Imperial College London

Characterization of the mosquito midgut immunity landscape and identification of novel antimicrobial peptides

Julia A. Cai

2023

Imperial College London
Department of Life Sciences

Thesis Submitted for the Degree of Doctor of Philosophy

Statement of Originality

All the work presented in this thesis is my own. Any work or contributions by others are duly noted.

Copyright Declaration

The copyright of this thesis rests with the author and its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND).

Under this licence, you may copy and redistribute the material in any medium or format on the condition that: you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

Despite all of the independent work I undertook, “it takes a village” is a quote that I think is very relevant to the completion of this degree. Over the last few years, many individuals have been necessary components of my Ph.D. journey and are deserving of my appreciation.

First, I would like to express my gratitude to Prof. George Christophides for providing me the opportunity to pursue this degree. I am extremely grateful for his invaluable insights, perspectives, and the many fruitful discussions we had, in addition to him entrusting me with numerous pedagogic responsibilities that I thoroughly enjoyed. I would also like to extend the gratitude to Dr. Dina Vlachou, who also provided crucial support throughout this degree, as well as the opportunity to become a research assistant.

In VigiLab, several other members were of instrumental help to me in which I am indebted to. Particularly, my colleagues Temesgen Kebede and Maria Grazia Inghilterra, for providing the majority of the mosquitoes used in my experiments (and at times rearing mosquitoes solely for me) and parasite cultures respectively. I would like to also acknowledge all other members of the lab for their support and advice during my time here, especially Alexander Bailey, who worked concurrently with me and made the experience much more enjoyable. Furthermore, I was fortunate enough to supervise and work alongside many students, who provided me the joy of teaching. In chronological order: Bow, Maria, Zoe, Ella, Lucy, Coco, Yuqing, Beth, Daisy, Sihao, and Nawal.

Beyond the lab, I am grateful for many of my peers and friends who provided support throughout the degree. In approximate chronological order: Emma, Amy, Clare, Helen, Ellie, Rodrigo, Daniela, Enoch, Jan, Yixin, James, Gwen, Shibo, Sima, Anna, Ignacio, Matteo, Jonathan, and many more I have gone for a pint with. Together we weathered many storms, including the COVID-19 pandemic.

Finally, I am deeply grateful for my family who have rallied behind me and provided a never-ending, steady stream of support at every moment. Their unwavering encouragement was a necessary component to me undertaking this degree, for without it would have been impossible.

Abstract

Malaria is a preventable disease which continues to impact a quarter of a billion people annually. It is caused by *Plasmodium* parasites, which can be transmitted to humans through the bites of female *Anopheles* mosquitoes. Numerous intervention strategies have been deployed to curb the spread of malaria, including the use of insecticides, drugs, and a vaccine. While these strategies have successfully reduced incidences of malaria, progress has stalled due to a variety of limitations, particularly the spread of insecticide resistance within mosquito populations. Therefore, novel intervention strategies must be explored and implemented. One promising area of research is the generation of malaria-refractory transgenic mosquitoes. Host-parasite interactions taking place in the mosquito midgut could be a critical component of developing such interventions, as the parasite undergoes a major bottleneck in this tissue during their development. In the midgut, a battery of immune responses contribute to this natural bottleneck, including the production of antimicrobial peptides (AMPs). Thus, better characterization of these midgut immune responses could enable the usage of these effectors and/or their loci in the blockade of malaria transmission. Transcriptome analyses were performed on mosquito tissues which enabled the characterization of spatiotemporal expression profiles for all genes in response to blood feeding and malaria infection, in addition to the identification of novel, putative AMPs. A clear dichotomy was detected in the midgut, where more anterior regions invest significant amounts of resources into immunity compared with digestion in the more posterior regions. These functions persist even in the absence of an immune stimulus. It is proposed that the functional polarity of the midgut is underpinned by concentration gradients of different transcription factors in a manner similar to development patterning along a tissue axis. These data can inform the generation of future transgenic mosquitoes for use in a novel malaria intervention strategy.

Table of Contents

Statement of Originality 2
Copyright Declaration 3
Acknowledgements 4
Abstract 5
Abbreviations & Acronyms 15
Chapter 1: Introduction 19
1.1 Biology of mosquitoes and malaria 20
1.2 Midgut microbiota and homeostasis 26
1.2.1 Benefits of the microbiota 26
1.2.2 Pathogenicity of the microbiota 29
1.3 Insect innate immunity 32
1.3.1 The Toll pathway 33
1.3.2 The Imd pathway 35
1.3.3 The JNK pathway 38
1.3.4 The JAK/STAT pathway 40
1.3.5 The RNAi pathway 43
1.3.6 Other defense mechanisms and immune molecules 45
1.3.7 Antimicrobial peptides 51
1.4 Malaria intervention strategies: the past, present, and future 56
1.5 Thesis aims 62
Chapter 2: Methods 63
2.1 Mosquito rearing 64
2.2 P. falciparum cultures and mosquito infections 65
2.2.1 Asexual cultures 65
2.2.2 Gametocyte cultures and mosquito infections 65
2.2.3 Oocyst enumeration 66
2.2.4 *P. falciparum* and treatment with artemisinin/MF/KIN1 67
2.3 Mosquito sample collection 69
2.3.1 Adult and midgut sample collection 69
2.3.2 Collection of life stages 71
2.3.3 *P. falciparum*-infected sample collection 71
2.4 RNA extraction 72
2.5 cDNA synthesis and qRT-PCR 72

2.6 *An. gambiae* RNA-sequencing data processing and analysis 74
2.6.1 RNA-seq data processing 74
2.6.2 Hierarchical clustering 74
2.6.3 Venn diagrams and enriched GO terms 74
2.7 *An. darlingi* data 75
2.8 Novel AMP identification 76
2.8.1 AMP prediction 76
2.8.2 *de novo* transcriptome assembly 76
2.9 RNAi 77
2.9.1 dsRNA synthesis 77
2.9.2 RNAi-mediated silencing of genes 78
2.9.3 Calculation of knockdown efficiency 78
2.10 Bacterial infections 79
2.10.1 Bacterial cultures 79
2.10.2 Bacterial infections 79
2.10.3 Bacterial survival assays 79
2.10.4 Survival assay analyses 80
2.10.5 Incubation of bacterial cultures with MF 80
2.11 Immunofluorescence assays 81
2.12 Whole-mount midgut RNA FISH 82
Chapter 3: Results I – Mosquito Transcriptomes 84
3.1 Introduction 85
3.1.1 Abstract 85
3.1.2 Background 85
3.1.3 Methods 86
3.2 RNA-sequencing data of infected midguts and whole carcasses 88
3.3 Responses to *Plasmodium* are dissimilar between the midgut and carcass 90
3.4 RNA-sequencing data of mosquito midgut tissues 103
3.5 Mosquito midgut tissues have distinct transcriptomes which are impacted by blood feeding 105
3.5.1 Genes enriched in more anterior tissues of the midgut 112
3.5.2 Blood meal-responsive genes in the more posterior regions 116
3.5.3 Sugar meal-responsive genes at earlier time points 119
3.5.4 The midgut shows a clear dichotomy of functions 124
3.6 Discussion 126

Chapter 4: Results II – Midgut Homeostasis and the Cardia 135
4.1 Introduction 136
4.1.1 Abstract 136
4.1.2 Background 137
4.1.3 Methods 138
4.2 Few genes make up the majority of the secreted transcriptome of the cardia 140
4.3 Developing a whole-mount midgut RNA FISH protocol 143
4.4 DVE controls the high AMP expression in the cardia 148
4.5 AMPs launched in response to bacterial infection are *DVE*-independent 159
4.6 High AMP expression maintains midgut sterility 164
4.7 KD of *DVE* and *REL2* reduces AMP expression throughout the midgut 168
4.8 FAMeTs in the cardia contribute to immunotolerance in the posterior midgut 170
4.9 Discussion 177
Chapter 5: Results III – Novel Antimicrobial Effectors 185
5.1 Introduction 186
5.1.1 Abstract 186
5.1.2 Background 186
5.1.3 Methods 188
5.2 Identification of novel, putative AMPs from RNA-seq datasets 190
5.3 Putative AMP screen: antibacterial activity 203
5.4 Putative AMP screen: induction post-bacterial challenge 210
5.5 Putative AMP screen: anti-*Plasmodium* activity 212
5.6 Putative AMP locus genes throughout mosquito development and tissues 217
5.7 Synthetic KIN1 shows anti-*Plasmodium* activity *in vivo* 222
5.8 MF significantly impacts asexual *Plasmodium* development *in vitro* 225
5.9 Discussion 230
Chapter 6: Conclusions and Perspectives 240
Chapter 7: References 247
Chapter 8: Appendices 267
Appendix A 268
Appendix B 271
Appendix C 291
Appendix D 295
Appendix E 300

Appendix F 301
Appendix G 307
Appendix H 308
Appendix I 310
Appendix J 313

List of Figures

Chapter 1: Introduction

Fig. 1.1.1 Malaria incidences
Fig. 1.1.2 The life cycle of *Plasmodium falciparum*
Fig. 1.1.3 The life cycle of *Anopheles gambiae* mosquitoes
Fig. 1.1.4 The adult female *Anopheles gambiae* midgut
Fig. 1.3.1 The three traditional immune pathways: the Toll, Imd (including the non-canonical branch of the JNK pathway), and JAK/STAT in *Drosophila*
Fig. 1.3.2 The RNA interference pathway
Fig. 1.3.3 Proposed models of AMP interactions with membranes
Fig. 1.3.4 Amphipathicity and folding structure of CEC1

Chapter 2: Methods

Fig. 2.3.1 Mosquito samples collected for RNA-seq

Chapter 3: Results I – Mosquito Transcriptomes

Fig. 3.2.1 PCA of the *Ag/Pf* RNA-seq samples
Fig. 3.3.1 DE genes in infected and control samples
Fig. 3.3.2 Average normalized counts of known AMPs in the *Ag/Pf* data
Fig. 3.3.3 Venn diagrams of significantly DE transcripts at each time point
Fig. 3.3.4 Venn diagrams of DE transcripts in midgut and carcass samples
Fig. 3.4.1 PCA of the midgut sections RNA-seq samples
Fig. 3.5.1 Hierarchical clustering of the midgut sections data
Fig. 3.5.2 Clusters of gene expression profiles in the midgut sections data
Fig. 3.5.3 Average expression profile of each cluster
Fig. 3.5.4 Midgut investment into expression of AMPs and digestive enzymes

Chapter 4: Results II – Midgut Homeostasis and the Cardia

Fig. 4.1.1 Graphical abstract: model of AMP expression in the midgut
Fig. 4.1.2 Cardia morphogenesis in the *Drosophila* embryo
Fig. 4.2.1 The secreted transcriptomes of the cardia
Fig. 4.3.1 Confocal imaging of *CEC2* and nuclei in the mosquito midgut
Fig. 4.3.2 Spatial visualization of highly expressed genes in the cardia
Fig. 4.3.3 3D image stacks of the cardia
Fig. 4.4.1 Midgut expression profiles of (putative) IMD pathway suppressors (WT) and relative expression of AMPs post-KD in the posterior midguts of CP-REL2 transgenic mosquitoes
Fig. 4.4.2 Midgut expression profiles of Iro-C members and relative expression of AMPs post-KD in the cardia/anterior midgut

Fig. 4.4.3 Midgut expression profile of transcription factor AGAP009771 and relative expression of AMPs and cardia genes post-KD in the cardia/anterior midgut
Fig. 4.4.4 Midgut expression profile of *DVE* and relative expression of AMPs post-KD in the cardia/anterior midgut
Fig. 4.4.5 Fluorescent imaging of *GAM1* expression in the cardia of GMM transgenic and WT mosquitoes
Fig. 4.5.1 Relative expression of AMPs and *DVE* post-*E. coli* infection
Fig. 4.5.2 Relative expression of AMPs and *DVE* post-*S. marcescens* infection
Fig. 4.6.1 Relative expression of *16S* in the posterior midgut post-*DVE* KD
Fig. 4.6.2 Relative expression of major bacterial families in the posterior midgut post-*DVE* KD
Fig. 4.7.1 Relative expression of AMPs and Eubacteria *16S* post-*DVE* and *REL2* co-silencing in the midgut
Fig. 4.8.1 Relative expression of AMPs and *16S* post-*FAMeT* KD in the midgut
Fig. 4.8.2 Average growth of *S. marcescens* and *E. coli* with MF or DMSO treatment
Fig. 4.8.3 Relative expression of AMPs, *16S*, and FAMeTs post-MF feeding in the midgut

Chapter 5: Results III – Novel Antimicrobial Effectors

Fig. 5.2.1 Expression profiles of putative and known AMPs in the midgut post-blood feeding
Fig. 5.2.2 Average normalized counts of putative AMPs in the *Ag/Pf* RNA-seq data
Fig. 5.2.3 Putative AMP selection strategies
Fig. 5.2.4 AlphaFold structural predictions of putative and known AMP proteins
Fig. 5.2.5 Charges and predicted structures of putative and known AMP proteins
Fig. 5.3.1 Kaplan–Meier survival curves of mosquitoes post-locus AMP KD
Fig. 5.3.2 Kaplan–Meier survival curves (10 days) of mosquitoes post-locus AMP KD
Fig. 5.3.3 Kaplan–Meier survival curves of mosquitoes post-cardia gene AMP KD
Fig. 5.3.4 Kaplan–Meier survival curves of mosquitoes post-putative AMP KD
Fig. 5.4.1 Relative expression of putative AMPs post-bacterial challenge
Fig. 5.5.1 The effect of putative AMP (locus and cardia genes) KD on *P. falciparum* oocyst development
Fig. 5.5.2 The effect of putative AMP KD on *P. falciparum* oocyst development
Fig. 5.6.1 Relative expression of *GAM1* and putative locus AMPs throughout mosquito development
Fig. 5.6.2 Relative expression of *GAM1* and putative locus AMPs in mosquito midgut tissues and the carcass before and after blood feeding
Fig. 5.7.1 The effect of exogenous, synthetic KIN1 on *P. falciparum* oocyst development
Fig. 5.8.1 The artemisinin and JH III biosynthetic pathways

