THE UNIVERSITY
of LIVERPOOL

The Impact of Symbiosis on Arthropods – a genomic perspective

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Aleksandra Yaroslavna Beliavskaia

SINGEK
training network

February 2022

Abstract

"The Impact of Symbiosis on Arthropods – a genomic perspective"

by Aleksandra Beliavskaia

Vector-borne diseases (VBDs) have always posed a very serious threat to humanity. Malaria, Lyme disease, leishmaniasis, dengue, plague, and many others are killing millions and incapacitating tens of millions each year. While some of these dreadful maladies were drastically reduced by means of vaccination, vector, natural reservoir control, and effective antibiotic treatments, other VBDs still pose a major threat for many countries due to their sophisticated capabilities to evade immune responses and develop resistance to treatments. Several proof-of-concept studies have shown that VBDs can be effectively controlled by breaking the transmission cycle between the vector and the vertebrate host. Some of the early methods are quite primitive and disruptive to the environment - such as using broad-spectrum pesticides. Luckily, many years and efforts of scientific research allowed for more advanced approaches of vector control that are more targeted and less harmful for biocenoses. The work presented in this thesis contributes to developing methods based on advanced technologies to study host-parasite interactions and work towards further VBDs prevention and elimination.

Chapter II of the thesis is dedicated to developing a method that would allow screening of arthropod populations in an effective and cheap manner. Most current methods for detection of symbiotic microeukaryotes (protists) in *Arthropoda* and other metazoan hosts rely on targeted PCR-based analysis, such as amplification of barcoding genes of particular species or genera. Such approach is effective for revealing certain well-known pathogens, but overlooks the unknown members of symbiotic communities. This chapter provides the analysis of different approaches for an integral analysis of protistan communities based on ribosomal genes.

Chapter III describes the sequencing, assembly, and analysis of two high-quality genomes of important blood-feeding vectors: tsetse fly *Glossina morsitans* and tropical bont tick *Amblyomma variegatum*. Additionally, genomes of symbiotic bacteria present in the two species were assembled and characterised. These assemblies were used to develop a computational approach to distinguish between living symbiotic bacteria and bacterial insertions into host nuclear genomes.

Chapters IV and V describe the study of interactions of ticks and bacterial symbionts using several established tick cell lines as a model of host-symbiont interaction. Tick cells were infected with several symbiotic bacteria to understand their transcriptomic response to various infection agents.

The genomic and transcriptomic data generated in this study provide a valuable resource for further studies of the vector biology and vector-pathogen interactions.

Dedication

This manuscript is dedicated to my father who sacrificed his scientific talents and ambitions to the prosperity of our family during tough times.

"I am utterly convinced that Science and Peace will triumph over Ignorance and War, that nations will eventually unite, not to destroy, but to edify, and that the future will belong to those who have done the most for the sake of suffering humanity."

Louis Pasteur

As quoted in
"Louis Pasteur, Free Lance of Science" (1950)
by René Jules Dubos

Acknowledgements

There are many people without whom this thesis would not have been possible. The list is long and I am truly grateful to each and every one.

First of all, I would like to express my gratitude to my supervisors, Prof Alistair Darby, Prof Benjamin Makepeace and Prof Steven Paterson for continuous support throughout my PhD. I am grateful for their constant guidance, knowledge and vast expertise, for their patience, kindness and support throughout my PhD journey.

I would additionally like to thank Dr Lesley Bell-Sakyi for being my informal supervisor and her kind and valuable participation in my journey and Dr Lee Haines for many tsetse flies and fun facts about their biology and her welcoming attitude.

Thank you to all the members of the Darby Lab: Sam, Mark, Lauren, Amber, Stefany, Gwen, Christina. Our office has been a consistent place of lively chat and mutual support, and I missed it so much during the last two years of a constant covid lockdown. Thanks, too, must go to the members of CGR, Margaret for sharing her outstanding experience in NGS and single-cell technique, Richard for bioinformatics help and cookies in the office.

This project was supported by generous funding and useful workshops from the SINGEK training network within the European Union's Horizon 2020 research and innovation programme.

I want to thank my parents, who always supported my interests and life choices and my "scientific mom" Dr Maria Rautian who played a very important role in the development of my scientific views. My life-long friends, Nastia, Natasha, Kostia provided me with support and discussions which helped me on the way.

This thesis would not have been possible without constant and all-encompassing love and support from my dearest husband.

Contents

Table of Figures

Table of Tables

List of Abbreviations

CI	cytoplasmic incompatibility
DPO	double-priming oligo
eLSU	eukaryotic LSU
ETS	external transcribed spacer
GEMs	Gel Bead-In Emulsions
HGT	Horizontal Gene Transfer
IGR	intergenic region
ITS	internal transcribed spacer
LNA	locked nucleic acid
LSTM	Liverpool School of Tropical Medicine
LSU	rRNA large subunit rRNA (23S or 28S)
NTS	nontranscribed spacer
ONT	Oxford Nanopore Technologies
PacBio	Pacific Biosciences
PNA	peptide nucleic acid
scRNA-seq	single-cell RNA sequencing
SCRV	Saint Croix River virus
SIT	sterile insect technique
SSU	rRNA small subunit rRNA (16S or 18S)
TCB	Tick Cell Biobank
UMIs	Unique Molecular Identifiers
VBDs	vector-borne diseases
WGA	whole genome amplification
WHO	World Health Organization

The term “symbiosis” in this thesis is used in its original meaning as it was introduced by Heinrich Anton de Bary: “the living together of unlike organisms” (de Bary 1879). Here, the term does not imply the mutualistic relationship.

1. Chapter I General introduction

1.1. The evolutionary importance of symbiosis

Symbiosis is one of the most important phenomena which drives evolution in many ways. It has played a crucial role in the evolution of life on our planet, from the origin of eukaryotic cell to the more recent and ongoing formation of new symbiotic systems (Douglas 2010). There are symbiotic systems which historically drove a lot of attention and are well-studied, but most of symbiotic partnerships are less-known (McKenna *et al.* 2021).

First of all, the knowledge on diversity of symbionts is often limited due to uneven distribution and gaps in research, many ecological niches are largely unstudied in this regard (Lewin *et al.* 2018). Better understanding of the diversity and distribution patterns of symbionts can answer many important questions such as how symbionts affect community biodiversity and consistency, how can these patterns change under the influence of global climate change and other big ecological shifts, what are the dynamics of pathogen spreading and their effects on communities and human economy in new ecological surroundings (Frainer *et al.* 2018; Bass *et al.* 2023). Evolution happen on a species level (Lewin *et al.* 2018) so the more detailed and ungapped is the knowledge about diversity of symbionts and their related species, the more insights could be inferred from the genomic information. For example, the evolutionary processes leading to pathogenicity could only be fully understood with the analysis of non-pathogenic sister taxa. Another issue which requires a study of yet unknown organisms is the resolving of the Tree of Life phylogeny which for now lacks a plethora of species and thus many branches cannot be resolved (F. Burki 2014; Archibald 2015).

Secondly, very few symbiotic systems are thoroughly studied from the point of view of interactions between participants, most of the mechanisms of symbiosis

formations are inferred from a scarce number of model organisms (Ruby 2008). This is partly due to the fact that studying symbiotic relationships can be challenging, particularly in complex natural ecosystems where many different species are interacting with one another. Also, symbiotic relationships can be difficult to study experimentally, as interactions between organisms are often difficult to replicate in a laboratory setting. Advances in genomic sequencing and other high-throughput technologies have enabled researchers to study symbiotic systems at a much larger scale, allowing for the exploration of previously understudied systems.

While model systems have advanced the understanding of symbiotic relationships, they may not always be representative of the diversity of symbiotic systems that exist in nature. Different approaches are being developed to gain a less biased picture of the diversity and distribution of symbionts (Bass *et al.* 2023; Hadziavdic *et al.* 2014; Bradley, Pinto, and Guest 2016; Hugerth *et al.* 2014; Wang *et al.* 2014; Bower *et al.* 2004; Vestheim and Jarman 2008; Belda *et al.* 2017), although none of these methods is effective and unbiased with symbiotic systems where the host genetic material outweighs symbiont DNA by several orders of magnitude (Bass *et al.* 2023; Schneider *et al.* 2014). This issue poses a significant obstacle in studying of diversity of eukaryotic symbionts.

When examining the history of molecular biology, it becomes apparent that the power of theory to explain and predict outcomes decreases as the complexity of the problem increases (Fry 2016). In such cases, experimental data and observations may become more dominant and necessary for a complete understanding of the problem (Fry 2016). Molecular and genomic approaches have already revolutionised the research of symbiosis, and the complexity and depth of modern research of symbiotic interactions requires advanced technologies and further development of new approaches (Ruby 2008). This thesis aims to contribute a number of specific methods that could provide answers to some concrete questions in the field of symbiosis.

1.2. The burden of vector-borne diseases

This thesis is mostly focused on interactions of vectors and different pathogens directly affecting human health and economy. Vector-borne diseases (VBDs) account for approximately 17% of all infectious diseases and are collectively responsible for more than 700,000 human deaths per year, according to WHO (World Health Organization 2020). Most VBDs are transmitted by blood-feeding arthropods, which ingest disease-producing microorganisms during a blood meal from an infected host and later inject it into a new host during a subsequent blood meal. Mosquitoes and malaria play the leading role in deaths from vector-borne disease, resulting in more than 400,000 fatalities a year, mostly in small children ('World Malaria Report 2020: 20 Years of Global Progress and Challenges' 2020). The problem is relevant not only for human health; VBDs also seriously affect livestock (Takken *et al.* 2018) and pet animals (Maggi and Krämer 2019), resulting in substantial economic losses. VBDs threaten many endangered species and were even reported to cause extinction events (Rushmore, Bisanzio, and Gillespie 2017).

Many VBDs are on the rise despite all scientific and medical efforts (K. F. Smith *et al.* 2014). Lyme disease accounts for at least 20,000 cases each year, and this number is increasing as the pathogens spread into previously unaffected areas (Steere, Coburn, and Glickstein 2004; Polishchuk, Zdolnik, and Smetanin 2017; Stone, Tourand, and Brissette 2017). Several thousand cases of human ehrlichiosis have occurred since the first case was recognised in 1986 (Goddard 2008). Viral diseases transmitted by mosquitoes (dengue, Zika, chikungunya, yellow fever) constitute a problem for African, South American and South Asian countries and are expanding their geographic range (Ferreira, Fairlie, and Moreira 2020). Although malaria and sleeping sickness rates have decreased lately ('World Malaria Report 2020: 20 Years of Global Progress and Challenges' 2020), they are an immediate menace for many countries, and the situation has worsened due to the Covid19 pandemic, which put on hold or hampered non-Covid

research and disrupted usual control measures such as installation of insecticide-treated bed nets (Nghochuzie *et al.* 2020).

The ongoing epidemic of SARS-CoV-2 is an excellent example of how crucial it is to carry out surveillance for potential pathogens before they become a global threat. Virologists and epidemiologists often predicted that an epidemic such as Covid19 would happen, although their alarming messages never resulted in any practical measures (Global Preparedness Monitoring Board 2019; Jones *et al.* 2008). After the emergence of SARS-CoV in 2003, a group of scientists published a review of potentially dangerous coronaviruses and highlighted the importance of collecting data on wildlife sources of dormant pathogens and biosecurity in farms and wet markets, which can serve as the source of emerging infections (V. C. C. Cheng *et al.* 2007). Quite stunningly, the authors almost literally foretold the pandemic thirteen years before it started: "The presence of a large reservoir of SARS-CoV-like viruses in horseshoe bats, together with the culture of eating exotic mammals in southern China, is a time bomb". Although the origin of SARS-CoV-2 from the Wuhan wet market is sometimes doubted (Bloom *et al.* 2021; Segreto and Deigin 2021), it remains the most likely scenario, according to the WHO and numerous experts (Andersen *et al.* 2020; Barh *et al.* 2020; Latinne *et al.* 2020; Frutos, Gavotte, and Devaux 2021; Holmes *et al.* 2021). Unfortunately, humankind can hardly hope for this pandemic to be the last in the near future; on the contrary, scientists predict that emergence of infectious diseases might become more frequent due to escalating climate change (Brooks and Hoberg 2007; Mirsaeidi *et al.* 2016; Rocklov and Dubrow 2020; K. F. Smith *et al.* 2014).

Thus, systematic study of existing and potential pathogens is crucial for understanding pathogen transmission patterns and trends, preventing or limiting pathogen spread, and preparing for quick suppression of emerging diseases (Taylor, Latham, and Woolhouse 2001; Zhu *et al.* 2020).

1.2.1. Current state of disease and vector control

The complex life cycle of many vector-borne diseases and the fact that they are often not transmissible from human to human make it possible to prevent the spread of VBDs before they become a global health threat. Interventions that reduce human–vector contact and control of vector populations are now considered the most promising and safe approaches to eradicating VBDs (Aksoy *et al.* 2005; McGraw and O'Neill 2013). Traditional ways to control vectors, such as insecticides, traps and elimination of breeding sites, are not able to eradicate the diseases, often are not environmentally friendly, and need to be repeated regularly, making them costly and ineffective (Hill and Wikel 2005). Vectors are also known to develop resistance to the insecticides used for their control (Hemingway *et al.* 2016; Turner and Golder 1987). Furthermore, insecticides used to eliminate vectors are often broad-spectrum and therefore harmful for many other insects, which are critical components of ecological communities (Carson 1962). In one study, Hallmann and co-authors estimated that pesticides are responsible for a significant decline in insect populations in Germany, jeopardising plant pollination and insect-feeding birds and animals (Hallmann *et al.* 2017). Thus, there is a clear demand for alternative control methods that do not rely on chemical insecticides.

As an alternative to classical vector control measures, sophisticated biological techniques of varying efficacy have been introduced recently to reduce vector populations. Some of the earliest attempts included releasing insects with a reduced lifespan or sterile males (reviewed by McGraw and O'Neill, 2013). Researchers are also trying to apply natural enemies of vectors, such as parasitoid wasps, but this approach meets difficulties of scale and parasitoid pupae persistence and has shown mixed results in field conditions (Garros *et al.* 2018). Fungal pathogens have received some attention as biological enemies of vectors and a source of anti-arthropod toxins (Garros *et al.* 2018). An important advantage of such approaches is the reduced burden on the environment, as biological control agents are usually more host-specific than pesticides

(Garros *et al.* 2018). Another promising approach is altering the interactions of pathogens with their vectors, making vectors unable to acquire the pathogen (McGraw and O'Neill 2013). The World Mosquito Program modified the microbiome of mosquitoes to make them incapable of carrying dengue, Zika, chikungunya and yellow fever viruses (O'Neill 2018).

The rise of 'omics' technologies brought new hope to the field of vector control (Hill and Wikel 2005; Holt *et al.* 2002) and has already yielded some practical results. Revealing mechanisms and the molecular bases of host-vector-parasite interactions is expected to help identify novel targets for insecticides, vaccines and genetic manipulation of vectors and pathogens (Hill and Wikel 2005). Nowadays, the affordable cost of next-generation sequencing allows such massive genomic studies as i5K (obtaining five thousand *Arthropoda* genomes), the Darwin Tree of Life (sequencing of 70,000 eukaryotic species in the UK and Ireland) and Ag1000G (*Anopheles gambiae* 1000 Genomes Project). The Ag1000G project proved particularly useful by revealing the insecticide resistance mechanism in mosquitoes based on copy number variation (CNV) of several genes (Lucas *et al.* 2019). This insight is invaluable for screening populations for these mutations and better insecticide distribution and rotation and also provides researchers with targets for insecticide development (Lucas *et al.* 2019).

Vaccination might also help to fight VBDs but is unlikely to eradicate them. For example, there has been an effective vaccine against yellow fever since 1937, although the disease is still active due to supply issues and the abundance of natural reservoirs of the virus (Barrett 2017). Developing vaccines against malaria or trypanosomiases is highly challenging; many groups have been working on this task for decades, but there are still no effective vaccines on the market (McGraw and O'Neill 2013). Only one vaccine against malaria – RTS,S – has passed four phases of clinical trials and is now being piloted in three African countries (Adepoju 2019; Coelho *et al.* 2017). Although the vaccine is promising in preventing the disease, it does not provide

a life-long immunity (Adepoju 2019), which means that the success of the immunisation programme would be largely dependent on political and economic stability. A vaccine candidate against sleeping sickness has been published only recently (Autheman *et al.* 2021).

1.3. Potential sources of new and newly-emerging VBDs

1.3.1. Climate change and VBDs

The spread of VBDs is determined by many environmental and social factors such as global travel and trade, urbanisation, and environmental changes (Baylis 2017). These factors impact pathogen transmission, changing seasonality and infection paths, making them more or less intense or causing diseases to emerge in previously unaffected regions (Rocklov and Dubrow 2020). Habitat shifts caused by consequences of climate change such as fires, floods and droughts would make the transfer of pathogens from animals to humans easier (Brooks and Hoberg 2007; Baylis 2017). It has already been shown that suburbanisation contributed to the rise of many known VBDs (Goddard 2008).

The geographical distribution of vectors is tightly dependent on climate – mainly temperature and humidity (Sutherst 1998). Due to climate change, warmer zones are shifting towards the north in the Northern hemisphere, and vectors are following (Siraj *et al.* 2014). Change of vector geographical range combined with global tourism provides a perfect scenario for pathogen expansion into new populations (Baylis 2017). Warmer conditions can also increase vector population density and thus the chance of humans meeting vectors; this has happened with *Ixodes ricinus* in Europe and is arguably the reason for rising rates of Lyme disease (Lindgren, Talleklint, and Polfeldt 2000). It has been shown that species non-native for the UK, such as *Aedes albopictus,* were recorded in southern England during the last decade, increasing the risk of bringing dengue and chikungunya fever to the UK (Tafilaku and Bunn 2019).

Some indirect effects of climate change include human activities such as deforestation, claiming territories for agriculture, or new human settlements that also disrupt ecosystems and increase VBD emergence (Gubler 2011; Rocklov and Dubrow 2020). Sociodemographic and economic factors such as poverty, wars and poor healthcare systems result in a higher rate of VBDs in populations ('Global Vector Control Response 2017–2030' 2017).

1.4. Transformation of life cycle as a potential source of new VBDs

Parasites can adapt to new vectors and thus acquire new territories and new vertebrate species by modifying their life cycle. Trypanosomatids could serve as a good example of such flexibility; these parasites have a strong tendency for adapting their life cycle and expanding to new hosts and vectors. Phylogenetic reconstruction of the *Kinetoplastida* shows that trypanosomatids are monophyletic and evolved from a free-living heterokont into parasites of insects, and the majority of known diversity remains monoxenous (Frolov, Kostygov, and Yurchenko 2021). Only three trypanosomatid lineages developed a heteroxenous lifecycle with a vertebrate host (*Trypanosoma* and *Leishmania*) or plants (*Phytomonas*) (Lukes *et al.* 2018). This implies that acquiring a second host happened independently at least three times, and there are some preadaptations for incorporating a second host into their life cycle.

Trypanosomatids are able not only to shift from the ancestral monoxenous to heteroxenous life cycle but can also expand to other vectors and different vertebrate hosts, lose the insect host and become secondary monoxenous, and, as a result, develop new transmission strategies (Frolov, Malysheva, and Kostygov 2015). An example of a secondary monoxenous trypanosomatid is *Trypanosoma equiperdum.* This species is closely related to *Trypanosoma brucei* but has lost part of the maxicircles from the kinetoplast (mitochondrial) DNA and thus has lost the ability to develop in *Glossina* flies which requires full mitochondrial activity. The parasite becomes "locked" in the

vertebrate host population, where it can survive by relying on glycolysis (Lai *et al.* 2008). This life cycle change led to the geographical expansion of *T. equiperdum* beyond the tsetse belt (Lai *et al.* 2008). Such life cycle transformations might look like a dead-end from the evolutionary point of view, but they could be of serious concern for humans, especially considering that evolutionarily young host/pathogen systems often cause severe diseases due to lack of coadaptation (Ewald 1983). *T. evansi* has also evolved from *T. brucei* and lost maxicircles but has retained the heteroxenous life cycle using different vectors (*Tabanidae*). Due to the lack of necessary genes lost with maxicircle DNA, the pathogen can only be mechanically transmitted without biological development and reproduction. However, this case further demonstrates the capability of these pathogens to adapt to a wide range of combinations of vectors and hosts (Lai *et al.* 2008).

These examples demonstrate the importance of surveillance of natural habitats and vectors as potential sources of new VBDs. It is also important to keep track of vector distribution changes as new contacts between parasites and possible hosts are an obvious prerequisite for vector shift.

1.4.1. Changing vectorial capacity in *Arthropoda*

Vectorial capacity describes a vector's ability to spread disease among hosts and is dependent on interactions between host, pathogen and vector. Initially designed for mosquitoes and malaria, the classical vectorial capacity formula includes four parameters: ratio of vectors to humans or livestock, the length of parasite incubation period, vector survival through one day, and biting rates (Macdonald 1952). Models based on these parameters advised that shortening vector lifespan is the most effective intervention into the vector/pathogen/host interactions. This conclusion implied the broad use of insecticides and larval development site destruction (Macdonald 1952). Unfortunately, such formulae consider all vectors, all hosts, and all pathogens as having the same traits and infectivity, therefore not considering many factors that influence pathogen distribution (Brady *et al.* 2016). Nowadays, vectorial capacity estimations

are expanded to consider pathogen genetic heterogeneity, development of pesticide resistance in vectors, differences in vector immune response and effects of vector microbiota on parasites (Kramer and Ciota 2015). Such formulae should also include harm caused by a pathogen to a vector: decreased lifespan and fecundity (Kramer and Ciota 2015).

The effects of microbiota on vectorial capacity have recently drawn a lot of attention as a promising vector control tool. The most successful project – the World Mosquito Program – exploits the ability of *Wolbachia* symbionts to modify host longevity, fecundity and susceptibility to pathogenic viruses such as dengue and Zika (McGraw and O'Neill 2013). Pilot releases of mosquitoes infected with *Wolbachia* demonstrated a decrease in viral transmission and that the technology was self-maintaining due to *Wolbachia*-induced cytoplasmic incompatibility which favours infected over uninfected insects (O'Neill 2018). Most of these studies focus on the bacterial part of the vector microbiomes (Angleró-Rodríguez *et al.* 2017), although eukaryotic symbionts also can alter vector/pathogen interactions (Cansado-Utrilla *et al.* 2021). Thus, a microsporidian protist restricts *Plasmodium* from developing in mosquitoes (Herren *et al.* 2020), and a fungus from the *Ae. Aegypti* gut, on the contrary, enhances mosquito ability to transmit dengue virus (Angleró-Rodríguez *et al.* 2017).

Eukaryotic pathogens can also be affected by vector microbiota; it has been shown that the mosquito gut microbiome interferes with *Plasmodium* cells (Romoli and Gendrin 2018). There are many ways the microbiota is involved in anti-pathogen tolerance. First of all, the heavy growth of bacteria in the mosquito gut triggers non-specific immune response; secondly, some gut symbionts produce metabolites that hinder *Plasmodium* development, such as reactive oxygen species or toxins (Dennison *et al.* 2016; Valzano *et al.* 2016).

1.5. Diversity of eukaryotic symbionts associated with blood-feeding arthropods

The research on eukaryotic microorganisms causing VBDs focuses on several dozen species affecting human and livestock health. Most studies deal with the *Apicomplexa* group members *Plasmodium* spp., *Hepatozoon* spp., *Babesia* spp., *Theileria* spp., and the *Kinetoplastida* group members *Leishmania* spp. and *Trypanosoma* spp. It is noteworthy that the diversity of these genera is very high, including pathogens affecting species not of economic importance to humans. Many new species are discovered accidentally; for example, a new basal subfamily of trypanosomatids was found during the screening of *Culex* mosquitoes for West Nile virus (Van Dyken *et al.* 2006). It is not clear whether these new trypanosomatids are mono- or heteroxenous (Van Dyken *et al.* 2006), but taking into account that they infect blood-feeding mosquitoes, they have an ample chance to jump into mosquito prey, including humans. Modern PCR-based methods allow large-scale screening to uncover the as-yet-unknown diversity of these groups (Flegontov *et al.* 2013).

Environmental sequencing techniques have broadened our knowledge about microeukaryotic diversity, but it should be noted that these methods have some weaknesses, such as uneven amplification of different organisms. Some groups are well represented in sequence datasets obtained using the 18S rRNA gene as a marker; others are often missed (*Microsporidia*, some *Alveolata*, minor groups of protists such as *Malawimonada,* etc.) (Bass and del Campo 2020). Primers targeted at specific groups overcome these limitations, but they result in longer protocols, and there is a chance of missing unsuspected species. Another constraint in studying eukaryotic symbiont diversity is the ratio between host and symbiont genetic material in the sample, which can cause problems with detection of the latter during the PCR and produce false-negative results (Vestheim and Jarman 2008).

1.6. Aims and objectives

This study aims to develop effective and high-throughput instruments for studying host-parasite interactions, gathering insights into interactions between pathogens and vectors.

Chapter II describes the evaluation and development of different methods for fuller surveying of microeukaryotic diversity in multicellular hosts to address described weaknesses of screening wild populations is crucial for predicting pathogen distribution, vectorial capacity, and further understanding of the co-evolution of pathogens with their vectors.

The main aim of Chapter III is to develop a bioinformatic pipeline to differentiate between living symbiotic bacteria and the integration of symbiotic genomic fragments into the host genome. The computational approach to finding the symbiotic insertions into arthropod genomes should allow a wider studying of the phenomenon which might help to reveal evolutionary patterns of how genomes or parts of genomes of bacterial symbionts of *Arthropoda* have become incorporated into host chromosomes.

Chapters IV and V explore transcriptomic responses of tick cells to different bacterial pathogens. Tick cell lines are widely used to study their interactions with pathogens, although knowledge of tick immunity is limited and needs to be further investigated. Chapter IV describes the design of single-cell transcriptomic experiment with two different tick cell lines and three bacterial pathogens, and Chapter V explores the transcriptomic responses of different tick cells to pathogens and compares transcriptional patterns between different pathogens and two different tick species.

Chapter VI summarises the study results, discusses the constraints encountered and suggests how the findings could be applied in future research.

2. Chapter II Design and evaluation of different strategies for finding hidden symbiont biodiversity in *Arthropoda*

2.1. Introduction

The impact of microeukaryotes on animal health is as potentially relevant as the effect that prokaryotes have, but there is little knowledge about it. The vast majority of microbiome studies focus on prokaryotes: 95% of studies describe the diversity of bacteria, and only 5% on that of eukaryotes (Campo, Bass, and Keeling 2020). Such bias is observed not only because there are more pathogens among bacteria, and thus they historically attract more attention, but also the technical limitations of microbiome research methods (Bass and del Campo 2020). The most widely-used approach is metabarcoding with the SSU rRNA gene; Carl Woese proposed this method in the late 1970s and proved its utility to reveal the three main domains of life (Woese and Fox 1977). Sequencing of the 16S rRNA gene is still the primary method for reconstructing microbial taxonomy and phylogenetic trees; however, a price reduction of high-throughput sequencing technologies has recently given rise to whole-genome metagenomics (Hiraoka, Yang, and Iwasaki 2016).

The same approach – obtaining sequences of the 18S rRNA gene – is often applied to assessing eukaryotic diversity (Pawlowski *et al.* 2012); environmental sequencing techniques and technologies have contributed massively to our understanding of the microbial world (Bass and del Campo 2020; Bradley, Pinto, and Guest 2016; Hadziavdic *et al.* 2014; Hugerth *et al.* 2014; Wang *et al.* 2014). Unfortunately, the metabarcoding approach has many limitations, such as uneven amplification of organisms from diverse evolutionary branches (Pawlowski *et al.* 2012), varying speed of evolution of the 18S rRNA gene between different branches (Philippe

2000), and the lack of consensus opinion about standard metabarcoding primers within the scientific community (Pawlowski *et al.* 2012). These issues limit the application of universal eukaryotic primers in studying natural communities and make it almost impossible to apply the same approach to study the eukaryotic part of animal microbiomes or any community with a dominant species (Vestheim and Jarman 2008), as well as mitochondrial and plastid 16S rRNA genes which contaminate sequencing of prokaryotic symbiont communities (Lundberg *et al.* 2012). The primary constraint in studying symbiont diversity is the ratio between host and symbiont genetic material in the sample. Considering the small size of the pathogen population within a vector, there is always substantially more genetic material from host cells than from symbiont cells, which causes multiple problems with its detection during the PCR (Vestheim and Jarman 2008). Previous studies of eukaryotic parasites of arthropods also showed that while some pathogens are abundant in vector populations, others are very sparse; for example, *Trypanosoma cruzi* is present in up to 50% of specimens (Browne *et al.* 2017; Kjos, Snowden, and Olson 2009), while *Babesia* species are found in approximately 1% of the vector population (Diuk-Wasser *et al.* 2014; Oines *et al.* 2012; Onyiche *et al.* 2021).

Although culture-independent molecular techniques for revealing the microeukaryotic part of communities have massively broadened our knowledge in the field, it is necessary to develop sensitive methods to target the sequences of interest and avoid amplifying the abundant host DNA. This study aims to find the most efficient and accurate method for the populational screening of arthropods for microeukaryotes.

2.1.1. Structure of the ribosomal cistron in eukaryotes

The 18S rRNA gene is a popular metabarcoding marker for *Eukaryota* (Pawlowski *et al.* 2012). The 18S, 5.8S and 28S rRNAs usually lie together in a single transcription unit called a ribosomal cistron or the main transcription unit; the 5S rRNA gene is not usually linked to the ribosomal cistron and is transcribed separately.

Figure 1 illustrates a general model of rRNA genes based on metazoan models (frog, mouse, and fruit fly) (Torres-Machorro *et al.* 2010). The cistron is organised as head-to-tail repeats of rDNA units consisting of coding and intergenic regions (Figure 1a). Each copy includes genes of three subunits – 18S, 5.8S and 28S, with two internal transcribed spacers transcribed in one piece (Figure 1b) and later processed into separate SSU and eLSU subunits (Sollner-Webb and Mougey 1991). Similarly, 5S rDNA is also present as multiple repeats in tandems but separately from the 18S-5.8S-28S regions and is regulated by its own promoters (Torres-Machorro *et al.* 2010).

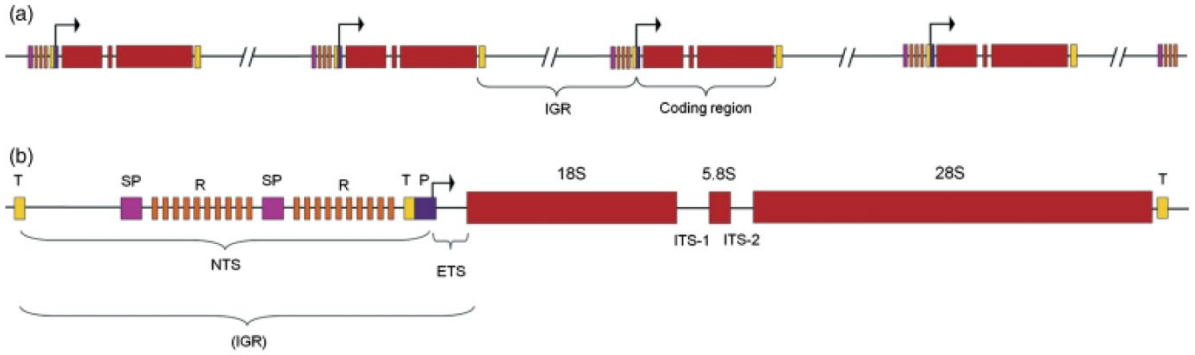

Figure 1 General organisation of ribosomal genes in Metazoa. (a) The ribosomal cistron with tandem repeats of 18S, 5.8S, 28S and IGR (intergenic region); (b) A single unit of the rDNA comprising NTS (nontranscribed spacer) and ETS (external transcribed spacer) and 18S, 5.8S and 28S subunits (from Torres-Machorro et al., 2010)

Considering that eukaryotes evolved as between five and eight independent branches according to different authors (Adl *et al.* 2012; Baldauf 2003; Fabien Burki *et al.* 2020), there is a considerable variation in the organisation and copy numbers of rRNA units among eukaryotes. Tandem repeats may be located on one or various chromosomes, linked or separated with 5S rDNA (Torres-Machorro *et al.* 2010). Cistron copy numbers can vary from a few copies to thousands (e.g., 4800 in green algae from the genus *Acetabularia* and 9000 in the ciliate *Tetrahymena*) (J. A. Eisen *et al.* 2006; Spring *et al.* 1978). Furthermore, the ratio of the main rDNA unit and the 5S gene may vary several-fold. The gene copy number is believed to be characteristic for each organism maintained at a constant level (Torres-Machorro *et al.* 2010).

Many microbial organisms from various phylogenetic branches share the typical organisation of the rRNA cistron illustrated in Figure 1; it can be found in all eukaryotic supergroups: in *Excavata* (e.g., *Leishmania*, *Trypanosoma*, *Giardi*a), *Amoebozoa* (e.g., *Acanthamoeba*), *Ophistokonta* (e.g., *Nosema*, *Schizosaccharomyces*, *Hansenula*), *Chromalveolata* (e.g., *Toxoplasma*, *Eimeria*) (Torres-Machorro *et al.* 2010). Many others have substantial variations. Unlinked and heterogeneous rDNAs are characteristic of the *Apicomplexa* group (Torres-Machorro *et al.* 2010). *Plasmodium* species have several types of rDNA units expressed during different life stages of the parasite: A-type is active in the vertebrate host during the asexual development, whilst S-type is active during the sexual phase in mosquitoes (Mercereau-Puijalon, Barale, and Bischoff 2002). A third O-type of the rDNA is found in *Plasmodium vivax* oocysts (J. Li *et al.* 1997). Another peculiar type of rDNA organisation is observed in *Microsporidia*: 22 copies of rDNA loci are located in telomeric regions of all 11 chromosomes in the obligate intracellular parasite *Encephalitozoon cuniculi* (Brugère *et al.* 2000).

In rare cases, the rDNA cistron is located in extrachromosomal molecules. *Dictyostelium discoideum* and *Physarum polycephalum* (both belonging to the *Amoebozoa* supergroup) have minichromosomes with several head-to-head repeats of rDNA genes and another copy of the same operon on chromosome IV (Sucgang *et al.* 2003). *Ciliates* such as *Euplotes crassus* and *Glaucoma chattoni* bear extrachromosomal rDNA copies as linear molecules in the macronucleus. *Paramecium tetraurelia* has rDNA molecules in both linear and circular forms (Torres-Machorro *et al.* 2010).

Typically, the SSU gene is approximately 2 kb long, although its size might measure up to 4.5 kb due to insertions into hypervariable regions, as shown for *some Kinetoplastida* and *Acanthamoeba* (Gunderson and Sogin 1986).

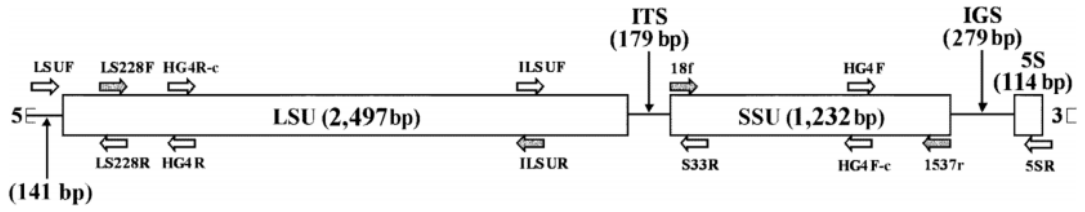

Figure 2 rDNA organisation in Nosema bombycis (Microsporidia). The 5.8S gene is linked with 23S rDNA, there is no ITS2, and the LSU fragment is upstream from the SSU (from Huang et al., 2004)

It is essential to consider the heterogeneity of organisation, location and size of rDNA genes in different microeukaryotes when designing universal or group-specific primers. One has to keep in mind the limitation of PCR product length, which might vary in size significantly due to insertions. The relative orientation of primers should be checked so that they elongate the DNA strand towards each other, which might be not the case in some genera such as *Nosema* (*Microsporidia*), where the LSU lies upstream from the SSU rRNA (Figure 2) (W.-F. Huang *et al.* 2004) or in circular minichromosomes of *Entamoeba histolytica* (*Amoebozoa*) or *Euglena gracilis* (*Excavata*) (Gunderson and Sogin 1986; Loftus *et al.* 2005).

2.1.2. Overview of existing metabarcoding methods

2.1.2.1. Overview of universal eukaryotic primers

The most commonly used markers for metabarcoding purposes are parts of the ribosomal RNA genes. Many sets of universal eukaryotic primers are designed to capture as many organisms from the environment as possible (Bradley, Pinto, and Guest 2016; Hadziavdic *et al.* 2014; Hugerth *et al.* 2014; Wang *et al.* 2014). Most primers used in metagenomic studies produce fragments of either approximately 500 bp for 454 sequencing or less than 200 bp for Illumina. Such limitations in size make the design of universal primers a tricky task as they should generate minimally biased, phylogenetically discriminating PCR products covering a wide range of species (Hugerth *et al.* 2014). The 18S rRNA gene has nine variable regions with conservative

regions in between (Figure 3). The ideal pair of primers should anneal within two conservative regions with variable sequence between the primers. Various authors came to different conclusions about the region that best represents the diversity of eukaryotes. Bradley and co-authors stated that the V8-V9 region provided the highest accuracy in their mock community, and Hugerth *et al.* suggested that V4 and V5 fragments are the most information-rich (Bradley, Pinto, and Guest 2016; Hugerth *et al.* 2014).

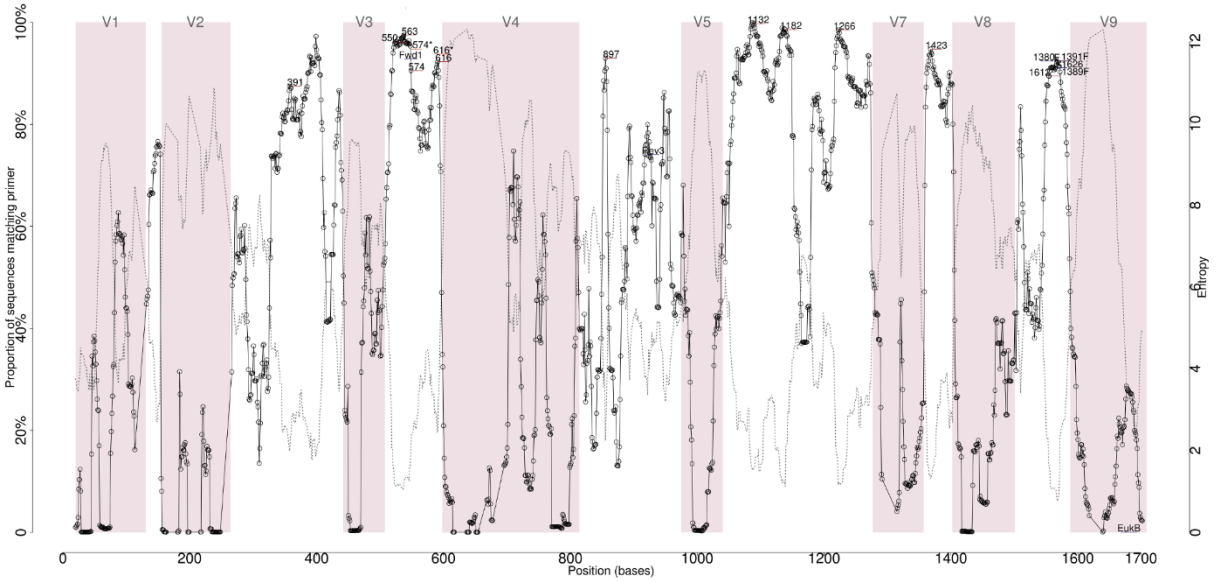

Figure 3 The schematic representation of variable and conservative regions of the 18S rRNA gene and candidate universal primers' position. The dotted grey line represents the entropy; the proportion of sequences matching primers is shown with circles connected with the black line. (from Hugerth et al., 2014)

Unfortunately, no primer pair covers all known phylogenetic branches of eukaryotes evenly. For example, *Fungi*, *Diplomonadida*, and *Parabasalia* are the groups that usually escape such assays (Wang *et al.* 2014). The use of a short fragment of the gene leads to inaccurate taxonomic assignment since it does not allow the assignment of a read to a particular species. Hugerth *et al.* stated that read assignment is only reliable at the genus level or higher (Hugerth *et al.* 2014). For example, the SSU rRNA has a low species-level resolution in *Fungi* and is rarely used to assess fungal diversity; researchers traditionally use more variable internal transcribed spacers (ITS1 and ITS2) (Schoch *et al.* 2012). Different SSU rRNA regions have different variability

between taxonomic groups, and the taxonomic composition of the sample hugely depends on the primers used. Several authors tested the same communities with primers for V4 and V8-9 regions separately, and the results between the two regions disagreed substantially (Bradley, Pinto, and Guest 2016; Stoeck *et al.* 2010). Bradley *et al.* showed that the taxonomic composition was dissimilar at the genus level and comparable only at the family level (Bradley, Pinto, and Guest 2016). Stoeck *et al.* demonstrated a diverse number of clusters detected (3993 clusters with V4 primers and only 2633 clusters with V9 primers) and different taxonomic profiles of dinoflagellate families even though they used longer PCR-fragments compatible with 454 sequencing (Stoeck *et al.* 2010). Issues with the low resolution of short PCR products should gradually disappear with the rise of long-read sequencing technologies. At the same time, the newly-developed approaches also have some limitations, which are discussed below.

2.1.3. Eukaryotic primers biased against *Metazoa*

PCR-based methods do not detect minor templates since rare sequences tend to be lost in the early stages of the PCR (Kalle, Kubista, and Rensing 2014). Considering that symbiont genetic material might comprise only a small fraction of the total DNA in the sample, it should result in many false-negative results (del Campo *et al.* 2019). Bower *et al.* attempted to find amplifying primers biased against *Metazoa* (uNonMet) (Bower *et al.* 2004). The forward primer they used was universal, while the reverse primer provides the specificity against *Metazoa* (Bower *et al.* 2004). The reverse primer (18s-EUK-1134-R) has several mismatches with the metazoan 18S rRNA gene, significantly reducing the metazoan template amplification (Bower *et al.* 2004). uNonMet primers recovered only 2.6% of the metazoan 18S rRNA reads present in the SILVA 132 RefNR database with fewer than 1% of reads from bilaterians in the *in silico* evaluation (del Campo *et al.* 2019). Del Campo *et al.* concluded that these primers could assess most animals' symbiont diversity, except sponges and perhaps ctenophores. However, it is worth mentioning that they tested these primers *in vitro*

with only ten genera from *Ctenophora*, *Cnidaria* (corals) and *Bilateria* (humans) (del Campo *et al.* 2019). Bower *et al.* initially tested the primers *in vitro* using 12 genera of marine metazoans (Bower *et al.* 2004).

These primers were initially designed for downstream analysis with 454 sequencing, and amplify a fragment of approximately 600 nucleotides (Bower *et al.* 2004). Most of the metabarcoding studies use Illumina sequencing, as it is cheaper and has a higher throughput, while 454 technology has been discontinued (Slatko, Gardner, and Ausubel 2018). The length of six hundred nucleotides is at the high end of the limit of most devices based on Illumina technology, and the ends of the amplicons tend to have a low quality which restricts the paired-end sequencing of longer fragments (Dohm *et al.* 2008); therefore, there is a need to perform a nested PCR to reduce the size of the amplicon and make it compatible with Illumina (del Campo *et al.* 2019). However, nested PCR carries a risk of increased contamination due to additional manipulation and might magnify the existing biases of the amplification (Marmiroli and Maestri 2007). Nowadays, the use of long-read technologies, such as Oxford Nanopore or Pacific Biosciences (PacBio), should solve the limitation in fragment size (Delgado *et al.* 2019; Seshadri *et al.* 2018).

2.1.4. Overview of approaches for blocking non-target templates

Another way to avoid dominant host DNA, and at the same time ensure that no potential symbiont is missed, is to use universal PCR primers accompanied by blocking oligonucleotides. The blocking oligonucleotides compete with the universal primers in a mix and block non-target DNA amplification. Blocking oligonucleotides can be DNA-based, locked nucleic acid (LNA)-based, peptide nucleic acid (PNA)-based, and Morpholino-based (Karkare and Bhatnagar 2006; Vestheim, Deagle, and Jarman 2011). Based on similar studies and guidelines (Belda *et al.* 2017; Leray *et al.* 2013; O'rorke, Lavery, and Jeffs 2012; Vestheim and Jarman 2008; von Wintzingerode *et al.* 2000) and due to the fact that LNA- and Morpholino-based oligonucleotides are

not easily available for custom synthesis, only DNA- and PNA-based were used in the present study.

2.1.4.1. DNA-based blockers

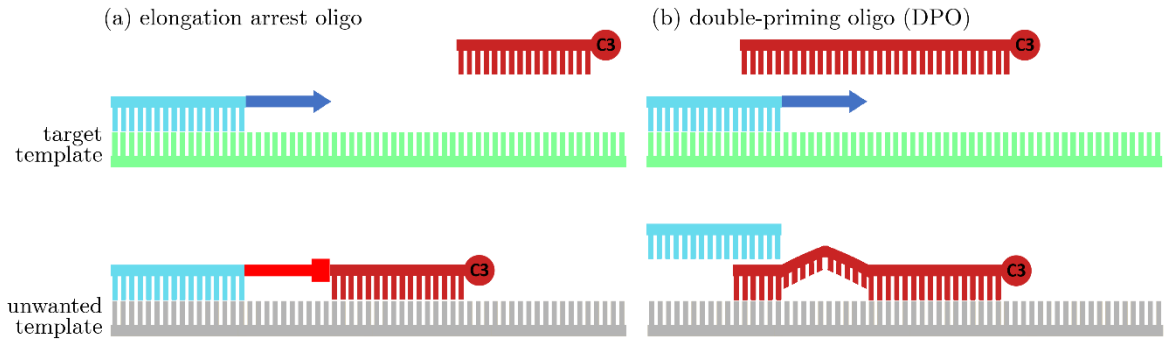

Figure 4 Schematic illustration of blocking oligonucleotide types. Amplifying primer shown in blue, blocking oligos shown in red.

DNA-based blockers are standard single-strand oligonucleotides with a C3-spacer at the 3'-end, which does not allow polymerases to extend and, therefore, aborts the PCR-product synthesis before reaching a reverse primer. C3-spacer modification is a short three-carbon chain attached to the oligonucleotide terminal 3' hydroxyl group, preventing the elongation during the PCR without noticeably influencing its annealing properties (Cradic *et al.* 2004; Vestheim and Jarman 2008). Figure 4a shows an elongation arrest blocker targeting some sequence pattern which is unique to the unwanted template between the amplifying primers, so that the DNA polymerase cannot extend the strand when it comes to the blocking oligo. Figure 4b shows a double-priming oligo (DPO) blocker: the start of a DPO overlaps with the amplifying primer and outcompetes it if the second part, unique to the unwanted template, finds the complementary sequence. DNA polymerase cannot use the DPO primer for extension. The design of both types of blockers requires a unique region within the amplified sequence to be blocked and a melting temperature of several degrees higher than that of the amplifying primers so that blocking primers would anneal first (Vestheim and Jarman 2008). The efficacy of this approach shows significant variability between various experiments: Vestheim *et al.* reported a 70% reduction of

a non-target template (Vestheim, Deagle, and Jarman 2011), while Belda *et al.* reported a similar approach to be ineffective (Belda *et al.* 2017).

2.1.4.2. PNA-based blockers

Peptide nucleic acid is an artificially synthesised polymer similar to DNA or RNA but with N-(2-aminoethyl)-glycine replacing the regular pentose-phosphate backbone, as shown in Figure 5 (Malhotra and Ali 2018). PNA-mediated PCR clamping relies on the two unique properties of PNA oligomers: PNA-DNA duplexes generally have higher thermal stability than the corresponding DNA-DNA duplexes, and DNA-polymerases do not recognise PNA oligomers and consequently do not amplify the blocked region (von Wintzingerode *et al.* 2000; Sforza *et al.* 2014).

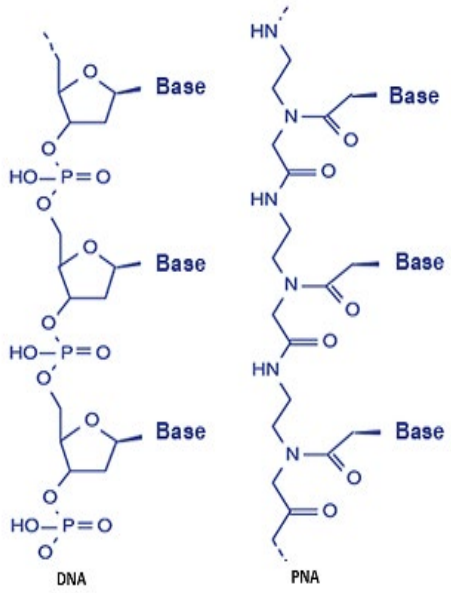

Figure 5 Chemical structure of DNA and PNA molecules (from Malhotra and Ali 2018)

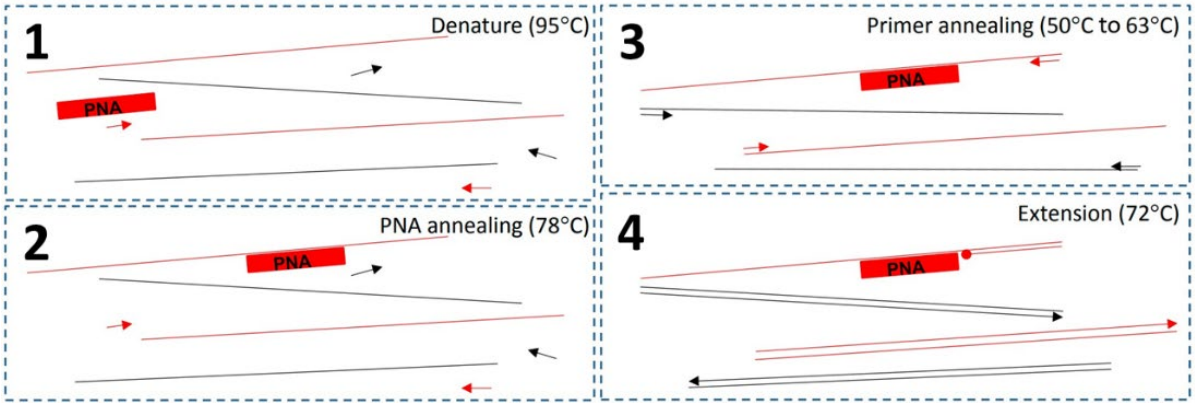

Figure 6 The principle of PCR with PNA blockers (adapted from Lundberg et al., 2013)

Figure 6 shows the principle of PCR reaction with PNA blockers. After denaturation, PNA anneals specifically to cognate templates via base pairing. Depending on the design, PNA either directly blocks primer annealing or blocks the elongation of the nascent strand (Lundberg *et al.* 2013). This approach was shown to be very effective in silencing the host DNA of *Anopheles* mosquitoes and allowed detection of symbiotic microeukaryotes, including *Plasmodium*, which was not possible without blocking or when using DNA-based blockers (Belda *et al.* 2017).

2.1.4.3. Removing non-target templates by restriction

Restriction enzymes can be applied before or after amplification. This method requires a unique cutting site in the unwanted sequence. Considering that restriction sites are usually short, spanning 4-8 nucleotides (Roberts 2005), finding a unique pattern for a target template is difficult. Furthermore, it would not work if the dominant DNA outweighed the minor templates totally in the PCR, resulting in loss of the latter (Vestheim and Jarman 2008); therefore, eliminating the unwanted DNA before amplification appears more promising. Green and Minz attempted to develop a method for enhancing amplification of the minor template using the Suicide polymerase Endonuclease restriction (SuPER) to cut only target DNA in a mix prior to PCR (Green and Minz 2005). This method has not achieved wide acceptance as it requires several extra handling steps and generally needs a unique restriction site specific to host DNA. The method can be augmented using a new CRISPR-Cas9 technology,

which allows choice of more extended target sites and thus is more precise (Jinek *et al.* 2012).

2.1.5. Aims and objectives

The aim of this chapter is to discover and create an effective and efficient method for detecting eukaryotic symbionts in *Arthropoda*. In order to achieve this goal, several objectives have been outlined. Firstly, a review of existing methods for silencing unwanted templates will be conducted. Secondary, experimental approaches for silencing unwanted templates will be tested. Finally, the sequences of the ribosomal genes will be analysed in detail, and primers biased against *Metazoa* will be developed.

2.2. Materials and methods

DNA extracted from different arthropod species (*Glossina* spp*., Ixodes ricinus, Lipoptena cervi, Ceratitis capitata, Macrosiphum euphorbiae*) available in the Darby lab were used for the initial testing, the full list of samples with the detailed information can be found in the Supplementary table 1.

Adult ticks (*Ixodes ricinus*) and deer keds (*Lipoptena cervi*) were collected in September 2017 from culled deer at Powys Castle, Wales, by kind permission of the National Trust. Ticks and keds were frozen at -20°C before DNA extraction.

Glossina morsitans flies bearing a putative eukaryotic symbiont thought to be a gregarine (Dr Lee Haines, personal communication) or *Trypanosoma brucei*, were provided by Dr Lee Haines of the Liverpool School of Tropical Medicine (LSTM). Some of the flies were fed on deer blood before DNA extraction; others were provided unfed after emerging. Some flies were homogenised in the lysis buffer using pestles prior to DNA extraction, other individuals were dissected, and only their internal organs were used for DNA extraction (full details can be found in Supplementary table 2). DNA was extracted by the following methods:

- DNeasy Blood & Tissue kit (Qiagen, UK) according to the manufacturer's protocol;
- Phenol-chloroform extraction (Sambrook, Fritsch, and Maniatis 1989);
- CTAB extraction protocol:
  - lysis in 2% CTAB buffer: 2% CTAB (hexadecyltrimethylammonium bromide), 100 mM Tris-HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl, 0.2% β-mercaptoethanol, 0.1 mg/mL proteinase K;
  - incubation at 60°C for 30 min;
  - phase separation, DNA precipitation and resuspension as in Phenol-chloroform extraction.

A female *G. morsitans* fly from the LSTM colony was dissected under a Zeiss laser microdissection microscope with PALM microbeam in the Single Cell Genomics Laboratory at the Centre for Genomic Research at the University of Liverpool, and ten cells of the putative eukaryotic symbiont were collected. These cells were used for DNA extraction with the DNeasy Blood & Tissue kit (Qiagen, UK) according to the manufacturer's protocol. The DNA was used for whole genome amplification (WGA) with a REPLI-g kit (Qiagen, UK) according to the manufacturer's protocol.

All extracted DNA was stored at -20°C. The full list of DNA samples used in this chapter can be found in Supplementary table 2.

To visualise the putative eukaryotic symbiont of *G. morsitans*, salivary glands and parts of midgut were dissected from several flies; dissected tissues were immediately placed on a microscopic slide in a small drop of sterile 1X PBS buffer and covered with a cover slip to flatten the tissues. Photomicrographs were taken using a Leica DM2500 camera with differential interference contrast (DIC).

2.2.1. Testing different sets of amplifying primers

2.2.1.1. Universal eukaryotic primers

PCR reactions were carried out in a volume of 20 µl containing 1µl of template DNA, 10 µl NEBNext HighFidelity Master Mix (New England Biolabs), 0.5µl of each primer at 10mM concentration and 8 µl PCR-clean water. Thermal cycling conditions were: 96°C for 5 min, 30 cycles of 96°C for 15 sec, 55°C for 15 sec, 72°C for 60 sec, final elongation at 72°C for 5 min. LongAmp *Taq* DNA polymerase (M0323S, New England Biolabs) was used with all the same parameters except for elongation temperature, which was 65°C. PCR with a gradient of temperatures from 45°C to 58°C was performed to find an optimal annealing temperature for EukA & EukB primers. Primer sequences are given in Table 2-1.

Table 2-1 Universal eukaryotic primers for the 18S rRNA gene used in the study

Primer	**5'-3' sequence**	**Reference**
EukA	AACCTGGTTGATCCTGCCAGT	(Medlin *et al.*, 1988)
EukB	TGATCCTTCTGCAGGTTCACCTAC	(Medlin *et al.*, 1988)
V4f	CCAGCASCYGCGGTAATTCC	(Bradley *et al.*, 2016)
V4r	ACTTTCGTTCTTGAT	(Bradley *et al.*, 2016)
V8f	ATAACAGGTCTGTGATGCCCT	(Bradley *et al.*, 2016)
V9r	CCTTCYGCAGGTTCACCTAC	(Bradley *et al.*, 2016)
F574	GCGGTAATTCCAGCTCCAA	(Hadziavdic *et al.*, 2014)
F1183	AATTTGACTCAACACGGG	(Hadziavdic *et al.*, 2014)
R1631	TACAAAGGGCAGGGACG	(Hadziavdic *et al.*, 2014)

PCR products were visualised by conventional electrophoresis in a 0.8% agarose gel with TAE 1X buffer (40 mM Tris-HCl pH 7.4, 20 mM sodium acetate, 1.0 mM Na2-EDTA) and stained with ethidium bromide. PCR products were purified with a

QIAquick Gel Extraction Kit (QIAGEN, Cat No. 28115) and quantified with the Qubit dsDNA High-Sensitivity assay (Life Technologies). Sanger sequencing was performed by a commercial company (Eurofins Genomics, Germany).

2.2.1.2. Primers biased against Metazoa (UNonMet)

PCR reactions were carried out in a volume of 20 µl containing 1µl of template DNA, 10 µl NEBNext 2x HighFidelity Master Mix (M0541S, New England Biolabs), 0.5µl of each primer at 10mM concentration and 8 µl PCR-clean water. Primers sequences were 5'-GTGCCAGCAGCCGCG-3' for the forward 18S-EUK581-F primer and 5'-TTTAAGTTTCAGCCTTGCG-3' for the reverse 18s-EUK-1134-R primer (Bower *et al.* 2004). Thermal cycling conditions were: 96°C for 5 min, 30 cycles of 96°C for 15 sec, 50°C for 15 sec, 72°C for 45 sec, final elongation at 72°C for 5 min. LongAmp *Taq* DNA polymerase (M0323S, New England Biolabs) was used with all the same parameters except for elongation temperature of 65°C. Gel electrophoresis, bands visualisation, extraction, and sequencing were performed as described in section 2.2.1.1.

2.2.2. Testing methods of blocking host DNA

2.2.2.1. Restriction-based approach

The online tool Restriction Mapper was used to find restriction sites within target sequences (*restrictionmapper.org*). The following sequences were used for restriction enzyme selection:

- Gregarines: FJ832160.1 *Gregarinidae* sp. from *Tubulanus polymorphus*, LN901450.1 *Lecudina tuzetae*, KR024702.1 *Lankesteria* sp., KJ736741.1 *Gregarina ormierei*, JX426618.1 *Trichotokara eunicae*, JX131300.1 *Ascogregarina taiwanensis*, FJ865355.1 *Psychodiella* sp., EU670240.1 *Lankesteria chelyosomae*;
- Tsetse fly: KC177312.1 *G. morsitans*.

Aliquots of 40 ng or 500 ng of total DNA were incubated at 37°C for 30 min with 1 µl of SmaI (ThermoFisher, FD0663) and/or 1 µl AanI (ThermoFisher, FD2064) restriction enzymes and inactivated at 65°C for 5 min. The restriction was followed by the PCR with EukA & EukB and V4f & V4r universal primers as described in 2.2.1.1.

2.2.2.2. DNA-blocking approach

DNA-blockers were designed based on the following sequences: KC177312.1 *G. morsitans*; AF322431.1 *Glossina palpalis*, AF322426.1 *L. cervi*, KC177300.1 *Ceratitis capitata*, KP322979.1 *Ustilago maydis*, XR_002989946.1 *Trypanosoma brucei*, so that blocking oligonucleotides were identical with the *Diptera* sequences and had at least three mismatches with *T. brucei* and *U. maydis*.

Two types of DNA-based blocking primers were designed as shown in Figure 7:

- V4-EA, elongation arrest blocker with C3 spacer on 3' end; CTTCATACGGGTAGTACAACTATA/3SpC3/
- V8-DPO, annealing inhibiting dual priming oligonucleotide, containing two separate priming regions joined by a polydeoxyinosine linker (ideoxyI); the left part is overlapping the universal primer, and the right part anneals to a tsetse-specific region and has a C3 spacer at the 3' end; CCCTTAGATGT/ideoxyI//ideoxyI//ideoxyI//ideoxyI//ideoxyI/CTAGACC GAGAG/3SpC3/

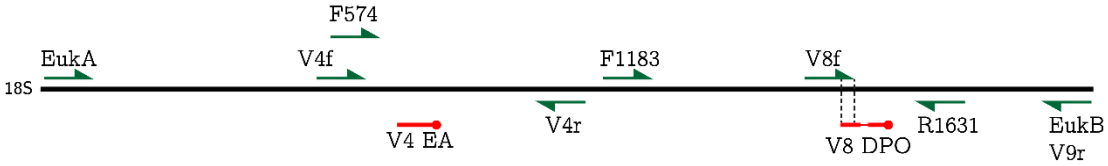

Figure 7 Schematic representation of universal amplifying primers and blocking oligonucleotides at the 18S rRNA gene. Green arrows show primers, red lines show blocking primers. Dashed lines between V8f and V8-DPO blocker show the overlap between them.

Blocking oligonucleotides were synthesised by a commercial company (Integrated DNA Technologies, Belgium). Seven different ratios of amplifying and blocking primers from 2X to 50X were tested. PCR reactions were used corresponding to the primers as described for amplifying primers in 2.2.1.1.

2.2.2.3. PNA-blocking approach

PNA oligonucleotides were designed based on the SILVA 132 NR99 database (Quast *et al.* 2013; Yilmaz *et al.* 2014). All prokaryotic sequences were removed, and all eukaryotic 18S rRNA sequences were divided into two groups – *Arthropoda* and all other *Eukaryota* except for *Metazoa* ("others"). All possible 15-mers were produced with jellyfish v. 2.1.3 (Marcais and Kingsford 2011) and aligned to two groups using bowtie2 with options --very-sensitive-local and --all (Langmead and Salzberg 2012). Samtools v.1.11 was used for sorting and filtering the alignments (H. Li *et al.* 2009). k-mers were sorted according to maximum matches in the "*Arthropoda*" group and minimum matches in the "others" group.

Melting temperature and quality control of PNA blockers were performed with the PNA tool from PNAbio (http://pnabio.com/support/PNA_Tool.htm). PNA oligonucleotides were synthesised by a commercial company (Biomers, Germany).

PNA blockers were tested in different concentrations: from 1X to 100X ratio to amplifying primers. Thermal cycling conditions were:

- (3-step PCR) 96°C for 5 min, 30 cycles of 96°C for 15 sec, 50°C for 15 sec, 72°C for 60 sec, final elongation at 72°C for 5 min;
- (4-step PCR) 96°C for 5 min, 30 cycles of 96°C for 15 sec, 65°C for 15 sec, 50°C for 15 sec, 72°C for 60 sec, final elongation at 72°C for 5 min.

Gel electrophoresis, band visualisation, extraction and sequencing were performed as described in the section 2.2.1.1.

2.2.3. Design of a new set of non-*Metazoa*n primers

Primers were designed with the assistance of Dr Alexander Predeus, Wellcome Sanger Institute, Hinxton, as follows. SSU database release 138 was used (Quast *et al.* 2013). The downloaded files contained 95286 sequences for the LSU gene and 510508 for the SSU gene; prokaryotic records were removed. Sequence names were standardised and parsed to reduce the impact of model organisms and make a unique set of species with one corresponding sequence. The dataset was divided into two parts: all *Metazoa* and "other eukaryotes". The tool fsm-lite was used to generate k-mers between 18 and 24 nucleotides (https://github.com/nvalimak/fsm-lite). The length of k-mers was chosen as the optimal for PCR primers. Pyseer GWAS suite was used to select k-mers associated with a group of non-*Metazoa*ns, and to filter them by occurrence frequency (Lees *et al.* 2018). The results were filtered by allele frequency greater than 0.3 and effect size lower than -1. The resulting selected k-mers overrepresented in non-*Metazoa* (688 for 18S and 824 for 28S) were aligned back to all 18S or 28S sequences using bowtie2 (Langmead and Salzberg 2012). The resulting alignments were overlapped to generate five distinct regions (3 regions in 18S and 2 regions in 28S). A consensus sequence for each region was extracted and mapped to all 18S and 28S sequences using the Smith-Waterman algorithm (water program from EMBOSS tools v6.6.0.0), in both forward and reverse mode (P. Rice, Longden, and Bleasby 2000). Alignments were then parsed using Perl script (parse_for_weblogo.pl) to extract the matching sequence from each reference, together with metazoan/non-metazoan assignment from the phenotype file and with taxonomy data. The resulting sequences were grouped into the following categories: *Bilateria*, other *Metazoa*, *Microsporidia*, other *Fungi*, *Choanoflagellata*, *Amoebozoa*, *Alveolata*, *Rhizaria*, *Stramenopila*, *Discicristatata*, *Metamonada*, *Archaeplastida*, *Ichthyosporea*, *Jakobida*, *Cryptista* and *Haptista*. Sequence consensus for each category was visualised using the weblogo package for R (Crooks *et al.* 2004).

2.2.4. Sequencing of PCR-products with Oxford Nanopore technology

Oxford Nanopore SQK-LSK109 kit with barcoding expansion kit EXP-NBD104 was used according to the manufacturer's protocol. Reads were basecalled and demultiplexed using ONT Guppy software v.2.3.5. Reads were taxonomically assigned using Centrifuge against the 'nt' database (Kim *et al.* 2016).

2.3. Results

2.3.1. Testing different sets of amplifying primers

2.3.1.1. Universal eukaryotic primers

Five pairs of universal primers were chosen from the literature and tested: EukA & EukB for the full-length 18S rDNA gene (Medlin *et al.* 1988), V4f & V4r and F574 & V4r for the V4 variable region, and V8f & V8r and F1183 & R1631 for the V8-V9 variable regions (Bradley *et al.*, 2016; Hadziavdic *et al.*, 2014). EukA & EukB amplify almost the whole SSU rRNA gene resulting in a 1800-2000 bp fragment. V4f & V4r, F574 & V4r, and V8f & V9r primers amplify approximately 400 bp, and F1183 & R1631 produce a 450 bp fragment.

Several species of tsetse flies from Burkina Faso (*G. morsitans, Glossina tabaniformis, Glossina palpalis gambiensis* and *Glossina medicorum*) were tested with EukA & EukB primers. Most of the tested samples gave only one band of approximately 2 kb, which corresponded to the length of the 18S gene; approximately one quarter of the samples showed more than one band on the gel, which might result from symbionts present in the samples. Figure 8 shows eight samples (details about the species of flies are available in Supplementary table 2), three of which resulted in the double band (indicated with yellow arrowheads). The observed bands were very close in size, so it was necessary to run the gels slowly at approximately 1 V/cm for 10-12 hours to separate them. Electrophoresis at a higher voltage did not allow the resolution of these bands, and they appeared as a single thick band. The bands which were visually distinct in the gel (TPf14, HSKm21, Hm17) were independently cut from

the gel; the PCR products were extracted and sent for Sanger sequencing, which revealed the presence of *Penicillium* sp. and *Trypanosoma* sp. in the samples.

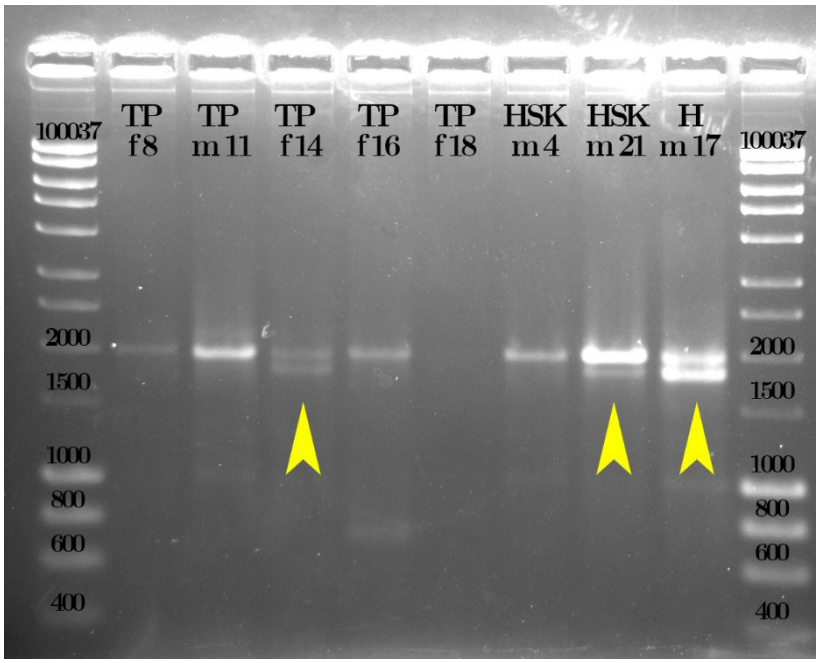

Figure 8 PCR products of eight samples of Glossina spp. amplified with universa eukaryotic primers EukA & EukB. HyperLadderI from BioLine was used. Samples indicated with yellow arrowheads gave a visible second band in addition to the main Glossina 18S band. The gel was run at 30V (approximately 1V/cm) for 10 hours in 0.7% agarose on 0.5X TBE.

2.3.1.2. Primers biased against metazoans

UNonMet primers should amplify a fragment of approximately 0.5 kb, which includes the V4 and V5 variable regions of the 18S rRNA gene (Figure 9). The size range of PCR products is higher than the theoretical expectation, spanning up to 2 kb.

Figure 9 Schematic representation of alignment of the UNonMet primers on the 18S rRNA gene.

Figure 10 shows PCR products from several insect samples: aphid *Macrosiphum euphorbiae* (Meuph), medfly *C. capitata* (Ccapi), deer ked *L. cervi* (Lcerv K10), and tsetse flies *G. morsitans* (Gmors W24), *Glossina medicorum* (Gmedi TPf8) and *G. palpalis* (Gpalp HSKf21 & Gpalp TPf14). Sanger sequencing revealed the following organisms: 1 – mixed products; 2 – *Myzus persicae*; 3 – *Myzus persicae*; 4 – *Penicillium* sp.; 5 – *C. capitata*; 6 – *Cutibacterium* sp.; 7 – *Propionibacterium* sp.; 8 – *Cutibaterium* sp.; 9 – *Dactyella sp.*; 10 – *L. cervi*; 11 – *Trypanosoma grayi*; 12 – *G. palpalis*; 13 – *Hepatozoon sp.*; 14 – *Glossina sp.*; 15 – *T. grayi*; 16 – *Penicillium sp.* The amplification resulted in multiple bands in all tested samples, including weak high-molecular bands, which could not be sequenced with the Sanger technology due to the small amount of material. These bands might be the result of non-specific annealing of the primers.

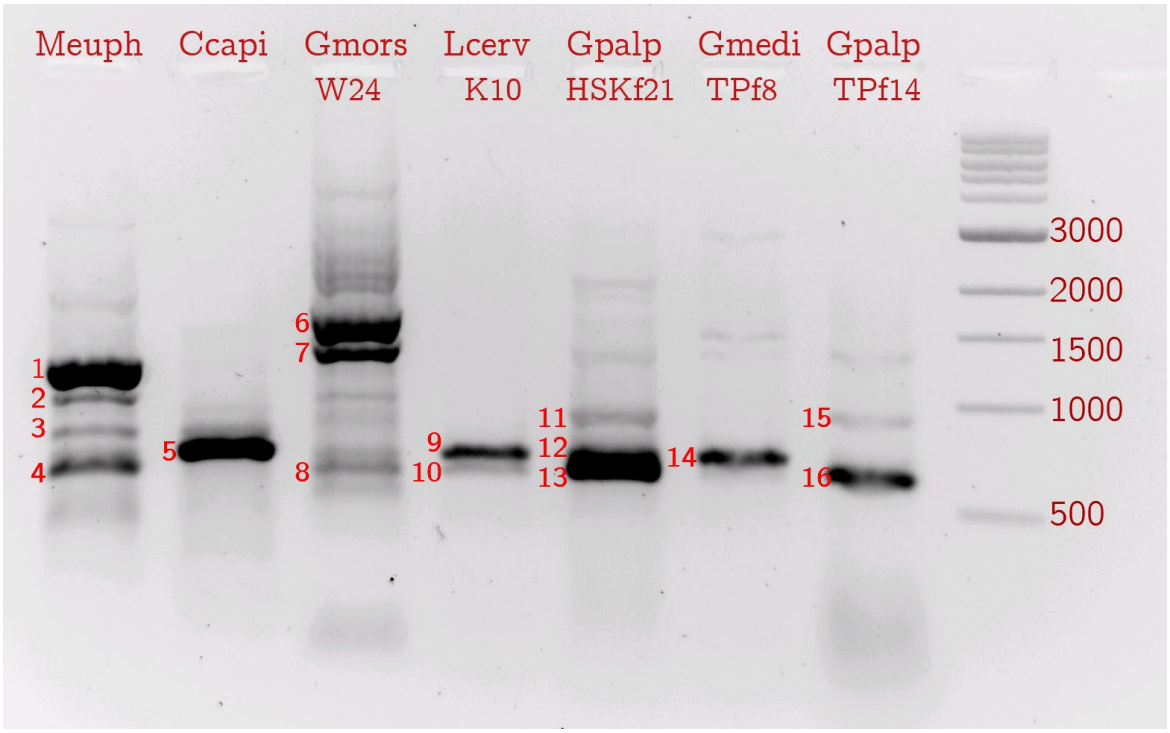

Figure 10 Results of the PCR with UNonmet primers. Quick-Load Purple 1 kb DNA Ladder from NEB is used. Numbered bands were extracted and sent for Sanger sequencing.

2.3.2. Testing methods of blocking host DNA

2.3.2.1. Restriction-based approach

The restriction enzymes *AanI* and *SmaI* were chosen; they target the 18S rRNA gene of *G. morsitans* and avoid several species of *Gregorina* (*Apicomplexa*). There are several possible restriction sites inside the 18S rRNA gene of *G. m. morsitans* (GenBank: KC177312.1), and only two do not have targets in the selected gregarine sequences.

Amplification of the tsetse fly DNA (flies from LSTM with the suspected symbiont of gregarine origin: G1, G2, G3, G6a, G13, G14, G16, details in Supplementary table 2) after restriction with one or both restriction enzymes yielded no products. This indicates that the restriction of the SSU rRNA gene templates was successful. There are two possible reasons for absence of symbiont products in the PCR: first, the concentration of the symbiont DNA could be below the level of PCR sensitivity, or there might be no symbionts present; second, as the restriction recognition sites are relatively short (6 nt), there is a chance that possible symbionts have the same pattern within their 18S rRNA genes as the tsetse flies and are also affected by the restriction.

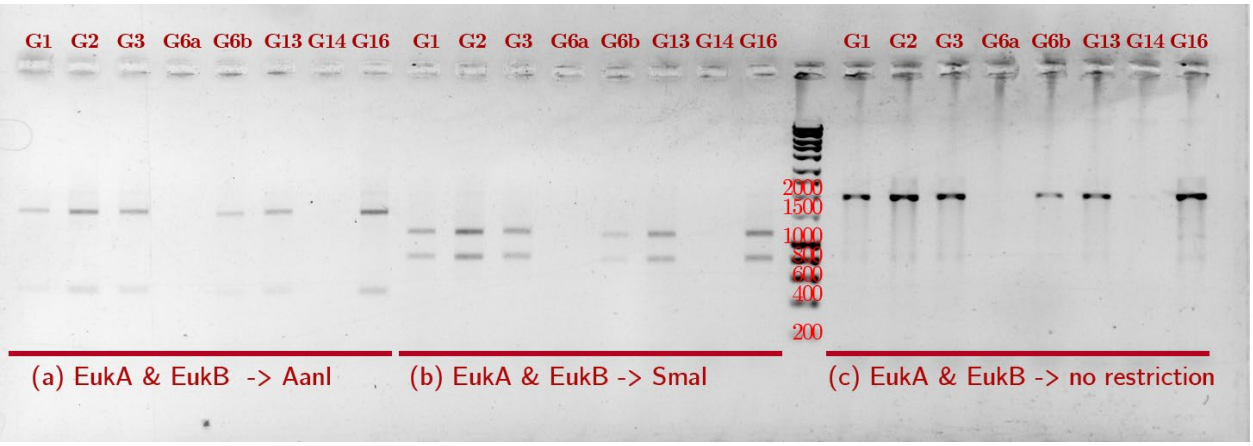

Figure 11 Results of the restriction of PCR products of EukA & EukB universal primers with AanI (a) and SmaI (b) enzymes applied to samples of Glossina morsitans from the LSTM. PCR-products without restriction shown in panel 'c'. HyperLadder I from BioLine was used. The description of the samples can be found in Supplementary table 1.

The opposite approach – cutting host templates after the PCR – was also not found to be effective. Figure 11 shows the results of cutting PCR products with two restriction enzymes, with the unmodified PCR products as a control. The size of PCR products after amplification with EukA & EukB primers is approximately 2 kb. Cutting PCR products with the *AanI* restriction enzyme resulted in 1500 and 500 bp fragments, and *SmaI* cut the products into 1200 and 800 kb fragments. The uniformity of the restricted bands in all tested samples most likely means that the PCR product is the same, namely the 18S rRNA of *G. morsitans*. PCR products of EukA & EukB primers produced fragments of a very similar length from different organisms, so two or more products might look like a single product, but the restriction should have amplified and revealed differences between heterogeneous sequences. Thus, we can conclude that there are no other sequences amplified apart from the host in this case. It could have resulted from absence of symbionts in the samples or a minimal amount of symbiont DNA, which was lost in the first rounds of the PCR.

2.3.2.2. DNA-blocking approach

The next approach was to block the amplification of host DNA with blocking primers, as shown in previous studies (Vestheim and Jarman 2008; Leray *et al.* 2013). Two types of blocking primers were designed: elongation arrest blocker and annealing inhibiting dual priming oligonucleotide. The elongation arrest blocker was designed to target the tsetse-specific region and ignore the SSU gene of symbionts. The dual priming oligonucleotide contains two separate priming regions joined by a polydeoxyinosine linker; its left part overlaps with the universal primer. The right part anneals to the tsetse-specific region and has the C3 spacer at the 3' end, which should arrest the elongation. Preliminary experiments showed this approach to be ineffective. The elongation arrest blocker designed for the V4 region of the SSU gene did not affect the outcome of PCRs. Figure 12 (a, b) shows PCR results with a template of DNA extracted from a tsetse fly infected with *T. brucei* and two pairs of primers for the V4 region and V4 elongation arrest blocker (V4-EA). All combinations of the primers and

oligonucleotides resulted in almost identical outcome – two bands corresponding to *G. morsitans* and *T. brucei*. There was no substantial or uniform reduction of the host signal.

