Epidemiological and bioinformatical analyses of tick-borne pathogens

（マダニ由来病原体に関する疫学ならびに生物情報科学的解析）

Yongjin QIU

Contents

Abbreviations

16S	16 small subunit
α’	Nominal significant level
B.C.	Before Christ
BLSOM	Batch-Learning Self-Organizing Map
CBP	Copperbelt Province, Zambia
CCHF	Crimean-Congo hemorrhagic fever
cDNA	Complementary deoxyribonucleic acid
CP	Central Province, Zambia
CTF	Colorado tick fever
DDBJ	DNA Data Bank of Japan
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
ds	Double strand
EDTA	Ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
EP	Eastern Province, Zambia
gltA	Citrate synthase gene
GS-junior	Genome Sequencer junior

HFf	*Haemaphysalis flava* female
HRT	Heartland
ID	Identification
IOf	*Ixodes ovatus* female
IOm	*Ixodes ovatus* male
IPf	*Ixodes persulcatus* female
IPm	*Ixodes persulcatus* male
KFD	Kyasanur forest disease
LP	Luapula Province, Zambia
LSK	Lusaka, Zambia
MG-RAST	Metagenomics-RAST
MID	Multiplex indicator
MLST	Multilocus sequence typing
NCBI	National Center for Biotechnology Information
NMWCO	Nominal molecular weight cut-off
No.	Number
NP	Northern Province, Zambia
nt	nucleotide
NWP	North-western Province, Zambia
ORF	Open reading frame
p	Difference of ratio

PBS	Phosphate-buffered Saline
PCA	Principal component analysis
PCO	Principal component
PCR	Polymerase chain reaction
POW	Powasan
rDNA	Ribosomal deoxyribonucleic acid
RD	Required difference
RDP	Ribosomal data project
RNA	Ribonucleic acid
S.D.	Standard deviations
SFTS	Sever fever with thrombocytopenia syndrome
SISPA	Sequence-independent single-primer amplification
SOM	Self-Organizing Map
SP	Southern Province, Zambia
ss	Single strand
TBE	Tick-borne encephalitis
U.S.A.	United States of America
UV	Ultraviolet
WP	Western Province, Zambia

Preface

Ticks (ixodida) are relatively large acarines. All are haematophagous arthropods that feed on the blood of vertebrates ranging from mammals and birds to reptiles. The order is divided conventionally into “hard” and “soft” ticks; the former, the Ixodidae, possess a dorsal scutum, whereas the latter, the Argasidae, does not. There are approximately 900 known tick species within three families: Argasidae (soft ticks), Ixodidae (hard ticks) and Nuttalliellidae (monotypic) [40,41]. Ixodidae ticks usually have a larger body and a longer life span than species in other families. The tick life cycle basically consists of four stages; egg, larva, nymph, and adult. They require a large quantity of blood meals for engorgement, molting, and egg production and long intervals off the host between each post-embryotic phase (Figure 1).

Relationships between ticks and humans have been recognized since ancient times as shown by their appearance in the artwork of the Egyptian papyrus scroll of Antef from the time of Thutmos III around 1,500 B.C., where hyaena-like animal is rendered with three excrescences with a round shape resembling ticks on its ear [4]. Aristotle (355 B.C.) in his *Historia Animalium* stated that “ticks are generated from couch grass”, which may be a reference to host-questing ticks [4]. People recognized ticks and their biological behavior at least before 355 B.C., and the historically recorded battle against ticks and tick-borne diseases might have started around that time.

Rocky Mountain spotted fever, which had affected people in the U.S.A. for over a century, is an infectious disease that is proved to be tick-associated. This was the first demonstration of a tick acting as a vector of a microbial disease in humans, and was soon followed by discoveries of many other tick-borne pathogens [89]. At present,

ticks are recognized as important parasitic arthropods in veterinary and medical sciences, because they can harbor and transmit various viruses, bacteria, and protozoan pathogens, which are often zoonotic [10,37,116]. In addition, several tick species can cause a non-infectious disease known as tick paralysis [27].

Opportunities for ticks to come into contact with humans and animals are increasing as their habitats are changing and their distribution is widening. Changes in human behavior such as outdoor recreational activities may also increase the chances of encounters with ticks. Reflecting these ecological, environmental and behavioral factors, the incidence of tick-borne diseases, including that have emerged recently, is on the rise [80,116]. Tick and tick-borne diseases seriously affect animal and human health worldwide with the highest economic loss occurring in livestock production.

Tick-borne bacterial pathogens occupy a considerable proportion of the prokaryotic domain. They include agents of zoonotic diseases that are caused by pathogens such as *Borrelia burgdorferi sensu lato,* the agent of Lyme disease, *Rickettsia spp.* the agents of rickettsioses, *Anaplasma phagocytophilum,* the agent of anaplasmosis, and *Erhlichia spp.,* the agents of erhlichiosis [12,80]. Major tick-borne diseases in humans are listed in Table 1. Ticks harbor not only pathogens but also symbionts, such as *Rickettsia*-, *Wolbachia*-, and *Coxiella*-like bacteria [56,61,90], some of which have potential to cause diseases in the mammalian hosts. In addition, recent studies found several bacterial organisms, such as *Leptospira* and Chlamydiae, which had not been detected in ticks previously [73,115].

On the other hand, tick-borne viral pathogens mainly consist of species from tree families; *Flaviviridae*, *Bunyaviridae*, and *Reoviridae* [6,78]. They sometimes cause fatal human diseases such as tick-borne encephalitis (TBE), Crimean-Congo

hemorrhagic fever (CCHF), and Colorado tick fever (CTF) [8,39,62]. TBE virus is distributed from Europe through Siberia to the Far East. Vector ticks of TBE virus are mainly *I. persulcatus* in the Far East and *I. ricinus* in other regions. Mortality from TBE is 1–30% depending on the subtype of the virus [62]. CCHF is a highly fatal disease that is enzootic in Palearctic, Oriental and Afrotropical regions [8]. Several tick species are considered as vectors or reservoirs of CCHF virus. CTF virus is endemic in western U.S.A. and western Canada. The major symptoms are fever, headache, and nausea, which are not specific to this disease. Fatal cases of CTF are rare. In addition to these well-recorded viral diseases, discoveries of new tick-borne viral diseases are on the rise. For example, new tick-borne viral diseases, such as severe fever with thrombocytopenia syndrome (SFTS) and heartland virus infection have emerged very recently [70,116].

Recently, reports of novel tick-borne pathogens are increasing [80,81]. Thus, ticks may have many potential pathogens that have not been reported previously. Analysis of entire tick microbial population which includes potential bacterial and viral pathogens may be one of practicable approaches to predict emerging tick-borne diseases. However, conventional methods in the detection of bacterial and viral pathogens have several limitations. For example, culturable bacteria are only a small fraction of the total population existing in nature. Upon DNA data analysis, identifying sequences that have low or no homologies with known DNA fragments registered in the DNA databases, is difficult for further taxonomic analyses. Thus, alternative strategies effective for microbial population analyses are required.

The present thesis consists of three chapters. In chapter I, the prevalence of *Coxiella burnetii*, which is a tick-borne zoonotic bacterial pathogen, has been investigated in domestic animals in Zambia. In chapter II, the microbial populations in

ticks using 16S rDNA amplicon pyrosequencing technology have been analyzed to reveal the bacterial organisms present in tick salivary glands. In chapter III, to elucidate tick viral flora or "virome", DNA/cDNA sequences have been analyzed by using shotgun sequencing technique and a unique bioinformatic procedure.

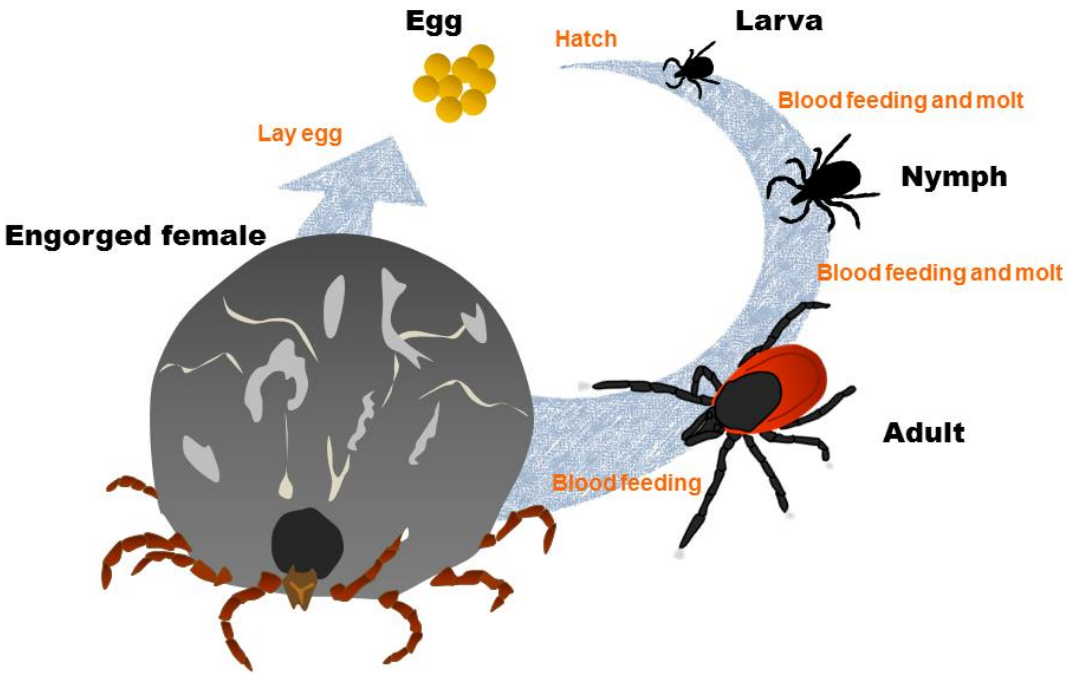

Figure 1. Ixodid tick life cycle.

Ticks have four distinct life stages; egg, larva, nymph, and adult.

Table 1. Major tick-borne diseases in human.

Disease	Pathogen	Main vector	Area	Reference
Rocky Mountain spotted fever	*Rickettsia ricketsii*	*Dermacentor andersoni, D. variabilis*	Continental America	51,83
Japanese spotted fever	*R. japaonica*	*Ixode ovatus, Haemaphysalis flava, Haemaphysalis longicornis*	Asia	65
African tick bite fever	*R. africae*	*Amblyomma. variegatum, A. hebraeum*	Africa	51,83
Mediterranean spotted fever	*R. conorii*	*Rhipicephalus sanguineus*	Europe, Africa, India, Oriental region	51,83
Queensland tick typhus	*R. australis*	*I. holocyclus*	Australia	95
Lyme disease	*Borrelia burgdorferi sensu lato*	*Ixodes persulcatus, Ixodes ricinus*	Worldwide	51
Anaplasmosis	*Anaplasma phagocytophilum*	*I. ricinus, I. scapularis, I. pacificus*	Europe, North America	51,83
Erhlichiosis	*Erhlichia chaffeensis*	*A. americanum*	America	51,83
Q fever	*Coxiella burnetii*	Many species of different genera	World wide	51
Crimean-Congo haemorrhagic fever	CCHF virus	*Hyalomma marginatum*	Africa, Balkans, Middle East and Asia	51,83
Tick-borne encephalitis	TBE virus	*I. persulcatus, I. ricinus*	Europe, Siberia, Far East	59,83
Kyasanur forest disease	KFD virus	*Haemaphysalis spinigera*	India, Saudi Arabia, China	51,83
Omsk haemorrhgic fever	OHF virus	*I. persulcatus, Dermacentor reticulatus*	Siberia	83
Powassan virus infection	POW virus	*I. cookei, I. scapularis*	North America, Canada	25,59
Severe fever with thrombocytopenia syndrome	SFTS virus	*H. longicornis*	East Asia	116
Heartland virus infection	HRT virus	*A. americanum*	America	70
Human babesiosis	*Babesia microti, B. divergens*	*I. ricinus, I. scapularis*	Europe, North America	51,83

Chapter I

First genetic detection of *Coxiella burnetii* in Zambian livestock

Introduction

Coxiella burnetii, an obligate intracellular gram-negative bacterium, is the causative agent of Q fever in humans and wide range of animals, including cattle, goat, cat, dog, and wild animals. It causes a variety of symptoms such as acute flu-like symptoms, pneumonia, hepatitis, and chronic endocarditis in humans [65,69]. It also causes abortion or infertility in animals [65,69]. The disease is a ubiquitous zoonosis with worldwide distribution [65]. From spring 2007–2011, a Q fever outbreak of unprecedented scale occurred in the Netherlands, involving 4,108 notified human cases including 24 fatal cases [107]. Epidemiological studies conducted to investigate the source(s) of infection, which clearly identified several sources. One of them was public visit of an ovine farm functioning as a healthcare farm for daily activities, and this kind of activities are popular in Netherlands during the lambing season [94]. The life cycle of *C. burnetii* is not fully understood, but humans are considered incidental hosts. *C. burnetii* has a variety of reservoir including domestic and wild animals and arthropods such as ticks [102]. However, tick bite is not considered as a transmission route of this disease to humans, although crushing an infected tick by fingers has resulted in Q fever [28]. The role of ticks as vectors and reservoirs has been discussed previously, but there are no reports of Q fever associated with tick bite [23].

Of note, domestic ruminants including cattle, goats, and sheep are often infected and serve as main sources of the human infections [65,113]. In several African

countries including Zambia, there is the report of the sero-prevalence of *C. burnetii* in humans [26,35,54,79]. There is also a report of genetic detection of *C. burnetii* DNA from febrile patients in the malaria endemic area in Senegal [86]. These data suggest that Q fever could distribute widespread in African countries. The aim of this study was to investigate the prevalence of *C. burnetii* in domestic animals in Zambia and to extrapolate the potential reservoir of *C. burnetii*.

Materials and Methods

Animal blood samples

In Chama (N = 295, 11°21'S, 33°16'E), Chongwe (N = 50, 15°33'S, 28°69'E), Monze (N = 80, 16°28'S, 27°49'E), and Petauke (N= 64, 14°24'S, 31°32'E) districts of Zambia, blood samples were collected from the Angoni breed cattle from 2008 to 2010 (Figure 2). Boer breed goat blood samples were also obtained in Chama district. In each district, sampling was conducted at 2–7 different sites where the pastured cattle and goats were gathered by the owners. In totally, 489 cattle and 53 goat blood samples were collected.

DNA extraction and Conventional PCR method

Genomic DNA was extracted using the DNA Isolation Kit for Mammalian Blood (Roche Molecular Biochemical, Boehringer, Germany). The DNA was extracted from 1 ml of EDTA-treated blood and was eluted in final volume of 200 μl according to the manufactures instructions. For the genomic detection of *C. burnetii* infection, polymerase chain reaction (PCR) was performed with previously designed primers (Trans1: 5'-TATGTATCCACCGTAGCCAGTC-3' and Trans2: 5'-CCCAACAACACCTCCTTATTC-3'), which target a repetitive transposon-like element of the *C. burnetii* genome, and the expected product of amplification of these primers was 687 bp in length [111]. The sensitivity and specificity of the assay have been well evaluated in the previous studies, and the target element exist at least 19 copies in *C. burnetii* Nine Mile I genome [11,106]. The PCR reaction was conducted in a final volume of 10 μl, containing 5 μl of KAPA Blood PCR Mix B (Kapa Biosystems,

Boston, MA), 1.25 pmol of each primer, and 1 μl of template DNA. The PCR conditions started with a denaturation step at 95°C for 5 min, followed by 40 cycles of 95°C for 30 sec, 60°C for 30 sec, and 72°C for 1 min, and final extension step at 72°C for 2 min. The resulting PCR products (approx. 687 bp) were electrophoresed on 1% agarose gel, stained with Gel-Red (Biotium, Hayward, CA), and visualized with a UV trans-illuminator.

Previous study of genotyping of *C. burnetii* showed that a correlation between genotype and duration of infections, acute or chronic [36]. To estimate what genotype of *C. burnetii* distributed in Zambia, Multilocus Sequence Typing (MLST) was performed with established primers, which targeted spacer namely Cox2, Cox5, Cox18, Cox20, Cox22, Cox37, Cox51, Cox56, Cox57, and Cox61 (Table 2) [36]. The PCR reaction was conducted in a final volume of 10 μl, containing 5 μl of PCR Buffer for KOD FX Neo (Toyobo, Tokyo, Japan), 2 μl of 2 mM of each deoxynucleoside triphosphate (dNTP) mixture, 3 pmol of each primer, 0.2 μl of KOD FX Neo, 0.75 μl of template DNA, and 1.45 μl of distilled water. The PCR conditions started with a denaturation step at 94°C for 2 min, followed by 40 cycles of 98°C for 10 sec, 57°C for 30 sec, and 68°C for 1 min. The resulting PCR products were electrophoresed on 1% agarose gel, stained with Gel-Red (Biotium, Hayward, CA), and visualized with a UV trans-illuminator.

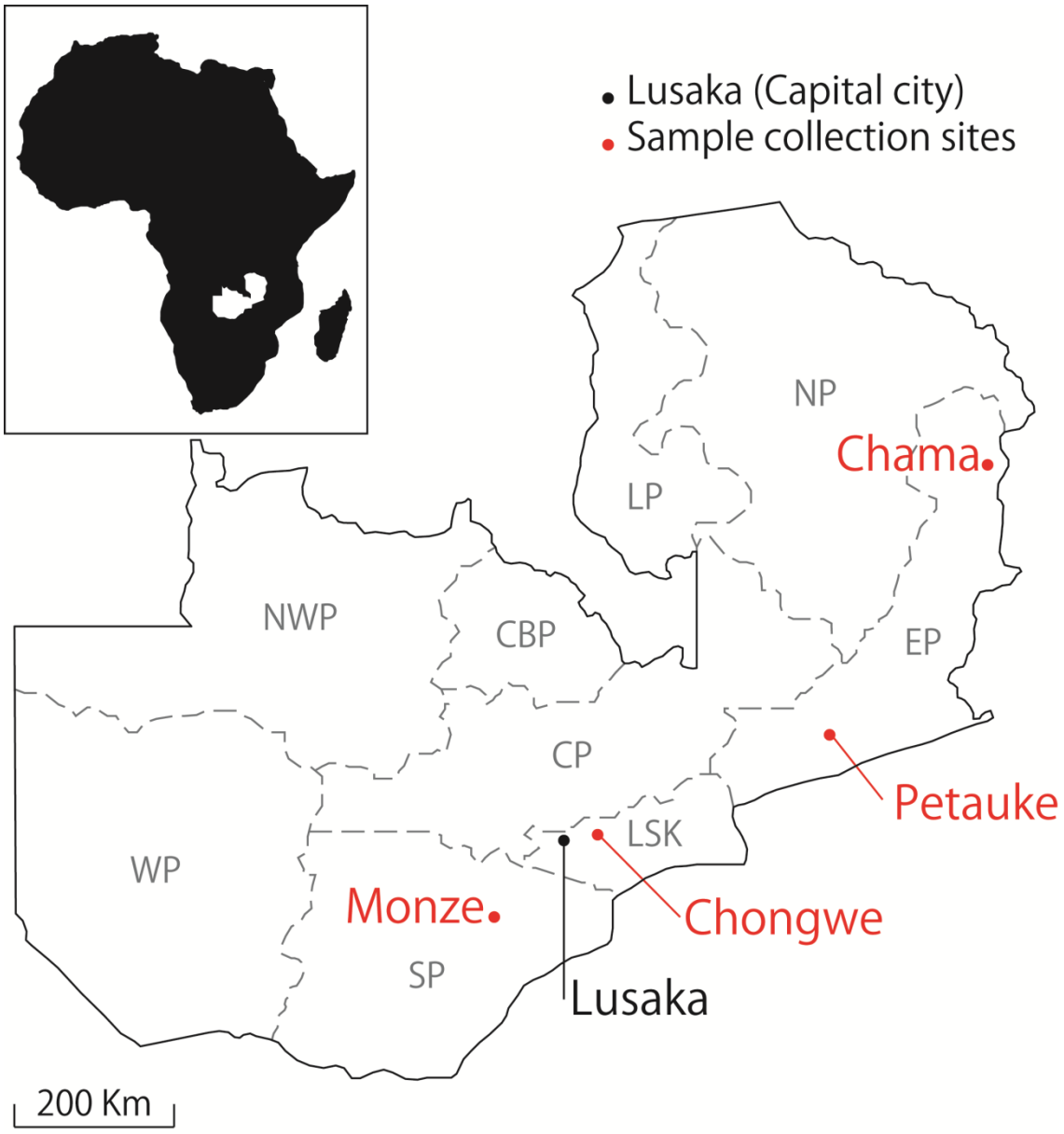

Figure 2: Map of Zambia showing its nine provinces: Northern (NP), Eastern (EP), Luapula (LP), Central (CP), Copperbelt (CBP), Lusaka (LSK), Southern (SP), Western (WP), and North-western (NWP) provinces. The *Coxiella burnetii* samples used in this study were obtained from Monze (SP), Chongwe (LSP), Petauke (EP), and Chama (EP) districts.

