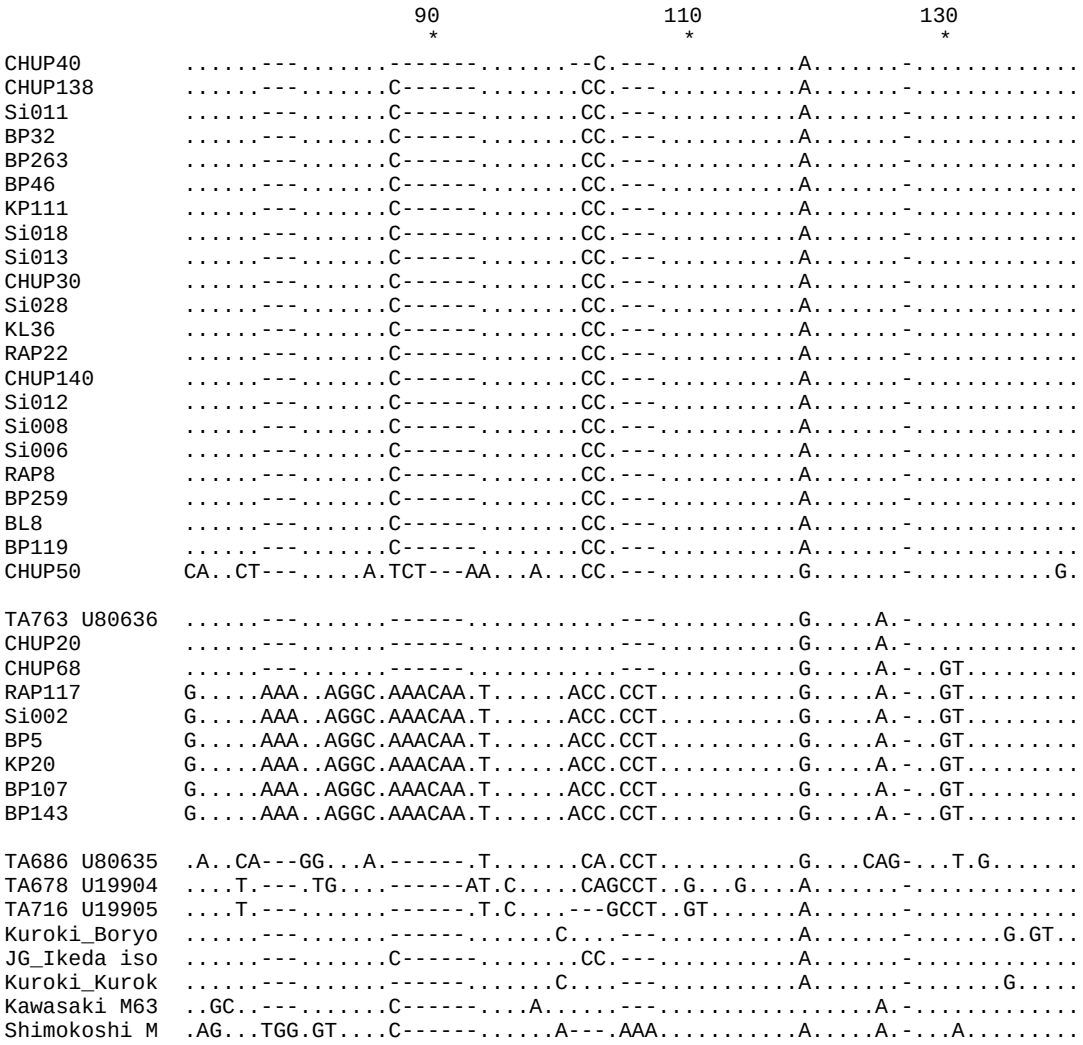
Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 1 of 14)

```
                              20                  40                  60
                               *                   *                   *
CHUP40          --------.GA...C-...AA.AAA..A-.A..G.AA.A.AGACAAA.T..G.TC..A.TATG-----AT
CHUP138         --------.GT...-...AG.......-GT..........------....G.T.....----------AT
Si011           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
BP32            --------.GT...-...AG.......-GT..........------....G.T.....----------AT
BP263           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
BP46            --------.GT...-...AG.......-GT..........------....G.T.....----------AT
KP111           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
Si018           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
Si013           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
CHUP30          --------.GT...-...AG.......-GT..........------....G.T.....----------AT
Si028           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
KL36            --------.GT...-...AG.......-GT..........------....G.T.....----------AT
RAP22           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
CHUP140         --------.GT...-...AG.......-GT..........------....G.T.....----------AT
Si012           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
Si008           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
Si006           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
RAP8            --------.GT...-...AG.......-GT..........------....G.T.....----------AT
BP259           --------.GT...-...AG.......-GT..........------....G.T.....----------AT
BL8             --------.GT...-...AG.......-GT..........------....G.T.....----------AT
BP119           --------.GT.....-.T.AG.......-GT..........------....G.T.....----------AT
CHUP50          --------AAC..GG-C...AG.C...---T.GT.CT.ACG-------T.ACT.A...T---------GC


TA763 U80636    ....C.....-....-....G.A..A..-...........------....G.....A..----------C
CHUP20          ....C.....-....-....G....A..-...........------....G.....A..----------C
CHUP68          ....C.....-....-....G....A..-...........------....G.....A..----------C
RAP117          .G.GC..AAT-..G.-....AT.T..CT-.A.GGTT.T...------G...CTAA..AT.ACACTGGTAC
Si002           .G.GC..AAT-..G.-....AT.T..CT-.A.GGTT.T...------G...CTAA..AT.ACACTGGTAC
BP5             .G.GC..AAT-..G.-....AT.T..CT-.A.GGTT.T...------G...CTAA..AT.ACACTGGTAC
KP20            GG.GC..AAT-..G.-....AT.T..CT-.A.GGTT.T...------G...CTAA..AT.ACACTGGTAC
BP107           .G.GC..AAT-..G.-....AT.T..CT-.A.GGTT.T...------G...CTAA..AT.ACACTGGTAC
BP143           .G.GC..AAT-..G.-....AT.T..CT-.A.GGTT.T...------G...CTAA..AT.ACACTGGTAC


TA686 U80635    .G.GC....T-....-....GCA.--------..T.ATA.C------A..ACT.A.CATG---------C
TA678 U19904    ..AC.--------.GC-....-------------.T..T.G.------A..AG..C..AT.----------C
TA716 U19905    ..AC.-------A.C-....-------------GTA.T...------A..AG..C..AT.----------C
Kuroki_Boryo    ...GCG...T-....-...AA........-.T.........------....G.....----------C
JG_Ikeda iso    --------.GT...-...AG.......-GT..........------....G.T....A---------AT
Kuroki_Kurok    ...GCG...T-....-...AA........-.T.........------....G.....----------C
Kawasaki M63    --------AAC..G.-....GT.C..C---T..T.--------------...CTAT..AT---------AC
Shimokoshi M    -----G.AAATG.G.-....AT.C.--------..TA.T.G.------A..ACT.C..ATG---------C
```

Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 2 of 14)



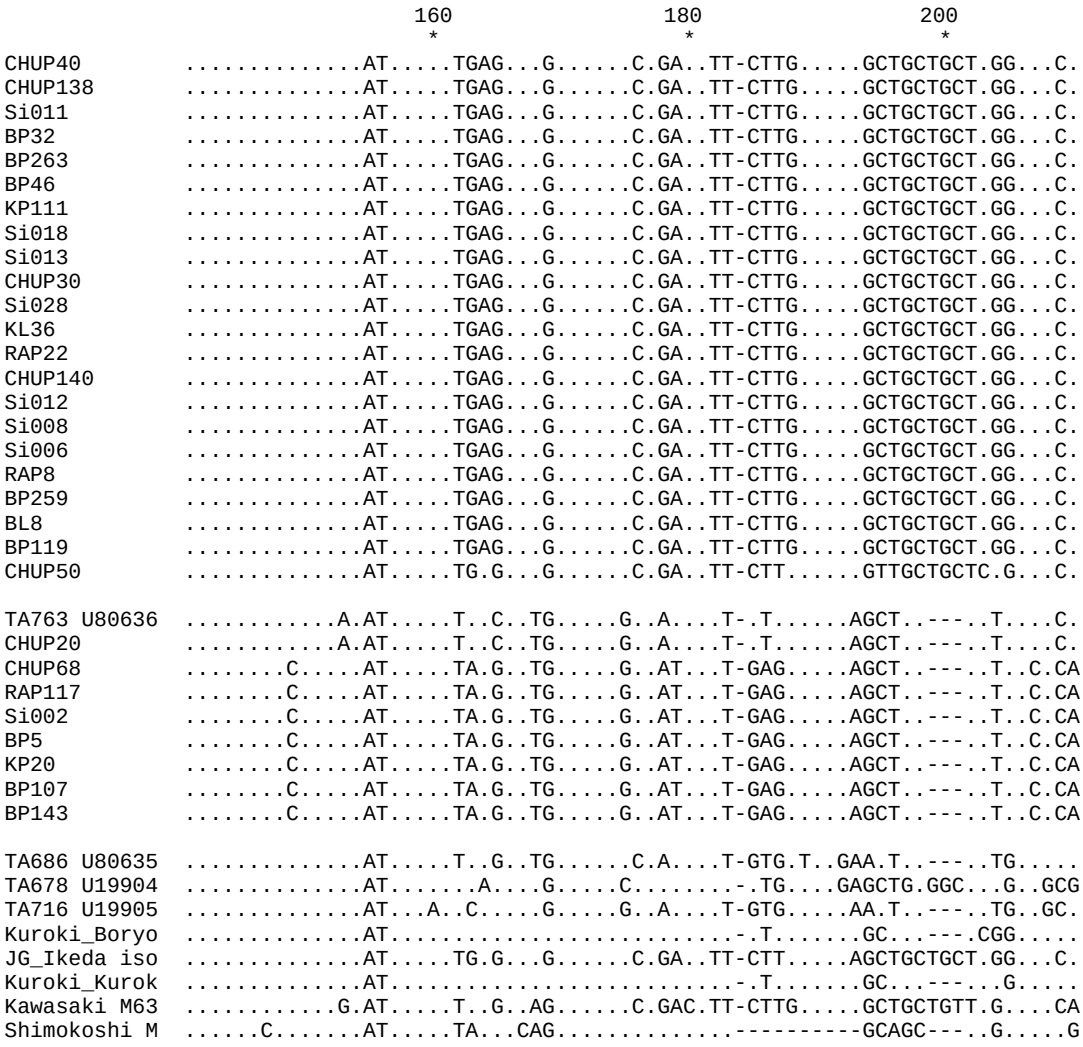
Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 3 of 14)



Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 4 of 14)

```
                                    160                 180                 200
           STRAIN_F▼                  *                   *                   *
Karp M33004   TATAAGTATAGCTGTACGTGACTTTGGGATTGATATTCCTAACCAGACCTCAGCAGCA---AGCACAAGC
Karp UT76     ..............AT......................G....-.T.......GC...---...TGGG--
Karp FPW2031  ..............AT...........................-.T............---.........
KP155         ..............AT...........................-.T............---.........
KL53          ..............AT...........................-.T............---.........
Si023         ..............AT...........................-.T............---.........
KP171         ..............AT...........................-.T............---.........
KP137         ..............AT...........................-.T............---.........
Si024         ..............AT...........................-.T............---.........
BP120         ..............AT...........................-.T............---.........
BL133         ..............AT...........................-.T............---.........
KP174         ..............AT...........................-.T............---.........
KP67          ..............AT...........................-.T............---.........
RAP25         ..............AT...........................-.T............---.........
BP288         ..............AT...........................-.T............---.........
Si015         ..............AT...........................-.T............---.........
Si016         .........T....AT...........................-.T..G...AGC...---...G.G...
BP151         .........T....AT...........................-.T..G...AGC...---...G.G...
KP66          .........T....AT...........................-.T..G...AGC...---...G.G...
KP159         .........T....AT...........................-.T..G...AGC...---...G.G...
BP64          .........T....AT...........................-.T..G...AGC...---...G.G...
BP100         ..............AT......................G....-.T.......GC...---...TGGG--
Si020         ..............AT......................G....-.T.......GC...---...TGGG--
Si026         ..............AT......................G....-.T.......GC...---...TGGG--
RAP39         ..............AT......................G....-.T.......GC...---...TGGG--
KP116         ..............AT......................G....-.T.......GC...---...TGGG--
KL102         ..............AT......................G....-.T.......GC...---...TGGG--
Si025         ..............AT......................G....-.T.......GC...---...TGGG--
Si001         ..............AT......................G....-.T.......GC...---...TGGG--
BL9           ..............AT......................G....-.T.......GC...---...TGGG--
BP269         ..............AT......................G....-.T.......GC...---...TGGG--
KP115         ..............AT......................G....-.T.......GC...---...TGGG--
BP41          ..............AT......................G....-.T.......GC...---...TGGG--
KP68          ..............AT...........................-.T.......GC...---...G.....
Si010         ..............AT.....................TG....-.T.......GC...---...G.....
RAP13         ..............AT.....................TG....-.T.......GC...---...G.....
Si019         ..............AT.....................TG....-.T.......GC...---...G.....


Kato M63382   .........T..G.AT......C.....G..............-GT......AGGAGGAGCTAAT..CCT
Si007         ..............AT......C.....G..............-GT......AGGAGGAGCTAAT..CCT
RAP64         ..............AT......C.....G..............-GT......AGGAGGAGCTAAT..CCT
BP13          ..............AT......C.....G..............-GT......AGGAGGAGCTAAT..CCT


Gilliam FPW2  ..............AT.....TGAG...G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
Gilliam UT19  ..............AT.....TGAG...G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
KP112         ..............AT.....TGAG...G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
KP128         ..............AT.....TGAG...G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
Si027         ..............AT.....TGAG...G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
Si021         ..............AT.....TGAG...G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
Si022         ..............AT.....TGAG...G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
RAP36         ..............AT.....TGAG...G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
Si009         ..............AT.....TGAG...G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
Si005         ..............AT.....TGAG...G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
RAP107        ..............AT.....TGA....G......C.GA..TT-CTTG.....GCTGCTGCT.GG...C.
```

Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection
(page 5 of 14)



Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 6 of 14)

```
                                230               250               270
                                 *                 *                 *
Karp M33004     --C--GCAGCCTCAGGCTTAATGATGAGCAACGTGCTGCAGCTAGGATCGCTTGGTTAAAGAATTGTGCT
Karp UT76       -------------.AT......A.......G.....................................
Karp FPW2031    GG.--..........-.....................................................
KP155           GG.--..........-.....................................................
KL53            GG.--..........-.....................................................
Si023           GG.--..........-.....................................................
KP171           GG.--..........-.....................................................
KP137           GG.--..........-.....................................................
Si024           GG.--..........-.....................................................
BP120           GG.--..........-.....................................................
BL133           GG.--..........-.....................................................
KP174           GG.--..........-.....................................................
KP67            GG.--..........-.....................................................
RAP25           GG.--..........-.....................................................
BP288           GG.--.........C.-.....................................................
Si015           GG.--.........C.-.....................................................
Si016           TG.--A.......C.-.....................................................A....
BP151           TG.--A.......C.-.....................................................A....
KP66            TG.--A.......C.-.....................................................A....
KP159           TG.--A.......C.-.....................................................A....
BP64            TG.--A.......C.-.....................................................A....
BP100           -------------.AT......A.......G.....................................
Si020           -------------.AT......A.......G.....................................
Si026           -------------.AT......A.......G.....................................
RAP39           -------------.AT......A.......G.....................................
KP116           -------------.AT......A.......G.....................................
KL102           -------------.AT......A.......G.....................................
Si025           -------------.AT......A.......G.....................................
Si001           -------------.AT......A.......G.....................................
BL9             -------------.AT......A.......G.....................................
BP269           -------------.AT......A.......G.....................................
KP115           -------------.AT......A.......G.....................................
BP41            -------------.AT......A.......G.....................................
KP68            TG.--.A.T.....-..........A..G.............A.....................
Si010           TG.--.A.T...GC.-..........A..G.....................................
RAP13           TG.--.A.T...GC.-..........A..G.....................................
Si019           TG.--.A.T...GC.-..........A..G................T....................


Kato M63382     GGGTGAT.A...TG.TGC.......ATT.GG.......AC.A.......A.......G.......A....
Si007           GGGTAAT.A...TG.TGC.......ATT.GG........T.A..............G.......A....
RAP64           GGGTAAT.A...TG.TGC.......ATT.GG........T.A..............G.......A....
BP13            GGGTAAT.A...TG.TGC.......ATT.GG........T.A..............G.......A....


Gilliam FPW2    A------------CA.....C..T......G..........AG.....T........G.......A....
Gilliam UT19    A------------CA.....C..T......G..........AG.....T........G.......A....
KP112           A------------CA.....C..T......G..........AG.....T........G.......A....
KP128           A------------CA.....C..T......G..........AG.....T........G.......A....
Si027           A------------CA.....C..T......G..........AG.....T........G.......A....
Si021           A------------CA.....C..T......G..........AG.....T........G.......A....
Si022           A------------CA.....C..T......G..........AG.....T........G.......A....
RAP36           A------------CA.....C..T......G..........A......T........G.......A....
Si009           A------------CA.....C..T......G..........AG.....T........G.......A....
Si005           A------------CA.....C..T......G..........AG.....T........G.......A....
RAP107          A------------CA.....C..T......G..........AG.....T........G.......A....
```

Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 7 of 14)

```
                              230                 250                 270
                               *                   *                   *
CHUP40          A------------CA.....C..T......G..........AG.....T........G.......A....
CHUP138         A------------CA.....C..T......G..........AG.....T........G.......A....
Si011           A------------CA.....C..T......G..........AG.....T........G.......A....
BP32            A------------CA.....C..T......G..........AG.....T........G.......A....
BP263           A------------CA.....C..T......G..........AG.....T........G.......A....
BP46            A------------CA.....C..T......G..........AG.....T........G.......A....
KP111           A------------CA.....C..T......G..........AG.....T........G.......A....
Si018           A------------CA.....C..T......G..........AG.....T........G.......A....
Si013           A------------CA.....C..T......G..........AG.....T........G.......A....
CHUP30          A------------CA.....C..T......G..........AG.....T........G.......A....
Si028           A------------CA.....C..T......G..........AG.....T........G.......A....
KL36            A------------CA.....C..T......G..........AG.....T........G.......A....
RAP22           A------------CA.....C..T......G..........AG.....T........G.......A....
CHUP140         A------------CA.....C..T......G..........AG.....T........G.......A....
Si012           A------------CA.....C..T......G..........AG.....T........G.......A....
Si008           A------------CA.....C..T......G..........AG.....T........G.......A....
Si006           A------------CA.....C..T......G..........AG.....T........G.......A....
RAP8            A------------CA.....C..T......G..........AG.T...T........G.......T....
BP259           A------------CA.....C..T......G..........AG.....T........G.......A....
BL8             A------------CA.....C..T......G..........AG.....T........G.......A....
BP119           A------------CA.....C..T......G..........AG.....T........G.......A....
CHUP50          A------------CA.....C..T......G..........A......T........G.......A....


TA763 U80636    GC.--T....AAGCA-AA.G..CC.CTTGTT....GA.T.CG......T...........AG.G.A....
CHUP20          GC.--T....AAGCA-AA.G..CC.CTTGTT....GA.T.CG......T...........AG.G.A....
CHUP68          AG.--T....AA----AA.G..CC.CTTGTT....GATT.CG......T...........AC.G.A....
RAP117          AG.--T....AA----AA.G..CC.CTTGTT....GATT.CG......T...........AC.G.A....
Si002           AG.--T....AA----AA.G..CC.CTTGTT....GATT.CG......T...........AC.G.A....
BP5             AG.--T....AA----AA.G..CC.CTTGTT....GATT.CG......T...........AC.G.A....
KP20            AG.--T....AA----AA.G..CC.CTTGTT....GATT.CG......T...........AC.G.A....
BP107           AG.--T....AA----AA.G..CC.CTTGTT....GATT.CG......T...........AC.G.A....
BP143           AG.--T....AA----AA.G..CC.CTTGTT....GATT.CG......T...........AC.G.A....


TA686 U80635    AAA---TCAGG.G.ATG---..CCGCTTGTT....GA.T.CG......T...........AG.A.A....
TA678 U19904    AAACATTCAAG.T.ATG---...C..TT..G....GA..GCG......T...........AC.G.A....
TA716 U19905    AAA---TCAGG.G.ATG---..CC.CTTGTT....GA.G.CG...A..T...........AG...A....
Kuroki_Boryo    TG.--A.......C.-.........C..A.G.....................................
JG_Ikeda iso    A------------CA.....C..T......G..G....T.CG...A..T..G.....G...G...A....
Kuroki_Kurok    TG.--A.......C.-.........C..A.G.....................................
Kawasaki M63    A------------CA.....C..T......G..G....A..A......T........G.......A....
Shimokoshi M    A-------------------....T...TGTG........T........T........G.......A....
```