The second blocker was designed to target the V8 region of the SSU gene as an annealing inhibiting dual priming oligonucleotide (V8-DPO). The V8-DPO primer seemed to block all amplification non-selectively. Figure 12 (c, d) shows PCR results with a template of DNA extracted from a tsetse fly and two pairs of primers (V8f & V9r and F1183 & R1631) for the V8/9 regions and V8-DPO blocker. Different amplifying and blocking primer ratios were tested, with ten concentration ratios from 2X to 50X, with the same result.

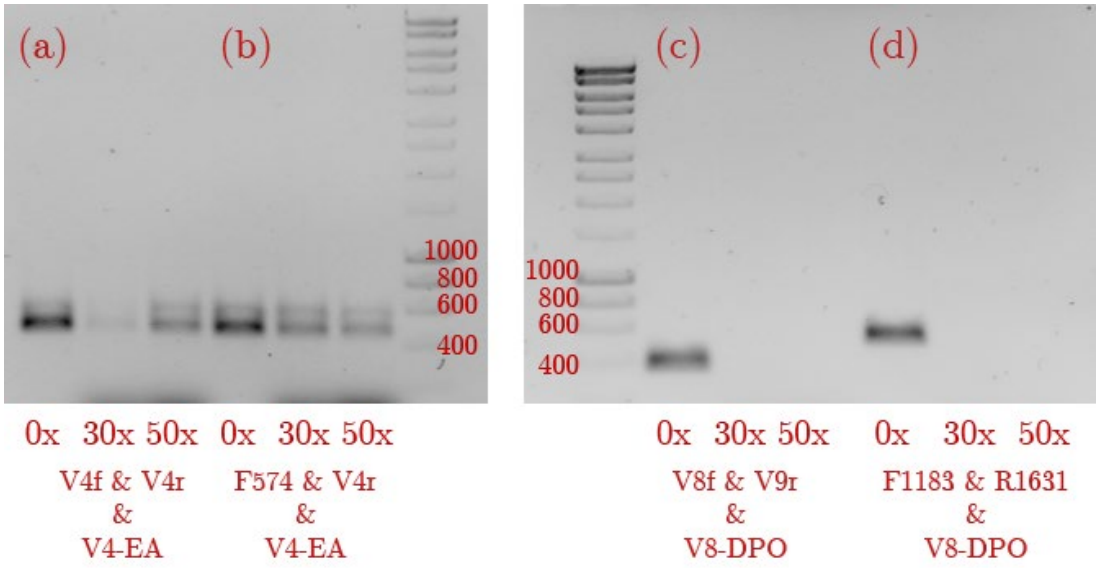

Figure 12 PCR amplification of sample W20 (Glossina morsitans with Trypanosoma brucei) with four pairs of primers and two DNA-blockers. All combinations were tested in 0X, 30X and 50X ratios of amplifying and blocking oligos. (a) amplifiying primers V4f & V4r with V4-EA blocker (b) primers F574 & V4r with V4-EA blocker; (c) primers V8f & V9r with V8-DPO blocker; (d) primers F1183 & R1631 with V8-DPO blocker.

2.3.2.3. PNA-blocking approach

Considering that PNA-based oligonucleotides are about 100 times more expensive than DNA-based oligonucleotides, blocking oligonucleotides were designed to be universal for most *Arthropoda* species (Table 2-2). Several short sequences were almost universal for arthropods (0 or 1 mismatch with up to 97.7% of arthropod species from the SILVA database) and almost non-existent in other eukaryotic lineages (approximately 6% of non-*Metazoa*n eukaryotic records in the SILVA database).

Table 2-2 PNA blockers designed for the study

Blocker	**Target**	**5'-3' sequence**	**DNA/PNA duplex melting temperature at 4µM, °C**	**Comments from the PNA-tool**	***Arthropoda* blocked, %**	**"others" blocked, %**
ART3	*Arthropoda*, V4 region	ATTCTTGGACCGTCG	72.1	Self-complementarity: Good	85.2	6.0
ART4	*Arthropoda*, V4 region	GCGGTATCTGATCGC	74.6	Self-complementarity: Good	97.7	6.9

Figure 13 shows the alignment of PNA blockers of species of interest available for testing in the lab: castor bean tick *I. ricinus*, deer ked *Lipoptena cervi*, medfly *Ceratitis capitata*, tsetse flies *G. morsitans* and *G. palpalis*, *Basidiomycota* fungi *U. maydis*, and *T. brucei*. Both blockers target the V4 variable region of the 18S rRNA gene and could be used with the following sets of universal amplifying primers:

- ART3: EukA & EukB, V4f & V4r, F574 & V4r
- ART4: EukA & EukB

Both blockers aligned perfectly to the target species and had more than three mismatches with non-arthropod species, which should be enough to reduce the annealing (Figure 14).

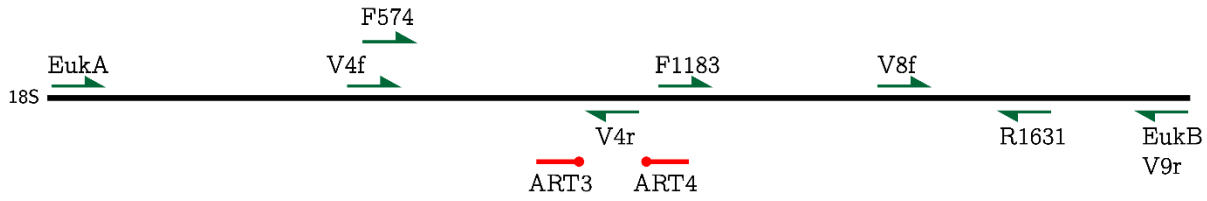

Figure 14 Schematic representation of PNA-blockers and amplifying primers for the 18S rRNA gene. Green arrows show primers, red lines show PNA blocking oligonucleotides.

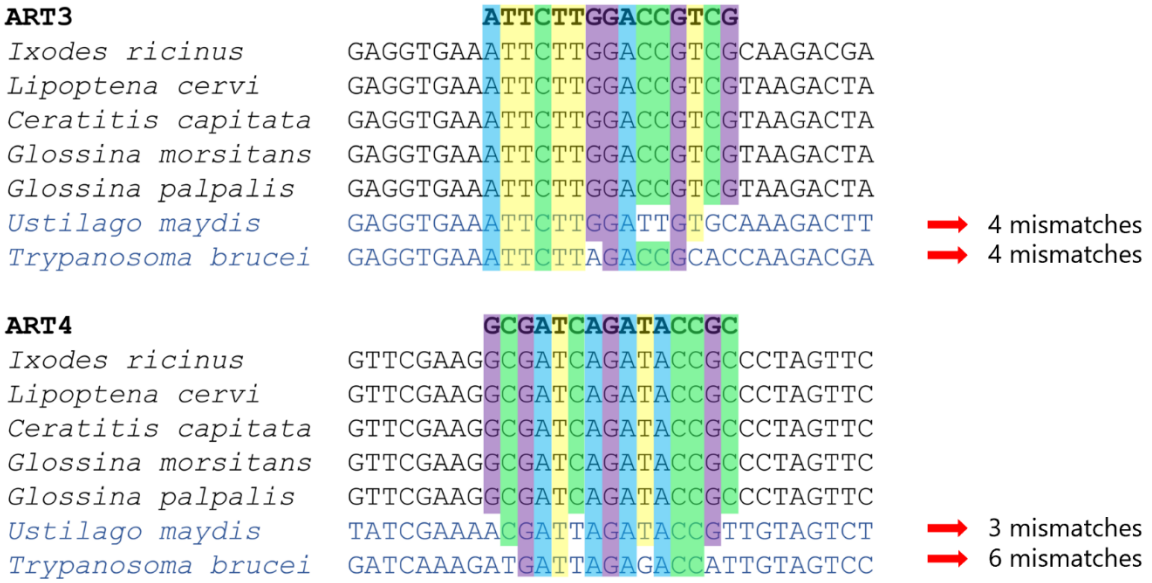

Figure 15 Alignment of PNA blockers with some arthropod and non-arthropod species of interest. Both blockers do not have mismatches with Arthropoda species and enough mismatches with non-arthropod species.

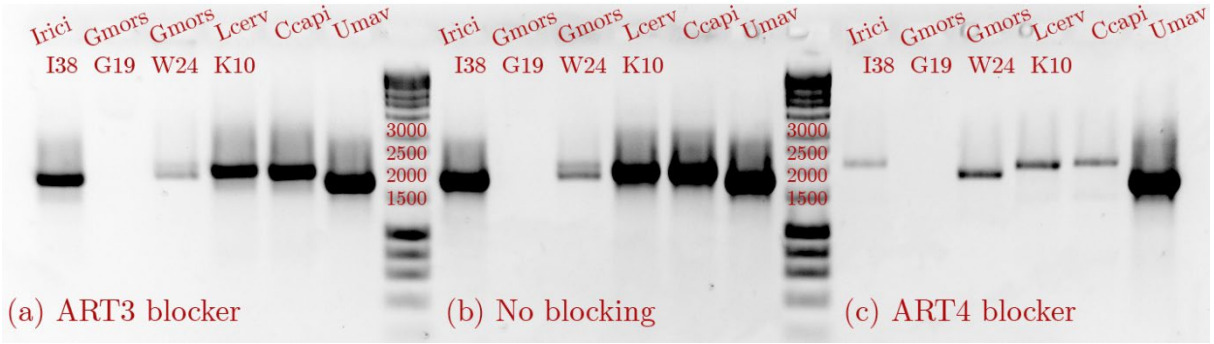

Figure 13 Results of PCR with universal eukaryotic primers EukA & EukB without blocking (b) and with ART3 blocker (a) and ART4 blocker (c). Umav (U. maydis, fungi) sample was used as a control

Both PNA blockers visually reduced arthropod bands on the gel. Figure 15 illustrates amplification of the 18S gene from five different samples of arthropods and one fungal sample as a control. There is a significant difference in band intensity for samples Irici I38, Gmors W24, Lcerv K10 and Ccapi between amplification without blocking and amplification with ART3 or ART4 blockers. According to the intensity of the bands, the ART4 blocker seems to be more effective than ART3, and in the case of Gmors W24, it eliminated the upper band, which came from *G. morsitans*. There was no difference in the amplification of the control fungal DNA (Umav). The G19 DNA sample did not yield bands in any condition, which is probably the result of DNA template degradation; the template could be amplified in previous experiments.

Three pairs of PCR products (without blocking and with ART4 blocker) were sequenced with long-read Oxford Nanopore technology to confirm the result. The top five taxa for each sample are given in Table 2-3. Samples Gmors W24 and Ccapi showed a significant reduction in reads of *Arthropoda* origin from 58% to 4% and from 7% to 1.5%, respectively. There was no decrease in arthropod reads in Lcerv K10, presumably because this sample did not contain any templates other than arthropod; less than 0.3% of reads belonged to four other taxa, which was below the threshold for ONT sequencing and may have been the result of high error rate and barcodes cross-talk. The decrease in band intensity was indirectly confirmed by decreasing the total number of reads: 12,798 reads without blocking and 9,881 reads when amplifying with ART4 blocker.

While these results seemed promising, in later experiments, PNA blockers produced very inconsistent results. For example, the same PCR mixtures with ART3 and ART4 blockers resulted in different bands depending on the particular PCR machine used or placement of the tube within the PCR machine (edge wells vs central wells).

Table 2-3 Sequencing results of PCR-products of samples Gmors W24, Lcerv K10 and Ccapi, without blocking and with ART4 blocker, showing the percentage of the total number of reads

	Gmors W24		**Lcerv K10**		**Ccapi**	
	no blocking	**ART4**	**no blocking**	**ART4**	**no blocking**	**ART4**
Arthropoda	58.01	4.11	98.88	98.72	7.03	1.56
Basidiomycota	21.18	46.25	-	-	-	-
Ascomycota	9.35	24.86	-	-	80.47	88.41
Mucoromycota	2.64	6.32	0.19	0.24	4.32	4.62
Apicomplexa	2.2	5.03	-	-	-	-
Proteobacteria	-	-	0.21	0.28	-	-

2.3.3. *In silico* evaluation of new primers biased against *Metazoa*

The biggest curated database of SSU and LSU rRNA genes for Eukaryotes is SILVA (Quast *et al.* 2013; Yilmaz *et al.* 2014). SILVA v138 was used for this analysis with 32270 sequences for the SSU gene and 9714 for the LSU gene after filtering and reducing redundancy. Three regions within the SSU rRNA gene and two regions within the LSU rRNA gene were considered promising as amplifying primers biased against *Metazoa*, or more specifically, *Bilateria* (Figure 16). Highlighted in yellow are positions where *Bilateria* has nucleotides different from all other eukaryotic groups. The summary of these positions is shown in Table 2-4. The following degenerate primers were designed based on the *in silico* analysis:

- SSU-rg1-F 5'-GGTGAAAWTMKYDGAYBH-3'
- SSU-rg2-F 5'- GGKASTAYGDHCGCAAG-3'
- LSU-rg1-R 5'-MSGWTCAWYCCKYDT-3'
- LSU-rg2-R 5'-RMARRTRTGCCKYCCCAGC-3'

Constraints of time precluded *in vitro* testing of these primers during the study.

Table 2-4 Summary of positions in the SSU and LSU rDNA genes discriminating Bilateria from other groups of eukaryotes

	Bilateria	**Other groups**	**Exceptions**
SSUrg1 nucl 3	T with minor G	C	G in *Microsporidia* and Haptista No consensus in *Metamonada* and *Amoebozoa*
SSUrg1 nucl 23	C	T	G in *Microsporidia* C in *Discicristata* and *Haptista* and some *Amoebozoa* and *Metamonada*
SSUrg1 nucl 24	G	T	A in *Microsporidia* and *Haptista* C in *Rhizaria* and some *Amoebozoa* G in *Discicristata*
SSUrg2 nucl 6	G/A	G	T in *Microsporidia*
SSUrg2 nucl 15	C/T	C	-
SSUrg3 nucl 1	G with minor A	mostly T	C in *Discicristata* G/A in *Haptista*
SSUrg3 nucl 2	G	A	G/A in other *Fungi* T/C in *Discicristata*
LSUrg1 nucl 21	C	A	-
LSUrg1 nucl 24	T	C	T in *Amoebozoa* A in *Alveolata* C/A in *Discicristata* No consensus in *Metamonada*
LSUrg1 nucl 35	A	G	A in *Amoebozoa* T in *Alveolata* No consensus in *Metamonada*
LSUrg2 nucl 10	A with minor G	G	-
LSUrg2 nucl 20	T with minor C	C	-

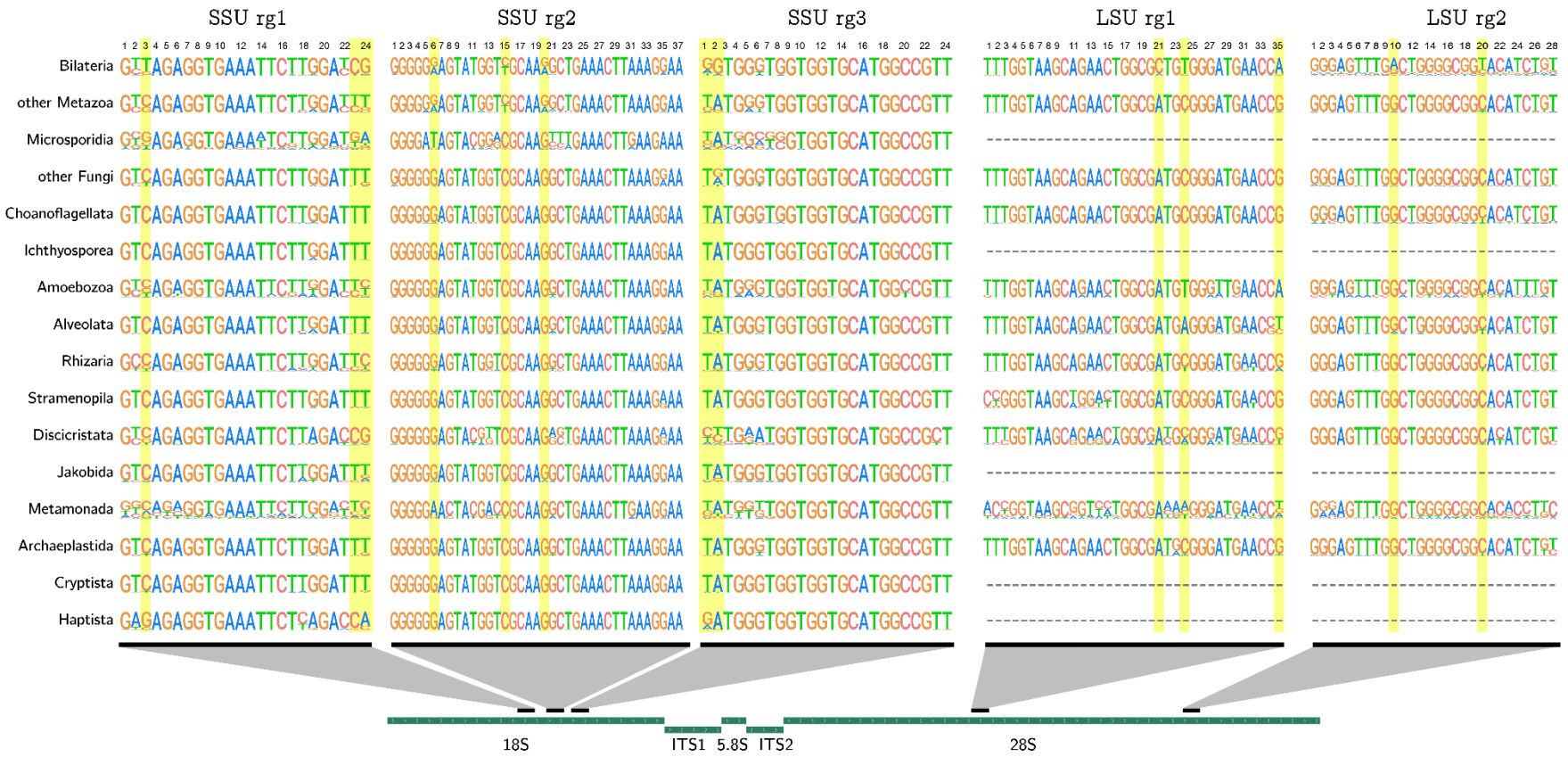

Figure 16 Consensus sequences for regions of interest for the SSU and LSU rRNA genes of the main groups of Eukaryotes. Positions where Bilateria has nucleotides different from all other groups of eukaryotes are highlighted in yellow.

2.4. Discussion

2.4.1. Restriction-based approach

Although the restriction of the non-target 18S rRNA gene of *G. morsitans* did not give a positive outcome, this result cannot be considered conclusive. The described experimental design was based on the assumption that *G. morsitans* samples from the LSTM colony carry a symbiont of supposedly apicomplexan origin (Dr Lee Haines, personal communication) (Figure 17). After several attempts to identify this symbiont, it is not certain that this assumption is reasonable, and the experiment should have been designed differently. Although there is a high probability of finding an applicable restriction enzyme for most host/symbiont systems, this method seems to have a limited scope. It is reasonable to design such an experiment if one deals with a large set of unified samples – looking for one or a few particular symbionts in one host when 18S rRNA gene sequences for them are known. If a researcher targets several symbionts, especially if some are obscure, results become unreliable. Considering the recent development of CRISPR-Cas9 technology that recognises patterns of approximately 20 bp (Nachmanson *et al.* 2018), that might be a more targeted and precise alternative to restriction.

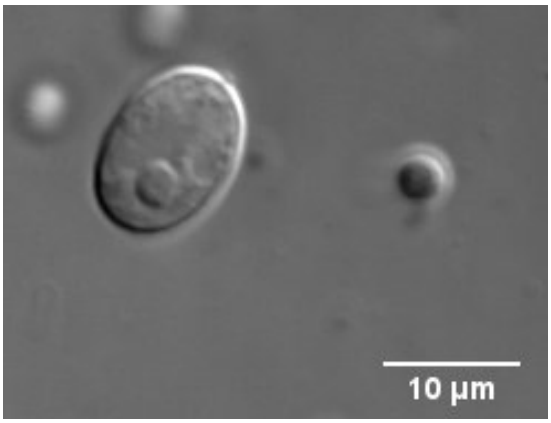

Figure 17 A cell seen in Glossina morsitans tissues that could be the putative eukaryotic symbiont of G. morsitans

2.4.2. Universal primers and UNonMet primers

Universal eukaryotic EukA & EukB primers produced single, double or triple bands in PCR products (Figure 8). Nevertheless, such approach does not seem practical. If the amount of different DNA templates differs by several orders of magnitude, as is the case in many symbiotic systems with a multicellular host, then minor templates tend to be lost during the first rounds of PCR (Kalle, Kubista, and Rensing 2014).

Non-*Metazoa*n primers UNonMet, designed by Bower *et al.* (Bower *et al.* 2004), resulted in higher variability of bands on the gel (Figure 10). Separate sequencing of UNonMet products revealed not only the host sequences but also DNA of minor symbionts, both prokaryotic and eukaryotic. Comparison of amplification of samples TPf8 and TPf14 with universal EukA & EukB (Figure 8) and UNonMet primers (Figure 10) showed similar results: one main band for the TPf8 and at least two bands for TPf14. However, the UNonMet PCR products were shorter and had a greater difference in size, making them more convenient for screening. On the other hand, the universal primers do not amplify prokaryotic sequences, which are not of interest when screening for eukaryotic symbionts.

2.4.3. Blocking approach

Two blocking methods were tested, DNA-based and PNA-based. Neither method yielded a satisfactory result. The DNA-blocker for the V4 region did not affect the PCR outcome, and the blocker for the V8 region non-specifically blocked all amplification. These results are indirectly confirmed in the recent article from Belda *et al.* on preferential suppression of *Anopheles gambiae* host sequences (Belda *et al.* 2017). The authors claimed ineffectiveness of DNA blocking primers using the example of *Anopheles* and its microbiome, and proposed another approach to outcompete host DNA with the help of peptide-nucleic acid (PNA) oligonucleotide blockers that

effectively reduced mosquito 18S rRNA gene sequences by more than 80% for the V4 hypervariable region (Belda *et al.* 2017).

Considering the high cost of PNA synthesis, it was decided to design blockers that can be used with any *Arthropoda* species. However, it was not possible to find a pattern that was unique and universal for all *Arthropoda* species. The best result was the ART4 blocker, which *in silico* blocked 97.7% of arthropod species from the SILVA database and only 6.9% of non-*Metazoa*n eukaryotic organisms (Table 2-2). The ART3 oligo should block a narrower range of arthropods (85.2%) and 6% of other eukaryotes.

Unfortunately, performance of PNA blockers was inconsistent, sometimes giving a noticeable reduction in a host DNA and completely failing at other times. Thus, this approach cannot be recommended as a method of choice.

2.4.4. Design of novel non-*Metazoa*n primers

2.4.4.1. Groups included in the analysis

The field of microeukaryotic phylogeny is rapidly changing nowadays, so it is impossible to find a definitive version of microeukaryotic taxonomy; thus, this analysis was based on the trees from Adl *et al.*, 2012 and Burki *et al.*, 2020 and the taxonomic assignment from SILVA which was not always consistent with the trees (Adl *et al.* 2012; Fabien Burki *et al.* 2020). The major supergroups included in the analysis were SAR (*Stramenopiles*, *Alveolates*, *Rhizaria*), *Archaeplastida*, *Excavata*, *Amoebozoa*, *Opisthokonta*, *Cryptista*, and *Haptista*.

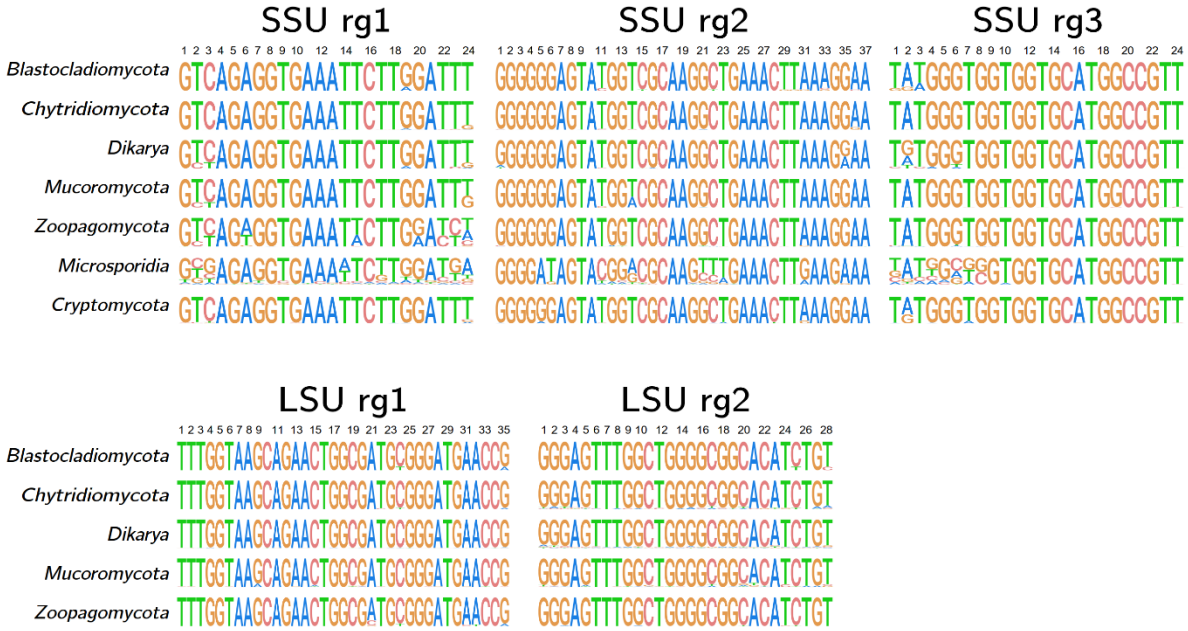

Figure 18 Alignment of the regions of interest for seven groups of Fungi. Only two groups – Microsporidia and Zoopagomycota – show noticeable variations within these sequences, as shown for the SSU. No 28S rRNA gene information for organisms from these two taxa is available in the SILVA database.

The supergroup *Opisthokonta* was later downgraded to be part of the *Amorphea* supergroup within the clade of *Obazoa* (Burki *et al.* 2020). However, considering that opisthokonts are disproportionately well-studied and include *Metazoa*, which is the group of particular interest in this study, it was included in the analysis not as one group but as six sub-groups. Another reason to examine the subdivisions of

Opisthokonta separately was that one of its branches – *Fungi* – is rich with diverse parasitic species. According to the SILVA database, the group *Fungi* has six branches: *Blastocladiomycota*, *Chytridiomycota*, *Dikarya*, *Zoopagomycota, Cryptomycota, and Microsporidia*. The first five branches have a very similar profile of the chosen regions of 18S and 28S rRNA genes (Figure 18). The only group of *Fungi* with a differing consensus of the SSU gene is *Microsporidia*, which is of great interest for the analysis as all major groups of animals host microsporidia, and they are very likely to be found in blood-feeding arthropods (Cali, Becnel, and Takvorian 2017).

According to SILVA and Adl *et al.* (2012), there are three clades within the SAR supergroup – *Stramenopiles*, *Alveolata* and *Rhizaria* (Adl *et al.*, 2012). Since 2020, the SAR supergroup has been united with *Telonemia* (Burki *et al.* 2020), but the latter was omitted from the analysis because the SILVA database contains only 78 records, 76 of which are derived from environmental samples, and this group of mainly marine heterotrophs is considered unlikely to contain symbiotic species (Bass and del Campo 2020; Strassert *et al.* 2019).

The *Amoebozoa* clade composition has changed since 2012, and some groups were taken away into the separate supergroup CRuMs, but these organisms were herein analysed within *Amoebozoa* following the SILVA taxonomy (Burki *et al.* 2020). Nevertheless, the three groups from CRuMs would not significantly impact the analysis, as they are poorly studied (*Mantamonas* has four records, *Diphylleida* – two records, and *Rigifilida* – three records) and have no symbiotic species described (Bass and del Campo 2020). The same applies to *Glaucophyta* – a clade within *Archaeplastida*: very few described species and even fewer records are available in SILVA, and all described species are free-living (Price *et al.* 2017).

The *Excavata* supergroup is one of the main groups of interest as it has many symbiotic protists described. Although its monophyly is nowadays doubted (Burki *et al.* 2020), it was included in the analysis as one group.

The 28S rRNA part of the SILVA database is much smaller than the 18S part. Figure 19 compares the number of records per group for the SSU and LSU rDNA genes. Many understudied groups have few records for the SSU gene and none for the LSU gene. The lack of records for some groups such as *Ctenophora* or *Haptista* is not crucial as no symbionts from them have been described to date (Bass and del Campo 2020), but many important groups are also missing, e.g., *Microsporidia*, *Jakobida*, and *Ichthyosporea*.

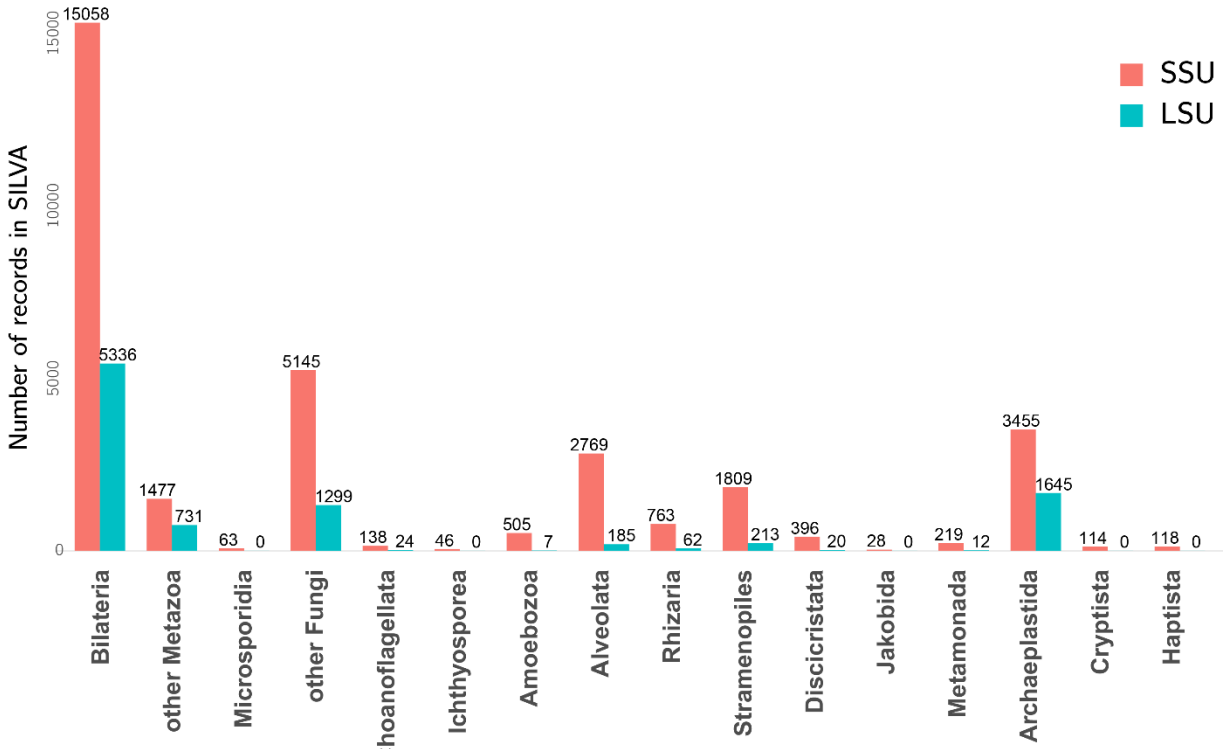

Figure 19 Numbers of records for the SSU and LSU rRNA genes for different taxa in the SILVA v138 database

2.4.4.2. SSU rg1 primers

There are three positions within the first SSU region (SSU rg1) (Figure 20, nucl 3, 23, and 24), which can discriminate *Bilateria* from most other groups. Unfortunately, these variable nucleotides are quite distant from each other. They could be included in one primer, but mismatches on the 5' end of primers do not affect the efficacy of the PCR (Stadhouders *et al.* 2010), so depending on the direction of the primer, either nucleotide three or nucleotides 23-24 will not affect the amplification outcome. This region is located in the middle of the SSU rRNA gene, so it is sensible to make it a forward primer so that the extension proceeds towards ITS regions and 5.8S and 28S subunit genes. Therefore, the proposed primer can start with the conserved region from position 7 to the end of the chosen region; the optimal primer length is 18 bp, starting with two purine bases ensuring stronger annealing (van Pelt-Verkuil, van Belkum, and Hays 2008).

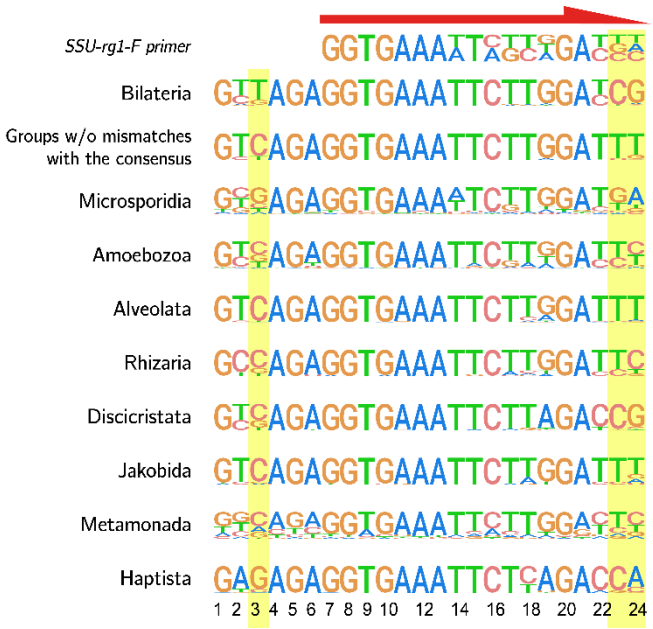

Figure 20 Consensus sequences for the SSU region1 of the 18S rRNA gene. SSU-rg1-F is the provisional non-Metazoan primer. Red arrow shows the direction of the primer. Positions which discriminate Bilateria from other eukaryotes are highlighted with yellow. All non-Metazoan groups that do not have much variation in the consensus are combined together. The detailed weblogo is shown in Figure 16.

The most frequent sequence among non-*Metazoan* groups is 5'-GGTGAAATTCTTGGATTT-3'. The majority of Bilateria have "CG" in the last two positions, while most of the other eukaryotes have "TT". However, there are five target groups – *Microsporidia, Amoebozoa, Discicristata, Metamonada,* and *Haptista* – which also have the two last nucleotides different from the consensus and therefore, there is a chance that this primer will neglect these groups of interest. Variations include "GA" in *Microsporidia*, "CC" in *Amoebozoa* and *Metamonada*, "TC" in *Rhizaria*, "CA" in *Haptista*, and, unfortunately, "CG" in *Discicristata*, which coincides with *Bilateria* and thus cannot be resolved. All other variations can be included in the reaction as degenerated nucleotides:

5'-GGTGAAA[T|A]T[C|A][T|G][T|C][G|T|A]GA[T|C][T|G|C][T|A|C]-3' or 5'-GGTGAAAWTMKYDGAYBH-3'. This primer would be unlikely to pick up any *Discicristata*.

2.4.4.3. SSU rg2 primers and analysis of the 18s-EUK-1134-R primer

The non-*Metazoa*n reverse primer 18s-EUK-1134-R designed by Bower *et al.* (Bower *et al.* 2004) lies within the second SSU region (SSU rg2) and, according to this analysis, has one or two nucleotide differences between *Metazoa* and other groups (Figure 21, positions 15 and 20). Position 15 has either "C" or "T" nucleotide in *Bilateria*, but strictly "C" in other groups; position 20 is either "G" or "A" in *Bilateria*, while other groups have a consensus "G". Consequently, the 18s-EUK-1134-R primer was designed to complement "C" and "G" nucleotides in these positions (Figure 21 shows the alignment of the reverse complement of the primer). A closer look at the taxonomic distribution of these altering nucleotides within *Bilateria* shows that most arthropods have "T" and "A" in these positions, which confidently discriminates them from other groups. Only 539 (3.3%) of *Metazoan* sequences in SILVA have a perfect match with the 18s-EUK-1134-R primer, 280 of which belong to *Porifera* and 206 belong to *Cnidaria,* while the present study focuses on blood-feeding *Arthropoda*. It is

worth mentioning that both mismatches are in the 3' end of the primer, making the discrimination more effective. It has been demonstrated that PCR efficiency significantly decreases with mismatches in the 3′ region and can even be prevented by a single mismatched base at the 3′ end (Bru, Martin-Laurent, and Philippot 2008).

Although the EUK-1134-R primer performed well in reducing the amplification of templates from *Arthropoda,* as shown in 2.3.1.2, it has several mismatches with *Microsporidia* and *Discicristata* (Figure 21); both of these groups were of significant interest to this analysis. Like many other parasites, microsporidia have genetically divergent 18S genes and are rarely detected in environmental samples (Bass and del Campo 2020; B. A. P. Williams *et al.* 2018). Four mismatches between the EUK-1134-R primer and *Microsporidia* SSU rRNA gene sequence could theoretically reduce amplification of these symbionts. The same applies to the two mismatches within the *Discicristata*. Taking into account these mismatches, a mixture of primers based on EUK-1134-R and targeting the following variations of the sequence could improve the amplification of these groups (positions different from the EUK-1134-R primer are shown in bold):

- CGCAAGGCTGAAACTTAAA (EUK-1134-R)
- CGCAAG**TT**TGAAACTT**G**AA (EUK-1134-R-1 favouring *Microsporidia*)
- CGCAAGGCTGAAACTT**G**AA (EUK-1134-R-2 favouring *Metamonada*)
- CGCAAG**AG**TGAAACTT**G**AA (EUK-1134-R-3 favouring *Discicristata*).

SSU rg2 has another variable position not included in the original EUK-1134-R primer (Figure 21, position 6). There is a possibility that a primer which included all three discriminating nucleotides might perform better than the original EUK-1134-R primer. Another reason for redesigning this primer is changing the direction. The EUK-1134-R primer is a reverse primer, which means that it can amplify a maximum 1kb fragment if used with the universal primer at the beginning of the 18S rRNA gene (e.g., EukA, Figure 7). If the primer from the same region is orientated in the other direction, theoretically it can amplify 800 bp of the 18S rRNA gene and both ITS

regions, 5.8S and part of the 28S rRNA gene. Such coverage should give a better resolution on several levels down to the strain level, even in groups with conserved SSU genes. The "G" nucleotide in position 20 (Figure 21) is the perfect candidate to be the 3'-end nucleotide in the new primer. The following 17 base pairs were chosen for the candidate primer 5'-GGGAGTATGGTCGCAAG-3'. Figure 21 shows the alignment of the original non-*Metazoan* EUK-1134-R primer and the proposed improved primer – SSU-rg2-F. SSU-rg2-F has one more mismatch with *Bilateria* and has an opposite direction, making it possible to pair it with a reverse primer located in the 28S rRNA gene. Some positions are made degenerate to improve the amplification of *Microsporidia*, *Discicristata* and *Metamonada*:

5'-GG[G|T]A[G|C]TA[T|C]G[G|T|A][T|A|C]CGCAAG-3' or

5'- GGKASTAYGDHCGCAAG-3'.

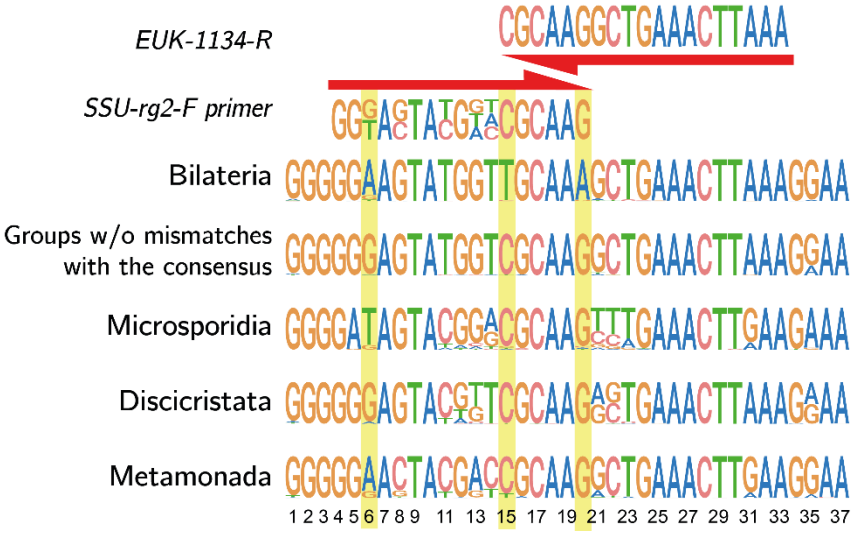

Figure 21 Consensus sequences for the SSU rg2 region of the 18S rRNA gene. Alignment of the reverse complement of the EUK-1134-R primer with Bilateria and some groups of interest is shown. SSU-rg2-F is the new primer designed in this study. Red arrows show the direction of primers. Positions, which discriminate Bilateria from other eukaryotes, are highligted with yellow. All non-Metazoan groups which don't have much variation in the consensus are combined together. The detailed weblogo is shown in Figure 16.

2.4.4.4. SSU rg3 primers

The third region of interest within the 18S rRNA gene starts with two nucleotides (Figure 16), which should discriminate *Bilateria* from all other groups, including other metazoans, and the following region is conserved in most groups except for *Microsporidia*, *Discicristata* and *Metamonada*. Unfortunately, these groups have high variability in the first two positions and the 4-5 following nucleotides, so it is challenging to design a degenerate primer. Moreover, this primer can only be reverse, as the variable nucleotides should be at the 3' end. Considering that EUK-1134-R is located close to this region, the fragment size and region will be very similar and, as EUK-1134-R has already proved to be effective, there is no need to design another primer.

2.4.4.5. LSU rg1 primers

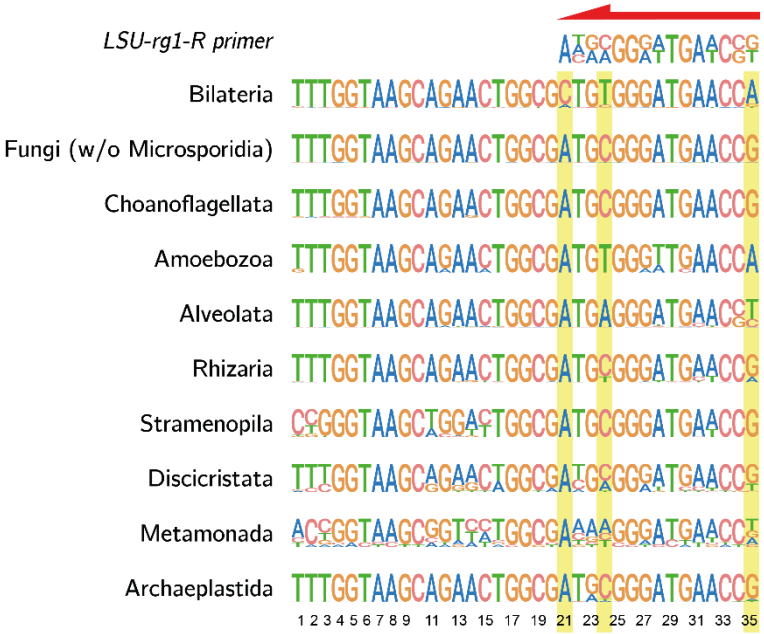

Figure 22 Consensus sequences for the LSU region1. LSU-rg1-R is the provisional non-Metazoan primer. Red arrow shows the direction of the primer. Positions, which discriminate Bilateria from other eukaryotes, are highlighted with yellow.

This first LSU region (LSU rg1) has three variable nucleotides (positions 21, 24 and 35 shown in Figure 22), which seem promising in discriminating *Bilateria*. Analysis of the structure of the 28S rRNA gene in different *Eukaryota* revealed that this gene

might be separated into different sub-parts and relocated to different sites in the genome; the second part of the LSU gene is more likely to contain introns and overall be more variable than the start of the gene (Torres-Machorro *et al.* 2010). Taking this into account, it seems rational to design the oligonucleotide as a reverse primer to amplify a part of the 28S rRNA gene and other parts of the ribosomal cistron. It is advantageous to place nucleotides 21 and 24 at the 3'-end of the reverse primer. The degenerate primer 5'- MSGWTCAWYCCKYDT -3' was designed to pick up groups with a variable sequence in this region – *Alveolata*, *Discicristata*, *Metamonada* (Figure 22).

2.4.4.6. LSU rg2 primers

Two variable nucleotides discriminating *Bilateria* are found in the second LSU region (LSU rg2) – positions 10 and 20 (Figure 23). Nucleotides from 21 to 28 are relatively conserved among all presented groups, so the primer is designed to be reverse

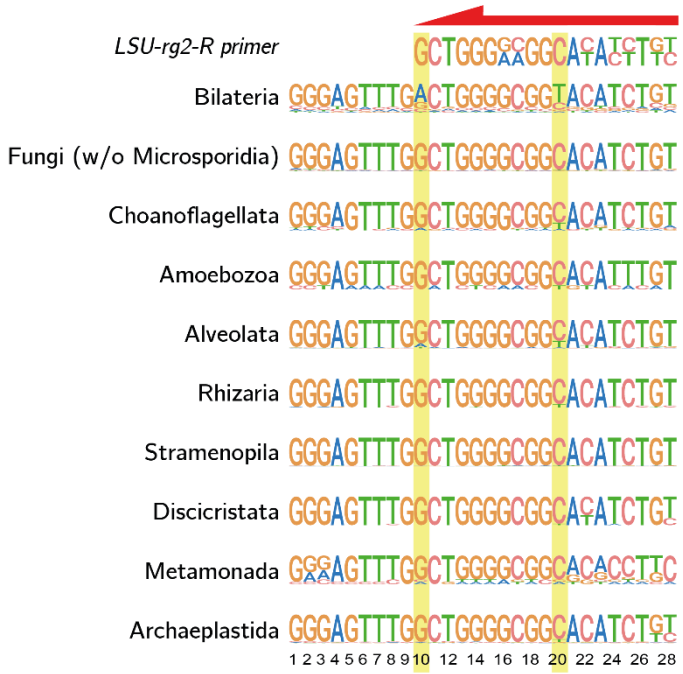

Figure 23 Consensus sequences for the LSU region 2. LSU-rg2-R is the proposed primer. Red arrow shows the direction of the primer. Positions, which discriminate Bilateria from other eukaryotes, are highligted with yellow.

and start at position 28 and have the 3'-end base pair different between *Bilateria* and all others.

This region is located close to the end of the 28S rRNA gene, making it less attractive, considering the multiformity of the second half of the 28S rRNA gene (Torres-Machorro *et al.*, 2010). Thus, this primer is less likely to result in a thorough coverage of diversity.

2.4.5. Method of choice for screening of arthropod populations

All described methods have their limitations. One major limitation is based on the scarcity, bias and irregularities of the genomic databases. Some branches and model species such as *Metazoa* and model organisms are well covered, while other divisions contain only a few records (such as *Microsporidia* or *Malawimonada*). Due to the explosion in new technologies, most of the genomic data has been acquired during the last decade. Understandably, the plethora of new data for phylogenetic reconstruction has led to many inconsistencies in taxonomy in the databases, which might also affect the analysis quality (Edgar 2018). The tree of life has been rearranged several times, especially in the eukaryotic part, where new technologies revealed new clades of microeukaryotes completely missed before the genomic era (Burki *et al.* 2020).

DNA- and PNA-blocking approaches do not seem to be effective in such experiments. PNA blocking has several disadvantages; first, PNA synthesis is expensive; second, the PCR with PNA blockers needs fine-tuning, including a 4-step PCR. Even with all the precautions, the results were inconsistent. Therefore, DNA- or PNA- blocking cannot be considered the method of choice for this study.

When designing the metabarcoding experiments, one must understand that it is impossible to find a pair of primers that gives complete and unbiased results. First of all, each step of the experiment introduces various biases. Different extraction methods might differ in efficacy, and some protists might have cell walls which would make them less susceptible to digestion (Santos *et al.* 2015). PCR templates have a

different affinity with primers depending on the number and position of mismatches and nature (G/C vs A/T). As in the 16S rRNA gene metabarcoding, such results cannot be used to assess the ratio of different organisms in the sample because of bias introduced with uneven DNA extraction and PCR amplification and different copy numbers of ribosomal genes in different organisms. It is impossible to find a pair of primers covering all eukaryotes except for *Bilateria*. One should find a balance between reducing the amplification of *Bilateria* and not missing organisms from other groups. The most difficult groups to amplify with universal primers are *Microsporidia*, *Discicristata* and *Metamonada*. One of the solutions to this issue is to use degenerate primers. If there is a position where a nucleotide is different between *Bilateria* and the other taxa, and alters between the three above-mentioned taxa, it should be possible to cover all desired diversity using a mixture of primers.

Some primers look more promising than others, but all the proposed primers designed *in silico* will need to be evaluated *in vitro*. The EUK-1134-R primer has already shown a satisfactory reduction of the host signal, so even if other proposed primers are not as effective, this approach is the most reliable and straightforward to access the symbiont diversity of arthropods and other animals. One of the enormous advantages of this method is its comparatively low price. The set of amplifying primers should cost no more than £100, and the following ONT sequencing run can be multiplexed up to 96 samples, which means that the analysis of one specimen should cost about £10-£15. This price can be further reduced with an additional check of the PCR-products with gel-electrophoresis and excluding specimens that produced a single band. However, this step is time-consuming and not always reliable, as a PCR product from the host and the same fragment from a symbiont might be very close in size and therefore be visualised as a single band.

3. Chapter III Bioinformatic approach to look for prokaryotic insertions in genomes of *Arthropoda*

3.1. Introduction

Horizontal gene transfer (HGT) in eukaryotes appears to be a rare event (Hotopp *et al.* 2007). The early segregation of germline cells during development is believed to protect eukaryotic genomes from HGT events, although there may be symbionts residing within a germline (Blaxter 2007). The perfect example of this is *Wolbachia* bacteria transmitted within host germline cells which might have enabled the migration of their genetic material to a host genome (Dunning Hotopp 2011; Serbus and Sullivan 2007). Bacteria from the genus *Wolbachia* are found in more than 40% of insect species, as well as in filarial nematodes, crustaceans and arachnids (Zug and Hammerstein 2012). Insertions of *Wolbachia* genomes into arthropod genomes have been described in pea aphids (Nikoh *et al.* 2008), mosquitoes (Klasson *et al.* 2009), beetles (Aikawa *et al.* 2009), fruit flies and parasitoid wasps (Dunning Hotopp 2011). These occurrences seem to be independent from the evolutionary perspective as the fraction of the *Wolbachia* genome, and the localisation in the host genome varies (Dunning Hotopp 2011). For example, the entire chromosome of *Wolbachia* has been integrated into chromosome 2 of *Drosophila ananassae* (Klasson *et al.* 2014), and only 30% of the *Wolbachia* genome is found in the X-chromosome of the bean beetle *Callosobruchus chinensis* (Nikoh *et al.* 2008).

The bovine lungworm *Dictyocaulus viviparus*, a parasitic nematode of cattle, is not known to bear *Wolbachia* bacteria, although its genome sequencing revealed extensive insertions of *Wolbachia* fragments into the nematode chromosomes (Koutsovoulos *et al.* 2014). The authors identified 193 contigs of *Wolbachia* origin

within the worm assembly. They claimed that these contigs do not belong to a living symbiont but are fossils of an old *Wolbachia* insertion based on the level of fragmentation of the sequenced fragments, their interspersion with the host contigs, essential genes missing or containing deleterious mutations. Most predicted *Wolbachia* genes had insertions, deletions, frameshifts or nonsense codons compared with living bacteria (Koutsovoulos *et al.* 2014).

A significant technical issue in recognising such insertions arises from the fact that a host might bear the both bacterial symbiont and the insertion (Boothby *et al.* 2015; Koutsovoulos *et al.* 2016; W. F. Martin 2017). Some research groups made a successful effort to get rid of symbiotic bacteria in order to pinpoint sequences incorporated into the host genome (Alam *et al.* 2011), and vice versa – remove the insertion to obtain a symbiont genome not contaminated by insertions (Gasser *et al.* 2019). Such an approach is not always possible and requires comparatively long laboratory cultivation of arthropods and complex genetic manipulations like introgressive hybridisation (Gasser *et al.* 2019; Klasson *et al.* 2014).

Many false claims of HGT have been made, which later were doubted and assigned to contamination (Boothby *et al.* 2015; Koutsovoulos *et al.* 2016; Martin 2017). Therefore, there is a strong demand to develop an assay that effectively distinguishes true pathogens and bacterial insertions in the host genome. This study attempts to solve this issue using state-of-the-art sequencing technologies and bioinformatic analysis.

Two species of *Arthropoda* were chosen for the study; first, a tsetse fly *Glossina morsitans* was used as a testbed for the pipeline formulation as it has been described in literature that *G. morsitans* bore large insertions of *Wolbachia* genome fragments into its nuclear genome (Brelsfoard *et al.* 2014; Doudoumis *et al.* 2012). Generally, *Wolbachia* is famous for its ability to integrate into host genomes and is quite thoroughly studied which makes it a good system for method developing.

Second object, a tropical bont tick *Amblyomma variegatum* cell line maintained in the Tick Cell Biobank at the University of Liverpool was suspected to have a rickettsial insertion in its nuclear genome (personal communication from Prof Makepeace, Prof Darby and Dr Bell-Sakyi). *Rickettsia* have been found to reside in ovaries of their hosts and can be vertically transmitted, but it is the first case of rickettsial insertion into the nuclear genome of the host.

3.1.1. The case of *Wolbachia* insertion into the *Glossina morsitans* genome

Wolbachia insertions in *G. morsitans* were first described in 2012 (Doudoumis *et al.* 2012) and then sequenced in 2014 along with the cytoplasmic *Wolbachia* (Brelsfoard *et al.* 2014). A thorough examination of sequences showed three different insertions found on the X, Y and B chromosomes. This result was confirmed by visualisation using *in situ* hybridisation of the fly's mitotic chromosomes with *Wolbachia*-specific probes (Brelsfoard *et al.* 2014). It has also been shown that insertions reside in heterochromatic regions, likely to protect them against negative selection. It has been reported for other species with *Wolbachia* insertions that some of their genes are actively transcribed (Hotopp *et al.* 2007; Klasson *et al.* 2009), although not in the case of the X-chromosome as in the bean beetle in which *Wolbachia* genes are transcriptionally inactive (Kondo *et al.* 2002). It is yet to be determined whether any of the insertion genes in *G. morsitans* are expressed. Considering the large size of insertions, A and B (527,507 and 484,123 bp) and their phylogenetic similarity, they likely originate from a comparatively recent single insertion event (Brelsfoard *et al.* 2014).

3.1.2. The hypothesis of the *Rickettsia africae* insertion into the *Amblyomma variegatum* genome

There is a strong suspicion that the cell lines of *A. variegatum,* created and maintained at the Tick Cell Biobank at the University of Liverpool by Dr Bell-Sakyi

(Lesley Bell-Sakyi *et al.* 2000; L Bell-Sakyi 2004), have a rickettsial genome or its parts inserted into the nuclear genome ((M. P. Alberdi, Dalby, *et al.* 2012), personal communication from Prof Makepeace and Dr Bell-Sakyi). Cell lines AVL/ CTVM13 and AVL/CTVM17 were *Rickettsia*-positive in the PCR assay using SSU rDNA, *sca4*, and *ompB* genes, but no sign of bacterial infection has been visually spotted in the cells by means of either light or electron microscopy (M. P. Alberdi, Dalby, *et al.* 2012). This might mean either an extremely low infection rate of the cell lines or the templates for the amplification are located within the host genome. Another indirect evidence for the insertion is the lack of amplification of genes *mutS* and *rpoH* (personal communication from Dr Al-Khafaji, Prof Makepeace, and Dr Bell-Sakyi). These genes are important for rickettsial metabolism: *mutS* is responsible for mismatch repair (S. G. E. Andersson *et al.* 1998) and *rpoH*, RNA polymerase sigma-32 factor, is believed to play a role in the heat shock response (Ge *et al.* 2004); the genes are present in all published assemblies of *Rickettsia* species. Results of qPCR for the single-copy genes (*gltA* gene of *R. africae* and the *rpl6* gene of *A. variegatum*) showed an approximate 1:1 ratio between these genes in the samples (personal communication from Prof Makepeace), reinforcing the hypothesis about the insertion.

3.1.3. Overview of the species of *Arthropoda* used in the present study

3.1.3.1. Tsetse fly G. morsitans and its symbionts

There are 31 species and sub-species (depending on classification) of tsetse flies on the African continent, and six of them are recognised as vectors of sleeping sickness and thirteen species as vector of nagana diseases of livestock (Krinsky 2019). Given the medical, veterinary and agricultural significance of the *Glossina* genus, knowledge of the genomic aspects of the vector and vector-pathogen interactions are a high priority. Male and female tsetse flies are the cyclical vectors of the trypanosomes that cause African sleeping sickness in humans and nagana in animals (Fevre *et al.* 2008). Considering that there are no vaccines for trypanosomiasis and drugs used for sleeping

sickness treatment are often dangerous and not practical due to increasing resistance of *Trypanosoma* spp., tsetse control remains the most effective option (Aksoy *et al.* 2005). Current methods to control tsetse flies are insect trapping and insecticides, which are not very effective due to the difficulties in reaching some geographic regions and environmental concerns about insecticide safety (Bouyer *et al.* 2013). Insecticides, e.g., DDT, dieldrin, and γ-BHC, affected populations of reptiles, fish, birds, and insects other than tsetse flies in South Africa (Grant 2001). Some tsetse fly control methods, such as tsetse habitat destruction or elimination of tsetse reservoir hosts, have been discontinued because of ecological and environmental concerns (Kebede, Ashenafi, and Daya 2015).

A less environmentally harmful and highly effective method to reduce the vector population is the sterile insect technique (SIT); any mating with a sterile male will prevent females from giving birth to any offspring (Dame and Schmidt 1970). The technique effectively eradicated *Glossina austeni* from Unguja island in Zanzibar (Vreysen *et al.* 2000). However, the cost of SIT is exorbitant, and the mass rearing of the flies is a significant problem as large numbers of male flies are needed to be released to out-compete the wild population (Vreysen *et al.* 2000).

When studying the genomics of a tsetse fly, one must remember that it is never just one fly; it comprises genomic material from many organisms, including bacterial symbionts, trypanosomes, nematodes, phages and baculoviruses (Welburn *et al.* 2001). *Wigglesworthia* is described in all *G. morsitans* flies as an obligate symbiont necessary for providing vitamins that the tsetse fly diet lacks and that affects host fecundity, larval development and the immune system (Pais *et al.* 2008; Weiss, Wang, and Aksoy 2011; Aksoy 1995). *Wigglesworthia* bacteria are essential for tsetse flies to neutralise the dietary limitations of a hematophagous lifestyle. The facultative symbiont *Sodalis glossinidae* is not present in all natural populations and has been recently linked to the ability of tsetse flies to transmit trypanosomes and therefore is a promising target for disease control (Farikou *et al.* 2010). *Wolbachia* symbionts are also not essential for

tsetse flies; examined natural populations of *G. morsitans* may have from 37.5% to 100% *Wolbachia* infection rate (Doudoumis *et al.* 2012). *Wolbachia* are known to control the reproductive abilities of insects induing individuals with cytoplasmic incompatibility (CI) of infected individuals with aposymbiotic flies (Brelsfoard *et al.* 2014). A better understanding of tsetse genomics and genomics of their symbionts can improve existing control strategies aimed directly at the fly or its parasite transmission ability (Aksoy *et al.* 2005).

3.1.3.2. Current genomic data available for G. morsitans

There are two published assemblies of the *G. morsitans* genome: the UC Berkeley group used short-read technology resulting in a 348 Mb genome assembled into 32,924 scaffolds with N50 of 9,802 bp (Vicoso and Bachtrog 2015), and the Wellcome Trust Sanger Institute produced an assembly of 366 Mb in 13,807 contigs with N50 of 120,413 bp (International Glossina Genome Initiative *et al.* 2014). The latter assembly is considered representative at the moment. The genome is exceptionally useful for understanding the biology of *G. morsitans* and all experiments involving genomic analysis, such as transcriptomic experiments. However, there is room for improvement. None of the listed assemblies dealt with a male individual of *G. morsitans*; one group used solely female flies, the other used a mixture of male and female flies, so obtaining an assembly of a male individual might lead to new insights in sex determination in tsetse flies as well as more details of a *Wolbachia* insertion which has been shown in the Y chromosome (Brelsfoard *et al.* 2014). It is worth mentioning that modern long-read technologies allow better assembling of genomes aiming for chromosome-level assemblies.

3.1.3.3. Overview of the tropical bont tick A. variegatum

Amblyomma variegatum ticks are of high importance for human and veterinary medicine as vectors of several pathogens, mainly *Ehrlichia ruminantium* (previously known as *Cowdria ruminantium*) causing heartwater disease in ruminants (Allsopp

2015), and *R. africae* causing African tick bite fever in humans, as well as benign African theileriosis (*Theileria mutans* and *Theileria velifera*), bovine dermatophilosis (*Dermatophilus congolensis*), and several viral diseases (Jongejan and Uilenberg 2004; Pagel Van Zee *et al.* 2007). *A. variegatum* are three-host ticks; they can acquire pathogens either from a blood meal or transstadial from the nymphal to adult stage (Yonow 1995). Transovarial transmission has not been recorded for *E. ruminantium*, although it has been demonstrated for *R. africae* (Socolovschi *et al.* 2009). *A. variegatum* is causing serious concerns because of its geographical expansion. It originated in sub-Saharan Africa (J. B. Walker and Olwage 1987), then invaded the Caribbean Basin with livestock trade in the 19th century, and recently has been detected in Madagascar and Yemen (Barre and Uilenberg 2010) and Italy (Pintore *et al.* 2021). There is a potential for spread into large areas of South and North America (Leger *et al.* 2013; Estrada-Peña *et al.* 2007).

There are no whole genomes sequenced from *A. variegatum* or any *Amblyomma* species at the time of writing. There are only published genomes for nine species of ixodid ticks to date, six of which were only published two years ago (Gulia-Nuss *et al.* 2016; Barrero *et al.* 2017; Cramaro *et al.* 2017; Guerrero *et al.* 2019; Jia *et al.* 2020). The genomics of hard ticks is poorly studied mainly because of their genome size – those that have already been sequenced have enormous genomes spanning from 1.7 to 2.8 Gb (Table 3-1). Pregenomic genome size estimations for several hard ticks approach 7-9 Gb; so Ullmann and co-authors analysed reassociation kinetics of DNA and estimated the genome of *R. microplus* to be 7.1 Gb and *I. scapularis* to be 2.1 Gb (Ullmann *et al.* 2005). The current high-quality assembly of *R. microplus* is much smaller – only 2.53 Gb (Jia *et al.* 2020); this discrepancy might result from the erroneous estimation, difficulties in assembling repetitive regions, or both. The assembly size of 2.53 Gb corresponds better to the known range of hard tick genome sizes (Table 3-1). The haploid genome of *I. scapularis* was also estimated to be slightly above 2 Gb by flow cytometry (Geraci *et al.* 2007), and this size was later confirmed by whole-genome sequencing from two independent groups (Miller *et al.* 2018; Gulia-

Nuss *et al.* 2016). There are several genome size estimations for members of the *Amblyomma* genus: 3.1 Gb for *Amblyomma americanum*, 2.7 Gb for *Amblyomma cajennense*, and 2.9 Gb for *Amblyomma maculatum* (Geraci *et al.* 2007). Previous attempts to sequence the *A. variegatum* genome indicated the genome size to be approximately 5-6 Gb (personal communication from Prof Alistair Darby).

Table 3-1 Published assemblies for Ixodidae ticks

Tick	**Genome size, Gb**	**Reference**
Dermacentor silvarum	2.47 (ASM1333974v1)	(Jia *et al.* 2020)
Haemaphysalis longicornis	2.56 (ASM1333976v1) or 7.36 (HLAgriLifeRun1)	(Jia *et al.* 2020) (Guerrero *et al.* 2019)
Hyalomma asiaticum	1.71 (ASM1333968v1)	(Jia *et al.* 2020)
Ixodes persulcatus	1.9 (BMI_IPER_1.0)	(Jia *et al.* 2020)
Ixodes scapularis	2.23 (ASM1692078v2), 2.08 (ISE6_asm2.2_deduplicated)	(direct submission to NCBI, unpublished) (Miller *et al.* 2018)
Ixodes ricinus	0.52 (partial assembly ASM97304v2)	(Cramaro *et al.* 2015)
Rhipicephalus sanguineus	2.37 (ASM1333969v1)	(Jia *et al.* 2020)
Rhipicephalus microplus	2.53 (ASM1333972v1) or 2.01 (Rmi2.0)	(Jia *et al.* 2020) (Barrero *et al.* 2017)
Rhipicephalus annulatus	2.76 (TxGen Rann)	(direct submission to NCBI, unpublished)

3.1.4. Aims and objectives

This chapter aims to develop a bioinformatic pipeline to differentiate between living symbiotic bacteria and the integration of symbiotic genomic fragments into the host genome. The current approach involves complex experiments which require substantial laboratory resources and highly skilled personnel; the bioinformatic approach aims to make the evaluation faster and cheaper.

The confusion between living symbiotic bacteria and bacterial insertion can introduce biases populational studies of *Rickettsia* in ticks as such studies use genes that are likely to be relocated to the host genome in case of insertion. In particular, screening for *R. africae*, the agent of African tick-bite fever, usually utilise *ompB* and *gltA* genes (Lorusso *et al.* 2013; Pintore *et al.* 2021; Iweriebor, Nqoro, and Obi 2020; Popov *et al.* 2007; 2007), which might lead to erroneous conclusions about the presence and density of the pathogen in natural populations and affect the choice of control measures.

To accomplish this aim, a number of specific objectives have been outlined. Firstly, high quality assemblies of two arthropod species will be obtained using a long-read sequencing technology: a genome of a male *G. morsitans* fly and *A. variegatum* tick. Secondary, an algorithmic approach to differentiate symbiotic bacteria from insertions of bacterial sequences into the host genome will be developed based on these assemblies.

3.2. Methods

3.2.1. Material collection and preparation

3.2.1.1. Glossina morsitans

Dr Lee Haines of the LSTM kindly provided tsetse flies (*G. morsitans*). Flies were maintained in a laboratory colony at 26°C and 65–70% relative humidity and were allowed to feed on defibrinated horse blood every 48h using an artificial membrane system.

High molecular weight DNA for long-read sequencing was extracted from a single male tsetse fly using a modified phenol-chloroform protocol with phase-lock gel tubes (Quick 2008). Single fly individual of *G. morsitans* were crushed with a pestle to make the internal organs of the fly available for lysis. 600 µl of ATL buffer (QIAGEN) and 20 µl of RNase A (R4642-50MG, Sigma Aldrich) were added to the sample and incubated at 37°C for 30 min. Then 20 µl of proteinase K (20 mg/ml) were added and incubated at 56°C for 4 h with slow rotations performed every 30 min. The mixture was decanted into 2 ml centrifuge tubes containing MaXtract High-Density phase-lock gel (QIAGEN) along with 600 µl of phenol/chloroform/isoamyl alcohol mixture (25/24/1) (77617-100ML, Sigma Aldrich). The tubes were incubated at room temperature on a Hula mixer for 10 min and then centrifuged at 1500 rcf at 4°C for 10 min. The top aqueous phase was poured into a new tube with the phase-lock gel, while the phenol/chloroform phase was locked with a gel at the bottom of the tube. 600 µl of chloroform/isoamyl alcohol mixture (25:1) was added to the aqueous phase, rotated for 10 min at RT and centrifuged at 1500 rcf at 4°C for 10 min. The aqueous phase was poured into a clean 5 ml tube. 1.8 ml of ice-cold absolute ethanol and 240 µl of 5M ammonium acetate were added and mixed by slow tube rotation ten times. The jelly-like DNA precipitate was slowly transferred into a 1.7 ml tube with a wide-bore 1 ml tip and washed twice with 70% ethanol. The tube was centrifuged for 10 min at 1500 rcf to remove the ethanol residues and air-dried for 30 min. The DNA was

submerged in 50 µl of nuclease-free water and kept overnight at 4°C for resuspension. The DNA quality was assessed with Nanodrop ND-1000 Spectrophotometer (Thermofisher, USA), and quantity was estimated with Qubit using Quant-iT dsDNA HS Assay Kit (Thermofisher, USA).

Sequencing libraries for ONT sequencing were prepared according to SQK-LSK108 and SQK-RAD004 protocols with a starting amount of 1 – 1.2 µg DNA and sequenced with MIN106 (R 9.4.1) flowcells for MinION (Oxford Nanopore Technologies, UK). Raw nanopore fast5 files were basecalled with ONT Guppy basecalling software v.5.0.11 using dna_r9.4.1_450bps_sup configuration. Reads were analysed, filtered, trimmed and split according to quality using Filtlong tool v.0.2.1 (github.com/rrwick/Filtlong). Read quality and length distribution were assessed using NanoPlot v.1.38.1 (De Coster *et al.* 2018).

3.2.1.2. A. variegatum

Dr Lesley Bell-Sakyi of the Tick Cell Biobank (TCB) at the University of Liverpool kindly provided cultures of the *A. variegatum* cell line AVL/CTVM17 at passage 125-127, maintained as described previously (L Bell-Sakyi 2004).

High molecular weight DNA for long-read sequencing was extracted from approximately 2.2 ml of AVL/CTVM17 cell culture Dneasy Blood & Tissue kit (Qiagen, UK) according to the manufacturer's protocol. Sequencing libraries for ONT sequencing were prepared according to SQK-LSK109 protocol with a starting amount of 2 µg DNA and sequenced with MIN106 (R 9.4.1) flowcells for MinION (Oxford Nanopore Technologies, UK). Basecalling, quality control, trimming and filtering were performed as described in the section 3.2.1.1.

Prof Matthew Loose and Dr Inswasti Cahyani of the University of Nottingham kindly provided part of the ONT data for *A. variegatum*. The DNA was extracted as described in the protocol by Cahyani *et al.* (Cahyani 2021). Ultra-Long DNA Sequencing Kit SQK-ULK001 was used for the library preparation and sequenced on

GridION and PromethION (Oxford Nanopore Technologies, UK) devices at the University of Nottingham.

PacBio library preparation and sequencing were performed in the Centre for Genomic Research at the University of Liverpool (CGR). TruSeq PCR-free paired-end libraries (2x150 bp) with a 350 bp insert were generated and sequenced with HiSeq 4000. The CGR performed the following read curation: the raw fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt v.1.2.1 with option -O 3 (M. Martin 2011); the reads were further trimmed using Sickle v.1.200 with a minimum window quality score of 20 (github.com/najoshi/sickle); reads shorter than 20 bp after trimming were removed.

3.2.2. Genome assemblies and curation

3.2.2.1. Arthropod genomes

Genomes were assembled using the following software:

- ONT long-reads: flye v.2.8.3 (Kolmogorov *et al.* 2020), wtdbg2 v.2.5 (Ruan and Li 2020), raven v.1.1.10 (Vaser and Šikić 2021), shasta v.0.5.1 (Shafin *et al.* 2020)
- PacBio long-reads: flye v.2.8.3, wtdbg2 v.2.5, hifiasm v.0.15.1 (H. Cheng *et al.* 2021), hicanu v.2.1 (Nurk *et al.* 2020), IPA v.1.3.1 (https://github.com/PacificBiosciences/pbipa)
- Illumina short-reads: AbySS v.2.2.4 (Nielsen *et al.* 2009), Spades v.3.13.2 (Bankevich *et al.* 2012).

All long-read assemblers were used with default parameters; if the genome size was required on input, it was estimated as 300 Mb for *G. morsitans* and 5 Gb for *A. variegatum*. AbySS was run with k-mer sizes of 50, 60, 70, 80, 90, 100, 110 and 120.

The ONT-based assembly of *G. morsitans* was polished with ONT raw reads using Medaka (nanoporetech.github.io/Medaka) followed by five rounds of Racon polishing with Illumina reads (Vaser *et al.* 2017). Each round of Racon polishing was evaluated with BUSCO score using the 'diptera_odb10' dataset to assess the genome completeness (Seppey, Manni, and Zdobnov 2019). The assembly showing the best BUSCO score was then polished for 15 rounds with Illumina reads using Pilon (B. J. Walker *et al.* 2014), which was then evaluated with the BUSCO score and the number of changes made by Pilon.

Reads were taxonomically assigned using blastn v.2.11.0 (Camacho *et al.* 2009) versus 'nt' database built on April 28th, 2021. Blobtools package v.1 was used to visualise the assembly with the taxonomical distribution of contigs and their coverage (Laetsch and Blaxter 2017).

Bowtie2 v.2.2.4 (Langmead and Salzberg 2012) and minimap2 v.2.17 (Heng Li 2018) were used to align short and long reads, respectively. Samtools toolset was used for sorting and indexing alignments (H. Li *et al.* 2009). Alignments were visualised in Integrative Genome Viewer (IGV) (Thorvaldsdottir, Robinson, and Mesirov 2013). Illumina reads were aligned to the polished assembly, and contigs with zero Illumina coverage were removed.

Mitochondrial genomes were assembled using the following approach: raw reads were aligned to the closest publicly available mitochondrial genome of the species (KY457513.1 *Amblyomma hebraeum* for *A. variegatum*, NC_037368.1 *Melophagus ovinus* for *G. morsitans*), aligned reads were extracted with seqtk toolset (github.com/lh3/seqtk) and assembled with flye v.2.8.3 (Kolmogorov *et al.* 2019) 2019b) in case of PacBio reads, and unicycler (Wick *et al.* 2017) for ONT and Illumina reads. The mitochondrial genomes were annotated on the MITOS2 webserver (Donath *et al.* 2019) using the RefSeq 81 *Metazoa* reference, the genetic code 5 for Invertebrata and the other parameters set to default.

Pairwise comparisons of assemblies were performed with minimap2 v.2.17 (Li 2018), mashmap (Jain *et al.* 2018) and visualised with dotPlotly.R scripts (github.com/tpoorten/dotPlotly) or with the custom script generateDotPlot.pl provided by the Mashmap authors (github.com/marbl/MashMap/blob/master/scripts/generateDotPlot).

3.2.2.2. Symbiotic bacteria genomes

G. morsitans ONT reads were taxonomically classified using kraken2 v.2.1.2 (Wood, Lu, and Langmead 2019) versus the 'nt' database downloaded on 27.01.2020. Reads assigned to the genera *Wolbachia*, *Sodalis* and *Wigglesworthia* were taken out into three separate files and used for the further assembly with flye v.2.8.3 (Kolmogorov *et al.* 2019). The resulting assemblies were compared with the published assemblies of these symbionts (*Sodalis glossinidius* str. 'morsitans' (GCA_000010085.1), *Wigglesworthia glossinidia* (GCF_000247565.1), *Wolbachia* endosymbiont of *G. morsitans* wGmm (GCF_000689175)) using the D-Genies tool (Cabanettes and Klopp 2018).

A. variegatum ONT and PacBio HiFi reads were taxonomically classified using kraken2 v.2.1.2 (Wood, Lu, and Langmead 2019) versus the 'nt' database downloaded on 27.01.2020. Reads assigned to genus *Rickettsia* were taken out into a separate file and used for the further assembly with flye v.2.8.3 (Kolmogorov *et al.* 2019). The resulting assemblies were compared with the reference *R. africae* genome (GCF_000023005.1) using the D-Genies tool (Cabanettes and Klopp 2018). Raw ONT and PacBio reads were aligned back to assemblies using minimap2 v.2.24 (Li 2018).

The resulting assemblies were annotated with Prokka v.1.14.6 (Seemann 2014). Gene lengths were collected from Prokka annotation files and plotted in R using ggplot2 package (Wickham 2016).

3.2.3. Searching for sequences of mixed origin

3.2.3.1. Searching for R. africae insertion in the A. variegatum assembly

All of the published complete genomes of *Rickettsia* (73 assemblies as for June 2021, see Supplementary table 4) and the only assembly of *R. africae* (GCF_000023005.1) were used for the analysis. All genomes were sliced with 1 bp step and 30 bp window using the seqkit tool v.0.16.1 (Shen *et al.* 2016). The resulting datasets were aligned to the curated assemblies and raw reads obtained from the AVL/CTVM17 cell DNA using bowtie2 v.2.2.4 (Langmead and Salzberg 2012); samtools toolkit v.1.9-47 (Li *et al.* 2009) was used for filtering, sorting and indexing resulting alignments. Igvtools v.2.5.3 (Thorvaldsdottir, Robinson, and Mesirov 2013) was used for average coverage density calculation. Contigs and reads covered with bacterial fragments were chosen for further analysis. Alignments were visualised with the IGV browser (Robinson *et al.* 2011).

Raw ONT and PacBio reads were filtered by length using the Filtlong tool v.0.2.1 (github.com/rrwick/Filtlong); only reads longer than 1000 bp were used for the analysis. All complete *Rickettsia* assemblies (Supplementary table 4) were aligned to raw reads using minimap2 v.2.17 (Li 2018).

3.2.3.2. Searching for Wolbachia insertion in the G. morsitans assembly

All of the published complete genomes of *Wolbachia* (46 assemblies as for June 2021, see Supplementary table 5), as well as contig-level assemblies of *S. glossinidius* str. 'morsitans' (GCA_000010085.1), *W. glossinidia* (GCF_000247565.1), and *Wolbachia* endosymbiont of *G. m. morsitans* *w*Gmm (GCF_000689175) were used for the analysis. All genomes were sliced and aligned to the *G. morsitans* assembly as described for *A. variegatum* (see 3.2.3.1).

Raw ONT reads were filtered by length using the Filtlong tool v.0.2.1 (github.com/rrwick/Filtlong); only reads longer than 5000 bp were used for the analysis. *G. m. morsitans* *w*Gmm (GCF_000689175) and 73 complete genomes of *Wolbachia spp.* (see Supplementary table 5) were aligned to raw reads using minimap2 v.2.17 (Li 2018). Reads were manually separated into three groups: (1) evenly covered with bacterial sequences; (2) left part of a read covered with bacterial sequences whilst right part had no matches in them; (3) right part of a read covered with bacterial sequences whilst left part had no matches in them.

3.3. Results

3.3.1. Sequencing and assembly of *Glossina morsitans* genome

The single male tsetse fly used for sequencing yielded approximately 3.7 mg of DNA with a 74 µg/ml concentration and 260/280 and 260/230 ratios of 1.95 and 2.4, respectively. 15.6 Gb of long-read sequences were obtained from three ONT runs with N50 read length of 12,395 bp. The longest read was 190,894 bp. Reads longer than 10 kb comprised 8.9 Gb, 405 reads were longer than 100 kb and comprised 47 Mb (Table 3-2). The latest estimation of *G. morsitans* genome size is 0.366 Gb (International Glossina Genome Initiative *et al.* 2014), which means that this data amounted to roughly 42X coverage, with at least 24X coverage with reads longer than 10 kb and 1X coverage with ultra-long reads (>100 kb).

