

Interactions of trypanosomatid *Herpetomonas muscarum*
with dipteran *Drosophila melanogaster*



A thesis submitted for the degree of Doctor of Philosophy
in the University of Oxford Hilary Term 2019

Megan Sloan

St. Edmund Hall

Department of Biochemistry

Abstract

Trypanosomatid parasites are causative agents of important human and animal diseases such as sleeping sickness and leishmaniasis. Most trypanosomatids are transmitted to their mammalian hosts by insects, often belonging to Diptera (or true flies). With resistance to both vector-targeted pesticides and trypanocidal drugs being reported, there is a need for novel transmission blocking strategies to be developed. Studies using the blood-feeding vectors themselves are not broadly accessible, as such, new model systems are being developed to unpick insect-trypanosomatids interactions. In this thesis the interactions between model dipteran *Drosophila melanogaster* and its own natural trypanosomatid *Herpetomonas muscarum* are investigated from the perspective of the parasite. This work describes the sequencing and annotation of the *H. muscarum* genome, which was shown to be highly syntenic to disease-causing trypanosomatids, especially with those of the *Leishmania* genus. Infections of *D. melanogaster* were shown to be confined to adults, where *H. muscarum* persisted in the crop and midgut of the flies for 3-4 days. RNA-sequencing analysis of infected flies showed that rapid, large-scale transcriptomic changes, associated with autophagy, nutrient acquisition and cell surface remodelling, are triggered in *H. muscarum* after ingestion by the fly. Much of the transcriptomic changes observed reflect what is known for disease-causing trypanosomatids. As such, several of the identified surface proteins are of interest as putative transmission blocking targets. Further characterisation of these genes will be greatly aided by genetic manipulation methods developed for *H. muscarum* as part of this work. This thesis has described multiple, evolutionarily conserved, trypanosomatid responses to ingestion by dipterans by using a *Drosophila-Herpetomonas* model. Future studies of these conserved responses will inform development of novel, broad-spectrum transmission blocking strategies.

Acknowledgements

I would like to use this space to convey my sincere thanks to my supervisors: Prof. Petros Ligoxygakis, Dr. Helen Farr and Prof. Keith Gull, for their patience, guidance and encouragement. I am also very grateful to Dr. Eva Gluenz who was never too busy to sit down and troubleshoot an experiment with me or give me advice.

I greatly enjoyed collaborating with Dr. James Cotton, an all-round bioinformatics ‘guru’, and other members of the Berriman Group from the Wellcome Sanger Centre. I would like to sincerely thank them for guiding me through the murky waters of genome annotation and making my visit a great experience.

I feel extremely fortunate to have had excellent scientists for colleagues, lab mates, and friends, during my time in Oxford. I would like to specifically thank Jack Dorling, Magda Atilano, Laura Makin and Jessica Valli for being there through the late nights and failed experiments, as well as the good days.

Finally, I owe my sanity to my brilliant wife Laura and my wonderful family. Without their unwavering support, and occasional food packages, neither this thesis nor its author would exist. Thank you for believing in me, especially when I didn’t myself.

It does truly take a village.

This work was funded by the Wellcome Trust.

Abbreviations

AMP	Antimicrobial Peptide
BLAST	Basic Local Alignment Search Tool
CDS	Coding Sequence
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DD	Destabilisation Domain
DEG	Differentially Expressed Gene
DNA	Deoxyribonucleic Acid
DNA-seq	Deoxyribonucleic Acid sequencing
DSB	Double Strand Break
FP	Fluorescent Protein
gDNA	genomic Deoxyribonucleic Acid
GFP	Green Fluorescent Protein
GNBP	Gram Negative Binding Protein
GO	Gene Ontology
GOI	Gene Of Interest
HAT	Human African trypanosomiasis
kDNA	kinetoplast Deoxyribonucleic Acid
KO	Knockout
LPG	Lipophosphoglycan
LRR	Leucine Rich Repeats
mRNA	messenger Ribonucleic Acid
NO	Nitric Oxide
NTD	Neglected Tropical Diseases

PAMP	Pathogen Associated Molecular Pattern
PCR	Polymerase Chain Reaction
PGRP	Peptidoglycan Recognition Protein
PM	Peritrophic Matrix
PPG	Proteophosphoglycan
PSA	Promastigote Surface Antigen
PSG	Promastigote Secretory Gel
qPCR	quantitative Polymerase Chain Reaction
rDNA	ribosomal Deoxyribonucleic Acid
RBP	RNA-binding protein
RNA	Ribonucleic Acid
RNA-seq	ribosomal ribonucleic acid sequencing
RNAi	Ribonucleic Acid interference
ROS	Reactive Oxygen Species
rRNA	ribosomal Ribonucleic Acid
scoDEG	single copy orthologue Differentially Expressed Gene
SLS	Splice Leader Sequence
SMRT	Single Molecule Real Time
UTR	Untranslated Region
VSG	Variant Surface Glycoprotein
VSP	Variant Surface Protein
WT	Wild Type

Table of Contents

CHAPTER ONE: INTRODUCTION	2
TRYPANOSOMATID PARASITES: AN OVERVIEW	2
Human African Trypanosomiasis	3
Leishmaniasis.....	6
Chagas Disease	10
Treatment and control of trypanosomatid disease	13
DIPTERAN-TRYPANOSOMATID INTERACTIONS	16
Innate immunity in Dipterans: an overview.....	24
Dipteran immune responses to trypanosomatid infection.....	32
The role of the vector microbiome in trypanosomatid infection	35
RATIONALE.....	39
Hurdles to new transmission blocking targets	40
A new dipteran-trypanosomatid model.....	40
AIMS AND STRATEGY.....	41
REFERENCES.....	42
 CHAPTER TWO: THE GENOME OF <i>HERPETOMONAS MUSCARUM</i>	 59
PREFACE	59
The genomes of trypanosomatids	61
RESULTS AND DISCUSSION	64
Genome assembly	64
Genome annotation	67
Genome structure	69
Genome Content	74
Trypanosomatid signatures	85
CONCLUDING REMARKS	91
REFERENCES.....	92
 CHAPTER THREE: THE INFECTION OF <i>DROSOPHILA MELANOGASTER</i> WITH <i>HERPETOMONAS MUSCARUM</i>	 100
PREFACE	100
<i>Herpetomonas</i>	101

<i>Drosophila</i> -trypanosomatid interactions	102
Spatiotemporal distribution of <i>Herpetomonas muscarum</i> in wild <i>Drosophila melanogaster</i> populations	104
RESULTS AND DISCUSSION	109
<i>Morphology of Herpetomonas muscarum</i>	109
<i>Per os</i> infection of <i>Drosophila melanogaster</i> adults	109
Infection of <i>Drosophila melanogaster</i> larvae	126
CONCLUDING REMARKS	127
REFERENCES.....	130
 CHAPTER FOUR: THE TRANSCRIPTOMES OF <i>HERPETOMONAS MUSCARUM</i>	 134
PREFACE	134
Trypanosomatid transcriptomes in their insect hosts.....	135
RESULTS AND DISCUSSION	145
The <i>Herpetomonas muscarum</i> transcriptomes during in vitro culture	145
The <i>Herpetomonas muscarum</i> transcriptome after ingestion by <i>Drosophila melanogaster</i>	154
Does the <i>Herpetomonas muscarum</i> response to ingestion by <i>Drosophila melanogaster</i> resemble the response of medically important trypanosomatids in their vectors?	180
CONCLUDING REMARKS	184
REFERENCES.....	186
 CHAPTER FIVE: DEVELOPMENT OF TOOLS FOR THE STUDY OF <i>HERPETOMONAS MUSCARUM</i>.....	 194
PREFACE	194
Genetic manipulation of trypanosomatids	194
CRISPR.....	197
Genome editing of <i>Herpetomonas muscarum</i>	200
RESULTS AND DISCUSSION	201
Fluorescent <i>Herpetomonas muscarum</i> cell lines	201
Towards a CRISPR-Cas9 genome editing system for <i>Herpetomonas muscarum</i> ...	208
CONCLUDING REMARKS	211

REFERENCES.....	213
CHAPTER SIX: CONCLUSIONS AND OUTLOOK	217
ADDRESSING THE THESIS AIMS	217
THE VIEW FROM THE HOST	218
STRENGTHS AND LIMITATIONS OF THE <i>DROSOPHILA-HERPETOMONAS</i> MODEL	220
FUTURE DIRECTIONS	225
Localisation of infection at ‘higher resolution’	225
Is there another host for <i>Herpetomonas muscarum</i> ?	226
Are the identified surface proteins of interest critical to infection outcome?	227
CONCLUDING REMARKS	228
REFERENCES.....	228
CHAPTER SEVEN: MATERIALS AND METHODS	232
ORGANISMS USED IN THIS WORK	232
<i>Drosophila melanogaster</i>	232
<i>Herpetomonas muscarum</i>	233
CONSTRUCTION OF <i>HERPETOMONAS MUSCARUM</i> EXPRESSION VECTORS	
MICROSCOPY.....	235
<i>DROSOPHILA MELANOGASTER</i> INFECTION PROTOCOLS	243
THE <i>HERPETOMONAS MUSCARUM</i> GENOME	248
Next-generation Genome Sequencing: an overview.....	248
Sequencing and assembly of the <i>Herpetomonas muscarum</i> genome	256
Annotation.....	257
THE <i>HERPETOMONAS MUSCARUM</i> TRANSCRIPTOME.....	258
RNA extraction	258
RNA-sequencing.....	259
REFERENCES.....	260
CHAPTER EIGHT: APPENDIX.....	263
LIST OF SUPPLEMENTARY MATERIALS	263

Figures and Tables

CHAPTER ONE: INTRODUCTION

Figure 1.1 The <i>Trypanosoma brucei</i> life cycle.	5
Figure 1.2 Incidence of new human African trypanosomiasis cases reported to the World Health Organisation (WHO) in Africa 1990-2017.	6
Figure 1.3 The <i>Leishmania</i> lifecycle.	11
Table 1.1 Dipteran vectored neglected tropical diseases and their causative agents.	17
Figure 1.4 The major innate signalling pathways of insects.	30

CHAPTER TWO: THE GENOME OF *HERPETOMONAS MUSCARUM*

Figure 2.1 Average coverage depth of <i>Herpetomonas muscarum</i> scaffolds > 100 kb.	65
Figure 2.2 Genome scope profile of <i>Herpetomonas muscarum</i> scaffolds (kmer size = 21).	66
Table 2.1 Summary statistics of three <i>Herpetomonas muscarum</i> genome annotations using gene models from <i>Leishmania major</i> , <i>Leishmania braziliensis</i> and <i>Trypanosoma brucei</i> and a maximal annotation which combines all three annotations.	69
Figure 2.3 Common motifs of <i>Herpetomonas muscarum</i> strand switch regions as identified by MEME (A) and GLAM2 (B).	71
Table 2.2 Repeat elements identified in the <i>Herpetomonas muscarum</i> genome scaffolds >100kb	73
Table 2.3 A summary of orthogroup assignments of representative trypanosomatid proteomes.	75
Figure 2.4A Venn diagram showing shared orthogroups between <i>Herpetomonas muscarum</i> , <i>Leishmania major</i> and <i>Trypanosoma brucei</i> .	76
Figure 2.4B Phylogenetic tree based on orthogroups containing a single gene from each trypanosomatid used in the analysis.	76

Table 2.4 Transfer RNAs and their codons identified in <i>Herpetomonas muscarum</i> by ARAGORN (Laslett and Canback 2002) and tRNA-ScanSE2 (Lowe and Chan 2016).	84
Table 2.5 Alignment of base pairs 1-40bp of the splice leader sequence of trypanosomatid	86
Figures 2.5 <i>Herpetomonas muscarum</i> scaffolds show synteny with other trypanosomatids.	88
Figure 2.6 Tubulin arrays of <i>Herpetomonas muscarum</i> .	90

CHAPTER THREE: THE INFECTION OF *DROSOPHILA MELANOGASTER* WITH *HERPETOMONAS MUSCARUM*

Figure 3.1A Trypanosomatid incidence in wild <i>Drosophila melanogaster</i> populations in the USA and Canada by month – based on <i>Herpetomonas muscarum</i> mapped DNA-seq reads.	107
Figure 3.1B Trypanosomatid incidence in wild <i>Drosophila melanogaster</i> populations in the USA and Canada by year – based on <i>Herpetomonas muscarum</i> mapped DNA-seq reads.	107
Figure 3.1C Trypanosomatid incidence in wild <i>Drosophila melanogaster</i> populations in the USA and Canada by year - based on <i>Herpetomonas muscarum</i> mapped DNA-seq reads.	108
Figure 3.2 Morphology of <i>H. muscarum</i> cells in axenic culture.	110
Figure 3.3 <i>Herpetomonas muscarum</i> load in <i>Drosophila melanogaster</i> after two hours feeding.	111
Figure 3.4 Pre-staining <i>Herpetomonas muscarum</i> before infection reduced initial infection load.	113
Figure 3.5A <i>Herpetomonas muscarum</i> in the fly digestive tract immediately after feeding.	115
Figure 3.5B <i>Herpetomonas muscarum</i> are found in the fly crop and midgut 6 hours after feeding.	116
Figure 3.5C <i>Herpetomonas muscarum</i> are clustered close to the crop epithelium 6 hours after feeding.	117
Figure 3.6 <i>Herpetomonas muscarum</i> ‘rosettes’ in the fly midgut 24 hours after feeding.	119
Table 3.1 Extraction of <i>Herpetomonas muscarum</i> from <i>Drosophila melanogaster</i> adults for use in ‘secondary’ infections.	122

Figure 3.7 <i>Herpetomonas muscarum</i> loads in <i>Drosophila melanogaster</i> during the ‘secondary infections’.	123
Figure 3.8 Growth curves of <i>Herpetomonas muscarum</i> extracted from <i>Drosophila melanogaster</i> (red) vs. <i>in vitro</i> cultured cells (blue).	123
Figure 3.9A Egg laying is reduced in <i>Drosophila melanogaster</i> fed <i>Herpetomonas muscarum</i> .	125
Figure 3.9B Egg laying is reduced in immune deficient <i>Drosophila melanogaster</i> fed <i>Herpetomonas muscarum</i> .	125
Figure 3.10 <i>Herpetomonas muscarum</i> load in <i>Drosophila melanogaster</i> L2 larvae after two hours feeding.	127

CHAPTER FOUR: THE TRANSCRIPTOMES OF *HERPETOMONAS MUSCARUM*

Table 4.1 Mapping summary of RNA-sequencing reads from <i>in vitro</i> axenic cultures of <i>Herpetomonas muscarum</i> .	145
Figure 4.1 Principal component analysis of differentially expressed <i>Herpetomonas muscarum</i> genes in log phase culture vs. stationary phase culture.	147
Table 4.2 Pfam domains significantly enriched in <i>Herpetomonas muscarum</i> genes upregulated upon entry into stationary phase during axenic culture (vs. log phase).	148
Table 4.3 Significantly differentially regulated cyclins and cyclin-related kinases between stationary and log phase <i>Herpetomonas muscarum</i> .	153
Figure 4.2 Number of RNA-seq reads extracted from whole flies which mapped to <i>Herpetomonas muscarum</i> .	155
Figure 4.3 Principal component analysis of differentially expressed <i>Herpetomonas muscarum</i> genes in log phase culture vs. samples isolated from infection flies at 6, 12 and 18 hours post ingestion.	157
Table 4.5 Significantly enriched Pfam domains in genes upregulated in <i>Herpetomonas muscarum</i> at six hours post ingestion by <i>Drosophila melanogaster</i> (vs log-phase axenic culture).	158
Table 4.6 Significantly enriched Pfam domains in genes downregulated in <i>Herpetomonas muscarum</i> at six hours post ingestion by <i>Drosophila melanogaster</i> (vs log-phase axenic culture).	159

Figure 4.4 Heat map of normalised, log transformed counts for <i>Herpetomonas muscarum</i> orthologues to the <i>Leishmania</i> promastigote surface antigens.	161
Figure 4.5 A Heat map of normalised, log transformed counts for differentially expressed <i>Herpetomonas muscarum</i> proteins with a <i>Giardia</i> variant surface protein (PF03302.13) domain.	162
Figure 4.6 Heat map of normalised, log transformed counts for differentially expressed <i>Herpetomonas muscarum</i> amastin genes.	166
Figure 4.7 Heat map of normalised, log transformed counts for differentially expressed <i>Herpetomonas muscarum</i> proteins with a Cytomegalovirus UL20A (PF05984.12) domain.	167
Table 4.7 Cell cycle-associated proteins differentially expressed in <i>Herpetomonas muscarum</i> upon ingestion by <i>Drosophila melanogaster</i> .	173
Table 4.8 <i>Herpetomonas muscarum</i> gene transcripts significantly upregulated at 12 hours post ingestion (vs. 6 hour post ingestion).	175
Table 4.9 <i>Herpetomonas muscarum</i> gene transcripts significantly upregulated at 12 hours post ingestion (vs. 18 hour post ingestion).	177
Figure 4.8 Venn diagram showing the numbers of genes differentially expressed in <i>Herpetomonas muscarum</i> between two <i>in vitro</i> culture conditions and after ingestion by <i>Drosophila melanogaster</i> .	178
Table 4.10 The top 10 represented Pfam domains in <i>Herpetomonas muscarum</i> ingested by <i>Drosophila melanogaster</i> vs. stationary phase axenic culture.	179
Table 4.11 Summary of co-regulated genes in <i>Herpetomonas muscarum</i> after ingestion by <i>Drosophila melanogaster</i> and <i>Leishmania major</i> after ingestion by <i>Phlebotomus duboscqi</i> .	181
Table 4.12 Summary of co-regulated genes in <i>Herpetomonas muscarum</i> after ingestion by <i>Drosophila melanogaster</i> and <i>Trypanosoma brucei</i> after ingestion by <i>Glossina morsitans</i> .	184

CHAPTER FIVE: DEVELOPMENT OF TOOLS FOR THE STUDY OF *HERPETOMONAS MUSCARUM*

Figure 5.1 Double stranded DNA repair by homologous recombination.	196
Figure 5.2 Knockout and tagging of a gene of interest (GOI) by homologous recombination.	199

Figure 5.3 Creation of double stranded breaks in DNA by Cas9	200
Figure 5.4 Transformation of <i>Herpetomonas muscarum</i> with pNUS-GFPcB (Tetaud et al. 2002).	202
Table 5.1 Conditions used for transfection of pNUS-GFPcB into <i>Herpetomonas muscarum</i> and <i>Leishmania mexicana</i> .	203
Figure 5.5 Transformation of <i>Herpetomonas muscarum</i> with pMS003.	206
Table 5.2 A novel set of plasmids for genetic manipulation of <i>Herpetomonas muscarum</i> .	207
Figure 5.6 Growth curves of wild type and transgenic tdTomato <i>Herpetomonas muscarum</i> cell lines cultured <i>in vitro</i> .	207
Figure 5.6 pMS007 for expression of a destabilised Cas9 in <i>Herpetomonas muscarum</i> .	210
Figure 5.7 pMS007 is maintained in <i>Herpetomonas muscarum</i> for > 10 passages.	210

CHAPTER SIX: CONCLUSIONS AND OUTLOOK

No figures or tables.

CHAPTER SEVEN: MATERIALS AND METHODS

Table 7.1 <i>Drosophila melanogaster</i> stocks used for this work	232
Table 7.2 Fly food component list	233
Table 7.3 Determining required concentration of antibiotics to kill <i>H. muscarum</i> culture, seeded at 1×10^6 cells ml^{-1} .	234
Table 7.4 Primers for construction of novel expression vector pMS003.	236
Table 7.5 Luria broth agar plate recipe.	237
Table 7.6 Primers for construction pMS004-7	239
Table 7.7 HindIII/SpeI digest of pMS003 to release tdTomato fragment.	240
Table 7.8 DNA fragment amounts and volumes used for NEBuilder reactions to change the tdTomato open reading frame of pMS003 for alternative genes.	240

Table 7.9 DNA fragment amounts and volumes used for NEBuilder reactions to produce pMS004-8	241
Table 7.10 Reaction mix (per sample) for quantification of <i>Herpetomonas muscarum</i> in <i>Drosophila melanogaster</i> from gDNA extractions.	246
Table 7.11 qPCR cycling condition for quantification of <i>Herpetomonas muscarum</i> in <i>Drosophila melanogaster</i> from gDNA extractions.	246
Figure 7.1 A representative standard curve from the <i>Herpetomonas muscarum</i> gDNA standards and an uninfected fly control.	247
Table 7.12 Apple juice agar plate reagents.	248
Figure 7.2 Illumina Sequencing.	249
Figure 7.3 Single molecule real time (SMRT) sequencing.	255

CHAPTER ONE

Introduction

CHAPTER ONE:

Introduction

Trypanosomatid parasites: an overview

Trypanosomatids are single-celled, flagellated, protozoan parasites of the class Kinetoplastea. The majority are confined to insect hosts, where they appear to have a high prevalence – at least 22% of Heteroptera ('true bugs') harbour trypanosomatids (Maslov *et al.*, 2007). However, several trypanosomatids are pathogenic to vertebrates and plants. In humans, trypanosomatid diseases are some of the leading causes of chronic disability and poverty in low- and middle-income countries (Hotez, 2008, 2009). There are two trypanosomatid genera responsible for human diseases: *Trypanosoma* and *Leishmania*. These parasites are transmitted to humans via the bite of an infected insect. Trypanosomatid parasites are not exclusively a human 'problem', they also infect a range of wild and domestic animals, which are important parasite reservoirs. Infections of cattle, pigs and horses cause major economic losses and are a significant burden on agriculture.

The first part of this chapter aims to provide an overview of trypanosomatid disease and how these parasites interact with their insect vectors. Insects play critical roles in the spread of trypanosomatid diseases. Without the ability to survive in the insect vector, trypanosomatid parasites would be significantly less of a burden to global health. As such, studies of insect-parasite interactions, and application of this knowledge to develop transmission blocking strategies, are vital to the eradication of these diseases.

Human African Trypanosomiasis

Human African Trypanosomiasis (HAT), also known as ‘sleeping sickness’, is caused by two subspecies of *Trypanosoma brucei*: *gambiense* and *rhodesiense*. Both subspecies are transmitted by flies of the *Glossina* genus, which are commonly known as tsetse. These hematophagous flies, native only to sub-Saharan Africa, are grouped into species and subspecies which correspond to their habitat niches: savanna (*morisitans*), forest (*fusca*) and river/lake habitats (*palpalis*) (Dorn *et al.*, 2017). The two *T. brucei* subspecies have different epidemiology and disease symptoms. *T. b. rhodesiense* infections, prevalent in East Africa, are acute and have a rapid onset, whereas, *T. b. gambiense* causes comparatively slower, chronic disease in West and Central Africa. *T. b. gambiense* is responsible for 98% of human sleeping sickness cases (WHO, 2018b).

The lifecycle of *T. brucei* is complex and involves coordinated differentiation between many cell morphotypes in both insect and mammalian hosts (Figure 1.1). Mammalian blood contains two cell morphologies: a slender actively replicating form and ‘stumpy’ form which has undergone cell cycle arrest. The ratio of these two forms is cell-density dependant, with accumulation of a ‘stumpy initiation factor’ (SIF) correlated with differentiation (Reuner *et al.*, 1997). The SIF factor signal appears to be relayed to the cell via a surface-expressed G protein-coupled receptor-like protein of the GPR89-family receptor, *TbGPR89* (Rojas *et al.* 2019). Only the stumpy type cells are able to survive in the tsetse midgut, where they differentiate to give procyclic trypomastigotes which actively proliferate by binary fission (MacGregor *et al.*, 2012). To avoid simply passing through the digestive tract and being voided with the faeces, the trypanosomes must escape the blood meal which is encased in a chitinous sac called the peritrophic matrix (discussed further below) and migrate anteriorly towards the salivary glands of the fly. A

week after ingestion, elongated, mesocyclic trypanosomes can be seen packed in the proventriculus in large numbers, though crossing the peritrophic matrix can much sooner – shortly after ingestion (Gibson and Bailey, 2003; Sharma *et al.*, 2008). The trypanosomes escape the peritrophic matrix and enter the ectoperitrophic space. The cells migrate further to the insect's salivary glands where they are able to attach themselves to the gland wall and replicate – with sexual genetic exchange reported in these epimastigote forms (Peacock *et al.*, 2011, 2014). Finally, a cell cycle arrested (detached) form, the metacyclic, is produced by a further differentiation. It is these cells which are able to infect vertebrates when deposited into the skin at subsequent tsetse biting events. The trypanosome lifecycle in the fly takes approximately 4 weeks, but infected flies continue to produce infective metacyclic trypanosomes for the remainder of their life span (Gibson and Bailey, 2003). From the skin the metacyclics can invade many tissues including the lymphatic system, adipocytes and bloodstream where they differentiate once more into slender blood stream forms (Casas-Sánchez and Acosta-Serrano 2016).

Early symptoms of *T. brucei* infection include localised inflammation at the insect bite site, headache, fever and enlarged lymph nodes. The circulatory system allows the trypanosomes to reach many sites around the body. Crossing the blood-brain barrier during later stages of infection, and entry into the central nervous system causes significant neurological problems which can lead to coma. Without treatment the disease is fatal. Vector control and human drug treatments (see below) have been largely successful in the control of HAT with efforts on track to eliminate human infections by 2030 (Figure 1.2). However, it is likely that infections of cattle and other animals will take much longer to eliminate. Furthermore, Leishmaniasis and Chagas disease, both also caused by trypanosomatids, remain a large burden to global health.

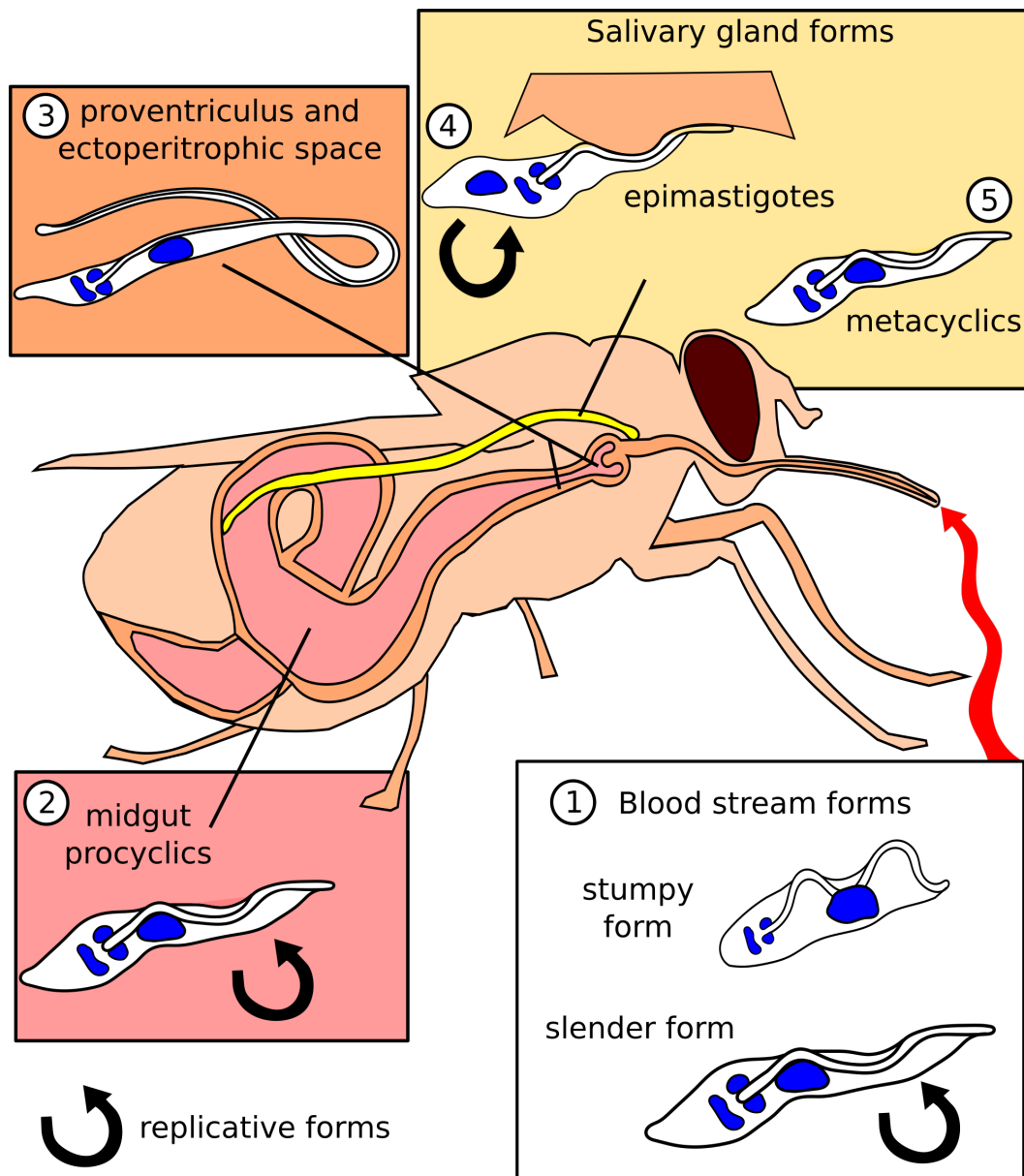


Figure 1.1 – The *Trypanosoma brucei* life cycle. Tsetse ingest the two *T. brucei* blood stream forms (1) with the blood meal. In the insect midgut the slender forms are killed, and the stumpy forms differentiate into replicative procyclics (2). The trypanosomes cross the newly formed peritrophic matrix in the proventriculus (3). They then migrate to the fly's salivary glands, via the ectoperitrophic space, where epimastigotes (4) attach themselves to the salivary gland walls and proliferate. Further differentiation gives infective metacyclic forms (5) which are able to infect vertebrates upon deposition into subsequent bite sites, thus completing the life cycle.

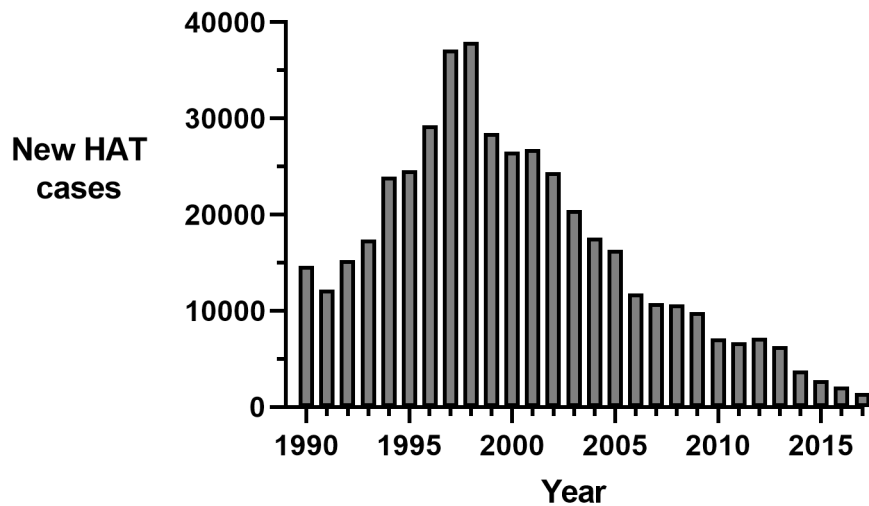


Figure 1.2 – Incidence of new human African trypanosomiasis cases reported to the World Health Organisation (WHO) in Africa 1990-2017. This graph was produced from data from the WHO global health observatory data repository and includes cases from Botswana, Burundi, Ethiopia, Kenya, Malawi, Mozambique, Namibia, Rwanda, Eswatini, Uganda, United Republic of Tanzania, Zambia and Zimbabwe.

Leishmaniasis

Leishmaniasis is a compendium of diseases which are endemic in 85 territories across the Americas, Asia, Northern Africa and Europe. Chronic infections, such as cutaneous and muco-cutaneous Leishmaniasis, are most common. During these infections the parasites cause lesions on the skin and/or erosion of the mucosal surfaces of the nose, mouth and throat. Whilst these lesions are often self-limiting, they can be extremely debilitating.

Visceral leishmaniasis (VL), or kala-azar, is a more acute form of disease where infections of the liver and/or spleen are fatal in 95% of untreated cases (WHO, 2017).

Occurrence of opportunistic visceral infections in immunocompromised individuals, such as HIV+ patients (Lindoso *et al.*, 2016), is increasing - with the most extreme cases

evident in Brazil where VL-HIV co-infections increased from 0.7% of total HIV infections in 2001 to 8.5% in 2012 (Lindoso *et al.*, 2014). These infections are difficult to cure due to the high toxicities of anti-trypanosomal drugs, which are unsuitable for use on immunocompromised patients. In contrast to the successes in the control of sleeping sickness, Leishmaniasis remains a significant burden to global health. The World Health Organisation estimates over 1.6 million new cases of Leishmaniasis, including 0.4 million cases of visceral disease, occur annually (Alvar *et al.* 2015, Georgiadou *et al.* 2015) and a further 1 billion people are at risk.

There are over 20 *Leishmania* species capable of causing disease in humans. These are transmitted by female sand flies from two genera, *Phlebotomus* and *Lutzomyia*. Whilst over 800 sand flies species have been described, only 98 species are proven or suspected vectors of *Leishmania* parasites (Maroli *et al.*, 2013). They, like several neglected tropical disease vectors (NTD) vectors, belong to the Nematocera. This insect suborder's name is derived from the Greek words for 'thread' and 'horns'- which describes this groups thin, segmented antennae (McAlister, 2017). These long, slender flies are widely distributed across the earth's subtropical and temperate regions. Adult sand flies of both sexes feed on sugar-rich plant fluids, with the females requiring a blood meal for egg production.

As with *T. brucei*, there are several discrete stages to the *Leishmania* life cycle.

Amastigotes enter the sand fly midgut in the blood meal which is pumped into the flies' midgut via the stomodeal valve (Figure 1.3). Here, the blood meal is enveloped by the peritrophic matrix which protects the gut epithelium from the recent meal. Within 18 hours of ingestion the *Leishmania* amastigotes differentiate into procyclic promastigotes which replicate rapidly as the meal is digested (Dostálová and Volf, 2012). It takes up to

72 hours for the blood meal to be digested and excreted. Promastigote cells differentiate into non-proliferative, elongated and highly motile nectomonad cells (Gossage, Rogers and Bates, 2003). To prevent being voided with the digested meal the nectonomads must circumvent the peritrophic matrix and attach to the microvilli of the midgut epithelium. The expression of chitinases may allow the *Leishmania* cells to break down and cross the peritrophic matrix (Coutinho-Abreu *et al.* 2010, Rogers *et al.* 2008), however Sádlová and Volf (2009) showed that the breakdown of the peritrophic matrix did not differ between *Leishmania* infected and uninfected sand flies. Recent work by Pruzinova *et al.* 2015 demonstrated the timing of the fly-mediated peritrophic matrix breakdown was crucial to the maturation of *Leishmania* infection (discussed further later in the Chapter).

They achieve this by intercalating their flagellum between the microvilli of the gut epithelium. Molecular components of these interactions are thought to include the major parasite surface glycoconjugate lipophosphoglycan (LPG) and galectin present at the cell surface of the cells of the sand fly epithelium (Pimenta *et al.*, 1992; Kamhawi *et al.*, 2004), although other works report non-LPG mediated attachment can be used by several *Leishmania* species (Rogers *et al.*, 2004; Svobodova *et al.*, 2006; Myskova *et al.*, 2007). Further work is needed to identify and characterise these putative receptor-ligand pairings during this critical stage of infection.

For the next several days the nectomonads migrate anteriorly towards the stomodeal valve, during this migration the cell body shortens into the leptomonad cell type which secrete a protective mucin-like gel (Stierhof *et al.*, 1999). At this stage several cells, termed haptomonads, can be found attached to the cuticle of the stomodeal valve. This attachment is different to midgut epithelium attachment, with expansion of the flagellar

tip into hemi-desmosome-like structures (Vickerman and Tetley, 1990; Wakid and Bates, 2004; Bates, 2007). The final *Leishmania* cell type found in sand flies are the highly motile metacyclic forms which have an elongated flagellum and a thin, short cell body. It is not known if the metacyclics originate from leptomonads or nectonomads cells (or both), but it is these cells which are deposited into the vertebrate skin at subsequent biting events. This deposition is thought to be via regurgitation triggered by the promastigote secretory gel (PSG) secreted by the leptomonads in the stomodeal valve (Rogers *et al.*, 2011). Interestingly, when the sand fly ingests another blood meal the metacyclic cells have been shown to revert back to the leptomonad cell stage (Serafim *et al.* 2018). These ‘retroleptonomads’ proliferate rapidly and then differentiate to produce increased numbers (>100-fold more) of infective metacyclics. Therefore, the availability of blood sources and the frequency of feeding in sand flies is likely to be a critical, and overlooked aspect of Leishmaniasis transmission.

Unlike *T. brucei*, *Leishmania* species now enter an intracellular stage of the life cycle – the amastigote. After entry into the vertebrate the *Leishmania* metacyclics are phagocytosed by mononuclear phagocytes, such as macrophages or neutrophils.

Enveloped in the parasitophorous vacuole, where the higher temperatures and acidic pH trigger differentiation into the amastigote cell cycle stage, the amastigotes undergo repeated rounds of replication (Berman, Dwyer and Wyler, 1979). Egression of the amastigotes allows infection of further phagocytes (Real *et al.*, 2014) spreading the infection.

Chagas Disease

Chagas disease, named after its discoverer physician Carlos Chagas, is caused by *Trypanosoma cruzi*. It is estimated that 6-7 million people are infected globally, with over 25 million at risk of infection (WHO, 2018a). The majority of cases are found in Latin America where the insect vectors, blood-feeding triatomine bugs from the *Triatoma*, *Rhodnius*, and *Panstrongylus* genera are commonly found. These insects are found in dark crevices in human dwellings during the day and emerge at night to blood-feed. Due to enhanced global travel, Chagas disease is being reported in several countries across Europe, with transmission events occurring without the insect vectors e.g. via blood donation, organ transplants and from mother-to-child (Strasen *et al.*, 2014). Recently implemented blood screening programs have been shown to be very effective and are vital to prevent the spread of disease (Requena-Méndez *et al.*, 2017).

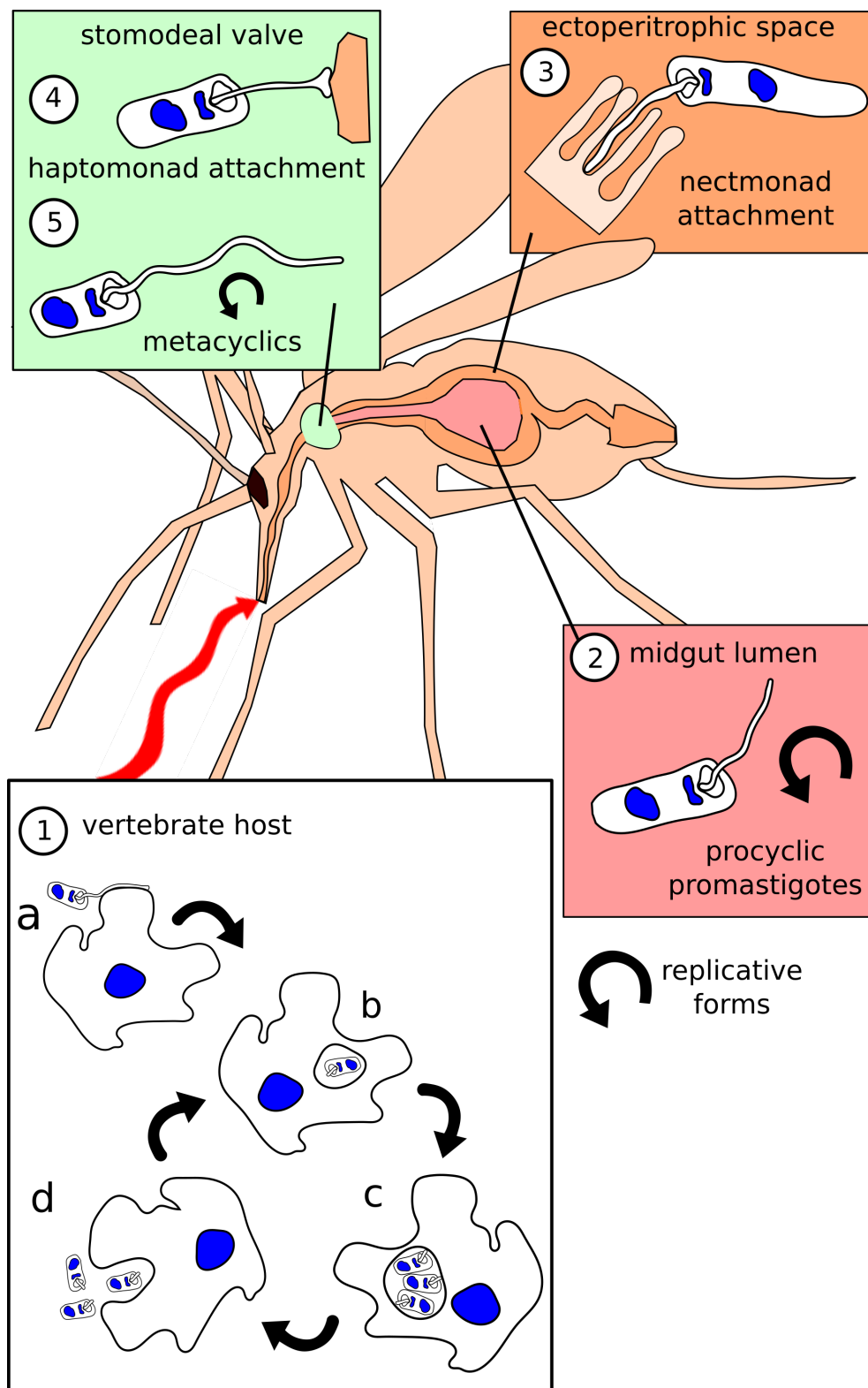


Figure 1.3 – The *Leishmania* lifecycle. Full legend on following page.

Figure 1.3 – The *Leishmania* lifecycle. *Leishmania* metacyclics deposited at the sand fly bite site are phagocytosed by mononuclear phagocytes, such as macrophages (1a). Here they differentiate into amastigote forms (1b) which replicate within parasitophorous vesicle (1c) before egression into the surrounding tissues (1d). Female sand flies ingest amastigotes in the blood meal, which differentiate into procyclic promastigotes that replicate in the fly midgut (2). Procyclics differentiate to elongated, motile nectonoms which escape into the ectoperitrophic space and attach to the midgut epithelium by the flagellum (3). Leptomonad cells (not shown), which secrete PSG (green), form as cells migrate anteriorly towards the stomodeal valve. In the stomodeal valve haptomonads attach to the stomodeal valve cuticle via the flagellar tip (4). Finally, cells differentiate to give the vertebrate-infective, proliferative metacyclics with a short cell body and elongated flagellum (5).

Chagas disease has two phases. For the first few months post-infection, there is an acute phase where *T. cruzi* cells can be found in the circulatory system. This phase is difficult to diagnose with common symptoms including localised inflammation at the bite sites and fever. Occasionally patients have inflammation in heart muscle and brain lining. The long-term, chronic phase of infection can be wholly asymptomatic with very few parasites found circulating in the host. This is a significant challenge to attempts to treat and eliminate disease, as many individuals are unaware they are infected. Several years after the initial infection around 40% of people enter a chronic stage of infection which is characterised by cardiomyopathies and/or intestinal issues, due to parasite presence and associated tissue damage and inflammation in the heart and digestive muscles (Benvenuti *et al.*, 2003).

In comparison to the other aforementioned trypanosomatids, *T. cruzi* has a simple life cycle. After ingestion and entry into the midgut of the insect vector, *T. cruzi* bloodstream stage cells differentiate into proliferative epimastigotes. It was thought that epimastigotes were unable to cause infection in humans owing to their susceptibility to complement-mediated killing (Romano *et al.*, 2012). However more recent studies have shown that epimastigotes remain infective for several days after differentiation (Kessler *et al.*, 2017).

The cells pass through the insect gut and into the rectum of the insect, here they undergo metacyclogenesis to give human-infective metacyclic trypomastigotes, which are excreted by the insect. The insect faeces often enter open wounds and insect bite sites, as the insects often defecate during, or soon after feeding (Reisenman *et al.*, 2011). This allows the trypomastigotes to enter the new vertebrate host. Similar to *Leishmania* species, *T. cruzi* has an intracellular amastigote cell type which proliferates within membrane-bound compartments inside phagocytes and muscle cells of vertebrate hosts.

Treatment and control of diseases caused by trypanosomatid parasites

Vaccines for trypanosomatid infections are as yet unavailable. However, candidates for transmission blocking vaccines - immunizations aimed preventing parasite development within the vector, are emerging (Coutinho-Abreu and Ramalho-Ortigao 2010). Sera from mice immunised with *Leishmania major* LPG and gp63, have all been shown to inhibit *L. major* development in *P. papatasi* (Tonui *et al.* 2001). Whilst encouraging, the reduced *L. major* development in this case was likely due to damage of the sand fly midgut epithelium caused by generic 'immune-active substances' present in the sera – rather than a specific response. The development of *L. major* in sand flies can also be 'blocked' using sera from mice immunized with sand-fly derived molecules, including the chitinase

PpChit1 (Ramalho-Ortigao *et al.* 2005) and a galectin (Kamhawi *et al.* 2004) from *Phlebotomus papatasi*. Encouragingly the PpChit1 anti-sera also demonstrated efficacy against other sand fly/*Leishmania* combinations. Therefore, these vaccines could provide an important, broad spectrum strategy to combat disease.

A small number of drug treatments are available for trypanosomatid diseases. Until recently there were four drugs available for the treatment of HAT: suramin, pentamidine, melarsoprol and eflornithine (Steverding, 2010). The acute stages of *T. cruzi* infections are largely treated by either benznidazole or nifurtimox (Bern *et al.*, 2007). Whilst paromomycin, miltefosine, and liposomal amphotericin B are currently used to treat *Leishmania* infections (Azevedo Cardoso *et al.*, 2014). These drugs are often expensive, require long-term intravenous dosage, are highly toxic and have variable efficacies against different trypanosomatid species and life cycle stages (Ponte-Sucre *et al.*, 2017). Given these diseases affect the some of the poorest regions globally, such properties reduce their accessibility and suitability to those who need them most. Furthermore, resistance is emerging to several drugs (Basselin *et al.*, 2002). Pentavalent antimonials, once the standard leishmaniasis treatment, are no longer efficacious in Bihar – where over half of the worlds visceral leishmaniasis cases are found (Chakravarty and Sundar, 2010; Nagle *et al.*, 2014; Ponte-Sucre *et al.*, 2017).

Treatment options are improving, with several new treatments in development by the Drugs for Neglected Diseases initiative (DNDi). Fexinidazole, an orally administered drug which is effective after 10 days of treatment, has been shown to be safe and effective for late-stage human *T. b. gambiense* infections (Mesu *et al.*, 2018). As this drug could potentially be taken in the home, it is hoped that fexinidazole will reduce the requirement

for hospitalisation of HAT patients. A new Phase II/III study, based in Malawi, has been approved, to assess fexinidazole to treat HAT caused by *T. b rhodesiense* (Baker and Welburn, 2018).

With few good treatment options available, there has been much focus on disease prevention by controlling and reduction of the vector insect populations. The majority of vector control strategies are reliant on spraying buildings with pesticides (or pesticide impregnated paints) with long residual half-lives. Whilst this is effective, long-lasting pesticides are expensive, application is labour intensive and requires proper training. ‘Chemical-free’ methods of vector population control include the release of sterilised male flies. More recently developed transmission blocking strategies are based around modifications to the vector microbiome to make the midgut environment less habitable for the trypanosomatids. ‘Paratransgenesis’ involves genetic modification of bacteria present in the vector midgut to highly express anti-pathogen proteins such as anti-microbial peptides and then reintroducing the modified vectors into wild populations (Medlock *et al.*, 2013; De Vooght, Van Keer and Van Den Abbeele, 2018). A recent study showed that inoculation of tsetse with *Kosakonia cowanii*, which confers resistance to the malaria parasite to mosquitos, made tsetse significantly more refractory to both *T. congolense* and *T. brucei* (Weiss *et al.*, 2019).

Utilization of the cytoplasmic incompatibility phenomenon caused by natural tsetse bacterial symbionts *Wolbachia* may improve the efficacy of these strategies. These bacteria are found in the reproductive organs of the flies and can cause developmental arrest of embryos when incompatible matings occur. *Wolbachia* infected males and uninfected females do not produce fertile offspring. *Wolbachia* infected females however,

are able to mate with uninfected males or those harbouring *Wolbachia* of the same genotype. As such reproductive advantages gained by *Wolbachia* infection could be used to drive the modified symbionts through natural populations (Alam *et al.*, 2011).

Other transmission blocking developments include feeding the insect vectors plant derived compounds which break down to produce trypanocidal toxins when digested (Ferreira *et al.*, 2018). Whilst these studies are in early stages, they have been shown to reduce trypanosomatid survival and could be an important transmission blocking strategy in the future (Hurwitz *et al.*, 2014; Gilbert *et al.*, 2016).

Dipteran-trypanosomatid interactions

As with a third of the world's neglected tropical diseases (NTDs), *Trypanosoma brucei* and the *Leishmania* are transmitted by dipteran, or 'true fly', vectors (Table 1.1).

Members of the order Diptera have a single pair of wings and are holometabolous – meaning they undergo a complete metamorphosis during their life cycle. There are three suborders which make up the Diptera: the Nematocera, the Brachycera and the Cyclorrhapha. As previously mentioned, the vectors for Leishmaniasis, the sand flies, belong to the Nematocera along with mosquitos and biting midges. Tsetse, on the other hand, belong to stockier, more robust flies of the Brachycera ('short horns') along with house flies, horse flies and fruit flies.

Disease	Causative Agent (s)	Vector(s)
Lymphatic filariasis	<i>Wuchereria bancrofti</i> , <i>Brugia malayi</i> , <i>Brugia timori</i>	<i>Aedes spp.</i> , <i>Anopheles spp.</i> , <i>Culex spp.</i> , <i>Mansonia spp.</i>
Onchocerciasis	<i>Onchocerca volvulus</i>	<i>Simulium spp.</i>
Human African trypanosomiasis (HAT)	<i>Trypanosoma brucei rhodesiense</i> , <i>Trypanosoma brucei gambiense</i>	<i>Glossina spp.</i>
Leishmaniasis	<i>Leishmania spp.</i>	<i>Phlebotomus sp.</i> <i>Lutzomyia sp.</i>
Dengue and Chikungunya	DEN-1, DEN-2, DEN-3, DEN-4, Chikungunya virus	<i>Aedes aegypti</i>

Table 1.1 – Dipteran vectored neglected tropical diseases and their causative agents.

Somewhat paradoxically, given their importance as vectors, both tsetse and sand flies are overwhelmingly resistant to trypanosomatid infection – largely thanks to conditions in the fly midgut (Lehane and Msangi, 1991; Lehane, 1997). Here, there are several significant barriers to survival which are so severe that the rate of successful infections established is often extremely low. Even at its most susceptible, at the first blood meal following eclosure from the puparium, only ~50% of *T. brucei* infections are able to become established. This proportion declines rapidly as the adult tsetse matures, from the third blood meal onwards less than 10% of challenged flies remain infected (Lehane, Aksoy and Levashina, 2004). This resistance to infection is mirrored in sand flies, typically fewer than 5% of wild sand flies sampled are found to contain *Leishmania* parasites (Akhoundi *et al.*, 2013; Tiwary *et al.*, 2013; Maia *et al.*, 2015) – though a higher prevalence has been reported in refugee camps in Greece (Fotakis *et al.*, 2019).

The insect defences, which confer such strong resistance to invading trypanosomatids and other pathogens, are complex and multi-layered. They involve not only host derived

factors but are greatly impacted by other organisms present in the insect's microbiome. I will give a brief overview of the dipteran defences to invading pathogens and discuss the key interactions between these defences and trypanosomatid parasites.

The peritrophic matrix acts as a structural barrier

Ingested meals are compartmentalised in the insect digestive system by the peritrophic matrix (PM), a semipermeable 'mesh' sac principally comprised of glycoproteins, chitin and gluco/galacto-saminoglycans (GAGs) (Lehane, Allingham and Weglicki, 1996; Hao and Aksoy, 2002; Rose *et al.*, 2014; Rogerson *et al.*, 2018). The PM is functionally analogous to the layer of mucus which coats the mammalian gut, protecting the gut epithelium from both mechanical damage, toxins and invading microorganisms (Pascoa *et al.*, 2002; Devenport *et al.*, 2006; Kuraishi *et al.*, 2011). There are two categories of PM. The type I PM is produced by delamination of epithelial cells along the length of the gut – often in response to feeding. Sand flies, the vectors of Leishmaniasis, exhibit the type I PM. The tsetse PM falls into a second category. Type II PMs are secreted continuously from a specialised group of cells within the proventriculus. The tsetse PM is approximately 300nm thick and is formed from three layers with the outer layers thought to contain high concentrations of GAGs (Miller and Lehane, 1993). Descriptions of the PM from sand flies indicate a two-part structure, with thin laminar layers close to the gut epithelium with thick amorphous layers in contact with the blood meal (Secundino *et al.*, 2005; Sádlová and Volf, 2009).

The structural core of the PM is formed from chitin fibres and proteins called peritrophins (Shen and Jacobs-Lorena, 1999). The chitin binding domains of peritrophins facilitate cross-linking of chitin fibres which determine PM strength, pore size and elasticity.

Several peritrophins also contain mucin domains which may serve to increase resistance to proteolytic enzymes or allow water retention. The importance of the PM during *Leishmania* infections was demonstrated by knockdown of peritrophin Pper1 in *Phlebotomus papatasi*. Knock-down of this single gene resulted in an approximately 40% increase in *Leishmania major* loads at 48 hours after feeding (Coutinho-Abreu *et al.*, 2013). The case is similar for trypanosome infections in tsetse. RNAi knock down of peritrophins pro1/2 and chitin synthase resulted in tsetse with compromised PMs, this phenotype was associated with increased susceptibility to trypanosome infection (Weiss *et al.* 2014).

As well as presenting a physical barrier, several PM components have been demonstrated to play a more active role in protection against invading trypanosomatids. Carbohydrate binding, calcium dependant (C-type) lectins and lectin-domain containing proteins with extensive glutamic-acid proline repeats, have been identified in aggregates of trimers in the *Glossina* peritrophic matrix and throughout the midgut (Haines *et al.*, 2005; Weiss *et al.*, 2013). The latter are structurally similar to the EP proteins which dominate the surface of *T. brucei* during the midgut stages of infection (discussed below). Tsetse EP protein expression is increased in both bacterial and trypanosome infections, and knock down of the proteins, by RNAi, resulted in significant increases in *T. brucei* midgut colonisation rate (Haines *et al.*, 2010). Additionally, tsetse EP protein levels decline when the fly is starved, which may explain increased susceptibility to trypanosome infection in starved tsetse (Kubi *et al.*, 2006). Whilst the specifics of the interaction are unknown, tsetse EP proteins are thought to interact with *T. brucei* surface glycans via a lectin domain and facilitate the aggregation of the trypanosomes, trapping them in the peritrophic matrix and endoperitrophic space (Rose *et al.*, 2014). This entrapment may

facilitate trypanosome killing by proteolytic digestive enzymes or simply allow the invaders to be expelled upon defecation. Weiss *et al.* (2013) also identified a strictidine synthase in the tsetse peritrophic matrix. In other insects, this enzyme has been shown to be involved in the synthesis of antimicrobial alkaloids (de A Gonzaga *et al.*, 2003).

Whilst the PM does provide an important protective layer for the midgut epithelium, it may also inhibit trypanosomatid detection and delay responses to infection. Weiss *et al.* (2014) showed that expression of the antimicrobial peptide attacin was delayed in trypanosomes with an intact PM, compared to tsetse whose PM was perturbed by RNAi knockdown of key structural components.

Tsetse continuously secrete peritrophic matrix, from cells in the proventriculus, regardless of digestion status (Sádlová and Volf, 2009; Rose *et al.*, 2014). This observation has been linked to the tsetse's bacterial endosymbionts (see below) as aposymbiotic flies have no, or a severely compromised, peritrophic matrix (Weiss *et al.*, 2013). Multiple routes to the ectoperitrophic space have been suggested for *Trypanosoma brucei* including: penetration of the newly secreted PM in the proventriculus (Freeman, 1970, 1973; Ellis and Evans, 1977), penetration of more developed or 'mature' PM in the midgut (Gibson and Bailey, 2003) or circumvention of the PM in a similar way to described in *Leishmania* in the posterior midgut (see below) (Taylor, 1932). There are multiple reports of trypanosomes embedded in the midgut PM, apparently in the process of penetrating through it. However, a recent study using high resolution microscopy techniques provides compelling evidence for a penetration of the still-forming PM in the proventriculus of the tsetse fly (Rose *et al.*, 2019). The group observed trypanosomes embedded in 'cysts' between intact PM layers in the tsetse midgut – indicating

incomplete ‘penetration’ through the chitinous PM layer. The morphology of the trapped trypanosomes was elongated and appeared similar to cells found in the proventriculus, rather than the midgut procyclic cell form. As such it appears the trypanosomes in these cysts have passed through the immature PM forming in the proventriculus, but then became trapped between layers forming a cyst. These cysts have then been moved more posteriorly by peristaltic movements in the gut which move the continuously secreted PM through the gut. Therefore, it appears trypanosomes in the midgut PM are not penetrating through to the ectoperitrophic space, but are simply cells that became trapped between the still forming PM layers in the proventriculus.

In sand flies, the PM is secreted by the entire midgut when triggered by the gut distention caused by ingestion of a meal. The timing of PM secretion, and its degradation after digestion of a meal, is species specific (Pruzinova *et al.*, 2015). It was previously thought that *Leishmania* used chitinases to break up the peritrophic matrix and escape into the ectoperitrophic space (Yosef *et al.*, 1991; Shakarian and Dwyer, 2000; Rogers *et al.*, 2008). Indeed, episomal over-expression of chitinases of *Leishmania mexicana* during *Lutzomyia longipalpis* infection did lead to early PM escape, increased stomodeal damage and parasite loads (Rogers *et al.*, 2008). However, timing of the PM opening in *Phlebotomus duboscqi* is the same in both *Leishmania major* infected and uninfected females (Sádlová and Volf, 2009), with parasite escape from the posterior of the PM coinciding with sand fly mediated PM break down (Ramalho-Ortigao and Traub-Cseko, 2003; Ramalho-Ortigao *et al.*, 2005). Furthermore, RNAi knockdown of the vector chitinase PpChit1 results in a thickened PM and reduced parasite load – presumably due to the inability of ‘trapped’ *L. major* to escape and attach to the midgut epithelium (Coutinho-Abreu *et al.*, 2013).

Sand flies of the genus *Sergentomyia*, which feed on reptiles, are refractory to several *Leishmania* species, unless the PM is prematurely degraded by addition of exogenous chitinases which then allowed mature infections to develop (Sadlova *et al.*, 2018). The refractory nature of *Sergentomyia* is thought to be due to the greater proportion of time that the PM is intact for during digestion with defecation quickly following PM breakdown. In other sand fly species, there is a longer window between PM degradation and defecation. This means parasites have an extended window in which they can access midgut epithelium, increasing chances of successful attachment and avoiding being excreted (Pruzinova *et al.*, 2015). Together, these observations suggest that insect-mediated PM breakdown is key to *Leishmania*'s escape into the ectoperitrophic space rather than parasite-derived chitinases alone.

Inside the ectoperitrophic space

Upon entry into the ectoperitrophic space, trypanosomatids encounter several more layers of the insect vector defences. Whilst the peritrophic matrix is permeable to digestive enzymes, the concentration of proteolytic enzymes is significantly higher outside of it. Depending on the species, the protease levels peak at 18–48h after ingestion of the blood meal in sand flies (Telleria *et al.*, 2010; Dostálová and Volf, 2012). Trypsin-like, serine proteases are the most common, though chymotrypsins, metallocarboxypeptidases, and alanyl aminopeptidase have also been reported (Dillon and Lane, 1993; Dillon *et al.*, 2006). These enzymes are a significant hurdle to parasite establishment by causing proteolytic damage to cells (Wigglesworth, 1929; Schlein and Jacobson, 1998). The digestive enzymes of the tsetse have been shown to recognise and cleave the EP and GPEET glycoproteins which dominate the *T. brucei* surface in midgut-dwelling forms (Acosta-Serrano *et al.* 2001, Liniger *et al.*, 2003). In the sand fly *P. papatasi*, addition of

trypsin inhibitors with the blood meal increased *Leishmania donovani* survival in the early stages of infection. Furthermore, RNAi knockdown of LITryp1, a feeding-induced trypsin, in *Lutzomyia longipalpis* lead to greater survival of *Leishmania mexicana* upon infection (Sant'Anna *et al.*, 2014). As such the digestive enzymes are an important, yet often overlooked, part of the dipteran defence to invading trypanosomatids.

Both *Trypanosoma* and *Leishmania* are thought to moderate the expression and activity of their vectors' proteolytic enzymes to promote survival. Aksoy *et al.* (2016) showed significant downregulation of major gut enzymes, trypsins and chymotrypsins, as well as peritrophins components of the peritrophic matrix in flies exposed to trypanosomes (Aksoy *et al.*, 2016). The same study also showed that variant surface glycoproteins (VSGs) from the ingested trypanosomes cell are taken up by the cells in the tsetse proventriculus. This uptake resulted in reduction of expression of tsetse mir-275 microRNA and interfered with both Wnt-signalling and the Iroquois/IRX family of transcription factors. This interference resulted in disrupted peritrophic matrix formation and promoted establishment of the trypanosomatids in the midgut. This work was the first to demonstrate a functional role for trypanosome VSG coat, which is more commonly associated with avoidance of triggering a humoral immune response in the mammalian blood stream, in the insect midgut.

Leishmania parasites appear have evolved similar strategies to promote survival in sand flies. Digestive enzymes are differentially regulated during *L. major* and *L. infantum* infection when compared to sand flies fed sterile blood (Ramalho-Ortigão *et al.*, 2007; Jochim *et al.*, 2008; Telleria *et al.*, 2010; Dostálová *et al.*, 2011). Furthermore, the *L. major* genome encodes several serine protease inhibitors shown to work against insect proteases *in vitro* (Morrison *et al.*, 2012; Alam *et al.*, 2016; Verma *et al.*, 2017), yet it

does not appear to contain the enzymes which these proteins target. As such, the function of these enzymes may be to promote survival in the sand fly through inhibition of digestive enzyme activities. Further studies of these protease inhibitors are of great interest for the development of novel transmission blocking strategies.

In addition to the insect's digestive enzymes and the formidable barrier presented by the peritrophic matrix, insects have multiple immune signalling pathways to combat infection. Many of these pathways are able to detect and respond to the presence of trypanosomatids, and other pathogens, in the midgut. An overview of the major innate immune pathways of dipterans, with a focus on gut immune responses, is given in the following section, but they are extensively reviewed elsewhere (Kounatidis and Ligoxygakis, 2012; Rosales, 2017).

Innate immunity in Dipterans: an overview

Insects lack the classical adaptive immune pathways of mammals, yet they are the most numerous and diverse group of organisms on earth, occupying practically every habitat. This demonstrates the strength of the innate immune responses of these organisms. Activation of these robust innate pathways is somewhat pathogen-specific and leads to the expression of a suite immune effector molecules, such as antimicrobial peptides, tailored to suit the threat faced. There are two major innate immune pathways in the dipteran humoral response, Toll and the immune deficiency (Imd) pathway. These main do not act alone, but interact with several others to form complex and robust response to infection.

Pathogen recognition

In order to produce an effective immune response, the host must recognise the not only that it is under attack, but also the nature of the attacker. There are several families of pattern recognition proteins, each recognising a specific set of pathogen-associated molecular patterns (PAMPs) including: peptidoglycan recognition proteins (PGRP), gram negative binding proteins (GNBPs), scavenger proteins and lectins. The peptidoglycan recognition receptors (PGRPs) are the best described. These proteins interact with peptidoglycan, a major bacterial cell wall component, via a domain at the C-terminus. Variation within this domain allows recognition of specific forms of peptidoglycan. The PGRPs are grouped into long ($LC_{a,x,y}$ and LE) and short forms (SA and SC). The short forms are found in the cuticle, hemolymph, fat body and cuticle. Whilst long forms are expressed in epithelia and hemocytes.

PGRPs can both stimulate and dampen an immune response (Myllymäki, Valanne and Rämetsä, 2014). PGRP-LB binds to PGRP-LCx, inhibiting Imd pathway activation. PGRP-LB, -SC1a/b, and -SC2 cleave DAP-type peptidoglycan into fragments which are unable to trigger an Imd response. This dampening is crucial to prevent overstimulation of the immune system due to the presence of commensal microbes in the fly digestive tract - which are also a source of peptidoglycan and other PAMPs. PGRPs are also able to act as effector molecules with PGRP-SB1 demonstrated to be bactericidal (Mellroth and Steiner, 2006).

The immune deficiency (Imd) pathway

The immune deficiency (Imd) pathway (Figure 1.4A) is a conserved innate response, similar to the mammalian tumour necrosis factor 1 (TNF1) pathway, which is activated in

response to, gram-negative and trypanosomatid pathogens. The pathway is especially crucial during gut infections. The pathway can be activated by the binding of DAP-peptidoglycan to PGRP-LC and -LE, which then interact with key adaptor protein IMD (Choe, Lee and Anderson, 2005). IMD recruits FADD (Fas-associated protein with death domain) and the caspase DREDD (death-related ced-3/Nedd2-like protein) (Georgel *et al.*, 2001; Leulier *et al.*, 2002). DREDD then cleaves IMD, which is then K63-ubiquitinated by IAP2 (inhibitor of apoptosis 2) (Paquette *et al.*, 2010; Meinander *et al.*, 2012). TAK1-associated binding protein 2 (TAB2) interacts with the IAP2's ubiquitin chains and recruits/activates TAK1 (transforming growth factor β -activated kinase 1) (Kleino *et al.*, 2005). TAK1 phosphorylates the I κ B kinase complex, which in turn phosphorylates the Relish transcription factor (Silverman *et al.*, 2000). Phosphorylated Relish is then cleaved by caspase-type protease (Stöven *et al.*, 2000), likely DREDD (Paquette *et al.*, 2010). Translocation of the N-terminal end of Relish into the cytoplasm activates AMP transcription, including dipterecin and cecropin.

The IMD pathway is tightly regulated with negative regulators evident for the majority pathway stages. These include the aforementioned PGRP-LB, as well as: Pirk (Kleino *et al.*, 2008), Caspar (Kim *et al.*, 2006), POSH (Plenty of SH3s) (Tsuda *et al.*, 2005), Trabid (Fernando, Kounatidis and Ligoxygakis, 2014), CYLD (Tsichritzis *et al.*, 2007), Caudal (Ryu *et al.*, 2008) and transglutaminase (Maki, Shibata and Kawabata, 2017). This control is vital to the homeostasis between the insect and commensal organisms in its gut microbiome. Additionally, the Imd pathway interacts with other signalling pathways during the production of an immune response, including the JNK (Boutros, Agaisse and Perrimon, 2002) and nitric oxide pathways (Foley and O'Farrell, 2003). These pathways

are associated with wound healing and AMP production in response to microbial challenge respectively (Ravanti and Kähäri, 2000).

IMD and DUOX help moderate gut microbiota and combat gut infections

Reactive oxygen species (ROS), including superoxide anions (O_2^-) and hydroxyl radicals ($OH^\cdot$), are able to directly kill pathogens through oxidative stress and DNA damage.

These molecules are formed at basal levels during respiration in insects, though higher levels are produced during gut infections (Moule, Monack and Schneider, 2010; Wu *et al.*, 2012).

The NADPH oxidase, dual-oxidase (DUOX) is responsible for production of ROS in the gut epithelium in response to microbes (Figure 1.4B). Upon infection, increased expression of DUOX is facilitated by transcription factor ATF2 (activating transcription factor 2) of the mitogen-activated protein (MAP) kinase p38 pathway (Ha, Lee, Seo, *et al.*, 2009). In order for ATF2 to effect transcriptomic change it must be phosphorylated by the p38 kinase, which in turn is activated by a MAP kinase cascade. This kinase cascade can be activated both by the IMD pathway, through IMD interaction with MEKK1, and by $G\alpha_q$ -phospholipase C- β (PLC- β).

PLC- β is also involved in the regulation of DUOX enzymatic activity via the $G\alpha_q$ -phospholipase C- β (PLC- β)- Ca^{2+} pathway (Ha, Lee, Park, *et al.*, 2009). In this pathway, microbe-derived molecules activate G-protein coupled receptors (GPCRs) at the epithelial cell surface, allowing the $G\alpha_q$ subunit to activate PLC β . This enzyme hydrolyzes phosphatidylinositol 4,5-bisphosphate (PIP₂), to give inositol 1,4,5-trisphosphate (IP₃) and diacylglycerol (DAG). The released IP₃ binds to its receptor (IP₃R) in the membrane

of the endoplasmic reticulum. This binding results in the release of Ca^{2+} , which increases DUOX activity via its EF-hand domain.

The Toll pathway

The second major insect innate response is coordinated by the Toll pathway (Figure 1.4C). Toll signalling can be activated by PGRP-SA (Chang *et al.*, 2004) and gram-negative binding proteins (GNBPs, Gobert *et al.*, 2003; Mishima *et al.*, 2009) in response to fungal and gram-positive bacterial pathogens. Pathogen recognition triggers protein cleavage cascades, which culminate in the cleavage and dimerization of the Toll ligand Spätzle (Gangloff *et al.* 2008, Valanne, Wang and Rämet, 2011). Spätzle-Toll binding results in the dimerization of the TIR (Toll/Interleukin-1 Receptor) domains in the cytoplasmic tail of the receptor. The TIR domains are then bound by adaptor protein MyD88, via its own TIR domain (Xiao *et al.*, 1999). The protein kinase Pelle is recruited to the growing protein complex by another adaptor protein, Tube. Tube interacts with both Pelle and MyD88 via death domains (Sun *et al.*, 2002). Autophosphorylation of Pelle results in phosphorylation of the I κ B inhibitor Cactus, causing its degradation (Wu and Anderson, 1998). In the absence of pathway activation, Cactus binds the NF κ B transcription factors Dorsal and Dif, inhibiting their translocation into the nucleus (Kidd, 1992). Degradation of Cactus allows translocation of Dorsal and Dif into the nucleus where they facilitate the transcription of several specific gene sets. Dorsal has been shown to activate expression of antimicrobial peptide (AMP) dipterecin (Gross *et al.*, 1996) and is produced in response to bacterial pathogens. Whereas Dif nuclear translocation is associated with drosomycin, an antifungal AMP, expression (Rutschmann *et al.*, 2000).