Table 2. MLST primers.

Spacer name	Open Reading Frame	Nucleotide sequence (5'-3')	Expected length (bp) of amplified fragment
Cox2	Hypothetical protein	Cox20766 CAACCCTGAATACCAAGGA	397
	Hypothetical protein	Cox21004 GAAGCTTCTGATAGGCGGGA	
Cox5	Surface domain protein	Cox77554 CAGGAGCAAGCTTGAATGCG	395
	Entericidin, putative	Cox77808 TGGTATGACAAGCTTGAATGCG	
Cox18	Ribonuclease H	Cox28060 CGCAGACGAATTAGCCAATC	557
	DNA polymerase III, epsilon subunit	Cox283490 TTCGATGATCCGATGGCCTT	
Cox20	Hypothetical protein	Cox365301 GATATTTATCAGCGTCAAAGCAA	631
	Hypothetical protein	Cox365803 TCTATTATTGCAATGCAAGTGG	
Cox22	Hypothetical protein	Cox378717 GGGAATAAGAGAGTTAGCTCA	383
	Amino acid permease family protein	Cox378965 CGCAAATTTCGGCACAGACC	
Cox37	Hypothetical protein	Cox657471 GGCTTGTCTGGTGTAACTGT	463
	Hypothetical protein	Cox657794 ATTCCGGGACCTTCGTTAAC	
Cox51	Replicative DNA helicase, intein-contaning	Cox824598 TAACGCCCGAGAGCTCAGAA	674
	Conserved hypothetical protein Uridine kinase	Cox825124 GCGAGAACCGAATTGCTATC	
Cox56	*OmpA*-like transmembrane domein protein	Cox886418 CCAAGCTCTCTGTGCCCAAT	479
	Conserved hypothetical protein	Cox886784 ATGCGCCAGAAACGCATAGG	
Cox57	Rhodanese-like domain protein	Cox892828 TGGAAATGGAAGGCGGATTC	617
	Hypothetical protein	Cox893316 GGTGGAAGGCGTAAGCCTTT	
Cox61	Dioxygenase, putative	Cox956825 GAAGATAGAGCGGCAAGGAT	611
	Hypothetical protein	Cox957249 GGGATTTCAACTTCCGATAGA	

Results

For cattle, 38 out of 489 samples were *C. burnetii*-positive by PCR (Table 3). The prevalence of *C. burnetii* infection differed among the four sampling sites, with the highest prevalence observed in Chama (33 of 295, 11.2%), followed by Chongwe (3 of 50, 6.0%), Petauke (2 of 64, 3.1%), and Monze (0 of 80, 0%). For goat, out of 53 samples from Chama, only 4 (7.5%) were positive for *C. burnetii* (Table 3). According to a multiple comparison analysis (Ryan's method) among the 4 areas, the prevalence in Chama was significantly higher than that of Monze (α':0.0083, *RD*:0.089, *p*:0.1119). Amplified fragments of each spacer for MLST were not obtained from *C. burnetii*-positive samples.

Table 3. Prevalence of *Coxiella burnetii* DNA in Zambian cattle and goats.

Animal	Area	Prevalence*	
Goat	Chama	4/53 (7.5%)	
Cattle	Chama	33/295 (11.2%)	*
	Chongwe	3/50 (6.0%)	
	Monze	0/80 (0%)	
	Petauke	2/64 (3.1%)	
	Sub-total	38/489 (7.8%)	

*No. of positives/No. of tested.

*Significantly different by multiple comparison analysis

Discussion

This study revealed that *C. burnetii* existed in Zambia, which is same as other African countries [26]. The major infection routes of *C. burnetii* to humans are considered through the inhalation of aerosol following parturition of an infected animal and the ingestion of contaminated raw milk or milk products [30]. *C. burnetii* DNA was detected from 38 cattle and 4 goats. A previous study reported that the prevalence of antibodies against *C. burnetii* in humans living extensive cattle-breeding areas (Eastern province and Western province) was higher than less breeding areas (North province) [79]. In this study, the highest detection of *C. burnetii* DNA was recorded in the samples collected from the Chama area (Eastern province). Both of the results conducted by serological and genetical methods indicates that this disease poses higher risk rather in the eastern part of Zambia.

Taken together, results indicated that domestic animals are one of the risk factors for human infection with *C. burnetii* in Zambia. Collectively, these data may also imply that Eastern province is endemic area for *C. burnetii* infection in Zambia. Therefore, the people should be aware of the infection of *C. burnetii* as a case of non-malarial febrile illness.

To clarify the epidemiology of Q fever, such as animal to animal or to human spread of the agents, transboundary movement of the disease, and risk analysis in public health, genotyping of the bacterium such as MLST is necessary. Unfortunately, the genotyping of circulating strains in Zambia was not successful. One possible explanation is that concentration of *C. burnetii* DNA in samples was too low to amplify targeted spacer regions. Further studies are required to expand the other

sampling areas and specimen such as vaginal swab, placenta, and milk, which are likely to contain the higher concentration of *C. burnetii* than blood samples [54].

In this study, domestic animals were considered as one of risk factor of Q fever in humans in Zambia. Some report suggested that ticks may play a role for the ecology of C. burnetii, especially in the animal to animal transmissions [102]. However, there is little evidence to support their role in bacterial transmission to humans, except one case report of the percutaneous infection following crushing of infected ticks between fingers [28]. For better understanding of *C. burnetii* ecology in Zambia, surveillance of *C. burnetii* in Zambian ticks is required.

Summary

Q fever is a widespread zoonosis caused by *Coxiella burnetii*, an obligate intercellular gram-negative bacterium. Investigation of *C. burnetii* infection in livestock animals in Zambia was carried out using molecular detection techniques. A total of 489 cattle and 53 goat blood samples were collected from 4 local sites (Chama, Chongwe, Monze, and Petauke). Molecular screening by polymerase chain reaction (PCR) was performed using *C. burnetii* species-specific primers. The prevalence of *C. burnetii* differed among four sites and the highest prevalence was observed in Chama which is located in Eastern province in Zambia. The present study reports the first genomic detection of *C. burnetii* in Zambia.

Chapter II

Microbial population analysis of the salivary glands of ticks; a possible strategy for the surveillance of bacterial pathogens

Introduction

In chapter I, prevalence of *C. burnetii* infection, one of the tick-borne diseases, was analyzed in domestic animals of Zambia. Better understanding of tick-borne microbes may improve our preparedness for tick-borne diseases and prediction of emerging tick-borne pathogens. Animal pathogens transmitted by ticks cover a variety of biological agents, bacteria, viruses, and protozoans [10,37,116]. Currently available methods for pathogen detection, including the most conventional microbiological procedures such as isolation and culture, morphological detections by electron microscopes, and molecular biological means such as PCR, have their technical limitations in detecting such a wide variety of organisms. In other words, there is no universal method for detection and identification of these pathogens, and effective methods to detect and characterize unknown microbes have not been established.

Concept of "Metagenomic" is firstly observed in 1998 [42]. The most advanced feature of this concept is massive detection and characterization of microorganisms even unculturable ones. Recent progress in high throughput DNA sequencing technologies gives us Giga base amounts of sequence information [17,75,76]. In addition, a fusion of the "Metagenomic" and high throughput sequencing technologies has led to analysis of microbial community and to discovery of organisms, which have not been known previously, in a variety of samples including soil and water

[112]. The diversity of microbes associated with ticks was also revealed by using high throughput sequencing technology, since unculturable microorganisms represent the huge majority of bacteria including pathogenic, commensal, and symbiotic microorganisms [3,17,72,73]. The analysis of 16S ribosomal DNA (16S rDNA) amplicons by pyrosequencing is the method specialized in analyzing bacterial communities that can be applicable to detect unculturable microorganisms, and reveal entire bacterial populations in samples.

In this chapter, 16S rDNA amplicon analysis method was applied to bacterial community in tick salivary glands. Because a lot of tick-borne pathogens, such as *Anaplasma, Ehrlichia*, and *Rickettsia*, concentrate within the salivary glands, and are transferred into the host animal during blood feeding term [34,84,93]. Additionally, the presence of non-pathogenic bacteria, such as *Coxiella*-like symbiont, has been found in tick salivary glands in previous studies done by using electron microscopy [56]. These facts suggest that tick salivary glands have rich bacterial community and these bacteria including potential pathogens have a chance of transition to animal host together with tick saliva. Thus, the investigation of the bacterial community in the tick salivary glands may help for a better understanding of the microbes including pathogens.

Materials and Methods

Sample collection and DNA preparation

Adult host-questing ticks were collected by flagging flannel sheets from vegetation area in the foothills of Mount Fuji Shizuoka Prefecture, Japan, where Japanese spotted fever is endemic from 2008 to 2010 [43]. Table 4 indicates information on the sampling sites. Tick species were identified morphologically using a taxonomical key. Three tick species were used for this study. The sample numbers of *I. ovatus*, *I. persulcatus*, and *H. flava* used for this study were 24 (14 female, 10 male), 12 (6 female, 6 male), and 5 (female only), respectively. Firstly, tick specimen was split into two parts (anterior and posterior) at the area between coxa 1 and coxa 2 as indicated with a red arrow in figure 3A. In the case of tick species used in this study, the midgut does not exceed the anterior edge of coxa 2, that is, the area between coxa 1 and coxa 2 is free of the midgut. The anterior part was then removed from the posterior part using sterile forceps (Figure 3B). Since the salivary glands were always attached with the anterior part, they could be removed from the tick carcass without damaging the midgut (Figure 3C). The salivary glands were then collected into a sterile 1.5 ml tube using sterile forceps, followed by washing with sterile PBS (pH 7.0) in order to minimize bacterial contamination. All dissection steps were performed under a stereomicroscope with great cares to avoid the contamination from the midgut fluid. Genomic DNA was individually extracted using QIAamp DNA Mini kit (QIAGEN, Tokyo, Japan) according to the manufacturer's instructions, and stored at -20°C. Samples of *I. ovatus* female and male, *I. persulcatus* female and male, and *H. flava* female are indicated by IOf, IOm, IPf, IPm, and HFf, respectively, throughout this chapter (Table 5).

Table 4. Longitude and latitude of sampling sites.

Site no.	North latitude	East longitude
1	35.24	138.74
2	34.89	138.95
3	35.31	138.77
4	35.32	138.73
5	35.20	138.77
6	35.34	138.56
7	35.26	138.53
8	35.30	138.73
9	35.09	138.88

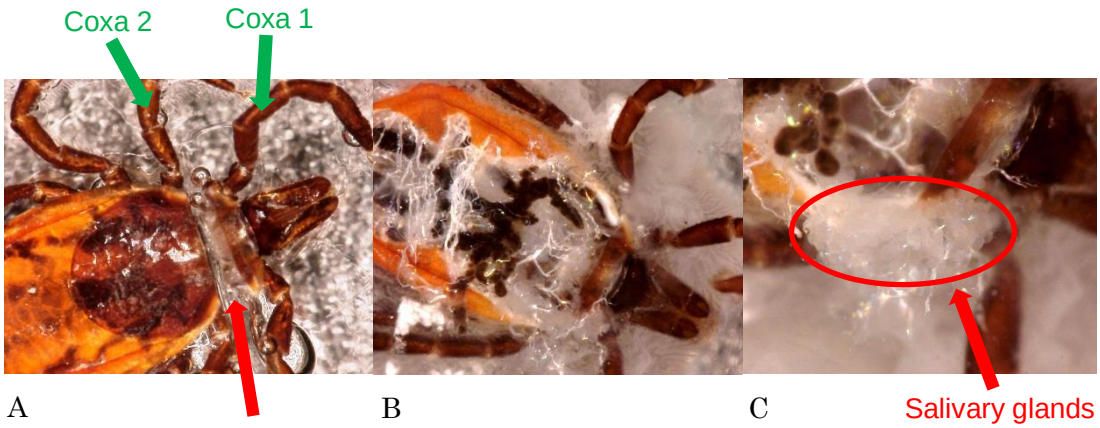

Figure 3. Dissection of ticks to collect salivary glands.

A: Split into two parts at the area between coxa 1 and coxa 2.

B: The anterior part was removed from the other part.

C: Salivary glands were removed without damaging the midgut.

PCR amplification of V1 to V3 regions for 16S rDNA amplicon libraries

The V1 to V3 hyper variable regions of bacterial 16S rDNA were amplified by PCR using the previously established universal primers 27F (5'-X-AGAGTTTGATCMTGGCTCAG-3') and 518R (5'-ATTACCGCGGCTGCTGG-3'), corresponding to positions 27 to 518 of the *Escherichia coli* 16S rDNA [18,55]. Ten bases of a multiplex indicator (MID) tag sequence denoted as 'X', was attached to the 27F primer. Primers 27F and 518R were modified with 5'-adapter A (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3') and 5'-adapter B (5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3') sequences, respectively, for pyrosequencing (Roche, Basel, Switzerland). PCR was performed in a total volume of 50 μl, containing PCR buffer, 0.2 μl of Platinum Taq DNA polymerase (Life technologies, Tokyo, Japan), 0.2 mM of each primer, 1μl of 10 mM dNTPs, 1.5 μl of 50 mM $MgCl_2$, and 1μl of template DNA. The PCR conditions started with a denaturation step at 94°C for 2 min, followed by 30 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min. The resulting PCR products (approx. 500 bp) were assessed by agarose 1% gel electrophoresis, followed by purification using the Wizard SV Gel and PCR Clean-Up System (Promega, Tokyo, Japan). Quality and concentration of the amplicon libraries were assessed with an Agilent 2100 BioAnalyzer (Agilent Technologies, Palo Alto, USA) using a DNA 1000 lab chip (Agilent Technologies).

Pyrosequencing and data analysis

Amplicon libraries with different MID tags were gathered and subjected to pyrosequencing using a 454 Genome Sequencer Junior (GS-Junior; Roche) following the manufacturer's protocol. The pyrosequencing data were deposited in the DNA Data

Bank of Japan (DDBJ) (http://www.ddbj.nig.ac.jp) with accession no. DRA001731. The resulting data files (standard flowgram format, .sff files) were converted to FASTA files and sorted according to sample-specific MID tags using CLC Genomics Workbench (CLC Bio, Tokyo, Japan). Primers and ten base MID tag barcode sequences were trimmed, then short reads (<150 bp) and low quality reads were removed. DECIPHER's Find Chimeras web tool (http://decipher.cee.wisc.edu/FindChimeras.html) was used to remove chimeric sequences [114]. The remaining reads were phylogenetically classified using Ribosomal Database Project (RDP) 16S Classifier version 10 (http://rdp.cme.msu.edu/index.jsp) with 80% of confidence threshold, which can accurately and rapidly provide assignments for domains to the genus level [21]. A comparative analysis of each sample was performed using the MG-RAST metagenomics analysis server employing the RDP dataset (http://metagenomics.anl.gov/). Alpha diversity of each sample was also calculated using the MG-RAST server. Data sets were represented as the mean 6 standard deviations (S.D.) after the Smirnov-Grubbs outlier test (a = 0.05).

Rickettsia-specific PCR

Rickettsia-specific PCR amplification of the citrate synthase gene (*gltA*) using the primers RpCS877p (5'-GGGGGCCTGCTCACGGCGG-3') and RpCS1273r (5'-CATAACCAGTGTAAAGCTG-3') was performed on 22 samples that were highlighted by RDP analysis containing the genus *Rickettsia* [91]. PCR was performed in a final volume of 25 μl containing PCR buffer for KOD-Plus-Neo, 0.5 μl of KOD-Plus-Neo DNA polymerase (Toyobo, Tokyo, Japan), 0.3 mM of each primer, 2.5 μl of 2 mM dNTPs, 1.5 μl of 25 mM $MgSO_4$, and 1 μl of template DNA. The PCR

condition started at 94℃ 2 min for denaturation, followed by 40 cycle of 94℃ for 15 sec, 54℃ for 30 sec, and 68℃ for 30 sec, and 68℃ for 2 min as a final extension step. ExoSap-IT (Affymetrix, Tokyo, Japan) were used for the purification of PCR products according to the manufacturer's instructions. Sequencing was performed using BigDye v3.1 terminator chemistry (Applied Biosystems, Tokyo, Japan) and the forward and revers primers. Sequence products were analyzed on a 3130xl Genetic Analyzer (Life Technologies, Tokyo, Japan) and using ATGC software (GENETYX Corporation, Tokyo, Japan).

Full-length 16S rDNA sequencing analysis

There are sequence reads that could not be classified into the genus level by RDP analysis. To characterize these sequence reads, the almost full-length 16S rDNA gene was amplified from four *I. persulcatus* female samples by PCR using the universal primers fD1 (5'-AGAGTTTGATCCTGGCTCAG-3') and Rp2 (5'-ACGGCTACCTTGTTACGACTT-3') [110]. PCR was performed in a total volume of 50 μl containing PCR Buffer for KOD-Plus-Neo, 1 μl of KOD-Plus-Neo DNA polymerase, 0.3 mM of each primer, 5 μl of 2 mM dNTP mixture, 3 μl of 25 mM $MgSO_4$, and 2 μl of DNA template. PCR conditions consisted of a denaturation step at 94℃ for 2 min, followed by 40 cycle of 94℃ for 15 sec, 55℃ for 30 sec, and 68℃ for 45 sec, and a final extension step at 68℃ for 2 min. Quality of the PCR products (approx. 1,400 bp) was assessed by agarose 1% gel electrophoresis, followed by purification of the products using the Wizard SV Gel and PCR Clean-Up System (Promega). PCR products were A-tailed and then cloned with TA cloning plasmids pGEM-T Easy (Promega). Ten clones per sample were randomly selected and

sequenced.