Table 3-2 Glossina morsitans ONT reads length and quality distribution

	Before filtering	**After filtering (Q>7)**
Mean read length (bp):	5,822	5,966
Mean read quality:	12.7	13.3
Median read length (bp):	2,892	2,974
Median read quality:	12.8	13.3
Number of reads:	2,676,241	2,458,373
Read length N50 (bp):	12,39	12,692
Total bases (bp):	15,583,390,495	14,666,828,975
Megabases of reads above quality cut-offs (and percentage of all reads)		
>Q5:	15301.2Mb (97.2%)	14666.8Mb (100.0%)
>Q7:	14666.8Mb (91.9%)	14666.8Mb (100.0%)
>Q10:	12422.4Mb (75.3%)	12422.4Mb (82.0%)
>Q12:	10052.3Mb (56.5%)	10052.3Mb (61.5%)
>Q15:	6192.2Mb (31.0%)	6192.2Mb (33.8%)
Top 5 longest reads (mean quality score)		
#1	190,894 bp (10.2)	190,894 bp (10.2)

#2	173,743 bp (9.9)	173,743 bp (9.9)
#3	173,287 bp (11.2)	173,287 bp (11.2)
#4	173,107 bp (11.0)	173,107 bp (11.0)
#5	169,316 bp (17.1)	169,316 bp (17.1)

Long reads were assembled with several assemblers (see Methods 3.2.2 and Supplementary table 3). Flye and raven assemblers resulted in the lowest number of contigs (708 and 941, respectively) and showed the best BUSCO scores (3177 complete BUSCOs in both assemblies). The length of both assemblies was slightly longer than the current representative genome, 370,719,651 bp for Flye and 388,584,165 bp for Raven. Such metrics as N50 and L50 and the longest contig were better in flye assembly than in raven: N50 of 9,302,271 bp vs 3,246,851 bp, L50 of 12 contigs vs 32 contigs, and the largest contig of 30,116,467 bp vs 20,010,946 bp. The flye assembly was chosen for further genome polishing.

Polishing with ONT raw reads using Medaka reduced the number of missing BUSCOs from 56 to 36. The following rounds of Racon polishing with Illumina reads decreased this number to 17-19. The assembly after the first and third rounds of Racon polishing was exhaustively polished with Pilon. After seven rounds of Pilon polishing, the best BUSCO score was obtained. The complete statistics on described steps of assembly and polishing can be found in Supplementary table 3.

The polished assembly was 370,683,298 bp long and assembled in 723 contigs with N50 of 8,605,315 bp and L50 of 13 contigs. Two largest contigs were 24 Mb long (scaffold_279_segment0 and scaffold_304_segment0). A total of 3259 complete BUSCOs were found, of which 3200 were single-copy and 59 duplicated, nine genes were fragmented, and 17 were missing. The currently available RefSeq assembly GCA_001077435.1 has 3188 complete BUSCOs and 49 missing.

Manual curation of the assembly

Sixty-five contigs had zero coverage with Illumina reads, implying that these sequences did not belong to the assembly and were removed. Very short contigs (145

contigs less than 1000 bp) were aligned against the rest of the assembly, and 128 of them had a match in the rest of the assembly. This might mean that these fragments belonged to repetitive or heterogeneous regions of the genome and, combined with a high ONT error rate, these parts were misassembled. These contigs were removed from the assembly as they were unlikely to contain information valuable for this study. The seventeen short contigs which did not match the rest of the assembly were searched with blast, and the results are presented in Table 3-3. Four of them did not have any match in the 'nt' database and were also removed from the dataset. Ten contigs clearly belonged to *W. glossinidia*, and thus were kept for further analysis. Two contigs belonged to *Bacteria* and, considering their short length and low identity with the hit, these contigs might have had originated from conserved genomic regions of the tsetse fly bacterial symbionts. One contig was assigned to *Drosophila melanogaster* and most likely originated from the region of the *G. morsitans* genome which is conserved among *Diptera*.

Table 3-3 Blast search results for short contigs which had no match within the rest of the G. morsitans assembly

	length, bp	**Top hit**	**Top hit name**	**per. Ident, %**	**e-value**	**score**	**query cover, %**
contig_1000_segment0	772	CP003315.1	*Wigglesworthia glossinidia* of *Glossina morsitans*	100	0	1426	100
contig_1050_segment0	607	CP003315.1	*W. glossinidia* of *G. morsitans*	100	0	1122	100
contig_1060_segment0	837	CP003315.1	*W. glossinidia* of *G. morsitans*	99.9	0	1541	100
contig_1061_segment0	550	CP003315.1	*W. glossinidia* of *G. morsitans*	100	0	1016	100
contig_1062_segment0	823	CP003315.1	*W. glossinidia* of *G. morsitans*	100	0	1520	100
contig_1063_segment0	608	CP003315.1	*W. glossinidia* of *G. morsitans*	100	0	1123	100
contig_164_segment0	180	-	-	-	-	-	-
contig_202_segment0	207	CP023338.1	*Drosophila melanogaster*	97.1	2.28 E-39	174	49
contig_236_segment1	111	-	-	-	-	-	-

contig_311_ segment0	644	-	-	-	-	-	-
contig_656_ segment0	516	CP045970.1	*Listeria monocytogenes*	72.5	3.89 E-21	115	74
contig_659_ segment0	750	CP034900.1	*Buchnera aphidicola*	74	4.50 E-17	102	35
contig_824_ segment0	230	-	-	-	-	-	-
contig_877_ segment0	846	CP003315.1	*W. glossinidia* of *G. morsitans*	100	0	1563	100
contig_926_ segment0	673	CP003315.1	*W. glossinidia* of *G. morsitans*	100	0	1243	100
contig_994_ segment0	725	CP003315.1	*W. glossinidia* of *G. morsitans*	100	0	1339	100
contig_999_ segment0	695	CP003315.1	*W. glossinidia* of *G. morsitans*	100	0	1284	100

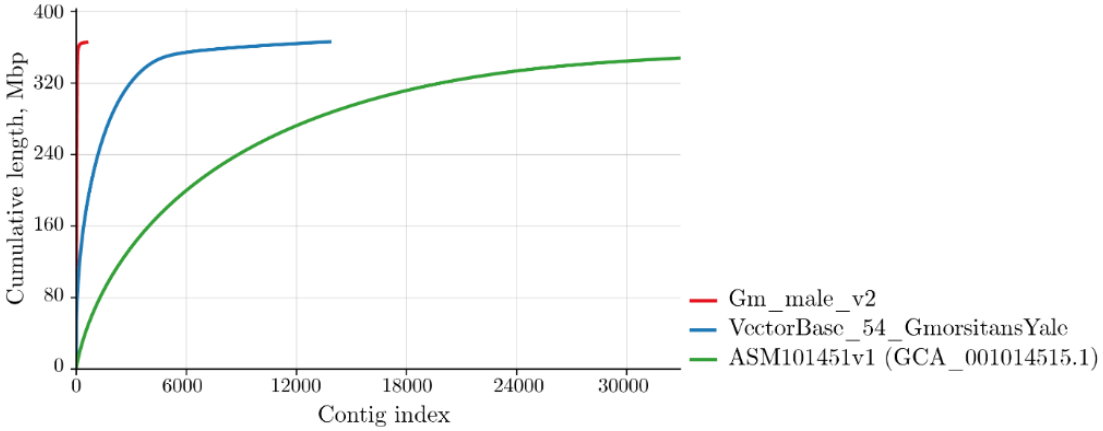

Figure 24 The cumulative plot of the curated assembly from this study (Gm_male_v2) with the published G. morsitans assemblies.

Table 3-4 shows the main characteristics of the polished and curated assembly used for further analysis compared with published assemblies. The overall curated assembly with the working title "Gmors_asm_v2" spanned 365,558,824 bp in 566 contigs, with the largest contig of 24,922,338 bp and GC content of 34.25%. N50 and N75 were 9,323,236 and 4,442,315, respectively. Figure 24 demonstrates the cumulative length of contigs for the assembly generated in this study ("Gm_male_v2"), the representative assembly from the VectorBase and another assembly available from GenBank. It is obvious from the plot that most of the "Gm_male_v2" assembly comprised long contigs; L50 and L75 were 12 and 27 fragments, respectively. Sequences of bacterial symbionts were kept in the assembly.

The nucleotide sequence of the curated *G. morsitans* "Gm_male_v2" assembly can be found in Digital supplementary material.

Table 3-4 Comparison of published assemblies of G. morsitans and the assembly generated in this study ("Gmors_asm_v2")

	Gmors_asm_v2	**GCA_001077435.1_ASM1077743v1**	**GCA_001014515.1_ASM1014515v1**	**VectorBase-54_GmorsitansYale**
Number of contigs	566	24071	32924	13807
Total length, bp	365,558,824	363,107,242	348,062,779	366,195,856
Largest contig, bp	24,922,338	538,223	309,339	25,362,821
GC, %	34.25	34.12	34.03	34.12
N50, bp	9,323,236	49,788	20,491	120,413
L50	12	2,011	4,597	570
N's per 100 kbp	0.01	0	3665.11	842.94
Complete BUSCOs	3240	3188	3036	3216
Single-copy BUSCOs	3186	3147	3006	3168
Duplicated BUSCOs	54	41	30	48
Fragmented BUSCOs	16	48	124	30
Missing BUSCOs	29	49	125	39

3.3.2. Targeted assemblies of symbiotic bacteria from *G. morsitans*

The nucleotide sequences of assemblies of symbiotic bacteria from *G. morsitans* can be found in Digital supplementary material.

Taxonomical assignment revealed 47,755 (1.78% of all ONT reads), 29,750 (1.11%), and 16,550 (0.62%) reads belonging to *Sodalis*, *Wigglesworthia* and *Wolbachia* genera, respectively. These subsets of reads were separately assembled with flye resulting in three assemblies shown in Table 3-5.

The genome of *S. glossinidius* was assembled in 16 fragments, of which six were circular, the biggest contig (contig_3_len_4208665_cov_33_circular) spanned 4.2 Mb and matched the chromosome of *S. glossinidius* str. 'morsitans' (GCA_000010085.1) (Figure 25a). Three circular fragments were very closely matched with pSG2, pSG3 and pSG4 plasmids (see Table 3-6). The pSG1 plasmid was also

present but was not assembled correctly, and it matched with two non-circular fragments of 80 kb and 0.8 kb. One of the circular fragments of 36.4 kb matched with part of the *S. glossinidius* chromosome; perhaps it was a misassembly due to a repetitive region. Another eight non-circular fragments belonged to the *S. glossinidius* chromosome and consisted of different repetitive regions.

Table 3-5 Assembly statistics for three G. morsitans bacterial symbionts

	Total length, bp	**Fragments**	**N50, bp**	**Largest contig, bp**	**Mean coverage**
Sodalis glossinidius	4,465,608	16	4,208,665	4,208,665	43
Wigglesworthia glossinidia	718,367	1	718,367	718,367	320
"Wolbachia sp."	1,721,369	50	91,660	386,408	58

Table 3-6 Blast results for Sodalis glossinidius contigs

Contig	**Blast top hit**	**Top hit ID**	**Query coverage, %**	**Identity, %**
contig_2_len_23236_cov_407	*Sodalis glossinidius* str. 'morsitans' isolate B4, chr: SgGMMB4	LN854557.1	100	99.76
contig_3_len_4208665_cov_33_circular	n/a	n/a	n/a	n/a
contig_5_len_1103_cov_40	*S. glossinidius* str. 'morsitans' isolate B4, chr: SgGMMB4	LN854557.1	100	100
contig_6_len_806_cov_9	*S. glossinidius* str. 'morsitans' isolate B4, chr: SgGMMB4	LN854557.1	100	100
contig_7_len_2002_cov_37	*S. glossinidius* str. 'morsitans' isolate B4, chr: SgGMMB4	LN854557.1	100	99.95
contig_8_len_512_cov_10	*S. glossinidius* str. 'morsitans' isolate B4, chr: SgGMMB4	LN854557.1	100	99.61
contig_9_len_36475_cov_542_circular	*S. glossinidius* str. 'morsitans' isolate B4, chr: SgGMMB4	LN854557.1	100	99.97
contig_12_len_892_cov_86	*S. glossinidius* str. 'morsitans' isolate B4, pSG1	LN854558.1	100	100
contig_13_len_21631_cov_124_circular	*S. glossinidius* pSG4 plasmid from *Glossina palpalis*	AJ868438.1	100	99.97
contig_14_len_24017_cov_203	*S. glossinidius* str. 'morsitans' isolate B4, chr: SgGMMB4	LN854557.1	54	92.89
contig_15_len_10086_cov_80	*S. glossinidius* str. 'morsitans' isolate B4, chr: SgGMMB4	LN854557.1	100	99.94
contig_16_len_4248_cov_121	*S. glossinidius* str. 'morsitans' isolate B4, chr: SgGMMB4	LN854557.1	100	100
contig_17_len_27219_cov_195_circular	*S. glossinidius* pSG2 from *Glossina austini*	AJ868436.1	100	99.91
contig_18_len_19503_cov_224_circular	*S. glossinidius* pSG3 plasmid from *G. palpalis*	AJ868437.1	98	99.99
contig_19_len_80607_cov_91	*S. glossinidius* str. 'morsitans' isolate B4, pSG1	LN854558.1	100	99.95

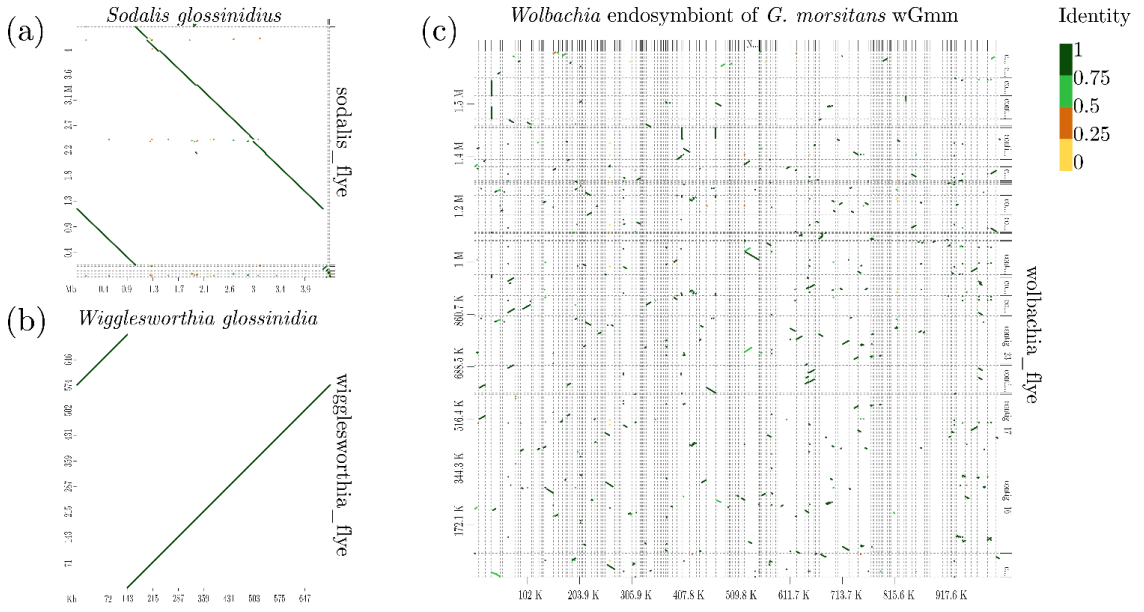

Figure 25 Dot-plot comparison of reference assemblies of three bacterial symbionts of Glossina morsitans with assemblies produced in this study. (a) Sodalis glossinidius str. 'morsitans' (GCA_000010085.1); (b) Wigglesworthia glossinidia (GCF_000247565.1); (c) Wolbachia endosymbiont of G. morsitans wGmm (GCF_000689175).

The *W. glossinidia* genome was assembled into one contig, which had a 99.82% match with the reference assembly ASM24756v1 (Figure 25b).

The *Wolbachia* genome was assembled into fifty contigs, none of which matched the known *Wolbachia* endosymbiont of *G. morsitans* wGmm (Figure 25c). This implies that the sequenced tsetse fly did not bear a cytoplasmic *Wolbachia* and that the contigs belonged to a *Wolbachia* insertion into the host genome.

The analysis of genes used for the MLST (multi-locus sequence typing) analysis revealed that the assembled *Wolbachia* sequences most likely originated from a *Wolbachia* belonging to the group A, which are found in flies and bees. The top hit for a *coxA* gene were group A *Wolbachia* from different bees (99.25% identical with *coxA* genes from *Sphecodes monilicornis*, *Nomada fabriciana*, *Andrena fbphaemorrhoa*, *Lasioglossum morio*, accession numbers OX366400.1, OX366351.1, OX366326.1, OX366318.1 respectively). Cell division protein *ftsZ* was 99.41% identical with 18 sequences from group A *Wolbachia* including four species listed above and several

Wolbachia from *Drosophila* flies. The *gatB* gene for glutamyl-tRNA amidotransferase subunit B was split into two parts and the top hit for it was another group A *Wolbachia* from *Gymnosoma rotundatum*, OX366347.1 with 96.6% identity. Fructose-biphosphate aldolase *fbpA* gene was 99.44% identical with the group A *Wolbachia* from *Drosophila pseudotakahashii* and other species mentioned above. Surface protein *wsp* gene sequence was 99.71% identical with *Wolbachia* from *G. rotundatum* and *Phyllotreta cruciferae*. No *hcpA* genes was found in the contigs. *Wolbachia* strain *w*Gmm from *G. morsitans* (NZ_AWUH01000190.1) was not in the top 100 hits for the described marker genes.

The 16S rRNA gene was substantially shorter than it supposed to be, Figure 26 shows the comparison of 16S rRNA gene from the closest match, *Wolbachia* strain *w*Irr, only two fragments of the gene are present.

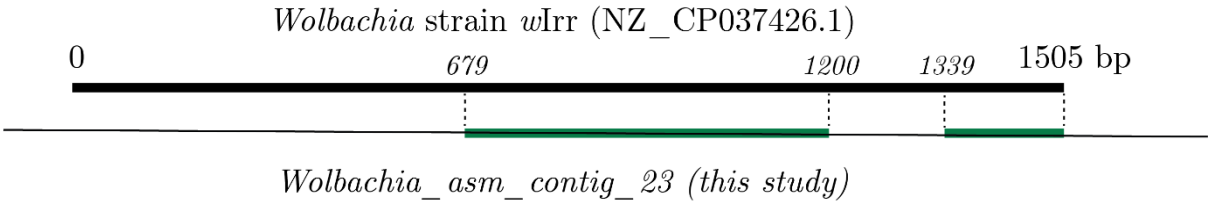

Figure 26 Comparison of 16S rRNA gene from Wolbachia strain wIrr and fragments of 16S rRNA gene from Wolbachia assembly obtained in this study. Black lines represent the missing region.

3.3.3. Sequencing and assembly of *A. variegatum* genome

The *A. variegatum* genome was sequenced in three batches:

- 4.4 ml of AVL/CTVM17 culture yielded approximately 40 µg of DNA with a 9.9 µg/ml concentration. Approximately 10 µg of the DNA were sent to the CGR for the PacBio Sequel II sequencing yielding 92.5 Gb of HiFi reads with the read quality above Q20 and median accuracy Q32. Mean and median read lengths were 14,295 bp and 13,532 bp, respectively;
- 2 µg of the same DNA extraction were used for ONT sequencing yielding 22.3 Gb of long-reads with N50 read length of 14,489 bp. The longest read was 220,142 bp.

Reads longer than 10 kb comprised 13.3 Gb, reads longer than 100 kb comprised 54.1 Mb;

- AVL/CTVM17 cell cultures were sent to the University of Nottingham, where Prof Matthew Loose and Dr Inswasti Cahyani extracted high-molecular-weight DNA with a concentration of 56 ng/μg and 260/280 and 260/230 readings of 2.4 and 1.8, respectively. Sequencing runs yielded 44.1 Gb of data with N50 93,314 bp and the longest read spanning 2,536,912 bp. Reads longer than 10 kb comprised 38 Gb, and ultra-long reads longer than 100 kb comprised 21 Gb.

Table 3-7 contains characteristics of the resultant data.

The datasets were used for genome assembly with several different assemblers, and four of them produced reasonably good assemblies shown in Table 3-8. IPA and hifiasm assemblers produced the assembly with the least number of contigs and highest N50. The high number of duplicated BUSCO genes in the "flye" assembly might have resulted from haplotypes that were not merged.

The nucleotide sequence of the *A. variegatum* assemblies can be found in Digital supplementary material.

Table 3-7 Amblyomma variegatum cell line raw data length and quality distribution. Data from the two ONT batches were combined (22.3 Gb from the University of Liverpool and 44.1 Gb from University of Nottingham). PacBio HiFi reads did not require filtering by quality.

	ONT data before filtering	**ONT data after filtering (Q>7)**	**PacBio data**
Mean read length (bp):	5,191.2	5,835	14,295
Mean read quality:	11.7	13.4	32
Median read length (bp):	665	516	13,532
Median read quality:	12.3	13.2	31.6
Number of reads:	12,795,503	6,254,812	6,467,236
Read length N50 (bp):	45,249	99,278	14,283
Total bases (bp):	66,424,146,478	36,500,413,644	92,448,861,707
Megabases of reads above quality cut-offs (and percentage of all reads)			
>Q5:	64621.0 Mb (96.1%)	36500.4 Mb (100%)	n/a
>Q7:	60852.3 Mb (89.1%)	36500.4 Mb (100%)	n/a
>Q10:	52927.7 Mb (74.1%)	36500.2 Mb (100%)	n/a
>Q12:	42466.7 Mb (53.5%)	30510.6 Mb (73.4%)	n/a
>Q15:	14513.0 Mb (12.8%)	11394.3 Mb (16.4%)	92448.9Mb (100%)
Top 5 longest reads (mean quality score)			
#1	2,536,912 (14.4)	2,536,912 (14.4)	50,290 (21.1)
#2	1,547,154 (6.5)	1,270,861 (11.8)	50,189 (20.5)
#3	1,270,861 (11.8)	1,210,236 (14.3)	50,022 (20.1)
#4	1,219,144 (6.9)	1,128,272 (12.3)	49,899 (20.1)
#5	1,210,236 (14.3)	1,123,993 (15.5)	49,853 (21.8)

Table 3-8 Amblyomma variegatum cell line assembly characteristics and BUSCO scores. The best values are shown in bold.

Assembler	flye	IPA	Hifiasm	wtdbg2
Number of contigs	97,461	**9,692**	21,134	87,197
Assembly length, bp	9,411,593,811	4,963,502,940	8,352,857,305	3,868,746,216
Largest contig, bp	10,815,541	7,368,898	**39,392,054**	3,204,649
GC (%)	47.4	47.38	47.38	47.47
N50, bp	244,846	871,606	**874,674**	119,803
N75, bp	99,942	**451,084**	388,716	36,103
L50	9,454	**1,605**	2,133	7,176
L75	24,435	**3,576**	5,726	22,388
N's per 100 kb	0.39	0.6	0	0
Complete BUSCOs	2,698	2,068	2,711	**2,758**
Single-copy BUSCOs	1,889	1,940	**2,548**	2,235
Duplicated BUSCOs	809	128	163	523
Fragmented BUSCOs	106	86	**80**	**80**
Missing BUSCOs	130	780	143	**96**

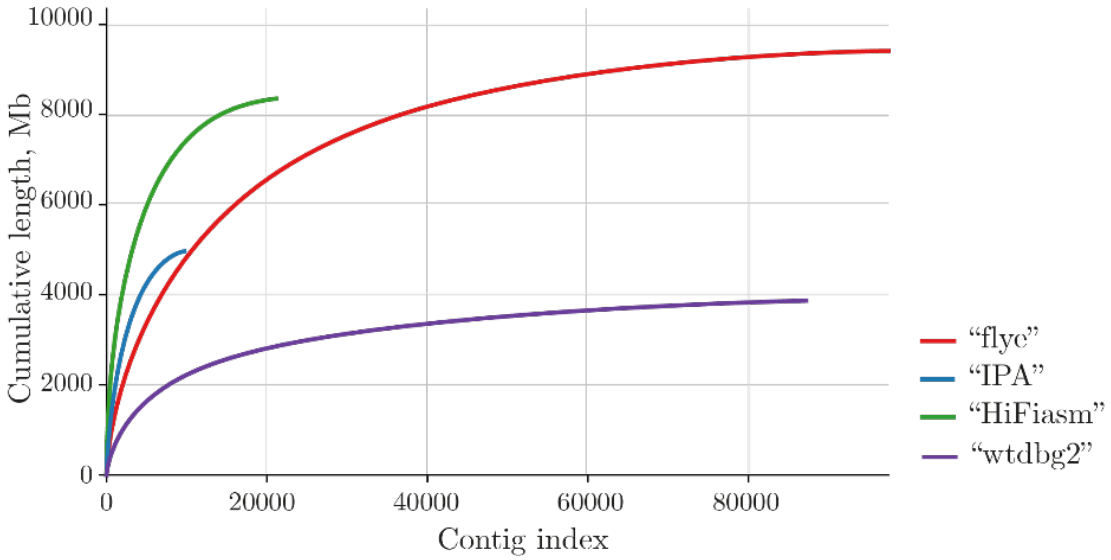

Figure 27 The cumulative plot of four assemblies of Amblyomma variegatum cell line generated in this study.

3.3.4. Targeted assemblies of *R. africae*

All ONT and PacBio HiFi reads were classified with kraken and reads assigned to the genus *Rickettsia* were extracted into separate files and assembled with Flye. The assembly of ONT reads resulted in an assembly slightly bigger than the reference genome (1.52 Mb vs 1.29 Mb), although the assembly was quite similar to the reference genome in terms of nucleotide identity and contigs arrangement (Figure 28). Red dashed lines in Figure 28 delineate the region of the reference genome missing from the assembly, which is approximately 55 Kb. PacBio reads yielded an assembly almost twice as big as the expected size (2.4 Mb) due to numerous repetitive regions as seen in Figure 28b. The same region of the reference genome was missing (Figure 28b).

There was an approximately 230 Kb region of the assembly (part of "scaffold_11") which was missing from the reference genome (Figure 28a, between blue dashed lines). Figure 29 shows the "scaffold_11" in more details, focusing on transition of the left "non-rickettsial" part of the scaffold to the right "rickettsial" part of the scaffold. Figure 29a shows the overview of the "scaffold_11"; the top track demonstrates that the right part of the scaffold had a good match within the reference *R. africae* genome, and the left part had no matches in it. Figure 29b focuses on the central region of the scaffold where the transition occurred. The top track shows the alignment of raw ONT reads which were taxonomically assigned to *Rickettsia* spp. and used to produce the *R. africae* assembly; the resulting assembly was 5-25 times covered by such reads. There were reads which continuously covered the discussed region. The second track shows the full set of raw ONT reads aligned to the *R. africae* assembly; there was an obvious drop in coverage between the left and the right parts of the assembly. The majority of reads, which were aligned to the left part of the scaffold, had been soft-clipped at the same position (Figure 29b). The reason for such alignment pattern is not clear, there is a chance that this scaffold was chimeric and resulted from the erroneous assembly. A search for matches of the left part of the scaffold in the 'nt' database did not result in any hits, except for six predicted genes (see below).

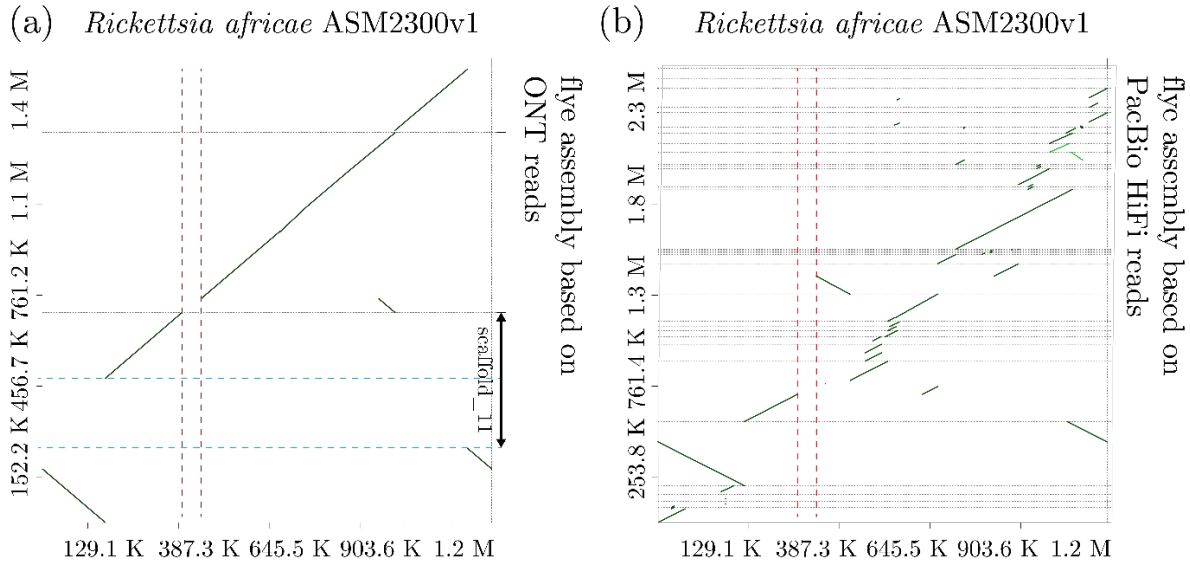

Figure 28 Dot-plot comparison of the reference Rickettsia africae genome with two assemblies based on ONT reads (a) and PacBio HiFi reads (b). Red dashed lines delineate the region of the reference genome missing from both assemblies. On the contrary, blue dashed lines in the panel 'a' confine the region in the "scaffold_11" of the assembly from ONT reads missing from the reference genome.

The annotation of the assembly on the ONT reads revealed 2113 CDSs in 1.52 Mb assembly. There were 127 annotated genes split by stop codons into two or more parts in the assembly of ONT reads, which might have resulted from the fact that the assembly was unpolished, thus this is not a final result. The same annotation approach found only 29 broken genes in the reference genome of *R. africae*.

Interestingly, the section missing from the reference genome region of "scaffold_11" (Figure 28a, between blue dashed lines, and Figure 29a) contained 154 predicted proteins according to Prokka annotation, although all of them were annotated as "hypothetical" (Figure 29a, red arrow region). This region had noticeably less dense gene coverage compared with the right part of the scaffold, and the genes were shorter (Figure 29c). The mean length of the subset of the genes from the left part of the "scaffold_11" was 373 bp, whereas the mean length of the genes from the right part was 497 bp (Figure 29c, dashed lines). Only six of these 154 hypothetical proteins found matches in the 'nt' database, and top hits for these six genes were from either tick or insect genomes (Table 3-9).

Table 3-9 Blast 'nt' hits of six hypothetical proteins from the R. africae assembly based on ONT reads

Predicted gene	**Blast top hit**	**Top hit ID**	**Identity, %**	**e-value**
ONT_Ra_01680	XM_037426675.1	*Rhipicephalus microplus* uncharacterized LOC119175721	76.1	0
ONT_Ra_01796	EU018131.1	*Rhipicephalus appendiculatus* isolate RAHD_87 Ruka SINE elements	82.6	1.70E-31
ONT_Ra_01797	XM_037414893.1	*Rhipicephalus microplus* calcium-dependent secretion activator	83.3	1.71E-65
ONT_Ra_01798	XR_005181519.1	*Rhipicephalus sanguineus* uncharacterized LOC119382390	83.7	2.00E-34
ONT_Ra_01809	XR_005191321.1	*Dermacentor silvarum* uncharacterized LOC119450099	89.0	8.40E-14
ONT_Ra_01819	OV277349.1	*Sicus ferrugineus* chromosome 2	100	0.006

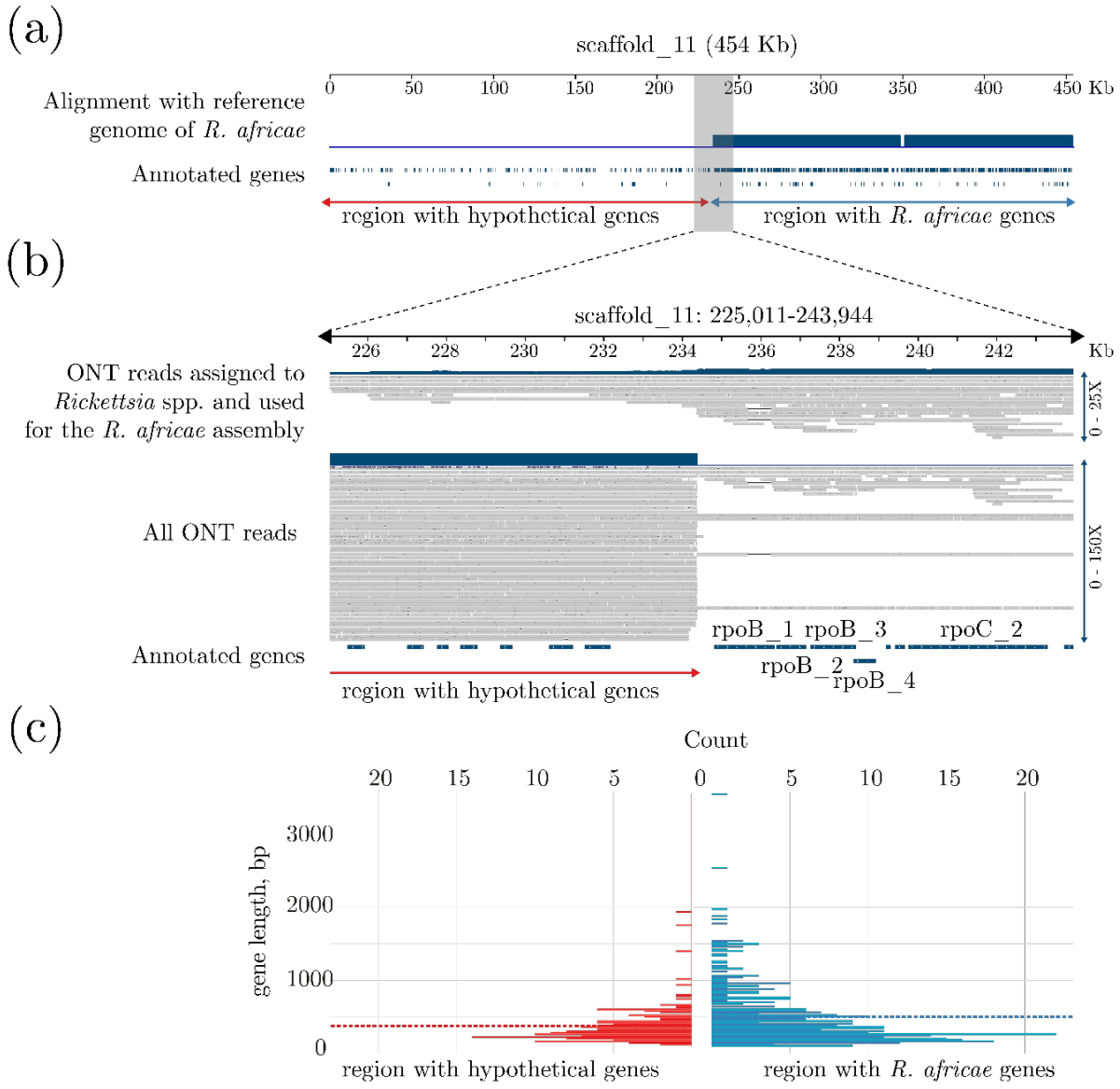

Figure 29 The schematic representation of "scaffold_11" from the assembly based on the ONT reads.

(a) Overview of the 454 Kb scaffold; top track demonstrates the alignment of the assembly with the reference Rickettsia africae genome; red arrow shoes the region with no match with the reference R. africae genome and only hypothetical proteins in the annotation; green arrow shows the region which matched the reference genome and its annotation.

(b) The central region of the "scaffold_11"; top track shows the alignment of ONT reads which were used for the assembly, the coverage range shown is 0-25X; the bottom track shows the alignment of all ONT reads, the coverage range shown is 0-150X.

(c) Comparison of gene length distribution between the left region with hypothetical genes and right region with R. africae genes. The dashed lines show the mean value for each dataset.

3.3.5. Searching for bacterial insertions within the resultant assemblies

3.3.5.1. Wolbachia and other bacterial sequences within the G. morsitans assembly

Figure 30 represents a schematic overview of contigs from the curated *G. morsitans* assembly "Gm_male_v2". The position of circles depends on the contig coverage and GC-content, color corresponds to the taxonomic affiliation of the contig, and the size is proportional to the size of the contig. The biggest contig assigned to *Anaplasmataceae (Wolbachia)* was 17 Mb, whereas the typical *Wolbachia* genome size should be 6-20 times smaller. The blast algorithm sometimes assigns the whole contig to the taxon even if only parts of the contigs correspond to the taxon if other parts of the sequence have no good hits. This contig contained three regions of *Wolbachia* motifs 3 Kb, 4 Kb and 130 Kb in size (Figure 31a). There were 37 contigs from 1.1 Kb to 17 Mb assigned to *Anaplasmataceae* by blastn algorithms; none of them contained *Wolbachia* regions longer than 130 Kb while previous studies claimed bigger segments (528 Kb and 484 Kb) incorporated into the tsetse fly genome. The third insertion was reported as only 2 Kb (Brelsfoard *et al.* 2014).

Another attempt to understand the nature of the *Wolbachia* sequences within the *G. morsitans* assembly was to assemble the bacterial reads separately from the host reads. This approach – classify reads according to taxonomical assignment, extract reads of taxon of interest, assemble these reads separately – proved to result in high quality complete or almost complete assemblies with many symbiotic bacteria. For example, *S. glossinidius* and *W. glossinidia* were assembled in full into 16 and 1 fragments, respectively (see Figure 25(a, b) and section 3.3.2 for details). At the same time reads assigned to *Wolbachia* were assembled in 50 contigs spanning 1.72 Mb with the 58X mean coverage, and this fragmented assembly was not similar to the *w*Gmm genome (Figure 25c) or any other published *Wolbachia* genome.

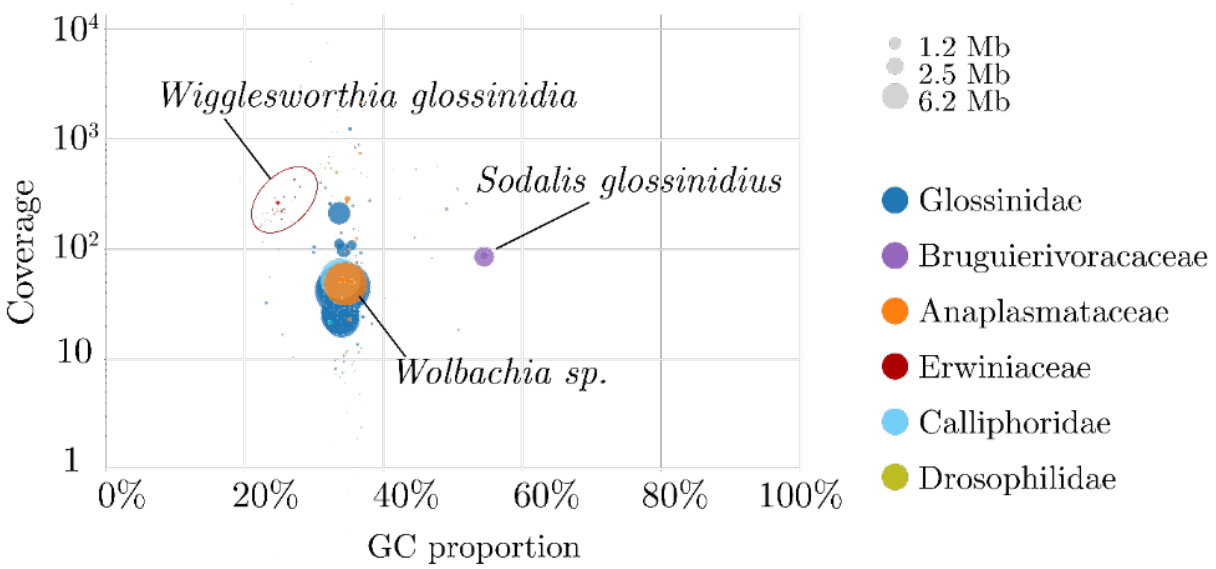

Figure 31 'Blobplot' representation of the "Gm_male_v2" assembly, contigs are placed in the graph depending on the coverage and GC content. Circle sizes represent size of contigs. Circle colours represent taxonomic assignment of contigs on the family level.

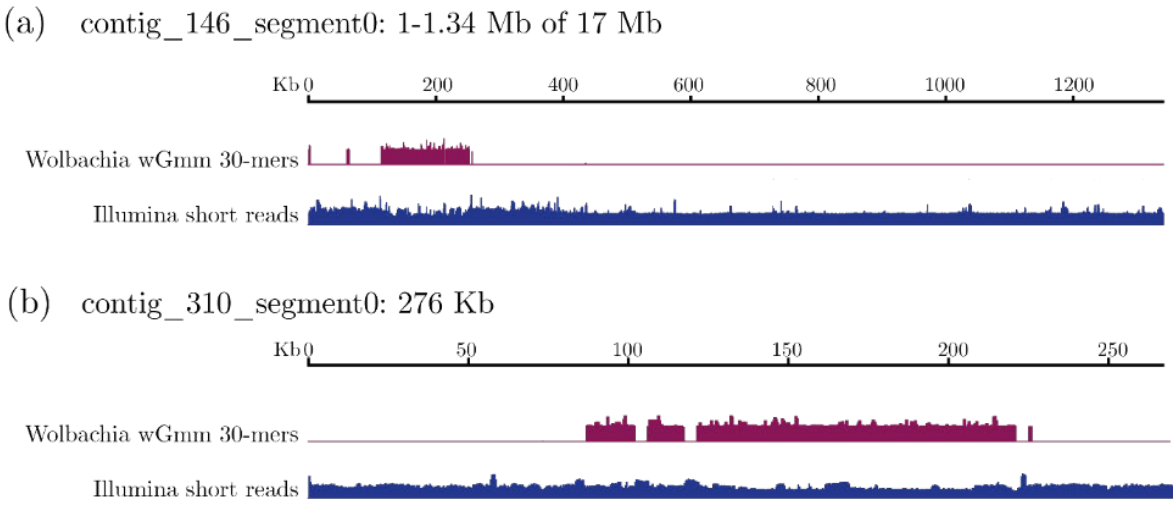

Figure 30 (a) Left flank of the 17 Mb contig from G. morsitans "flye" assembly with three Wolbachia motifs; the top dark-red track shows coverage of the contig with 30-mers derived from Wolbachia wGmm reference assembly, the bottom blue track shows coverage with short Illumina reads. (b) The 276 Kb contig with the Wolbachia motif in the middle.

The fragmented nature of the *Wolbachia* assembly, its dispersion across *G. morsitans* contigs, and low identity with the *Wolbachia* assemblies suggest that the *Wolbachia* sequences did not belong to the living symbiont and were incorporated into host genome. However, these arguments can serve only as indirect evidence and further experiments are required to understand the nature of these *Wolbachia* fragments. The pattern of *Wolbachia* sequence distribution is different from those described by Brelsfoard and coauthors (Brelsfoard *et al.* 2014), which might indicate a further degradation of insertions within the host genome.

3.3.5.2. R. africae sequences within the A. variegatum assembly

Two assemblies and raw reads were aligned to the reference *R. africae* assembly as shown in the Figure 32. Most of the assembly matched the reference genome except for one region shown as two flanking regions highlighted with grey boxes (the reference *R. africae* genome is circular, and is split within this region for the illustration purposes). This region was approximately 55 kb and contained genes *tolB*, *rpoH*, *tolQ*, *tolR*, *asd*, *hslU*, *lpxB*, *mltG*, *topA*, *ruvX*, *mutS* and *rpiB*, as well as genes without any assigned function (not labelled in Figure 32). The lack of *rpoH* and *mutS* genes (highlighted in red) corresponded well with the lack of amplification of these genes in *A. variegatum* AVL/CTVM13 and AVL/CTVM17 cell lines as mentioned in the introduction to this chapter (see section 3.1.2).

Aligning 30-mers derived from the *R. africae* reference genome and the collection of 72 published complete genomes of *Rickettsia* spp. to the "flye" assembly revealed 28 contigs, from 30 kb to 323 kb, which were entirely or almost entirely covered with rickettsial sequences. Such contigs could be equally a part of the symbiotic *Rickettsia* genome or of the rickettsial insertion, so they were of no use for the current analysis. Another two contigs (contig_69393, 275 kb, and contig_16902, 977 kb) had a distinct pattern with only a fraction of the contig matching rickettsial sequences; the rest of the contig only had matches to different genes of hard ticks (Figure 33, tracks "*R. africae* k-mers" and "*Rickettsia* spp. k-mers"). There was no substantial change in

coverage between the “tick” part of these contigs and the “rickettsial” part (Figure 33, tracks “ONT reads” and “PacBio reads”).

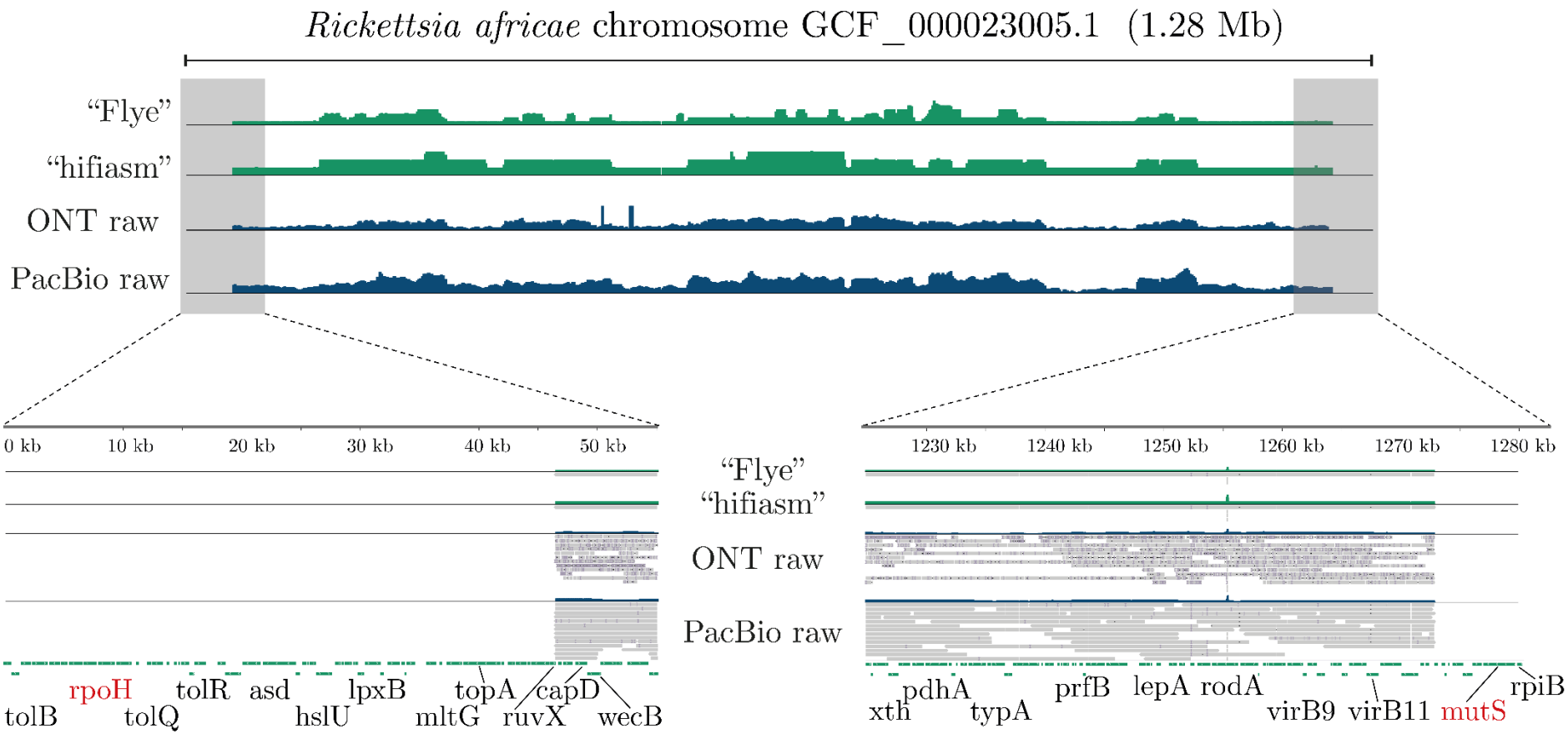

Figure 32 Rickettsia africae reference assembly (GCF_000023005.1) aligned with two assemblies ("flye" and "hifiasm") and raw ONT and PacBio reads derived from the A. variegatum cell line AVL/CTVM17. The overview of the full R. africae chromosome with four coverage tracks is shown on the top. The bottom row shows left and right flanking regions with a drop of coverage in all four tracks. Two genes in red – rpoH and mutS – could not be amplified with PCR assays on tick cells and were clearly not present in the genomic data.

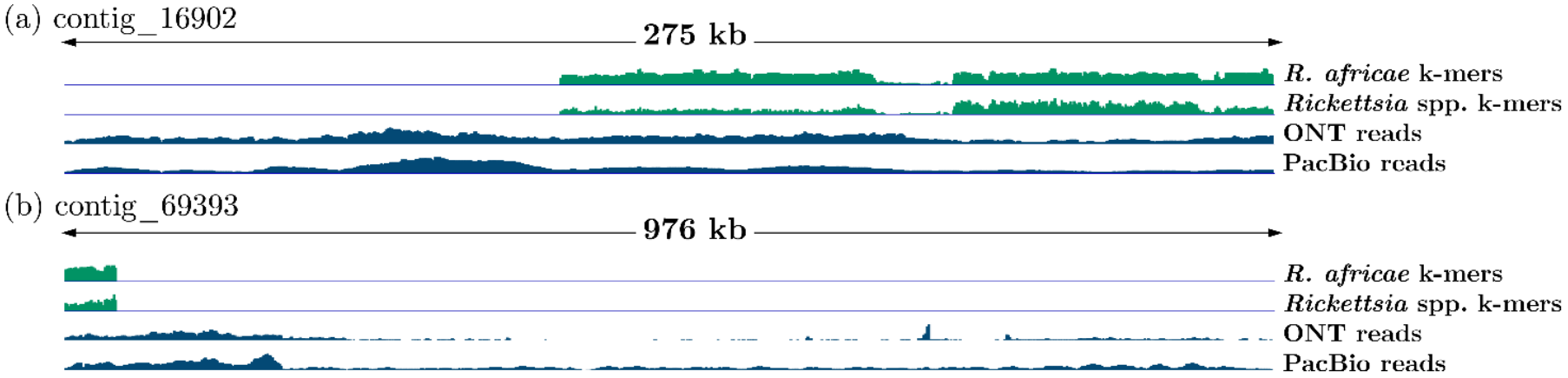

Figure 33 Two contigs from the "flye" A. variegatum cell line assembly which had indicated a partial match to rickettsial genome compendium. Top track on each plot indicates k-mers matched to the published genome of R. africae, and second from top track indicates k-mers matched to a compendium of all complete Rickettsia genomes from NCBI. Two bottom tracks indicate the coverage by ONT and PacBio HiFi reads.

3.4. Discussion

Two genome assemblies of important *Arthropoda* species were generated in this study. The chromosome level assembly of *G. morsitans*, a vector of trypanosomiases, had a higher continuity and completeness than currently available in the public databases. The first assembly for *A. variegatum* and for the genus *Amblyomma*, vectors of numerous human and livestock pathogens, was obtained. Due to the large size of the genome and its complex structure, the *A. variegatum* assembly needs further curation, although formal quality assessment (the best BUSCO score was 94%) indicates that the assembly can already be used as a reference genome for tasks such as transcriptomic experiments or pathway reconstruction. In this study the assemblies were used to confirm or reject the hypotheses of the rickettsial insertion in the *A. variegatum* nuclear genome and *Wolbachia* insertion in the *G. morsitans* genome.

While the transfer of genetic material from *Wolbachia* to its hosts has been confirmed in multiple symbiotic systems (e.g., tsetse flies (Doudoumis *et al.* 2012), pea aphids (Nikoh *et al.*, 2008), mosquitoes (Klasson *et al.*, 2009), beetles (Aikawa *et al.*, 2009)), *Rickettsia* bacteria are not known to undergo integration into the host nuclear genomes. The *A. variegatum* cell line did not show any physical presence of *R. africae*, although was positive in molecular screening assay based on amplification of marker genes. This study attempted to confirm the nature of rickettsial sequences in the tick cells by aims of bioinformatic analysis.

3.4.1. Rickettsial insertion in the *A. variegatum* nuclear genome

Two approaches were used to analyse *R. africae* sequences from the *A. variegatum* assembly. First, all reads were filtered according to their taxonomical assignment and only "rickettsial" reads were used for the assembly. The resulting assembly was compared with the reference genome of *R. africae*. The comparison

revealed two interesting differences between the assembly obtained in this study and the reference assembly of *R. Africae*: an approximately 55 Kb region of the reference genome which was absent from our assembly and an approximately 230 Kb region which was on the contrary present in our assembly and not present in the reference *R. africae* genome (Figure 28).

The 55 Kb region absent from our assembly contained several gene essential for *Rickettsia*, which are found in all other complete assemblies of *Rickettsia* spp. The genes *tolB*, *tolR* and *tolQ* are described as integral membrane proteins of the Tol–Pal system essential for active transport across the outer membrane and to maintain cell envelope integrity (Cascales, Lloubès, and Sturgis 2008). Although some authors claimed that *Rickettsia prowazekii* lacked this system (Sturgis 2001), a later complete assembly of *R. prowazekii* of a better quality contains three listed genes (GCA_000277165.1 ASM27716v1). All other complete rickettsial assemblies (70 assemblies listed in Supplementary table 4) also contained two copies of all three genes with the exception of two assemblies of *Rickettsia parkeri*, which lacked *tolR*. The *rpiB* gene was absent only in three assemblies – two *Rickettsia canadensis* and *Rickettsia* sp. from *Bemisia tabaci*. All other genes which fall into the absent region (*rpoH*, *asd*, *hslU*, *lpxB*, *mltG*, *topA*, *ruvX*, and *mutS*) are present in all complete rickettsial assemblies. Such result suggests that these genes are essential for rickettsial metabolism and are unlikely to be absent in fully-functional symbiotic bacteria.

The 230 Kb region which is present in our assembly and not present in the reference genome of *R. africae* had no matches within bacterial assemblies in the 'nt' database. The annotation revealed only hypothetical genes in this region, of which only six had any similarities with known genes (Figure 29a). These six genes were similar to *Arthropoda* genes, and not bacterial genes (Table 3-9), which might indicate that this scaffold captured the transition between host genome and bacterial insertion. However, it must be noted that there were many reads which aligned to this region

and were clipped at the point of the transition, so theoretically this scaffold might be the result of misassembly.

The second approach was to find rickettsial fragments in the whole *A. variegatum* assembly. Two *A. variegatum* assemblies were aligned with the *R. africae* reference genome and all possible 30-mers from complete *Rickettsia* assemblies to find contigs or parts of contigs with bacterial motifs. Several contigs wholly matched rickettsial sequences, and thus were of no interest for further analysis, they could equally possibly come from the living symbiont or from the insertion. Two contigs of 275 Kb and 976 Kb were found were found to comprise tick sequences flanked by rickettsial sequences (Figure 33). It is important to note that there was no abrupt change in raw reads coverage within these contigs as is often the case between the host genome and symbiont genome; bacteria are often present in several or many copies per host cell. A good example of such difference in coverage is the *S. glossinidius* and *W. glossinidia* and the host sequences (Figure 30). Most of the fly contigs lay in the range of 20-40X coverage with exception of several outliers, whereas *S. glossinidius* contigs were covered approximately 85 times, and *W. glossinidia* contigs were covered in the range of 200-400X (Figure 30). This means that in case of a misassembly when bacterial sequence is erroneously joined with the host sequences, there is likely to be a clear discrepancy in coverage.

Many indirect arguments have been collected in this study supporting the hypothesis that rickettsial sequences are incorporated into the *A. variegatum* cell line genome. Such features as the lack of essential genes, contigs consisting of bacterial and tick parts, similar coverage of bacterial and tick fragments indicate that the genetic material was more likely to originate from the insertion rather than the living symbiont. However, there is a need to experimentally confirm the presence of the rickettsial insertion in the *A. variegatum* genome, which could be done with *in situ* hybridisation of insertion-specific probes with the tick chromosomes.

3.4.2. *Wolbachia* insertion in the *G. morsitans* nuclear genome

Wolbachia fragments were likely to originate from the insertion based on such features as the high level of fragmentation of *Wolbachia* assembly, the failure to assemble a *Wolbachia* genome of the quality similar to other two cytoplasmic symbionts, the dissimilarity of the assembled *Wolbachia* contigs from the *Wolbachia* *w*Gmm assembly and all other published *Wolbachia* assemblies. Interestingly, the pattern of the *Wolbachia* fragments placement in the *G. morsitans* assembly was different from the described by Brelsfoard and coauthors (Brelsfoard *et al.* 2014): there were more *Wolbachia* fragments and they were substantially shorter. The higher level of dispersion of the *Wolbachia* motifs within the *G. morsitans* assembly might indicate an ongoing degradation of the inserts within the host genome. The analysis of the main marker genes used for taxonomical assignment of *Wolbachia* revealed that the obtained sequences originated from some *Wolbachia* from the supergroup A, but not from *w*Gmm as was described by Brelsfoard and co-authors (Brelsfoard *et al.* 2014).

The original idea was to use the improved, more contiguous *G. morsitans* assembly as a testbed for bioinformatic search for bacterial insertions, since the *Wolbachia* insertion into the *G. morsitans* genome was well known and described in literature (Brelsfoard *et al.* 2014; Doudoumis *et al.* 2013; 2012). Unfortunately, it turned out that the sequences of the described insertions are not publicly available. This made the direct detection of the *Wolbachia* insertion impossible. Thus, the analysis of the *Wolbachia* sequences within the *G. morsitans* assembly was based on the sequence comparison of cytoplasmic *Wolbachia*, *Wolbachia* symbiont from *G. morsitans* (*w*Gmm) in particular, and the localisation of bacterial patterns within the host assembly.

It is important to mention that the tsetse fly colony maintained at the LSTM and used in this study was affected by heat and went through a bottle neck (personal communication with Dr Lee Haines). This heat stress may have affected the *Wolbachia* presence in the colony as environmental stresses can reduce or kill *Wolbachia*

populations (Ulrich *et al.* 2016). There is a hypothesis that it is the induction of phage WO that triggers loss of *Wolbachia* (Sarah R. Bordenstein and Bordenstein 2011) and *Wolbachia* chromosome undergoes degradation during the bacteriophage lytic cycle (Seth R. Bordenstein *et al.* 2006). The heat shock is known to cause DNA breaks (Velichko *et al.* 2012) and alter membrane integrity (Niu and Xiang 2018), which couples with the *Wolbachia* DNA degradation might lead to further incorporation of *Wolbachia* genetic material into host genome. So the heat stress event might explain the differences between *Wolbachia* insertion patterns observed in this study and those published previously (Brelsfoard *et al.* 2014).

3.4.3. Limitations of the bioinformatic approach for insertion search

The bioinformatic approach for bacterial insertions analysis has its limitations. First of all, this analysis depends on the quality of reference assemblies. For example, neither the *R. africae* nor the *Wolbachia w*Gmm assemblies were derived from axenic cultures but instead were derived from mixed samples: infected Vero cells in the case of *R. africae* (Fournier *et al.* 2009) and *G. morsitans* ovaries in the case of *Wolbachia* (Brelsfoard *et al.* 2014). These assemblies were based on short-read data, which means a high probability of misassemblies and contamination by host sequences. Such contaminations can be minimised by relying on complete genomes of other species of *Rickettsia* and *Wolbachia*, as complete circularised bacterial assemblies have a smaller chance of contamination.

Considering that more and more genomes of animals are sequenced nowadays, the described analysis of bacterial fragments within host assemblies should be helpful for at least preliminary assessment of the data, although the results of the *in silico* analysis should be further confirmed by experimental techniques such as *in situ* hybridisation.

3.4.4. Future work and the potential use of the resultant assemblies

By employing a bioinformatic assessment technique for identifying the source of bacterial sequences, the investigation of bacterial insertions into host genomes can be simplified, leading to a better understanding of the occurrence rate and patterns of such insertions, as well as the evolutionary destiny of the inserted bacterial fragments.

The high quality assemblies will produce a better genomic annotation and could also result in the discovery of novel genomic structures such as tandem repeats (Torresen *et al.* 2017), better resolution of coding and non-coding regions, improving gene models (Satou *et al.* 2008) and identification of previously unknown genes and pathways (E. S. Rice *et al.* 2017). Assemblies obtained from a single individual and based on long reads are helpful for resolving haplotigs and thus reducing erroneous duplications from genomes (Guan *et al.* 2020).

The assemblies produced here can be further improved. While we were able to produce chromosome-size scaffolds for *G. morsitans*, they remain unlocalised to a particular chromosome. The chromosome set within the *G. morsitans* species complex is not stable (Gariou-Papalexiou *et al.* 2002). The karyotype of *G. morsitans* consists of two pairs of autosomes (L1 and L2), a pair of sex chromosomes (XX for female individuals and XY for males) and from zero to seven S-chromosomes (supernumerary) or B-chromosomes (Gariou-Papalexiou *et al.* 2002; Itard 1973).

The assembly produced for *G. morsitans* in this project should be helpful for insights about sex determination in tsetse flies and *Diptera* in general. Currently, available assemblies originate either from a mixture of sexes or from female flies; the assembly generated in this study was obtained from a single male fly, so the comparison between female and male assemblies should lead to better understanding of the Y chromosome structure.

4. Chapter IV Single-cell transcriptomics of tick cell lines: preparing materials and data

4.1. Introduction

Single-cell sequencing experiments require a lot of prior preparation and data collection for successful analysis, especially when working with non-model organisms (Lafzi *et al.* 2018) For this study, it was necessary to choose appropriate tick cell cultures and symbiotic bacteria capable of growing in the selected cells. It was necessary to take into account such factors as availability and variability of cell cultures, the economic and epidemiological importance of tick species, and the range of symbionts maintained in the culture.

4.1.1. Cell lines in tick-pathogen research

The research using field ticks or laboratory tick populations has many limitations. In the case of using field ticks, the most serious difficulty is the limited accessibility of material and the seasonality of material collection; this is why field collection is mainly used in population screening studies. Maintaining and propagating tick colonies in laboratory conditions requires a complicated setup and is therefore costly and time-consuming (Salata *et al.* 2021). In order to be successfully colonised, live ticks require either live animals to feed on or artificial feeding assays. Using live animals for feeding is expensive and raises ethical concerns, as they might suffer from skin inflammation, anaemia, and other consequences of tick infestation (Krober and Guerin 2007). Designing an *in vitro* assay for tick feeding is challenging because it has to be developed separately for each species because of differences in feeding time and structure of mouthparts (Krober and Guerin 2007). Furthermore, ticks require particular chemical and physical stimuli to start the feeding process (Guerin *et al.* 2000) and a system that permits some ticks to feed and detach when replete and not

at the same time for all (Krober and Guerin 2007). For example, Kuhnert and coauthors induced attachment of ticks to a feeding membrane "with combinations of host hair, tick faeces, a bovine pelage extract and a synthetic aggregation-attachment pheromone mixture" (Kuhnert, Diehl, and Guerin 1995). It should also be noted that ticks feed and develop slowly, some ticks are able to complete their life cycle in six weeks under favourable conditions (Senbill *et al.* 2018), for other species (e.g., *I. ricinus* in cold areas of Europe) it might take 3-11 days for each blood meal and 3–6 years for the whole life cycle in nature, limiting the scope of experiments on live ticks (Vechtova *et al.* 2020).

Tick cell lines are cheaper to maintain; they grow faster, in many cases doubling the number of cells approximately every week, and can be easily transferred between labs. Tick cell lines also do not require a considerable amount of paperwork for the ethical justification of experiments. Therefore, tick cell lines play a significant role in tick and tick-borne pathogen research, including tick biology, host-pathogen relationships, genetic manipulation, genomics and proteomics (Lesley Bell-Sakyi *et al.* 2018).

The first continuous tick cell lines were established in 1975 (Varma, Pudney, and Leake 1975). Nowadays, there are more than 70 tick cell lines available, and it is possible to use them to propagate a wide range of pathogens vectored by ticks: 61 viruses, 18 bacteria and at least three protists have been described (Salata *et al.* 2021; Lesley Bell-Sakyi *et al.* 2012). This expanded the tick-borne disease research as many agents (e.g., some *Anaplasma*, *Rickettsia* and *Borrelia* species) cannot be grown *in vitro* in any other culture system (Lesley Bell-Sakyi *et al.* 2007). Although not a replacement for the whole organism, tick cell lines enable studies at the cellular and molecular level and provide a more accessible, more ethical, and less expensive *in vitro* alternative to *in vivo* tick feeding experiments (Al-Rofaai and Bell-Sakyi 2020). Of course, there are limitations to research using tick cell lines, such as lack of information about pathogen dynamics, the location and route of a pathogen within a host, and symptoms of infection.

Tick cell lines are established from whole embryos, larvae or nymphs without attempts to select particular tissue types, so cell lines might consist of several cell types (Bell-Sakyi 1991; Bell-Sakyi *et al.* 2007). Different cell types are believed to be essential for cell line survival as all attempts to clone cells from cell lines or individual organs failed (Munderloh *et al.* 1994). Cells in cultures divide relatively slowly, grow to high densities (10^6 – 10^7 cells per ml), and many lines do not require regular subculture, making them particularly suitable for the isolation of slow-growing microorganisms (Lesley Bell-Sakyi *et al.* 2007).

Tick cell lines are widely used in tick-pathogen research and have drastically broaden the field (Lesley Bell-Sakyi *et al.* 2007). There are studies that have shown that cell lines may serve as an *in vitro* method for examining the ways in which tick-borne pathogens interact with tick physiology (Mateos-Hernandez *et al.* 2021) and participate in bacterial interplay (Solyman *et al.* 2022; Skinner *et al.* 2022).

4.1.2. Overview of species used in the study

Ixodid ticks have a rigid chitinous shield covering the entire dorsal surface of the adult male, reflected in their common name – hard ticks. *Ixodidae* may be one-, two- or three-host species depending on how many species of host animals they attach on to feed. There are at least 702 species of *Ixodidae* ticks described to date (Guglielmone *et al.* 2010). Seventy-five species affect the human economy, being either a vector of human and livestock diseases or causing severe reactions such as allergies or toxicosis (Jongejan and Uilenberg 2004). Tick-borne diseases affect at least three aspects of the human economy: human health, *e.g.,* Lyme disease, tick-borne encephalitis; damage to livestock, *e.g.,* bovine anaplasmosis, babesiosis and theileriosis; and diseases of pets, such as cats and dogs, who can suffer from numerous pathogens transmitted by ticks (Pfaffle *et al.* 2013). Most genomic and transcriptomic studies are carried out with *Ixodes* species as they are widespread and thus important for North America (*I. scapularis*) and Europe (*I. ricinus*) as well as their cell lines have been widely distributed since (Jongejan and Uilenberg 2004).

One of the limiting factors when choosing tick cell lines is the availability of a high-quality reference genome assembly and a well-curated annotation. During the planning of the experiments described below, the availability of public genomic data was a very important criterion as hard ticks have very large genomes spanning from 1.7 to 2.8 Gb, and the project budget did not include *de novo* tick genome assembly and annotation on such a scale. To date, the genomes of only nine species of ticks of the family *Ixodidae* have been sequenced (see Table 3-1).

4.1.2.1. Ixodes scapularis

I. scapularis, also known as the deer tick or black-legged tick, is the primary tick vector in North America of Lyme disease as well as the relapsing fever spirochete *Borrelia miyamotoi*, causative agents of human granulocytic anaplasmosis and babesiosis; and Powassan encephalitis virus (R. J. Eisen, Eisen, and Beard 2016; Thapa, Zhang, and Allen 2019). This tick has a three-host life cycle which might last for up to four years (Vandyk *et al.* 1996). Lyme disease has been on the rise during the last decades with 30-40 thousand cases reported each year (Kugeler *et al.* 2021). Even this number is believed to be an underestimate, with the real number of cases being up to 10 times higher (Kugeler *et al.* 2021). *I. scapularis* is also spreading to the north into Canadian habitats where it has not previously occurred; the number of areas with *I. scapularis* has increased from one to seven in ten years from 1993 to 2003 (Ogden *et al.* 2005).

Due to its importance, *I. scapularis* was the first hard tick chosen to be sequenced; the sequencing project was initiated in 2004 (Hill and Wikel 2005), some intermediate results were made public a few years later (Pagel Van Zee *et al.* 2007) and the assembly was completed in 2016 (Gulia-Nuss *et al.* 2016). This assembly has supported many molecular studies such as transcriptomic experiments which are not possible without a good-quality reference genome. Despite its importance, no mitochondrion contig was present in the assembly.

4.1.2.2. Ixodes ricinus

I. ricinus, or the castor bean tick, is the most important tick vector in Europe and Northern Africa (Cramaro *et al.* 2015), and its distribution has recently expanded to the north and to higher altitudes (Rizzoli *et al.* 2014; Jaenson *et al.* 2012). It transmits many pathogens relevant to humans and livestock: *Borrelia burgdorferi sensu lato*, the agent of Lyme disease, tick-borne encephalitis virus, *Anaplasma* spp, *Babesia* spp. and *Rickettsia* spp. (Rizzoli *et al.* 2014). *I. ricinus* has a three-host life cycle, involving a wide range of vertebrate hosts, including birds, lizards and mammals of all sizes (Mysterud *et al.* 2015). The genome assembly of *I. ricinus* is smaller compared to other hard ticks, only 514 Mb according to the only available assembly as for November 2021 (Cramaro *et al.* 2015), although the flow cytometry-based analysis revealed a haploid genome size of 2.65 Gb for *I. ricinus* ticks and 3.80 Gb for the *I. ricinus* cell line" (Cramaro *et al.* 2017).

4.1.2.3. Rhipicephalus microplus (formerly Boophilus microplus)

R. microplus is considered the most important livestock pest in tropical and subtropical areas of the world, where it infests various livestock species, including cattle, horses, goats, sheep, and pigs (Jongejan and Uilenberg 2004). It is known as the Asian blue tick, Australian cattle tick, southern cattle tick, Cuban tick, Madagascar blue tick, and Porto Rican Texas fever tick; the variety of names represents its wide geographical distribution. *R. microplus* co-evolved with Asian bovines and spread across tropical and sub-tropical regions due to the global migration of *Bos taurus* cattle breeds for the dairy industry during the 18th and 19th centuries (Tabor *et al.* 2017). *R. microplus* transmits anaplasmosis and babesiosis and yearly global losses were estimated at US$13–18 billion at the end of the last century (de Castro 1997). *R. microplus* has been eradicated from the US since the 1940s and is controlled by Permanent Quarantine Zone at the border of the US and Mexico (Knolhoff and Onstad 2014). Still, due to the global warming, *R. microplus* is expected to migrate to the

north more intensively and sneak into northern territories via wildlife corridor, making the species more relevant for US economics (Marques *et al.* 2020; Showler, Pérez de León, and Saelao 2021). Despite its high impact on many countries, less omics research is carried out with this species partly because the high-quality genome assembly was only published last year (Jia *et al.* 2020). The first version of *R. microplus* genome was published in 2012 although due to the large size and high complexity of the genome the assembly was not widely used for omics research; even the coding part of the genome wasn't sequenced with the satisfactory coverage (0.9X) (Bellgard *et al.* 2012). The high quality assembly based on long reads was published recently (Jia *et al.* 2020). There are two independent submissions of mitochondrial genomes for *R. microplus*. One comes from the recent genome assembly, CM023492.1 (Jia *et al.* 2020), the second was sequenced previously separately from the nuclear genome, NC_023335.1, which is included in the RefSeq database (Burger, Shao, and Barker 2014).

4.1.3. Tick-borne pathogens

4.1.3.1. Ehrlichia minasensis

Bacteria from the genus *Ehrlichia* are obligate intracellular pathogens of mammals; some species are capable of infecting humans (Cabezas-Cruz *et al.* 2016). The natural cycle of *Ehrlichia* spp. pathogens involves an arthropod vector and a definitive mammalian host and might include incidental hosts such as humans (T. B. Saito and Walker 2016). Human ehrlichioses usually appear as non-specific fever and malaise, although might result in serious complications such as meningoencephalitis, pneumonia, hepatocellular death, erythrophagocytosis, myeloid hyperplasia, and others (Saito and Walker 2016).

E. minasensis is a pathogen of cattle initially isolated into IDE8 cells, and can be maintained in tick cultures of *I. scapularis, I. ricinus, R. microplus, Rhipicephalus decloratus, Ornithodoros moubata* or canine macrophages (Cabezas-Cruz *et al.* 2016; Zweygarth *et al.* 2013). Bacteria form small colonies inside vacuoles in the cytoplasm

surrounded by mitochondria and cisternae of rough endoplasmic reticulum (Cabezas-Cruz *et al.* 2013). The genome of *E. minasensis* is 1.4 Mb with 30.36% GC-content (Cabezas-Cruz *et al.* 2015). There are two public draft assemblies available as for November 2021; both are based on short reads and are fragmented (Cabezas-Cruz *et al.* 2015; Aguiar *et al.* 2019).

4.1.3.2. Spiroplasma sp.

The genus *Spiroplasma* belongs to the phylum *Tenericutes,* class *Mollicutes*. The cells are pleomorphic, of various shape and size, typically thin filaments of 100-200 nm in diameter and 3-5 µm in length; they form colonies within vacuoles in host cells (Williamson *et al.* 2015). *Spiroplasma*s are found in a wide range of hosts among plants and arthropods; some of them are considered to be pathogenic, others seem to be harmless or mutualistic (Binetruy *et al.* 2019). The first *Spiroplasma* from ticks was isolated in in 1964 from *Haemaphysalis leporispalustris* (Clark 1964), later in 1981 found in *Ixodes pacificus* and was named *S. ixodetis* (Tully *et al.* 1981). Since then, this species was noticed in many other arthropods (Binetruy *et al.* 2019). *S. ixodetis* is transovarially transmitted and might induce early male killing, although the male killing trait was only observed in Coleoptera and Lepidoptera and not in ticks (Engelstädter and Hurst 2009). The *Spiroplasma sp.* DMAR11 used in this study was isolated from *Dermacentor marginatus* ticks from Spain; based on 16S rDNA, ITS, and rpoB gene sequences, this strain is closely related to *Spiroplasma sp.* Bratislava1 from *I. ricinus* and *S. ixodetis* (Palomar *et al.* 2019).

4.1.3.3. Rickettsia raoultii

R. raoultii was first isolated from *Dermacentor* ticks collected in Russia and France, later found in many Asian countries such as China, Thailand and Kazakhstan (Mediannikov *et al.* 2008; Parola *et al.* 2013). These bacteria are Gram-negative, non-motile, rod-shaped obligate intracellular endosymbionts from the spotted fever group rickettsias (Mediannikov *et al.* 2008). At first, the species was not linked to any human

diseases, although later it was found to be associated with tick-borne lymphadenopathy in Europe and China (Switaj *et al.* 2012; Foldvari, Rigo, and Lakos 2013; Jia *et al.* 2014) and meningeal syndrome and neurological abnormalities (Igolkina *et al.* 2018; Dong *et al.* 2019). *R. raoultii* has varying growth rates in cell lines derived from different tick species (*R. microplus*, *Rhipicephalus sanguineus* and *I. scapularis*) (Husin *et al.* 2021).

4.1.3.4. St. Croix River virus

St. Croix River virus (SCRV) is an endogenous tick virus belonging to the genus *Orbivirus* within the family *Reoviridae*. It can only infect tick cells and was first detected in an *I. scapularis* cell line (Attoui *et al.* 2001), although later its presence was reported in cell lines derived from another tick genus from a different continent (Lesley Bell-Sakyi and Attoui 2016; M. P. Alberdi, Dalby, *et al.* 2012). The genome consists of 10 segments spanning 18.5 Kb, and phylogenetic analysis revealed that SCRV is ancestral to other tick- and insect-borne viruses (Lesley Bell-Sakyi and Attoui 2013). Since recently it is believed that most tick cell lines bear endogenous viruses, although most of them remain uncharacterised (Bell-Sakyi and Attoui 2013).

4.1.4. Tick cell lines and bacterial pathogens available at the Tick Cell Biobank

The TCB maintains a wide range of tick cell lines including those for the above-mentioned species. At the time of writing, there are five cell lines derived from *I. scapularis*, four from *I. ricinus* and nine from *R. microplus* according to the TCB website ('Tick Cell Lines - Liverpool Shared Research Facilities - University of Liverpool' 2021).

Table 4-1 shows which bacteria can be maintained in each of the three tick cell lines used in the present study. The *R. microplus* cell line BME/CTVM23 (M. P. Alberdi, Nijhof, *et al.* 2012) and the *I. scapularis* IDE8 (Munderloh *et al.* 1994) cell lines both support growth of at least three of four pathogens (M. P. Alberdi, Dalby, *et*

al. 2012; Palomar *et al.* 2019); L. Bell-Sakyi, personal communication), while IRE/CTVM20 (Lesley Bell-Sakyi *et al.* 2007) is known to support three of the bacteria (Palomar *et al.*, 2019; L. Bell-Sakyi personal communication), and there is no published data on *R. raoultii* in this cell line. *A. marginale* and *E. minasensis* are phylogenetically and phenotypically very close, they both belong to the same family *Anaplasmataceae* within the *Alphaproteobacteria*, and form a similar pattern of infection – both species reside within vacuoles in the cytoplasm where they form colonies. *R. raoultii* also belongs to the *Alphaproteobacteria* although its infection process differs from *A. marginale* and *E. minasensis* as the bacterial cells lie freely in the cytoplasm of the host cell. *Spiroplasma* spp. are phylogenetically distant from the other listed bacteria, they also reside within vacuoles in the host cell cytoplasm forming large colonies.

Table 4-1 The ability of the bacteria to grow in tick cell lines. '+' means that pathogens can be cultivated in the cell line, '—' means that this pathogen cannot be grown in the cell line, '?' means that there is no data for such combination.