Fig. 5.8.2 Average parasitemia of *in vitro P. falciparum* asexual cultures following co-incubation with various concentrations of artemisinin (AR), MF, or DMSO (control)
Fig. 5.8.3 The effect of MF on mosquito midgut stages of *P. falciparum*

Chapter 6: Conclusions and Perspectives
Fig. 6.1 Graphical summary of the proposed model of midgut compartmentalization

Chapter 8: Appendices
Fig. B.1 PCA of the *Ag/Pf* data prior to outlier removal
Fig. C.1 Confocal imaging of AGAP009313 mRNA and nuclei in the mosquito midgut
Fig. C.2 Confocal imaging of AGAP013543 mRNA and nuclei in the mosquito midgut
Fig. D.1 Relative expression of AMPs post-*DVE* KD in the posterior midgut
Fig. D.2 Fluorescent imaging of *GAM1* expression in the midgut of GMM transgenic and WT mosquitoes
Fig. D.3 Expression profile of FAMeTs and expression of *GAM1* post-*FAMeT* KD in the midgut
Fig. D.4 Expression profiles of *MET* and *TAI* in the midgut
Fig. D.5 Relative expression of AMPs post-MF feeding in the cardia and anterior midgut
Fig. F.1 Kaplan–Meier survival curves (10 days) of mosquitoes post-putative and known AMP KD
Fig. F.2 Kaplan–Meier survival curves (10 days) of mosquitoes post-cardia gene KD
Fig. G.1 Relative expression of remaining putative AMPs post-bacterial challenge

List of Tables

Chapter 2: Methods

Table 2.7.1 Infection intensities of *An. darlingi* midguts infected with *P. vivax* from gametocyte-carrying patients in the Amazonian rainforest of Peru

Chapter 3: Results I – Mosquito Transcriptomes

Table 3.2.1 *Ag/Pf* data infection statistics

Table 3.3.1 Significantly DE transcripts between infected and control samples

Table 3.5.1 Significantly enriched KEGG terms associated with gene clusters identified in the midgut sections RNA-seq data

Chapter 5: Results III – Novel Antimicrobial Effectors

Table 5.2.1 Expression profile and average counts of selected putative AMPs

Table 5.5.1 *P. falciparum* infection statistics and average KD efficiencies of putative AMP KD mosquitoes

Table 5.6.1 Wilcoxon rank-sum test statistics of locus AMP expression comparisons between blood- and sugar-fed tissues

Table 5.7.1 Summary of *P. falciparum* infection statistics following the addition of KIN1 or water (control) to an infectious blood meal

Chapter 8: Appendices

Table A.1 List of primers

Table B.1 Significantly enriched GO terms from DE transcripts between infected and control samples

Table B.2 Significantly enriched GO terms from DE transcripts unique to or shared between infected and control midgut samples at 24 hours

Table B.3 Significantly enriched GO terms from DE transcripts unique to or shared between infected and control carcass samples at 7 days

Table C.1 Transcripts making up the top two-thirds of the secreted transcriptome of each sample

Table C.2 Significantly enriched GO terms of genes making up the top two-thirds of the secreted transcriptome

Table E.1 Novel transcripts which fulfilled the initial putative AMP selection criteria identified from *de novo* transcriptome assemblies of the midgut sections data

Table F.1 Average KD efficiencies and Cox proportional hazard models of bacterial survival assays following putative AMP KD

Table H.1 *F*-test statistics of GLMs with a quasi-Poisson distribution examining the effect of treatment condition and replicate on *P. falciparum* infection intensity

Table I.1 Kruskal–Wallis and Dunn's test statistics with Benjamini–Hochberg false discovery rate corrections comparing relative gene expression between various mosquito life stages
Table I.2 Kruskal–Wallis and Dunn's test statistics with Benjamini–Hochberg false discovery rate corrections comparing relative gene expression in mosquito tissues of the same feeding condition
Table J.1 Wilcoxon signed-rank test statistics of the *P. falciparum* asexual stage parasitemia over time following treatment with artemisinin or MF *in vitro* compared to that of controls (DMSO-treated)

Abbreviations & Acronyms

AGO2	Argonaute 2
ALAR	alanine-rich
AMP	antimicrobial peptide
ANN	artificial neural network
APL1	*Anopheles Plasmodium*-responsive leucine-rich repeat 1
ara	araucan
ATT1	attacin 1
Bcd	Bicoid
Bruce	baculovirus IAP repeat-containing ubiquitin-conjugating enzyme
Cact	Cactus
Cad	Caudal
CASPL1	long caspase 1
cDNA	complementary DNA
CEC	cecropin
CI	confidence interval
CLIP	Clip-domain serine protease
CLK	Clock
CL3	containment level 3 laboratory
CP	zinc carboxypeptidase A1
cRBC	concentrated red blood cells
CTL	C-type lectin
DA	discriminant analysis
DAP	diaminopimelic acid
DE	differentially expressed
DEF	defensin
df	degrees of freedom
dFADD	*Drosophila* Fas-associated death domain-containing protein
Dif	Dorsal-related immunity factor
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
Dome	Domeless
Dredd	death-related ced-3/Nedd2-like caspase
DSCAM	Down syndrome cell adhesion molecule
dsRNA	double-stranded RNA
DUOX	dual oxidase
Dve	defective proventriculus
ELO	elongase of long-chain fatty acids
EPIP	enolase-plasminogen interaction peptides
ER	endoplasmic reticulum

FA	farnesoic acid/farnesoate
FAMeT	farnesoic acid O-methyl transferase
FBN	fibrinogen-related protein
FILM	Facility for Imaging by Light Microscopy
FISH	fluorescence *in situ* hybridization
Fkh	forkhead
FOXO	Forkhead box protein O
FREP	fibrinogen-related protein
GALE	galectin
GAM1	gambicin 1
GFP	green fluorescent protein
GLYR	glycine-rich
GLM	generalized linear model
GMM	GAM1-Mag-Mel
GNBP	Gram-negative binding proteins
GO	gene ontology
GST	glutathione S-transferase
Hop	Hopscotch
HPX	heme peroxidase
HR	hazard ratio
IAP	inhibitor of apoptosis
IκB	inhibitor of κB
IKK	IκB kinase
Imd	immune deficiency
IMPer	immunomodulatory peroxidase
IRD5	immune response deficient 5
Iro	Iroquois
JAK	Janus kinase
JH III	juvenile hormone III
JNK	c-Jun N-terminal kinase
KD	knockdown
KEGG	Kyoto Encyclopedia of Genes and Genomes
Ken	Ken & Barbie
KIN1	Kininogen 1
KOG1	kontroller of growth 1
LPS	lipopolysaccharide
LRIM	leucine-rich repeat immune protein
LYSC1	class C lysozyme 1/lysozyme c-1
Met	Methoprene-tolerant
MF	methyl farnesoate

mirr	mirror
ModSP	modular serine protease
NO	nitric oxide
NOS	nitric oxide synthase
NOX5	NADPH oxidase 5
NPC	Niemann-Pick type C
PAMP	pathogen-associated molecular pattern
PBS	phosphate-buffered saline
PBT	PBS, 0.1% Tween® 20
PC	principal component
PCA	principal component analysis
PCR	polymerase chain reaction
PE	primer efficiency
PGN	peptidoglycan
PGRP	peptidoglycan recognition protein
Pias	protein inhibitors of activated STAT
PIMMS43	*Plasmodium* infection of the mosquito midgut screen 43
PLA2	phospholipase A2
PM	peritrophic matrix
PO	phenoloxidase
PPP	pentose phosphate pathway
prd	paired
PRR	pattern recognition receptor
P47Rec	P47 receptor
qRT-PCR	quantitative reverse-time polymerase chain reaction
RBC	red blood cell
REL	Relish
RHEB	Ras homolog enriched in brain
RF	random forest
RISC	RNA-induced silencing complex
ROS	reactive oxygen species
RNA	ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA-sequencing
RNAse	ribonuclease
scFv	short-chain antibody
SCR	scavenger receptor
SCRBQ	SCR class B Croquemort
SGU	secreted glycoconjugate of unknown function
siRNA	short interfering RNA

SM1	salivary gland and midgut peptide 1
SMFA	standard membrane feeding assay
Socs	suppressor of cytokine signaling
SPE	Spätzle processing enzyme
Spirit	serine protease immune response integrator
SRPN	serine protease inhibitor (serpin)
STAT	signal transducer and activator of transcription
SVM	support vector machine
TAB2	TAK1-associated binding protein 2
Tai	Taiman
TAK1	transforming growth factor β-activated kinase 1
TCT	tracheal cytotoxin
TEP1	thioester-containing protein 1
TIR	Toll/IL-1R
TO	takeout
TOR	target of rapamycin
TORC	target of rapamycin complex
TPM	transcripts per million
TRAF6	tumor necrosis factor receptor-associated factor 6
Upd	Unpaired
VG	vitellogenin
VRI	Vrille
WHO	World Health Organization
WT	wildtype
20E	20-hydroxy-ecdysone

Chapter 1: Introduction

1.1 Biology of mosquitoes and malaria

Malaria is a disease which has burdened humanity for millennia, with the earliest reference dating back to an ancient Chinese document from 2,700 BC (Cox, 2010). Despite being preventable and treatable, malaria persists as a global issue and impacts over 200 million people each year, with the majority of incidences in sub-Saharan Africa (WHO, 2023; Fig. 1.1.1). The disease is caused by unicellular *Plasmodium* parasites and transmitted to humans by female *Anopheles* mosquitoes. There are five known species of *Plasmodium* which cause malaria in humans, with *P. falciparum* being the deadliest (WHO, 2023).

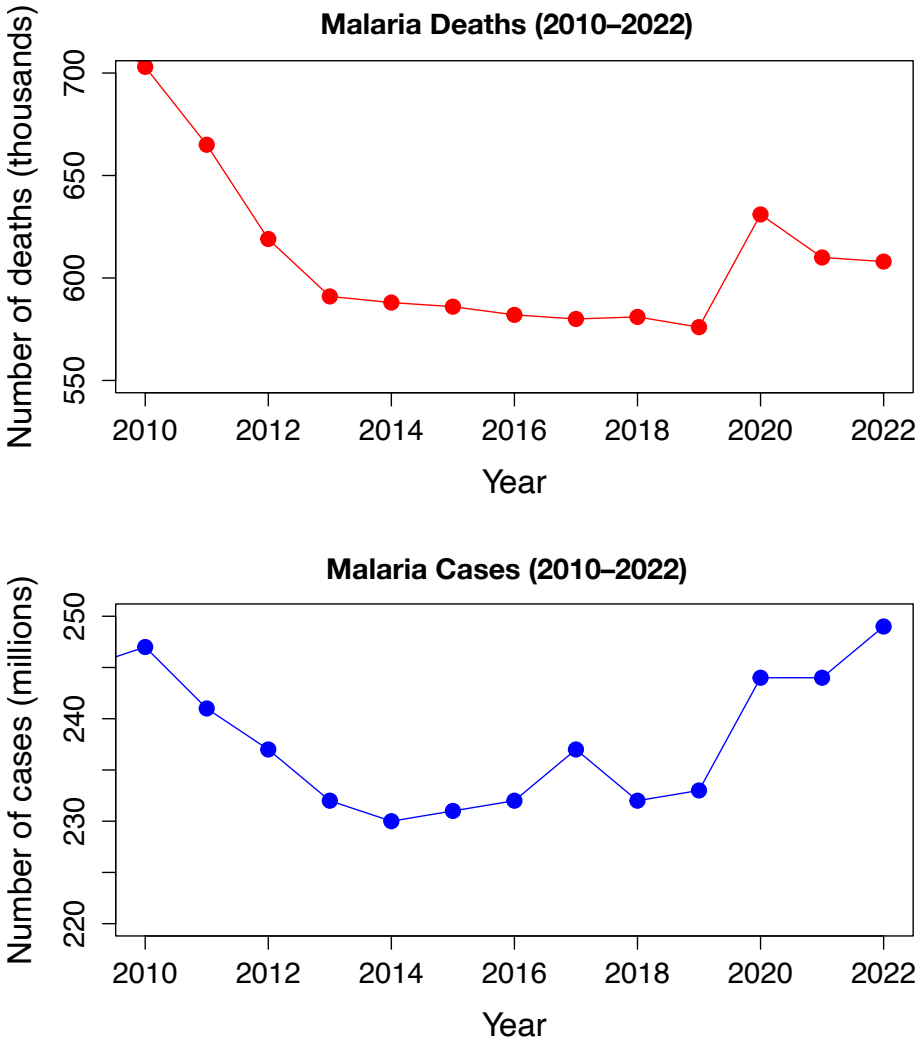


Figure 1.1.1. Malaria incidences. *The number of malaria deaths and cases between 2010 and 2022. Data obtained from WHO, 2023.*

The transmission cycle begins when a female *Anopheles* mosquito takes a blood meal from a malaria-infected human. In the mosquito midgut, male *P. falciparum* gametocytes exflagellate, while female gametocytes emerge from the host red blood cells (RBCs). Male and female gametes fuse to form a zygote, which soon differentiates into a motile ookinete. Around 24 hours from the start of the infection, ookinetes invade and traverse the midgut epithelia to develop into oocysts upon emergence from the basal side. Each ookinete develops into a single oocyst, and within each oocyst sporogonic development produces thousands of sporozoites. After about a week, the oocysts rupture and release the sporozoites which migrate to and invade the salivary glands, after which subsequent mosquito bites enable the disease to be transmitted to other human hosts.