Antimicrobial peptides

The major outcome of Imd/Toll signaling is the production of suite of AMPs. These effector molecules allow the killing and elimination of invading pathogens. In dipterans, many AMPs are constitutively expressed at low levels, but higher expression can be induced upon infection (Hao *et al.*, 2001). Their structures and target organisms vary. However, their primary mode of action is pathogen cell membrane disruption, either directly or by inhibiting its synthesis, causing osmotic stress and cell death. AMP synthesis occurs primarily in the insect fat body, however they are also produced by hemocytes and the midgut epithelium (Wu, Patočka and Kuča, 2018). Additionally, AMPs have been shown to lead hemocytes to sites of infection (Bartholomay *et al.*, 2004) and provide protection from oxidative stress (Zhao, Zhou and Haddad, 2011). These secondary roles are not yet fully understood and are yet to be investigated in the context of trypanosomatid infections.

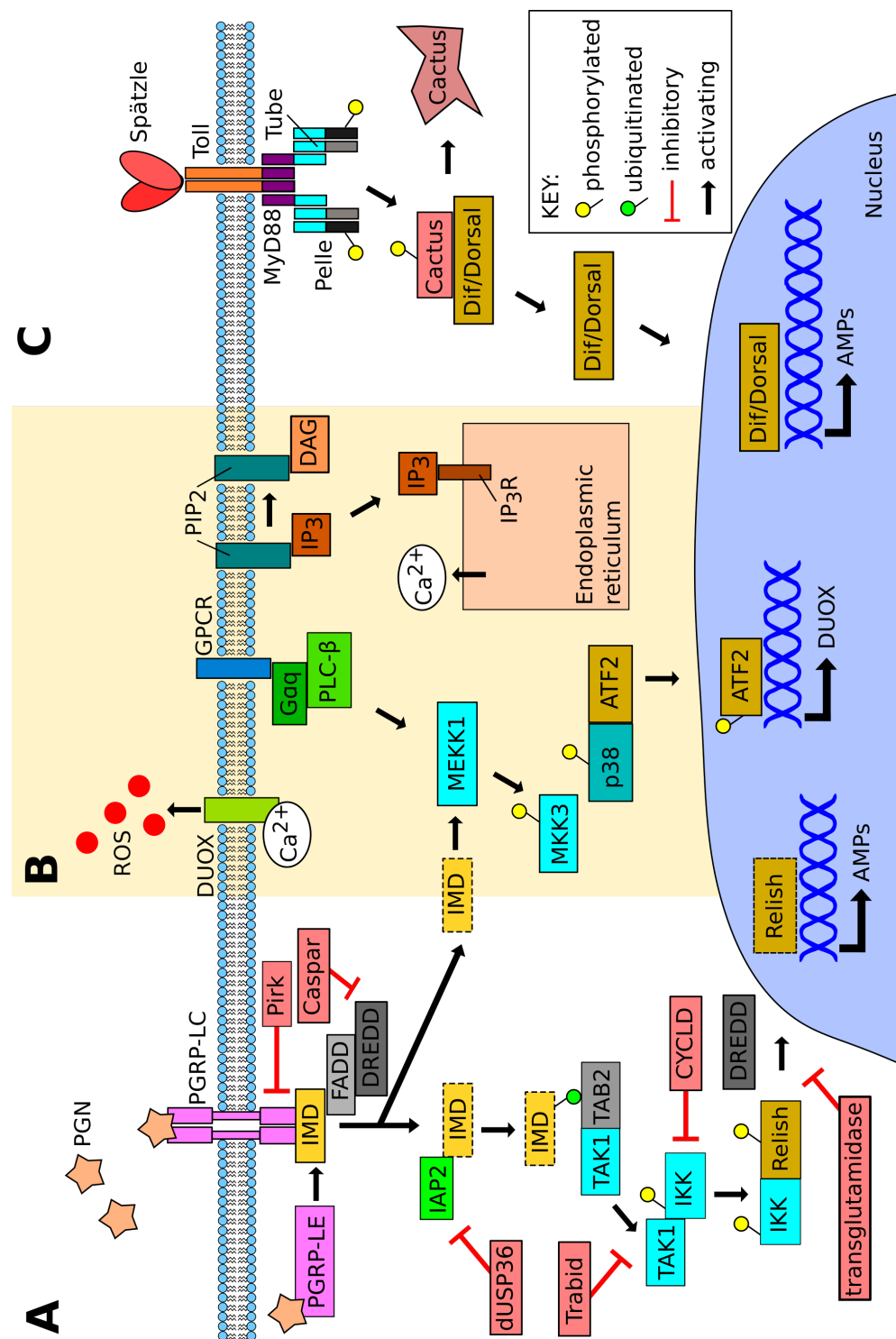


Figure 1.4 – The major innate signalling pathways of insects. Please see the main text for details.

Dipteran immune responses to trypanosomatid infection

The immune responses discussed above, amongst others, allow dipterans to respond to a broad range of pathogens. The responses induced vary based on the invading organisms. Fly responses to trypanosomatids appear to be distinct to those in response to bacterial infections indicating trypanosomatid parasite recognition at some level (Beschin *et al.*, 2014). Here I will discuss what is known about the innate immune responses to *Trypanosoma* and *Leishmania* parasites in their dipteran vectors.

Trypanosomatid recognition

Several proteins which recognise trypanosomatid molecular patterns have been identified including galectins, peptidoglycan recognition receptors and scavenger receptors. The specifics of the interaction between these surveillance molecules are roles and not yet fully understood. In tsetse, peptidoglycan binding protein, PGRP-LB, has been demonstrated to recognise and kill trypanosomes (Wang and Aksoy, 2012). This trypanocidal activity was shown to be independent of the amidase activity of PGRP-LB, which allows the protein to cleave bacterial peptidoglycan. As such the mode of action in this case is unknown. PGRP-LB levels were much lower in flies susceptible to *T. brucei*, compared to flies able to clear infection (Wang *et al.*, 2009). This highlights the molecules importance in the response to trypanosome infection.

In sand flies, Dillon *et al.* (2006) identified three putative scavenger class B receptors (SCRs) with CD36 domains which were expressed by *Lutzomyia longipalpis* upon ingestion of *Leishmania* parasites. Scavenger receptors are a large family of proteins with diverse structures and functions. Class B scavenger receptors (SR-BI) are embedded in membranes at both N and C termini with a serine/threonine rich extracellular region of

~400 amino acids. These receptors have roles in cell signalling, fatty-acid uptake and responses to bacterial, fungal and protozoan pathogens in a diverse range of organisms (PrabhuDas *et al.*, 2017). The *Drosophila* homologue for one of these SCRs, CG2736, has also been shown to be upregulated in response to protozoan infections, but not bacterial infections (Roxstrom-Lindquist, Terenius and Faye, 2004). Another study showed CD36-like scavenger receptors on the surface of *Drosophila* hemocytes were critical for control of *Leishmania* injected into the fly hemolymph by phagocytosis (Okuda *et al.*, 2016). As such these receptors are of great interest for further studies.

Galectins, galactose-binding lectins, have also been implicated in the detection of *Leishmania* infection in sand flies. These proteins are thought to interact with lipophosphoglycans (LPG) which are major surface molecules of *Leishmania* parasites (Dimopoulos *et al.*, 1998; Kamhawi *et al.*, 2004; Dillon *et al.*, 2006). Upon binding a pathogen or pathogen-derived molecule, insect galectins are able to facilitate both uptake of pathogens by phagocytosis locally, and the activation of a broader immune response via cell signalling pathways (Chen *et al.*, 1999; Yu, Gan and R. Kanost, 1999; Dimopoulos *et al.*, 2001). However, *Leishmania* parasites appear to take advantage of galectin binding, as attachment to the midgut epithelium allows the parasites to persist in the insect and avoid being excreted. The interaction may be species-specific as the side-chains of LPG, which have exposed galactose residues to which the galectins are thought to bind, vary between parasite species (Kamhawi *et al.*, 2004).

Trypanosomatid infections are associated with an IMD response in insects

The trypanosomatid infections of both tsetse and sand flies are associated with an IMD pathway response. IMD negative regulator, Caspar, is downregulated during *Leishmania*

infection of *L. longipalpis* (Telleria *et al.*, 2012). Interestingly this effect of *Leishmania* infection was not observed in a recently developed *L. longipalpis* embryonic cell line when infected with *Leishmania infantum chagasi*. However the group did observe upregulation IMD pathway transcription factor Relish, and the antimicrobial peptides defensin, attacin and cecropin (Tinoco-Nunes *et al.*, 2016). The role of the Toll pathway in response to *Leishmania* challenge is unclear. Cells challenged with the trypanosomatid showed upregulation of both transcription factor Dorsal and it's negative regulator Cactus. It is unclear if this is a prioritisation of an IMD response by the fly or an attempt to modulate host responses by *L. i. chagasi*. Similarly, infection of tsetse with *T. brucei* causes upregulation the IMD-regulated antimicrobial peptides attacin and cecropin (Hu and Aksoy, 2006). The same study found RNAi knockdown of IMD pathway transcription factor Relish increased both prevalence and intensity of trypanosomatid infections. Knock down of attacin alone in the flies had a similar, but less pronounced, influence on infection rate.

Several IMD-regulated AMPs have been demonstrated to have trypanocidal activities including the cecropins, defensins, stomoxyn, attacin and dipteracin (Boulanger, Bulet and Lowenberger, 2006). The coordinated expression of specific combinations of these AMPs are crucial to infection outcome. The bumblebee parasite *Crithidia bombi*, a member of the Leishmanii clade, show strain-dependent sensitivities to different insect AMP combinations (Marxer, Vollenweider and Schmid-Hempel, 2016). In tsetse, *T. brucei* establishment in the midgut causes upregulation of antimicrobial peptides, attacin and cecropin, in the fat body (Hao *et al.*, 2001; Boulanger *et al.*, 2002). Knock down of the two antimicrobial peptides also resulted in increased infection intensity, underlining the role these proteins play in the tsetse response to trypanosome infection (Hu and Aksoy, 2006).

In the sand fly *P. dubosqi*, challenge with *L. major* resulted in an increase of defensin encoding mRNAs (Boulanger *et al.*, 2004). In contrast to this, neither ingestion of *L. mexicana*, nor injection of the trypanosomatids to the *L. longipalpis* haemocoel, significantly altered expression of AMP defensin (Telleria *et al.*, 2013). As such it is possible that different AMPs are required for the control of *Leishmania* infection in ‘New World’ vector *L. longipalpis* to those in ‘Old World’ vectors like *P. dubosqi* (Telleria *et al.*, 2018). Alternatively, *L. mexicana* may have mechanisms to moderate the expression of defensin in their sand fly hosts.

Reactive oxygen species (ROS) in trypanosomatid infection

Whilst ROS have been demonstrated to kill *Leishmania*, the parasites appear to be able to avoid triggering increases in ROS productions in sand flies (Sant’Anna *et al.*, 2014). It is not known whether this reflects active suppression of the host immune response, or the avoidance of detection by the parasite. The antioxidant protein peroxidoxin is upregulated in *Leishmania* promastigotes (Diaz-Albiter *et al.*, 2012) and many protect the parasites from the ROS produced at basal levels in the sand fly. Unlike *Leishmania*, *T. brucei* infection does result in increased hydrogen peroxide levels in tsetse (Beschlin *et al.*, 2014) and treatment of the blood meal with antioxidants prior to feeding increased the rate of infection (MacLeod *et al.*, 2007). As such these molecules play important roles in trypanosome clearance.

The role of the vector microbiome in trypanosomatid infection

Tsetse have three associated bacterial endosymbionts; *Wigglesworthia glossinidia*, *Sodalis glossinidius* and *Wolbachia* sp. These symbionts provide the fly with vital nutrients which are not provided by the hematophagous diet, e.g. dominant tsetse

symbiont, *W. glossinidia*, provides B vitamins in exchange for its protected niche inside midgut cells (Akman *et al.*, 2002; Rio, Attardo and Weiss, 2016; Bing *et al.*, 2017). The symbionts are crucial to the integrity of the peritrophic matrix, with aposymbiotic flies having a severely compromised PM (Weiss *et al.*, 2013). Each symbiont appears to have a different impact on the outcome of trypanosome infection. *W. glossinidia*, which is unable to survive outside the tsetse midgut environment, is required for the maturation of the tsetse immune system, specifically for hemocyte-mediated immunity (Heller, 2011; Weiss, Wang and Aksoy, 2011). Tsetse lacking *Wigglesworthia* are more susceptible to trypanosome infection than their wild-type counterparts (Pais *et al.*, 2008), indicating a protective influence of the symbiont. On the other hand, presence of the (non-essential) symbiont *S. glossinidius*, significantly increased the likelihood of *T. brucei* infection (Farikou *et al.*, 2010). *S. glossinidius* is found mostly in the midgut though has been isolated from a diverse range of tsetse tissues (Cheng and Aksoy, 1999). *S. glossinidius* releases N-acetyl-D-glucosamine (GlcNAc) through chitinolytic activity and may inhibit insect lectins, reducing the fly's ability to recognize the trypanosomatid, promoting *T. brucei* survival (Welburn *et al.*, 1993; Rio, Attardo and Weiss, 2016). Finally, *Wolbachia* are a group of maternally inherited, intracellular α -proteobacteria present in 65% of insects and numerous other arthropods (Bandi *et al.*, 1998). Incidence of *Wolbachia* in tsetse in Cameroon was not statistically linked to trypanosomatid infection rate (Kame- Ngasse *et al.*, 2018). However, *Wolbachia* have been shown to confer increased resistance to viruses and malaria in mosquitos (Moreira *et al.*, 2009; van den Hurk *et al.*, 2012), as such further investigations into the symbionts role in trypanosome infection are warranted.

In addition to the endosymbionts, several bacterial species are present in the tsetse midgut. These bacteria vary according to the fly species and its geographic location (Lindh and Lehane, 2011; Geiger *et al.*, 2013; Aksoy *et al.*, 2014; Hamidou Soumana *et al.*, 2014). Members of the *Enterobacter*, *Enterococcus* and *Acinetobacter* genera are common with *Chryseobacterium*, *Sphingobacterium*, *Providencia*, *Lactococcus*, *Pseudomonas* and *Staphylococcus* have also been observed. The persistence of these species in tsetse and any roles they may play in *T. brucei* infections have yet to be fully investigated. However, work in other insects suggests a protective role for some bacterial species. *Serratia marcescens* and *Pseudomonas fluorescens* strains have both been shown to be able to lyse *T. cruzi* in the triatome bug *Rhodnius prolixus* (Mercado and Colón-Whitt, 1982; Azambuja, Feder and Garcia, 2004)

Extensive work has been undertaken to describe the microbiome of multiple sand fly species, and life cycle stages, from both laboratory and wild populations (reviewed in Telleria *et al.*, 2018). This explosion of recent work is too vast to review in detail here. Generally, there is a diverse range of bacterial genera regularly found in sand flies, but few appear common at the (sand fly) species level, and fewer still are common across sand fly genera. This is likely to reflect the diverse diet of sandflies which feed on many organic substrates as larvae and then both plant material and blood as adults (Fraihy *et al.*, 2017). In the context of *Leishmania* infection there is much left to investigate in the complex interactions of the fly, its microbiome and the trypanosomatid. It has been shown that while the quantities of bacteria increase after ingestion of a blood meal in female sand flies, species diversity decreases (Kelly *et al.*, 2017; Pires *et al.*, 2017). This recovered after feeding in uninfected *L. longipalpis*, but not in *L. infantum* infected flies where *Acetobacteraceae* spp. became dominant at the expense of *Pseudomonadaceae*

spp. Antibiotic treatment of the flies, and the subsequent perturbation of the microbiota prevented maturation of the *Leishmania* infection. A similar study of *L. major* infections in *P. dubosqi*, also suggested the microbiome may act to buffer the osmolarity of the sand fly midgut, thus promoting trypanosomatid survival (Louradour *et al.*, 2017). As such, the microbiota of *L. longipalpis* appear to be crucial to infection establishment and therefore vector competence.

The findings above appear to contrast earlier work, which had suggested a protective role for the sand fly microbiome against *Leishmania* infections. Sant'Anna *et al.* (2014) showed that feeding *L. longipalpis* with *Asaia sp.* and *Ochrobactrum intermedium*, which the group had isolated from laboratory and wild sand fly populations, prevented the establishment of *L. mexicana* in the flies. In this study the large influx of bacteria may act to trigger insect immune responses, thus 'priming' the fly prior to *Leishmania* ingestion and promoting parasite clearance.

The insect microbiome may continue to be beneficial to *Leishmania* cells after their deposition at the vertebrate bite site. Gut microbes are known to be deposited into the vertebrate blood stream along with *Leishmania* parasites, where they appear to promote neutrophil infiltration of the tissue (Dey *et al.*, 2018). These cells phagocytose the *Leishmania* metacyclics allowing continuation of the *Leishmania* lifecycle. In addition to bacterial species, fungal species from *Aspergillus*, *Penicillium*, *Geotrichum*, *Fusarium*, *Acremonium* and *Candida* have been reported in *Phlebotomus* in Iran (Akhoundi *et al.*, 2012). Whilst studies of the sand fly fungal component of the microbiota are limited, there is some evidence that fungal presence makes sand flies more resistant to *L. major* (Adler and Theodor, 1929; Schlein, Polacheck and Yuval, 1985).

In summary, interactions between trypanosomatids and their dipteran vectors are complex with several factors influencing the outcome of infection. Insect defenses are robust, multilayered and largely successful in the prevention of trypanosomatid establishment – as evident from low trypanosomatid prevalence reported in wild insect populations. Yet trypanosomatid diseases remain a significant burden to global health. This is likely due to additional factors, such as the close proximity of large numbers of insect vectors to human and domestic animal dwellings. The successes of vector control programs and the development of novel transmission blocking techniques, such as paratransgenesis, give cause for hope. These techniques have stemmed from intense research into the interactions between the trypanosomatids and their vectors, however much of the molecular detail of these interactions is not fully understood. With growing reports of resistance of both trypanosomatids and insects to drug treatments, further studies will be important for the elimination of trypanosomatid diseases.

Rationale

Whilst several drugs are available for the treatment of trypanosomatid diseases, these drugs have a number of shortcomings including high toxicity, variable efficacy and high costs. These attributes often exclude them from those who need them most. Furthermore, resistance to some treatments is emerging in India (Ponte-Sucre *et al.*, 2017) and Sudan (Basselin *et al.*, 2002), where Leishmaniasis is endemic. These issues are mirrored in strategies used to control the disease vectors. Pesticides used to reduce vector insect populations are expensive, labour intensive to apply and are often harmful to other organisms. Their use has proved effective in the control of HAT, however there is debate as to whether their use is sustainable and resistance to a number of pesticides is being

reported in *Phlebotomus papatasi* - an important vector for *Leishmania major* (Hassan *et al.*, 2012; Dhiman and Yadav, 2016; Fawaz *et al.*, 2016). Therefore, there is still an urgent need for new anti-trypanosomatid drug treatments and transmission blocking strategies.

Hurdles to new transmission blocking targets

Studies using hematophagous insects come with not insignificant challenges, including the requirement for dedicated facilities (category 3 laboratories) and blood, which means working with these organisms is not broadly accessible. Whilst the trypanosomatids themselves are genetically tractable with several well-established tools for their study, including recently developed CRISPR genome editing techniques (Dean *et al.*, 2015; Dyer, Dean and Sunter, 2016; Beneke *et al.*, 2017), there are relatively few well-developed tools available for working with their vector insects. RNA interference methodology is available (Sant'Anna *et al.*, 2008; Walshe *et al.*, 2009), however the production of transgenic *Glossina* lines is challenging due to the intrauterine development of tsetse larvae. This mechanism of reproduction also means the female only produces one larva every 10 days or so, which is comparatively slow, with other flies able to lay hundreds of eggs per day.

A new dipteran-trypanosomatid model

The model organism *Drosophila melanogaster* presents a valuable opportunity to unpick the molecular interactions between disease-causing trypanosomatids and their insect vectors. As a member of the Brachycera, *Drosophila* is closer to tsetse than the sand flies, though all three insects belong to the Diptera. Fruit flies are simple and inexpensive to rear, have a short generation time and present less risk to the researcher as they lack the

skin-piercing mouthparts of trypanosomatid disease vectors. Furthermore, *Drosophila melanogaster* has been extensively studied and as such has many well-established research tools and techniques including extensive mutant libraries and a well-described microbiome. Crucially, *Drosophila melanogaster* is also a natural host for trypanosomatids, specifically *Herpetomonas muscarum*.

Aims and strategy

Previous work in the Ligoxygakis lab had identified and isolated the trypanosomatid *Herpetomonas muscarum* from wild *Drosophila* populations in Oxfordshire. The group had begun to investigate the response of the fly to ingestion of the trypanosomatid in a laboratory setting which was recently published in (Wang, Sloan and Ligoxygakis, 2019). However, little was known about the trypanosomatid side of this interaction. As such I wanted to investigate the trypanosomatid response to being ingested by the fly and compare these responses to those of disease-causing trypanosomatids in their respective vectors. Given that several of the responses to trypanosomatids appear to be conserved across the Diptera, I hypothesized that the trypanosomatid responses to the flies would be similarly conserved.

Specific aims of this project were:

- the sequencing and annotation of the *H. muscarum* genome (see Chapter Two),
- the identification of important stages and sites of interaction between *H. muscarum* and *D. melanogaster* (see Chapter Three),
- categorisation of the *H. muscarum* transcriptomic response to ingestion by *D. melanogaster* (see Chapter Four).

I addressed these aims using *per os* infections of *D. melanogaster*, light and electron microscopy, nucleic acid sequencing and bioinformatics analyses. To assist the addressing of these aims, I developed new tools for the interrogation of the *Drosophila-Herpetomonas* model. In addition to an annotated *H. muscarum* genome sequence, I produced a set of plasmids enabling the generation of *H. muscarum* mutant cell lines (see Chapter Five). Comparison of the outputs of the aims above, to interactions of disease-causing trypanosomatids and their respective vectors, allowed me to identify several evolutionarily conserved aspects of dipteran-trypanosomatid interactions. Several genes with roles in these conserved responses are of interest further study in the context of broad-spectrum transmission blocking methods. Additionally, the existence of these conserved responses also underlines the value of the *Drosophila-Herpetomonas* model.

References

- de A Gonzaga, W. *et al.* (2003) ‘Antibacterial Alkaloids from *Zanthoxylum rhoifolium*’, *Planta Med*, 69(04), pp. 371–374. doi: 10.1055/s-2003-38882.
- Acosta-Serrano, A., Vassella, E., Liniger, M., Kunz Renggli, C., Brun, R., Roditi, I. and Englund, P. T. (2001) The surface coat of procyclic *Trypanosoma brucei*: Programmed expression and proteolytic cleavage of procyclin in the tsetse fly. *Proceedings of the National Academy of Sciences*. National Academy of Sciences 98(4): 1513–1518.
- Adler, S. and Theodor, O. (1929) ‘Attempts to Transmit *Leishmania Tropica* by Bite: The Transmission of *L. Tropica* by *Phlebotomus: Sergenti*’, *Annals of Tropical Medicine & Parasitology*. Taylor & Francis, 23(1), pp. 1–18. doi: 10.1080/00034983.1929.11684589.
- Akhoundi, M. *et al.* (2012) ‘Diversity of the Bacterial and Fungal Microflora from the Midgut and Cuticle of Phlebotomine Sand Flies Collected in North-Western Iran’, *PLOS ONE*. Public Library of Science, 7(11), p. e50259.
- Akhoundi, M. *et al.* (2013) ‘Molecular characterization of *Leishmania* infection from naturally infected sand flies caught in a focus of Cutaneous Leishmaniasis (Eastern Iran)’, *Journal of arthropod-borne diseases*. Tehran University of Medical Sciences, 7(2), pp. 122–131.

- Akman, L. *et al.* (2002) 'Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*', *Nature Genetics*. Nature Publishing Group, 32, p. 402.
- Aksoy, E. *et al.* (2014) 'Analysis of Multiple Tsetse Fly Populations in Uganda Reveals Limited Diversity and Species-Specific Gut Microbiota', *Applied and Environmental Microbiology*. Edited by H. Goodrich-Blair, 80(14), p. 4301 LP-4312. doi: 10.1128/AEM.00079-14.
- Aksoy, E. *et al.* (2016) 'Mammalian African trypanosome VSG coat enhances tsetse's vector competence', *Proceedings of the National Academy of Sciences of the United States of America*. 2016/05/16. *National Academy of Sciences*, 113(25), pp. 6961–6966. doi: 10.1073/pnas.1600304113.
- Alam, M. N. *et al.* (2016) 'Identification and characterization of a *Leishmania donovani* serine protease inhibitor: Possible role in regulation of host serine proteases.', *Life sciences*. Netherlands, 144, pp. 218–225. doi: 10.1016/j.lfs.2015.12.004.
- Alam, U. *et al.* (2011) '*Wolbachia* Symbiont Infections Induce Strong Cytoplasmic Incompatibility in the Tsetse Fly *Glossina morsitans*', *PLOS Pathogens*. Public Library of Science, 7(12), p. e1002415.
- Alvar, J. *et al.* (2012) 'Leishmaniasis worldwide and global estimates of its incidence.', *PloS one*. United States, 7(5), p. e35671. doi: 10.1371/journal.pone.0035671.
- Azambuja, P., Feder, D. and Garcia, E. S. (2004) 'Isolation of *Serratia marcescens* in the midgut of *Rhodnius prolixus*: impact on the establishment of the parasite *Trypanosoma cruzi* in the vector', *Experimental Parasitology*, 107(1), pp. 89–96. doi: <https://doi.org/10.1016/j.exppara.2004.04.007>.
- Azevedo Cardoso, E. *et al.* (2014) 'Leishmaniasis: history, evolution of treatment and the need for new drugs', *Current Biotechnology*. Bentham Science Publishers, 3(4), pp. 279–288.
- B. Wigglesworth, V. (1929) *Digestion in the Tsetse-Fly: A Study of Structure and Function*, *Parasitology*. doi: 10.1017/S0031182000022988.
- Baker, C. H. and Welburn, S. C. (2018) 'The Long Wait for a New Drug for Human African Trypanosomiasis.', *Trends in parasitology*. England, 34(10), pp. 818–827. doi: 10.1016/j.pt.2018.08.006.
- Bandi, C. *et al.* (1998) 'Phylogeny of *Wolbachia* in filarial nematodes', *Proceedings. Biological sciences*, 265(1413), pp. 2407–2413. doi: 10.1098/rspb.1998.0591.
- Bartholomay, L. C. *et al.* (2004) 'Reassessing the role of defensin in the innate immune response of the mosquito, *Aedes aegypti*', *Insect Molecular Biology*. John Wiley & Sons, Ltd (10.1111), 13(2), pp. 125–132. doi: 10.1111/j.0962-1075.2004.00467.x.
- Basselin, M. *et al.* (2002) 'Resistance to pentamidine in *Leishmania mexicana* involves exclusion of the drug from the mitochondrion', *Antimicrobial agents and chemotherapy*. American Society for Microbiology, 46(12), pp. 3731–3738. doi: 10.1128/AAC.46.12.3731-3738.2002.

- Bates, P. A. (2007) 'Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies', *International Journal for Parasitology*, 37(10), pp. 1097–1106. doi: <https://doi.org/10.1016/j.ijpara.2007.04.003>.
- Beneke, T. *et al.* (2017) 'A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids', *Royal Society open science*. The Royal Society Publishing, 4(5), p. 170095. doi: 10.1098/rsos.170095.
- Benvenuti, L. A. *et al.* (2003) 'Pathophysiology of the heart in Chagas' disease: current status and new developments', *Cardiovascular Research*, 60(1), pp. 96–107. doi: 10.1016/S0008-6363(03)00361-4.
- Berman, J. D., Dwyer, D. M. and Wyler, D. J. (1979) 'Multiplication of *Leishmania* in human macrophages *in vitro*.', *Infection and immunity*. United States, 26(1), pp. 375–379.
- Bern, C. *et al.* (2007) 'Evaluation and Treatment of Chagas Disease in the United States A Systematic Review', *JAMA*, 298(18), pp. 2171–2181. doi: 10.1001/jama.298.18.2171.
- Beschin, A. *et al.* (2014) 'African trypanosome control in the insect vector and mammalian host', *Trends in Parasitology*, 30(11), pp. 538–547. doi: <https://doi.org/10.1016/j.pt.2014.08.006>.
- Bing, X. *et al.* (2017) 'Unravelling the relationship between the tsetse fly and its obligate symbiont *Wigglesworthia*: transcriptomic and metabolomic landscapes reveal highly integrated physiological networks', *Proceedings. Biological sciences*. 2017/06/28. The Royal Society, 284(1857), p. 20170360. doi: 10.1098/rspb.2017.0360.
- Boulanger, N. *et al.* (2002) 'Immunopeptides in the defense reactions of *Glossina morsitans* to bacterial and *Trypanosoma brucei brucei* infections.', *Insect biochemistry and molecular biology*. England, 32(4), pp. 369–375.
- Boulanger, N. *et al.* (2004) 'Characterization of a defensin from the sand fly *Phlebotomus duboscqi* induced by challenge with bacteria or the protozoan parasite *Leishmania major*', *Infection and immunity*. American Society for Microbiology, 72(12), pp. 7140–7146. doi: 10.1128/IAI.72.12.7140-7146.2004.
- Boulanger, N., Bulet, P. and Lowenberger, C. (2006) 'Antimicrobial peptides in the interactions between insects and flagellate parasites', *Trends in Parasitology*, 22(6), pp. 262–268. doi: <https://doi.org/10.1016/j.pt.2006.04.003>.
- Boutros, M., Agaisse, H. and Perrimon, N. (2002) 'Sequential Activation of Signaling Pathways during Innate Immune Responses in *Drosophila*', *Developmental Cell*, 3(5), pp. 711–722. doi: [https://doi.org/10.1016/S1534-5807\(02\)00325-8](https://doi.org/10.1016/S1534-5807(02)00325-8).
- Casas-Sánchez, A. and Acosta-Serrano, Á. (2016) Skin deep. *eLife*. eLife Sciences Publications, Ltd 5: e21506.

- Chakravarty, J. and Sundar, S. (2010) 'Drug resistance in leishmaniasis', Journal of global infectious diseases. Medknow Publications, 2(2), pp. 167–176. doi: 10.4103/0974-777X.62887.
- Chang, C.-I. *et al.* (2004) 'A *Drosophila* pattern recognition receptor contains a peptidoglycan docking groove and unusual L, D-carboxypeptidase activity', PLoS biology. Public Library of Science, 2(9), p. e277.
- Chen, C. *et al.* (1999) 'Identification, purification and properties of a β -1,3-glucan-specific lectin from the serum of the cockroach, *Blaberus discoidalis* which is implicated in immune defence reactions', Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 122(3), pp. 309–319. doi: [https://doi.org/10.1016/S0305-0491\(99\)00020-6](https://doi.org/10.1016/S0305-0491(99)00020-6).
- Cheng, Q. and Aksoy, S. (1999) 'Tissue tropism, transmission and expression of foreign genes *in vivo* in midgut symbionts of tsetse flies.', Insect molecular biology. England, 8(1), pp. 125–132.
- Choe, K.-M., Lee, H. and Anderson, K. V (2005) '*Drosophila* peptidoglycan recognition protein LC (PGRP-LC) acts as a signal-transducing innate immune receptor', Proceedings of the National Academy of Sciences. National Acad Sciences, 102(4), pp. 1122–1126.
- Coutinho-Abreu, I. V. and Ramalho-Ortigao, M. (2010) Transmission blocking vaccines to control insect-borne diseases: a review. Memórias do Instituto Oswaldo Cruz. scielo, 1–12.
- Coutinho-Abreu, I. V *et al.* (2013) 'Characterization of *Phlebotomus papatasi* Peritrophins, and the Role of PpPer1 in *Leishmania major* Survival in its Natural Vector', PLOS Neglected Tropical Diseases. Public Library of Science, 7(3), p. e2132.
- Dean, S. *et al.* (2015) 'A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids', Open biology. The Royal Society, 5(1), p. 140197. doi: 10.1098/rsob.140197.
- Devenport, M. *et al.* (2006) 'Identification of the *Aedes aegypti* peritrophic matrix protein AeIMUC1 as a heme-binding protein.', Biochemistry. United States, 45(31), pp. 9540–9549. doi: 10.1021/bi0605991.
- Dey, R. *et al.* (2018) 'Gut Microbes Egested during Bites of Infected Sand Flies Augment Severity of Leishmaniasis via Inflammasome-Derived IL-1 β ', Cell host & microbe. 2017/12/28, 23(1), p. 134–143.e6. doi: 10.1016/j.chom.2017.12.002.
- Dhiman, R. C. and Yadav, R. S. (2016) 'Insecticide resistance in phlebotomine sandflies in Southeast Asia with emphasis on the Indian subcontinent', Infectious Diseases of Poverty. London: BioMed Central, 5, p. 106. doi: 10.1186/s40249-016-0200-3.
- Diaz-Albiter, H. *et al.* (2012) 'Reactive Oxygen Species-mediated Immunity against *Leishmania mexicana* and *Serratia marcescens* in the Phlebotomine Sand Fly *Lutzomyia longipalpis*', Journal of Biological Chemistry , 287(28), pp. 23995–24003. doi: 10.1074/jbc.M112.376095.

- Dillon, R. J. *et al.* (2006) 'Analysis of ESTs from *Lutzomyia longipalpis* sand flies and their contribution toward understanding the insect–parasite relationship', *Genomics*, 88(6), pp. 831–840. doi: <https://doi.org/10.1016/j.ygeno.2006.06.011>.
- Dillon, R. J. and Lane, P. (1993) 'Bloodmeal digestion in the midgut of *Phlebotomus papatasi* and *Phlebotomus langeroni*', *Medical and Veterinary Entomology*. John Wiley & Sons, Ltd (10.1111), 7(3), pp. 225–232. doi: 10.1111/j.1365-2915.1993.tb00681.x.
- Dimopoulos, G. *et al.* (1998) 'Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during critical transition stages of the parasite life cycle', *The EMBO journal*, 17(21), pp. 6115–6123. doi: 10.1093/emboj/17.21.6115.
- Dimopoulos, G. *et al.* (2001) 'Innate immune defense against malaria infection in the mosquito', *Current Opinion in Immunology*, 13(1), pp. 79–88. doi: [https://doi.org/10.1016/S0952-7915\(00\)00186-2](https://doi.org/10.1016/S0952-7915(00)00186-2).
- Dorn, P. L. *et al.* (2017) 'Genetics and Evolution of Infectious Diseases.' Edited by M. Tibayrenc. Oxford, UK: Elsevier.
- Dostálová, A. *et al.* (2011) 'The midgut transcriptome of *Phlebotomus (Larroussius) perniciosus*, a vector of *Leishmania infantum*: comparison of sugar fed and blood fed sand flies', *BMC Genomics*, 12(1), p. 223. doi: 10.1186/1471-2164-12-223.
- Dostálová, A. and Volf, P. (2012) '*Leishmania* development in sand flies: parasite-vector interactions overview', *Parasites & Vectors*, 5(1), p. 276. doi: 10.1186/1756-3305-5-276.
- Dyer, P., Dean, S. and Sunter, J. (2016) 'High-throughput Gene Tagging in *Trypanosoma brucei*', *Journal of visualized experiments: JoVE*. MyJove Corporation, (114), p. 54342. doi: 10.3791/54342.
- Ellis, D. S. and Evans, D. A. (1977) 'Passage of *Trypanosoma brucei rhodesiense* through the peritrophic membrane of *Glossina morsitans morsitans*.', *Nature*. England, 267(5614), pp. 834–835.
- Farikou, O. *et al.* (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), pp. 115–121. doi: <https://doi.org/10.1016/j.meegid.2009.10.008>.
- Fawaz, E. Y. *et al.* (2016) 'Pyrethroid insecticide resistance mechanisms in the adult *Phlebotomus papatasi* (diptera: Psychodidae)', *Journal of Medical Entomology*, 53(3), pp. 620–628. doi: 10.1093/jme/tjv256.
- Fernando, M. D. A., Kounatidis, I. and Ligoxygakis, P. (2014) 'Loss of Trabid, a New Negative Regulator of the *Drosophila* Immune-Deficiency Pathway at the Level of TAK1, Reduces Life Span', *PLOS Genetics*. Public Library of Science, 10(2), p. e1004117.

- Ferreira, T. N. *et al.* (2018) 'Transmission blocking sugar baits for the control of *Leishmania* development inside sand flies using environmentally friendly beta-glycosides and their aglycones', *Parasites & Vectors*, 11(1), p. 614. doi: 10.1186/s13071-018-3122-z.
- Foley, E. and O'Farrell, P. H. (2003) 'Nitric oxide contributes to induction of innate immune responses to gram-negative bacteria in *Drosophila*', *Genes & development*. Cold Spring Harbor Lab, 17(1), pp. 115–125.
- Fotakis, E. A. *et al.* (2019) 'Identification of *Leishmania* Species in Naturally Infected Sand Flies from Refugee Camps, Greece', *Emerging Infectious Disease journal*, 25(2), p. 361. doi: 10.3201/eid2502.181359.
- Fraihi, W. *et al.* (2017) 'An integrated overview of the midgut bacterial flora composition of *Phlebotomus perniciosus*, a vector of zoonotic visceral leishmaniasis in the Western Mediterranean Basin', *PLOS Neglected Tropical Diseases*. Public Library of Science, 11(3), p. e0005484.
- Freeman, J. C. (1970) 'The presence of trypanosomes in the ecto-peritrophic space of tsetse flies, half an hour after ingestion of the infective blood meal.', *Transactions of the Royal Society of Tropical Medicine and Hygiene*. England, 64(1), pp. 187–188.
- Freeman, J. C. (1973) 'The penetration of the peritrophic membrane of the tsetse flies by trypanosomes.', *Acta tropica*. Netherlands, 30(4), pp. 347–355.
- Gangloff, M. *et al.* (2008) 'Structural insight into the mechanism of activation of the Toll receptor by the dimeric ligand Spätzle', *Journal of Biological Chemistry*. ASBMB, 283(21), pp. 14629–14635.
- Geiger, A. *et al.* (2013) '*Glossina* spp. gut bacterial flora and their putative role in fly-hosted trypanosome development', *Frontiers in Cellular and Infection Microbiology*, p.34.
- Georgel, P. *et al.* (2001) '*Drosophila* immune deficiency (IMD) is a death domain protein that activates antibacterial defense and can promote apoptosis', *Developmental cell*. Elsevier, 1(4), pp. 503–514.
- Georgiadou, S.P., Makaritsis, K. P. and Dalekos, G. N. (2015) Leishmaniasis revisited: Current aspects on epidemiology, diagnosis and treatment (2015/06/30). *Journal of translational internal medicine*. De Gruyter Open 3(2): 43–50.
- Gibson, W. and Bailey, M. (2003) 'The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes', *Kinetoplastid biology and disease*. BioMed Central, 2(1), p. 1. doi: 10.1186/1475-9292-2-1.
- Gilbert, J. A. *et al.* (2016) 'Determinants of Human African Trypanosomiasis Elimination via Paratransgenesis', *PLOS Neglected Tropical Diseases*. Public Library of Science, 10(3), p. e0004465.

- Gobert, V. *et al.* (2003) 'Dual activation of the *Drosophila* toll pathway by two pattern recognition receptors', *Science*. American Association for the Advancement of Science, 302(5653), pp. 2126–2130.
- Gossage, S. M., Rogers, M. E. and Bates, P. A. (2003) 'Two separate growth phases during the development of *Leishmania* in sand flies: implications for understanding the life cycle', *International journal for parasitology*, 33(10), pp. 1027–1034.
- Gross, I. *et al.* (1996) '*Drosophila* immunity: a comparative analysis of the Rel proteins dorsal and Dif in the induction of the genes encoding dipterin and cecropin', *Nucleic acids research*. Oxford University Press, 24(7), pp. 1238–1245.
- Ha, E.-M., Lee, K.-A., Seo, Y. Y., *et al.* (2009) 'Coordination of multiple dual oxidase–regulatory pathways in responses to commensal and infectious microbes in *Drosophila* gut', *Nature Immunology*. Nature Publishing Group, 10, p. 949.
- Ha, E.-M., Lee, K.-A., Park, S. H., *et al.* (2009) 'Regulation of DUOX by the Gαq-phospholipase Cβ-Ca²⁺ pathway in *Drosophila* gut immunity', *Developmental cell*. Elsevier, 16(3), pp. 386–397.
- Haines, L. R. *et al.* (2005) 'Increased expression of unusual EP repeat-containing proteins in the midgut of the tsetse fly (*Glossina*) after bacterial challenge', *Insect Biochemistry and Molecular Biology*, 35(5), pp. 413–423. doi: <https://doi.org/10.1016/j.ibmb.2005.01.005>.
- Haines, L. R. *et al.* (2010) 'Tsetse EP protein protects the fly midgut from trypanosome establishment', *PLoS pathogens*. Public Library of Science, 6(3), pp. e1000793–e1000793. doi: 10.1371/journal.ppat.1000793.
- Hamidou Soumana, I. *et al.* (2014) 'Midgut expression of immune-related genes in *Glossina palpalis gambiense* challenged with *Trypanosoma brucei gambiense*', *Frontiers in Microbiology*, p. 609.
- Hao, Z. *et al.* (2001) 'Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis', *Proceedings of the National Academy of Sciences of the United States of America*. 2001/10/09. The National Academy of Sciences, 98(22), pp. 12648–12653. doi: 10.1073/pnas.221363798.
- Hao, Z. and Aksoy, S. (2002) 'Proventriculus-specific cDNAs characterized from the tsetse, *Glossina morsitans morsitans*', *Insect Biochemistry and Molecular Biology*, 32(12), pp. 1663–1671. doi: [https://doi.org/10.1016/S0965-1748\(02\)00106-6](https://doi.org/10.1016/S0965-1748(02)00106-6).
- Hassan, M. M. *et al.* (2012) 'Insecticide resistance in the sand fly, *Phlebotomus papatasi* from Khartoum State, Sudan', *Parasites & Vectors*, 5(1), p. 46. doi: 10.1186/1756-3305-5-46.
- Heller, K. (2011) 'Tsetse Flies Rely on Symbiotic *Wigglesworthia* for Immune System Development', *PLOS Biology*. Public Library of Science, 9(5), p. e1001070.
- Hotez, P. J. (2008) 'Stigma: The Stealth Weapon of the NTD', *PLOS Neglected Tropical Diseases*. Public Library of Science, 2(4), p. e230.

- Hotez, P. J. (2009) 'The Neglected Tropical Diseases and Their Devastating Health and Economic Impact on the Member Nations of the Organisation of the Islamic Conference', PLOS Neglected Tropical Diseases. Public Library of Science, 3(10), p. e539.
- Hu, C. and Aksoy, S. (2006) 'Innate immune responses regulate trypanosome parasite infection of the tsetse fly *Glossina morsitans morsitans*', Molecular Microbiology. John Wiley & Sons, Ltd (10.1111), 60(5), pp. 1194–1204. doi: 10.1111/j.1365-2958.2006.05180.x.
- van den Hurk, A. F. *et al.* (2012) 'Impact of *Wolbachia* on infection with chikungunya and yellow fever viruses in the mosquito vector *Aedes aegypti*.' PLoS neglected tropical diseases. United States, 6(11), p. e1892. doi: 10.1371/journal.pntd.0001892.
- Hurwitz, I. *et al.* (2014) 'Paratransgenic Control of Leishmaniasis: New Developments BT - Pathogenesis of Leishmaniasis: New Developments in Research', in Satoskar, A. and Durvasula, R. (eds). New York, NY: Springer New York, pp. 25–43. doi: 10.1007/978-1-4614-9108-8_3.
- Jochim, R. C. *et al.* (2008) 'The midgut transcriptome of *Lutzomyia longipalpis*: comparative analysis of cDNA libraries from sugar-fed, blood-fed, post-digested and *Leishmania infantum* chagasi-infected sand flies', BMC Genomics, 9(1), p. 15. doi: 10.1186/1471-2164-9-15.
- Kame-Ngasse, G. I. *et al.* (2018) 'Prevalence of symbionts and trypanosome infections in tsetse flies of two villages of the "Faro and Déo" division of the Adamawa region of Cameroon', BMC microbiology. BioMed Central, 18(Suppl 1), p. 159. doi: 10.1186/s12866-018-1286-5.
- Kamhawi, S. *et al.* (2004) 'A role for insect galectins in parasite survival.', Cell. United States, 119(3), pp. 329–341.
- Kelly, P. H. *et al.* (2017) 'The Gut Microbiome of the Vector *Lutzomyia longipalpis* Is Essential for Survival of *Leishmania infantum*', mBio. Edited by S. M. S. Beverley Alan, 8(1), pp. e01121-16. doi: 10.1128/mBio.01121-16.
- Kessler, R. L. *et al.* (2017) 'Recently differentiated epimastigotes from *Trypanosoma cruzi* are infective to the mammalian host', Molecular Microbiology. John Wiley & Sons, Ltd (10.1111), 104(5), pp. 712–736. doi: 10.1111/mmi.13653.
- Kidd, S. (1992) 'Characterization of the *Drosophila* cactus locus and analysis of interactions between cactus and dorsal proteins', Cell. Elsevier, 71(4), pp. 623–635.
- Kim, M. *et al.* (2006) 'Caspar, a suppressor of antibacterial immunity in *Drosophila*.' Proceedings of the National Academy of Sciences of the United States of America. United States, 103(44), pp. 16358–16363. doi: 10.1073/pnas.0603238103.
- Kleino, A. *et al.* (2005) 'Inhibitor of apoptosis 2 and TAK1-binding protein are components of the *Drosophila* Imd pathway', The EMBO journal. EMBO Press, 24(19), pp. 3423–3434.

- Kleino, A. *et al.* (2008) 'Pirk is a negative regulator of the *Drosophila* Imd pathway.', *Journal of immunology* (Baltimore, Md.:1950). United States, 180(8), pp. 5413–5422.
- Kounatidis, I. and Ligoxygakis, P. (2012) '*Drosophila* as a model system to unravel the layers of innate immunity to infection', *Open biology*. The Royal Society, 2(5), p. 120075. doi: 10.1098/rsob.120075.
- Kubi, C. *et al.* (2006) 'The effect of starvation on the susceptibility of teneral and non-teneral tsetse flies to trypanosome infection.', *Medical and veterinary entomology*. England, 20(4), pp. 388–392. doi: 10.1111/j.1365-2915.2006.00644.x.
- Kuraishi, T. *et al.* (2011) 'Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in *Drosophila melanogaster*.'., *Proceedings of the National Academy of Sciences of the United States of America*. United States, 108(38), pp. 15966–15971. doi: 10.1073/pnas.1105994108.
- Lehane, M. J. (1997) 'Peritrophic Matrix Structure and Function', *Annual Review of Entomology*. Annual Reviews, 42(1), pp. 525–550. doi: 10.1146/annurev.ento.42.1.525.
- Lehane, M. J., Aksoy, S. and Levashina, E. (2004) 'Immune responses and parasite transmission in blood-feeding insects', *Trends in Parasitology*, 20(9), pp. 433–439. doi: <https://doi.org/10.1016/j.pt.2004.07.002>.
- Lehane, M. J., Allingham, P. G. and Weglicki, P. (1996) 'Composition of the peritrophic matrix of the tsetse fly, *Glossina morsitans morsitans*', *Cell and Tissue Research*, 283(3), pp. 375–384. doi: 10.1007/s004410050548.
- Lehane, M. J. and Msangi, A. R. (1991) 'Lectin and peritrophic membrane development in the gut of *Glossina m. morsitans* and a discussion of their role in protecting the fly against trypanosome infection', *Medical and Veterinary Entomology*. John Wiley & Sons, Ltd (10.1111), 5(4), pp. 495–501. doi: 10.1111/j.1365-2915.1991.tb00578.x.
- Leulier, F. *et al.* (2002) 'Inducible expression of double-stranded RNA reveals a role for dFADD in the regulation of the antibacterial response in *Drosophila* adults', *Current Biology*. Elsevier, 12(12), pp. 996–1000.
- Lindh, J. M. and Lehane, M. J. (2011) 'The tsetse fly *Glossina fuscipes fuscipes* (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts', *Antonie van Leeuwenhoek*, 99(3), pp. 711–720. doi: 10.1007/s10482-010-9546-x.
- Lindoso, J. A. *et al.* (2014) 'Visceral leishmaniasis and HIV coinfection in Latin America.', *PLoS neglected tropical diseases*. United States, 8(9), p. e3136. doi: 10.1371/journal.pntd.0003136.
- Lindoso, J. A. L. *et al.* (2016) 'Leishmaniasis-HIV coinfection: current challenges', *HIV/AIDS (Auckland, N.Z.)*. Dove Medical Press, 8, pp. 147–156. doi: 10.2147/HIV.S93789.

- Liniger, M. *et al.* (2003) 'Cleavage of trypanosome surface glycoproteins by alkaline trypsin-like enzyme(s) in the midgut of *Glossina morsitans*', International Journal for Parasitology, 33(12), pp. 1319–1328. doi: [https://doi.org/10.1016/S0020-7519\(03\)00182-6](https://doi.org/10.1016/S0020-7519(03)00182-6).
- Louradour, I. *et al.* (2017) 'The midgut microbiota plays an essential role in sand fly vector competence for *Leishmania major*.', Cellular microbiology. England, 19(10). doi: 10.1111/cmi.12755.
- MacGregor, P. *et al.* (2012) 'Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act', Nature Reviews Microbiology. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 10, p. 431.
- MacLeod, E. T. *et al.* (2007) 'Antioxidants promote establishment of trypanosome infections in tsetse.', Parasitology. England, 134(Pt 6), pp. 827–831. doi: 10.1017/S0031182007002247.
- Maia, C. *et al.* (2015) 'Molecular detection of *Leishmania* DNA and identification of blood meals in wild caught phlebotomine sand flies (Diptera: Psychodidae) from southern Portugal', Parasites & Vectors, 8(1), p. 173. doi: 10.1186/s13071-015-0787-4.
- Maki, K., Shibata, T. and Kawabata, S. (2017) 'Transglutaminase-catalyzed Incorporation of Polyamines Masks the DNA-binding Region of the Transcription Factor Relish', Journal of Biological Chemistry. doi: 10.1074/jbc.M117.779579 .
- Maroli, M. *et al.* (2013) 'Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern.', Medical and veterinary entomology. England, 27(2), pp. 123–147. doi: 10.1111/j.1365-2915.2012.01034.x.
- Marxer, M., Vollenweider, V. and Schmid-Hempel, P. (2016) 'Insect antimicrobial peptides act synergistically to inhibit a trypanosome parasite', Philosophical transactions of the Royal Society of London. Series B, Biological sciences. The Royal Society, 371(1695), p. 20150302. doi: 10.1098/rstb.2015.0302.
- Maslov, D. A. *et al.* (2007) 'Discovery and Barcoding by Analysis of Spliced Leader RNA Gene Sequences of New Isolates of Trypanosomatidae from Heteroptera in Costa Rica and Ecuador', Journal of Eukaryotic Microbiology. John Wiley & Sons, Ltd (10.1111), 54(1), pp. 57–65. doi: 10.1111/j.1550-7408.2006.00150.x.
- McAlister, E. (2017) The secret life of flies. London: Natural History Museum, London.
- Medlock, J. *et al.* (2013) 'Evaluating paratransgenesis as a potential control strategy for African trypanosomiasis.', PLoS neglected tropical diseases. United States, 7(8), p. e2374. doi: 10.1371/journal.pntd.0002374.
- Meinander, A. *et al.* (2012) 'Ubiquitylation of the initiator caspase DREDD is required for innate immune signalling', The EMBO journal. EMBO Press, 31(12), pp. 2770–2783.

- Mellroth, P. and Steiner, H. (2006) 'PGRP-SB1: an *N*-acetylmuramoyl L-alanine amidase with antibacterial activity', *Biochemical and biophysical research communications*. Elsevier, 350(4), pp. 994–999.
- Mercado, T. I. and Colón-Whitt, A. (1982) 'Lysis of *Trypanosoma cruzi* by *Pseudomonas fluorescens*.', *Antimicrobial Agents and Chemotherapy*, 22(6), p. 1051 LP-1057. doi: 10.1128/AAC.22.6.1051.
- Mesu, V. K. B. K. *et al.* (2018) 'Oral fexinidazole for late-stage African *Trypanosoma brucei gambiense* trypanosomiasis: a pivotal multicentre, randomised, non-inferiority trial', *The Lancet*. Elsevier, 391(10116), pp. 144–154. doi: 10.1016/S0140-6736(17)32758-7.
- Miller, N. and Lehane, M. J. (1993) 'Ionic environment and the permeability properties of the peritrophic membrane of *Glossina morsitans morsitans*', *Journal of Insect Physiology*, 39(2), pp. 139–144. doi: [https://doi.org/10.1016/0022-1910\(93\)90105-Z](https://doi.org/10.1016/0022-1910(93)90105-Z).
- Mishima, Y. *et al.* (2009) 'The N-terminal domain of *Drosophila* Gram-negative binding protein 3 (GNBP3) defines a novel family of fungal pattern recognition receptors', *Journal of Biological Chemistry*. ASBMB, 284(42), pp. 28687–28697.
- Moreira, L. A. *et al.* (2009) 'A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and *Plasmodium*.', *Cell*. United States, 139(7), pp. 1268–1278. doi: 10.1016/j.cell.2009.11.042.
- Morrison, L. S. *et al.* (2012) 'Ecotin-like serine peptidase inhibitor ISP1 of *Leishmania major* plays a role in flagellar pocket dynamics and promastigote differentiation', *Cellular Microbiology*. John Wiley & Sons, Ltd (10.1111), 14(8), pp. 1271–1286. doi: 10.1111/j.1462-5822.2012.01798.x.
- Moule, M. G., Monack, D. M. and Schneider, D. S. (2010) 'Reciprocal Analysis of *Francisella novicida* Infections of a *Drosophila melanogaster* Model Reveal Host-Pathogen Conflicts Mediated by Reactive Oxygen and imd-Regulated Innate Immune Response', *PLOS Pathogens*. Public Library of Science, 6(8), p. e1001065.
- Myllymäki, H., Valanne, S. and Rämet, M. (2014) 'The *Drosophila* Imd Signaling Pathway', *The Journal of Immunology*, 192(8), p. 3455 LP-3462. doi: 10.4049/jimmunol.1303309.
- Myskova, J. *et al.* (2007) 'A lipophosphoglycan-independent development of *Leishmania* in permissive sand flies', *Microbes and Infection*, 9(3), pp. 317–324. doi: <https://doi.org/10.1016/j.micinf.2006.12.010>.
- Nagle, A. S. *et al.* (2014) 'Recent developments in drug discovery for leishmaniasis and human African trypanosomiasis', *Chemical reviews*. 2014/11/03. American Chemical Society, 114(22), pp. 11305–11347. doi: 10.1021/cr500365f.
- Okuda, K. *et al.* (2016) '*Leishmania amazonensis* Engages CD36 to Drive Parasitophorous Vacuole Maturation', *PLOS Pathogens*. Public Library of Science, 12(6), p. e1005669.

- Pais, R. *et al.* (2008) 'The Obligate *Mutualist Wigglesworthia glossinidia* Influences Reproduction, Digestion, and Immunity Processes of Its Host, the Tsetse Fly', *Applied and Environmental Microbiology*, 74(19), p. 5965 LP-5974. doi: 10.1128/AEM.00741-08.
- Paquette, N. *et al.* (2010) 'Caspase-mediated cleavage, IAP binding, and ubiquitination: linking three mechanisms crucial for *Drosophila* NF- κ B signaling', *Molecular cell*. Elsevier, 37(2), pp. 172–182.
- Pascoa, V. *et al.* (2002) '*Aedes aegypti* peritrophic matrix and its interaction with heme during blood digestion.', *Insect biochemistry and molecular biology*. England, 32(5), pp. 517–523.
- Peacock, L. *et al.* (2011) 'Identification of the meiotic life cycle stage of *Trypanosoma brucei* in the tsetse fly.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 108(9), pp. 3671–3676. doi: 10.1073/pnas.1019423108.
- Peacock, L. *et al.* (2014) 'Meiosis and haploid gametes in the pathogen *Trypanosoma brucei*', *Current biology : CB*. Cell Press, 24(2), pp. 181–186. doi: 10.1016/j.cub.2013.11.044.
- Pimenta, P. F. *et al.* (1992) 'Stage-specific adhesion of *Leishmania* promastigotes to the sandfly midgut', *Science*, 256(5065), p. 1812 LP-1815. doi: 10.1126/science.1615326.
- Pimenta, P. F. *et al.* (1997) 'A novel role for the peritrophic matrix in protecting *Leishmania* from the hydrolytic activities of the sand fly midgut.', *Parasitology*. England, 115 (Pt 4, pp. 359–369.
- Pires, A. C. A. M. *et al.* (2017) 'Bacterial diversity of wild-caught *Lutzomyia longipalpis* (a vector of zoonotic visceral leishmaniasis in Brazil) under distinct physiological conditions by metagenomics analysis', *Parasites & Vectors*, 10(1), p. 627. doi: 10.1186/s13071-017-2593-7.
- Ponte-Sucre, A. *et al.* (2017) 'Drug resistance and treatment failure in leishmaniasis: A 21st century challenge', *PLOS Neglected Tropical Diseases*. 1160 Battery Street, STE 100, San Francisco, CA 94111 USA: Public Library Science, 11(12). doi: 10.1371/journal.pntd.0006052.
- PrabhuDas, M. R. *et al.* (2017) 'A Consensus Definitive Classification of Scavenger Receptors and Their Roles in Health and Disease', *The Journal of Immunology*, 198(10), p. 3775 LP-3789. doi: 10.4049/jimmunol.1700373.
- Pruzinova, K. *et al.* (2015) 'Comparison of Bloodmeal Digestion and the Peritrophic Matrix in Four Sand Fly Species Differing in Susceptibility to *Leishmania donovani*', *PLOS ONE*. Public Library of Science, 10(6), p. e0128203.
- Ramalho-Ortigao, J. M. *et al.* (2005) 'Characterization of a blood activated chitinolytic system in the midgut of the sand fly vectors *Lutzomyia longipalpis* and *Phlebotomus papatasi*.', *Insect molecular biology*. England, 14(6), pp. 703–712. doi: 10.1111/j.1365-2583.2005.00601.x.

- Ramalho-Ortigao, J. M. and Traub-Cseko, Y. M. (2003) 'Molecular characterization of Llchit1, a midgut chitinase cDNA from the leishmaniasis vector *Lutzomyia longipalpis*.' Insect biochemistry and molecular biology. England, 33(3), pp. 279–287.
- Ramalho-Ortigão, M. *et al.* (2007) 'Exploring the midgut transcriptome of *Phlebotomus papatasi*: comparative analysis of expression profiles of sugar-fed, blood-fed and *Leishmania-major*-infected sandflies', BMC genomics. BioMed Central, 8, p. 300. doi: 10.1186/1471-2164-8-300.
- Ravanti, L. and Kähäri, V.-M. (2000) 'Matrix metalloproteinases in wound repair', International journal of molecular medicine. Spandidos Publications, 6(4), pp. 391–798.
- Real, F. *et al.* (2014) 'Cell-to-cell transfer of *Leishmania amazonensis* amastigotes is mediated by immunomodulatory LAMP-rich parasitophorous extrusions', Cellular microbiology. 2014/06/13. John Wiley & Sons Ltd, 16(10), pp. 1549–1564. doi: 10.1111/cmi.12311.
- Reisenman, C. E. *et al.* (2011) 'Feeding and defecation behavior of *Triatoma rubida* (Uhler, 1894) (Hemiptera: Reduviidae) under laboratory conditions, and its potential role as a vector of Chagas disease in Arizona, USA', The American journal of tropical medicine and hygiene. The American Society of Tropical Medicine and Hygiene, 85(4), pp. 648–656. doi: 10.4269/ajtmh.2011.11-0137.
- Requena-Méndez, A. *et al.* (2017) 'Cost-effectiveness of Chagas disease screening in Latin American migrants at primary health-care centres in Europe: a Markov model analysis', The Lancet Global Health. Elsevier, 5(4), pp. e439–e447. doi: 10.1016/S2214-109X(17)30073-6.
- Reuner, B. *et al.* (1997) 'Cell density triggers slender to stumpy differentiation of *Trypanosoma brucei* bloodstream forms in culture.', Molecular and biochemical parasitology. Netherlands, 90(1), pp. 269–280.
- Rio, R. V. M., Attardo, G. M. and Weiss, B. L. (2016) 'Grandeur Alliances: Symbiont Metabolic Integration and Obligate Arthropod Hematophagy', Trends in Parasitology, 32(9), pp. 739–749. doi: <https://doi.org/10.1016/j.pt.2016.05.002>.
- Rogers, M. B. *et al.* (2011) 'Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*', Genome research. Cold Spring Harbor Laboratory Press, 21(12), pp. 2129–2142. doi: 10.1101/gr.122945.111.
- Rogers, M. E. *et al.* (2004) 'Transmission of cutaneous leishmaniasis by sand flies is enhanced by regurgitation of fPPG', Nature, 430(6998), pp. 463–467. doi: 10.1038/nature02675.
- Rogers, M. E. *et al.* (2008) '*Leishmania* chitinase facilitates colonization of sand fly vectors and enhances transmission to mice', Cellular microbiology. Blackwell Publishing Ltd, 10(6), pp. 1363–1372. doi: 10.1111/j.1462-5822.2008.01132.x.

- Rogerson, E. *et al.* (2018) 'Variations in the Peritrophic Matrix Composition of Heparan Sulphate from the Tsetse Fly, *Glossina morsitans morsitans*', *Pathogens* (Basel, Switzerland). MDPI, 7(1), p. 32. doi: 10.3390/pathogens7010032.
- Rojas, F., Silvester, E., Young, J., Milne, R., Tettey, M., Houston, D. R., Walkinshaw, M. D., Pérez-Pi, I., Auer, M., Denton, H., Smith, T. K., Thompson, J. and Matthews, K. R. (2019) Oligopeptide Signalling through *TbGPR89* Drives Trypanosome Quorum Sensing. *Cell. Elsevier* 176(1): 306–317.e16.
- Romano, P. S. *et al.* (2012) 'Molecular and cellular mechanisms involved in the *Trypanosoma cruzi*/host cell interplay', *IUBMB Life*. John Wiley & Sons, Ltd, 64(5), pp. 387–396. doi: 10.1002/iub.1019.
- Rosales, C. (2017) 'Cellular and Molecular Mechanisms of Insect Immunity', in Sheilds, V. D. C. (ed.) *Insect Physiology and Ecology*. doi: 10.577267107.
- Rose, C. *et al.* (2014) 'An Investigation into the Protein Composition of the Teneral *Glossina morsitans morsitans* Peritrophic Matrix', *PLOS Neglected Tropical Diseases*. Public Library of Science, 8(4), p. e2691.
- Rose, C. *et al.* (2019) '*Trypanosoma brucei* colonises the tsetse gut via an immature peritrophic matrix in the proventriculus', *bioRxiv*, p. 513689. doi: 10.1101/513689.
- Roxstrom-Lindquist, K., Terenius, O. and Faye, I. (2004) 'Parasite-specific immune response in adult *Drosophila melanogaster*: a genomic study.', *EMBO reports*. England, 5(2), pp. 207–212. doi: 10.1038/sj.embor.7400073.
- Rutschmann, S. *et al.* (2000) 'The Rel protein DIF mediates the antifungal but not the antibacterial host defense in *Drosophila*', *Immunity*. Elsevier, 12(5), pp. 569–580.
- Ryu, J.-H. *et al.* (2008) 'Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*.', *Science* (New York, N.Y.). United States, 319(5864), pp. 777–782. doi: 10.1126/science.1149357.
- Sadlova, J. *et al.* (2018) 'Refractoriness of *Sergentomyia schwetzi* to *Leishmania* spp. is mediated by the peritrophic matrix', *PLOS Neglected Tropical Diseases*. Public Library of Science, 12(4), p. e0006382.
- Sádlová, J. and Volf, P. (2009) 'Peritrophic matrix of *Phlebotomus duboscqi* and its kinetics during *Leishmania major* development', *Cell and Tissue Research*, 337(2), p. 313. doi: 10.1007/s00441-009-0802-1.
- Sant'Anna, M. R. *et al.* (2008) 'Gene silencing in phlebotomine sand flies: Xanthine dehydrogenase knock down by dsRNA microinjections', *Insect biochemistry and molecular biology*. 2008/04/04, 38(6), pp. 652–660.
- Sant'Anna, M. R. V *et al.* (2014) 'Colonisation resistance in the sand fly gut: *Leishmania* protects *Lutzomyia longipalpis* from bacterial infection', *Parasites & Vectors*, 7(1), p. 329. doi: 10.1186/1756-3305-7-329.

- Schlein, Y. and Jacobson, R. L. (1998) 'Resistance of *Phlebotomus papatasi* to infection with *Leishmania donovani* is modulated by components of the infective bloodmeal.', *Parasitology*. England, 117 Pt 5, pp. 467–473.
- Schlein, Y., Polacheck, I. and Yuval, B. (1985) 'Mycoses, bacterial infections and antibacterial activity in sandflies (Psychodidae) and their possible role in the transmission of leishmaniasis', *Parasitology*. 2009/04/06. Cambridge University Press, 90(1), pp. 57–66. doi: DOI: 10.1017/S0031182000049015.
- Secundino, N. F. C. *et al.* (2005) '*Lutzomyia longipalpis* Peritrophic Matrix: Formation, Structure, and Chemical Composition', *Journal of Medical Entomology*, 42(6), pp. 928–938. doi: 10.1093/jmedent/42.6.928.
- Serafim, T. D., Coutinho-Abreu, I. V., Oliveira, F., Meneses, C., Kamhawi, S. and Valenzuela, J. G. (2018) Sequential blood meals promote *Leishmania* replication and reverse metacyclogenesis augmenting vector infectivity. *Nature Microbiology* 3(5): 548–555.
- Shakarian, A. M. and Dwyer, D. M. (2000) 'Pathogenic *Leishmania* Secrete Antigenically Related Chitinases Which Are Encoded by a Highly Conserved Gene Locus', *Experimental Parasitology*, 94(4), pp. 238–242. doi: <https://doi.org/10.1006/expr.2000.4493>.
- Sharma, R. *et al.* (2008) 'Asymmetric cell division as a route to reduction in cell length and change in cell morphology in trypanosomes.', *Protist*. Germany, 159(1), pp. 137–151. doi: 10.1016/j.protis.2007.07.004.
- Shen, Z. and Jacobs-Lorena, M. (1999) 'Evolution of chitin-binding proteins in invertebrates.', *Journal of molecular evolution*. Germany, 48(3), pp. 341–347.
- Silverman, N. *et al.* (2000) 'A *Drosophila* I κ B kinase complex required for Relish cleavage and antibacterial immunity', *Genes & development*. Cold Spring Harbor Lab, 14(19), pp. 2461–2471.
- Steverding, D. (2010) 'The development of drugs for treatment of sleeping sickness: a historical review', *Parasites & Vectors*, 3(1), p. 15. doi: 10.1186/1756-3305-3-15.
- Stierhof, Y. D. *et al.* (1999) 'Filamentous proteophosphoglycan secreted by *Leishmania* promastigotes forms gel-like three-dimensional networks that obstruct the digestive tract of infected sandfly vectors.', *European journal of cell biology*. Germany, 78(10), pp. 675–689. doi: 10.1016/S0171-9335(99)80036-3.
- Stöven, S. *et al.* (2000) 'Activation of the *Drosophila* NF- κ B factor Relish by rapid endoproteolytic cleavage', *EMBO reports*. EMBO Press, 1(4), pp. 347–352.
- Strasen, J. *et al.* (2014) 'Epidemiology of Chagas disease in Europe: many calculations, little knowledge.', *Clinical research in cardiology : official journal of the German Cardiac Society*. Germany, 103(1), pp. 1–10. doi: 10.1007/s00392-013-0613-y.
- Sun, H. *et al.* (2002) 'A heterotrimeric death domain complex in Toll signaling', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 99(20), pp. 12871–12876.