Tick cell line | *Symbiont*	*Rickettsia raoultii*	*Ehrlichia minasensis*	*Spiroplasma spp.*
R. microplus BME/CTVM23	+	+	+
I. ricinus IRE/CTVM20	?	+	+
I. scapularis IDE8	—	+	+

4.1.5. Aims and objectives

The primary aim of the chapter is to develop an experimental design that involves the assessment of tick cell lines, selection of bacterial symbionts, and determination of the appropriate time points for infection. In order to achieve this goal, several objectives have been outlined. Fisrtly, the phenotypic heterogeneity of tick cell lines will be evaluated. Secondary, bacterial growth pattens in the chosen tick lines will be studied. Finally, the design of the transcriptomic experiment will be formulated using the collected information.

Additionally, the chapter outlines the collection of all genomic information required for subsequent bioinformatic analysis.

4.2. Materials & methods

4.2.1. Tick cell line maintenance

The TCB at the University of Liverpool provided all tick cell lines and bacterial species (Lesley Bell-Sakyi *et al.* 2007; 2018). The tick cell lines and bacteria used in the study are described in Table 4-2 and Table 4-3. The IDE8 cell line was used by kind permission of Prof. Ulrike Munderloh, University of Minnesota.

Culture medium was prepared freshly on the day of use. The BME/CTVM23 cells were maintained in L15 (Leibovitz) medium supplemented with 10% tryptose phosphate broth (TPB), 20% fetal bovine serum (FBS), 2 mM l-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin (complete L-15). IDE8 cells were maintained in L15-B medium (Munderloh and Kurtti 1989) supplemented with 10% TPB, 10% FBS, 0.1% bovine lipoprotein (MP Biomedicals) and L-glutamine and antibiotics as above (complete L-15B). IRE/CTVM20 cells were maintained in a 1:1 mixture of complete L-15 and complete L-15B. Cell lines were maintained in 10 ml flat-sided culture tubes (Nunc) in 2.2 ml of complete medium and incubated at 28°C or 32°C. Medium was changed weekly and subcultures were carried out as required every 2-4 weeks.

Table 4-2 Embryo-derived tick cell lines used in the experiment

	Tick species	**Origin**	**Incubation temperature**	**Passage**	**Reference**
IDE8	*Ixodes scapularis*	USA	32°C - 34°C	84	(Munderloh *et al.* 1994)
IRE/CTVM20	*Ixodes ricinus*	UK	28°C	180	(Bell-Sakyi *et al.* 2007)
BME/CTVM23	*Rhipicephalus microplus*	Mozambique	32°C	73-77	(Alberdi, Nijhof, *et al.* 2012)

Table 4-3 Bacterial species used in the experiment

Species	**Strain**	**Host species**	**Origin**	**Isolated in**	**Reference**
Ehrlichia minasensis	UFMG-EV	*Rhipicephalus microplus*	Brazil	IDE8	(Cabezas-Cruz *et al.* 2016)
Rickettsia raoultii	Białystok	*Dermacentor reticulatus*	Poland	BME/CTVM23	(Palomar *et al.* 2019)
Spiroplasma sp.	DMAR11	*Dermacentor marginatus*	Spain	BME/CTVM23	(Palomar *et al.* 2019)

4.2.2. Tick cell morphometry

Pictures of tick cell cultures were obtained using an inverted Zeiss Axio Observer microscope with 40X magnification. Pictures were analysed using Fiji software (Schindelin *et al.* 2012). Two to three hundred cells for each cell line were measured and classified as one of three shapes – round, spindle-shaped or dendritic.

4.2.3. Evaluating the infection rate of pathogens within the tick cells

4.2.3.1. Preparation of Giemsa-stained cytocentrifuge smears

Approximately 50 µl of resuspended tick cell culture were put inside a cytospin well and centrifuged at 1000 rpm for 5 min (Shandon Cytospin 3). Slides were air-dried for 1 min, fixed in methanol for 3 min, stained for 40 min in 5% Giemsa stain with added Azur II (2 g/l of undiluted stain; Merck), and rinsed in two changes of distilled water buffered to pH 7.2 (Shute 1966). The smears were examined using a Leitz Orthoplan microscope at x1000 magnification with oil immersion; at least 100 cells per slide were examined for the presence of bacteria.

4.2.3.2. Evaluation of culture infection rates by qPCR

DNA extraction was performed from 200 µl of infected, resuspended tick cell cultures (BME/CTVM23 infected with *R. raoultii*, BME/CTVM23 infected with

E. minasensis, BME/CTVM23 infected with *Spiroplasma sp.*, and IDE8 infected with *E. minasensis*) using a DNeasy Blood & Tissue extraction kit (QIAGEN, UK) according to the manufacturer's protocol. All DNA samples were diluted to the concentration of 10 ng/µl. The qPCR assay was carried out with a LightCycler 480 (Roche, Basel, Switzerland). The reactions were performed in 20 µl volumes containing 1x SensiMix SYBR (Bioline), 0.5 µM of each primer, and 1 µl (10 ng) of DNA template. Thermocycling conditions were as follows: 94°C for 300 s, then 45 cycles of 94°C for 10 s, 60°C for 10 s and 72°C for 10 s. Samples and standards were analysed in triplicate. Quantification was performed using the synthetic standards in concentrations from 5×10^6 to 5×10^{-1} copies/µl in 100 ng/µl yeast tRNA (Invitrogen) (Table 4-4).

Table 4-4 Primers and standards used in the study

	Forward, 5'-3'	Reverse, 5'-3'	Amplicon length, bp	Standard, 5'-3'	Reference
Ixodes scapularis ribosomal protein L6 nuclear gene, *rpl6*	CCGGTCCAAGATGTTCCA	TGCGCTTCCTCTTCTCCTTG	77	CCGGTCCAAGATGTTCCACAAGCGGGGCCTGTTTAAGGTGAAGCACGCCCCTCCGACCAAGGAGAAGAGGAAGCGCA	(Al-Khafaji 2018)
Spiroplasma sp. 16S rDNA gene	AACCGTGCTTTAATGGGAGC	ACTCAACACCCGTACCAACA	88	AACCGTGCTTTAATGGGAGCTAATATGCAACGTCAAGCAATTCCATTATTAAAACCCCAAACACCAATTGTTGGTACGGGTGTTGAGT	This study
Rickettsia rickettsii citrate synthase gene, *gltA*	TCGCAAATGTTCACGGTACTTT	TCGTGCATTTCTTTCCATTGTG	74	TCGCAAATGTTCACGGTACTTTTTGCAATAGCAAGAACCGTAGGCTGGATGGCACAATGGAAAGAAATGCACGA	(Stenos, Graves, and Unsworth 2005)
Ehrlichia minasensis SSU rDNA gene	TTGCTATTAGATGAGCCTATATTAG	GTGTGGCTGATCATCCTCT	97	GCTATTAGATGAGCCTACGTTAGATTAGCTAGTTGGTAAGGTAATGGCTTACCAAGGCTATGATCTATAG CTGGTCTGAGAGGACGATCAGCCACAC	(Makepeace, Rodgers, and Trees 2006)

4.2.4. Detecting the presence of SCRV in cell lines

RNA was extracted from 0.5 µl of resuspended tick cell cultures with RNeasy Mini kit (QIAGEN), purified with a TURBO DNase-free kit (Invitrogen) and first-strand cDNA synthesis was carried out with a SuperScript III Reverse Transcriptase kit (ThermoFisher) according to the manufacturer's protocols. PCR amplification was performed with primers targeting a 358 bp region of segment 2 of the SCRV genome as described previously (Alberdi, Dalby, *et al.* 2012).

4.2.5. Generating reference genomes and annotations

4.2.5.1. Sequencing of Spiroplasma sp. genome

Two samples of 2.2 ml suspension of BME/CTVM23 cell heavily infected with *Spiroplasma sp.* (DMAR11) were each semi-purified as follows: the cell suspension was centrifuged at 1500 × g for 5 min to pellet the cells, the supernatant containing extracellular bacteria was saved, the pellet was mixed with 0.5 ml of trypsin solution (500 µg/ml in 1X PBS) and incubated at 37°C for 20 min. After incubation the mixture was passed 10 times through a bent 26G needle and the saved supernatant was added to it to deactivate the trypsin, the mixture was centrifuged at 1500 × g for 5 min to remove the cell debris, the supernatant was transferred to a clean tube and centrifuged at 21000 × g for 10 min to pellet the bacteria. The bacterial pellet was kept at -20°C until DNA extraction. DNA was extracted with a DNeasy Blood & Tissue extraction kit (QIAGEN, UK) according to the manufacturer's protocol. One microliter of DNA was measured with the Qubit dsDNA HS Assay Kit (QIAGEN). The DNA was then checked for the integrity by running overnight 10 µl of each DNA extraction on a 0.7% agarose gel in 1X TAE buffer, and submitted to the Centre for Genomic Research at the University of Liverpool for library preparation and sequencing with PacBio technology (Pacific Biosciences Inc.). Initial data quality control was performed by the bioinformatics team of the Centre for Genomic Research. The genome was assembled

with flye v.2.8.3 (Kolmogorov *et al.* 2019). The resulting assembly was annotated using Prokka v1.14.6 against the generic bacterial database (Seemann 2014).

4.2.5.2. Sequencing the E. minasensis genome

Two milliliters of IDE8 cell suspension infected with *E. minasensis* were processed as described in 4.2.5.1. Briefly, semipurification, DNA extraction, and quality assessment (except for gel DNA integrity check) were carried out as described in the section 4.2.5.1. The sequencing library for ONT sequencing was prepared according to the manufacturer's protocol for SQK-LSK108 ligation, and sequenced using MIN106 (R 9.4.1) flowcells for MinION (Oxford Nanopore Technologies). The raw signal was then basecalled using Guppy v5.0.11 with the "super-accuracy" basecalling model (dna_r9.4.1_450bps_sup.cfg).

Kraken v2.1.2 with 'nt' database was used to classify the reads (Wood, Lu, and Langmead 2019). Bacterial reads were selected using the custom-made script and seqtk toolkit (https://github.com/lh3/seqtk). Filtlong v0.2.0 was used to select longest/highest quality reads with options "--min_length 1000 --target_bases 280000000" (https://github.com/rrwick/Filtlong). These reads were assembled using Flye 2.9-b1768 (Kolmogorov *et al.* 2019). To polish the assembly, 500,000 faux error-free 500 bp single-end reads from the assembled genome, using randomreads.sh utility from BBmap v38.07, with options "len=500 reads=500000 adderrors=f" (Bushnell 2014). Errors were polished using Racon v1.4.20 for two iterative rounds, with "-w 100" option (Vaser *et al.* 2017). The resulting assembly was annotated using Prokka v1.14.6 against the generic bacterial database (Seemann 2014).

4.2.5.3. Amending the mitochondrial genome of R. microplus

The RefSeq mitochondrial assembly (NC_023335.1 *Rhipicephalus microplus* isolate BomiB mitochondrion) was polished using the combined R2 reads from the *R. microplus* single cell RNA sequencing (described in Chapter V, section 5.2.2) and Racon v1.4.20. The assembly was then rotated to start with the small rRNA subunit.

The final 14,905 bp long assembly was annotated with MITOS2 using the RefSeq 81 *Metazoa* reference, the genetic code 5 for *Invertebrata* and the other parameters set to default (Donath *et al.* 2019). Only rRNA and protein coding genes were retained for the final annotation. Additionally, one of the chromosomes of the *R. microplus* nuclear genome contained a 9,000 bp fragment exactly matching mitochondria. This chromosomal mitochondrial insert was hard masked.

4.2.5.4. Assembling the mitochondrial genome of I. scapularis

The mitochondrial sequence of *I. scapularis* was absent from the published RefSeq genome, and it was necessary to assemble it separately. This was done using a two-step process: 1) identification of short and long reads that originated from mitochondria; 2) hybrid assembly of the resulting selection. To this end, a crude mitogenome assembly was first generated using Novoplasty (Dierckxsens, Mardulyn, and Smits 2016) and the mitochondrion of *I. ricinus* as a seed (GenBank: JN248424.2 (Montagna *et al.* 2012)) with Illumina reads obtained from the *I. scapularis* ISE6 cell line (SRR5252948 (Miller *et al.* 2018)). All Illumina reads were mapped back to the resulting draft assembly using bowtie2 package (Langmead and Salzberg, 2012), and successfully aligned reads were extracted with seqtk package (github.com/lh3/seqtk). The same was done with PacBio reads obtained from the ISE6 cell line (SRX3593705 (Miller *et al.* 2018)) using the minimap2 (Li 2018) long read aligner. Finally, the resulting selection of mitochondrion-derived short and long reads were assembled with Unicycler v0.4.8, a hybrid assembler known to produce high-accuracy circularized bacterial genomes (Wick *et al.* 2017).

Raw reads from the scRNA-seq *I. scapularis* IDE8 samples (described in Chapter V, section 5.2.2) were then aligned to the assembled mitochondrion using bowtie2 and samtools (Li *et al.* 2009) packages. Successfully aligned reads were extracted using samtools and used for three rounds of polishing the mitochondrial sequence with Racon v1.4.12 (Vaser *et al.* 2017) and one round of polishing with Pilon (B. J. Walker *et al.* 2014). The resulting mitochondrial sequence was annotated with

the online tool MITOS2 using the RefSeq 81 *Metazoa* reference, the genetic code 5 for *Invertebrata* and the other parameters set to default (Donath *et al.* 2019). Only rRNA and protein coding genes were retained for the final annotation.

4.3. Results

4.3.1. Characterization of cell lines

Three cell lines (Table 4-2) were initially considered as acceptable for the experiment based on availability in the TCB, growth rate, and the range of bacteria that could be grown in each cell line. Cell lines derived from *R. microplus* (BME/CTVM23) and *I. scapularis* (IDE8) showed a high level of phenotypic heterogeneity while all cells from *I. ricinus* IRE/CTVM20 were uniformly round (Figure 34). Figure 36 shows size and shape distribution of cells. BME/CTVM23 and IDE8 cell lines demonstrated a very similar distribution with round cells with mean diameters of 30 and 27 µm respectively, while IRE/CTVM20 cells had a mean diameter of 24 µm. Spindle-shaped and dendritic cells were more diverse in size, although the range was quite similar between BME/CTVM23 and IDE8 cells.

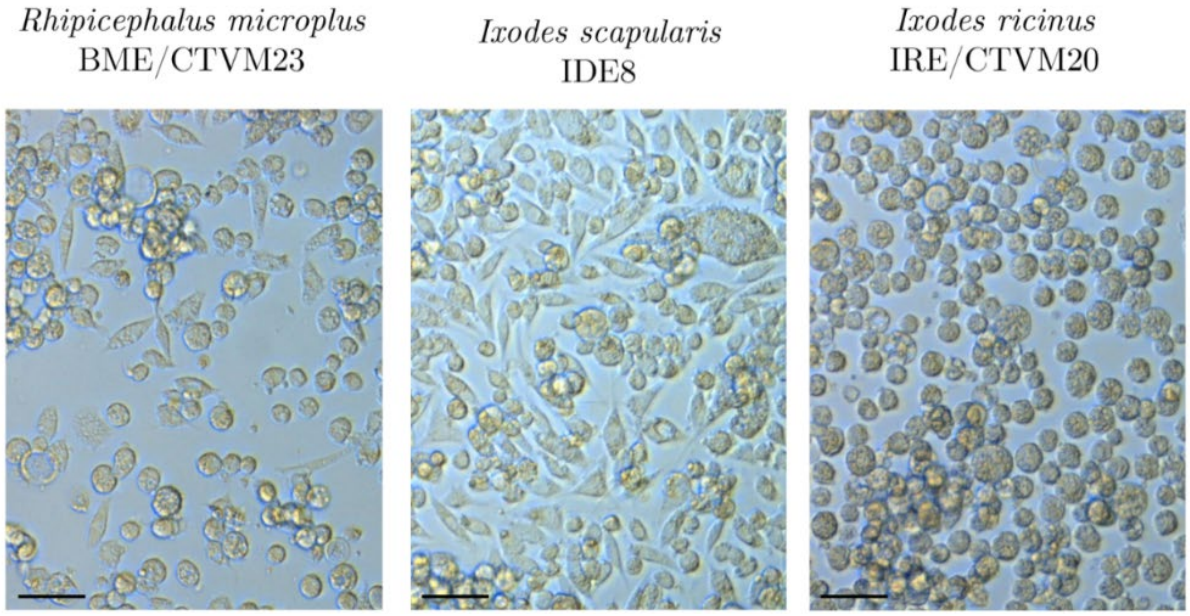

Figure 34 Cell phenotypes in three tick cell lines. The size range and shapes of cells can be observed. Live, inverted microscope. Scale bars = 50 µm.

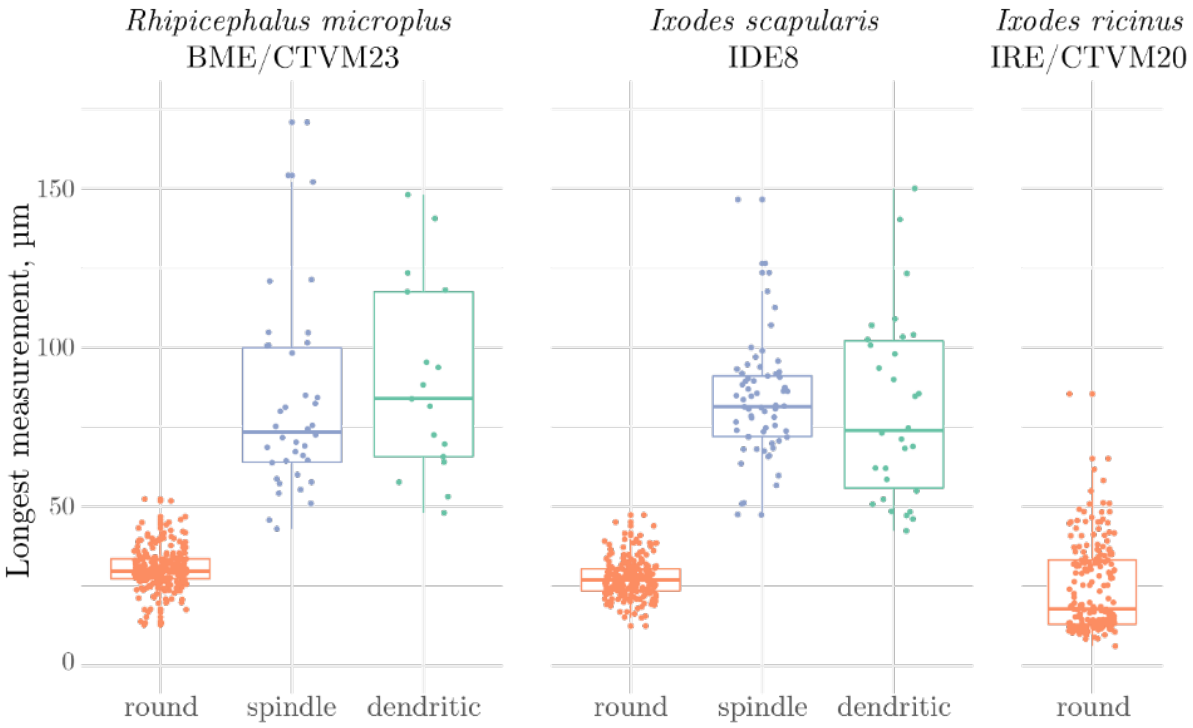

Figure 35 Cell size and shape distribution in three tick cell lines.

4.3.2. Evaluating the infection rate of pathogens within the tick cells

In order to find the optimal conditions for scRNA-seq experiments, it was necessary to build growth curves of each studied bacterium in each cell line. If the bacterial load was too low, RNA-seq would not be able to detect the bacterial genes that would be lost among the more abundant tick RNA. On the other hand, if the load was too high, cells would become fragile and hard to sort effectively. Thus, infection rate estimation was crucial for successful scRNA-seq.

Bacterial quantification using qPCR was performed on replicate cultures on days 9 and 15 after infection for *E. minasensis*, days 11 and 15 for *R. raoultii* and days 11 and 18 for *Spiroplasma sp.* Figure 36 shows the results of the quantification. *E. minasensis* in the *I. scapularis* IDE8 cell line showed a comparatively consistent growth rate (Figure 36, top left), whereas the same pathogen strain transferred to the *R. microplus* BME/CTVM23 cell line demonstrated bacterial counts differing by two

orders of magnitude on day 15 (Figure 36, top right). *R. raoultii* growth rates were quite similar on day 11, although by day 15 the rates diverged between replicates (Figure 36, bottom right). *Spiroplasma sp.* achieved higher counts than *E. minasensis* and *R. raoultii*, reaching several thousand bacteria per tick cell by day 11 and tens of thousands by day 18 (Figure 36, bottom left). *E. minasensis* and *R. raoultii* only reached several thousand bacteria per tick cell, before the host cells started to disintegrate as seen by microscopy, thereby flattening the growth curves (dates after the 15th day after infection not shown on the graphs).

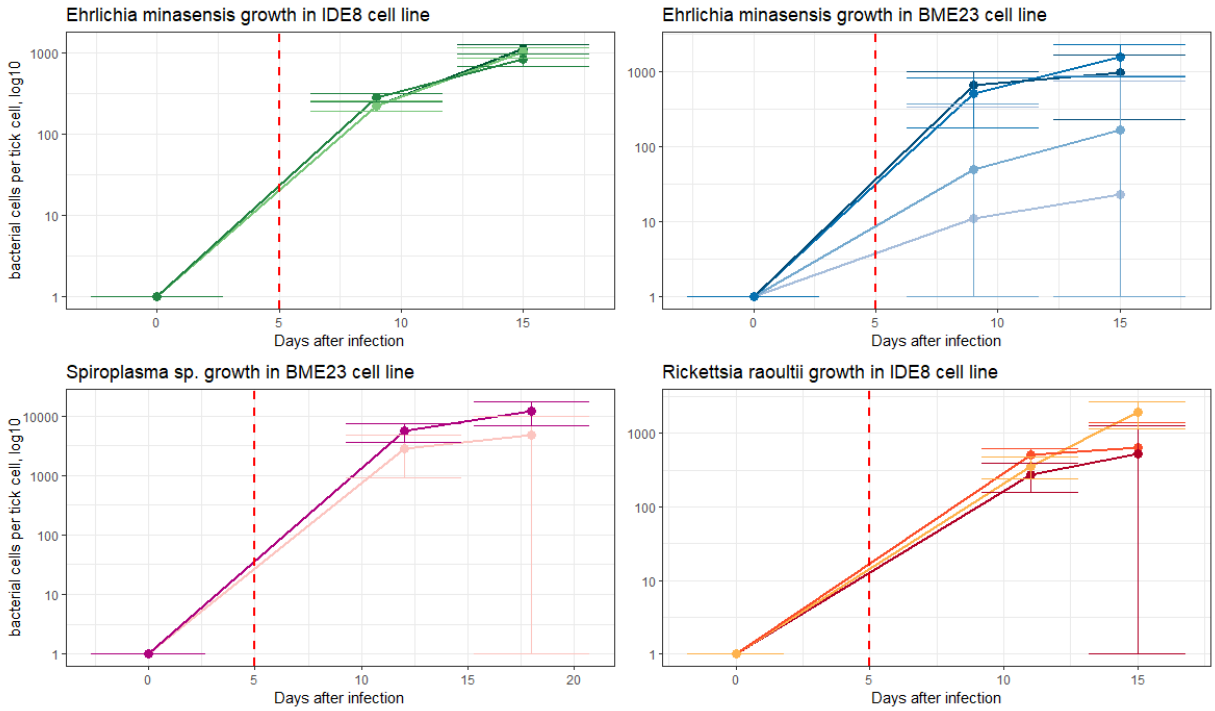

Figure 36 Growth curves for three bacterial pathogens in two tick cell lines according to qPCR quantification. Red line marks the fifth day after infection which was later chosen for the experiment. Y-axis is shown in log_{10} scale.

4.3.3. Studying the presence of SCRV in cell lines

In addition to bacterial growth profiling, it was necessary to evaluate the presence of the SCRV virus in the selected cell lines in case the presence of SCRV affected the response of the tick cell to bacterial infection. Five samples were tested for the presence of SCRV: intact IDE8 and BME/CTVM23 cell lines, IDE8 infected with *E. minasensis*, BME/CTVM23 infected with *E. minasensis* derived from IDE8 cells, and BME/CTVM23 with added medium from IDE8 culture. Table 4-5 shows the yield of cDNA from five samples of uninfected and *E. minasensis*-infected tick cells, Figure 37 shows the results of the cDNA amplification with SCRV specific primers from the same samples. All samples except for uninfected BME/CTVM23 cell culture were positive for SCRV, which implies that the virus, which persistently infects the IDE8 cell line, was transferred to the BME/CTVM23 cells with both *E. minasensis* and the medium from the IDE8 culture.

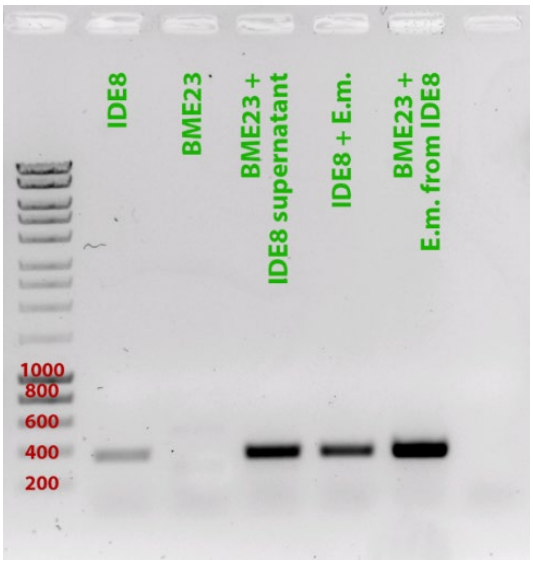

Figure 37 Results of amplification for SCRV detection.

"IDE8" – uninfected cell line; "BME23" – uninfected BME/CTVM23 cell line; "BME23 + IDE8 supernatant" – BME/CTVM23 cell line with added medium from IDE8 culture; "IDE8 + E.m." – IDE8 cell line infected with E. minasensis; "BME23 + E.m. from IDE8" – BME/CTVM23 cell line infected with E. minasensis derived from IDE8 cell culture.

Table 4-5 cDNA yields. "IDE8" – uninfected cell line; "BME23" – uninfected BME/CTVM23 cell line; "BME23 + IDE8 supernatant" – BME/CTVM23 cell line with added medium from IDE8 culture; "IDE8 + E.m." – IDE8 cell line infected with E. minasensis; "BME23 + E.m. from IDE8" – BME/CTVM23 cell line infected with E. minasensis derived from IDE8 cell culture.

Sample	*Concentration, ng/µl*	*Volume, µl*	*Total cDNA, ng*
BME/CTVM23	112	5	560
IDE8	6	10	60
BME23 + IDE8 supernatant	140	5	700
IDE8 + E.m.	98	5	490
BME23 + E.m. from IDE8	112	5	560

4.3.4. Improving reference genomes and annotations

4.3.4.1. Sequencing and assembly of the Spiroplasma sp. genome

The two cultures yielded, respectively, 2 µg of DNA with concentration 20 ng/µl (sample Sp1) and 1.2 µg with concentration 12 ng/µl (sample Sp2). The electrophoresis showed that both samples contained long non-degraded DNA approximately 50Kb long (Figure 38).

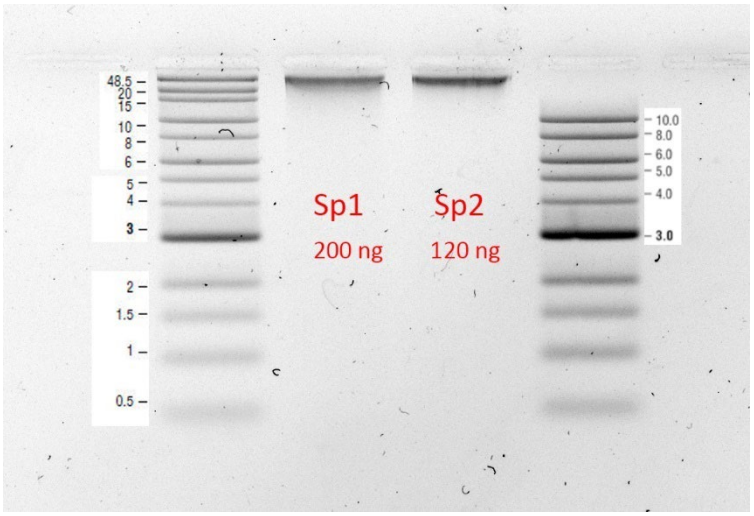

Figure 38 Overnight gel with two DNA extractions from Spiroplasma sp. (Sp1 and Sp2). The total DNA bands are at the same level as the top ladder band of 48.5 Kb.

Sequencing generated 779,545 reads that spanned 12.7 Gb of nucleotide sequence. Approximately 75% of reads were classified as eukaryotic and 15% were classified as bacterial (120,316 reads). Using the latter set, 11,677 longest/highest quality reads of approximately 200 Mb overall length were selected and assembled into a single 1,740,586 bp circular chromosome. The assembly was annotated in Prokka resulting in 2212 CDSs, 27 tRNAs, 6 rRNAs (two ribosomal operons) and 1 tmRNA.

4.3.4.2. Sequencing and assembly of the E. minasensis genome

DNA extraction yielded 1 µg of DNA with concentration 4.7 ng/µl, and 230 ng were used for the ONT library preparation. Sequencing generated 1,414,988 reads that spanned 3.1 Gb. Approximately 50% of reads were classified as eukaryotic and another 15% were classified as bacterial (219,742 reads). Using the latter set, 11,463 longest/highest quality reads of approximately 280 Mb overall length were selected and assembled into a single 1,346,938 bp circular chromosome. The assembly was annotated in Prokka resulting in 931 CDSs, 36 tRNAs, 2 rRNAs and 1 tmRNA.

4.3.4.3. Assembling the mitochondrial genome of I. scapularis

None of the available genome assemblies of *I. scapularis* included a mitochondrial sequence, which was essential for further analysis of the scRNA-seq data. Authors of the Wikel strain assembly (GCA_000208615.1) claimed to have mitochondrion assembled and removed from the main genomic sequence (Gulia-Nuss *et al.* 2016), although it was not available in any public databases. Authors of the second published genome of the ISE6 cell line (GCF_002892825.2) have not analysed the mitochondrial sequence, claiming that mitochondrial contigs were submitted within the nuclear genome (Miller *et al.* 2018); however, our analysis did not confirm the presence of a mitochondrial genome with the published assembly.

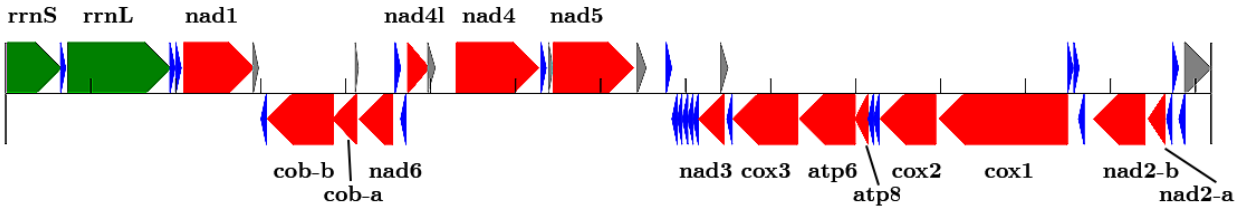

Figure 39 Schematic representation of mitochondrial genome of Ixodes scapularis and its genes. Green arrows show rrnS and rrnL genes, red – protein coding genes, blue – tRNA genes, grey – OH (origin of heavy strand replication).

To mitigate this issue, a *de novo* assembly of the *I. scapularis* mitochondrial genome was performed (see section 4.2.5.4 for exact protocol). A hybrid assembly from pre-selected mitochondrion-derived raw Illumina and PacBio reads resulted in a single circular 14,546 bp contig. The annotation using Mitos2 web server identified all typical mitochondrial genes, including 22 tRNA genes; the only missing feature was OL (origin of light strand replication) (Figure 39). Finally, the sequence was rotated to start with the rrnS gene.

4.4. Discussion

4.4.1. Choosing cell lines and pathogens for the scRNA-seq experiment

An important factor for choosing a cell line was the phenotypic heterogeneity of cells and diversity of bacteria maintained in the culture. These observations led to a decision that BME/CTVM23 and IDE8 cell lines were more suitable than IRE/CTVM20 for the planned scRNA-seq experiment.

The *R. microplus* BME/CTVM23 cell line was chosen to be the main focus of the scRNAseq experiment described in Chapter V, as it can support the most diverse set of symbionts. *R. raoultii*, *E. minasensis* and *Spiroplasma sp.* were chosen to infect the BME/CTVM23 cell line (Figure 40, 1-4). One complication arose from the fact that *E. minasensis* was transferred from the IDE8 cell line, which bears the endogenous tick virus SCRV, and, therefore, the recipient cell line also received the virus. Thus, the comparison of the single cell transcriptomes of the uninfected BME/CTVM23 cell line with those of the cells infected with *E. minasensis* (and SCRV) was not directly possible, as there would be two parameters altered. As the presence of the virus was confirmed by PCR (see section 4.3.3), it became obvious that viral particles were transferred with the medium from the source culture, and it was impossible to get rid of them when transferring bacteria. In order to distinguish between the tick immune responses against *E. minasensis* and against the virus, another sample was included in the experiment – BME/CTVM23 cells infected with SCRV alone (Figure 40(5)).

Considering different growth patterns of *E. minasensis* in two different cell lines, two *I. scapularis* cell samples were also included in the experiment – IDE8 cells bearing the SCRV (IDE8 only exists in SCRV-infected form), and IDE8 cells infected with *E. minasensis* (and SCRV). Figure 40 summarises the seven samples included in the scRNA-seq experiment described in Chapter V.

4.4.2. Choosing the appropriate infection time point

Two methods of assessing bacterial load in tick cells were used – visual observation of Giemsa-stained cytocentrifuge smears of infected cultures and qPCR measurement of the tick/bacterial DNA ratio. Both approaches had their limitations. Visual counting of infected cells was only done on a small subset of cells (100-200 cells) and even the highest magnification of the light microscopy did not allow counting of separate bacteria in the aggregates, thus, the estimation of bacterial number was approximate and dealt with categories such as "not infected", "infected", and "over-infected" (cells which had ruptured because of the excessive bacterial load). In case of *Spiroplasma sp.*, in which the cells are long and thin, it was impossible to recognize single bacteria and the infection became obvious only when bacteria formed colonies. The qPCR quantification on the other hand was not informative for the distribution of bacteria in tick cells, only estimating the mean of the distribution, whereas the visual control clearly identified infection heterogeneity between cells. Another complication in estimation of the growth rates of symbiotic bacteria was that the bacteria/host system did not always demonstrate the same rate of development of the infection and individual replicates differed substantially, as shown in Figure 36. *E. minasensis* demonstrated different patterns of development in the two cell lines, with relatively consistent growth curves in the *I. scapularis* IDE8 cell line (Figure 36, top left), and widely variable rates in the *R. microplus* BME/CTVM23 line (Figure 36, top right). *R. raoultii* and *Spiroplasma sp.* were only used in the *R. microplus* cell line and their infection rates were comparatively reproducible during the first ten days (Figure 36, bottom left and right).

For the purposes of planning scRNA-seq experiments, infection conditions were chosen according to several crucial criteria:

- Tick cell cultures must not contain over-infected cells which easily fall apart contaminating the medium with free-floating bacteria and cell debris. This is very important for single-cell experiments, where each cell should be enclosed

in a separate droplet, and thus medium contamination might hugely affect the result.

- Tick cells from several stages of infection should be present. The stages are, roughly: 1) uninfected cells, 2) cells which recently captured a single bacterium, 3) cells which have small colonies of dividing bacteria, and 4) cells which have a substantial number of bacteria in the cytoplasm.
- Tick cell lines should be infected with the different bacteria and then analysed on the same number of days after subculturing to reduce the sources of variation.
- Tick cell lines should be passaged, infected and taken into the experiment at the same time to minimise the differences which might arise from the different stages of tick cell lines.

Combining the data from qPCR evaluation and visual, day five after infection was chosen as the most suitable for the experiment. According to qPCR data, the mean infection rate at this timepoint should be approximately 100 bacterial cells per tick cell for all pathogens. According to visual counting of infected and uninfected cells at this timepoint, approximately 20% of *R. microplus* cells appeared to be infected with *R. raoultii*, approximately 30% cells with *Spiroplasma sp.*, and between 20 and 50% of cells infected with *E. minasensis*. The rate of *E. minasensis* in *I. scapularis* IDE8 cells was approximately 20%. These figures enabled the assumption that all stages of infection – from single bacteria to large colonies – would be present in cell lines at the moment of the experiment, and medium contamination from over-infected cells would be minimal. Visual checking of the infected cultures revealed very low levels or absence of cells that were disintegrating due to a very heavy infection.

4.4.3. Design of the transcriptomic experiment on tick cell lines infected with different symbionts

Figure 40 summarises the samples chosen for the experiment. Tick cell lines BME/CTVM23 (*R. microplus*) and IDE8 (*I. scapularis*) cell lines over IRE/CTVM20

(*I. ricinus*) due to their phenotypic heterogeneity and diversity of bacteria maintained in these lines (samples 1 and 3 in Figure 40). *I. scapularis* cell samples were infected with *E. minasensis* to investigate the immune response to invasion of a gram-negative bacterium (sample 2 in Figure 40). The *R. microplus* BME/CTVM23 cell line could support the more diverse set of symbionts: *R. raoultii*, *E. minasensis*, and *Spiroplasma* sp. (samples 4, 6 and 7 in Figure 40). To differentiate the immune response to *E. minasensis* and SCRV as they cannot be separated in IDE8 cell line, BME/CTVM23 cells infected with SCRV alone was included in the experiment (sample 5 in Figure 40).

Appropriate infection time points for the experiment were evaluated in the scope of the chapter. The fifth day after infection was chosen as the most suitable since it exhibited a similar mean infection rate for all bacterial pathogens and stages of infection present in the cell line.

The results of the experiment are presented in Chapter V.

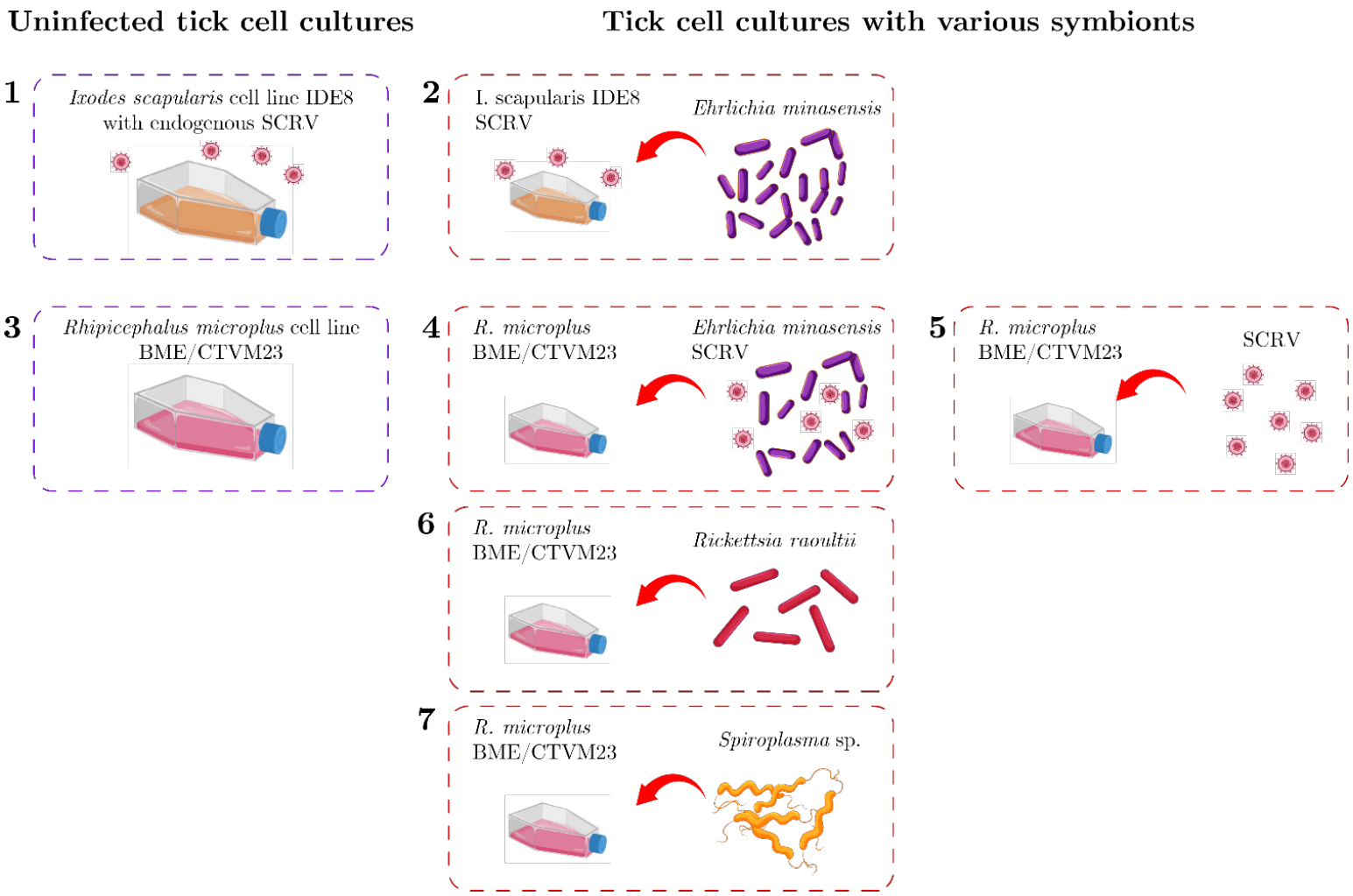

Figure 40 The schematic overview of seven samples involved into the experiment. Elements of the picture have been downloaded from Freepik.com (vector image created by brgfx).

5. Chapter V Study of the transcriptomic response of tick cells to bacterial pathogens

5.1. Introduction

Ticks are obligate blood-feeders highly adapted to the blood-feeding mode of living; all active stages require blood as a source of nutrition and, in the case of adults, for sperm or egg production. Thus ticks have a great capacity to be vectors of pathogens like other blood-feeding arthropods. Ticks transmit a wide variety of pathogens to vertebrates, including viruses, bacteria, protozoa and helminths, which act as infectious agents in vertebrate hosts and cause a significant burden on public and livestock health. The costs associated with mortality, relapse, treatments, and decreased production yields are economically significant (Jose de la Fuente 2008) and result from the heavy burden of ticks and pathogenic diseases spread by ticks (Rodriguez-Vivas, Jonsson, and Bhushan 2018).

Although ticks are vectors of many human and livestock diseases, their interactions with parasites are under-investigated and knowledge of how ticks react to the invasion and their immune response is still vague (Brites-Neto, Duarte, and Martins 2015). The research on arthropod-borne diseases has mainly focused on the vertebrate part of the life cycle as this part greatly affects human economics. Secondly, research on vertebrates was easier for *in vitro* studies due to popularity and availability of cell lines. In contrast, the cultivation of arthropods or establishing arthropod cell lines has been associated with considerable difficulties (Bell-Sakyi *et al.* 2007; Al-Rofaai and Bell-Sakyi 2020).

In this work, cell lines of two species of hard ticks were used, *Ixodes scapularis* and *Rhipicephalus microplus*. The main characteristics of these two species were provided in the previous chapter.

5.1.1. Current measures to control tick-borne diseases (TBD)

The spread of tick-borne pathogens depends on three factors: the density of the tick population, the infection rate of ticks, and the frequency of contacts of humans or livestock with wild animal reservoirs (Rahlenbeck, Fingerle, and Doggett 2016). Thus, TBD dissemination can be controlled by manipulating three factors: reducing the size of the tick population, trying to eliminate pathogens from the livestock population or applying prevention techniques on the relevant territories with a high chance of transferring pathogens to humans or livestock (Rodriguez-Vivas, Jonsson, and Bhushan 2018).

Nowadays, control of ticks is primarily achieved by applying chemical or plant-based acaricides and insecticides (Ostfeld *et al.* 2006; Rodriguez-Vivas, Jonsson, and Bhushan 2018). However, many tick species develop resistance to these chemicals in 4-12 years (Knolhoff and Onstad 2014), making it necessary to find new compounds or find an alternative approach to control ticks and/or their pathogens. (Al-Rofaai and Bell-Sakyi 2020; Rodriguez-Vivas, Jonsson, and Bhushan 2018) Furthermore, a concern about the broad application of acaricides and insecticides is the environmental harm that they cause, e.g., honeybee population decrease (Ansari, Moraiet, and Ahmad 2014; Kevan and Viana 2003). Unfortunately, acaricides, which are the first choice of defence against ticks, are environmentally harmful, and more and more tick populations are developing resistance to them (Gasparotto *et al.* 2020; Rodriguez-Vivas, Jonsson, and Bhushan 2018; Ceraul, Sonenshine, and Hynes 2002).

Livestock can be vaccinated against ticks. Anti-tick immunity was described almost a century ago in 1939 by William Trager, who showed that the ability of ticks to engorge sequentially on the same guinea pig is notably reduced, suggesting that animals are developing some kind of anti-tick immunity following infestations (Trager 1939). The first tick vaccine was designed in 1989 using the protective antigen Bm86 isolated from the *R. microplus* gut; it induced a protective response in vaccinated hosts, which damaged the tick gut affecting tick and egg viability (Lew-Tabor and

Rodriguez Valle 2016; Willadsen *et al.* 1989). TickGARD and Gavac vaccines based on Bm86 antigen were released to Australian and Latin American markets in 1994 (Jose de la Fuente, Rodriguez, and Garcia-Garcia 2000). However, these vaccines were not entirely protective against all *R. microplus* strains and sometimes were misused, making anti-tick vaccination not very popular, and TickGARD has been discontinued since 2010 (Lew-Tabor and Rodriguez Valle 2016). The progress of 'omics technologies has boosted tick vaccine development, and new genomics and transcriptomics data provide researchers with many targets for new, more effective vaccines (José de la Fuente 2018); however, none of the currently-used vaccines is sufficient to eliminate ticks, they can only reduce the number of ticks and require regular boosts, which is not always logistically and economically feasible (Garcia *et al.* 2020). Another way of lowering tick abundance is by using biological agents (e.g., parasitoid wasps, nematodes, bacteria or fungi), but these approaches are still under development (Ostfeld *et al.* 2006).

5.1.2. Arthropod innate immune system

The tick immune system plays a major role in vector competence, but our knowledge of tick immunity is limited. Most of the studies of arthropod immunity focus on insects. Considering that ticks are evolutionarily distant from insects and developed a blood-feeding lifestyle independently of insects (Giribet and Edgecombe 2019), there is a certain need for research focused on tick immunity. Little is known about how transmitted pathogens can escape the immune responses of ticks, and the vector-pathogen interface is the least understood site of the system of the tick-pathogen-vertebrate host.

5.1.2.1. Cellular immune mechanisms

There are two types of antigen deactivation mechanisms in ticks observed at the cellular level – encapsulation and nodulation. Both of them involve haemocyte cells forming a multilayer capsule around an invader. The encapsulation was experimentally

shown with artificial pieces of Epon-Araldite placed under the cuticle of the tick *Dermacentor variabilis*, followed by forming a capsule made of haemocytes and fibrous matrix (Eggenberger, Lamoreaux, and Coons 1990). Direct injection of *E. coli* cells into haemolymph resulted in a rapid aggregation of bacteria in clumps surrounded by haemocytes (Ceraul, Sonenshine, and Hynes 2002). The human pathogen *Borrelia burgdorferi*, which causes Lyme disease, was shown to be phagocytosed by tick cells in a "coiling" manner (Mattila, Munderloh, and Kurtti 2007).

There are several types of haemocytes: plasmatocytes and granulocytes I, which show phagocytic activity, and granulocytes II without phagocytic activity (Kuhn and Haug 1994)(Kuhn and Haug 1994). In some tick species, authors report spherulocytes (Borovickova and Hypsa 2005) All haemocytes are believed to differentiate from stem cells called prohaemocytes, which are only rarely detected in the haemolymph (Borovickova and Hypsa 2005; Söderhäll 2010). The capability of plasmatocytes to phagocytise foreign material was shown in several species of ticks and with different material (bacteria, yeast, artificial beads) and the use of reactive oxygen species and H_2O_2 for antigen deactivation was demonstrated (Söderhäll 2010; Pereira *et al.* 2001).

No tick cell lines derived exclusively from haemocytes are available at the moment, but some (proven for IDE12 & AAE2 and DAE15) show the phagocytic reaction to a pathogen (Mattila, Munderloh, and Kurtti 2007; Kurtti and Keyhani 2008). Many tick cell lines might contain haemocytes and demonstrate a phagocytic behaviour, including AVL/CTVM13 and AVL/CTVM17, BME/CTVM23 (Esteves *et al.* 2008).

5.1.2.2. Molecular mechanisms of defense described in ticks

Knowledge about molecular mechanisms in tick immunity is still patchy. Some work has been done on the salivary proteins in ticks, which is a central part of the tick-host interactions and effect of the tick on the host immune system, limited information is available on how tick proteins shape vector immunity and influence

pathogen persistence (A. A. Smith and Pal 2014). These authors looked for immune gene homologues in the *I. scapularis* genome and summarised molecules shown to play a role in the immune defence of ticks. They divided them into several groups: gut-microbe homeostasis (17 genes), agglutination (37), leucine-rich repeat (LRR) proteins (21), proteases (33), coagulation (11), non-self recognition and signal transduction via Toll, IMD, and JAK-STAT pathways (55), free radical defence (13), phagocytosis (33), and antimicrobial peptides (14) (Smith and Pal 2014).

Gut Microbe Homeostasis

The gut microbiome greatly affects the host immune system in all animals including ticks (Hooper, Littman, and Macpherson 2012). Mechanisms which distinguish between intrinsic microbiota and pathogenic microorganisms are not studied in ticks, but are well characterised in *Drosophila melanogaster* (Buchon, Broderick, and Lemaitre 2013). The *Drosophila* immune system has two routes to recognise pathogenic agents: the first is initiated by recognition of non-self molecules, the second relies on signals released by the fly's own damaged cells (Lazzaro and Rolff 2011). Some studies have shown that dual oxidase and peroxidases participate in maintaining gut homeostasis and mucosal immunity (Bae, Choi, and Lee 2010). In mosquitoes, dual oxidase, along with a specific heme-peroxidase, catalyses the formation of an acellular molecular barrier, the dityrosine network, which builds up in the luminal space along the gut epithelium during feeding (Kumar *et al.* 2010). The dityrosine network prevents extraneous microorganisms from entering the gut cells.

Signal Transduction Pathways

Three major pathways, namely Toll, immune deficiency (IMD), and Janus kinase (JAK) signalling transducer and activator of transcription (STAT) pathways, contribute to the activation of the immune response within arthropods (Smith and Pal 2014). Toll pathways are activated by bacterial, viral, or fungal invasion, the IMD pathway is induced by Gram-negative bacteria, and JAK-STAT has been shown to act in the presence of bacterial and protistan pathogens (L. Liu *et al.* 2012; Gupta *et*

al. 2009). The Toll pathway is stimulated by cell wall components of Gram-positive bacteria and, via several transcription factors, boosts the production of antimicrobial peptides (Smith and Pal 2014).

Complement-Related Molecules

Thioester-containing proteins (TEPs) form a conserved and phylogenetically distinct family of secreted proteins that play central roles in the innate immune response. There are two families of TEPs – complement factors and α_2-macroglobulins, which have been known in vertebrates for a long time, and have been recently found in arthropods (M. Williams and Baxter 2014). Complement factors are monomeric and located on surfaces when activated, while α_2-macroglobulins are usually multimeric protease inhibitors that encapsulate targets once cleaved in a protease-sensitive "bait region". Complement deposition on pathogen surfaces is responsible for enhanced phagocytosis, attraction of phagocytes, and direct lysis, while α_2-macroglobulins inactivate protease virulence factors (Williams and Baxter 2014). TAM and IrAM are two α_2-macroglobulins described for the soft tick *O. moubata* and the hard tick *I. ricinus*, respectively (Kopacek *et al.* 2010; Buresova *et al.* 2009); nine gene-candidates belonging to α_2-macroglobulins were found in the *I. scapularis* genome (IsAm) (Kopacek *et al.* 2000). A functional study based on RNAi-silencing linked with a phagocytic assay revealed that IrAM is involved in phagocytosis of the soft tick pathogen *Chryseobacterium indologenes* by *I. ricinus* haemocytes. The phagocytosis activity was dependent on an active metalloprotease secreted by the bacteria, indicating that interaction of tick α_2-macroglobulins with a protease from an invading pathogen is linked with a cellular immune response. Interestingly, this mechanism is not responsible for eliminating every pathogen, so the phagocytosis of *Borrelia burgdorferi* or a commensal bacteria *Staphylococcus xylosus* was not affected by IrAM silencing (Buresova *et al.* 2009).

Homologues of the Factor D clip-domain serine proteases were found in haemocytes of *Dermacentor variabilis* and *Dermacentor andersoni* and in the genome

of *I. scapularis* (Simser *et al.* 2004; Kopacek *et al.* 2010). They are closely related to Limulus factor D, an enzyme from the horseshoe crab *Tachypleus tridentatus*, which has an antimicrobial activity against Gram-negative bacteria (Kawabata *et al.* 1996).

Agglutination

Tick fibrinogen-related proteins (FRePs) are immune molecules (or lectins) which are most likely recognising self/non-self-carbohydrates within the tick haemolymph, distinguishing pathogen-associated molecular patterns on the surface of invading pathogens. They are often produced in the gut, haemocytes, or fat body (Grubhoffer, Kovář, and Rudenko 2004; A. A. Smith and Pal 2014). Soft ticks possess Dorin M, which is mainly expressed in haemocytes and salivary glands (R. Rego *et al.* 2006). The closely-related tachylectins from the horseshoe crab share this feature, and function as pattern recognition molecules (Kawabata and Tsuda 2002).

A family of fibrinogen-related proteins with 27 genes was identified in *I. ricinus* (Honig Mondekova *et al.* 2017). The authors separated these 27 genes into three groups based on the phylogenetic divergence – Ixoderin A, Ixoderin B and Ixoderin C. Ixoderin A is close to Dorin M and tachylectins, while the two other ixoderins cluster with less well-characterised genes from other arthropods. Three forms of ixoderins showed different patterns of expression: ixoderin A was mostly expressed in haemocytes and Malpighian tubules and was notably upregulated after a blood meal. Ixoderin B was active in salivary glands, and ixoderin C was expressed in all studied tissues with higher transcription in the gut and trachea (Honig Mondekova *et al.* 2017; R. O. M. Rego *et al.* 2005). Ixoderins appear to be important as opsonins involved in the phagocytosis of different bacteria and yeasts by tick haemocytes.

A galectin-like protein, OmGalec, was isolated from *O. moubata*. It is expressed at all stages of the tick cell cycle and was abundant in haemocytes, midgut, and reproductive organs (X. Huang *et al.* 2007).

Antimicrobial peptides

Antimicrobial peptides are known in arthropods; at least eight different classes have been described for *Drosophila,* mainly produced in the fat body in response to infection and secreted into the haemolymph (Imler and Bulet 2005). A diverse repertoire of antimicrobial peptides and lysozyme-like enzymes has been described for ticks as well. Antimicrobial production is triggered by the activation of the Toll, IMD, or JNK-STAT pathway in the presence of bacteria, fungi, or viruses (Smith and Pal 2014).

The tick defensin family is the most described; defensin orthologues have been found in at least twenty hard and soft tick species (Chrudimska *et al.* 2010). In addition to the haemolymph, tick defensins were also reported to be expressed in the midgut and other tissues, including salivary glands and fat body. C-type lysozymes were characterised in 1990s and it was shown that they are upregulated in the presence of *E. coli* (Podboronov 1991; Y. Saito *et al.* 2009). Most of the antimicrobial peptides have a particular specificity either against Gram-positive bacteria, or Gram-negative bacteria, or fungi. The precise scope of each peptide is not yet defined and is limited by the list of pathogens which were tested in experiments. For example, ixodidin showed in experiments the ability to inhibit *Micrococcus luteus* and decrease the activity of *E. coli* and had some inhibitory activity against chymotrypsin and elastase serine proteases. The molecular mechanisms of these actions remain unclear (Fogaca *et al.* 2006).

Several antimicrobial peptides were discovered by transcriptome sequencing of infected ticks. DvKPI (*Dermacentor variabilis* Kunitz protease inhibitor) was upregulated six-fold in ticks infected with *Rickettsia montanensis*, thus preventing the pathogen from massive colonisation, which would be harmful to the host (Ceraul *et al.* 2008). Similarly, *I. ricinus* serine protease inhibitor, IrSPI, was discovered by comparing transcriptomes of uninfected salivary glands with *Bartonella henselae*-infected glands; IrSPI had the highest expression level in the infected sample (X. Y.

Liu *et al.* 2014). Persulcatusin is another antimicrobial peptide discovered in the midgut of *Ixodes persulcatus* and is drawing increased attention because of its ability to inhibit multi-drug resistant strains of *Staphylococcus aureus* (Miyoshi *et al.* 2017).

5.1.3. Limitations of microarray and bulk transcriptome experiments

Tick cell lines are established from tick embryos, without selecting particular tissue types, so it is believed that cell lines consist of two or more cell types which are essential for the cell line survival as attempts to clone cells from cell lines failed (Lesley Bell-Sakyi *et al.* 2007). Researchers noticed the variability of shapes and sizes of cells within cell lines, but there is no data regarding which tissues these cell types came from, how many cell types are observed in each cell line and how different are the various cell lines, which is essential if we want to compare experiments performed with different cell lines. Alberdi *et al.* attempted to match two cell lines from two *Ixodes* species with tick tissues, but their results do not seem convincing (P. Alberdi *et al.* 2016). The single-cell approach is promising for describing the composition of tick cell lines and studying the immune response of each of the assumed cell types to different pathogens.

5.1.4. Aims and objectives

The main aim of the chapter is to acquire a better understanding about the heterogeneity of tick cell lines and their immune response in the presence of bacterial symbionts. In order to achieve this goal, several objectives have been outlined. Firstly, the composition of cell lines derived from two tick species will be analysed using single-cell transcriptomics approach. Secondary, transcriptomic responses of tick cells to bacterial symbionts will be studied.

5.2. Materials & methods

5.2.1. Infecting tick cells with bacterial symbionts

5.2.1.1. Infection of BME/CTVM23 culture

Infections of *E. minasensis*, *R. raoultii*, and *Spiroplasma sp.*, maintained as described in Chapter IV, were passaged into fresh uninfected BME/CTVM23 cells every two weeks. 0.5 ml of source culture was used for *R. raoultii* and *E. minasensis* infections, and 0.1 ml for *Spiroplasma sp.* infection. Tubes containing uninfected BME/CTVM23 cells were pooled and divided equally between five new tubes before infection to minimise the possible differences in conditions in tubes.

An uninfected IDE8 cell culture was used as a source of the endogenous virus SCRV; 2 ml of cell suspension were centrifuged at 10000 rcf to separate tick cells and cell debris from SCRV particles, then 0.5 ml of supernatant were transferred to a tube of uninfected BME/CTVM23 cell culture.

5.2.1.2. Infection of IDE8 culture

Five ml of IDE8 cell culture infected with *E. minasensis* at a rate >50% were harvested by pipetting, the cell suspension was centrifuged at room temperature for 5 min at 200 rcf, the supernatant was transferred to a new tube, the cell pellet was resuspended in 500 µl of trypsin (500 µg/ml in PBS) and incubated for 20 min at 37 °C. The trypsin was deactivated by adding the supernatant back to the tube, and the cell suspension was passed ten times through a bent 26G needle to rupture the cells mechanically and release the intracellular bacteria. The suspension was centrifuged at room temperature for 5 min at 1500 x *g*. The supernatant containing cell-free bacteria was collected, and 200 µl was added to uninfected cell cultures growing in flat-sided tubes. Tubes containing uninfected IDE8 cells were mixed and separated into two new tubes before infection to minimise the possible differences in conditions in different tubes.

5.2.2. 10X sorting and sequencing

BME/CTVM23 and IDE8 cell cultures were used on the fifth day after infection. Seven cultures, as described in Table 5-1, were included in the experiment.

Table 5-1 Tick cell cultures and microorganisms used in the scRNAseq experiment

Sample ID	Tick species and cell line ID	Symbiont	Intended number of cells
Is_uninf	*Ixodes scapularis* IDE8	endogenous SCRV	8000
Is_Em	*I. scapularis* IDE8	*Ehrlichia minasensis,* SCRV	10000
Rm_uninf	*Rhipicephalus microplus* BME/CTVM23	none	8000
Rm_Em	*R. microplus* BME/CTVM23	*E. minasensis*, SCRV	10000
Rm_Rr	*R. microplus* BME/CTVM23	*Rickettsia raoultii*	10000
Rm_Sp	*R. microplus* BME/CTVM23	*Spiroplasma sp.*	10000
Rm_SCRV	*R. microplus* BME/CTVM23	SCRV	8000

0.5 ml of each cell culture was centrifuged at 300 rcf for 5 min, the supernatant was removed, and 500 µl of 1X PBS with 0.04% BSA was added, and the cell pellet was pipetted with a wide-bore tip. The procedure was repeated. Then each cell culture was pushed through a strainer with a pore size of 70 µm. If the culture was clumping the strainer, the culture was pushed through several strainers (up to three) until the culture went through easily.

The following steps of cell sorting, library preparation and sequencing were performed by Dr Margaret Hudges from the Centre for Genomic Research at the University of Liverpool. The density and viability of cells were calculated using a Luna Automated Cell Counter (Logos Biosystems Inc., USA). Chips were loaded after calculating the accurate volumes using the 'Cell Suspension Volume Calculator Table'. Once GEMs (Gel Bead-In Emulsions) were obtained, reverse transcription and cDNA amplification steps were performed. scRNA-seq libraries were prepared using the

Chromium Next GEM Single Cell 3′ v3.1 (10X Genomics), according to the manufacturer's protocol. Sequencing was done on Illumina NovaSeq, generating paired-end reads.

5.2.3. 10X data processing

The reference for *I. scapularis* samples consisted of *I. scapularis* assembly (nuclear and mitochondrial), *E. minasensis* and SCRV assemblies. The reference for *R. microplus* samples consisted of *R. microplus* assembly (nuclear and mitochondrial), *E. minasensis*, *R. raoultii*, *Spiroplasma sp.*, and SCRV assemblies (Table 5-2). All bacterial, mitochondrial and viral references were reformatted to include "exon" and "gene" features, and to start with the same tag for the convenience of downstream analysis. Thus, gene IDs were reformatted to start with "MT" for all mitochondrial genomes; "Em" for *E. minasensis*; "Rs" for *R. raoultii*; "Sp" for *Spiroplasma sp.*, and "StCRV" for St Croix River virus.

The STAR reference index was prepared using STAR v2.7.9a (Dobin *et al.* 2013). Following this, STARsolo alignment and quantification was run on the experiments with the following options: "--outSAMtype BAM SortedByCoordinate --outBAMsortingThreadN 2 --limitBAMsortRAM 60000000000 --outMultimapperOrder Random --runRNGseed 1 --outSAMattributes NH HI AS nM CB UB GX GN --soloMultiMappers Uniform PropUnique EM --soloType CB_UMI_Simple --soloCBwhitelist 3M-february-2018.txt --soloBarcodeReadLength 0 --soloUMIlen 12 --soloStrand Forward --soloUMIdedup 1MM_CR --soloCBmatchWLtype 1MM_multi_Nbase_pseudocounts --soloUMIfiltering MultiGeneUMI_CR --soloCellFilter EmptyDrops_CR --clipAdapterType CellRanger4 --outFilterScoreMin 30 --soloFeatures Gene GeneFull Velocyto --soloOutFileNames output/ genes.tsv barcodes.tsv matrix.mtx".

These options were selected to reproduce the workflow of latest Cell Ranger versions (4+), with some important differences. To elaborate,

- since the experiment is 3' scRNA-seq that use v3 of 10x chemistry, the following setting were used: "--soloCBwhitelist 3M-february-2018.txt --soloBarcodeReadLength 0 --soloUMIlen 12 --soloStrand Forward". These settings define the whitelists of barcodes (3M-february-2018.txt file available from Cell Ranger github repository at https://github.com/10XGenomics/cellranger/tree/master/lib/python/cellranger/barcodes), and expect 16 bp cell barcodes and 12 bp UMIs;
- the options "--outSAMtype BAM SortedByCoordinate --outBAMsortingThreadN 2 --limitBAMsortRAM 60000000000 --outMultimapperOrder Random --runRNGseed 1 --outSAMattributes NH HI AS nM CB UB GX GN" generate the sorted BAM file that can be used for read alignment visualisation;
- the options "--soloUMIdedup 1MM_CR --soloCBmatchWLtype 1MM_multi_Nbase_pseudocounts --soloUMIfiltering MultiGeneUMI_CR --soloCellFilter EmptyDrops_CR --clipAdapterType CellRanger4 --outFilterScoreMin 30" were set to mimic the trimming, cell filtering, and UMI/barcode demultiplexing algorithms used by Cell Ranger version 4 and later;
- the option "--soloMultiMappers Uniform PropUnique EM" allowed generation of quantification output that uses both uniquely mapping and multi-mapping reads, with different versions of the algorithm, EM being the most accurate;
- the option "--soloFeatures Gene GeneFull Velocyto" allowed generation of quantification matrices using either normally spliced ("Gene") or both spliced and unspliced ("GeneFull") reference transcripts; "Velocyto" had the matrices necessary to calculate RNA velocities.

The matrices produced by CellRanger were processed with Seurat v.4.0.4 (Butler *et al.* 2018; Stuart *et al.* 2019). A log-normalisation was applied to normalise gene expression of individual cells to the total gene expression of each dataset. Datasets were integrated with the Seurat 'IntegrateData' function with 10000 anchor features (Hafemeister and Satija 2019).

Table 5-2 Reference genomes used in the 10X data processing

Species	**Assembly**	**Genome size, bp**	**Number of contigs**	**N50**	**Number of genes**	**Reference**
I. scapularis strain PalLab	GCA_016920785.2	2,226,883,318	3,313	1,735,392	38,656	Unpublished
I. scapularis mitochondrion	this study	14,184	1	14,184		N/A
E. minasensis strain UFMG-EV	this study	1,347,208	1	1,347,208	1,031	N/A
R. raoultii strain Khabarovsk	GCA_000940955.1	1,483,284	4 (chr + 3 plasmids)	1,344,642	1,702	(El Karkouri *et al.* 2016)
Spiroplasma sp. DMAR11	this study	1,740,586	1	1,740,586	2,246	N/A
R. microplus strain BME/CTVM23	GCA_013339725.1	2,529,930,304	8,624	183,350,851	28,973	(Jia *et al.* 2020)
R. microplus mitochondrion	GCA_013339725.1 (Seq3413)	14,905	1	14,905		(Jia *et al.* 2020)
St Croix River Virus (SCRV)	GCF_000856145.1	18,445	10	N/A	10	(Attoui *et al.* 2001)

5.3. Results

5.3.1. Sequencing yields and data processing

Table 5-3 shows sorting and sequencing yields for the seven submitted samples. The cell lines had comparatively low percentage of viable cells, especially uninfected *I. scapularis* ("Is_uninf"), *R. microplus* infected with *E. minasensis* ("Rm_Em") and *R. microplus* infected with SCRV ("Rm_SCRV"). The percentage of viable cells did not correlate with the infection condition, as the most viable samples were *R. microplus* and *R. microplus* infected with *Spiroplasma sp.*, and one of the least viable was uninfected *I. scapularis*.

Table 5-3 Description of samples taken into sorting and sequencing and sequencing yields

Sample ID	*Is_uninf*	*Is_Em*	*Rm_uninf*	*Rm_Em*	*Rm_Rr*	*Rm_Sp*	*Rm_SCRV*
Density, cell/ml	$5.3*10^5$	$1.5*10^6$	$2.2*10^6$	$2.5*10^6$	$2.5*10^6$	$3.3*10^6$	$1.9*10^6$
Viable cells, %	0.65	0.76	0.81	0.65	0.7	0.79	0.62
Millions of reads	266	291	210	316	234	255	237

Figure 41 illustrates the main quality characteristics of single-cell sequencing such as the percentage of mitochondrial counts per cell, number of molecules (transcripts) per cell and number of expressed genes per cell (Figure 41A), as well as the ratio between the first two values (Figure 41B). *R. microplus* samples had a substantially higher mitochondrial reads count than *I. scapularis* and what is considered normal for a healthy cell (Ilicic *et al.* 2016). This difference between cell lines did not correlate with the observed viability of cells taken into sorting or with the infection state. The high mitochondrial counts in *R. microplus* might mean that these cells are more fragile and more susceptible to the physical impact of filtering and sorting, although the difference could be due to biological differences in mitochondrial

counts between cell types (Medini, Cohen, and Mishmar 2021). Two samples – *R. microplus* infected with *E. minasensis* and *R. microplus* infected with SCRV had especially long distribution of mitochondrial reads from 3% up to 30%. A typical pipeline of single-cell data analysis includes the step of filtration according to number of mitochondrial reads, where all cells with more than 5% mitochondrial counts are removed from the following analysis (Ilicic *et al.* 2016). However, such pipelines were developed for either model organisms, or human tissues, and tick cell lines demonstrated a clearly different pattern, so different thresholds were chosen for these samples: for *I. scapularis*, cells with more than 10% of mitochondrial reads were removed, and for *R. microplus* only those cells with more than 30% of mitochondrial reads were removed. Such mild filtering also saved more than a third part of all cells in *R. microplus* samples (Table 5-4 shows the drastic drop of cell number if filtered with less than 10%). Figure 41B shows the ratio between mitochondrial count and number of transcripts per cell; for *I. scapularis* these were -0.22 and -0.12, while for *R. microplus* samples they were between -0.38 and -0.47. The graphs demonstrate that most of the *I. scapularis* cells with high mitochondrial counts had few transcripts, forming an L-shaped distribution, while graphs for *R. microplus* had distributions of a more triangular shape, when cells with many transcripts also had a high number of mitochondrial transcripts.

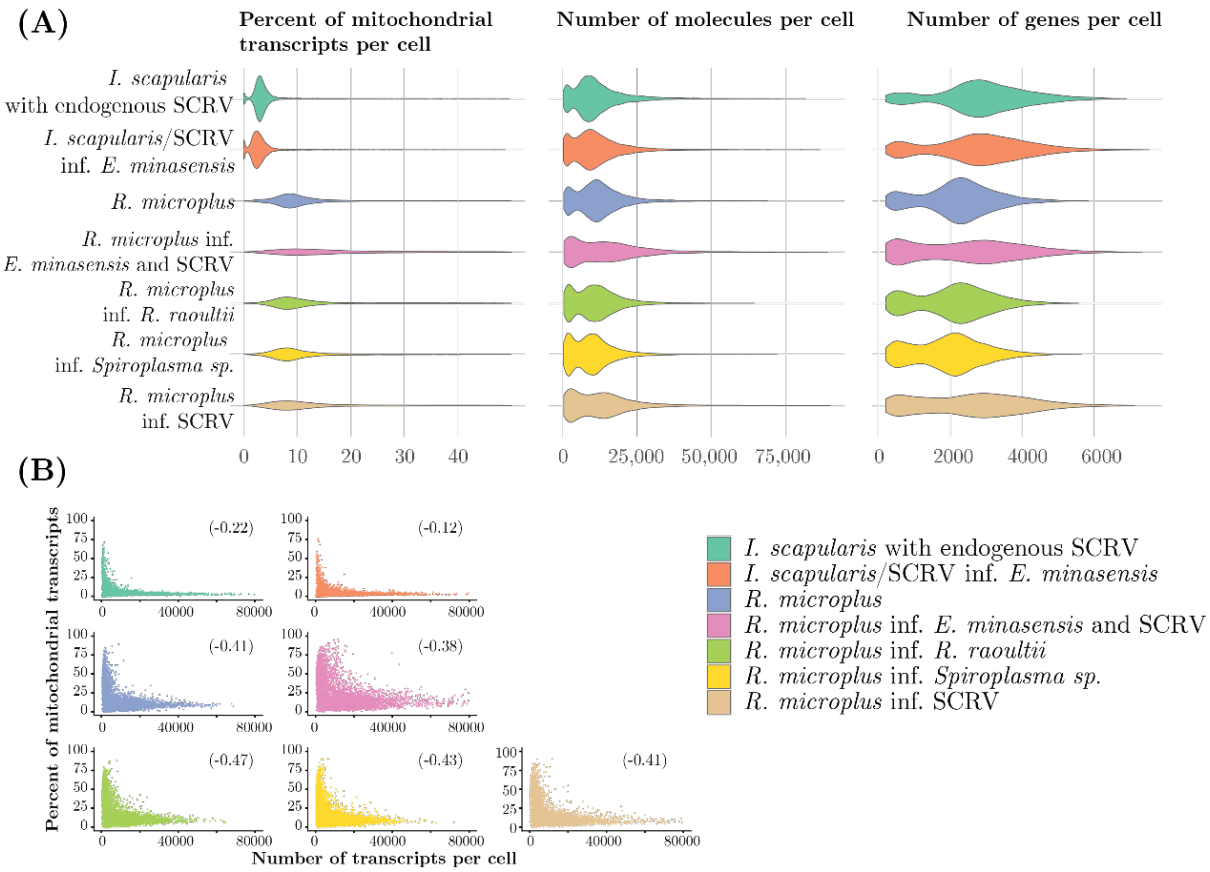

Figure 41 (A) Mitochondrial count, number of RNA molecules and number of detected expressed genes per cell in the samples. (B) Ratio of number of transcripts to mitochondrial counts.

Table 5-4 Number of cells in three of the samples depending on the mitochondrial read count threshold (mt)

	After initial barcodes counts	**After filtering mt<10**	**After filtering mt<20**	**After filtering mt<30**
Rhipicephalus microplus	6361	3068	4641	4719
R. microplus inf. *Ehrlichia minasensis* and SCRV	6186	1310	2836	3262
R. microplus inf. SCRV	5751	2118	3350	3517

Table 5-5 contains a summary of the different steps of single-cell RNA-seq analysis, including counting valid barcodes, UMIs (Unique Molecular Identifiers) and final number of cells. Almost all reads (>97%) were assigned to a valid barcode whitelisted by 10X protocols (Table 5-5, row 2). Between 69% and 88% of all reads were successfully mapped to the reference genomes depending on use of multimapping reads (Table 5-5, rows 3, 4). Approximately half of all reads were matched with annotated genes within the reference genomes (Table 5-5, rows 5, 6). If considering introns of the genes, these figures rose up by 10% (Table 5-5, rows 8, 9). After removing a possible systemic bias caused by contamination by exogeneous or endogenous ambient transcripts using CellBender software, the number of cells in each sample ranged between 5751 and 8708 cells (Table 5-5, row 11).

When cells are squashed during the cell sorting, damaged RNA leaks out and gets inside droplets together with actual cells (Luecken and Theis 2019). This can lead to the distortion of the clustering, integration, and other factors. CellBender is a deep learning-based tool that: 1) learns the model of empty droplets and "certain" cells, defining a new list of filtered cells, which is usually slightly bigger than that of other algorithms; and 2) removes the ambient RNA reads from these cells (Fleming, Marioni, and Babadi 2019). CellBender on average removed about 1% of all UMIs as background (Table 5-5, row 14).

Count matrices of the described data can be found in Digital supplementary material (Chapter V, h5_count_matrices).

Table 5-5 Sequencing outcomes and results of preliminary data analysis

Row	Sample ID	**Is_uninf**	**Is_Em**	**Rm_uninf**	**Rm_Em**	**Rm_Rr**	**Rm_Sp**	**Rm_SCRV**
1	**Millions of reads**	**266**	**291**	**210**	**316**	**234**	**255**	**237**
2	% reads with valid 10x barcodes	97.2	97.3	97.2	97.1	97.2	97.1	97.1
3	% reads mapped to genome, unique + multimappers	80.1	78.1	88.6	86.1	87.9	86.8	86.9
4	% reads mapped to genome, unique	71.4	69.1	79.3	76.6	78.5	77.2	77.7
5	% reads assigned to gene, unique + multimappers	48.1	45.1	57.1	53	56.3	55.5	55.8
6	% reads assigned to gene, unique	46.9	44	54.7	50.4	53.6	52.7	53.4
7	Millions of reads in cells mapped to gene	107	114	102	137	108	114	104
8	% reads assigned to gene w/introns, unique + multimappers	60	57	68.1	65.7	67.4	66.6	65.9
9	% reads assigned to gene w/introns, unique	55.9	53.6	65.1	62.5	64.1	63	62.8
10	Millions of reads in cells mapped to gene w/introns	129	140	123	172	131	138	124
11	**Number of cells by CellBender**	**7296**	**7572**	**6361**	**6186**	**7641**	**8708**	**5751**
12	Millions of UMIs in cells before background correction	85	88	76	81	81	85	71
13	**Millions of UMIs in cells after background correction**	**84**	**87**	**76**	**80**	**81**	**84**	**70**
14	% UMI removed as background	0.79	0.31	0.49	1.13	1	1.56	1.24

5.3.2. Effect of dataset filtering according to mitochondrial counts

It is common practice to filter datasets to remove cells with high mitochondrial counts, usually more than two or three standard deviation measures, as they are believed to be dying cells with leaking membranes (Ilicic *et al.* 2016). However, it should be noted that most of the scRNA-seq guidelines were designed for human or mouse tissues and focused on describing different cell types in samples, where dying cells do not add up to the understanding of a dataset. In the case of the current experiment there was a particular interest in how infection affected the condition of cells including their viability which could be roughly assessed by the number of mitochondrial transcripts. Mitochondrial counts varied between the two tick cell lines as described in section 5.3.1 and Figure 41A: most cells in *I. scapularis* samples had up to 5% of mitochondrial transcripts, whereas the distribution of mitochondrial counts in *R. microplus* cells was much wider at up to 25% (Figure 41A).

Different filtering could have an effect on clustering of the datasets and, depending on whether high bacterial load coincided with high bacterial counts, excessive filtering could bias the analysis of bacterial infection. Figure 42 demonstrates how removal of cells overloaded with mitochondrial transcripts affected the datasets.

The only dataset which largely changed the clustering pattern after removing cells with more than 30% of mitochondrial reads was *R. microplus* infected with *E. minasensis* and SCRV. Clusters on the right side of the cell cloud (2, 4, 12, 13, 14, 15) had up to 100% mitochondrial reads (Figure 42A) and removing these cells reduced the number of clusters from 16 to 12 and made the distribution of mitochondrial reads more even across clusters. Unfortunately, the filtering also removed a fraction of cells with high bacterial load (cluster 12); the other clusters had more uniform distribution of bacterial infection (Figure 42B). A small portion of *I. scapularis* cells with detectable *E. minasensis* signal (cluster 11) was also removed after mitochondrial read filtering

although it did not strongly affect the clustering pattern; the number and size of the clusters remained very similar.