Within the human host, sporozoites injected into the bloodstream during a mosquito bite migrate to the liver and infect hepatocytes. Within each infected hepatocyte, the parasite replicates and matures into a liver schizont, causing the host hepatocyte to rupture and release merozoites. The merozoites enter the blood stream to infect RBCs, where they undergo a continuous asexual replication, cycling from ring (immature trophozoite) to mature trophozoite stages to schizonts, which rupture the host cell, releasing merozoites which invade other RBCs. This continuous erythrocytic cycle is responsible for the known disease pathology, namely the characteristic cyclic fevers when the schizonts burst out of host cells in synchrony. Some parasites exit the cycle at the ring stage and differentiate into male and female gametocytes over a period of many days. These gametocytes can initiate a new transmission cycle upon uptake by a female mosquito taking a blood meal. The life cycle of *P. falciparum* is summarized in Figure 1.1.2.

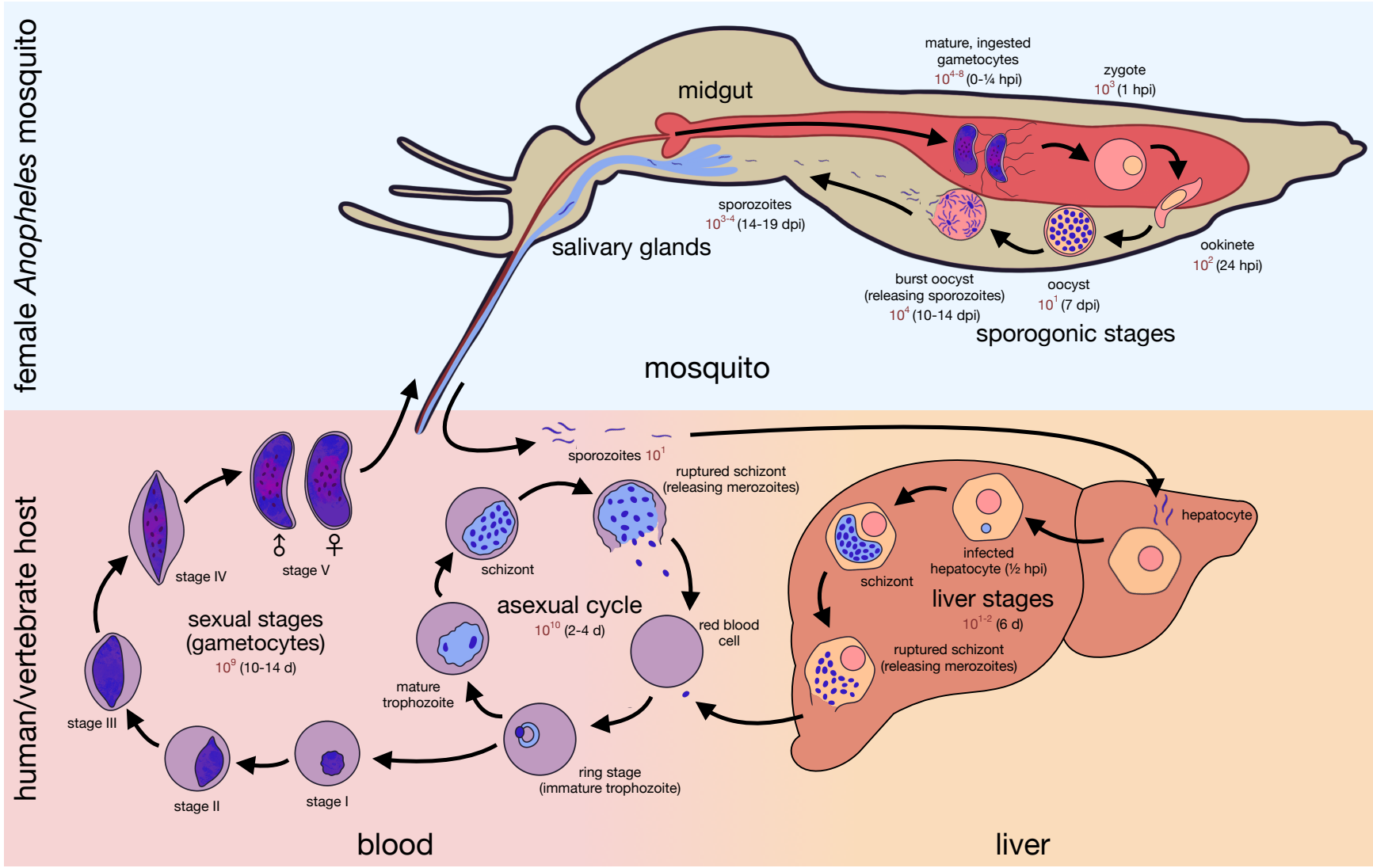


***Figure 1.1.2. The life cycle of* Plasmodium falciparum**. *Approximate parasite numbers and timings are listed in red and parentheses respectfully. d = days, dpi = days post-infection, hpi = hours post-infection. Parasite numbers obtained from Sinden, 2010.*

To date, around 500 mosquito species have been identified as belonging to the genus *Anopheles*, with approximately 50 of them known to be human malaria vectors (WHO, 2023). Nine of these species make up the *An. gambiae*[*] *senso lato* species complex, including three of the most important vectors of *P. falciparum*: *An. gambiae sensu stricto* (formerly known as the *An. gambiae* S molecular form), *An. coluzzii* (formerly known as the *An. gambiae* M molecular form), and *An. arabiensis* (O'Loughlin, 2020). The classification of the two former *An. gambiae* molecular forms as two distinct species occurred within the last decade (Coetzee et al, 2013). Members of this complex are closely related and are nearly morphologically identical, but can differ greatly in their behavior (White, 1974).

Anopheline mosquitoes undergo four major life stages in their life cycle (Fig. 1.1.3). Mosquito eggs are laid in a body of water, where they hatch into larvae 24 to 72 hours later. Larvae are aquatic and feed on organic detritus and microorganisms at the water surface. Here, larvae develop through four larval instar stages over the course of about a week. Following the final instar, larvae develop into pupae where metamorphosis takes place before adult emergence around 24 hours later. Adults typically rest in cool places during the day and are mostly active during the night or at twilight, when they form mating swarms (Mosquito Taxonomic Inventory, 2008). Adult males feed only on plant sugars (*e.g.* nectars) as an energy source, whereas females feed on both sugars and blood from warm-blooded vertebrates (Foster et al, 1995). Hematophagy is necessary for mated female mosquitoes to develop eggs, where they lay up to a few hundred eggs in a clutch two to three days post-blood meal (Clements, 1992; Foster et al, 1995). After laying a clutch of eggs, females seek a fresh blood source and begin new gonotrophic cycles, a process repeated every two to three days throughout their lifetime (Gillies, 1953).

[*] The genus in the scientific names of organisms belonging to the Culicidae family is often abbreviated with two letters as outlined by Reinert (1975).

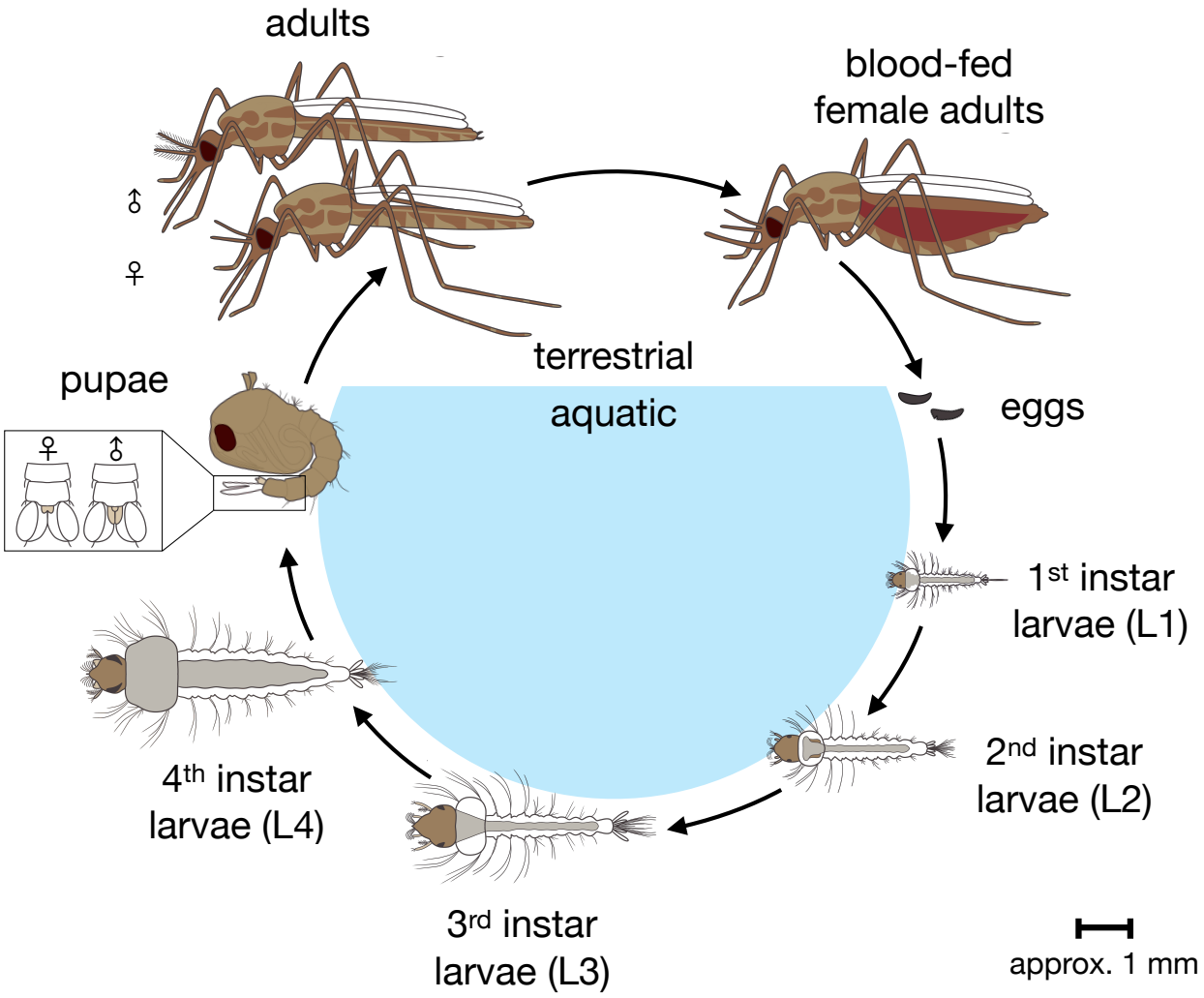


Figure 1.1.3. The life cycle of* Anopheles gambiae *mosquitoes. *Blood-fed, female adults lay eggs in the water, which hatch and develop through four larval stages. The pupal stage is the final aquatic stage, following which adults emerge and occupy terrestrial habitats. Scale bar represents ~1 mm.*

The mosquito gut is highly compartmentalized and consists of several distinct regions: the foregut, cardia (also known as the proventriculus), anterior midgut, posterior midgut, and hindgut (Fig. 1.1.4). Sugars are stored in the foregut (Clements, 1992), made up of the pharynx and three colorless sac-like structures: the ventral diverticulum (crop), and two smaller dorsal diverticula. These sugars are gradually released into the midgut, where they are absorbed and digested primarily in the anterior midgut (Billingsley & Hecker, 1991; Hixson et al, 2022). In females, blood meals are contained and digested in the posterior compartment of the midgut, which also harbors a community of commensal microbes (Rodgers et al, 2020, Hixson et al, 2022).