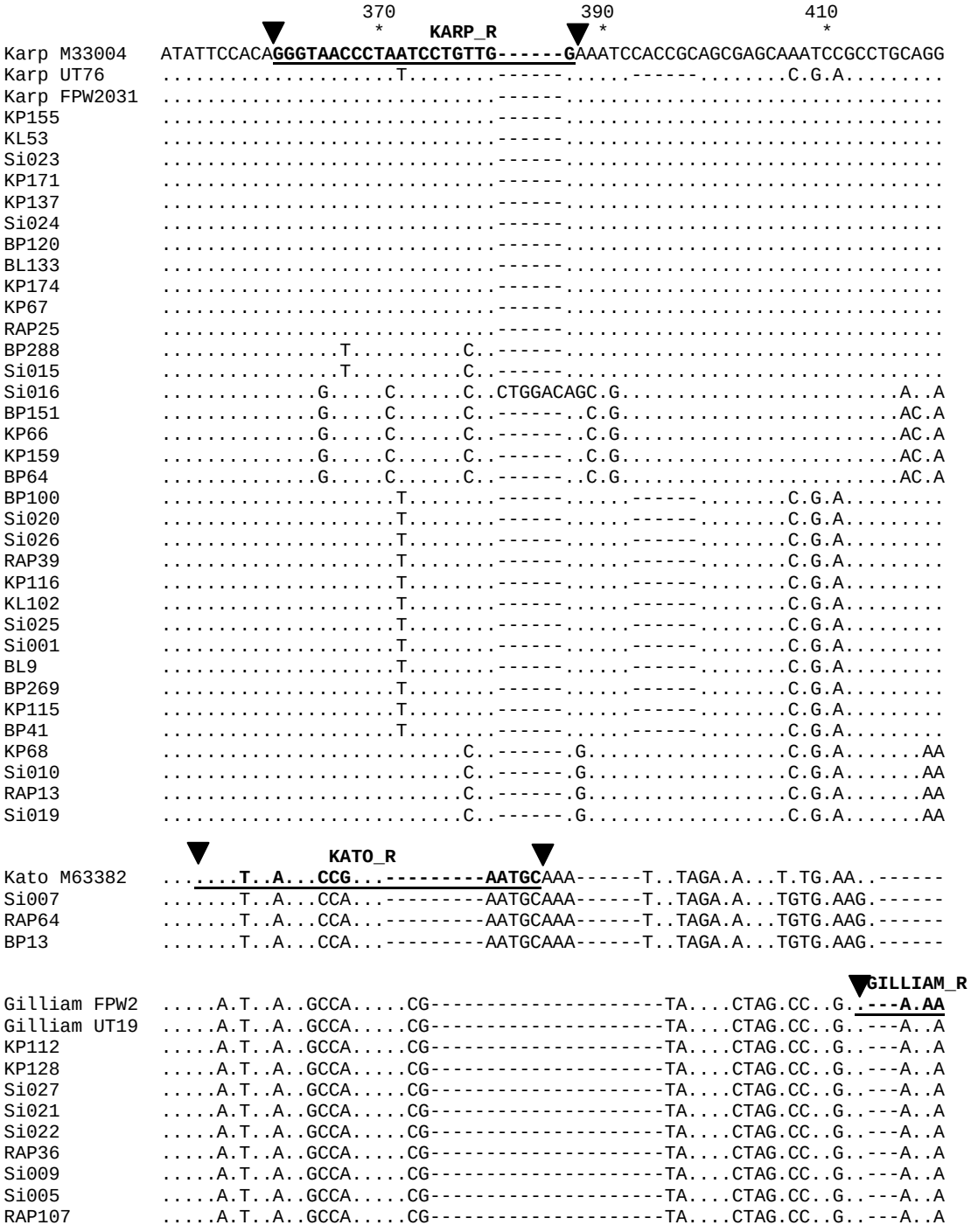
Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 8 of 14)

```
                                300                 320                 340
                                 *                   *                   *
Karp M33004   GGTATTGACTATAGGGTAAAAAACCCTAATGATCCTAATGGGCCTATGGTTATAAATCCGATATTGTTAA
Karp UT76     ....................G.T......A.....................................
Karp FPW2031  ....................................................................
KP155         ....................................................................
KL53          ....................................................................
Si023         ....................................................................
KP171         ....................................................................
KP137         ....................................................................
Si024         ....................................................................
BP120         ....................................................................
BL133         ....................................................................
KP174         ....................................................................
KP67          ....................................................................
RAP25         ....................................................................
BP288         ....................................................................
Si015         ....................................................................
Si016         ....................................................................
BP151         ....................................................................
KP66          ....................................................................
KP159         ........................................G...........................
BP64          ....................................................................
BP100         ....................G.T......A.....................................
Si020         ....................G.T......A.....................................
Si026         ....................G.T......A.....................................
RAP39         ....................G.T......A.....................................
KP116         ....................G.T......A.....................................
KL102         ....................G.T......A.....................................
Si025         ....................G.T......A.....................................
Si001         ....................G.T......A.....................................
BL9           ....................G.T......A.....................................
BP269         ....................G.T......A.....................................
KP115         ....................G.T......A.....................................
BP41          ....................G.T......A.....................................
KP68          ....................G.T......A.....................................
Si010         ....................G.T......A.....................................
RAP13         ....................G.T......A.....................................
Si019         ....................G.T......A.....................................


Kato M63382   ...G.........T...TCC.G.T......A.....---CA.G...GAA..G.......AG.GC.A....
Si007         ............T...TCC.G.T...C.GA.....---CA.G...GA...G.......AG.GC.A...G
RAP64         ............T...TCC.G.T...C.GA.....---CA.G...GA...G.......AG.GC.A...G
BP13          ............T...TCC.G.T...C.GA.....---CA.G...GA...G.......AG.GC.A...G


Gilliam FPW2  ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
Gilliam UT19  ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
KP112         ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
KP128         ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
Si027         ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
Si021         ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
Si022         ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
RAP36         ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
Si009         ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
Si005         ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
RAP107        ............T...CCC.G.T...C.GA.....---CATG...GA...........TG.G.......
```