There was a difference in the level of bacterial signal from *E. minasensis* between the two tick cell lines (Figure 42A). Some cells in the *I. scapularis* sample had up to 40% of *E. minasensis* reads, while the maximum fraction of *E. minasensis* reads in *R. microplus* cells was less than 1%.

Cultures infected with *R. raoultii* and *Spiroplasma* sp. also had high mitochondrial counts (up to 50%), however not as high as in the *R. microplus* with *E. minasensis* sample (Figure 42A). Despite the fact that the mitochondrial signal was not regressed it did not seem to bias the clustering - most of the clusters in the *R. raoultii* and *Spiroplasma* sp. samples had similar mitochondrial counts. Removing cells with more than 30% of mitochondrial reads did not change the pattern of clusters much; the *R. raoultii* infection remained even across clusters and *Spiroplasma* sp. infected cells remained within one cluster (cluster 9 before filtering and cluster 7 after filtering).

The following analysis is based on the filtered datasets if not stated otherwise.

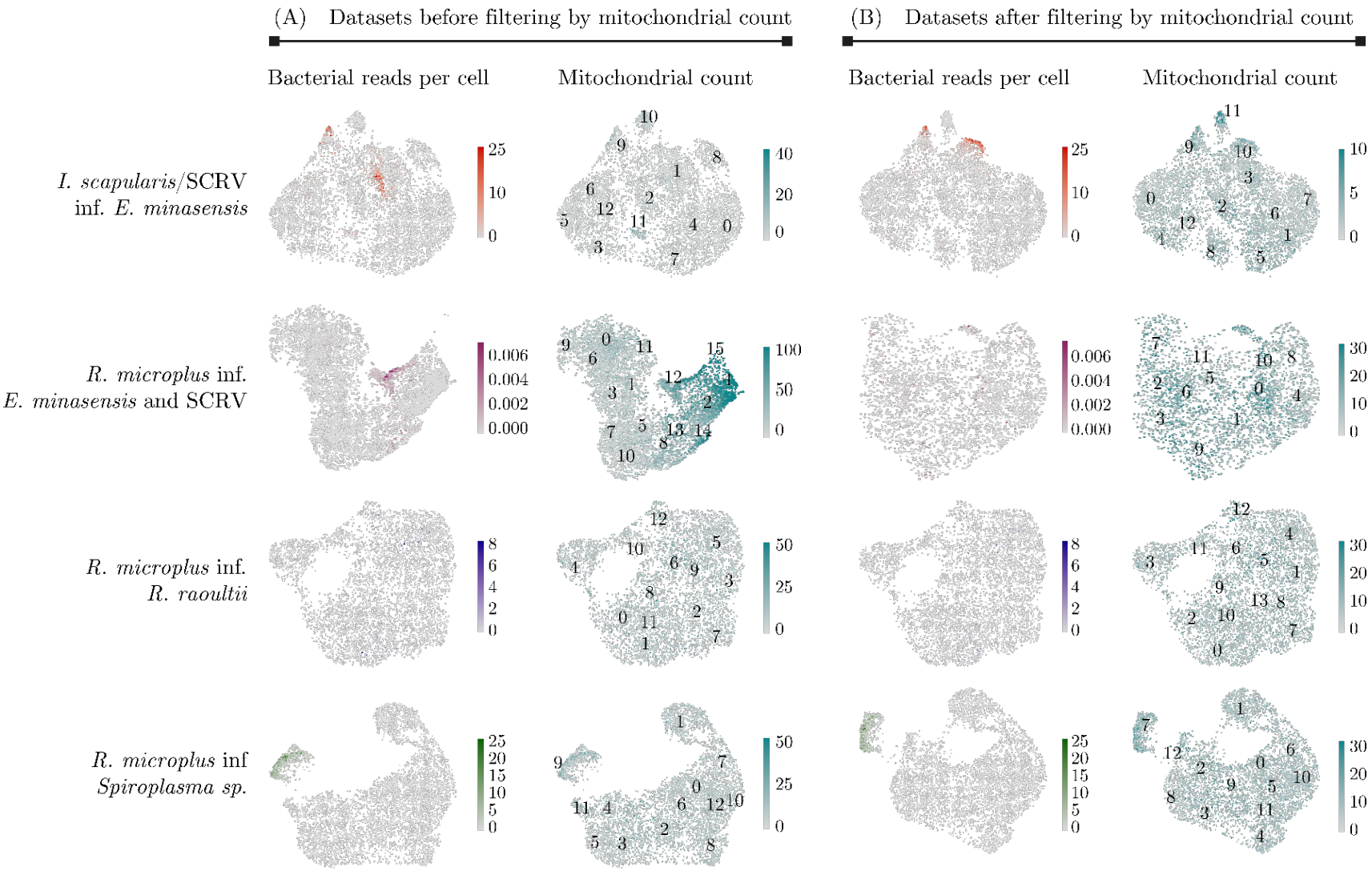

Figure 42 UMAP plots of datasets infected with bacteria before (A) and after (B) filtering by mitochondrial counts.

5.3.3. Analysis of *I. scapularis* samples

The main characteristics of the integrated *I. scapularis* samples are shown in Figure 43. Figure 43A demonstrates that some of the clustering might be explained by the cell cycle. Clusters 0, 4 and 12 were predominantly in G2M phase, clusters 10 and 11 had more cells in G1 phase, and cluster 5 had a substantial bias towards S phase. From Figure 43B it is obvious that the cell cycle distribution was not affected by infection at the time of experiment, and clusters had very similar ratios between uninfected and infected samples.

The dataset was clustered into thirteen groups (Figure 43C, D). The biggest cluster consisted of approximately two thousands cells, the smallest of only 135 cells (Figure 43E). Clusters 0, 1, 6, 7 and 12 contained almost equal number of cells from the infected and uninfected samples, clusters 2, 4, 5 and 8 were slightly biased towards infected cells, and clusters 3, 9, 10 and 11 had more cells coming from the uninfected sample (Figure 43D).

The mitochondrial reads load was not regressed, although it did not affect clustering and all clusters had quite similar mitochondrial count distributions (Figure 43F). Only cluster 11 had a higher percentage of mitochondrial reads. Interestingly, the same cluster was predominantly formed by cells from the infected sample (Figure 43C), although there was no substantial bacterial signal (Figure 43G), which indicated that the poor state of these cells was not directly linked to the level of infection. Clusters 9 and 10 had high counts of bacterial transcripts (Figure 43G).

The level of viral transcripts detected was very low (Figure 43H). Most of the cells in both samples had no detectable viral signal, although there were more cells with detectable SCRV in the uninfected sample, which is shown in Figure 44.

Genes defining clusters are available in the Digital supplementary material (Chapter V, "IDE8_SCRV_Em_cluster_markers.xlsx").

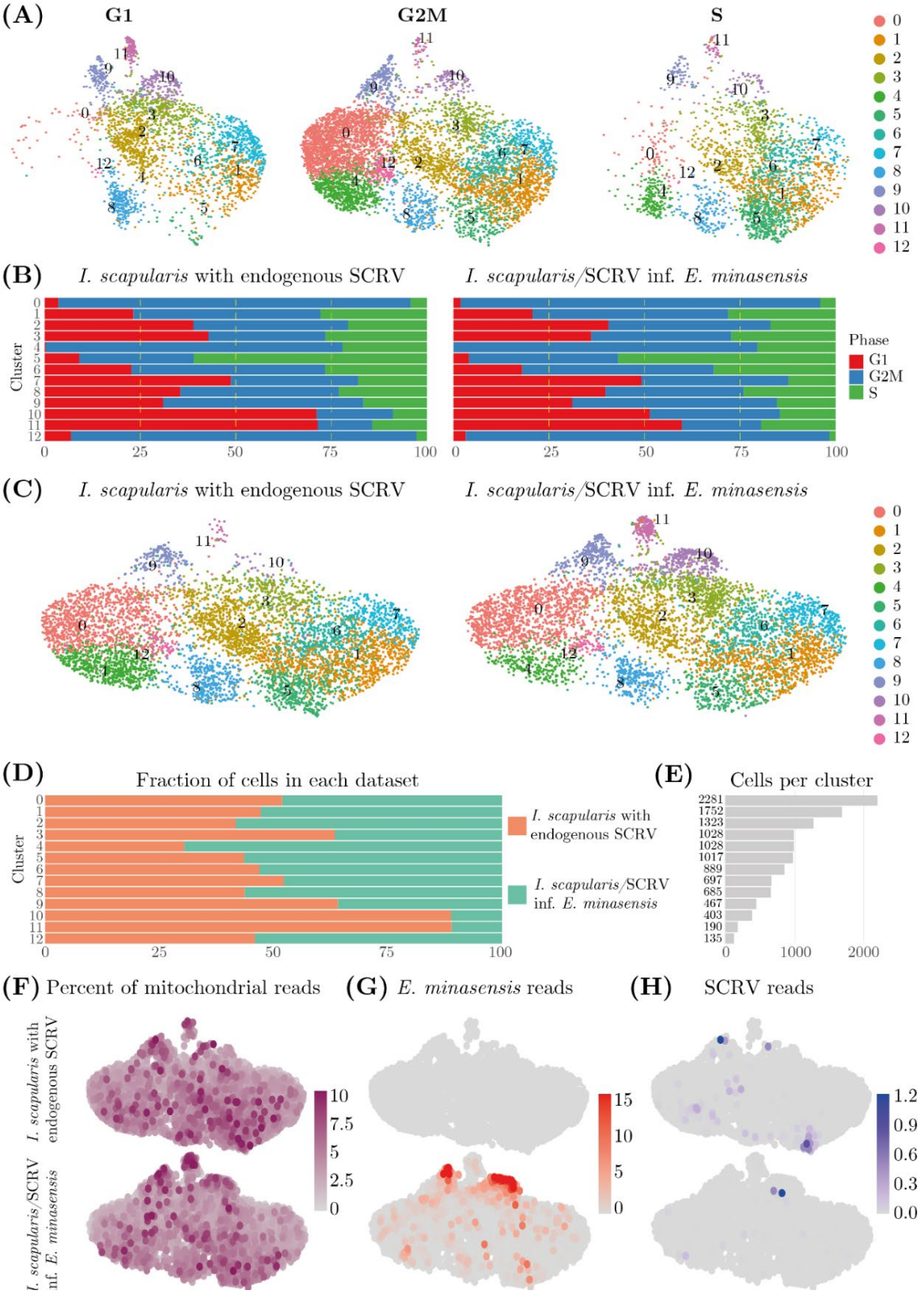

Figure 43 Ixodes scapularis, uninfected and infected with Ehrlichia minasensis, integrated samples. (A) Clusters by cell cycle phase coloured by clustering; (B) Bar plots comparing the ratios of cell cycle phases per cluster between samples; (C) Distribution of clusters between samples coloured by clustering; (D) Bar plot comparing the distribution of cells in clusters between samples; (E) number of cells in each cluster; (F) percentage of mitochondrial reads per cell within infected and uninfected I. scapularis; (G) percentage of reads originating from Ehrlichia minasensis per cell; (H) percentage of reads originating from SCRV per cell.

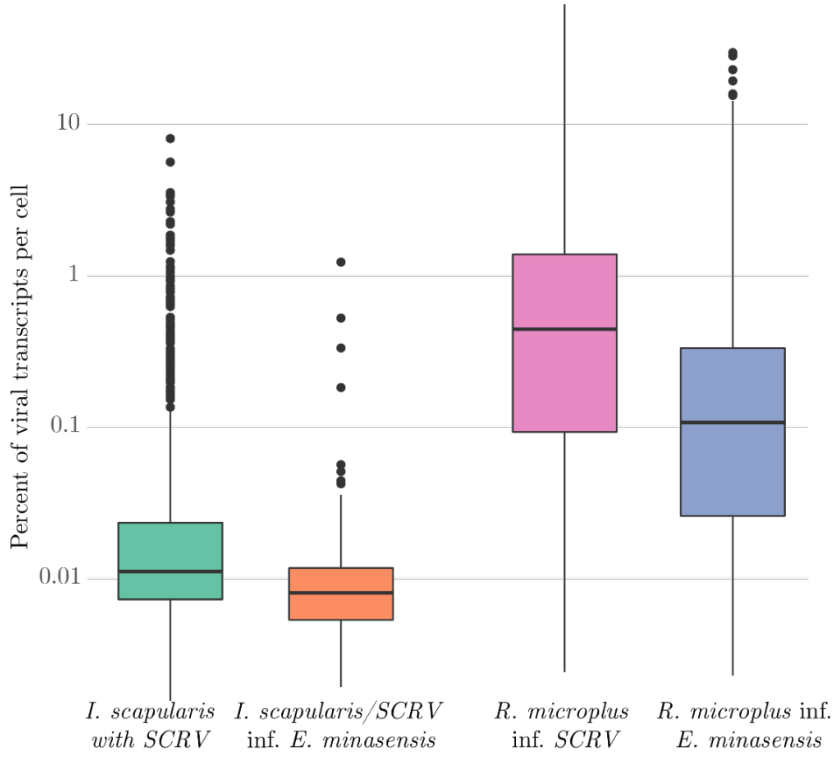

Figure 44 Percent of viral transcripts per cell in the I. scapularis and R. microplus samples. Y-axis is shown with a log_{10} scale.

5.3.3.1. Differentially expressed genes in the infected and uninfected I. scapularis samples

The bacterial signal was not regressed in this analysis thus the strongest marker of the infected sample came from *E. minasensis*. The first two rows of the heatmap (Figure 45) demonstrate the expression of 16S and 23S rDNA genes of *E. minasensis*. Not all cells in the infected sample were actually infected which is reflected in the heatmap (red vertical lines correspond to the cells with detected bacterial signal, and white and light-blue lines correspond to the cells with no bacterial signal); 3667 out of 5937 cells (62%) had no *E. minasensis* transcripts. The level of expression of two *E. minasensis* ribosomal genes should roughly reflect the number of bacteria in a cell. This implied level of infection did not correlate with the level of expression of other host cell marker genes: only actin-5C and ERF3A genes had a similar pattern of

expression among cells, while mitochondrial ribosomal signal (MT-*rrnS* and MT-*rrnS*) negatively correlated with the intensity of the bacterial ribosomal signal (Figure 45).

Another two markers of infection were mitochondrial ribosomal genes *rrnS* and *rrnL*, which were upregulated in the infected sample (Figure 45). Interestingly, while two mitochondrial ribosomal genes were upregulated in the infected sample, some other mitochondrial genes such as *cox2*, *cox1 and nad3* were clearly downregulated in the same sample and were more expressed in the uninfected cells (Figure 45, Supplementary table 6).

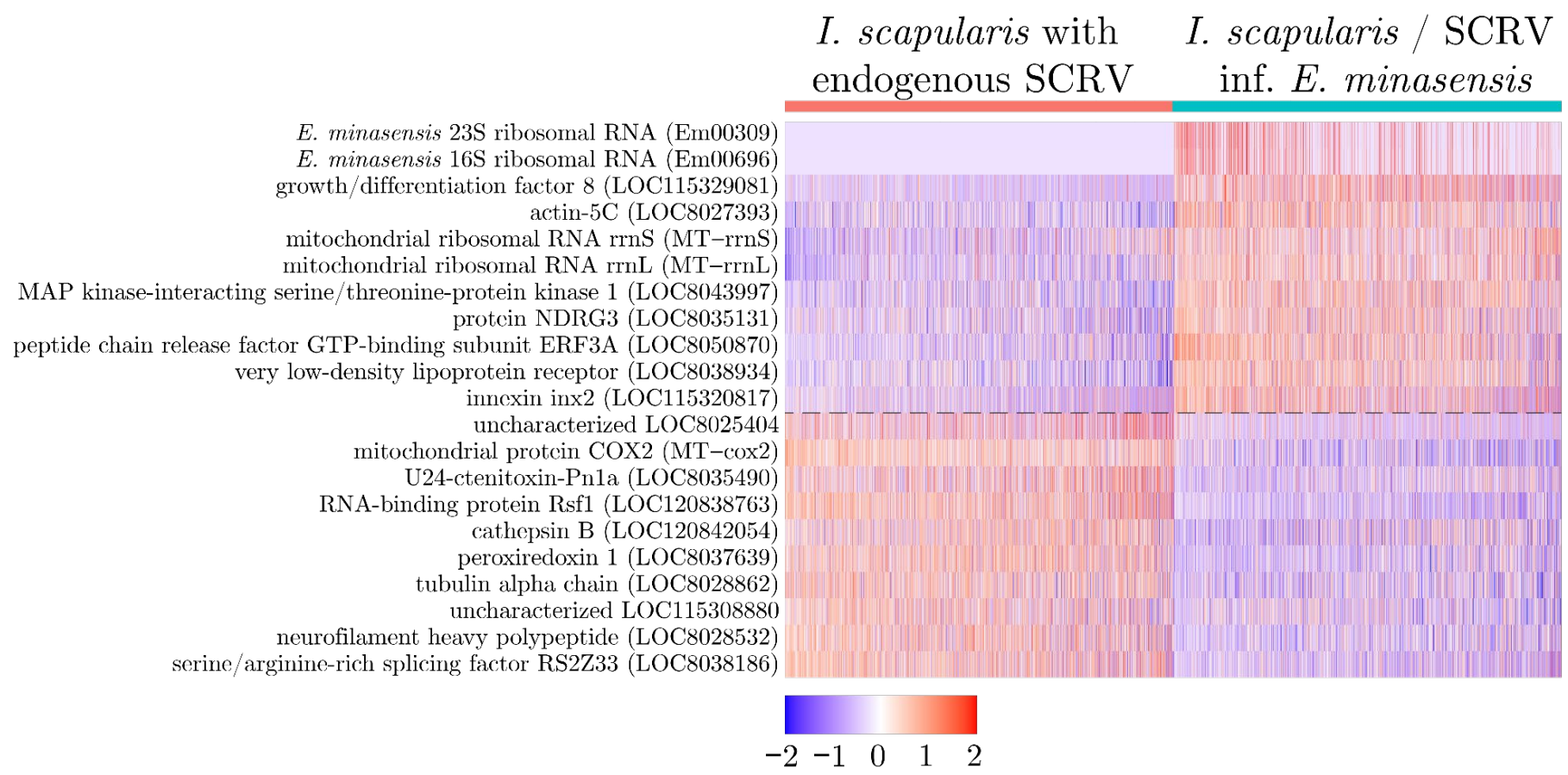

Figure 45 Heatmap of the most differentially expressed genes in Ixodes scapularis samples. Dashed line separates genes up- and downregulated in the two samples. Full list of differentially expressed genes can be found in Supplementary table 5.

5.3.4. Analysis of *R. microplus* samples

There were five samples of *R. microplus* with and without different pathogens. These samples were compared in three groups using the same uninfected control in case of *Spiroplasma sp.* and *R. raoultii* infection, and two controls – *R. microplus* and *R. microplus* infected with SCRV – in case of *E. minasensis* infection (Figure 40).

5.3.4.1. R. microplus infected with E. minasensis and SCRV

The main characteristics of the three integrated *R. microplus* samples (uninfected, infected with SCRV and infected with *E. minasensis* and SCRV) are shown in Figure 46. Figure 46A demonstrates that some clusters might be explained by the cell cycle. So, cluster 7 contains cells in G2M and S phases, most of cluster 2 cells are in S phase, and clusters 1,4 and 8 on the contrary have almost no cells in the S phase. Clusters 1 and 9 have substantially more cells in G1 phase than in the other two phases. The cell cycle seems to be approximately the same between different samples and clusters had very similar ratios between uninfected and infected samples (Figure 46B).

The dataset was clustered into twelve groups (Figure 46C, D). The biggest cluster consisted of 1428 cells, the smallest of only 396 cells (Figure 46E). All clusters had quite similar distributions of mitochondrial counts (Figure 46F). The *R. microplus* infected with *E. minasensis* had more mitochondrial reads than the other two samples (Figure 41A, Figure 46F). The *E. minasensis* infection was distributed more evenly among cells than in the case of *I. scapularis* (Figure 46G): only 796 out of 3262 (24%) cells had no bacterial signal whereas this number was 62% for *I. scapularis* infected with *E. minasensis*. The level of viral transcripts detected was an order of magnitude higher than in *I. scapularis* samples and followed the same pattern that the viral transcription was lower in the presence of *E. minasensis* (Figure 44, Figure 46H). Genes defining clusters are available in the Digital supplementary material (Chapter V, "BME23_SCRV_Em_cluster_markers.xlsx").

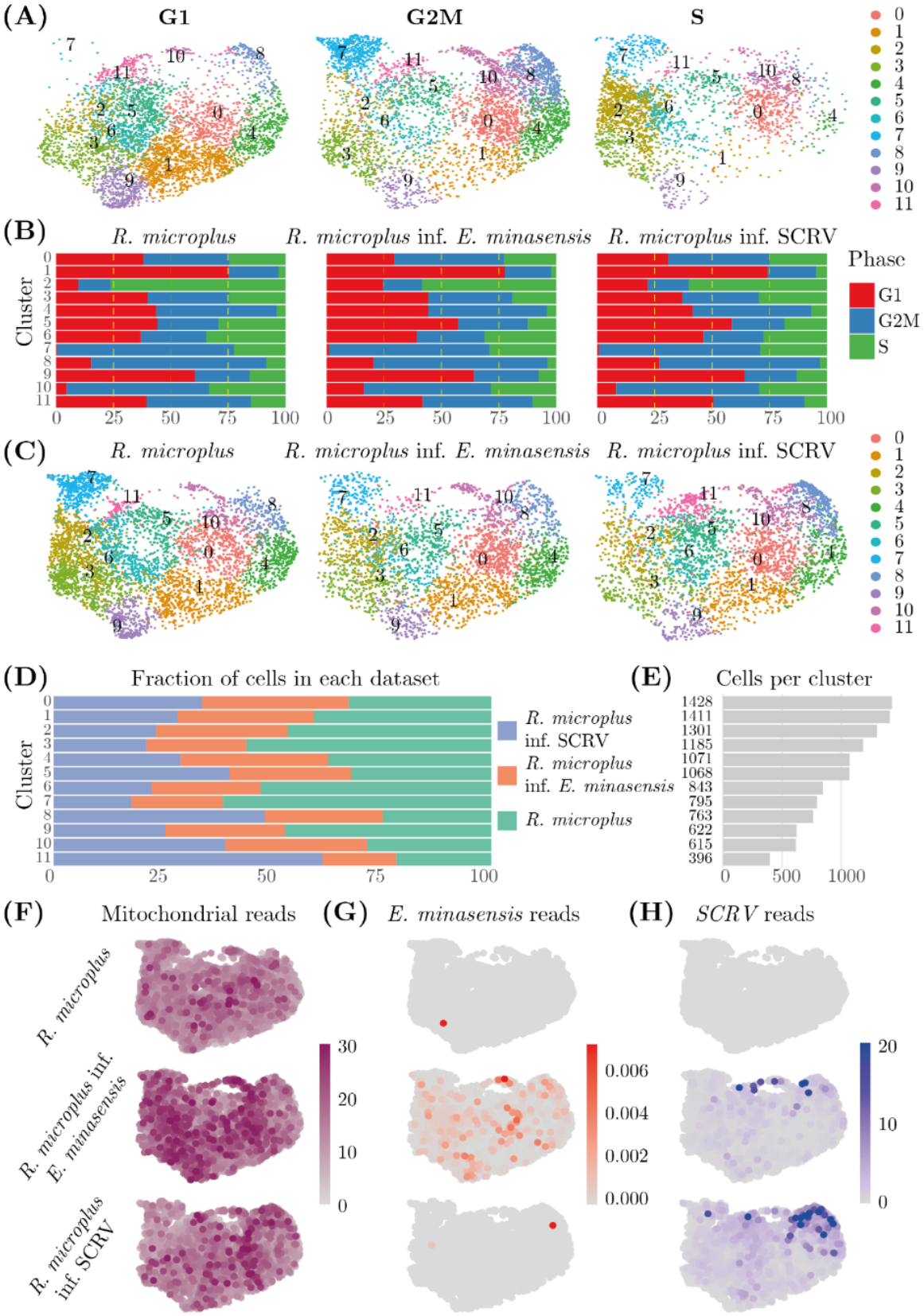

Figure 46 Three R. microplus samples integrated: uninfected, infected with SCRV and infected with E. minasensis and SCRV. (A) Clusters by cell cycle phase coloured by clustering; (B) Bar plots comparing the ratios of cell cycle phases per cluster between samples; (C) Distribution of clusters between samples coloured by clustering; (D) Bar plot comparing the distribution of cells in clusters between samples; (E) number of cells in each cluster; (F) percentage of mitochondrial reads per cell within infected and uninfected I. scapularis; (G) percentage of reads originating from Ehrlichia minasensis per cell; (H) percentage of reads originating from SCRV per cell.

5.3.4.2. Differentially expressed genes in the uninfected R. microplus samples and samples infected with E. minasensis and SCRV

Neither bacterial nor viral signals were regressed in this analysis thus the strongest markers of the infected samples came from *E. minasensis* and SCRV (Figure 47A). Only three genes were distinctively upregulated in the *R. microplus* infected with *E. minasensis* and SCRV sample: two ribosomal genes of *E. minasensis* and the tick gene γ-interferon-inducible lysosomal thiol reductase. The latter was also upregulated in the *R. microplus* with SCRV sample although the signal was weaker.

There were several main types of marker genes in the *R. microplus* infected with SCRV sample: viral transcripts, stress response proteins and spliceosomal RNAs (Figure 47A, C, Supplementary table 7). Viral genes were also expressed in the *R. microplus* with *E. minasensis* and SCRV sample to a lesser extent (Figure 47A, Figure 44).

The top ten genes which were upregulated in the uninfected sample compared with both infected samples are shown in Figure 47A (LOC119161184 to LOC119170592). Other upregulated genes with lower fold change rates are shown in Supplementary table 7. Many of these genes belonged to the family of ribosomal proteins, and their expression was mostly concentrated in clusters 2, 3, 6, 7, 9 and 11 (Figure 47B).

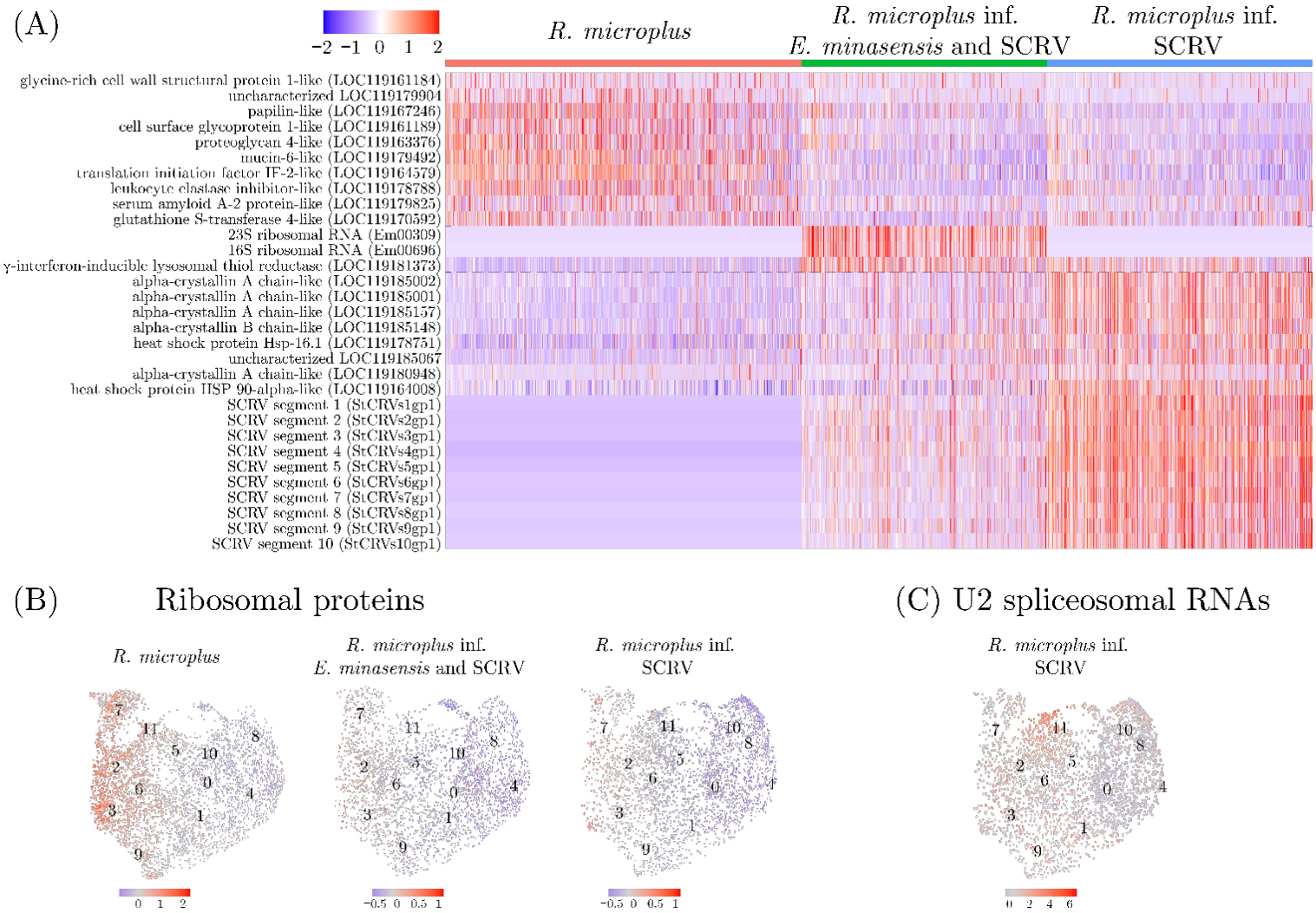

Figure 47 (A) Heatmap of differentially expressed genes in the samples of uninfected Rhipicephalus microplus, R. microplus infected with SCRV and R. microplus infected with Ehrlichia minasensis and SCRV. Dashed lines separate genes up- and downregulated between samples. Full list can be found in Supplementary table 6. (B) Expression of ribosomal proteins in three different samples. (C) Expression of U2 spliceosomal RNAs in the R. microplus infected with SCRV cells.

5.3.4.3. R. microplus infected with R. raoultii

The main characteristics of the *R. microplus* and *R. microplus* infected with *R. raoultii* integrated dataset are shown in Figure 48. Figure 48A demonstrates that the distribution of cells and clusters among cell phases was not even. Most cells were in the G1 phase and least cells in the S phase. Clusters 3, 11 and 12 were predominantly built of dividing cells (G2M), clusters 0 and 2 had equal numbers of cells in each phase, other clusters consisted mostly of G1 cells (more than 50%) (Figure 48B). The cell phase distribution between infected and uninfected samples was very similar (Figure 48B).

The dataset was clustered into fourteen groups with the biggest cluster of 1221 cells and the smallest of only 156 cells, with all clusters almost equally divided between infected and uninfected samples (Figure 48D). All clusters had quite high mitochondrial counts (Figure 48F), and the distribution of *R. raoultii* reads was also quite even across cells and clusters (Figure 48G). Only approximmately a third part of cells (1571 of 5043) had no bacterial signal.

Genes defining clusters are available in the Digital supplementary material (Chapter V, "BME23_Rr_cluster_markers.xlsx").

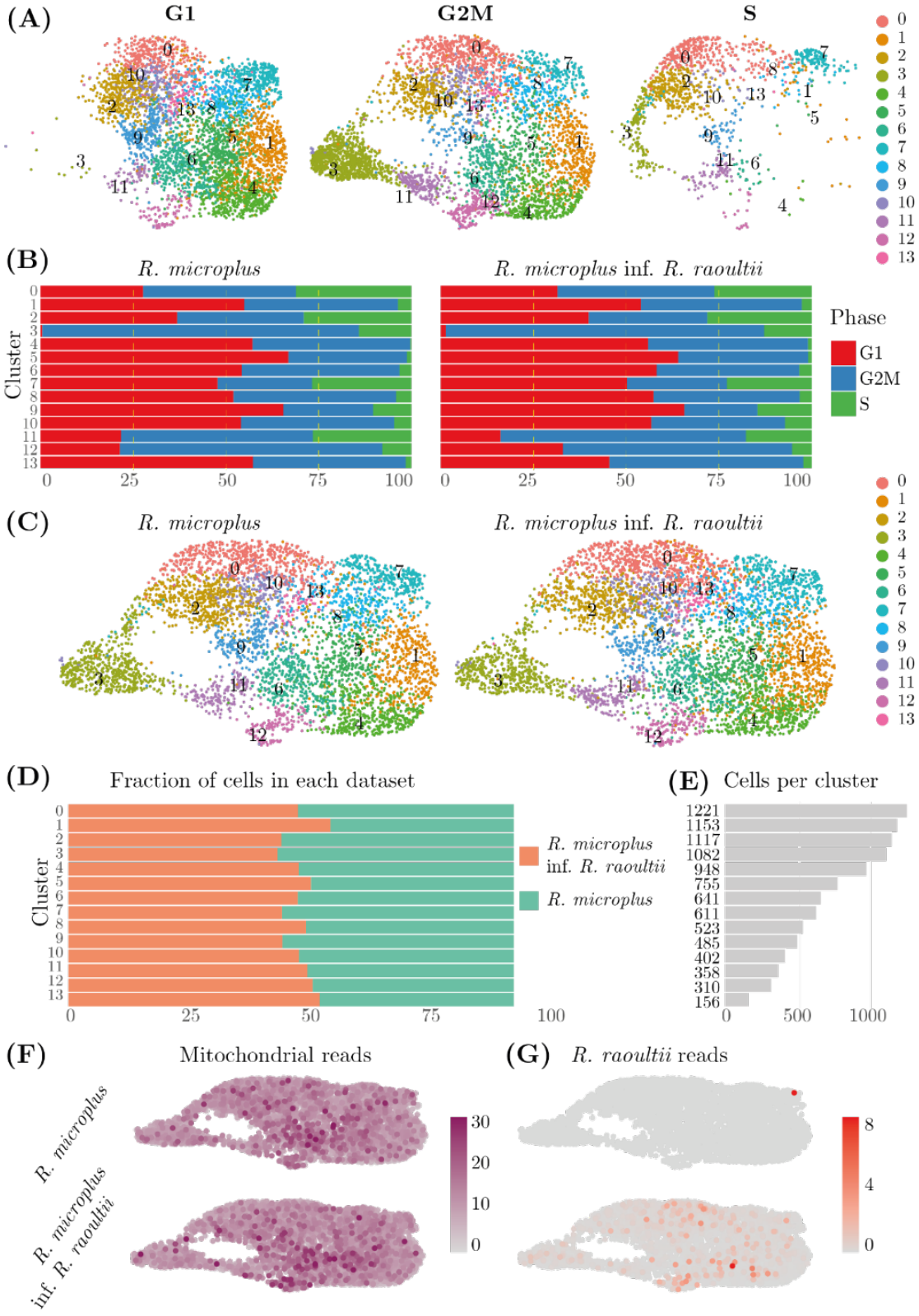

Figure 48 Two Rhipicephalus microplus samples integrated: uninfected, infected with Rickettsia raoultii. (A) Clusters by cell cycle phase coloured by clustering; (B) Bar plots comparing the ratios of cell cycle phases per cluster between samples; (C) Distribution of clusters between samples coloured by clustering; (D) Bar plot comparing the distribution of cells in clusters between samples; (E) number of cells in each cluster; (F) percentage of mitochondrial reads per cell within infected and uninfected R. microplus; (G) percentage of reads originating from R. raoultii per cell.

5.3.4.4. Differentially expressed genes in the uninfected R. microplus samples and infected with R. raoultii

There were not many genes differentially expressed between the uninfected *R. microplus* sample and *R. microplus* with *R. raoultii* (Figure 49, Supplementary table 8). The most certain marker came from *R. raoultii* – two ribosomal genes. Another two genes which were upregulated compared with the uninfected sample were fatty acid synthase and insulin-induced gene 1 protein. Another seven genes were slightly downregulated in the infected sample, three of which had no assigned function.

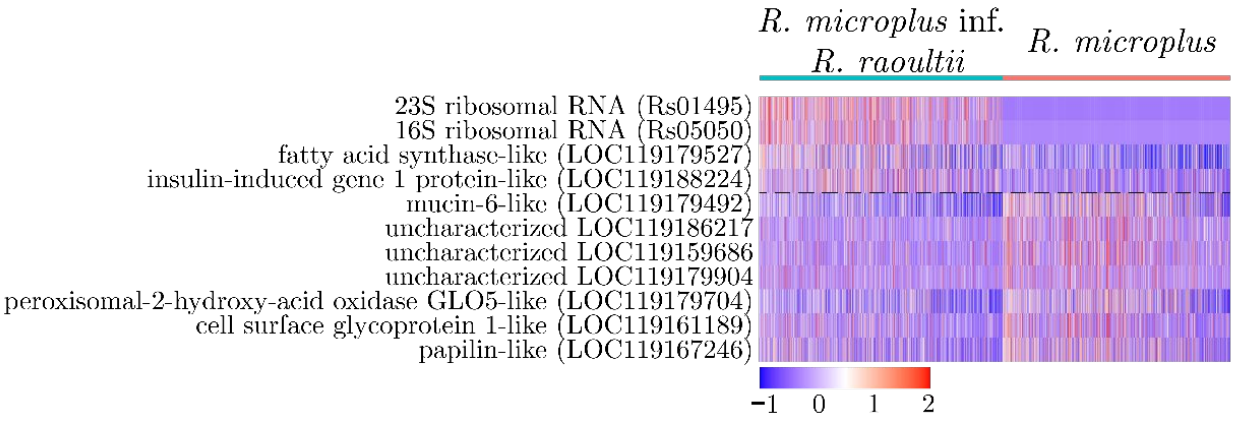

Figure 49 Heatmap of differentially expressed genes in samples of uninfected Rhipicephalus microplus and R. microplus infected with Rickettsia raoultii

5.3.4.5. R. microplus infected with Spiroplasma sp.

The main characteristics of the *R. microplus* and *R. microplus* infected with *Spiroplasma* sp. integrated dataset are shown in Figure 51. Figure 51A demonstrates that the distribution of cells and clusters among cell phases was not even. Most of the clusters had cells either in G1 or G2M states, except for two clusters: cluster 1 was predominantly built with dividing cells (G2M), and cluster 6 had more than half of the cells in the synthetic phase whereas all other clusters had maxima of 10% of cells in the S phase (Figure 51B). The cell phase distribution between infected and uninfected samples was almost identical (Figure 51B).

The dataset was clustered into thirteen groups with the biggest cluster of 1386 cells and the smallest of only 231 cells (Figure 51D, E). Clusters 7 and 11 were biased towards infected cells, all other clusters had almost even parts of infected and uninfected cells. 1744 cells out of 5586 (31%) had no bacterial signal. All clusters had quite high mitochondrial counts with cluster 7 having a slightly higher level of mitochondrial expression (Figure 51F), which coincided with the distribution of *Spiroplasma* sp. transcripts; most of the signal from *Spiroplasma* came from cluster 7 (Figure 51G). Genes defining clusters are available in the Digital Appendix (Chapter V, "BME23_Sp_cluster_markers.xlsx").

5.3.4.6. Differentially expressed genes in the uninfected R. microplus samples and infected with Spiroplasma sp.

Spiroplasma infection also did not change the expression profile of tick cells very much, similarly to the *R. microplus* infected with *R. raoultii.* Only nine genes were identified as differentially expressed, eight of which were upregulated in the infected sample (Figure 50). The highest numbers of fold change were exhibited by two 23S rDNA *Spiroplasma* sp. genes (*Spiroplasma* sp. had two ribosomal operons). Similar to *R. microplus* infected with *E. minasensis*, a ribosomal protein and small nuclear U3 RNA involved in eukaryotic mRNA processing were upregulated in the infected sample. The fatty acid synthase gene was upregulated in the infected sample, similar to the *R. microplus* sample with *R. raoultii*.

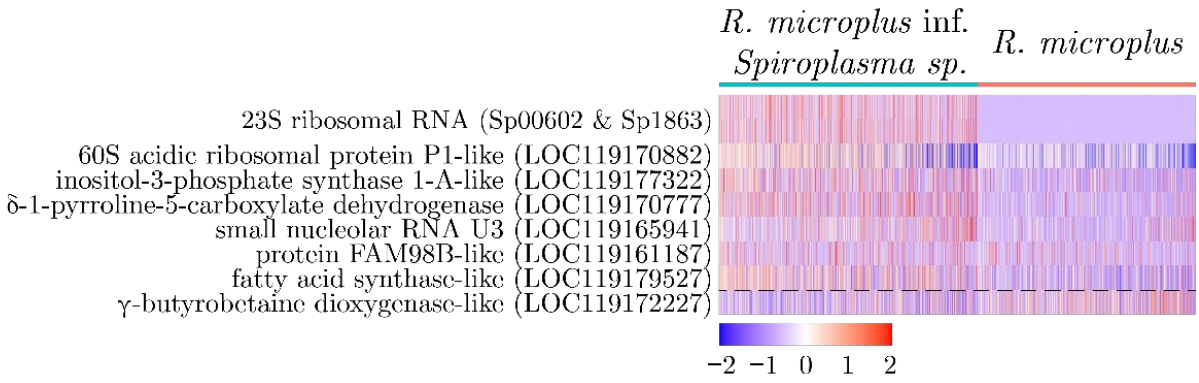

Figure 50 Heatmap of differentially expressed genes in samples of uninfected Rhipicephalus microplus and R. microplus infected with Spiroplasma sp.

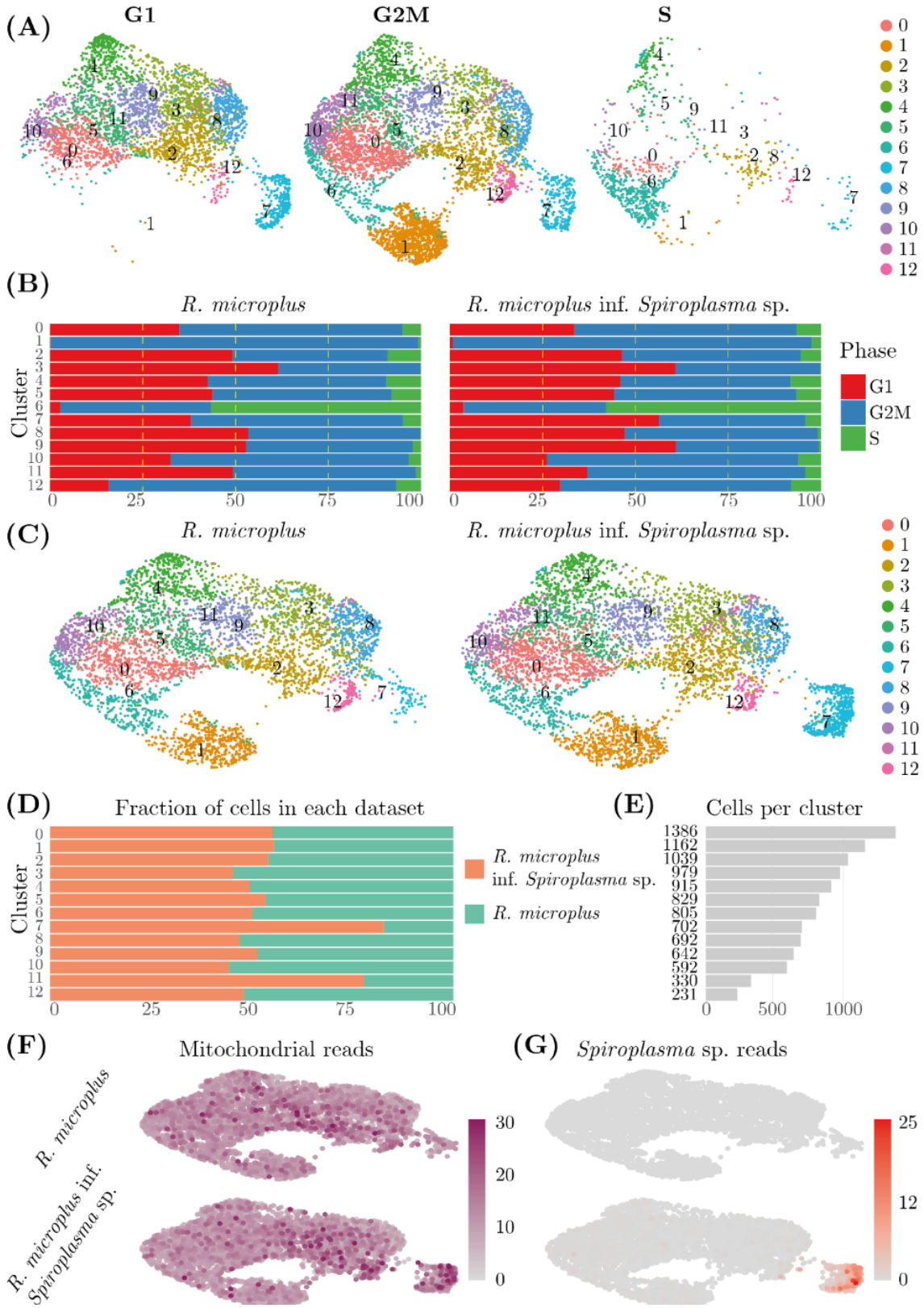

Figure 51 Two Rhipicephalus microplus samples integrated: uninfected and infected with Spiroplasma sp. (A) Clusters by cell cycle phase coloured by clustering; (B) Bar plots comparing the ratios of cell cycle phases per cluster between samples; (C) Distribution of clusters between samples coloured by clustering; (D) Bar plot comparing the distribution of cells in clusters between samples; (E) number of cells in each cluster; (F) percentage of mitochondrial reads per cell within infected and uninfected R. microplus; (G) percentage of reads originating from Spiroplasma sp. per cell.

5.4. Discussion

5.4.1. Data processing

The field of single-cell analysis technologies has drastically increased during the last decade, and there are hundreds of tools to analyse single-cell datasets (385 tools as of March 2019) (Luecken and Theis 2019). Three characteristics are commonly used to assess experiment quality: counts per barcode (count depth), the number of genes per barcode, and the fraction of counts from mitochondrial genes (Ilicic *et al.* 2016). Outliers with a higher count depth usually correspond to doublets; low count depth corresponds to an empty droplet with an ambient RNA. If few genes or a high percentage of mitochondrial reads are detected, it indicates the leaking of the outer membrane and a dying state of the cell (Ilicic *et al.* 2016). As counts and the number of transcripts vary considerably between cell types, these QC covariates should be analysed jointly (Luecken and Theis 2019). Another important filtering step is removing genes that are not expressed in more than a few cells and therefore are not informative for cluster formation (Luecken and Theis 2019). Here, the probability of dropout events and the expectation of small clusters in the dataset have to be considered. For example, filtering out genes expressed in fewer than N cells will most likely eliminate a cluster with N cells. Based on these considerations and several iterations of preliminary analysis, genes present in less than three cells and cells with less than 200 transcripts were filtered out. This filter should have removed dying cells and empty droplets with ambient RNA. Cells with a high fraction of mitochondrial genes were not entirely removed from this analysis as dying cells were also interesting for the study and often correlated with cells with a high load of bacteria. Such cells might give an insight into which factor was the most detrimental for tick cells (bacterial infection, viral infection, or something else).

Normalisation is a crucial step for scRNA-seq experiments. Raw read counts cannot be directly used to compare gene expression between cells because of cell-to-

cell variations in a number of molecules detected in each cell which is the result of bias introduced during cell lysis and reverse transcription and stochastic molecular sampling (Hebenstreit 2012). In a well-normalised dataset, the variance of a gene should be independent of gene abundance and sequencing depth of a cell.

Factors which could be considered confounding are usually regressed to reveal a "true" signal. These are mitochondrial counts, pathogen read counts, ribosomal genes, stress response genes and cell cycle genes, and regressing these factors theoretically should reveal cell type composition in a sample. However, the regression of confounding factors inexplicably removed the prominent marker genes and produced a very weak signal. Supplementary table 10 and Supplementary table 11 show differentially expressed genes with a highest log fold change after the regression of signals from mitochondria and symbionts. The majority of genes exhibited only minor variations in expression, while those that were notably upregulated were primarily connected to stress and linked to groups of cells undergoing cell death. The reason behind such behaviour is not entirely clear; it might be caused either by the complexity of the data and imperfections of the particular analysis pipeline, or by the fact that there is no underlying cell diversity except for the one created by cell cycle progression, stress response of a cell and bacterial and viral loads. For this reason, the main analysis presented in the study is done without regression of confounders.

5.4.2. Transcriptomic response to various bacterial infections

Without regression of bacterial signal, the most prominent markers were expression of bacterial ribosomal genes (23 and 16S rDNA). Such result was not obvious because it was not clear whether the bacterial signal would be strong enough to detect or the eukaryotic signal would outcompete bacterial expression. Moreover, it was not clear what minimal level of bacterial infection was necessary for signal detection. It is very likely that there were cells in the infected samples with a low number of bacteria which did not display a bacterial signal due to the dropout effect on minor templates.

The patterns of infection and transcriptomic responses to the same pathogenic bacteria in two different tick cell lines were quite different. First of all, the distribution of bacterial signal varied: *E. minasensis* observably infected only 38% of cells in the *I. scapularis* sample whereas 76% of cells in the *R. microplus* sample demonstrated a clear bacterial signal even after filtering the dataset and removing a fraction of infected cells as they also had a high mitochondrial count. Secondly, the level of bacterial infection per cell drastically differed: *I. scapularis* cells infected with *E. minasensis* had up to 25% of reads coming from the pathogen, while *R. microplus* infected cells had less than one percent of bacterial reads (Figure 42). This might result from the fact that *E. minasensis* had been maintained in *I. scapularis* cells for a long time, whereas *R. microplus* cells were introduced to *E. minasensis* only ten passages before the experiment.

R. raoultii and *Spiroplasma* sp. infection of *R. microplus* cell lines did not change the transcriptomic portrait of the samples despite the substantial bacterial load in parts of the cell populations. Rickettsial reads were detected in 69% of cells with bacterial load up to 8% (Figure 42). *Spiroplasma* sp. signal was detected in the same fraction of cells with even higher load up to 25% (Figure 42). However, very few genes were defined as differentially expressed in these samples. Apart from the bacterial ribosomal signal, two tick genes were upregulated and seven downregulated in the *R. microplus* infected with *R. raoultii* samples (Figure 49) and six genes up- and one downregulated in the sample with *Spiroplasma* sp. infection (Figure 50).

One of the genes which was downregulated in the *R. microplus* sample infected with *R. raoultii* and *E. minasensis*/SCRV was papilin. This gene is associated with antibacterial activity according to the UniProt record G0WRZ9. Some experiments in *Drosophila* flies showed that papilin is excreted by haemocytes as a part of defence and is associated with the phagocytosis of apoptotic cells (Kramerova *et al.* 2000).

The fatty acid synthase gene was upregulated in the *R. microplus* samples infected with *R. raoultii* and *Spiroplasma* sp. (Figure 49, Figure 50). This gene was

described as part of the extrinsic apoptosis pathway in *I. scapularis* and shown to be upregulated in the presence of *Anaplasma phagocytophilum* (Ayllon *et al.* 2015). *A. phagocytophilum* infection inhibits cell apoptosis which is beneficial for the increase of infection (P. Alberdi *et al.* 2016); here the same effect occurred: the upregulated fatty acid synthase should have inhibited cellular apoptosis. Strangely, this gene was not differentially expressed in either of the cultures with *E. minasensis* infection.

5.4.1. No differential expression of classic immunity pathways

Two of the three bacteria used in the present study are classified as Gram-negative (*E. minasensis* and *R. raoultii*), while *Spiroplasma* sp. is Gram-positive. According to the literature, the Toll pathway is triggered by cell wall components of Gram-positive bacteria (Smith and Pal 2014), and the IMD pathway is induced by Gram-negative bacteria such as *A. phagocytophilum* (L. Liu *et al.* 2012). No genes associated with these pathways were expressed in the samples. The reason behind such behaviour is not clear. One possible explanation is the assumption that these pathways are activated immediately after the exposure of a pathogen and their signals are weakened after several days of infection. Another reason might be the strong stress response of the cells which outcompeted less prominent signals. During the library preparation of scRNA-seq samples minor templates might be lost causing so-called dropout events (Lun, Bach, and Marioni 2016).

In other transcriptomic experiments, where a clear signal from classic immune pathways was detected, cells were collected for the analysis 24 h after adding the infecting agent (Conceicao *et al.* 2021). Experiments with tick cells and bacterial pathogens surveyed on the 7th day after infection also showed differential expression of sets of genes not directly related to described immune pathways (P. Alberdi *et al.* 2016).

5.4.2. Suppression of viral expression in the presence of *E. minasensis*

Two samples of tick cells were infected with *E. minasensis* together with SCRV to try to separate the effects of bacterial infection and viral infection. The *I. scapularis* culture contained the virus as a chronic infection, while the *R. microplus* cells were introduced to the virus together with the *E. minasensis* bacteria. Both samples showed a similar expression pattern in which the viral signal was less prominent in the presence of bacteria (Figure 44). Such a phenomenon when two pathogens are competing with each other and inhibiting the competitor is known. It is not known whether such effect is nonspecific and relies on the more active immune response or such interaction is specific between SCRV and *E. minasensis*. Nonspecific interactions might be explained similar to what was observed in the human immune system and called "trained immunity" or "innate immune memory" when innate immunity can display adaptive characteristics after challenge with pathogens (Netea *et al.* 2016). If the suppression of SCRV infection appears to be specific for *E. minasensis* the mechanism might be similar to the one in the symbiotic system of mosquitoes infected with *Wolbachia* bacteria that restrict the ability of the mosquitoes to transfer dengue, Zika, chikungunya and yellow fever viruses (Lu *et al.* 2012; Kamtchum-Tatuene *et al.* 2016).

The *R. microplus* cell culture which was introduced to SCRV shortly before the experiment expressed a clear U2 spliceosomal RNAs signal (Figure 47C) indicating that the cells were mounting a defence response against the virus as previously described for adeno-associated viruses (Schreiber *et al.* 2015).

The results of the experiment described in this chapter raise more questions than give answers, and the possible underlying reasons of such transcriptomic responses and future directions are discussed in the concluding discussion (section 6.4).

6. Chapter VI Concluding discussion

6.1. The context of the studies

The diversity of symbiotic interactions results from evolutionary transitions between different lifestyles, which range from mutualistic to parasitic interactions and can involve a variety of transmission modes. Organisms that demonstrate diverse transmission biology are crucial for advancing our understanding of the ecology and evolution of symbiotic relationships, and for gaining insights into the biological processes that shape the diversity and functioning of ecosystems. However, the study of transitions in transmission mode and lifestyles is often limited by a lack of clades or groups of organisms that encapsulate diverse transmission biology (Drew *et al.* 2021). The research of eukaryotic symbionts of arthropods is mainly limited by pathogens affecting humans, thus the diversity of symbionts is underinvestigated and the knowledge is biased. Therefore, more even and continuous knowledge about symbiont diversity and functionality is crucial for further symbiotic lifestyle research.

Studying of the symbiotic diversity is tightly linked with the development of modern methods of molecular biology. So the high throughput sequencing led to the boom in protist research and discovery a lot of new symbiotic protists (Adl *et al.* 2012; Campo, Bass, and Keeling 2020). Although there are still niches where all existing methods fail to detect symbiotic organisms if the symbiotic load is quite low (Bass *et al.* 2023; Schneider *et al.* 2014). Chapter II of this thesis addressed this issue and offered a new solution.

The interactions of symbiosis happen on many levels, and some symbionts are able to exchange genetic material with each other. Gene transfers into eukaryotic genome from symbionts or organelles were considered a rare evolutionary event although recent studies claim that such transfers may happen more frequently than previously thought (Bruto *et al.* 2014; Puigbò *et al.* 2011). Transfers are typically

detected by their consequences – similar genes are present in distantly related organisms and mechanisms by which gene transfers happen are the open question (J. O. Andersson 2005). The study described in Chapter III aimed to develop a method which should help to understand the frequency of such transfer, although if the further studies will show that the rickettsial insertion had only happened in the cell line and not in wild population of ticks, it might provide a practical *in vitro* model for exploring the molecular mechanisms of genomic transfers from prokaryotic symbionts to eukaryotic hosts (more on this in section 6.3).

Another important factor in interactions in symbiosis is physiological control of one partner by the other. The immune system plays a key role in balancing of symbiotic relationships; the ability to successfully establish symbiosis is tightly coupled with recognizing, responding and controlling symbionts by hosts and the ability to cope with immune reactions by symbiotic invaders (Anbutsu and Fukatsu 2010). Because all host animals must protect themselves against colonization by inappropriate or pathogenic microorganisms, a central theme in bacteria–host interactions is how the symbiont either avoids damage by the defenses of the host or communicates with host cells to modulate them (Ruby 2008).

Recent advances in genetics and molecular biology have greatly expanded our understanding of the details of the immune mechanisms in *Arthropoda*, but some groups were less studied than others and most of the research is focused on model organisms such as *Drosophila* flies (Shaw *et al.* 2018). Ticks, despite their huge importance for humans, are poorly studied from the immunological point of view. Ticks are evolutionary far from insects and have developed under very different environmental conditions, thus their immune reactions might be very different from those described in insect models. Transcriptomic experiment described in Chapters IV and V intended to contribute to knowledge on tick immunity.

Overall, this thesis aimed to develop approaches which will allow to look into symbiotic interactions beyond well-known model systems and extend the current knowledge about symbiosis.

6.2. Improved 18S metagenetic survey to reveal hidden protistan diversity in *Metazoa*

Protists (unicellular microeukaryotes) are abundant members of almost all Earth communities. According to the newest phylogenetic tree of eukaryotes, only three branches have multicellular organisms: *Opisthokonta* branch contains all animals and fungi, *Archaeplastida* contain green plants and red algae, *Phaeophytes* or brown algae reside within the SAR branch (Adl *et al.* 2012; Burki 2014). All other branches (from 4 to 8 according to different classifications) solely contain protists. The estimations of true protistan diversity are often discordant and subjective (Pawlowski *et al.* 2012); however, all authors agree that the group has been hugely neglected until recently due to the complexity of their cultivation in laboratory conditions and the lack of appropriate methods to assess this diversity. Recent developments in high-throughput sequencing techniques and targeted efforts to describe previously neglected communities such as deep ocean waters revealed a huge number of new species, genera and families of unicellular eukaryotes and the point of saturation is still to reach (Seeleuthner *et al.* 2018; Sieracki *et al.* 2019; Vernette *et al.* 2021).

Modern approaches to characterise free-living communities of protists using universal primers such as 18S rDNA proved to be effective and relatively unbiased (Hugerth *et al.* 2014; Bradley, Pinto, and Guest 2016); at the same time, methods for detection of symbiotic protists in *Arthropoda* and other metazoan hosts rely on targeted PCR-based analysis, such as amplification of particular species or genera. This approach effectively reveals specific well-known pathogens but does not provide the whole composition of symbiotic communities. Current PCR-based methods have two limitations preventing the complete characterisation of protistan communities. The

first problem is using different barcoding genes for different taxa: while some taxonomic groups rely on ribosomal genes or ITS regions, other groups are assessed using mitochondrial cytochrome c oxidase genes (Pawlowski *et al.* 2012). The second problem arises because DNA ratios of a metazoan host and unicellular symbionts differ by several orders of magnitude, and minor templates are often lost during the first rounds of PCR (Kalle, Kubista, and Rensing 2014). Such dropout events may lead to false-negative screening results.

This study analyses different approaches for an integral analysis of protistan communities based on ribosomal genes and attempts to develop an integral method for revealing a true composition of the protistan part of microbiomes. The main conclusion of the chapter was that most of the methods which were tested in the laboratory (restriction, DNA- and PNA-blocking) are either unreliable or ineffective. The most promising approach was using the non-*Metazoa*n primers designed by Bower and co-authors (Bower *et al.* 2004). Considering that the amount of available genomic data in 2004 was by orders of magnitude less than what is available nowadays, we decided to use this advantage and attempt to design a more efficient pair of primers biased against *Metazoa*. Bower *et al.* aimed to use their primers with 454 sequencing, which allowed amplicons of approximately 500 bp in length. Later, the protocol was modified to make it compatible with The Illumina platform (del Campo *et al.* 2019). Nowadays, the rise of long-read sequencing technologies allows longer amplicons and therefore gain a higher resolution of a community structure. Bower's non-*Metazoa*n primers only amplify a fragment of the 18S rDNA gene, whereas primers designed in this study might span over the large fragment of the 18S rDNA gene, both highly variable ITS regions and a small 5.8S rDNA and capture the beginning of the 28S rDNA gene (Figure 16). Considering the different mutation speeds of ribosomal genes and internal spacers, such fragments should resolve communities' taxonomic structure on several levels down to subspecies and strains (Pawlowski *et al.* 2012). Time constraints prevented us from testing the newly designed primers *in vitro*, although *in silico* validation showed the high discriminating potential of developed primers.

The initial aim of the chapter was to develop a method to avoid DNA of arthropod hosts in samples, although the design of the primers was not limited to *Arthropoda,* and the primers should be applicable to studies of eukaryotic symbionts in all metazoan hosts.

6.3. Revealing symbiotic bacteria and bacterial insertions in arthropod genomes

Lateral gene transfer in eukaryotes is considered a rare event (J. O. Andersson 2005). However, with the accumulation of genomic data, the scientific community has now an opportunity to look for LGT events and confirm that it is more widespread and important evolutionary process than previously thought.

Chapter III was dedicated to finding the computational approach to confidently distinguish between living bacterial symbionts and bacterial genomes or fragments incorporated into the host nuclear genome. There are many cases of such insertions described in the literature, and most of such investigations required complex experiments such as in situ hybridisation to establish the location of a questionable sequence (Brelsfoard *et al.* 2014). An accurate method to reveal the origin of bacterial sequences should accelerate analyses of bacterial insertions into host genomes and provide insights on the frequency of such insertions, patterns of the process, and the evolutionary fate of inserted bacterial fragments.

Genomes of two vectors, *G. morsitans* and *A. variegatum*, were assembled and analysed from the perspective of distinguishing between symbiotic bacteria and bacterial insertions. Several bioinformatic approaches were employed, and even though none of the used methods can be used as a definitive answer individually, the collection of the arguments, which all point in one direction, allowed to confirm or reject hypotheses about bacterial insertions in arthropod genomes.

The outstanding question is the evolutionary origin of the transferred *Wolbachia* sequences in the *G. morsitans* genome, as it is not similar to the previously described by Brelsfoard *et al.* (Brelsfoard *et al.* 2014). It is yet to explore whether the two cases of *Wolbachia* insertions were independent evolutionary events or the case of a rapid degeneration of incorporated sequences in closed laboratory population of tsetse flies.

It is accepted fact that transfer of genes from prokaryotes is mostly possible if they reside in germ lines (J. O. Andersson 2005). *Rickettsia* are often found in ovaries of *Arthropoda* (Kurtti *et al.* 2015; Harris *et al.* 2018; Dally *et al.* 2020; Thongprem *et al.* 2020; Al-Khafaji *et al.* 2020) and transovarial inheritance is described of these symbionts (Burgdorfer 1963; Thongprem *et al.* 2020), which makes their genetic material as susceptible for LGT as *Wolbachia*.

Given the integrity of the rickettsial insertion in the *A. variegatum* genome, the comparative recency of the integration could be assumed. The lack of pseudogenisation supported this hypothesis. No other rickettsial insertions are described in literature to date, so there was no possibility to compare the insertion with similar systems. The important question which remains unanswered within the scope of this study is whether the natural populations of *A. variegatum* ticks have the same insertion or it is the artefact of the cultivation of the cell line. If further research reveals that the rickettsial insertion occurred solely in the cell line rather than in the natural tick population, it could present a useful in vitro model for investigating the molecular mechanisms involved in genomic transfers.

6.4. Transcriptomics of tick cells lines: main implications and future work

The knowledge about the tick immune system is still patchy, and many pathways and genes which are theoretically involved in the development of immune response are inferred from the better studied model organisms such as *Drosophila melanogaster* or *Caenorhabditis elegans*. Given a huge evolutionary distance between

ticks and insects (evolutionary paths of *Chelicerata* and *Mandibulata* parted at least 500 Mya ago (Giribet and Edgecombe 2019)), different lifestyle and feeding strategies such approach is not always feasible and independent studies of tick immunity are needed. For many years omics studies on ticks were limited due to the lack of genomic reference data; the situation has improved in the recent decade with the advent of more affordable and efficient sequencing technologies. Now at least nine genomic assemblies are available in public databases.

This transcriptomic experiment aimed to gain a bird's-eye view of the immune responses of two species of hard ticks and their interactions with several pathogenic bacteria.

6.4.1. Revealing cell types tick lines

Several studies attempted to link tick cell lines to particular tissues. So Alberdi *et al.* claimed that *I. ricinus* IRE/CTVM20 cell line has similarities with cells from midgut and *I. scapularis* ISE6 cell line resemble haemocytes (P. Alberdi *et al.* 2016). At the same time Oliver *et al.* found that ISE6 has neuron-like phenotype (Oliver *et al.* 2015). These claims are made on a limited of genes and cannot be used as a basis for assignment cell lines to a particular cell type, these can only hint towards the tissue or organ of origin of cells (Oliver *et al.* 2015).

It is also important to mention that tick cell lines are comprised of cells with different phenotypes: they might have very different size and shape, react differently to pathogens (e.g., phagocytosing pathogens or not, different paces of infection spreading between cells), which might indicate the presence of different cell types in cell lines. Bulk transcriptomic experiments can only capture the generalised transcriptomic response of the set of cells; thus, a single cell approach was chosen to get a higher resolution of possible transcriptomic responses.

Our results didn't reveal distinct cell types in *I. scapularis* IDE8 and *R. microplus* BME/CTVM23 cell lines. Most of the cell-to-cell variations observed in the

single-cell data could be explained by the phases of the cell cycle and viability of cells. The fact that no cell types could be observed in the transcriptomic data does not mean that there are no cell types, although the used approach clearly showed itself unsuitable for the purpose. Possible reason for the lack of defined transcriptional cell clusters could be that strong stress impacts the cells during the filtering and sorting process, and the non-specific stress response might mask less prominent signals such as cell types differences. Another possible explanation might be that the various phenotypes of cells are not based on their type but depend on a current state of a cell: growing phase or interphase, whether a cell is attached to the tube surface or free floating. The results of the described experiment do not allow to make definitive conclusions on this matter, and further investigations are necessary.

This issue can be resolved in several different ways. The most straightforward and definite approach to reveal the cell type composition in the cell lines is to make a single-cell atlas of the whole tick and characterise tick tissues composition. With the help of obtained cell type signatures, it will be easy to determine which cell types comprise different cell lines using label transfer (Stuart *et al.* 2019). This approach has several drawbacks. First of all, creating a single-cell atlas of a whole animal, even a small one, is a complex project requiring careful planning and vast resources. Only model organisms such as *Homo sapiens* (Rozenblatt-Rosen *et al.* 2017), *Mus musculus* (Schaum *et al.* 2017), *D. melanogaster* (Hongjie Li *et al.* 2021) and *C. elegans* (Cao *et al.* 2017) are sequenced at the single-cell level with high resolution and good characterisation of sequenced organs and tissues.

Another possible problem that might arise when inferring transcriptomic labels from tick tissues to cell cultures is that the cells in cultures might not belong to a particular cell type. Cell cultures can be immortalised and not resemble any existing cell type of an animal and express unique gene patterns not found in any cell type *in vivo* (Carter and Shieh 2010). Such a scenario is entirely possible given that some tick cell lines tend to change their chromosome number (Bell-Sakyi *et al.* 2007), which is a

common feature of immortalised cell lines (Carter and Shieh 2010). This might explain the disagreement between characterisation of ISE6 cell line as haemocytes or neuron-like cells by two groups of authors – the cells might be capable of demonstrating both signatures based on the conditions and methods used.

An indirect approach to link transcriptomic profiles of cells with cell types and tissues is label transfer using existing cell atlases of related species. The organisms which are well characterised on a single-cell level and are closest to ticks in the evolutionary sense are *D. melanogaster* and *C. elegans*. Despite the evolutionary distance between ticks, insects and worms, this approach might give insights into basic cell types shared among many animal species.

6.4.2. Tick-bacteria interactions. Next steps to decoding tick immune response

Three bacteria species and one virus were used to trigger immune response in tick cells. Based on the previous works, different responses were expected to Gram-negative and Gram-positive bacteria (Söderhäll 2010), although the experiment results did not reveal the expected Toll and IMD pathways, and the differences between bacteria were not pronounced. Only some elements of previously described reactions to pathogens were noticed, such as regulation of cell apoptosis (Ayllón *et al.* 2015).

The transcriptomic response to *R. raoultii* and *Spiroplasma sp.* was very similar: both bacteria did not trigger a strong reaction, and only a few genes were differentially expressed between infected and uninfected samples. *E. minasensis* showed a stronger effect on tick cells of both species, although this response could not be related to any known immune pathways. The strongest signal came from genes of cell cycle and stress response, which might be accounted either to pathogen invasion or cell culture handling during the experiment stress. Many differentially expressed genes were of a very broad function, and it could not be unambiguously interpreted

Due to a limited selective pressure, many cell lines are known to be deficient in some regulatory or metabolic pathways. For example, Vero cells derived from African green monkeys are interferon-deficient, which makes them unable to respond to viral infections (Desmyter, Melnick, and Rawls 1968); C6/36 mosquito cell line is shown to lack RNA interference response, which makes the cells much more susceptible to some viruses (Brackney *et al.* 2010). Some authors suspect the lack of IMD pathway in triatomine bug *Rhodnius prolixus*, which suggest that this pathway can be reorganised or lost (Mesquita *et al.* 2015). A similar lack of some components of IMD pathway is observed in pea aphids *Acyrthosiphon pisum* (Gerardo *et al.* 2010) and, arguably, *I. scapularis* (Sonenshine and Hynes 2008). This fact might serve as another possible explanation why the cell lines lacked the canonical immune response in the presence of Gram-negative *Spiroplasma sp.* While this hypothesis is definitely worth further investigation, it also raises several important questions. Firstly, it was shown for *D. melanogaster* that Toll and IMD pathways work synergistically and are able to trigger the same downstream immune components (Tanji *et al.* 2007), so theoretically, the IMD deficient organism should retain some ability to resist pathogens via an alternative pathway. Secondly, only one of the used bacteria was Gram-negative, the other two – *R. raoultii* and *E. minasensis* – should have stimulated the immune response via the Toll pathway; however, there was no detectable upregulation of Toll genes.

One of the reasons for choosing used tick cell lines and bacterial symbionts (section 4.4.1) was that these bacteria could successfully infect chosen cell cultures and grow in these cells. This feature is invaluable for studying bacteria in laboratory conditions as it can provide researchers with affordable and easy access to large amounts of material. However, such easy maintenance of symbionts in the cell lines might mean the lack of immune response of tick cells to bacterial infection, which allows pathogens to thrive. Obtaining the transcriptomic response of a cell line that can avoid bacterial invasion might provide valuable insight into the mechanisms of tick immune response.

During the planning of the experiment, the chain infection was observed in cell lines: a small proportion of tick cells became infected after mixing with bacteria, the infection progressed and led to cell death and disintegration, letting the bacterial population into the medium where they infect new cells. The hypothesis was that each cell would demonstrate the immune response upon its infection, although the phenomenon of immune priming was not taken into account. It was observed in a wide range of invertebrates, including *Insecta* and *Crustacea* (Contreras-Garduño *et al.* 2016), and is quite possible in ticks as well. In insects, this process is activated by a sublethal dose of pathogens or pathogen parts and is regulated by increasing of haemocytes and antimicrobial production (Contreras-Garduño *et al.* 2016). Given that bacterial cell debris is floating in the medium between cells from the first moment of infection, it might have happened that all cells in the tube experienced the same immune response shortly after the infection and the transcriptomic response captured on day 5 already represent the reaction of primed cells. If this guess is correct, then there was no chance to capture Toll or IMD or other quick immune responses as they develop within 24 hours after infection (Lemaitre *et al.* 1995), and the current study used 5-days old infections.

6.4.3. Biases of single-cell experiments

Despite the undeniable advantages of experiments on a single cell level, some unavoidable biases are introduced on cell and library preparation steps. One of the main biases which might have affected this experiment is the preparation of cells for sorting. Transcriptomic response of cells is very quick and might be triggered by even slight physical stress applied to cells; such actions as shaking animals or cells triggers a substantial change in expressed genes (Tessler *et al.* 2020). The used tick cell cultures underwent the following steps before the RNA extraction: shaking of tubes during transportation from Tick Cell Biobank to the Centre for Genomic Research; centrifugation and the following substitute of growing medium with cold PBS; pipetting cells to bring them back to suspension; pushing the suspension through the

strainer; and, finally pushing cells through the 10X sorter. The captured transcriptomic profile represents the state of cells after the listed procedures.

It would have been useful to compare the transcriptomic profile obtained in this single cell experiment with bulk RNA-seq results that do not require such multistep preparation. Such comparison might reveal the part of the response which comes from stress and can be subtracted from the results.

6.5. Pros and cons of using cell cultures in vector research

One of the purposes of cell lines is to model vector-pathogen interactions providing insights into biology of such interactions which later might be used for alternating the symbiotic system for human benefit. For any model, it is crucial to understand the similarities and differences between the model and living organism to know the scope of the model application. The model comprised of a tick cell culture and a single pathogen is far from the real-life situation in a lot of aspects such as cell type composition, lack of microenvironmental conditions, lack of the microbiota community, although using such reductionist models for studying basics of host-pathogen interactions is art-recognised approach (Barrila *et al.* 2018).

The large part of this study employed cell cultures of ticks (see sections 3.2.1.2, 4.2.1, 5.2.1) and it is worth discussing whether living ticks can be safely substituted by cheaper and more practical cell cultures. Despite its limitations such as the lack of biological context and different microenvironmental conditions the approach helped to advance the understanding of mechanisms that underlie infection and disease (Al-Rofaai and Bell-Sakyi 2020; Lesley Bell-Sakyi *et al.* 2007). The advantages of using cell lines in tick research are described in section 4.1.1.

The results of studies performed on tick cell lines have a limited physiological relevance and it is necessary to confirm findings using *in vivo* models. First of all, cell

cultures lack the complexity of *in vivo* systems, and their responses to vectors may differ from those observed in living organisms. Secondly, continuous culturing of cells can lead to genetic drift, resulting in changes in the characteristics of the cell line (Noronha *et al.* 2020; Lesley Bell-Sakyi *et al.* 2007), which may alter the results of studies. Cell cultures cannot fully replicate the complexity of *in vivo* systems; in case of tick-pathogen research the interactions between ticks and pathogens might be affected by a mammalian host. One of the recent studies found that a mammalian interferon molecule is required for triggering of the JAK-STAT immune pathway in ticks, these observations were only possible in experiments with living ticks (Rana *et al.* 2023)

In vitro research on host pathogen interaction can be improved in terms of microenvironmental and biophysical conditions by using 3D model systems such as organoids or organs-on-the-chip (Barrila *et al.* 2018). There are some advances in developing 3D models for ticks to produce spheroids and organoids provide advantages in biomedical research that more closely resemble those in tissues (Suderman *et al.* 2021)

6.6. Prospects and possible implications of the data

The described experiments do not limit the scope and the value of the data produced in this study. Generating high-quality assemblies of important vectors can provide valuable information about vector biology.

The first *G. morsitans* genome was sequenced a while ago (International Glossina Genome Initiative *et al.* 2014) and has been used in numerous studies since then. This reference assembly is of undoubtful importance for the studies of *G. morsitans* biology, although it has a couple of limitations. First of all, the assembly is comparatively fragmented due to the limitations of sequencing technologies available at the time of the project. The assembly generated in this study is based on long-read technology and consist of much longer fragments, some of which are comparable with chromosome

sizes (17-24 Mb). This offers the opportunity to map chromosome loci onto nucleotide sequences and analyse possible genomic rearrangements between *G. morsitans* and close species. Secondly, the current reference assembly is based on only female individuals, which does not allow to directly analyse sex determination in tsetse flies. In this study, a male individual was used for sequencing, providing a background for comparing sex regions of the genome. Furthermore, last but not least, only one fly was used for the assembly to minimize polymorphism within the assembly, and the sequencing library was not amplified to reduce possible biases. These factors make the *G. morsitans* assembly generated in this study a valuable resource for further investigation of its biology, features and traits.

There are no sequenced genomes for any of the *Amblyomma* species available in the public databases to date. Here, the assembly of *A. variegatum* was generated; this assembly was also based on long reads, PacBio and ONT technologies. The assembly could not achieve such continuity as in the case of *G. morsitans* due to the large size of the tick genome and a high number of repeats. Despite some imperfections, the completeness of the assembly in terms of coding genes makes it ready to use as a reference assembly for most 'omics' applications.

Two bacterial assemblies were generated for the study of tick immunity response. The assembly of *E. minasensis,* which is currently available from the GenBank, is largely contaminated by tick sequences; thus, the high-quality decontaminated complete assembly was generated. *Spiroplasma sp.* reference was also necessary for the tick transcriptomic experiment, and no public assembly was available for this strain. The PacBio technology was used for the task, and the complete genome in a single chromosome was generated. These two high-quality assemblies can be used for many other projects which involve these bacterial species.

References

Adepoju, Paul. 2019. 'RTS,S Malaria Vaccine Pilots in Three African Countries'. *The Lancet* 393 (10182): 1685. https://doi.org/10.1016/S0140-6736(19)30937-7.

Adl, Sina M., Alastair G. B. Simpson, Christopher E. Lane, Julius Lukeš, David Bass, Samuel S. Bowser, Matthew W. Brown, *et al.* 2012. 'The Revised Classification of Eukaryotes'. *Journal of Eukaryotic Microbiology* 59 (5): 429–514. https://doi.org/10.1111/j.1550-7408.2012.00644.x.

Aguiar, Daniel M., João P. Araujo, Luciano Nakazato, Emilie Bard, and Alejandro Cabezas-Cruz. 2019. 'Complete Genome Sequence of an *Ehrlichia Minasensis* Strain Isolated from Cattle'. Edited by Julie C. Dunning Hotopp. *Microbiology Resource Announcements* 8 (15): e00161-19, /mra/8/15/MRA.00161-19.atom. https://doi.org/10.1128/MRA.00161-19.

Aikawa, Takuya, Hisashi Anbutsu, Naruo Nikoh, Taisei Kikuchi, Fukashi Shibata, and Takema Fukatsu. 2009. 'Longicorn Beetle That Vectors Pinewood Nematode Carries Many *Wolbachia* Genes on an Autosome'. *Proceedings of the Royal Society B: Biological Sciences* 276 (1674): 3791–98. https://doi.org/10.1098/rspb.2009.1022.

Aksoy, Serap. 1995. 'Wigglesworthia Gen. Nov. and Wigglesworthia Glossinidia Sp. Nov., Taxa Consisting of the Mycetocyte-Associated, Primary Endosymbionts of Tsetse Flies'. *International Journal of Systematic and Evolutionary Microbiology* 45 (4): 848–51. https://doi.org/10.1099/00207713-45-4-848.