- Svobodova, M. *et al.* (2006) ‘Distinct transmission cycles of *Leishmania tropica* in 2 adjacent foci, Northern Israel’, *Emerging infectious diseases*. Centers for Disease Control and Prevention, 12(12), pp. 1860–1868. doi: 10.3201/eid1212.060497.
- Taylor, A. W. (1932) ‘The Development of West African Strains of *Trypanosoma gambiense* in *Glossina tachinoides* under Normal Laboratory Conditions, and at Raised Temperatures’, *Parasitology*. 2009/04/06. Cambridge University Press, 24(3), pp. 401–418. doi: DOI: 10.1017/S0031182000020801.
- Telleria, E. L. *et al.* (2010) ‘Trypsin-Like Serine Proteases in *Lutzomyia longipalpis* – Expression, Activity and Possible Modulation by *Leishmania infantum chagasi*’, *PLOS ONE*. Public Library of Science, 5(5), p. e10697.
- Telleria, E. L. *et al.* (2012) ‘Caspar-like Gene Depletion Reduces *Leishmania* Infection in Sand Fly Host *Lutzomyia longipalpis*’, *Journal of Biological Chemistry*, 287(16), pp. 12985–12993. doi: 10.1074/jbc.M111.331561.
- Telleria, E. L. *et al.* (2013) ‘Bacterial feeding, *Leishmania* infection and distinct infection routes induce differential defensin expression in *Lutzomyia longipalpis*’, *Parasites & Vectors*, 6(1), p. 12. doi: 10.1186/1756-3305-6-12.
- Telleria, E. L. *et al.* (2018) ‘*Leishmania*, microbiota and sand fly immunity’, *Parasitology*. 2018/06/20. Cambridge University Press, 145(10), pp. 1336–1353. doi: 10.1017/S0031182018001014.
- Tinoco-Nunes, B. *et al.* (2016) ‘The sandfly *Lutzomyia longipalpis* LL5 embryonic cell line has active Toll and Imd pathways and shows immune responses to bacteria, yeast and *Leishmania*’, *Parasites & vectors*. BioMed Central, 9, p. 222. doi: 10.1186/s13071-016-1507-4.
- Tiwary, P. *et al.* (2013) ‘Seasonal Variation in the Prevalence of Sand Flies Infected with *Leishmania donovani*’, *PLOS ONE*. Public Library of Science, 8(4), p. e61370.
- Tonui, W. K., Mbatia, P. A., Anjili, C. O., Orago, A. S., Turco, S. J., Githure, J. I., Koech, D. K. 2001. Transmission blocking vaccine studies in leishmaniasis: I. Lipophosphoglycan is a promising transmission blocking vaccine molecule against cutaneous leishmaniasis. *East Afr Med J* 78: 84-89.
- Tsichritzis, T. *et al.* (2007) ‘A *Drosophila* ortholog of the human cylindromatosis tumor suppressor gene regulates triglyceride content and antibacterial defense.’, *Development* (Cambridge, England). England, 134(14), pp. 2605–2614. doi: 10.1242/dev.02859.
- Tsuda, M. *et al.* (2005) ‘The RING-finger scaffold protein Plenty of SH3s targets TAK1 to control immunity signalling in *Drosophila*.’, *EMBO reports*. England, 6(11), pp. 1082–1087. doi: 10.1038/sj.embor.7400537.
- Valanne, S., Wang, J.-H. and Rämet, M. (2011) ‘The *Drosophila* Toll Signaling Pathway’, *The Journal of Immunology*, 186(2), p. 649 LP-656. doi: 10.4049/jimmunol.1002302.

- Verma, S. *et al.* (2017) 'Role of inhibitors of serine peptidases in protecting *Leishmania donovani* against the hydrolytic peptidases of sand fly midgut', *Parasites & vectors*. BioMed Central, 10(1), p. 303. doi: 10.1186/s13071-017-2239-9.
- Vickerman, K. and Tetley, L. (1990) 'Flagellar Surfaces of Parasitic Protozoa and Their Role in Attachment BT - Ciliary and Flagellar Membranes', in Bloodgood, R. A. (ed.). Boston, MA: Springer US, pp. 267–304. doi: 10.1007/978-1-4613-0515-6_11.
- De Vooght, L., Van Keer, S. and Van Den Abbeele, J. (2018) 'Towards improving tsetse fly paratransgenesis: stable colonization of *Glossina morsitans morsitans* with genetically modified *Sodalis*', *BMC Microbiology*, 18(1), p. 165. doi: 10.1186/s12866-018-1282-9.
- Wakid, M. H. and Bates, P. A. (2004) 'Flagellar attachment of *Leishmania* promastigotes to plastic film *in vitro*', *Experimental Parasitology*, 106(3), pp. 173–178. doi: <https://doi.org/10.1016/j.exppara.2004.03.001>.
- Walshe, D. P. *et al.* (2009) 'Prolonged gene knockdown in the tsetse fly *Glossina* by feeding double stranded RNA.', *Insect molecular biology*. England, 18(1), pp. 11–19. doi: 10.1111/j.1365-2583.2008.00839.x.
- Wang, J. *et al.* (2009) 'Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission', *Proceedings of the National Academy of Sciences*, 106(29), p. 12133 LP-12138. doi: 10.1073/pnas.0901226106.
- Wang, J. and Aksoy, S. (2012) 'PGRP-LB is a maternally transmitted immune milk protein that influences symbiosis and parasitism in tsetse's offspring', *Proceedings of the National Academy of Sciences*, 109(26), p. 10552 LP-10557. doi: 10.1073/pnas.1116431109.
- Wang, L., Sloan, M. A. and Ligoxygakis, P. (2019) 'Intestinal NF- κ B and STAT signalling is important for uptake and clearance in a *Drosophila-Herpetomonas* interaction model', *PLOS Genetics*. Public Library of Science, 15(3), p. e1007931.
- Weiss, B. L. *et al.* (2013) 'Trypanosome Infection Establishment in the Tsetse Fly Gut Is Influenced by Microbiome-Regulated Host Immune Barriers', *PLOS Pathogens*. Public Library of Science, 9(4), p. e1003318.
- Weiss, B. L. *et al.* (2019) 'Colonization of the tsetse fly midgut with commensal *Kosakonia cowanii* Zambiae inhibits trypanosome infection establishment', *PLOS Pathogens*. Public Library of Science, 15(2), p. e1007470.
- Weiss, B. L., Wang, J. and Aksoy, S. (2011) 'Tsetse Immune System Maturation Requires the Presence of Obligate Symbionts in Larvae', *PLOS Biology*. Public Library of Science, 9(5), p. e1000619.
- Welburn, S. C. *et al.* (1993) 'Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies', *Parasitology*. 2009/04/06. Cambridge University Press, 107(2), pp. 141–145. DOI: 10.1017/S003118200006724X.

- WHO (2017) Leishmaniasis fact sheet. [WWW document]
- WHO (2018a) Chagas disease (American trypanosomiasis) Fact Sheet. [WWW document]
- WHO (2018b) Trypanosomiasis, human African (sleeping sickness). [WWW document]
- Wu, L. P. and Anderson, K. V (1998) 'Regulated nuclear import of Rel proteins in the *Drosophila* immune response', *Nature*. Nature Publishing Group, 392(6671), p. 93.
- Wu, Q., Patočka, J. and Kuča, K. (2018) 'Insect Antimicrobial Peptides, a Mini Review', *Toxins*. MDPI, 10(11), p. 461. doi: 10.3390/toxins10110461.
- Wu, S.-C. *et al.* (2012) 'Infection-Induced Intestinal Oxidative Stress Triggers Organ-to-Organ Immunological Communication in *Drosophila*', *Cell Host & Microbe*, 11(4), pp. 410–417. doi: <https://doi.org/10.1016/j.chom.2012.03.004>.
- Xiao, T. *et al.* (1999) 'Three-dimensional structure of a complex between the death domains of Pelle and Tube', *Cell*. Elsevier, 99(5), pp. 545–555.
- Yan, J. *et al.* (2002) 'Cloning and functional expression of a fat body-specific chitinase cDNA from the tsetse fly, *Glossina morsitans morsitans*', *Insect Biochemistry and Molecular Biology*, 32(9), pp. 979–989. doi: [https://doi.org/10.1016/S0965-1748\(02\)00034-6](https://doi.org/10.1016/S0965-1748(02)00034-6).
- Yosef, S. *et al.* (1991) 'Chitinase secreted by *Leishmania* functions in the sandfly vector', *Proceedings of the Royal Society of London. Series B: Biological Sciences*. Royal Society, 245(1313), pp. 121–126. doi: 10.1098/rspb.1991.0097.
- Yu, X.-Q., Gan, H. and R. Kanost, M. (1999) 'Immulectin, an inducible C-type lectin from an insect, *Manduca sexta*, stimulates activation of plasma prophenol oxidase', *Insect Biochemistry and Molecular Biology*, 29(7), pp. 585–597. doi: [https://doi.org/10.1016/S0965-1748\(99\)00036-3](https://doi.org/10.1016/S0965-1748(99)00036-3).
- Zhao, H. W., Zhou, D. and Haddad, G. G. (2011) 'Antimicrobial peptides increase tolerance to oxidant stress in *Drosophila melanogaster*', *Journal of Biological Chemistry*. ASBMB, 286(8), pp. 6211–6218

CHAPTER TWO

The genome of *Herpetomonas muscarum*

CHAPTER TWO:

The genome of *Herpetomonas muscarum*

The work presented in this chapter was done in collaboration with, and under the guidance of, Dr. James Cotton (Wellcome Sanger Institute, Hinxton, UK). The DNA extraction from *Herpetomonas muscarum* was performed by Dr. Lihui Wang. DNA sequencing and genome assembly was done at the Wellcome Sanger Institute by James Cotton, Karen Brooks and Mandy Sanders. All other work, including RNA extraction from *H. muscarum*, genome annotation and comparative genomics is my own. All sequencing data was submitted to the European Nucleotide Archive (ENA) under accession numbers ERP008869 and ERP023447.

Preface

To misquote William Shakespeare, the genome is the window to how an organism functions. Despite many notable biological scientific discoveries in the pre-genomics era, having the genome sequence for the organism under study is rapidly becoming essential. Not only does the genome sequence provide a blue print for genetic manipulations and transcriptomics studies, but comparative genomics can give insights into evolutionary history and an organism's 'life style'. In the context of insect-parasite interactions, we are discovering that the genotypes of both organisms are crucial in deciding infection outcome. Therefore, it is likely that the answers to many questions about these systems lie within the genome.

The genomes of trypanosomatids

Early work on trypanosomatid genomes was focused on the dioxenous, disease-causing trypanosomatids *Trypanosoma brucei* (Berriman *et al.*, 2005), *Trypanosoma cruzi* (El-Sayed, Myler, Bartholomeu, *et al.*, 2005) and species of the *Leishmania* genus (Ivens *et al.*, 2005; Downing *et al.*, 2011; Real *et al.*, 2013). However, increasing numbers of monoxenous trypanosomatid genomes are being investigated (Motta *et al.*, 2013) including: bee parasites from the *Crithidia* genus (Schmid-Hempel *et al.*, 2018), *Lotmaria passim* (Runckel, DeRisi and Flenniken, 2014) and *Leptomonas pyrrhocoris* (Flegontov *et al.*, 2016) a globally disseminated parasite isolated from fire bugs (*Pyrrhocoris apterus*). These genomic studies, and their predecessors, have revealed a number of distinctive features in trypanosomatids (El-Sayed, Myler, Blandin, *et al.*, 2005; Peacock *et al.*, 2007; Teixeira *et al.*, 2012).

The ancestral state of the trypanosomatid genomes is believed to be similar to the genomes of *Leishmania* and *Trypanosoma cruzi* – made of > 30 small, mostly diploid chromosomes. The 11 chromosomes of *Trypanosoma brucei* appear to have been formed from fusions of smaller chromosomes. This observation is based on the high level of synteny, or conserved gene order, between trypanosomatids. For example, even though they were predicted to have diverged before the existence of mammals, the *Leishmania major* genome is 75% syntenic to that of *Trypanosoma brucei* (El-Sayed, Myler, Blandin, *et al.*, 2005).

The genomic DNA (gDNA) is arranged into polycistronic (multi-gene) transcriptional units, made up of seemingly unrelated genes. These units can be transcribed bi-directionally from the divergent strand switch regions (SSR) which contain several

transcriptional start sites (Martínez-Calvillo *et al.*, 2003). However these regions are not essential for chromosome stability through DNA replication or gene expression, as transcripts were still produced when the strand switch region was deleted from *Leishmania major* chromosome 1 (Dubessay *et al.*, 2002). Puechberty *et al.* (2007) analysed the strand switch region of chromosome 1 for several species of *Leishmania*. They identified a conserved poly(C) tract in the otherwise highly polymorphic region containing the putative transcription initiation site. A number of poly(C)-binding proteins have been implicated in transcriptional control in other organisms (Blyn *et al.*, 1997; Collier *et al.*, 1998; Andino *et al.*, 1999; Ostareck *et al.*, 2001; Eid *et al.*, 2009), as such this region likely plays an important role in transcriptional control in *Leishmania*.

Unlike other eukaryotes, trypanosomatid genes do not typically have introns. There are just four genes show to be *cis*-spliced in trypanosomatids: poly-(A) polymerase (Mair *et al.*, 2000), the ATP-dependant RNA helicase DBP2B, an RNA-binding protein (LmjF29.200), the hypothetical protein LmjF32.08550 (Ivens *et al.*, 2005) and the mini-exon gene (or splice leader sequence). The splice leader sequence, a further curiosity of trypanosomatid genomes, is a short nucleotide sequence which is *trans*-spliced to the 5' of each mRNA (Agabian, 1990). The *trans*-splicing of a splice leader sequence has been observed in several other organisms including nematodes and chordates (Vandenberghe, Meedel and Hastings, 2001; Ganot *et al.*, 2004; Lidie and Van Dolah, 2007; Zhang *et al.*, 2007). In trypanosomatids, splice leader is encoded by the mini-exon genes which are found throughout the genome in tandem arrays. Each mini-exon has two components; the highly conserved 39bp sequence trans-spliced on to mRNAs (the exon) and a less well conserved intronic sequence. Between each mini-exon gene there is a variable spacer region which is not transcribed.

Trans-splicing of splice leader sequences (SLS) precedes, and is coupled to, the maturation of the 3' end of mRNAs (Ullu, Matthews and Tschudi, 1993). The SLS is spliced in at specific acceptor sites, a polypyrimidine tract followed by di-nucleotide AG, in the 5' untranslated region (UTR) of the mRNA. Then mRNA can be polyadenylated by *cis*-splicing. This process differs from the 'eukaryotic norm', where polyadenylation is targeted to a conserved AAUAAA motif followed by U- or GU-rich sequence in the 3' UTR. Trypanosomatid genes lack the highly conserved AAUAAA motif and downstream GU-rich sequences. Instead, polyadenylation is targeted to pyrimidine-rich sequences upstream of the adjacent gene's trans-splice site, rather than being defined to a single site/motif. As such, pyrimidine-rich sequences of UTRs site are crucial to both capping and polyadenylation of mRNA in trypanosomatids (LeBowitz *et al.*, 1993; Matthews, Tschudi and Ullu, 1994; Benz *et al.*, 2005).

In addition to their genomic DNA, trypanosomatids, like all kinetoplastids, have a membrane-bound organelle containing mitochondrial DNA (kDNA) called the kinetoplast. The kDNA is arranged in interlocking 'maxi-circles' which encode typical mitochondrial proteins (Chen *et al.*, 1995; Lukeš *et al.*, 2002; Borghesan *et al.*, 2013). Each maxi circle in turn has thousands of associated 'mini-circles', identifiable by the presence of the universal minicircle sequence (Liu *et al.*, 2005), which contain RNA-editing genes (Lukeš *et al.*, 2002).

The content of trypanosomatid genomes seems to reflect the lifestyle of the organism. *Trypanosoma brucei*, for example, has no intercellular life cycle stages and has a much-reduced metabolic repertoire compared to *Leishmania* (Berriman *et al.*, 2005). This likely reflects the greater nutrient availability in blood serum in contrast to within the

parasitophorous vacuoles that *Leishmania* inhabit during infection of vertebrates.

Similarly, *Leishmania* are capable of hydrolysing disaccharides whereas *T. brucei* cannot (Berriman *et al.*, 2005) – reflecting the (partially) sap-feeding vs. the obligate blood-feeding life styles of their respective insect hosts sand flies and tsetse.

Here, I report the genome of *Herpetomonas muscarum* which was isolated from a wild population of *Drosophila melanogaster* in Oxfordshire, UK. This is the first report of a *Herpetomonas* whole genome sequence. The genome was sequenced using a hybrid sequencing approach, where single molecule real-time (SMRT) sequencing was performed and the outputs then corrected using Illumina sequencing reads. For full details of this methodology, please see Chapter Seven.

Results and Discussion

Genome assembly

DNA sequencing reads were assembled using HGAP3 (Chin *et al.*, 2013) to give a 41.7 Mbp genome with a GC content of 57.8%. The assembly was then corrected using ICORN2 (Otto *et al.*, 2010) and REAPR (Hunt *et al.*, 2013) to give 264 scaffolds (N50=707455 bp, n=21). The median read coverage per scaffold, based on sliding 1kb windows, was 114 reads per base. There are populations of scaffolds coverage with at approximately 50x and 160x which may represent monosomic and trisomic scaffolds (Figure 2.1). It is of interest that many of the *H. muscarum* scaffolds with a high copy number map closely to the *Leishmania* chromosome 31. This chromosome is known to be tetrasomic in *L. major* (Akopyants *et al.*, 2009) and even hexasomic in *L. braziliensis* (Rogers *et al.*, 2011). Using comparative genomics of more trypanosomatid genomes to find common elements of these high copy chromosomes, it may be possible to resolve the

source of their supernumeracy in the future – though this is beyond the scope of this work.

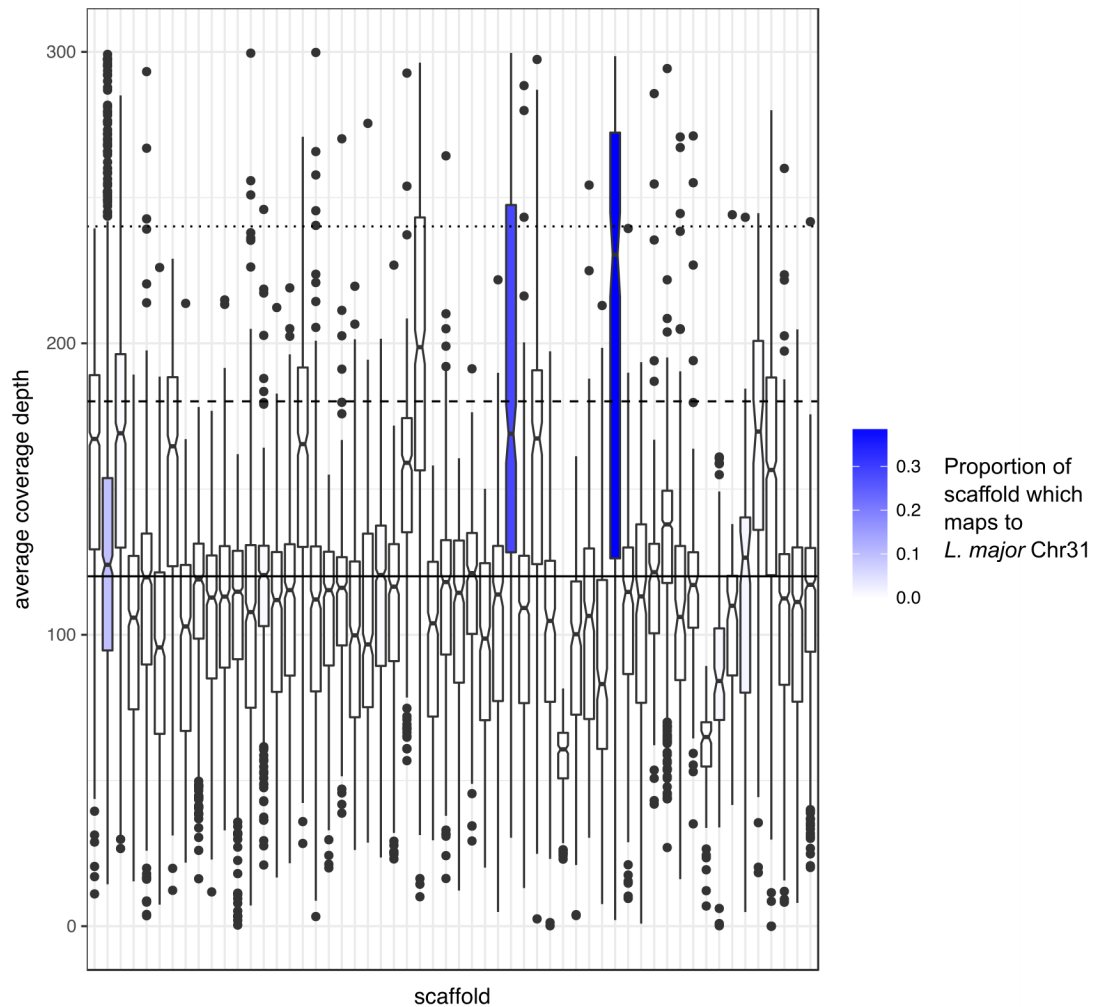


Figure 2.1 - Average coverage depth of *Herpetomonas muscarum* scaffolds > 100 kb. The solid line is the global median coverage. The dashed line is 1.5x and the dotted line is 2x of the global median respectively. In blue are scaffolds which were mapped, by PROmer (Kurtz et al. 2004), to the *Leishmania major* chromosome 31 - sequences of 300bp which map with > 70% identity. The shade of blue represents the proportion of the scaffold which was mapped.

To assess the *H. muscarum* assembly, a kmer profile of the sequencing reads was produced using JellyFish. (Marçais and Kingsford, 2011). A kmer size of 21bp was used to generate the profile. The kmer distribution was then analysed by Genome Scope v1.0 (Vurture *et al.*, 2017), which gave an estimated haploid genome length of approximately 35.2 Mbp with an error rate of 0.92% (Figure 2.2). This indicates the assembly is close to the correct size.

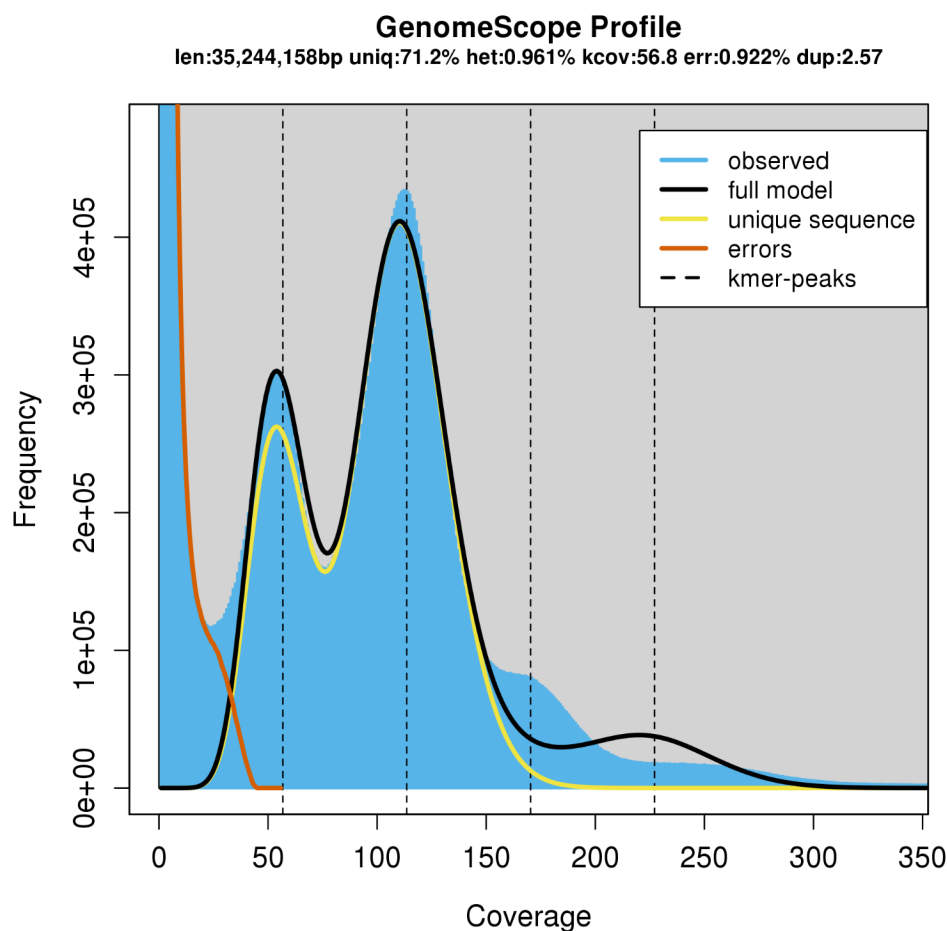


Figure 2.2 - Genome scope profile of *Herpetomonas muscarum* scaffolds (kmer size = 21). **Uniq** – percentage of genome that is not repetitive. **Kcov** – average kmer coverage. **Dup** – average rate of read duplicates. **Err** – error rate.

Using the Argus Optical Mapping System (OpGen), an optical map was generated from high molecular weight genomic DNA which was digested with KpnI and BamHI and captured in agarose plugs. The optical map consisted of 37-38 chromosomes with approximately half being contiguous.

Genome annotation

Three separate annotations were generated using online Companion (COMprehensive Parasite ANnotatION) software (Steinbiss *et al.*, 2016). The Companion automated parasite genome annotation pipe line utilizes five gene prediction tools to ensure the most complete annotations possible; RATT (Otto *et al.*, 2011), SNAP (Korf, 2004), INFERNAL (Nawrocki and Eddy, 2013), AUGUSTUS (Stanke *et al.*, 2006) and ARAGORN (Laslett and Canback, 2004). RATT is a homology-based gene prediction tool which will transfer gene annotations to a new genome based on its synteny to a reference genome sequence. Additionally, RATT is able to detect whether the gene models have changed between the two sequences and can correct changed start and stop codons, or frameshifts. The remaining gene prediction tools utilised by Companion work *ab initio* and do not require a reference sequence. These tools predict genomic features, such as exons, introns, and splice sites, in the DNA sequence based on what is known about these features in eukaryotes e.g. each gene must have a start codon and a stop codon.

A non-redundant set of genes is produced from each of the five annotation platforms by choosing the ‘best explanation’ for each locus depending on the source of information, gene lengths and semantic properties e.g. splice site canonicity. The Companion tool has

an additional filter for kinetoplastid genomes. Given that genes in these organisms are groups into uni-directional clusters (polycistrons) the software uses a implemented a filtering method to eliminate overprediction of genes on the opposing strand to a polycistron. Once the final annotation model has been created functional information (GO terms, orthologue information and putative gene names) are added to the gene set from OrthoMCL data (Li, Stoeckert Jr and Roos, 2003). In the absence of known orthologs the best Pfam-A hit is used to assign a putative function. To annotate putative pseudogenes protein-DNA alignments (which allow for frame shifts) are generated by the LAST aligner are used.

The three annotations were produced using gene models from *Leishmania braziliensis*, *Leishmania major* and *Trypanosoma brucei* – a summary of the individual annotations can be found in Table 2.1. These gene models were then manually combined produce a ‘maximal annotation’ containing 12, 687 genes. Overall, 52.2% of the genome is covered by coding sequences (CDSs). The average CDS length was 1779 bp. This annotation is conservative as I kept all unique open reading frames produced by the gene models, even in cases where the gene prediction was not strongly supported by RNA-sequencing evidence. It is therefore likely that this annotation contains a higher number of genes than the ‘true annotation’. However, ~12, 500 annotated genes is close to that reported for other trypanosomatid species, e.g. the *Trypanosoma brucei* (Tb927_v5.1) has 11, 984 annotated genes, and as such is not infeasibly large.

Feature	Gene model			<i>H. muscarum</i> v1.0
	<i>L. major</i>	<i>L. braziliensis</i>	<i>T. brucei</i>	
genes	11094	11214	11341	12687
mRNAs	10569	10689	10816	12162
CDSs	10580	10715	10816	12175
polypeptides	11341	11398	11368	12934
pseudogenes	772	709	552	772
rRNAs	168	168	168	168
snRNAs	3	3	3	3
snoRNAs	181	181	181	181
tRNAs	173	173	173	173

Table 2.1 - Summary statistics of three *Herpetomonas muscarum* genome annotations using gene models from *Leishmania major*, *Leishmania braziliensis* and *Trypanosoma brucei* and a maximal annotation which combines all three annotations.

Genome structure

As seen in other trypanosomatid genomes, open reading frames are found on both strands. Genes appear to be arranged in groups, with all members of one group on the same strand and in the same direction. This is indicative of polycistronic transcripts which are typical in trypanosomatids. The regions between polycistrons, commonly referred to as strand switch regions, are thought to contain transcriptional start sites. As such, I investigated these regions in *H. muscarum* further in an attempt to identify conserved motifs which may be associated with transcription. To identify the strand switch region sequences, the genes where the downstream open reading frame is on the opposing strand of the same

scaffold were identified. This highlighted 386 genes which were found on 112 different scaffolds. These putative strand switches were manually inspected and could be grouped into different three situations. There were 128 *bona fide* strand switches which were either divergent (72 cases) or convergent (56 cases) (Table S2.1). There were 166 cases where a single gene (or small group of < 3 genes) had become inverted within a polycistron (166 cases). Finally, there were 92 cases where a strand switch does occur, but the precise locus was unclear. These cases tended to be at where a single gene at the end of a scaffold was on the opposing strand to all other genes on the scaffold – as such it was unclear if this represented a *bona fide* strand switch or a single gene inversion.

The divergent strand switch regions, which are likely to encode transcription initiation sequences, varied from 72-10954bp with a mean of 3184bp. The sequences were analysed by MEME (Bailey and Elkan, 1994) and GLAM2 (Frith *et al.*, 2008) to look for enriched DNA motifs within the divergent strand switch regions. This analysis identified 16 enriched motifs 5-50bp in length which were enriched in the 72 divergent strand switch sequences compared to a control set of the same sequences which had been randomly ‘scrambled’. Of these motifs, four sequences were present in over half of the strand switch regions (Figure 2.3A and B). These motifs included poly(C) and poly(G) runs, which have also been identified in the only strand switch region of chromosome 1 of *Leishmania* species (Puechberty *et al.*, 2007). Furthermore, the strand switch region of *L. major* chromosome 1, as described by Puechberty *et al.* (2007), shares 47.43% identity to the strand switch region on the *H. muscarum* scaffold 40 (Figure S2.1). This identity increases to 57.8% in the 73bp conserved region of *L. major* strand switch region containing the bi-directional transcriptional start site (Martínez-Calvillo *et al.*, 2003). This was surprisingly high as, across the *Leishmania* genus the sequence identity of this

conserved region is 54% and the whole strand switch region only shows 30% identity.

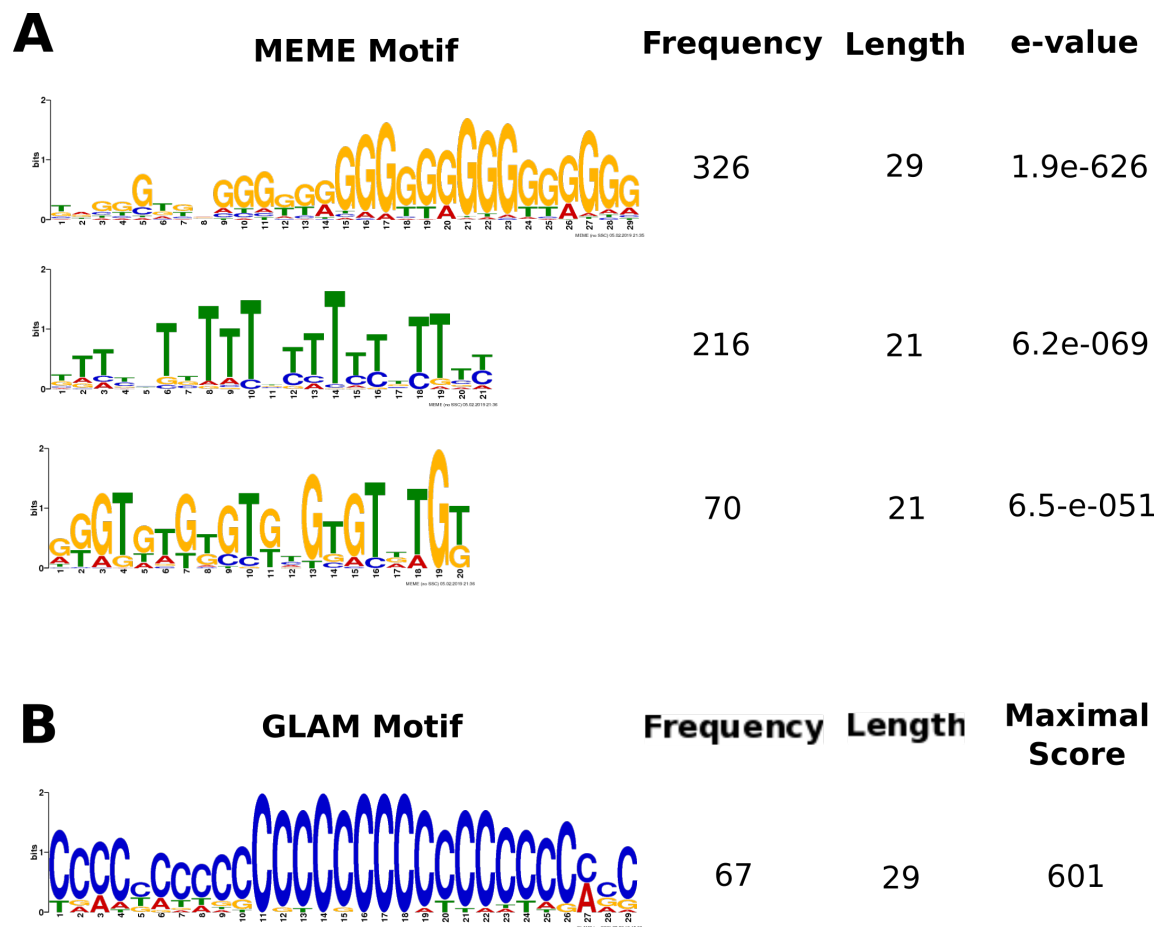


Figure 2.3 - Common motifs of *Herpetomonas muscarum* strand switch regions as identified by MEME (A) and GLAM2 (B).

Repetitive Regions

There are a diverse range of repetitive sequences in eukaryotic genomes. They are commonly grouped into tandem repeats (including centromeres and telomeres) and transposable elements. The latter of which have the ability to move between loci in the genome. Repetitive sequences within the genome, originally thought to be ‘junk DNA’ are now known to place important roles in genomic structure and stability (Shammas, 2011).

RepeatMasker (Smit, A. F. A., Hubley, R., Green, 2017) was used to gain an overview of the repeat content of the *H. muscarum* genome (Table 2.2). For this analysis only scaffolds greater than 100kb (56 scaffolds, 34.98 Mb) were used - to filter out kDNA mini and maxi circles. This identified that ~25% of the *H. muscarum* genome is comprised of interspersed repeats and low complexity sequences. Almost 9% of the *H. muscarum* genome was found to be interspersed repeats, including: long interspersed elements (LINEs), short interspersed elements (SINEs), and long terminal repeats (LTRs). Few studies have explored the repeat content of trypanosomatid genomes in detail. Recent work by Pita *et al.* (2019) demonstrated the variability in repeat content between *T. cruzi*, *T. brucei* and *L. major* (Pita *et al.*, 2019). A staggering 51% of the *T. cruzi* genome was found to be repetitive sequence, whilst *T. brucei* and *L. major* had much lower proportions of repetitive content – 21% and 8.8% respectively. Therefore, in terms of the amount of repetitive content, the *H. muscarum* genome appears closest to *T. brucei*. However, the proportions of specific repeat elements differ between the two organisms. For example, in *T. brucei* 5.67% of the gDNA was found to be LINEs, whereas I observed LINEs made up just 0.14% of *H. muscarum* gDNA. The reasons for the disparity in repeat content between repeat content in trypanosomatids is unclear.

Further investigation into this largely overlooked aspects of trypanosomatid genomes is needed.

Repeat Element Class	Frequency	Total Bases Occupied	Percent of total sequence (~35Mb)
SINEs	31	3385	0.01%
LINEs	331	50542	0.14%
LTRs	173	431969	1.23%
LINE 1	24	9828	0.03%
DNA elements	2622	837675	2.39%
hAT-Charlie	12	32476	0.09%
Unclassified	1855	1780710	5.09%
Total interspersed repeats		3104281	8.87%
Small RNAs	31	186806	0.53%
Satellites	6	7354	0.02%
Simple Repeats	52893	2048169	5.86%
Low complexity	5914	280207	0.80%

Table 2.2 – Repeat elements identified in the *Herpetomonas muscarum* genome scaffolds >100kb

Telomeres

To locate putative telomeric regions, the *H. muscarum* scaffolds were ‘BLASTed’ against a sequence of three tandem telomeric hexamer repeats – 5’-TTAGGG-3’ (Moyzis *et al.*, 1988; da Silva *et al.*, 2010). This hexamer repeat is highly conserved amongst eukaryotes. The BLAST alignments showed 65 scaffolds contained the triple hexamer repeats at least once with 100% identity. Manual inspection of the alignments revealed 85 genomic regions sized 18-53, 264bp comprised of the telomeric hexamer repeat (Table S2.2). Through comparison of the coordinates of these regions with the scaffold sizes, I predict that scaffolds 1, 4, 11-13, 15,16 and 29 represent (approximately) complete

chromosomes, in that they have two predicted telomeric regions at either end of the scaffold.

Genome Content

Orthofinder (Emms and Kelly, 2015) was used to identify orthologous proteins from other trypanosomatids in *H. muscarum*. For the analysis, protein coding genes from the following organisms were used: *Trypanosoma rangeli*, *Trypanosoma grayi*, *Trypanosoma brucei brucei*, *Trypanosoma brucei gambiense*, *Trypanosoma vivax*, *Trypanosoma congolense*, *Leishmania donovani*, *Leishmania major*, *Leishmania mexicana*, *Leptomonas pyrrhocoris*, *Leptomonas seymori*, *Phytomonas EMI*, *Crithidia fasciculata* and finally, *Bodo saltans* as an outgroup. From these 14 species, 85.7% of proteins were assigned to 10, 219 orthogroups (for a summary see Table 2.3, for the full set of orthogroups see Table S2.3). I found 7, 349 of these orthogroups contained *H. muscarum* genes. There were 1, 254 *H. muscarum* protein sequences unable to be assigned to an orthogroup (10.3% of the proteome). Additionally, there were 48 orthogroups containing exclusively *H. muscarum* protein sequences, these groups contain 313 genes for which no conserved domains or significant BLAST hits could be found.

H. muscarum shared more orthogroups with *L. major* (6616), than *T. brucei* (5955) - which is more distantly related (Figure 2.4A). Within the Leishmaniiae clade, *H. muscarum* and two species of ‘old world’ *Leishmania*, *L. major* and *L. donovani*, shared 80% of their orthogroups (Figure S2.2A). However, *H. muscarum* had marginally more orthogroups in common, 6803, with the two *Leptomonas sp.* used in the analysis (Figure S2.2B). The phylogenetic tree produced by Orthofinder, based on sequences from proteins in orthogroups which contain a single gene from every species used in the

analysis (Figure 2.4B), is consistent with others published for the trypanosomatids (Maslov *et al.*, 2013).

Total number of genes	169 196
Number of genes in orthogroups	144 984
Number of unassigned genes	24 212
Percentage of genes in orthogroups	85.70%
Percentage of unassigned genes	14.30%
Number of orthogroups	10 219
Number of species-specific orthogroups	228
Number of genes in species-specific orthogroups	2 543
Percentage of genes in species-specific orthogroups	1.50%
Mean orthogroup size	14.2
Median orthogroup size	15
Number of orthogroups with all species present	153
Number of single-copy orthogroups	8

Table 2.3 – A summary of orthogroup assignments of representative trypanosomatid proteomes.

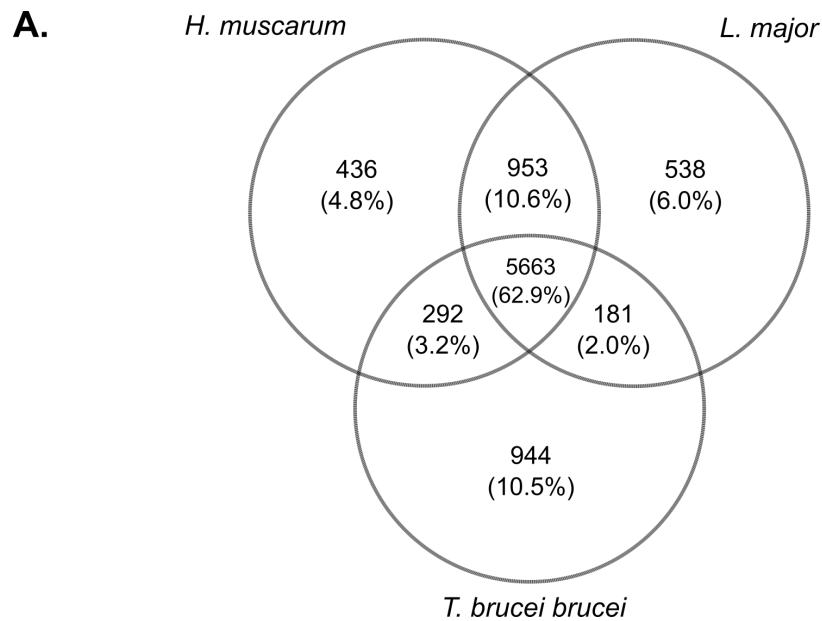


Figure 2.4A - Venn diagram showing shared orthogroups between *Herpetomonas muscarum*, *Leishmania major* and *Trypanosoma brucei*.
Numbers of orthogroups (percentage of orthogroups).

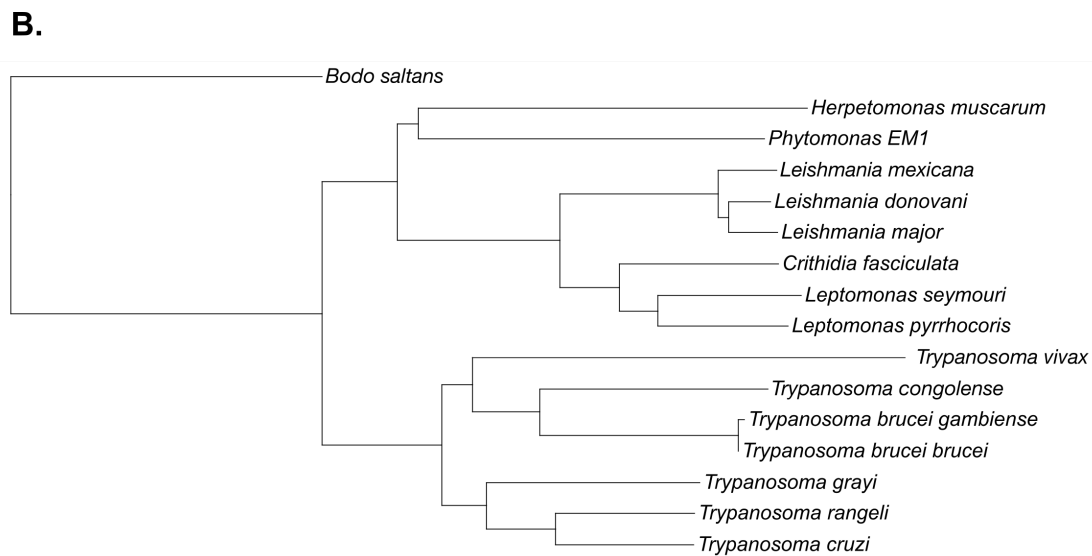


Figure 2.4B - Phylogenetic tree based on orthogroups containing a single gene from each trypanosomatid used in the analysis.

I went on to compare the gene content in *H. muscarum* to that of other trypanosomatids by comparing occurrence of specific groups genes of interest e.g. metabolic pathway genes. For this analysis the gene tables from *T. brucei* were used, as this genome is the best described. A summary of findings can be found Table S2.4 and discussion of some of the orthologues of interest, including ‘missing’ orthologues, follows.

Metabolism

H. muscarum appears to be missing the sphingolipid biosynthesis genes SLS1-4, which includes the inositol phosphorylceramide synthase and two choline phosphorylceramide synthases. Given that sphingolipids are thought to be essential to eukaryotic membranes (Sutterwala *et al.*, 2008), this was surprising. However *L. major* promastigotes do not require *de novo* sphingolipid synthesis - a mutant devoid of sphingolipids (SL) was shown to be viable (Zhang *et al.*, 2003). However, the SL-free mutant was unable to differentiate into a metacyclic stage *in vitro* and showed severe defects in vesicular trafficking. As such it is likely that, like *L. major*, *H. muscarum* must scavenge its sphingolipids from its environment. Furthermore, differentiation of *H. muscarum*, either *in vitro* or *in vivo*, was not observed during this work. The absence of this pathway may therefore be a requirement for trypanosomatid metacyclogenesis.

H. muscarum does not have orthologues for the carnitine O-acetyltransferase (CAT) (Tb927.11.2230) and L-threonine 3-dehydrogenase (Tb927.6.2790) genes of the acetate metabolism pathway. We were also unable to find an orthologue to these genes in the species from the Leishmanii clade used in the analysis. As such these genes may have been lost sometime after the group diverged from the *Trypanosoma*.

Additionally, three *T. brucei* respiratory chain genes do not appear to have orthologues in *H. muscarum*, including mitochondrial NADH-ubiquinone oxidoreductase flavoprotein 2 (Tb927.7.6350) which had orthologues in all species used in the analysis apart from *H. muscarum*. Similarly, the only genomes in the analysis without an orthologue for the cytochrome c oxidase assembly protein (Tb927.10.3120) were *H. muscarum* and *Phytomonas EMI*. Given the importance of these genes, this likely indicates an important gap in the *H. muscarum* annotation. Finally, no orthologue was identified for the *T. brucei* alternative oxidase (AOX) (Tb927.10.7090) which is found in *Trypanosoma* and is upregulated in bloodstream forms. This oxidase is thought to enhance organisms ability to cope with stress associated with temperature change, infections and oxidative stress (Vanlerberghe and McIntosh, 1997).

Additionally, I observed that for several *T. brucei* proteins, there are multiple *H. muscarum* orthologues. Two of the most extreme examples of this being AAT13 and membrane-bound phosphatase 2 (MBAP2) which have 38 and 18 orthologous genes in *H. muscarum* respectively. The increased copy number of these genes hints at their importance. Furthermore, both AAT13 and MBAP2 have been shown to be highly upregulated in *Leishmania* during sand flies (Martin *et al.*, 2014; Inbar *et al.*, 2017). These genes are associated with purine acquisition and endosomal trafficking may be indicative of key metabolic requirements of *H. muscarum*.

Differentiation

RNA-binding proteins (RBPs) have emerged as key modulators of gene expression in trypanosomatids and are especially interesting in the context of trypanosome development and differentiation (Kolev *et al.* 2014). Orthologues were found for 72/75 *T.*

brucei RNA-binding proteins (please see Table S2.4 for a summary). Furthermore, several of these proteins had multiple orthologues in *H. muscarum* including RBP10 (4 orthologues, Tb927.8.2780). The finding of 4 orthologues for RBP10 was surprising as this RBP is known to be highly expressed in bloodstream forms of *T. brucei*. Over expression of RBP10 in procyclics led to an increase of many bloodstream-form specific mRNAs as well as transcripts associated with sugar transport, the flagellum and cytoskeleton (Wurst et al., 2012). The role for this protein in *H. muscarum* is unclear as *H. muscarum* does not appear to have a bona-fide vertebrate host, however given this proteins links to sugar transport, it may play a more general role in metabolism in *H. muscarum*. RNA-binding proteins with no orthologues found in *H. muscarum* were: chromatin-remodelling-associated RRM2 (Tb927.6.2550, (Naguleswaran et al. 2015) or the pre-RNA processing protein RBSR1 (Tb927.9.6870, Wippel *et al.* 2019) and a hypothetical RBP (Tb927.10.14950). The apparent ‘completeness’ of *H. muscarum*’s repertoire of these key differentiation regulators is intriguing, given that I did not observe clear differentiation in *H. muscarum* in either *in vitro* assays or during *D. melanogaster* infections. It is likely these RBPs can play additional roles in the cell. Future comparisons of their expression levels/timings and interaction partners in *H. muscarum* and other trypanosomatids may reveal why *H. muscarum* do not appear to have differentiated forms.

Additionally, I was unable to find any orthologues for the hydrophilic acylated surface proteins (HASPs) or small hydrophilic endoplasmic reticulum-associated proteins (SHERPs) which are associated with metacyclogenesis in *Leishmania*. I also note that the repressor of differentiation kinase 1 (RDK1, Tb927.11.14070) has 6 orthologues in *H.*

muscarum. RDK1 acts with the PTP1/PIP39 phosphatase cascade to prevent uncontrolled differentiation from bloodstream to procyclic form (Jones *et al.*, 2014).

Surface proteins

No orthologues were found for the EP procyclins which are known to be expressed highly *T. brucei* procyclic whilst in the tsetse vectors and are thought to provide protection from the digestive enzymes in the insect midgut (Acosta-Serrano *et al.*, 2001; Haines *et al.*, 2010). As such *H. muscarum* likely relies on other surface proteins for protection in the insect midgut, this is discussed further in Chapter Four.

Additionally, Orthofinder was unable to find an orthologue to major surface proteins of salivary gland forms of *T. brucei* BARPs (bloodstream alanine-rich proteins). The BARPs are GPI-anchored proteins required for tsetse salivary gland colonisation (Urwyler *et al.* 2007, Fragoso *et al.* 2009). I was also unable to find orthologues for the *T. brucei* metacyclic invariant surface proteins (MISPs) which are found extending above the VSG coat in salivary gland metacyclic forms (Casas-Sánchez *et al.* 2018). The lack of these genes, which are crucial for tsetse salivary gland colonisation by trypanosomes, in *H. muscarum* may partially explain why I do not see *H. muscarum* invading the salivary glands of *D. melanogaster*.

Finally, the 13 *T. brucei* GP63 genes were grouped with 28 *H. muscarum* genes. GP63 is a major surface protease in *L. major* promastigotes. The comparatively high copy number of GP63 in *H. muscarum* may highlight its importance. Furthermore, GP63 has been implicated in *Leishmania* virulence (Pereira. *et al.* 2010), and as such these genes are of great interest for further study.

Nuclear proteome

Kinetochore interacting protein 3 (KKIP3, Tb927.10.6700) and SR protein (Tb927.9.6870) had no orthologues in *H. muscarum* or other species from the Leishmanii clade used in the analysis and as such they appear to be *Trypanosoma* specific. This is intriguing as, RNAi of KKIP3 in *T. brucei* resulted in defects in DNA segregation and reduced population growth (D'Archivio and Wickstead, 2017). Therefore, whilst this gene is not essential, it is clearly important for accurate replication. Additionally, *T. brucei*'s kinetochore interacting protein 1 (KKIP1), PHF5-like protein (Tb927.10.7390) and U1 small nuclear ribonucleoprotein 24 kDa (Tb927.3.1090) had orthologues in all species used in the analysis apart from *H. muscarum*. Similar to KKIP3, RNAi knock down of KKIP1 caused defects in DNA replication, though in the case of KKIP these defects were more severe – resulting in the loss of entire chromosomes (D'Archivio and Wickstead, 2017). It is unclear if these genes have been lost in *H. muscarum* or this indicates a gap in the current annotation. Based on the importance of KKIP1 and the fact these genes have orthologues in all other species analysed, it is likely to be the latter.

Finally, *H. muscarum* appears to have a 'full set' of the *T. brucei* RNA interference pathway genes including an orthologue for TbARGO1 (Tb927.10.10850). The presence of these genes means RNAi could be utilized in future studies with *H. muscarum* for gene expression manipulation. Genes from this well-conserved (in metazoans) pathway have been lost in several trypanosomatids including: *L. major*, *L. donovani* and *T. cruzi* (Robinson and Beverley 2003, DaRocha *et al.* 2004). It has been suggested that the RNAi pathway has been lost in these organisms due to viral perturbation (Beverley 2003, Lye *et al.* 2010). Given that the one of the roles of the RNAi pathway is to protect the cell from

viral attack, it is unsurprising that viruses with RNAi inhibitors have evolved (e.g. Sullivan *et al.* 2005) – though this has not been explicitly demonstrated for *Leishmania* RNA viruses (LRVs). There is also likely selection pressure to maintain populations of *Leishmania* carrying LRVs, as many have been shown to have increased rates of survival (and importantly increased virulence) during vertebrate infections (Hartley *et al.* 2012). It would be interesting to look for evidence of viruses akin to the LRVs in *H. muscarum* to test the link between RNAi and virus infection in trypanosomatids.

The presence of a functional RNAi pathway has also been linked to transposon activity in *Leishmania* – with RNA-negative species lacking active transposable elements (TEs), and RNAi competent *L. braziliensis* harbouring several classes of active TEs (Peacock *et al.* 2007, Lye *et al.* 2010). Given this, it is possible that the loss/lack of active TEs in *L. major* and *L. donovani* have lifted the requirement of the RNAi pathway to protect against TE-associated genomic perturbations. I do observe expression of transcripts corresponding to the telomere associated mobile elements (TATEs) in all samples of *H. muscarum* assayed in Chapter Four. This observation would lend support to the above hypothesis.

Transfer ribonucleic acids (tRNAs)

ARAGON v1.2.36 (Laslett and Canback, 2004) was able to annotate 173 tRNAs, which represent 50 different isoacceptors and all 20 standard amino acids, in the *H. muscarum* genome (Table 2.4). Three tRNAs were identified which ARAGON was unable to identify a unique amino acid for. Two have the anticodon ATCA and as such could decode asparagine (ATC) or selenocysteine (TCA). The other anticodon, CCCG, may encode arginine (CCC) or glycine (CCG). No suppressor tRNAs, which are mutated

tRNAs which allow the translation of stop codons into amino acids rather than signalling translation termination, were found. *H. muscarum* appears to have more encoded tRNAs than were found in *T. brucei* (Tan *et al.*, 2002), where just 50 tRNAs (representing 40 isoacceptors) were described. As in *T. brucei* (Rubio *et al.*, 2013) the only tRNA for the tyrosine GTA isoacceptor contains an intron in *H. muscarum*. This sole tRNA must decode all tyrosine residues. Two further intron-containing tRNAs were identified by tRNA-ScanSE2 (Lowe and Chan, 2016), these decode isoleucine with the anticodon GAT.

Amino acid	Anticodon Counts						Total tRNAs
Ala	AGC: 6	GGC: 8	CGC: 1	TGC: 2			17
Arg	ACG: 9	GCG: 1	CCG: 1	TCG: 2	CCT: 2	TCT: 2	17
Asn	ATT: 0	GTT: 4					4
Asp	ATC: 0	GTC: 9					9
Cys	ACA: 0	GCA: 2					2
Gln			CTG: 4	TTG: 2			6
Glu			CTC: 11	TTC: 2			13
Gly	ACC: 0	GCC: 8	CCC: 2	TCC: 2			12
His	ATG: 1	GTG: 3					4
Ile	AAT: 7	GAT: 0	CAT: 0	TAT: 2			9
Leu	AAG: 4	GAG: 0	CAG: 3	TAG: 2	CAA: 1	TAA: 1	11
Lys			CTT: 7	TTT: 1			8
Met/iMet			CAT: 6				6
Phe	AAA: 0	GAA: 3					3
Pro	AGG: 3	GGG: 0	CGG: 4	TGG: 1			8
SelCys				TCA: 4			4
Ser	AGA: 2	GGA: 0	CGA: 2	TGA: 2	ACT: 0	GCT: 5	11
Supressor		CTA: 0	TTA: 0	TCA: 0			0
Thr	AGT: 3	GGT: 0	CGT: 4	TGT: 2			9
Trp			CCA: 4				4
Tyr	ATA: 0	GTA: 1					1
Val	AAC: 6	GAC: 1	CAC: 3	TAC: 2			12
							170

Table 2.4 – Transfer RNAs and their codons identified in *Herpetomonas muscarum* by ARAGORN (Laslett and Canback 2002) and tRNA-ScanSE2 (Lowe and Chan 2016).

Trypanosomatid signatures

Spliced leader sequence

To locate the splice leader sequence (SLS) for *H. muscarum*, the conserved 39bp SL sequence from *Phytomonas serpens* (L42381.1) was 'BLASTed' against the *H. muscarum* scaffolds. This gave 259 hits over 24 scaffolds. I used these hits to identify 19 clusters of mini-exon gene repeats containing 3-43 copies of the mini-exon gene (Table S2.5). The first 111bp of the gene are common to all copies of the mini-exon gene and contain a 40bp splice leader sequence and what I predict to be the intron. The splice leader sequence (1-40bp) and the putative intronic region (41-111bp) were then aligned with mini-exon sequences of several other trypanosomatids in the Leishmaniinae clade and 9 other *Herpetomonas* isolates. Whilst the splice leader sequence is well-conserved across the clade (Table 2.5), I observe variability in the A/T-rich region between bases 11-19bp which appears genus specific, with the exception of the *Herpetomonas* sequences. *H. rotimani* and *H. nabiculae* have identical sequence across the 11-19bp region. However, the *H. muscarum* and *H. nabiculae* differ from each other, and the other *Herpetomonas* sequences over this variable region. It has been recorded that trypanosomatids can have multiple SL sequences within the same genome, as such this may explain the differences in the *Herpetomonas* SLs.

Additionally, compared to other trypanosomatids, the *Herpetomonas* sequences have an 'additional' adenosine between bases 10 and 11. The intronic region from *H. muscarum* shows high similarity to that of previously reported *Herpetomonas sp.* sequences. The first 15bp of the intronic sequence appear to be conserved in other species from the Leishmaniinae clade, however the sequence becomes more variable thereafter in both in terms of base content and length.

Species	Accession #	Splice leader sequence (bases 1-40)
<i>Herpetomonas muscarum</i>	(Wang, Sloan, Ligoxygakis 2019)	AACTAACGCT AAAAATTGTT TACAGTTTCTGTACATTATTG
<i>Herpetomonas muscarum</i>	EU095980-4.1*, EU095979.1*, EU095983.1	AACTAACGCT AAAAATTGTT TACAGTTTCTGTACTATATTG
<i>Herpetomonas</i> sp. <i>TCC263</i>	EU095976.1	AACTAAAGCA TTATATATAGAT ACAGTTTCTGTACTATATTG
<i>Herpetomonas</i> sp. <i>TCC263</i>	EU095977.1	AACTAAAGCA TTATATATAGAT ACAGTTTCTGTACTATATTG
<i>Herpetomonas roitmani</i>	EU095978.1	AACTAAAGCA TTATATATAGAT ACAGTTTCTGTACTTTATTG
<i>Herpetomonas nabiculae</i>	KF054153.1	AACTAACGCT AT-TATTGTT TACAGTTTCTGTACTTTATTG
<i>Phytomonas EM1</i>	X87138.1	AACTAACGCT- ATTCTAGAT ACAGTTTCTGTACTTTATTG
<i>Phytomonas serpens</i>	L42381.1, L42378.1, L42377.1, L42382.1, L42376.1	AACTAACGCT- ATTCTAGAT ACAGTTTCTGTACTTTATTG
<i>Phytomonas</i> sp. <i>Mar8</i>	AF250993.1	AACTAACGCT- ATTCTAGAT ACAGTTTCTGTACTTTATTG
<i>Phytomonas</i> sp. <i>Alp1</i>	AF250967.1	AACTAACGCT- ATTCTAGAT ACAGTTTCTGTACTTTATTG
<i>Leishmania braziliensis</i>	MG010484.1	AACTAACGCT- ATATAAGTAT ACAGTTTCTGTACTTTATTG
<i>Leishmania tarentolae</i>	AY100201.1	AACTAACGCT- ATATAAGTAT ACAGTTTCTGTACTTTATTG
<i>Leishmania hoogstraali</i>	AY100197.1, AY100200.1	AACTAACGCT- ATATAAGTAT ACAGTTTCTGTACTTTATTG
<i>Leishmania gymnodactyli</i>	AY100195.1, AY100196.1	AACTAACGCT- ATATAAGTAT ACAGTTTCTGTACTTTATTG
<i>Leishmania adleri</i>	AY100199.1, AY100194.1	AACTAACGCT- ATATAAGTAT ACAGTTTCTGTACTTTATTG
<i>Leishmania major</i>	XR_002460055.1	AACTAACGCT- ATATAAGTAT ACAGTTTCTGTACTTTATTG
<i>Leishmania mexicana</i>	Agami and Shapira 1992	AACTAACGCT- ATATAAGTAT ACAGTTTCTGTACTTTATTG
<i>Leishmania donovani</i>	CP022617.1	AACTAACGCT- ATATAAGTAT ACAGTTTCTGTACTTTATTG
<i>Leishmania infantum</i>	AF097653.1	AACTAACGCT- ATATAAGTAT ACAGTTTCTGTACTTTATTG
<i>Blastocrithidia culicis</i>	DQ860204.1	AACTAACGCT- ATATTGTT TACAGTTTCTGTACTATATTG
<i>Blastocrithidia culicis</i>	DQ860203.1	AACTAACGCT- ATATTGTT TACAGTTTCTGTACTTTATTG

Table 2.5 – Alignment of base pairs 1-40bp of the splice leader sequence of trypanosomatid

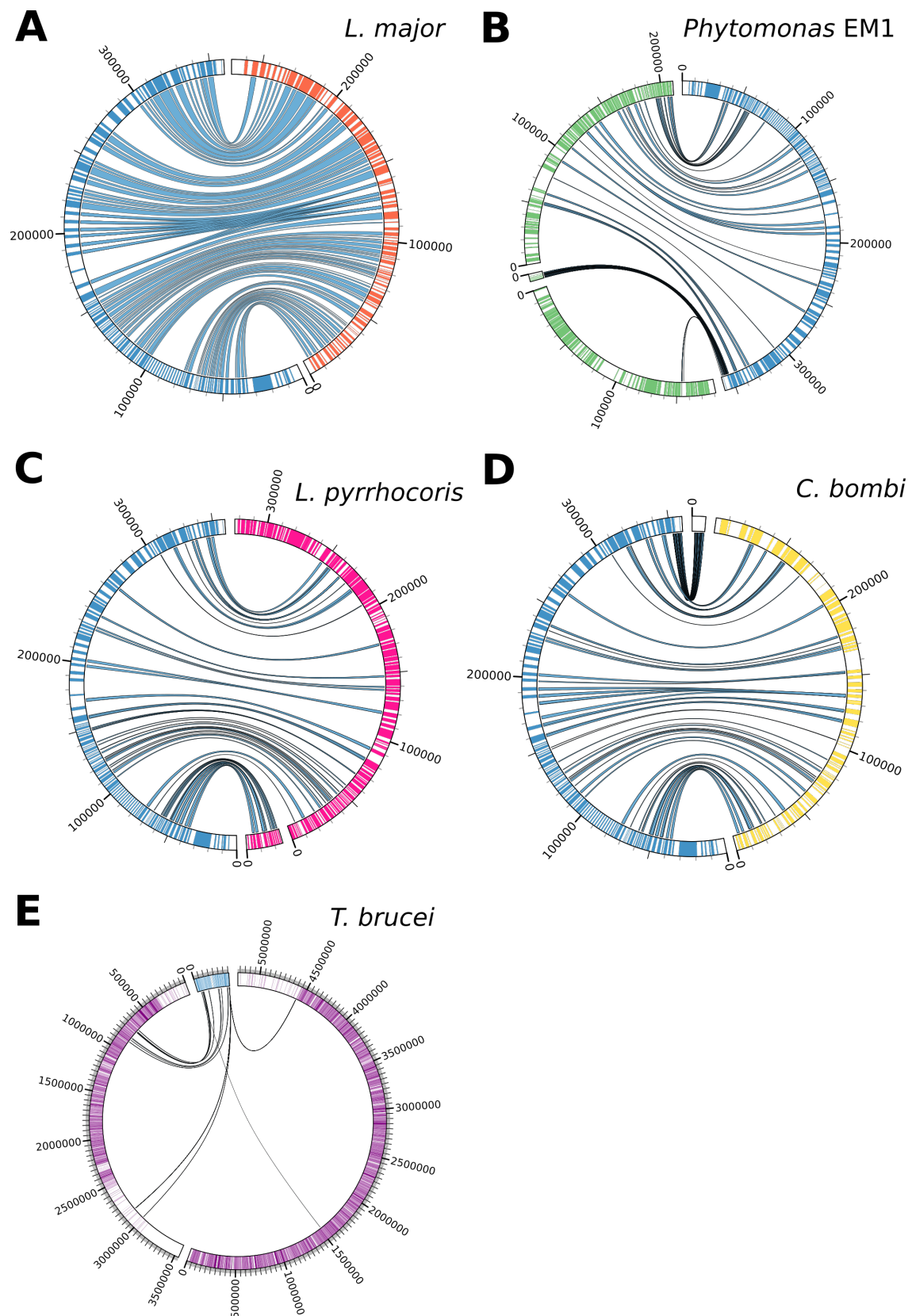
Synteny

Despite diverging before the existence of mammals (El-Sayed, Myler, Blandin, *et al.*, 2005), trypanosomatids show high gene order conservation across the genome. *H. muscarum* is no exception as it shows strong synteny with a trypanosomatids across the group (Figure 2.5A-E). The synteny between *H. muscarum* and *L. major* was particularly striking, with high levels of synteny observed across all chromosomes/scaffolds (Figure S2.3). This high-level similarity between the *H. muscarum* and *L. major* provides support to the validity of using the monoxenous trypanosomes to model their dioxenous, disease-causing relatives.

Tubulin Loci

The architecture of the tubulin arrays have been described in a number of trypanosomatids (Jackson, Vaughan and Gull, 2006), with two mutually exclusive formats being defined – monotypic and alternating. Monotypic tubulin arrays consist of either α -tubulin **or** β -tubulin. Alternating arrays contain both α -tubulin **and** β -tubulin genes which alternate along the array.

The *H. muscarum* orthologues of *Trypanosoma brucei* α - and β -tubulin genes were found using the orthologue tables generated by Orthofinder and used to locate the tubulin arrays. I identified three genomic loci containing *H. muscarum* tubulin genes. Two of these loci consist of alternating arrays (Figure 2.6A) and the third locus consists of four copies of β -tubulin (Figure 2.6B). The alternating arrays are consistent with previous findings (reported as *Herpetomonas megaseliae* in Jackson et al. 2006). However, the presence of a monotypic beta tubulin array appears to contrast the established alternating or monotypic model.



Figures 2.5 - *Herpetomonas muscarum* scaffolds show synteny with other trypanosomatids. For the full legend see the following page.

Figures 2.5 - *Herpetomonas muscarum* scaffolds show synteny with other trypanosomatids. As an example, these plots shows co-linearity between *H. muscarum* genes (genes highlighted in blue) on scaffold 40 and: **A.** *L. major* chromosome 1 (genes highlighted in red). **B.** *Phytomonas EMI* scaffolds HF955082, HF955140 and HF955140 (genes highlighted in green) **C.** *Leptomonas pyrrhocoris* scaffolds LpyrH10_33 and LpyrH10_41 (genes highlighted in pink) **D.** *Crithidia bombi* scaffolds (genes highlighted in yellow) OESO01000125 and OESO01000148 **E.** *Trypanosoma brucei* chromosomes 9 and 11 (genes highlighted in purple). Scaffold/Chromosome labels show length in bp. This data was produced using Promer alignments (Delcher et al. 2002). Ribbons between scaffolds show windows of >100 amino acid (translated) align with at least 50% identity. This data was visualised using Circos (Krzywinski et al. 2009).

The genes surrounding the monotypic β -tubulin locus are (mostly) syntenic with regions of *T. brucei* chromosome 4 and chromosome 8 of *L. major* (gene numbers Tb927.5.970 - Tb927.927.5.3090 and Lmj.08.1090 - Lmj.08.11140). This region of *L. major* chromosome 8 is the one of two singleton β -tubulin loci in the species. As such the tubulin configuration of *H. muscarum* appears to be intermediate between the tubulin array configurations of *T. brucei* and *L. major*.

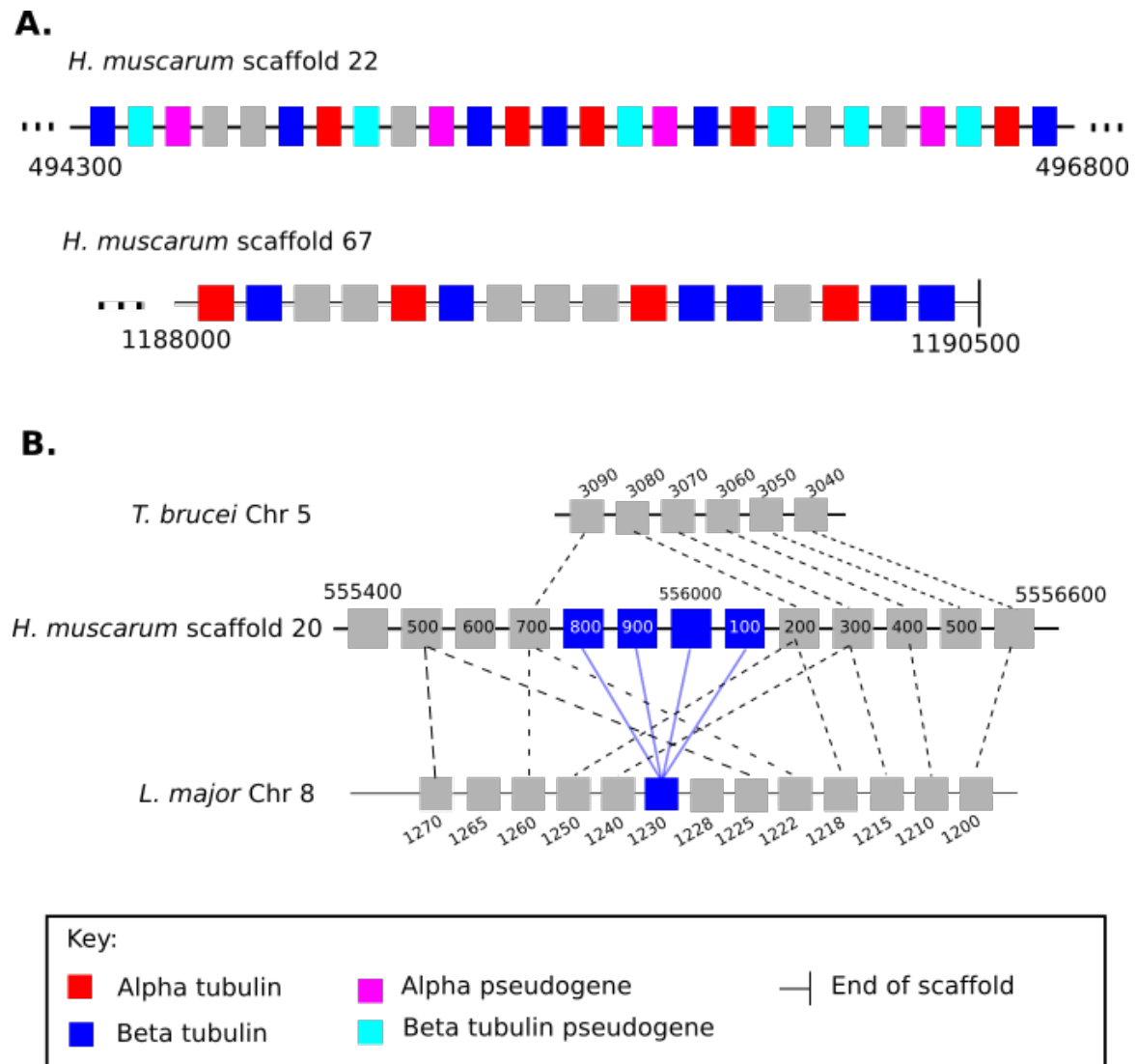


Figure 2.6 – Tubulin arrays of *Herpetomonas muscarum*. A - Alternating arrays. Scaffolds 22 and 67 have two loci containing alternating α - (red) and β - (blue) tubulin genes. Several of these genes we predict to be pseudogenes (α -tubulin - pink, β -tubulin - light blue) as they contain tubulin domains but also contain sequence consistent with non-LTR transposons. **B - A monotypic β -tubulin locus.** Four copies of a β -tubulin (blue) were found in tandem on scaffold 20. This locus is syntenic to the β -tubulin locus on *L. major* chromosome 8. This region also shares some synteny with a locus on *T. brucei* on chromosome 5, however the β -tubulin gene is absent. Dotted lines indicate orthologous genes. Blue lines indicate orthologous β -tubulin genes.