Sanger sequencing data analysis

Sanger sequencing data were analyzed using GENETYX version 9.1 (GENETYX Corporation, Tokyo, Japan). The GenBank accession numbers for the *gltA* sequences are AB911107 to AB911109, and the 16S rDNA sequences AB906824 to AB906829. Sequences were compared with those in public databases using nucleotide BLAST at NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Phylogenetic analysis was conducted using MEGA version 6.05 [100]. The universal 16S rDNA sequences were aligned with those of closely related bacteria in GenBank using ClustalW and a maximum likelihood phylogram was constructed.

Results

Classification and quantification of bacterial taxa

Between 3,351 and 9,788 sequence reads were obtained for individual *I. ovatus*, of which almost 98% were assigned to the genus level (Table 5 and Figure 4A). A total of 71 bacterial genera were detected in *I. ovatus*, with 59 found in males and 37 in females. The two dominant bacterial genera were *Spiroplasma* and *Coxiella*, and these accounted for more than 90% of the bacterial community in ticks, except for a single *I. ovatus* female and 3 *I. ovatus* males (Figure 4A). *Rickettsia* (genus contains known tick-borne pathogens *R. japonica* and *R. helvetica*) was detected in ten samples, *Ehrlichia* (genus contains known tick-borne pathogens *E. chaffeensis* and *E. muris*) was detected two samples.

Between 3,599 and 16,117 sequence reads were recorded for individual *I. persulcatus*, with almost 82% assigned to the genus level, except for those of four *I. persulcatus* females (Table 5 and Figure 4B). By the RDP classifier, these reads were classified as the phylum of Proteobacteria, while some of them were classified as the class of Alphaproteobacteria. These were 127 different bacterial genera detected in *I. persulcatus*, of which 92 were detected in males, and 81 in females. *Rickettsia* was detected in nine *I. persulcatus* (4 female, 5 male) individuals and *Ehrlichia* was detected in a single *I. persulcatus* male (IPm5).

Between 6,204 and 10,017 sequence reads were obtained for individual *H. flava*, of which almost 97% were identified to the genus level (Table 5 and Figure 4C). A total of 59 different bacterial genera were detected, and *Coxiella* accounted for more than 90% of the microbial population in all samples. *Spiroplasma* was not detected in

any individuals of *H. flava*, while appearing in all *Ixodes* samples. *Rickettsia* spp. were detected in three *H. flava* females, and no sample contained *Ehrlichia* sp.

A summarized diagram of the number of bacterial genera detected in each tick group is presented in Figure 5. Out of 163 different genera identified, 18 were detected in all tick groups. These were *Acinetobacter*, *Arcicella*, *Burkholderia*, *Corynebacterium*, *Coxiella*, *Cryobacterium*, *Curvibacter*, *Flavobacterium*, *Limnohabitans*, *Methylobacterium*, *Novosphingobium*, *Polynucleobacter*, *Propionibacterium*, *Pseudomonas*, *Rickettsia*, *Sphingomonas*, *Staphylococcus*, and *Streptophyta*. Some bacterial genera were uniquely associated with tick species or sex, i.e., IOf (1 genus), IOm (13 genera), IPf (24), IPm (35) and HFf (19).

Table 5. Sequence results and number of detected genera.

Sample ID	Tick species	Sex	No. of sequence reads	No. of genera
IOf1	*I. ovatus*	female	5,664	10
IOf2	*I. ovatus*	female	4,484	10
IOf3	*I. ovatus*	female	3,498	6
IOf4	*I. ovatus*	female	4,712	7
IOf5	*I. ovatus*	female	5,591	3
IOf6	*I. ovatus*	female	4,200	5
IOf7	*I. ovatus*	female	5,030	6
IOf8	*I. ovatus*	female	5,634	23
IOf9	*I. ovatus*	female	7,643	8
IOf10	*I. ovatus*	female	5,636	7
IOf11	*I. ovatus*	female	4,049	2
IOf12	*I. ovatus*	female	7,275	14
IOf13	*I. ovatus*	female	3,351	3
IOf14	*I. ovatus*	female	7,048	7
IOm1	*I. ovatus*	male	4,986	13
IOm2	*I. ovatus*	male	3,790	12
IOm3	*I. ovatus*	male	7,916	22
IOm4	*I. ovatus*	male	3,844	5
IOm5	*I. ovatus*	male	6,340	18
IOm6	*I. ovatus*	male	7,130	22
IOm7	*I. ovatus*	male	6,176	22
IOm8	*I. ovatus*	male	9,788	28
IOm9	*I. ovatus*	male	9,628	12
IOm10	*I. ovatus*	male	7,170	17
IPf1	*I. persulcatus*	female	8,964	38
IPf2	*I. persulcatus*	female	3,599	16
IPf3	*I. persulcatus*	female	7,085	42
IPf4	*I. persulcatus*	female	8,242	40
IPf5	*I. persulcatus*	female	7,943	25
IPf6	*I. persulcatus*	female	10,506	18
IPm1	*I. persulcatus*	male	7,414	55
IPm2	*I. persulcatus*	male	8,173	19
IPm3	*I. persulcatus*	male	10,803	26
IPm4	*I. persulcatus*	male	10,144	34
IPm5	*I. persulcatus*	male	16,117	29
IPm6	*I. persulcatus*	male	9,221	40
HFf1	*H. flava*	female	6,438	3
HFf2	*H. flava*	female	8,339	44
HFf3	*H. flava*	female	6,204	5
HFf4	*H. flava*	female	10,017	18
HFf5	*H. flava*	female	8,294	24

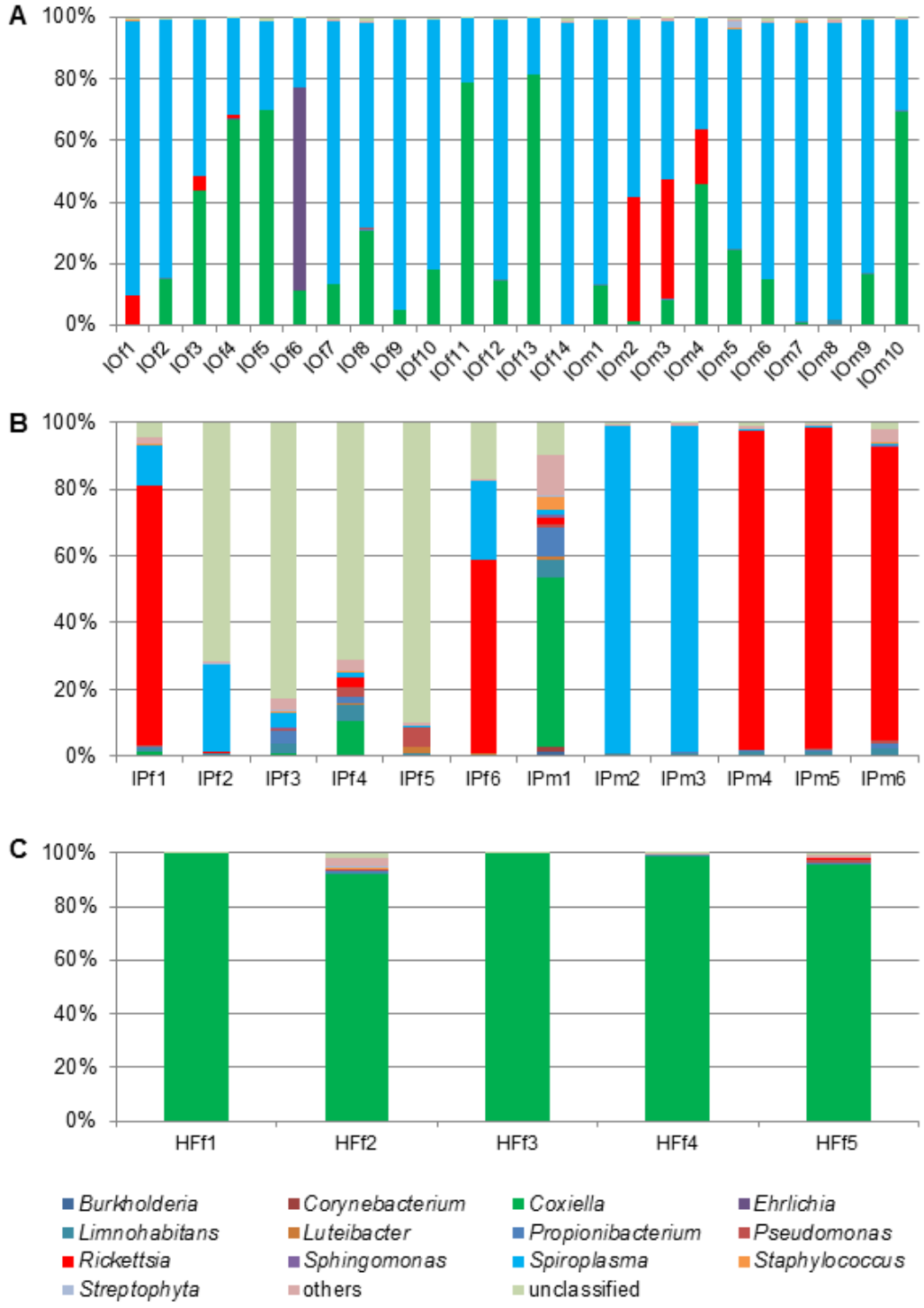

Figure 4. Relative abundances of different bacterial genera in the salivary glands of (A) *I. ovatus*, (B) *I. persulcatus* and (C) *H. flava*. All genera with less than 1.0% contribution were pooled into one group and labeled ‘‘others’’.

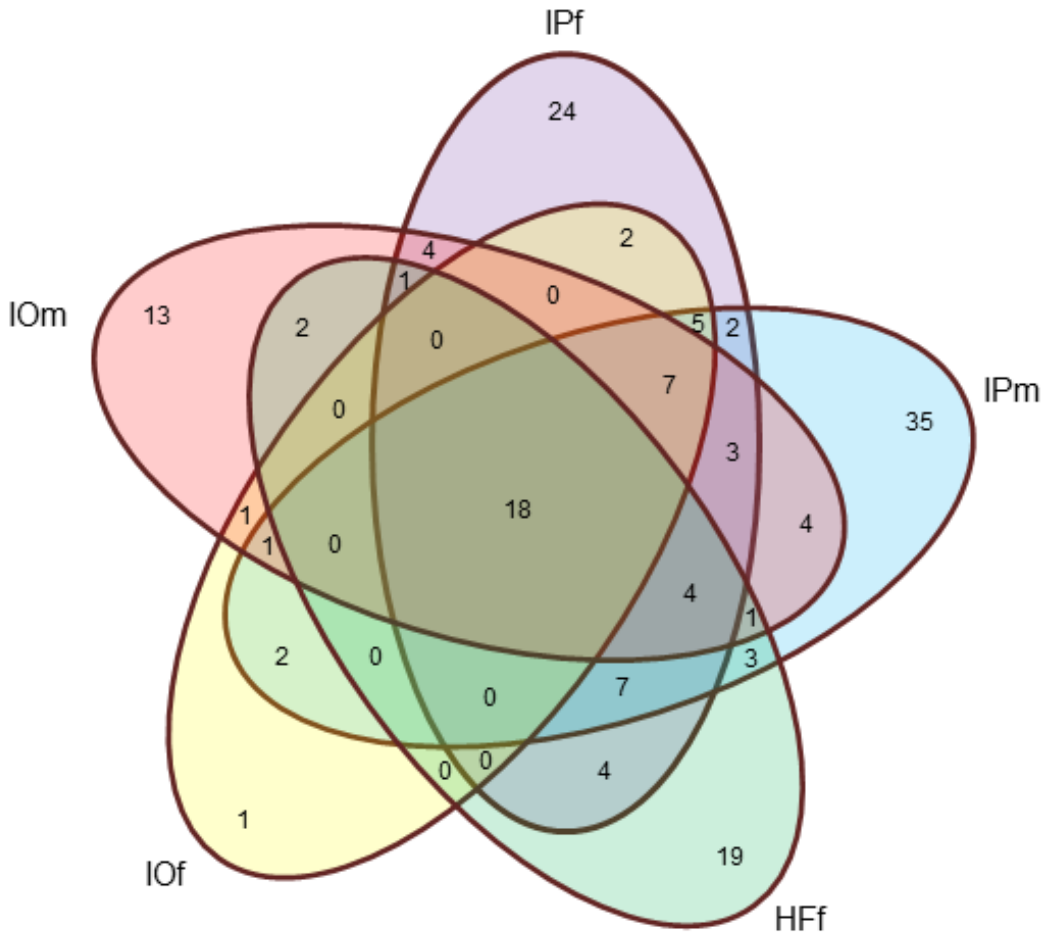

Figure 5. Venn diagram of all 163 identified genera distributed across the tick species and sex.

Comparison of microbiomes in salivary glands between tick species

Principal component analysis (PCA) was performed using the MG-RAST server with normalized values and Bray-Curtis distance (Figure 6) for each tick sample. The microbial community composition of each sample clustered approximately according to tick species. The microbial populations of *Ixodes* and *Haemaphysalis* were completely separated by Principal component 2 (PCO2). The microbial community composition of *Haemaphysalis* ticks was broadly distributed along PCO1; however, in *I. ovatus* and *I. persulcatus* microbial populations were more distinct, but with some overlap within this component.

Alpha diversity (Shannon diversity index) for each sample was calculated using the MGRAST server (Figure 7). Smirnov-Grubbs's outlier test (a= 0.05) was used before the calculation of means and S.D. IPm1 was identified as an outlier and removed in the calculation for the mean value of IPm alpha diversity. Mean values were 5.75±1.19 (IOf), 5.33±0.72 (IOm), 4.97±1.25 (IPf), 3.11±0.55 (IPm) and 2.14±0.32 (HFf).

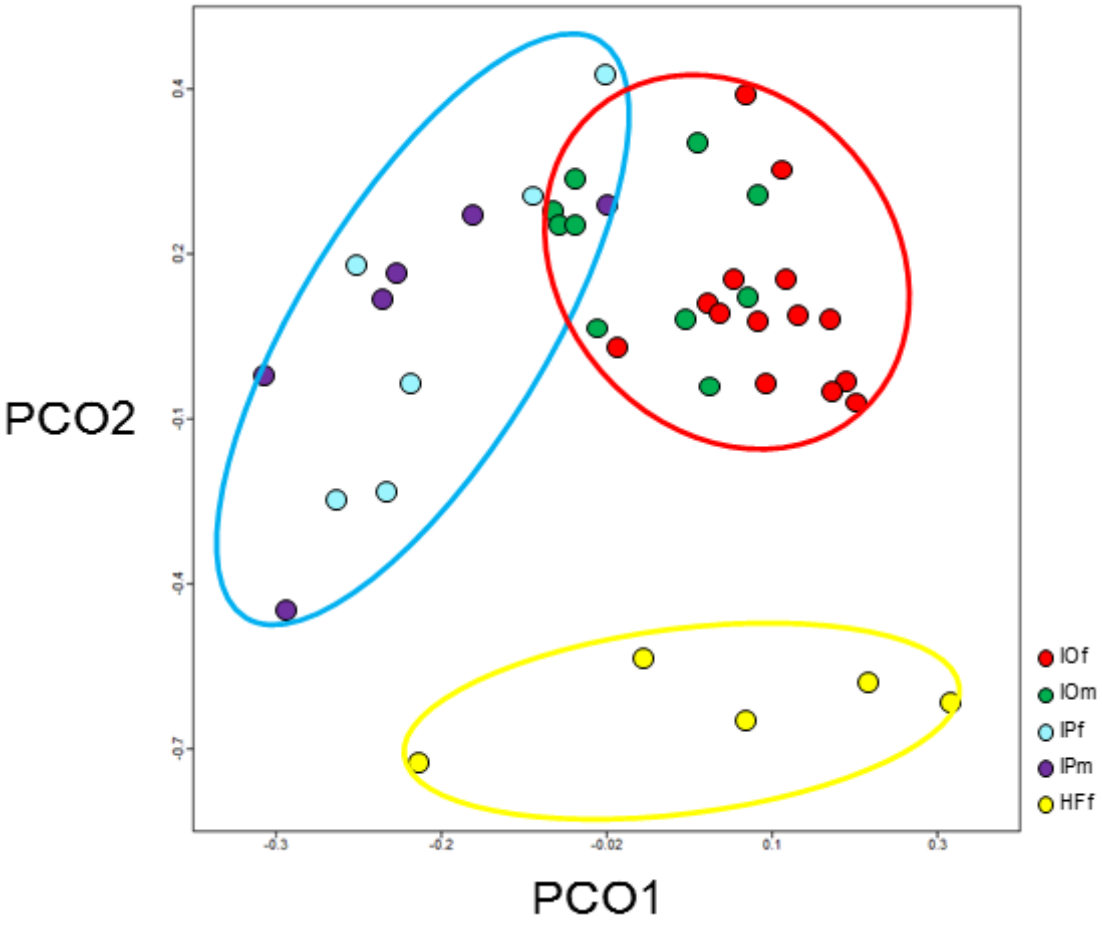

Figure 6. Principal component analysis of the bacterial composition in each tick sample.

The plots were generated using the MG-RAST server. Each tick sample is shown in a different color depending on the species and sex of the tick; IOf, IOm, IPf, IPm, and HFf are represented in red, green, blue, purple, and yellow, respectively. The plots derived from the same tick species are highlighted in circles; *I. ovatus* (IO), *I. persulcatus* (IP), and *H. flava* (HF) are, respectively, highlighted in red, blue, and yellow circles.

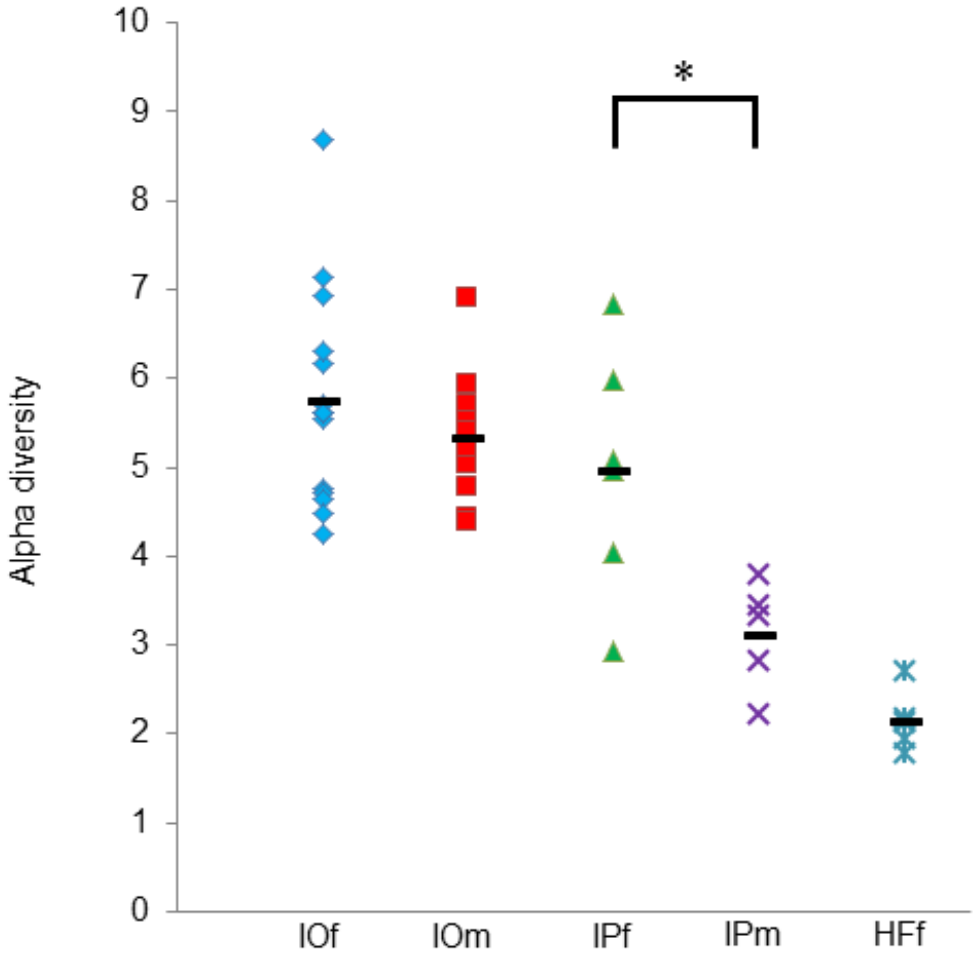

Figure 7. Alpha diversity calculated for each tick sample. The alpha diversity of each tick sample was calculated using the MG-RAST server.