Aksoy, Serap, Matt Berriman, Neil Hall, Masahira Hattori, Winston Hide, and Michael J. Lehane. 2005. 'A Case for a Glossina Genome Project'. *Trends in Parasitology* 21 (3): 107–11. https://doi.org/10.1016/j.pt.2005.01.006.

Alam, Uzma, Jan Medlock, Corey Brelsfoard, Roshan Pais, Claudia Lohs, Séverine Balmand, Jozef Carnogursky, *et al.* 2011. 'Wolbachia Symbiont Infections Induce Strong Cytoplasmic Incompatibility in the Tsetse Fly Glossina Morsitans'. *PLOS Pathogens* 7 (12): e1002415. https://doi.org/10.1371/journal.ppat.1002415.

Alberdi, M. Pilar, Matthew J. Dalby, Julio Rodriguez-Andres, John K. Fazakerley, Alain Kohl, and Lesley Bell-Sakyi. 2012. 'Detection and Identification of Putative Bacterial Endosymbionts and Endogenous Viruses in Tick Cell Lines'. *Ticks and Tick-Borne Diseases* 3 (3): 137–46. https://doi.org/10.1016/j.ttbdis.2012.05.002.

Alberdi, M. Pilar, Ard M. Nijhof, Frans Jongejan, and Lesley Bell-Sakyi. 2012. 'Tick Cell Culture Isolation and Growth of *Rickettsia raoultii* from Dutch Dermacentor Reticulatus Ticks'. *Ticks and Tick-Borne Diseases* 3 (5–6): 349–54. https://doi.org/10.1016/j.ttbdis.2012.10.020.

Alberdi, Pilar, Karen L. Mansfield, Raúl Manzano-Román, Charlotte Cook, Nieves Ayllón, Margarita Villar, Nicholas Johnson, Anthony R. Fooks, and José de la Fuente. 2016. 'Tissue-Specific Signatures in the Transcriptional Response to Anaplasma Phagocytophilum Infection of Ixodes Scapularis and Ixodes Ricinus Tick Cell Lines'. *Frontiers in Cellular and Infection Microbiology* 6 (February). https://doi.org/10.3389/fcimb.2016.00020.

Al-Khafaji, Alaa M. 2018. 'Interactions between Pathogenic and Nonpathogenic Rickettsiales and the Tick Host'. University of Liverpool.

Al-Khafaji, Alaa M., Stuart D. Armstrong, Ilaria Varotto Boccazzi, Stefano Gaiarsa, Amit Sinha, Zhiru Li, Davide Sassera, Clotilde K. S. Carlow, Sara Epis, and Benjamin L. Makepeace. 2020. 'Rickettsia Buchneri, Symbiont of the Deer Tick Ixodes Scapularis, Can Colonise the Salivary Glands of Its Host'. *Ticks and Tick-Borne Diseases* 11 (1): 101299. https://doi.org/10.1016/j.ttbdis.2019.101299.

Allsopp, B.A. 2015. 'Heartwater – Ehrlichia Ruminantium Infection'. *Revue Scientifique et Technique de l'OIE* 34 (2): 557–68. https://doi.org/10.20506/rst.34.2.2379.

Al-Rofaai, Ahmed, and Lesley Bell-Sakyi. 2020. 'Tick Cell Lines in Research on Tick Control'. *Frontiers in Physiology* 11: 152. https://doi.org/10.3389/fphys.2020.00152.

Anbutsu, H., and T. Fukatsu. 2010. 'Evasion, Suppression and Tolerance of Drosophila Innate Immunity by a Male-Killing Spiroplasma Endosymbiont'. *Insect Molecular Biology* 19 (4): 481–88. https://doi.org/10.1111/j.1365-2583.2010.01008.x.

Andersen, Kristian G., Andrew Rambaut, W. Ian Lipkin, Edward C. Holmes, and Robert F. Garry. 2020. 'The Proximal Origin of SARS-CoV-2'. *Nature Medicine* 26: 450–55. https://doi.org/10.1038/s41591-020-0820-9.

Andersson, J. O. 2005. 'Lateral Gene Transfer in Eukaryotes'. *Cellular and Molecular Life Sciences CMLS* 62 (11): 1182–97. https://doi.org/10.1007/s00018-005-4539-z.

Andersson, Siv G. E., Alireza Zomorodipour, Jan O. Andersson, Thomas Sicheritz-Pontén, U. Cecilia M. Alsmark, Raf M. Podowski, A. Kristina Näslund, Ann-Sofie Eriksson, Herbert H. Winkler, and Charles G. Kurland. 1998. 'The

Genome Sequence of Rickettsia Prowazekii and the Origin of Mitochondria'. *Nature* 396 (6707): 133–40. https://doi.org/10.1038/24094.

Angleró-Rodríguez, Yesseinia I, Octavio AC Talyuli, Benjamin J Blumberg, Seokyoung Kang, Celia Demby, Alicia Shields, Jenny Carlson, Natapong Jupatanakul, and George Dimopoulos. 2017. 'An Aedes Aegypti-Associated Fungus Increases Susceptibility to Dengue Virus by Modulating Gut Trypsin Activity'. Edited by Dominique Soldati-Favre. *ELife* 6 (December): e28844. https://doi.org/10.7554/eLife.28844.

Ansari, M. Shafiq, Maher Ahmed Moraiet, and Salman Ahmad. 2014. 'Insecticides: Impact on the Environment and Human Health'. In *Environmental Deterioration and Human Health*, edited by Abdul Malik, Elisabeth Grohmann, and Rais Akhtar, 99–123. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7890-0_6.

Archibald, John M. 2015. 'Endosymbiosis and Eukaryotic Cell Evolution'. *Current Biology* 25 (19): R911–21. https://doi.org/10.1016/j.cub.2015.07.055.

Attoui, Houssam, Julie M. Stirling, Ulrike G. Munderloh, Frédérique Billoir, Sharon M. Brookes, J. Nicholas Burroughs, Philippe de Micco, Peter P. C. Mertens, and Xavier de Lamballerie. 2001. 'Complete Sequence Characterization of the Genome of the St Croix River Virus, a New Orbivirus Isolated from Cells of Ixodes ScapularisThe GenBank Accession Numbers of the Sequences Reported in This Paper Are AF133431, AF133432 and AF145400–AF145407.' *Journal of General Virology,* 82 (4): 795–804. https://doi.org/10.1099/0022-1317-82-4-795.

Autheman, Delphine, Cécile Crosnier, Simon Clare, David A. Goulding, Cordelia Brandt, Katherine Harcourt, Charlotte Tolley, *et al.* 2021. 'An Invariant Trypanosoma Vivax Vaccine Antigen Induces Protective Immunity'. *Nature* 595 (7865): 96–100. https://doi.org/10.1038/s41586-021-03597-x.

Ayllon, Nieves, Margarita Villar, Ruth C. Galindo, Katherine M. Kocan, Radek Šíma, Juan A. López, Jesús Vázquez, *et al.* 2015. 'Systems Biology of Tissue-Specific Response to Anaplasma Phagocytophilum Reveals Differentiated Apoptosis in the Tick Vector Ixodes Scapularis'. Edited by Man-Wah Tan. *PLOS Genetics* 11 (3): e1005120. https://doi.org/10.1371/journal.pgen.1005120.

Bae, Yun Soo, Myoung Kwon Choi, and Won-Jae Lee. 2010. 'Dual Oxidase in Mucosal Immunity and Host–Microbe Homeostasis'. *Trends in Immunology* 31 (7): 278–87. https://doi.org/10.1016/j.it.2010.05.003.

Baldauf, Sandra L. 2003. 'Phylogeny for the Faint of Heart: A Tutorial'. *Trends in Genetics* 19 (6): 345–51. https://doi.org/10.1016/S0168-9525(03)00112-4.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, *et al.* 2012. 'SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing'. *Journal of Computational Biology* 19 (5): 455–77. https://doi.org/10.1089/cmb.2012.0021.

Barh, Debmalya, Bruno Silva Andrade, Sandeep Tiwari, Marta Giovanetti, Aristóteles Góes-Neto, Luiz Carlos Junior Alcantara, Vasco Azevedo, and Preetam Ghosh. 2020. 'Natural Selection versus Creation: A Review on the Origin of SARS-COV-2'. *Le Infezioni in Medicina* 28 (3): 302–11.

Barre, N, and G Uilenberg. 2010. 'Spread of Parasites Transported with Their Hosts: Case Study of Two Species of Cattle Tick'. *Scientific and Technical Review of the Office International Des Epizooties* 29 (1): 12.

Barrero, Roberto A., Felix D. Guerrero, Michael Black, John McCooke, Brett Chapman, Faye Schilkey, Adalberto A. Pérez de León, *et al.* 2017. 'Gene-Enriched Draft Genome of the Cattle Tick *Rhipicephalus microplus*: Assembly by the Hybrid Pacific Biosciences/Illumina Approach Enabled Analysis of the Highly Repetitive Genome'. *International Journal for Parasitology* 47 (9): 569–83. https://doi.org/10.1016/j.ijpara.2017.03.007.

Barrett, Alan D. T. 2017. 'Yellow Fever Live Attenuated Vaccine: A Very Successful Live Attenuated Vaccine but Still We Have Problems Controlling the Disease'. *Vaccine* 35 (44): 5951–55. https://doi.org/10.1016/j.vaccine.2017.03.032.

Barrila, Jennifer, Aurélie Crabbé, Jiseon Yang, Karla Franco, Seth D. Nydam, Rebecca J. Forsyth, Richard R. Davis, *et al.* 2018. 'Modeling Host-Pathogen Interactions in the Context of the Microenvironment: Three-Dimensional Cell Culture Comes of Age'. Edited by Anthony T. Maurelli. *Infection and Immunity* 86 (11): e00282-18. https://doi.org/10.1128/IAI.00282-18.

Bary, Anton de. 1879. *Die Erscheinung Der Symbiose: Vortrag Gehalten Auf Der Versammlung Deutscher Naturforscher Und Aerzte Zu Cassel*. Trübner. https://books.google.co.uk/books?id=7oQ_AAAAYAAJ.

Bass, David, and Javier del Campo. 2020. 'Microeukaryotes in Animal and Plant Microbiomes: Ecologies of Disease?' *European Journal of Protistology* 76 (October): 125719. https://doi.org/10.1016/j.ejop.2020.125719.

Bass, David, Kevin W. Christison, Grant D. Stentiford, Lauren S.J. Cook, and Hanna Hartikainen. 2023. 'Environmental DNA/RNA for Pathogen and Parasite Detection, Surveillance, and Ecology'. *Trends in Parasitology*, February, S1471492222003154. https://doi.org/10.1016/j.pt.2022.12.010.

Baylis, Matthew. 2017. 'Potential Impact of Climate Change on Emerging Vector-Borne and Other Infections in the UK'. *Environmental Health* 16 (1): 112. https://doi.org/10.1186/s12940-017-0326-1.

Belda, Eugeni, Boubacar Coulibaly, Abdrahamane Fofana, Abdoul H. Beavogui, Sekou F. Traore, Daryl M. Gohl, Kenneth D. Vernick, and Michelle M. Riehle. 2017. 'Preferential Suppression of *Anopheles Gambiae* Host Sequences Allows Detection of the Mosquito Eukaryotic Microbiome'. *Scientific Reports* 7 (1): 3241. https://doi.org/10.1038/s41598-017-03487-1.

Bellgard, Matthew I., Paula M. Moolhuijzen, Felix D. Guerrero, David Schibeci, Manuel Rodriguez-Valle, Daniel G. Peterson, Scot E. Dowd, *et al.* 2012. 'CattleTickBase: An Integrated Internet-Based Bioinformatics Resource for Rhipicephalus (Boophilus) Microplus'. *International Journal for Parasitology* 42 (2): 161–69. https://doi.org/10.1016/j.ijpara.2011.11.006.

Bell-Sakyi, L. 1991. 'Continuous Cell Lines from the Tick Hyalomma Anatolicum Anatolicum'. *The Journal of Parasitology* 77 (6): 1006–8.

Bell-Sakyi, L. 2004. 'Ehrlichia Ruminantium Grows in Cell Lines from Four Ixodid Tick Genera'. *Journal of Comparative Pathology* 130 (4): 285–93. https://doi.org/10.1016/j.jcpa.2003.12.002.

Bell-Sakyi, Lesley, and Houssam Attoui. 2013. 'Endogenous Tick Viruses and Modulation of Tick-Borne Pathogen Growth'. *Frontiers in Cellular and Infection Microbiology* 3: 25. https://doi.org/10.3389/fcimb.2013.00025.

———. 2016. 'Article Commentary: Virus Discovery Using Tick Cell Lines'. *Evolutionary Bioinformatics* 12s2 (January): EBO.S39675. https://doi.org/10.4137/EBO.S39675.

Bell-Sakyi, Lesley, Alistair Darby, Matthew Baylis, and Benjamin L. Makepeace. 2018. 'The Tick Cell Biobank: A Global Resource for in Vitro Research on Ticks, Other Arthropods and the Pathogens They Transmit'. *Ticks and Tick-Borne Diseases* 9 (5): 1364–71. https://doi.org/10.1016/j.ttbdis.2018.05.015.

Bell-Sakyi, Lesley, Alain Kohl, Dennis A. Bente, and John K. Fazakerley. 2012. 'Tick Cell Lines for Study of Crimean-Congo Hemorrhagic Fever Virus and Other Arboviruses'. *Vector-Borne and Zoonotic Diseases* 12 (9): 769–81. https://doi.org/10.1089/vbz.2011.0766.

Bell-Sakyi, Lesley, Erich Zweygarth, Edmour F. Blouin, Ernest A. Gould, and Frans Jongejan. 2007. 'Tick Cell Lines: Tools for Tick and Tick-Borne Disease Research'. *Trends in Parasitology* 23 (9): 450–57. https://doi.org/10.1016/j.pt.2007.07.009.

Binetruy, Florian, Xavier Bailly, Christine Chevillon, Oliver Y. Martin, Marco V. Bernasconi, and Olivier Duron. 2019. 'Phylogenetics of the Spiroplasma Ixodetis Endosymbiont Reveals Past Transfers between Ticks and Other Arthropods'. *Ticks and Tick-Borne Diseases* 10 (3): 575–84. https://doi.org/10.1016/j.ttbdis.2019.02.001.

Blaxter, Mark. 2007. 'Symbiont Genes in Host Genomes: Fragments with a Future?' *Cell Host & Microbe* 2 (4): 211–13. https://doi.org/10.1016/j.chom.2007.09.008.

Bloom, Jesse D., Yujia Alina Chan, Ralph S. Baric, Pamela J. Bjorkman, Sarah Cobey, Benjamin E. Deverman, David N. Fisman, *et al.* 2021. 'Investigate the Origins of COVID-19'. Edited by Jennifer Sills. *Science* 372 (6543): 694.1-694. https://doi.org/10.1126/science.abj0016.

Boothby, Thomas C., Jennifer R. Tenlen, Frank W. Smith, Jeremy R. Wang, Kiera A. Patanella, Erin Osborne Nishimura, Sophia C. Tintori, *et al.* 2015. 'Evidence for Extensive Horizontal Gene Transfer from the Draft Genome of a Tardigrade'. *Proceedings of the National Academy of Sciences* 112 (52): 15976–81. https://doi.org/10.1073/pnas.1510461112.

Bordenstein, Sarah R., and Seth R. Bordenstein. 2011. 'Temperature Affects the Tripartite Interactions between Bacteriophage WO, Wolbachia, and Cytoplasmic Incompatibility'. *PLOS ONE* 6 (12): e29106. https://doi.org/10.1371/journal.pone.0029106.

Bordenstein, Seth R., Michelle L. Marshall, Adam J. Fry, Ulandt Kim, and Jennifer J. Wernegreen. 2006. 'The Tripartite Associations between Bacteriophage, Wolbachia, and Arthropods'. *PLOS Pathogens* 2 (5): e43. https://doi.org/10.1371/journal.ppat.0020043.

Borovickova, Barbara, and Vaclav Hypsa. 2005. 'Ontogeny of Tick Hemocytes: A Comparative Analysis of Ixodes Ricinus and Ornithodoros Moubata'. *Experimental and Applied Acarology* 35 (4): 317–33. https://doi.org/10.1007/s10493-004-2209-8.

Bouyer, Jérémy, Fanny Bouyer, Meritxell Donadeu, Tim Rowan, and Grant Napier. 2013. 'Community- and Farmer-Based Management of Animal African Trypanosomosis in Cattle'. *Trends in Parasitology* 29 (11): 519–22. https://doi.org/10.1016/j.pt.2013.08.003.

Bower, Susan M., Ryan B. Carnegie, Benjamin Goh, Simon R. M. Jones, Geoffrey J. Lowe, and Michelle W. S. Mak. 2004. 'Preferential PCR Amplification of Parasitic Protistan Small Subunit RDNA from Metazoan Tissues'. *The Journal of Eukaryotic Microbiology* 51 (3): 325–32. https://doi.org/10.1111/j.1550-7408.2004.tb00574.x.

Brackney, Doug E., Jaclyn C. Scott, Fumihiko Sagawa, Jimmy E. Woodward, Neil A. Miller, Faye D. Schilkey, Joann Mudge, *et al.* 2010. 'C6/36 Aedes Albopictus Cells Have a Dysfunctional Antiviral RNA Interference Response'. *PLOS Neglected Tropical Diseases* 4 (10): e856. https://doi.org/10.1371/journal.pntd.0000856.

Bradley, Ian M., Ameet J. Pinto, and Jeremy S. Guest. 2016. 'Design and Evaluation of Illumina MiSeq-Compatible, 18S RRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities'. *Applied and Environmental Microbiology* 82 (19): 5878–91. https://doi.org/10.1128/AEM.01630-16.

Brady, Oliver J., H. Charles J. Godfray, Andrew J. Tatem, Peter W. Gething, Justin M. Cohen, F. Ellis McKenzie, T. Alex Perkins, *et al.* 2016. 'Vectorial Capacity and Vector Control: Reconsidering Sensitivity to Parameters for Malaria Elimination'. *Transactions of The Royal Society of Tropical Medicine and Hygiene* 110 (2): 107–17. https://doi.org/10.1093/trstmh/trv113.

Brelsfoard, Corey, George Tsiamis, Marco Falchetto, Ludvik M. Gomulski, Erich Telleria, Uzma Alam, Vangelis Doudoumis, *et al.* 2014. 'Presence of Extensive Wolbachia Symbiont Insertions Discovered in the Genome of Its Host Glossina Morsitans Morsitans'. Edited by Jesus G. Valenzuela. *PLoS Neglected Tropical Diseases* 8 (4): e2728. https://doi.org/10.1371/journal.pntd.0002728.

Brites-Neto, José, Keila Maria Roncato Duarte, and Thiago Fernandes Martins. 2015. 'Tick-Borne Infections in Human and Animal Population Worldwide'. *Veterinary World* 8 (3): 301–15. https://doi.org/10.14202/vetworld.2015.301-315.

Brooks, Daniel R., and Eric P. Hoberg. 2007. 'How Will Global Climate Change Affect Parasite–Host Assemblages?' *Trends in Parasitology* 23 (12): 571–74. https://doi.org/10.1016/j.pt.2007.08.016.

Browne, Annie J., Carlos A. Guerra, Renato Vieira Alves, Veruska Maia da Costa, Anne L. Wilson, David M. Pigott, Simon I. Hay, Steve W. Lindsay, Nick Golding, and Catherine L. Moyes. 2017. 'The Contemporary Distribution of *Trypanosoma Cruzi* Infection in Humans, Alternative Hosts and Vectors'. *Scientific Data* 4 (1): 170050. https://doi.org/10.1038/sdata.2017.50.

Bru, D., F. Martin-Laurent, and L. Philippot. 2008. 'Quantification of the Detrimental Effect of a Single Primer-Template Mismatch by Real-Time PCR Using the 16S RRNA Gene as an Example'. *Applied and Environmental Microbiology* 74 (5): 1660–63. https://doi.org/10.1128/AEM.02403-07.

Brugère, Jean-François, Emmanuel Cornillot, Guy Méténier, Aaron Bensimon, and Christian P. Vivarès. 2000. '*Encephalitozoon Cuniculi* (Microspora) Genome:

Physical Map and Evidence for Telomere-Associated RDNA Units on All Chromosomes'. *Nucleic Acids Research* 28 (10): 2026–33.

Bruto, Maxime, Claire Prigent-Combaret, Patricia Luis, Yvan Moënne-Loccoz, and Daniel Muller. 2014. 'Frequent, Independent Transfers of a Catabolic Gene from Bacteria to Contrasted Filamentous Eukaryotes'. *Proceedings of the Royal Society B: Biological Sciences* 281 (1789): 20140848. https://doi.org/10.1098/rspb.2014.0848.

Buchon, Nicolas, Nichole A. Broderick, and Bruno Lemaitre. 2013. 'Gut Homeostasis in a Microbial World: Insights from Drosophila Melanogaster'. *Nature Reviews Microbiology* 11 (9): 615–26. https://doi.org/10.1038/nrmicro3074.

Buresova, Veronika, Ondrej Hajdusek, Zdenek Franta, Daniel Sojka, and Petr Kopacek. 2009. 'IrAM—An A2-Macroglobulin from the Hard Tick Ixodes Ricinus: Characterization and Function in Phagocytosis of a Potential Pathogen Chryseobacterium Indologenes'. *Developmental & Comparative Immunology* 33 (4): 489–98. https://doi.org/10.1016/j.dci.2008.09.011.

Burgdorfer, Willy. 1963. 'Investigation of "Transovarial Transmission" of Rickettsia Rickettsii in the Wood Tick, Dermacentor Andersoni'. *Experimental Parasitology* 14 (2): 152–59. https://doi.org/10.1016/0014-4894(63)90019-5.

Burger, Thomas D., Renfu Shao, and Stephen C. Barker. 2014. 'Phylogenetic Analysis of Mitochondrial Genome Sequences Indicates That the Cattle Tick, Rhipicephalus (Boophilus) Microplus, Contains a Cryptic Species'. *Molecular Phylogenetics and Evolution* 76 (July): 241–53. https://doi.org/10.1016/j.ympev.2014.03.017.

Burki, F. 2014. 'The Eukaryotic Tree of Life from a Global Phylogenomic Perspective'. *Cold Spring Harbor Perspectives in Biology* 6 (5): a016147–a016147. https://doi.org/10.1101/cshperspect.a016147.

Burki, Fabien, Andrew J. Roger, Matthew W. Brown, and Alastair G. B. Simpson. 2020. 'The New Tree of Eukaryotes'. *Trends in Ecology & Evolution* 35 (1): 43–55. https://doi.org/10.1016/j.tree.2019.08.008.

Bushnell, Brian. 2014. 'BBMap: A Fast, Accurate, Splice-Aware Aligner'. LBNL-7065E. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner.

Butler, Andrew, Paul Hoffman, Peter Smibert, Efthymia Papalexi, and Rahul Satija. 2018. 'Integrating Single-Cell Transcriptomic Data across Different Conditions, Technologies, and Species'. *Nature Biotechnology* 36 (5): 411–20. https://doi.org/10.1038/nbt.4096.

Cabanettes, Floréal, and Christophe Klopp. 2018. 'D-GENIES: Dot Plot Large Genomes in an Interactive, Efficient and Simple Way'. *PeerJ* 6 (June): e4958. https://doi.org/10.7717/peerj.4958.

Cabezas-Cruz, Alejandro, Marie Vancová, Erich Zweygarth, Mucio Flavio Barbosa Ribeiro, Libor Grubhoffer, and Lygia Maria Friche Passos. 2013. 'Ultrastructure of Ehrlichia Mineirensis, a New Member of the Ehrlichia Genus'. *Veterinary Microbiology* 167 (3): 455–58. https://doi.org/10.1016/j.vetmic.2013.08.001.

Cabezas-Cruz, Alejandro, Erich Zweygarth, Marzena Broniszweska, Lygia M. F. Passos, Múcio Flávio Barbosa Ribeiro, Marina Manrique, Raquel Tobes, and José de la Fuente. 2015. 'Complete Genome Sequence of Ehrlichia Mineirensis, a Novel Organism Closely Related to Ehrlichia Canis with a New Host Association'. *Genome Announcements* 3 (1). https://doi.org/10.1128/genomeA.01450-14.

Cabezas-Cruz, Alejandro, Erich Zweygarth, Marie Vancová, Marzena Broniszewska, Libor Grubhoffer, Lygia Maria Friche Passos, Múcio Flávio Barbosa Ribeiro, Pilar Alberdi, and José de la Fuente. 2016. 'Ehrlichia Minasensis Sp. Nov., Isolated from the Tick *Rhipicephalus microplus*'. *International Journal of Systematic and Evolutionary Microbiology,* 66 (3): 1426–30. https://doi.org/10.1099/ijsem.0.000895.

Cahyani, Inswasti. 2021. 'FindingNemo: A Toolkit of CoHex- and Glass Bead-Based Protocols for Ultra-Long Sequencing on ONT Platforms', September. https://doi.org/10.17504/protocols.io.bxwrppd6.

Cali, Ann, James J. Becnel, and Peter M. Takvorian. 2017. 'Microsporidia'. In *Handbook of the Protists*, edited by John M. Archibald, Alastair G.B. Simpson, and Claudio H. Slamovits, 1559–1618. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-28149-0_27.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. 2009. 'BLAST+: Architecture and Applications'. *BMC Bioinformatics* 10 (1): 421. https://doi.org/10.1186/1471-2105-10-421.

Campo, Javier del, David Bass, and Patrick J. Keeling. 2020. 'The Eukaryome: Diversity and Role of Microeukaryotic Organisms Associated with Animal Hosts'. *Functional Ecology* 34 (10): 2045–54. https://doi.org/10.1111/1365-2435.13490.

Campo, Javier del, Maria J. Pons, Maria Herranz, Kevin C. Wakeman, Juana del Valle, Mark J. A. Vermeij, Brian S. Leander, and Patrick J. Keeling. 2019. 'Validation of a Universal Set of Primers to Study Animal-associated

Microeukaryotic Communities'. *Environmental Microbiology*, July, 1462-2920.14733. https://doi.org/10.1111/1462-2920.14733.

Cansado-Utrilla, Cintia, Serena Y. Zhao, Philip J. McCall, Kerri L. Coon, and Grant L. Hughes. 2021. 'The Microbiome and Mosquito Vectorial Capacity: Rich Potential for Discovery and Translation'. *Microbiome* 9 (1): 111. https://doi.org/10.1186/s40168-021-01073-2.

Cao, Junyue, Jonathan S. Packer, Vijay Ramani, Darren A. Cusanovich, Chau Huynh, Riza Daza, Xiaojie Qiu, *et al.* 2017. 'Comprehensive Single-Cell Transcriptional Profiling of a Multicellular Organism'. *Science* 357 (6352): 661–67. https://doi.org/10.1126/science.aam8940.

Carson, Rachel. 1962. *Silent Spring*. Boston: Houghton Mifflin Company.

Carter, Matt, and Jennifer C. Shieh. 2010. 'Chapter 13 - Cell Culture Techniques'. In *Guide to Research Techniques in Neuroscience*, edited by Matt Carter and Jennifer C. Shieh, 281–96. New York: Academic Press. https://doi.org/10.1016/B978-0-12-374849-2.00013-6.

Cascales, Eric, Roland Lloubès, and James N Sturgis. 2008. 'The TolQ-TolR Proteins Energize TolA and Share Homologies with the Flagellar Motor Proteins MotA-MotB: TolQ-TolR Are Needed to Energize TolA'. *Molecular Microbiology* 42 (3): 795–807. https://doi.org/10.1046/j.1365-2958.2001.02673.x.

Castro, Julio J. de. 1997. 'Sustainable Tick and Tickborne Disease Control in Livestock Improvement in Developing Countries'. *Veterinary Parasitology* 71 (2–3): 77–97. https://doi.org/10.1016/S0304-4017(97)00033-2.

Ceraul, Shane M., Sheila M. Dreher-Lesnick, Albert Mulenga, M. Sayeedur Rahman, and Abdu F. Azad. 2008. 'Functional Characterization and Novel Rickettsiostatic Effects of a Kunitz-Type Serine Protease Inhibitor from the Tick Dermacentor Variabilis'. *Infection and Immunity* 76 (11): 5429–35. https://doi.org/10.1128/IAI.00866-08.

Ceraul, Shane M., Daniel E. Sonenshine, and Wayne L. Hynes. 2002. 'Resistance of the Tick <I>Dermacentor Variabilis</I> (Acari: Ixodidae) Following Challenge with the Bacterium <I>Escherichia Coli</I> (Enterobacteriales: Enterobacteriaceae)'. *Journal of Medical Entomology* 39 (2): 376–83. https://doi.org/10.1603/0022-2585-39.2.376.

Cheng, Haoyu, Gregory T. Concepcion, Xiaowen Feng, Haowen Zhang, and Heng Li. 2021. 'Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm'. *Nature Methods* 18 (2): 170–75. https://doi.org/10.1038/s41592-020-01056-5.

Cheng, Vincent C. C., Susanna K. P. Lau, Patrick C. Y. Woo, and Kwok Yung Yuen. 2007. 'Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection'. *Clinical Microbiology Reviews* 20 (4): 660–94. https://doi.org/10.1128/CMR.00023-07.

Chrudimska, Tereza, Tomás Chrudimský, Marina Golovchenko, Nataliia Rudenko, and Libor Grubhoffer. 2010. 'New Defensins from Hard and Soft Ticks: Similarities, Differences, and Phylogenetic Analyses'. *Veterinary Parasitology* 167 (2–4): 298–303. https://doi.org/10.1016/j.vetpar.2009.09.032.

Clark, H. F. 1964. 'Suckling Mouse Cataract Agent'. *Journal of Infectious Diseases* 114 (5): 476–87. https://doi.org/10.1093/infdis/114.5.476.

Coelho, Camila Henriques, Justin Yai Alamou Doritchamou, Irfan Zaidi, and Patrick E. Duffy. 2017. 'Advances in Malaria Vaccine Development: Report from the 2017 Malaria Vaccine Symposium'. *Npj Vaccines* 2 (1): 1–5. https://doi.org/10.1038/s41541-017-0035-3.

Conceicao, Christiano Calixto, Jhenifer Nascimento da Silva, Angélica Arcanjo, Cíntia Lopes Nogueira, Leonardo Araujo de Abreu, Pedro Lagerblad de Oliveira, Katia C. Gondim, *et al.* 2021. 'Aedes Fluviatilis Cell Lines as New Tools to Study Metabolic and Immune Interactions in Mosquito-Wolbachia Symbiosis'. *Scientific Reports* 11 (1): 19202. https://doi.org/10.1038/s41598-021-98738-7.

Contreras-Garduño, Jorge, Humberto Lanz-Mendoza, Bernardo Franco, Adriana Nava, Mario Pedraza-Reyes, and Jorge Canales-Lazcano. 2016. 'Insect Immune Priming: Ecology and Experimental Evidences'. *Ecological Entomology* 41 (4): 351–66. https://doi.org/10.1111/een.12300.

Cradic, Kendall, Jason Wells, Lindsay Allen, Kent Kruckeberg, Ravinder Singh, and Stefan Grebe. 2004. 'Substitution of 3′-Phosphate Cap with a Carbon-Based Blocker Reduces the Possibility of Fluorescence Resonance Energy Transfer Probe Failure in Real-Time PCR Assays'. *Clinical Chemistry* 50 (6): 1080–82. https://doi.org/10.1373/clinchem.2004.033183.

Cramaro, Wibke J., Oliver E. Hunewald, Lesley Bell-Sakyi, and Claude P. Muller. 2017. 'Genome Scaffolding and Annotation for the Pathogen Vector Ixodes Ricinus by Ultra-Long Single Molecule Sequencing'. *Parasites & Vectors* 10 (1): 71. https://doi.org/10.1186/s13071-017-2008-9.

Cramaro, Wibke J., Dominique Revets, Oliver E. Hunewald, Regina Sinner, Anna L. Reye, and Claude P. Muller. 2015. 'Integration of Ixodes Ricinus Genome Sequencing with Transcriptome and Proteome Annotation of the Naïve Midgut'. *BMC Genomics* 16 (1): 871. https://doi.org/10.1186/s12864-015-1981-7.

Crooks, Gavin E., Gary Hon, John-Marc Chandonia, and Steven E. Brenner. 2004. 'WebLogo: A Sequence Logo Generator'. *Genome Research* 14 (6): 1188–90. https://doi.org/10.1101/gr.849004.

Dally, Maria, Maya Lalzar, Eduard Belausov, Yuval Gottlieb, Moshe Coll, and Einat Zchori-Fein. 2020. 'Cellular Localization of Two Rickettsia Symbionts in the Digestive System and within the Ovaries of the Mirid Bug, Macrolophous Pygmaeus'. *Insects* 11 (8): 530. https://doi.org/10.3390/insects11080530.

Dame, D. A., and C. H. Schmidt. 1970. 'The Sterile-Male Technique Against Tsetse Flies, Glossina Spp.' *Bulletin of the Entomological Society of America* 16 (1): 24–30. https://doi.org/10.1093/besa/16.1.24.

De Coster, Wouter, Svenn D'Hert, Darrin T Schultz, Marc Cruts, and Christine Van Broeckhoven. 2018. 'NanoPack: Visualizing and Processing Long-Read Sequencing Data'. *Bioinformatics* 34 (15): 2666–69. https://doi.org/10.1093/bioinformatics/bty149.

Delgado, Beatriz, Magdalena Serrano, Carmen Gonzalez, Alex Bach, and Oscar Gonzalez-Recio. 2019. 'Long Reads from Nanopore Sequencing as a Tool for Animal Microbiome Studies'. Biorxiv. 2019. https://www.biorxiv.org/content/10.1101/2019.12.21.886028v1.

Dennison, Nathan J., Raúl G. Saraiva, Chris M. Cirimotich, Godfree Mlambo, Emmanuel F. Mongodin, and George Dimopoulos. 2016. 'Functional Genomic Analyses of Enterobacter, Anopheles and Plasmodium Reciprocal Interactions That Impact Vector Competence'. *Malaria Journal* 15 (1): 425. https://doi.org/10.1186/s12936-016-1468-2.

Desmyter, Jan, Joseph L. Melnick, and William E. Rawls. 1968. 'Defectiveness of Interferon Production and of Rubella Virus Interference in a Line of African Green Monkey Kidney Cells (Vero)'. *Journal of Virology* 2 (10): 955.

Diuk-Wasser, Maria A., Yuchen Liu, Tanner K. Steeves, Corrine Folsom-O'Keefe, Kenneth R. Dardick, Timothy Lepore, Stephen J. Bent, *et al.* 2014. 'Monitoring Human Babesiosis Emergence through Vector Surveillance New England, USA'. *Emerging Infectious Diseases* 20 (2): 225–31. https://doi.org/10.3201/eid2002.130644.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. 'STAR: Ultrafast Universal RNA-Seq Aligner'. *Bioinformatics (Oxford, England)* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dohm, Juliane C., Claudio Lottaz, Tatiana Borodina, and Heinz Himmelbauer. 2008. 'Substantial Biases in Ultra-Short Read Data Sets from High-Throughput DNA

Sequencing'. *Nucleic Acids Research* 36 (16): e105. https://doi.org/10.1093/nar/gkn425.

Donath, Alexander, Frank Jühling, Marwa Al-Arab, Stephan H Bernhart, Franziska Reinhardt, Peter F Stadler, Martin Middendorf, and Matthias Bernt. 2019. 'Improved Annotation of Protein-Coding Genes Boundaries in Metazoan Mitochondrial Genomes'. *Nucleic Acids Research* 47 (20): 10543–52. https://doi.org/10.1093/nar/gkz833.

Dong, Zhihui, Yicheng Yang, Qian Wang, Songsong Xie, Shanshan Zhao, Wenbo Tan, Wumei Yuan, and Yuanzhi Wang. 2019. 'A Case with Neurological Abnormalities Caused by *Rickettsia raoultii* in Northwestern China'. *BMC Infectious Diseases* 19 (1): 796. https://doi.org/10.1186/s12879-019-4414-4.

Doudoumis, Vangelis, George Tsiamis, Florence Wamwiri, Corey Brelsfoard, Uzma Alam, Emre Aksoy, Stelios Dalaperas, *et al.* 2012. 'Detection and Characterization of Wolbachia Infections in Laboratory and Natural Populations of Different Species of Tsetse Flies (Genus Glossina)'. *BMC Microbiology* 12 (1): S3. https://doi.org/10.1186/1471-2180-12-S1-S3.

Douglas, A. E. 2010. *The Symbiotic Habit*. Princeton, N.J: Princeton University Press.

Drew, Georgia C., Giles E. Budge, Crystal L. Frost, Peter Neumann, Stefanos Siozios, Orlando Yañez, and Gregory D. D. Hurst. 2021. 'Transitions in Symbiosis: Evidence for Environmental Acquisition and Social Transmission within a Clade of Heritable Symbionts'. *The ISME Journal* 15 (10): 2956–68. https://doi.org/10.1038/s41396-021-00977-z.

Dunning Hotopp, Julie C. 2011. 'Horizontal Gene Transfer between Bacteria and Animals'. *Trends in Genetics* 27 (4): 157–63. https://doi.org/10.1016/j.tig.2011.01.005.

Edgar, Robert. 2018. 'Taxonomy Annotation and Guide Tree Errors in 16S RRNA Databases'. *PeerJ* 6 (June): e5030. https://doi.org/10.7717/peerj.5030.

Eggenberger, Lisa R., William J. Lamoreaux, and Lewis B. Coons. 1990. 'Hemocytic Encapsulation of Implants in the TickDermacentor Variabilis'. *Experimental & Applied Acarology* 9 (3–4): 279–87. https://doi.org/10.1007/BF01193434.

Eisen, Jonathan A, Robert S Coyne, Martin Wu, Dongying Wu, Mathangi Thiagarajan, Jennifer R Wortman, Jonathan H Badger, *et al.* 2006. 'Macronuclear Genome Sequence of the Ciliate Tetrahymena Thermophila, a Model Eukaryote'. Edited by Mikhail Gelfand. *PLoS Biology* 4 (9): e286. https://doi.org/10.1371/journal.pbio.0040286.

Eisen, Rebecca J., Lars Eisen, and Charles B. Beard. 2016. 'County-Scale Distribution of *Ixodes Scapularis* and *Ixodes Pacificus* (Acari: Ixodidae) in the Continental United States'. *Journal of Medical Entomology* 53 (2): 349–86. https://doi.org/10.1093/jme/tjv237.

El Karkouri, Khalid, Oleg Mediannikov, Catherine Robert, Didier Raoult, and Pierre-Edouards Fournier. 2016. 'Genome Sequence of the Tick-Borne Pathogen *Rickettsia raoultii*'. *Genome Announcements* 4 (2): e00157-16, /ga/4/2/e00157-16.atom. https://doi.org/10.1128/genomeA.00157-16.

Engelstädter, Jan, and Gregory D.D. Hurst. 2009. 'The Ecology and Evolution of Microbes That Manipulate Host Reproduction'. *Annual Review of Ecology, Evolution, and Systematics* 40 (1): 127–49. https://doi.org/10.1146/annurev.ecolsys.110308.120206.

Esteves, Eliane, Flavio A. Lara, Daniel M. Lorenzini, Gustavo H.N. Costa, Aline H. Fukuzawa, Luis N. Pressinotti, José Roberto M.C. Silva, *et al.* 2008. 'Cellular and Molecular Characterization of an Embryonic Cell Line (BME26) from the Tick Rhipicephalus (Boophilus) Microplus'. *Insect Biochemistry and Molecular Biology* 38 (5): 568–80. https://doi.org/10.1016/j.ibmb.2008.01.006.

Estrada-Peña, A., R. G. Pegram, N. Barré, and José M. Venzal. 2007. 'Using Invaded Range Data to Model the Climate Suitability for *Amblyomma variegatum* (Acari: Ixodidae) in the New World'. *Experimental and Applied Acarology* 41 (3): 203–14. https://doi.org/10.1007/s10493-007-9050-9.

Ewald, P W. 1983. 'Host-Parasite Relations, Vectors, and the Evolution of Disease Severity'. *Annual Review of Ecology and Systematics* 14 (1): 465–85. https://doi.org/10.1146/annurev.es.14.110183.002341.

Farikou, Oumarou, Flobert Njiokou, Jean A. Mbida Mbida, Guy R. Njitchouang, Hugues Nana Djeunga, Tazoacha Asonganyi, Pere P. Simarro, Gérard Cuny, and Anne Geiger. 2010. 'Tripartite Interactions between Tsetse Flies, Sodalis Glossinidius and Trypanosomes—An Epidemiological Approach in Two Historical Human African Trypanosomiasis Foci in Cameroon'. *Infection, Genetics and Evolution* 10 (1): 115–21. https://doi.org/10.1016/j.meegid.2009.10.008.

Ferreira, Alvaro Gil, Shane Fairlie, and Luciano Andrade Moreira. 2020. 'Insect Vectors Endosymbionts as Solutions against Diseases'. *Current Opinion in Insect Science* 40 (August): 56–61. https://doi.org/10.1016/j.cois.2020.05.014.

Fevre, Eric M., Beatrix v. Wissmann, Susan C. Welburn, and Pascal Lutumba. 2008. 'The Burden of Human African Trypanosomiasis'. Edited by Simon Brooker. *PLoS Neglected Tropical Diseases* 2 (12): e333. https://doi.org/10.1371/journal.pntd.0000333.

Flegontov, Pavel, Jan Votýpka, Tomáš Skalický, Maria D. Logacheva, Aleksey A. Penin, Goro Tanifuji, Naoko T. Onodera, *et al.* 2013. 'Paratrypanosoma Is a Novel Early-Branching Trypanosomatid'. *Current Biology* 23 (18): 1787–93. https://doi.org/10.1016/j.cub.2013.07.045.

Fleming, Stephen J., John C. Marioni, and Mehrtash Babadi. 2019. 'CellBender Remove-Background: A Deep Generative Model for Unsupervised Removal of Background Noise from ScRNA-Seq Datasets'. *BioRxiv* 791699 (October). https://doi.org/10.1101/791699.

Fogaca, Andréa C., Igor C. Almeida, Marcos N. Eberlin, Aparecida S. Tanaka, Philippe Bulet, and Sirlei Daffre. 2006. 'Ixodidin, a Novel Antimicrobial Peptide from the Hemocytes of the Cattle Tick Boophilus Microplus with Inhibitory Activity against Serine Proteinases'. *Peptides* 27 (4): 667–74. https://doi.org/10.1016/j.peptides.2005.07.013.

Foldvari, Gábor, Krisztina Rigo, and András Lakos. 2013. 'Transmission of Rickettsia Slovaca and *Rickettsia raoultii* by Male Dermacentor Marginatus and Dermacentor Reticulatus Ticks to Humans'. *Diagnostic Microbiology and Infectious Disease* 76 (3): 387–89. https://doi.org/10.1016/j.diagmicrobio.2013.03.005.

Fournier, Pierre-Edouard, Khalid El Karkouri, Quentin Leroy, Catherine Robert, Bernadette Giumelli, Patricia Renesto, Cristina Socolovschi, Philippe Parola, Stéphane Audic, and Didier Raoult. 2009. 'Analysis of the Rickettsia Africae Genome Reveals That Virulence Acquisition in Rickettsia Species May Be Explained by Genome Reduction'. *BMC Genomics* 10 (1): 166. https://doi.org/10.1186/1471-2164-10-166.

Frainer, André, Brendan G. McKie, Per-Arne Amundsen, Rune Knudsen, and Kevin D. Lafferty. 2018. 'Parasitism and the Biodiversity-Functioning Relationship'. *Trends in Ecology & Evolution* 33 (4): 260–68. https://doi.org/10.1016/j.tree.2018.01.011.

Frolov, Alexander O., Alexei Y. Kostygov, and Vyacheslav Yurchenko. 2021. 'Development of Monoxenous Trypanosomatids and Phytomonads in Insects'. *Trends in Parasitology* 37 (6): 538–51. https://doi.org/10.1016/j.pt.2021.02.004.

Frolov, Alexander O., M.N. Malysheva, and Alexei Yu. Kostygov. 2015. 'Transformations of Life Cycles in the Evolutionary History of Trypanosomatidae. Macrotransformations'. *Parazitologiya* 49 (4). https://www.elibrary.ru/item.asp?id=23944751.

Frutos, Roger, Laurent Gavotte, and Christian A. Devaux. 2021. 'Understanding the Origin of COVID-19 Requires to Change the Paradigm on Zoonotic Emergence

from the Spillover to the Circulation Model'. *Infection, Genetics and Evolution*, March, 104812. https://doi.org/10.1016/j.meegid.2021.104812.

Fry, Michael. 2016. *Landmark Experiments in Molecular Biology*. London, UK ; San Diego, CA, USA: Academic Press is an imprint of Elsevier.

Fuente, Jose de la. 2008. 'Overview: Ticks as Vectors of Pathogens That Cause Disease in Humans and Animals'. *Frontiers in Bioscience* Volume (13): 6938. https://doi.org/10.2741/3200.

Fuente, José de la. 2018. 'Controlling Ticks and Tick-Borne Diseases...looking Forward'. *Ticks and Tick-Borne Diseases* 9 (5): 1354–57. https://doi.org/10.1016/j.ttbdis.2018.04.001.

Fuente, Jose de la, Manuel Rodriguez, and Josec Garcia-Garcia. 2000. 'Immunological Control of Ticks through Vaccination with Boophilus Microplus Gut Antigens'. *Annals of the New York Academy of Sciences* 916 (1): 617–21.

Garcia, Gustavo R., José Marcos Chaves Ribeiro, Sandra Regina Maruyama, Luiz Gustavo Gardinassi, Kristina Nelson, Beatriz R. Ferreira, Thales Galdino Andrade, and Isabel K. Ferreira de Miranda Santos. 2020. 'A Transcriptome and Proteome of the Tick *Rhipicephalus microplus* Shaped by the Genetic Composition of Its Hosts and Developmental Stage'. *Scientific Reports* 10 (1): 12857. https://doi.org/10.1038/s41598-020-69793-3.

Gariou-Papalexiou, A, G Yannopoulos, A Zacharopoulou, and R H Gooding. 2002. 'Photographic Polytene Chromosome Maps for *Glossina Morsitans Submorsitans* (Diptera: Glossinidae): Cytogenetic Analysis of a Colony with Sex-Ratio Distortion'. *Genome* 45 (5): 871–80. https://doi.org/10.1139/g02-057.

Garros, Claire, Jérémy Bouyer, Willem Takken, and Renate C. Smallegange, eds. 2018. *Pests and Vector-Borne Diseases in the Livestock Industry*. Vol. 5. Ecology and Control of Vector-Borne Diseases. The Netherlands: Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-863-6.

Gasparotto, Paulo Henrique Gilio, Carlos Alexandre Fernandes dos Santos, Jerônimo Vieira Dantas Filho, Rauane Cristine Santos Ferraz, Flavio Roberto Chaves da Silva, and Cíntia Daudt. 2020. 'Resistance of Rhipicephalus (Boophilus) Microplus (Canestrini, 1888) to Acaricides Used in Dairy Cattle of Teixeirópolis, Rondônia, Brazil'. *Acta Veterinaria Brasilica* 14 (2): 99–105. https://doi.org/10.21708/avb.2020.14.2.9123.

Gasser, Mark T., Matthew Chung, Robin E. Bromley, Suvarna Nadendla, and Julie C. Dunning Hotopp. 2019. 'Complete Genome Sequence of WAna, the Wolbachia Endosymbiont of Drosophila Ananassae'. Edited by Christina A. Cuomo.

Microbiology Resource Announcements 8 (43): MRA.01136-19, e01136-19. https://doi.org/10.1128/MRA.01136-19.

Ge, Hong, Yao-Yu Eric Chuang, Shuping Zhao, Min Tong, Mong-Hsun Tsai, Joseph J. Temenak, Allen L. Richards, and Wei-Mei Ching. 2004. 'Comparative Genomics of *Rickettsia Prowazekii* Madrid E and Breinl Strains'. *Journal of Bacteriology* 186 (2): 556–65. https://doi.org/10.1128/JB.186.2.556-565.2004.

Geraci, Nicholas S., J. Spencer Johnston, J. Paul Robinson, Stephen K. Wikel, and Catherine A. Hill. 2007. 'Variation in Genome Size of Argasid and Ixodid Ticks'. *Insect Biochemistry and Molecular Biology* 37 (5): 399–408. https://doi.org/10.1016/j.ibmb.2006.12.007.

Gerardo, Nicole M., Boran Altincicek, Caroline Anselme, Hagop Atamian, Seth M. Barribeau, Martin de Vos, Elizabeth J. Duncan, *et al.* 2010. 'Immunity and Other Defenses in Pea Aphids, Acyrthosiphon Pisum'. *Genome Biology* 11 (2): R21. https://doi.org/10.1186/gb-2010-11-2-r21.

Giribet, Gonzalo, and Gregory D. Edgecombe. 2019. 'The Phylogeny and Evolutionary History of Arthropods'. *Current Biology* 29 (12): R592–602. https://doi.org/10.1016/j.cub.2019.04.057.

Global Preparedness Monitoring Board. 2019. 'A World at Risk: Annual Report on Global Preparedness for Health Emergencies.' In . Geneva: World Health Organization.

'Global Vector Control Response 2017–2030'. 2017. In . Geneva: World Health Organization.

Goddard, Jerome. 2008. *Infectious Diseases and Arthropods*. 2nd ed. Infectious Disease. Totowa, N.J: Humana Press.

Grant, Ian F. 2001. 'Insecticides for Tsetse and Trypanosomiasis Control: Is the Environmental Risk Acceptable?', 5.

Green, S. J., and D. Minz. 2005. 'Suicide Polymerase Endonuclease Restriction, a Novel Technique for Enhancing PCR Amplification of Minor DNA Templates'. *Applied and Environmental Microbiology* 71 (8): 4721–27. https://doi.org/10.1128/AEM.71.8.4721-4727.2005.

Grubhoffer, L., V. Kovář, and N. Rudenko. 2004. 'Tick Lectins: Structural and Functional Properties'. *Parasitology* 129 (S1): S113–25. https://doi.org/10.1017/S0031182004004858.

Guan, Dengfeng, Shane A McCarthy, Jonathan Wood, Kerstin Howe, Yadong Wang, and Richard Durbin. 2020. 'Identifying and Removing Haplotypic Duplication

in Primary Genome Assemblies'. *Bioinformatics* 36 (9): 2896–98. https://doi.org/10.1093/bioinformatics/btaa025.

Gubler, Duane J. 2011. 'Dengue, Urbanization and Globalization: The Unholy Trinity of the 21st Century'. *Tropical Medicine and Health* 39 (4SUPPLEMENT): S3–11. https://doi.org/10.2149/tmh.2011-S05.

Guerin, Patrick M., Thomas Krober, Conor McMahon, Pablo Guerenshtein, Stoyan Grenacher, Michele Vlimant, Peter-Allan Diehl, Pascal Steullet, and Zainulabeudin Syed. 2000. 'Chemosensory and Behavioural Adaptations of Ectoparasitic Arthropods'. *Nova Acta Leopoldina* 83 (316): 213–30.

Guerrero, Felix D., Kylie G. Bendele, Noushin Ghaffari, Joseph Guhlin, Kristene R. Gedye, Kevin E. Lawrence, Peter K. Dearden, *et al.* 2019. 'The Pacific Biosciences de Novo Assembled Genome Dataset from a Parthenogenetic New Zealand Wild Population of the Longhorned Tick, Haemaphysalis Longicornis Neumann, 1901'. *Data in Brief* 27 (December): 104602. https://doi.org/10.1016/j.dib.2019.104602.

Guglielmone, Alberto A., Richard G. Robbins, Dmitry A. Apanaskevich, Trevor N. Petney, Agustín Estrada-Peña, Ivan G. Horak, Renfu Shao, and Stephen C. Barker. 2010. 'The Argasidae, Ixodidae and Nuttalliellidae (Acari: Ixodida) of the World: A List of Valid Species Names'. *Zootaxa* 2528 (1): 1. https://doi.org/10.11646/zootaxa.2528.1.1.

Gulia-Nuss, Monika, Andrew B. Nuss, Jason M. Meyer, Daniel E. Sonenshine, R. Michael Roe, Robert M. Waterhouse, David B. Sattelle, *et al.* 2016. 'Genomic Insights into the Ixodes Scapularis Tick Vector of Lyme Disease'. *Nature Communications* 7 (1): 10507. https://doi.org/10.1038/ncomms10507.

Gunderson, J. H., and M. L. Sogin. 1986. 'Length Variation in Eukaryotic RRNAs: Small Subunit RRNAs from the Protists *Acanthamoeba Castellanii* and *Euglena Gracilis*'. *Gene* 44 (1): 63–70. https://doi.org/10.1016/0378-1119(86)90043-0.

Gupta, Lalita, Alvaro Molina-Cruz, Sanjeev Kumar, Janneth Rodrigues, Rajnikant Dixit, Rodolfo E. Zamora, and Carolina Barillas-Mury. 2009. 'The STAT Pathway Mediates Late-Phase Immunity against Plasmodium in the Mosquito Anopheles Gambiae'. *Cell Host & Microbe* 5 (5): 498–507. https://doi.org/10.1016/j.chom.2009.04.003.

Hadziavdic, Kenan, Katrine Lekang, Anders Lanzen, Inge Jonassen, Eric M. Thompson, and Christofer Troedsson. 2014. 'Characterization of the 18S RRNA Gene for Designing Universal Eukaryote Specific Primers'. *PLoS ONE* 9 (2): e87624. https://doi.org/10.1371/journal.pone.0087624.

Hafemeister, Christoph, and Rahul Satija. 2019. 'Normalization and Variance Stabilization of Single-Cell RNA-Seq Data Using Regularized Negative Binomial Regression'. *Genome Biology* 20 (1): 296. https://doi.org/10.1186/s13059-019-1874-1.

Hallmann, Caspar A., Martin Sorg, Eelke Jongejans, Henk Siepel, Nick Hofland, Heinz Schwan, Werner Stenmans, *et al.* 2017. 'More than 75 Percent Decline over 27 Years in Total Flying Insect Biomass in Protected Areas'. *PLOS ONE* 12 (10): e0185809. https://doi.org/10.1371/journal.pone.0185809.

Harris, Emma K., Krit Jirakanwisal, Victoria I. Verhoeve, Chanida Fongsaran, Chanakan Suwanbongkot, Matthew D. Welch, and Kevin R. Macaluso. 2018. 'Role of Sca2 and RickA in the Dissemination of Rickettsia Parkeri in Amblyomma Maculatum'. *Infection and Immunity* 86 (6): e00123-18. https://doi.org/10.1128/IAI.00123-18.

Hebenstreit, Daniel. 2012. 'Methods, Challenges and Potentials of Single Cell RNA-Seq'. *Biology* 1 (3): 658–67. https://doi.org/10.3390/biology1030658.

Hemingway, Janet, Hilary Ranson, Alan Magill, Jan Kolaczinski, Christen Fornadel, John Gimnig, Maureen Coetzee, *et al.* 2016. 'Averting a Malaria Disaster: Will Insecticide Resistance Derail Malaria Control?' *The Lancet* 387 (10029): 1785–88. https://doi.org/10.1016/S0140-6736(15)00417-1.

Herren, Jeremy K., Lilian Mbaisi, Enock Mararo, Edward E. Makhulu, Victor A. Mobegi, Hellen Butungi, Maria Vittoria Mancini, *et al.* 2020. 'A Microsporidian Impairs Plasmodium Falciparum Transmission in Anopheles Arabiensis Mosquitoes'. *Nature Communications* 11 (1): 2187. https://doi.org/10.1038/s41467-020-16121-y.

Hill, Catherine A., and Stephen K. Wikel. 2005. 'The Ixodes Scapularis Genome Project: An Opportunity for Advancing Tick Research'. *Trends in Parasitology* 21 (4): 151–53. https://doi.org/10.1016/j.pt.2005.02.004.

Hiraoka, Satoshi, Ching-chia Yang, and Wataru Iwasaki. 2016. 'Metagenomics and Bioinformatics in Microbial Ecology: Current Status and Beyond'. *Microbes and Environments* 31 (3): 204–12. https://doi.org/10.1264/jsme2.ME16024.

Holmes, Edward C., Stephen A. Goldstein, Angela L. Rasmussen, David L. Robertson, Alexander Crits-Christoph, Joel O. Wertheim, Simon J. Anthony, *et al.* 2021. 'The Origins of SARS-CoV-2: A Critical Review'. *Cell* 0 (0). https://doi.org/10.1016/j.cell.2021.08.017.

Holt, Robert A., G. Mani Subramanian, Aaron Halpern, Granger G. Sutton, Rosane Charlab, Deborah R. Nusskern, Patrick Wincker, *et al.* 2002. 'The Genome

Sequence of the Malaria Mosquito Anopheles Gambiae'. *Science* 298 (5591): 129–49. https://doi.org/10.1126/science.1076181.

Honig Mondekova, Helena, Radek Sima, Veronika Urbanova, Vojtech Kovar, Ryan O. M. Rego, Libor Grubhoffer, Petr Kopacek, and Ondrej Hajdusek. 2017. 'Characterization of Ixodes Ricinus Fibrinogen-Related Proteins (Ixoderins) Discloses Their Function in the Tick Innate Immunity'. *Frontiers in Cellular and Infection Microbiology* 7 (December). https://doi.org/10.3389/fcimb.2017.00509.

Hooper, L. V., D. R. Littman, and A. J. Macpherson. 2012. 'Interactions Between the Microbiota and the Immune System'. *Science* 336 (6086): 1268–73. https://doi.org/10.1126/science.1223490.

Hotopp, J. C. D., M. E. Clark, D. C. S. G. Oliveira, J. M. Foster, P. Fischer, M. C. M. Torres, J. D. Giebel, *et al.* 2007. 'Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes'. *Science* 317 (5845): 1753–56. https://doi.org/10.1126/science.1142490.

Huang, Wei-Fone, Shu-Jen Tsai, Chu-Fang Lo, Yamane Soichi, and Chung-Hsiung Wang. 2004. 'The Novel Organization and Complete Sequence of the Ribosomal RNA Gene of *Nosema Bombycis*'. *Fungal Genetics and Biology* 41 (5): 473–81. https://doi.org/10.1016/j.fgb.2003.12.005.

Huang, Xiaohong, Naotoshi Tsuji, Takeharu Miyoshi, Sachiko Nakamura-Tsuruta, Jun Hirabayashi, and Kozo Fujisaki. 2007. 'Molecular Characterization and Oligosaccharide-Binding Properties of a Galectin from the Argasid Tick Ornithodoros Moubata'. *Glycobiology* 17 (3): 313–23. https://doi.org/10.1093/glycob/cwl070.

Hugerth, Luisa W., Emilie E. L. Muller, Yue O. O. Hu, Laura A. M. Lebrun, Hugo Roume, Daniel Lundin, Paul Wilmes, and Anders F. Andersson. 2014. 'Systematic Design of 18S RRNA Gene Primers for Determining Eukaryotic Diversity in Microbial Consortia'. Edited by Christian R. Voolstra. *PLoS ONE* 9 (4): e95567. https://doi.org/10.1371/journal.pone.0095567.

Husin, Nurul Aini, Jing Jing Khoo, Mulya Mustika Sari Zulkifli, Lesley Bell-Sakyi, and Sazaly AbuBakar. 2021. 'Replication Kinetics of *Rickettsia raoultii* in Tick Cell Lines'. *Microorganisms* 9 (7): 1370. https://doi.org/10.3390/microorganisms9071370.

Igolkina, Y., E. Krasnova, V. Rar, M. Savelieva, T. Epikhina, A. Tikunov, N. Khokhlova, V. Provorova, and N. Tikunova. 2018. 'Detection of Causative Agents of Tick-Borne Rickettsioses in Western Siberia, Russia: Identification of *Rickettsia raoultii* and Rickettsia Sibirica DNA in Clinical Samples'. *Clinical*

Microbiology and Infection 24 (2): 199.e9-199.e12. https://doi.org/10.1016/j.cmi.2017.06.003.

Ilicic, Tomislav, Jong Kyoung Kim, Aleksandra A. Kolodziejczyk, Frederik Otzen Bagger, Davis James McCarthy, John C. Marioni, and Sarah A. Teichmann. 2016. 'Classification of Low Quality Cells from Single-Cell RNA-Seq Data'. *Genome Biology* 17 (1): 29. https://doi.org/10.1186/s13059-016-0888-1.

Imler, Jean-Luc, and Philippe Bulet. 2005. 'Antimicrobial Peptides in Drosophila: Structures,Activities and Gene Regulation'. In *Chemical Immunology and Allergy*, edited by D. Kabelitz and J.-M. Schröder, 1–21. Basel: KARGER. https://doi.org/10.1159/000086648.

International Glossina Genome Initiative, G. M. Attardo, P. P. Abila, J. E. Auma, A. A. Baumann, J. B. Benoit, C. L. Brelsfoard, *et al.* 2014. 'Genome Sequence of the Tsetse Fly (Glossina Morsitans): Vector of African Trypanosomiasis'. *Science* 344 (6182): 380–86. https://doi.org/10.1126/science.1249656.

Itard, Jacques. 1973. 'Revue Des Connaissances Actuelles Sur La Cytogénétique Des Glossines (Diptera)'. *Revue d'élevage et de Médecine Vétérinaire Des Pays Tropicaux* 26 (2): 151. https://doi.org/10.19182/remvt.7848.

Iweriebor, Benson Chuks, Ayabulela Nqoro, and Chikwelu Larry Obi. 2020. 'Rickettsia Africae an Agent of African Tick Bite Fever in Ticks Collected from Domestic Animals in Eastern Cape, South Africa'. *Pathogens (Basel, Switzerland)* 9 (8). https://doi.org/10.3390/pathogens9080631.

Jaenson, Thomas GT, David GE Jaenson, Lars Eisen, Erik Petersson, and Elisabet Lindgren. 2012. 'Changes in the Geographical Distribution and Abundance of the Tick Ixodes Ricinus during the Past 30 Years in Sweden'. *Parasites & Vectors* 5 (1): 8. https://doi.org/10.1186/1756-3305-5-8.

Jain, Chirag, Sergey Koren, Alexander Dilthey, Adam M Phillippy, and Srinivas Aluru. 2018. 'A Fast Adaptive Algorithm for Computing Whole-Genome Homology Maps'. *Bioinformatics* 34 (17): i748–56. https://doi.org/10.1093/bioinformatics/bty597.

Jia, Na, Jinfeng Wang, Wenqiang Shi, Lifeng Du, Yi Sun, Wei Zhan, Jia-Fu Jiang, *et al.* 2020. 'Large-Scale Comparative Analyses of Tick Genomes Elucidate Their Genetic Diversity and Vector Capacities'. *Cell* 182 (5): 1328-1340.e13. https://doi.org/10.1016/j.cell.2020.07.023.

Jia, Na, Yuan-Chun Zheng, Lan Ma, Qiu-Bo Huo, Xue-Bing Ni, Bao-Gui Jiang, Yan-Li Chu, Rui-Ruo Jiang, Jia-Fu Jiang, and Wu-Chun Cao. 2014. 'Human Infections with *Rickettsia raoultii*, China'. *Emerging Infectious Diseases* 20 (5): 866–68. https://doi.org/10.3201/eid2005.130995.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. 'A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity'. *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Jones, Kate E., Nikkita G. Patel, Marc A. Levy, Adam Storeygard, Deborah Balk, John L. Gittleman, and Peter Daszak. 2008. 'Global Trends in Emerging Infectious Diseases'. *Nature* 451 (7181): 990–93. https://doi.org/10.1038/nature06536.

Jongejan, F., and G. Uilenberg. 2004. 'The Global Importance of Ticks'. *Parasitology* 129 (S1): S3–14. https://doi.org/10.1017/S0031182004005967.

Kalle, Elena, Mikael Kubista, and Christopher Rensing. 2014. 'Multi-Template Polymerase Chain Reaction'. *Biomolecular Detection and Quantification* 2 (December): 11–29. https://doi.org/10.1016/j.bdq.2014.11.002.

Kamtchum-Tatuene, Joseph, Benjamin L. Makepeace, Laura Benjamin, Matthew Baylis, and Tom Solomon. 2016. 'The Potential Role of Wolbachia in Controlling the Transmission of Emerging Human Arboviral Infections': *Current Opinion in Infectious Diseases*, November, 1. https://doi.org/10.1097/QCO.0000000000000342.

Karkare, Shantanu, and Deepak Bhatnagar. 2006. 'Promising Nucleic Acid Analogs and Mimics: Characteristic Features and Applications of PNA, LNA, and Morpholino'. *Applied Microbiology and Biotechnology* 71 (5): 575–86. https://doi.org/10.1007/s00253-006-0434-2.

Kawabata, Shun-ichiro, Fuminori Tokunaga, Yoshie Kugi, Shiho Motoyama, Yoshiki Miura, Michimasa Hirata, and Sadaaki Iwanaga. 1996. 'Limulus Factor D, a 43-KDa Protein Isolated from Horseshoe Crab Hemocytes, Is a Serine Protease Homologue with Antimicrobial Activity'. *FEBS Letters* 398 (2–3): 146–50. https://doi.org/10.1016/S0014-5793(96)01224-0.

Kawabata, Shun-ichiro, and Ryoko Tsuda. 2002. 'Molecular Basis of Non-Self Recognition by the Horseshoe Crab Lectins'. *Journal of Endotoxin Research* 8 (6): 437–39. https://doi.org/10.1179/096805102125001037.

Kebede, Assefa, Hagos Ashenafi, and Terzu Daya. 2015. 'A Review on Sterile Insect Technique (SIT) and Its Potential towards Tsetse Eradication in Ethiopia'. *Advances in Life Science and Technology*, 21.

Kevan, Peter G., and Blandina F. Viana. 2003. 'The Global Decline of Pollination Services'. *Biodiversity* 4 (4): 3–8. https://doi.org/10.1080/14888386.2003.9712703.

Kim, Daehwan, Li Song, Florian P. Breitwieser, and Steven L. Salzberg. 2016. 'Centrifuge: Rapid and Sensitive Classification of Metagenomic Sequences'. *Genome Research* 26 (12): 1721–29. https://doi.org/10.1101/gr.210641.116.

Kjos, Sonia A., Karen F. Snowden, and Jimmy K. Olson. 2009. 'Biogeography and *Trypanosoma Cruzi* Infection Prevalence of Chagas Disease Vectors in Texas, USA'. *Vector-Borne and Zoonotic Diseases* 9 (1): 41–50. https://doi.org/10.1089/vbz.2008.0026.

Klasson, Lisa, Zakaria Kambris, Peter E. Cook, Thomas Walker, and Steven P. Sinkins. 2009. 'Horizontal Gene Transfer between Wolbachia and the Mosquito Aedes Aegypti'. *BMC Genomics* 10 (1): 33. https://doi.org/10.1186/1471-2164-10-33.

Klasson, Lisa, Nikhil Kumar, Robin Bromley, Karsten Sieber, Melissa Flowers, Sandra H Ott, Luke J Tallon, Siv G E Andersson, and Julie C Dunning Hotopp. 2014. 'Extensive Duplication of the Wolbachia DNA in Chromosome Four of Drosophila Ananassae'. *BMC Genomics* 15 (1): 1097. https://doi.org/10.1186/1471-2164-15-1097.

Knolhoff, Lisa M., and David W. Onstad. 2014. 'Resistance by Ectoparasites'. In *Insect Resistance Management*, 185–231. Elsevier. https://doi.org/10.1016/B978-0-12-396955-2.00006-0.

Kolmogorov, Mikhail, Derek M. Bickhart, Bahar Behsaz, Alexey Gurevich, Mikhail Rayko, Sung Bong Shin, Kristen Kuhn, *et al.* 2020. 'MetaFlye: Scalable Long-Read Metagenome Assembly Using Repeat Graphs'. *Nature Methods* 17 (11): 1103–10. https://doi.org/10.1038/s41592-020-00971-x.

Kolmogorov, Mikhail, Jeffrey Yuan, Yu Lin, and Pavel A. Pevzner. 2019. 'Assembly of Long, Error-Prone Reads Using Repeat Graphs'. *Nature Biotechnology* 37 (5): 540–46. https://doi.org/10.1038/s41587-019-0072-8.

Kondo, Natsuko, Naruo Nikoh, Nobuyuki Ijichi, Masakazu Shimada, and Takema Fukatsu. 2002. 'Genome Fragment of Wolbachia Endosymbiont Transferred to X Chromosome of Host Insect'. *Proceedings of the National Academy of Sciences* 99 (22): 14280–85. https://doi.org/10.1073/pnas.222228199.

Kopacek, Petr, Ondrej Hajdusek, Veronika Buresova, and Sirlei Daffre. 2010. 'Tick Innate Immunity'. In *Invertebrate Immunity*, edited by Kenneth Soderhall. Springer Science.

Kopacek, Petr, Christoph Weise, Thangamani Saravanan, Katerina Vitova, and Libor Grubhoffer. 2000. 'Characterization of an α-Macroglobulin-like Glycoprotein Isolated from the Plasma of the Soft Tick Ornithodoros Moubata: Tick α-

Macroglobulin'. *European Journal of Biochemistry* 267 (2): 465–75. https://doi.org/10.1046/j.1432-1327.2000.01020.x.

Koutsovoulos, Georgios, Sujai Kumar, Dominik R. Laetsch, Lewis Stevens, Jennifer Daub, Claire Conlon, Habib Maroon, Fran Thomas, Aziz A. Aboobaker, and Mark Blaxter. 2016. 'No Evidence for Extensive Horizontal Gene Transfer in the Genome of the Tardigrade Hypsibius Dujardini'. *Proceedings of the National Academy of Sciences* 113 (18): 5053–58. https://doi.org/10.1073/pnas.1600338113.

Koutsovoulos, Georgios, Benjamin Makepeace, Vincent N. Tanya, and Mark Blaxter. 2014. 'Palaeosymbiosis Revealed by Genomic Fossils of Wolbachia in a Strongyloidean Nematode'. Edited by Cédric Feschotte. *PLoS Genetics* 10 (6): e1004397. https://doi.org/10.1371/journal.pgen.1004397.

Kramer, Laura D, and Alexander T Ciota. 2015. 'Dissecting Vectorial Capacity for Mosquito-Borne Viruses'. *Current Opinion in Virology*, SI: 15 Virus–vector interactions•Viral immunology, 15 (December): 112–18. https://doi.org/10.1016/j.coviro.2015.10.003.

Kramerova, I.A., N. Kawaguchi, L.I. Fessler, R.E. Nelson, Y. Chen, A.A. Kramerov, M. Kusche-Gullberg, *et al.* 2000. 'Papilin in Development; a Pericellular Protein with a Homology to the ADAMTS Metalloproteinases'. *Development* 127 (24): 5475–85. https://doi.org/10.1242/dev.127.24.5475.

Krinsky, William L. 2019. 'Tsetse Flies (Glossinidae)'. In *Medical and Veterinary Entomology*, 369–82. Elsevier. https://doi.org/10.1016/B978-0-12-814043-7.00018-2.

Krober, Thomas, and Patrick M. Guerin. 2007. 'In Vitro Feeding Assays for Hard Ticks'. *Trends in Parasitology* 23 (9): 445–49. https://doi.org/10.1016/j.pt.2007.07.010.

Kugeler, Kiersten J., Amy M. Schwartz, Mark J. Delorey, Paul S. Mead, and Alison F. Hinckley. 2021. 'Estimating the Frequency of Lyme Disease Diagnoses, United States, 2010–2018'. *Emerging Infectious Diseases* 27 (2): 616–19. https://doi.org/10.3201/eid2702.202731.

Kuhn, Karl Heinz, and Tilman Haug. 1994. 'Ultrastructural, Cytochemical, and Immunocytochemical Characterization of Haemocytes of the Hard Tick Ixodes Ricinus (Acari; Chelicerata)'. *Cell and Tissue Research* 277 (3): 493–504. https://doi.org/10.1007/BF00300222.

Kuhnert, F., P. A. Diehl, and P. M. Guerin. 1995. 'The Life-Cycle of the Bont Tick Amblyomma Hebraeum in Vitro'. *International Journal for Parasitology* 25 (8): 887–96. https://doi.org/10.1016/0020-7519(95)00009-q.

Kumar, S., A. Molina-Cruz, L. Gupta, J. Rodrigues, and C. Barillas-Mury. 2010. 'A Peroxidase/Dual Oxidase System Modulates Midgut Epithelial Immunity in Anopheles Gambiae'. *Science* 327 (5973): 1644–48. https://doi.org/10.1126/science.1184008.

Kurtti, Timothy J., Roderick F. Felsheim, Nicole Y. Burkhardt, Jonathan D. Oliver, Chan C. Heu, and Ulrike G. Munderloh. 2015. 'Rickettsia Buchneri Sp. Nov., a Rickettsial Endosymbiont of the Blacklegged Tick Ixodes Scapularis'. *International Journal of Systematic and Evolutionary Microbiology* 65 (Pt_3): 965–70. https://doi.org/10.1099/ijs.0.000047.

Kurtti, Timothy J., and Nemat O. Keyhani. 2008. 'Intracellular Infection of Tick Cell Lines by the Entomopathogenic Fungus Metarhizium Anisopliae'. *Microbiology* 154 (6): 1700–1709. https://doi.org/10.1099/mic.0.2008/016667-0.

Laetsch, Dominik R., and Mark L. Blaxter. 2017. 'BlobTools: Interrogation of Genome Assemblies'. *F1000Research* 6 (July): 1287. https://doi.org/10.12688/f1000research.12232.1.

Lafzi, Atefeh, Catia Moutinho, Simone Picelli, and Holger Heyn. 2018. 'Tutorial: Guidelines for the Experimental Design of Single-Cell RNA Sequencing Studies'. *Nature Protocols* 13 (12): 2742–57. https://doi.org/10.1038/s41596-018-0073-y.

Lai, De-Hua, Hassan Hashimi, Zhao-Rong Lun, Francisco J. Ayala, and Julius Lukeš. 2008. 'Adaptations of Trypanosoma Brucei to Gradual Loss of Kinetoplast DNA: Trypanosoma Equiperdum and Trypanosoma Evansi Are Petite Mutants of T. Brucei'. *Proceedings of the National Academy of Sciences* 105 (6): 1999–2004. https://doi.org/10.1073/pnas.0711799105.

Langmead, Ben, and Steven L. Salzberg. 2012. 'Fast Gapped-Read Alignment with Bowtie 2'. *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

Latinne, Alice, Ben Hu, Kevin J. Olival, Guangjian Zhu, Libiao Zhang, Hongying Li, Aleksei A. Chmura, *et al.* 2020. 'Origin and Cross-Species Transmission of Bat Coronaviruses in China'. *Nature Communications* 11 (1): 4235. https://doi.org/10.1038/s41467-020-17687-3.

Lazzaro, B. P., and J. Rolff. 2011. 'Danger, Microbes, and Homeostasis'. *Science* 332 (6025): 43–44. https://doi.org/10.1126/science.1200486.

Lees, John A, Marco Galardini, Stephen D Bentley, Jeffrey N Weiser, and Jukka Corander. 2018. 'Pyseer: A Comprehensive Tool for Microbial Pangenome-Wide Association Studies'. Edited by Oliver Stegle. *Bioinformatics* 34 (24): 4310–12. https://doi.org/10.1093/bioinformatics/bty539.

Leger, Elsa, Gwenael Vourc'h, Laurence Vial, Christine Chevillon, and Karen D. McCoy. 2013. 'Changing Distributions of Ticks: Causes and Consequences'. *Experimental and Applied Acarology* 59 (1): 219–44. https://doi.org/10.1007/s10493-012-9615-0.

Lemaitre, B., E. Kromer-Metzger, L. Michaut, E. Nicolas, M. Meister, P. Georgel, J. M. Reichhart, and J. A. Hoffmann. 1995. 'A Recessive Mutation, Immune Deficiency (Imd), Defines Two Distinct Control Pathways in the Drosophila Host Defense.' *Proceedings of the National Academy of Sciences* 92 (21): 9465–69. https://doi.org/10.1073/pnas.92.21.9465.

Leray, Matthieu, Natalia Agudelo, Suzanne C. Mills, and Christopher P. Meyer. 2013. 'Effectiveness of Annealing Blocking Primers versus Restriction Enzymes for Characterization of Generalist Diets: Unexpected Prey Revealed in the Gut Contents of Two Coral Reef Fish Species'. Edited by Bernd Schierwater. *PLoS ONE* 8 (4): e58076. https://doi.org/10.1371/journal.pone.0058076.

Lewin, Harris A., Gene E. Robinson, W. John Kress, William J. Baker, Jonathan Coddington, Keith A. Crandall, Richard Durbin, *et al.* 2018. 'Earth BioGenome Project: Sequencing Life for the Future of Life'. *Proceedings of the National Academy of Sciences* 115 (17): 4325–33. https://doi.org/10.1073/pnas.1720115115.

Lew-Tabor, A.E., and M. Rodriguez Valle. 2016. 'A Review of Reverse Vaccinology Approaches for the Development of Vaccines against Ticks and Tick Borne Diseases'. *Ticks and Tick-Borne Diseases* 7 (4): 573–85. https://doi.org/10.1016/j.ttbdis.2015.12.012.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Li, Heng. 2018. 'Minimap2: Pairwise Alignment for Nucleotide Sequences'. Edited by Inanc Birol. *Bioinformatics* 34 (18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, Hongjie, Jasper Janssens, Maxime De Waegeneer, Sai Saroja Kolluru, Kristofer Davie, Vincent Gardeux, Wouter Saelens, *et al.* 2021. 'Fly Cell Atlas: A Single-Cell Transcriptomic Atlas of the Adult Fruit Fly'. *BioRxiv*. https://doi.org/10.1101/2021.07.04.451050.

Li, J., R. R. Gutell, S. H. Damberger, R. A. Wirtz, J. C. Kissinger, M. J. Rogers, J. Sattabongkot, and T. F. McCutchan. 1997. 'Regulation and Trafficking of Three Distinct 18 S Ribosomal RNAs during Development of the Malaria Parasite'.

Journal of Molecular Biology 269 (2): 203–13. https://doi.org/10.1006/jmbi.1997.1038.

Lindgren, E., L. Talleklint, and T. Polfeldt. 2000. 'Impact of Climatic Change on the Northern Latitude Limit and Population Density of the Disease-Transmitting European Tick Ixodes Ricinus.' *Environmental Health Perspectives* 108 (2): 119–23. https://doi.org/10.1289/ehp.00108119.

Liu, Lei, Jianfeng Dai, Yang O. Zhao, Sukanya Narasimhan, Ying Yang, Lili Zhang, and Erol Fikrig. 2012. 'Ixodes Scapularis JAK-STAT Pathway Regulates Tick Antimicrobial Peptides, Thereby Controlling the Agent of Human Granulocytic Anaplasmosis'. *The Journal of Infectious Diseases* 206 (8): 1233–41. https://doi.org/10.1093/infdis/jis484.