Concluding Remarks

This chapter has described the first whole genome sequence for a *Herpetomonas* species using a hybrid Illumina/SMRT approach. The 41Mb genome is currently resolved to 264 scaffolds, though optical mapping indicates there to be 37-38 chromosomes. Of the current scaffolds eight appear to represent complete chromosomes. Future sequencing of closely related genomes will allow greater chromosome resolution. Comparisons of the *H. muscarum* genome to those of disease-causing trypanosomatids revealed strong similarities in genome structure and content. This was most striking with the *Leishmania* species with which it shares approximately 80% of protein orthogroups. Given these similarities, there is likely much that can be learned from *H. muscarum*, and other monoxenous species, about the disease-causing trypanosomatids. For example, conservation of the *Leishmania* chromosome 31 supernumeracy appears to be mimicked in *H. muscarum* and as such may be in other monoxenous trypanosomatids. With increasing numbers of monoxenous trypanosomatid genomes becoming available the source of the supernumeracy could be discovered.

Several important differences were also evident between the *H. muscarum* genome and the genomes of disease causing trypanosomatids. *H. muscarum* lacks several genes associated with differentiation, including the SL-synthesis genes and metacyclic surface proteins, HASPs and SHERPs. This suggests that *H. muscarum* may completely lack the ability to differentiate. This is supported by ‘wet lab’ experiments which were unable to induce differentiation in *H. muscarum* promastigotes *in vitro*. Additionally, I was unable to find orthologues to several genes which are described for other trypanosomatids. Some of these absences are likely to reflect ‘gaps’ in the current annotation e.g. the case of KKIP3. However, others, such as for the AOX gene, may highlight genuine gene losses

in this species. This work also revealed large expansions in several gene families in *H. muscarum*, including: the GP63, MBAP2 and amino acid transporters. Trypanosomatids are known to expand the copy number of important genes, as such these observations provide insight into the ‘life style’ of *H. muscarum*.

Finally, the genome sequence is an important tool for the study of *H. muscarum* and its interactions with natural host *Drosophila melanogaster*. It has provided a blueprint allowing genetic manipulation (see Chapters 3 and 5) and transcriptomic analyses (see Chapter 4).

References

- Acosta-Serrano, A. *et al.* (2001) ‘The surface coat of procyclic *Trypanosoma brucei*: Programmed expression and proteolytic cleavage of procyclin in the tsetse fly’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 98(4), pp. 1513–1518. doi: 10.1073/pnas.041611698.
- Agabian, N. (1990) ‘*Trans* splicing of nuclear pre-mRNAs’, *Cell*. Elsevier, 61(7), pp. 1157–1160. doi: 10.1016/0092-8674(90)90674-4.
- Akopyants, N. S. *et al.* (2009) ‘Demonstration of Genetic Exchange During Cyclical Development of *Leishmania* in the Sand Fly Vector’, *Science*, 324(5924), p. 265 LP-268. doi: 10.1126/science.1169464.
- Andino, R. *et al.* (1999) ‘Intracellular determinants of picornavirus replication’, *Trends in Microbiology*. Elsevier, 7(2), pp. 76–82. doi: 10.1016/S0966-842X(98)01446-2.
- Benz, C. *et al.* (2005) ‘Messenger RNA processing sites in *Trypanosoma brucei*’, *Molecular and Biochemical Parasitology*, 143(2), pp. 125–134. doi: <https://doi.org/10.1016/j.molbiopara.2005.05.008>.
- Berriman, M. *et al.* (2005) ‘The Genome of the African Trypanosome *Trypanosoma brucei*’, *Science*, 309(5733), p. 416 LP-422. doi: 10.1126/science.1112642.
- Beverley, S.M., 2003. Protozomics: trypanosomatid parasite genetics comes of age. *Nature Reviews Genetics*, 4(1), p.11
- Blyn, L. B. *et al.* (1997) ‘Requirement of poly(rC) binding protein 2 for translation of poliovirus RNA’, *Journal of virology*, 71(8), pp. 6243–6246.

- Borghesan, T. C. *et al.* (2013) 'Molecular Phylogenetic Redefinition of *Herpetomonas* (Kinetoplastea, Trypanosomatidae), a Genus of Insect Parasites Associated with Flies', *Protist*. Urban & Fischer, 164(1), pp. 129–152. doi: 10.1016/J.PROTIS.2012.06.001.
- Casas-Sánchez, A., Perally, S., Ramaswamy, R., Haines, L. R., Rose, C., Yunta, C., Aguilera-Flores, M., Lehane, M. J., Almeida, I. C., Boulanger, M. J. and Acosta-Serrano, A. (2018) The crystal structure and localization of *Trypanosoma brucei* invariant surface glycoproteins suggest a more permissive VSG coat in the tsetse-transmitted metacyclic stage. *bioRxiv* 477737.
- Chen, J. *et al.* (1995) 'The topology of the kinetoplast DNA network', *Cell*. Cell Press, 80(1), pp. 61–69. doi: 10.1016/0092-8674(95)90451-4.
- Chin, C.-S. *et al.* (2013) 'Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data', *Nature Methods*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 10, p. 563.
- Collier, B. *et al.* (1998) 'Translational Inhibition *in Vitro* of Human Papillomavirus Type 16 L2 mRNA Mediated through Interaction with Heterogenous Ribonucleoprotein K and Poly(rC)-binding Proteins 1 and 2 ', *Journal of Biological Chemistry* , 273(35), pp. 22648–22656. doi: 10.1074/jbc.273.35.22648.
- D'Archivio, S. and Wickstead, B. (2017) 'Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes', *The Journal of Cell Biology*, 216(2), p. 379 LP-391. doi: 10.1083/jcb.201608043.
- DaRocha, W.D., Otsu, K. and Teixeira, S.M., 2004. Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracyclineinducible T7 promoter system in *Trypanosome cruzi*. *Mol Biochem Parasitol*, 133(2), pp.175-186
- Downing, T. *et al.* (2011) 'Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance', *Genome research*. Cold Spring Harbor Laboratory Press, 21(12), pp. 2143–2156. doi: 10.1101/gr.123430.111.
- Dubessay, P. *et al.* (2002) 'The switch region on *Leishmania major* chromosome 1 is not required for mitotic stability or gene expression, but appears to be essential', *Nucleic acids research*. Oxford University Press, 30(17), pp. 3692–3697.
- Eid, J. *et al.* (2009) 'Real-Time DNA Sequencing from Single Polymerase Molecules', *Science*, 323(5910), p. 133 LP-138. doi: 10.1126/science.1162986.
- El-Sayed, N. M., Myler, P. J., Blandin, G., *et al.* (2005) 'Comparative Genomics of Trypanosomatid Parasitic Protozoa', *Science*, 309(5733), p. 404 LP-409.
- El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., *et al.* (2005) 'The Genome Sequence of *Trypanosoma cruzi*, Etiologic Agent of Chagas Disease', *Science*, 309(5733), p. 409 LP-415. doi: 10.1126/science.1112631.
- Ellis, M. *et al.* (2002) 'Processing and Trafficking of *Leishmania Mexicana* GP63: Analysis using GPI8 Mutants Deficient in Glycosylphosphatidylinositol Protein

- Anchoring', *Journal of Biological Chemistry*, 277(31), pp. 27968–27974. doi: 10.1074/jbc.M202047200.
- Emms, D. M. and Kelly, S. (2015) 'OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy', *Genome Biology*, 16(1), p. 157. doi: 10.1186/s13059-015-0721-2.
- Flegontov, P. *et al.* (2016) 'Genome of *Leptomonas pyrrhocoris*: A high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*', *Scientific Reports*, 6. doi: 10.1038/srep23704.
- Frith, M. C. *et al.* (2008) 'Discovering Sequence Motifs with Arbitrary Insertions and Deletions', *PLOS Computational Biology*. Public Library of Science, 4(5), p. e1000071.
- Ganot, P. *et al.* (2004) 'Spliced-leader RNA *trans* splicing in a chordate, *Oikopleura dioica*, with a compact genome', *Molecular and cellular biology*. American Society for Microbiology, 24(17), pp. 7795–7805. doi: 10.1128/MCB.24.17.7795-7805.2004.
- Haines, L. R. *et al.* (2010) 'Tsetse EP protein protects the fly midgut from trypanosome establishment', *PLoS pathogens*. Public Library of Science, 6(3), pp. e1000793–e1000793. doi: 10.1371/journal.ppat.1000793.
- Hartley, M-A, Ronet, C., Zangger, H., Beverley, S. and Fasel, N. (2012) *Leishmania* RNA virus: when the host pays the toll. *Frontiers in Cellular and Infection Microbiology*, 99.
- Hunt, M. *et al.* (2013) 'REAPR: a universal tool for genome assembly evaluation', *Genome Biology*, 14(5), p. R47. doi: 10.1186/gb-2013-14-5-r47.
- Inbar, E. *et al.* (2017) 'The Transcriptome of *Leishmania major* Developmental Stages in Their Natural Sand Fly Vector', *mBio*, 8(2). doi: 10.1128/mBio.00029-17.
- Ivens, A. C. *et al.* (2005) 'The Genome of the Kinetoplastid Parasite, *Leishmania major*', *Science*, 309(5733), p. 436 LP-442. doi: 10.1126/science.1112680.
- Jackson, A. P., Vaughan, S. and Gull, K. (2006) 'Evolution of Tubulin Gene Arrays in Trypanosomatid parasites: genomic restructuring in *Leishmania*', *BMC Genomics*. London: BioMed Central, 7, p. 261. doi: 10.1186/1471-2164-7-261.
- Jones, N. G. *et al.* (2014) 'Regulators of *Trypanosoma brucei* cell cycle progression and differentiation identified using a kinome-wide RNAi screen', *PLoS pathogens*. Public Library of Science, 10(1), pp. e1003886–e1003886. doi: 10.1371/journal.ppat.1003886.
- Kolev, N. G., Ullu, E. and Tschudi, C. (2014) The emerging role of RNA-binding proteins in the life cycle of *Trypanosoma brucei*. *Cellular microbiology* 16(4): 482–489.
- Korf, I. (2004) 'Gene finding in novel genomes', *BMC bioinformatics*. BioMed Central, 5, p. 59. doi: 10.1186/1471-2105-5-59.

- Laslett, D. and Canback, B. (2004) 'ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences', *Nucleic acids research*. Oxford University Press, 32(1), pp. 11–16. doi: 10.1093/nar/gkh152.
- LeBowitz, J. H. *et al.* (1993) 'Coupling of poly(A) site selection and *trans*-splicing in *Leishmania*.', *Genes & Development*, 7(6), pp. 996–1007. doi: 10.1101/gad.7.6.996.
- Li, L., Stoeckert Jr, C. J. and Roos, D. S. (2003) 'OrthoMCL: identification of ortholog groups for eukaryotic genomes', *Genome research*. Cold Spring Harbor Laboratory Press, 13(9), pp. 2178–2189. doi: 10.1101/gr.1224503.
- Lidie, K. B. and Van Dolah, F. M. (2007) 'Spliced Leader RNA-Mediated *trans*-Splicing in a Dinoflagellate, *Karenia brevis*', *Journal of Eukaryotic Microbiology*. John Wiley & Sons, Ltd (10.1111), 54(5), pp. 427–435. doi: 10.1111/j.1550-7408.2007.00282.x.
- Liu, B. *et al.* (2005) 'Fellowship of the rings: the replication of kinetoplast DNA', *Trends in parasitology*. Elsevier, 21(8), pp. 363–369.
- Lowe, T. M. and Chan, P. P. (2016) 'tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes', *Nucleic Acids Research*. Oxford University Press, 44(Web Server issue), pp. W54–W57. doi: 10.1093/nar/gkw413.
- Lukeš, J. *et al.* (2002) 'Kinetoplast DNA Network: Evolution of an Improbable Structure', *Eukaryotic Cell*, 1(4), pp. 495–502. doi: 10.1128/EC.1.4.495-502.2002.
- Mair, G. *et al.* (2000) 'A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA', *RNA (New York, N.Y.)*, 6(2), pp. 163–169.
- Marçais, G. and Kingsford, C. (2011) 'A fast, lock-free approach for efficient parallel counting of occurrences of k-mers', *Bioinformatics (Oxford, England)*. 2011/01/07. Oxford University Press, 27(6), pp. 764–770. doi: 10.1093/bioinformatics/btr011.
- Martin, J. L. *et al.* (2014) 'Metabolic reprogramming during purine stress in the protozoan pathogen *Leishmania donovani*', *PLoS pathogens*. Public Library of Science, 10(2), p. e1003938.
- Martínez-Calvillo, S. *et al.* (2003) 'Transcription of *Leishmania major* Friedlin Chromosome 1 Initiates in Both Directions within a Single Region', *Molecular Cell*. Elsevier, 11(5), pp. 1291–1299. doi: 10.1016/S1097-2765(03)00143-6.
- Maslov, D. A. *et al.* (2013) 'Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed', *Trends in Parasitology*. Elsevier, 29(1), pp. 43–52. doi: 10.1016/j.pt.2012.11.001.
- Matthews, K. R., Tschudi, C. and Ullu, E. (1994) 'A common pyrimidine-rich motif governs *trans*-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes.', *Genes & Development*, 8(4), pp. 491–501. doi: 10.1101/gad.8.4.491.

- Motta, M. C. M. *et al.* (2013) 'Predicting the Proteins of *Angomonas deanei*, *Strigomonas culicis* and Their Respective Endosymbionts Reveals New Aspects of the Trypanosomatidae Family', PLOS ONE. Public Library of Science, 8(4), p. e60209.
- Moyzis, R. K. *et al.* (1988) 'A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes', Proceedings of the National Academy of Sciences, 85(18), p. 6622 LP-6626. doi: 10.1073/pnas.85.18.6622.
- Naguleswaran, A., Gunasekera, K., Schimanski, B., Heller, M., Hemphill, A., Ochsenreiter, T. and Roditi, I. (2015) *Trypanosoma brucei* RRM1 Is a Nuclear RNA-Binding Protein and Modulator of Chromatin Structure. mBio 6(2): e00114-15.
- Nawrocki, E. P. and Eddy, S. R. (2013) 'Infernal 1.1: 100-fold faster RNA homology searches', Bioinformatics (Oxford, England). 2013/09/04. Oxford University Press, 29(22), pp. 2933–2935. doi: 10.1093/bioinformatics/btt509.
- Ostareck, D. H. *et al.* (2001) 'Lipoxygenase mRNA Silencing in Erythroid Differentiation: The 3'UTR Regulatory Complex Controls 60S Ribosomal Subunit Joining', Cell. Elsevier, 104(2), pp. 281–290. doi:10.1016/S0092-8674(01)00212-4.
- Otto, T. D. *et al.* (2010) 'Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology', Bioinformatics (Oxford, England). 2010/06/18. Oxford University Press, 26(14), pp. 1704–1707. doi: 10.1093/bioinformatics/btq269.
- Otto, T. D. *et al.* (2011) 'RATT: Rapid Annotation Transfer Tool', Nucleic acids research. 2011/02/08. Oxford University Press, 39(9), pp. e57–e57. doi: 10.1093/nar/gkq1268.
- Peacock, C. S. *et al.* (2007) 'Comparative genomic analysis of three *Leishmania* species that cause diverse human disease', Nature genetics. 2007/06/17, 39(7), pp. 839–847. doi: 10.1038/ng2053.
- Pita, S. *et al.* (2019) 'The Tritryps Comparative Repeatome: Insights on Repetitive Element Evolution in Trypanosomatid Pathogens', Genome biology and evolution. Oxford University Press, 11(2), pp. 546–551. doi: 10.1093/gbe/evz017.
- Puechberty, J. *et al.* (2007) 'Compared genomics of the strand switch region of *Leishmania* chromosome 1 reveal a novel genus-specific gene and conserved structural features and sequence motifs', BMC genomics. BioMed Central, 8, p. 57. doi: 10.1186/1471-2164-8-57.
- Real, F. *et al.* (2013) 'The genome sequence of *Leishmania (Leishmania) amazonensis*: functional annotation and extended analysis of gene models', DNA research : an international journal for rapid publication of reports on genes and genomes. 2013/07/15. Oxford University Press, 20(6), pp. 567–581. doi: 10.1093/dnares/dst031.
- Robinson, K.A. and Beverley, S.M., 2003. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. Molecular and biochemical parasitology, 128(2), pp.217-228.

- Rogers, M. B. *et al.* (2011) ‘Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*’, *Genome research*. Cold Spring Harbor Laboratory Press, 21(12), pp. 2129–2142. doi: 10.1101/gr.122945.111.
- Rubio, M. A. T. *et al.* (2013) ‘Unusual Noncanonical Intron Editing Is Important for tRNA Splicing in *Trypanosoma brucei*’, *Molecular Cell*. Elsevier, 52(2), pp. 184–192. doi: 10.1016/j.molcel.2013.08.042.
- Runckel, C., DeRisi, J. and Flenniken, M. L. (2014) ‘A Draft Genome of the Honey Bee Trypanosomatid Parasite *Crithidia Mellificae*’, *PLoS ONE*, 9(4). doi: 10.1371/journal.pone.0095057.
- Schmid-Hempel, P. *et al.* (2018) ‘The genomes of *Crithidia bombi* and *C. expoeki*, common parasites of bumblebees’, *PLoS ONE*, 13(1). doi: 10.1371/journal.pone.0189738.
- Shammas, M. A. (2011) ‘Repetitive sequences, genomic instability and Barrett’s esophageal adenocarcinoma’, *Mobile genetic elements*. 2011/09/01. Landes Bioscience, 1(3), pp. 208–212. doi: 10.4161/mge.1.3.17456.
- da Silva, M. S. *et al.* (2010) ‘The *Leishmania amazonensis* TRF (TTAGGG repeat-binding factor) homologue binds and co-localizes with telomeres’, *BMC microbiology*. BioMed Central, 10, p. 136. doi: 10.1186/1471-2180-10-136.
- Smit, A. F. A., Hubley, R., Green, P. (2017) Repeat Masker v1.0.9, 2017.
- Stanke, M. *et al.* (2006) ‘Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources’, *BMC bioinformatics*. BioMed Central, 7, p. 62. doi: 10.1186/1471-2105-7-62.
- Steinbiss, S. *et al.* (2016) ‘Companion: a web server for annotation and analysis of parasite genomes’, *Nucleic Acids Research*. Oxford University Press, 44(Web Server issue), pp. W29–W34. doi: 10.1093/nar/gkw292.
- Sutterwala, S. S. *et al.* (2008) ‘Developmentally Regulated Sphingolipid Synthesis in African Trypanosomes’, *Molecular microbiology*, 70(2), pp. 281–296. doi: 10.1111/j.1365-2958.2008.06393.x.
- Tan, T. H. P. *et al.* (2002) ‘tRNAs in *Trypanosoma brucei*: Genomic Organization, Expression, and Mitochondrial Import’, *Molecular and Cellular Biology*. American Society for Microbiology, 22(11), pp. 3707–3716. doi: 10.1128/MCB.22.11.3707-3716.2002.
- Teixeira, S. M. *et al.* (2012) ‘Trypanosomatid comparative genomics: Contributions to the study of parasite biology and different parasitic diseases’, *Genetics and Molecular Biology*. Ribeirão Preto, SP, Brazil: Sociedade Brasileira de Genética, 35(1), pp. 1–17.
- Timothy L. Bailey and Charles Elkan (1994) ‘Fitting a mixture model by expectation maximization to discover motifs in biopolymers’, in *Proceedings of the Second*

International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, California, pp. 28–36.

- Ullu, E., Matthews, K. R. and Tschudi, C. (1993) ‘Temporal order of RNA-processing reactions in trypanosomes: rapid *trans* splicing precedes polyadenylation of newly synthesized tubulin transcripts’, *Molecular and cellular biology*, 13(1), pp. 720–725.
- Vandenberghe, A. E., Meedel, T. H. and Hastings, K. E. (2001) ‘mRNA 5’-leader trans-splicing in the chordates’, *Genes & development*. Cold Spring Harbor Laboratory Press, 15(3), pp. 294–303. doi: 10.1101/gad.865401.
- Vanlerberghe, G. C. and McIntosh, L. (1997) ‘Alternative Oxidase: From Gene to Function’, *Annual Review of Plant Physiology and Plant Molecular Biology*. Annual Reviews, 48(1), pp. 703–734. doi: 10.1146/annurev.arplant.48.1.703.
- Vurture, G. W. *et al.* (2017) ‘GenomeScope: fast reference-free genome profiling from short reads’, *Bioinformatics*, 33(14), pp. 2202–2204.
- Wang, L., Sloan, M. A. and Ligoxygakis, P. (2019) ‘Intestinal NF- κ B and STAT signalling is important for uptake and clearance in a *Drosophila-Herpetomonas* interaction model’, *PLOS Genetics*. Public Library of Science, 15(3), p. e1007931.
- Wippel, H. H., Malgarin, J. S., Martins, S. de T., Vidal, N. M., Marcon, B. H., Miot, H. T., Marchini, F. K., Goldenberg, S. and Alves, L. R. (2019) The Nuclear RNA-binding Protein RBSR1 Interactome in *Trypanosoma cruzi*. *Journal of Eukaryotic Microbiology*. John Wiley & Sons, Ltd (10.1111) 66(2): 244–253.
- Wurst, M., Seliger, B., Jha, B. A., Klein, C., Queiroz, R., Clayton, C. Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in *Trypanosoma brucei*. *Mol Microbiol*. 2012; 83:1048–1063.
- Zhang, H. *et al.* (2007) ‘Spliced leader RNA *trans*-splicing in dinoflagellates’, *Proceedings of the National Academy of Sciences of the United States of America*. 2007/03/02. National Academy of Sciences, 104(11), pp. 4618–4623. doi: 10.1073/pnas.0700258104.
- Zhang, K. *et al.* (2003) ‘Sphingolipids are essential for differentiation but not growth in *Leishmania*’, *The EMBO Journal*. Oxford, UK: Oxford University Press, 22(22), pp. 6016–6026. doi: 10.1093/emboj/cdg584.

CHAPTER THREE

The infection of *Drosophila melanogaster* with
Herpetomonas muscarum

CHAPTER THREE:

The infection of *Drosophila melanogaster* with

Herpetomonas muscarum

Herpetomonas muscarum was isolated by Prof. Petros Ligoxygakis. Data presented herein includes figures published in Wang, Sloan and Ligoxygakis 2019. Larval and repeat infections were performed with James Harvey, an undergraduate student at the University of Oxford. All other work is my own unless stated.

Preface

It has long been known that *Drosophila* harbour kinetoplastids (Chatton, E. and Alilaire, 1908), with numerous studies showing there is a diverse array of trypanosomatids present in wild insect populations (Votýpka *et al.*, 2010; Maslov *et al.*, 2013; Yurchenko *et al.*, 2016). Ebbert *et al.* (2001) reported a trypanosomatid prevalence of up to 17%, in eight species of *Drosophila* surveyed from five sites in Ohio, USA (Ebbert, Burkholder and Marlowe, 2001). During the two-year study, the group also noted that prevalence was dependant on the *Drosophila* species, rather than location or sex. More recently, discovery of these trypanosomatid-fly interactions has moved away from pain-staking dissection, isolation and study of morphology. DNA sequencing technologies allow rapid and far reaching screening of wild *Drosophila* populations for their resident trypanosomatids. Wilfert *et al.* (2011) used PCR-based screening to study the diversity of trypanosomatids from *Drosophila* populations in Greece, Portugal and the United Kingdom (Wilfert *et al.*, 2011). The group used primers complementary to the conserved region of trypanosomatid splice leader sequences. All five species of *Drosophila*

collected were found to harbour trypanosomatids, though fewer than 5% of the flies in each species harboured trypanosomatids. Of the flies found to harbour trypanosomatids, ~30% were found to harbour multiple trypanosomatid species. The study was unable to provide detailed phylogenetic information about the trypanosomatids present. However, the group were able to amplify trypanosomatid GAPDH (glyceraldehyde-3-phosphate dehydrogenase) genes from three samples from UK populations of *D. melanogaster*, *Drosophila tristis* and *Drosophila obscura*. These sequences were closest to those of *Herpetomonas muscarum* in *D. melanogaster* samples and to *Blastocrithidia culicis* in *D. tristis* and *D. obscura*.

Herpetomonas

Members of *Herpetomonas*, a monophyletic genus containing 13 species, are mainly found in dipterans. The global dispersal of *Herpetomonas* is broad, with species described in Europe, the Americas, Asia and Africa (Tyc *et al.*, 2013). *Herpetomonas* species have also been isolated from biting midges and cockroaches (Podlipaev *et al.*, 2004). Despite being defined as monoxenous and found mostly within insects, *Herpetomonas* have also been isolated from plants (Fiorini *et al.*, 2001; Barclay McGhee and Postell, 2007; Marín *et al.*, 2007) and appear to be able to infect vertebrates (Pacheco *et al.*, 1998; Chicharro and Alvar, 2003; Morio *et al.*, 2008). However, vertebrate infections appear to be opportunistic and only in immunocompromised individuals. In 2013, the genera was phylogenetically re-defined; based on the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and small subunit rRNA sequences of over 400 *Herpetomonas* isolates (Borghesan *et al.*, 2013). The study included the re-assignment of *H. megaseliae* to *H. muscarum* and the addition of two species previously identified as *Leptomonas* into the genus.

Drosophila-trypanosomatid interactions

The value of using *Drosophila* species to study insect-trypanosomatids is beginning to be appreciated. This body of work was pioneered by Rowton and McGhee who described interaction between *Herpetomonas ampelophilae* and *Drosophila melanogaster* which had been isolated in Clarke County, Georgia, USA (Rowton and Barclay McGhee, 1978). *Herpetomonas ampelophilae* cell were identified in the midgut, both endotrophically and peritrophically, as well as the Malpighian tubules of up to 30% of wild flies sampled. Prevalence in laboratory-housed flies ranged from 5-100% of flies (Rowton and Barclay McGhee, 1983). The trypanosomatids were able to be transmitted between adults, presumably via the food after deposition in faeces/vomit from infected individuals. Infection prevalence was greatly reduced by deviation in temperature from 25°C or the presence of food mites in the *Drosophila* food (Rowton and Barclay McGhee, 1983). This highlights the delicate balance of these interactions.

More recent studies of *Drosophila*-trypanosomatid interactions were based on ‘non-natural’ pairings. For example, Boulanger *et al.* (2001) analysed the immune responses in *Drosophila melanogaster* when challenged with two species of *Crithidia* – members of this genus naturally infect mosquitoes and bees (Boulanger *et al.*, 2001; Runckel, DeRisi and Flenniken, 2014). The group showed that *per os* feeding of *Drosophila* with *Crithidia* caused the expression of several immune factors, including antimicrobial peptides (AMPs) and *Drosophila* immune induced molecules (DIMs), in the insect fat body. Injection of the trypanosomatids into the insect hemolymph resulted in comparatively lower AMP expression and, depending on the *Crithidia* species, could be lethal for the fly.

Hamilton *et al.* (2016) were the first to characterise a ‘natural’ *Drosophila*-trypanosomatid infection using *Drosophila falleni* and *Jaenimonas drosophilae*. The trypanosomatid, *J. drosophilae*, was discovered in laboratory fly stocks routinely used for culturing the nematode *Howardula aoronymphium* (a parasite of *Drosophila*). Phylogenetic analysis indicated the closest species to *J. drosophilae* are from the genera *Herpetomonas* and *Stringomonas*. The group showed that after feeding *D. falleni* larvae with *J. drosophilae* from *in vitro* culture, the trypanosomatids were able to persist throughout development of the fly and that infection led to a reduction in host fecundity. Co-housing experiments showed that infected adults could transmit the trypanosomatid to uninfected flies with approximately 44% efficiency. The trypanosomatid *J. drosophilae* appears to have a wide host range as it was able to successfully infect both *D. neotestacea* and *D. melanogaster* - though infectivity was decreased in *D. melanogaster* which is comparatively more distant to *D. falleni* than *D. neotestacea*. As such these fly-trypanosomatid interactions are strongly impacted by the genotypes of fly and trypanosomatid. This is the case for several other described insect-trypanosomatid interactions, including the aforementioned *Drosophila*-*Crithidia* study (Boulanger *et al.*, 2001) and for infections of bumble bee *Bombus terrestris* by *Crithidia bombi*. (Sadd and Barribeau, 2013). Finally, as observed with *Crithidia fasciculata* (Boulanger *et al.*, 2001), when the group injected the trypanosomatid into *D. melanogaster* haemolymph, this resulted in fly death. This result was shown to be independent of Imd and Toll immune responses as no genes from these pathways were upregulated and pathway mutants were not more susceptible to infection. The biological significance of this result is not yet clear as it is not known if *J. drosophilae* are naturally able to enter the insect haemolymph.

This chapter describes the characterisation of the infection of *Drosophila melanogaster* by *Herpetomonas muscarum*. Whilst the presence of trypanosomatids in *Drosophila* has long been known, little is known about the distribution of specific trypanosomatid species in these populations, including *H. muscarum*. As such, I also sought to describe the distribution of *H. muscarum* in *Drosophila* populations across the USA by using publicly available sequencing data.

Spatiotemporal distribution of Herpetomonas muscarum in wild Drosophila melanogaster populations

Herpetomonas muscarum used in this study was isolated from *Drosophila melanogaster* in Oxfordshire, UK. Of the flies sampled, 5% of males and 6% of females were found to harbour *H. muscarum*. Beyond this the distribution of *H. muscarum* in *Drosophila* populations is unknown. To further investigate the distribution of *H. muscarum* in wild *Drosophila* populations I made use of publicly available DNA sequencing data from wild *Drosophila melanogaster* populations. Data is available for several sites around the United States and Canada, with sites sampled up to three times a year between 2012-2015. I sought to look for evidence of *H. muscarum* in these populations and to investigate the trypanosomatids spatiotemporal distribution. To do this, I mapped the DNA sequencing reads to the *D. melanogaster* and *H. muscarum* genomes concatenated together, then later extracted the reads which mapped to *H. muscarum* genome. On average 78% of reads mapped to the concatenated *D. melanogaster* and *H. muscarum* genomes (range 69-89%) (Table S3.1). The proportions of these reads mapping to the *Herpetomonas muscarum* genome were low ($< 0.04\%$). Given each DNA sample is extracted from a pool of 50 flies, it is unclear if this indicates low prevalence across all flies, or that only a small proportion of flies harbour the trypanosomatid.

When manually inspecting where in the *H. muscarum* genome these reads were mapped to, I found the majority mapped to the ribosomal DNA (rDNA) loci.

Given the conservation of the rDNA sequences it was possible that these reads were not *H. muscarum* specific but may have also mapped to other trypanosomatids, even other eukaryotes. To test this, rDNA mapped reads were isolated and then re-mapped to a broad panel of eukaryotic 18S rDNA sequences from: *Drosophila* associated yeasts (*Hanseniaspora occidentalis*, *Hanseniaspora uvarum*, *Saccharomyces paradoxus*, *Brettanomyces naardenensis* and *Debaryomyces hansenii* (Hoang, Kopp and Chandler, 2015)), kinetoplastids (*Euglena gracilis*, *Diplonema papillatum*), *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Dictyostelium discoideum*, *Plasmodium falciparum* and finally several other trypanosomatid species (with representatives from genera *Trypanosoma*, *Leishmania*, *Leptomonas*, *Phytomonas*, *Herpetomonas* and *Crithidia*). From this, reads were found to map to many of the trypanosomatids - but not the other eukaryotes. Therefore, these reads are not specific to *H. muscarum*, however they do appear to come from a trypanosomatid.

It is unlikely these trypanosomatids are *Trypanosoma brucei* as they are their natural vectors, tsetse, are restricted to sub-Saharan Africa. Furthermore, the trypanosomatids are unlikely to be from *Leishmania* genus, as they are unable to establish infection in *Drosophila* (unpublished observations, R. Dillon). However, there is significant overlap between the habitat in many insects which harbour trypanosomatids - including the sand flies, triatomine bugs, cockroaches and midges. This would provide opportunities for trypanosomatids to be passed between insects through food contamination, albeit briefly if the infections cannot be fully established. Additionally, the presence of multiple

trypanosomatid species within in the same fly has been reported in other studies (Wilfert *et al.* 2011). As such, further studies would be required to resolve the trypanosomatid prevalence to a species level.

By comparison of the proportions of *H. muscarum* mapped reads between samples at different locations and times I investigated the distribution of trypanosomatids in these wild *Drosophila* populations. There were no statistically significant differences in occurrence of trypanosomatid rDNA in these samples when comparing by month (Figure 3.1A, Kruskal-Wallis $p=0.32$), year (Figure 3.1B, Kruskal-Wallis $p=0.54$) or US state (Figure 3.1C, Kruskal-Wallis $p=0.0562$). However, the sample size for these analyses was small and there were differences in prevalence between states - e.g. samples taken from Massachusetts tended to have fewer trypanosomatid mapped reads. There does appear to be a general trend with more *H. muscarum* mapped reads found in samples from southern/mid-USA compared to those in the north-east of the country – though many states have no available data at present. This work supports earlier observations of the widespread occurrence of trypanosomatids in *Drosophila* populations and this methodology could be applied to other insect populations where there are large amounts of DNA-sequencing data available from field populations.

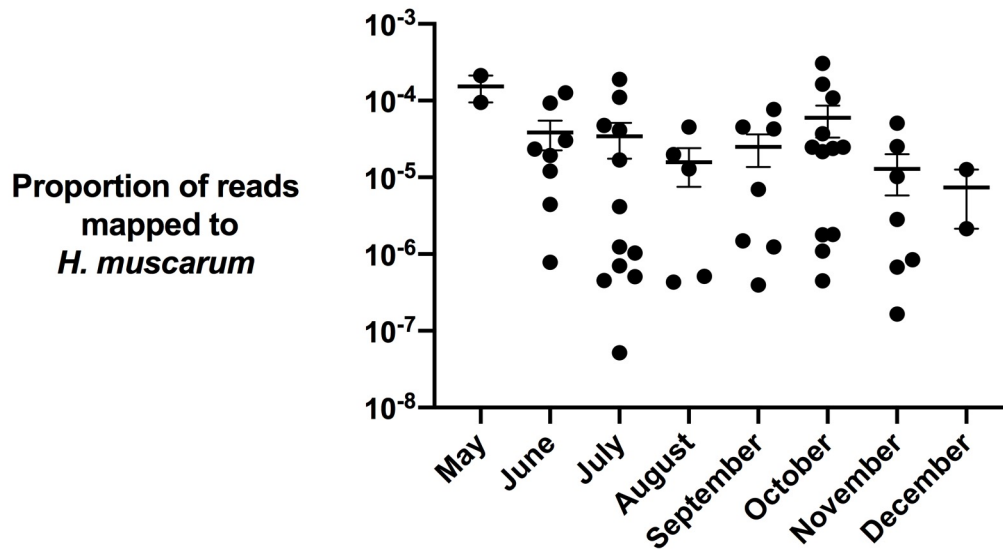


Figure 3.1A - Trypanosomatid incidence in wild *Drosophila melanogaster* populations in the USA and Canada by month – based on *Herpetomonas muscarum* mapped DNA-seq reads. Proportion of DNA-seq reads per sample which were mapped to the *H. muscarum* genome from publicly available data collected 2012-2015. Error bars show the standard error of the mean.

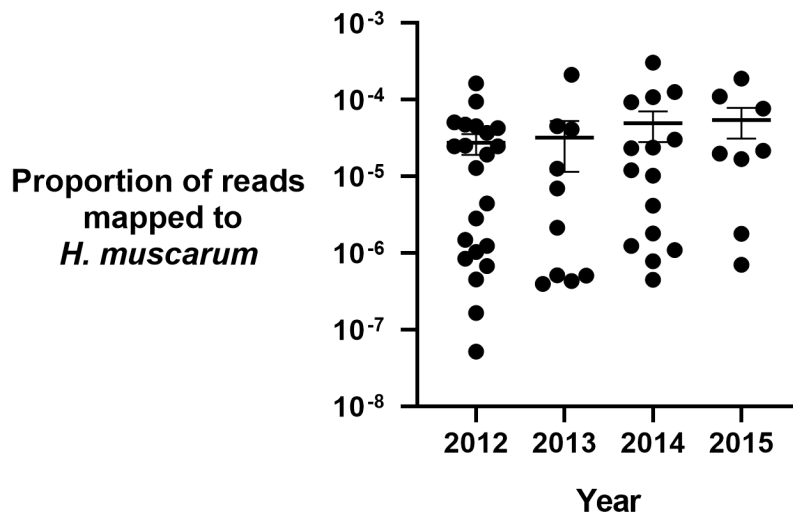


Figure 3.1B - Trypanosomatid incidence in wild *Drosophila melanogaster* populations in the USA and Canada by year – based on *Herpetomonas muscarum* mapped DNA-seq reads. Proportion of DNA-seq reads per sample which were mapped to the *H. muscarum* genome from publicly available data collected 2012-2015. Error bars show the standard error of the mean.

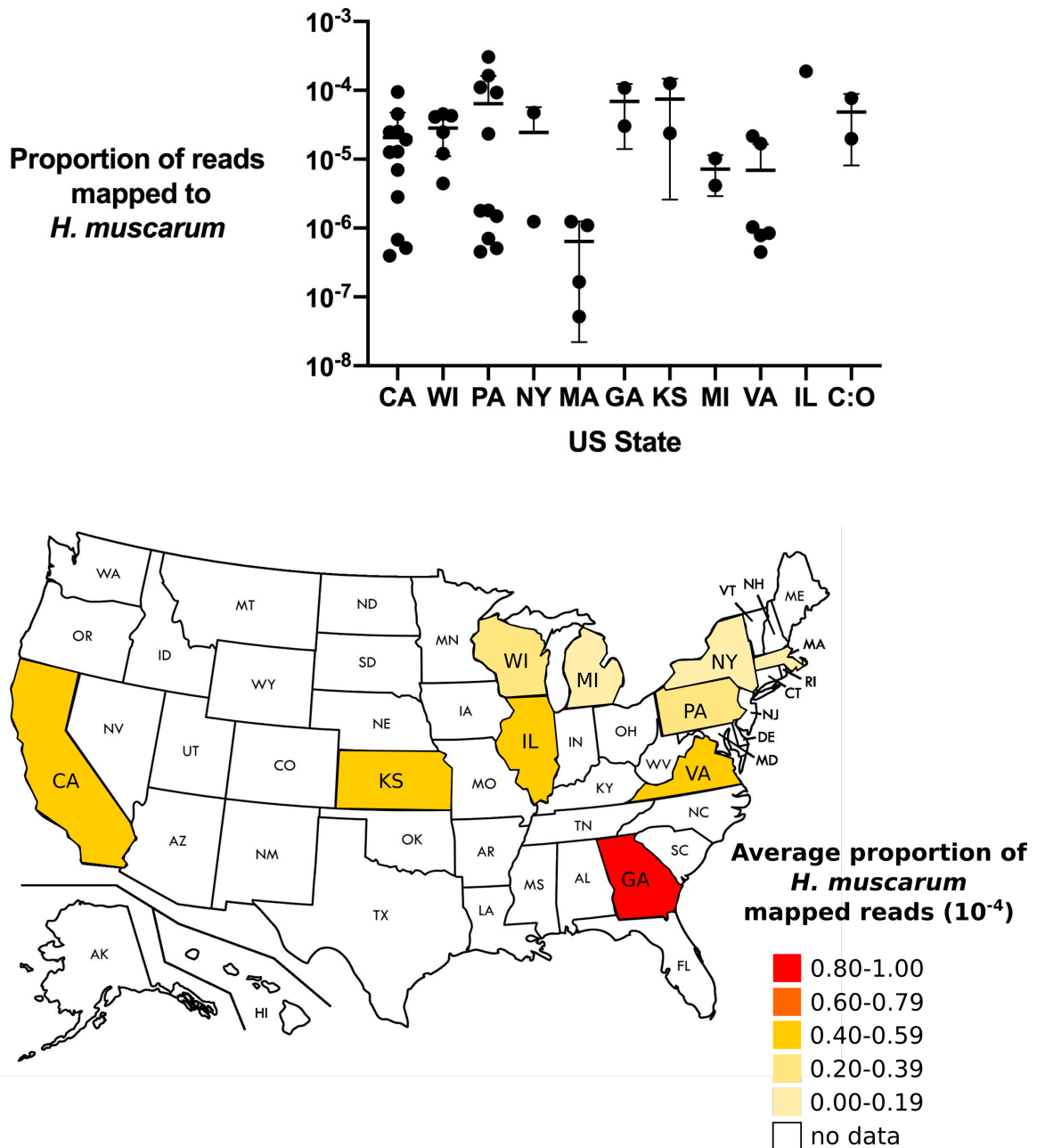


Figure 3.1C. Trypanosomatid incidence in wild *Drosophila melanogaster* populations in the USA and Canada by year - based on *Herpetomonas muscarum* mapped DNA-seq reads. Top: Proportion of DNA-seq reads per sample which were mapped to the *H. muscarum* genome from publicly available data collected 2012-2015. Error bars show the standard error of the mean. CA – California, WI – Wisconsin, PA – Pennsylvania, NY – New York, MA – Massachusetts, GA – Georgia, KS – Kansas, MI – Michigan, VA – Virginia, IL – Illinois and C:O – Ontario, Canada. **Bottom:** Map showing average proportion of *H. muscarum* mapped reads (10^{-4}) per state.

Results and Discussion

Morphology of Herpetomonas muscarum

Little was known about the isolated *H. muscarum* cell line prior to this work, as such I first sought to describe the general morphology of the cells. This greatly aided the recognition and ‘tracking’ of *H. muscarum* in the *D. melanogaster* midgut. Light and transmission electron microscopy revealed *H. muscarum* cells during *in vitro* culture share the typical trypanosomatid promastigote morphology (Wheeler, Gluenz and Gull, 2013) (Figure 3.2). The cell body is encased by the plasma membrane with a corset of subpellicular microtubules and cells had one, or two, flagella connected to the main cell body via the flagella pocket at the anterior of the cell. The majority of cells have a single flagellum with a well-developed paraflagellar rod during routine *in vitro* culture. Cells with two flagella, are likely undergoing cell replication.

Per os infection of Drosophila melanogaster adults

The following experiments were performed by feeding suspensions of log-phase *H. muscarum* culture to week-old adult *D. melanogaster* via filter paper discs. Adults were starved overnight, and sucrose was added to the *H. muscarum* suspension to encourage feeding. After feeding flies were returned to conventional fly food for the duration of experiments. Methods are described in detail in Chapter Seven.

Duration of infection

The number of *Herpetomonas muscarum* present in the flies after feeding was assayed quantitative PCR (qPCR) using genomic DNA extracts from whole flies as a template. The qPCR was performed using primers designed to amplify 162bp of the *H. muscarum*

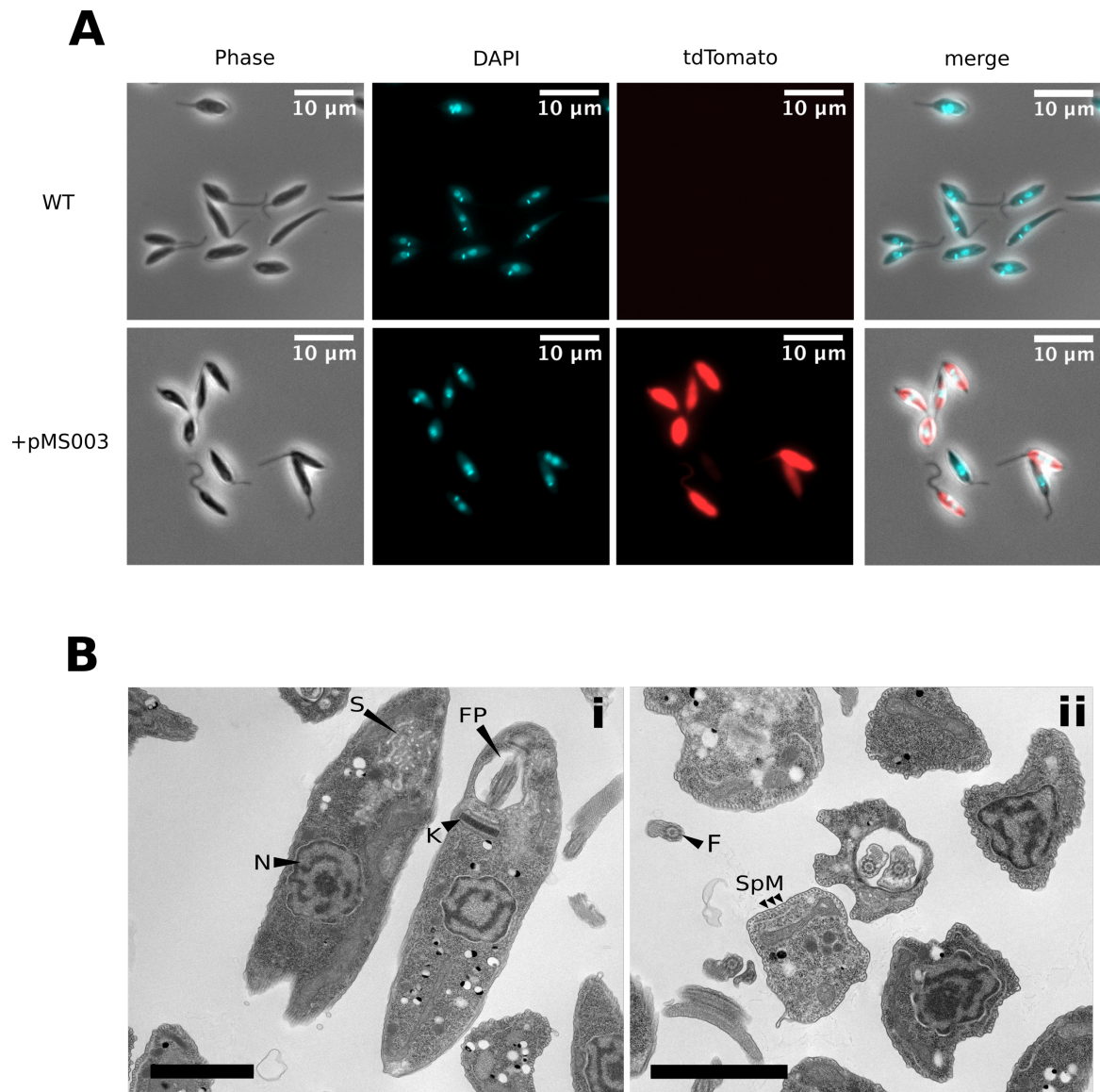


Figure 3.2 – Morphology of *H. muscarum* cells in axenic culture. A. Light microscopy of wild type *H. muscarum* and a tdTomato expressing cell line. Cells were fixed and stained with the DNA-binding dye DAPI – see Chapter Seven for details of method. **B. Transmission electron microscopy images of *Herpetomonas muscarum* *in vitro* cultured cells.** **i** – Two *H. muscarum* cell bodies in transverse section, **ii** - Cross-sections of several *H. muscarum* cells. N – nucleus, K – kinetoplast, FP – flagella pocket, F – flagellum, S – spongiome, SpM – subpellicular microtubules.

orthologue of the *Leishmania* paraflagellar rod protein 1D (HMUS01147600, for primer sequences see Chapter Seven). A ‘control’ experiment was performed in parallel for each of the three biological repeats, where flies were fed culture medium only, supplemented with sucrose. After two hours of feeding with *H. muscarum* culture, 100% of females had ingested the trypanosomatids. After initial ingestion, the numbers of *H. muscarum* per fly dramatically declined after 24 hours, though *H. muscarum* DNA was still detectable in flies up to 96 hours, after feeding (Figure 3.3). However, only in one experiment were *H. muscarum* still detectable at 96 hours. The majority of *H. muscarum* were ‘cleared’ at 72 hours or earlier as reported previously (Wang, Sloan and Ligoxygakis, 2019). As such, whilst infections longer than 2-3 days are possible, the majority of flies are able to clear the trypanosomatids quickly.

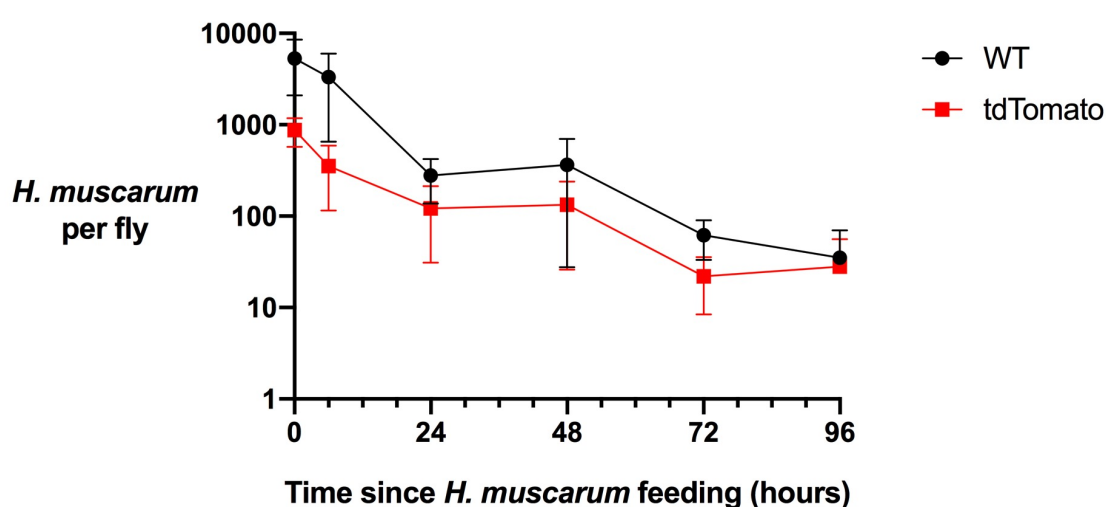


Figure 3.3 – *Herpetomonas muscarum* load in *Drosophila melanogaster* after two hours feeding. The average number of *H. muscarum* per fly for three biological repeat infections and error bars show the standard error of the mean. WT – wild type, tdTomato – *H. muscarum* expressing the tdTomato fluorescence protein.

Localisation of *Herpetomonas muscarum* in the *Drosophila melanogaster* gut

To determine where the *H. muscarum* cells go to after ingestion by the fly, I examined the digestive tract of flies which were fed *H. muscarum* either stained with the fluorescent dye Mitotracker™, or expressing the fluorescent tdTomato protein. For details about the production of tdTomato expressing line see Chapter Five.

Infections using Mitotracker™ stained *H. muscarum* showed that the trypanosomatids are found in the midgut of the fly, closely associated with the gut epithelium six hours after ingestion (Figure S3.1). This appeared to be similar to observations of *Leishmania* binding to the sand fly midgut to avoid being voided with the remains of the blood meal. A caveat to these observations is that pre-staining *H. muscarum* with Mitotracker™ dye significantly impacted on the dynamics of infection (Kruskal-Wallis, $p < 0.0001$). The *H. muscarum* ‘load’ per fly at 6 hours post feeding, was shown to be inversely proportional to the percentage of stained cells used for feeding (Figure 3.4). Unstained infections were significantly different to those where 10%, 50% or 100% of the *H. muscarum* were stained (Dunn’s multiple comparisons test (DMC), $p < 0.0175$) but there was no significant difference between the three infections where Mitotracker™ was used (DMC, $p > 0.5633$).

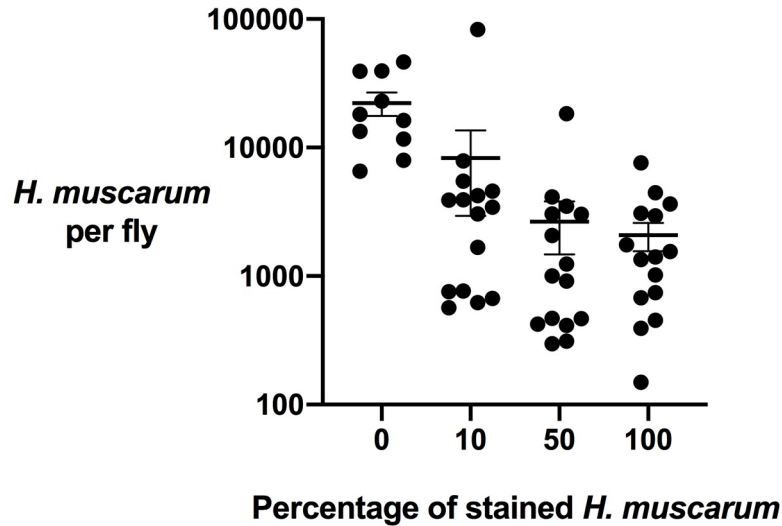


Figure 3.4 – Pre-staining *Herpetomonas muscarum* before infection reduced initial infection load. Data points represent the *H. muscarum* load in individual flies at 6 hours after feeding. Bars show the mean and standard error of the mean.

Given the apparent impact of the mitotracker dye on the dynamics of infection, I repeated these infections using a *H. muscarum* line which expressing the fluorescent protein – tdTomato (Figure 3.2). The production of this transgenic line is detailed in Chapter Five. The *H. muscarum* growth rate was unaffected by tdTomato expression during *in vitro* culture and infections using the tdTomato cells line closely resembled the wild type infections (see Chapter Five). Immediately following feeding, *H. muscarum* cells were found in the crop and midgut of the flies in large numbers (Figure 3.5A).

At six hours post feeding there were still many trypanosomatids in the crop of the flies – an understudied organ analogous to the mammalian stomach. The *Drosophila* crop is comprised of a single layer of epithelial cells surrounded cuticle which is highly folded – allowing substantial expansion to accommodate larger volumes of food (Dimitriadis VK,

1992). Food can then be pushed through to the midgut where the majority of digestion is thought to occur. The trypanosomatids were no longer dispersed throughout the crop but appeared restricted to a smaller area close to the crop wall (Figure 3.5B and C). The majority of *H. muscarum* cells in these tightly packed clusters are rounded – a phenotype commonly associated with stress in trypanosomatids. During imaging of live (dissected, unfixed) gut preparations these cells have beating flagella and as such appear to be alive. In contrast to immediately after feeding, there were very few cells visible in the midgut at six hours, often in small groups of cells (Figure 3.5B). These groups of cells were found both endo- and ecto-peritrophically in the anterior and middle midgut.

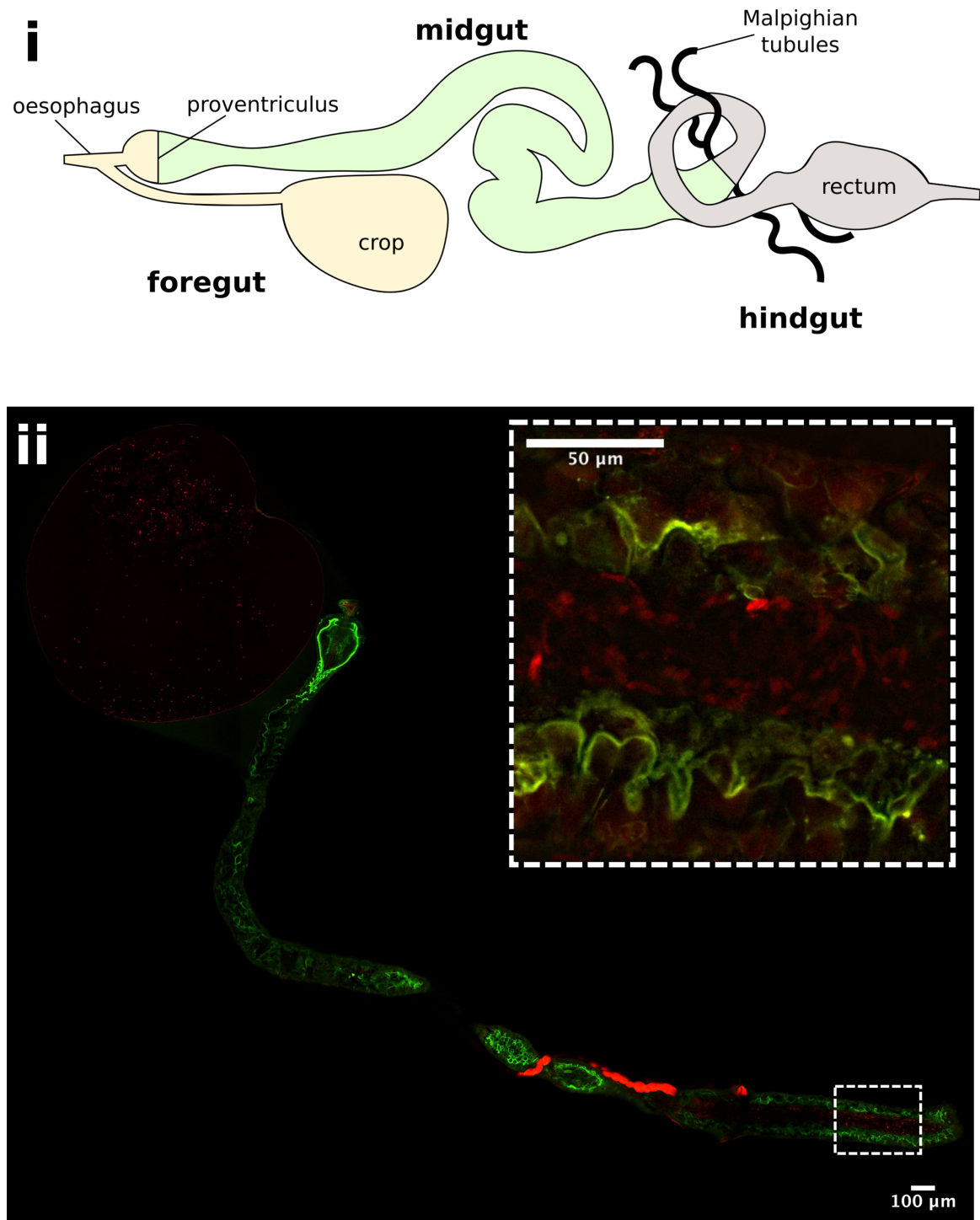


Figure 3.5A – *Herpetomonas muscarum* in the fly digestive tract immediately after feeding. i - Schematic of the *D. melanogaster* digestive tract. ii – The foregut and midgut of *D. melanogaster* two hours after feeding with tdTomato expressing *H. muscarum*. The flies used for this work express a myosin-GFP fusion protein to allow the gut epithelial border to be visualised. *H. muscarum* can be seen in the crop and the midgut (inset) of the fly.

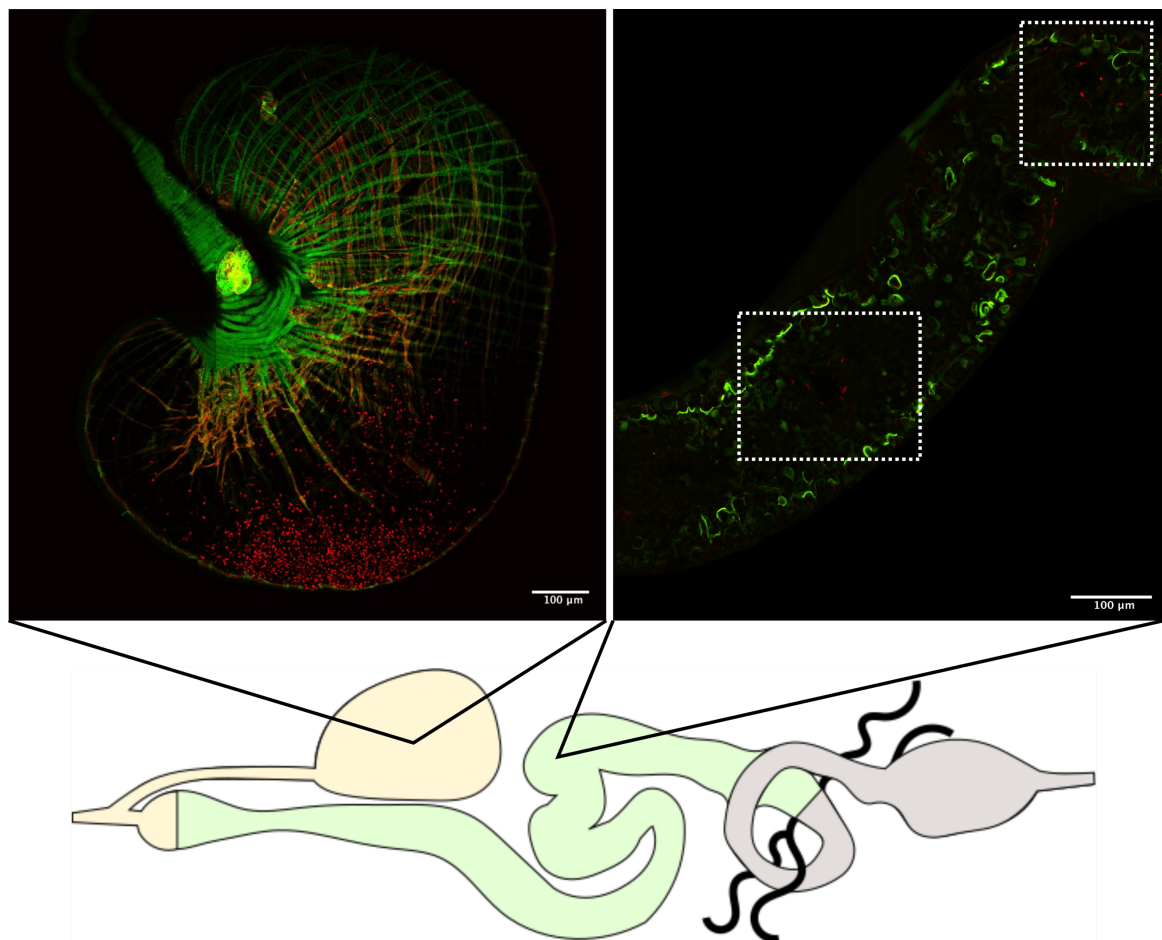


Figure 3.5B - *Herpetomonas muscarum* are found in the fly crop and midgut 6 hours after feeding. *H. muscarum* cells expressing tdTomato were fed to *D. melanogaster* expressing a myosin-GFP fusion protein to allow the gut epithelial border to be visualised. *H. muscarum* can be seen mainly in the crop, clustered, close to the epithelial layer. Small numbers of *H. muscarum* cells can be seen in the midgut (highlighted by the white-dashed boxes).

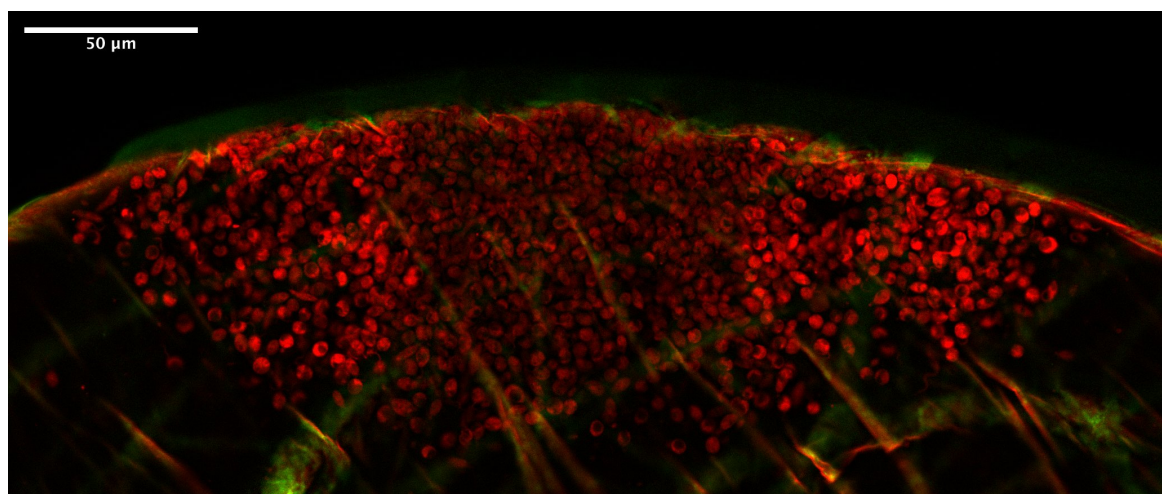


Figure 3.5C - *Herpetomonas muscarum* are clustered close to the crop epithelium 6 hours after feeding. *H. muscarum* cells expressing tdTomato were fed to *D. melanogaster* expressing a myosin-GFP fusion protein.

At 24 hours there were very few *H. muscarum* expressing the fluorescent tdTomato protein marker visible in the *Drosophila* digestive tracts with the majority of dissected flies appearing to have cleared the trypanosomatids. At this time point the majority of food has been pushed out of the crop and through the digestive system. The emptied, tightly folded crop was highly auto-fluorescent and as such it is not possible to rule out the presence of *H. muscarum* in the crop at this time point. Future studies using sectioning prior to microscopy would allow better study of the crop in this state.

When these infections were performed with unstained, ‘wild-type’ *H. muscarum* cells, rosettes of cells which appeared to be surrounded by an ‘exclusion zone’ (Figure 3.6) were observed in the midgut of the fly 24 hours after feeding. It is unclear what this ‘zone’ is, it could be an air bubble in the sample or a secreted substance of trypanosomatid or fly origin. Putatively, the rosettes of cells resemble ‘cysts’ of *T. brucei* which have become trapped during penetration of newly secreted peritrophic matrix in the tsetse proventriculus (Rose *et al.* 2019). Like tsetse, *D. melanogaster* has a type II peritrophic matrix and as such *H. muscarum* may have a similar strategy to trypanosomes of circumventing this midgut barrier. Alternatively, zone around the *H. muscarum* rosettes may be similar to the proteophosphoglycan (PPG) matrix secreted by *Leishmania* promastigotes which acts as a protective coating for the cells.

Given the rosettes of cells were not observed in infections using transgenic *H. muscarum* cells, expression of the tdTomato protein may reduce *H. muscarum* fitness or it may be that cells are able to moderate expression of the tdTomato – similar to what has been observed in *T. brucei* expressing GFP during experimental tsetse infection (Gibson and Bailey, 2003).

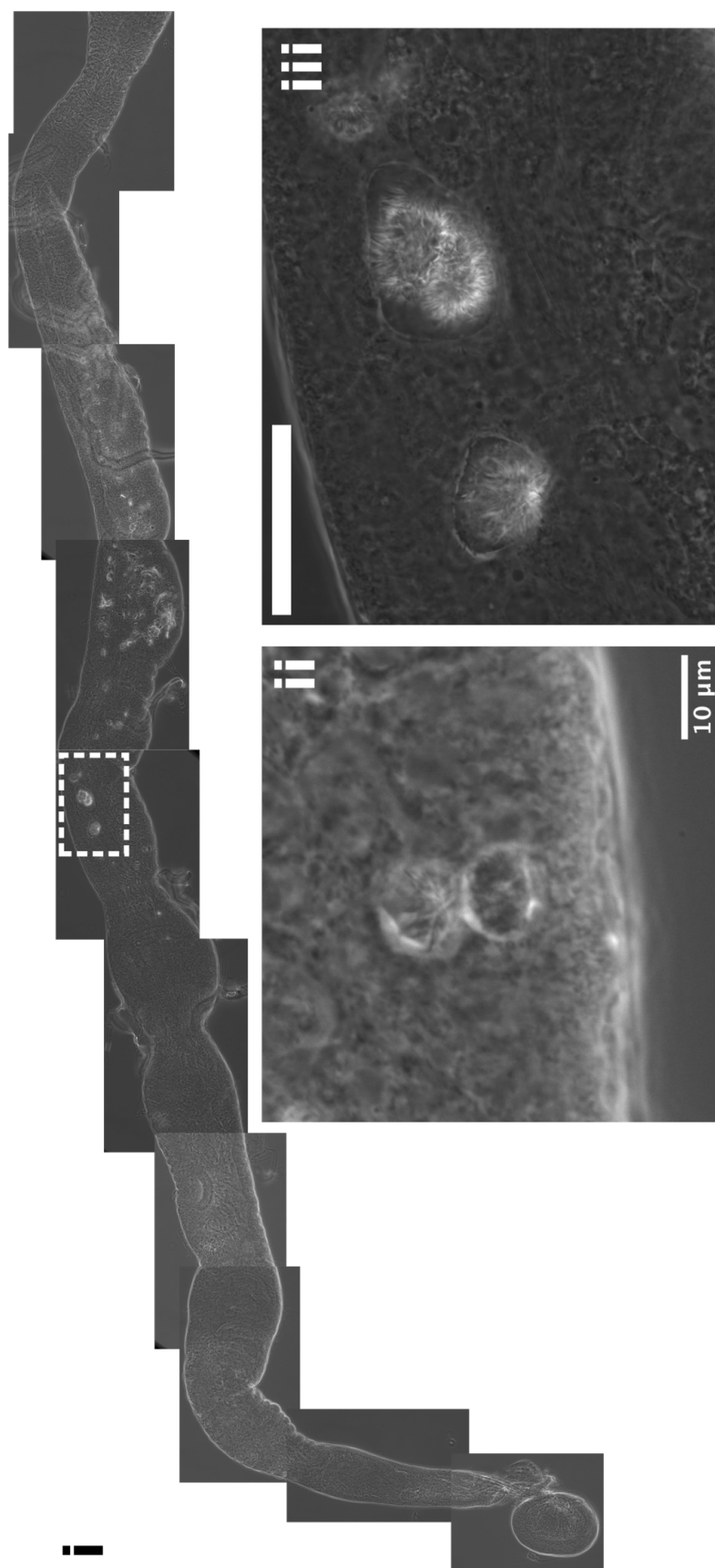


Figure 3.6 - *Herpetomonas muscarum* ‘ rosettes’ in the fly midgut 24 hours after feeding. Wild type *H. muscarum* were fed to Oregon^R *D. melanogaster*. Cells were found in the anterior of the midgut (i) in what appear to be ‘ rosettes’ of cells (ii, iii). Scale bar in (iii) represents 50μ m.