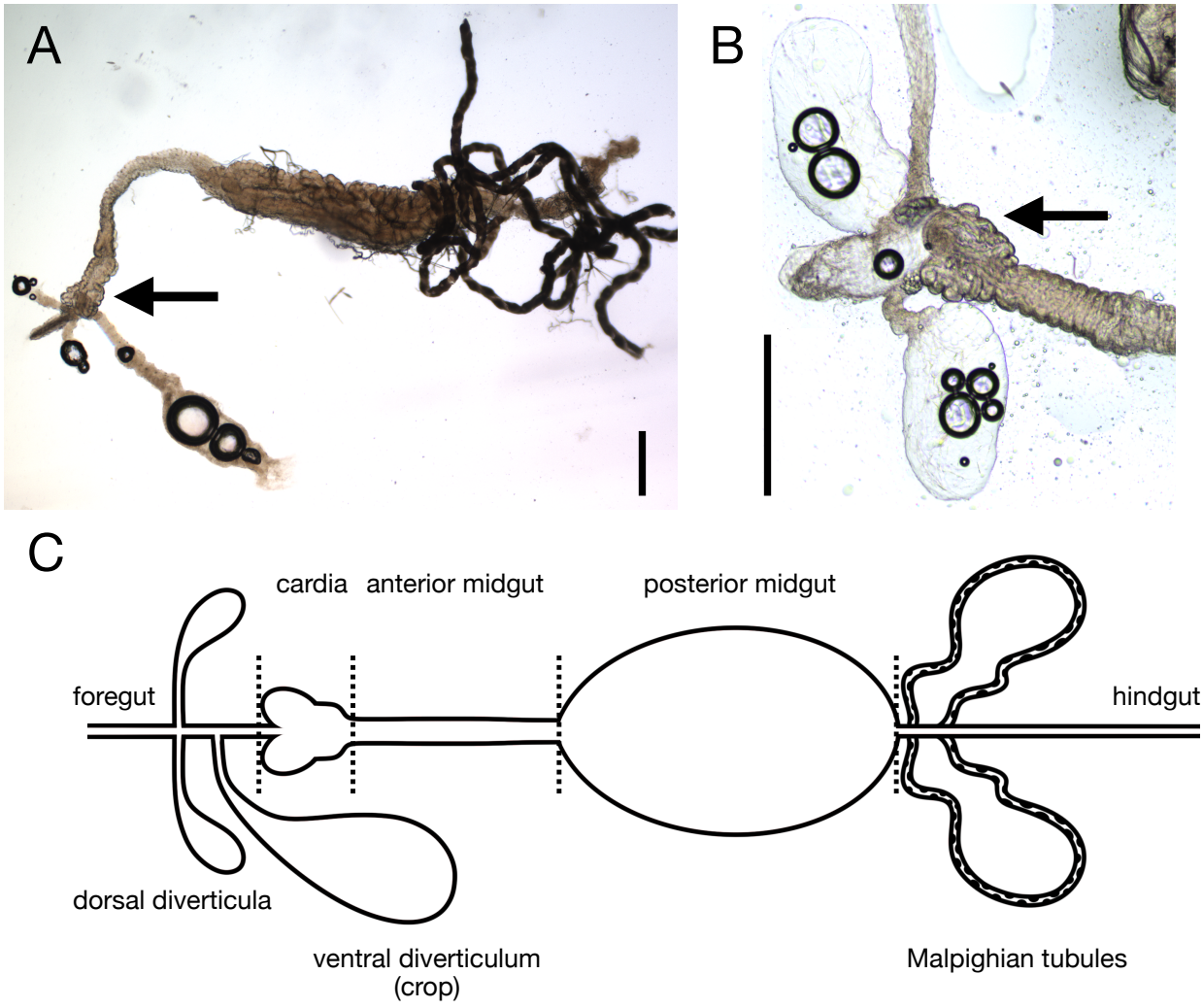


***Figure 1.1.4**. **The adult female* Anopheles gambiae *midgut.** A) A light microscope image of the entire midgut axis. The cardia is indicated with a black arrow. B) A light microscope image of the foregut and midgut junction. The cardia is indicated with a black arrow. Scale bars represent 300 µm. C) Diagram of the entire midgut axis.*

1.2 Midgut microbiota and homeostasis

Over the course of mosquito development, the environment changes drastically from aquatic to open air, where mosquitoes are exposed to variety of microbes including bacteria, fungi, viruses, and protozoans. Some of these microbes form symbiotic relationships with the mosquito host, whereas others cause damage or contribute to infection. *Anopheles* mosquitoes harbor a community of mostly Gram-negative bacteria in their midguts which are necessary for their development and physiology (Weiss & Aksoy 2011; Boissière et al, 2012; Osei-Poku et al, 2012; Coon et al, 2014; Gendrin et al, 2015; Romoli et al, 2021). Whilst most bacteria are acquired from the environment throughout development, largely via the diet, some are acquired through mating or vertically from the parents (Favia et al, 2007; Damiani et al, 2010; Weiss & Aksoy, 2011; Engel & Moran, 2013). The microbiota population is dynamic throughout the mosquito lifetime (Wang et al, 2011; Coon et al, 2014; Tchioffo et al, 2016); microbiota loads vary greatly across developmental stages and all microbiota acquired in larval stages are purged during adult emergence (Moll et al, 2001; Romoli et al, 2021). The midgut is then re-colonized with microbiota encountered later in adulthood.

1.2.1 Benefits of the microbiota

The benefits of microbiota to their host have been widely described in the animal kingdom (reviewed in Peixoto, Harkins & Nelson, 2021). This is no different in mosquitoes, where the microbiota also form crucial mutualisms with their host. Microbiota enable the acquisition of nutrients that hosts do not naturally synthesize, aid in blood meal digestion, and contribute to host development (Weiss & Aksoy, 2011; Kukutla et al, 2014; Romoli et al, 2021). In a study using *An. stephensi* mosquitoes, antibiotic clearance of *Asaia* bacteria in larvae significantly delayed development, which could be rescued upon administration of antibiotic-resistant *Asaia* (Chouaia et al, 2012). Similarly, over-colonization of the *An. gambiae* larval midgut with *Asaia* drastically increases the larval developmental rate (Mitraka et al, 2013). In addition, two studies in *Ae. aegpyti* have shown that while larvae are able to hatch from surface-sterilized eggs, axenic larvae are unable to develop past the first larval stage (Coon et al, 2014; Romoli et al, 2021). This phenotype could be rescued with the introduction of a single species of bacteria, including

one that has not been characterized before in the microbiota of wild mosquitoes (Coon et al, 2014; Romoli et al, 2021).

In a recent study by Romoli et al, a novel method to generate germ-free and gnotobiotic mosquitoes at each specific life stage was developed without the use of antibiotics (2021). Traditionally, antibiotics were used to study the effect of the microbiota on hosts, but this only reduces and modifies the microbiota load and composition respectively (Hughes et al, 2014), and can have toxic effects on the host (Chabanol et al, 2020). Romoli and colleagues found that the microbiota are necessary for folate biosynthesis, and a reduction in microbiota load at various larval stages leads to a significant detriment to larval development (Romoli et al, 2021). Together, these striking results highlight the necessity of commensal endosymbionts in the midguts of mosquitoes throughout development.

Another characterized benefit of the microbiota involves defense against other pathogenic microbes. Defensive symbioses can be achieved through a variety of mechanisms involving both direct and indirect interactions between the microbiota and pathogens. Microbiota can outcompete pathogenic microbes for space and nutrients, indirectly causing a reduction or elimination of pathogens (Weiss & Aksoy, 2011; Engel & Moran, 2013; Clay, 2014). The microbiota also aid in the immune priming of their hosts, where subsequent pathogen exposure leads to an improved response despite the lack of adaptive immunity in insects (Schmid-Hempel, 2005; Brown et al, 2019; Kulkarni et al, 2021). Following exposure to endogenous strains of *Serratia* or *Enterobacter* in sugar meals, *An. gambiae* survived better following systemic challenge of the same bacterial species (Kulkarni et al, 2021). This effect was lost when *Serratia*-primed mosquitoes were infected with *Enterobacter* and vice versa, where a more complex transcriptomic response was launched instead which did not confer the same protection (Kulkarni et al, 2021). Despite this evidence, other studies have also showed that immune priming can extend to subsequent infections of the different species or strains, or can occur as a result of physical injury (Brown et al, 2019; Cappelli et al, 2019; Powers et al, 2020).

Immune priming by the microbiota can also provide protection against infection of other pathogens. This is typically achieved through the elicitation of basal host immune responses by the microbiota to control their proliferation which in turn can also impact

other pathogens (Ryu et al, 2008; Cappelli et al, 2019; Bai et al, 2019). In fact, a considerable overlap in host immune responses to bacteria and *Plasmodium* have been characterized (Dong et al, 2006; Dong, Manfredini & Dimopoulos, 2009; Meister et al, 2009). Several studies have outlined the importance of microbiota in protecting against *Plasmodium* infection in anopheline mosquitoes. Antibiotic clearance of the microbiota in *An. gambiae* led to increased incidences of *P. falciparum* infection (Beier et al, 1994; Dong, Manfredini & Dimopoulos, 2009). Priming of laboratory-reared *An. gambiae* to seven different bacterial species isolated from wild mosquitoes significantly reduced *P. falciparum* infection intensity (Tchioffo et al, 2013). Likewise, colonization of the midgut by *Asaia* activates basal immunity in *An. stephensi*, which in turn reduces *P. berghei* infection intensities (Cappelli et al, 2019). Concomitant infection of bacteria and *Plasmodium* inhibited oocyst formation in *Anopheles*, in some cases to completion (Pumpuni et al, 1993; Gonzalez-Ceron et al, 2003; Dong, Manfredini & Dimopoulos, 2009; Meister et al, 2009; Cirimotich et al 2011; Ramirez et al, 2014).

A study by Rodrigues et al experimentally demonstrated the profound effect of immune priming by the microbiota on *Plasmodium* infection, where initial pre-exposure to *Plasmodium* ookinetes elicited a stronger immune response during a second exposure and led to a significant reduction in infection intensity (2010). During the initial infection, ookinetes traverse the midgut epithelia, compromising a barrier which also prevents bacteria from coming into contact with the midgut epithelial cells and hemolymph. This in turn causes the differentiation of prohemocytes to granulocytes in the hemolymph, which target parasites upon emergence on the basal side of the midgut after a subsequent infection. Remarkably, priming could be transferred to naïve mosquitoes when injected with the hemolymph of primed mosquitoes (Rodrigues et al, 2010). Removal of the microbiota led to the loss of all priming events, confirming the role of the microbiota in conferring this response. Taken together, these data highlight the ability of the microbiota conferring protection against *Plasmodium*.

Direct interactions between *Plasmodium* and bacteria have also been characterized. In a study by Dong, Manfredini & Dimopoulos, co-infection of *An. gambiae* with bacteria (either live or heat-inactivated) and *P. falciparum* also achieved improved levels of malaria resistance, with the effect being greater with live bacteria (2009).

Similarly, Cirimotich and colleagues found that the *Enterobacter* Esp_Z strain isolated from wild *An. arabiensis* secretes a reactive oxygen species (ROS) which inhibits multiple stages of *Plasmodium* development in a dose-dependent manner (2011). *In vitro* studies have also shown anti-plasmodial activity by *Serratia* and *Chromobacterium*, although the metabolites involved are still uncharacterized (Bahia et al, 2014; Ramirez et al, 2014).

The ability of the microbiota to alter vectorial capacity warrants further investigation into the mechanisms in which they may confer refractoriness. Currently, microbiota studies in *Anopheles* and many other organisms have focused on bacteria, whereas other organisms which make up the microbiota remain relatively uncharacterized. A study conducted by Valzano and colleagues found that an *Anopheles*-associated yeast *Wickerhamomyces anomalus* secretes a killer toxin which alters *P. berghei* membranes and kills ookinetes *in vitro* (2016). This study is just an example of the untapped potential of lesser-studied organisms making up the mosquitoes' microbiota which may have major implications for malaria transmission. These tripartite interactions between the mosquito host, microbiota, and malaria parasite highlight the need to account for microbiota in studying malaria in vectors, and warrant further investigation into their potential use as a malaria control strategy.

1.2.2 Pathogenicity of the microbiota

While the microbiota provide numerous benefits to their host, they can also become pathogenic, causing damage or even death if populations are left uncontrolled. Following a blood meal, the bacteria load in female midguts can drastically expand up to 16,000 fold as a result of the influx of nutrients (Demaio et al, 1996; Dong et al, 2009; Rodgers et al, 2020). This explosion in microbiota population can mimic an oral infection, overstimulating the mosquito immune response and increasing oxidative stress, both of which have negative impacts on mosquito fitness (DeJong et al, 2007; Oliveira et al, 2011; Gendrin et al, 2015; Champion & Xu, 2018). Reducing this proliferation via mild antibiotic treatment of the blood meal was experimentally shown to increase fecundity and survival in *An. coluzzii* mosquitoes (Gendrin et al, 2015), demonstrating the detriment of microbiota expansion. Therefore, the microbiota population needs to be tightly controlled

to allow some microbiota to persist in order for the host to reap their benefits, whilst also resisting against excessive proliferation to avoid damage/death.

In a study by Ryu and colleagues, midgut homeostasis of microbial communities was examined in *Drosophila* through the alteration of AMP expression (2008). Overexpression of specific (cecropin or diptericin) or all AMPs controlled by an immune pathway was achieved through transgenic modification(s) or depletion of pathway inhibitor Caudal (Cad) respectively, which in turn altered microbial composition. While it is unclear how total microbiota load was affected in this study, the five dominant species remained the same but there was a clear trade-off between two species of microbiota. In control flies, a novel Acetobacteraceae strain A911 dominated the midgut microbiota and a novel *Gluconobacter* sp. strain G707 was a minor member whereas the opposite was true in AMP overexpressing flies. This shift towards the domination of *Gluconobacter* in the modified flies contributed to apoptosis of midgut cells and host mortality. In germ-free flies, introduction of *Gluconobacter* also led to a severe detriment on the host, but did not impact conventionally-reared control flies. However, when germ-free flies were first colonized with A911, G707 was unable to establish dominance in the midguts. Reintroduction of *Cad* in the *Cad*-depleted flies restored host health and a wildtype (WT) commensal structure, and it was also found that overexpression of cecropin alone was enough to selectively deplete A911 presence in the microbiota. These data highlight the complexities of midgut homeostasis, where host factors and microbes themselves can exert selection pressures on specific species. Thus, maintaining midgut homeostasis can have major implications on host health.

In mosquitoes, a type I peritrophic matrix (PM) is secreted by midgut epithelial cells in response to midgut distension (*i.e.* following a blood meal), physically separating the blood bolus from epithelial cells within the midgut lumen. This semipermeable PM is made up of chitin, and its full formation is microbiota-dependent (Rodgers et al, 2017). Within this acellular sac, blood meal digestion occurs and the microbiota proliferate. Blood meal digestion typically takes place over two to three days, after which the PM is excreted from the mosquito, eliminating remnants of the blood meal and the majority of the (proliferated) microbiota. Some microbiota still remain in the ectoperitrophic space

which seed the microbiota for the next gonotrophic cycle, where a new PM will form upon blood feeding (Rodgers et al, 2017).