Liu, Xiang Ye, Jose de la Fuente, Martine Cote, Ruth C. Galindo, Sara Moutailler, Muriel Vayssier-Taussat, and Sarah I. Bonnet. 2014. 'IrSPI, a Tick Serine Protease Inhibitor Involved in Tick Feeding and Bartonella Henselae Infection'. Edited by Joseph M. Vinetz. *PLoS Neglected Tropical Diseases* 8 (7): e2993. https://doi.org/10.1371/journal.pntd.0002993.

Loftus, Brendan, Iain Anderson, Rob Davies, U. Cecilia M. Alsmark, John Samuelson, Paolo Amedeo, Paola Roncaglia, *et al.* 2005. 'The Genome of the Protist Parasite *Entamoeba Histolytica*'. *Nature* 433 (7028): 865–68. https://doi.org/10.1038/nature03291.

Lorusso, Vincenzo, Karolina Anna Gruszka, Ayodele Majekodunmi, Augustine Igweh, Susan C. Welburn, and Kim Picozzi. 2013. 'Rickettsia Africae in *Amblyomma variegatum* Ticks, Uganda and Nigeria'. *Emerging Infectious Diseases* 19 (10): 1705. https://doi.org/10.3201/eid1910.130389.

Lu, Peng, Guowu Bian, Xiaoling Pan, and Zhiyong Xi. 2012. 'Wolbachia Induces Density-Dependent Inhibition to Dengue Virus in Mosquito Cells'. Edited by Scott L. O'Neill. *PLoS Neglected Tropical Diseases* 6 (7): e1754. https://doi.org/10.1371/journal.pntd.0001754.

Lucas, Eric R., Alistair Miles, Nicholas J. Harding, Chris S. Clarkson, Mara K. N. Lawniczak, Dominic P. Kwiatkowski, David Weetman, Martin J. Donnelly, and The Anopheles gambiae 1000 Genomes Consortium. 2019. 'Whole-Genome Sequencing Reveals High Complexity of Copy Number Variation at Insecticide Resistance Loci in Malaria Mosquitoes'. *Genome Research* 29 (8): 1250–61. https://doi.org/10.1101/gr.245795.118.

Luecken, Malte D, and Fabian J Theis. 2019. 'Current Best Practices in Single-cell RNA-seq Analysis: A Tutorial'. *Molecular Systems Biology* 15 (6). https://doi.org/10.15252/msb.20188746.

Lukes, Julius, Anzhelika Butenko, Hassan Hashimi, Dmitri A. Maslov, Jan Votýpka, and Vyacheslav Yurchenko. 2018. 'Trypanosomatids Are Much More than Just Trypanosomes: Clues from the Expanded Family Tree'. *Trends in Parasitology* 34 (6): 466–80. https://doi.org/10.1016/j.pt.2018.03.002.

Lun, Aaron T. L., Karsten Bach, and John C. Marioni. 2016. 'Pooling across Cells to Normalize Single-Cell RNA Sequencing Data with Many Zero Counts'. *Genome Biology* 17 (1): 75. https://doi.org/10.1186/s13059-016-0947-7.

Lundberg, Derek S., Sarah L. Lebeis, Sur Herrera Paredes, Scott Yourstone, Jase Gehring, Stephanie Malfatti, Julien Tremblay, *et al.* 2012. 'Defining the Core *Arabidopsis Thaliana* Root Microbiome'. *Nature* 488 (7409): 86–90. https://doi.org/10.1038/nature11237.

Lundberg, Derek S, Scott Yourstone, Piotr Mieczkowski, Corbin D Jones, and Jeffery L Dangl. 2013. 'Practical Innovations for High-Throughput Amplicon Sequencing'. *Nature Methods* 10 (10): 999–1002. https://doi.org/10.1038/nmeth.2634.

Macdonald, G. 1952. 'The Analysis of the Sporozoite Rate.' *Tropical Diseases Bulletin* 49 (6). https://www.cabdirect.org/cabdirect/abstract/19541000114.

Maggi, Ricardo G., and Friederike Krämer. 2019. 'A Review on the Occurrence of Companion Vector-Borne Diseases in Pet Animals in Latin America'. *Parasites & Vectors* 12 (1): 145. https://doi.org/10.1186/s13071-019-3407-x.

Makepeace, Benjamin L., Lisa Rodgers, and Alexander J. Trees. 2006. 'Rate of Elimination of Wolbachia Pipientis by Doxycycline In Vitro Increases Following Drug Withdrawal'. *Antimicrobial Agents and Chemotherapy* 50 (3): 922–27. https://doi.org/10.1128/AAC.50.3.922-927.2006.

Malhotra, Bansi Dhar, and Md. Azahar Ali. 2018. 'Chapter 8 - Nanostructured Materials for DNA Biochip'. In *Nanomaterials for Biosensors*, edited by Bansi Dhar Malhotra and Md. Azahar Ali, 221–62. Micro and Nano Technologies. William Andrew Publishing. https://doi.org/10.1016/B978-0-323-44923-6.00008-X.

Marcais, Guillaume, and Carl Kingsford. 2011. 'A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of k-Mers'. *Bioinformatics* 27 (6): 764–70. https://doi.org/10.1093/bioinformatics/btr011.

Marmiroli, Nelson, and Elena Maestri. 2007. 'Chapter 6 - Polymerase Chain Reaction (PCR)'. In *Food Toxicants Analysis*, edited by Yolanda Picó, 147–87. Amsterdam: Elsevier. https://doi.org/10.1016/B978-044452843-8/50007-9.

Marques, Roberta, Rodrigo F. Krüger, A. Townsend Peterson, Larissa F. de Melo, Natália Vicenzi, and Daniel Jiménez-García. 2020. 'Climate Change Implications for the Distribution of the Babesiosis and Anaplasmosis Tick Vector, Rhipicephalus (Boophilus) Microplus'. *Veterinary Research* 51 (1): 81. https://doi.org/10.1186/s13567-020-00802-z.

Martin, Marcel. 2011. 'Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads'. *EMBnet.Journal* 17 (1): 10. https://doi.org/10.14806/ej.17.1.200.

Martin, William F. 2017. 'Too Much Eukaryote LGT'. *BioEssays* 39 (12): 1700115. https://doi.org/10.1002/bies.201700115.

Mateos-Hernandez, Lourdes, Natália Pipová, Eléonore Allain, Céline Henry, Clotilde Rouxel, Anne-Claire Lagrée, Nadia Haddad, *et al.* 2021. 'Enlisting the Ixodes Scapularis Embryonic ISE6 Cell Line to Investigate the Neuronal Basis of Tick—Pathogen Interactions'. *Pathogens* 10 (1): 70. https://doi.org/10.3390/pathogens10010070.

Mattila, Joshua T., Ulrike G. Munderloh, and Timothy J. Kurtti. 2007. 'Phagocytosis of the Lyme Disease Spirochete, *Borrelia Burgdorferi* , by Cells from the Ticks, *Ixodes Scapularis* and *Dermacentor Andersoni* , Infected with An Endosymbiont, *Rickettsia Peacockii*'. *Journal of Insect Science* 7 (58): 1–12. https://doi.org/10.1673/031.007.5801.

McGraw, Elizabeth A., and Scott L. O'Neill. 2013. 'Beyond Insecticides: New Thinking on an Ancient Problem'. *Nature Reviews Microbiology* 11 (3): 181–93. https://doi.org/10.1038/nrmicro2968.

McKenna, Victoria, John M. Archibald, Roxanne Beinart, Michael N. Dawson, Ute Hentschel, Patrick J. Keeling, Jose V. Lopez, *et al.* 2021. 'The Aquatic Symbiosis Genomics Project: Probing the Evolution of Symbiosis across the Tree of Life'. *Wellcome Open Research* 6 (October): 254. https://doi.org/10.12688/wellcomeopenres.17222.1.

Mediannikov, Oleg, Kotaro Matsumoto, Irina Samoylenko, Michel Drancourt, Véronique Roux, Elena Rydkina, Bernard Davoust, Irina Tarasevich, Philippe Brouqui, and Pierre-EdouardYR 2008 Fournier. 2008. '*Rickettsia raoultii* Sp. Nov., a Spotted Fever Group Rickettsia Associated with Dermacentor Ticks in Europe and Russia'. *International Journal of Systematic and Evolutionary Microbiology* 58 (7): 1635–39. https://doi.org/10.1099/ijs.0.64952-0.

Medini, H., T. Cohen, and D. Mishmar. 2021. 'Mitochondrial Gene Expression in Single Cells Shape Pancreatic Beta Cells' Sub-Populations and Explain Variation in Insulin Pathway'. *Scientific Reports* 11 (1): 466. https://doi.org/10.1038/s41598-020-80334-w.

Medlin, Linda, Hille J. Elwood, Shawn Stickel, and Mitchell L. Sogin. 1988. 'The Characterization of Enzymatically Amplified Eukaryotic 16S-like RRNA-Coding Regions'. *Gene* 71 (2): 491–99. https://doi.org/10.1016/0378-1119(88)90066-2.

Mercereau-Puijalon, Odile, Jean-Christophe Barale, and Emmanuel Bischoff. 2002. 'Three Multigene Families in *Plasmodium* Parasites: Facts and Questions'. *International Journal for Parasitology* 32 (11): 1323–44. https://doi.org/10.1016/s0020-7519(02)00111-x.

Mesquita, Rafael D., Raquel J. Vionette-Amaral, Carl Lowenberger, Rolando Rivera-Pomar, Fernando A. Monteiro, Patrick Minx, John Spieth, *et al.* 2015. 'Genome of Rhodnius Prolixus, an Insect Vector of Chagas Disease, Reveals Unique Adaptations to Hematophagy and Parasite Infection'. *Proceedings of the National Academy of Sciences* 112 (48): 14936–41. https://doi.org/10.1073/pnas.1506226112.

Miller, Jason R., Sergey Koren, Kari A. Dilley, Derek M. Harkins, Timothy B. Stockwell, Reed S. Shabman, and Granger G. Sutton. 2018. 'A Draft Genome Sequence for the Ixodes Scapularis Cell Line, ISE6'. *F1000Research* 7 (March): 297. https://doi.org/10.12688/f1000research.13635.1.

Mirsaeidi, Mehdi, Hooman Motahari, Mojdeh Taghizadeh Khamesi, Arash Sharifi, Michael Campos, and Dean E. Schraufnagel. 2016. 'Climate Change and Respiratory Infections'. *Annals of the American Thoracic Society* 13 (8): 1223–30. https://doi.org/10.1513/AnnalsATS.201511-729PS.

Miyoshi, Naruhide, Emiko Isogai, Keiichi Hiramatsu, and Takashi Sasaki. 2017. 'Activity of Tick Antimicrobial Peptide from Ixodes Persulcatus (Persulcatusin) against Cell Membranes of Drug-Resistant Staphylococcus Aureus'. *The Journal of Antibiotics* 70 (2): 142–46. https://doi.org/10.1038/ja.2016.101.

Montagna, Matteo, Davide Sassera, Francesca Griggio, Sara Epis, Claudio Bandi, and Carmela Gissi. 2012. 'Tick-Box for 3′-End Formation of Mitochondrial Transcripts in Ixodida, Basal Chelicerates and Drosophila'. Edited by Christos A. Ouzounis. *PLoS ONE* 7 (10): e47538. https://doi.org/10.1371/journal.pone.0047538.

Munderloh, Ulrike G., and Timothy J. Kurtti. 1989. 'Formulation of Medium for Tick Cell Culture'. *Experimental & Applied Acarology* 7 (3): 219–29. https://doi.org/10.1007/BF01194061.

Munderloh, Ulrike G., Yan Liu, Maming Wang, Chunsheng Chen, and Timothy J. Kurtti. 1994. 'Establishment, Maintenance and Description of Cell Lines from the Tick Ixodes Scapularis'. *The Journal of Parasitology* 80 (4): 533. https://doi.org/10.2307/3283188.

Mysterud, Atle, Ragna Byrkjeland, Lars Qviller, and Hildegunn Viljugrein. 2015. 'The Generalist Tick Ixodes Ricinus and the Specialist Tick Ixodes Trianguliceps on Shrews and Rodents in a Northern Forest Ecosystem– a Role of Body Size Even among Small Hosts'. *Parasites & Vectors* 8 (1): 639. https://doi.org/10.1186/s13071-015-1258-7.

Nachmanson, Daniela, Shenyi Lian, Elizabeth K. Schmidt, Michael J. Hipp, Kathryn T. Baker, Yuezheng Zhang, Maria Tretiakova, *et al.* 2018. 'Targeted Genome Fragmentation with CRISPR/Cas9 Enables Fast and Efficient Enrichment of Small Genomic Regions and Ultra-Accurate Sequencing with Low DNA Input (CRISPR-DS)'. *Genome Research* 28 (10): 1589–99. https://doi.org/10.1101/gr.235291.118.

Netea, Mihai G., Leo A. B. Joosten, Eicke Latz, Kingston H. G. Mills, Gioacchino Natoli, Hendrik G. Stunnenberg, Luke A. J. O'Neill, and Ramnik J. Xavier. 2016. 'Trained Immunity: A Program of Innate Immune Memory in Health and Disease'. *Science* 352 (6284): aaf1098. https://doi.org/10.1126/science.aaf1098.

Nghochuzie, Nora Nganyewo, Charles Ochieng' Olwal, Aniefiok John Udoakang, Lucas Naam-Kayagre Amenga-Etego, and Alfred Amambua-Ngwa. 2020. 'Pausing the Fight Against Malaria to Combat the COVID-19 Pandemic in Africa: Is the Future of Malaria Bleak?' *Frontiers in Microbiology* 11: 1476. https://doi.org/10.3389/fmicb.2020.01476.

Nielsen, Cydney B., Shaun D. Jackman, Inanç Birol, and Steven J.M. Jones. 2009. 'ABySS-Explorer: Visualizing Genome Sequence Assemblies'. *IEEE Transactions on Visualization and Computer Graphics* 15 (6): 881–88. https://doi.org/10.1109/TVCG.2009.116.

Nikoh, Naruo, Kohjiro Tanaka, Fukashi Shibata, Natsuko Kondo, Masahiro Hizume, Masakazu Shimada, and Takema Fukatsu. 2008. 'Wolbachia Genome Integrated in an Insect Chromosome: Evolution and Fate of Laterally Transferred Endosymbiont Genes'. *Genome Research* 18 (2): 272–80. https://doi.org/10.1101/gr.7144908.

Niu, Yue, and Yun Xiang. 2018. 'An Overview of Biomembrane Functions in Plant Responses to High-Temperature Stress'. *Frontiers in Plant Science* 9 (July): 915. https://doi.org/10.3389/fpls.2018.00915.

Noronha, Nandita, Grégory Ehx, Marie-Christine Meunier, Jean-Philippe Laverdure, Catherine Thériault, and Claude Perreault. 2020. 'Major Multilevel Molecular Divergence between THP-1 Cells from Different Biorepositories'. *International Journal of Cancer* 147 (7): 2000–2006. https://doi.org/10.1002/ijc.32967.

Nurk, Sergey, Brian P. Walenz, Arang Rhie, Mitchell R. Vollger, Glennis A. Logsdon, Robert Grothe, Karen H. Miga, Evan E. Eichler, Adam M. Phillippy, and

Sergey Koren. 2020. 'HiCanu: Accurate Assembly of Segmental Duplications, Satellites, and Allelic Variants from High-Fidelity Long Reads'. *Genome Research* 30 (9): 1291–1305. https://doi.org/10.1101/gr.263566.120.

Ogden, N.H., M. Bigras-Poulin, C.J. O'Callaghan, I.K. Barker, L.R. Lindsay, A. Maarouf, K.E. Smoyer-Tomic, D. Waltner-Toews, and D. Charron. 2005. 'A Dynamic Population Model to Investigate Effects of Climate on Geographic Range and Seasonality of the Tick Ixodes Scapularis'. *International Journal for Parasitology* 35 (4): 375–89. https://doi.org/10.1016/j.ijpara.2004.12.013.

Oines, Oivind, Jana Radzijevskaja, Algimantas Paulauskas, and Olav Rosef. 2012. 'Prevalence and Diversity of *Babesia* Spp. in Questing *Ixodes Ricinus* Ticks from Norway'. *Parasites & Vectors* 5 (1): 156. https://doi.org/10.1186/1756-3305-5-156.

Oliver, Jonathan D., Adela S. Oliva Chávez, Roderick F. Felsheim, Timothy J. Kurtti, and Ulrike G. Munderloh. 2015. 'An Ixodes Scapularis Cell Line with a Predominantly Neuron-like Phenotype'. *Experimental and Applied Acarology* 66 (3): 427–42. https://doi.org/10.1007/s10493-015-9908-1.

O'Neill, Scott L. 2018. 'The Use of Wolbachia by the World Mosquito Program to Interrupt Transmission of Aedes Aegypti Transmitted Viruses'. In *Dengue and Zika: Control and Antiviral Treatment Strategies*, edited by Rolf Hilgenfeld and Subhash G. Vasudevan, 1062:355–60. Advances in Experimental Medicine and Biology. Singapore: Springer. https://doi.org/10.1007/978-981-10-8727-1_24.

Onyiche, ThankGod E., Cristian Răileanu, Susanne Fischer, and Cornelia Silaghi. 2021. 'Global Distribution of *Babesia* Species in Questing Ticks: A Systematic Review and Meta-Analysis Based on Published Literature'. *Pathogens* 10 (2): 230. https://doi.org/10.3390/pathogens10020230.

O'rorke, R., S. Lavery, and A. Jeffs. 2012. 'PCR Enrichment Techniques to Identify the Diet of Predators'. *Molecular Ecology Resources* 12 (1): 5–17. https://doi.org/10.1111/j.1755-0998.2011.03091.x.

Ostfeld, Richard S., Amber Price, Victoria L. Hornbostel, Michael A. Benjamin, and Felicia Keesing. 2006. 'Controlling Ticks and Tick-Borne Zoonoses with Biological and Chemical Agents'. *BioScience* 56 (5): 383. https://doi.org/10.1641/0006-3568(2006)056[0383:CTATZW]2.0.CO;2.

Pagel Van Zee, J., N.S. Geraci, F.D. Guerrero, S.K. Wikel, J.J. Stuart, V.M. Nene, and C.A. Hill. 2007. 'Tick Genomics: The Ixodes Genome Project and Beyond'. *International Journal for Parasitology* 37 (12): 1297–1305. https://doi.org/10.1016/j.ijpara.2007.05.011.

Pais, Roshan, Claudia Lohs, Yineng Wu, Jingwen Wang, and Serap Aksoy. 2008. 'The Obligate Mutualist Wigglesworthia Glossinidia Influences Reproduction, Digestion, and Immunity Processes of Its Host, the Tsetse Fly'. *Applied and Environmental Microbiology* 74 (19): 5965–74. https://doi.org/10.1128/AEM.00741-08.

Palomar, Ana M., Shonnette Premchand-Branker, Pilar Alberdi, Oxana A. Belova, Anna Moniuszko-Malinowska, Olaf Kahl, and Lesley Bell-Sakyi. 2019. 'Isolation of Known and Potentially Pathogenic Tick-Borne Microorganisms from European Ixodid Ticks Using Tick Cell Lines'. *Ticks and Tick-Borne Diseases* 10 (3): 628–38. https://doi.org/10.1016/j.ttbdis.2019.02.008.

Parola, Philippe, Christopher D. Paddock, Cristina Socolovschi, Marcelo B. Labruna, Oleg Mediannikov, Tahar Kernif, Mohammad Yazid Abdad, *et al.* 2013. 'Update on Tick-Borne Rickettsioses around the World: A Geographic Approach'. *Clinical Microbiology Reviews* 26 (4): 657–702. https://doi.org/10.1128/CMR.00032-13.

Pawlowski, Jan, Stéphane Audic, Sina Adl, David Bass, Lassaâd Belbahri, Cédric Berney, Samuel S. Bowser, *et al.* 2012. 'CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms'. *PLOS Biology* 10 (11): e1001419. https://doi.org/10.1371/journal.pbio.1001419.

Pelt-Verkuil, Elizabeth van, Alex van Belkum, and John P. Hays. 2008. 'PCR Primers'. In *Principles and Technical Aspects of PCR Amplification*, 63–90. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-6241-4_5.

Pereira, L. S., P. L. Oliveira, C. Barja-Fidalgo, and S. Daffre. 2001. 'Production of Reactive Oxygen Species by Hemocytes from the Cattle Tick Boophilus Microplus'. *Experimental Parasitology* 99 (2): 66–72. https://doi.org/10.1006/expr.2001.4657.

Pfaffle, Miriam, Nina Littwin, Senta V. Muders, and Trevor N. Petney. 2013. 'The Ecology of Tick-Borne Diseases'. *International Journal for Parasitology* 43 (12–13): 1059–77. https://doi.org/10.1016/j.ijpara.2013.06.009.

Philippe, Hervé. 2000. 'Opinion: Long Branch Attraction and Protist Phylogeny'. *Protist* 151 (4): 307–16. https://doi.org/10.1078/S1434-4610(04)70029-2.

Pintore, Elisabetta, Emanuela Olivieri, Anna Maria Floriano, Davide Sassera, Nino Sanna, and Giovanni Garippa. 2021. 'First Detection of *Amblyomma variegatum* and Molecular Finding of Rickettsia Africae in Sardinia, Italy'. *Ticks and Tick-Borne Diseases* 12 (1): 101561. https://doi.org/10.1016/j.ttbdis.2020.101561.

Podboronov, VM. 1991. 'Antibacterial Protective Mechanisms of Ixodoid Ticks'. In *Modern Acarology*, edited by Frantisek Dusbábek and Vladimir Bukva, 2:375–80. Prague: Academia.

Polishchuk, M. V., T. D. Zdolnik, and V. N. Smetanin. 2017. 'Ixodes Tick-Borne Borrelioses: Modern Epidemiological Situation in The Center of The European Part of Russia'. *I.P. Pavlov Russian Medical Biological Herald* 25 (2): 202–8. https://doi.org/10.23888/PAVLOVJ20172202-208.

Popov, Vsevolod L., Edward I. Korenberg, Valentina V. Nefedova, Violet C. Han, Julie W. Wen, Yurii V. Kovalevskii, Natalia B. Gorelova, and David H. Walker. 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 and Zoonotic Diseases* 7 (4): 699–716. https://doi.org/10.1089/vbz.2007.0148.

Price, Dana C., Jürgen M. Steiner, Hwan Su Yoon, Debashish Bhattacharya, and Wolfgang Löffelhardt. 2017. 'Glaucophyta'. In *Handbook of the Protists*, edited by John M. Archibald, Alastair G.B. Simpson, Claudio H. Slamovits, Lynn Margulis, Michael Melkonian, David J. Chapman, and John O. Corliss, 1–65. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-32669-6_42-1.

Puigbò, Pere, Sergei Mekhedov, Yuri I. Wolf, and Eugene V. Koonin. 2011. 'A Comprehensive Census of Horizontal Gene Transfers from Prokaryotes to Unikonts'. *Genome Biology* 12 (1): P20. https://doi.org/10.1186/1465-6906-12-S1-P20.

Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. 'The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools'. *Nucleic Acids Research* 41 (D1): D590–96. https://doi.org/10.1093/nar/gks1219.

Rahlenbeck, Sibylle, Volker Fingerle, and Stephen Doggett. 2016. 'Prevention of Tick-Borne Diseases: An Overview'. *British Journal of General Practice* 66 (650): 492–94. https://doi.org/10.3399/bjgp16X687013.

Rana, Vipin S., Chrysoula Kitsou, Shraboni Dutta, Michael H. Ronzetti, Min Zhang, Quentin Bernard, Alexis A. Smith, *et al.* 2023. 'Dome1–JAK–STAT Signaling between Parasite and Host Integrates Vector Immunity and Development'. *Science* 379 (6628): eabl3837. https://doi.org/10.1126/science.abl3837.

Rego, R, V Kovar, P Kopacek, C Weise, P Man, I Sauman, and L Grubhoffer. 2006. 'The Tick Plasma Lectin, Dorin M, Is a Fibrinogen-Related Molecule☆'. *Insect*

Biochemistry and Molecular Biology 36 (4): 291–99. https://doi.org/10.1016/j.ibmb.2006.01.008.

Rego, Ryan O. M., Ondrej Hajdusek, Vojtech Kovár, Petr Kopácek, Libor Grubhoffer, and Václav Hypsa. 2005. 'Molecular Cloning and Comparative Analysis of Fibrinogen-Related Proteins from the Soft Tick Ornithodoros Moubata and the Hard Tick Ixodes Ricinus'. *Insect Biochemistry and Molecular Biology* 35 (9): 991–1004. https://doi.org/10.1016/j.ibmb.2005.04.001.

Rice, Edward S., Satomi Kohno, John St John, Son Pham, Jonathan Howard, Liana F. Lareau, Brendan L. O'Connell, *et al.* 2017. 'Improved Genome Assembly of American Alligator Genome Reveals Conserved Architecture of Estrogen Signaling'. *Genome Research* 27 (5): 686–96. https://doi.org/10.1101/gr.213595.116.

Rice, Peter, Ian Longden, and Alan Bleasby. 2000. 'EMBOSS: The European Molecular Biology Open Software Suite'. *Trends in Genetics* 16 (6): 276–77. https://doi.org/10.1016/S0168-9525(00)02024-2.

Rizzoli, Annapaola, Cornelia Silaghi, Anna Obiegala, Ivo Rudolf, Zdeněk Hubálek, Gábor Földvári, Olivier Plantard, *et al.* 2014. 'Ixodes Ricinus and Its Transmitted Pathogens in Urban and Peri-Urban Areas in Europe: New Hazards and Relevance for Public Health'. *Frontiers in Public Health* 2: 251. https://doi.org/10.3389/fpubh.2014.00251.

Roberts, Richard J. 2005. 'How Restriction Enzymes Became the Workhorses of Molecular Biology'. *Proceedings of the National Academy of Sciences* 102 (17): 5905–8. https://doi.org/10.1073/pnas.0500923102.

Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov. 2011. 'Integrative Genomics Viewer'. *Nature Biotechnology* 29 (1): 24–26. https://doi.org/10.1038/nbt.1754.

Rocklov, Joacim, and Robert Dubrow. 2020. 'Climate Change: An Enduring Challenge for Vector-Borne Disease Prevention and Control'. *Nature Immunology* 21 (5): 479–83. https://doi.org/10.1038/s41590-020-0648-y.

Rodriguez-Vivas, Roger I., Nicholas N. Jonsson, and Chandra Bhushan. 2018. 'Strategies for the Control of *Rhipicephalus microplus* Ticks in a World of Conventional Acaricide and Macrocyclic Lactone Resistance'. *Parasitology Research* 117 (1): 3–29. https://doi.org/10.1007/s00436-017-5677-6.

Romoli, Ottavia, and Mathilde Gendrin. 2018. 'The Tripartite Interactions between the Mosquito, Its Microbiota and Plasmodium'. *Parasites & Vectors* 11 (1): 200. https://doi.org/10.1186/s13071-018-2784-x.

Rozenblatt-Rosen, Orit, Michael J. T. Stubbington, Aviv Regev, and Sarah A. Teichmann. 2017. 'The Human Cell Atlas: From Vision to Reality'. *Nature* 550 (7677): 451–53. https://doi.org/10.1038/550451a.

Ruan, Jue, and Heng Li. 2020. 'Fast and Accurate Long-Read Assembly with Wtdbg2'. *Nature Methods* 17 (2): 155–58. https://doi.org/10.1038/s41592-019-0669-3.

Ruby, Edward G. 2008. 'Symbiotic Conversations Are Revealed under Genetic Interrogation'. *Nature Reviews Microbiology* 6 (10): 752–62. https://doi.org/10.1038/nrmicro1958.

Rushmore, Julie, Donal Bisanzio, and Thomas R. Gillespie. 2017. 'Making New Connections: Insights from Primate–Parasite Networks'. *Trends in Parasitology* 33 (7): 547–60. https://doi.org/10.1016/j.pt.2017.01.013.

Saito, Tais B., and David H. Walker. 2016. 'Ehrlichioses: An Important One Health Opportunity'. *Veterinary Sciences* 3 (3): 20. https://doi.org/10.3390/vetsci3030020.

Saito, Y., S. Konnai, S. Yamada, S. Imamura, H. Nishikado, T. Ito, M. Onuma, and K. Ohashi. 2009. 'Identification and Characterization of Antimicrobial Peptide, Defensin, in the Taiga Tick, *Ixodes Persulcatus*'. *Insect Molecular Biology* 18 (4): 531–39. https://doi.org/10.1111/j.1365-2583.2009.00897.x.

Salata, Cristiano, Sara Moutailler, Houssam Attoui, Erich Zweygarth, Lygia Decker, and Lesley Bell-Sakyi. 2021. 'How Relevant Are in Vitro Culture Models for Study of Tick-Pathogen Interactions?' *Pathogens and Global Health* 0 (0): 1–19. https://doi.org/10.1080/20477724.2021.1944539.

Santos, Susana S., Tue Kjærgaard Nielsen, Lars H. Hansen, and Anne Winding. 2015. 'Comparison of Three DNA Extraction Methods for Recovery of Soil Protist DNA'. *Journal of Microbiological Methods* 115 (August): 13–19. https://doi.org/10.1016/j.mimet.2015.05.011.

Satou, Yutaka, Katsuhiko Mineta, Michio Ogasawara, Yasunori Sasakura, Eiichi Shoguchi, Keisuke Ueno, Lixy Yamada, *et al.* 2008. 'Improved Genome Assembly and Evidence-Based Global Gene Model Set for the Chordate Ciona Intestinalis: New Insight into Intron and Operon Populations'. *Genome Biology* 9 (10): R152. https://doi.org/10.1186/gb-2008-9-10-r152.

Schaum, Nicholas, Jim Karkanias, Norma F Neff, Andrew P. May, Stephen R. Quake, Tony Wyss-Coray, Spyros Darmanis, *et al.* 2017. 'Single-Cell Transcriptomic Characterization of 20 Organs and Tissues from Individual Mice Creates a *Tabula Muris*', December. https://doi.org/10.1101/237446.

Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, *et al.* 2012. 'Fiji: An Open-Source Platform for Biological-Image Analysis'. *Nature Methods* 9 (7): 676–82. https://doi.org/10.1038/nmeth.2019.

Schneider, Daniela I., Lisa Klasson, Anders E. Lind, and Wolfgang J. Miller. 2014. 'More than Fishing in the Dark: PCR of a Dispersed Sequence Produces Simple but Ultrasensitive Wolbachia Detection'. *BMC Microbiology* 14 (1): 121. https://doi.org/10.1186/1471-2180-14-121.

Schoch, Conrad L., Keith A. Seifert, Sabine Huhndorf, Vincent Robert, John L. Spouge, C. André Levesque, Wen Chen, and Fungal Barcoding Consortium. 2012. 'Nuclear Ribosomal Internal Transcribed Spacer (ITS) Region as a Universal DNA Barcode Marker for Fungi'. *Proceedings of the National Academy of Sciences* 109 (16): 6241–46. https://doi.org/10.1073/pnas.1117018109.

Schreiber, Claire A., Toshie Sakuma, Yoshihiro Izumiya, Sara J. Holditch, Raymond D. Hickey, Robert K. Bressin, Upamanyu Basu, Kazunori Koide, Aravind Asokan, and Yasuhiro Ikeda. 2015. 'An SiRNA Screen Identifies the U2 SnRNP Spliceosome as a Host Restriction Factor for Recombinant Adeno-Associated Viruses'. *PLOS Pathogens* 11 (8): e1005082. https://doi.org/10.1371/journal.ppat.1005082.

Seeleuthner, Yoann, Tara Oceans Coordinators, Samuel Mondy, Vincent Lombard, Quentin Carradec, Eric Pelletier, Marc Wessner, *et al.* 2018. 'Single-Cell Genomics of Multiple Uncultured Stramenopiles Reveals Underestimated Functional Diversity across Oceans'. *Nature Communications* 9 (1): 310. https://doi.org/10.1038/s41467-017-02235-3.

Seemann, T. 2014. 'Prokka: Rapid Prokaryotic Genome Annotation'. *Bioinformatics* 30 (14): 2068–69. https://doi.org/10.1093/bioinformatics/btu153.

Segreto, Rossana, and Yuri Deigin. 2021. 'The Genetic Structure of SARS-CoV-2 Does Not Rule out a Laboratory Origin'. *BioEssays* 43 (3): 2000240. https://doi.org/10.1002/bies.202000240.

Senbill, Haytham, Lakshmi Kanta Hazarika, Aiswarya Baruah, Deepak Kumar Borah, Badal Bhattacharyya, and Sahidur Rahman. 2018. 'Life Cycle of the Southern Cattle Tick, Rhipicephalus (Boophilus) Microplus Canestrini 1888 (Acari: Ixodidae) under Laboratory Conditions'. *Systematic and Applied Acarology* 23 (6): 1169. https://doi.org/10.11158/saa.23.6.12.

Seppey, Mathieu, Mosè Manni, and Evgeny M. Zdobnov. 2019. 'BUSCO: Assessing Genome Assembly and Annotation Completeness'. In *Gene Prediction: Methods and Protocols*, edited by Martin Kollmar, 227–45. Methods in Molecular

Biology. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-9173-0_14.

Serbus, Laura, and Willaim Sullivan. 2007. 'A Cellular Basis for Wolbachia Recruitment to the Host Germline'. *PLOS Pathogens* 3 (12). https://doi.org/10.1371/journal.ppat.0030190.

Seshadri, Rekha, Sinead C. Leahy, Graeme T. Attwood, Koon Hoong Teh, Suzanne C. Lambie, Adrian L. Cookson, Emiley A. Eloe-Fadrosh, *et al.* 2018. 'Cultivation and Sequencing of Rumen Microbiome Members from the Hungate1000 Collection'. *Nature Biotechnology* 36 (4): 359–67. https://doi.org/10.1038/nbt.4110.

Sforza, Stefano, Tullia Tedeschi, Mariangela Bencivenni, Alessandro Tonelli, Roberto Corradini, and Rosangela Marchelli. 2014. 'Use of Peptide Nucleic Acids (PNAs) for Genotyping by Solution and Surface Methods'. In *Peptide Nucleic Acids*, edited by Peter E Nielsen and Daniel H. Appella, 1050:143–57. Methods in Molecular Biology. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-553-8_12.

Shafin, Kishwar, Trevor Pesout, Ryan Lorig-Roach, Marina Haukness, Hugh E. Olsen, Colleen Bosworth, Joel Armstrong, *et al.* 2020. 'Nanopore Sequencing and the Shasta Toolkit Enable Efficient de Novo Assembly of Eleven Human Genomes'. *Nature Biotechnology* 38 (9): 1044–53. https://doi.org/10.1038/s41587-020-0503-6.

Shaw, Dana K., Ann T. Tate, David S. Schneider, Elena A. Levashina, Jonathan C. Kagan, Utpal Pal, Erol Fikrig, and Joao H. F. Pedra. 2018. 'Vector Immunity and Evolutionary Ecology: The Harmonious Dissonance'. *Trends in Immunology* 39 (11): 862–73. https://doi.org/10.1016/j.it.2018.09.003.

Shen, Wei, Shuai Le, Yan Li, and Fuquan Hu. 2016. 'SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation'. *PLOS ONE* 11 (10): e0163962. https://doi.org/10.1371/journal.pone.0163962.

Showler, Allan T, Adalberto Pérez de León, and Perot Saelao. 2021. 'Biosurveillance and Research Needs Involving Area-Wide Systematic Active Sampling to Enhance Integrated Cattle Fever Tick (Ixodida: Ixodidae) Eradication'. Edited by Sarah Hamer. *Journal of Medical Entomology* 58 (4): 1601–9. https://doi.org/10.1093/jme/tjab051.

Shute, P.G. 1966. 'The Staining of Malaria Parasites'. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 60 (3): 412–16. https://doi.org/10.1016/0035-9203(66)90311-7.

Sieracki, M. E., N. J. Poulton, O. Jaillon, P. Wincker, C. de Vargas, L. Rubinat-Ripoll, R. Stepanauskas, R. Logares, and R. Massana. 2019. 'Single Cell Genomics Yields a Wide Diversity of Small Planktonic Protists across Major Ocean Ecosystems'. *Scientific Reports* 9 (1): 6025. https://doi.org/10.1038/s41598-019-42487-1.

Simser, J. A., A. Mulenga, K. R. Macaluso, and A. F. Azad. 2004. 'An Immune Responsive Factor D-like Serine Proteinase Homologue Identified from the American Dog Tick, Dermacentor Variabilis'. *Insect Molecular Biology* 13 (1): 25–35. https://doi.org/10.1111/j.1365-2583.2004.00455.x.

Siraj, A. S., M. Santos-Vega, M. J. Bouma, D. Yadeta, D. Ruiz Carrascal, and M. Pascual. 2014. 'Altitudinal Changes in Malaria Incidence in Highlands of Ethiopia and Colombia'. *Science* 343 (6175): 1154–58. https://doi.org/10.1126/science.1244325.

Skinner, Kalin M., Jacob Underwood, Arnab Ghosh, Adela S. Oliva Chavez, and Corey L. Brelsfoard. 2022. 'Wolbachia Impacts Anaplasma Infection in Ixodes Scapularis Tick Cells'. *International Journal of Environmental Research and Public Health* 19 (3): 1051. https://doi.org/10.3390/ijerph19031051.

Slatko, Barton E., Andrew F. Gardner, and Frederick M. Ausubel. 2018. 'Overview of Next-Generation Sequencing Technologies'. *Current Protocols in Molecular Biology* 122 (1): e59. https://doi.org/10.1002/cpmb.59.

Smith, Alexis A., and Utpal Pal. 2014. 'Immunity-Related Genes in Ixodes Scapularis: Perspectives from Genome Information'. *Frontiers in Cellular and Infection Microbiology* 4 (August). https://doi.org/10.3389/fcimb.2014.00116.

Smith, Katherine F., Michael Goldberg, Samantha Rosenthal, Lynn Carlson, Jane Chen, Cici Chen, and Sohini Ramachandran. 2014. 'Global Rise in Human Infectious Disease Outbreaks'. *Journal of The Royal Society Interface* 11 (101): 20140950. https://doi.org/10.1098/rsif.2014.0950.

Socolovschi, C., T. P. Huynh, B. Davoust, J. Gomez, D. Raoult, and P. Parola. 2009. 'Transovarial and Trans-Stadial Transmission of Rickettsiae Africae in *Amblyomma variegatum* Ticks'. *Clinical Microbiology and Infection* 15 (December): 317–18. https://doi.org/10.1111/j.1469-0691.2008.02278.x.

Söderhäll, Kenneth, ed. 2010. *Invertebrate Immunity*. Advances in Experimental Medicine and Biology, v. 708. New York, N.Y. : Austin, Tex: Springer Science+Business Media ; Landes Bioscience.

Sollner-Webb, Barbara, and Edward Mougey. 1991. 'News from the Nucleolus: RRNA Gene Expression'. *Trends in Biochemical Sciences* 16 (2): 58–62. https://doi.org/doi: 10.1016/0968-0004(91)90025-q.

Solyman, Muna Salem M., Jessica Ujczo, Kelly A. Brayton, Dana K. Shaw, David A. Schneider, and Susan M. Noh. 2022. 'Iron Reduction in Dermacentor Andersoni Tick Cells Inhibits Anaplasma Marginale Replication'. *International Journal of Molecular Sciences* 23 (7): 3941. https://doi.org/10.3390/ijms23073941.

Sonenshine, Daniel E., and Wayne L. Hynes. 2008. 'Molecular Characterization and Related Aspects of the Innate Immune Response in Ticks'. *Frontiers in Bioscience-Landmark* 13 (18): 7046–63. https://doi.org/10.2741/3209.

Spring, H., D. Grierson, V. Hemleben, M. Stöhr, G. Krohne, J. Stadler, and W. W. Franke. 1978. 'DNA Contents and Numbers of Nucleoli and Pre-RRNA-Genes in Nuclei of Gametes and Vegetative Cells of *Acetabularia Mediterranea*'. *Experimental Cell Research* 114 (1): 203–15. https://doi.org/10.1016/0014-4827(78)90054-X.

Stadhouders, Ralph, Suzan D. Pas, Jeer Anber, Jolanda Voermans, Ted H.M. Mes, and Martin Schutten. 2010. 'The Effect of Primer-Template Mismatches on the Detection and Quantification of Nucleic Acids Using the 5′ Nuclease Assay'. *The Journal of Molecular Diagnostics : JMD* 12 (1): 109–17. https://doi.org/10.2353/jmoldx.2010.090035.

Steere, Allen C., Jenifer Coburn, and Lisa Glickstein. 2004. 'The Emergence of Lyme Disease'. *The Journal of Clinical Investigation* 113 (8): 1093–1101. https://doi.org/10.1172/JCI21681.

Stenos, John, Stephen R. Graves, and Nathan B. Unsworth. 2005. 'A Highly Sensitive and Specific Real-Time PCR Assay for the Detection of Spotted Fever and Typhus Group Rickettsiae'. *The American Journal of Tropical Medicine and Hygiene* 73 (6): 1083–85.

Stoeck, Thorsten, David Bass, Markus Nebel, Richard Christen, Meredith D. M. Jones, Hans-Werner Breiner, and Thomas A. Richards. 2010. 'Multiple Marker Parallel Tag Environmental DNA Sequencing Reveals a Highly Complex Eukaryotic Community in Marine Anoxic Water'. *Molecular Ecology* 19 Suppl 1 (March): 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x.

Stone, Brandee L., Yvonne Tourand, and Catherine A. Brissette. 2017. 'Brave New Worlds: The Expanding Universe of Lyme Disease'. *Vector-Borne and Zoonotic Diseases* 17 (9): 619–29. https://doi.org/10.1089/vbz.2017.2127.

Strassert, Jürgen F H, Mahwash Jamy, Alexander P Mylnikov, Denis V Tikhonenkov, and Fabien Burki. 2019. 'New Phylogenomic Analysis of the Enigmatic Phylum *Telonemia* Further Resolves the Eukaryote Tree of Life'. *Molecular Biology and Evolution* 36 (4): 757–65. https://doi.org/10.1093/molbev/msz012.

Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2019. 'Comprehensive Integration of Single-Cell Data'. *Cell* 177 (7): 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

Sturgis, J. N. 2001. 'Organisation and Evolution of the Tol-Pal Gene Cluster'. *Journal of Molecular Microbiology and Biotechnology* 3 (1): 113–22.

Sucgang, Richard, Guokai Chen, Wen Liu, Ryan Lindsay, Jing Lu, Donna Muzny, Gad Shaulsky, William Loomis, Richard Gibbs, and Adam Kuspa. 2003. 'Sequence and Structure of the Extrachromosomal Palindrome Encoding the Ribosomal RNA Genes in *Dictyostelium*'. *Nucleic Acids Research* 31 (9): 2361–68.

Suderman, Michael T., Kevin B. Temeyer, Kristie G. Schlechte, and Adalberto A. Pérez de León. 2021. 'Three-Dimensional Culture of Rhipicephalus (Boophilus) Microplus BmVIII-SCC Cells on Multiple Synthetic Scaffold Systems and in Rotating Bioreactors'. *Insects* 12 (8): 747. https://doi.org/10.3390/insects12080747.

Sutherst, R. W. 1998. 'Implications of Global Change and Climate Variability for Vector-Borne Diseases: Generic Approaches to Impact Assessments'. *International Journal for Parasitology* 28 (6): 935–45. https://doi.org/10.1016/S0020-7519(98)00056-3.

Switaj, Karolina, Tomasz Chmielewski, Piotr Borkowski, Stanislawa Tylewska-Wierzbanowska, and Maria Olszynska-Krowicka. 2012. 'Spotted Fever Rickettsiosis Caused by *Rickettsia raoultii*--Case Report'. *Przeglad Epidemiologiczny* 66 (2): 347–50.

Tabor, Ala E., Abid Ali, Gauhar Rehman, Gustavo Rocha Garcia, Amanda Fonseca Zangirolamo, Thiago Malardo, and Nicholas N. Jonsson. 2017. 'Cattle Tick *Rhipicephalus microplus*-Host Interface: A Review of Resistant and Susceptible Host Responses'. *Frontiers in Cellular and Infection Microbiology* 7 (December): 506. https://doi.org/10.3389/fcimb.2017.00506.

Tafilaku, Jonida, and Sarah Bunn. 2019. 'Climate Change and Vector-Borne Diseases in Humans in the UK'. 597. UK: Parlamentary Office of Science and Technology.

Takken, Willem, Jérémy Bouyer, Renate C. Smallegange, and Claire Garros. 2018. 'Livestock Pests and Vector-Borne Diseases – a Much Neglected Subject'. In *Ecology and Control of Vector-Borne Diseases*, edited by Claire Garros, Jérémy Bouyer, Willem Takken, and Renate C. Smallegange, 5:11–14. The Netherlands: Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-863-6_1.

Tanji, Takahiro, Xiaodi Hu, Alexander N. R. Weber, and Y. Tony Ip. 2007. 'Toll and IMD Pathways Synergistically Activate an Innate Immune Response in Drosophila Melanogaster'. *Molecular and Cellular Biology*, June. https://doi.org/10.1128/MCB.01814-06.

Taylor, Louise H., Sophia M. Latham, and Mark E.J. Woolhouse. 2001. 'Risk Factors for Human Disease Emergence'. Edited by M. E. J. Woolhouse and C. Dye. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 356 (1411): 983–89. https://doi.org/10.1098/rstb.2001.0888.

Tessler, Michael, Mercer R. Brugler, John A. Burns, Nina R. Sinatra, Daniel M. Vogt, Anand Varma, Madelyne Xiao, Robert J. Wood, and David F. Gruber. 2020. 'Ultra-Gentle Soft Robotic Fingers Induce Minimal Transcriptomic Response in a Fragile Marine Animal'. *Current Biology* 30 (4): R157–58. https://doi.org/10.1016/j.cub.2020.01.032.

Thapa, Santosh, Yan Zhang, and Michael S. Allen. 2019. 'Bacterial Microbiomes of Ixodes Scapularis Ticks Collected from Massachusetts and Texas, USA'. *BMC Microbiology* 19 (1): 138. https://doi.org/10.1186/s12866-019-1514-7.

Thongprem, Panupong, Sophie E. F. Evison, Gregory D. D. Hurst, and Oliver Otti. 2020. 'Transmission, Tropism, and Biological Impacts of Torix Rickettsia in the Common Bed Bug Cimex Lectularius (Hemiptera: Cimicidae)'. *Frontiers in Microbiology* 11. https://www.frontiersin.org/articles/10.3389/fmicb.2020.608763.

Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov. 2013. 'Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration'. *Briefings in Bioinformatics* 14 (2): 178–92. https://doi.org/10.1093/bib/bbs017.

'Tick Cell Lines - Liverpool Shared Research Facilities - University of Liverpool'. 2021. November 2021. https://www.liverpool.ac.uk/liverpool-shared-research-facilities/facilities/bio-resources/tick-cell-biobank/tick-cell-lines/.

Torresen, Ole K., Bastiaan Star, Sissel Jentoft, William B. Reinar, Harald Grove, Jason R. Miller, Brian P. Walenz, *et al.* 2017. 'An Improved Genome Assembly Uncovers Prolific Tandem Repeats in Atlantic Cod'. *BMC Genomics* 18 (1): 95. https://doi.org/10.1186/s12864-016-3448-x.

Torres-Machorro, Ana Lilia, Roberto Hernández, Ana María Cevallos, and Imelda López-Villaseñor. 2010. 'Ribosomal RNA Genes in Eukaryotic Microorganisms: Witnesses of Phylogeny?' *FEMS Microbiology Reviews* 34 (1): 59–86. https://doi.org/10.1111/j.1574-6976.2009.00196.x.

Trager, William. 1939. 'Acquired Immunity to Ticks'. *The Journal of Parasitology* 25 (1): 57. https://doi.org/10.2307/3272160.

Tully, Joseph G., David L. Rose, Conrad E. Yunker, Jack Cory, Robert F. Whitcomb, and David L. Williamson. 1981. 'Helical Mycoplasmas (Spiroplasmas) from Ixodes Ticks'. *Science* 212 (4498): 1043–45. https://doi.org/10.1126/science.7233197.

Turner, D. A., and T. K. Golder. 1987. 'Is Resistance Developing in Tsetse Flies? Susceptibility to Two Chlorinated Hydrocarbon Insecticides in Sprayed and Unsprayed Populations of Glossina Pallidipes in Kenya'. In *Integrated Tse-Tse Fly Control: Methods and Strategies*. CRC Press.

Ullmann, A. J., C. M. R. Lima, F. D. Guerrero, J. Piesman, and W. C. Black. 2005. 'Genome Size and Organization in the Blacklegged Tick, Ixodes Scapularis and the Southern Cattle Tick, Boophilus Microplus'. *Insect Molecular Biology* 14 (2): 217–22. https://doi.org/10.1111/j.1365-2583.2005.00551.x.

Ulrich, Jill N., John C. Beier, Gregor J. Devine, and Leon E. Hugo. 2016. 'Heat Sensitivity of WMel Wolbachia during Aedes Aegypti Development'. *PLoS Neglected Tropical Diseases* 10 (7): e0004873. https://doi.org/10.1371/journal.pntd.0004873.

Valzano, Matteo, Valentina Cecarini, Alessia Cappelli, Aida Capone, Jovana Bozic, Massimiliano Cuccioloni, Sara Epis, *et al.* 2016. 'A Yeast Strain Associated to Anopheles Mosquitoes Produces a Toxin Able to Kill Malaria Parasites'. *Malaria Journal* 15 (1): 21. https://doi.org/10.1186/s12936-015-1059-7.

Van Dyken, Meg, Bethany G. Bolling, Chester G. Moore, Carol D. Blair, Barry J. Beaty, William C. Black, and Brian D. Foy. 2006. 'Molecular Evidence for Trypanosomatids in Culex Mosquitoes Collected during a West Nile Virus Survey'. *International Journal for Parasitology* 36 (9): 1015–23. https://doi.org/10.1016/j.ijpara.2006.05.003.

Vandyk, John K., David M. Bartholomew, Wayne A. Rowley, and Kenneth B. Platt. 1996. 'Survival of Ixodes Scapularis (Acari: Ixodidae) Exposed to Cold'. *Journal of Medical Entomology* 33 (1): 6–10. https://doi.org/10.1093/jmedent/33.1.6.

Varma, M. G. R., Mary Pudney, and C. J. Leake. 1975. 'The Establishment of Three Cell Lines from the Tick Rhipicephalus Appendiculatus (Agari: Ixodidae) and Their Infection with Some Arboviruses'. *Journal of Medical Entomology* 11 (6): 698–706. https://doi.org/10.1093/jmedent/11.6.698.

Vaser, Robert, and Mile Šikić. 2021. 'Time- and Memory-Efficient Genome Assembly with Raven'. *Nature Computational Science* 1 (5): 332–36. https://doi.org/10.1038/s43588-021-00073-4.

Vaser, Robert, Ivan Sović, Niranjan Nagarajan, and Mile Šikić. 2017. 'Fast and Accurate de Novo Genome Assembly from Long Uncorrected Reads'. *Genome Research* 27 (5): 737–46. https://doi.org/10.1101/gr.214270.116.

Vechtova, Pavlina, Zoltan Fussy, Radim Cegan, Jan Sterba, Jan Erhart, Vladimir Benes, and Libor Grubhoffer. 2020. 'Catalogue of Stage-Specific Transcripts in Ixodes Ricinus and Their Potential Functions during the Tick Life-Cycle'. *Parasites & Vectors* 13 (1): 311. https://doi.org/10.1186/s13071-020-04173-4.

Velichko, Artem K., Nadezhda V. Petrova, Omar L. Kantidze, and Sergey V. Razin. 2012. 'Dual Effect of Heat Shock on DNA Replication and Genome Integrity'. Edited by Mark J. Solomon. *Molecular Biology of the Cell* 23 (17): 3450–60. https://doi.org/10.1091/mbc.e11-12-1009.

Vernette, Caroline, Nicolas Henry, Julien Lecubin, Colomban Vargas, Pascal Hingamp, and Magali Lescot. 2021. 'The Ocean Barcode Atlas: A Web Service to Explore the Biodiversity and Biogeography of Marine Organisms'. *Molecular Ecology Resources* 21 (4): 1347–58. https://doi.org/10.1111/1755-0998.13322.

Vestheim, Hege, Bruce E. Deagle, and Simon N. Jarman. 2011. 'Application of Blocking Oligonucleotides to Improve Signal-to-Noise Ratio in a PCR'. In *PCR Protocols*, edited by Daniel J. Park, 265–74. Methods in Molecular Biology. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-60761-944-4_19.

Vestheim, Hege, and Simon N Jarman. 2008. 'Blocking Primers to Enhance PCR Amplification of Rare Sequences in Mixed Samples – a Case Study on Prey DNA in Antarctic Krill Stomachs'. *Frontiers in Zoology* 5 (1): 12. https://doi.org/10.1186/1742-9994-5-12.

Vicoso, Beatriz, and Doris Bachtrog. 2015. 'Numerous Transitions of Sex Chromosomes in Diptera'. Edited by Harmit S. Malik. *PLOS Biology* 13 (4): e1002078. https://doi.org/10.1371/journal.pbio.1002078.

Vreysen, Marc J. B., Khalfan M. Saleh, Mashavu Y. Ali, Abdulla M. Abdulla, Zeng-Rong Zhu, Kassim G. Juma, V. Arnold Dyck, Atway R. Msangi, Paul A. Mkonyi, and H. Udo Feldmann. 2000. 'Glossina Austeni (Diptera: Glossinidae) Eradicated on the Island of Unguja, Zanzibar, Using the Sterile Insect Technique'. *Journal of Economic Entomology* 93 (1): 123–35. https://doi.org/10.1603/0022-0493-93.1.123.

Walker, Bruce J., Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, *et al.* 2014. 'Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement'. Edited by Junwen Wang. *PLoS ONE* 9 (11): e112963. https://doi.org/10.1371/journal.pone.0112963.

Walker, J. B., and A. Olwage. 1987. 'The Tick Vectors of Cowdria Ruminantium (Ixodoidea, Ixodidae, Genus Amblyomma) and Their Distribution'. *The Onderstepoort Journal of Veterinary Research* 54 (3): 353–79.

Wang, Yong, Ren Mao Tian, Zhao Ming Gao, Salim Bougouffa, and Pei-Yuan Qian. 2014. 'Optimal Eukaryotic 18S and Universal 16S/18S Ribosomal RNA Primers and Their Application in a Study of Symbiosis'. Edited by Newton C. M. Gomes. *PLoS ONE* 9 (3): e90053. https://doi.org/10.1371/journal.pone.0090053.

Weiss, Brian L., Jingwen Wang, and Serap Aksoy. 2011. 'Tsetse Immune System Maturation Requires the Presence of Obligate Symbionts in Larvae'. *PLOS Biology* 9 (5): e1000619. https://doi.org/10.1371/journal.pbio.1000619.

Welburn, Susan C, Eric M Fèvre, Paul G Coleman, Martin Odiit, and Ian Maudlin. 2001. 'Sleeping Sickness: A Tale of Two Diseases'. *Trends in Parasitology* 17 (1): 19–24. https://doi.org/10.1016/s1471-4922(00)01839-0.

Wick, Ryan R., Louise M. Judd, Claire L. Gorrie, and Kathryn E. Holt. 2017. 'Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads'. Edited by Adam M. Phillippy. *PLOS Computational Biology* 13 (6): e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Willadsen, P., G. A. Riding, R. V. McKenna, D. H. Kemp, R. L. Tellam, J. N. Nielsen, J. Lahnstein, G. S. Cobon, and J. M. Gough. 1989. 'Immunologic Control of a Parasitic Arthropod. Identification of a Protective Antigen from Boophilus Microplus'. *Journal of Immunology (Baltimore, Md.: 1950)* 143 (4): 1346–51.

Williams, Bryony A. P., Kristina M. Hamilton, Meredith D. Jones, and David Bass. 2018. 'Group-Specific Environmental Sequencing Reveals High Levels of Ecological Heterogeneity across the Microsporidian Radiation'. *Environmental Microbiology Reports* 10 (3): 328–36. https://doi.org/10.1111/1758-2229.12642.

Williams, Marni, and Richard Baxter. 2014. 'The Structure and Function of Thioester-Containing Proteins in Arthropods'. *Biophysical Reviews* 6 (3–4): 261–72. https://doi.org/10.1007/s12551-014-0142-6.

Williamson, David L., Gail E. Gasparich, Laura B. Regassa, Collette Saillard, Joël Renaudin, Joseph M. Bové, and Robert F. Whitcomb. 2015. 'Spiroplasma'. In *Bergey's Manual of Systematics of Archaea and Bacteria*, edited by William B Whitman, Fred Rainey, Peter Kämpfer, Martha Trujillo, Jonsik Chun, Paul DeVos, Brian Hedlund, and Svetlana Dedysh, 1–46. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118960608.gbm01262.

Wintzingerode, F. von, O. Landt, A. Ehrlich, and U. B. Gobel. 2000. 'Peptide Nucleic Acid-Mediated PCR Clamping as a Useful Supplement in the Determination of Microbial Diversity'. *Applied and Environmental Microbiology* 66 (2): 549–57. https://doi.org/10.1128/AEM.66.2.549-557.2000.

Woese, C. R., and G. E. Fox. 1977. 'Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms'. *Proceedings of the National Academy of Sciences* 74 (11): 5088–90. https://doi.org/10.1073/pnas.74.11.5088.

Wood, Derrick E., Jennifer Lu, and Ben Langmead. 2019. 'Improved Metagenomic Analysis with Kraken 2'. *Genome Biology* 20 (1): 257. https://doi.org/10.1186/s13059-019-1891-0.

World Health Organization. 2020. 'Vector-Borne Diseases'. 2020.

'World Malaria Report 2020: 20 Years of Global Progress and Challenges'. 2020. In . Geneva: World Health Organization.

Yilmaz, Pelin, Laura Wegener Parfrey, Pablo Yarza, Jan Gerken, Elmar Pruesse, Christian Quast, Timmy Schweer, Jörg Peplies, Wolfgang Ludwig, and Frank Oliver Glöckner. 2014. 'The SILVA and "All-Species Living Tree Project (LTP)" Taxonomic Frameworks'. *Nucleic Acids Research* 42 (D1): D643–48. https://doi.org/10.1093/nar/gkt1209.

Yonow, Tania. 1995. 'The Life-Cycle of *Amblyomma variegatum* (Acari: Ixodidae): A Literature Synthesis with a View to Modelling'. *International Journal for Parasitology* 25 (9): 1023–60. https://doi.org/10.1016/0020-7519(95)00020-3.

Zhu, Pingfen, Paul A. Garber, Ling Wang, Meng Li, Katherine Belov, Thomas R. Gillespie, and Xuming Zhou. 2020. 'Comprehensive Knowledge of Reservoir Hosts Is Key to Mitigating Future Pandemics'. *The Innovation* 1 (3): 100065. https://doi.org/10.1016/j.xinn.2020.100065.

Zug, Roman, and Peter Hammerstein. 2012. 'Still a Host of Hosts for Wolbachia: Analysis of Recent Data Suggests That 40% of Terrestrial Arthropod Species Are Infected'. *PLOS ONE* 7 (6): e38544. https://doi.org/10.1371/journal.pone.0038544.

Zweygarth, E., H. Schöl, K. Lis, A. Cabezas Cruz, C. Thiel, C. Silaghi, M. F. B. Ribeiro, and L. M. F. Passos. 2013. '*In Vitro* Culture of a Novel Genotype of *Ehrlichia* Sp. from Brazil'. *Transboundary and Emerging Diseases* 60 (November): 86–92. https://doi.org/10.1111/tbed.12122.

Appendices

Digital supplementary material

You can access the digital materials by:

- clicking the active link to shared Dropbox folder
- typing a short link in the address bar of your browser: bit.ly/35fIDqE
- scanning the QR code

Chapter II Supplementary materials

Supplementary table 2 List of DNA extractions used in the experiments in Chapter II. All samples were from single specimens unless otherwise indicated. Where indicated, DNA samples were extracted by colleagues at the University of Liverpool.

Sample ID	**species**	**Eukaryotic symbionts**	**sex**	**tissue**	**extraction method**	**ng/ul Nanodrop**	**ug/ml QUBIT**	**notes**
G1	*Glossina morsitans*	“Gregarines”	F	whole fly with blood	CTAB	-	379.00	-
G2	*G. morsitans*	“Gregarines”	M	whole fly with blood	CTAB	571.00	309.00	-
G3	*G. morsitans*	“Gregarines”	M	whole fly	phenol-chloroform	147.00	169.00	-
G6a	*G. morsitans*	“Gregarines”	F	abdomen lining w/o intestine	CTAB	468.00	-	-
G6b	*G. morsitans*	“Gregarines”	F	intestine with some blood	CTAB	74.00	25.10	-
G13	*G. morsitans*	"Gregarines"	F	intestine	CTAB	264.00	186.00	-
G14	*G. morsitans*	“Gregarines”	F	intestine	phenol-chloroform	33.00	10.10	-
G16	*G. morsitans*	“Gregarines”	F	intestine	CTAB	1259.00	355.00	-
G19	*G. morsitans*	"Gregarines", *Trypanosoma brucei*	F	tissues with *T. brucei* and blood	phenol-chloroform	75.00	27.00	-

G20	*G. morsitans*	"Gregarines", *T. brucei*	mixed	salivary glands from several flies	phenol-chloroform	18.00	1.49	-
W20	*G. morsitans*	"Gregarines", *T. brucei*	mixed	Whole genome amplified of G20 DNA	n/a	450	37.6	-
G24	*G. morsitans*	10 laser dissected cells of "Gregarines"	F	n/a	DNeasy Blood & Tissue kit	Below Nanodrop range	0.07	-
W24	*G. morsitans*	10 laser dissected cells of "Gregarines"	F	Whole genome amplified of G24 DNA	n/a	397	168.00	-
TPf8	*Glossina medicorum*	Unknown	F	Whole fly	DNeasy Blood & Tissue kit	-	20.60	DNA extracted by Tom Palmer
TPm11	*Glossina palpalis*	Unknown	M	Whole fly	DNeasy Blood & Tissue kit	-	16.20	DNA extracted by Tom Palmer
TPf14	*G. palpalis*	Unknown	F	Whole fly	DNeasy Blood & Tissue kit	-	36.20	DNA extracted by Tom Palmer
TPf16	*Glossina tabaniformes*	Unknown	F	Whole fly	DNeasy Blood & Tissue kit	-	16.00	DNA extracted by Tom Palmer
TPf18	*G. tabaniformes*	Unknown	F	Whole fly	DNeasy Blood & Tissue kit	-	19.30	DNA extracted by Tom Palmer

HSKm4	*G. palpalis*	Unknown	M	Whole fly	DNeasy Blood & Tissue kit	-	40.90	DNA extracted by Hafsa XXX
HSKm21	*G. palpalis*	Unknown	M	Whole fly	DNeasy Blood & Tissue kit	-	19.00	DNA extracted by Hafsa XXX
HSKf21	*G. palpalis*	Trypanosoma sp.	F	Whole fly	DNeasy Blood & Tissue kit	-	35.10	DNA extracted by Hafsa XXX
Hm17	*G. palpalis*	Unknown	M	Whole fly	DNeasy Blood & Tissue kit	-	16.80	DNA extracted by Hafsa XXX
I38	*Ixodes ricinus*	Unknown	F	Internal organs	DNeasy Blood & Tissue kit	-	4.66	DNA extracted by Yichuan Bao
K10	*Lipoptena cervi*	Unknown	F	Whole fly	DNeasy Blood & Tissue kit	121.5	10.7	DNA extracted by Emily Podmore
Meuph	*Macrosiphum euphorbiae*	Unknown	n/a	Ten adult aphids	Zymo gDNA miniprep Kit	-	17.80	DNA extracted by Dr Mark Whitehead
Ccapi	*Ceratitis capitata*	Unknown	Unknown	Whole fly	DNeasy Blood & Tissue kit	-	17.00	DNA extracted by Dr Mark Whitehead
Umay	*Ustilago maydis*	No	n/a	Cell culture	DNeasy Blood & Tissue kit	-	18.90	DNA extracted by Dr Stefany Solano-Gonzales

Chapter III Supplementary materials

Supplementary table 3 Glossina morsitans assembly characteristics and BUSCO scores. The table describes working assemblies at different stages of assembly and polishing. The scores of the final assembly are given in the Chapter III results.

	Number of contigs	Total length, bp	Largest contig, bp	GC, %	N50, bp	L50	N's per 100 Kb	Complete BUSCOs	Single-copy BUSCOs	Duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs
Flye asm	708	370,720,840	30,116,467	34.53	9,302,271	12	0.62	3177	3126	51	53	55
Flye asm, Medaka polish	723	371,237,569	24,972,289	34.52	8,621,799	13	0	3220	3165	55	29	36
Flye asm, Medaka and 1 round of Racon polish	723	371,015,312	24,942,703	34.49	8,611,007	13	0.01	3260	3201	59	8	17
Flye asm, Medaka, 1 round of Racon and 1 round of Pilon polish	723	370,835,829	24,922,836	34.49	8,605,577	13	0.01	3257	3198	59	10	18
Flye asm, Medaka, 1 round of Racon and 4 rounds of Pilon polish	723	370,754,599	24,923,714	34.49	8,605,905	13	0.01	3258	3198	60	9	18
Flye asm, Medaka, 1 round of Racon and 7 rounds of Pilon polish	723	370,704,018	24,922,338	34.49	8,605,315	13	0.01	3259	3200	59	9	17
Flye asm, Medaka, 1 round of Racon and 10 rounds of Pilon polish	723	370,691,536	24,923,698	34.49	8,605,907	13	0.01	3258	3199	59	9	18
Flye asm, Medaka, 1 round of Racon and 13 rounds of Pilon polish	723	370,653,901	24,922,384	34.49	8,605,315	13	0.01	3258	3199	59	9	18
Flye asm, Medaka, 1 round of Racon and 15 rounds of Pilon polish	723	370,644,355	24,922,386	34.49	8,605,312	13	0.01	3258	3201	57	9	18
Flye asm, Medaka and 2 rounds of Racon polish	723	370,929,740	24,936,105	34.49	8,609,262	13	0.01	3255	3194	61	12	18

Flye asm, Medaka and 3 rounds of Racon polish	723	370,880,668	24,932,734	34.49	8,608,833	13	0.01	3257	3196	61	9	19
Flye asm, Medaka, 3 rounds of Racon and 1 round of Pilon polish	723	370,737,313	24,914,199	34.49	8,604,667	13	0.01	3258	3199	59	10	17
Flye asm, Medaka, 3 rounds of Racon and 2 rounds of Pilon polish	723	370,704,328	24,914,853	34.49	8,604,872	13	0.01	3258	3199	59	10	17
Flye asm, Medaka, 3 rounds of Racon and 3 rounds of Pilon polish	723	370,704,328	24,914,853	34.49	8,604,872	13	0.01	3258	3199	59	10	17
Flye asm, Medaka, 3 rounds of Racon and 4 rounds of Pilon polish	723	370,704,328	24,914,853	34.49	8,604,872	13	0.01	3258	3199	59	10	17
Flye asm, Medaka, 3 rounds of Racon and 5 rounds of Pilon polish	723	370,704,328	24,914,853	34.49	8,604,872	13	0.01	3258	3199	59	10	17
Flye asm, Medaka and 4 rounds of Racon polish	723	370,845,365	24,931,989	34.49	8,608,518	13	0.01	3258	3199	59	8	19
Flye asm, Medaka and 5 rounds of Racon polish	723	370,827,419	24,930,732	34.49	8,608,541	13	0.01	3258	3197	61	8	19
Raven asm	941	388,584,165	20,010,946	34.37	3,246,851	32	0	3177	3135	42	52	56
Raven asm, Medaka polish	1213	387,052,166	20,058,138	34.47	2,910,651	35	0	3223	3166	57	23	39
Raven asm, Medaka and 1 round of Racon polish	1213	386,643,635	20,036,519	34.46	2,908,532	35	0.01	3257	3197	60	8	20
Raven asm, Medaka and 2 rounds of Racon polish	1213	386,430,761	20,031,947	34.46	2,906,687	35	0.01	3258	3196	62	8	19
Raven asm, Medaka and 3 rounds of Racon polish	1213	386,288,383	20,030,577	34.46	2,906,492	35	0.01	3257	3196	61	8	20
Raven asm, Medaka and 4 rounds of Racon polish	1213	386,191,376	20,030,120	34.46	2,992,243	34	0.01	3259	3197	62	7	19
Raven asm, Medaka and 5 rounds of Racon polish	1213	386,128,656	20,029,236	34.46	2,992,031	34	0.01	3258	3197	61	8	19
wtdbg2 (redbean) asm	2600	381,011,320	25,624,761	34.45	3,921,972	20	0	2855	2823	32	162	268
wtdbg2 asm, Medaka polish	2543	375,687,188	15,924,437	34.6	3,574,665	25	0	3101	3044	57	50	134
Shasta asm	1891	320,133,519	10,591,914	34.41	2,061,877	42	0	2352	2330	22	247	686
Shasta asm, Medaka polish	1757	318,555,427	10,549,191	34.62	2,053,215	42	0	2759	2722	37	90	436

Supplementary table 4 List of Rickettsia spp. complete assemblies used for the analysis

1	NC_003103.1	Rickettsia conorii str. Malish 7

2	NC_006142.1	Rickettsia typhi str. Wilmington
3	NC_007109.1	Rickettsia felis URRWXCal2
4	NC_007940.1	Rickettsia bellii RML369-C
5	NC_009879.1	Rickettsia canadensis str. McKiel
6	NC_009900.1	Rickettsia massiliae MTU5
7	NC_010263.3	Rickettsia rickettsii str. Iowa
8	NC_009881.1	Rickettsia akari str. Hartford
9	NC_009882.1	Rickettsia rickettsii str. 'Sheila Smith'
10	NC_009883.1	Rickettsia bellii OSU 85-389
11	NC_012730.1	Rickettsia peacockii str. Rustic
12	NC_017560.1	Rickettsia prowazekii str. Rp22
13	NC_012633.1	Rickettsia africae ESF-5
14	NC_000963.1	Rickettsia prowazekii str. Madrid E
15	NC_015866.1	Rickettsia heilongjiangensis 054
16	NC_016639.1	Rickettsia slovaca 13-B
17	NC_017065.1	Rickettsia slovaca str. D-CWPP
18	NC_017049.1	Rickettsia prowazekii str. Chernikova
19	NC_017050.1	Rickettsia prowazekii str. Katsinyian
20	NC_017056.1	Rickettsia prowazekii str. BuV67-CWPP
21	NC_017048.1	Rickettsia prowazekii str. GvV257
22	NC_017057.1	Rickettsia prowazekii str. RpGvF24
23	NC_017066.1	Rickettsia typhi str. TH1527
24	NC_017062.1	Rickettsia typhi str. B9991CWPP
25	NC_016050.1	Rickettsia japonica YH
26	NC_016908.1	Rickettsia rickettsii str. Colombia
27	NC_016909.1	Rickettsia rickettsii str. Arizona
28	NC_016914.1	Rickettsia rickettsii str. Hino

29	NC_016929.1	Rickettsia canadensis str. CA410
30	NC_016915.1	Rickettsia rickettsii str. Hlp#2
31	NC_016913.1	Rickettsia rickettsii str. Brazil
32	NC_016930.1	Rickettsia philipii str. 364D
33	NC_017028.1	Rickettsia amblyommatis str. GAT-30V
34	NC_017042.1	Rickettsia rhipicephali str. 3-7-female6-CWPP
35	NC_017058.1	Rickettsia australis str. Cutlack
36	NC_017043.1	Rickettsia montanensis str. OSU 85-930
37	NC_017044.1	Rickettsia parkeri str. Portsmouth
38	NC_020992.1	Rickettsia prowazekii str. NMRC Madrid E
39	NC_020993.1	Rickettsia prowazekii str. Breinl
40	NZ_LN794217.1	Rickettsia monacensis strain IrR/Munich
41	NZ_CP006009.1	Rickettsia rickettsii str. R
42	NZ_CP006010.1	Rickettsia rickettsii str. Morgan
43	NZ_CP012420.1	Rickettsia amblyommatis strain Ac37
44	NZ_CP013133.1	Rickettsia rhipicephali strain HJ#5
45	NZ_CP014865.1	Rickettsia prowazekii strain Naples-1
46	NZ_CP018913.1	Rickettsia rickettsii strain Iowa isolate Large Clone
47	NZ_CP018914.1	Rickettsia rickettsii strain Iowa isolate Small Clone

48	NZ_CP015010.1	Rickettsia bellii isolate An04
49	NZ_CP015012.1	Rickettsia amblyommatis isolate An13
50	NZ_CP016305.1	Rickettsia sp. MEAM1 (Bemisia tabaci)
51	NZ_AP017602.1	Rickettsia japonica strain YH_M
52	NZ_AP017581.1	Rickettsia japonica strain Tsuneishi
53	NZ_AP017580.1	Rickettsia japonica strain PO-1
54	NZ_AP017579.1	Rickettsia japonica strain OHH-1
55	NZ_AP017578.1	Rickettsia japonica strain Nakase
56	NZ_AP017577.1	Rickettsia japonica strain MZ08014
57	NZ_AP017576.1	Rickettsia japonica strain HH07167
58	NZ_AP017575.1	Rickettsia japonica strain HH07124
59	NZ_AP017574.1	Rickettsia japonica strain HH06154
60	NZ_AP017573.1	Rickettsia japonica strain HH-1
61	NZ_AP017572.1	Rickettsia japonica strain DT-1

62	NZ_CP032049.1	Rickettsia japonica strain LA4/2015
63	NZ_CP040325.1	Rickettsia parkeri strain Atlantic Rainforest
64	NZ_AP019563.1	Rickettsia asiatica strain Maytaro1284
65	NZ_AP019864.1	Rickettsia heilongjiangensis strain Sendai-29
66	NZ_AP019865.1	Rickettsia heilongjiangensis strain Sendai-58
67	NZ_AP019863.1	Rickettsia heilongjiangensis strain HCN-13
68	NZ_AP019862.1	Rickettsia heilongjiangensis strain CH8-1
69	NZ_CP047359.1	Rickettsia japonica strain LA16/2015
70	NZ_LS992663.1	Rickettsia typhi isolate TM2540

Supplementary table 5 List of Wolbachia spp. complete assemblies used for the analysis

1	NC_002978.6	*Wolbachia* of *Drosophila melanogaster*
2	NC_006833.1	*Wolbachia* strain TRS of Brugia malayi
3	NC_012416.1	*Wolbachia* sp. wRi
4	NC_010981.1	*Wolbachia* of Culex quinquefasciatus Pel
5	NC_018267.1	*Wolbachia* of Onchocerca ochengi
6	NC_021084.1	*Wolbachia* of Drosophila simulans wNo
7	NC_021089.1	*Wolbachia* of Drosophila simulans wHa
8	NZ_HG810405.1	*Wolbachia* of Onchocerca volvulus str. Cameroon
9	NZ_AP013028.1	*Wolbachia* of Cimex lectularius strain wCle
10	NZ_LK055284.1	*Wolbachia* of Drosophila simulans wAu genome assembly
11	NZ_CM003641.1	*Wolbachia* of Trichogramma pretiosum strain wTpre
12	NZ_CP011148.1	*Wolbachia* of Drosophila incompta strain wInc_Cu
13	NZ_CP015510.2	*Wolbachia* of Folsomia candida strain Berlin
14	NZ_CP041924.1	*Wolbachia* pipientis strain wAlbB-HN2016
15	NZ_CP041923.1	*Wolbachia* pipientis strain wAlbB-FL2016
16	NZ_CP016430.1	*Wolbachia* of Bemisia tabaci strain China 1
17	NZ_CP031221.1	*Wolbachia* pipientis wAlbB
18	NZ_CP034333.1	*Wolbachia* of Brugia malayi isolate TRS
19	NZ_CP034334.1	*Wolbachia* of Drosophila mauritiana strain wMau
20	NZ_CP034335.1	*Wolbachia* of Drosophila mauritiana strain wMau
21	NZ_CP041215.1	*Wolbachia* of Carposina sasakii isolate wCauA
22	NZ_CP042446.1	*Wolbachia* pipientis strain wMel_N25
23	NZ_CP042444.1	*Wolbachia* pipientis strain wMel_I23
24	NZ_CP042445.1	*Wolbachia* pipientis strain wMel_ZH26

25	NZ_CP042904.1	*Wolbachia* of Drosophila ananassae strain W2.1
26	NZ_CP021120.1	*Wolbachia* of Chrysomya megacephala isolate wMeg
27	NZ_CP037426.1	*Wolbachia* pipientis strain wIrr
28	NZ_CP050521.1	*Wolbachia* of Brugia pahangi isolate FR3
29	NZ_CP051156.1	*Wolbachia* of Ctenocephalides felis wCfeT
30	NZ_CP051157.1	*Wolbachia* of Ctenocephalides felis wCfeJ
31	NZ_CP051264.2	*Wolbachia* of Diaphorina citri isolate KPSwDI15P40
32	NZ_CP051265.2	*Wolbachia* of Di. citri isolate KPSwDI10P38
33	NZ_CP051266.2	*Wolbachia* of Di. citri isolate KPSwDI05P26
34	NZ_CP046577.1	*Wolbachia* of *Litomosoides sigmodontis* strain wLsig
35	NZ_CP046578.1	*Wolbachia* of Dirofilaria (Dirofilaria) immitis strain FR3
36	NZ_CP046579.1	*Wolbachia* of Cruorifilaria tuberocauda strain 55YT
37	NZ_CP046580.1	*Wolbachia* of *Dipetalonema caudispina* strain 362YU
38	NZ_CP051608.1	*Wolbachia* of *Di. citri* isolate dawsonii
39	NZ_CP050530.1	*Wolbachia pipientis* isolate wNik
40	NZ_CP050531.1	*Wolbachia pipientis* isolate wStv
41	NZ_CP061738.1	Candidatus *Wolbachia massiliensis* isolate PL13
42	NZ_CP046921.1	*Wolbachia* of *Drosophila melanogaster* isolate wMelpop
43	NZ_CP046922.1	*Wolbachia* of *D. melanogaster* isolate wMelPop2
44	NZ_CP046923.1	*Wolbachia* of *D. melanogaster* isolate wMelOctoless
45	NZ_CP046924.1	*Wolbachia* of *D. melanogaster* isolate wMelCS_b
46	NZ_CP046925.1	*Wolbachia* of *D. melanogaster* isolate wMel

Chapter V Supplementary materials

Supplementary table 6 Differentially expressed genes in Ixodes scapularis cell samples (I. scapularis with endogenous SCRV and infected with Ehrlichia minasensis). "Product" reflects on the gene function, "Gene" stands for the gene ID in the RefSeq annotation ('Em' genes come from the E. minasensis genome, 'MT' – tick mitochondrial genes, 'LOC' – nuclear tick genome), "avg_log2FC" describes change in expression level of the gene in the sample compared with the other sample, "pct.1" means the fraction of cells in the current sample in which the gene was detected, "pct.2" means the fraction of cells in the other samples in which the gene was detected, "p_val" – adjusted p-value, based on Bonferroni correction using all genes in the dataset. Genes are sorted by "Sample" and "avg_log2FC" in descending order. Lower log2FC threshold was 0.5.