***Herpetomonas muscarum* transmission between adults**

Co-housing of flies fed *H. muscarum* with naïve flies did not lead to the transmission of the trypanosomatids, even though live *H. muscarum* were isolated from the faeces of infected adults. Similarly, the addition of 0.5ml of log-phase *H. muscarum* into vials of conventional food did not produce a continual infection in the fly populations incubated in the vials. Live *H. muscarum* cells were found in food smears on microscope slides for 2-3 days after artificial food inoculation, suggesting they were unable to survive in conventional fly food for longer periods. As such, it appears that *D. melanogaster* are unable to transmit *H. muscarum* between adult flies *in vitro*. It is unclear if this is due to laboratory conditions or if this is reflective of wild *Drosophila* populations. Alternatively, additional factors, such as microbiome constituents or specific nutrients, present in wild populations of *Drosophila* which are not replicated in the laboratory could be important to *H. muscarum* survival. Other organisms in the insect landscape, including endosymbionts, have been shown to influence trypanosomatid survival in tsetse (Farikou *et al.*, 2010).

Alternatively, it is possible that there is an as-yet unknown host or reservoir for *H. muscarum* which is required for transmission between fly hosts. Given wild *D. melanogaster*'s major food source is decaying fruits, and that *Herpetomonas* have been isolated from fruits (Marín *et al.*, 2007), plants seem to be a feasible reservoir of trypanosomatids. Furthermore, *Herpetomonas* is evolutionarily close to the plant pathogens of the *Phytomonas* genus. Finally, *D. melanogaster* may simply represent a 'dead-end' in for *H. muscarum* and there is another 'true' host. However, the comparatively high prevalence observed in *D. melanogaster* populations (5%) in Oxfordshire, vs. < 1% prevalence of *Leishmania* in sand fly populations, seems to refute this hypothesis.

Repeat exposure of *Herpetomonas muscarum* to *Drosophila melanogaster*

Given that other insect-trypanosomatid infections are known to persist in excess of 20 days, it seemed surprising that *H. muscarum* was only detectable in *H. muscarum* for 2-3 after feeding for the majority of infections. It is well established that pathogens can lose infectivity when cultured *in vitro*. Given that *H. muscarum* used in the above studies had been passaged for over 5 years *in vitro*, I hypothesized that the *H. muscarum* cells would be less effective at causing an infection compared to cells which had not been cultured *in vitro*. To assess whether the dynamics of infection changed when *H. muscarum* which had more recently been exposed to *D. melanogaster* I designed a ‘repeat’ infection experiment. For the primary infection, seven day-old, mated, adult Oregon^R females were fed log-phase *H. muscarum* culture as described for the previous experiments. After feeding, 100 females were homogenized to release any ingested *H. muscarum* and the resultant slurry sieved through a 40µm mesh to remove larger debris from the fly carcasses. To allow an estimation of *H. muscarum* recovery efficiency after sieving, five infected adults were snap frozen to assess the infection loads before sieving by gDNA extraction and qPCR. The flow through was centrifuged to pellet any *H. muscarum* cells. These cells rinsed in sterile 1xPBS before being resuspended into 500µL *H. muscarum* media with sucrose. At this stage 50µL of cell suspension was kept back to allow estimates of how many *H. muscarum* were recovered after sieving by gDNA extraction and qPCR. Extraction efficiency was low (~15-40%). As such the remaining cells were ‘concentrated’ by centrifugation and removal of supernatant to a final volume of into 50µL for the secondary infection – so as to match the conditions for the primary infection.

<i>H. muscarum</i> per fly (primary infection)	<i>H. muscarum</i> extracted per fly	Extraction efficiency	Cell density used for secondary infection (<i>H. muscarum</i> ml ⁻¹)	<i>H. muscarum</i> per fly (secondary infection)
1.46E+04	5.86E+03	40.0%	1.06E+07	57
1.23E+05	2.01E+04	16.4%	3.62E+07	97
6.10E+04	9.06E+03	14.9%	1.63E+07	861

Table 3.1 – Extraction of *Herpetomonas muscarum* from *Drosophila melanogaster* adults for use in ‘secondary’ infections.

The cells were then fed to a second set of seven-day-old adult female Oregon^R flies (secondary infection). From the secondary infection flies were snap frozen at 0, 6, 24, 48 and 72 hours after feeding and the *H. muscarum* loads quantified by gDNA extraction and qPCR. The infection dynamics did not appear to change upon the second exposure of the ‘primed’ trypanosomatid to a naïve fly, with *H. muscarum* numbers rapidly declining following feeding (Figure 3.7). To investigate whether the isolation protocol was impacting the trypanosomatids ability to cause a secondary infection, *H. muscarum* extracted from the flies were cultured *in vitro* and their growth recorded daily (Figure 3.8). I did not observe a significant change in the growth rate of the cells compared to cells only grown in culture. As such, if there is a loss of fitness when *H. muscarum* is cultured *in vitro*, it may require multiple passages through flies or a longer exposure to rescue the loss of fitness and change the dynamics of interaction.

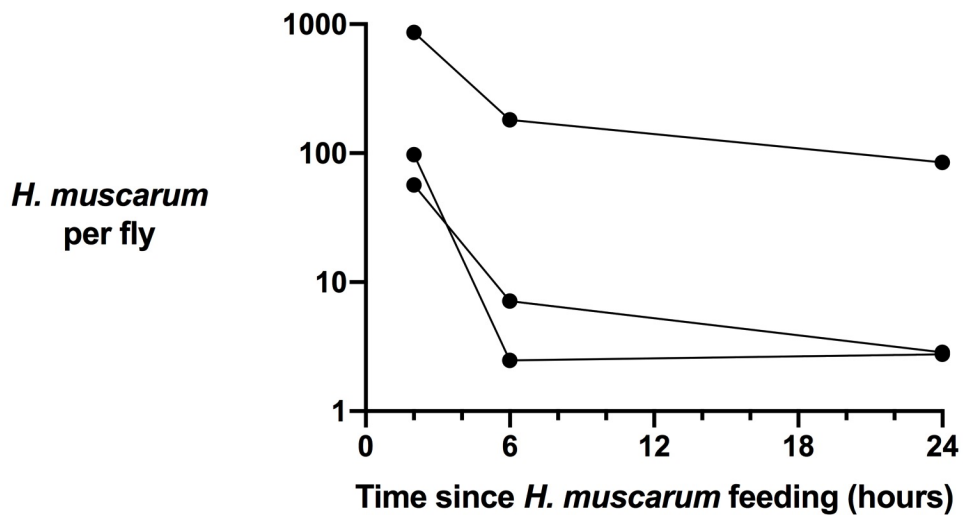


Figure 3.7 – *Herpetomonas muscarum* loads in *Drosophila melanogaster* during the ‘secondary infections’. Three biological repeats are shown separately.

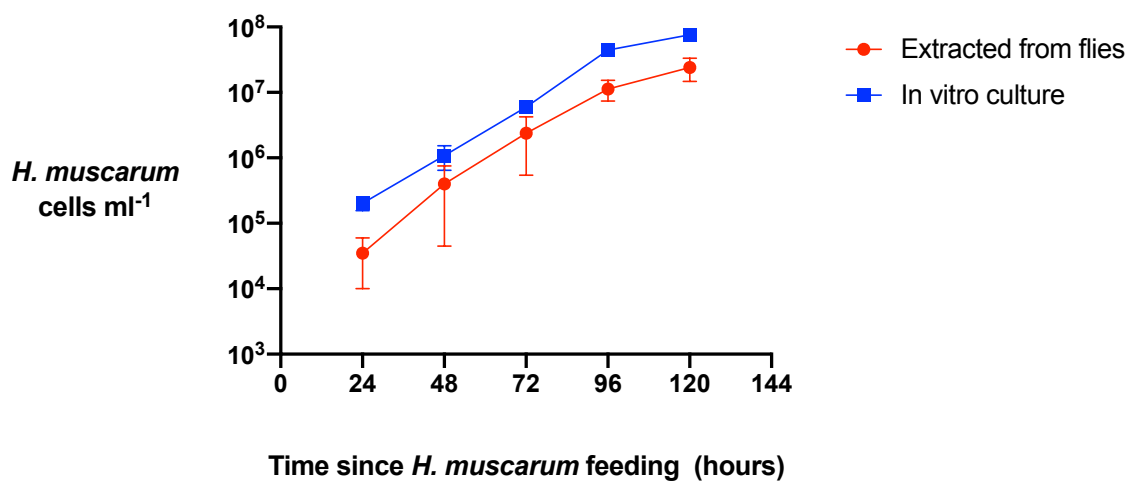


Figure 3.8 – Growth curves of *Herpetomonas muscarum* extracted from *Drosophila melanogaster* (red) vs. *in vitro* cultured cells (blue). The graph shows the average cell density for three biological repeats, with error bars showing the standard error of the mean.

Herpetomonas muscarum* exposure negatively impacts egg-laying in *Drosophila melanogaster

H. muscarum feeding reduced egg deposition of *D. melanogaster* w¹¹¹⁸ females in the first week after ingestion (Figure 3.9A). This result was not statistically significant (ANOVA, $p = 0.7337$). The reduction in egg laying associated with *H. muscarum* ingestion was more pronounced in immune deficient Relish^{E20} mutant flies, which lack the transcription factor for the Imd innate immune pathway, Relish (Figure 3.9B). However, this difference was not quite statistically significant different (ANOVA, $p=0.0814$). There was also no significant difference between the uninfected w¹¹¹⁸ and Relish^{E20} lines (ANOVA, $p=0.963$). Reduced fecundity as a result of parasite infection is a well-documented phenomenon, especially in invertebrates (Hurd, 2001). There is thought to be a trade off in the allocation of a limited pool of resources between reproduction, growth and maintenance. The host may reduce egg-production to devote those resources to the resolution of infection. This compromise can actually extend the lifespan of the infected insect, compared to uninfected individuals, as a result of the reduced reproductive efforts. Indeed, Dr. Lihui Wang observed the lifespan extension of *D. melanogaster* infected with *H. muscarum* when compared to uninfected controls (Wang, Sloan and Ligoxygakis, 2019).

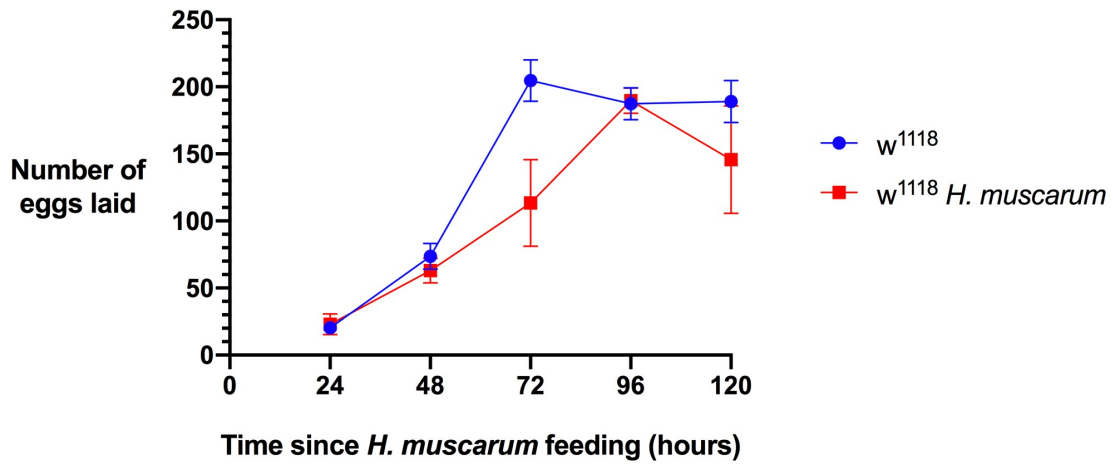


Figure 3.9A – Egg laying is reduced in *Drosophila melanogaster* fed *Herpetomonas muscarum*. Eggs laid by 20 mated females were counted daily for *w¹¹¹⁸* flies which had been fed *H. muscarum* (red) or 10% sucrose solution (blue). Graph shows averages for three repeat infections and error bars show standard error of the mean.

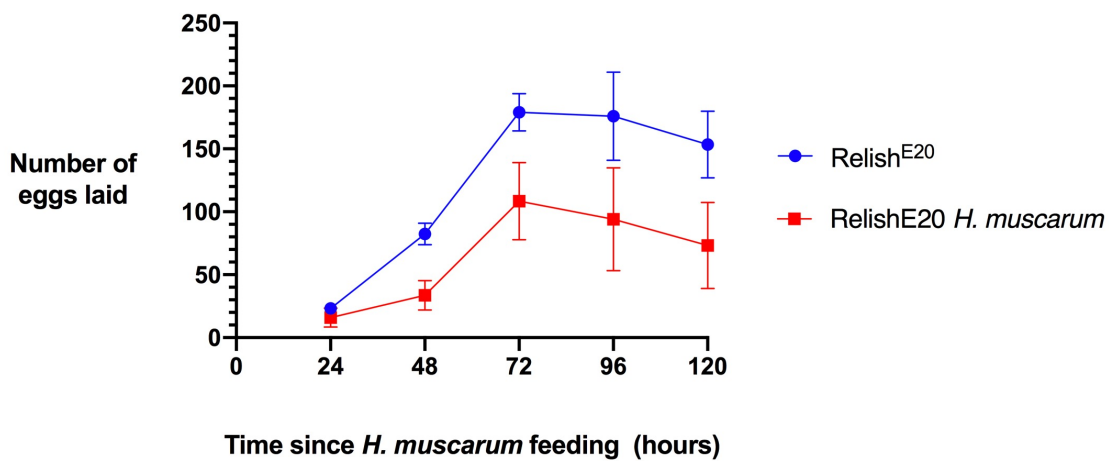


Figure 3.9B – Egg laying is reduced in immune deficient *Drosophila melanogaster* fed *Herpetomonas muscarum*. Eggs laid by 20 mated females were counted daily for *Relish^{E20}* flies which had been fed *H. muscarum* (red) or 10% sucrose solution (blue). Graph shows averages for three repeat infections and error bars show standard error of the mean.

Infection of Drosophila melanogaster larvae

D. melanogaster spends approximately five days as a larva. The majority of this time is spent feeding, fully immersed in the food/substrate. As such larvae are exposed to many micro-organisms both in a laboratory setting and in wild populations. Given their ‘lifestyle’, the larvae stages of the fly could represent a major reservoir for trypanosomatids. In order to assess whether the infection dynamics were different when larvae ingested *H. muscarum* (in comparison to adult flies) I designed a *H. muscarum* feeding protocol for *Drosophila* larvae based on those described in (Hamilton *et al.*, 2015). The infections were performed using Oregon^R L₂ stage larvae reared in bottles for 2 days after mated females were allowed to deposit eggs into the food. Larvae were incubated at 25°C in a 1.5ml Eppendorf tube containing thirty homogenised adult flies which had been ‘spiked’ with 1×10^7 *H. muscarum* cells from log-phase culture. Larvae were allowed to feed for two hours before being moved back into vials containing conventional *Drosophila* food. Larvae were snap frozen at 0, 6, 24 and 48 hours after feeding and the *H. muscarum* loads quantified by gDNA extraction and qPCR. An infection ‘control’ experiment was performed in parallel where larvae were fed *H. muscarum* medium with sucrose. Larvae were able to clear the *H. muscarum* within 24 hours of ingestion (Figure 3.10). The constant feeding behaviour of fly larvae may prevent *H. muscarum* from becoming established in the digestive tract, with the constant ‘flow’ of food acting as a mechanical barrier to establishment. Furthermore, larvae do not have a crop organ, which is where the majority of *H. muscarum* were observed in adults. As such the crop may represent a key site of interaction between *H. muscarum* and *D. melanogaster*.

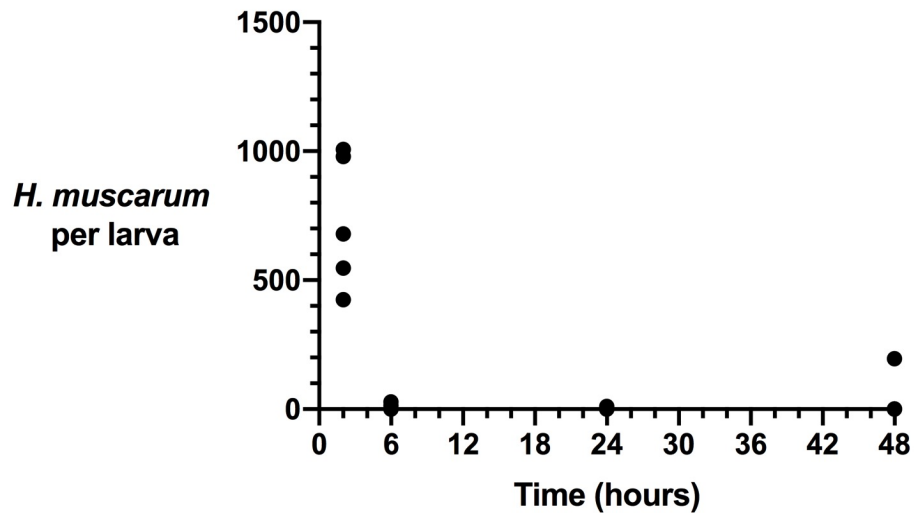


Figure 3.10 - *Herpetomonas muscarum* load in *Drosophila melanogaster* L2 larvae after two hours feeding. The graph shows the average number of ‘wild-type’ *H. muscarum* per larvae (from five larvae pooled) for five repeat infections.

Concluding Remarks

Work in this chapter has described the infection dynamics between *Drosophila melanogaster* and a trypanosomatid found to exist in wild fruit fly populations, *Herpetomonas muscarum*. The trypanosomatids are able to persist within adult flies for at least 96 hours after ingestion in a laboratory setting; though previous work has shown that many flies clear the infection much earlier, often as quickly as 12 hours post-ingestion (Wang, Sloan and Ligoxygakis, 2019). Presence of the trypanosomatids in the *Drosophila* midgut negatively affected egg laying in infected females whilst increasing the median lifespan (Wang, Sloan and Ligoxygakis, 2019). This follows the common observation

that immune challenge in insects results in reductions in fecundity, probably due to the re-allocation of energy stores from reproduction to combatting infection.

After ingestion, *H. muscarum* were found clustered together in the crop of the fly, analogous to the mammalian stomach, throughout the infection. I also observed putative interactions between *H. muscarum* and the fly middle midgut epithelium soon after ingestion and the formation of rosettes of cells in the anterior midgut which appear to be surrounded by an exclusion zone. It is tempting to hypothesize that this exclusion zone is peritrophic matrix encasing *H. muscarum* cells which have become trapped whilst attempting to enter the ectoperitrophic space – similar to has been observed in *T. brucei* (Rose *et al.* 2019). These observations appear to resemble important stages of vector infection by ‘medically important’ trypanosomatids. Namely, the circumvention of the tsetse peritrophic matrix by *T. brucei* and the attachment of *Leishmania* to the sand fly midgut to prevent exclusion with the remains of the digested blood meal. These parallels further underline the potential of the *Drosophila-Herpetomonas* model for the study of medically-important trypanosomatid parasites and their vectors.

In contrast to previous works with *Drosophila* and the trypanosomatids *H. ampelophilae* (Rowton and Barclay McGhee, 1978, 1983) and *J. drosophilae* (Hamilton *et al.*, 2015), I do not observe transmission of *H. muscarum* between adult flies or the persistence of *H. muscarum* in larvae beyond 6 hours. This may be due to the lack of a crop organ in the larvae which appears to be an important interaction site for *H. muscarum* in adults.

Given the observations that 5% of wild *Drosophila melanogaster* sampled (Oxfordshire, UK) were found to harbour the *H. muscarum*, the low ‘infectivity’ and apparent lack of

transmission mechanism were surprising. A possible reason for this disparity in observations could be the loss of infectivity after *in vitro* culturing which is commonly observed not only for trypanosomatid parasites but many other pathogenic organisms. However, I did not observe an increase in infectivity when the *H. muscarum* were ‘primed’ after passage through a fly prior to infecting a second fly. Alternative hypotheses include; the laboratory setting is impacting on the dynamics of infection and/or there is an unknown reservoir of *H. muscarum* in wild populations.

Life in the laboratory differs from the outside world in many important ways (Markow, 2015). The flies’ food source has been restricted to a few standardised recipes which often contain anti-microbial agents. While these food mixes are sufficient to maintain healthy *Drosophila*, they do not replicate the diversity of food sources which wild flies feed on. This is reflected in the flies’ gut microbiome (Staubach *et al.*, 2013). Given the nature of trypanosomatid infections, these differences in the gut landscape explain the seemingly low infectivity I observed in laboratory reared flies. For example, *H. muscarum* may have an ‘infection partner’ in the insect microbiome that assists/creates the conditions required for a more substantial infection which are absent from the laboratory environment.

In regard to a potentially unknown reservoir for *H. muscarum*, *Herpetomonas* sits closely to the plant-dwelling trypanosomatids of the *Phytomonas* genera. Furthermore, *H. muscarum* have been isolated from plants, specifically fruits, which are a staple of the *Drosophila* diet. As such, a plant life cycle stage for *Herpetomonas muscarum* is possible and may be essential for transmission of the trypanosomatids.

References

- Barclay McGhee, R. and Postell, F. J. (2007) 'Axenic Cultivation of *Phytomonas davidi* Lafont (Trypanosomatidae), a Symbiote of Laticiferous Plants (Euphorbiaceae)*', The Journal of Protozoology. Wiley/Blackwell (10.1111), 23(2), pp. 238–241. doi: 10.1111/j.1550-7408.1976.tb03761.x.
- Borghesan, T. C. *et al.* (2013) 'Molecular Phylogenetic Redefinition of *Herpetomonas* (Kinetoplastea, Trypanosomatidae), a Genus of Insect Parasites Associated with Flies', Protist. Urban & Fischer, 164(1), pp. 129–152. doi: 10.1016/J.PROTIS.2012.06.001.
- Boulanger, N. *et al.* (2001) 'Immune response of *Drosophila melanogaster* to infection with the flagellate parasite *Crithidia* spp.', Insect Biochemistry and Molecular Biology, 31(2), pp. 129–137. doi: [https://doi.org/10.1016/S0965-1748\(00\)00096-5](https://doi.org/10.1016/S0965-1748(00)00096-5).
- Chatton, E. and Alilaire, E. (1908) 'Coexistence d'un *Leptomonas* (*Herpetomonas*) et d'un Trypanosoma chez un muscide non vulnérant, *Drosophila* confuse Staeger', Comptesrendus des Seances de la Societe de Biologie, 64, pp. 1004–1006.
- Chicharro, C. and Alvar, J. (2003) 'Lower trypanosomatids in HIV/AIDS patients', Annals of Tropical Medicine & Parasitology. Taylor & Francis, 97(sup1), pp. 75–78. doi: 10.1179/000349803225002552.
- Dimitriadis VK, P. E. (1992) 'Functional morphology of the crop of *Drosophila auraria*.' Cytobios, (69), pp. 278–279.
- Ebbert, M. A., Burkholder, J. J. and Marlowe, J. L. (2001) 'Trypanosomatid Prevalence and Host Habitat Choice in Woodland *Drosophila*', Journal of Invertebrate Pathology, 77(1), pp. 27–32. doi: <https://doi.org/10.1006/jipa.2000.4989>.
- Fiorini, J. E. *et al.* (2001) 'Morphological, Biochemical and Molecular Characterization of *Herpetomonas samuelpessoai camargoi* n. subsp., a Trypanosomatid Isolated from the Flower of the Squash *Cucurbita moschata*' Journal of Eukaryotic Microbiology, 48(1), pp. 62–69. doi: 10.1111/j.1550-7408.2001.tb00416.x.
- Gibson, W. and Bailey, M. (2003) 'The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes', Kinetoplastid biology and disease. BioMed Central, 2(1), p. 1. doi: 10.1186/1475-9292-2-1.
- Hamilton, P. T. *et al.* (2015) 'Infection Dynamics and Immune Response in a Newly Described *Drosophila*-Trypanosomatid Association', mBio. Edited by K. Gull, 6(5), pp. e01356-15. doi: 10.1128/mBio.01356-15.
- Hoang, D., Kopp, A. and Chandler, J. A. (2015) 'Interactions between *Drosophila* and its natural yeast symbionts-Is *Saccharomyces cerevisiae* a good model for studying the fly-yeast relationship?', PeerJ. PeerJ Inc., 3, pp. e1116–e1116. doi: 10.7717/peerj.1116.
- Hurd, H. (2001) 'Host fecundity reduction: a strategy for damage limitation?', Trends in Parasitology, 17(8), pp. 363–368. doi: [https://doi.org/10.1016/S1471-4922\(01\)01927-4](https://doi.org/10.1016/S1471-4922(01)01927-4).

- Marín, C. *et al.* (2007) '*Herpetomonas* spp. isolated from tomato fruits (*Lycopersicon esculentum*) in southern Spain', *Experimental Parasitology*. Academic Press, 116(1), pp. 88–90. doi: 10.1016/J.EXPPARA.2006.11.003.
- Markow, T. A. (2015) 'The secret lives of *Drosophila* flies', *eLife*. eLife Sciences Publications, Ltd, 4, p. e06793. doi: 10.7554/eLife.06793.
- Maslov, D. A. *et al.* (2013) 'Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed', *Trends in Parasitology*. Elsevier, 29(1), pp. 43–52. doi: 10.1016/j.pt.2012.11.001.
- Morio, F. *et al.* (2008) 'Isolation of a protozoan parasite genetically related to the insect trypanosomatid *Herpetomonas samuelpessoai* from a human immunodeficiency virus-positive patient', *Journal of Clinical Microbiology*, 46(11), pp. 3845–3847. doi: 10.1128/JCM.01098-08.
- Pacheco, R. S. *et al.* (1998) 'Parasite Genotypically Related to a Monoxenous Trypanosomatid of Dog's Flea Causing Opportunistic Infection in an HIV Positive Patient', *Memórias do Instituto Oswaldo Cruz*. scielo, pp. 531–537.
- Podlipaev, S. *et al.* (2004) '*Herpetomonas ztiplika* n. sp. (Kinetoplastida: Trypanosomatidae): a parasite of the blood-sucking biting midge *Culicoides kibunensis* Tokunaga, 1937 (Diptera: Ceratopogonidae).', *The Journal of parasitology*, 90(2), pp. 342–347. doi: 10.1645/GE-156R.
- Rose, C. *et al.* (2019) '*Trypanosoma brucei* colonises the tsetse gut via an immature peritrophic matrix in the proventriculus', *bioRxiv*, p. 513689. doi: 10.1101/513689.
- Rowton, E. D. and Barclay McGhee, R. (1978) 'Population Dynamics of *Herpetomonas ampelophilae*, with a Note on the Systematics of *Herpetomonas* from *Drosophila* spp.', *The Journal of Protozoology*. John Wiley & Sons, Ltd (10.1111), 25(2), pp. 232–235. doi: 10.1111/j.1550-7408.1978.tb04402.x.
- Rowton, E. D. and Barclay McGhee, R. (1983) 'Transmission of *Herpetomonas* in Laboratory Populations of *Drosophila melanogaster*', *The Journal of Protozoology*. John Wiley & Sons, Ltd (10.1111), 30(4), pp. 669–671. doi: 10.1111/j.1550-7408.1983.tb05341.x.
- Sadd, B. M. and Barribeau, S. M. (2013) 'Heterogeneity in infection outcome: lessons from a bumblebee-trypanosome system', *Parasite Immunology*. John Wiley & Sons, Ltd (10.1111), 35(11), pp. 339–349. doi: 10.1111/pim.12043.
- Staubach, F. *et al.* (2013) 'Host species and environmental effects on bacterial communities associated with *Drosophila* in the laboratory and in the natural environment', *PloS one*. Public Library of Science, 8(8), p. e70749.
- Tyc, J. *et al.* (2013) 'Growing diversity of trypanosomatid parasites of flies (Diptera: Brachycera): frequent cosmopolitanism and moderate host specificity.', *Molecular phylogenetics and evolution*. United States, 69(1), pp. 255–264. doi: 10.1016/j.ympev.2013.05.024.

- Votýpka, J. *et al.* (2010) 'Probing into the diversity of trypanosomatid flagellates parasitizing insect hosts in South-West China reveals both endemism and global dispersal', *Molecular Phylogenetics and Evolution*, 54(1), pp. 243–253. doi: <https://doi.org/10.1016/j.ympev.2009.10.014>.
- Wang, L., Sloan, M. A. and Ligoxygakis, P. (2019) 'Intestinal NF- κ B and STAT signalling is important for uptake and clearance in a *Drosophila-Herpetomonas* interaction model', *PLOS Genetics. Public Library of Science*, 15(3), p. e1007931.
- Wheeler, R. J., Gluenz, E. and Gull, K. (2013) 'The Limits on Trypanosomatid Morphological Diversity', *PLOS ONE. Public Library of Science*, 8(11), p. e79581.
- Wilfert, L. *et al.* (2011) 'Trypanosomatids are common and diverse parasites of *Drosophila*', *Parasitology*. 2011/04/18. Cambridge University Press, 138(7), pp. 858–865. doi: DOI: 10.1017/S0031182011000485.
- Yurchenko, V. *et al.* (2016) 'Diversity of Trypanosomatids in Cockroaches and the Description of *Herpetomonas tarakana* sp. n.', *Journal of Eukaryotic Microbiology*, 63(2), pp. 198–209. doi: 10.1111/jeu.12268.

CHAPTER FOUR

The transcriptomes of *Herpetomonas muscarum*

CHAPTER FOUR:

The transcriptomes of *Herpetomonas muscarum*

Much of the work in this chapter is published in Sloan *et al.* 2019. The RNA extraction from infected flies and cDNA library synthesis were performed by Dr. Lihui Wang. The RNA sequencing was done by Oxford Gene Technology on the HiSeq2000 platform. For the in vitro samples, I performed the RNA extraction from *Herpetomonas muscarum* cultures. The cDNA library synthesis and sequencing were performed by Mandy Sanders and Karen Brooks (Wellcome Sanger Institute). All analysis is my own unless stated.

Preface

One way of looking at an organism's response to a change in the environment is to look at changes in its transcriptome – genes under active transcription and presumed translation. In the context of host-parasite interactions, both host and parasite must adapt rapidly to the changing conditions. The host must protect itself and attempt to overcome the invading parasite, which usually involves production of an immune response. In turn the parasite must adapt to survive in the hostile host environment, which may include changes to temperature and pH, as well as the presence of immune factors. There are many, diverse strategies employed by both host and parasite to tip the balance of these interactions in their favour. What these strategies have in common is that they are made possible by changes to gene expression. There are a number of techniques available to detect these, often subtle, changes. RNA sequencing employs next generation sequencing techniques to determine which mature messenger RNAs (mRNAs) are present at specific time points in an organism, cell or tissue. A gene's expression level can then be inferred

from the amount of the gene-specific mRNA present. For an account of RNA-sequencing methodology see Chapter Eight.

Trypanosomatid transcriptomes in their insect hosts

Many studies of the trypanosomatid transcriptome are focused on differentiation between the final life cycle stage found in the insect vector and those found in vertebrates – with a view to identify genes which are important in the context of human infections. There are fewer studies devoted to the insect-stage transcriptomes of these parasites. The studies which are available, often use axenic *in vitro* trypanosomatid cultures, owing to the challenges associated with working with hematophagous insects and the low masses of trypanosomatids present in each infected insect. The latter restricts the amount of trypanosomatid RNA which can be purified from infected flies. This issue can in some ways be circumvented by mRNA amplification prior to sequencing. However, this introduces PCR-associated amplification biases to the study, as the efficiency of nucleic amplification varies according to the sequence. Features such as hairpins, GC content and lengths of poly-A and poly-T residues can significantly alter amplification (Duftner *et al.*, 2008). A further caveat to these studies is that, while axenic *in vitro* cultures of trypanosomatids have been shown to correlate reasonably well to their *in vivo* counterparts, there are still many differences in their transcriptomes –as much as 13% of differentially regulated genes in some studies (Holzer, McMaster and Forney, 2006; Alcolea *et al.*, 2014).

The dixenous trypanosomatids of the *Leishmania* and *Trypanosoma* genera have several life-cycle stages within their insect hosts. Each stage has a distinct morphology and unique transcriptome traits. Much of the following account of *Leishmania* and

Trypanosoma transcriptomes whilst in the insect host is based on the landmark studies by Alcolea *et al.* (2014), Savage *et al.* (2016) and Inbar *et al.* (2017).

Early stages of insect infection: responding to a changing environment

Many gene expression changes seen in insect stage trypanosomatids are associated with allowing the cell to respond to the new, challenging environment of the insect digestive tract - including genes associated with chromatin remodelling and protein production. Histones (and their associated modifying genes), ribosomal subunits, spliceosome apparatus components and RNA polymerases have all been shown to be differentially regulated in insect-stage trypanosomatids. There is a close association between histone mRNA abundance and the cell type in *L. major*, with histone upregulation in the replicative procyclic promastigotes but reduced expression in nectonomads – which have not been observed to undergo cell division. The patterns of expression in histones were typical of cell cycle associated mRNA transcripts including cyclins and DNA polymerases in *L. major*.

There are several major changes to a trypanosomatid's environment when they are ingested by the insect vector at the bite site of the infected vertebrate. The parasites must sense and trigger a response to major changes in temperature, pH, exposure to the insect microbiome, digestive enzymes and immune factors. The exclusively extracellular African trypanosomes have an expanded repertoire of transmembrane adenylate-cyclases which has not yet been observed in other trypanosomatids. Adenylate cyclases convert ATP to the 'second messenger' cAMP which is vital to signal transduction in many responses to external stimuli. In *Trypanosoma brucei* there are three sets of these adenylate-cyclases which are differentially regulated in the insect stages – all of which

are found in the flagellum membrane (Savage *et al.*, 2016). Fourteen of these signal relaying proteins were associated with midgut procyclics, thirteen were upregulated in proventriculus-dwelling procyclics and two were shown to be specific to the metacyclic cell stage. As such the three sets of adenylate-cyclases may be important responding to specific micro-environments within the tsetse.

Cell surface genes

During host-parasite interaction, some of the most important proteins are likely to be those found on the surface of the invading cell. These are what the invader displays to the host and it is these molecules, along with any secreted molecules, which allow the detection and response to the foreign species by the insect. Accordingly, surface protein genes display some of the most dynamic expression in the trypanosomatid genome. Almost 10% of differentially expressed transcripts in *Trypanosoma brucei* during infection of *Glossina morsitans* were from the cell surface phylome (Savage *et al.*, 2016). Surface proteins are typically the most studied and therefore some of the best understood proteins in host-parasite interactions.

Trypanosoma brucei's surface proteome is carefully regulated at each stage of tsetse infection. Procyclins, glycoproteins bearing internal Glutamic acid-Proline (EP) or Glycine-Proline-Glutamic acid-Glutamine acid-Threonine (GPEET) repeats, are known to be the dominant surface protein of midgut forms of *Trypanosoma brucei*. These proteins are important for cell fitness as they provide protection against the fly's digestive enzymes (Haines *et al.*, 2010). However, they are not essential for completion of the lifecycle as infectious metacyclics were still able to be produced in tsetse infected with a procyclin-negative *T. brucei* line (Vassella *et al.*, 2009). Procyclins are expressed in a

specific order with GPEET2 dominating the cell surface early after ingestion by tsetse, followed by a switch to EP-repeat procyclins (Acosta-Serrano *et al.*, 2001; Savage *et al.*, 2016).

Like *T. brucei*, the *Leishmania* cell surface proteome is carefully regulated. Several of these surface proteins have been shown to play important roles during infection including: the promastigote surface antigens (PSAs), the major surface protease GP63 and lipophosphoglycan (LPG). The PSAs are a large family of secreted and membrane bound proteins hallmarked by tandem leucine rich repeat sequences flanked by cysteine rich regions (Devault and Bañuls, 2008). Members of this family are the most strongly upregulated genes in the *Leishmania major* procyclic cells during infections of *Phlebotomus papatasi*. The PSAs are grouped into eight orthogroups, seven of which are slow evolving and as such are speculated to have functions in basic interactions with the insect vector. The final class of PSAs were shown by Devault and Bañuls (2008) to be evolving under positive selection, and as such are thought to be involved in more specialised adaptations to specific vectors. The PSAs have been demonstrated to provide resistance to complement-mediated lysis during the vertebrate stages of infection (Lincoln *et al.*, 2004). However, they may have additional roles during the infection of the sand flies. Leucine-rich repeats are often associated with protein-protein or protein-glycolipid interactions and many have been implicated in binding of bacterial pathogens to mammalian and plant hosts (Seepersaud *et al.*, 2005; Pizarro-Cerdá and Cossart, 2006).

Major surface proteases, the GP63s, are also highly expressed early on in *L. major* sand fly infections. The GP63s can be attached to cell surface proteins via a glycosylphosphatidylinositol (GPI)-linker. While the significance of GP63 in insect

infections is unclear, it has been shown that GP63 is able to bind to and suppress the proliferation of human natural killer cells (Lieke *et al.*, 2008). There are seven GP63 genes in the *Leishmania major* genome. Their expression is stage-specific and are thought to provide protection against the sand fly immune factors and digestive enzymes. LP3, which has a key role in the synthesis and assembly of GP63, LPG and many other glycosphospholipid (GPI)-anchored proteins, is initially upregulated in procyclic promastigotes and then subsequently down regulated in nectonomads forms.

LPG has been implicated in establishment of infection in the insect midgut by several studies in *Leishmania* (Lang *et al.*, 1991; McConville *et al.*, 1992; Pimenta *et al.*, 1992, 1994; Kamhawi *et al.*, 2004). Inbar *et al.* (2017) did not observe differential expression of LPG biosynthesis genes, though the galactosyltransferases responsible for modification of the LPG sidechains were differentially expressed. The majority of these transferases are down regulated in the early stages of sand fly infection, however one side-chain galactosyltransferase (SCG5) was shown to be upregulated in procyclic promastigotes. As such SCG5 specific modifications to LPG may be important to the survival of *Leishmania* cells in the sand fly.

Metabolism and stress

It has been shown in *Leishmania* that insect stage cells utilize mainly sugar and amino acids (Hart and Coombs 1982, Oppendoes and Coombs 2007). Accordingly, Inbar *et al.* (2017) showed upregulation of three glucose transporters (GT1-3) in *L. major* promastigotes, which were then further upregulated in nectonomad forms. Furthermore, the majority of glyceraldehyde-3-phosphate (GAPDH) and several other glycolysis gene mRNA transcripts peak during the midgut stages of sand fly infection in *L. major*. The

amino acid transporter, AAP24, was also shown to be upregulated in promastigotes compared to the amastigote form - which favour use of fatty acids as a carbon source. AAP24 is responsible for the uptake of proline and alanine (Inbar *et al.*, 2013). Proline is a major carbon source for insect dwelling parasites and it can reach 60mM concentrations in insect haemolymph (Marchese *et al.*, 2018). Other amino acid transporter mRNAs in *L. major* appear to gradually accumulate throughout the sand fly infection stages, peaking at the metacyclic stage. However, transcripts from the aATP11 family of transporters and the arginine transporter, AAP3, peaked sharply in nectonomads. These transporters have been linked to stress responses (Aoki *et al.*, 2017), as such their upregulation highlights the hostile environment of the insect digestive tract.

Protein recycling by autophagy, the controlled self-degradation pathway allowing the recycling of superfluous or non-functional cell components, is a cellular process normally associated with a stressed phenotype. It has been shown that autophagy is also essential for metacyclogenesis, the differentiation to give the metacyclic cell type, in *Trypanosoma cruzi* and *Leishmania* (Besteiro *et al.*, 2006; Williams, Mottram and Coombs, 2013; Vanrell *et al.*, 2017). The major marker for autophagy, ATG8, is strongly upregulated in *L. major* nectonomads (Inbar *et al.*, 2017). ATG8 mediates the production and lipidation of autophagosomes, whose contents are to be recycled (Shpilka *et al.*, 2011). The mature autophagosomes then fuse with lysosomes which contain hydrolytic enzymes, allowing their contents to be degraded and recycled. The ATG8 transcript levels tail off again in metacyclic *L. major* cells. It is difficult to resolve whether autophagy is triggered as a stress response, as metabolic indicators tell us that nectonomads are likely under nutrient stress (as discussed above), or in relation to differentiation. It may be both, with differentiation acting as a form of stress response in trypanosomatids.

Late stage insect infection: pre-adaption for the vertebrate environment

Several studies suggest that the metacyclic trypanosomatids show ‘pre-adaptation’ towards the vertebrate life-cycle stages (Sacks, 1989; Depledge *et al.*, 2009; Alcolea *et al.*, 2010, 2014). There have also been reports of *Leishmania* cells with amastigote-like morphologies in the stomodaeal valve of the sand fly (Añez *et al.*, 2003).

Alcolea *et al.* (2014) compared the transcriptomes of *Leishmania infantum* metacyclic promastigotes isolated from the stomodaeal valves of sand flies with amastigotes extracted from *in vitro* macrophage infections. This study highlighted the up-regulation of the biotin/lipoate liagse (BLPL) genes in sand-fly dwelling promastigotes. The BLPL genes are known to be involved in fatty acid biosynthesis via an acetyl-CoA carboxylase. Inbar *et al.* (2017) also reported the upregulation of fatty-acid biosynthesis genes in *L. major*, though slightly earlier - during in nectonomads in the later stages of midgut infection. In *L. infantum*, there was also upregulation of membrane lipid modification genes; sphingolipid desaturase and glycosylinositol phospholipid galactofuranose transferase (GIPL-galf) in the metacyclics. As such the group hypothesized there was an increased requirement for fatty-acids in the metacyclic in the sand fly and that this could be due to remodelling of the cell surface sphingolipids and phospholipids. Given that GIPLs are known to be crucial for uptake into the macrophage during vertebrate infection this change in gene expression is thought to represent ‘pre-adaptation’ to the amastigote stage (Suzuki *et al.*, 2002).

Additionally, stationary phase *L. infantum* promastigotes in culture, which show gene expression correlation to *in vivo* metacyclic promastigotes isolated from the stomodaeal valve, show high levels of amastin gene expression when compared to log phase axenic

culture (Alcolea *et al.*, 2010) - though the highest levels of amastin expression are reached during the amastigote stage. These surface proteins are known virulence factors have been shown to mediate membrane-membrane contacts between the *Leishmania* amastigote and the parasitophorous vacuole in the macrophage (de Pavia *et al.* 2015). Metacyclic cells also show upregulation amastigote-associated of the hydrophilic acylated surface proteins (HASPs). These proteins are required for differentiation to the amastigote stage (Doehl *et al.*, 2017), where they appear to interact with proteins of macrophage origin and enhance the production of reactive oxygen species (Kumar *et al.*, 2017). Upon entry into the fly they are dramatically downregulated in procyclic and nectonomads forms. Observation of upregulation of amastins and HASPs in metacyclic-like promastigotes provides further evidence of pre-adaptation to amastigote life.

Similar ‘pre-adaptive’ gene expression changes have also been recorded in the *Trypanosoma*. Savage *et al.* (2016) showed upregulation of a glucose transporter, which is highly expressed cells during infection of vertebrates, in salivary gland-dwelling metacyclic cells. Glucose is a major carbon source for blood stream inhabiting parasites, where the concentration can reach ~ 5mM with 95-99% oxygen saturation levels (Smith *et al.*, 2017). In contrast, tsetse is a glucose-poor environment with amino acids becoming the most accessible carbon sources (Seed and Wenck, 2003). As such upregulation of the blood stream glucose transporter could represent pre-adaption to the blood stream environment.

Trypanosomes also appear to pre-emptively change their surface coat prior to leaving their insect vector and entering a vertebrate host. During the latter stages of *T. brucei*'s infection of *G. morsitans*, epimastigotes inhabiting the salivary glands change their

dominant surface molecule from procyclin to BARP (bloodstream alanine-rich) proteins. An 87-fold upregulation of BARP compared to midgut forms was reported by Savage *et al.* (2016). The final tsetse stage of *T. brucei*, the metacyclic forms, express a single variant specific glycoprotein (VSG) per cell. It is this diversity which allows *T. brucei* to evade the vertebrate immune system. Savage *et al.* (2016) showed upregulation of twelve invariant surface glycoproteins in metacyclic, two VSGs lacking cysteine residues at the c-terminal (atypical VSGs). Metacyclic cells also express invariant surface proteins, known as MISPs, a recently described group of GPI-linked surface molecules which are thought to extend beyond the dense VSG coat (Casas-Sánchez *et al.* 2018). The knock down of these proteins did not affect the trypanosomes ability to infect the vector, however they may be important during the early stages of vertebrate infections.

Finally, an important process during the late stages of insect infection appears to be sexual replication (Akopyants *et al.*, 2009; Peacock *et al.*, 2014). The genetic diversity provided by meiosis greatly improves the likelihood of survival in the subsequent vertebrate hosts. In *L. major*, transcripts for the meiosis associated genes: SPO11, HOP1 and DMC1, were shown to be most abundant in metacyclic cells. Similarly the meiosis-specific genes, MND1, DMC1, and HOP1, were detected in *Trypanosoma brucei* in the salivary glands of tsetse at 14 and 38 days after ingestion with a peak of expression at day 20 (Peacock *et al.*, 2011).

Are responses to ingestion by dipterans conserved between trypanosomatids?

By identifying genes which are crucial to infection in multiple trypanosomatid-dipteran interactions, it may be possible to develop new broad-spectrum transmission blocking strategies for trypanosomatid diseases. In order to look for such genes, I sought to

compare the transcriptomic changes in *H. muscarum* after ingestion by *D. melanogaster*, to those already described for disease-causing trypanosomatids in their respective vectors.

RNA-sequencing had already been used by myself and others in the Ligoxygakis lab to investigate the transcriptome of *D. melanogaster* after ingestion of *H. muscarum* (Wang, Sloan and Ligoxygakis, 2019). Given that the sequenced RNAs were extracted from infected flies, this pool of RNA contained mRNAs from both host and trypanosomatid. As such I was able to investigate the transcriptome of *H. muscarum* from the same sequencing runs. By comparing the transcriptome of *H. muscarum* before and after ingestion I hoped to identify which genes were highly upregulated and therefore may be important for survival in the fly. This chapter also describes the transcriptome of *H. muscarum* during *in vitro* culture by examining the transcriptomes of cells during log and stationary phases of growth. It was important to examine these transcriptomes as the flies were fed *H. muscarum* from *in vitro* culture and therefore these transcriptomes would provide a baseline to compare with the in-fly transcriptomes with.

The work in this chapter provides novel insights into the transcriptome of a monoxenous trypanosomatid after ingestion by an insect. Importantly, it also identified transcriptomic responses to infection which are conserved between *H. muscarum* and parasites of the *Leishmania* and *Trypanosoma*. These included predicted changes to the cell surface proteome, as well as transcriptomic changes associated with increased levels of transcription, nutrient stress and autophagy. These observations highlight the potential of the *Drosophila-Herpetomonas* to assist in unpicking the molecular interactions between disease-causing dipterans and trypanosomatids.

Results and Discussion

The Herpetomonas muscarum transcriptomes during *in vitro* culture

RNA sequencing

I wished to examine the transcriptome of *H. muscarum* during *in vitro* axenic culture as the log-phase *in vitro* transcriptome is the ‘pre-infection’ transcriptome and stationary phase axenic culture transcriptomes have been shown to resemble those of insect stages in other trypanosomatids (Alcolea *et al.*, 2016). A total of 1×10^9 RNA-sequencing reads were produced from six *in vitro* culture samples (three biological replicates of each growth phase). On average there were 1.64×10^7 reads per sample, and 96.1% of reads mapped to the *H. muscarum* genome (see Table 4.1).

Sample	Total Reads	Mapped Reads	Percentage of reads mapped	Unmapped reads	Percentage of reads unmapped
Log phase #1	1.94E+08	1.87E+08	96.69%	6.42E+06	3.31%
Log phase #2	1.97E+08	1.91E+08	96.86%	6.17E+06	3.14%
Log phase #3	1.93E+08	1.86E+08	96.73%	6.29E+06	3.27%
Stationary phase #4	1.47E+08	1.40E+08	95.70%	6.32E+06	4.30%
Stationary phase #5	1.58E+08	1.51E+08	95.72%	6.76E+06	4.28%
Stationary phase #6	1.38E+08	1.31E+08	94.41%	7.73E+06	5.59%

Table 4.1 – Mapping summary of RNA-sequencing reads from *in vitro* axenic cultures of *Herpetomonas muscarum*.

Differential expression analysis

The relationship between log phase and stationary phase samples was studied using the DESeq2 package (Love *et al.* 2014). Principal component analysis (PCA) (Figure 4.1) showed that the first principal component explains 68% of the variance between log and

stationary phase samples and that stationary-phase replicates were more variable than the log-phase replicates. Overall, there was a high degree of difference between the samples from log and stationary phase with the dataset dividing into two distinct clusters.

There were 4044 genes which were significantly differentially regulated (Wald test, p -adjusted < 0.05) between log phase and stationary phase cells (Table S4.1) – approximately a third of the genome. The majority of the gene expression changes were modest, with just 264 genes upregulated ≥ 2 -fold in stationary phase cells (Table S4.2) and 811 downregulated ≥ 2 -fold (Table S4.3). GO enrichment analysis, using Ontologizer (Vingron et al., 2007), did not identify any significantly enriched GO terms associated with differentially regulated genes. However only 62% of *H. muscarum* genes have associated GO terms. As such, I also looked for enrichment in Pfam domains. There were 26 Pfam domains significantly enriched in the genes upregulated in stationary phase (Table 4.2) and 73 Pfam domains significantly enriched among downregulated transcripts (Table S4.4).

Transcriptomic changes associated with entry into stationary phase

Autophagy and stress. *H. muscarum* begins to enter the stationary phase of growth when the population becomes dense, toxic waste products build up and fewer nutrients are available per cell. It was therefore unsurprising that we observed several gene expression changes indicating metabolic change and nutrient starvation. Genes containing the Pfam domain associated with major autophagy marker ATG8 were significantly enriched in stationary phase transcripts (33 in total). As previously discussed, autophagy is a vital

process for survival in nutrient poor environments and involves the segregation of the cell components to be recycled into double membrane-bound vesicles called autophagosomes.

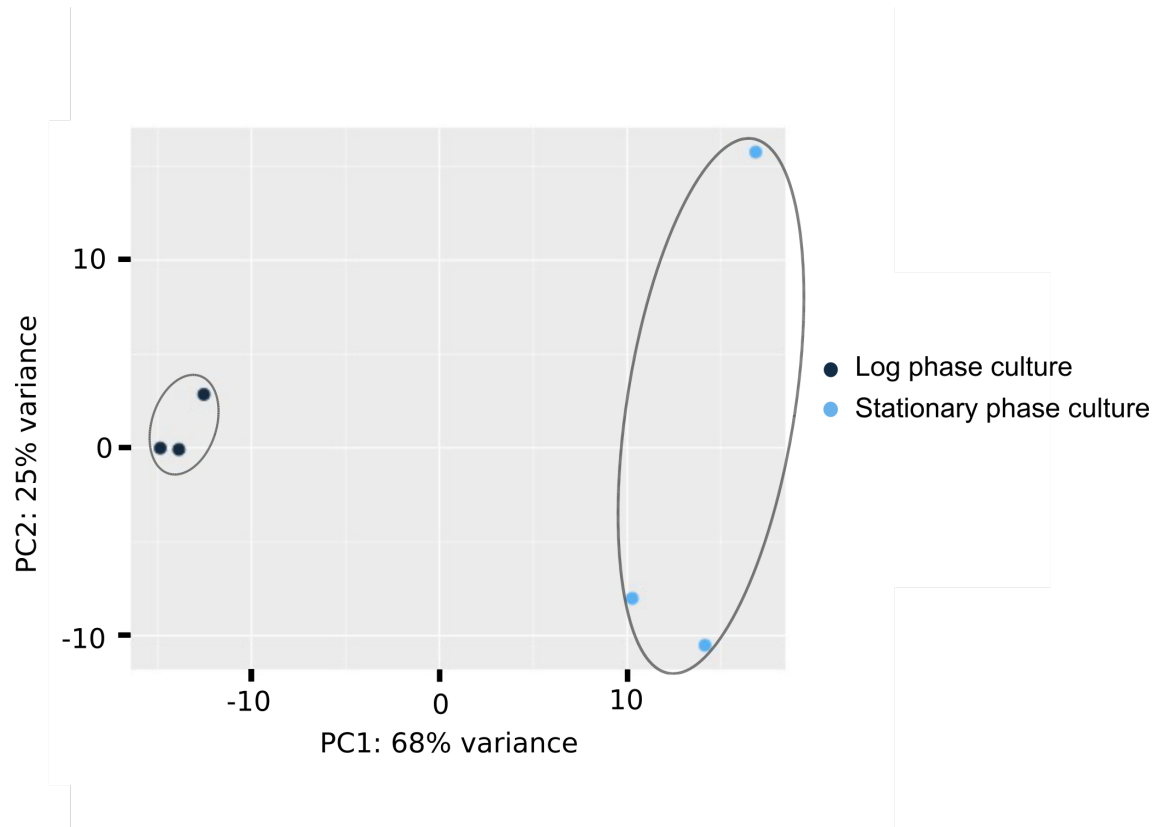


Figure 4.1 - Principal component analysis of differentially expressed *Herpetomonas muscarum* genes in log phase culture vs. stationary phase culture. There are two clear sample groupings (circled) which correspond to RNA each condition (n=3 per condition). Dark blue = log phase samples and light blue = stationary phase samples.

Pfam Description	Pfam_ID	Frequency	Chi-squared value	p-value
Protein kinase domain	PF00069.25	126	66.87	2.89E-16
Autophagy protein Atg8-like	PF02991.16	33	38.29	6.10E-10
RNA-recognition motif 1	PF00076.22	33	5.36	2.06E-02
Kinesin	PF00225.23	24	7.71	5.50E-03
Zinc finger CCCH	PF00642.24	23	7.59	5.87E-03
Mitochondrial carrier protein	PF00153.27	19	9.56	1.98E-03
Histone	PF00125.24	18	9.03	2.66E-03
AMP-binding	PF00501.28	17	13.05	3.03E-04
Calcineurin-like phosphoesterase	PF00149.28	14	5.46	1.95E-02
Sugar transporter	PF00083.24	14	5.91	1.51E-02
Giardia variant-specific surface protein	PF03302.13	13	19.75	8.84E-06
Cytochrome b5	PF00173.28	12	4.87	2.74E-02
Protein tyrosine phosphatase	PF00102.27	12	5.84	1.57E-02
E1-E2 ATPase	PF00122.20	11	8.53	3.50E-03
Cyclic nucleotide-binding domain	PF00027.29	11	8.53	3.50E-03
Polyadenylation specificity factor A subunit	PF03178.15	10	10.91	9.59E-04
Calpain family cysteine protease	PF00648.21	9	4.25	3.94E-02
Placenta-specific gene 8 family	PF04749.17	9	5.37	2.04E-02
Phosphoinositide 3-kinase	PF00454.27	8	7.57	5.93E-03
Cyclin	PF00134.23	8	4.85	2.77E-02
Alpha kinase family	PF02816.18	6	8.88	2.88E-03
Ion transport protein	PF00520.31	5	4.02	4.50E-02
Adenosine/AMP deaminase	PF00962.22	5	4.02	4.50E-02
FtsX-like permease family	PF02687.21	4	5.92	1.50E-02
BTB/POZ domain	PF00651.31	4	5.92	1.50E-02
membrane-bound O-acyltransferase family	PF03062.19	3	4.44	3.51E-02

Table 4.2 – Pfam domains significantly enriched in *Herpetomonas muscarum* genes upregulated upon entry into stationary phase during axenic culture (vs. log phase).

Autophagy requires increased amounts of cell membranes, this may partially explain why I also observed the upregulation of fatty-acid synthesis related genes in stationary phase, as fatty acids are crucial components of cell membranes. Three lipases, two putative lipase precursor-like proteins, fatty-acyl-CoA synthase 1 and putative fatty acid elongase (ELO) protein were upregulated upon entry into stationary phase. This is consistent with observations of *Trypanosoma cruzi* cultures (Lee *et al.* 2006).

Whilst the upregulation of autophagy-related genes is an indicator of cell stress, I also observed the downregulation of several genes with domains associated with responding to oxidative stress including: thioredoxin, glutathione S-transferase and alkyl hydroperoxide reductase (AhpC) and thiol specific antioxidant (TSA). As such, cells do not appear to be under significant oxidative stress. Other forms of stress, such as reduced nutrient availability or pH changes, may be driving the predicted increases in autophagy. Additionally, transcripts bearing the heat shock protein 60 HSP60 domain (PF00118) were also significantly enriched in the downregulated transcripts, which is another indicator of cell stress.

Surface proteins. Proteins sharing a domain (cl28643) with the variant surface protein (VSP) proteins of the *Giardia lamblia*, a flagellated intestinal pathogen, were highly represented amongst genes upregulated in stationary phase *H. muscarum*. In *G. lamblia*, these VSPs are integral membrane proteins rich in cysteine residues, often in CxxC repeats. They have a highly conserved C-terminal membrane spanning region which has a hydrophilic cytoplasmic tail with a conserved five amino acid CRGKA signature sequence, and an extended polyadenylation signal (Svård *et al.*, 1998; Adam, 2001). One VSP, of hundreds in the *Giardia* genome, is expressed per *Giardia* cell and they are

thought to protect the cells from proteolysis (Nash *et al.*, 1988). A similar strategy of surface protein expression is utilised by blood stage *T. brucei* cells (Aitcheson *et al.*, 2005). This method of antigen switching plays a major role in immune system avoidance and survival in vertebrate hosts. In *H. muscarum* the VSP domain-containing genes are predicted, by Phobius (Käll, Krogh and Sonnhammer, 2004), to encode proteins with 8-9% cysteine residues, and a single predicted transmembrane domain predicted at the C-terminus. Notably there were also ten VSP domaining proteins were down regulated upon entry into stationary phase.

In addition to the VSP domain containing genes, several other putative surface proteins were differentially regulated upon entry to stationary phase; two putative amastin genes were highly upregulated, and eight transcripts which encode for proteins with the cytomegalovirus UL20A protein domain (PF05984), were downregulated in stationary phase *H. muscarum* cells. The functions of proteins with UL20a domains, including the domains namesake, are largely unknown. Deletion of UL20a from the human cytomegalovirus genome resulted in reduced viral production in infected fibroblasts (Van Damme and Van Loock, 2014). Further study will be required to elucidate the role of these proteins in trypanosomatids. Each of these putative surface protein classes were also highly upregulated *H. muscarum* after ingestion by *D. melanogaster* and are further discussed in later in the chapter.

Transcription. The bias towards downregulated transcripts in the stationary phase cells as compared to log phase suggests a reduction of transcription and translation during stationary phase. Furthermore, five tRNA-synthase Pfam domains (PF00133.22, PF00749.21, PF00152.20, PF00587.25, PF01411.19) were significantly enriched in

downregulated transcripts (chi-squared, $p < 0.05$) and I observe downregulation of RNA polymerase III subunits.

Cell cycle-associated proteins. The Pfam domain associated with cyclins was significantly enriched in genes upregulated in stationary phase cells. From this, I investigated the expression profiles of the cyclins, and their associated kinases. Eleven were found to be differentially regulated between the two cell populations (Table 4.3). There was significant downregulation of the mitosis-associated cyclin 8, CRK3 and several mitochondrial DNA polymerase subunits in stationary phase cells. Knockdown of CRK3 in *T. brucei* is associated with a reduction in cell growth (Tu and Wang, 2004). Furthermore, there was upregulation of the G₁-associated cyclins 7, 4 and 11. These observations would suggest a reduction in cell replication. Consistent with this, and with a reduction in cell growth, there were also significant reductions in transcripts for α - and β -tubulins, DNA polymerases and several protein synthesis-related genes including: 40S ribosomal subunits, 28S rRNAs and five putative elongation factor 2 genes. However, there was also upregulation of mitosis-associated cyclin 2 in the stationary phase cells. Cyclin 2 has two roles in *T. brucei* procyclics: cell cycle progression through G₁ and the maintenance of correct cell morphology at the posterior end of the cell (Hammarton, Engstler and Mottram, 2004). The CRKs 10 and 12, which were also upregulated in stationary phase cells, have been shown to interact with cyclin 2 and their knock-down results in growth defects (Liu, Hu and Li, 2013). CRK12 is also essential to survival of *T. brucei* in mice and its depletion by RNAi lead to defects in endocytosis, an enlarged flagellar pocket and abnormal kinetoplast localisation (Monnerat *et al.*, 2013). Given the relative abundance of many transcripts associated with reduced replication in stationary

phase cells, the upregulation of cyclin 2 and its associated CRKs may be more relevant to the maintenance of correct cell morphology than mitosis.

Overall, gene expression change associated with cell surface remodelling, autophagy and reductions in transcription were observed in cells entering stationary phase. Cyclin expression patterns appear to suggest a bias in cells at G₁ phase. This is similar to what is reported for in vitro culture of *T. brucei* procyclics (Matthews, 2005). I will now discuss transcriptomic changes in *H. muscarum* associated with ingestion by *D. melanogaster*.

Gene Name	<i>H. muscarum</i> orthologue ID	log2FoldChange	adjusted p-value
CRK4	HMUS00195900.1	1.3	8.89E-10
cyclin 11	HMUS01322900.1	1.2	4.32E-05
cyclin 2	HMUS00751100.1	1.2	2.72E-19
cyclin 4	HMUS00787500.1	1.1	1.53E-17
cyclin 7	HMUS00475100.1	0.8	2.41E-14
CRK10	HMUS01143000.1	0.7	6.49E-09
cyclin 5	HMUS00580100.1	0.7	2.02E-12
cyclin 10	HMUS01323000.1	0.5	0.001028281
CRK12	HMUS00986000.1	0.3	0.014946163
DNA-directed RNA polymerase III subunit, putative	HMUS00638800.1	-0.3	0.031652615
mitochondrial DNA polymerase I protein C	HMUS00828800.1	-0.5	0.005919747
mitochondrial DNA polymerase I protein D	HMUS00617400.1	-0.5	0.018358548
mitochondrial DNA polymerase I protein B,	HMUS01100200.1	-0.6	0.007351177
DNA polymerase alpha/epsilon subunit B	HMUS00740000.1	-0.7	0.004342921
DNA polymerase delta catalytic subunit	HMUS00566500.1	-0.7	0.006068252
CRK3	HMUS00914500.1	-1.0	1.06E-40
cyclin 8	HMUS00524500.1	-1.0	1.40E-39

Table 4.3 – Significantly differentially regulated cyclins and cyclin-related kinases between stationary and log phase *Herpetomonas muscarum*.

RNA sequencing reads

To identify genes which are likely to be important to the outcome of *H. muscarum* infection of *D. melanogaster*, I sought to analyse the transcriptome of the trypanosomatid using RNA-sequencing analysis. RNA was purified from whole flies at 6, 12, 18 and 54 hours post ingestion of *H. muscarum*. The cDNA produced from the extracted RNA was then sequenced on the HiSeq2000 platform (see Chapter Seven for details). The resulting sequencing reads were then mapped to the combined genomes of the fly and the trypanosomatid.

The number of sequencing reads which mapped to the *H. muscarum* genome ranged from 6949 to approx. 16.2 million reads per sample (Table S4.5). At six hours post ingestion, 40% of the total mapped reads were shown to map to *H. muscarum* (average of three biological replicates). This decreased to 20% in samples from 12 hours post ingestion and 9% at 18 hours post ingestion. This is unsurprising as the amount of *H. muscarum* RNA present in the flies decreases as the trypanosomatids are cleared by *D. melanogaster*, which can occur as early as 12 hours post ingestion (see Chapter Three). For differential expression analysis, only data up to and including 18 hours post infection was used, as at 54 hours the number of sequencing reads mapping to the *H. muscarum* genome dropped below 1% of the total number of mapped reads (Figure 4.2).

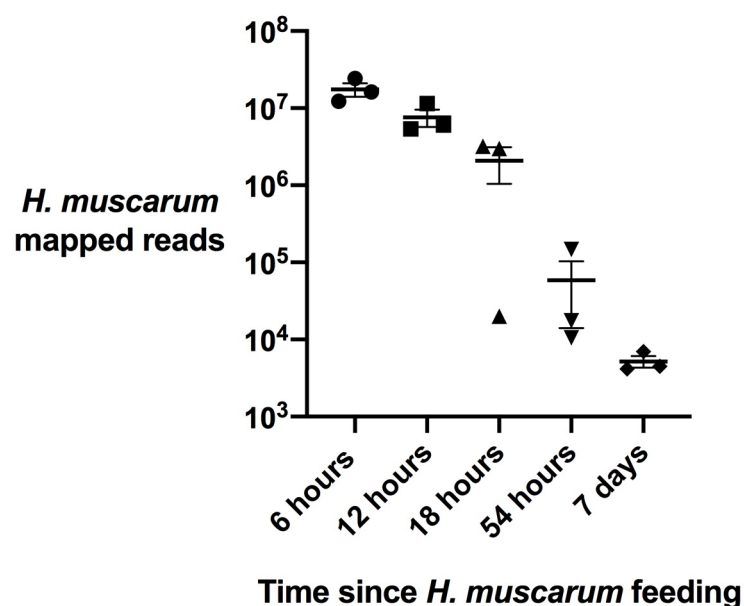


Figure 4.2 – Number of RNA-seq reads extracted from whole flies which mapped to *Herpetomonas muscarum*. Error bars show the standard error of the mean.

Differential expression analysis

The relationships between log phase samples and samples isolated from *D. melanogaster* were studied using the DESeq2 package (Love, Huber and Anders, 2014). Principal component analysis (PCA) shows that the first two principal components of variation in mRNAs between *H. muscarum* from *in vitro* culture and *H. muscarum* after ingestion by *D. melanogaster* were PC1 = 58% and PC2 = 10% (Figure 4.3). The PCA plot shows a high degree of difference between the *in vitro* samples and samples isolated from infected flies. The level of change in expression was much higher than between the two *in vitro* conditions discussed above.

For the infections, log phase *H. muscarum* cultures were used to feed the flies. In order to identify transcriptomic changes in *H. muscarum* associated with being ingested by the fly, I compared the transcriptome of *H. muscarum* cells from log phase *in vitro* culture to the in-fly transcriptomes. Over a third of the genome, 4633 genes, was significantly differentially regulated (Wald test, adjusted p-value < 0.05) between log phase axenic culture samples and samples from infected flies (Table S4.6). Comparisons of gene expression between sequential time points over the course of infection revealed that there was a large initial transcriptomic change upon ingestion - with 4662 genes differentially regulated between log phase culture and six hours post ingestion. This large initial transcriptomic shift was followed by more subtle transcriptomic changes between 6-12 (204 genes) and 12-18 hours (25 genes) (adjusted p-values < 0.05).

Herpetomonas muscarum* genes differentially regulated at six hours post-ingestion by *Drosophila melanogaster

Approximately a third of the *H. muscarum* genome is significantly differentially expressed between log phase axenic culture and six hours post ingestion by *D. melanogaster* (p < 0.05). Of this subset, 640 genes had a fold change of ≥ 4 between the time points – highlighting the magnitude of the trypanosomatids response to ingestion. GO enrichment analysis, using Ontologizer (Vingron et al., 2007), identified two significantly enriched GO terms in the 346 transcripts comparatively enriched at six hours post ingestion; OG0000045 (autophagosome assembly, p = 0.0014) and OG0003333 (amino acid transporters, p= 0.0002). Given the aforementioned lack of annotated GO terms in *H. muscarum*, I also looked at Pfam enrichment in the *H. muscarum* genes significantly upregulated upon ingestion by the fly. The top 15 represented Pfam domains in genes upregulated ≥ 4 -fold at six hours post-ingestion are all significantly enriched

compared to the full gene set (Table 4.5). Additionally, there were several Pfam domains enriched in the downregulated transcripts (Table 4.6). I will discuss these genes further below.