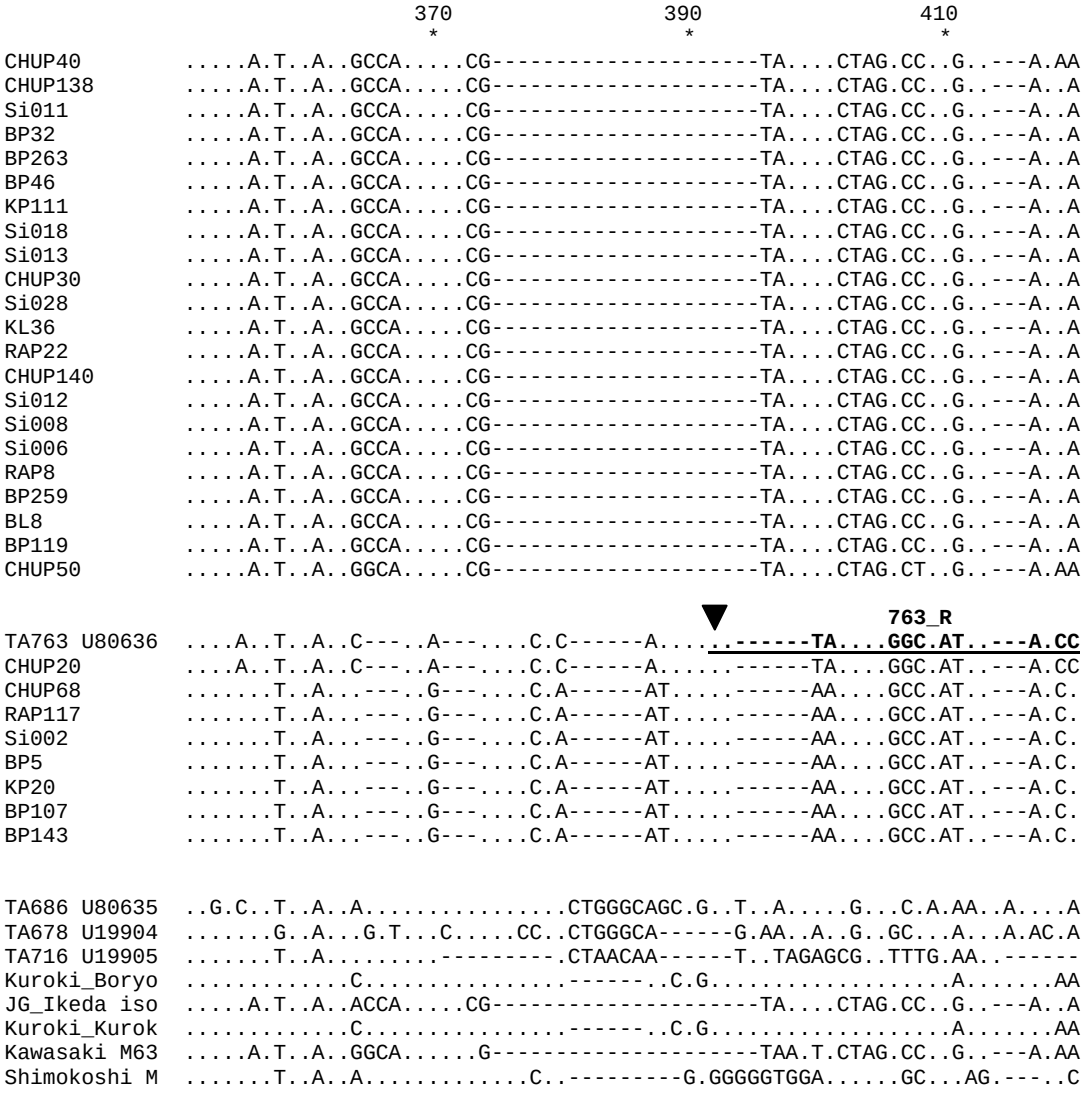
Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 9 of 14)

```
                              300                 320                 340
                               *                   *                   *
CHUP40          ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
CHUP138         ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
Si011           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
BP32            ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
BP263           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
BP46            ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
KP111           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
Si018           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
Si013           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
CHUP30          ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
Si028           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
KL36            ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
RAP22           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
CHUP140         ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
Si012           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
Si008           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
Si006           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
RAP8            ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
BP259           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
BL8             ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
BP119           ............T...CCC.G.T...C.GA.....---CATG...GA..........TG.G.......
CHUP50          ............T...TCC.G.T...C.GA.....---CATG...GA..........TG.........

TA763 U80636    ............T...G..GG.T......A.........---.G....A.GG.......AG.G.......
CHUP20          ............T...G...G.T......A.........---.G....A.GG.......AG.G.......
CHUP68          ............T...G..GG.T......A.........---.AG...A.GG.......TG.G.......
RAP117          ............T...G..GG.T......A.........---.AG...A.GG.......TG.G.......
Si002           ............T...G..GG.T......A.........---.AG...A.GG.......TG.G.......
BP5             ............T...G..GG.T......A.........---.AG...A.GG.......TG.G.......
KP20            ............T...G..GG.T......A.........---.AG...A.GG.......TG.G.......
BP107           ............T...G..GG.T......A.........---.AG...A.GG.......TG.G.......
BP143           ............T...G..GG.T......A.........---.AG...A.GG.......TG.G.......

TA686 U80635    ............T...G..GG.T......A.........---.AA....A.GG........G.G.......
TA678 U19904    ............CA....CGGG.T---...A..-------..CAA....CG.........AG.G..AGC.G
TA716 U19905    ........A....T...G..GG.T......A.........---.AA....A.GG........G.G....C..
Kuroki_Boryo    ............T...G..GG.T......A.........---.A....A.GG........G.G.......
JG_Ikeda iso    ............T...TCC.G.T...C.GA.....---AATG...GA...G.......TG.........
Kuroki_Kurok    ............T...G..GG.T......A.........---.A....A.GG........G.G.......
Kawasaki M63    ............T...CCC.G.TT..C.GA.....---AATG...GA...G.......TG.........
Shimokoshi M    ............TAT..GCCGG.TT.....A.....---CA.GG..GA...G.......TG.G.......
```

Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 10 of 14)



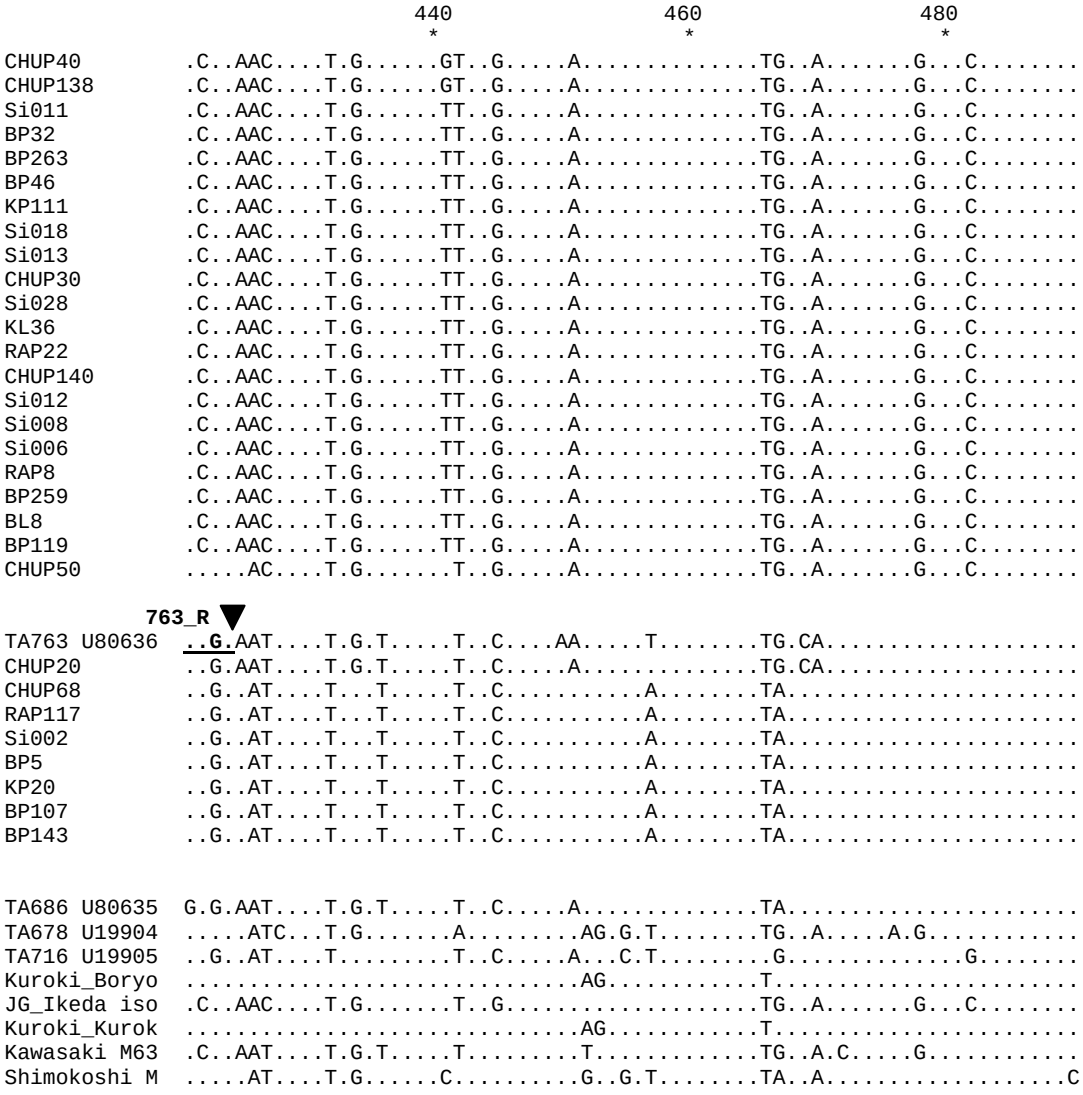
Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 11 of 14)



Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 12 of 14)

```
                                  440               460               480
                                   *                 *                 *
Karp M33004     TTTTGCGATACATAACCATGAGCAATGGAGGCATTTGGTAGTTGGGCTTGCTGCATTATCAAATGCTAAT
Karp UT76       G..........G.............T........AC...T.....................
Karp FPW2031    ..........................................................
KP155           ..........................................................
KL53            ..........................................................
Si023           ..........................................................
KP171           ..........................................................
KP137           ..........................................................
Si024           ..........................................................
BP120           ..........................................................
BL133           ..........................................................
KP174           ..........................................................
KP67            ..........................................................
RAP25           ..........................................................
BP288           ..........................................................
Si015           ..........................................................
Si016           .....A.......G.............T..............................
BP151           .....A.......G.............................................
KP66            .....A.......G.............................................
KP159           .....A.......G.............................................
BP64            .....A.......G.............................................
BP100           G..........G.............T........AC...T.....................
Si020           G..........G.............T........AC...T.....................
Si026           G..........G.............T........AC...T.....................
RAP39           G..........G.............T........AC...T.....................
KP116           G..........G.............T........AC...T.....................
KL102           G..........G.............T........AC...T.....................
Si025           G..........G.............T........AC...T.....................
Si001           G..........G.............T........AC...T.....................
BL9             G..........G.............T........AC...T.....................
BP269           G..........G.............T........AC...T.....................
KP115           G..........G.............T........AC...T.....................
BP41            G..........G.............T........AC...T.....................
KP68            ..................T..C.....T........AC.......................
Si010           ..................T.........................................
RAP13           ..................T.........................................
Si019           ..................T.........................................


Kato M63382     ..G.AGT....T.........T..C.........C.T........TA..A.....A.G.....
Si007           ..G.AGA....T.G.......T..C.........C.T........TA..A.A...........
RAP64           ..G.AGA....T.G.......T..C.........C.T........TA..A.A...........
BP13            ..G.AGA....T.G.......T..C.........C.T........TA..A.A...........

                  GILLIAM_R
                           ▼
Gilliam FPW2    .C..AAC....T.G...GT..G.....A.............TG..A.......G...C.....
Gilliam UT19    .C..AAC....T.G.....GT..G.....A.............TG..A.......G...C.....
KP112           .C..AAC....T.G.....GT..G.....A.............TG..A.......G...C.....
KP128           .C..AAC....T.G.....GT..G.....A.............TG..A.......G...C.....
Si027           .C..AAC....T.G.....GT..G.....A.............TG..A.......G...C.....
Si021           .C..AAC....T.G.....GT..G.....A.............TG..A.......G...C.....
Si022           .C..AAC....T.G.....GT..G.....A.............TG..A.......G...C.....
RAP36           .C..AAC....T.G.....GT..G.....A.............TG..A.......G...C.....
Si009           .C..AAC....T.G.....GT..G.....A.............TG..A.......G...C.....
Si005           .C..AAC....T.G.....GT..G.....A.............TG..A.......G...C.....
RAP107          .C..AAC....T.G.....GT..G.....A.............TG..A.......G...C.....
```

Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 13 of 14)



Multiple sequence alignment of 56-kDa protein encoding gene of 76 sequence samples with the reference sequences of *O. tsutsugamushi* for the strain-specific region selection (page 14 of 14)

BIOGRAPHY

NAME	Mrs. Wichittra Anukool
DATE OF BIRTH	25 February 1979
PLACE OF BIRTH	Lampang, Thailand
INSTITUTION ATTENDED	Kasetsart University, 1996-2002: Doctor of Veterinary Medicine Mahidol University, 2008-2011: Master of Science (Immunology)
RESEARCH GRANTS	Siriraj Graduate Thesis Scholarship
HOME ADDRESS	174 M.8, Lampangluang, Kokha, Lampang. Thailand Tel. 0-5483-2715 E-mail: wiwanvet@yahoo.com
EMPLOYMENT ADDRESS	Veterinary Research and Development Center (Upper Northern Zone) 221 M.6, Wiengtarn, Hangchat, Lampang. Thailand Tel. 0-5483-0195 E-mail: vrd_np@dld.go.th

PUBLICATION / PRESENTATION

1. Duong V, Mai TT, Blasdell K, Lo LV, Morvan C, Lay S, Anukool W, Wongprompitak P, Suputtamongkol Y, Laurent D, Richner B, Ra C, Chien BT, Frutos R, Buchy P. Molecular epidemiology of *Orientia tsutsugamushi* in Cambodia and Central Vietnam reveals a broad region-wide genetic diversity. Infect Genet Evol 2011 (accepted)

PUBLICATION / PRESENTATION (cont.)

2. Wongprompitak P, Anukool W, Wongsawat E, Silpasakorn S, Duong V, Buchy P, Morand S, Frutos R, Ekpo P and Suputtamongkol Y. Country-wide molecular epidemiology of *Orientia tsutsugamushi* in Thailand (in prep)