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
23S ribosomal RNA	Em00309	*I. scapularis*/SCRV inf. *E. minasensis*	4.60	0.28	0.00	0
16S ribosomal RNA	Em00696	*I. scapularis*/SCRV inf. *E. minasensis*	3.83	0.27	0.00	0
growth/differentiation factor 8	LOC115329081	*I. scapularis*/SCRV inf. *E. minasensis*	1.24	0.72	0.36	0
actin-5C	LOC8027393	*I. scapularis*/SCRV inf. *E. minasensis*	0.84	1.00	1.00	0
mitochondrial ribosomal RNA rrnS	MT-rrnS	*I. scapularis*/SCRV inf. *E. minasensis*	0.81	1.00	0.99	0
MAP kinase-interacting serine/threonine-protein kinase 1	LOC8043997	*I. scapularis*/SCRV inf. *E. minasensis*	0.81	1.00	0.99	0
uncharacterized LOC121834122	LOC121834122	*I. scapularis*/SCRV inf. *E. minasensis*	0.79	0.85	0.76	3.29E-227
protein NDRG3	LOC8035131	*I. scapularis*/SCRV inf. *E. minasensis*	0.76	0.97	0.93	3.22E-273

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
eukaryotic peptide chain release factor GTP-binding subunit ERF3A	LOC8050870	*I. scapularis*/SCRV inf. *E. minasensis*	0.73	0.93	0.86	0
very low-density lipoprotein receptor	LOC8038934	*I. scapularis*/SCRV inf. *E. minasensis*	0.71	0.99	0.98	0
uncharacterized LOC120842277	LOC120842277	*I. scapularis*/SCRV inf. *E. minasensis*	0.71	0.65	0.56	5.94E-117
FYVE and coiled-coil domain-containing protein 1	LOC8030671	*I. scapularis*/SCRV inf. *E. minasensis*	0.70	0.90	0.80	7.16E-241
solute carrier family 2, facilitated glucose transporter member 1	LOC8043339	*I. scapularis*/SCRV inf. *E. minasensis*	0.69	0.89	0.78	1.26E-247
terminal nucleotidyltransferase 5B	LOC8053604	*I. scapularis*/SCRV inf. *E. minasensis*	0.61	0.97	0.91	0
gamma-butyrobetaine dioxygenase	LOC8038768	*I. scapularis*/SCRV inf. *E. minasensis*	0.61	0.71	0.47	6.86E-191
glutamine synthetase	LOC8051141	*I. scapularis*/SCRV inf. *E. minasensis*	0.60	0.56	0.29	2.95E-252
sestrin homolog	LOC121836714	*I. scapularis*/SCRV inf. *E. minasensis*	0.58	0.98	0.95	2.79E-256
mitochondrial ribosomal RNA rrnL	MT-rrnL	*I. scapularis*/SCRV inf. *E. minasensis*	0.58	1.00	1.00	1.14E-251
innexin inx2	LOC115320817	*I. scapularis*/SCRV inf. *E. minasensis*	0.57	0.73	0.55	1.45E-198
ATP-dependent RNA helicase dbp2	LOC8033189	*I. scapularis*/SCRV inf. *E. minasensis*	0.56	0.91	0.82	1.57E-292
growth factor receptor-bound protein 14	LOC8025683	*I. scapularis*/SCRV inf. *E. minasensis*	0.56	0.89	0.80	7.98E-206

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
calreticulin	LOC120846343	*I. scapularis*/SCRV inf. *E. minasensis*	0.56	1.00	1.00	5.23E-216
tubulin-folding cofactor B	LOC8035307	*I. scapularis*/SCRV inf. *E. minasensis*	0.56	0.86	0.71	8.46E-262
AF4/FMR2 family member 1-like	LOC121835564	*I. scapularis*/SCRV inf. *E. minasensis*	0.55	0.80	0.66	7.81E-211
nuclear transcription factor Y subunit alpha	LOC8028254	*I. scapularis*/SCRV inf. *E. minasensis*	0.55	0.76	0.58	8.40E-238
L-asparaginase	LOC115314979	*I. scapularis*/SCRV inf. *E. minasensis*	0.54	0.76	0.58	2.65E-212
heat shock protein HSP 90-alpha	LOC8041558	*I. scapularis*/SCRV inf. *E. minasensis*	0.53	1.00	1.00	0.000257208
uncharacterized LOC8023940	LOC8023940	*I. scapularis*/SCRV inf. *E. minasensis*	0.52	0.75	0.58	5.29E-220
X-box-binding protein 1	LOC8032599	*I. scapularis*/SCRV inf. *E. minasensis*	0.52	1.00	0.99	4.05E-271
GTP-binding protein 2	LOC8043850	*I. scapularis*/SCRV inf. *E. minasensis*	0.52	0.70	0.51	2.74E-179
protein capicua homolog	LOC8040671	*I. scapularis*/SCRV inf. *E. minasensis*	0.51	0.94	0.89	6.77E-251
SPRY domain-containing protein 3	LOC8052623	*I. scapularis*/SCRV inf. *E. minasensis*	0.50	0.90	0.83	8.29E-171
uncharacterized LOC8025404	LOC8025404	*I. scapularis* with endogenous SCRV	1.07	0.63	0.41	1.24E-172
mitochondrial protein COX3	MT-cox3	*I. scapularis* with endogenous SCRV	1.00	1.00	1.00	0

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
U24-ctenitoxin-Pn1a	LOC8035490	*I. scapularis* with endogenous SCRV	0.97	0.96	0.89	1.30E-261
RNA-binding protein Rsf1	LOC120838763	*I. scapularis* with endogenous SCRV	0.90	0.94	0.80	0
cathepsin B	LOC120842054	*I. scapularis* with endogenous SCRV	0.79	0.99	0.97	6.40E-258
peroxiredoxin 1	LOC8037639	*I. scapularis* with endogenous SCRV	0.78	1.00	0.99	0
tubulin alpha chain	LOC8028862	*I. scapularis* with endogenous SCRV	0.74	0.99	0.96	0
uncharacterized LOC8027774	LOC8027774	*I. scapularis* with endogenous SCRV	0.74	0.42	0.26	1.46E-77
uncharacterized LOC115308880	LOC115308880	*I. scapularis* with endogenous SCRV	0.72	0.98	0.97	5.44E-222
neurofilament heavy polypeptide	LOC8028532	*I. scapularis* with endogenous SCRV	0.72	0.97	0.92	0
serine/arginine-rich splicing factor RS2Z33	LOC8038186	*I. scapularis* with endogenous SCRV	0.70	0.91	0.74	0
cystatin-B	LOC115324653	*I. scapularis* with endogenous SCRV	0.69	0.99	0.97	0
uncharacterized LOC115311024	LOC115311024	*I. scapularis* with endogenous SCRV	0.69	0.71	0.39	0
doublesex- and mab-3-related transcription factor 1Y	LOC8030900	*I. scapularis* with endogenous SCRV	0.68	0.98	0.97	1.63E-260
protein Dicer	LOC8039763	*I. scapularis* with endogenous SCRV	0.65	1.00	0.99	0

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
heterogeneous nuclear ribonucleoprotein 27C	LOC8041736	*I. scapularis* with endogenous SCRV	0.62	0.96	0.88	0
mitochondrial protein COX1	MT-cox1	*I. scapularis* with endogenous SCRV	0.62	1.00	1.00	0
high mobility group protein HMG-I/HMG-Y-like	LOC121833528	*I. scapularis* with endogenous SCRV	0.61	0.83	0.70	6.32E-170
myosin regulatory light chain sqh	LOC8037988	*I. scapularis* with endogenous SCRV	0.60	0.98	0.97	0
mitochondrial protein COX2	MT-cox2	*I. scapularis* with endogenous SCRV	0.59	1.00	1.00	0
dynein light chain 1, cytoplasmic	LOC8030084	*I. scapularis* with endogenous SCRV	0.59	0.99	0.98	0
26S proteasome complex subunit SEM1	LOC8042785	*I. scapularis* with endogenous SCRV	0.58	0.97	0.91	0
RNA-binding protein lark	LOC8043087	*I. scapularis* with endogenous SCRV	0.57	1.00	0.99	0
serine/arginine-rich splicing factor 7	LOC8052109	*I. scapularis* with endogenous SCRV	0.56	0.97	0.90	0
copper transport protein ATOX1	LOC115319407	*I. scapularis* with endogenous SCRV	0.56	0.97	0.92	0
mitochondrial protein NAD3	MT-nad3	*I. scapularis* with endogenous SCRV	0.55	1.00	0.99	0
interferon alpha-inducible protein 27-like protein 2B	LOC120835946	*I. scapularis* with endogenous SCRV	0.55	1.00	0.99	0
histone H2A.V	LOC121837043	*I. scapularis* with endogenous SCRV	0.55	0.94	0.88	3.31E-281

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
protein deadpan	LOC8025328	*I. scapularis* with endogenous SCRV	0.55	0.49	0.31	5.40E-111
PRA1 family protein 3	LOC120842230	*I. scapularis* with endogenous SCRV	0.54	0.71	0.45	2.06E-283
polyubiquitin-C	LOC8036289	*I. scapularis* with endogenous SCRV	0.54	1.00	0.99	0
death-associated protein 1	LOC8030152	*I. scapularis* with endogenous SCRV	0.54	0.88	0.74	5.82E-295
B-cell receptor-associated protein 31	LOC8033203	*I. scapularis* with endogenous SCRV	0.53	0.98	0.93	6.83E-262
splicing factor 3B subunit 5	LOC8041725	*I. scapularis* with endogenous SCRV	0.52	0.97	0.93	0
interferon alpha-inducible protein 27-like protein 2	LOC121837059	*I. scapularis* with endogenous SCRV	0.52	0.92	0.84	2.34E-155
histone H2A	LOC8029157	*I. scapularis* with endogenous SCRV	0.52	0.92	0.87	1.99E-159
ATP-binding cassette sub-family G member 1	LOC8051430	*I. scapularis* with endogenous SCRV	0.52	0.98	0.97	1.54E-47
microtubule-associated protein Jupiter	LOC8034169	*I. scapularis* with endogenous SCRV	0.51	1.00	0.99	9.72E-264
uncharacterized LOC8036807	LOC8036807	*I. scapularis* with endogenous SCRV	0.51	1.00	0.98	2.94E-293
integrin beta-PS-like	LOC121835038	*I. scapularis* with endogenous SCRV	0.51	1.00	0.98	0
serine/arginine-rich splicing factor 3	LOC115314463	*I. scapularis* with endogenous SCRV	0.51	0.87	0.76	2.77E-225

Supplementary table 7 Differentially expressed genes in Rhipicephalus microplus cell samples (uninfected, infected with St Croix River Virus (SCRV) and infected with Ehrlichia minasensis and SCRV). "Product" reflects on the gene function, "Gene" stands for the gene ID in the RefSeq annotation ('Em' genes come from the E. minasensis genome, 'MT' – tick mitochondrial genes, 'LOC' – nuclear tick genome, 'StCRV' – virus genome segments), "avg_log2FC" describes change in expression level of the gene in the sample compared with the other sample, "pct.1" means the fraction of cells in the current sample in which the gene was detected, "pct.2" means the fraction of cells in two other samples in which the gene was detected, "p_val" – adjusted p-value, based on Bonferroni correction using all genes in the dataset. Genes are sorted by "Sample" and "avg_log2FC" in descending order. Lower log2FC threshold was 0.5.

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
glycine-rich cell wall structural protein 1-like	LOC119161184	*R. microplus*	1.63	0.31	0.28	0.135234439
uncharacterized LOC119179904	LOC119179904	*R. microplus*	1.27	0.36	0.23	1.33E-75
papilin-like	LOC119167246	*R. microplus*	1.26	0.70	0.63	1.38E-126
cell surface glycoprotein 1-like	LOC119161189	*R. microplus*	1.20	0.56	0.49	5.22E-71
proteoglycan 4-like	LOC119163376	*R. microplus*	1.17	0.73	0.61	3.60E-153
mucin-6-like	LOC119179492	*R. microplus*	1.12	0.78	0.59	0
translation initiation factor IF-2-like	LOC119164579	*R. microplus*	1.10	0.94	0.90	0
leukocyte elastase inhibitor-like	LOC119178788	*R. microplus*	1.07	0.68	0.53	7.75E-172
serum amyloid A-2 protein-like	LOC119179825	*R. microplus*	1.04	0.58	0.53	6.86E-32
glutathione S-transferase 4-like	LOC119170592	*R. microplus*	1.00	0.77	0.68	1.31E-95
uncharacterized LOC119159686	LOC119159686	*R. microplus*	0.98	0.47	0.32	1.11E-108
uncharacterized LOC119170170	LOC119170170	*R. microplus*	0.93	0.51	0.35	1.79E-147
60S ribosomal protein L29-like	LOC119186032	*R. microplus*	0.91	1.00	1.00	0
isoinhibitor K-like	LOC119172095	*R. microplus*	0.90	0.32	0.14	1.93E-146
peroxisomal (S)-2-hydroxy-acid oxidase GLO5-like	LOC119179704	*R. microplus*	0.89	0.76	0.56	2.53E-269

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
60S ribosomal protein L37a-like	LOC119184682	*R. microplus*	0.85	1.00	1.00	0
uncharacterized LOC119186217	LOC119186217	*R. microplus*	0.85	0.44	0.41	4.12E-24
ribosomal protein rpl-36.A	LOC119164389	*R. microplus*	0.85	1.00	1.00	0
40S ribosomal protein S21	LOC119163311	*R. microplus*	0.84	1.00	1.00	0
uncharacterized LOC119163560	LOC119163560	*R. microplus*	0.84	0.94	0.93	1.20E-157
40S ribosomal protein S25-like	LOC119171793	*R. microplus*	0.83	1.00	1.00	0
uncharacterized LOC119184974	LOC119184974	*R. microplus*	0.79	1.00	1.00	0
uncharacterized LOC119187547	LOC119187547	*R. microplus*	0.79	0.99	0.93	0
alpha-1-macroglobulin-like	LOC119159557	*R. microplus*	0.78	0.90	0.86	5.26E-187
60S ribosomal protein L37-like	LOC119180885	*R. microplus*	0.76	1.00	1.00	0
glycine-rich protein DOT1-like	LOC119164380	*R. microplus*	0.74	0.55	0.65	0.077201642
60S ribosomal protein L30	LOC119170647	*R. microplus*	0.74	1.00	1.00	0
60S ribosomal protein L39	LOC119170875	*R. microplus*	0.73	1.00	1.00	0
beta-hexosaminidase subunit alpha-like	LOC119178235	*R. microplus*	0.72	0.77	0.68	8.33E-151
40S ribosomal protein SA-like	LOC119179491	*R. microplus*	0.72	1.00	1.00	0
40S ribosomal protein S26-like	LOC119159407	*R. microplus*	0.71	1.00	1.00	0
40S ribosomal protein S9-like	LOC119179503	*R. microplus*	0.69	1.00	1.00	0
60S ribosomal protein L13a-like	LOC119174431	*R. microplus*	0.69	1.00	1.00	0
laminin subunit beta-1-like	LOC119163525	*R. microplus*	0.67	0.91	0.84	2.89E-148
60S ribosomal protein L36	LOC119159745	*R. microplus*	0.66	1.00	1.00	0
40S ribosomal protein S5	LOC119161554	*R. microplus*	0.64	1.00	1.00	0
hornerin-like	LOC119167389	*R. microplus*	0.63	0.36	0.28	5.92E-35
60S acidic ribosomal protein P2-like	LOC119176942	*R. microplus*	0.62	1.00	1.00	0
60S ribosomal protein L22-like	LOC119168026	*R. microplus*	0.62	1.00	1.00	0

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
40S ribosomal protein S27	LOC119172085	*R. microplus*	0.61	1.00	1.00	0
dermonecrotic toxin SPH-like	LOC119179817	*R. microplus*	0.61	0.52	0.32	5.60E-170
uncharacterized LOC119165065	LOC119165065	*R. microplus*	0.61	0.82	0.79	2.39E-91
60S ribosomal protein L4-like	LOC119172451	*R. microplus*	0.60	0.99	0.98	0
cytochrome b-c1 complex subunit 6, mitochondrial-like	LOC119163767	*R. microplus*	0.60	1.00	0.98	0
high mobility group protein B1-like	LOC119170112	*R. microplus*	0.60	0.92	0.90	2.27E-101
uncharacterized LOC119175929	LOC119175929	*R. microplus*	0.58	0.40	0.30	1.43E-58
60S ribosomal protein L8	LOC119178768	*R. microplus*	0.58	1.00	1.00	0
uncharacterized LOC119175100	LOC119175100	*R. microplus*	0.58	0.89	0.76	0
60S ribosomal protein L6-like	LOC119174999	*R. microplus*	0.57	1.00	1.00	0
60S ribosomal protein L5-A-like	LOC119170407	*R. microplus*	0.57	1.00	1.00	0
60S ribosomal protein L27-like	LOC119159736	*R. microplus*	0.57	1.00	1.00	0
60S ribosomal protein L27a-like	LOC119169155	*R. microplus*	0.56	1.00	1.00	0
60S ribosomal protein L19-like	LOC119170639	*R. microplus*	0.56	1.00	1.00	0
uncharacterized LOC119170033	LOC119170033	*R. microplus*	0.56	0.98	0.95	0
phenoloxidase-activating factor 2-like	LOC119166682	*R. microplus*	0.55	0.58	0.51	2.07E-47
40S ribosomal protein S17-like	LOC119186846	*R. microplus*	0.55	1.00	1.00	0
60S ribosomal protein L31-like	LOC119170440	*R. microplus*	0.55	1.00	1.00	0
10 kDa heat shock protein, mitochondrial-like	LOC119178178	*R. microplus*	0.54	0.93	0.87	5.97E-198
60S ribosomal protein L21-like	LOC119170108	*R. microplus*	0.54	1.00	1.00	0
polyadenylate-binding protein 4-like	LOC119167073	*R. microplus*	0.54	1.00	1.00	0
elongation factor 1-beta-like	LOC119178699	*R. microplus*	0.54	1.00	1.00	0
40S ribosomal protein S11-like	LOC119180341	*R. microplus*	0.53	1.00	1.00	0

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
60S ribosomal protein L10-like	LOC119180409	*R. microplus*	0.53	1.00	1.00	0
60S ribosomal protein L35a-like	LOC119178752	*R. microplus*	0.53	1.00	1.00	0
40S ribosomal protein S4	LOC119174474	*R. microplus*	0.51	1.00	1.00	0
40S ribosomal protein S20-like	LOC119163611	*R. microplus*	0.51	1.00	1.00	0
protein disulfide-isomerase A6 homolog	LOC119170493	*R. microplus*	0.51	0.77	0.64	3.20E-183
laminin subunit alpha-like	LOC119164831	*R. microplus*	0.51	0.85	0.75	4.33E-139
40S ribosomal protein S10-like	LOC119170126	*R. microplus*	0.51	1.00	1.00	0
keratin-associated protein 19-2-like	LOC119171469	*R. microplus*	0.51	0.75	0.72	5.01E-31
60S ribosomal protein L32-like	LOC119181396	*R. microplus*	0.50	1.00	1.00	0
60S ribosomal protein L9-like	LOC119170048	*R. microplus*	0.50	1.00	1.00	0
40S ribosomal protein S16	LOC119170128	*R. microplus*	0.50	1.00	1.00	0
23S ribosomal RNA	Em00309	*R. microplus* inf. *E. minasensis* and SCRV	5.23	0.69	0.00	0
16S ribosomal RNA	Em00696	*R. microplus* inf. *E. minasensis* and SCRV	4.20	0.52	0.00	0
gamma-interferon-inducible lysosomal thiol reductase-like	LOC119181373	*R. microplus* inf. *E. minasensis* and SCRV	1.02	0.77	0.54	1.92E-181
MAP kinase-interacting serine/threonine-protein kinase 1-like	LOC119180780	*R. microplus* inf. *E. minasensis* and SCRV	0.99	0.99	0.98	0
60S acidic ribosomal protein P1-like	LOC119170882	*R. microplus* inf. *E. minasensis* and SCRV	0.96	0.89	0.83	0

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
Y+L amino acid transporter 2-like	LOC119165189	*R. microplus* inf. *E. minasensis* and SCRV	0.91	0.88	0.68	2.63E-267
ras-related protein Rab-27A-like	LOC119171711	*R. microplus* inf. *E. minasensis* and SCRV	0.90	0.86	0.74	6.68E-225
integrator complex subunit 2-like	LOC119172777	*R. microplus* inf. *E. minasensis* and SCRV	0.89	0.95	0.95	5.37E-53
very low-density lipoprotein receptor-like	LOC119181006	*R. microplus* inf. *E. minasensis* and SCRV	0.82	0.84	0.61	2.75E-231
sodium-dependent proline transporter-like	LOC119163699	*R. microplus* inf. *E. minasensis* and SCRV	0.77	0.84	0.63	9.48E-196
lachesin-like	LOC119185044	*R. microplus* inf. *E. minasensis* and SCRV	0.75	0.67	0.52	5.46E-110
mitochondrial protein NAD2	MT-nad2	*R. microplus* inf. *E. minasensis* and SCRV	0.75	0.99	0.98	5.96E-259
mitochondrial ribosomal RNA rrnL	MT-rrnL	*R. microplus* inf. *E. minasensis* and SCRV	0.75	1.00	1.00	0

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
serine/threonine-protein kinase NLK-like	LOC119172958	*R. microplus* inf. *E. minasensis* and SCRV	0.72	0.89	0.75	2.46E-217
neuron navigator 2-like	LOC119179109	*R. microplus* inf. *E. minasensis* and SCRV	0.71	0.65	0.38	2.05E-209
cAMP-specific 3',5'-cyclic phosphodiesterase 4C-like	LOC119180732	*R. microplus* inf. *E. minasensis* and SCRV	0.69	0.73	0.53	1.35E-183
ets DNA-binding protein pokkuri-like	LOC119170723	*R. microplus* inf. *E. minasensis* and SCRV	0.68	0.67	0.42	5.31E-208
P-selectin-like	LOC119163798	*R. microplus* inf. *E. minasensis* and SCRV	0.66	0.76	0.50	9.42E-262
neurocalcin homolog	LOC119163734	*R. microplus* inf. *E. minasensis* and SCRV	0.66	0.91	0.82	5.20E-189
solute carrier family 12 member 8-like	LOC119174967	*R. microplus* inf. *E. minasensis* and SCRV	0.65	0.74	0.62	3.02E-86
uncharacterized LOC119176801	LOC119176801	*R. microplus* inf. *E. minasensis* and SCRV	0.65	0.67	0.46	1.97E-136

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
caskin-2-like	LOC119180292	*R. microplus* inf. *E. minasensis* and SCRV	0.64	0.90	0.78	5.41E-186
NADP-dependent malic enzyme-like	LOC119174261	*R. microplus* inf. *E. minasensis* and SCRV	0.64	0.85	0.68	2.31E-164
NADPH--cytochrome P450 reductase-like	LOC119180268	*R. microplus* inf. *E. minasensis* and SCRV	0.63	0.91	0.73	6.85E-214
protein spitz-like	LOC119161327	*R. microplus* inf. *E. minasensis* and SCRV	0.63	0.81	0.56	8.65E-240
3 beta-hydroxysteroid dehydrogenase type 7-like	LOC119169799	*R. microplus* inf. *E. minasensis* and SCRV	0.63	0.94	0.82	3.59E-218
uncharacterized LOC119177235	LOC119177235	*R. microplus* inf. *E. minasensis* and SCRV	0.63	0.83	0.68	1.21E-161
la-related protein 4-like	LOC119188156	*R. microplus* inf. *E. minasensis* and SCRV	0.63	0.97	0.91	1.25E-277
scavenger receptor class B member 1-like	LOC119177309	*R. microplus* inf. *E. minasensis* and SCRV	0.62	0.60	0.39	3.06E-111

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
plexin domain-containing protein 1-like	LOC119178571	*R. microplus* inf. *E. minasensis* and SCRV	0.62	0.95	0.89	7.44E-261
solute carrier family 2, facilitated glucose transporter member 1-like	LOC119183081	*R. microplus* inf. *E. minasensis* and SCRV	0.61	0.82	0.59	2.75E-162
excitatory amino acid transporter-like	LOC119164459	*R. microplus* inf. *E. minasensis* and SCRV	0.61	0.71	0.61	4.14E-70
leupaxin-like	LOC119173083	*R. microplus* inf. *E. minasensis* and SCRV	0.61	0.86	0.67	1.88E-191
actin, clone 403	LOC119180016	*R. microplus* inf. *E. minasensis* and SCRV	0.61	1.00	1.00	5.01E-101
BTB/POZ domain-containing protein 6-B-like	LOC119167227	*R. microplus* inf. *E. minasensis* and SCRV	0.61	0.67	0.39	3.64E-214
protein Smaug homolog 1-like	LOC119166854	*R. microplus* inf. *E. minasensis* and SCRV	0.61	0.90	0.84	4.64E-172
sodium- and chloride-dependent glycine transporter 1-like	LOC119170435	*R. microplus* inf. *E. minasensis* and SCRV	0.60	0.74	0.51	1.07E-165

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
prominin-like protein	LOC119179321	*R. microplus* inf. *E. minasensis* and SCRV	0.60	0.80	0.66	1.16E-168
protein TMEPAI-like	LOC119176772	*R. microplus* inf. *E. minasensis* and SCRV	0.59	0.56	0.35	1.24E-141
fructose-bisphosphate aldolase A-like	LOC119176078	*R. microplus* inf. *E. minasensis* and SCRV	0.59	1.00	0.99	8.61E-117
microtubule-associated serine/threonine-protein kinase 3-like	LOC119181309	*R. microplus* inf. *E. minasensis* and SCRV	0.59	0.73	0.53	6.33E-162
Down syndrome cell adhesion molecule-like protein 1 homolog	LOC119174636	*R. microplus* inf. *E. minasensis* and SCRV	0.58	0.47	0.32	1.66E-79
pneumococcal serine-rich repeat protein-like	LOC119178306	*R. microplus* inf. *E. minasensis* and SCRV	0.58	0.88	0.78	1.40E-125
CD9 antigen-like	LOC119170582	*R. microplus* inf. *E. minasensis* and SCRV	0.58	0.86	0.79	3.85E-107
dual specificity protein phosphatase 10-like	LOC119167068	*R. microplus* inf. *E. minasensis* and SCRV	0.57	0.81	0.59	1.68E-133

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
excitatory amino acid transporter 3-like	LOC119181405	*R. microplus* inf. *E. minasensis* and SCRV	0.57	0.63	0.52	1.16E-55
small nucleolar RNA U3	LOC119165941	*R. microplus* inf. *E. minasensis* and SCRV	0.57	0.71	0.53	1.94E-63
transmembrane protein 47-like	LOC119170732	*R. microplus* inf. *E. minasensis* and SCRV	0.56	0.60	0.33	9.62E-191
transcription factor kayak-like	LOC119167126	*R. microplus* inf. *E. minasensis* and SCRV	0.55	0.91	0.74	2.66E-167
forkhead box protein K1-like	LOC119180236	*R. microplus* inf. *E. minasensis* and SCRV	0.55	0.85	0.67	2.16E-177
MIF4G domain-containing protein A-like	LOC119166653	*R. microplus* inf. *E. minasensis* and SCRV	0.55	0.69	0.45	1.03E-193
LIM domain transcription factor LMO4.2-like	LOC119170328	*R. microplus* inf. *E. minasensis* and SCRV	0.54	0.76	0.63	8.29E-100
alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 5-like	LOC119180255	*R. microplus* inf. *E. minasensis* and SCRV	0.54	0.87	0.71	2.03E-96

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
oxidation resistance protein 1-like	LOC119170503	*R. microplus* inf. *E. minasensis* and SCRV	0.54	0.82	0.64	2.55E-163
uncharacterized LOC119172285	LOC119172285	*R. microplus* inf. *E. minasensis* and SCRV	0.53	0.70	0.52	5.91E-152
eukaryotic translation initiation factor 4 gamma 1-like	LOC119161595	*R. microplus* inf. *E. minasensis* and SCRV	0.52	0.87	0.76	3.83E-144
CDK5 and ABL1 enzyme substrate 2-like	LOC119163724	*R. microplus* inf. *E. minasensis* and SCRV	0.52	0.81	0.66	2.99E-120
ABC transporter G family member 21-like	LOC119176581	*R. microplus* inf. *E. minasensis* and SCRV	0.52	0.82	0.74	2.52E-65
probable ATP-dependent RNA helicase DDX17	LOC119180324	*R. microplus* inf. *E. minasensis* and SCRV	0.52	1.00	1.00	1.46E-148
cGMP-dependent 3',5'-cyclic phosphodiesterase-like	LOC119180013	*R. microplus* inf. *E. minasensis* and SCRV	0.52	0.74	0.45	1.56E-197
alkaline phosphatase-like	LOC119178734	*R. microplus* inf. *E. minasensis* and SCRV	0.52	0.89	0.87	2.77E-39

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
ETS-like protein pointed	LOC119171764	*R. microplus* inf. *E. minasensis* and SCRV	0.51	0.80	0.72	2.09E-65
muscleblind-like protein 2	LOC119164409	*R. microplus* inf. *E. minasensis* and SCRV	0.51	0.89	0.81	2.24E-124
tyrosine-protein kinase SRK3-like	LOC119163644	*R. microplus* inf. *E. minasensis* and SCRV	0.51	0.86	0.75	1.19E-100
protein BTG2-like	LOC119174910	*R. microplus* inf. *E. minasensis* and SCRV	0.51	0.99	0.97	1.53E-109
membralin-like	LOC119181726	*R. microplus* inf. *E. minasensis* and SCRV	0.51	0.73	0.46	1.46E-222
poly(rC)-binding protein 3-like	LOC119172413	*R. microplus* inf. *E. minasensis* and SCRV	0.51	0.88	0.70	1.05E-164
receptor-type tyrosine-protein phosphatase N2-like	LOC119178278	*R. microplus* inf. *E. minasensis* and SCRV	0.51	0.85	0.78	1.92E-81
pleckstrin homology-like domain family B member 1	LOC119171819	*R. microplus* inf. *E. minasensis* and SCRV	0.50	0.82	0.67	1.69E-147

product	**gene**	**sample**	**avg_log2FC**	**pct.1**	**pct.2**	**p_val_adj**
aquaporin-9-like	LOC119188255	*R. microplus* inf. *E. minasensis* and SCRV	0.50	0.46	0.28	2.97E-74
actin-related protein 2/3 complex subunit 1A-A-like	LOC119180131	*R. microplus* inf. *E. minasensis* and SCRV	0.50	0.91	0.77	8.75E-189
StCRVs7gp1	StCRVs7gp1	*R. microplus* inf. SCRV	3.18	0.95	0.32	0
StCRVs5gp1	StCRVs5gp1	*R. microplus* inf. SCRV	2.28	0.81	0.24	0
alpha-crystallin A chain-like	LOC119185002	*R. microplus* inf. SCRV	2.27	0.82	0.51	0
StCRVs3gp1	StCRVs3gp1	*R. microplus* inf. SCRV	2.24	0.75	0.21	0
alpha-crystallin A chain-like	LOC119185001	*R. microplus* inf. SCRV	2.22	0.88	0.58	0
alpha-crystallin A chain-like	LOC119185157	*R. microplus* inf. SCRV	2.21	0.86	0.52	0
StCRVs6gp1	StCRVs6gp1	*R. microplus* inf. SCRV	1.99	0.76	0.22	0
StCRVs2gp1	StCRVs2gp1	*R. microplus* inf. SCRV	1.99	0.79	0.25	0
StCRVs1gp1	StCRVs1gp1	*R. microplus* inf. SCRV	1.98	0.78	0.24	0
alpha-crystallin B chain-like	LOC119185148	*R. microplus* inf. SCRV	1.79	0.90	0.66	0
StCRVs4gp1	StCRVs4gp1	*R. microplus* inf. SCRV	1.65	0.72	0.20	0
StCRVs8gp1	StCRVs8gp1	*R. microplus* inf. SCRV	1.56	0.65	0.17	0
heat shock protein HSP 90-alpha-like	LOC119164008	*R. microplus* inf. SCRV	1.52	1.00	0.98	0
StCRVs9gp1	StCRVs9gp1	*R. microplus* inf. SCRV	1.48	0.66	0.20	0
heat shock protein Hsp-16.1/Hsp-16.11-like	LOC119178751	*R. microplus* inf. SCRV	1.42	0.96	0.85	6.17E-197
uncharacterized LOC119185067	LOC119185067	*R. microplus* inf. SCRV	1.41	0.75	0.45	1.45E-264
StCRVs10gp1	StCRVs10gp1	*R. microplus* inf. SCRV	1.30	0.63	0.16	0
alpha-crystallin A chain-like	LOC119180948	*R. microplus* inf. SCRV	1.20	0.48	0.26	3.52E-125
alpha-crystallin A chain-like	LOC119185153	*R. microplus* inf. SCRV	0.96	0.96	0.92	9.28E-103

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
cytohesin-1-like	LOC119179489	*R. microplus* inf. SCRV	0.93	0.91	0.78	0
inositol-3-phosphate synthase 1-A-like	LOC119177322	*R. microplus* inf. SCRV	0.90	0.79	0.60	8.06E-187
CCAAT/enhancer-binding protein delta-like	LOC119176174	*R. microplus* inf. SCRV	0.89	0.93	0.89	1.96E-242
uncharacterized LOC119177231	LOC119177231	*R. microplus* inf. SCRV	0.85	0.55	0.34	5.35E-119
uncharacterized LOC119171210	LOC119171210	*R. microplus* inf. SCRV	0.85	0.67	0.40	2.64E-218
uncharacterized LOC119185104	LOC119185104	*R. microplus* inf. SCRV	0.81	0.44	0.17	1.62E-228
mRNA cap guanine-N7 methyltransferase-like	LOC119172401	*R. microplus* inf. SCRV	0.80	0.52	0.34	1.88E-95
zwei Ig domain protein zig-2-like	LOC119175902	*R. microplus* inf. SCRV	0.80	0.53	0.34	1.18E-95
U2 spliceosomal RNA	LOC119184553	*R. microplus* inf. SCRV	0.80	0.64	0.43	1.58E-172
histone H1-delta-like	LOC119179514	*R. microplus* inf. SCRV	0.78	0.98	0.97	1.73E-100
alpha-crystallin A chain-like	LOC119185154	*R. microplus* inf. SCRV	0.78	0.67	0.48	6.76E-146
dnaJ protein homolog 1-like	LOC119181188	*R. microplus* inf. SCRV	0.76	0.67	0.45	2.52E-181
uncharacterized LOC119181069	LOC119181069	*R. microplus* inf. SCRV	0.76	0.68	0.46	1.69E-162
heat shock 70 kDa protein cognate 4	LOC119172195	*R. microplus* inf. SCRV	0.75	0.95	0.92	9.12E-104
copper transport protein ATOX1-like	LOC119170213	*R. microplus* inf. SCRV	0.75	0.77	0.59	4.27E-189
choline/ethanolamine kinase-like	LOC119170411	*R. microplus* inf. SCRV	0.74	0.88	0.72	1.43E-248
perilipin-2-like	LOC119188251	*R. microplus* inf. SCRV	0.74	0.97	0.93	1.55E-116
motile sperm domain-containing protein 2-like	LOC119159774	*R. microplus* inf. SCRV	0.74	0.83	0.61	1.75E-230
uncharacterized LOC119182515	LOC119182515	*R. microplus* inf. SCRV	0.73	0.80	0.58	1.10E-160
putative aminopeptidase W07G4.4	LOC119177170	*R. microplus* inf. SCRV	0.73	0.93	0.86	1.28E-184
uncharacterized LOC119161405	LOC119161405	*R. microplus* inf. SCRV	0.72	0.75	0.55	1.39E-179
uncharacterized LOC119170319	LOC119170319	*R. microplus* inf. SCRV	0.71	0.95	0.90	2.04E-129

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
nuclear hormone receptor E75-like	LOC119176056	*R. microplus* inf. SCRV	0.69	0.78	0.57	1.36E-175
uncharacterized LOC119180403	LOC119180403	*R. microplus* inf. SCRV	0.69	0.96	0.89	2.61E-244
phenylacetaldehyde reductase-like	LOC119178640	*R. microplus* inf. SCRV	0.68	0.96	0.90	2.63E-239
WD repeat domain phosphoinositide-interacting protein 2-like	LOC119180019	*R. microplus* inf. SCRV	0.68	0.92	0.78	1.81E-240
uncharacterized LOC119161549	LOC119161549	*R. microplus* inf. SCRV	0.67	1.00	0.99	3.01E-203
facilitated trehalose transporter Tret1-like	LOC119172134	*R. microplus* inf. SCRV	0.66	0.81	0.64	5.17E-182
UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase-like	LOC119171686	*R. microplus* inf. SCRV	0.66	0.78	0.59	3.82E-181
sestrin-3-like	LOC119174647	*R. microplus* inf. SCRV	0.66	0.95	0.84	4.25E-238
complex I assembly factor ACAD9, mitochondrial-like	LOC119174792	*R. microplus* inf. SCRV	0.63	0.87	0.69	2.24E-234
transcription factor AP-1-like	LOC119180583	*R. microplus* inf. SCRV	0.63	0.98	0.96	1.27E-183
CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase, mitochondrial-like	LOC119179589	*R. microplus* inf. SCRV	0.62	0.72	0.53	2.16E-124
store-operated calcium entry-associated regulatory factor-like	LOC119180723	*R. microplus* inf. SCRV	0.61	0.93	0.80	3.75E-144
uncharacterized LOC119163100	LOC119163100	*R. microplus* inf. SCRV	0.61	0.79	0.58	9.29E-194
serine hydrolase-like protein	LOC119159538	*R. microplus* inf. SCRV	0.60	0.82	0.62	1.36E-178
U2 spliceosomal RNA	LOC119184099	*R. microplus* inf. SCRV	0.60	0.64	0.43	2.49E-163
cytochrome P450 3A8-like	LOC119172114	*R. microplus* inf. SCRV	0.59	0.71	0.57	3.58E-67
U2 spliceosomal RNA	LOC119183045	*R. microplus* inf. SCRV	0.59	0.64	0.43	8.31E-163
U2 spliceosomal RNA	LOC119183048	*R. microplus* inf. SCRV	0.59	0.64	0.43	8.31E-163
U2 spliceosomal RNA	LOC119183049	*R. microplus* inf. SCRV	0.59	0.64	0.43	8.31E-163
U2 spliceosomal RNA	LOC119183050	*R. microplus* inf. SCRV	0.59	0.64	0.43	8.31E-163

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
U2 spliceosomal RNA	LOC119183052	*R. microplus* inf. SCRV	0.59	0.64	0.43	8.31E-163
U2 spliceosomal RNA	LOC119183046	*R. microplus* inf. SCRV	0.59	0.64	0.43	8.31E-163
U2 spliceosomal RNA	LOC119183047	*R. microplus* inf. SCRV	0.59	0.64	0.43	8.31E-163
peroxisomal membrane protein 11A-like	LOC119187945	*R. microplus* inf. SCRV	0.59	0.75	0.56	1.25E-126
multi-drug resistance efflux pump PmrA homolog	LOC119175031	*R. microplus* inf. SCRV	0.58	0.71	0.50	1.77E-158
eukaryotic translation initiation factor 2 subunit 2-like	LOC119176726	*R. microplus* inf. SCRV	0.58	0.99	0.98	1.45E-259
cyclin-I-like	LOC119179355	*R. microplus* inf. SCRV	0.58	0.96	0.90	1.39E-130
ets DNA-binding protein pokkuri-like	LOC119171668	*R. microplus* inf. SCRV	0.58	0.48	0.28	3.56E-115
ras-related GTP-binding protein A-like	LOC119178117	*R. microplus* inf. SCRV	0.58	0.91	0.80	5.85E-246
proline-rich protein 5-like	LOC119172020	*R. microplus* inf. SCRV	0.58	0.57	0.31	5.54E-194
mitochondrial fission 1 protein-like	LOC119159618	*R. microplus* inf. SCRV	0.58	0.99	0.96	6.58E-288
glucose-6-phosphatase 2-like	LOC119172192	*R. microplus* inf. SCRV	0.58	0.52	0.39	1.84E-61
putative nuclease HARBI1	LOC119178280	*R. microplus* inf. SCRV	0.57	0.78	0.66	8.12E-117
stearoyl-CoA desaturase 5-like	LOC119187539	*R. microplus* inf. SCRV	0.56	0.86	0.69	1.58E-150
inositol-tetrakisphosphate 1-kinase-like	LOC119163988	*R. microplus* inf. SCRV	0.56	0.85	0.64	2.17E-213
30S ribosomal protein S11-like	LOC119159434	*R. microplus* inf. SCRV	0.56	0.96	0.89	5.77E-170
immediate early response gene 2 protein-like	LOC119185999	*R. microplus* inf. SCRV	0.55	0.96	0.93	1.68E-117
general transcription factor 3C polypeptide 3-like	LOC119168498	*R. microplus* inf. SCRV	0.55	0.85	0.72	1.98E-175
phosphoserine phosphatase-like	LOC119184205	*R. microplus* inf. SCRV	0.55	0.72	0.57	3.31E-116
uncharacterized LOC119163216	LOC119163216	*R. microplus* inf. SCRV	0.55	0.90	0.83	3.36E-114
src kinase-associated phosphoprotein 2-like	LOC119174947	*R. microplus* inf. SCRV	0.55	0.84	0.67	7.73E-183

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
28S ribosomal protein S36, mitochondrial-like	LOC119188289	*R. microplus* inf. SCRV	0.54	0.88	0.75	1.33E-113
transcription factor Sp8-like	LOC119174788	*R. microplus* inf. SCRV	0.54	0.80	0.70	2.06E-99
ADP-ribose glycohydrolase ARH3-like	LOC119175995	*R. microplus* inf. SCRV	0.54	0.95	0.90	7.00E-188
adenylate kinase isoenzyme 6-like	LOC119170030	*R. microplus* inf. SCRV	0.54	0.89	0.79	4.73E-190
aldehyde dehydrogenase, mitochondrial-like	LOC119187513	*R. microplus* inf. SCRV	0.53	0.84	0.68	6.40E-155
U2 spliceosomal RNA	LOC119165935	*R. microplus* inf. SCRV	0.53	0.44	0.33	3.53E-47
sulfotransferase 1C2-like	LOC119178464	*R. microplus* inf. SCRV	0.52	0.76	0.55	4.18E-152
fatty acid-binding protein-like	LOC119185260	*R. microplus* inf. SCRV	0.52	0.86	0.79	1.10E-123
mid1-interacting protein 1-B-like	LOC119178642	*R. microplus* inf. SCRV	0.52	0.91	0.87	3.27E-109
uncharacterized LOC119171669	LOC119171669	*R. microplus* inf. SCRV	0.52	0.26	0.12	5.23E-81
ADP-sugar pyrophosphatase-like	LOC119163799	*R. microplus* inf. SCRV	0.52	0.94	0.88	4.26E-57
solute carrier family 35 member F6-like	LOC119168869	*R. microplus* inf. SCRV	0.52	0.86	0.75	1.33E-172
ubiquitin-like protein 7	LOC119166988	*R. microplus* inf. SCRV	0.52	0.85	0.70	5.09E-215
muscle-specific protein 20-like	LOC119181440	*R. microplus* inf. SCRV	0.51	0.97	0.95	1.45E-171
phosphatidylinositol 3-kinase regulatory subunit alpha-like	LOC119181020	*R. microplus* inf. SCRV	0.51	0.87	0.74	1.78E-171
ras-related C3 botulinum toxin substrate 2-like	LOC119178269	*R. microplus* inf. SCRV	0.50	0.96	0.94	7.80E-149
gamma-aminobutyric acid receptor-associated protein	LOC119176268	*R. microplus* inf. SCRV	0.50	1.00	1.00	1.66E-244
ras-related protein Rab-11A-like	LOC119172043	*R. microplus* inf. SCRV	0.50	0.92	0.85	3.10E-191
cytosolic purine 5'-nucleotidase-like	LOC119175815	*R. microplus* inf. SCRV	0.50	0.82	0.69	3.39E-152

Supplementary table 8 Differentially expressed genes in samples of Rhipicephalus microplus uninfected and infected with Rickettsia raoultii. "Product" reflects on the gene function, "Gene" stands for the gene ID in the RefSeq annotation ('Rs' genes come from the R. raoultii genome, 'MT' – tick mitochondrial genes, 'LOC' – nuclear tick genome), "avg_log2FC" describes change in expression level of the gene in the sample compared with the other sample, "pct.1" means the fraction of cells in the current sample in which the gene was detected, "pct.2" means the fraction of cells in the other sample in which the gene was detected, "p_val" – adjusted p-value, based on Bonferroni correction using all genes in the dataset. Genes are sorted by "sample" and "avg_log2FC" in descending order. Lower log2FC threshold was 0.5.

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
23S ribosomal RNA	Rs01495	*R. microplus* inf. *R. raoultii*	2.90	0.64	0.00	0
16S ribosomal RNA	Rs05050	*R. microplus* inf. *R. raoultii*	1.80	0.52	0.00	0
fatty acid synthase-like	LOC119179527	*R. microplus* inf. *R. raoultii*	0.77	0.85	0.73	1.93E-149
insulin-induced gene 1 protein-like	LOC119188224	*R. microplus* inf. *R. raoultii*	0.64	0.68	0.48	3.70E-146
mucin-6-like	LOC119179492	*R. microplus*	0.71	0.78	0.70	3.54E-96
uncharacterized LOC119186217	LOC119186217	*R. microplus*	0.66	0.44	0.34	6.16E-32
uncharacterized LOC119159686	LOC119159686	*R. microplus*	0.63	0.47	0.35	2.22E-45
uncharacterized LOC119179904	LOC119179904	*R. microplus*	0.60	0.36	0.28	1.81E-19
peroxisomal (S)-2-hydroxy-acid oxidase GLO5-like	LOC119179704	*R. microplus*	0.52	0.76	0.68	2.49E-71
cell surface glycoprotein 1-like	LOC119161189	*R. microplus*	0.52	0.56	0.52	4.59E-08
papilin-like	LOC119167246	*R. microplus*	0.51	0.70	0.65	7.68E-25

Supplementary table 9 Differentially expressed genes in Rhipicephalus microplus uninfected and infected with Spiroplasma sp. "Product" reflects on the gene function, "Gene" stands for the gene ID in the RefSeq annotation ('Sp' genes come from the Spiroplasma sp. genome, 'MT' – tick mitochondrial genes, 'LOC' – nuclear tick genome), "avg_log2FC" describes change in expression level of the gene in the sample compared with the other sample, "pct.1" means the fraction of cells in the current sample in which the gene was detected, "pct.2" means the fraction of cells in the other sample in which the gene was detected, "p_val" – adjusted p-value, based on Bonferroni correction using all genes in the dataset. Genes are sorted by "sample" and "avg_log2FC" in descending order. Lower log2FC threshold was 0.5.

product	gene	sample	avg_log2FC	pct.1	pct.2	p_val_adj
23S ribosomal RNA	Sp01863	*R. microplus* inf. *Spiroplasma sp.*	2.94	0.51	0.00	0
23S ribosomal RNA	Sp00602	*R. microplus* inf. *Spiroplasma sp.*	2.43	0.47	0.00	0
uncharacterized LOC119185522	LOC119185522	*R. microplus* inf. *Spiroplasma sp.*	0.79	0.95	0.88	0
60S acidic ribosomal protein P1-like	LOC119170882	*R. microplus* inf. *Spiroplasma sp.*	0.72	0.86	0.85	7.35E-218
uncharacterized LOC119187546	LOC119187546	*R. microplus* inf. *Spiroplasma sp.*	0.69	1.00	0.99	0
inositol-3-phosphate synthase 1-A-like	LOC119177322	*R. microplus* inf. *Spiroplasma sp.*	0.63	0.68	0.51	1.89E-120
delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial-like	LOC119170777	*R. microplus* inf. *Spiroplasma sp.*	0.63	0.53	0.34	1.05E-127
small nucleolar RNA U3	LOC119165941	*R. microplus* inf. *Spiroplasma sp.*	0.62	0.57	0.43	1.05E-64
protein FAM98B-like	LOC119161187	*R. microplus* inf. *Spiroplasma sp.*	0.60	0.41	0.31	1.34E-26

fatty acid synthase-like	LOC119179527	*R. microplus* inf. *Spiroplasma sp.*	0.54	0.84	0.73	9.62E-102
gamma-butyrobetaine dioxygenase-like	LOC119172227	*R. microplus*	0.55	0.90	0.85	8.45E-51

Supplementary table 10 Differentially expressed genes in integrated Ixodes scapularis cell samples with mitochondrial counts and pathogen read counts signal regressed. "Product" reflects on the gene function, "Gene" stands for the gene ID in the RefSeq annotation "avg_log2FC" describes change in expression level of the gene in the sample compared with the other sample, "pct.1" means the fraction of cells in the current sample in which the gene was detected, "pct.2" means the fraction of cells in the other samples in which the gene was detected. Table is sorted by "Cluster' and "avg_log2FC" in descending order.

product	gene	cluster	avg_log2FC	pct.1	pct.2
cathepsin B	LOC120842054	0	1.19	1.00	0.90
kinesin-like protein KIF11	LOC8053770	1	1.43	0.97	0.50
cell division cycle protein 20 homolog	LOC115325879	1	1.40	0.92	0.35
myosin-6	LOC8042605	1	1.37	0.96	0.49
uncharacterized LOC115308880	LOC115308880	1	1.04	1.00	0.92
tubulin alpha chain	LOC8044135	1	1.04	0.99	0.76
growth/differentiation factor 8	LOC115329081	3	1.48	0.91	0.79
uncharacterized LOC8027774	LOC8027774	4	2.09	0.81	0.29
zwei Ig domain protein zig-4	LOC8036660	4	1.59	1.00	0.85
stromelysin-3	LOC8051455	4	1.53	0.88	0.54
E3 ubiquitin-protein ligase UBR4-like	LOC121834581	4	1.27	0.73	0.32
laminin subunit gamma-1	LOC8034493	4	1.26	0.99	0.89
uncharacterized LOC121048110	LOC121048110	4	1.13	0.91	0.56
uncharacterized LOC115314989	LOC115314989	5	1.29	0.96	0.85

product	gene	cluster	avg_log2FC	pct.1	pct.2
uncharacterized LOC8028338	LOC8028338	5	1.16	0.82	0.62
protein BTG2	LOC8051179	5	1.05	1.00	0.97
uncharacterized LOC120850041	LOC120850041	5	1.05	0.66	0.38
mothers against decapentaplegic homolog 6	LOC8031782	6	1.99	0.99	0.82
heat shock protein HSP 90-alpha	LOC8041558	6	1.66	1.00	0.97
uncharacterized LOC8051439	LOC8051439	6	1.63	0.97	0.73
uncharacterized LOC8030032	LOC8030032	6	1.59	0.95	0.65
mitochondrial ribosomal RNA rrnS	MT-rrnS	6	1.50	1.00	0.93
microtubule-associated protein futsch	LOC8033651	6	1.40	0.96	0.72
Krueppel-like factor 6	LOC8028313	6	1.40	0.98	0.82
ubiquitin carboxyl-terminal hydrolase 2	LOC8034595	6	1.35	1.00	0.95
BMP and activin membrane-bound inhibitor homolog	LOC115310360	6	1.17	0.99	0.89
retroviral integration site protein Fli-1 homolog	LOC8023218	6	1.07	0.98	0.90
nuclear transcription factor Y subunit beta	LOC8035603	6	1.06	0.44	0.48
U2 spliceosomal RNA	LOC115314205	7	1.59	0.61	0.47
small nucleolar RNA U3	LOC115327235	7	1.59	0.57	0.45
U1 spliceosomal RNA	LOC115327552	7	1.17	0.73	0.48
U2 spliceosomal RNA	LOC120848116	7	1.11	0.60	0.47
#N/A	StCRVs7gp1	8	1.30	0.25	0.08
#N/A	StCRVs2gp1	8	1.13	0.34	0.08
U2 spliceosomal RNA	LOC115319233	8	1.07	0.96	0.42

product	gene	cluster	avg_log2FC	pct.1	pct.2
U2 spliceosomal RNA	LOC120849918	8	1.07	0.96	0.42
#N/A	StCRVs1gp1	8	1.00	0.22	0.07
nidogen-1	LOC8038711	9	1.66	0.98	0.93
uncharacterized LOC121834122	LOC121834122	9	1.61	0.88	0.84
hypoxia up-regulated protein 1	LOC8040011	9	1.58	0.98	0.94
very low-density lipoprotein receptor	LOC8038934	9	1.56	0.94	0.93
endoplasmic reticulum chaperone BiP	LOC8051659	9	1.55	0.97	0.94
protein disulfide-isomerase A3	LOC8024633	9	1.49	0.96	0.96
uncharacterized LOC121835386	LOC121835386	9	1.49	0.53	0.40
complement C3	LOC8039362	9	1.49	0.83	0.83
acetylcholinesterase	LOC8032357	9	1.41	0.34	0.23
uncharacterized LOC121048595	LOC121048595	9	1.37	0.70	0.69
fibroblast growth factor receptor 4	LOC8051224	9	1.31	0.85	0.78
cation-independent mannose-6-phosphate receptor	LOC8043753	9	1.31	0.89	0.86
protein disulfide-isomerase A6 homolog	LOC115320057	9	1.31	0.87	0.92
uncharacterized LOC121046660	LOC121046660	9	1.28	0.58	0.49
uncharacterized LOC120842277	LOC120842277	9	1.28	0.67	0.73
stearoyl-CoA desaturase 5	LOC8051438	9	1.27	0.97	0.96
protein disulfide-isomerase	LOC120837176	9	1.26	0.96	0.96

product	gene	cluster	avg_log2FC	pct.1	pct.2
dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	LOC115319911	9	1.20	0.83	0.80
protein O-mannosyl-transferase TMTC2	LOC8051218	9	1.20	0.72	0.69
endoplasmic reticulum resident protein 29	LOC8051554	9	1.19	0.83	0.87
teneurin-m-like	LOC120848855	9	1.19	0.65	0.75
mitochondrial ribosomal RNA rrnS	MT-rrnS	9	1.18	0.86	0.94
endoplasmin	LOC8039070	9	1.18	0.90	0.92
mitochondrial ribosomal RNA rrnL	MT-rrnL	9	1.17	0.99	0.95
long-chain-fatty-acid--CoA ligase 3	LOC8029783	9	1.17	0.87	0.92
protein draper	LOC8033158	9	1.15	0.96	0.93
23 kDa integral membrane protein	LOC8029816	9	1.12	0.85	0.88
BTB/POZ domain-containing protein 9-like	LOC115315686	9	1.11	0.56	0.44
dnaJ homolog subfamily C member 3	LOC8025291	9	1.10	0.80	0.81
teneurin-m	LOC8024278	9	1.09	0.66	0.63
uncharacterized LOC8027088	LOC8027088	9	1.09	0.81	0.82
uncharacterized LOC121837265	LOC121837265	9	1.08	0.57	0.55
mitochondrial protein COX2	MT-cox2	9	1.07	0.95	0.92
SPRY domain-containing protein 3	LOC8052623	9	1.07	0.85	0.87
protein ERGIC-53	LOC8023926	9	1.06	0.76	0.74
innexin inx2	LOC115320817	9	1.06	0.76	0.79
uncharacterized LOC8029656	LOC8029656	9	1.05	0.81	0.83
sodium/calcium exchanger 3	LOC8028603	9	1.04	0.74	0.82
fatty acid hydroxylase domain-containing protein 2	LOC8032220	9	1.03	0.79	0.83
neuron navigator 2	LOC8040291	9	1.03	0.72	0.77

product	gene	cluster	avg_log2FC	pct.1	pct.2
sodium-dependent proline transporter	LOC8042754	9	1.03	0.74	0.84
integrin alpha-PS1	LOC8052228	9	1.02	0.69	0.78
multidrug resistance-associated protein 1	LOC8040491	9	1.02	0.75	0.77
heat shock protein 68	LOC8043999	10	5.29	0.98	0.31
heat shock protein 68	LOC115323694	10	5.23	0.97	0.37
heat shock protein 68-like	LOC115323617	10	5.05	0.98	0.33
protein lethal(2)essential for life	LOC115312592	10	4.85	0.98	0.37
heat shock protein 68-like	LOC8042924	10	4.40	0.95	0.32
heat shock protein 68-like	LOC115323629	10	4.38	0.90	0.24
alpha-crystallin B chain	LOC8025742	10	4.18	0.91	0.32
heat shock protein beta-6-like	LOC115312538	10	4.00	0.95	0.60
alpha-crystallin A chain	LOC8025894	10	3.92	0.97	0.53
alpha-crystallin B chain	LOC8043348	10	3.75	0.99	0.79
heat shock protein beta-6-like	LOC115312540	10	3.63	0.89	0.46
dnaJ homolog subfamily B member 4	LOC115324594	10	3.32	0.97	0.65
heat shock protein beta-6-like	LOC115323622	10	2.96	0.72	0.32
heat shock protein beta-6-like	LOC115312542	10	2.95	0.87	0.20
heat shock protein 68-like	LOC115312541	10	2.69	0.68	0.22

product	gene	cluster	avg_log2FC	pct.1	pct.2
heat shock protein HSP 90-alpha	LOC8041558	10	2.65	1.00	0.97
mitochondrial ribosomal RNA rrnL	MT-rrnL	10	2.62	1.00	0.95
alpha-crystallin A chain	LOC8025743	10	2.59	0.92	0.77
heat shock protein 68-like	LOC121835372	10	2.44	0.76	0.09
mitochondrial ribosomal RNA rrnS	MT-rrnS	10	2.42	0.99	0.93
U2 spliceosomal RNA	LOC115314205	10	2.08	0.62	0.48
heat shock protein 68-like	LOC115321088	10	1.93	0.74	0.07

Supplementary table 11 Differentially expressed genes in integrated Rhipicephalus microplus cell samples with mitochondrial counts and pathogen read counts signal regressed. "Product" reflects on the gene function, "Gene" stands for the gene ID in the RefSeq annotation "avg_log2FC" describes change in expression level of the gene in the sample compared with the other sample, "pct.1" means the fraction of cells in the current sample in which the gene was detected, "pct.2" means the fraction of cells in the other samples in which the gene was detected. Table is sorted by "Cluster" and "avg_log2FC" in descending order.

product	gene	cluster	avg_log2FC	pct.1	pct.2
papilin-like	LOC119167246	0	1.84	0.98	0.82
uncharacterized LOC119179904	LOC119179904	0	1.83	0.86	0.73
proteoglycan 4-like	LOC119163376	0	1.78	0.99	0.84
uncharacterized LOC119170170	LOC119170170	0	1.55	0.97	0.74
cell surface glycoprotein 1-like	LOC119161189	0	1.55	0.91	0.77
leukocyte elastase inhibitor-like	LOC119178788	0	1.54	0.97	0.81
uncharacterized LOC119159686	LOC119159686	0	1.32	0.92	0.81
uncharacterized LOC119186217	LOC119186217	0	1.26	0.87	0.74
translation initiation factor IF-2-like	LOC119164579	0	1.26	1.00	0.94
laminin subunit beta-1-like	LOC119163525	0	1.21	1.00	0.91
uncharacterized LOC119159679	LOC119159679	0	1.21	0.87	0.74
serum amyloid A-2 protein-like	LOC119179825	0	1.20	0.81	0.79
alpha-1-macroglobulin-like	LOC119159557	0	1.20	1.00	0.91
uncharacterized LOC119159685	LOC119159685	0	1.14	0.85	0.69
uncharacterized LOC119175929	LOC119175929	0	1.12	0.88	0.63
hornerin-like	LOC119167389	0	1.11	0.89	0.70
mucin-6-like	LOC119179492	0	1.02	0.99	0.90
uncharacterized LOC119182515	LOC119182515	1	1.71	0.88	0.74
nucleoside diphosphate kinase 6-like	LOC119163657	1	1.35	0.95	0.83
gamma-butyrobetaine dioxygenase-like	LOC119172227	1	1.01	0.99	0.94

product	gene	cluster	avg_log2FC	pct.1	pct.2
glyoxylate reductase/hydroxypyruvate reductase-like	LOC119167003	2	1.93	0.99	0.93
cytochrome c	LOC119187542	2	1.54	1.00	0.97
store-operated calcium entry-associated regulatory factor-like	LOC119180723	2	1.52	0.97	0.84
perilipin-2-like	LOC119188251	2	1.42	0.98	0.95
cysteine and histidine-rich domain-containing protein 1-like	LOC119174704	2	1.35	0.97	0.89
myeloid leukemia factor 1-like	LOC119168723	2	1.34	0.99	0.95
heat shock protein Hsp-16.1/Hsp-16.11-like	LOC119178751	2	1.28	0.98	0.86
muscle LIM protein 1-like	LOC119169195	2	1.24	0.92	0.82
uncharacterized LOC119171777	LOC119171777	2	1.21	0.70	0.55
fructose-bisphosphate aldolase A-like	LOC119176078	2	1.18	0.99	0.98
pyruvate kinase PKM-like	LOC119172236	2	1.16	0.84	0.58
AN1-type zinc finger protein 2A-like	LOC119170011	2	1.16	0.95	0.81
28S ribosomal protein S36, mitochondrial-like	LOC119188289	2	1.12	0.95	0.80
aquaporin-9-like	LOC119188255	2	1.12	0.78	0.48
persulfide dioxygenase ETHE1, mitochondrial-like	LOC119171673	2	1.10	0.93	0.72
alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 5-like	LOC119180255	2	1.09	0.93	0.72
general transcription and DNA repair factor IIH helicase subunit XPB-like	LOC119178783	2	1.08	0.94	0.81
yrdC domain-containing protein, mitochondrial-like	LOC119174804	2	1.08	0.83	0.66
glutathione S-transferase 4-like	LOC119170592	3	1.31	0.97	0.90
isoinhibitor K-like	LOC119172095	3	1.04	0.88	0.77
uncharacterized LOC119163560	LOC119163560	4	1.88	1.00	0.96
high mobility group protein B1-like	LOC119170112	4	1.71	1.00	0.95
mucin-17-like	LOC119175957	4	1.71	0.99	0.68
histone H2A-like	LOC119184461	4	1.57	0.88	0.66
histone H1-delta-like	LOC119179560	4	1.57	0.98	0.97
glutathione S-transferase 4-like	LOC119170592	4	1.53	0.99	0.90
uncharacterized LOC119165065	LOC119165065	4	1.39	0.99	0.90
cell division cycle protein 20 homolog	LOC119170477	4	1.28	0.98	0.70

product	gene	cluster	avg_log2FC	pct.1	pct.2
tubulin alpha-1C chain	LOC119181077	4	1.22	1.00	0.99
uncharacterized LOC119169166	LOC119169166	4	1.20	1.00	0.88
disks large-associated protein 5-like	LOC119177262	4	1.12	0.93	0.60
serine/threonine-protein kinase PLK1-like	LOC119181289	4	1.10	0.92	0.54
secreted acidic protein 1A-like	LOC119164388	4	1.08	1.00	0.96
tubulin beta chain-like	LOC119164422	4	1.07	0.99	0.91
cyclin-dependent kinases regulatory subunit-like	LOC119187792	4	1.06	0.95	0.60
G2/mitotic-specific cyclin-B2-like	LOC119174253	4	1.03	0.94	0.57
uncharacterized LOC119179212	LOC119179212	4	1.02	0.93	0.53
POC1 centriolar protein homolog A-like	LOC119163199	4	1.01	0.93	0.64
kinesin-like protein KIF11	LOC119174001	4	1.00	0.95	0.67
proliferating cell nuclear antigen-like	LOC119187753	5	2.01	0.99	0.73
uncharacterized LOC119163560	LOC119163560	5	1.33	1.00	0.96
DNA replication licensing factor MCM3-like	LOC119170594	5	1.21	0.97	0.67
DNA replication licensing factor mcm7-like	LOC119185306	5	1.21	0.98	0.66
alpha-crystallin A chain-like	LOC119185153	5	1.18	1.00	0.95
neuromodulin-like	LOC119170637	5	1.16	1.00	0.99
DNA replication licensing factor mcm5-like	LOC119166889	5	1.12	0.95	0.61
histone H2A-like	LOC119184461	5	1.04	0.81	0.67
uncharacterized LOC119165065	LOC119165065	5	1.01	1.00	0.90
alpha-crystallin A chain-like	LOC119185157	6	3.29	0.94	0.69
alpha-crystallin A chain-like	LOC119185001	6	3.18	0.94	0.76
alpha-crystallin A chain-like	LOC119185002	6	3.11	0.91	0.73
alpha-crystallin B chain-like	LOC119185148	6	3.05	0.96	0.77
histone H2B	LOC119182249	6	2.73	0.94	0.79
histone H2B	LOC119182246	6	2.38	0.88	0.73
histone H2B	LOC119182243	6	2.35	0.89	0.73
alpha-crystallin A chain-like	LOC119180948	6	2.06	0.83	0.55

product	gene	cluster	avg_log2FC	pct.1	pct.2
histone H2B	LOC119184476	6	2.06	0.86	0.72
uncharacterized LOC119182239	LOC119182239	6	1.94	0.92	0.75
alpha-crystallin A chain-like	LOC119185153	6	1.92	0.98	0.96
heat shock protein HSP 90-alpha-like	LOC119164008	6	1.85	0.99	0.98
heat shock protein 68-like	LOC119180646	6	1.74	0.76	0.51
transcription factor HES-4-B-like	LOC119181002	6	1.70	0.90	0.63
histone H3-like	LOC119182250	6	1.61	0.81	0.68
histone H2A	LOC119166861	6	1.54	0.93	0.89
heat shock protein 68-like	LOC119181076	6	1.51	0.68	0.38
uncharacterized LOC119185067	LOC119185067	6	1.50	0.86	0.64
dual specificity protein phosphatase 10-like	LOC119167068	6	1.42	0.92	0.74
heat shock protein Hsp-16.1/Hsp-16.11-like	LOC119178751	6	1.41	0.96	0.87
histone H4	LOC119167733	6	1.38	0.83	0.72
uncharacterized LOC119177903	LOC119177903	6	1.33	0.87	0.70
5.8S ribosomal RNA	LOC119180443	6	1.33	0.83	0.61
5.8S ribosomal RNA	LOC119187104	6	1.33	0.82	0.61
5.8S ribosomal RNA	LOC119180442	6	1.33	0.82	0.61
5.8S ribosomal RNA	LOC119180445	6	1.33	0.82	0.62
5.8S ribosomal RNA	LOC119180444	6	1.33	0.82	0.62
5.8S ribosomal RNA	LOC119180446	6	1.33	0.82	0.62
histone H2B	LOC119184462	6	1.33	0.75	0.65
histone H4	LOC119166863	6	1.28	0.98	0.99
alpha-crystallin A chain-like	LOC119185154	6	1.25	0.84	0.70
uncharacterized LOC119171672	LOC119171672	6	1.24	0.98	0.96
uncharacterized LOC119183321	LOC119183321	6	1.23	0.92	0.82
histone H2B	LOC119184395	6	1.17	0.80	0.64
uncharacterized LOC119179274	LOC119179274	6	1.17	0.86	0.76
tubulin beta chain	LOC119176503	6	1.13	0.97	0.96

product	gene	cluster	avg_log2FC	pct.1	pct.2
dnaJ protein homolog 1-like	LOC119181188	6	1.11	0.84	0.65
histone H1-delta-like	LOC119179514	6	1.07	0.97	0.98
putative protein TPRXL	LOC119172398	6	1.07	0.97	0.95
replication stress response regulator SDE2-like	LOC119163294	6	1.06	0.94	0.81
motile sperm domain-containing protein 2-like	LOC119159774	6	1.03	0.87	0.71
histone H2A-like	LOC119184461	6	1.03	0.83	0.67
histone H2A-like	LOC119167034	6	1.01	0.96	0.94
glycine-rich cell wall structural protein-like	LOC119164382	7	5.32	0.97	0.70
glycine-rich cell wall structural protein 1-like	LOC119161184	7	5.09	0.99	0.74
glycine-rich cell wall structural protein 2-like	LOC119161168	7	4.68	0.96	0.68
glycine-rich protein DC9.1-like	LOC119161178	7	4.33	0.93	0.65
acanthoscurrin-2-like	LOC119161176	7	4.19	0.99	0.78
glycine-rich protein DOT1-like	LOC119164380	7	3.89	0.99	0.80
abscisic acid and environmental stress-inducible protein-like	LOC119161179	7	3.72	0.99	0.71
protein FAM98B-like	LOC119161187	7	3.55	0.96	0.65
glycine-rich cell wall structural protein 2-like	LOC119164379	7	3.40	0.99	0.81
retinitis pigmentosa 1-like 1 protein	LOC119171236	7	3.36	0.86	0.62
uncharacterized LOC119174219	LOC119174219	7	2.99	1.00	0.82
chorion peroxidase-like	LOC119187450	7	2.87	0.92	0.73
glycine-rich protein DOT1-like	LOC119165325	7	2.51	0.84	0.53
transcriptional regulatory protein LGE1-like	LOC119161185	7	2.25	0.98	0.78
uncharacterized LOC119165569	LOC119165569	7	2.10	0.93	0.68
keratin, type II cytoskeletal 1-like	LOC119161186	7	1.92	0.88	0.52
hemocytin-like	LOC119178529	7	1.80	0.91	0.57
34 kDa spicule matrix protein-like	LOC119187696	7	1.79	1.00	0.96
phenoloxidase-activating factor 2-like	LOC119166682	7	1.66	0.97	0.82
uncharacterized LOC119163773	LOC119163773	7	1.59	0.91	0.67
uncharacterized LOC119178241	LOC119178241	7	1.57	0.79	0.49

product	gene	cluster	avg_log2FC	pct.1	pct.2
putative defense protein	LOC119180975	7	1.55	0.96	0.77
uncharacterized LOC119174349	LOC119174349	7	1.53	0.91	0.65
glycine-rich protein DOT1-like	LOC119164383	7	1.40	0.82	0.53
uncharacterized LOC119172699	LOC119172699	7	1.35	0.84	0.63
isoinhibitor K-like	LOC119172095	7	1.27	0.85	0.78
uncharacterized LOC119173543	LOC119173543	7	1.23	0.99	0.80
U-scoloptoxin(18)-Er1a-like	LOC119163628	7	1.19	0.92	0.74
endochitinase-like	LOC119170761	7	1.15	0.95	0.71
serine protease inhibitor swm-1-like	LOC119163852	7	1.12	0.99	0.85
keratin-associated protein 19-2-like	LOC119171469	7	1.08	0.89	0.84
leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 3	LOC119187185	7	1.05	0.97	0.74
gelsolin, cytoplasmic-like	LOC119159447	7	1.02	0.96	0.85
clotting factor B-like	LOC119169313	7	1.01	0.95	0.67
actin-5C-like	LOC119159731	7	1.00	1.00	1.00
ribonucleoside-diphosphate reductase subunit M2-like	LOC119163338	8	2.12	0.99	0.92
glutathione S-transferase 4-like	LOC119170592	8	1.76	0.99	0.90
cyclin-H-like	LOC119171644	8	1.67	0.98	0.82
uncharacterized LOC119175085	LOC119175085	8	1.57	0.94	0.77
histone-lysine N-methyltransferase eggless-like	LOC119170383	8	1.51	0.96	0.77
succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial-like	LOC119179326	8	1.48	0.99	0.93
GTP-binding protein RHO1-like	LOC119187329	8	1.41	0.92	0.70
uncharacterized LOC119179839	LOC119179839	8	1.31	0.98	0.86
E3 ubiquitin-protein ligase Mdm2-like	LOC119164580	8	1.29	0.98	0.85
cyclin-I-like	LOC119179355	8	1.22	1.00	0.92
KAT8 regulatory NSL complex subunit 2-like	LOC119174764	8	1.21	0.96	0.77
histone H4	LOC119167733	8	1.21	0.80	0.72
glutathione peroxidase-like	LOC119181298	8	1.17	1.00	0.98

product	gene	cluster	avg_log2FC	pct.1	pct.2
cytochrome c oxidase assembly protein cox20, mitochondrial-like	LOC119164187	8	1.07	0.97	0.83
trichohyalin-like	LOC119164848	8	1.06	0.64	0.62
endoplasmic reticulum chaperone BiP-like	LOC119161602	10	3.52	0.97	0.84
mesencephalic astrocyte-derived neurotrophic factor homolog	LOC119163195	10	3.04	0.99	0.81
uncharacterized LOC119186656	LOC119186656	10	3.03	0.97	0.82
endoplasmin-like	LOC119163344	10	2.71	1.00	0.92
endoplasmic reticulum resident protein 29-like	LOC119181310	10	2.64	1.00	0.91
protein disulfide-isomerase A3-like	LOC119159421	10	1.88	1.00	0.96
calreticulin-like	LOC119181385	10	1.57	1.00	0.98
store-operated calcium entry-associated regulatory factor-like	LOC119180723	10	1.54	0.99	0.86
dnaJ homolog subfamily B member 9-like	LOC119173752	10	1.50	0.91	0.62
uncharacterized LOC119177319	LOC119177319	10	1.45	0.86	0.55
protein disulfide-isomerase A6 homolog	LOC119170493	10	1.41	0.97	0.88
BET1 homolog	LOC119181063	10	1.40	0.94	0.78
stearoyl-CoA desaturase 5-like	LOC119187539	10	1.30	0.96	0.76
muscle LIM protein 1-like	LOC119169195	10	1.24	0.93	0.83
uncharacterized LOC119161549	LOC119161549	10	1.24	0.99	0.99
uncharacterized LOC119172334	LOC119172334	10	1.20	0.89	0.77
cytochrome c	LOC119187542	10	1.06	1.00	0.98
hypoxia up-regulated protein 1-like	LOC119172039	10	1.01	0.87	0.59
small nucleolar RNA U3	LOC119165941	11	3.49	0.89	0.65
5.8S ribosomal RNA	LOC119180442	11	2.40	0.76	0.62
5.8S ribosomal RNA	LOC119187104	11	2.40	0.76	0.62
5.8S ribosomal RNA	LOC119180443	11	2.40	0.76	0.62
5.8S ribosomal RNA	LOC119180445	11	2.39	0.76	0.63
5.8S ribosomal RNA	LOC119180444	11	2.39	0.76	0.63
5.8S ribosomal RNA	LOC119180446	11	2.39	0.76	0.63
U2 spliceosomal RNA	LOC119184553	11	2.36	0.97	0.65

product	gene	cluster	avg_log2FC	pct.1	pct.2
U2 spliceosomal RNA	LOC119183045	11	2.26	0.97	0.66
U2 spliceosomal RNA	LOC119183048	11	2.26	0.97	0.66
U2 spliceosomal RNA	LOC119183049	11	2.26	0.97	0.66
U2 spliceosomal RNA	LOC119183050	11	2.26	0.97	0.66
U2 spliceosomal RNA	LOC119183052	11	2.26	0.97	0.66
U2 spliceosomal RNA	LOC119183046	11	2.26	0.97	0.66
U2 spliceosomal RNA	LOC119183047	11	2.26	0.97	0.66
U2 spliceosomal RNA	LOC119184099	11	2.26	0.97	0.66
histone H2B	LOC119182249	11	2.08	0.78	0.80
U6atac minor spliceosomal RNA	LOC119168800	11	1.91	0.76	0.43
U4 spliceosomal RNA	LOC119165942	11	1.72	0.71	0.46
U1 spliceosomal RNA	LOC119165931	11	1.68	0.79	0.47
U1 spliceosomal RNA	LOC119180451	11	1.67	0.73	0.38
U6atac minor spliceosomal RNA	LOC119168799	11	1.61	0.73	0.43
large subunit ribosomal RNA	LOC119180437	11	1.56	1.00	1.00
histone H3-like	LOC119182250	11	1.55	0.70	0.69
U1 spliceosomal RNA	LOC119180452	11	1.53	0.71	0.32
uncharacterized LOC119182239	LOC119182239	11	1.52	0.74	0.76
U1 spliceosomal RNA	LOC119180455	11	1.52	0.73	0.32
U1 spliceosomal RNA	LOC119180453	11	1.52	0.71	0.32
U1 spliceosomal RNA	LOC119180456	11	1.52	0.71	0.32
U1 spliceosomal RNA	LOC119185701	11	1.52	0.71	0.32
U1 spliceosomal RNA	LOC119186176	11	1.52	0.71	0.32
U1 spliceosomal RNA	LOC119187163	11	1.52	0.71	0.32
histone H4	LOC119167733	11	1.43	0.79	0.72
histone H1-delta-like	LOC119179514	11	1.41	0.98	0.98
small nucleolar RNA U3	LOC119165939	11	1.40	0.72	0.59
U1 spliceosomal RNA	LOC119160192	11	1.40	0.77	0.38

product	gene	cluster	avg_log2FC	pct.1	pct.2
U1 spliceosomal RNA	LOC119160191	11	1.40	0.77	0.38
U1 spliceosomal RNA	LOC119186994	11	1.40	0.77	0.39
small nucleolar RNA U3	LOC119170904	11	1.38	0.65	0.43
U2 spliceosomal RNA	LOC119165935	11	1.23	0.72	0.53
alpha-crystallin A chain-like	LOC119180948	11	1.19	0.62	0.57
transcription factor HES-4-B-like	LOC119181002	11	1.17	0.86	0.65
5.8S ribosomal RNA	LOC119173020	11	1.13	0.63	0.54
protein bric-a-brac 1-like	LOC119186484	11	1.07	0.91	0.83
U4 spliceosomal RNA	LOC119165944	11	1.06	0.67	0.47
5.8S ribosomal RNA	LOC119180444	12	5.20	1.00	0.63
5.8S ribosomal RNA	LOC119180445	12	5.20	1.00	0.63
5.8S ribosomal RNA	LOC119180446	12	5.20	1.00	0.63
5.8S ribosomal RNA	LOC119180442	12	5.19	1.00	0.62
5.8S ribosomal RNA	LOC119187104	12	5.19	1.00	0.62
5.8S ribosomal RNA	LOC119180443	12	5.19	1.00	0.62
5.8S ribosomal RNA	LOC119173020	12	4.00	0.91	0.54
small nucleolar RNA U3	LOC119165941	12	3.54	0.84	0.65
large subunit ribosomal RNA	LOC119180437	12	2.52	1.00	1.00
high mobility group protein B1-like	LOC119170112	12	1.68	0.95	0.95
uncharacterized LOC119165065	LOC119165065	12	1.53	0.93	0.90
cell division cycle protein 20 homolog	LOC119170477	12	1.21	0.83	0.72
tubulin alpha-1 chain-like	LOC119186059	12	1.16	0.94	0.96