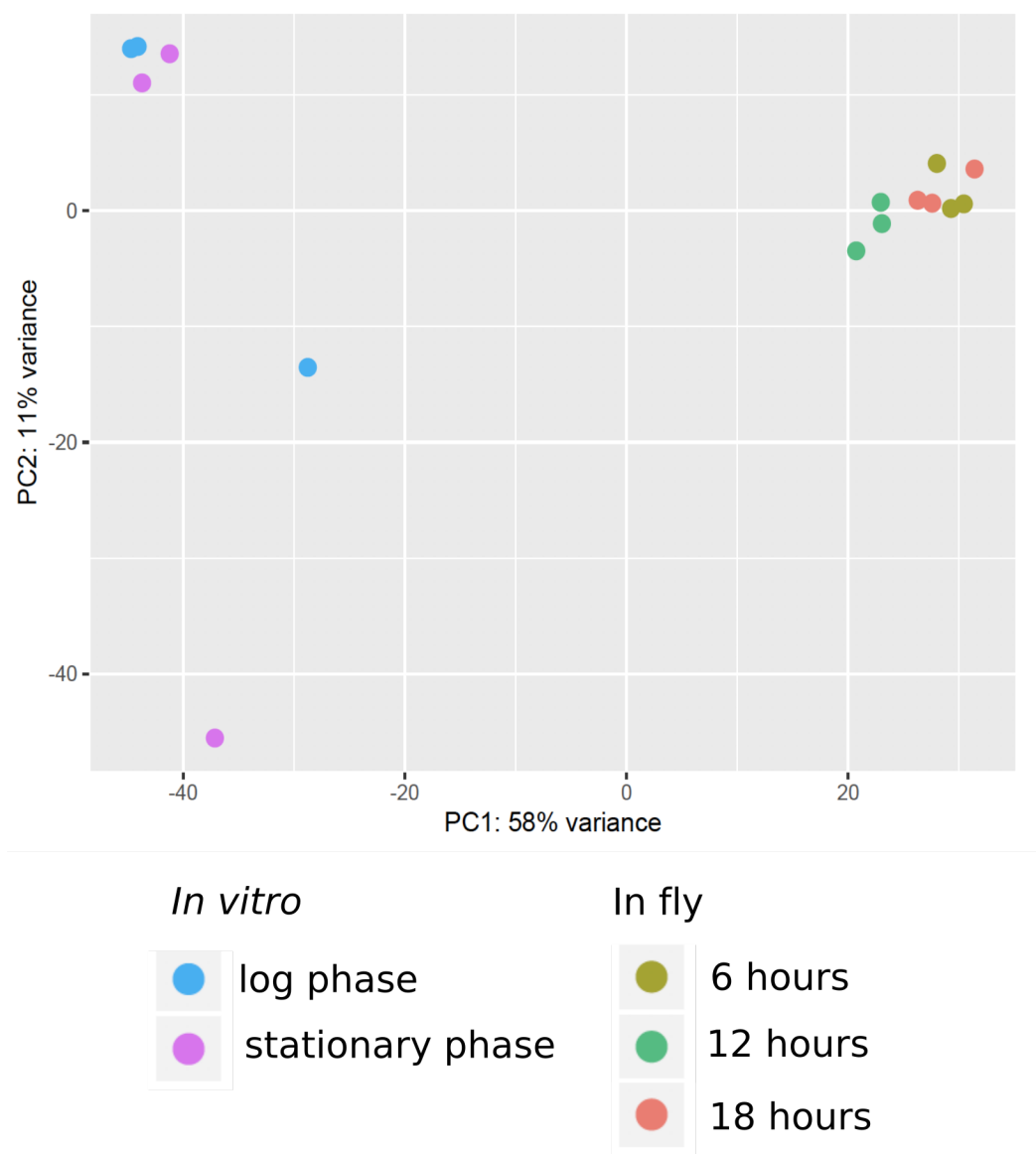


Figure 4.3 - Principal component analysis of differentially expressed *Herpetomonas muscarum* genes in log phase culture vs. samples isolated from infection flies at 6, 12 and 18 hours post ingestion. There are two clear sample groupings which correspond to RNA from *in vitro* axenic culture and RNA isolated from infected flies. N=3 per condition.

Pfam Description	Pfam_ID	Frequency	Chi-squared value	p-value
Leucine-rich repeat proteins	PF12799.7, PF13516.6, PF13855.6, PF14580.6, PF08263.12	65	26.8	2.26E-07
Autophagy protein Atg8-like	PF02991.16	26	324.6	1.41E-72
<i>Giardia</i> variant-specific surface protein	PF03302.13	21	36.9	1.23E-09
Amastin	PF07344.11	21	164.7	1.05E-37
Amino acid transporter	PF01490.18	11	21.2	4.17E-06
RNA recognition motif 1	PF00076.22	10	16.7	4.30E-05
polyadenylation specificity factor domain A	PF03178.15	9	119.0	1.06E-27
Zinc finger, C3HC4 type	PF13920.6	7	8.1	4.55E-03
Histone	PF00125.24	6	20.0	7.64E-06
BT1 family (pteridine transport)	PF03092.16	6	18.7	1.53E-05
Ligand-gated ion channel	PF00060.26	5	61.5	4.37E-15
SecA DEAD-like domain	PF07517.14	5	48.2	3.80E-12
Tubulin	PF00091.25	5	16.5	4.71E-05
Sugar transporters	PF00083.24	4	9.8	1.78E-03
Ribosomal subunit L36e	PF01158.18	3	41.8	1.03E-10
Ribosomal subunit L44	PF00935.19	3	41.8	1.03E-10

Table 4.5 – Significantly enriched Pfam domains in genes upregulated in *Herpetomonas muscarum* at six hours post ingestion by *Drosophila melanogaster* (vs log-phase axenic culture). The table shows the top 10 represented Pfam domains. Chi-squared tests were performed to test for statistically significant enrichment of the Pfams frequency in upregulated genes vs. the Pfams in the whole genome.

Pfam description	Pfam ID	Frequency	Chi-squared value	p-values
Papain family cysteine protease	PF00112.23	11	110.18	8.94E-26
Triglyceride lipases	PF01764.25	7	20.08	7.43E-06
Biotin/lipoate A/B protein ligase family	PF03099.19	5	54.12	1.89E-13
Cytomegalovirus UL20A protein domain	PF05984.12	4	22.62	1.97E-06
Forkhead-associated domain	PF00498.26	3	17.28	3.22E-05
Ig-like, plexins, transcription factors	PF01833.24	3	11.43	7.23E-04
Glucokinase	PF02685.16	2	30.79	2.88E-08
Initiation factor eIF2 gamma, C terminal	PF09173.11	2	20.90	4.83E-06
Abnormal spindle-like microcephaly-assoc'd, ASPM-SPD-2-Hydin	PF15780.5	2	20.90	4.83E-06
Aldehyde dehydrogenase family	PF00171.22	2	11.97	5.39E-04

Table 4.6 - Significantly enriched Pfam domains in genes downregulated in *Herpetomonas muscarum* at six hours post ingestion by *Drosophila melanogaster* (vs log-phase axenic culture). The table shows the top 10 represented Pfam domains. Chi-squared tests were performed to test for statistically significant enrichment of the Pfams frequency in downregulated genes vs. the Pfams in the whole genome.

Leucine-rich repeat proteins. The most represented Pfam domain in genes upregulated at 6 hours post ingestion were the leucine-rich repeat (LRR) domains. As previously discussed, LRRs are primarily known to be involved in protein-protein and protein-glycolipid interactions and are the major domain in the *Leishmania* protein surface antigens (PSAs), which are known virulence factors. Ten of the upregulated LRR-containing genes encode orthologues of the *Leishmania* PSAs (Figure 4.4). The predicted

protein structures for 8/10 of these transcripts consists of a single transmembrane domain at the N-terminus, with the majority of the protein predicted to be on the external face of the cell (Table S4.7). One transcript encodes a protein with no predicted transmembrane domains and could therefore be a secreted protein. The remaining transcript encodes a protein with two predicted transmembrane domains, with the region between these domains on the external face of the cell. The remaining upregulated LRR-containing transcripts are putative adenylate cyclases. As previously discussed, these proteins also feature prominently in the *T. brucei* genes which are differentially regulated upon ingestion by tsetse (Savage *et al.*, 2016). These genes likely assist in the coordination of the trypanosomatids' responses to the changing environment.

Cell surface genes. Seven of the top fifteen genes, 21/346 overall, upregulated in *H. muscarum* at six hours post ingestion by *D. melanogaster* contained the *Giardia* variant-specific surface protein (VSP) domain (PF03302.13). These genes are members of three orthogroups: OG11, OG450 and OG6613 (see Table S2.3). A heatmap showing the normalised read counts for these genes across all samples is shown in Figure 4.5.

Transmembrane domain prediction tools (Krogh *et al.*, 2001; Käll, Krogh and Sonnhammer, 2004) were able to predict a single transmembrane domain at the N-terminus in the majority of predicted protein sequence for these genes. However, there were also eight transcripts without predicted transmembrane domains, which are predicted to be secreted proteins. The majority of these putative surface antigens are 769-781 amino acids in length, have a single predicted transmembrane helix at residues 7-29 (Table S4.7). As previously mentioned, many of these proteins are also upregulated by the cells upon entry into stationary phase, though not to the same levels.

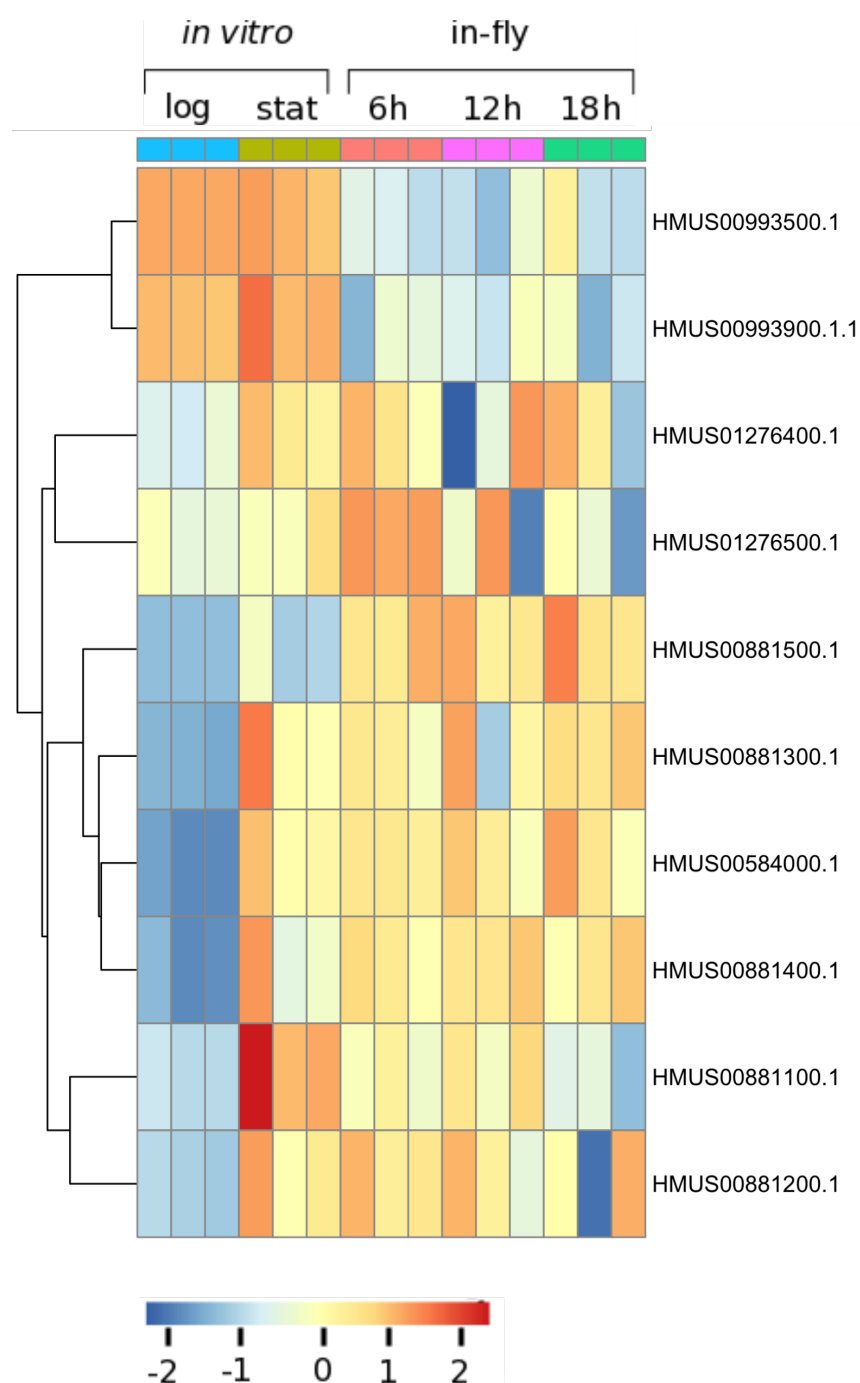


Figure 4.4 – Heat map of normalised, log transformed counts for *Herpetomonas muscarum* orthologues to the *Leishmania* promastigote surface antigens. Log = log phase axenic culture samples, Stat = stationary phase axenic culture samples. 6h = six hours post ingestion by *D. melanogaster*, 12h = twelve hours post ingestion by *D. melanogaster*, 18h = eighteen hours post ingestion by *D. melanogaster*.

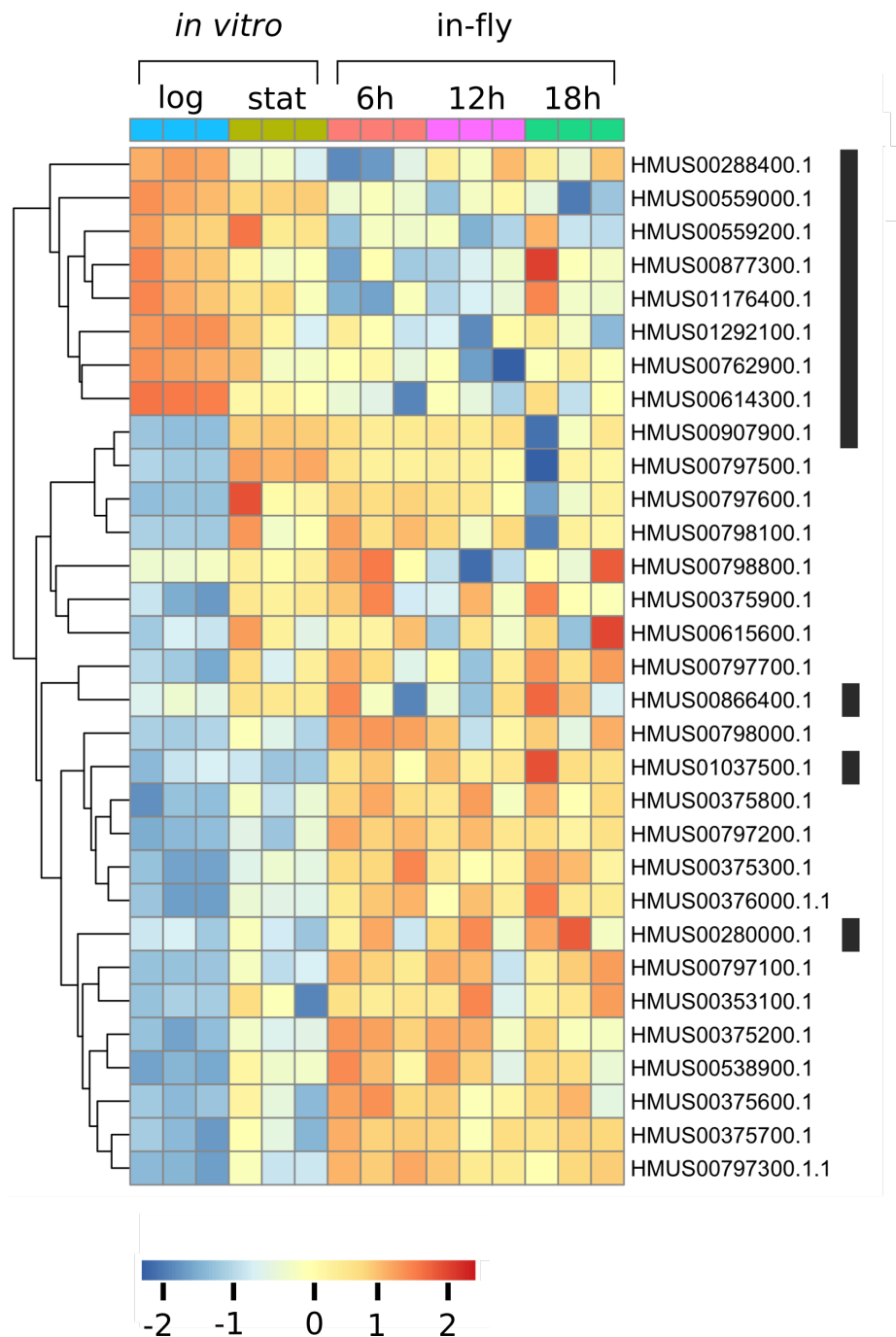


Figure 4.5A - Heat map of normalised, log transformed counts for differentially expressed *Herpetomonas muscarum* proteins with a *Giardia* variant surface protein (PF03302.13) domain. The black bar indicates the genes from orthogroup 11 which are mostly downregulated upon ingestion of *H. muscarum* by the fly. Log = log phase axenic culture samples, Stat = stationary phase axenic culture samples. 6h = six hours post ingestion by *D. melanogaster*, 12h = twelve hours post ingestion by *D. melanogaster*, 18h = eighteen hours post ingestion by *D. melanogaster*.

Additionally, several transcripts for VSP-containing proteins are downregulated in *H. muscarum* upon entry into the fly. These thirteen proteins are generally smaller than those upregulated at the same time point (95-501 amino acids) and tended to be part of orthogroup 11.

Thirty amastins, from 11 different orthogroups, were differentially regulated in *H. muscarum* at 6 hours post ingestion (Figure 4.6). The majority (21) were upregulated upon entry into the fly, though 14 transcripts were also upregulated during stationary phase *in vitro* culture. Each orthogroup represented contained both up- and down-regulated genes. Amastins are a family of glycoproteins made up of approximately 200 amino acids grouped into four subfamilies (α , β , δ , γ) with distinct molecular structures (Jackson, Vaughan and Gull, 2006). Generally, they have four transmembrane domains separated by serine-threonine rich extracellular domains (Rochette *et al.*, 2005). Their functions are largely unknown. They are most commonly associated with amastigote forms of *Leishmania* and implicated in interaction with macrophages. RNAi knock down of delta-amastins reduced survival and virulence in *Leishmania* during macrophage infections (de Pavia *et al.* 2015). The same study also showed ultrastructural changes and microtubule disorganisation at the cell membrane of both organisms. There were also fewer regions of contact between *Leishmania* and the macrophages.

However, it has also been shown that β -amastins are upregulated during the insect stages of the life cycle in *T. cruzi* (Kangussu-Marcolino *et al.*, 2013). The *H. muscarum* amastins from orthogroup 18 share only 25-30% identity (across the whole sequence) to the two pairs of *T. cruzi* β -amastin alleles highlighted in this study. This may initially seem to be quite low, however the β -amastins have been shown to be highly divergent

(18-25% identity) between *T. cruzi* strains (Kangussu-Marcolino *et al.*, 2013). Therefore, based on sequence alone, it is unclear which proteins may have parallel roles in the two trypanosomatid species. From the aforementioned work by de Pavia *et al.* (2015), it is tempting to hypothesize the *H. muscarum* amastins may be involved in cell-cell contacts with the host, its microbiome or between *H. muscarum* cells. While this is speculative, microscopy of *H. muscarum* cells in the crop of *D. melanogaster* six hours after ingestion showed that cells were tightly clustered together in one area close to the crop wall (Figure 3.6) - as such there is likely cell-cell contact during at this time point.

Several other classes of surface protein genes were differentially expressed between log-phase axenic culture and six hours post-ingestion. Transcripts for proteins containing the Cytomegalovirus UL20A protein domain (PF05984) were significantly down regulated upon ingestion (Figure 4.7). Five of these genes were from orthogroup 11 – the same group as many of the down regulated VSP domain containing genes.

Finally, sixteen (of the twenty-eight in the genome) *H. muscarum* orthologues to known *Leishmania* virulence factor, GP63, were significantly differentially regulated in the first six hours post ingestion by the fly. All but one of the differentially regulated GP63 orthologues were predicted to be GPI anchored at the cell surface (GPI-SOM online tool, Fankhauser and Mäser 2005). The exception, HMUS00892600.1, is predicted (THTMM v2.0, Krogh *et al.* 2001) to have a single transmembrane domain and the majority of the protein to be cytosolic. Most GP63 transcripts were upregulated in *H. muscarum* after ingestion (log2 foldchanges 0.29-2.73), however two putative GP63 genes, HMUS01311000 and HMUS01311200, were downregulated with log2 foldchanges of -1.94 and -1.58 respectively.

Stress-related genes. The insect gut is a hostile environment. The presence of digestive enzymes, changes in pH and the insect's gut microbiota make surviving a difficult challenge for any invading organisms. In correlation with this, a number of stress-associated genes and pathways are upregulated in *H. muscarum* upon entry into the fly. As previously mentioned, autophagy is an important process for survival in stressful conditions where fewer nutrients are available - such as in the midgut of an insect. Similar to observed in stationary phase axenic culture, twenty-six putative ATG8 genes were upregulated in *H. muscarum* at six hours post ingestion compared to log-phase axenic culture – suggesting extensive protein recycling is occurring in the cells.

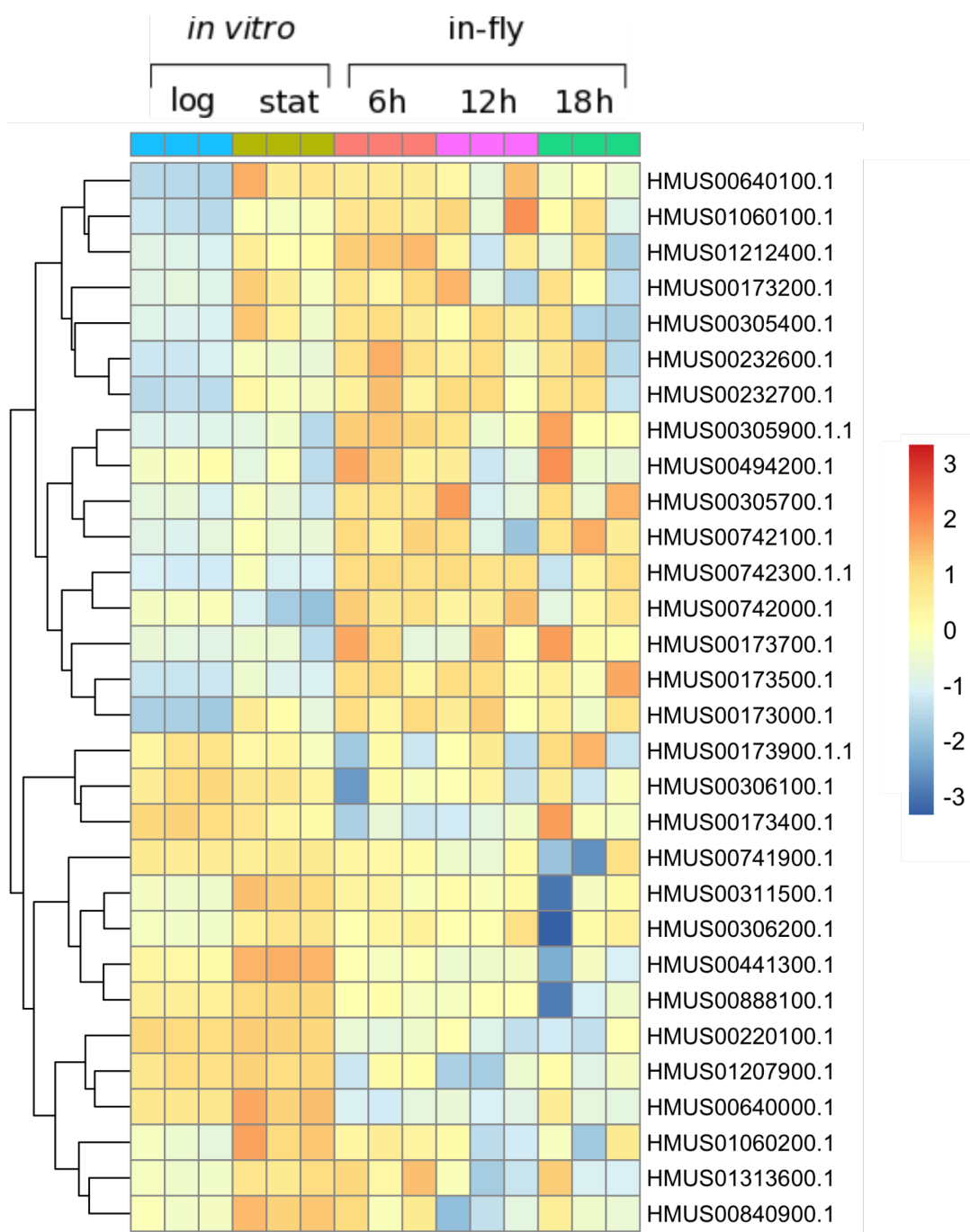


Figure 4.6 - Heat map of normalised, log transformed counts for differentially expressed *Herpetomonas muscarum* amastin genes. Log = Log phase axenic culture samples, Stat = stationary phase axenic culture samples. 6h = six hours post ingestion by *D. melanogaster*, 12h = twelve hours post ingestion by *D. melanogaster*, 18h = eighteen hours post ingestion by *D. melanogaster*.

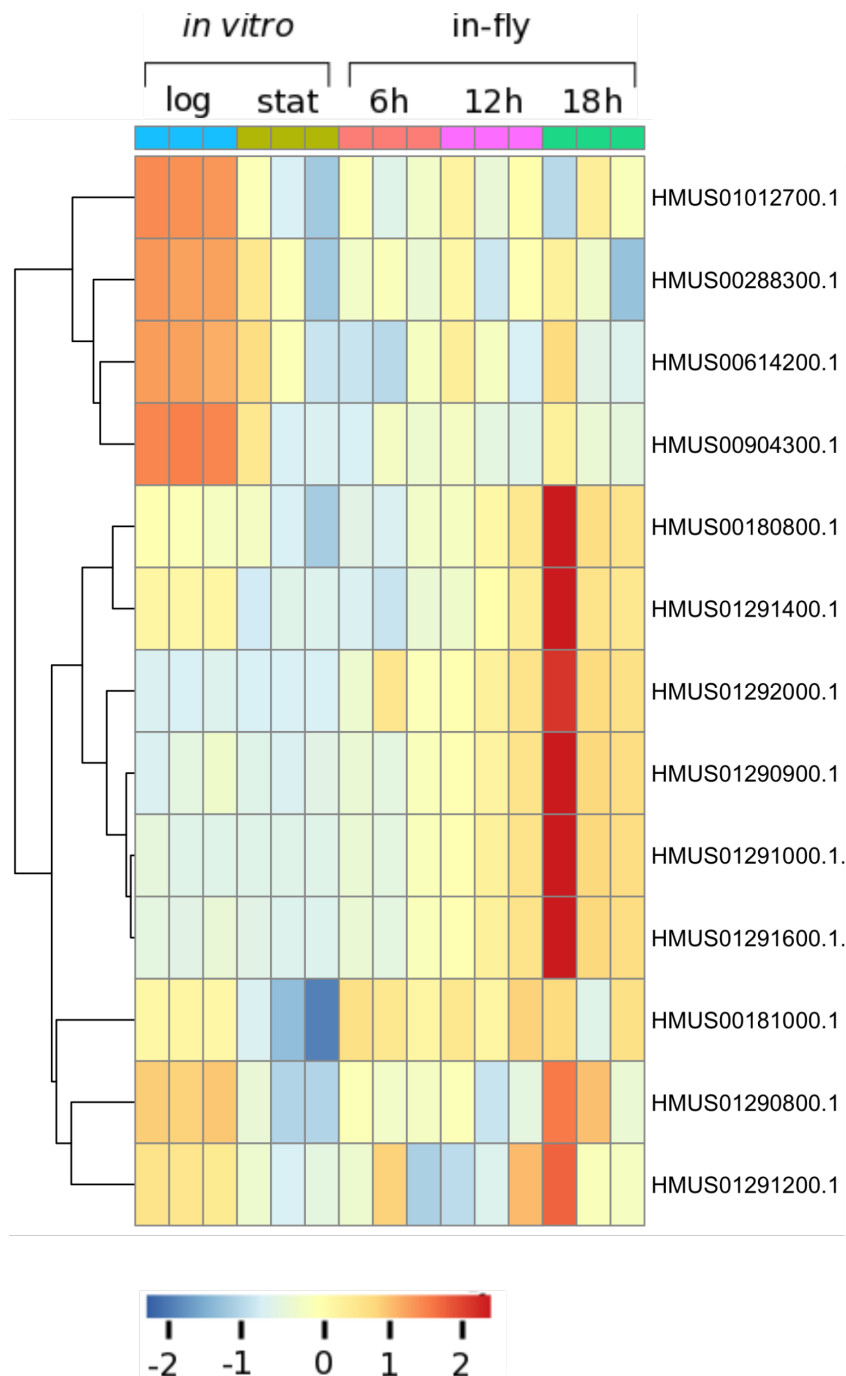


Figure 4.7 - Heat map of normalised, log transformed counts for differentially expressed *Herpetomonas muscarum* proteins with a Cytomegalovirus UL20A (PF05984.12) domain. Log = log phase axenic culture samples, Stat = stationary phase axenic culture samples. 6h = six hours post ingestion by *D. melanogaster*, 12h = twelve hours post ingestion by *D. melanogaster*, 18h = eighteen hours post ingestion by *D. melanogaster*.

Additionally, 40 transcripts for heat shock protein 83 were shown to be upregulated at six hours after ingestion. Heat shock proteins act as molecular chaperones which stabilise other proteins, help them to fold correctly and be regulated after damage in stressful conditions. The upregulation of these genes provides further evidence that these cells are in a stressed state.

Metabolism. There was significant enrichment of putative amino acid, pteridine and sugar transporters in the upregulated transcripts. These included the amino acid transporters (AATs) orthologous to the *Leishmania* amino acid permease 3 (AAP3), AAT11, AAT12 and AAT20. AAP3 has been shown to be arginine specific and is linked to virulence in *Leishmania donovani* infections in humans (Darlyuk *et al.*, 2009). AAT11 is upregulated in during stress responses associated with response to purine starvation (Marchese *et al.*, 2018). In *Leishmania major*, AAP3 and AAT20 were strongly upregulated in nectomonad forms (Inbar *et al.*, 2017). These transporters have been shown to transport neutral amino acids across the cell membrane, notably proline and alanine, which can be used as alternative carbon sources by trypanosomatids and are abundant in insect vectors haemolymph.

Six putative pteridine transporters were also upregulated in *H. muscarum* at 6 hours post ingestion. Pteridines are needed by trypanosomatids to produce enzyme cofactors such as bipterin. *Leishmania* parasites are unable to synthesize their own pteridines (Cunningham and Beverly 2001) and as such must scavenge them from their environment. It is not currently known if *H. muscarum* is also a pteridine auxotroph, however like the *Leishmania* species, the cells appear to scavenge from the environment upon entry into the fly.

Several transcripts putatively involved in lipid metabolism were downregulated in *H. muscarum* following ingestion by *D. melanogaster*, including triglyceride lipases and members of the biotin/lipoate protein ligase (BLPL) family. This contrasts what has been observed in *L. major* in the midgut of sand flies where genes from these families were upregulated. Therefore, whilst upregulation of pteridine and amino acid transporters appears to be a conserved trypanosomatid response to being ingested by insects, lipid metabolism during insect infection may differ between trypanosomatid genera.

Gene expression-related transcripts. Consistent with the differential expression of many genes upon entry into the fly, and therefore a predicted increase in chromatin remodelling and translation activity, there was upregulation of histones (2A, 3 and 4), RNA polymerase subunits 1 and 2, putative 40S/60S ribosomal proteins and putative 28S beta rRNAs in *H. muscarum* after ingestion by the fly. This result is consistent with what has been reported in other *T. brucei* where the 40S and 60S ribosomal subunits were amongst the most highly upregulated genes in cells isolated from the midgut and proventriculus of *G. morsitans* (Savage *et al.* 2016).

Cell cycle-associated genes. Upon ingestion by the fly there was strong upregulation of putative G₁-associated cyclins 4, 7 and 11 as well as the G₁ associated cyclin-related kinase 1 (CRK1) (Tu and Wang, 2004, 2005). Cyclin 6, cyclin 8 and CRK9, which are associated with the G₂/mitosis transition (Hammarton *et al.*, 2003; Gourguechon and Wang, 2009), were slightly downregulated suggesting a reduction in cell replication at six hours post ingestion (Table 4.7). I also observed the downregulation of S-phase associated RNA-binding protein PUF9 (*Drosophila* Pumilio and *C. elegans* Fem-3-binding factor domain 9) transcripts which further suggests a reduction, or lack of,

replication of *H. muscarum* after ingestion by *D. melanogaster* (Archer *et al.* 2009). Furthermore, there was also downregulation of putative DNA polymerase kappa, the theta DNA polymerase subunit, mitochondrial DNA polymerase subunits and the meiosis-associated genes NBS1, Rad50 and SPO11.

Given the apparent reduction replication rate in *H. muscarum* cells at six hours after ingestion, the upregulation of nine tubulin genes (3 α - and 6 β -tubulins) is likely to accommodate the changes in cell morphology, rather than to produce new daughter cells. This is consistent with my observations of *H. muscarum* in dissected *D. melanogaster* digestive tracts. I observed that cells had a more rounded cell body in the fly than during log phase axenic culture (Figure 3.6). Tubulin upregulation is also observed in *T. brucei* isolated from the midgut and proventriculus of *Glossina morsitans* (Savage *et al.*, 2016), though these cells are replicative – as such the ‘motivation’ for increased tubulin gene expression may be different.

Differentiation and RNA-binding proteins. It is well documented that (human) disease-causing trypanosomatids have several life cycle stages within their respective vectors. Coordinated differentiation between these discrete stages requires a suite of RNA-binding proteins (RBPs) which regulate parasite gene expression (Kolev *et al.* 2014). Despite the lack of observed differentiated forms in infections of *D. melanogaster* (Chapter Three), the *H. muscarum* genome contains an almost-complete (relative to *T. brucei*) repertoire of these differentiation regulating RBPs (Chapter Two). As such, I investigated the expression pattern of differentiation-associated RBPs in *H. muscarum*. I was surprised to find that orthologues, RBP10 and hnRNP F/H, were upregulated in *H. muscarum* after ingestion by the fly (Wurst *et al.* 2012, Gupta *et al.* 2013). These proteins have been

shown to regulate gene expression in *T. brucei* blood-stream forms (Wurst *et al.* 2012, Gupta *et al.* 2013). This was intriguing, given apparent lack of a blood-stage form in *H. muscarum*. Three out of the four orthologues of *TbRBP10* were strongly (> 4-fold) upregulated in *H. muscarum* cells after ingestion by the fly. RNAi knockdown of RBP10 in bloodstream trypanosomes resulted in the down-regulation of a large number of bloodstream form mRNAs (Wurst *et al.* 2012). The same study showed that overexpression of the protein in procyclics led to an increase of many bloodstream-form specific mRNAs, including genes involved in sugar transport. This is likely owing to the fact blood is a glucose-rich environment and the cell will attempt to utilize this ready carbon source (Smith *et al.* 2017). During feeding experiments sucrose is added to the *H. muscarum* culture media to encourage the flies to feed. As such these genes may be unregulated in response to increased sugars available in the environment.

However, several other cell-cycle regulating RBPs associated with blood-stream form trypanosomes were also upregulated in *H. muscarum* after ingestion by the fly, including zinc-finger domain-containing RBPs ZC3H11 and ZC3H18. The former is essential in bloodstream-form trypanosomes and is involved in protection from heat shock, whilst depletion of ZC3H18 delayed blood-stream-to-procyclic differentiation in *T. brucei* (Benz *et al.* 2011, Droll *et al.* 2013). As such the situation may be more complex than solely metabolism-driven expression changes.

In addition to parallels with blood-stream trypanosomes, transcripts for ALBA3/4 proteins (named for their ‘acetylation lowers binding affinity’ domain) were significantly downregulated in *H. muscarum* upon entry into the fly. In trypanosomes, these proteins are expressed in all stages, except those found in the tsetse proventriculus. RNAi

knockdown of these proteins in *T. brucei* axenic procyclics resulted in elongation of the cell body and repositioning of the nucleus and the kinetoplast to resemble the epimastigote cell-stage (Subota *et al.* 2011). As such the reduction in ALBA3/4 transcripts suggests there may be parallels between trypanosomes during the latter stages of tsetse infection and *H. muscarum* during *D. melanogaster* infection.

Other differentially regulated RNA-binding proteins with as yet unclear roles in differentiation included: the essential gene expression regulation protein RBP42 and ZC3H12, a protein associated with differentiation (Kolev *et al.* 2014).

Herpetomonas muscarum* genes differentially regulated between six and twelve hours post-ingestion by *Drosophila melanogaster

There were 204 genes which were differentially regulated between six and twelve hours post ingestion (p-adjusted < 0.05), 161 of these had a fold change of ≥ 2 with just 31 genes upregulated at the latter timepoint (Table 4.8). Hypothetical proteins lacking functional information dominated the highly upregulated genes. The most enriched transcript at 12 hours post ingestion encodes a putative surface protein, the top blastp hit for which was the *Giardia* variant-specific surface protein VSP136-4. This suggests VSP domain-containing proteins continue to be important throughout infection of the fly. Two DNA replication and repair associated transcripts were also upregulated at 12 hours post ingestion: an orthologue of *T. brucei* cell division cycle protein 45 (CDC45), and tyrosyl-DNA phosphodiesterase-like protein. CDC45 is part of the CMG (Cdc45·Mcm2–7·GINS) complex which functions as a helicase during DNA replication (Dang and Li, 2011) and may also play a role in DNA repair (DeBrot, Lancaster and Bjornsti, 2016). Furthermore, Tyrosyl-DNA phosphodiesterases are involved in the repair of

topoisomerase-related DNA damage (Kawale and Povirk, 2018). These observations indicate that *H. muscarum* cells are under genotoxic stress after ingestion by *D. melanogaster*.

Gene Name	<i>H. muscarum</i> orthologue ID	log2foldchange	adjusted p-value
cyclin 11	HMUS01322900.1	-3.31	6.54E-27
cyclin 4	HMUS00787500.1	-1.15	3.62E-13
CRK4	HMUS00195900.1	-0.95	3.46E-03
CRK1	HMUS01116400.1	-0.84	9.95E-08
CRK8	HMUS00385600.1	-0.49	2.32E-02
cyclin 7	HMUS00475100.1	-0.44	2.21E-02
cyclin 8	HMUS00524500.1	0.36	1.94E-02
mitochondrial DNA polymerase I protein D	HMUS00617400.1	0.57	9.16E-03
cyclin 6	HMUS00719100.1	0.74	2.14E-02
cyclin 5	HMUS00580100.1	0.85	9.34E-05
CRK9	HMUS01274200.1	0.87	1.45E-03
DNA polymerase theta catalytic subunit	HMUS00097200.1	1.15	1.51E-07
mitochondrial DNA polymerase I protein C	HMUS00828800.1	1.25	7.27E-09
DNA polymerase kappa	HMUS01207400.1	1.36	4.93E-03
CRK11	HMUS00452900.1	1.46	8.20E-04
CRK12	HMUS00986000.1	2.07	6.68E-18

Table 4.7 – Cell cycle-associated proteins differentially expressed in *Herpetomonas muscarum* upon ingestion by *Drosophila melanogaster*. Fold changes show are at 6 hours post ingestion compared to log phase axenic culture.

Herpetomonas muscarum* genes differentially regulated between twelve and eighteen hours post-ingestion by *Drosophila melanogaster

In the 23 genes found to be upregulated at 18 hours post ingestion (compared to at 12 hours, Table 4.9) genes involved in binding to damaged DNA (OG00033330) were significantly enriched. Only two of these transcripts were able to be assigned putative functions: eukaryotic replication factor A and a structure specific endonuclease. This observation provides further evidence of genotoxic stress in *H. muscarum* after ingestion, as indicated by other upregulated DNA repair genes at 12 hours post-ingestion.

The most highly upregulated transcript at 18 hours post ingestion was an orthologue of the *Leishmania major* UDP-galactose transporter LPG5B. This protein allows import of UDP-galactose into the golgi body where they are used to synthesize phosphoglycans. Capul *et al.* (2007) showed that, in *L. major*, loss of LPG5b resulted in cells with defects in proteophosphoglycans (PPG) (Capul *et al.*, 2007). PPGs are known virulence factors and are found in membrane bound, filamentous and secreted forms. As such, upregulation of putative PPG synthesis pathway genes is of great interest. It is possible that *H. muscarum* may secrete a similar PPG gel to that of *Leishmania*, however currently this is speculative and further study is required. I did observe rosettes of *H. muscarum* cells with ‘exclusion zones’ (Figure 3.7) which could be a PPG-like gel. These rosettes could be probed with available anti-*Leishmania* PPG antibodies to test this hypothesis.

<i>H. muscarum</i> gene ID	fold change	adjusted p-value	Putative function
HMUS00288400.1	5.08	2.12E-02	Giardia variant surface protein
HMUS00718700.1	3.98	1.72E-02	hypothetical protein
HMUS00177100.1	3.76	8.65E-03	tyrosyl-DNA phosphodiesterase-like protein
HMUS01087000.1	3.64	4.87E-03	hypothetical protein
HMUS00515400.1	3.56	2.38E-02	hypothetical protein
HMUS01269600.1	3.46	2.96E-03	hypothetical protein
HMUS01173100.1	3.45	1.72E-03	CDC45
HMUS00835400.1	3.25	2.96E-03	hypothetical protein
HMUS00814800.1	3.04	4.84E-02	hypothetical protein
HMUS01067800.1	2.94	5.28E-03	hypothetical protein
HMUS00191000.1	2.93	4.03E-02	RING-H2 zinc finger domain containing protein
HMUS00888000.1	2.87	4.39E-02	LC8-type dynein light chain
HMUS00040500.1	2.85	1.72E-02	hypothetical protein
HMUS01076600.1	2.75	1.15E-02	hypothetical protein
HMUS01216900.1	2.71	1.06E-02	hypothetical protein
HMUS01258800.1	2.65	1.03E-02	adenylate kinase 2
HMUS00333000.1	2.63	4.01E-02	hypothetical protein
HMUS00672100.1	2.63	7.76E-03	hypothetical protein
HMUS01324700.1	2.63	3.92E-02	hypothetical protein
HMUS00663800.1.1	2.61	4.41E-02	hypothetical protein
HMUS00221400.1	2.60	1.09E-02	dephospho-CoA kinase
HMUS00191800.1	2.56	4.17E-02	hypothetical protein
HMUS00161000.1	2.52	3.41E-02	hypothetical protein
HMUS00397500.1	2.51	3.38E-02	no significant hits
HMUS00368900.1	2.45	2.02E-02	hypothetical protein
HMUS01089800.1	2.41	3.53E-02	COP9 signalosome complex subunit 4
HMUS00752900.1	2.36	2.37E-03	GDP-L-fucose synthase
HMUS00882400.1	2.33	1.72E-02	molybdenum cofactor biosynthesis protein C
HMUS00364400.1	2.25	1.70E-03	1,2-Dihydroxy-3-keto-5-methylthiopentene dioxygenase
HMUS00836800.1	2.18	4.92E-02	adenosine 5'-monophosphoramidase
HMUS00862500.1	2.04	3.53E-02	RNA triphosphatase

Table 4.8 – *Herpetomonas muscarum* gene transcripts significantly upregulated at 12 hours post ingestion (vs. 6 hour post ingestion).

***Herpetomonas muscarum* genes differentially regulated between stationary phase *in vitro* culture and in-fly samples**

Comparisons between stationary phase *in vitro* culture and in-fly samples revealed 5102 differentially expressed genes (adjusted p-value < 0.05). Approximately 55% of the genes differentially regulated between *in vitro* and in-fly samples were the same for log phase vs in-fly and stationary phase vs. in-fly comparisons (Figure 4.8). However, 1639 genes were only significantly differentially regulated in stationary phase vs in-fly comparisons (Table S4.8). Genes differentially regulated between log phase *in vitro* culture and in-fly samples have already been discussed, I will now outline the genes only differentially regulated when the transcriptomes of stationary phase *in vitro* samples of *H. muscarum* are compared with those after ingestion by *D. melanogaster*. Of the 1639 genes, 750 had a fold change of ≥ 2 – approximately a third of which were upregulated in *H. muscarum* after ingestion by *D. melanogaster*.

Half of the top ten in-fly enriched transcripts were telomere associated mobile elements (TATE) DNA transposons and among the most represented Pfam hits in the fly-enriched transcripts were reverse transcriptase (PF00078.27) and phage integrase (PF00589.22) domains (Table 4.10). Though TATE DNA transposons comprise 1.32% of the *Leishmania major* genome, very little is known about these transposable elements, other than that they contain a tyrosine recombinase (Pita *et al.*, 2019). It is possible that these transposable elements are more mobile in cells ingested by the fly, however it may also be representative of predicted reductions in transcription in cells reaching stationary phase in axenic culture.

<i>H. muscarum</i> gene ID	Fold change	adjusted p-value	Putative function
HMUS00969300.1	19.89	3.20E-02	UDP-galactose transporter (LPG5B)
HMUS00674800.1	15.86	2.13E-02	Zinc knuckle
HMUS01014100.1	9.96	1.18E-03	chaperone DnaJ protein
HMUS00310500.1	6.70	1.11E-02	myo-inositol-1(or 4)-monophosphatase 1
HMUS00437100.1	6.09	5.43E-03	structure-specific endonuclease
HMUS01313800.1	5.78	5.43E-03	conserved hypothetical
HMUS00184000.1.1	5.40	3.20E-02	conserved hypothetical
HMUS00542900.1	5.28	2.42E-04	ribosomal protein L2
HMUS01053400.1	4.68	6.99E-03	conserved hypothetical
HMUS00338800.1	4.12	4.98E-02	no significant hits
HMUS00969100.1	4.10	2.86E-03	tubulin tyrosine ligase protein
HMUS00838700.1	4.03	1.40E-02	conserved hypothetical
HMUS00770900.1	3.76	4.79E-02	conserved hypothetical
HMUS00758800.1	3.73	3.69E-02	conserved hypothetical
HMUS00473100.1.1	3.67	1.18E-03	conserved hypothetical
HMUS00008700.1	3.53	3.38E-02	putative protein phosphatase
HMUS00309000.1	3.32	4.63E-02	organic solute carrier partner 1
HMUS00546600.1	3.11	3.20E-02	tetratricopeptide repeat protein
HMUS00572300.1	3.03	4.09E-02	membrane transporter
HMUS00739400.1	3.00	3.08E-02	conserved hypothetical
HMUS00922000.1	2.95	3.08E-02	endonuclease/exonuclease/phosphatase
HMUS00669200.1.1	2.84	1.40E-02	endosomal integral membrane protein
HMUS01054800.1	2.41	4.98E-02	aminomethyltransferase
HMUS00023100.1	2.11	3.47E-02	conserved hypothetical
HMUS00788100.1	0.30	4.09E-02	Kinesin-like protein

Table 4.9 - *Herpetomonas muscarum* gene transcripts significantly upregulated at 18 hours post ingestion (vs. 12 hour post ingestion).

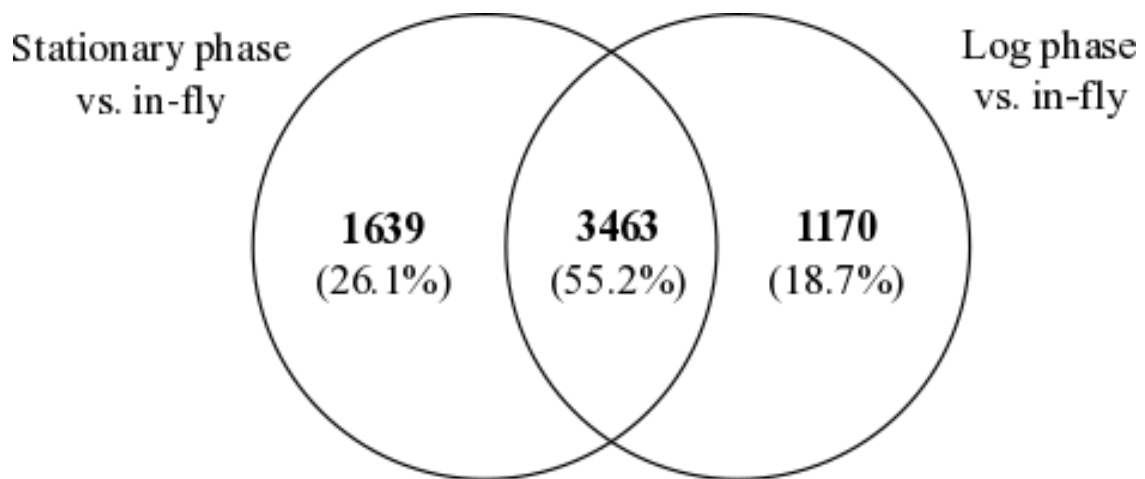


Figure 4.8 – Venn diagram showing the numbers of genes differentially expressed in *Herpetomonas muscarum* between two *in vitro* culture conditions and after ingestion by *Drosophila melanogaster*.

As previously discussed, transcripts for several proteins containing a *Giardia* VSP domain are enriched in stationary phase compared to log phase in axenic culture. However, five were shown to be even more abundant in the *H. muscarum* cells ingested by *D. melanogaster*. Two other putative surface antigens were also enriched in ingested *H. muscarum* which contained a domain similar to Cytomegalovirus UL20A glycoprotein and the domain of unknown function DUF4148.

Pfam description	Frequency	Pfam Accession	Chi-squared	p-value
Reverse transcriptase 1	6	PF00078.27	8.93	0.002811
Giardia variant surface protein	5	PF03302.13	0.57	0.450115
Thioredoxin	4	PF00085.20	86.36	1.50E-20
Histone	4	PF00125.24	15.62	7.74E-05
Glutathione S-transferase, C-terminal domain	3	PF00043.25	19.54	9.83E-06
emp24/gp25L/p24/GOLD domain family	3	PF01105.24	12.72	3.62E-04
alkyl hydroperoxide reductase (AhpC) and thiol specific antioxidant (TSA)	3	PF00578.21	11.57	6.70E-04
Tubulin	3	PF00091.25	16.44	5.02E-05
Phage integrase	3	PF00589.22	2.01	1.57E-01
Ulp1 protease family, C-terminal catalytic domain	3	PF02902.19	2.01	1.57E-01

Table 4.10 – The top 10 represented Pfam domains in *Herpetomonas muscarum* ingested by *Drosophila melanogaster* vs. stationary phase axenic culture.

Transcripts encoding for putative anti-oxidant proteins were significantly enriched in *H. muscarum* after ingestion by the fly. Enriched Pfam domains in the upregulated gene set included thioredoxin, glutathione S-transferase and alkyl hydroperoxide reductase (AhpC)/thiol specific antioxidant (TSA) domains. Previous work in the Ligoxygakis lab has shown that the *D. melanogaster* response to *H. muscarum* ingestion included the production of reactive oxygen species (Wang, Sloan and Ligoxygakis, 2019), as such the upregulation of these anti-oxidant proteins is likely an attempt to cope with this insect immune response.

Does the Herpetomonas muscarum response to ingestion by Drosophila melanogaster resemble the response of medically important trypanosomatids in their vectors?

For any model to be useful it must be representative, to some extent, of what is happening ‘in the field’. As such I wanted to directly compare the *H. muscarum* response to ingestion by *D. melanogaster* to that of a medically important trypanosomatid in its respective vector. To this end, I compared the *H. muscarum* in-fly transcriptome recently published datasets for *Leishmania major*, *Leishmania mexicana* and *Trypanosoma brucei*. I took the lists of differentially expressed genes (DEGs) from each study and, using the table of orthologues generated in Chapter Two (Table S2.3), I identified genes which had a single-copy orthologue in *H. muscarum* (scoDEGs). Whilst this maximises the chances of only comparing ‘true’ orthologues, this excludes members of the larger gene families, many of which are of great interest e.g. the VSP domain-containing genes described earlier in this chapter. For these genes, I compared the direction of the fold changes in the aforementioned trypanosomatids with the fold change *H. muscarum* between *in vitro* culture and all in-fly samples.

Comparisons with *Leishmania major* in *Phlebotomus duboscqi*

Inbar *et al.* (2017) performed an RNA-sequencing study on the transcriptomes of four distinct *L. major* lifecycle stages during infection of a natural vector species of sand fly *Phlebotomus duboscqi*. The group dissected infected sand flies at two, four and fifteen days after feeding with *L. major* amastigotes which had been derived from the footpads of infected mice. The midguts which were found to contain homogenous populations of; procyclic, nectomonad or metacyclic promastigotes were collected, pooled and had the total RNA extracted for sequencing. The group then performed pairwise comparisons of

differentially expressed genes in the different lifecycle stages. The DEGs I observe in *H. muscarum* after ingestion by *D. melanogaster* most closely resemble those upregulated in nectonomads compared with procyclic promastigotes with 72.6% of scoDEGs with the showing the same fold change pattern in both cases (Table 4.11). There were also strong similarities between the *H. muscarum* DEGs and genes upregulated in metacyclic cells in comparison to procyclic promastigotes (52.6% of scoDEGs). The full lists of scoDEGs shared between *H. muscarum* and the *L. major* transcriptomes are in Tables S4.9-S4.11).

	<i>Leishmania major</i> life cycle stage		
	Procyclic promastigotes (early midgut)	Nectonomads (late midgut)	Metacyclics (stomodeal valve)
Total number of <i>L. major</i> DEGs from Inbar <i>et al.</i> (2017)	1213	562	471
DEGs with single-copy <i>H. muscarum</i> orthologues (scoDEGs)	383	190	135
Percentage of DEGs with single-copy <i>H. muscarum</i> orthologues (scoDEGs)	31.6%	33.8%	28.7%
scoDEGs with matching fold change	128	138	71
scoDEGs with non-matching fold change	255	52	64
Percentage scoDEGs with matching fold change	33.4%	72.6%	52.6%

Table 4.11 – Summary of co-regulated genes in *Herpetomonas muscarum* after ingestion by *Drosophila melanogaster* and *Leishmania major* after ingestion by *Phlebotomus duboscqi*. The first data row of the table shows the number of differentially expressed genes (DEGs) in 3 major life cycle stages procyclic promastigotes (vs. amastigotes), nectonomads (vs. procyclics) and metacyclics (vs. nectonomads) from Inbar *et al.* (2017). Subsequent table rows are generated by comparing genes differentially expressed in *H. muscarum* after ingestion by *D. melanogaster* to the data from Inbar *et al.* (2017).

Of the 128 *H. muscarum* genes with a matching fold change in *L. major* nectonomads, 16 gene transcripts are upregulated in both trypanosomatids. The majority lack functional information, though amino acid transporter AAT15 was included in the co-upregulated transcripts. Transcripts downregulated in both *L. major* nectonomads and *H. muscarum* were dominated by DNA replication-associated proteins including: histones, proliferative cell nuclear antigen (PCNA), DNA-replication licensing factor MCM5 and the mitochondrion-associated enzymes DNA ligase α and mitochondrial DNA polymerase. Furthermore, both species downregulated cell cycle proteins cyclin 6 and RBP33. Knock down of cyclin 6 in *T. brucei* was shown to cause mitotic arrest and cessation of growth (Hammarton, 2007). These observations suggest that both *L. major* neptomonad cells and *H. muscarum* at six hours post-ingestion reduce cell replication in their respective insects. In contrast to the *L. major* nectonomads, cytochrome c oxidase subunit VI putative (COX6) was upregulated in *H. muscarum*. Interestingly, COX6 is also upregulated in *L. mexicana* promastigotes cultured *in vitro* and *T. brucei* midgut forms (discussed below).

Comparisons with *Leishmania mexicana*

Holzer *et al.* (2006) described 288 genes which were differentially expressed between *Leishmania mexicana in vitro* cultured promastigotes and amastigotes derived from mice lesions (Holzer, McMaster and Forney, 2006). The group used high-density, whole-genome oligonucleotide microarrays based on the *L. major* genome – as the open reading frames from the two species share high sequence identity on some chromosomes (>90%). The group found 208 genes which were enriched in promastigotes compared to lesion-derived amastigotes, approximately half of which have single copy orthologues in *H. muscarum*. Sixty-six of the *H. muscarum* single copy orthologues were differentially

regulated at six hours post ingestion by *D. melanogaster* with 17 genes upregulated in both *H. muscarum* and *L. mexicana* insect stage cells (Table S4.12). Whilst the majority of these genes encode hypothetical proteins, there is also co-enrichment in: an 20S proteasome subunit, cytochrome c oxidase subunit 6, histone2A, glutathione peroxidase, thioredoxin and long-chain-fatty-acid-coA ligase-like protein.

Comparisons with *Trypanosoma brucei* in *Glossina morsitans*

Savage *et al.* (2016) performed an RNA-seq study on *T. brucei* infected midguts, proventriculus and salivary glands of the *Glossina morsitans*. The group then compared the transcriptomes of the trypanosomatids in the three tissues. The surface coat proteins and ribosomal subunits dominated the top differentially expressed genes found in this study. For the top 200 transcripts in *T. brucei* cells in the three tissues, I identified the genes with single copy orthologues in *H. muscarum* that were up regulated in at six hours post ingestion. This identified 12 genes co-enriched in *H. muscarum* and *T. brucei* midgut-dwelling cells, 8 genes co-enriched in *H. muscarum* and *T. brucei* proventriculus-dwelling cells and 6 co-enriched in *H. muscarum* and *T. brucei* salivary gland forms (summarised in Table 4.12, full tables can be found in supplementary tables S4.13-S4.15). As such, it would appear there are more transcriptional parallels between the insect stages of *H. muscarum* with *Leishmania* than *T. brucei*. Cytochrome oxidase subunits, 60S ribosomal L39 subunits, inhibitor of cysteine peptidase and the paraflagellar rod PF16 protein were enriched in *H. muscarum* and all three *T. brucei* forms – though only the cytochrome oxidase subunit VI was upregulated in the salivary gland form. Additionally, cyclin 7 is enriched in both *H. muscarum* and *T. brucei* cells in the proventriculus. Cyclin 7 has been associated with G₁ growth phase and the regulation

of cytoskeleton morphogenesis at the posterior of the cells – knock down of cyclin 7 resulted in ‘branched’ cell posteriors (Liu, Hu and Li, 2013).

	Top 200 upregulated genes in <i>Trypanosoma brucei</i> life cycle stages		
	Midgut	Proventriculus	Salivary Glands
Genes with single copy orthologues	28	20	22
Number of differentially expressed single copy orthologues	15	11	12
Number of genes with a matching fold change	12	8	6
Number of genes with a non-matching fold change	3	3	6
Percentage of genes with a matching fold change	80.0%	72.7%	50.0%

Table 4.12 - Summary of co-regulated genes in *Herpetomonas muscarum* after ingestion by *Drosophila melanogaster* and *Trypanosoma brucei* after ingestion by *Glossina morisitans*.

Concluding remarks

This RNA-sequencing study discussed in this chapter has shown that *H. muscarum* undergoes large scale transcriptomic change in response to ingestion with approximately half of the genome differentially regulated at six hours after ingestion. This transcriptome is comparatively stable for the remaining time in the fly with only 204 and 25 genes differentially expressed at 12 and 18 hours after ingestion. Surface proteins dominated the top differentially regulated genes, with VSP domain-containing protein transcripts, amastins and orthologues to the *Leishmania* PSAs being highly expressed by

Herpetomonas muscarum soon after ingestion by *D. melanogaster*. Transcriptomic changes associated with increased transcription, nutrient stress and autophagy were also observed, much of which was consistent with what is known about the transcriptomes of disease causing trypanosomatids in their dipteran vectors.

More direct comparisons of the *H. muscarum* transcriptome with those of disease-causing trypanosomatids from the *Leishmania* and *Trypanosoma* genera showed that the in-fly *Herpetomonas muscarum* transcriptome most closely resembles that of late-stage midgut infections namely the *Leishmania major* nectonomads and midgut/proventriculus-dwelling *Trypanosoma brucei*. There were several parallels between the trypanosomatid in-insect transcriptomes. Conserved aspects of these transcriptomes included extensive cell surface remodelling and upregulation transcription-associated genes, amino acid transporters and autophagy pathway genes. As such the *H. muscarum* and *D. melanogaster* represents an exciting model to investigate trypanosomatid interactions with the vector midgut.

There is an important caveat to these conclusions – *H. muscarum* mRNAs purified from infected flies can originate from both live and dead cells. Given the context, it is likely that there are many dead/dying *H. muscarum* cells in the *D. melanogaster* gut during the course of infection, and these cells will likely have a different transcriptomic signal to live cells. The effect of ‘dead cell’ bias on the RNA pool is also likely to change over the course of infection, as the ratio of live/dead parasites changes. It will be important to verify that which changes in transcript prevalence I have observed, equate to changes in protein expression in live cells in the *Drosophila* gut. This could be achieved by tagging the proteins of interest either with antibodies or creating genetic mutant lines and then

performing ‘live’ imaging on dissected fly digestive tracts. Alternatively, *H. muscarum* cells could have been purified from infected flies, dead cells removed/depleted, and then only live cells used for RNA-seq analysis.

Finally, this work has identified several genes of interest for future studies (Table S4.16). These proteins are highly expressed and predicted to be located on the surface of the cell. As such they are available for interaction with host factors and could play critical roles in the direct interaction between trypanosomatid and insect. Importantly, the expression patterns of these genes closely resemble those of their orthologues in disease-causing relatives. Manipulation of the expression of these genes and evaluation of the ability of the transgenic cells to infect *D. melanogaster* will identify how essential these genes are to survival in the context of infection. Those that are found to be critical to infection outcome will be of great interest as they could be used as targets for broad spectrum transmission blocking strategies in the future.

References

- Acosta-Serrano, A. *et al.* (2001) ‘The surface coat of procyclic *Trypanosoma brucei*: Programmed expression and proteolytic cleavage of procyclin in the tsetse fly’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 98(4), pp. 1513–1518. doi: 10.1073/pnas.041611698.
- Adam, R. D. (2001) ‘Biology of *Giardia lamblia*’, *Clinical Microbiology Reviews*, 14(3), p. 447 LP-475. doi: 10.1128/CMR.14.3.447-475.2001.
- Aitcheson, N. *et al.* (2005) ‘VSG switching in *Trypanosoma brucei*: antigenic variation analysed using RNAi in the absence of immune selection’, *Molecular microbiology*, 57(6), pp. 1608–1622. doi: 10.1111/j.1365-2958.2005.04795.x.
- Akopyants, N. S. *et al.* (2009) ‘Demonstration of Genetic Exchange During Cyclical Development *Leishmania* in the Sand Fly Vector’, *Science*, 324(5924), p. 265 LP-268. doi: 10.1126/science.1169464.

- Alcolea, P. J. *et al.* (2010) 'Transcriptomics throughout the life cycle of *Leishmania infantum*: High down-regulation rate in the amastigote stage', *International Journal for Parasitology*, 40(13), pp. 1497–1516.
doi: <https://doi.org/10.1016/j.ijpara.2010.05.013>.
- Alcolea, P. J. *et al.* (2014) 'Stage-specific differential gene expression in *Leishmania infantum*: from the foregut of *Phlebotomus perniciosus* to the human phagocyte', *BMC genomics*. BioMed Central, 15(1), p. 849. doi: 10.1186/1471-2164-15-849.
- Alcolea, P. J. *et al.* (2016) 'Influence of the Microenvironment in the Transcriptome of *Leishmania infantum* Promastigotes: Sand Fly versus Culture', *PLOS Neglected Tropical Diseases*. Public Library of Science, 10(5), p. e0004693.
- Añez, N. *et al.* (2003) 'Detection of amastigote-like forms in the valve of *Phlebotomus papatasi* infected with *Leishmania major*', *Memórias do Instituto Oswaldo Cruz*. scielo, pp. 495–498.
- Aoki, J. I. *et al.* (2017) 'L-arginine availability and arginase activity: Characterization of amino acid permease 3 in *Leishmania amazonensis*', *PLoS neglected tropical diseases*. Public Library of Science, 11(10), pp. e0006025–e0006025.
doi: 10.1371/journal.pntd.0006025.
- Archer, S. K., Luu, V. D., de Queiroz, R. A., Brems, S., Clayton, C. (2009) *Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. *PLoS Pathog.* 5:e1000565.
- Benz, C., Mulindwa, J., Ouna, B., Clayton, C. (2011) The *Trypanosoma brucei* zinc finger protein ZC3H18 is involved in differentiation. *Mol Biochem Parasitol.* 177:148–151.
- Besteiro, S. *et al.* (2006) 'Endosome Sorting and Autophagy Are Essential for Differentiation and Virulence of *Leishmania major*', *Journal of Biological Chemistry*, 281(16), pp. 11384–11396. doi: 10.1074/jbc.M512307200.
- Capul, A. A. *et al.* (2007) 'Two Functionally Divergent UDP-Gal Nucleotide Sugar Transporters Participate in Phosphoglycan Synthesis in *Leishmania major*', *Journal of Biological Chemistry*, 282(19), pp. 14006–14017.
- Casas-Sánchez, A. and Acosta-Serrano, Á. (2016) Skin deep. *eLife*. eLife Sciences Publications, Ltd 5: e21506.
- Van Damme, E. and Van Loock, M. (2014) 'Functional annotation of human cytomegalovirus gene products: an update', *Frontiers in microbiology*. Frontiers Media S.A., 5, p. 218. doi: 10.3389/fmicb.2014.00218.
- Dang, H. Q. and Li, Z. (2011) 'The Cdc45·Mcm2-7·GIN5 protein complex in trypanosomes regulates DNA replication and interacts with two Orc1-like proteins in the origin recognition complex', *The Journal of biological chemistry*. 2011/07/28. American Society for Biochemistry and Molecular Biology, 286(37), pp. 32424–32435. doi: 10.1074/jbc.M111.240143.

- Darlyuk, I. *et al.* (2009) 'Arginine Homeostasis and Transport in the Human Pathogen *Leishmania donovani*', *Journal of Biological Chemistry*, 284(30), pp. 19800–19807.
- DeBrot, A., Lancaster, C. and Bjornsti, M.A. (2016) 'Function of Cdc45 in DNA Replication and in Response to Genotoxic Stress', *The FASEB Journal. Federation of American Societies for Experimental Biology*, 30(1_supplement), p. 798.2-798.2. doi: 10.1096/fasebj.30.1_supplement.798.2.
- Depledge, D. P. *et al.* (2009) 'Comparative expression profiling of *Leishmania*: modulation in gene expression between species and in different host genetic backgrounds', *PLoS neglected tropical diseases. Public Library of Science*, 3(7), pp. e476–e476. doi: 10.1371/journal.pntd.0000476.
- Devault, A. and Bañuls, A. L. (2008) 'The promastigote surface antigen gene family of the *Leishmania* parasite: differential evolution by positive selection and recombination', *BMC evolutionary biology. BioMed Central*, 8, p. 292. doi: 10.1186/1471-2148-8-292.
- Doehl, J. S. P. *et al.* (2017) '*Leishmania* HASP and SHERP Genes Are Required for In Vivo Differentiation, Parasite Transmission and Virulence Attenuation in the Host', *PLOS Pathogens. Public Library of Science*, 13(1), p. e1006130.
- Droll, D., Minia, I., Fadda, A., Singh, A., Stewart, M., Queiroz, R., Clayton, C. (2013) Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. *PLoS Pathog.* 9:e1003286.
- Duftner, N. *et al.* (2008) 'Efficacy of RNA amplification is dependent on sequence characteristics: implications for gene expression profiling using a cDNA microarray', *Genomics*. 2007/11/14, 91(1), pp. 108–117. doi: 10.1016/j.ygeno.2007.09.004.
- Fankhauser, N. and Mäser, P. (2005) Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics*. 21(9), pp. 1846–1852.
- Gourguechon, S. and Wang, C. C. (2009) 'CRK9 contributes to regulation of mitosis and cytokinesis in the procyclic form of *Trypanosoma brucei*', *BMC cell biology. BioMed Central*, 10, p. 68. doi: 10.1186/1471-2121-10-68.
- Gupta, S. K., Kosti, I., Plaut, G., Pivko, A., Tkacz, I. D., Cohen-Chalamish, S., *et al.* (2013) The hnRNP F/H homologue of *Trypanosoma brucei* is differentially expressed in the two life cycle stages of the parasite and regulates splicing and mRNA stability. *Nucleic Acids Res.* 41:6577–6594.
- Haines, L. R. *et al.* (2010) 'Tsetse EP protein protects the fly midgut from trypanosome establishment', *PLoS pathogens. Public Library of Science*, 6(3), pp. e1000793–e1000793. doi: 10.1371/journal.ppat.1000793.

- Hammarton, T. C. *et al.* (2003) 'Stage-specific Differences in Cell Cycle Control in *Trypanosoma brucei* Revealed by RNA Interference of a Mitotic Cyclin', *Journal of Biological Chemistry*, 278(25), pp. 22877–22886.
- Hammarton, T. C. (2007) 'Cell cycle regulation in *Trypanosoma brucei*', *Molecular and Biochemical Parasitology*. Elsevier, 153(1), pp. 1–8. doi: 10.1016/J.MOLBIOPARA.2007.01.017.
- Hammarton, T. C., Engstler, M. and Mottram, J. C. (2004) 'The *Trypanosoma brucei* Cyclin, CYC2, Is Required for Cell Cycle Progression through G1 Phase and for Maintenance of Procyclic Form Cell Morphology', *Journal of Biological Chemistry*, 279(23), pp. 24757–24764.
- Holzer, T. R., McMaster, W. R. and Forney, J. D. (2006) 'Expression profiling by whole-genome interspecies microarray hybridization reveals differential gene expression in procyclic promastigotes, lesion-derived amastigotes, and axenic amastigotes in *Leishmania mexicana*', *Molecular and Biochemical Parasitology*, 146(2), pp. 198–218. doi: <https://doi.org/10.1016/j.molbiopara.2005.12.009>.
- Inbar, E. *et al.* (2013) 'A versatile proline/alanine transporter in the unicellular pathogen *Leishmania donovani* regulates amino acid homeostasis and osmotic stress responses', *Biochemical Journal*, 449(2), p. 555 LP-566. doi: 10.1042/BJ20121262.
- Inbar, E. *et al.* (2017) 'The Transcriptome of *Leishmania major* Developmental Stages in Their Natural Sand Fly Vector', *mBio*, 8(2). doi: 10.1128/mBio.00029-17.
- Jackson, A. P., Vaughan, S. and Gull, K. (2006) 'Evolution of Tubulin Gene Arrays in Trypanosomatid parasites: genomic restructuring in *Leishmania*', *BMC Genomics*. London: BioMed Central, 7, p. 261. doi: 10.1186/1471-2164-7-261.
- Käll, L., Krogh, A. and Sonnhammer, E. L. L. (2004) 'A Combined Transmembrane Topology and Signal Peptide Prediction Method', *Journal of Molecular Biology*, 338(5), pp. 1027–1036. doi: <https://doi.org/10.1016/j.jmb.2004.03.016>.
- Kamhawi, S. *et al.* (2004) 'A role for insect galectins in parasite survival.', *Cell*. United States, 119(3), pp. 329–341.
- Kawale, A. S. and Povirk, L. F. (2018) 'Tyrosyl-DNA phosphodiesterases: rescuing the genome from the risks of relaxation', *Nucleic acids research*. 2017/12/04. Oxford University Press, 46(2), pp. 520–537. doi: 10.1093/nar/gkx1219.
- Kolev, N. G., Ullu, E. and Tschudi, C. (2014) The emerging role of RNA-binding proteins in the life cycle of *Trypanosoma brucei*. *Cellular microbiology* 16(4): 482–489.
- Krogh, A. *et al.* (2001) 'Predicting transmembrane protein topology with a hidden markov model: application to complete genomes', *Journal of Molecular Biology*, 305(3), pp. 567–580. doi: <https://doi.org/10.1006/jmbi.2000.4315>.