The function of the PM is therefore two-fold: 1) to compartmentalize the blood meal and control digestive enzyme flow for efficient nutrient acquisition (Hegedus et al, 2009); and 2) to protect against microbial dissemination beyond the midgut and prevent the continuous activation of epithelial immune responses (Kumar et al, 2010; Rodgers et al, 2017). In the ectoperitrophic space between the PM and midgut epithelia, cross-linkage of secreted mucins relies on a peroxidase/dual oxidase system made up of immunomodulatory peroxidase (IMPer, also known as heme peroxidase [HPX] 15) and dual oxidase (DUOX) to create a dityrosine network, further protecting the epithelia (Kumar et al, 2010). Silencing of either IMPer or DUOX increases the interactions between bacteria and pathogen recognition receptors on the surface of midgut epithelial cells, which in turn elicits immune responses, indicating that the mucin layer promotes tolerance of the microbiota (Kumar et al, 2010). Incidentally, depletion of IMPer and DUOX also increases the interactions between *Plasmodium* and mosquito immune responses which reduces their survival (Kumar et al, 2010). This interaction involves immune responses beyond those that control the microbiota; parasites induce nitric oxide synthase (NOS) expression, which produce the antiparasitic molecule nitric oxide (NO) (Luckhart et al, 1998; Lim et al, 2005; Akman-Anderson, Olivier & Luckhart, 2007; Kumar et al, 2010). While the PM protects the midgut epithelia and prevents immune activation from the microbiota, it also benefits *Plasmodium* parasites as they are able to develop undetected. Sexual stages of *Plasmodium* development occurs within the PM, after which ookinetes must first cross the PM (and dityrosine network) to be able to traverse the midgut epithelia. *Plasmodium* spp. overcome this barrier by secreting high titers of at least one chitinase, which are activated by mosquito-derived proteases in the midgut (Shahabuddin et al, 1993; Tsai et al, 2001). Additionally, parasites have co-opted fibrinogen-related protein (FREP) 1, a pathogen recognition receptor (PRR) secreted by the mosquito midgut and integrated into the PM, to aid PM traversal (Li et al, 2013; Zhang et al, 2015). Gametocytes and ookinetes are able to bind to FREP1, anchoring themselves to the PM which facilitates penetration (Zhang et al, 2015).

1.3 Insect innate immunity

In nature, insects are continuously exposed to a variety of pathogens in which they must defend against. Insects lack an adaptive immunity, meaning all components of immunity are encoded by the germline (Janeway & Medzhitov, 2002). During an infection, host outcomes are determined by two main immunological responses: resistance and tolerance. Resistance against an infection involves the eradication or reduction of the number of invading microbes, achieved through the launch of effectors to directly attack the pathogen. Tolerance entails the minimization of impact caused by infection, whether it is direct damage from the pathogen itself or indirect self-harm caused by immune responses launched against invaders. Alternatively, tolerance is also defined as the ability to withstand an infection without activating immune responses, which is a more relevant definition in the context of immunity.

In mosquitoes, the hemocoel and midgut are the two main compartments of the body which are susceptible to infection. The hemocoel, which houses the fat body, is an immune organ that is typically sterile. Here, infection manifests as a result of cuticular damage or pathogens breaching the midgut epithelium. Circulating and sessile hemocytes within the hemocoel perform many immune functions, including immunosurveillance and phagocytosis of microbes. In the midgut, infection can occur as a result of microbial ingestion and/or excessive microbiota proliferation, and midgut epithelial cells initiate immune responses. In the absence of circulating immune cells within the midgut lumen, immune responses typically consist of three main events: recognition of non-self, signal transduction, and the launch of a defensive response (Zhang et al, 2021).

Groups of microbes possess a variety of conserved molecules, known as pathogen-associated molecular patterns (PAMPs), which are not present in the host. For example, peptidoglycan (PGN) is found in bacteria of both Gram types, albeit of slightly different composition, and β-glucans are found in fungi, both of which are essential for the microbe and absent in eukaryotes (Medzhitov & Janeway, 1997). These PAMPs are recognized as non-self by PRRs such as short (S) and long (L) PGN recognition proteins (PGRPs), which activate a signaling cascade upon engagement. In insects, the main intracellular signal transduction pathways are the Toll, immune deficiency (Imd), and the

Janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathways, which have been characterized in detail in model organism *D. melanogaster*. Current understanding of pathway components in *An. gambiae* typically stems from the comparison of homology between the two organisms, as they are highly conserved (Christophides et al, 2002).

1.3.1 The Toll pathway

In *Drosophila*, activation of the Toll pathway begins with the recognition of PAMPs from Gram-positive bacteria and fungi by extracellular, circulating PRRs (Fig. 1.3.1). PRRs Gram-negative binding proteins (GNBPs) 1 and 3, which are poorly named, recognize lysine-type PGN of Gram-positive bacteria and β-glucan from fungal cell walls respectively, and PGRP-SA also recognizes Lys-type PGN (Gobert et al, 2003; Gottar et al, 2006). PGRP-SD has been traditionally attributed to the Toll pathway, but recent evidence has shown it recognizes Gram-negative bacteria (Leone et al, 2008) and mediates activation of the Imd pathway (Iatsenko et al, 2016).

While different Toll pathway PRRs are involved in the detection of different PAMPs or even microbes, the recognition signals merge and trigger activation of the same downstream signaling pathway (reviewed in Lindsay & Wasserman, 2014). This signaling involves a series of serine protease cascades, where zymogens undergo proteolytic cleavage to become their active forms: modular serine protease (ModSP) integrates signals from recognition molecules and activates Grass (Chamy et al, 2008), which in turn activates serine protease immune response integrator (Spirit), Sphinx 1 and 2, or Spheroide, which then activate Spätzle processing enzyme (SPE). SPE proteolytically cleaves and activates Toll receptor ligand Spätzle, which binds to Toll and causes a conformational change to allow dimerization and activation of the Toll receptor (Weber et al, 2003; Arnot, Gay & Gangloff, 2010; Lindsay & Wasserman, 2014). The dimerized Toll receptor has two intracellular Toll/IL-1R (TIR) domains, each of which bind to an adaptor protein MyD88 upon activation (Horng & Medzhitov, 2001; Tauszig-Delamasure et al, 2002). This leads to the recruitment of adaptor protein Tube, which recruits kinase Pelle to form a MyD88-Tube-Pelle heterotrimeric complex (Sun et al, 2002). Formation of this complex is mediated by two death domains on Tube, which form separate pairwise

interactions with the death domains of MyD88 and Pelle (Sun et al, 2002). This complex initiates the autophosphorylation of Pelle, which induces the phosphorylation and degradation of IκB factor Cactus (Cact), the inhibitor of NF-κB transcription factors Dorsal and Dorsal-related immunity factor (Dif) (Wu & Anderson, 1998; Towb, Bergmann & Wasserman, 2001). Dissociation of Cact from the transcription factors enables their nuclear localization and the subsequent transcription of several immune effectors such as the AMP metchnikowin (Hanson & Lemaitre, 2020). Strikingly, the deployment of effectors appear to be adapted to the type of pathogen detected (Lemaitre, Reichhart & Hoffman, 1997).

In addition to the recognition of PAMPs, danger signals such as virulence factors also trigger the Toll pathway. For example, proteases secreted by fungi and Gram-positive bacteria to facilitate penetration of the epithelia or insect cuticle also cleave host zymogen Persephone, activating this host serine protease (Ligoxygakis et al, 2002; Gottar et al, 2006; Chamy et al, 2008). Activated Persephone subsequently proteolytically cleaves and activates SPE and the signal is transduced through the Toll pathway as described above.

The Toll pathway is a key regulator of the dorsal–ventral patterning of *Drosophila* embryos through the regulation of Dorsal (Anderson & Nüsslein-Volhard, 1984), and appears to have been secondarily co-opted by the immune system. Through a cascade of different serine proteases than those involved in immunity, Spätzle is processed as a ligand for Toll, which leads to the nuclear translocation of Dorsal (Anderson & Nüsslein-Volhard, 1984). Toll signaling exists as a gradient over the dorsal-ventral axis of syncytial embryos, which creates a concentration gradient of Dorsal across the embryo (Ip et al, 1991; Stathopoulos et al, 2002). As a result, expression of different developmental genes is elicited in different regions causing dorsal-ventral patterning (Stathopoulos et al, 2002).

In *Anopheles*, the TOLL* pathway is elicited mainly by fungi, but the PRRs responsible are still unknown (Rhodes, Thomas & Michel, 2018). Previously however, all six *Anopheles* GNBPs have been shown to contain a conserved β-1,3-glucan-binding domain, but functional characterization experiments elucidated a role in antibacterial defense (Warr et al, 2008). Most of the intracellular components of the *Anopheles* TOLL pathway have a 1:1 ortholog with *Drosophila*, suggesting a conserved role of Toll pathway in immunity and embryo development (Christophides et al, 2002). The exception is Dif, where no ortholog is found in *Anopheles* (Christophides et al, 2002). The ortholog of Dorsal in *Anopheles* is known as Relish 1 (REL1) or Gambif (Barillas-Mury et al, 1996), and has been shown to regulate AMPs, thioester-containing protein 1 (TEP1) and leucine-rich repeat immune protein (LRIM) 1 (Frolet et al, 2006; Luna et al, 2006).

Evidence for a role of the TOLL pathway in bacterial infection is limited (Meister et al, 2005), although a recent study has also shown its importance in defense against Gram-negative bacteria (Bai et al, 2019). This suggests the potential of bacterial PRRs, especially GNBPs, being involved in both pathways, although further experimentation is required to facilitate their pathway assignment (Warr et al, 2008). Preliminary experiments have shown that some are upregulated following bacterial and *Plasmodium* infection (Christophides et al, 2002).

1.3.2 The Imd pathway

Activation of the Imd pathway in *Drosophila* begins upon recognition of Gram-negative bacteria (Fig. 1.3.1). This activation is mediated by the binding of diaminopimelic acid (DAP)-type PGN of Gram-negative bacteria to PRRs PGRP-LC or -LE (Kaneko et al, 2006). PGRP-LE binds monomeric DAP-type PGN (also known as tracheal cytotoxin, TCT) and is found intracellularly (Kaneko et al, 2006). PGRP-LC is a transmembrane PRR and recognizes extracellular TCT and polymeric DAP-type PGN (Choe, Lee &

* In order to differentiate between (homologous) genes/proteins in mosquitoes and other organisms with the same name, the gene nomenclature outlined in Christophides et al (2002) and Waterhouse et al (2007) were utilized. All *An. gambiae* and *Ae. aegypti* gene and protein names are fully capitalized, with gene/RNA names italicized. In order to differentiate between *Anopheles* and *Aedes* gene names, the first letter of the genus and species is added in front of the *Aedes* gene name, *e.g.* '*Aa*' for *Ae. aegypti*. Genes/proteins of other organisms referred to and discussed in this thesis *(e.g.* those from *Drosophila*) follow standard gene nomenclature, *i.e.* gene/RNA names are lowercase and italicized, protein names are capitalized and not italicized.

Anderson, 2005; Kaneko et al, 2006). Additionally, a truncated form of PGRP-LE enhances PGRP-LC-mediated recognition extracellularly (Takehana et al, 2002). PGRP-LF, another transmembrane PRR, is a negative regulator of the Imd pathway and prevents constitutive activation of immunity, likely through the sequestration of PGN (Maillet et al, 2008).

Upon PGN binding, these PGRPs dimerize or multimerize (Mellroth et al, 2005; Lim et al, 2006), and their cytoplasmic domain initiates signal transduction and the intracellular recruitment of adaptor protein Imd (Choe et al, 2002). Imd in turn recruits *Drosophila* Fas-associated death domain-containing protein (dFADD) (Leulier et al, 2002; Naitza et al, 2002), which associates with death-related ced-3/Nedd2-like caspase (Dredd) (Leulier et al, 2000). Imd is further activated after cleavage by Dredd, exposing an interaction site for ubiquitination by inhibitor of apoptosis (IAP) 2 (Kleino & Silverman, 2014). This ubiquitination recruits and activates transforming growth factor β-activated kinase 1 (TAK1) and its regulatory adapter protein TAK1-associated binding protein 2 (TAB2) (Silverman et al, 2003). Here, the pathway branches into two, as TAK1 is also involved in the regulation of the c-Jun N-terminal kinase (JNK) signaling pathway (Silverman et al, 2003; Park et al, 2004). The TAK1/TAB2 complex activates the IκB kinase (IKK) complex, which is made up of a catalytic subunit immune response deficient 5 (IRD5) and a regulatory subunit Kenny (Rutschmann et al, 2000; Silverman et al, 2003). The IKK complex phosphorylates transcription factor Relish, which then becomes a target for cleavage by Dredd (Kleino & Silverman, 2014). It is also thought that IAP2 has a role in the activation of Dredd (Meinander et al, 2012). Dredd cleaves the self-inhibiting inhibitor of κB (IκB) ankyrin domain of Relish, exposing its nuclear localization signal and allowing its nuclear translocation, which activates transcription of effectors such as AMPs (Leulier et al, 2000; Stöven et al, 2003; Kleino & Silverman, 2014). At this stage, the Imd pathway can be suppressed by Caspar, which inhibits the Dredd-dependent cleavage of Relish (Kim et al, 2006). Two other negative regulators of the Imd pathway are poor Imd response upon knock-in (Pirk) and developmental transcription factor Cad, which inhibit the PGRP-LC/Imd interaction (Kleino et al, 2008) and repress midgut AMP transcription respectively (Ryu et al, 2008; Clayton et al, 2013).