The mean value obtained for each tick group is represented by the horizontal line. Mean alpha diversity values: IOf (5.75), IOm (5.33), IPf (4.97), IPm (3.11), and HFf (2.14).

∗ (p-value < 0.01)

Sequencing of *gltA*

The *gltA* gene was detected in 11 out of 22 samples previously identified as containing the genus *Rickettsia*. Samples that were *gltA*-positive tended to have a greater abundance of rickettsial bacteria than those that were negative (Figure 8). All *gltA*-positive samples were subjected to sequencing analysis. Each *gltA*-positive sample contained only one sequence type, indicating that individual ticks harbored bacteria carrying a single *gltA* allele. From 11 tick samples, three different *gltA* sequences were identified, and BLAST searches showed the highest identities (99.8% to 100%) with *R. asiatica*, *R. helvetica*, and uncultured *Rickettsia* sp. (Table 6).

Sequencing of unclassified bacterial 16S rDNA

PCR products (1,400 bp) were generated using universal primers to resolve the identities of sequence reads detected in four individuals of *I. persulcatus*. Between six and nine clones per sample were classified into Alphaproteobacteria (data not shown) based on BLASTn similarity searches. All of the clones analyzed from two individual ticks were the same sequence type. There were four different sequence types in one individual, and two in another. These showed highest identities (99.5% to 99.7%) with uncultured Rickettsiales previously reported from *I. persulcatus* (GenBank accession number AF497583).

Molecular phylogenetic analysis revealed that the Alphaproteobacteria from four *I. persulcatus* females clustered together within a single clade. This clade contains *Candidatus* Lariskella arthropodarum identified in several stinkbug species (*Arocatus melanostomus*, *Nysius plebeius*, and *Physopelta gutta*) and Rickettsiales derived from flea (*Xenopsylla cheopis*) and ticks (*I. ovatus* and *I. persulcatus*) (Figure 9)

[29,33,68,71].

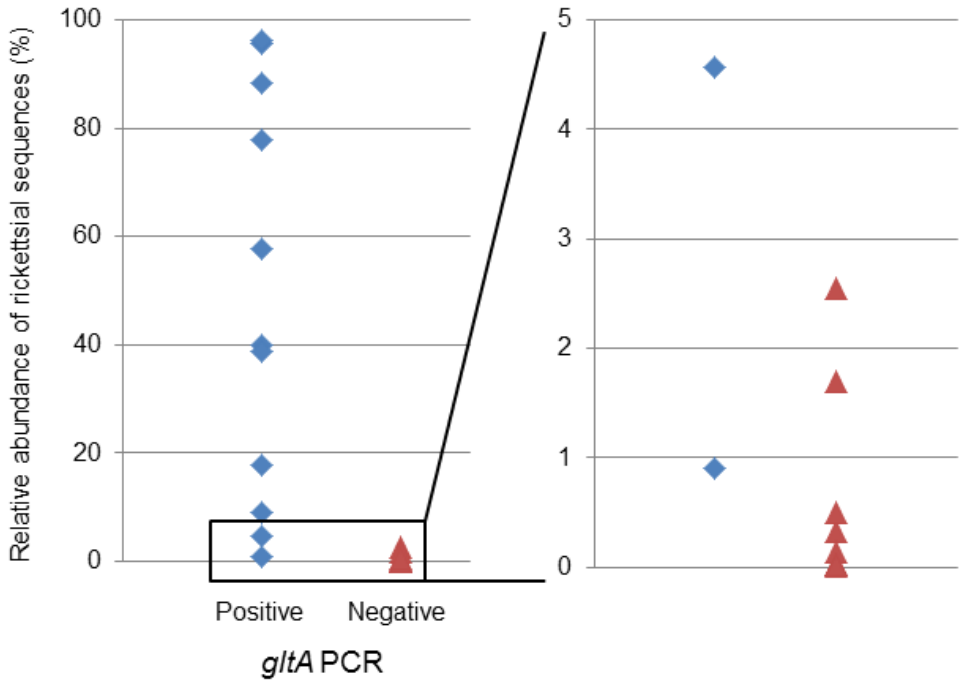

Figure 8. Comparison of the relative abundance of rickettsial sequences estimated by 16S amplicon analysis and the results of *gltA* PCR.

Vertical axis represents the relative abundance of rickettsial sequences calculated from the data obtained from 16S amplicon analysis. Blue dots represent samples in which Rickettsia was detected by both 16S amplicon analysis and *gltA* PCR. Red dots represent samples in which Rickettsia was detected by 16S amplicon analysis but not by *gltA* PCR. The plots with relative abundance values between 0% and 5% are shown in the magnified graph provided in the right column.

Table 6. Summary of *gltA* sequencing.

Sequence ID	Tick sample ID	Identity with reference (no. matched/no. nucleotides)	Reference GenBank no.	Rickettsia species	GenBank no.
gltA_IOf1	IOf1	99.8% (438/439)	AB297808	*R. asiatica*	AB911107
gltA_IOf3	IOf3	99.8% (438/439)	AB297808	*R. asiatica*	AB911107
gltA_IOf4	IOf4	99.8% (438/439)	AB297808	*R. asiatica*	AB911107
gltA_IOm2	IOm2	99.8% (438/439)	AF394901	*R. asiatica*	AB911107
gltA_IOm3	IOm3	99.8% (438/439)	AF394901	*R. asiatica*	AB911107
gltA_IOm4	IOm4	99.8% (438/439)	AF394901	*R. asiatica*	AB911107
gltA_IPf1	IPf1	99.8% (438/439)	U59723	*R. helvetica*	AB911108
gltA_IPf6	IPf6	100% (394/394)	JN849396	Uncultured *Rickettsia* sp.	AB911109
gltA_IPm4	IPm4	99.8% (438/439)	U59723	*R. helvetica*	AB911108
gltA_IPm5	IPm5	99.8% (438/439)	U59723	*R. helvetica*	AB911108
gltA_IPm6	IPm6	99.8% (438/439)	U59723	*R. helvetica*	AB911108

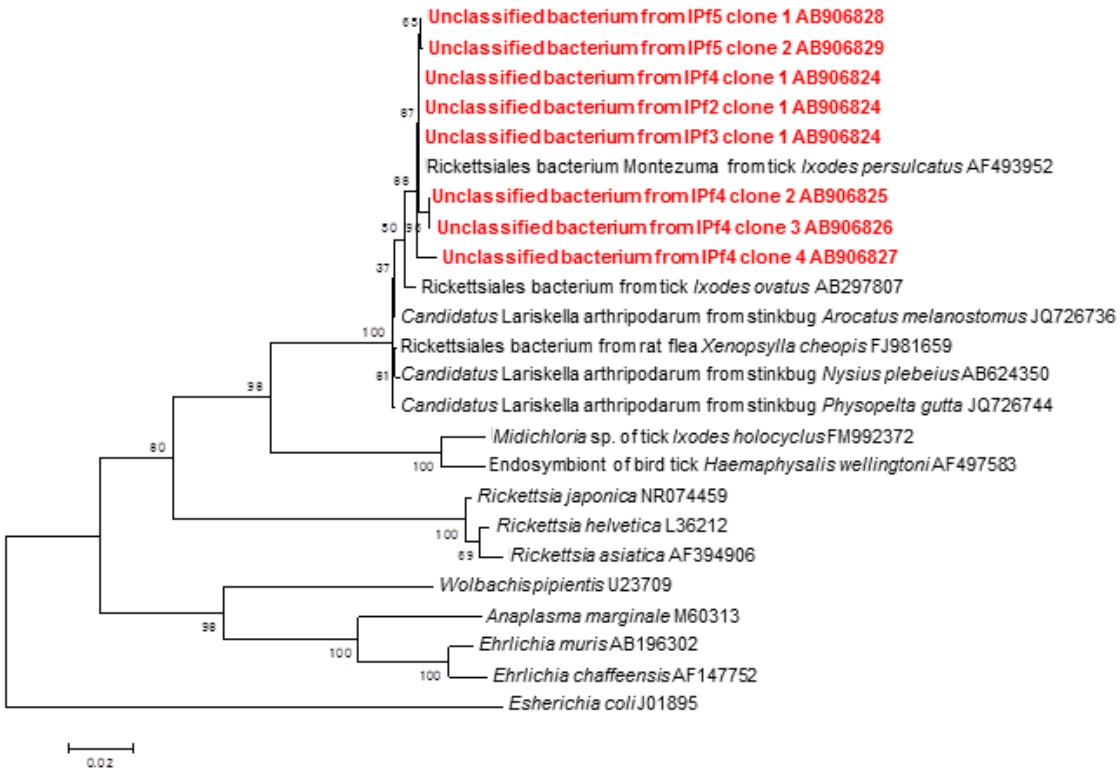

Figure 9. Phylogenetic analysis of the 16S rDNA sequences of unclassified bacteria from IPf2, IPf3, IPf4, and IPf5 using maximum likelihood method.

The tree is rooted with the Escherichia coli. All bootstrap values from 1000 replications are shown on interior branch nodes.

Discussion

The aim of this study was to assess and compare the diversity of bacterial populations within the salivary glands of *I. ovatus*, *I. persulcatus*, and *H. flava*. This metagenomic approach revealed bacterial populations totaling 163 different genera found in tick salivary glands (Figure 5). These included the genera of tick-borne pathogens such as *Ehrlichia* and *Rickettsia*. Further identification using species-specific PCR would be needed to clarify the presence of the tick-borne bacterial pathogens, such as *E. muris*, *E. chaffeensis*, *R. japonica* and *R. helvetica*, in the ticks used in this study [31,53,64,96]. This combination of detection approaches may be useful for the screening and detection of possible pathogens in arthropod vectors.

Rickettsia was detected in 22 of the 41 (53.6%) samples by 16S rDNA amplicon pyrosequencing; however, only half of the 22 positive samples were positive with *gltA* PCR. This may be attributed to the relative amounts of rickettsial DNA in the PCR templates, where *gltA* PCR-positive samples tended to contain a higher proportion of rickettsial DNA than those that were negative (Figure 8). However, there were two *gltA*-negative samples (IPm1 and IPf4) that had higher proportions of rickettsial DNA than a *gltA*-positive sample (IOf4). This result suggests that the sensitivity of conventional *gltA* PCR may be affected by the other factors such as the resolving power of agarose gel electrophoresis and the presence of PCR inhibitory components in samples [74,82]. Thus, a 16S rDNA amplicon pyrosequencing approach is a more sensitive method to detect specific pathogens.

Analysis of the *gltA* gene sequences from *I. ovatus* and *I. persulcatus* revealed that they belonged to *R. asiatica* and *R. helvetica*, respectively (Table 6). This result

agrees with previous findings on the potential of the ticks to act as vectors for these rickettsia in Japan [32]. *R. helvetica* belongs to the spotted fever group of rickettsia and is a causative agent of febrile illness. A human case associated with this pathogen has been reported elsewhere [48,77]. There was a high abundance (>70%) of this rickettsial species in some *I. persulcatus* samples (Figure 4B), suggesting that it is well adapted to the salivary glands of ticks, and waiting for the transmission to mammalian hosts. In addition to pathogenic strains, the genus *Rickettsia* also contains symbionts associated with ticks. *Rickettsia*-like symbionts can influence the tick physiology, population dynamics, and the transmission of other pathogenic *Rickettsia* spp [81,97].

Coxiella burnetii and *Coxiella*-like endosymbionts have been identified in several tick genera, including *Dermacentor*, *Ixodes*, *Haemaphysalis* and *Rhipicephalus* [9,13,22,60]. *Coxiella*-like endosymbionts have been located at high densities in the salivary glands of the lone star tick (*Amblyomma americanum*) using fluorescence *in situ* hybridization [56]. The findings in this study also highlighted the presence of *Coxiella* in the salivary glands of three species of tick. The dominant presence of *Coxiella* in the salivary glands of ticks warrants further investigation to resolve their potential roles in tick biology, particularly blood-sucking behavior, and their interaction with other microbes. The genus *Spiroplasma* contains a wide diversity of often unnamed or poorly characterized species, including non-pathogenic, symbiotic, and pathogenic organisms associated with a wide variety of arthropods. Symbiotic *Spiroplasma* has a close association with, and can affect the behavior of, their host arthropods. For example, Hurst *et al.* (2000) reported the preferential killing of males by *Spiroplasma*; when female insects (e.g., the butterfly *Danaus chrysippus*) are infected, the broods are female-biased because the infected male progeny die during embryogenesis [47]. One

Spiroplasma sp. has been reported in ticks, and it has also been associated with transmissible spongiform encephalopathy in humans and ruminants, although its role in the pathology of the host has not been clarified [5,46,104]. In this study, *Spiroplasma* was detected in *Ixodes* ticks, and not in *H. flava* (Figure 4). Previous research reported the genera *Spiroplasma* and the closely related *Mycoplasma* in several tick species in Japan [101]. The pathogenicity of *Spiroplasma* harbored in ticks in Japan is not known yet.

Results from the PCA of sequences indicated that microbial population structures in the salivary glands of ticks were different, and that samples from the same species of tick clustered together (Figure 6). Ticks can acquire microorganisms through a variety of ways, such as transovarial transmission, and from the environment, host animals during blood feeding, and mating partners. For microorganisms to exist in the salivary glands, they need to migrate from the midgut and enter the glands. The establishment of microorganisms within ticks can depend on the interactions between particular microbes, ticks and other symbioses [63,81]. The differences in the microbial populations within the salivary glands of tick species in this study were attributed to these complicated factors.

Previous studies revealed that tick microbial populations were different between developmental stages (egg, nymph, and adult) [3,72]. The bacterial compositions also differed between organs, such as between midgut and ovary [3]. Some bacterial species, for instance *Borrelia burgdorferi* that is a causative agent of Lyme disease, exist in the midgut of the tick, moving into the salivary glands when stimulated by feeding on blood [7,58]. For better understanding of microbial interactions with ticks as well as the potential pathogens transmitted by ticks, further

study should include the comparison of the microbes between salivary glands and other organs.

The analysis of the dynamics of microbial community composition during the process of feeding on blood may also uncover the roles of tick microbes. The mean alpha diversity value (Figure 6) was greater for the female *I. ovatus* (5.61) than that of male (5.31). This rank order was also recorded for female (5.02) and male (3.38) *I. persulcatus* ticks. Moreover, the diversity of *I. persulcatus* female was significantly higher than that in the male (p-value = 0.01). This rank order may imply that some bacterial species preferentially select the gender of ticks. There may be some strategic biological relevance in the transmission of bacteria to mammalian hosts because female ticks feed for a longer period of time than males.

The total number of bacterial genera (Table 3) detected in *I. persulcatus* (127) was greater than in *I. ovatus* (71). This suggested that *I. persulcatus* can harbor and transmit a wide range of bacteria than *I. ovatus.* Eighteen bacterial genera were commonly detected in all tick species, which indicate their strong biological relationships with the tick host, and essential roles in tick physiology.

Several *I. persulcatus* females contained unclassified bacteria belonging to the Proteobacteria and Alphaproteobacteria (Figure 4B). Based on the analysis of the nearly complete 16S rDNA sequences, the unclassified bacterial were classified into a single phylogenetic clade, which was recently proposed as a ‘‘*Candidatus* L. arthropodarum’’ clade [68]. This clade also includes Rickettsiales bacterium previously found in blood and biopsy samples of the patients with an acute fever disease, etiologically linked with tick bites [71]. The relationships between these microorganisms and their arthropod hosts are not clear, and their potential to act as causative agents of emerging tick-borne

mammalian diseases warrants further investigation.

Summary

Ticks are one of the most important blood-sucking vectors for infectious microorganisms in humans and animals. When feeding, they inject saliva, containing microbes, into the host to facilitate the uptake of blood. An understanding of the microbial populations within their salivary glands would provide a valuable insight when evaluating the vectorial capacity of ticks. Three tick species (*Ixodes ovatus*, *I. persulcatus* and *Haemaphysalis flava*) were collected in Shizuoka Prefecture of Japan between 2008 and 2011. Each tick was dissected and the salivary glands were collected. Bacterial communities in each salivary gland were characterized by 16S amplicon pyrosequencing using a 454 GS-Junior Next Generation Sequencer. The Ribosomal Database Project (RDP) Classifier was used to classify sequence reads at the genus level. The composition of the microbial populations of each tick species were assessed by principal component analysis (PCA) using the Metagenomics RAST (MG-RAST) metagenomic analysis tool. *Rickettsia*-specific PCR was used for the characterization of rickettsial species. Almost full length of 16S rDNA was amplified in order to characterize unclassified bacterial sequences obtained in *I. persulcatus* female samples. The numbers of bacterial genera identified for the tick species were 71 (*I. ovatus*), 127 (*I. persulcatus*) and 59 (*H. flava*). Eighteen bacterial genera were commonly detected in all tick species. The predominant bacterial genus observed in all tick species was *Coxiella*. *Spiroplasma* was detected in *Ixodes*, but not in *H. flava*. PCA revealed that microbial populations in tick salivary glands were different between tick species, indicating that host specificities may play an important role in determining the microbial complement. Four female *I. persulcatus* samples contained a high abundance

of several sequences belonging to Alphaproteobacteria symbionts. This study revealed the microbial populations within the salivary glands of three species of ticks, and the results will contribute to the knowledge and prediction of emerging tick-borne diseases.

Chapter III

Exploring the diversity of viruses in ticks (*Ixodes persulcatus*) using a high throughput sequencing technique

Introduction

In chapter II, bacterial communities were analyzed in tick salivary glands and revealed to consist of a variety of bacterial genera. However, ticks harbor not only bacteria but also viruses, including those pathogenic to higher animals. Thus, viral population analysis in ticks is also required to predict and preempt tick-borne emerging viral diseases.

To exploit viral populations or "viromes", a catch-all approach to detect and identify a wide range of viruses is required. For this purpose, recent studies have employed effective viral particle enrichment methods together with high throughput shotgun sequencing protocols to analyze DNA or cDNA in virus-enriched fractions. Using this combination of techniques, novel viruses have been found from a wide variety of biological and environmental samples including water, mosquito, and white fly [75,76,112]. Conventionally, these studies employ BLAST analysis to search for homologous sequences in viral genomes databases. However, a major disadvantage of this approach is that only a limited proportion of the sequence reads can be assigned to those of known viruses; significantly similar sequences to those used in the queries are quite often not found. Thus, it is highly possible that yet-unknown organisms cannot be identified by this BLAST-based search.