- Kumar, M. *et al.* (2017) 'Hydrophilic Acylated Surface Protein A (HASPA) of *Leishmania donovani*: Expression, Purification and Biophysico-Chemical Characterization.', *The protein journal*. Netherlands, 36(4), pp. 343–351. doi: 10.1007/s10930-017-9726-x.
- Lang, T. *et al.* (1991) 'Transmission and scanning EM-immunogold labeling of *Leishmania major* lipophosphoglycan in the sandfly *Phlebotomus papatasi*.', *European journal of cell biology*. Germany, 55(2), pp. 362–372.
- Lieke, T. *et al.* (2008) '*Leishmania* surface protein gp63 binds directly to human natural killer cells and inhibits proliferation', *Clinical and experimental immunology*. Blackwell Science Inc, 153(2), pp. 221–230. doi: 10.1111/j.1365-2249.2008.03687.x.
- Lincoln, L. M. *et al.* (2004) 'Genetic complementation of *Leishmania* deficient in PSA (GP46) restores their resistance to lysis by complement', *Molecular and Biochemical Parasitology*, 137(1), pp. 185–189. doi: <https://doi.org/10.1016/j.molbiopara.2004.05.004>.
- Liu, Y., Hu, H. and Li, Z. (2013) 'The cooperative roles of PHO80-like cyclins in regulating the G1/S transition and posterior cytoskeletal morphogenesis in *Trypanosoma brucei*', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 90(1), pp. 130–146. doi: 10.1111/mmi.12352.
- Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.
- Marchese, L. *et al.* (2018) 'The Uptake and Metabolism of Amino Acids, and Their Unique Role in the Biology of Pathogenic Trypanosomatids', *Pathogens*. doi: 10.3390/pathogens7020036.
- Matthews, K. R. (2005) 'The developmental cell biology of *Trypanosoma brucei*', *Journal of cell science*, 118(Pt 2), pp. 283–290. doi: 10.1242/jcs.01649.
- McConville, M. J. *et al.* (1992) 'Developmental modification of lipophosphoglycan during the differentiation of *Leishmania major* promastigotes to an infectious stage.', *The EMBO journal*. England, 11(10), pp. 3593–3600.
- Monnerat, S. *et al.* (2013) 'Identification and Functional Characterisation of CRK12:CYC9, a Novel Cyclin-Dependent Kinase (CDK)-Cyclin Complex in *Trypanosoma brucei*', *PloS one*. Public Library of Science, 8(6), pp. e67327–e67327. doi: 10.1371/journal.pone.0067327.
- Nash, T. E. *et al.* (1988) 'Antigenic variation in *Giardia lamblia*.', *The Journal of Immunology*, 141(2), p. 636 LP-641.
- de Paiva, R. M. C. *et al.* (2015) 'Amastin Knockdown in *Leishmania braziliensis* Affects Parasite-Macrophage Interaction and Results in Impaired Viability of Intracellular Amastigotes', *PLOS Pathogens*. Public Library of Science, 11(12), p. e1005296.

- Peacock, L. *et al.* (2011) 'Identification of the meiotic life cycle stage of *Trypanosoma brucei* in the tsetse fly.', *Proceedings of the National Academy of Sciences of the United States of America*, United States, 108(9), pp. 3671–3676. doi: 10.1073/pnas.1019423108.
- Peacock, L. *et al.* (2014) 'Meiosis and haploid gametes in the pathogen *Trypanosoma brucei*', *Current biology : CB*, Cell Press, 24(2), pp. 181–186. doi: 10.1016/j.cub.2013.11.044.
- Pimenta, P. F. *et al.* (1992) 'Stage-specific adhesion of *Leishmania* promastigotes to the sand fly midgut', *Science*, 256(5065), p. 1812 LP-1815. doi: 10.1126/science.1615326.
- Pimenta, P. F. *et al.* (1994) 'Evidence that the vectorial competence of phlebotomine sand flies for different species of *Leishmania* is controlled by structural polymorphisms in the surface lipophosphoglycan', *Proceedings of the National Academy of Sciences of the United States of America*, 91(19), pp. 9155–9159.
- Pita, S. *et al.* (2019) 'The Tritryps Comparative Repeatome: Insights on Repetitive Element Evolution in Trypanosomatid Pathogens', *Genome biology and evolution*. Oxford University Press, 11(2), pp. 546–551. doi: 10.1093/gbe/evz017.
- Pizarro-Cerdá, J. and Cossart, P. (2006) 'Bacterial Adhesion and Entry into Host Cells', *Cell*. Elsevier, 124(4), pp. 715–727. doi: 10.1016/j.cell.2006.02.012.
- Rochette, A. *et al.* (2005) 'Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania spp.*', *Molecular and Biochemical Parasitology*, 140(2), pp. 205–220. doi: <https://doi.org/10.1016/j.molbiopara.2005.01.006>.
- Sacks, D. L. (1989) 'Metacyclogenesis in *Leishmania* promastigotes', *Experimental Parasitology*, 69(1), pp. 100–103. doi: [https://doi.org/10.1016/0014-4894\(89\)90176-8](https://doi.org/10.1016/0014-4894(89)90176-8).
- Savage, A. F. *et al.* (2016) 'Transcriptome Profiling of *Trypanosoma brucei* Development in the Tsetse Fly Vector *Glossina morsitans*', *PloS one*. Public Library of Science, 11(12), pp. e0168877–e0168877. doi: 10.1371/journal.pone.0168877.
- Seed, J. R. and Wenck, M. A. (2003) 'Role of the long slender to short stumpy transition in the life cycle of the African trypanosomes', *Kinetoplastid biology and disease*. BioMed Central, 2(1), p. 3. doi: 10.1186/1475-9292-2-3.
- Seepersaud, R. *et al.* (2005) 'Characterization of a novel leucine-rich repeat protein antigen from group B streptococci that elicits protective immunity', *Infection and immunity*. American Society for Microbiology, 73(3), pp. 1671–1683. doi: 10.1128/IAI.73.3.1671-1683.2005.
- Shpilka, T. *et al.* (2011) 'Atg8: an autophagy-related ubiquitin-like protein family', *Genome biology*. BioMed Central, 12(7), p. 226. doi: 10.1186/gb-2011-12-7-226.

- Smith, T. K. *et al.* (2017) 'Metabolic reprogramming during the *Trypanosoma brucei* life cycle', F1000Research. F1000Research, 6, p. F1000 Faculty Rev-683. doi: 10.12688/f1000research.10342.2.
- Subota, I., Rotureau, B., Blisnick, T., Ngwabyt, S., Durand-Dubief, M., Engstler, M., Bastin P. (2011) ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation. *Mol Biol Cell*. 22:4205–4219.
- Suzuki, E. *et al.* (2002) 'Role of beta-D-galactofuranose in *Leishmania major* macrophage invasion', *Infection and immunity*. American Society for Microbiology, 70(12), pp. 6592–6596. doi: 10.1128/IAI.70.12.6592-6596.2002.
- Svärd, S. G. *et al.* (1998) 'Differentiation-associated surface antigen variation in the ancient eukaryote *Giardia lamblia*', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 30(5), pp. 979–989. doi: 10.1046/j.1365-2958.1998.01125.x.
- Tu, X. and Wang, C. C. (2004) 'The Involvement of Two cdc2-related Kinases (CRKs) in *Trypanosoma brucei* Cell Cycle Regulation and the Distinctive Stage-specific Phenotypes Caused by CRK3 Depletion', *Journal of Biological Chemistry*, 279(19), pp. 20519–20528. doi: 10.1074/jbc.M312862200.
- Tu, X. and Wang, C. C. (2005) 'Pairwise knockdowns of cdc2-related kinases (CRKs) in *Trypanosoma brucei* identified the CRKs for G1/S and G2/M transitions and demonstrated distinctive cytokinetic regulations between two developmental stages of the organism', *Eukaryotic cell*. American Society for Microbiology, 4(4), pp. 755–764. doi: 10.1128/EC.4.4.755-764.2005.
- Vanrell, M. C. *et al.* (2017) 'The regulation of autophagy differentially affects *Trypanosoma cruzi* metacyclogenesis', *PLOS Neglected Tropical Diseases*. Public Library of Science, 11(11), p. e0006049.
- Vassella, E. *et al.* (2009) 'Major surface glycoproteins of insect forms of *Trypanosoma brucei* are not essential for cyclical transmission by tsetse', *PloS one*. Public Library of Science, 4(2), pp. e4493–e4493. doi: 10.1371/journal.pone.0004493.
- Vingron, M. *et al.* (2007) 'Improved detection of overrepresentation of Gene-Ontology annotations with parent–child analysis', *Bioinformatics*, 23(22), pp. 3024–3031. doi: 10.1093/bioinformatics/btm440.
- Wang, L., Sloan, M. A. and Ligoxygakis, P. (2019) 'Intestinal NF-κB and STAT signalling is important for uptake and clearance in a *Drosophila-Herpetomonas* interaction model', *PLOS Genetics*. Public Library of Science, 15(3), p. e1007931.
- Williams, R. A. M., Mottram, J. C. and Coombs, G. H. (2013) 'Distinct Roles in Autophagy and Importance in Infectivity of the Two ATG4 Cysteine Peptidases of *Leishmania major*', *Journal of Biological Chemistry*, 288(5), pp. 3678–3690. doi: 10.1074/jbc.M112.415372.

Wurst, M., Seliger, B., Jha, B. A., Klein, C., Queiroz, R., Clayton, C. (2012) Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in *Trypanosoma brucei*. Mol Microbiol. 83:1048–1063.

CHAPTER FIVE

Development of tools for the study of

Herpetomonas muscarum

CHAPTER FIVE:

Development of tools for the study of *Herpetomonas muscarum*

Preface

The ability to alter the sequence of an organism's DNA precisely and at specific loci has proven invaluable across the spectrum of biological research – from basic science to biotechnology. Generation of mutant cell lines, including gene knockout mutants and cells expressing labelled proteins, enable the determination of a gene's function and spatial location in the cell. In the context of host-pathogen interactions, use of transgenic cell lines during *in vivo* infection assays can be used to determine which genes are critical to infection outcome – the essential precursor to the development of new disease therapies. Since the discoveries of bacterial restriction enzymes in the 1970s (Kelly and Smith, 1970), our ability to edit DNA sequences has developed rapidly. Assisted by the advent of whole genome sequencing, we now have the ability to edit entire genomes at an unprecedented speed and scale, through technologies such as CRISPR. The application of these techniques will undoubtedly accelerate research in numerous fields, including studies of the interactions of trypanosomatids and their vector insects.

Genetic manipulation of trypanosomatids

It has long been known that transgenic trypanosomatid lines can be produced by the introduction of foreign DNA into the cells by electroporation; application of an electromagnetic field to cells increases membrane permeability allowing macromolecules, including DNA, to enter the cell. The first, unambiguous, demonstration of the

trypanosomatid transformation was reported by Bellofatto and Cross (1989). The group demonstrated the transient expression of chloramphenicol acetyl transferase (CAT), from *Escherichia coli*, after electroporation of *Leptomonas seymouri* cells in presence of a plasmid containing the CAT gene fused to splice leader sequences (Bellofatto and Cross, 1989). Since then, many trypanosomatids have been shown to express genes from plasmids, which can be maintained as extra-chromosomal elements under antibiotic selection (Kapler, Coburn and Beverley, 1990; Tetaud *et al.*, 2002).

Trypanosomatids can be more stably transfected by integration of DNA at specific sites in by homologous recombination (Cruz and Beverley, 1990). Homologous recombination (HR) is a natural process used by all living organisms to seamlessly repair damaged DNA. In eukaryotes, HR also allows production of genetic variety during meiotic division. Though the technicalities of HR vary between organisms, the core stages of resection, strand invasion and DNA synthesis are conserved (reviewed in Wright, Shah and Heyer 2018, Figure 5.1). Briefly, following a double stranded break in the DNA, proteins bind to the ends of the broken DNA. The 5' ends of the DNA are then trimmed back, creating 3' overhangs, in a process called resection. The resulting single stranded overhangs invades DNA with sequence complementarity forming a displacement loop (D-loop) structure. This complementary DNA template is usually provided by the undamaged, identical sister chromatid. DNA polymerase extends the invading strand changing the D-loop to a cross-like structure termed the Holliday junction. Following extension, the invading strand is subsequently released from the Holliday junction, where it can anneal to the DNA at the other side of the double stranded break by complementary base pairing. The gaps in the DNA 'backbone' are then sealed by DNA ligase to complete the repair.

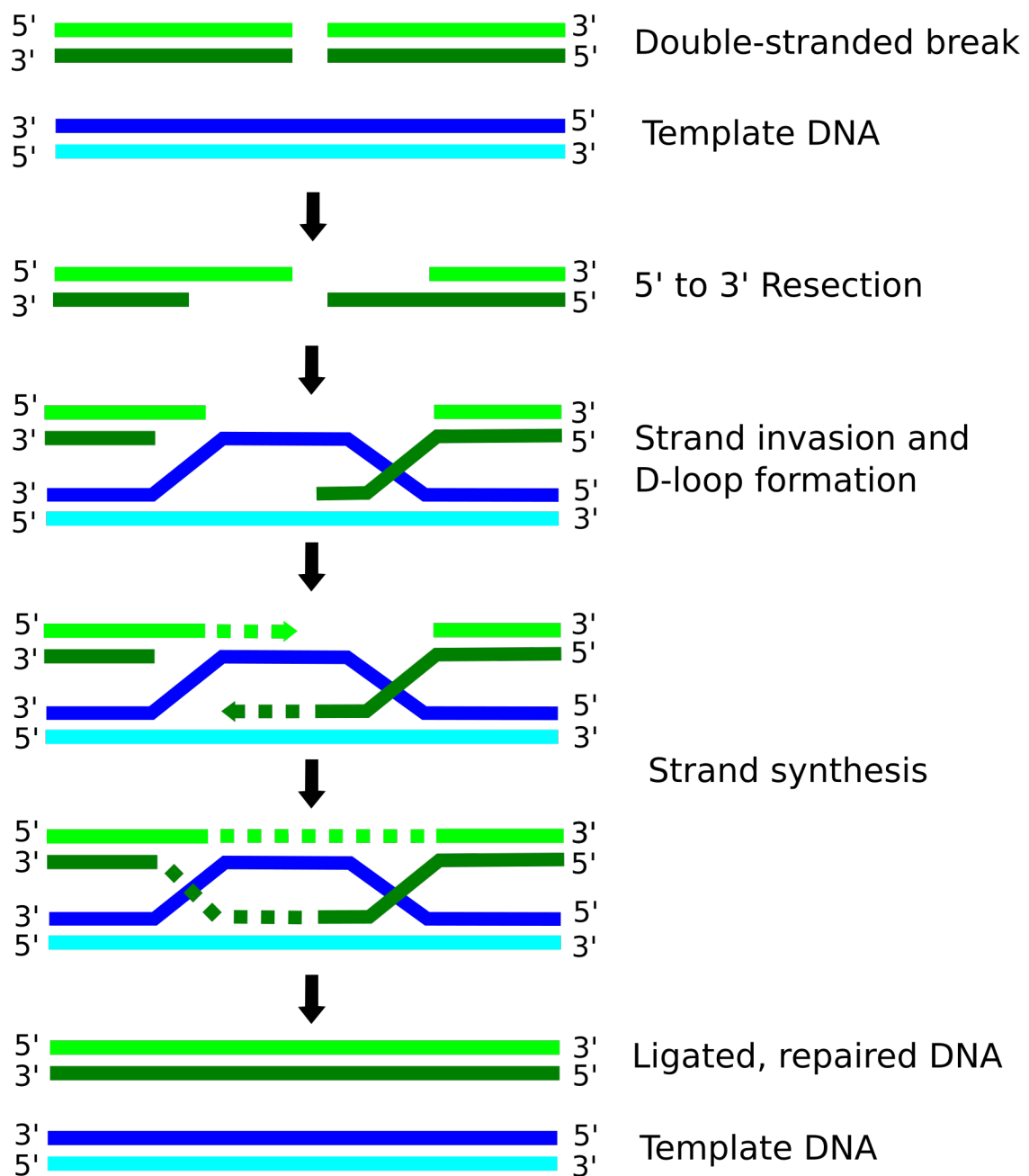


Figure 5.1 – Double stranded DNA repair by homologous recombination.

First described in yeast (Rothstein, 1983), HR can be utilised to generate knockout parasites, via targeted gene replacement, or to label proteins at the N- or C-terminus

(Figure 5.2). When provided with donor DNA flanked with sequences complementary to a specific genomic locus of interest, the cell will incorporate the donor DNA using HR. Donor DNA cassettes can be delivered into cells as PCR products or cloned into plasmid vectors as described in Dean *et al.* (2015). The efficiency of recombination of the donor DNA is variable between species and dependant on the length of the homologous flanks. For example, *Trypanosoma brucei* only requires 50bp homology flanks for successful recombination, but recombination efficiency appears to be lower in *Leishmania mexicana* which requires homology flanks in excess of 300bp (Dean *et al.*, 2015). This homology-based genome editing technique is scalable and has been developed to allow high-throughput tagging of *Trypanosoma brucei* genes. The TrypTag project, which aimed to tag and localise every gene product, with the exception of the variant surface glycoproteins, in the *Trypanosoma brucei* genome is nearing completion (Dean, Sunter and Wheeler, 2017).

CRISPR

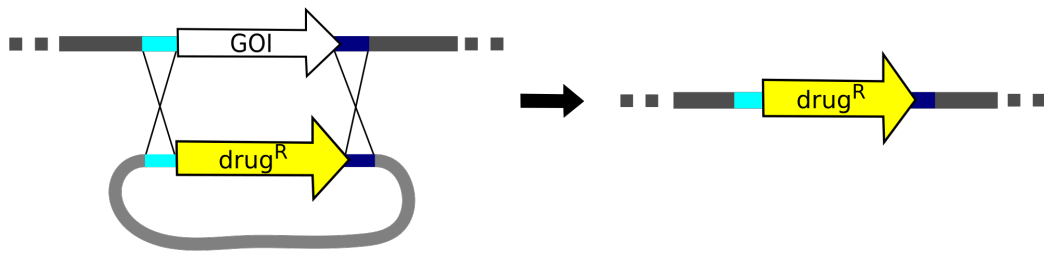
The development of the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) DNA editing technology has revolutionised the field of genetics. The technology utilises a natural defence mechanism of bacteria (and archaea) against viruses, whereby the DNA of the invading virus is cleaved by bacterial nuclease – Cas9 (CRISPR-associated gene 9) (Garneau *et al.*, 2010). This cleavage is non-random, as this could result in damage to the host DNA. To prevent this, the nuclease is targeted to specific viral sequences by guiding RNAs and requires a specific sequence, known as the protospacer adjacent motif (PAM), adjacent to the cut site for cleavage to occur (Figure 5.3). The specific PAM motif is dependent on the Cas9's organism of origin. For

example, the Cas9 of *Streptococcus pyogenes* will create double strand breaks (DSBs) upstream of any nucleotide followed by two guanine residues (NGG).

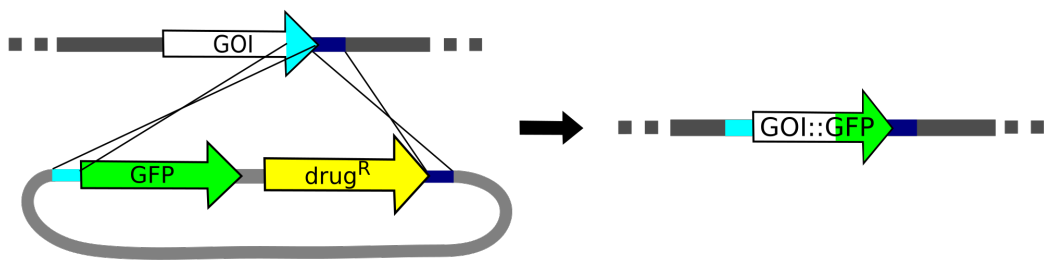
Researchers initially utilized this system to allow targeted genome editing in bacteria (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012), before demonstrating its efficacy in eukaryotes (Cong *et al.*, 2013; Jinek *et al.*, 2013; Mali *et al.*, 2013). By altering the sequence of the RNA which guides the Cas9 nuclease, precise double stranded breaks can be made anywhere in the genome, providing there is a nearby PAM motif. These breaks are then repaired by the cell's natural mechanisms. Providing donor DNA templates, similar to those shown in Figure 5.2, guides the repair of the DSB genes and allows the integration of knockout or tagging cassettes into the genome by homologous recombination. In the absence of donor DNA, DSBs are repaired by other methods such as non-homologous end joining (NHEJ). During NHEJ the cut ends of the DNA are 'trimmed back' before re-ligation and as such introduces short deletions at the break site. Depending on the size and location of these short deletions, this can result in frame shift mutations.

CRISPR has a several advantages over targeted gene replacement in the absence of Cas9-mediated DSBs. The efficiency of HR is significantly higher in the presence of a DSB and requires much shorter homology flanks (Dean *et al.*, 2015). Furthermore, by the provision of multiple guide RNAs, CRISPR allows the targeting of multiple genes simultaneously. This is crucial for genes present in multiple copies such as those in tandem arrays or found on supernumerary chromosomes – both of which are common in trypanosomatids. Previously, the generation of knockouts of such genes was slow and laborious as they required extensive cloning and multiple transfection steps.

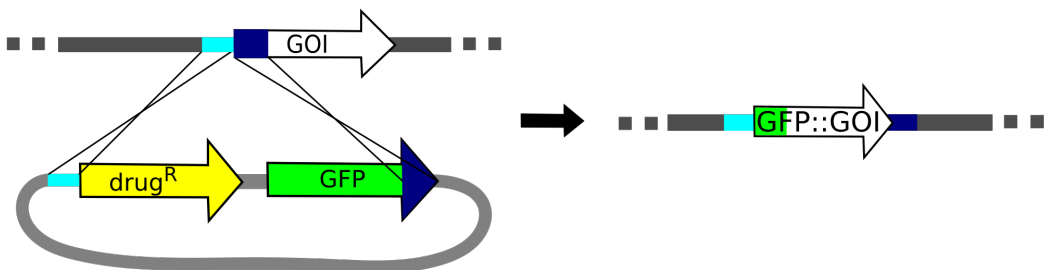
Gene knock out



C-terminal tagging



N-terminal tagging



Key

5' homology flank

3' homology flank



plasmid backbone



homologous recombination

Figure 5.2 – Knockout and tagging of a gene of interest (GOI) by homologous recombination. GFP = green fluorescent protein. The plasmid backbone contains the ampicillin resistance gene and bacterial origin of replication to allow cloning in *Escherichia coli*.

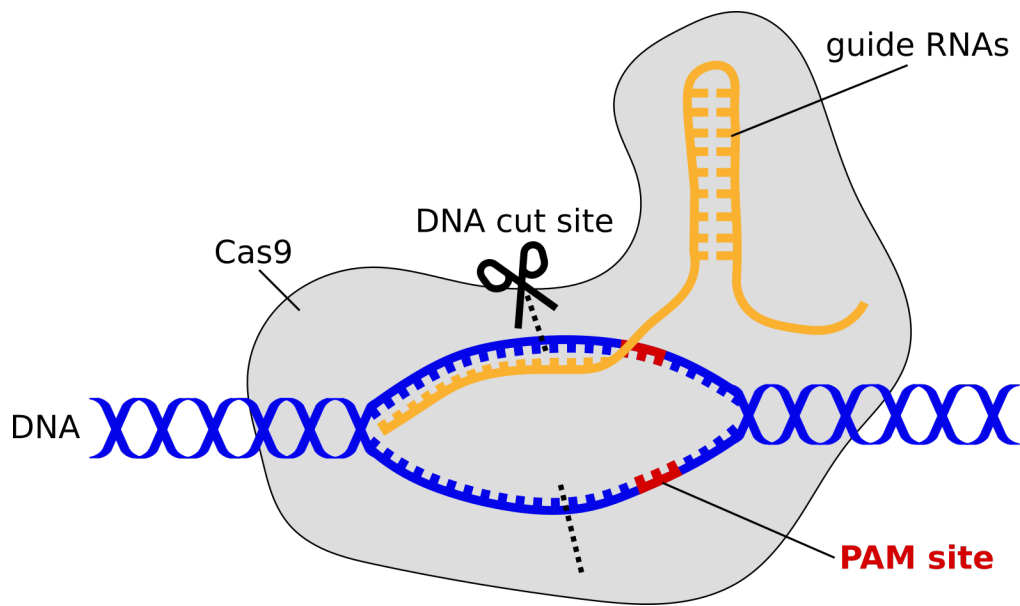


Figure 5.3 – Creation of double stranded breaks in DNA by Cas9

Recently, CRISPR systems have been developed for use in a number of trypanosomatids including those from *Leishmania*, *Trypanosoma* and *Crithidia* (Peng *et al.*, 2014; Beneke *et al.*, 2017; Liu and Kadowaki, 2018; Romagnoli *et al.*, 2018). Applications of this technology to are broad and will play an important role in the discovery of new therapeutic targets for the treatment and prevention of trypanosomatid diseases.

Genome editing of Herpetomonas muscarum

One of the strengths of the *Drosophila*– *Herpetomonas* model is the plethora of knowledge, tools and resources available for the study of *D. melanogaster*. A vital tool for the new model system is a means to genetically manipulate *H. muscarum*. However, despite extensive work in other trypanosomatids, there are no published examples of *H. muscarum* transgenic lines. As such, I developed genome editing tools for use in this

understudied trypanosomatid based on those described for *Leishmania* (Dean *et al.*, 2015; Beneke *et al.*, 2017).

Specifically, I produced a set of plasmids allowing exogenous expression of proteins in *H. muscarum*. The plasmids I made allowed the production of fluorescent *H. muscarum* cell lines to assist in the localisation of cells in the fly gut during infection by microscopy. Furthermore, these plasmids are modular and would allow, after one-step cloning, the integration of genes of interest into the beta-tubulin locus of *H. muscarum*. As such they are powerful tool for the *Drosophila-Herpetomonas* model.

Results and Discussion

Fluorescent Herpetomonas muscarum cell lines

In order to follow the pathway of *H. muscarum* after ingestion by *D. melanogaster* using microscopy (as shown in Chapter Three), I sought to produce transgenic *H. muscarum* cell lines which express a fluorescent marker. There are a number of expression vectors for trypanosomatids available (LeBowitz *et al.*, 1990; Tetaud *et al.*, 2002; Madeira da Silva *et al.*, 2009; Sugino and Niimi, 2012; Dean *et al.*, 2015). Initial attempts were made to use the pNUS-GFPcB expression vector, a generous gift from Dr. Eva Gluenz, to produce a green fluorescent protein (GFP) expressing *H. muscarum* line (Figure 5.4A). This vector was chosen as it had been successfully used in species from both the *Leishmania* and *Crithidia* genera (Tetaud *et al.*, 2002).

Using several electroporation conditions based on protocols for *Leishmania* (Dean *et al.* 2015) and *Leptomonas* transfections (Bellofatto, Torres-Munoz and Cross, 1991), I was

able to generate drug resistant *Herpetomonas muscarum* cell lines using the pNUS-GFP vector (Table 5.1, for the full protocol see Chapter Seven). The presence of the plasmid in the cells several passages after electroporation was also confirmed (Figure 5.4B).

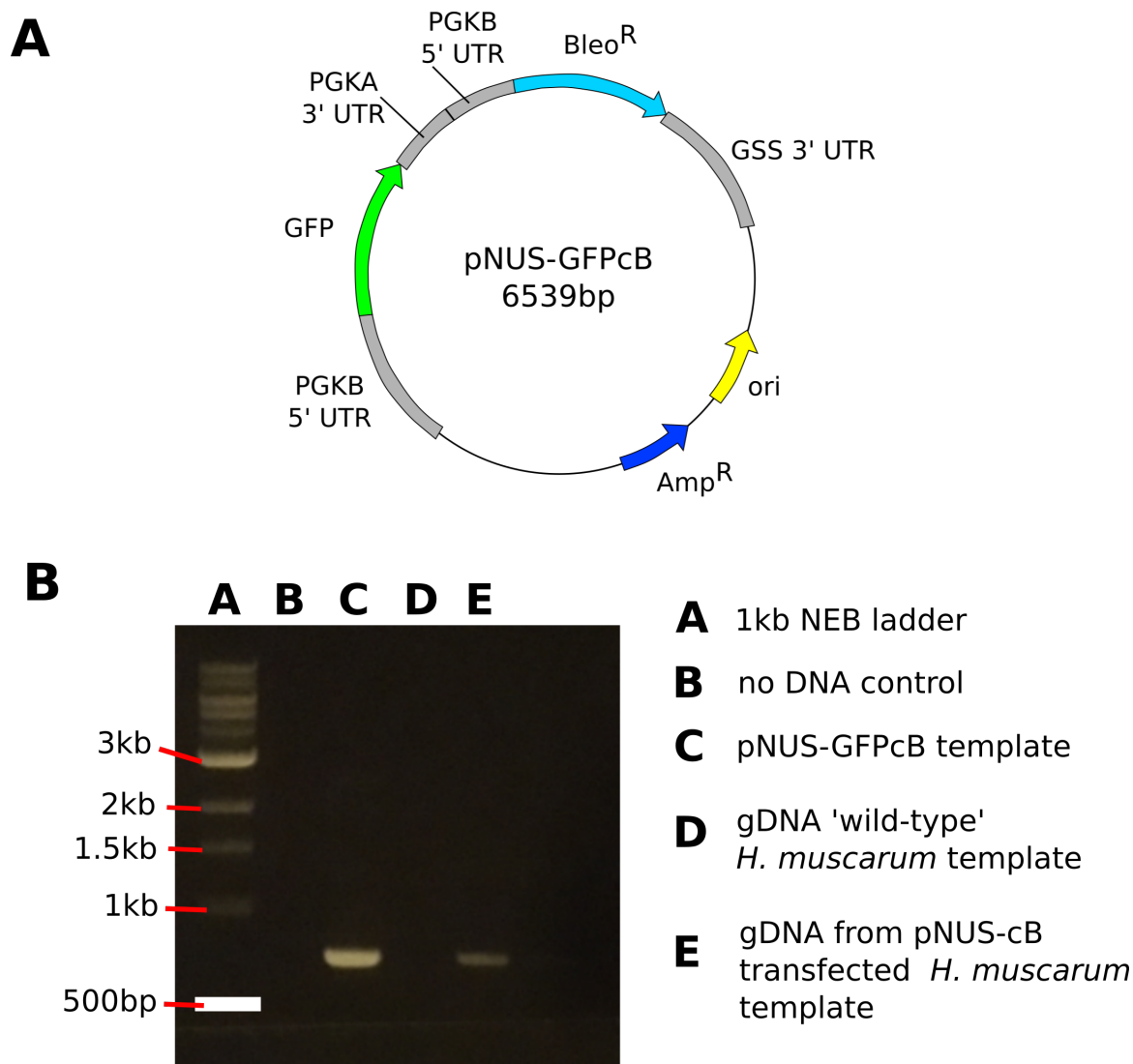


Figure 5.4 – Transformation of *Herpetomonas muscarum* with pNUS-GFPcB (Tetaud *et al.* 2002). **A.** Plasmid map for pNUS-GFPcB. Abbreviations: GFP – green fluorescent protein, PGKA/B - phosphoglycerate kinases A and B, UTR – untranslated region, Ori - bacterial origin of replication site, BleoR - phleomycin resistance gene. **B.** PCR products from reaction to amplify the GFP open reading frame from gDNAs of transfected cells. 1% Agarose gel.

However, the cells did not appear to express the fluorescence protein. Control transfections performed with *Leishmania mexicana* resulted in drug resistant and GFP expressing cells for all conditions used (Figure S5.1). I hypothesized that the reason for the lack of visible GFP expression in *H. muscarum* was due to the untranslated regions (UTRs) flanking the GFP gene. Gene expression in trypanosomatids is heavily dependent on the flanking UTRs which do not always ‘translate’ between trypanosomatid species (Aly *et al.*, 1994; Pradeep *et al.*, 1994). As such I produced a new expression vector using native *Herpetomonas muscarum* sequences.

Electroporation Condition	Buffer	pNUS-GFPcB mass	Drug resistant cells?	
			<i>L. mexicana</i>	<i>H. muscarum</i>
Amaza X-001	Roditi Buffer	10µg	yes	no
Amaza X-001	Amaza parasite buffer 1	60µg	yes	no
		20µg	yes	yes
BTX: 2 X 1500V, 200ms gap	ZPFM	60µg	yes	yes
		20µg	yes	yes

Table 5.1 – Conditions used for transfection of pNUS-GFPcB into *Herpetomonas muscarum* and *Leishmania mexicana*.

A novel expression vector for *Herpetomonas muscarum*

I constructed a new expression vector based on those developed in Dean *et al.* 2015.

These plasmids contain the pLENT ‘backbone’ enabling replication and cloning in *Escherichia coli*, and an expression ‘cassette’ flanked by the whole intergenic regions of *H. muscarum* beta-tubulin gene (HMUS00495500). The whole intergenic regions were used as the specific sequences required for gene expression in these regions are not known in *H. muscarum*. The expression cassette contains the gene for fluorescent protein tdTomato, the intergenic region between phosphoglycerate kinases A and B (PGKAB) and the neomycin resistance gene (Neo^R) (Figure 5.5A and S5.2). For full descriptions on the construction of this plasmid, and others in this chapter, see Chapter Seven.

Drug resistant cells, expressing tdTomato protein (Figure 5.5B) were successfully produced after electroporation with two pulses of 1500V for 100µs (with a 200ms gap between pulses) in the presence of pMS003 which had been linearised by XhoI restriction digest. The minimum mass of plasmid required to produce transgenic *H. muscarum* using the pMS003 vector was found to be between 10-20µg of linearised plasmid DNA – approximately 1.75-3.5 pmol of DNA. I observed a slight growth phenotype in the transgenic cells (Figure 5.6). However, the growth rate of the transgenic line was not significantly different to that of the wild type cells - whose doubling time approximately 12 hours with cells reaching the stationary phase of growth at 10⁸ cells per ml.

To determine if the plasmid is maintained as an episome or has been integrated into the genomic DNA, I extracted DNA from the tdTomato expressing cells, then used the DNA to transform into competent *E. coli* (NEB, DH5-alpha high efficiency). I was unable to recover any ampicillin resistant bacterial clones from transformation with *H. muscarum*

tdTomato DNAs nor DNA extracted from wild type *H. muscarum*. Positive controls of tdTomato plasmid DNA and wild type *H. muscarum* DNA ‘spiked’ with tdTomato plasmid, both yielded positive colonies. As such it is likely that the pMS003 is integrated into the genome. To my knowledge this is the first known genetic manipulation of a *Herpetomonas* species. Iterations of pMS003 were made with different fluorescent proteins and drug resistance genes (Table 5.2) to allow production of double mutants (see Concluding remarks).

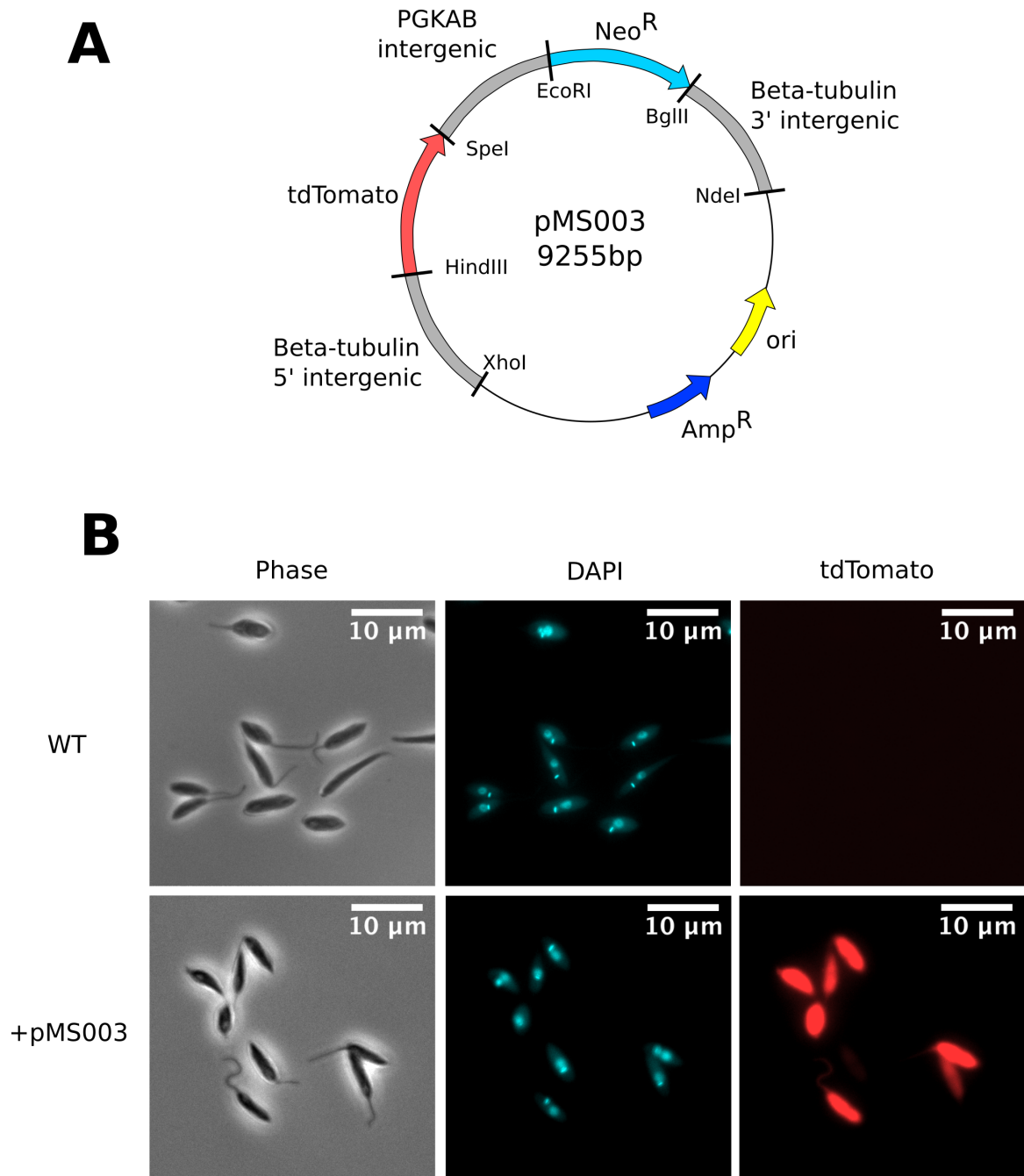


Figure 5.5 - Transformation of *Herpetomonas muscarum* with pMS003. **A.** The expression cassette from pMS003. Restriction enzymes sites for plasmid manipulation shown (not all enzyme sites shown). ori - bacterial origin of replication site, Amp^R - ampicillin resistance gene. **B.** Light microscopy of *H. muscarum* transfected with pMS003 and wild type *H. muscarum*. Cells were fixed and stained with the DNA-binding dye DAPI – see Chapter Seven for details of method.

Plasmid Name	Protein of interest	Selection Drug
pMS003	tdTomato	Neomycin (G418)
pMS004	tdTomato	Hygromycin
pMS005	GFP	Phleomycin
pMS006	Cas9 (<i>P. aeruginosa</i>)	Neomycin (G418)
pMS007	dd-Cas9 (<i>P. aeruginosa</i>)	Neomycin (G418)
pMS008	Cyan FP	Blasticidin

Table 5.2 – A novel set of plasmids for genetic manipulation of *Herpetomonas muscarum*.

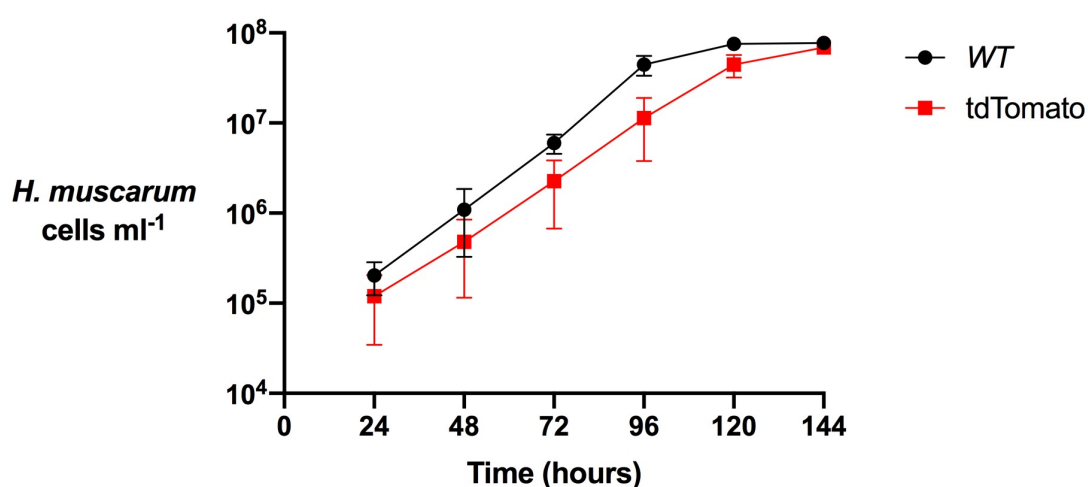


Figure 5.6 – Growth curves of wild type and transgenic tdTomato *Herpetomonas muscarum* cell lines cultured *in vitro*. The graph shows the average number of *H. muscarum* per fly (pooled samples of 5 flies) measured for three consecutive weeks and error bars show the standard error of the mean. WT – wild type, tdTomato – *H. muscarum* expressing the tdTomato fluorescence protein.

RNA-sequencing of the *H. muscarum* transcriptome after ingestion by *D. melanogaster* (Chapter Four) identified several genes of interest for further study. Generation of cells where these genes are knocked out or tagged will be important to understanding the roles these genes play during infection. As previously discussed, CRISPR-Cas9 genome editing tools allow the quick and efficient generation of mutant cell lines. There are several CRISPR-Cas9 systems described for trypanosomatids (Peng *et al.*, 2014; Beneke *et al.*, 2017; Liu and Kadowaki, 2018; Romagnoli *et al.*, 2018), as such I aimed to replicate these in *H. muscarum* which would provide an invaluable tool for future study of *H. muscarum*.

The previously described pMS003 plasmid was modified to contain the Cas9 nuclease gene from *Pseudomonas aeruginosa*, in place of tdTomato open reading frame, to produce pMS006 (Figure S5.3). No neomycin-resistant cells were recovered after pMS006 was electroporated into *H. muscarum*. Given the reports of Cas9 toxicity in other trypanosomatids (Peng *et al.*, 2014), I made a second Cas9 vector containing the Cas9 gene with a N-terminal FKBP destabilization domain (FKBP-DD) (pMS007, Figure 5.7 and S5.4). Proteins containing the FKBP-DD are targeted for degradation in the absence of the stabilising ligand Shield-1 (Armstrong and Goldberg, 2007; Madeira da Silva *et al.*, 2009). As such by controlled exposure to Shield-1 the Cas9 expression can be switched on/off in short intervals reducing the impact of any Cas9 toxicity. This approach has been used successfully in *Trichomonas vaginalis* (Janssen *et al.*, 2018). Additionally, I had successfully produced a *Leishmania mexicana* cell line expressing DD::Cas9 and used this line to knockout a copy of alpha-tubulin in my previous work (unpublished).

Transfection of *H. muscarum* with linearised pMS007 yielded drug resistance cell populations. The presence of the DD::Cas9 in *H. muscarum* after 10 passages was confirmed by PCR (Figure 5.8). When DNA extracted from the DD::Cas9 *H. muscarum* cells was used to transform chemically competent *E. coli* (as described above for pMS003), ampicillin-resistant bacterial clones were produced, from which I was able to recover pMS007 plasmid. As such it appears that at least some cells are maintaining pMS007 as an episome.

I was unable to produce cell lines with a construct designed to express wild type Cas9 under the control of the β -tubulin and phosphoglycerate kinase UTRs (pMS006). However, successful transfectants were produced when the Cas9 protein is fused at the N-terminus to a destabilisation domain (pMS007). This indicates that Cas9 expression can be toxic to *H. muscarum* – as has been reported in other organisms (Peng *et al.*, 2014; Sidik *et al.*, 2016). In these plasmids, the Cas9 gene is flanked by intergenic regions from beta-tubulin and phosphoglycerate kinase, both of which are highly expressed genes. As such I would predict high expression of Cas9 due to the adjacent UTRs – though this has not been proven in this work. A high level of Cas9 in the cell may be toxic if it is able to cleave the DNA unchecked. By using alternative intergenic regions, from genes with lower expression it may be possible to constitutively express Cas9 at low levels in the future.

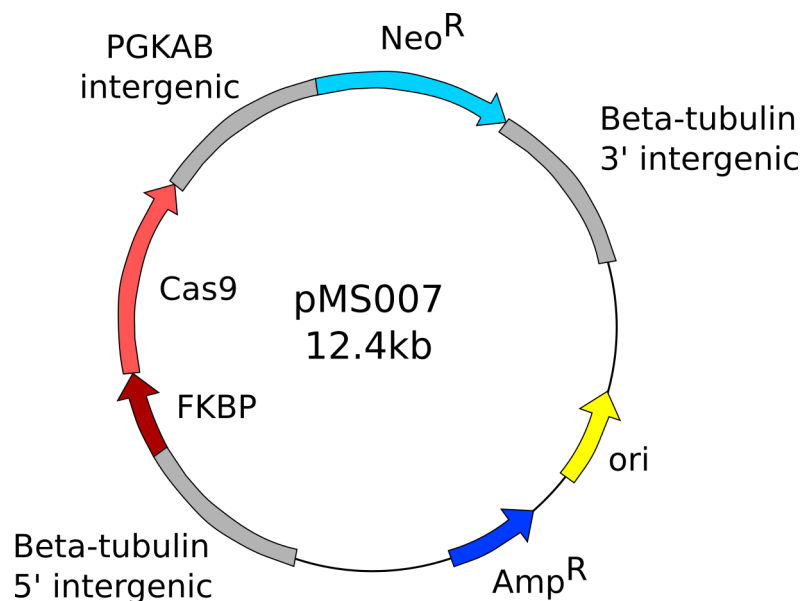


Figure 5.7 – pMS007 for expression of a destabilised Cas9 in *Herpetomonas muscarum*. Abbreviations: Neo^R – neomycin resistance gene, FKBP – destabilisation domain FKBP_L106P, ori - bacterial origin of replication site, Amp^R - ampicillin resistance gene.

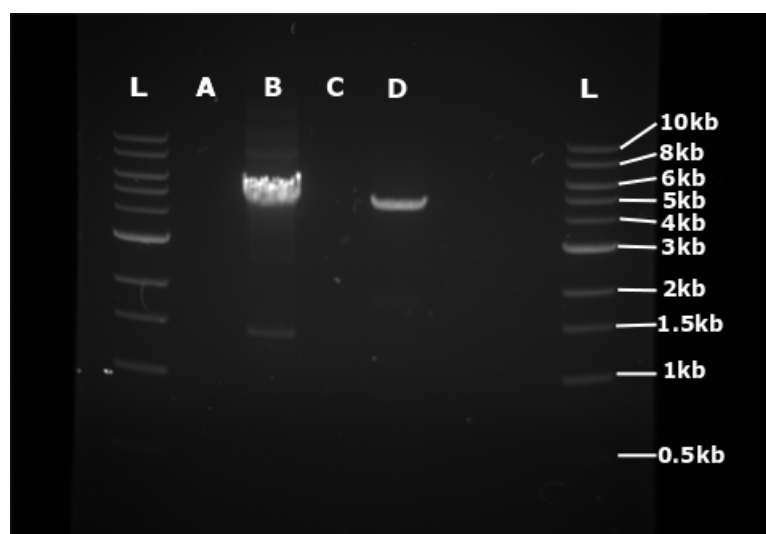


Figure 5.8 – pMS007 is maintained in *Herpetomonas muscarum* for > 10 passages. 1% agarose gel showing PCR reactions amplifying the dd-Cas9 gene fusion (4.6kb) from *H. muscarum* gDNA extracts. **A** – no DNA control, **B** – pLENT-ddCas9 plasmid (positive control), **C** – wild type *H. muscarum* gDNA and **D** – *H. muscarum* dd-Cas9 gDNA. **L** – NEB 1kb ladder. Unlabelled lanes did not contain samples.

Concluding Remarks

The development of tools for the manipulation of gene expression in *H. muscarum* will be crucial to unpick the trypanosomatids interactions with its fly host. Several genes of interest have already been identified by transcriptomics approaches described in Chapter Four. Prior to this work there had been no documented efforts to produce transgenic *Herpetomonas* cell lines, despite extensive work in other trypanosomatids. I have described the development of the first transgenic *Herpetomonas* cell lines. Cells expressing fluorescent marker proteins have allowed me to ‘follow’ *H. muscarum* through the digestive tract of *D. melanogaster* (Chapter Three).

The expression vectors used to express fluorescence proteins in *H. muscarum* have dual function. They will also enable the knockout and tagging of proteins at the C-terminus by homologous recombination. Using fusion PCR, like that described by Dean *et al.* 2015, homology flanks from any gene can be added to the expression cassette after it has been excised or amplified from pMS003. The additional plasmids derived from pMS003, which contain alternative drug resistance cassettes and fluorescent proteins (Table 5.1), will allow the knockout or tagging of multiple proteins in the same cell line under different drug selection markers and as such are a powerful new tool for the *Drosophila-Herpetomonas* model. Additionally, the production of a destabilised-Cas9 cell line is an important step towards a CRISPR genome editing system in *H. muscarum*. Further work is required to establish this technology in *H. muscarum* fully. This work has not yet demonstrated expression of a functional Cas9 in *H. muscarum* and further optimisation will be required to fine-tune the conditional expression of the destabilised Cas9 with its

ligand Shield-1. Specifically, the concentration of stabilising ligand, and the duration of exposure, required to stabilise the expression of Cas9 will require optimisation.

RNAi vs CRISPR technologies

Another strategy for manipulating the expression of genes of interest is RNA-interference (RNAi). Rather than generating gene knockouts, it is possible to reduce expression of genes by targeting their mRNA transcripts for degradation using complementary RNA molecules. This technique is invaluable for the study of essential genes and is well established for *Trypanosoma brucei*. This technique also has the benefit of being ‘reversible’ allowing the knock-down phenotype verified by restoring protein expression to normal in the same cells.

There are weaknesses to using an RNAi approach, e.g. if only a small amount of a protein is required by the cell then important aspects of its role may not be revealed by knock-down alone. As such the generation of knockouts can be a more powerful approach to studies of gene function. CRISPR allows the high-throughput generation of knockouts, including single-step, multi-gene knockouts. Additionally, by providing donor DNAs CRISPR can be used to ‘knock in’ genes in trypanosomatids. This strategy is useful for protein tagging and generation of over-expression mutants. Furthermore, recent development in ‘CRISPRi’ technology means it can be used to reversibly block gene expression – which is one of the strengths of RNAi. By using a Cas9 which reversibly binds to specific DNA sequences and blocks RNA polymerase progression a similar effect to RNAi knockdown can be achieved, albeit at the DNA level rather than at the level of mRNA.

Both RNAi and CRISPR have challenges in terms target specificity – where expression can be altered in genes which were not targeted by the study. However, a recent study by Smith *et al.* 2017 showed that CRISPR has fewer off target effects compared to RNAi. Overall, I chose to use CRISPR for this project based on its efficiency and its versatility. Additionally, the *H. muscarum* genome sequence was not available and it could have been that, like some *Leishmania* species (Robinson and Beverley 2003), *H. muscarum* lacked RNAi pathway genes. However, this work has shown that *H. muscarum* does encode orthologues to the major RNAi apparatus (see Chapter Two) and as such this technique could be developed for future studies.

References

- Aly, R. *et al.* (1994) ‘A regulatory role for the 5’ and 3’ untranslated regions in differential expression of hsp83 in *Leishmania*’, *Nucleic acids research*, 22(15), pp. 2922–2929.
- Armstrong, C. M. and Goldberg, D. E. (2007) ‘An FKBP destabilization domain modulates protein levels in *Plasmodium falciparum*’, *Nature Methods*. Nature Publishing Group, 4, p. 1007.
- Bellofatto, V. and Cross, G. A. (1989) ‘Expression of a bacterial gene in a trypanosomatid protozoan’, *Science*, 244(4909), p. 1167 LP-1169. doi: 10.1126/science.2499047.
- Bellofatto, V., Torres-Munoz, J. E. and Cross, G. A. M. (1991) ‘Stable Transformation of *Leptomonas seymouri* by Circular Extrachromosomal Elements’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 88(15), pp. 6711–6715.
- Beneke, T. *et al.* (2017) ‘A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids’, *Royal Society open science*. The Royal Society Publishing, 4(5), p. 170095. doi: 10.1098/rsos.170095.
- Cong, L. *et al.* (2013) ‘Multiplex Genome Engineering Using CRISPR/Cas Systems’, *Science*, 339(6121), p. 819 LP-823. doi: 10.1126/science.1231143.
- Cruz, A. and Beverley, S. M. (1990) ‘Gene replacement in parasitic protozoa’, *Nature*, 348(6297), pp. 171–173. doi: 10.1038/348171a0.

- Dean, S. *et al.* (2015) 'A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids', *Open biology*. The Royal Society, 5(1), p. 140197. doi: 10.1098/rsob.140197.
- Dean, S., Sunter, J. D. and Wheeler, R. J. (2017) 'TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource', *Trends in Parasitology*. Elsevier, 33(2), pp. 80–82. doi: 10.1016/j.pt.2016.10.009.
- Garneau, J. E. *et al.* (2010) 'The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 468, p. 67.
- Gasiunas, G. *et al.* (2012) 'Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria', *Proceedings of the National Academy of Sciences*, 109(39), p. E2579 LP-E2586. doi: 10.1073/pnas.1208507109.
- Janssen, B. D. *et al.* (2018) 'CRISPR/Cas9-mediated gene modification and gene knock out in the human-infective parasite *Trichomonas vaginalis*', *Scientific reports*. Nature Publishing Group UK, 8(1), p. 270. doi: 10.1038/s41598-017-18442-3.
- Jinek, M. *et al.* (2012) 'A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337(6096), p. 816 LP-821. doi: 10.1126/science.1225829.
- Jinek, M. *et al.* (2013) 'RNA-programmed genome editing in human cells', *eLife*. Edited by D. Weigel. eLife Sciences Publications, Ltd, 2, p. e00471. doi: 10.7554/eLife.00471.
- Kapler, G. M., Coburn, C. M. and Beverley, S. M. (1990) 'Stable transfection of the human parasite *Leishmania major* delineates a 30-kilobase region sufficient for extrachromosomal replication and expression', *Molecular and cellular biology*, 10(3), pp. 1084–1094.
- Kelly, T. J. and Smith, H. O. (1970) 'A restriction enzyme from *Hemophilus influenzae*: II. Base sequence of the recognition site', *Journal of Molecular Biology*, 51(2), pp. 393–409. doi: [https://doi.org/10.1016/0022-2836\(70\)90150-6](https://doi.org/10.1016/0022-2836(70)90150-6).
- LeBowitz, J. H. *et al.* (1990) 'Development of a stable *Leishmania* expression vector and application to the study of parasite surface antigen genes', *Proceedings of the National Academy of Sciences*, 87(24), p. 9736 LP-9740. doi: 10.1073/pnas.87.24.9736.
- Liu, Q. and Kadowaki, T. (2018) 'Gene knock out of honey bee trypanosomatid parasite, *Lotmaria passim*, by CRISPR/Cas9 system', *bioRxiv*, p. 478198. doi: 10.1101/478198.
- Madeira da Silva, L. *et al.* (2009) 'Regulated expression of the *Leishmania major* surface virulence factor lipophosphoglycan using conditionally destabilized fusion proteins', *Proceedings of the National Academy of Sciences*, 106(18), p. 7583 LP-7588. doi: 10.1073/pnas.0901698106.

- Mali, P. *et al.* (2013) 'RNA-Guided Human Genome Engineering via Cas9', *Science*, 339(6121), p. 823 LP-826. doi: 10.1126/science.1232033.
- Peng, D. *et al.* (2014) 'CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*', *mBio. American Society of Microbiology*, 6(1), p. e02097. doi: 10.1128/mBio.02097-14.
- Pradeep, K. *et al.* (1994) 'An episome of *Trypanosoma brucei* can exist as an extrachromosomal element in a broad range of trypanosomatids but shows different requirements for stable replication', *Molecular and Biochemical Parasitology*, 66(1), pp. 153–156.
- Robinson, K.A. and Beverley, S.M., 2003. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Molecular and biochemical parasitology*, 128(2), pp.217-228.
- Romagnoli, B. A. A. *et al.* (2018) 'Improvements in the CRISPR/Cas9 system for high efficiency gene disruption in *Trypanosoma cruzi*', *Acta Tropica*, 178, pp. 190–195. doi: <https://doi.org/10.1016/j.actatropica.2017.11.013>.
- Rothstein, R. J. B. T.-M. in E. (1983) 'One-step gene disruption in yeast', in *Recombinant DNA Part C*. Academic Press, pp. 202–211. doi: [https://doi.org/10.1016/0076-6879\(83\)01015-0](https://doi.org/10.1016/0076-6879(83)01015-0).
- Shi, H. *et al.* (2009) 'RNA interference in *Trypanosoma brucei*: role of the n-terminal RGG domain and the polyribosome association of argonaute', *The Journal of biological chemistry*. 2009/10/30. American Society for Biochemistry and Molecular Biology, 284(52), pp. 36511–36520. doi: 10.1074/jbc.M109.073072.
- Sidik, S. M. *et al.* (2016) 'A Genome-wide CRISPR Screen in *Toxoplasma* Identifies Essential Apicomplexan Genes.', *Cell*. United States, 166(6), p. 1423–1435.e12. doi: 10.1016/j.cell.2016.08.019.
- Smith, I., Greenside, P. G., Natoli, T., Lahr, D. L., Wadden, D., *et al.* (2017) Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map. *PLOS Biology* 15(11): e2003213.
- Sugino, M. and Niimi, T. (2012) 'Expression of Multisubunit Proteins in *Leishmania tarentolae* BT -Recombinant Gene Expression', in Lorence, A. (ed.). Totowa, NJ: Humana Press, pp. 317–325. doi: 10.1007/978-1-61779-433-9_16.
- Tetaud, E. *et al.* (2002) 'A new expression vector for *Crithidia fasciculata* and *Leishmania*', *Molecular and Biochemical Parasitology*, 120(2), pp. 195–204. doi: [https://doi.org/10.1016/S0166-6851\(02\)00002-6](https://doi.org/10.1016/S0166-6851(02)00002-6).
- Wright, W. D., Shah, S. S. and Heyer, W.-D. (2018) 'Homologous recombination and the repair of DNA Double-Strand Breaks', *Journal of Biological Chemistry*.

CHAPTER SIX

Conclusions and Outlook

CHAPTER SIX:

Conclusions and Outlook

Addressing the thesis aims

The overarching aim of this thesis was to investigate the interactions of *Herpetomonas muscarum*, a trypanosomatid of wild *Drosophila melanogaster* populations, with its insect host. *Drosophila* and *Herpetomonas* provide a new model system to identify and investigate evolutionarily conserved aspects of dipteran-trypanosomatid interactions. These studies are important as Dipteran-transmitted trypanosomatid diseases, human African trypanosomiasis and Leishmaniasis, are a significant burden on global health. By identifying the conserved insect and trypanosomatid responses to infection, novel, and crucially, broad-spectrum transmission blocking strategies can be developed.

This project: produced the first high-quality, annotated genome sequence for a *Herpetomonas* species, identified interactions between the fly and trypanosomatid in the fly crop and midgut, and characterised the *H. muscarum* transcriptomic response to ingestion by *D. melanogaster*. Several *H. muscarum* transcriptomic responses to ingestion were consistent with disease-causing trypanosomatids in their respective vectors. These included significant changes in cell surface protein transcripts, upregulation of sugar/amino acid transporters putative increases in autophagy. Overall, the in-fly *H. muscarum* transcriptome most closely resembled that of *L. major* nectonomads and midgut dwelling *T. brucei*. As such the model is likely to be most useful for the latter stages of trypanosomatid midgut infections.

The view from the host

The research herein complements a recently published study performed by the Ligoxygakis group (including myself) which characterised the *D. melanogaster* responses to trypanosomatid ingestion (Wang, Sloan and Ligoxygakis 2019). Briefly, the study RNA-seq study highlighted the importance of the IMD pathway, its associated AMPs, reactive oxygen species, and STAT signalling in controlling *H. muscarum* loads and clearance throughout infection. Whilst the importance (and power) of the robust AMP and ROS responses are clear from this work, the upregulation of STAT-signalling in response to infection was particularly interesting. The upregulation of STAT increased turnover of cells in the midgut epithelium in response in *H. muscarum* ingestion. In addition to this, we observed the transcripts for lectins 28Ca (CG7106) and galC1(CG9976) were downregulated in response to parasite ingestion. These carbohydrate binding proteins present in the cell membrane of gut epithelial cells are known to interact with trypanosomatid surface molecules and have been implicated in the establishment of *Leishmania* infections (Kamhawi *et al.* 2004, Dillon *et al.* 2006). Work described in Chapter Four highlighted the changes in cell surface protein transcription in ingested *H. muscarum* cells. It is possible that some of the upregulated proteins' may interact with molecules on the surface of the fly epithelium – including lectins. From these data, I hypothesize that the described fly epithelial responses act to prevent *H. muscarum* attachment to the epithelium and therefore prevent *H. muscarum* from persisting in the fly midgut for extended periods by binding to the epithelial cells. The increased epithelial turnover may facilitate the shedding of already attached *H. muscarum* as well as promoting the 'flushing' of the trypanosomatids out of the gut with faeces.

The trypanosomatids are not alone in the fly digestive tract. There is also potential to interact with many constituents of the insect's microbiome. These interactions and their impact on infection outcome are yet to be fully considered in the context of the *Drosophila-Herpetomonas* model.

It is known that the resident bacterial and fungal species in the insect gut can greatly influence the establishment and outcome of infection. Interactions between trypanosomatids and the microbiota interactions can be indirect, e.g. competing for nutrients/space, the activation/repression of insect pathogen surveillance/detection and immune responses. The microbiota can also modify the local environment to make it more/less habitable to the invading trypanosomatids - it has been suggested that bacterial species in the sand fly midgut buffer the osmolarity in the midgut which is important for *Leishmania* establishment (Louradour *et al.* 2017). Alternatively, the microbiome may have a more direct involvement through production of toxins or other anti-microbial factors. Some constituents are 'protective' to the insects, as in the case of *Wigglesworthia* in tsetse trypanosome infections (Pais *et al.* 2018), however others can also increase the likelihood of infection as in the case of *S. glossinidius* another tsetse symbiont (Welburn *et al.* 1993; Rio, Attardo and Weiss 2016, discussed further in Chapter One).

As such the microbiome of *D. melanogaster* and its interactions with both insect and trypanosomatid likely also have a role in *Drosophila-Herpetomonas* infections - though it is unclear at this stage to what extent. Differences between the microbiomes of wild flies and laboratory flies could explain the short infection durations observed during these experiments – both in larvae and adult flies. Additionally, a constituent of the wild fly microbiome may be required for *H. muscarum* to be transmitted between flies via their

food sources - in a similar way to how the sand fly microbiome can aid the establishment of *Leishmania* vertebrate infections (Dey *et al.* 2018). Clear elucidation of the role of the *Drosophila* microbiome during infection, in both wild and laboratory flies, will be crucial in the understanding and application of this new model - especially given the importance of vector microbiomes and disease transmission.

Strengths and limitations of the *Drosophila-Herpetomonas* model

Model organism *Drosophila melanogaster* is evolutionarily close to tsetse flies, and more distantly to the sand flies – though all are Dipterans. Studies with the hematophagous flies are not broadly accessible and relatively few tools and techniques have been fully developed for these organisms, though RNAi and increasing number of vector genomes are being sequenced. New technologies are being developed constantly and this gap in technology will undoubtedly close with time. In the meantime, *Drosophila melanogaster* has been extensively studied and has many well-established research tools and techniques, including extensive mutant libraries and a well-described microbiome. There are also several practical advantages to working with fruit flies: they are simple and inexpensive to rear, have a short generation time and present less risk to the researcher as they lack the skin-piercing mouthparts of trypanosomatid disease vectors.

Drosophila have been found to naturally harbour several trypanosomatids, including the subject of this work *Herpetomonas muscarum*. Comparisons of previous work by Ligoxygakis group with other works, shows there are several similarities in the way *D. melanogaster*, sand flies and tsetse respond to trypanosomatid ingestion (Wang, Sloan and Ligoxygakis, 2019). The major immune response responsible for resolution of

infections appears to be the immunodeficiency (Imd) pathway (Hu and Aksoy, 2006; Telleria *et al.*, 2012; Matetovici, Caljon and Van Den Abbeele, 2016; Tinoco-Nunes *et al.*, 2016). This pathway results in the activation of specific antimicrobial peptides, with upregulation of defensin, cecropin and attacin common to the majority of fly-trypanosomatid infections (Hao *et al.*, 2001; Boulanger *et al.*, 2002, 2004). Reactive oxygen species also play a role in trypanosomatid clearance in both tsetse and *Drosophila* (Beschlin *et al.*, 2014). However, this effect is not observed after sand flies ingest *Leishmania* (Diaz-Albiter *et al.*, 2012; Sant'Anna *et al.*, 2014). It is believed these species are able to modulate or avoid triggering this response. As well as being trypanocidal and able to trigger AMP production, ROS have roles in intestinal cell renewal. In *Drosophila*, *H. muscarum* ingestion triggered JAK/STAT-signalling and stem cell proliferation (Wang, Sloan and Ligoxygakis, 2019). Increased turnover of cells in the gut epithelium may assist in the 'flushing' of trypanosomatids out of the gut with faeces. Several JAK/STAT signalling pathway transcripts are also differentially regulated in *G. m. morisitans* after ingestion of *T. brucei* (Matetovici, Caljon and Van Den Abbeele, 2016).

The seemingly well-conserved dipteran responses to trypanosomatids, as well as the practical advantages of working with *Drosophila* are powerful advocates for the *Drosophila-Herpetomonas* model. Similarly, there are strong parallels between *H. muscarum* and disease-causing trypanosomatids. In terms of both genome structure and content, *H. muscarum* appears very similar to *Leishmania*. Similarities in content mean these species are likely to have similar tools and strategies for the avoidance of the fly immune responses and establishment of infection. This has, to some extent, been demonstrated by strong parallels between the *Leishmania* and *H. muscarum* transcriptomic responses to ingestion by flies. These conserved transcriptomic changes

were seen in transcripts related to several important cellular processes including autophagy, nutrient acquisition, metabolism and surface remodelling.

Whilst the interactions of *H. muscarum* and *D. melanogaster* resemble that of disease-causing trypanosomatids and their vectors in several ways, there are fundamental differences – one of the most crucial of which is the blood meal. Sand flies and tsetse are hematophagous, trypanosomes and *Leishmania* are ingested when suspended in a ‘warm’, (mostly) sterile, solution containing blood cells, immune cells and factors, proteins, lipids etc. In contrast, *H. muscarum* is likely acquired by *D. melanogaster* from its food source of fermenting fruit/plant product. These environments present very different challenges in term of pH, temperature, nutrient availability, competitors and harmful molecules. It is well known that the composition of the blood meal, likely serum components including complement factors, influences establishment of infection in tsetse-trypanosome infections (Maudlin *et al.* 1984, Aksoy *et al.* 2003, Walsh *et al.* 2011, Dyer *et al.* 2013). The repeated exposure to fresh blood meals during established infections as the vector provides additional stresses to trypanosomatids infecting hematophagous insects. The differences in meal not only effect the parasite but also the vector insect. For example, the temperature stress of a blood meal is significant to the insect and can alter body temperature by as much as 10°C (McAlister 2017). This in turn likely has a great impact on host-parasite interactions and disease transmission. Despite the inclusion of serum in the axenic culture media for *H. muscarum*, the *Drosophila-Herpetomonas* model does not fully consider, or model, this important aspect of trypanosomatid-vector interactions.

Several additional differences between model and subject were highlighted by this work.