For the majority of the Imd pathway components, 1:1 orthologs between *Drosophila* and *Anopheles* have been identified (Christophides et al, 2002). Both PGRP-LC and PGRP-LA are transmembrane surface receptors that activate the Imd pathway, where PGRP-LA is thought to contribute to the constitutive expression of AMPs (Gendrin et al, 2017; Rodgers et al, 2020). In contrast, *Drosophila* PGRP-LA plays a regulatory role (Gendrin et al, 2013). In both species, PGRP-LB is secreted and scavenges PGN to prevent Imd activation (Zaidman-Rémy et al, 2006; Gendrin et al, 2017; Rodgers et al, 2020). No orthologs for TAB2, PGRP-LE, and Pirk have been identified (Christophides et al, 2002; Kriventseva et al, 2007). Additionally, IRD5, Kenny, and Dredd are known as IKK1, IKK2, and long caspase 1 (CASPL1) in *Anopheles* respectively. Relish 2 (REL2), the homolog of Relish, exists as two alternatively spliced isoforms in *Anopheles*. The full-length form (REL2-F) contains a self-limiting IκB ankyrin domain which is cleaved following immune stimulation as described above, while the short form (REL2-S) lacks this domain and is thought to contribute constitutive expression of immune genes alongside PGRP-LA (Meister et al, 2005; Luna et al, 2006; Gendrin et al, 2017).

The *Anopheles* IMD pathway is heavily implicated in controlling bacteria, where it is activated by and defends against bacteria of both Gram types (Meister et al, 2005; Meister et al, 2009; Dong et al, 2011; Hixson et al, 2023). PGRP-LC is capable of binding both DAP- and Lys-type PGN (Meister et al, 2009; Hixson et al, 2023). The potency of this pathway against bacteria also implicates it in the resistance and tolerance of the microbiota. PGRP-LC is mainly responsible for resisting against the microbiota proliferation post-sugar and blood meal (Meister et al, 2009), whereas PGRP-LB is upregulated post-blood meal (Vlachou et al, 2005) and is thought to sequester PGN to dampen immune responses and allow for a certain degree of microbiota to persist (Gendrin et al, 2017). Target genes of REL2 include CEC1, gambicin 1 (GAM1), and LRIM1, some of which overlap with the downstream targets of the TOLL pathway (Meister et al, 2005).

Both the TOLL and IMD pathways have been implicated in antiplasmodial defense, but it remains unclear whether any *Plasmodium*-specific recognition events take place (Dong et al, 2006; Riehle et al, 2008; Garver, Dong & Dimopoulos, 2009; Mitri et al, 2009; Meister et al, 2009; Dong et al, 2011; Dong et al, 2012; Garver et al, 2012; Clayton et al,

2013; Ramphul et al, 2015). Silencing of various immune effectors such as *TEP1* led to increased incidences of *P. falciparum* and *P. berghei* infection (Dong et al, 2006; Garver et al, 2012). Depletion of Imd pathway factors *IMD*, *FADD*, *CASPL1*, and *REL2* individually also increased *P. falciparum* oocysts, and silencing of negative regulator *CASPAR* led to the absence of infection (Garver et al, 2012). The co-silencing of each factor with *CASPAR* reversed the phenotype observed following the knockdown (KD) of *CASPAR* alone (Garver et al, 2012). The KD of *REL2* and *IMD* in *An. gambiae* also significantly increased *P. berghei* infection intensities (Meister et al, 2005), and *CACT* and *CASPAR* silencing significantly reduced the number of *P. berghei* and *P. falciparum* oocysts compared to controls (Garver, Dong & Dimopoulos, 2009). While silencing of the negative regulators of each pathway reduced infection intensities of both *Plasmodium* species, *CASPAR* KD was more effective in reducing incidences of *P. falciparum* infection compared to *P. berghei*, and *CACT KD* showed the reverse phenotype (Garver, Dong & Dimopoulos, 2009). KD of *CACT* has also been reported to confer total refractoriness against *P. berghei* (Frolet et al, 2006), and silencing of *Anopheles Plasmodium*-responsive leucine-rich repeat 1 (APL1) paralog C, an effector under the control of REL1, increased *P. berghei* infection intensities (Riehle et al, 2008). Interestingly, APL1C also appears to confer resistance against fungi (Bukhari et al, 2022). Another member of the APL1 family, APL1A, is under the control of REL2-S and is potent against *P. falciparum* but not *P. berghei* (Mitri et al, 2009). Taken together, these studies suggest that the immune pathways are more potent against different species of *Plasmodium*, where the IMD is more adept in *P. falciparum* killing, and the TOLL pathway kills *P. berghei* more efficiently.

1.3.3 The JNK pathway

The *Drosophila* JNK pathway is activated in response to a broad range of stimuli, including stress (*e.g.* ROS-induced), wounding, and bacterial LPS (Kim et al, 2005; Tafesh-Edwards & Eleftherianos, 2020; Ogienko et al, 2022). Signaling through the pathway promotes a variety of responses, including cell proliferation, apoptosis, wound healing, and antimicrobial defense (Boutros, Agaisse & Perrimon, 2002; Buchon et al, 2009; Tafesh-Edwards & Eleftherianos, 2020). In response to stressors, part of the

pathway can be activated downstream of Imd, which is mediated by TAK1 and known as the non-canonical pathway (Silverman et al, 2003; Park et al, 2004; Ogienko et al, 2022). TAK1 activates a kinase cascade, where TAK1 phosphorylates JNK kinase Hemipterous (Hep), which in turn phosphorylates and activates JNK Basket (Bsk), which activates Forkhead box protein O (FOXO) (Wang, Bohmann & Jasper, 2005; Mishra et al, 2021; Ogienko et al, 2022), or phosphorylates Jun and Fos (Sluss et al, 1996; Kleino & Silverman, 2014). Together, Jun and Fos form transcription factor heterodimer activating protein 1 (AP-1), which initiates the production of targets including phosphatase Puckered (Puc), which dephosphorylates Bsk and acts as a negative regulator of the pathway, and Unpaired (Upd) cytokines (Yu et al, 2022). Another negative regulator is PGRP-LF, which is shared with the Imd pathway (Maillet et al, 2008). Upstream of TAK1, pathway components involve a series of kinase cascades and other ligand–receptor interactions that do not overlap with the Toll/Imd pathways (reviewed in Ogienko et al, 2022). Additionally, the JNK pathway can signal independently of TAK1 through other molecules (known as the canonical pathway), and is particularly relevant during embryonic development and dorsal closure (Ríos-Barrera & Riego-Escovar, 2013; Ogienko et al, 2022). Both the canonical and non-canonical pathways convene at Hep (Ríos-Barrera & Riego-Escovar, 2013).

Characterization of the JNK pathway in *Anopheles* has been very limited so far (Cirimotich et al, 2010). Orthologs for Hep, Bsk, Jun, Fos, and Puc have been identified in *Anopheles*, with two Bsk homologs identified and annotated as JNKa and JNKb respectively (Horton et al, 2011; Garver, de Almeida Oliveira & Barillas-Mury, 2013). JNKb shares more homology with Bsk (Horton et al, 2011). Nevertheless, the few studies that have investigated the *Anopheles* JNK pathway have shed light on its importance in anti-*Plasmodium* defenses (Jaramillo-Gutierrez et al, 2010; Garver, de Almeida Oliveira & Barillas-Mury, 2013; Ramphul et al, 2015). Silencing of JNK pathway positive regulatory components increases susceptibility of *An. gambiae* to *P. berghei* infection, and silencing of negative regulator *PUC* reduces infection intensities (Garver, de Almeida Oliveira & Barillas-Mury, 2013). The JNK pathway controls the expression of HPX2 and NADPH oxidase 5 (NOX5), two enzymes which are involved in the production of reactive nitrogen species which target and tag the parasite for downstream killing mechanisms

(de Almeida Oliveira, Lieberman & Barillas-Mury, 2012; Garver, de Almeida Oliveira & Barillas-Mury, 2013; Ramphul et al, 2015). Other effectors downstream of JNKa/b include TEP1 and FREP9 (Garver, de Almeida Oliveira & Barillas-Mury, 2013).

1.3.4 The JAK/STAT pathway

Initiation of the JAK/STAT pathway occurs upon binding of cytokine-like ligands Upd, Upd2, or Upd3 to the transmembrane receptor Domeless (Dome) (Fig. 1.3.1; Brown, Hu & Hombría, 2001). Expression of these ligands are induced as a response to tissue damage (Pastor-Pareja, Wu & Xu, 2008), midgut bacterial challenge (Buchon et al, 2009; Cronin et al, 2009), or viral infection (Dostert et al, 2005). Ligand binding induces dimerization of Dome and activation of the JAK Hopscotch (Hop), which are constitutively associated with the cytoplasmic tails of Dome (Myllymäki & Rämet, 2014). Activated Hop kinases phosphorylate each other and Dome to become a docking site for cytoplasmic STAT transcription factor Stat92E (Myllymäki & Rämet, 2014). The STAT molecules are also phosphorylated by Hop, inducing STAT dimerization and translocation into the nucleus, where they initiate transcription of their target genes (Kiu & Nicholson, 2012). Negative regulators of the pathway include suppressors of cytokine signaling (Socs) (Arbouzova & Zeidler, 2006), protein inhibitors of activated STAT (Pias) (Betz et al, 2001), and Ken & Barbie (Ken), which recognizes a deoxyribonucleic acid (DNA) consensus sequence overlapping that of Stat92E (Arbouzova, Bach & Zeidler, 2006). Like the Toll pathway, the JAK/STAT pathway also has roles in *Drosophila* development (Myllymäki & Rämet, 2014).

Two STAT transcription factors are encoded by the *An. gambiae* genome (Christophides et al, 2002; Waterhouse et al, 2007), where STAT-B regulates the transcription of *STAT-A* (Gupta et al, 2009). In *Aedes*, this pathway exerts antiviral activity against Dengue (Souza-Neto, Sim & Dimopoulos, 2009) and Zika viruses (Angleró-Rodríguez et al, 2017). In *Anopheles*, the JAK/STAT pathway is also implicated in antiviral responses, where it has been shown to both restrict (Carissimo et al, 2015) and promote viral infection (Nanfack-Minkeu et al, 2019). During bacterial infections, evidence has suggested that while the pathway is responsive, it is not essential for survival (Barillas-Mury et al, 1999; Gupta et al, 2009). Following bacterial infection, the

DNA binding activity of STAT-A is increased and its target *NOS* is upregulated, however *STAT-A* expression itself is not upregulated (Barillas-Mury et al, 1999; Gupta et al, 2009; Hillyer & Estévez-Lao, 2010). Silencing of either *STAT* does not change survival outcomes of *An. gambiae* following oral or systemic bacterial infections (Gupta et al, 2009). On the contrary, inhibition of NOS leads to uncontrolled proliferation of pathogenic bacteria and increased mortality (Hillyer & Estévez-Lao, 2010), suggesting other pathways also regulate *NOS* expression.

Increasing evidence suggests major implications of JAK/STAT signaling against *Plasmodium* infection. In the *An. gambiae/P. berghei* infection model, silencing of either *STAT-A* or *STAT-B* increases oocyst infection intensity at 7–8 days post-infection (Gupta et al, 2009). Upon detailed examination of this phenotype, oocyst killing is mediated by NOS produced by the midgut epithelia, and is regulated by STAT-A. Interestingly, Gupta and colleagues observed that NOS enables survival of early stage (2–3 days) oocysts, elucidating contrasting roles of this pathway in affecting *Plasmodium* development (2009). *An. stephensi* infected with *P. berghei* also showed increased levels of *NOS* and NO, and inhibition of NOS led to increased late-stage oocyst infection intensities (Luckhart et al, 1998). Here, the authors postulated that NOS acts on life stages before oocyst formation (Luckhart et al, 1998). In preliminary *P. falciparum* infection experiments, STAT-A silencing reduced late-stage oocyst intensities, but whether early-stage oocysts are affected similarly is still unknown (Gupta et al, 2009). In the *An. aquasalis/P. vivax* model, a South American malaria vector/parasite pair, *STAT* expression is induced upon *P. vivax* infection and is necessary to confer protection against the parasite (Bahia et al, 2011).