Batch Learning Self-Organizing Map (BLSOM) is a bioinformatics tool that

makes the learning process and resulting map independent of the order of data input [1,52]. This algorithm recognizes taxon-specific profiles of oligonucleotide frequencies and permits taxonomic clustering of genomic DNA fragments according to species without the need for species information. Since BLSOM does not require orthologous sequence data sets and sequence alignments, it is useful for the phylogenetic estimation of novel genome sequences in metagenomic libraries containing sequences from a wide variety of uncultured microorganisms [2,105]. In a previous study, this approach was successfully applied to analyze the bacterial flora of ticks, which resulted in the identification of over a hundred different genera, including novel Chlamydiae, that had not previously been found in ticks [73]. In addition, phylotype-specific classification methods theoretically similar to BLSOM were applied to metagenomic studies [19,24,67,109].

The aim of this chapter is to show that the virome, including pathogens in *Ixodes persulcatus* ticks, can be determined using a combination of a high throughput sequencing method, an effective viral particle purification method, and a bioinformatics tool based on BLSOM. This tick species is a vector of tick-borne encephalitis (TBE) virus in the Far East [44]. In Hokkaido, one human case of TBE has been reported [98], though, the responsible vector of TBE in that case was supposed to be *I. ovatus* [99]. Since no other tick-borne viral diseases have been reported in Hokkaido, the potency of *I. persulcatus* as a vector of viral pathogens remains to be established.

Materials and Methods

Sample collection

Adult *I. persulcatus* ticks were collected in the Hidaka region (42.97 N 142.68 E) in Hokkaido, Japan by a flagging method. Live adult ticks were separated according to sex and stored in an incubator at 4°C with over 80% humidity until used in the next preparation step.

Preparation of viral particle-enriched fractions from ticks

Stored ticks were washed with 70% of ethanol and rinsed with SM buffer with gelatin (1L of SM buffer containing 5.8g NaCl, 1.2g $MgSO_4$, 50 ml of 1 M Tris-HCl pH 7.5, and 0.1g gelatin) several times. Homogenates were prepared from 10 whole ticks by placing them in a 2 ml microtube containing two stainless balls (Tomy, Tokyo, Japan, Cat No. SUB-50) with 100 μl of SM buffer and then shaking the tubes in a beads homogenizer (Tomy) at 3,000 rpm for 30 seconds. After homogenization, 300 μl of SM buffer was added into each tube and the homogenates were remixed with fresh SM buffer. To remove tick debris and intact cells, the homogenates were centrifuged at 5,800 *g* for 30 minutes. Supernatants obtained from 10 tubes of the same gender were pooled and filtered through 0.45 and 0.22 μm pore-size polyether membranes (Whatman, UK Cat No. SLHV033RB and SLGVJ13SL). Small particles in the filtrates were concentrated with a tangential flow filtration cassette with a 30 kDa nominal molecular weight cut-off (NMWCO) regenerated cellulose membrane (Millipore, USA, Cat No. UFC803096). The virus-enriched fraction was recovered from the filter-retained part and resuspended in DNase buffer. Then the virus-enriched fraction

was treated with 2.5 U/μl DNase 1 (New England Biolabs, USA, Cat No. M0303L) and 2 U/μl RNase I_f (New England Biolabs, USA, Cat No. MM0243S) at 37°C for 60 min. After enzyme inactivation at 75°C for 20 min, viral nucleic acids were extracted using NucleoSpin RNA XS (Takara, Tokyo, Japan) according to the manufacturer's instructions.

Reverse transcription and amplification

Single-stranded DNA synthesis was performed on the extracted viral nucleic acids with SuperScript III reverse transcriptase (Invitrogen, USA) using previous published methods [16,49,50]. The random primer A (GTTTCCCAGTCACGATCNNNNNNNNN) used consisted of two parts; GTTTCCCAGTCACGATC corresponding to the sequence of primer B used in the following step, and randomly arranged 9 nucleotides (NNNNNNNNN). 100 pmol of random primer A was heated at 65^{o}C for 5 min and cooled on ice for 2min to denature secondary structure, and then the following components were added: 4 μl of 5× first strand buffer, 1 μl of 10mM dNTPs mix, 40 units of RNase OUT, 1 μl of 0.1M dithiothreitol (DTT), and 200 units of SuperScript III reverse transcriptase (Invitrogen, USA). The reaction mixture was incubated at 25°C for 5 min. 50°C for 60 min, and then at 75°C for 15min to inactivate the transcriptase. To synthesize double-strand DNA, 2.5 units of Klenow Fragment (3'-5' exo-) (New England Biolabs, Beijing, China) was added to the cDNA mixture containing random primer A. After incubation at 37°C for 60 min, the enzyme was inactivated at 75°C for 10 min.

The synthesized double-strand DNA was amplified by employing sequence-independent single-primer amplification (SISPA) methods established in a

previous study [88]. PCR of the first step products was performed using 20 μl of the reaction described above in a total volume of 50 μl containing 5 μl of 10× EX taq buffer, 4 μl of 10mM dNTPs, 5 units of EX taq, and 500 pmol of primer B (GTTTCCCAGTCACGATC). The reaction mixture was incubated under the following conditions: 40 cycles of 94°C for 21 sec, 40°C for 30 sec, 50°C for 30 sec, and 72°C for 1 min. Amplified products were purified using Wizard® SV Gel and PCR Clean-Up System (Promega Corporation, Madison, WI, USA) The concentrations and quality of the amplified products were assessed on an Agilent 2100 Bioanalyzer using a DNA1000 lab chip (Agilient).

Pyrosequencing and data analysis

Sequencing of the amplified products was performed on a 454 pyrosequencing Genome Sequencer Junior (GS Junior) (Roche, Basel, Switzerland) according to the manufacturer's protocol. The raw sequencing data file in standard fogram format (.sff) was converted into a FASTA file, the primer sequence was trimmed, and low quality and short (< 150 bp) reads were removed using CLC Genomics Workbench version 7.5.1 (Qiagen Inc., Valencia, CA). *De novo* assembly was also performed using this software. The contigs were subjected to a homology search using BLASTn with the GenBank nt database, and a phylogenic tree was constructed with Mega 6.05 software. Furthermore, the contigs consisting of more than 300 bases were used to identify viruses using BLSOM analysis.

BLSOM analysis

Self-Organizing Map (SOM) is a neural network algorithm based on

unsupervised learning that carries out a characteristic nonlinear projection from the high-dimensional space of input data onto a two-dimensional array of weight vectors [57]. Abe *et al.* (2003) modified conventional SOM for genome informatics to make the learning process and resulting map independent of the order of data input by employing the Batch Learning SOM, "BLSOM" [1]. Instead of random values, they defined the initial weight vectors by Principal Component Analysis (PCA). BLSOM learning was conducted as described previously [1].

In advance, two types of large-scale BLSOMs, namely Kingdom- and Virus group-BLSOM, were constructed to identify viruses from metagenomic sequences using sequences deposited in DDBJ/EMBL/GenBank as previously described [1]. Kingdom-BLSOM was constructed with tetranucleotide frequencies in all 5-kb sequences derived from the whole-genome sequences of 111 eukaryotes, 2,813 prokaryotes, 1,728 mitochondria, 110 chloroplasts, and 31,486 viruses. Virus group-BLSOM was constructed with a total of 602,951 1-kb sequences from 97 families.

After *de novo* assembly, contigs longer than 300 bp were mapped using Virus group-BLSOM. The mapping was conducted by finding the lattice point with the minimum Euclidean distance in the multidimensional space and was assigned to Virus group-BLSOM on the basis of statistical tests. To identify the contigs that could not be assigned using Virus group-BLSOM, they were mapped on Kingdom-BLSOM.

To investigate the accuracy of BLSOM analysis, three datasets (A, B, and C) were prepared from the viral sequences deposited in GenBank. The datasets A, B, and C contained BLAST-identified viral sequences with lengths ranging between 300-1,000 bp, 500-1,000 bp, and 750-1,000 bp, respectively. When the dataset A was tested,

approximately 70% of the fragments were correctly classified into the viruses by Kingdom-BLSOM at kingdom level, while the remaining sequences were assigned into either eukaryotes or prokaryotes. Furthermore, about 80% of these viral sequences were assigned to the corresponding taxa at the family level with accuracy (Table 7).

Table 7. Estimation accuracy of BLSOM-based classification using deposited viral sequence data.

Dataset	Range of sequence lengths (bp)	No. of reads	Virus detection rate (%) at Kingdom level	Coincidence ratio (%) of viral families
A	300-1,000	709,987	66.39	79.89
B	500-1,000	539,996	93.31	79.90
C	750-1,000	201,636	95.87	81.82

Estimation accuracies of BLSOM-based classification were calculated using all deposited sequences which have a sequence length in three ranges.

Results

Results of pyrosequencing

A total of 133,932 and 175,545 sequence reads were obtained from female and male samples, respectively. After trimming tag sequences and removing short (< 150 bp) and low quality reads, 100,634 and 156,837 sequence reads with average lengths of 351 bp and 429 bp from female and male samples, respectively, were finally used for further analyses.

De novo assembly and classification

Sequence reads were assembled by using CLC Genomic Workbench. After *de novo* assembly, 577 and 386 contigs were obtained from female and male samples, respectively. The longest contigs were 4,291 and 8,972 bp in length from female and male samples, respectively. Taxonomical classifications of contigs were performed with BLASTn analysis (cut-off e-value < 10^{-5}). In the female samples, 3.5% (20/577) of contigs were assigned to viruses but 56% of contigs could not be classified (Figure 10). In the male sample, 6.7% (26/386) of contigs were assigned to viruses while 50% remained unclassified (Figure 10). The remaining sequences were thought to be derived either from eukaryotes, most probably the host tick, or from bacteria. At lower taxonomic levels, these viruses were classified as members of the order *Mononegavirales*, or members of the families *Bunyaviridae* and *Rhabdviridae* (Figure 11).

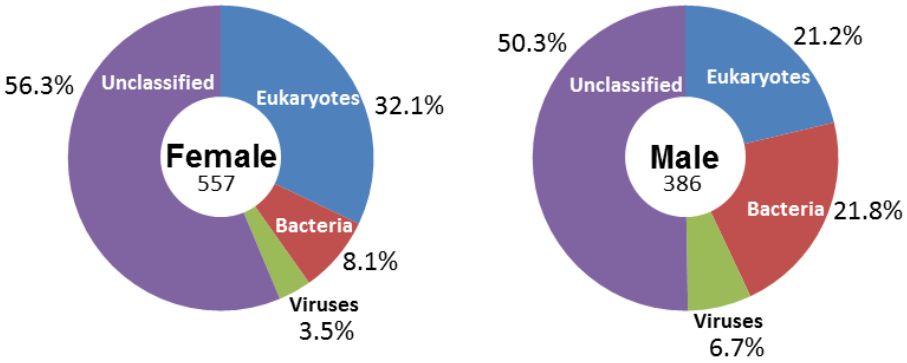

Figure 10. Kingdom classification of the contigs from female and male samples using BLASTn analysis.

Contigs were classified into eukaryotes, bacteria, and viruses using BLASTn analysis.

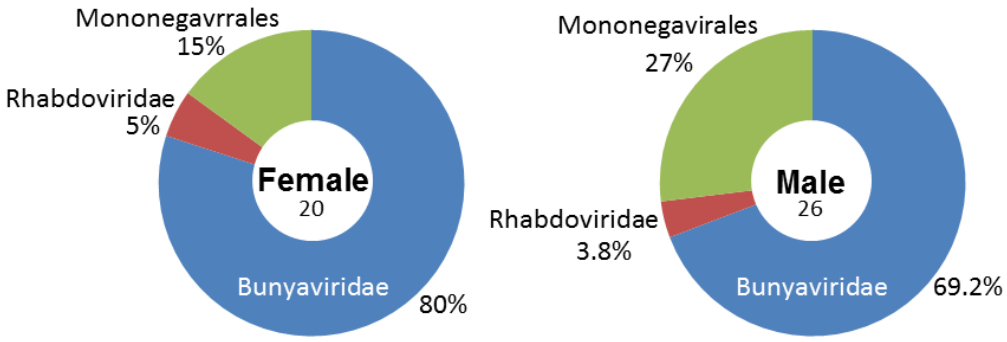

Figure 11. Order and family classification of the viral contigs from female and male samples using BLASTn analysis.

The contigs assigned to viruses by BLASTn analysis are listed in Table 8A and 8B. Some contigs were identified as a part of the RNA-dependent RNA polymerase gene of the family *Bunyaviridae* (Tables 8A and 8B). In addition, some contigs showed homology to the sequences of South Bay virus. The longest L segment-like contigs (IPf_95 and IPm_68) from both samples were clustered with sequences from viruses belonging to the genus *Nairovirus* in the phylogenetic tree (Figure 12). The similar sequences of South Bay virus S segments were found in both samples. The longest S segment-like contigs in both samples (IPf_11 and IPm_57) located in a clade of the genus *Nairovirus* in the phylogenetic tree (Figure 13). The contigs similar to Blacklegged tick *Phlebovirus* and Deer tick Mononegavirales-like virus were also detected in both male and female samples (Table 8A and 8B).

Female and male samples yielded 507 and 324 contigs, respectively, with a length of over 300 bp. BLSOM analysis of these contigs showed that over half of the contigs mapped to viruses (Figure 14). Only 2.0% (10/507) and 1.5% (5/324) from female and male samples, respectively, could not be assigned to any organisms (Figure 14). The viral contigs found from ticks were assigned to several viral groups (Figure 15). Double strand (ds) DNA viruses were occupied nearly 50% of viral contigs from female and male samples (Figure 15). Female and male ticks carried 37 and 30 different viruses at the family level, and the total viral taxa detected by this analysis covered 43 families (Table 9). These families were containing viruses infecting vertebrates, insects, plants and bacteriophages. Over 10 contigs were assigned each family of *Bunyaviridae*, *Herpesviridae*, *Siphoviridae*, and *Myoviridae* (Table 9). Among the contigs assigned to viruses by BLASTn analysis, 61% (11/18) and 56% (14/25) of them, respectively, from females and males were also classified to viruses by BLSOM analysis. In addition,

54.5% (6/11) and 42.8% (6/14) of viral contigs in females and males, respectively, were assigned to the same viral families.

Table 8A. Contigs from female sample assigned to viruses by BLASTn analysis.

Contig ID	Length (bp)	Total reads	BLASTn	Identity (%)	Mach with BLSOM
IPf_54	219	66	South Bay virus isolate H38 segment S, complete sequence	81	
IPf_18	246	408	South Bay virus isolate H38 segment S, complete sequence	80	
IPf_230	559	55	South Bay virus isolate H38 segment L, complete sequence	77	**
IPf_4	534	6,827	South Bay virus isolate H38 segment S, complete sequence	80	*
IPf_47	718	165	South Bay virus isolate H38 segment L, complete sequence	76	*
IPf_48	718	90	South Bay virus isolate H38 segment L, complete sequence	76	*
IPf_487	760	6	Blacklegged tick phlebovirus-2 isolate H5 segment L, complete sequence	76	
IPf_11	892	3,093	South Bay virus isolate H38 segment S, complete sequence	72	
IPf_223	1,073	34	Dugbe virus L protein gene, complete cds	78	*
IPf_127	1,331	313	South Bay virus isolate H38 segment L, complete sequence	78	*
IPf_322	1,492	40	Blacklegged tick phlebovirus-2 isolate H5 segment L, complete sequence	69	**
IPf_10	1,628	511	South Bay virus isolate H38 segment L, complete sequence	70	**
IPf_248	1,770	76	Crimean-Congo hemorrhagic fever virus isolate SPU 48/90 segment L, complete sequence	77	**
IPf_19	2,038	226	Deer tick mononegavirales-like virus isolate FI3 polymerase gene, complete cds	75	
IPf_86	3,079	917	South Bay virus isolate H38 segment L, complete sequence	76	**
IPf_125	3,576	408	Deer tick mononegavirales-like virus isolate FI3 polymerase gene, complete cds	91	
IPf_169	3,619	462	Crimean-Congo hemorrhagic fever virus strain TADJ/HU8966 segment L, complete sequence	81	**
IPf_95	4,291	1,419	South Bay virus isolate H38 segment L, complete sequence	75	

** Contig correctly assigned to the same viral families by BLSOM.

* Contig assigned to viruses by BLSOM at the kingdom level.

Table 8B. Contigs from male sample assigned to viruses by BLASTn analysis.

Contig ID	Length (bp)	Total reads	BLASTn	Identity (%)	Mach with BLSOM
IPm_362	299	2	*Ixodes scapularis* associated virus 1 isolate K13, partial genome	75	
IPm_19	347	154	South Bay virus isolate H38 segment S, complete sequence	81	
IPm_20	347	126	South Bay virus isolate H38 segment S, complete sequence	84	
IPm_122	442	180	South Bay virus isolate H38 segment L, complete sequence	86	**
IPm_123	442	161	South Bay virus isolate H38 segment L, complete sequence	73	**
IPm_207	450	4	Blacklegged tick phlebovirus-2 isolate RTS2 segment L, complete sequence	80	*
IPm_135	502	9	Deer tick mononegavirales-like virus isolate FI3 polymerase gene, complete cds	79	
IPm_176	535	20	Deer tick mononegavirales-like virus isolate DTM1 polymerase gene, complete cds	86	*
IPm_8	553	1,827	South Bay virus isolate H38 segment S, complete sequence	76	*
IPm_54	687	8,909	South Bay virus isolate H38 segment S, complete sequence	71	*
IPm_134	742	217	Deer tick mononegavirales-like virus isolate FI3 polymerase gene, complete cds	72	
IPm_21	743	3,055	South Bay virus isolate H38 segment S, complete sequence	80	**
IPm_286	762	16	Blacklegged tick phlebovirus-2 isolate RTS2 segment L, complete sequence	86	
IPm_2	892	9,761	South Bay virus isolate H38 segment S, complete sequence	71	*
IPm_178	956	93	Blacklegged tick phlebovirus-2 isolate H5 segment L, complete sequence	73	**
IPm_223	999	60	Crimean-Congo hemorrhagic fever virus strain AP92 segment L, complete sequence	76	
IPm_45	1,034	1,773	South Bay virus isolate H38 segment L, complete sequence	81	*
IPm_173	1,092	58	Deer tick mononegavirales-like virus isolate FI3 polymerase gene, complete cds	73	
IPm_235	1,177	18	Issyk-Kul virus strain LEIV-315K segment L, complete sequence	72	
IPm_57	1,190	2,976	South Bay virus isolate H38 segment S, complete sequence	71	**
IPm_161	1,696	244	Crimean-Congo hemorrhagic fever virus isolate SPU 383/87 segment L, complete sequence	75	
IPm_143	1,873	3,171	Deer tick mononegavirales-like virus isolate FI3 polymerase gene, complete cds	76	
IPm_170	2,083	187	*Ixodes scapularis* associated virus 1 isolate K13, partial genome	76	*
IPm_141	2,375	260	*Ixodes scapularis* associated virus 2 isolate A1, partial genome	72	*
IPm_111	4,206	1,304	Deer tick mononegavirales-like virus isolate FI3 polymerase gene, complete cds	71	
IPm_68	8,972	5,565	South Bay virus isolate H38 segment L, complete sequence	71	**

** Contig correctly assigned to the same viral families by BLSOM.

* Contig assigned to viruses by BLSOM at the kingdom level.