I observed that oral feeding of *D. melanogaster* with *H. muscarum* cultured *in vitro*

resulted in short lived infections lasting 3-4 days. Additionally, parasites were not transmitted between infected and uninfected adults which were cohoused. Infected tsetse, once an infection has become established, can continue to produce infective metacyclic trypanosomes for the duration of the life span – approximately 40 days (Gibson and Bailey, 2003). Similarly *Leishmania* infections of sand flies span several weeks (Stamper *et al.*, 2011). It is commonly observed the pathogens show reduced fitness, or lose the ability to cause establish infection entirely after periods of *in vitro* culture. However, the duration of infection did not change when *H. muscarum* were serially passaged through flies, rather than feeding flies axenic *H. muscarum* cultures. As such it is possible that during the 5+ years of *in vitro* culture that *H. muscarum* has irreversibly lost the ability to establish infection in *D. melanogaster*. In future works it would be useful to go back into the field and assess the infectious loads/duration of infections in wild flies. I would hypothesize that *H. muscarum* infections may be more persistent in wild flies that I have observed in the laboratory. Additionally, the ability of *H. muscarum* which has not been cultured *in vitro* to infect laboratory strains of flies could then also be assayed. This is an important consideration as the outcomes of infection can be influenced by both host and trypanosomatid genotypes. It may be that *H. muscarum* remains able to establish infection in some flies, but not the laboratory strains. It is well known that fly lines which have been laboratory lines for many years have many differences to their wild relatives including the microbiome (Staubach *et al.* 2013).

An important difference between *H. muscarum* and the disease-causing trypanosomatids is that, unlike *Trypanosoma* and *Leishmania*, *H. muscarum* does not appear to differentiate after ingestion by *D. melanogaster*. Nor do they differentiate in response to temperature and pH changes *in vitro*. I have only observed *H. muscarum* with cell

morphology typical of promastigotes. Whilst the *H. muscarum* genome encodes a relatively complete repertoire of differentiation-regulating ribosomal binding proteins, it lacks orthologues for several genes associated with differentiated forms including: metacyclic-associated small hydrophilic endoplasmic reticulum-associated proteins (SHERPs), hydrophilic acylated surface proteins (HASPs), and bloodstream alanine-rich proteins (BARPs) (Doehl *et al.*, 2017). Furthermore, *H. muscarum* contains an expanded repertoire of repressor of differentiation kinase 1 (RDK1). As such *H. muscarum* may lack the ability to differentiate entirely. The absence of differentiation events restricts the model to only representing trypanosomatid promastigote forms in their respective vectors. Analysis of the *H. muscarum* transcriptome after ingestion by the fly suggests the model best represents the later stage midgut dwelling *Leishmania* and *Trypanosoma* cells found in (or migrating to) the insect proventriculus, though there were also clear parallels to earlier midgut stages.

Utilization of other trypanosomatid species within the *Drosophila* model may allow better representation of other stages of trypanosomatid-vector infections. Several other species of trypanosomatid have been shown to naturally infect *Drosophila* species, each with different infection timings and sites. Expansion of the model system to include multiple ‘monoxenous’ trypanosomatids of *Drosophila* may strengthen the model considerably. *Herpetomonas ampelophilae* is able to invade the *Drosophila* salivary glands (Rownton and McGhee, 1978) and as such may provide a better model for salivary gland dwelling *Leishmania* and trypanosomes such as haptomonads, epimastigotes and metacyclic forms. *Drosophila* larvae are strongly infected by *J. drosophilidae* with many infections persisting through metamorphosis (Hamilton *et al.*, 2015). Additionally, *Crithidia*-*Drosophila* infections have been investigated (Boulanger *et al.*, 2001). Though *Crithidia*

are not explicitly reported in *Drosophila* populations, there is some overlap in habitat between *Drosophila* and their usual hosts, bees and mosquitos, as such it is not inconceivable that they sporadically interact. In laboratory infections, *Crithidia* are normally found in the fly rectum. Studies of *Drosophila*-*Crithidia* interactions may be informative for *Trypanosoma cruzi* research as they inhabit the rectum of triatomine bugs. In this case the insect vector is not a dipteran, and as such only distantly related to *Drosophila*. Therefore, the interactions in these two systems may have fewer conserved features.

Finally, in its current state, the model lacks a clearly developed method for efficient and scalable genetic manipulation of *H. muscarum*. This is still being addressed. I have demonstrated that *H. muscarum* is amenable to genetic manipulation using homologous recombination-based strategies and produced a small set of plasmids enabling exogenous expression of foreign genes. Furthermore, progress towards a CRISPR-Cas9 genome editing tool has been made with the production of inducible Cas9 *H. muscarum* lines. As such this technology is likely to be available for use in *H. muscarum* quickly.

Future directions

This work has shed light on the infection dynamics between *Herpetomonas muscarum* and *Drosophila melanogaster*. Importantly, it has also highlighted several evolutionarily conserved trypanosomatid responses to ingestion. However, results described herein have also prompted several new questions to be addressed in future studies. I will use this space to comment on and suggest future directions for this work.

Localisation of infection at 'higher resolution'

As previously discussed, close association between clusters of *H. muscarum* cells and the fly epithelial surfaces was observed in dissected fly guts. High resolution microscopy, such as transmission electron microscopy, would confirm and assist the characterisation of putative physical interactions between *H. muscarum* cells, and between *H. muscarum* and the fly epithelial surface. In addition, qPCR detected *H. muscarum* DNA in infected flies for 3-4 days after infection, however very few trypanosomatids were seen in the dissected digestive tracts beyond 24 hours after feeding. Difficulty in finding *H. muscarum* may be due to the relatively small numbers of trypanosomatids and the comparatively large volume of the fly digestive tract. However, it also raises the possibility that *H. muscarum* leaves the digestive tract of the fly and may enter other fly tissues. Further dissection and imaging/qPCR of other fly tissue, especially those with connections to the digestive tract such e.g. salivary glands, Malpighian tubules and sexual organs, may shed more light on this.

*Is there another host for *Herpetomonas muscarum*?*

Investigation of the *Herpetomonas-Drosophila* infection dynamics raised interesting questions about the *Herpetomonas* life cycle. Given the apparent prevalence of *H. muscarum* in wild *D. melanogaster* populations (5%), it was surprising that experiments infections of adults were short lived, and the trypanosomatids were unable to become established in larvae. It is possible that there is an as-yet unknown reservoir for *H. muscarum*. As trypanosomatids are largely found in insect digestive organs, and infections are acquired by ingestion, the food source of the insect is a logical place to look for a trypanosomatid reservoir. The major food source of wild *D. melanogaster* is plant-based, e.g. decaying fruits and sap. *Herpetomonas* is evolutionarily close to the plant

pathogens of the *Phytomonas* genus and have been isolated from fruits (Marín *et al.*, 2007). As such plants seem to be a feasible reservoir of trypanosomatids. This hypothesis could be tested experimentally in future studies.

Are the identified surface proteins of interest critical to infection outcome?

Several genes of interest for future studies were identified in this work (Figure S4.15).

The transcripts for these proteins are highly upregulated in *H. muscarum* upon ingestion by the fly and predicted to be located on the surface of the cell. As such they are available for interaction with host factors and could play critical roles in the direct interaction between trypanosomatid and insect. Each has orthologues in *Leishmania* and several are upregulated during their infections of the sand fly.

Manipulation of the expression of these genes, and evaluation of the ability of the transgenic cells to infect *D. melanogaster*, will identify how critical these genes are to survival in the context of infection. These experiments are now achievable as this work also produced several plasmids to allow the tagging and knockout of these genes.

Furthermore, significant strides towards a CRISPR-Cas9 genome editing system for use in *Herpetomonas* were made by this project. Development of this system would greatly enhance the ability to manipulate these genes as many are found in large genomic arrays.

It would also then be important to identify why these genes are important during infection. To this end, their interaction partners could be elucidated using a pull down assay based approach. Labelling the proteins with a suitable tag, e.g. HaloTag®, would allow their purification. Then, using the purified proteins as bait, interaction partners could be ‘pulled down’ from *D. melanogaster* digestive tract protein lysates. ‘Captured’

proteins could then be analysed by mass spectrometry. Thanks to the extensive tools and techniques available for the *D. melanogaster* model the role/importance of any identified insect proteins could then be assessed.

Concluding Remarks

Results presented in this thesis have highlighted multiple, evolutionarily conserved, trypanosomatid responses to ingestion by dipterans by using a new *Drosophila-Herpetomonas* model. Additionally, important tools enabling further studies with this model were developed during this work - a high-quality, annotated *H. muscarum* genome sequence and several transgenic *H. muscarum* lines. Further studies of these responses, and the specific genes of interest highlighted, are now required to elucidate which are critical to infection outcome.

References

- Aksoy, S. *et al.* (2003). Interactions between tsetse and trypanosomes with implications for the control of trypanosomiasis. *Adv. Parasitol.* 53, 1–83.
- Beschin, A. *et al.* (2014) ‘African trypanosome control in the insect vector and mammalian host’, *Trends in Parasitology*, 30(11), pp. 538–547. doi: <https://doi.org/10.1016/j.pt.2014.08.006>.
- Boulanger, N. *et al.* (2001) ‘Immune response of *Drosophila melanogaster* to infection with the flagellate parasite *Crithidia* spp.’, *Insect Biochemistry and Molecular Biology*, 31(2), pp. 129–137. doi: [https://doi.org/10.1016/S0965-1748\(00\)00096-5](https://doi.org/10.1016/S0965-1748(00)00096-5).
- Boulanger, N. *et al.* (2002) ‘Immuno peptides in the defense reactions of *Glossina morsitans* to bacterial and *Trypanosoma brucei brucei* infections.’, *Insect biochemistry and molecular biology*. England, 32(4), pp. 369–375.
- Boulanger, N. *et al.* (2004) ‘Characterization of a defensin from the sand fly *Phlebotomus duboscqi* induced by challenge with bacteria or the protozoan parasite *Leishmania major*’, *Infection and immunity*. American Society for Microbiology, 72(12), pp. 7140–7146. doi: 10.1128/IAI.72.12.7140-7146.2004.

- Dey, R. *et al.* (2018) 'Gut Microbes Egested during Bites of Infected Sand Flies Augment Severity of Leishmaniasis via Inflammasome-Derived IL-1 β ', *Cell host & microbe*. 2017/12/28, 23(1), p. 134–143.e6.
- Diaz-Albiter, H. *et al.* (2012) 'Reactive Oxygen Species-mediated Immunity against *Leishmania mexicana* and *Serratia marcescens* in the Phlebotomine Sand Fly *Lutzomyia longipalpis*', *Journal of Biological Chemistry*, 287(28), pp. 23995–24003. doi: 10.1074/jbc.M112.376095.
- Dillon, R. J. *et al.* (2006) 'Analysis of ESTs from *Lutzomyia longipalpis* sand flies and their contribution toward understanding the insect–parasite relationship', *Genomics*, 88(6), pp. 831–840.
- Doehl, J. S. P. *et al.* (2017) '*Leishmania* HASP and SHERP Genes Are Required for *In Vivo* Differentiation, Parasite Transmission and Virulence Attenuation in the Host', *PLOS Pathogens*. Public Library of Science, 13(1), p. e1006130.
- Dyer NA, Rose C, Ejeh NO and Acosta-Serrano A (2013) Flying tryps: survival and maturation of trypanosomes in tsetse flies. *Trends in Parasitology* 29(4): 188–196.
- Gibson, W. and Bailey, M. (2003) 'The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes', *Kinetoplastid biology and disease*. BioMed Central, 2(1), p. 1. doi: 10.1186/1475-9292-2-1.
- Hamilton, P. T. *et al.* (2015) 'Infection Dynamics and Immune Response in a Newly Described *Drosophila*-Trypanosomatid Association', *mBio*. Edited by K. Gull, 6(5), pp. e01356-15. doi: 10.1128/mBio.01356-15.
- Hao, Z. *et al.* (2001) 'Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis', *Proceedings of the National Academy of Sciences of the United States of America*. 2001/10/09. The National Academy of Sciences, 98(22), pp. 12648–12653. doi: 10.1073/pnas.221363798.
- Hu, C. and Aksoy, S. (2006) 'Innate immune responses regulate trypanosome parasite infection of the tsetse fly *Glossina morsitans morsitans*', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 60(5), pp. 1194–1204. doi: 10.1111/j.1365-2958.2006.05180.x.
- Kamhawi, S. *et al.* (2004) 'A role for insect galectins in parasite survival.', *Cell*. United States, 119(3), pp. 329–341.
- Louradour, I. *et al.* (2017) 'The midgut microbiota plays an essential role in sand fly vector competence for *Leishmania major*.', *Cellular microbiology*. England, 19(10). doi: 10.1111/cmi.12755.
- Marín, C. *et al.* (2007) '*Herpetomonas* spp. isolated from tomato fruits (*Lycopersicon esculentum*) in southern Spain', *Experimental Parasitology*. Academic Press, 116(1), pp. 88–90. doi: 10.1016/J.EXPPARA.2006.11.003.
- Matetovici, I., Caljon, G. and Van Den Abbeele, J. (2016) 'Tsetse fly tolerance to *T. brucei* infection: transcriptome analysis of trypanosome-associated changes in the

- tsetse fly salivary gland', BMC genomics. BioMed Central, 17(1), p. 971. doi: 10.1186/s12864-016-3283-0.
- Maudlin, I., Kabayo, J. P., Flood, M. E. T., Evans, D. A. (1984) Serum factors and the maturation of *Trypanosoma congolense* infections in *Glossina morsitans*. Z Parasitenkd 70: 11–19.
- McAlister, E. (2017) The secret life of flies. London: Natural History Museum, London.
- Pais, R. *et al.* (2008) 'The Obligate *Mutualist Wigglesworthia glossinidia* Influences Reproduction, Digestion, and Immunity Processes of Its Host, the Tsetse Fly', Applied and Environmental Microbiology, 74(19), p. 5965 LP-5974. doi: 10.1128/AEM.00741-08.
- de Paiva, R. M. C. *et al.* (2015) 'Amastin Knockdown in *Leishmania braziliensis* Affects Parasite-Macrophage Interaction and Results in Impaired Viability of Intracellular Amastigotes', PLOS Pathogens. Public Library of Science, 11(12), p. e1005296.
- Rio, R. V. M., Attardo, G. M. and Weiss, B. L. (2016) 'Grandeur Alliances: Symbiont Metabolic Integration and Obligate Arthropod Hematophagy', Trends in Parasitology, 32(9), pp. 739–749.
- Rowton, E. D. and Barclay McGhee, R. (1978) 'Population Dynamics of *Herpetomonas ampelophilae*, with a Note on the Systematics of *Herpetomonas* from *Drosophila* spp.', The Journal of Protozoology. John Wiley & Sons, Ltd (10.1111), 25(2), pp. 232–235. doi: 10.1111/j.1550-7408.1978.tb04402.x.
- Sant'Anna, M. R. V *et al.* (2014) 'Colonisation resistance in the sand fly gut: *Leishmania* protects *Lutzomyia longipalpis* from bacterial infection', Parasites & Vectors, 7(1), p. 329. doi: 10.1186/1756-3305-7-329.
- Stamper, L. W. *et al.* (2011) 'Infection parameters in the sand fly vector that predict transmission of *Leishmania major*', PLoS neglected tropical diseases. Public Library of Science, 5(8), pp. e1288–e1288. doi: 10.1371/journal.pntd.0001288.
- Staubach, F. *et al.* (2013) 'Host species and environmental effects on bacterial communities associated with *Drosophila* in the laboratory and in the natural environment', PloS one. Public Library of Science, 8(8), p. e70749.
- Telleria, E. L. *et al.* (2012) 'Caspar-like Gene Depletion Reduces *Leishmania* Infection in Sand Fly Host *Lutzomyia longipalpis*', Journal of Biological Chemistry , 287(16), pp. 12985–12993. doi: 10.1074/jbc.M111.331561.
- Tinoco-Nunes, B. *et al.* (2016) 'The sandfly *Lutzomyia longipalpis* LL5 embryonic cell line has active Toll and Imd pathways and shows immune responses to bacteria, yeast and *Leishmania*', Parasites & vectors. BioMed Central, 9, p. 222. doi: 10.1186/s13071-016-1507-4.
- Walshe, D. P. *et al.* (2011) Post eclosion age predicts the prevalence of midgut trypanosome infections in *Glossina*. PLoS ONE 6, e26984

- Wang, L., Sloan, M. A. and Ligoxygakis, P. (2019) 'Intestinal NF- κ B and STAT signalling is important for uptake and clearance in a *Drosophila-Herpetomonas* interaction model', PLOS Genetics. Public Library of Science, 15(3), p. e1007931.
- Welburn, S. C. *et al.* (1993) 'Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies', Parasitology. 2009/04/06. Cambridge University Press, 107(2), pp. 141–145.

CHAPTER SEVEN

Materials and Methods

CHAPTER SEVEN:

Materials and methods

Organisms used in this work

Drosophila melanogaster

Drosophila melanogaster used in these experiments (for stock information see Table 7.1)

were maintained under standard conditions; 25°C, 33% humidity with a 12-hour light/12-hour dark cycle. ‘Back up’ stocks of each line used were maintained at 18°C, 57% humidity under the same light cycle. Flies were raised on conventional fly food, for reagents see Table 7.2.

Line	Flybase ID	Description
Oregon ^R	FBsn0000276	The parental flies for this stock were collected from wild populations in Roseburg, Oregon by D.E. Lancefield in 1925.
W1118	FBal0018186	Flies are white-eyed due to a spontaneous loss of function mutation in the gene CG2759. Discovered by R. Levis.
Relish ^{E20}	FBal0101572	In these flies the transcription start sites for the Imd/NFκB transcription factor Relish, have been perturbed by the P{lacW} element. As such flies do not express Relish.
moesin-GFP	FBal0083341	Flies express a moesin-GFP fusion protein. Moesin binds to and stabilises the cortical cytoskeleton in epithelial tissues and as such was used to visualise the gut epithelium during microscopy of the digested tract of flies fed <i>H. muscarum</i> .

Table 7.1 – *Drosophila melanogaster* stocks used for this work

Reagent	Mass/Volume
95% propionic acid (Sigma-Aldrich P5561), 5% phosphoric acid (Sigma-Aldrich P5811) solution	140ml
Agar (Fisher Scientific, BPE2641-1)	169g
Ethanol (VWR)	703ml
Maize Flour (Brian Drewitt)	1.8kg
Malt Extract (Brian Drewitt)	1.8kg
Methyl 4-hydroxybenzoate (Nipagin, Sigma-Aldrich W271004)	74g
Molasses (Rayners, Brian Drewitt)	500g
Soya Flour (Brian Drewitt)	216g
Water	28L
Yeast Powder (Brian Drewitt)	366g

Table 7.2–Fly food component list

Herpetomonas muscarum

In vitro culture

Herpetomonas muscarum are cultured in supplemented BHI (3% brain heart infusion broth, 2.5mg/ml haemin, 1% FCS) and incubated at 28°C. Cells were maintained in a log phase of growth by splitting every 4 days.

Determining sensitivity of *Herpetomonas muscarum* to antibiotics

H. muscarum log phase cultures (1E6 cells ml⁻¹) were incubated in serial dilutions antibiotic, in a 24-well plate, and monitored to look for growth. The minimum concentration needed to prevent growth, which also resulted in the death of all cells after 1 week, was used for drug selection post-electroporation (Table 7.3). The ranges tested were decided upon based on the *Leishmania mexicana* lethal doses already determined by the labs of Prof. Keith Gull and Dr. Eva Gluenz. This work determined that the concentrations of antibiotic required for selection in *H. muscarum* were: 60µg ml⁻¹

Hygromycin, 40µg ml⁻¹ Neomycin or 15µg ml⁻¹ Phleomycin. When these experiments were run for blasticidin, it was found that cells were still able to grow in excess of 150µg ml⁻¹blasticidin. As such it is not very cost-effective to use this antibiotic for selection with *H. muscarum*. The antibiotic puromycin was not assayed in this work.

Concentration (µg ml ⁻¹)	Cell density after 24 hours for a seeding density of 1.1E6 cells ml (cells ml ⁻¹)		Were there live cells after 5 days?
	Repeat 1	Repeat 2	
No drug control			
0	1.10E+07	1.14E+07	yes
Hygromycin			
30	2.94E+06	2.74E+06	yes
40	1.14E+06	2.14E+06	yes
50	1.39E+06	1.79E+06	yes
60	8.98E+05	1.52E+06	no
Neomycin			
5	9.67E+06	9.78E+05	yes
10	7.78E+06	8.30E+06	yes
20	4.93E+06	5.23E+06	yes
30	1.12E+06	1.10E+06	no
40	1.06E+06	1.05E+06	no
50	1.10E+06	1.08E+06	no
60	1.12E+06	1.18E+06	no
Phleomycin			
5	1.59E+06	1.11E+06	yes
10	9.59E+05	8.48E+05	yes
15	1.22E+06	1.06E+06	no
25	1.38E+06	1.04E+06	no
50	8.59E+05	7.96E+05	no

Table 7.3 – Determining required concentration of antibiotics to kill *H. muscarum* culture, seeded at 1x10⁶ cells ml⁻¹.

Transfection of *Herpetomonas muscarum* with plasmid DNA

Cells were taken from log phase *in vitro* culture and pelleted by centrifugation at 800xg for 3 minutes. The supernatant was removed and the cells resuspended in either; Amaxa parasite 1 (Lonza), Zimmerman's (Zimmermann and Vienken, 1982) or Roditi's (Burkard, Fragoso and Roditi, 2007), electroporation buffer to a concentration of 8×10^7 cells ml^{-1} . For each electroporation 200 μL of cell suspension was used. Cells were added to 0.4cm gap electroporation cassettes (Biorad) with plasmid DNA and electroporated (various conditions, see Chapter Five). After electroporation cells were added to pre-warmed media and incubated at 28°C for 12 hours before the addition of the selection drug.

Construction of *Herpetomonas muscarum* expression vectors

pMS003 construction

The components of the expression cassette were amplified by PCR so that they contained an 18bp overlap with the adjacent fragment in the cassette – for details of primer sequences and templates see Table 7.4. This allowed the fragments to be joined to each other, and the pLENT backbone using the NEBuilder Assembly kit (Table 7.5). After assembly reaction, 2 μL of the resulting mix was transformed into chemically competent *Escherichia coli* (High efficiency DH5-alpha, New England Biolabs) according to manufacturer's instructions and cultured for 16 hours on Luria broth (LB) agar plates with 100 $\mu\text{g ml}^{-1}$ ampicillin. Individual clones were then assayed by colony PCR to check for the presence of the fluorescent protein (or gene of interest). 'Positive' colonies were used to inoculate 50ml of LB with 100 $\mu\text{g ml}^{-1}$ ampicillin which was cultured, shaking at 37°C for 16 hours. Plasmid was purified from the overnight culture using the Qiagen

Midi Prep kit (cat # 12143) and the expression cassette sequenced prior to use for transfection of *H. muscarum*.

Fragment Name	Template	Primers lowercase = adjacent fragment overhang, capitals = sequence-specific primer, added restriction enzyme sites
Beta-tubulin 5' intergenic sequence	<i>H. muscarum</i> gDNA	Forward: taaaacgacggccagtccaagctc ctcgag GAACATAAAGCGTACCTCCC Reverse: cttgctcaccata aagctt GGTGCTGTTGTGTGGGTATG
tdTomato	pLENTv2-TN (Dean <i>et al.</i> 2015)	Forward: acacaacagcacc aagctt ATGGTGAGCAAGGGCGAG Reverse: aagtagaatgcata actagt CTACTTGTACAGCTCGTCCATGC
PGKA-B intergenic sequence	<i>H. muscarum</i> gDNA	Forward: gctgtacaagtag actagt ATGCATTCTACTTTTCCCG Reverse: ctgttcaatcat gaattc GTTTCTTTTGAATTGAAAAATTG
Neomycin resistance gene(neo ^R)	pLENTv2-TN (Dean <i>et al.</i> 2015)	Forward: ttcaaaaagaac gaattc ATGATTGAACAAGATGGATTG Reverse: ctcgggctct agatct TCAGAAGAACTCGTCAAGAAG
Beta-tubulin 3' intergenic sequence	<i>H. muscarum</i> gDNA	Forward: cgagttcttctga agatct AGAGCCCGAGCGAGCGAG Reverse: cctctttctcgcctgctggatctc catatg GGTGAAAGTGGGGGGGAAG

Table 7.4 – Primers for construction of novel expression vector pMS003.

Restriction sites (red): CTCGAG = XhoI, AAGCTT = HindIII, ACTAGT = SpeI, GAATTC = EcoRI, CATATG = NdeI.

Reagent	Quantity
Tryptone	10g
Yeast extract	5g
Sodium chloride	10g
Agar	20g
Distilled water	1L

Table 7.5 – Luria broth agar plate recipe. Reagents were mixed, autoclaved, cooled to 50°C and poured into 90mm petri dishes. Where relevant, antibiotics were added after cooling. When set, plates were stored for up to 3 weeks at 4°C until use.

Construction of pMS004-8

The same strategy as described above was used to modify pMS003 to include alternative genes of interest, under different selection drugs (plasmids pMS004-8). The primer sequences and templates used to amplify the different genes, with Phusion high fidelity polymerase (Nw) are detailed in Table 7.6. After thermocycling according to the manufacturer's instructions (final volume of 50µL), 10% of the reaction mix was run out on a 1% agarose gel to check the DNA products' sizes were correct, the fragments were purified from the remaining PCR reaction mix using the Qiagen PCR purification kit (cat # 28104).

To insert GFP, Cyan FP, Cas9 or DD::Cas9 into pMS003, the tdTomato open reading frame was excised by HindIII/SpeI digestion (Table 7.7) to give 1431bp and 7824bp fragments corresponding to the tdTomato open reading frame and the remainder of

pMS003 respectively. The larger fragment was then purified from the agarose gel (Qiagen Gel Extraction kit cat # 28704) for insert ligation using NEBuilder (Table 7.8). Each fragment to be inserted was added to the linearised pMS003 ‘backbone’ and then the fragments were added to 10µL of 2X NEBuild reaction mix and the volume made up to 20µL with distilled water. Reactions were incubated for 15 minutes at 50°C and 2µL of the resulting mix used to transform DH5-alpha competent *E. coli* cells (NEB). Resulting colonies were screened by PCR to look for ‘correct’ clones – these clones were then used to inoculate 50ml of LB broth with 100µg ml⁻¹ Ampicillin. After 18 hours of incubation at 37°C, plasmids were harvested from the bacteria using the Qiagen Midi Prep kit (cat # 12143). This produced several intermediate plasmids containing fluorescent proteins upstream of a neomycin resistance gene. To allow production of selectable double mutants, I then sought to change the resistance gene in these intermediate plasmids. To do this, the neomycin resistance gene was excised from the plasmids by EcoRI/BglII digestion, giving 795bp (neo^R) and 8460bp fragments. The larger fragment was then purified from the agarose gel (Qiagen Gel Extraction kit cat # 28704) for insert ligation using NEBuilder (Table 7.9).

Fragment Name	Template	Primers (lowercase = adjacent fragment overhang, capitals = sequence-specific primer, added restriction enzyme sites)
GFP	pNUS-GFPcB (Tetaud <i>et al.</i> 2002)	Forward: ataccacacacagcaccacaa agctt ATGAGATCTGGTACCATG Reverse: cgggaaaagtagaatagcataactagttCACTTGTACAGCTCGTC
Cyan FP	pENT6-RP45-CFP (gift from K. Gull)	Forward: ataccacacacagcaccacaa agctt ATGGTGAGCAAGGGCGAG Reverse: cgggaaaagtagaatagcataactagttACTTGTACAGCTCGTCCATG
Hygromycin resistance gene	pTB007 (Beneke <i>et al.</i> 2017)	Forward: ttaatttcaaaaagaaacg gaattc ATGAAAAAGCCTGAACTCAC Reverse: gctcgctcgctgggctct agaatct CTATTCCCTTTGCCCTCGG
Phleomycin resistance gene	pNUS-GFPcB (Tetaud <i>et al.</i> 2002)	Forward: ttaatttcaaaaagaaacg gaattc ATGGCCAAAGTTGACCCAGTG Reverse: gctcgctcgctgggctct agaatct TCAGTCCTGCTCCTCGGC
Cas9	pTB007 (Beneke <i>et al.</i> 2017)	Forward: tcgcacacacacacacacacacataccacacacacacacacacac actagt ACCATGGACTATAAGGACCAC Reverse: aagtagaatgcatt gatca TTCTCAGTGGTGGTGGTG
DD-Cas9	pMS001	Forward: ataccacacacagcaccacATGGGAGTGCAGGTGGAAAC Reverse: cgggaaaagtagaatagcataTTACTTTTCTTTTTCCTTGGC

Table 7.6 - Primers for construction pMS004-7

Restriction sites (red): AAGCTT=HindIII, ACTAGT=SpeI, GAATTC=EcoRI, TGATCA=BclI and AGATCT= BglII.

Reagent	Volume (μL)
NEB 10X Buffer 2.1	4
pMS003 (100ng ml ⁻¹)	30
HindIII (NEB)	1
SpeI (NEB)	1
Distilled water	4
Total Volume:	40

Table 7.7 – HindIII/SpeI digest of pMS003 to release tdTomato fragment.

Reaction mix was incubated for 37°C for 2.5 hours.

DNA fragment	Concentration (ng ml ⁻¹)	Volume (μL)	Amount (fmol)	Mass (ng)
HindIII/SpeI linearised pMS003	17.8	3	21	50
GFP	74.9	2	63	16
Cyan FP	39.8	2	63	16
Cas9	106	1	75	100
HindIII/SpeI linearised pMS003 (2)	37.4	2	15	74
DD::Cas9	40.7	1	30	43

Table 7.8 – DNA fragment amounts and volumes used for NEBuilder reactions to change the tdTomato open reading frame of pMS003 for alternative genes.

For GFP, Cyan FP and Cas9 reactions the same pMS003 digest was used, however a further pMS003 digest was required to product the DD::Cas9 construct. As such they are separated in the table by a blank row.

DNA fragment	Concentration (ng ml ⁻¹)	Volume (μL)	Amount (fmol)	Mass (ng)
EcoRI/BglII linearised plasmids	47	2	18	94
Hygromycin resistance gene	9	6	80	54
Phleomycin resistance gene	8	3	97	24
Blasticidin resistance gene	5	4	80	20

Table 7.9 – DNA fragment amounts and volumes used for NEBuilder reactions to produce pMS004-8

Microscopy

Light microscopy of *Herpetomonas muscarum* from *in vitro* culture

Cells were pelleted and washed three times in sterile phosphate buffered saline (PBS) before being resuspended in 100μL of PBS per half slide. The cell suspension was pipetted onto a glass slide, whose borders were marked out with a hydrophobic barrier (DAKO pen, Agilent Technologies Ltd.), and the cells allowed to settle for 10 minutes. The slides were moved to a dark, humidified chamber for subsequent steps. Cells were then fixed by adding 100μL of 4% paraformaldehyde to the slide, which was left at room temperature for 10 minutes. The slides were rinsed by three successive immersions in clean PBS then permeabilised by incubation in ice cold methanol for 30 minutes. The cells were then re-hydrated in PBS for 5 minutes. Nucleic acids were stained with DAPI(4',6-diamidino-2-phenylindole), a blue-fluorescent DNA stain that exhibits enhanced fluorescence when bound to double stranded DNA, by covering the cells with 100μL of 1μg ml⁻¹DAPI solution in PBS for 10 minutes. The slides were rinsed by three successive immersions in clean PBS and sealed with a coverslip and nail varnish. All

slides were stored at 4°C in the dark. Cells were imaged using the Leica DM5500 B and Andor Neo sCMOS camera system.

Transmission electron microscopy (TEM) of *Herpetomonas muscarum* cells from *in vitro* culture

Log phase cells from *in vitro* culture, 5ml of 1E6 cells ml⁻¹, were fixed by resuspension in 2ml of 16% paraformaldehyde (PFA) (w/v) and 0.8ml of 25% glutaraldehyde (v/v) in PBS at 28°C. Cells were pelleted at 10,000rpm for 5minutes and the fixatives removed. Cell pellets were washed twice in 1ml 0.1M PIPES pH 7.2 and once in 1ml 0.1M PIPES pH 7.2 with 50mM glycine. After removal of the supernatant, the cell pellets were embedded in 0.5ml 2.5% agarose (w/v). Embedded pellets were fixed in 1% osmium tetroxide (OsO₄) for 1 hour at 4°C, washed three times in 2ml milli-Q water and then incubated overnight in 0.5% uranyl acetate in milli-Q water at 4°C. Following this, cell pellets were washed for 10 minutes in 2ml of milli-Q water at room temperature on a rotor before a 30minute incubation in 30% ethanol (EtOH) on ice. The pellets were then dehydrated by immersion in increasing concentrations of 2ml EtOH (50%, 70%, 80%, 90% and 95%) for 10 minutes each on ice. The three final dehydrations were done for 30minutes in 2ml 100% EtOH each. Pellets were then incubated in 2ml of the following mixes of (v/w) ethanol: agar100 resin (Agar Scientific); 3:1 (1hour with rotation), 1:1 (2.5hours with rotation), 1:3 (1 hour with rotation). Pellets were then incubated in 2ml 100% agar100 resin for 24 hours with the resin refreshed at 18 hours and 22 hours. Finally, pellets were transferred into the base of beam capsules with 1mm² tips (Agar Scientific) and the resin set at 60°C for 18 hours. Resin blocks were ‘faced up’ using a glass knife before sectioning with a diamond knife on the Reichert MT2 Ultracut T7 Microtome. Images were taken using the FEI Technai 12 Transmission electron

microscope with the Gatan Ultrascan 1000 CCD camera and Gatan Digital micrograph and SerialEM image acquisition platforms. ImageJ software was used for image handling and file type conversions.

Imaging the *Drosophila melanogaster* digestive tract

Flies were anaesthetised using CO₂ prior to dissection. Immediately post-dissection the fly guts were fixed in 200µL of 4% paraformaldehyde in PBS for 1 hour. The guts were then rinsed three times in 500µL of PBS, on the bench at room temperature, for 15 minutes per wash. The guts were then mounted on a glass slide in a drop of halocarbon oil (VWR) and sealed with a coverslip and nail varnish. The slides were then stored at 4°C and imaged with 48 hours. Mounted digestive tracts were imaged with the Olympus FV3000 confocal system (fluorescence images) and the Zeiss Imager.Z2 microscope with Hamamatsu Orca Flash4.0 camera (phase contrast images). ImageJ software was used for image handling and file-type conversions.

***Drosophila melanogaster* infection protocols**

Staining *Herpetomonas muscarum* with Mitotracker™

H. muscarum were pelleted from log phase culture, at 1E7 cells ml⁻¹ and resuspended in PBS with 5 nM Mitotracker™ Red dye for 30 minutes. Cells were rinsed three times in sterile PBS before resuspension in *H. muscarum* growth media with 10% sucrose for feeding to *D. melanogaster*.

Per os* infection of adult *Drosophila melanogaster* with *Herpetomonas muscarum

Prior to infection seven-day old (from eclosion) adult *D. melanogaster* were taken off conventional fly food and given access only to water for 12 hours prior to infection. Log-phase *H. muscarum* cultures were pelleted at 800xg for 3 minutes, the supernatant removed and then resuspended in standard *H. muscarum* culture media with 10% sucrose to a cell density of 2×10^7 cells ml⁻¹. Half a millilitre of this cell suspension was pipetted onto two 22mm filter paper discs (Whatman™) placed in the base of a clean, dry fly vial. *Drosophila melanogaster* were added to the vial and incubated under standard conditions for up to 6 hours, depending on the experiment. After this, the flies were transferred back onto conventional fly food for the remainder of the experiment.

Per os* infection of *Drosophila melanogaster* larvae with *Herpetomonas muscarum

L₃ larvae were collected from conventional fly food three days after mated females were allowed to lay eggs in the food by immersion of food in PBS and selecting the L₃ larvae using forceps. Larvae were then added to a 1.5ml Eppendorf tube containing homogenised adult flies ‘spiked’ with 100µL of *H. muscarum* culture at a density of 2×10^7 cells ml⁻¹ prepared in the same way as for the adults (see above).

Extraction of *Herpetomonas muscarum* from adult flies

One hundred adult flies were fed *H. muscarum* as described above. Flies were anaesthetized and homogenised in 200µL of PBS, in a 1.5ml Eppendorf, using a sterile disposable needle (BD Precision Glide™) and motorised pestle (GE Healthcare). The resulting ‘slurry’ was passed through a 0.4µm sieve (Corning®) to remove larger debris. The flow through was centrifuged at 800xg for 3 minutes to pellet any *H. muscarum* and the supernatant removed. The cell pellet was washed three times with sterile PBS before

cells were resuspended in *H. muscarum* growth media for either *in vitro* culture or secondary infections.

Determination of *Herpetomonas muscarum* ‘load’ by qPCR

To determine the number of *H. muscarum* in *D. melanogaster* quantitative PCR methodology was used as follows; 1ml of *H. muscarum* cell suspension used for feeding was retained and had the gDNA extracted using the Norgen Biotek Genomic DNA Isolation kit (cat # 24700) according to the manufacturer's instructions. The purified gDNA was serially diluted to provide ‘standards’ of gDNA from 10^1 - 10^6 cells which when used as qPCR templates (see Table 7.10 for reaction mix and Table 7.11 for cycling conditions). Primers designed to amplify 162 bp of *H. muscarum* paraflagellar rod protein 1D (Figure 7.1) were used for qPCRs. Standard curves of *H. muscarum* gDNA verses Ct values were produced for every experimental infection. By measuring the Ct value of the gDNA samples from infected flies, the absolute number of parasite cells could be interpolated from the standard curve. PCR controls with no gDNA template or gDNA from uninfected *D. melanogaster* were also run for each experiment.

Reagent	Volume
gDNA Template	4µL
2X Sensi-Fast SYBR® No-ROX Mix	10µL
10µM <i>H. muscarum</i> PFR2 Forward Sequence 5'-3': GGACTGCTGGAACAAGATC	0.8µL
10µM <i>H. muscarum</i> PFR2 Reverse Sequence 5'-3': AGCTTCTTGTGCTGGGAG	0.8µL
Distilled H ₂ O	4.4µL

Table 7.10 – Reaction mix (per sample) for quantification of *Herpetomonas muscarum* in *Drosophila melanogaster* from gDNA extractions. Primers amplify a 162 bp piece of the *H. muscarum* PFR1D orthologue (HMUS01147600).

Cycles	Temperature	Time
1	95°C	3 min
40	95°C	5 seconds
	65°C	10 seconds
	72°C	15 seconds

Table 7.11 – qPCR cycling condition for quantification of *Herpetomonas muscarum* in *Drosophila melanogaster* from gDNA extractions.

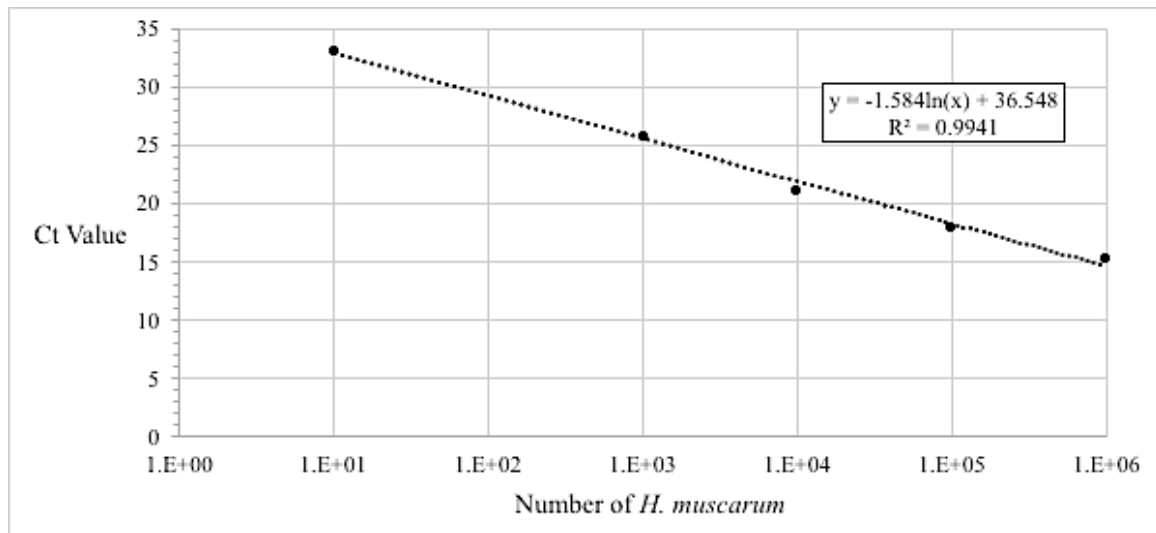


Figure 7.1 – A representative standard curve from the *Herpetomonas muscarum* gDNA standards and an uninfected fly control.

***Drosophila* egg-laying assay**

Flies were infected as described earlier in this section, then 10 females and 5 males were incubated in cages with apple juice agar plates (Table 7.12) sprinkled with baker's yeast. Uninfected controls were run in parallel. Plates were changed daily, and the number of eggs laid counted, for a total of five days. These experiments were performed in triplicate over three consecutive weeks.

Reagent	Quantity
Agar	18g
Apple Juice	200ml
Sucrose	20g
Distilled water	600ml

Table 7.12 – Apple juice agar plate reagents.

Reagents are autoclaved before pouring into 90mm petri dishes. Set plates were stored at 4°C until use.

The *Herpetomonas muscarum* genome

Next-generation Genome Sequencing: an overview

This section aims to give an overview of the sequencing methodology used to sequence the *H. muscarum* genome. Specific details of the sequencing performed are provided in the next section.

Early methods of genome sequencing, e.g. Sanger (Sanger et al 1977), are were costly, slow and labour intensive as they required individual gDNA fragments to be cloned into bacterial expression vectors which were then sequenced individually. As such the sequencing of an entire genome took many years and many individual sequencing machines. The development of next generation sequencing technologies has triggered an explosion in the numbers of sequenced genomes. At the forefront of these technologies is

illumina sequencing (Figure 7.2). There are four stages to Illumina sequencing: library preparation, cluster amplification, sequencing by synthesis and sequencing assembly.

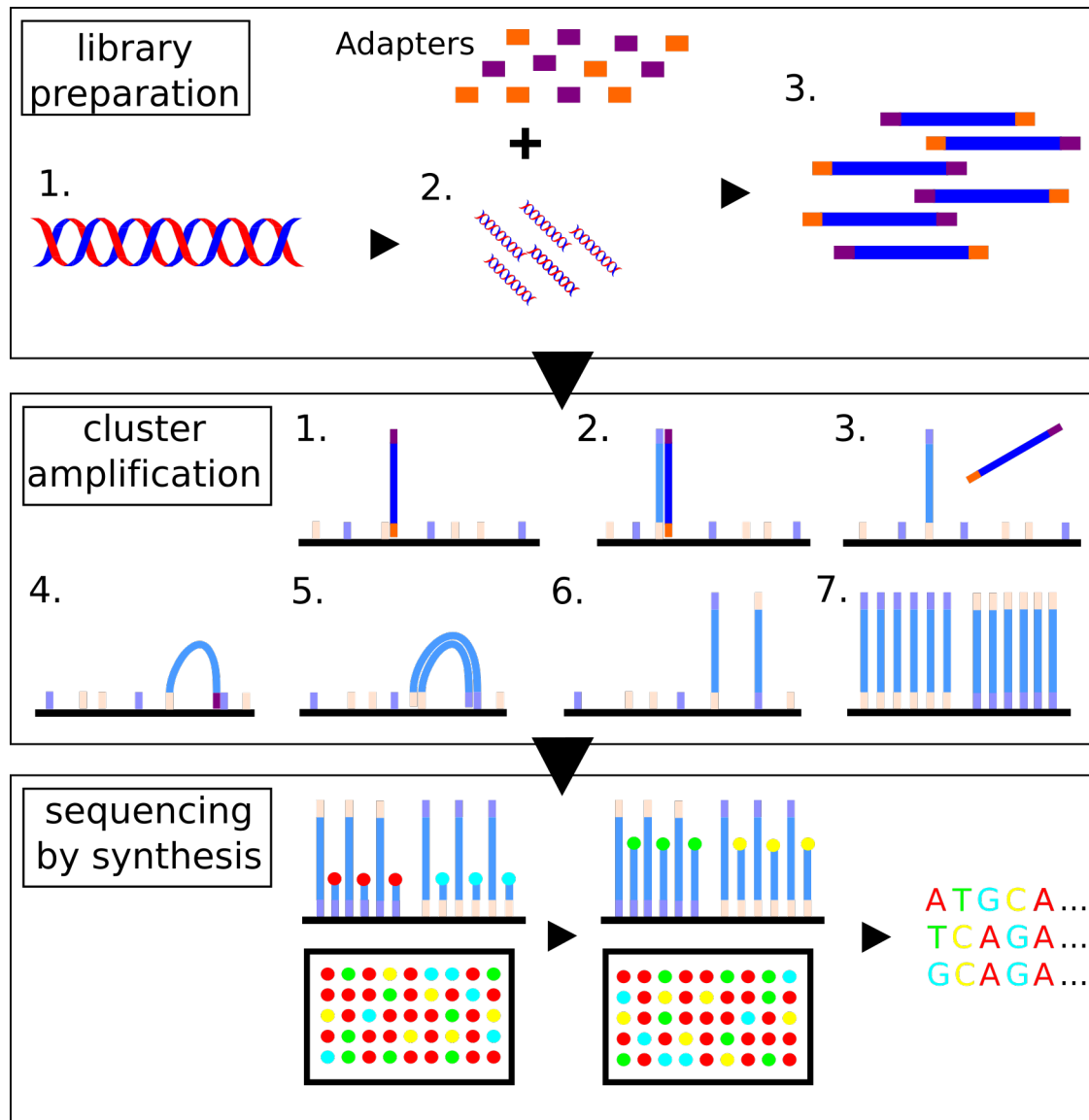


Figure 7.2 – Illumina Sequencing. See the following page for the full legend.

Figure 7.2 – Illumina Sequencing. Library preparation. 1. The genomic DNA (gDNA) is sheared into smaller fragments by sonication or similar. 2-3. Adapters are annealed to either end of the gDNA such that either end has a different adapter.

Cluster amplification. 1. Single stranded fragments anneal to adapter sequences bound to the flow cell surface (black) by complementary base pairing. 2. Synthesis of the second strand upwards from the adapter sequence bound to the flow cell surface. 3. Unbound DNA is washed from the flow cell. 4. The newly synthesized, now single stranded, fragment anneals to adapter sequence at free end by fragment bending. 5. Synthesis of the second strand upwards from the adapter sequence bound to the flow cell surface. To give sequences complementary to the original gDNA now bound to the flow cell surface (6). This process is repeated to produce clusters of clonal DNA fragments bound to the flow cell surface (7). **Sequencing by synthesis.** Strands are synthesized using fluorescently labelled nucleotides (red, blue, green, yellow circles) which give off a signal as the nucleotide is incorporated into the growing strand. The signal is after each nucleotide addition. These signals can later be resolved to give the fragments sequence.

Library Preparation

Purified genomic DNA (gDNA) is fragmented by ultrasonication into 200-500bp fragments. This fragmentation will produce DNA fragments whose ends are either blunt, have 3' an overhang or a 5' overhang. The non-blunt DNA ends are 'repaired' by incubating fragments with DNA polymerases and polynucleotide kinases to ensure all fragments blunt ended and have phosphorylated 5' termini. A single adenine is added to the 3' end of each fragment to allow adapter sequences with a 5' thymine to be ligated on.

Each fragment requires a different adapter at either end to be hybridised to a flow cell, amplified and sequenced in the downstream reactions.

Additional sequences can be added to the gDNA fragments during this stage of library preparation. By including a library-specific index sequences (or barcodes) as well as the flow cell adaptors, multiple libraries can be sequenced simultaneously in the same sequencing run – a process called multiplexing. Using these barcodes, the DNA sequencing reads can later be grouped according their respective source libraries prior to alignment and analysis. Multiplexing has dramatically reduced the both the cost and time scales required to sequence a genome. It does however come with the caveat that additional steps are needed to resolve the sequencing reads to the correct libraries and that this step has an associated error rate. Sequencing reads which ‘hop’ between libraries can cause misalignments and inaccurate sequencing results.

After a ‘clean-up’ step to remove adapter dimers and select fragments of the correct size; successfully ligated fragments are then amplified by PCR using primers complementary to the adapter sequences. This increases the quantity of templates to enable quantification, enriches the proportion of successfully ligated fragments and allows the incorporation an additional adapter sequence to allow the fragments to bind to a flow cell for the sequencing reaction. Recent improvements (Kozarewa et al 2009) to this protocol have removed the need for this PCR amplification step by including the flow cell adapter sequences in the initial ligation stages. In doing so this reduces the amplification bias (reduced amplification of AT-rich regions) and base misincorporation associated with PCR.

An important stage is quantification of the DNA library. When the fragments are bound to the surface of the flow cell and amplified they will form clonal DNA clusters. The concentration of DNA added to the flow cell determines the cluster density. If the density is too high, then the clusters can overlap; too low and the data yield suffers.

Cluster amplification

As a result of the PCR reaction, the gDNA fragments have an adapter at one end that is identical to one of the primers bound to the surface of the flow cell. At the opposite end they have the reverse complementary sequence to the other flow cell primer. When the single stranded fragments are added to the flow cell complementary fragment and flow cell adapter sequences hybridise. Polymerases then extend upwards from the flow cell primer to generate a reverse-complementary sequence of the original gDNA fragment which is tethered to the flow cell surface. The original gDNA fragments are not bound to the flow cell surface and are subsequently washed from the flow cell. The process is repeated, as the free-end of the newly synthesized strand is complementary to another flow cell-bound primer. This generates a double stranded piece of DNA which is bound at either end to the flow cell. The strands of this product are denatured using formamide, giving two single strands of DNA, bound at one end to the flow cell. The free ends are free-one again to producing anneal to another primer on the flow cell surface. Repeated cycles of formamide denaturation, annealing, and extension give a cluster in the order of thousands of clonal DNA strands.

Sequencing-by-synthesis

The clusters of DNA strands are then sequenced by synthesis using fluorescently labelled nucleotides. As each nucleotide is added the flow cell is imaged capturing the light

emitted by each cluster. From the wavelength of captured light after each nucleotide addition the DNA sequence can be determined.

Sequence assembly

The output of the sequencing by synthesis is a collection of DNA sequences, or reads, which must then be assembled together into longer sequences to give the finished genome. If a reference genome is available, the sequencing reads can be assembled into a genome by aligning the sequencing reads to the reference and using it as a guide to assemble the pool of sequencing reads. However, there is often no reference genome available. In these cases, the reads must be assembled *de novo*. For *de novo* assembly reads the shorter reads are assembled into larger sequences using regions which overlap between reads.

The quality of an assembly can be gauged by kmer analysis, which allows estimation of the ‘true’ genome size. This analysis is based on the principal that for any given length of DNA sequence, L , the number of possible kmers (DNA sequence of length, k) is given by the formula: $(L-k) + 1$. As such by analysing the kmer content of a newly assembled genome it is possible to estimate how close to the ‘true’ genome size an assembly is. Kmer distributions can also provide information as to repeat content and heterozygosity of a genome based on the frequencies of individual kmers.

Hybrid sequencing strategies

Whilst Illumina sequencing is capable of producing millions of reads in a single run. Genomes which contain long repetitive regions, regions with an extreme nucleotide bias (e.g. GC- or AT-rich regions) or palindromic sequences are difficult to assemble from

short reads. To circumvent these challenges, Illumina sequencing can be combined with technologies capable of producing longer sequencing reads, such as the single molecule real-time (SMRT) sequencing for *de novo* genome sequencing.

SMRT sequencing, like Illumina sequencing, utilises labelled nucleotides to ‘sequence by synthesis’ using fragments of DNA which can be in excess of 20kb. A single DNA polymerase is immobilised to the bottom of a well called a zero-mode wave guide (ZMV, Figure 7.3). ZMVs are approximately 100nm deep and 70nm in diameter their size relative to the wavelength of light which illuminates the chamber from below means that light is only detectable from 30nm at the base of the well. This allows the real-time the visualisation of individual labelled nucleotides being incorporated into the growing DNA strand. As each nucleotide is incorporated a pulse of light is given off with a wavelength specific to each nucleotide. During incorporation the fluorophore is cleaved off and diffuses outside the detectable volume in the bottom of the ZMV. The tiny detectable volumes, approx. 20 zeptolitres, of ZMV dramatically reduces background ‘noise’ from other unincorporated fluorophores. SMRT can produce average lengths of 30-45 kb though read lengths in excess of 300 kb have been reported. These longer read lengths are easier to assemble *de novo* than shorter Illumina reads.

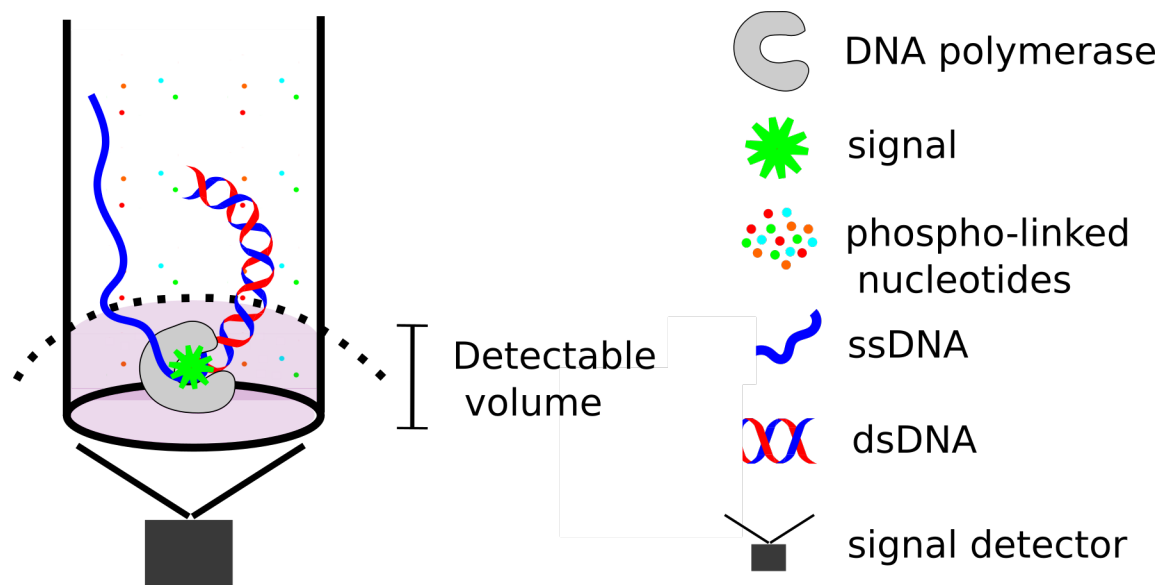


Figure 7.3 - Single molecule real time (SMRT) sequencing. A single zero-wave wave guide (ZMV). Showing sequencing by synthesis on a single DNA strand by a polymerase bound to the base of the ZMV using fluorescently labelled nucleotides. Only signal in the base of the ZMV is detectable, this this significantly reduced the ‘background signal’ from other unincorporated nucleotides/fluorophores which are shed after nucleotide incorporation.

Sequencing and assembly of the Herpetomonas muscarum genome

To sequence the *H. muscarum* genome both Illumina and Pacific Biosciences sequencing platforms were utilized. For Illumina sequencing 1 µg of genomic DNA was sheared into 300–500 base pair (bp) fragments by focused ultrasonication (Covaris Adaptive Focused Acoustics technology, AFA Inc., Woburn, USA). An amplification-free Illumina library was prepared using the methodology from (Kozarewa *et al.*, 2009) and 150 bp paired-end reads were generated on an Illumina MiSeq following the manufacturer's standard sequencing protocols (Bronner *et al.*, 2014).

For the Pacific Biosciences SMRT technology, 8 µg of genomic DNA was sheared to 20–25 kb by passing through a 25 mm blunt ended needle. A SMRT bell template library was generated using the Pacific Biosciences issued protocol for 20 kb Template Preparation using BluePippin™ Size-Selection System (Sage Science, Beverly, MA). After a greater than 7 kb size-selection the library was sequenced using P6 polymerase and chemistry version 4 (P6C4) on 6 single-molecule real-time (SMRT) cells (Eid *et al.*, 2009).

All generated reads were assembled with HGAP3 (Hierarchical Genome Assembly Process 3) (Chin *et al.*, 2013) with the genome parameter size set to 25 MB. This produced an assembly of 285 contigs which was then corrected with ICORN2 (Otto *et al.*, 2010), for five iterations. Briefly, ICORN2 (Iterative Correction of Reference Nucleotides 2) maps the short sequencing reads against the assembled scaffolds and looks for single nucleotide polymorphisms and short insertions and deletions (indel's) with a maximum length of 3 bp. These are assumed to be errors in the assembled scaffolds. Errors are corrected, and the coverage of perfect mapping reads is calculated. If the coverage after

correction is not less than the original coverage the correction is accepted. The algorithm runs until no more errors are found.

Using the Argus Optical Mapping System (OpGen), an optical map was generated from high molecular weight genomic DNA which was digested with KpnI and BamHI and captured in agarose plugs. The data was analysed with associated MapManager and MapSolver software tools (OpGen). The optical map consisted of 37-38 chromosomes with approximately half being contiguous. With the information obtained from the optical map and REAPR (Hunt *et al.*, 2013), manual genome improvement was performed on the PacBio assembly to produce a final genome assembly of 181 contigs.

Annotation

RNA-sequencing information (see relevant section below) was used to inform annotation of the *H. muscarum* sequence produced as described above. CRAM output files containing RNA sequencing reads from both *H. muscarum in vitro* culture and infected *D. melanogaster* were converted to fastq format using SAMtools (Li *et al.*, 2009), then mapped to the *H. muscarum* scaffolds using the next generation sequencing reads alignment package HISAT2 version 2.1.0 (Kim, Langmead and Salzberg, 2015). The mapped reads from each sample were assembled into transcripts with the Cufflinks package version 2.2.1 (Trapnell *et al.*, 2012) and merged to form a single set of transcripts for all generated reads.

The Companion annotation tool (Steinbiss *et al.*, 2016) was then used to generate several genome annotation files based on: the *H. muscarum* gDNA sequence, RNA sequencing transcripts and pre-existing gene models from three other trypanosomatids - *Leishmania*

braziliensis, *Leishmania major* and *Trypanosoma brucei*. These models were evaluated and combined to produce the final genome annotation (discussed in Chapter Two).

The *Herpetomonas muscarum* transcriptome

RNA extraction

***In vitro Herpetomonas muscarum* samples**

RNA was extracted from *H. muscarum in vitro* cell cultures in the both log and stationary phases of growth. For log phase cells; 45ml of culture at 1.18×10^6 cells per ml was taken from each flask and the cells pelleted by centrifugation for 10 minutes at 1000xg. The supernatant was discarded and the Norgen Biotek RNA Purification kit (cat # 17200) was used, according to manufacturer's instructions, to purify RNA from the cell pellet. This process was repeated for 5.3ml of the remaining culture three days later when the cell density was 1.21×10^7 cells per ml. The resulting RNA was eluted at concentrations 97-170 ng per μL with a $\frac{260}{230}$ absorbance of 1.86-2.19.

Infected *Drosophila melanogaster* samples

RNA extraction from infected *Drosophila melanogaster* was performed by Dr. Lihui Wang. The total RNA of 8–10 flies at; 6 hours, 12 hours, 18 hours, 30 hours, 54 hours and seven days post *H. muscarum* feeding, was extracted with the Norgen Biotek RNA Purification kit (cat # 17200) according to the manufacturer's instructions. Total RNA was also extracted from sucrose fed controls experiments done in parallel. Triplicate of samples were used for each time point.

RNA-sequencing

***In vitro Herpetomonas muscarum* samples**

Polyadenylated mRNAs were purified from total RNA using oligodT magnetic beads and strand-specific indexed libraries were prepared using the KAPA Stranded RNA-Seq kit (KR0934, v2.16) followed by ten cycles of amplification using KAPA HiFi DNA polymerase (KAPA Biosystems). The resulting libraries were quantified and pooled based on a post-PCR Agilent Bioanalyzer scores. The Illumina HiSeq v4 was then used to generate 75 bp paired-end reads following the manufacturer's standard sequencing protocols (Bronner *et al.*, 2014).

***Infected Drosophila melanogaster* samples**

cDNA libraries were prepared with the Illumina TruSeq RNA Sample Prep Kit v2. All sequencing was performed on Illumina HiSeq 2000 platform using TruSeq v3 chemistry (Oxford Gene Technology, OGT). All sequencing was paired-end and performed over 100 cycles. Read files in fastq format were generated from the sequencing platform via the manufacturer's proprietary software. An average of 26.3 million paired-end reads were sequenced per sample with a total of 788.4 million reads generated. All data can be found in the European Nucleotide Archive (ENA) with accession number PRJEB30020.

GO enrichment analysis with Ontologizer

The frequency of gene ontology (GO) term occurrence within sets of differentially expressed genes (the test sets) was compared to the number of genes associated with that GO term across the entire genome (the population set). These comparisons were carried out with Ontologizer (Vingron *et al.* 2007) using the parent-child algorithm with Bonferroni correction for multiple testing.

Pfam enrichment analysis

In order to look for Pfam domains which were enriched amongst sets of differentially expressed genes (DEGs) I performed Chi-squared statistical tests on the number of times any given domain appeared in the set of DEGs. For these tests, the number of times a domain occurs across the entire genome was used to calculate the ‘expected’ occurrence of that domain in the DEG subset. For example, if domain A occurs 300 times in the full gene set of 12 000 genes, the expected frequency for that domain in a subset of 200 genes would be 5 gene hits ($\frac{300}{12000} \times 200$).

References

- Bronner, I. F. *et al.* (2014) ‘Improved Protocols for Illumina Sequencing’, Current Protocols in Human Genetics. John Wiley & Sons, Ltd, 18(1), p. 18.2.1-18.2.42. doi: 10.1002/0471142905.hg1802s80.
- Burkard, G., Fragoso, C. M. and Roditi, I. (2007) ‘Highly efficient stable transformation of bloodstream forms of *Trypanosoma brucei*’, Molecular and Biochemical Parasitology, 153(2), pp. 220–223. doi: <https://doi.org/10.1016/j.molbiopara.2007.02.008>.
- Chin, C.-S. *et al.* (2013) ‘Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data’, Nature Methods. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 10, p. 563.
- Eid, J. *et al.* (2009) ‘Real-Time DNA Sequencing from Single Polymerase Molecules’, Science, 323(5910), p. 133 LP-138. doi: 10.1126/science.1162986.
- Hunt, M. *et al.* (2013) ‘REAPR: a universal tool for genome assembly evaluation’, Genome Biology, 14(5), p. R47. doi: 10.1186/gb-2013-14-5-r47.
- Kim, D., Langmead, B. and Salzberg, S. L. (2015) ‘HISAT: a fast spliced aligner with low memory requirements’, Nature Methods. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 12, p. 357.
- Kozarewa, I. *et al.* (2009) ‘Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes’, Nature Methods. Nature Publishing Group, 6, p. 291.
- Li, H. *et al.* (2009) ‘The Sequence Alignment/Map format and SAMtools’, Bioinformatics (Oxford, England). 2009/06/08. Oxford University Press, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

- Otto, T. D. *et al.* (2010) 'Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology', *Bioinformatics* (Oxford, England). 2010/06/18. Oxford University Press, 26(14), pp. 1704–1707. doi: 10.1093/bioinformatics/btq269.
- Sanger F, Nicklen S and Coulson AR (1977) 'DNA sequencing with chain-terminating inhibitors.' *Proceedings of the National Academy of Sciences of the United States of America* 74 (12): 5463–5467.
- Steinbiss, S. *et al.* (2016) 'Companion: a web server for annotation and analysis of parasite genomes', *Nucleic Acids Research*. Oxford University Press, 44(Web Server issue), pp. W29–W34. doi: 10.1093/nar/gkw292.
- Trapnell, C. *et al.* (2012) 'Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks', *Nature Protocols*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 7, p. 562.
- Vingron, M. *et al.* (2007) 'Improved detection of overrepresentation of Gene-Ontology annotations with parent–child analysis', *Bioinformatics*, 23(22), pp. 3024–3031. doi: 10.1093/bioinformatics/btm440.
- Zimmermann, U. and Vienken, J. (1982) 'Electric field-induced cell-to-cell fusion.', *The Journal of membrane biology*. United States, 67(3), pp. 165–182.

CHAPTER EIGHT

Appendix

CHAPTER EIGHT:

Appendix

List of supplementary materials

Chapter Two

Figure S2.1 – Alignments of the strand switch region of *Leishmania major* chromosome 1 with the strand switch region of *Herpetomonas muscarum* scaffold 40.

A. Alignment of the sequences corresponding to the putative bidirectional transcription initiation site (TIS) from the strand switch regions (SSR) of *L. major* chromosome 1 and *H. muscarum* scaffold 40. These regions share 57.8% identity. **B.** Alignment of the whole strand-switch region from *L. major* and *H. muscarum* shows 47.43% identity. The transcription initiation site from *L. major* is included for reference.

Figure S2.2A – Venn diagram showing orthogroups shared by *Herpetomonas muscarum* and two ‘old world’ species of *Leishmania*. Numbers of orthogroups (percentage of orthogroups).

Figure S2.2B - Orthogroups shared by *Herpetomonas muscarum* and two species of *Leptomonas*. Numbers of orthogroups (percentage of orthogroups).

Table S2.1 – Coordinates of strand switch regions of *Herpetomonas muscarum*.

Table S2.2 - Coordinates of telomeric repeat regions of the *Herpetomonas muscarum* genome.

Table S2.3 – Table of orthologous groups of genes from 14 trypanosomatid genomes of the *Trypanosoma*, *Leishmania*, *Leptomonas*, *Phytomonas*, *Crithidia* and *Herpetomonas* genera and the kinetoplastid *Bodo saltans*.

Table S2.4 – A table of *Herpetomonas muscarum* orthologues to *Trypanosoma brucei* genes of interest.

Table S2.5 - Coordinates of mini-exon gene arrays of *Herpetomonas muscarum*

Chapter three

Figure S3.1 – Mitotracker™ stained *Herpetomonas muscarum* in the midgut of *Drosophila melanogaster* at 6 hours post ingestion. Fixed guts were stained with DAPI (blue) to show epithelial cell nuclei.

Table S3.1 – DNA-seq from wild *Drosophila melanogaster* populations in the USA and Canada mapping information.

Chapter Four

Table S4.1 – Table of genes significantly differentially expressed between log and stationary phase axenic cultures of *Herpetomonas muscarum*, p-adjusted < 0.05.

Table S4.2 -Table of genes significantly upregulated greater than 2-fold in stationary phase vs. log phase axenic cultures of *Herpetomonas muscarum*, p-adjusted < 0.05.

Table S4.3 - Table of genes significantly downregulated greater than 2-fold in stationary phase vs. log phase axenic cultures of *Herpetomonas muscarum*, p-adjusted < 0.05.

Table S4.4 - Table of Pfam domains significantly enriched in transcripts significantly down regulated greater than 2-fold in stationary phase vs. log phase axenic cultures of *Herpetomonas muscarum*, p-adjusted < 0.05.

Table S4.5 – RNA-sequencing read mapping summary for reads extracted from *Drosophila melanogaster* fed with *H. muscarum*.

Table S4.6 - Table of genes significantly upregulated greater in *Herpetomonas muscarum* at six hours post ingestion by *Drosophila melanogaster* vs. log phase axenic culture, p-adjusted < 0.05.

Table S4.7 – Structural information for differentially expressed *H. muscarum* surface proteins. Structural information was acquired using the TMHMM1.0 online tool (Krogh *et al.*, 2001).

Table S4.8 - Table of genes differentially regulated between *Herpetomonas muscarum* after ingestion by *Drosophila melanogaster* vs. stationary phase axenic culture and not log phase axenic culture, p-adjusted < 0.05.

Table S4.9 – Comparison of differentially expressed genes in *Leishmania major* procyclic promastigotes during *Phlebotomus duboscqi* infection (Inbar *et al.* 2017) and *Herpetomonas muscarum* in *Drosophila melanogaster*.

Table S4.10 - Comparison of differentially expressed genes in *Leishmania major* nectonomads during *Phlebotomus duboscqi* infection (Inbar *et al.* 2017) and *Herpetomonas muscarum* in *Drosophila melanogaster*

Table S4.11 - Comparison of differentially expressed genes in *Leishmania major* metacyclics during *Phlebotomus duboscqi* infection (Inbar *et al.* 2017) and *Herpetomonas muscarum* in *Drosophila melanogaster*.

Tables S4.12 Comparison of differentially expressed genes in *Leishmania mexicana* *in vitro* cultured promastigotes (Holzer *et al.* 2006) and *Herpetomonas muscarum* in *Drosophila melanogaster*.

Table S4.13 - Comparison of differentially expressed genes in *Trypanosoma brucei* isolated from the *Glossina morsitans* midgut (Savage *et al.* 2016) and *Herpetomonas muscarum* in *Drosophila melanogaster*.

Table S4.14 - Comparison of differentially expressed genes in *Trypanosoma brucei* isolated from the *Glossina morsitans* proventriculus (Savage *et al.* 2016) and *Herpetomonas muscarum* in *Drosophila melanogaster*.

Table S4.15 - Comparison of differentially expressed genes in *Trypanosoma brucei* isolated from the *Glossina morsitans* salivary glands (Savage *et al.* 2016) and *Herpetomonas muscarum* in *Drosophila melanogaster*.

Table S4.16 – *Herpetomonas muscarum* putative surface proteins highly upregulated after ingestion by *Drosophila melanogaster* which are of interest for further study.

Chapter Five

Figure S5.1 – *Leishmania mexicana* cells transfected with pNUS-GFPcB plasmid as described in Chapter Five.

Figure S5.2 – The plasmid map and sequence for pMS003. This plasmid allows the expression of the fluorescent protein tdTomato in *H. muscarum* under selection of the antibiotic neomycin. It may also be used to knock out or tag the C-terminus of *H. muscarum* genes of interest.

Figure S5.3 – The plasmid map and sequence for pMS006. This plasmid allows the expression of the Cas9 gene from *Pseudomonas aeruginosa* in *H. muscarum* under selection of the antibiotic neomycin.

Figure S5.4 –The plasmid map and sequence for pMS007. This plasmid allows the expression of the Cas9 gene from *Pseudomonas aeruginosa*, which has an FKBP destabilisation domain fused to its N-terminus, in *H. muscarum* under selection of the antibiotic neomycin. The destabilisation domain will target the Cas9 for degradation unless the ligand Shield-1 is present.