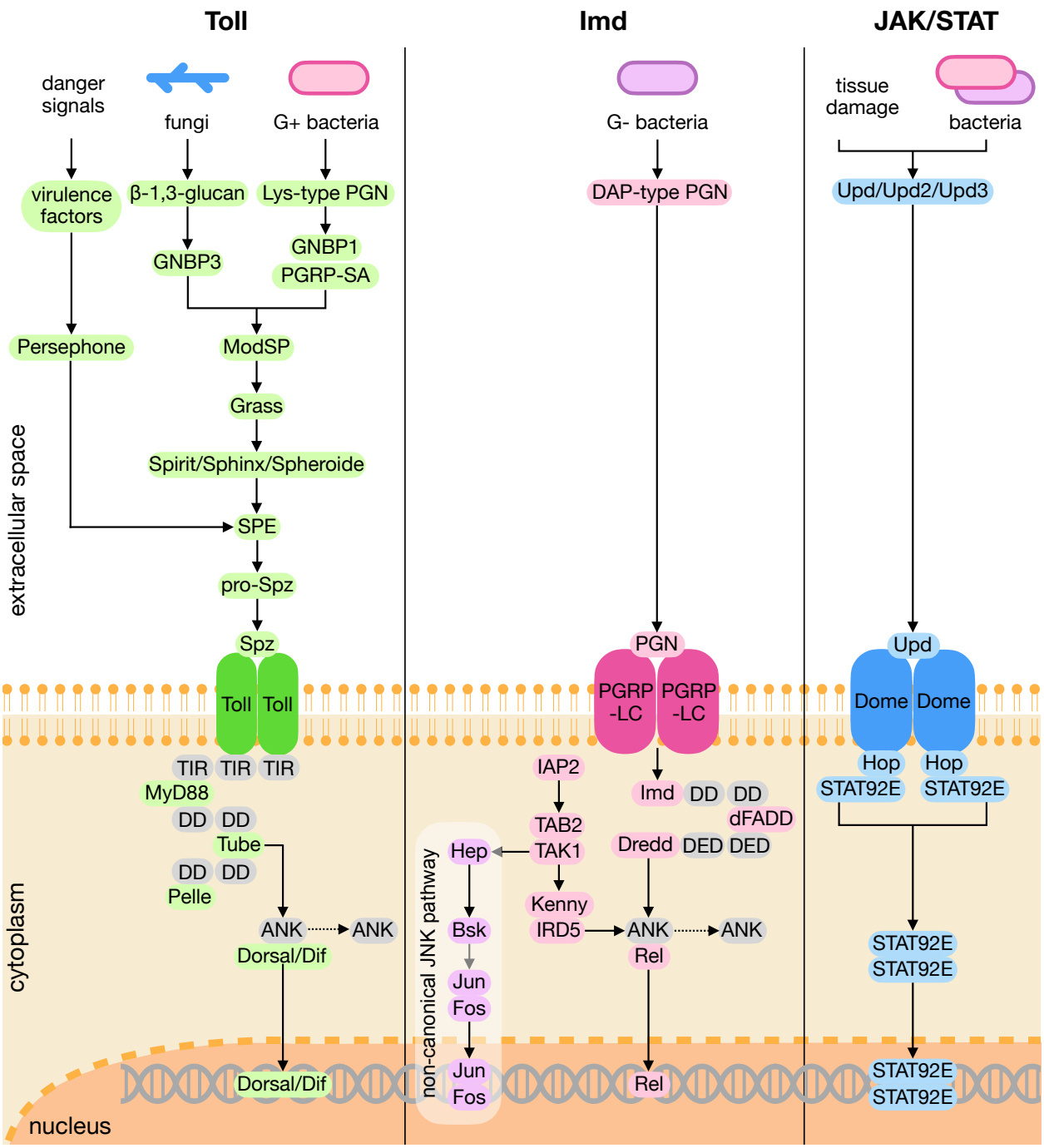


***Figure 1.3.1. The three traditional immune pathways: the Toll, Imd (including the non-canonical branch of the JNK pathway), and JAK/STAT in* Drosophila.** *The recognition of pathogens/danger signals initiates the activation of the pathways, ultimately leading to the translocation of transcription factors into the nucleus, inducing expression of target effector genes.*

1.3.5 The RNAi pathway

The ribonucleic acid interference (RNAi) pathway was initially identified in a nematode but has been found to be highly conserved in all eukaryotic organisms (Agrawal et al, 2003). While it is believed to have evolved as an antiviral defense mechanism (Cullen, 2014), it has been co-opted as a powerful tool to investigate gene functions through the silencing of genes. Briefly, one of the RNA-based silencing pathways involves the recognition and processing of perfectly base-paired double-stranded RNAs (dsRNA), such as that from a virus, into smaller (~22 nucleotides long) short interfering RNA (siRNA) duplexes by Dicer 2, a ribonuclease (RNAse) III enzyme (Hammond, 2005). After dicing, these duplexes are loaded into the RNA-induced silencing complex (RISC), which contains a slicing enzyme Argonaute 2 (AGO2) (Mongelli & Saleh, 2016). One strand of the RNA duplex is degraded, whereas the other remains and is methylated, resulting in a mature and active RISC (Ding & Voinnet, 2007). The remaining strand is known as the guide RNA strand and mediates sequence-specific recognition of target RNA cleavage through the slicing activity of AGO2 (Mongelli & Saleh, 2016). Thus, translation of a full protein is inhibited. The RNAi pathway is summarized in Figure 1.4.2.

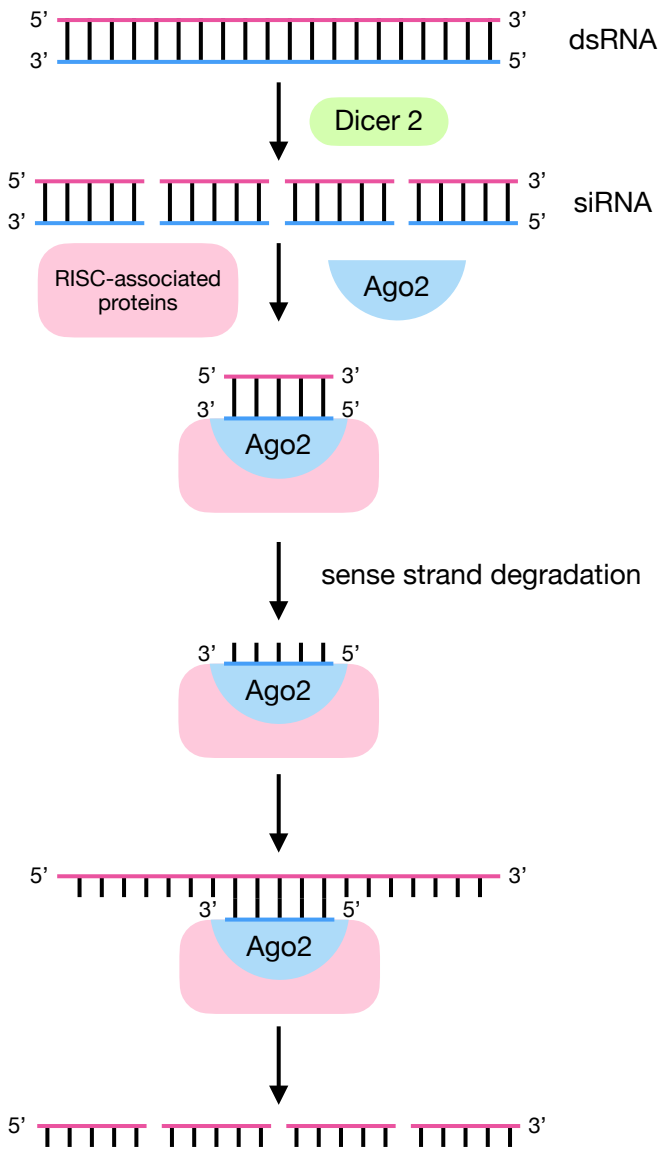


Figure 1.3.2. The RNA interference pathway.

Through the injection of dsRNA, individual or a selection of genes can be targeted and silenced in a highly specific manner. The majority of KD experiments in *Anopheles* have utilized the RNAi pathway, including all of the silencing experiments outlined in the previous sections. Furthermore, this pathway is important in the sensing and control of viral infections in mosquitoes (Keene et al, 2004; Carissimo et al, 2015), and was described to be the only effective pathway against Chikungunya infections in *Ae. aegypti* (McFarlane et al, 2014).

1.3.6 Other defense mechanisms and immune molecules

Circulating and sessile hemocytes in the hemolymph contribute to a variety of immune defense mechanisms. Three populations of hemocytes have been characterized in mosquitoes: granulocytes, oenocytoids, and prohemocytes, each of which contribute to different immune functions (Hillyer & Strand, 2014). In adult mosquitoes, approximately 25% of hemocytes are sessile and are attached to various membranes including the midgut, legs, and Malpighian tubules (Hillyer & Strand, 2014). The remaining majority of hemocytes are circulating, of which the most abundant cell population are granulocytes (Castillo, Robertson & Strand, 2006; King & Hillyer, 2013). Granulocytes are highly phagocytic and rapidly internalize pathogens for digestion (Hillyer & Strand, 2014). This is achieved through surface PRRs such as PGRP-LC (Moita et al, 2005) and facilitated by humoral factors such as TEP1 (Levashina et al, 2001), which directly recognize or opsonize pathogens respectively (Hillyer & Strand, 2014). Phagocytosis only occurs in the hemolymph and is particularly effective against bacteria and foreign particles (King & Hillyer, 2013), and also *Plasmodium* sporozoites to a lesser extent (Hillyer, Schmidt & Christensen, 2003; Hillyer, Barreau & Vernick, 2007). Prohemocytes also display some phagocytic activity (King & Hillyer, 2013; Hillyer & Strand, 2014), and oenocytoids generate large quantities of phenoloxidase (PO), the rate-limiting enzyme in the melanization pathway (Christensen et al, 2005; Castillo, Robertson & Strand, 2006).

In the melanization pathway, POs are activated by a cascade of Clip-domain serine proteases (CLIPs) upon pathogen recognition and catalyze the conversion of tyrosine to melanin (Christensen et al, 2005; Volz et al, 2006). Melanin is deposited on the surface of pathogens leading to their death or targeted phagocytosis by hemocytes (Hillyer & Strand, 2014). Targets of melanization include bacteria, filarial worms, *Plasmodium* parasites, and fungi (Collins et al, 1986; Christensen et al, 2005; Baton et al, 2009; Yassine, Kamareddine & Osta, 2012). Moreover, toxic intermediates of the PO cascade have been shown to have antiviral effects (Zhao et al, 2011; Rodriguez-Andres et al, 2012). In mosquito and malaria parasite interactions, melanization typically occurs between allopatric species combinations (Collins et al, 1986; Schwartz & Koella, 2002; Michel et al, 2006), but recent research has suggested melanization occurs in sympatric species if the infection intensity is high (Simões et al, 2017a). Melanization is also

involved in other aspects of mosquito physiology, including wound healing and cuticle hardening (Christensen et al, 2005).

Hemocytes also produce TEP1, a soluble protein that is involved in the complement-like response and is analogous to C3 found in vertebrates (Levashina et al, 2001). In the hemolymph, two forms of TEP1 circulate: a full-length (TEP1-F) and a cleaved form $TEP1_{cut}$ with a reactive thioester motif (Levashina et al, 2001; Baxter et al, 2010). $TEP1_{cut}$ forms a complex with leucine-rich repeat-containing proteins LRIM1 and APL1C, which localizes to bacterial or ookinete surfaces on the midgut basal lamina (Moita et al, 2005; Fraiture et al, 2009; Baxter et al, 2010; Williams et al, 2019). Attachment to a pathogenic surface enables recruitment of TEP1-activated proteases and $TEP1_{cut}$, which act as a convertase to cleave other TEP1-F molecules (Povelones et al, 2013). This accumulation of TEP1 leads to opsonization, targeting the invader for phagocytosis and/or melanization, or direct killing mechanisms via lysis, asphyxiation, and/or attack by ROS (Levashina et al, 2001; Blandin et al, 2004; Moita et al, 2005; Nappi & Christensen, 2005; Fraiture et al, 2009; Povelones et al, 2009). TEP1 is also regulated by the TOLL and IMD pathways (Frolet et al, 2006), and has major implications in *P. falciparum* killing (Dong et al, 2006; Garver, Dong & Dimopoulos, 2009).

Nitration responses also aid targeted killing of *Plasmodium* and activate the complement-like system. As ookinetes transit through the midgut epithelium, nitration responses are induced in the invaded epithelial cells through the upregulation of NOS, HPX2, and NOX5 (Luckhart et al, 1998; Han et al, 2000; Kumar et al, 2004; Gupta et al, 2009; de Almeida Oliveira, Lieberman & Barillas-Mury, 2012; Garver, de Almeida Oliveira & Barillas-Mury, 2013). NOS is also produced by granulocytes post-bacterial infection and is regulated by the JAK/STAT pathway (Gupta et al, 2009; Hillyer & Estévez-Lao, 2010), and the latter two enzymes are regulated by the JNK pathway (Garver, de Almeida Oliveira & Barillas-Mury, 2013). These enzymes aid in the generation of NO, a free radical which negatively impacts *Plasmodium* (Peterson, Gow & Luckhart, 2007; de Almeida Oliveira, Lieberman & Barillas-Mury, 2012) and bacteria (Hillyer & Estévez-Lao, 2010). It has been proposed that these nitration responses modify ookinetes during their invasion of the midgut epithelia, which facilitates TEP1 attachment, opsonization, and killing (de Almeida Oliveira, Lieberman & Barillas-Mury, 2012). Thus, the rapid traversal of the

midgut epithelium favors parasite development in the mosquito (Kumar et al, 2004), as the parasite encounters fewer nitration responses and other oxidative reactions, and is hence less likely to be detected by the complement-like system upon emergence on the basal side. Indeed, a *Plasmodium*-refractory strain of *Anopheles* known as L3-5 (Collins et al, 1986) has been characterized to have increased rates of H_2O_2 generation (Kumar et al, 2003; Molina-Cruz et al, 2008; Oliveira et al, 2011), and reduced detoxification rates have been shown to promote ookinete lysis (Jaramillo-Gutierrez et al, 2010). Congruently, refractoriness of this strain is conferred through the melanization and killing of all *Plasmodium* ookinetes (Molina-Cruz et al, 2012). Increased ROS production has also been linked to improved resistance against bacteria (Molina-Cruz et al, 2008). Furthermore, JNK pathway components are more abundantly expressed in this refractory strain (Garver, de Almeida Oliveira & Barillas-Mury, 2013).