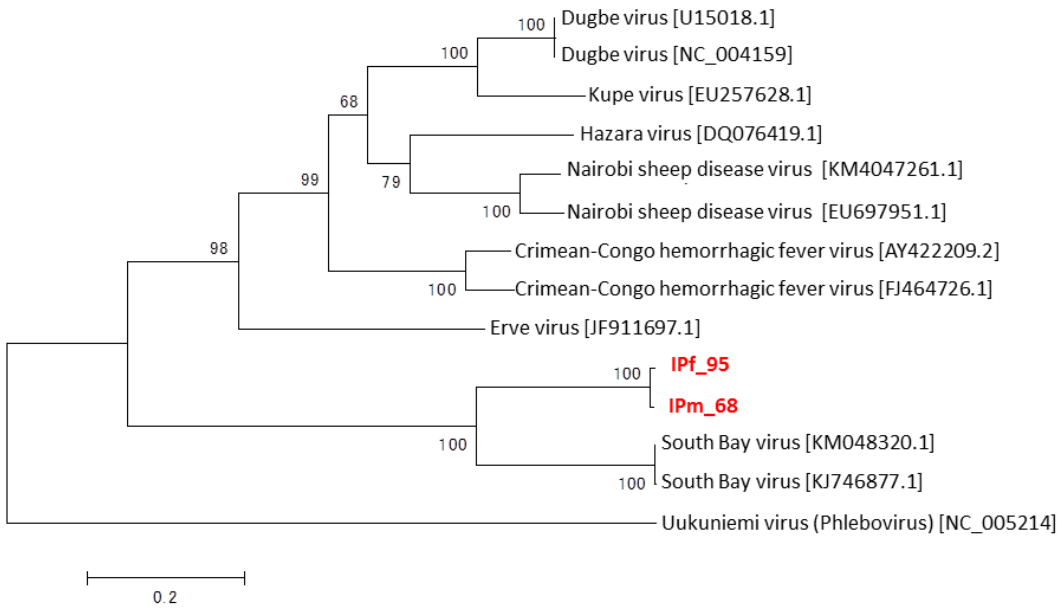

Figure 12. Maximum-likelihood phylogenetic tree based on the nucleotide sequences of the longest contigs mapped to *Bunyaviridae* L segment.

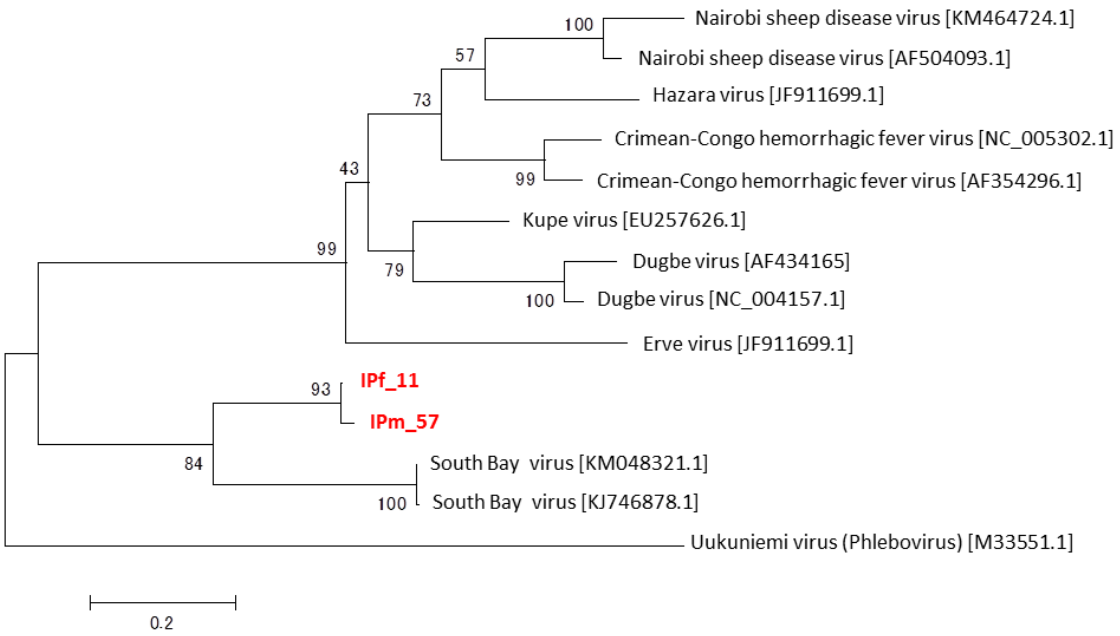

Figure 13. Maximum-likelihood phylogenetic tree based on the nucleotide sequence of the longest contigs mapped to *Bunyaviridae* S segment.

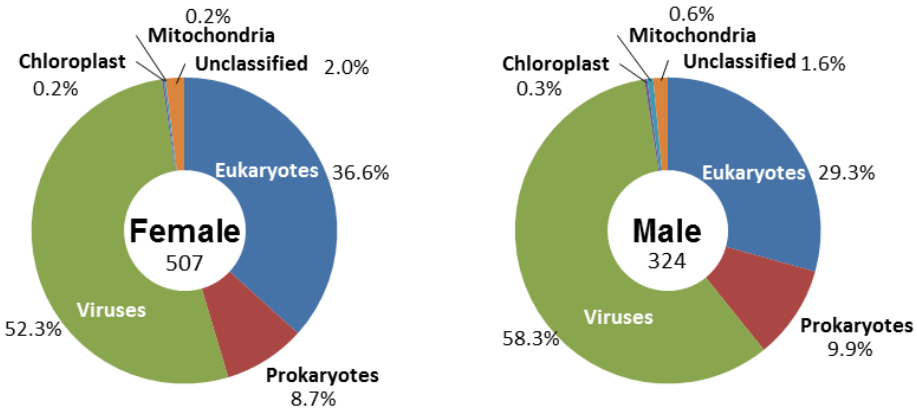

Figure 14. Results of Kingdom-BLSOM of the contigs from female and male samples.

Only contigs with a length of over 300 bp were used.

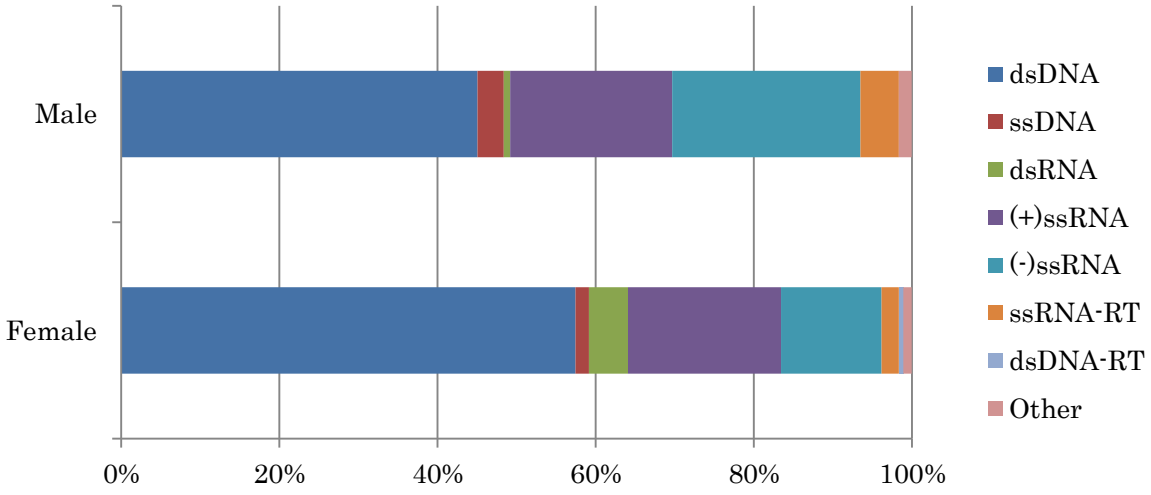

Figure 15. Virus group of the contigs from female and male samples.

Contigs classified by using Virus group-BLSOM.

Table 9. Results of BLSOM of the contigs from female and male samples.

Viral group	Family	No. of contigs		Total reads	
		Female	Male	Female	Male
dsDNA	*Adenoviridae*	11	2	6,086	195
dsDNA	*Alloherpesviridae*	3	2	68	359
dsDNA	*Ascoviridae*	1		11	
dsDNA	*Asfarviridae*	2		10	
dsDNA	*Baculoviridae*	6	2	81	44
dsDNA	*Herpesviridae*	30	13	589	1,572
dsDNA	*Iridoviridae*	2	3	6	13
dsDNA	*Mimiviridae*	1		7	
dsDNA	*Myoviridae*	16	14	909	2,880
dsDNA	*Papillomaviridae*	3	2	2,013	30
dsDNA	*Phycodnaviridae*	3	2	782	168
dsDNA	*Podoviridae*	5	4	866	316
dsDNA	*Polydnaviridae*	1		381	
dsDNA	*Polyomaviridae*	2			
dsDNA	*Poxviridae*	2		45	
dsDNA	*Siphoviridae*	16	11	402	574
ssDNA	*Anelloviridae*		2		10,736
ssDNA	*Circoviridae*		1		169
ssDNA	*Geminiviridae*	2	1	13,317	20
ssDNA	*Parvoviridae*	1		29	
dsRNA	*Birnaviridae*	3		214	
dsRNA	*Reoviridae*	6	1	4,875	15
(+)ssRNA	*Alphaflexiviridae*	2		140	
(+)ssRNA	*Arteriviridae*		1		7
(+)ssRNA	*Betaflexiviridae*	4	1	61	437
(+)ssRNA	*Bromoviridae*	2	1	16	31
(+)ssRNA	*Closteroviridae*	2	5	824	1,041
(+)ssRNA	*Coronaviridae*	6	1	78	16
(+)ssRNA	*Flaviviridae*	5	5	8,999	1,915
(+)ssRNA	*Hepeviridae*	1		19	
(+)ssRNA	*Luteoviridae*	1		4	
(+)ssRNA	*Picornaviridae*	5	7	218	321
(+)ssRNA	*Potyviridae*	6		160	
(+)ssRNA	*Secoviridae*		2		2,080
(+)ssRNA	*Togaviridae*	1	2	4	219
(-)ssRNA	*Arenaviridae*		1		1,773
(-)ssRNA	*Bornaviridae*		1		43
(-)ssRNA	*Bunyaviridae*	14	13	4,257	13,969
(-)ssRNA	*Filoviridae*	1	1	2	24
(-)ssRNA	*Orthomyxoviridae*	4	8	215	32,658
(-)ssRNA	*Paramyxoviridae*	4	5	541	860
ssRNA-RT	*Retroviridae*	4	6	521	9,989
dsDNA-RT	*Caulimoviridae*	1		31	
-	Other	2	2	354	83

Discussions

In this study, tick viromes are constituted of a variety of viral families including those containing human and animal pathogens, such as members of the families *Bunyaviridae*, *Flaviviridae*, *Reoviridae*, and *Orthomyxoviridae*. It is interesting to understand the relationships between viruses found in ticks and known pathogenic viruses from the viewpoint of viral evolution. It is also urgent to analyze whether ticks can transmit those yet-unknown viruses to animals and cause emerging diseases.

In BLAST analysis, some contigs were associated with South Bay virus (Figures 12 and 13). This virus was originally detected in *I. scapularis* using a high throughput sequencing technique [103], but its pathogenic potential in animals is still unknown. Contigs similar to *I. scapularis*-associated virus, Deer tick Mononega-like virus, and Blacklegged tick Phlebovirus were also identified (Tables 8A and 8B). Though these viruses were firstly detected in *I. scapularis* collected in New York [103], the present study demonstrated the presence of similar viruses in tick populations in Japan. Further studies are warranted to investigate the pathogenic potential of these newly identified viruses. The results obtained by BLAST analysis indicate that metagenomics approach coupled with viral purification steps is robust for the detection and characterization of a wide range of viruses, especially previously unknown viruses from arthropods.

In BLSOM analysis, sequences related to several insect virus families were identified; that is, *Ascoviridae*, *Baculoviridae*, *Bornaviridae*, *Closteroviridae*, *Iridoviridae*, *Polydnaviridae* and *Poxviridae* (Table 9). Biological interactions between human or animal pathogens, their vector arthropods and their own viruses, have been

reported [15,20,92], and such interactions can be utilized for disease control in agriculture and medicine. For example, mosquito-derived Densoviruses belonging to the family *Baculoviridae*, are used as stable vectors for the transformation of mosquitoes, which has created interest in using these viruses for mosquito and malaria control, either directly as insect-killing agents or as carriers of transgenes whose products interfere with parasite development [87]. Members of the *Baculoviridae* family are utilized for the control of insect pests [14]. The insect viruses detected in this study may have potential as tools for the biological control of ticks and tick-borne diseases.

Myoviridae, *Podoviridae*, and *Siphoviridae*, bacteriophage families, were also detected (Table 9). Phages detected in the tick viromes may be infectious to bacterial hosts residing in ticks or to those derived from vertebrate hosts. Bacteria belonging to the genus *Bacillus* and *Pseudomonas*, the hosts of *Bacillus phage G* and *Pseudomonas phages* belonging to the *Myoviridae*, *Podoviridae*, and *Siphoviridae* family, are common bacterial species found in the microbiome of *I. persulcatus* [73]. A metagenomic study of bacterial communities associated with ticks has revealed high bacterial diversity in ticks [73,85]. It is of interest to investigate the interactions between bacteria and bacteriophages in ticks, as phages may affect the physiology and fitness of ticks through their interaction with host bacteria.

Herpesviridae was detected from female and male samples. Main host of this viral family is vertebrates except that a herpesvirus was found in pacific oyster as an invertebrate host [108]. However, there is no report about arthropods as hosts of this viral family. Thus, possible explanations are that the herpesvirus genome detected in this study was mechanically acquired by ticks through the blood feeding on vertebrate hosts and the virus cannot replicate in ticks, or that ticks are potential hosts of

Herpesviridae.

About 5% of the contigs were assigned to viruses using BLAST approach, whereas 50% of contigs could not be classified (Figure 10). The classification methods for microbes based on sequence similarity such as BLASTn have inherent limitations when used for metagenomic analyses [75,76,112]. This is because current genome databases do not cover the sequences of the genomes of all living organisms, including viruses, despite the rapid increase in DNA entries. It is also possible that many of the sequence reads are too divergent from the sequence data deposited in reference databases, resulting in the difficulty of finding similar sequences. On the other hands, BLSOM does not require orthologous sequence data sets for phylogenetic classification of sequences [2, 105]. It is therefore possible to find taxonomical relationships of never-reported organisms to known, well-established organisms. This is one of the advantages of BLSOM, when it is applied to microbiomes composed of poorly characterized and highly diversified organisms [2]. Over half of the contigs were assigned to viruses using BLSOM, and only a small percentage of the contigs could not be assigned to any organisms (Figure 14) , which supports that BLSOM is theoretically advantageous in detecting and classifying previously unknown viruses over the homology-based search.

It should be mentioned, however, that BLSOM-based classification has limitations especially in estimation accuracy as demonstrated in the table 7. In fact, only about 60% of viral contigs (> 300 bp) identified by BLAST were allocated to viruses identified by BLSOM (Tables 8A and 8B). Moreover, about half of the pairwise comparisons between two methods were not in accord at the family level (Tables 8A and 8B). These results indicate that some viral sequences might be overlooked in

BLASTn analysis, while certain part of viral populations might be mislocated in BLSOM analysis. This discrepancy might be minimized if more entries are added to microbial sequence databases, especially those covering unexplored viral world. Currently, BLSOM-based classification is one of the bioinformatics tools which can redeem the homology-based methods.

This study is the first to perform tick viral population analysis using the nucleotide composition-based classification method, BLSOM. Since this approach can be applied to other vector arthropods of medical and veterinary importance, it might have great potential for mounting effective programs against vector-borne emerging infectious diseases. Both experimental and epidemiological studies are required to assess the risks of the identified viruses for human and animal health. Further identification of those viruses, at the species level or entire genome analyses could be achieved by using conventional methods, such as viral isolation from cell cultures or susceptible animals, or other molecular methods such as species-specific primer extension. Much deeper sequencing using larger amounts of nucleic acid than that used in this study might yield enough sequence data to assemble entire genomes of unknown viruses.

Summary

Ticks can transmit a wide range of viral, bacterial, and protozoan pathogens, which are often zoonotic. Novel tick-borne viral pathogens have been reported during the past few years. The aim of this study was to investigate the diversity of tick viral populations, which may contain as-yet unidentified pathogens, using a combination of high throughput pyrosequencing and BLAST analysis or a BLSOM program that can provide phylogenetic information based on oligonucleotide fingerprint similarity.

Viral particles were concentrated from tick homogenates by using membrane-filtration, centrifugation, and nucleases. Shot gun sequencing was performed by the 454 GS junior sequencer. In BLAST analysis, viral contigs were assigned to the *Mononegavirales*, *Bunyaviridae*, and *Rhabdviridae* families. Among these, sequences similar to tick associated viruses recently identified using next generation sequencing technology were also detected. On the other hand, the BLSOM method showed that ticks harbored a wide variety of viral taxa including 43 viral families, some of which have been previously reported to be associated with human and animal diseases, such as species belonging to the families *Bunyaviridae*, *Flaviviridae,* and *Reoviridae*.

Therefore, this approach is more sensitive for the screening of "yet-unknown" viral pathogens than conventional methods, and would thus allow the prediction of emerging tick-borne viral diseases. Both experimental and epidemiological studies are necessary to assess the risks of these viruses for human and animal health.

Conclusion

Ticks can transmit a wide range of microorganisms, such as viruses, bacteria, and protozoa, and its distribution is expanding mainly due to climate changes. In addition, emerging tick-borne diseases have recently been increasingly reported worldwide. Therefore, risks caused by ticks and tick-borne diseases are elevating. It is highly suspected that ticks still possess unrevealed pathogens which may threaten human and animal health. In this regard, epidemiological and bioinformatic studies were carried out on tick-borne pathogens, focusing on epidemiology of *Coxiella burnetii*, the causative agent of Q fever in livestock in Zambia and characterization of tick bacterial and viral populations.

In chapter I, *C. burnetii* DNA was detected in Zambian livestock using a PCR assay performed with primers based on a repetitive, transposon-like element. Blood samples of cattle and goats were collected in four areas, Monze, Chongwe, Petauke, and Chama. Samples from Chama area in the Eastern Province which is an extensive cattle-raising area showed the highest prevalence of *C. burnetii* DNA in cattle, which agree with the result of a previous serological study of humans showing that samples from Eastern and Western Provinces showed higher positive ratios than in other areas. These results suggested that livestock is one of the risk factors of infection with *C. burnetii* in Zambia.

In chapter II, bacterial flora was analyzed in tick salivary glands by 16S rDNA amplicon analysis with a next generation sequencer. Totally 163 different bacterial genera, including those known as tick-borne pathogens such as *Ehrlichia* and *Rickettsia*, were identified in this study. The principal component analysis revealed that

tick bacterial communities in salivary glands had differences in tick species. When compared with a conventional *Rickettsia*-specific PCR assay, this high throughput sequencing approach had higher sensitivity in the detection of rickettsial sequences. Thus, the strategy used in this study makes it feasible to detect both known and as-yet unknown pathogens, and therefore is useful for the surveillance of tick-borne pathogens.

In chapter III, viral community was analyzed in ticks by shot gun sequencing followed by BLASTn and Batch Learning Self-Organizing Map (BLSOM). BLASTn search of the resulting contig data identified 3 viral order and families, including *Mononegavirales, Bunyaviridae*, and *Rhabdviridae*. BLSOM is a composition-based data processing method which was designed to separate and cluster sequence fragments based on the similarity of oligonucleotide frequencies without any other taxonomical information. By applying this method, 43 different viral families were found from the same contig data sets used for BLASTn. This approach is useful for the screening of potential viral pathogens without prior knowledge, thus allowing the prediction of the emergence of yet-known tick-borne diseases. Both experimental and epidemiological studies are necessary to assess the ricks of these viruses for human and animal health.

The findings obtained from this study can provide valuable basic information for the prediction of and our preparedness against emerging tick-borne diseases.

Acknowledgements

This thesis would not have been realized without the support and cooperation from a great many people over the years and I would like to take this opportunity to thank them all.

First and foremost I would like to express the deepest and sincerest gratitude to my supervisor Prof. Chihiro Sugimoto from Division of Collaboration and Education, Hokkaido University Research Center for Zoonosis Control (Sapporo, Japan), whose innumerable accurate advice, valuable suggestions, greatest patience and encouragements helped me during the research for and writing of this thesis. In addition, the amount of freedom you have given me over past years is matchless and was highly appreciated. Prof. Sugimoto, your trust, network, and vision allowed me to complete this PhD with great pleasure.

I extend my heartfelt acknowledgment to Assistant Prof. Ryo Nakao (Unite of Risk Analysis and Management, Hokkaido University Research Center for Zoonosis Control, Sapporo, Japan) for giving me so many helpful supports and for tirelessly teaching me. My research project would not have been a success without his presence. He also taught me what is researcher by showing an attitude without any words and how to drink Japanese Sake. I would like to thank him again for being a great senior, friend, and drinking collaborator.

I am really thankful to collaborators, Associate Prof. Takashi Abe (Department of information engineering, Niigata University, Niigata, Japan), Ms. Aiko Ohnuma (Administration Office, Hokkaido University Research Center for Zoonosis Control, Sapporo, Japan), Dr. Fumihiko Kawamori (Department of Microbiology,

Shizuoka Prefectural Institute of Public Health and Environmental Science, Shizuoka, Japan), and Prof. Boniface Namangala (Department of Paraclinical Studies, School of Veterinary Medicine, University of Zambia, Lusaka, Zambia) for their technical and intellectual supports.

Sincerely, I appreciate supports and guidance for my research and studies from Prof. Kazuhiko Ohashi (Laboratory of Infectious Disease, Graduate School of Veterinary Medicine, Hokkaido University, Sapporo, Japan), Prof. Kimihito Ito (Division of Bioinformatics, Hokkaido University Research Center for Zoonosis Control, Sapporo, Japan), and Associate Prof. Junya Yamagishi (Division of Collaboration and Education, Hokkaido University Research Center for Zoonosis Control, Sapporo, Japan)

I would like to thank Associate Prof. Manabu Igarashi (Division of Global Epidemiology, Hokkaido University Research Center for Zoonosis Control, Sapporo, Japan), and Ms. Akina Mori (Division of Global Epidemiology, Hokkaido University Research Center for Zoonosis Control, Sapporo, Japan) for their help and encouragement during my research and studies.

I would like to express my appreciation to Dr. Junji Seto, Dr. Naota Monma, Associate Prof. Nariaki Nonaka, Assistant Prof. Yumi Kirino, and Dr. Seigo Yamamoto for their helping my field activities.

I thank my family, more especially my parents, Yongzhi Qiu and Huimei Wang, for being supportive and patient of my career choice, my close friends and colleagues, Ms. Kanjana Changkaew (It was always a pleasure lunch time talk), Mr. Makoto Kuroda (I will not forget the time we sheared), Dr. Mieko Muramatsu (Thank you for nice talk), Ms. Chiho Kaneko (Good luck with research in your PhD!), Dr.

Chonchanok Theethakaew (Khin khao mai?), Mr. Fumiaki Atarashi (Thank “le coq” for every time coming and talking together in Yamagata), and Mr. Dongcheol Hyeon (Thank you for your rooting for me from Okinawa) who have shared in the struggle for a brighter future and whose kind words and encouragement gave me strength to face daily challenges. I would also like to express my appreciation to secretaries of our laboratory, Ms. Rie Tanno (Chairman of our drinking club) and Ms. Miwako Kurouji (Thank you for giving wonderful taste of the homemade sweets).

Finally, I would like to take pleasure in expressing my gratitude to the all members of Hokkaido University Research Center for Zoonosis Control for their warm and kind supports. I would also like to pray for the soul of ticks used in this study, may they rest in peace.

References

1. Abe, T., Kanaya, S., Kinouchi, M., Ichiba, Y., Kozuki, T. and Ikemura, T. 2003. Informatics for unveiling hidden genome signatures. *Genome Res.*, **13**: 693-702.

2. Abe, T., Sugawara, H., Kinouchi, M., Kanaya, S. and Ikemura, T. 2005. Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples. *DNA Res.*, **12**: 281-290.

3. Andreotti, R., Pérez de León, A. A., Dowd, S. E., Guerrero, F. D., Bendele, K. G. and Scoles, G. A. 2011. Assessment of bacterial diversity in the cattle tick *Rhipicephalus* (*Boophilus*) *microplus* through tag-encoded pyrosequencing. *BMC Microbiol.*, **11**: 6.

4. Arthur, D. R. 1965. Ticks in Egypt in 1500 B.C.? *Nature*, **206**: 1060-1061.

5. Bastian, F. O., Sanders, D. E., Forbes, W. A., Hagius, S. D., Walker, J. V., Henk, W. G., Enright, F. M. and Elzer, P. H. 2007. *Spiroplasma* spp. from transmissible spongiform encephalopathy brains or ticks induce spongiform encephalopathy in ruminants. *J. Med. Microbiol.*, **56**: 1235-1242.

6. Belhouchet, M., Mohd Jaafar, F., Tesh, R., Grimes, J., Maan, S., Mertens, P. P. and Attoui, H. 2010. Complete sequence of Great Island virus and comparison with the T2 and outer-capsid proteins of Kemerovo, Lipovnik and Tribec viruses (genus Orbivirus, family Reoviridae). *J. Gen. Virol.*, **91**: 2985-2993.

7. Benach, J. L., Coleman, J. L., Skinner, R. A. and Bosler, E. M. 1987. Adult *Ixodes dammini* on rabbits: a hypothesis for the development and transmission of Borrelia burgdorferi. *J. Infect. Dis.*, **155**: 1300-1306.

8. Bente, D. A., Forrester, N. L., Watts, D. M., McAuley, A. J., Whitehouse, C. A. and Bray, M. 2013. Crimean-Congo hemorrhagic fever: history, epidemiology, pathogenesis, clinical syndrome and genetic diversity. *Antiviral. Res.*, **100**: 159-189.

9. Bernasconi, M. V., Casati, S., Péter, O. and Piffaretti, J. C. 2002. *Rhipicephalus* ticks infected with *Rickettsia* and *Coxiella* in Southern Switzerland (Canton Ticino). *Infect. Genet. Evol.*, **2**: 111-120.

10. Berrada, Z. L. and Telford, S. R. 2009. Burden of tick-borne infections on American companion animals. *Top. Companion. Anim. Med.*, **24**: 175-181.

11. Berri, M., Laroucau, K. and Rodolakis, A. 2000. The detection of *Coxiella burnetii* from ovine genital swabs, milk and fecal samples by the use of a single touchdown polymerase chain reaction. *Vet. Microbiol.*, **72**: 285-293.

12. Biesiada, G., Czepiel, J., Leśniak, M. R., Garlicki, A. and Mach, T. 2012. Lyme disease: review. *Arch. Med. Sci.*, **8**: 978-982.

13. Bonnet, S., de la Fuente, J., Nicollet, P., Liu, X., Madani, N., Blanchard, B., Maingourd, C., Alongi, A., Torina, A., Fernández de Mera, I. G., Vicente, J., George, J. C., Vayssier-Taussat, M. and Joncour, G. 2013. Prevalence of tick-borne pathogens in adult *Dermacentor* spp. ticks from nine collection sites in France. *Vector Borne Zoonotic Dis.*, **13**: 226-236.

14. Bonning, B. C. and Hammock, B. D. 1996. Development of recombinant baculoviruses for insect control. *Annu. Rev. Entomol.*, **41**: 191-210.

15. Carlson, J., Suchman, E. and Buchatsky, L. 2006. Densoviruses for control and genetic manipulation of mosquitoes. *Adv. Virus Res.*, **68**: 361-392.

16. Carmona-Vicente, N., Buesa, J., Brown, P. A., Merga, J. Y., Darby, A. C., Stavisky, J., Sadler, L., Gaskell, R. M., Dawson, S. and Radford, A. D. 2013. Phylogeny and prevalence of kobuviruses in dogs and cats in the UK. *Vet. Microbiol.*, **164**: 246-252.

17. Carpi, G., Cagnacci, F., Wittekindt, N. E., Zhao, F., Qi, J., Tomsho, L. P., Drautz, D. I., Rizzoli, A. and Schuster, S. C. 2011. Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks. *PLoS One*, **6**: e25604.

18. Carroll, I. M., Ringel-Kulka, T., Siddle, J. P. and Ringel, Y. 2012. Alterations in composition and diversity of the intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome. *Neurogastroenterol. Motil.*, **24**: 521-530, e248.

19. Chan, C. K., Hsu, A. L., Tang, S. L. and Halgamuge, S. K. 2008. Using growing self-organising maps to improve the binning process in environmental whole-genome shotgun sequencing. *J. Biomed. Biotechnol.*, **2008**: 513701.

20. Cheng, L., Chen, S., Zhou, Z. H. and Zhang, J. 2007. Structure comparisons of *Aedes albopictus* densovirus with other parvoviruses. *Sci. China. C. Life Sci.*, **50**: 70-74.

21. Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R. and Tiedje, J. M. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.*, **42**: D633-642.

22 Cooper, A., Stephens, J., Ketheesan, N. and Govan, B. 2013. Detection of *Coxiella burnetii* DNA in wildlife and ticks in northern Queensland, Australia. *Vector Borne Zoonotic Dis.*, **13**: 12-16.

23 Derrick, E. H. 1983. "Q" fever, a new fever entity: clinical features, diagnosis and laboratory investigation. *Rev. Infect. Dis.*, **5**: 790-800.

24. Dick, G. J., Andersson, A. F., Baker, B. J., Simmons, S. L., Thomas, B. C., Yelton, A. P. and Banfield, J. F. 2009. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.*, **10**: R85.

25. Dobler, G. 2010. Zoonotic tick-borne flaviviruses. *Vet Microbiol*, **140**: 221-228.

26. Dupont, H. T., Brouqui, P., Faugere, B. and Raoult, D. 1995. Prevalence of antibodies to *Coxiella burnetti*, *Rickettsia conorii*, and *Rickettsia typhi* in seven African countries. *Clin. Infect. Dis.*, **21**: 1126-1133.

27. Edlow, J. A. 2010. Tick paralysis. *Curr. Treat. Options. Neurol.*, **12**: 167-177.

28. Eklund, C. M., Parker, R. R. and Lackman, D. B. 1947. A case of Q fever probably contracted by exposure to ticks in nature. *Public Health Rep.*, **62**: 1413-1416.

29. Erickson, D. L., Anderson, N. E., Cromar, L. M. and Jolley, A. 2009. Bacterial communities associated with flea vectors of plague. *J. Med. Entomol.*, **46**: 1532-1536.

30. Fishbein, D. B. and Raoult, D. 1992. A cluster of *Coxiella burnetii* infections associated with exposure to vaccinated goats and their unpasteurized dairy products. *Am. J. Trop. Med. Hyg.*, **47**: 35-40.

31. Fournier, P. E., Fujita, H., Takada, N. and Raoult, D. 2002. Genetic identification of rickettsiae isolated from ticks in Japan. *J. Clin. Microbiol.*, **40**:

2176-2181.

32. Fujita, H., Fournier, P. E., Takada, N., Saito, T. and Raoult, D. 2006. *Rickettsia asiatica* sp. nov., isolated in Japan. *Int. J. Syst. Evol. Microbiol.*, **56**: 2365-2368.

33. Fujita, H., Takada, N., Kawabata, H., Ishiguro, F., Yamamoto, S., Oikawa, Y., Yano, Y., Ma, X.-H. and Oh, H.-S. 2007. Some suggestive records of rickettsiae isolated from ticks in Korea and central China. *Annual Report of Ohara General Hospital*, **47**: 4.

34. Futse, J. E., Ueti, M. W., Knowles, D. P. and Palmer, G. H. 2003. Transmission of *Anaplasma marginale* by *Boophilus microplus*: retention of vector competence in the absence of vector-pathogen interaction. *J. Clin. Microbiol.*, **41**: 3829-3834.

35. Ghirotti, M., Semproni, G., De Meneghi, D., Mungaba, F. N., Nannini, D., Calzetta, G. and Paganico, G. 1991. Sero-prevalences of selected cattle diseases in the Kafue flats of Zambia. *Vet. Res. Commun.*, **15**: 25-36.

36. Glazunova, O., Roux, V., Freylikman, O., Sekeyova, Z., Fournous, G., Tyczka, J., Tokarevich, N., Kovacava, E., Marrie, T. J. and Raoult, D. 2005. *Coxiella burnetii* genotyping. *Emerg. Infect. Dis.*, **11**: 1211-1217.

37. Gorenflot, A., Moubri, K., Precigout, E., Carcy, B. and Schetters, T. P. 1998. Human babesiosis. *Ann. Trop. Med. Parasitol.*, **92**: 489-501.

38. Granberg, F., Vicente-Rubiano, M., Rubio-Guerri, C., Karlsson, O. E., Kukielka, D., Belák, S. and Sánchez-Vizcaíno, J. M. 2013. Metagenomic detection of viral pathogens in Spanish honeybees: co-infection by Aphid Lethal Paralysis, Israel Acute Paralysis and Lake Sinai Viruses. *PLoS One*, **8**: e57459.

39. Gritsun, T. S., Lashkevich, V. A. and Gould, E. A. 2003. Tick-borne encephalitis. *Antiviral Res.*, **57**: 129-146.

40. Guglielmone, A. A., Robbins, R. G., Apanaskevich, D. A., Petney, T. N., Estrada-Pena, A., Horak, I. G., Shao, R. and Barker, S. C. 2010. The Argasidae, Ixodidae and Nuttalliellidae (Acari: Ixodida) of the world : a list of valid species names. *Zootaxa*, **2538**: 28.

41. Guglielmone, A. A. R., Richard G., Apanaskevich, D. A., Petney, T. N., Estrada-Peña, A. and Horak, I. G. 2014. The hard ticks of the world, Springer,

New York - London.

42. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. and Goodman, R. M. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.*, **5**: R245-249.

43. Hashimoto, S., Kawado, M., Murakami, Y., Izumida, M., Ohta, A., Tada, Y., Shigematsu, M., Yasui, Y., Taniguchi, K. and Nagai, M. 2007. Epidemics of vector-borne diseases observed in infectious disease surveillance in Japan, 2000-2005. *J. Epidemiol.*, **17 Suppl**: S48-55.

44. Hayasaka, D., Ivanov, L., Leonova, G. N., Goto, A., Yoshii, K., Mizutani, T., Kariwa, H. and Takashima, I. 2001. Distribution and characterization of tick-borne encephalitis viruses from Siberia and far-eastern Asia. *J. Gen. Virol.*, **82**: 1319-1328.

45. He, B., Li, Z., Yang, F., Zheng, J., Feng, Y., Guo, H., Li, Y., Wang, Y., Su, N., Zhang, F., Fan, Q. and Tu, C. 2013. Virome profiling of bats from Myanmar by metagenomic analysis of tissue samples reveals more novel Mammalian viruses. *PLoS One*, **8**: e61950.

46. Henning, K., Greiner-Fischer, S., Hotzel, H., Ebsen, M. and Theegarten, D. 2006. Isolation of *Spiroplasma* sp. from an Ixodes tick. *Int. J. Med. Microbiol.*, **296 Suppl 40**: 157-161.

47. Hurst, G. D. and Jiggins, F. M. 2000. Male-killing bacteria in insects: mechanisms, incidence, and implications. *Emerg. Infect. Dis.*, **6**: 329-336.

48. Ishiguro, F., Takada, N., Fujita, H., Noji, Y., Yano, Y. and Iwasaki, H. 2008. Survey of the vectorial competence of ticks in an endemic area of spotted fever group rickettsioses in Fukui Prefecture, Japan. *Microbiol. Immunol.*, **52**: 305-309.

49. Jones, M. S., Harrach, B., Ganac, R. D., Gozum, M. M., Dela Cruz, W. P., Riedel, B., Pan, C., Delwart, E. L. and Schnurr, D. P. 2007. New adenovirus species found in a patient presenting with gastroenteritis. *J. Virol.*, **81**: 5978-5984.

50. Jones, M. S., Lukashov, V. V., Ganac, R. D. and Schnurr, D. P. 2007. Discovery of a novel human picornavirus in a stool sample from a pediatric patient presenting with fever of unknown origin. *J. Clin. Microbiol.*, **45**: 2144-2150.

51. Jongejan, F. and Uilenberg, G. 2004. The global importance of ticks. *Parasitology*, **129 Suppl**: S3-14.

52. Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T., Mori, H. and Ikemura, T. 2001. Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. *Gene*, **276**: 89-99.

53. Kawahara, M., Ito, T., Suto, C., Shibata, S., Rikihisa, Y., Hata, K. and Hirai, K. 1999. Comparison of *Ehrlichia muris* strains isolated from wild mice and ticks and serologic survey of humans and animals with *E. muris* as antigen. *J. Clin. Microbiol.*, **37**: 1123-1129.

54. Kelly, P. J., Matthewman, L. A., Mason, P. R. and Raoult, D. 1993. Q fever in Zimbabwe. A review of the disease and the results of a serosurvey of humans, cattle, goats and dogs. *S. Afr. Med. J.*, **83**: 21-25.

55. Kim, Y. S., Kim, M. C., Kwon, S. W., Kim, S. J., Park, I. C., Ka, J. O. and Weon, H. Y. 2011. Analyses of bacterial communities in meju, a Korean traditional fermented soybean bricks, by cultivation-based and pyrosequencing methods. *J. Microbiol.*, **49**: 340-348.

56. Klyachko, O., Stein, B. D., Grindle, N., Clay, K. and Fuqua, C. 2007. Localization and visualization of a *coxiella*-type symbiont within the lone star tick, *Amblyomma americanum*. *Appl. Environ. Microbiol.*, **73**: 6584-6594.

57. Kohonen, T. 1990. The self-organizing map. *Proceedings of the IEEE* pp. 1464-1480.

58. Kurtenbach, K., De Michelis, S., Etti, S., Schäfer, S. M., Sewell, H. S., Brade, V. and Kraiczy, P. 2002. Host association of *Borrelia burgdorferi* sensu lato-the key role of host complement. *Trends Microbiol.*, **10**: 74-79.

59. Labuda, M. and Nuttall, P. A. 2004. Tick-borne viruses. *Parasitology*, **129 Suppl**: S221-245.

60. Lee, J. H., Park, H. S., Jang, W. J., Koh, S. E., Park, T. K., Kang, S. S., Kim, B. J., Kook, Y. H., Park, K. H. and Lee, S. H. 2004. Identification of the *Coxiella* sp. detected from *Haemaphysalis longicornis* ticks in Korea. *Microbiol. Immunol.*, **48**: 125-130.

61. Lewis, D. 1979. The detection of rickettsia-like microorganisms within the ovaries of female *Ixodes ricinus* ticks. *Z. Parasitenkd.*, **59**: 295-298.

62. Lindquist, L. and Vapalahti, O. 2008. Tick-borne encephalitis. *Lancet*, **371**: 1861-1871.

63. Macaluso, K. R., Sonenshine, D. E., Ceraul, S. M. and Azad, A. F. 2002. Rickettsial infection in *Dermacentor variabilis* (Acari: Ixodidae) inhibits transovarial transmission of a second *Rickettsia*. *J. Med. Entomol.*, **39**: 809-813.

64. Mahara, F. 1997. Japanese spotted fever: report of 31 cases and review of the literature. *Emerg. Infect. Dis.*, **3**: 105-111.

65. Mahara, F. 2006. Rickettsioses in Japan and the far East. *Ann. N. Y. Acad. Sci.*, **1078**: 60-73.

66. Marrie, T. J. 1990. Q fever - a review. *Can. Vet. J.*, **31**: 555-563.

67. Martin, C., Diaz, N. N., Ontrup, J. and Nattkemper, T. W. 2008. Hyperbolic SOM-based clustering of DNA fragment features for taxonomic visualization and classification. *Bioinformatics*, **24**: 1568-1574.

68. Matsuura, Y., Kikuchi, Y., Meng, X. Y., Koga, R. and Fukatsu, T. 2012. Novel clade of alphaproteobacterial endosymbionts associated with stinkbugs and other arthropods. *Appl. Environ. Microbiol.*, **78**: 4149-4156.

69. Maurin, M. and Raoult, D. 1999. Q fever. *Clin. Microbiol. Rev.*, **12**: 518-553.

70. McMullan, L. K., Folk, S. M., Kelly, A. J., MacNeil, A., Goldsmith, C. S., Metcalfe, M. G., Batten, B. C., Albariño, C. G., Zaki, S. R., Rollin, P. E., Nicholson, W. L. and Nichol, S. T. 2012. A new phlebovirus associated with severe febrile illness in Missouri. *N. Engl. J. Med.*, **367**: 834-841.

71. Mediannikov, O., Ivanov, L. I., Nishikawa, M., Saito, R., Sidel'nikov, I. N., Zdanovskaia, N. I., Mokretsova, E. V., Tarasevich, I. V. and Suzuki, H. 2004. [Microorganism "Montezuma" of the order Rickettsiales: the potential causative agent of tick-borne disease in the Far East of Russia]. *Zh. Mikrobiol. Epidemiol. Immunobiol.*, 1: 7-13. (In Russian)

72. Menchaca, A. C., Visi, D. K., Strey, O. F., Teel, P. D., Kalinowski, K., Allen, M. S. and Williamson, P. C. 2013. Preliminary assessment of microbiome changes following blood-feeding and survivorship in the *Amblyomma americanum*

nymph-to-adult transition using semiconductor sequencing. *PLoS One*, **8**: e67129.

73. Nakao, R., Abe, T., Nijhof, A. M., Yamamoto, S., Jongejan, F., Ikemura, T. and Sugimoto, C. 2013. A novel approach, based on BLSOMs (Batch Learning Self-Organizing Maps), to the microbiome analysis of ticks. *ISME J.*, **7**: 1003-1015.

74. Nakao, R., Magona, J. W., Zhou, L., Jongejan, F. and Sugimoto, C. 2011. Multi-locus sequence typing of *Ehrlichia ruminantium* strains from geographically diverse origins and collected in *Amblyomma variegatum* from Uganda. *Parasit. Vectors*, **4**: 137.

75. Ng, T. F., Duffy, S., Polston, J. E., Bixby, E., Vallad, G. E. and Breitbart, M. 2011. Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. *PLoS One*, **6**: e19050.

76. Ng, T. F., Willner, D. L., Lim, Y. W., Schmieder, R., Chau, B., Nilsson, C., Anthony, S., Ruan, Y., Rohwer, F. and Breitbart, M. 2011. Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. *PLoS One*, **6**: e20579.

77. Noji, Y., Takada, N., Ishiguro, F., Fujino, S., Aoyama, T., Fujita, H., Yano, Y., Shiomi, S., Mitsuto, I., Takase, K., Haba, T. and Mabuchi, H. 2005. The first reported case of spotted fever in Fukui Prefecture, the northern part of central Japan. *Jpn. J. Infect. Dis.*, **58**: 112-114.

78. Nuttall, P. A. 2009. Molecular characterization of tick-virus interactions. *Front. Biosci. (Landmark Ed)*, **14**: 2466-2483.

79. Okabayashi, T., Hasebe, F., Samui, K. L., Mweene, A. S., Pandey, S. G., Yanase, T., Muramatsu, Y., Ueno, H. and Morita, C. 1999. Short report: prevalence of antibodies against spotted fever, murine typhus, and Q fever rickettsiae in humans living in Zambia. *Am. J. Trop. Med. Hyg.*, **61**: 70-72.

80. Parola, P., Davoust, B. and Raoult, D. 2005. Tick- and flea-borne rickettsial emerging zoonoses. *Vet. Res.*, **36**: 469-492.

81. Parola, P., Paddock, C. D., Socolovschi, C., Labruna, M. B., Mediannikov, O., Kernif, T., Abdad, M. Y., Stenos, J., Bitam, I., Fournier, P. E. and Raoult, D. 2013. Update on Tick-Borne Rickettsioses around the World: a Geographic

Approach. *Clin. Microbiol. Rev.*, **26**: 657-702.

82. Peter, T. F., Deem, S. L., Barbet, A. F., Norval, R. A., Simbi, B. H., Kelly, P. J. and Mahan, S. M. 1995. Development and evaluation of PCR assay for detection of low levels of *Cowdria ruminantium* infection in *Amblyomma* ticks not detected by DNA probe. *J. Clin. Microbiol.*, **33**: 166-172.

83. Piesman, J. and Eisen, L. 2008. Prevention of tick-borne diseases. *Annu. Rev. Entomol.*, **53**: 323-343.

84. Popov, V. L., Korenberg, E. I., Nefedova, V. V., Han, V. C., Wen, J. W., Kovalevskii, Y. V., Gorelova, N. B. and Walker, D. H. 2007. Ultrastructural evidence of the ehrlichial developmental cycle in naturally infected *Ixodes persulcatus* ticks in the course of coinfection with *Rickettsia*, *Borrelia*, and a flavivirus. *Vector Borne Zoonotic Dis.*, **7**: 699-716.

85. Qiu, Y., Nakao, R., Ohnuma, A., Kawamori, F. and Sugimoto, C. 2014. Microbial population analysis of the salivary glands of ticks; a possible strategy for the surveillance of bacterial pathogens. *PLoS One*, **9**: e103961.

86. Ratmanov, P., Bassene, H., Fenollar, F., Tall, A., Sokhna, C., Raoult, D. and Mediannikov, O. 2012. The Correlation of Q Fever and *Coxiella burnetii* DNA in Household Environments in Rural Senegal. *Vector Borne Zoonotic Dis.*, **13**: 70-72.

87. Ren, X., Hoiczyk, E. and Rasgon, J. L. 2008. Viral paratransgenesis in the malaria vector *Anopheles gambiae*. *PLoS Pathog.*, **4**: e1000135.

88. Reyes, G. R. and Kim, J. P. 1991. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol. Cell. Probes*, **5**: 473-481.

89. Ricketts, H. 1906. The transmission of rocky mountain spotted fever the bite of the wood-tick (*Dermacentor occidentalis*). *J. Amer. Med. Assoc.*, **47**: 458.

90. Roshdy, M. A. 1968. A rickettsialike microorganism in the tick Ornithodoros savignyi: Observations on its structure and distribution in the tissues of the tick. *J. Invertebr. Pathol.*, **11**: 155-169.

91. Roux, V., Rydkina, E., Eremeeva, M. and Raoult, D. 1997. Citrate synthase gene comparison, a new tool for phylogenetic analysis, and its application for the

rickettsiae. *Int. J. Syst. Bacteriol.*, **47**: 252-261.

92. Rwegoshora, R. T. and Kittayapong, P. 2004. Pathogenicity and infectivity of the Thai-strain densovirus (AThDNV) in *Anopheles minimus* S.L. *Southeast Asian J. Trop. Med. Public Health*, **35**: 630-634.

93. Santos, A. S., Bacellar, F., Santos-Silva, M., Formosinho, P., Grácio, A. J. and Franca, S. 2002. Ultrastructural study of the infection process of *Rickettsia conorii* in the salivary glands of the vector tick *Rhipicephalus sanguineus*. *Vector Borne Zoonotic Dis.*, **2**: 165-177.

94. Schneeberger, P. M., Wintenberger, C., van der Hoek, W. and Stahl, J. P. 2014. Q fever in the Netherlands - 2007-2010: what we learned from the largest outbreak ever. *Med. Mal. Infect.*, **44**: 339-353.

95. Sexton, D. J., Dwyer, B., Kemp, R. and Graves, S. 1991. Spotted fever group rickettsial infections in Australia. *Rev. Infect. Dis.*, **13**: 876-886.

96. Shibata, S., Kawahara, M., Rikihisa, Y., Fujita, H., Watanabe, Y., Suto, C. and Ito, T. 2000. New *Ehrlichia* species closely related to *Ehrlichia chaffeensis* isolated from *Ixodes ovatus* ticks in Japan. *J. Clin. Microbiol.*, **38**: 1331-1338.

97. Socolovschi, C., Mediannikov, O., Raoult, D. and Parola, P. 2009. The relationship between spotted fever group Rickettsiae and ixodid ticks. *Vet. Res.*, **40**: 34.

98. Takashima, I., Morita, K., Chiba, M., Hayasaka, D., Sato, T., Takezawa, C., Igarashi, A., Kariwa, H., Yoshimatsu, K., Arikawa, J. and Hashimoto, N. 1997. A case of tick-borne encephalitis in Japan and isolation of the the virus. *J. Clin. Microbiol.*, **35**: 1943-1947.

99. Takeda, T., Ito, T., Chiba, M., Takahashi, K., Niioka, T. and Takashima, I. 1998. Isolation of tick-borne encephalitis virus from *Ixodes ovatus* (Acari: Ixodidae) in Japan. *J. Med. Entomol.*, **35**: 227-231.

100. Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.*, **30**: 2725-2729.

101. Taroura, S., Shimada, Y., Sakata, Y., Miyama, T., Hiraoka, H., Watanabe, M., Itamoto, K., Okuda, M. and Inokuma, H. 2005. Detection of DNA of

'*Candidatus* Mycoplasma haemominutum' and *Spiroplasma* sp. in unfed ticks collected from vegetation in Japan. *J. Vet. Med. Sci.*, **67**: 1277-1279.

102. Thomas, J. M. 2006. Q fever, In : *Tropical Infectious Diseases*, 2nd ed., pp. 574-577, Richard, L. G., David, H. W. and Peter, F. W. eds. Churchill Livingstone, Philadelphia.

103. Tokarz, R., Williams, S. H., Sameroff, S., Sanchez Leon, M., Jain, K. and Lipkin, W. I. 2014. Virome analysis of *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis* ticks reveals novel highly divergent vertebrate and invertebrate viruses. *J. Virol.*, **88**: 11480-11492.

104. Tully, J. G., Rose, D. L., Yunker, C. E., Cory, J., Whitcomb, R. F. and Williamson, D. L. 1981. Helical mycoplasmas (spiroplasmas) from Ixodes ticks. *Science*, **212**: 1043-1045.

105. Uehara, H., Iwasaki, Y., Wada, C., Ikemura, T. and Abe, T. 2011. A novel bioinformatics strategy for searching industrially useful genome resources from metagenomic sequence libraries. *Genes Genet. Syst.*, **86**: 53-66.

106. Vaidya, V. M., Malik, S. V., Kaur, S., Kumar, S. and Barbuddhe, S. B. 2008. Comparison of PCR, immunofluorescence assay, and pathogen isolation for diagnosis of q fever in humans with spontaneous abortions. *J. Clin. Microbiol.*, **46**: 2038-2044.

107. van Loenhout, J. A., Paget, W. J., Vercoulen, J. H., Wijkmans, C. J., Hautvast, J. L. and van der Velden, K. 2012. Assessing the long-term health impact of Q-fever in the Netherlands: a prospective cohort study started in 2007 on the largest documented Q-fever outbreak to date. *BMC Infect. Dis.*, **12**: 280.

108. Wang, N., Baldi, P. F. and Gaut, B. S. 2007. Phylogenetic analysis, genome evolution and the rate of gene gain in the Herpesviridae. *Mol. Phylogenet. Evol.*, **43**: 1066-1075.

109. Weber, M., Teeling, H., Huang, S., Waldmann, J., Kassabgy, M., Fuchs, B. M., Klindworth, A., Klockow, C., Wichels, A., Gerdts, G., Amann, R. and Glöckner, F. O. 2011. Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics. *ISME J.*, **5**: 918-928.

110. Weisburg, W. G., Barns, S. M., Pelletier, D. A. and Lane, D. J. 1991. 16S

ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.*, **173**: 697-703.

111. Willems, H., Thiele, D., Frölich-Ritter, R. and Krauss, H. 1994. Detection of *Coxiella burnetii* in cow's milk using the polymerase chain reaction (PCR). *Zentralbl. Veterinarmed. B.*, **41**: 580-587.

112. Winter, C., Garcia, J. A., Weinbauer, M. G., DuBow, M. S. and Herndl, G. J. 2014. Comparison of deep-water viromes from the Atlantic ocean and the Mediterranean sea. *PLoS One*, **9**: e100600.

113. Woldehiwet, Z. 2004. Q fever (coxiellosis): epidemiology and pathogenesis. *Res. Vet. Sci.*, **77**: 93-100.

114. Wright, E. S., Yilmaz, L. S. and Noguera, D. R. 2012. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl. Environ. Microbiol.*, **78**: 717-725.

115. Wójcik-Fatla, A., Zając, V., Cisak, E., Sroka, J., Sawczyn, A. and Dutkiewicz, J. 2012. Leptospirosis as a tick-borne disease? Detection of *Leptospira* spp. in *Ixodes ricinus* ticks in eastern Poland. *Ann. Agric. Environ. Med.*, **19**: 656-659.

116. Yu, X. J., Liang, M. F., Zhang, S. Y., Liu, Y., Li, J. D., Sun, Y. L., Zhang, L., Zhang, Q. F., Popov, V. L., Li, C., Qu, J., Li, Q., Zhang, Y. P., Hai, R., Wu, W., Wang, Q., Zhan, F. X., Wang, X. J., Kan, B., Wang, S. W., Wan, K. L., Jing, H. Q., Lu, J. X., Yin, W. W., Zhou, H., Guan, X. H., Liu, J. F., Bi, Z. Q., Liu, G. H., Ren, J., Wang, H., Zhao, Z., Song, J. D., He, J. R., Wan, T., Zhang, J. S., Fu, X. P., Sun, L. N., Dong, X. P., Feng, Z. J., Yang, W. Z., Hong, T., Zhang, Y., Walker, D. H., Wang, Y. and Li, D. X. 2011. Fever with thrombocytopenia associated with a novel bunyavirus in China. *N. Engl. J. Med.*, **364**: 1523-1532.

和文要旨

マダニは、様々なウイルス、細菌、原虫を媒介する獣医学ならびに公衆衛生学上極めて重要な吸血性節足動物である。現在でも新興感染症の原因としてマダニから新規の病原体が発見されており、マダニ保有微生物叢の全貌解明はマダニ媒介性新興感染症の出現予測や診断法開発に重要である。しかし、自然界中の微生物の 99%以上は培養不可能あるいは困難であり、微生物の網羅的解析は通常の手法では難しい。

本研究では、マダニ関連細菌でもある Q 熱病原体(*Coxiella burnetii*)のアフリカにおける疫学的調査を手始めとして、次世代シーケンス技術を用いたマダニ保有微生物叢の網羅的解析を実施した。

第一章では、ザンビア共和国の Monze、Chongwe、Petauke、Chama の 4 ヶ所において採取したウシ並びにヤギの血液を用い *Coxiella burnetii* 特異的プライマーを使用した遺伝子増幅法(PCR)により保有調査を行った。その結果、ザンビア共和国の家畜において本菌遺伝子が確認され、その陽性率は、サンプル採取地で異なりザンビア東部の Chama 地域のウシで 11.2%と最も高かった。本結果は、1990 年代に実施された研究で明らかにされた人における本菌に対する抗体調査と同様に、畜産業が盛んな地域でより高い遺伝子陽性率を示す傾向にあった。

マダニは吸血時に唾液を宿主に注入することから、唾液腺に存在する微生物は、唾液と共に吸血宿主に移行する可能性があり、唾液腺中微生物叢の解析は新興のマダニ媒介性感染症対策の先回り対策に有用な知見をもたらす。そこで、第二章ではマダニ唾液腺における細菌叢の網羅的検索を 16S リボソーム DNA の PCR 増幅産物(16S アンプリコン)解析手法を用いて行った。その結果、

マダニ唾液腺から 163 属に及ぶ細菌が検出され、マダニ種により細菌叢の構成が異なることが明らかとなった。また、16S リボソーム DNA 全長における解析では、急性熱性疾患との関連が疑われるリケッチア目細菌に遺伝的に近い配列が検出された。加えて、既知のマダニ媒介性細菌であるリケッチアに関して、クエン酸合成酵素遺伝子の検出による *gltA*-PCR 法と 16S アンプリコン解析手法を比較したところ、後者の方が検出感度が高いことが認められた。本研究の結果、マダニ唾液腺には病原体を含む多様な細菌が存在していると考えられた。

上述のように、マダニ細菌叢解析の手法は確立できたが、マダニが保有・媒介する病原体にはウイルスも含まれる。そこで第三章では、ショットガンシーケンス技術を用いたマダニ保有ウイルス叢の網羅的解析を、現在広く用いられているシーケンスアラインメントに基づく相同配列検索法である BLAST に加えて、連続塩基組成により帰属生物群を推定する一括学習型自己組織化マップ(Batch-Learning Self-Organizing Map, BLSOM)手法と組み合わせて行った。まず、マダニ乳剤をフィルトレーション、遠心分画、核酸分解酵素処理により、ウイルスを含む画分を濃縮した。さらにその分画から調製された DNA あるいは cDNA 断片を次世代シーケンサーにより塩基配列を決定した。得られた配列を BLAST で解析した結果、メスで 3.5%、オスで 6.7%のコンティグがウイルス由来と推定され、それらはモノネガウイルス目、ブニヤウイルス科、ラブドウイルス科に属すると考えられた。その中には、2014 年にニューヨークで採集された *Ixodes scapularis* から検出されたブニヤウイルス科の South Bay virus ゲノムの L と S 分節に近い配列など次世代シーケンス技術により、その存在が明らかとなったマダニ関連ウイルスに類縁の配列が含まれていた。一方、BLSOM 解析では 50%以上のコンティグがウイルス由来配列と推定され、それは 43 種類のウイルス科に亘っていた。本研究の結果より、マダニ保有ウイルス叢は、極めて多様

性に富むウイルス種から構成されている可能性が示唆された。

このようにマダニが保有する細菌、ウイルス叢を網羅的に解析する技法を本研究で開発したが、この様なアプローチは蚊など他の節足動物が媒介する病原体の検索にも応用できる手法である。さらに、これらの微生物がヒトを含む哺乳動物に病原性を発揮するかどうかについては、動物接種実験等を実施しなければならないと考えられるが、遺伝子情報を大規模に収集し、データベース化しておけば、新興感染症の出現予測や何らかの未知の感染症が発生した場合、病原体の迅速同定や診断法の迅速な開発に役立てることができよう。