Despite its refractoriness to many *Plasmodium* species, the L3-5 strain is susceptible to certain *P. falciparum* strains, such as NF54 and 3D7 (Collins et al, 1986; Molina-Cruz et al, 2012). In *P. falciparum*, this is mediated by the protein P47 (also known as Pfs47), which binds *An. gambiae* P47 receptor (P47Rec) and suppresses nitration responses and TEP1-mediated lysis (Molina-Cruz et al, 2013; Ukegbu et al, 2017; Molina-Cruz et al, 2020; Alves e Silva et al, 2023). Thus, these parasites are able to evade the immune system and continue development into oocysts (Molina-Cruz et al, 2013; Alves e Silva et al, 2023). Interestingly, *Penicillium chrysogenum*, a fungus commonly part of *An. gambiae* microbiota, was found to increase mosquito susceptibility to *Plasmodium* through the manipulation of nitration responses (Angleró-Rodríguez et al, 2016). *P. chrysogenum* upregulates host ornithine decarboxylase, which sequesters arginine, a substrate for NOS in NO production, which in turn favors parasite survival (Angleró-Rodríguez et al, 2016). At least 42 *P47* alleles have been identified across the global *P. falciparum* populations and show strong geographical structure, which is compatible with sympatric mosquito-parasite co-adaptations and evasion of mosquito immune responses (Molina-Cruz et al, 2015; Molina-Cruz, Canepa & Barillas-Mury, 2017). In some cases, a single amino acid replacement mutation is sufficient to alter a parasite strain's ability to evade immune responses in a different mosquito strain/species (Molina-Cruz et al, 2015; Canepa, Molina-Cruz & Barillas-Mury, 2016). Another parasite protein, *Plasmodium*

infection of the mosquito midgut screen 43 (PIMMS43), possesses a similar function, where it enables evasion of the mosquito immune response and shows strong variation in genetic structure that is adapted to sympatric vector populations (Ukegbu et al, 2020). These studies suggest a potential for Pfs47 and PIMMS43 to be used as a target in blocking malaria transmission.

To date, mechanisms of *Plasmodium* recognition by the mosquito have been attributed to a few molecules and mostly through indirect means. A previous study has implicated the ingestion of human/mammalian cytokine transforming growth factor β1 during an infectious blood meal in inducing mosquito nitration responses (Luckhart et al, 2003). Subsequent studies have also found that parasite-derived molecules, namely glycosylphosphatidylinositols and hemozoin, and merozoites induce NOS expression through a variety of kinases that are also associated with the insulin pathway (Lim et al, 2005; Akman-Anderson, Olivier & Luckhart, 2007). Hemozoin was also found to activate TAK1, a kinase involved in the activation of the IMD pathway (Akman-Anderson, Olivier & Luckhart, 2007). This suggests that hemozoin is also eliciting the IMD pathway, the most potent pathway against *P. falciparum* (Garver, Dong & Dimopoulos, 2009; Mitri et al, 2009). Taken together, parasite-derived molecules are pro-inflammatory in the mosquito, which leads to the upregulation of responses that directly impact *Plasmodium* development.

Another group of immune molecules are the FREPs (also known as FBNs), which have fibrinogen-like domains. This PRR family consists of 59 members in *Anopheles* while only 14 members have been identified in *D. melanogaster*, indicating a massive expansion and diversification, perhaps reflective of strong selective pressures imposed by the variety of habitats mosquitoes occupy throughout their lifetime (Christophides et al, 2002; Zdobnov et al, 2002; Wang, Zhao & Christensen, 2005; Dong & Dimopoulos, 2009). In other invertebrates, different FREPs have been shown to recognize specific categories of pathogens (Gokudan et al, 1999; Zhang, Zeng & Loker, 2008), such as aslectin of *Armigeres subalbatus* mosquitoes which bind and recognize bacteria (Wang et al, 2004). So far, a handful of the FBNs have been functionally characterized in *Anopheles*.

In a study by Dong & Dimopoulos, the majority of the FREPs were found to be expressed in hemocytes and not the midgut, and sex-specific members of the family were identified (2009). Expression of FREPs 3, 6, 8, 9, 13, 22, 37, and 39 were induced by *P. falciparum* and/or *P. berghei* ookinetes (Dong et al, 2006; Dong & Dimopoulos, 2009). In response to bacterial infection, around 10 FREPs were induced (Christophides et al, 2002; Dong et al, 2006; Dong & Dimopoulos, 2009). Previously, it was observed that the KD of FREPs 8, 9, and 39 increased oocyst infection intensities of both *Plasmodium* species and reduced survival outcomes post-bacterial infection (Dong et al, 2006). These results show a clear implication of FREPs in mosquito immunity. Interestingly, in the study by Dong & Dimopoulos, a pool of FREPs with anti-*Plasmodium* and/or antibacterial activity were silenced in mosquitoes, which led to increased oocyst intensities post-*Plasmodium* infection more so than if any of the FREPs were silenced individually (2009). This suggests that FREPs may be functioning in synergy with each other. Moreover, it was observed that FREP9 interacts with the surface of bacterial cells and forms dimers (Dong & Dimopoulos, 2009). A strong co-localization in mosquito midguts was observed between ookinetes and FREP9, elucidating a potential direct interaction or association through other known factors (Dong & Dimopoulos, 2009). FREP1 was already previously shown to be utilized by the parasite to traverse the PM (Zhang et al, 2015). So far, knowledge of FREP regulation is limited to FREP9, which is under the control of the IMD and JNK pathways (Garver, Dong & Dimopoulos, 2009; Garver, de Almeida Oliveira & Barillas-Mury, 2013). *FREP9* overexpression was also investigated through the generation of a transgenic mosquito, where *FREP9* expression was driven by a blood feeding-induced fat body promoter (Simões et al, 2017b). This transgenic line was less susceptible to *P. berghei* and bacterial infection, but surprisingly not more resistant to *P. falciparum* than WT mosquitoes (Simões et al, 2017b). Despite having a strong effect in KD studies, the lack of *P. falciparum* resistance in this line suggests FREP9 may associate with other immune factors to kill parasites (Simões et al, 2017b). Further characterization of FREPs might enable their use as a tool in malaria intervention strategies.

Another immune effector of particular interest is Down syndrome cell adhesion molecule (DSCAM). This molecule contains numerous exons which are alternatively

spliced, theoretically enabling the generation of over 30,000 splice forms (Dong, Taylor & Dimopoulos, 2006). The alternative splicing of this molecule is proposed to be analogous to that of antibodies produced by vertebrates, although its specificity and affinity to different pathogens is not comparable to that of antibodies (Dong, Taylor & Dimopoulos, 2006). DSCAM associates with and acts as a receptor for pathogens, and is implicated in the phagocytosis/recognition of bacteria and *Plasmodium* (Dong, Taylor & Dimopoulos, 2006). Expression of different splice forms are induced following immune challenge of different bacterial and *Plasmodium* species, suggesting that different splice form repertoires are specialized in the recognition of specific pathogens (Dong, Taylor & Dimopoulos, 2006; Dong et al, 2012). These repertoires were also impacted by IMD and TOLL pathway activation, where each pathway led to changes in the abundance of different DSCAM splice forms through the activation of different splicing factors (Dong et al, 2012). In the context of *Plasmodium* infection, while depletion of all *DSCAM* splice forms increased *P. berghei* and *P. falciparum* oocyst loads, these effects were mediated by different exons/splice forms (Dong, Taylor & Dimopoulos, 2006; Dong et al, 2012). Activation of the REL2/IMD pathway led to the induction of a similar repertoire observed following *P. falciparum* infection, whereas TOLL pathway activation led to an abundance of splice forms associated with *P. berghei* infection (Dong et al, 2012). Indeed, the specific silencing of a *P. falciparum* infection-responsive but not a *P. berghei*-responsive exon led to increased *P. falciparum* oocyst loads comparable to the silencing of all *DSCAM* transcripts (Dong et al, 2012). The generation of transgenic mosquito lines with enhanced expression of different *DSCAM* exons associated with the two species of *Plasmodium* also confirmed this phenotype following *Plasmodium* infection (Dong et al, 2012). These results provide valuable insight into immune recognition of pathogens, in addition to supporting the notion that immune responses to different species of *Plasmodium* are distinct. Given that the IMD pathway is the primary pathway implicated in antibacterial responses, these results also indicate that antibacterial effectors impact *P. falciparum*.

1.3.7 Antimicrobial peptides

AMPs have been described in all domains of life and constitute an important part of immune responses as effectors. Their chemical structures, peptide conformations, and specificities show great diversity, however the majority are thought to have the following characteristics: an overall net positive charge, amphipathicity, short in length, secreted, and the propensity to form secondary structures (Epand & Vogel, 1999; Tossi, Sandri & Giangaspero, 2004; Ladram & Nicholas, 2016; Zhang & Gallo, 2016). These characteristics contribute to their affinity for microbial membranes, which are typically negatively-charged (Oren & Shai, 1998; Epand & Vogel, 1999; Zhang & Gallo, 2016). Many AMPs are rich in specific residues which affect secondary structures, *e.g.* small residues glycine, alanine, which are commonly associated with α-helices, and cysteine, which promotes the formation of β-sheets (Wang, Li & Wang, 2009). AMPs can be broadly categorized into three classes: 1) linear with amphipathic α-helices; 2) cysteine-rich (*i.e.* forming disulfide bonds), and 3) proline- or glycine-rich (Bulet et al, 1999; Oren & Shai, 1998). In insects, glycine, alanine, and lysine are the most common amino acids in AMPs (Wang, Li & Wang, 2009).

Rather than through interactions with surface receptors, AMPs are believed to exhibit antimicrobial activity through the interaction and disruption of microbial membranes through two main mechanisms: 1) pore formation and 2) carpet-rolling (Fig. 1.3.3; Oren & Shai, 1998; Huang, Chen & Lee, 2004; Tossi, Sandri & Giangaspero, 2004). At low peptide concentrations, AMPs adsorb to the hydrophilic head groups of the phospholipid bilayer and lie on the membrane surface, whereas at high concentrations, AMPs enter into an inserted state (Ludtke et al, 1996). In the pore formation mechanisms, barrel-stave pores can be formed through the insertion and reorientation of AMPs into the membrane (Christensen et al, 1998; Fig. 1.3.3). Alternatively, AMPs can also aggregate on the microbial surface, associating with the head groups of the phospholipid bilayer, leading to the thinning of the membrane and toroidal pore formation (Matsuzaki et al, 1996; Ludtke et al, 1996; Fig. 1.3.3). AMPs remain associated with the phospholipid head groups when forming toroidal pores. In the carpet-like model, the accumulation of AMPs causes disruption and lysis of the membrane, leading to the formation of micelles (Oren & Shai, 1998; Fig. 1.3.3). Following membrane disruption, some AMPs can further inhibit

other cellular processes such as DNA and protein synthesis (Tossi, Sandri & Giangaspero, 2004; Brogden, 2005; Ladram & Nicholas, 2016).

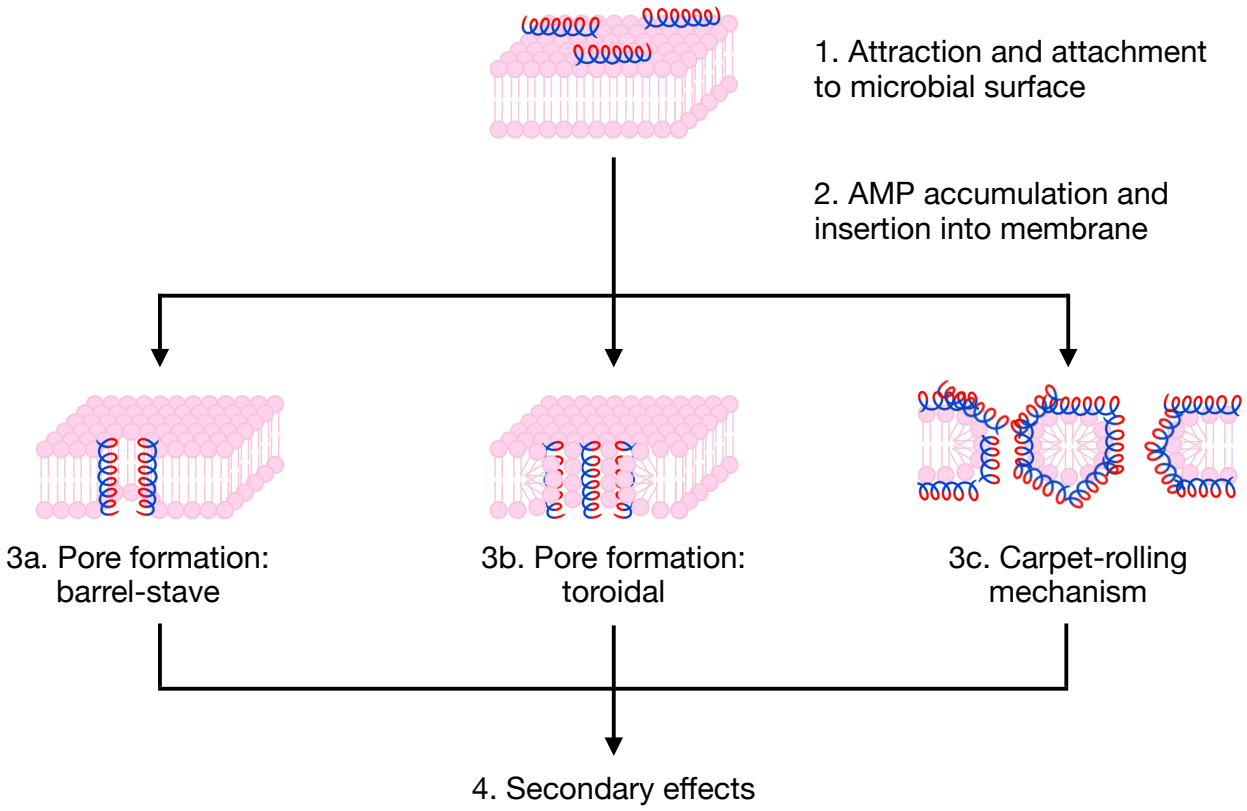


Figure 1.3.3. Proposed models of AMP interactions with membranes. *Circles represent membrane lipid head groups, lines represent fatty acid tails. Blue regions of AMPs represent hydrophobic regions, red regions represent hydrophilic regions. Adapted from Brogden, 2005.*

To date, approximately two dozen AMPs belonging to seven classes have been characterized in *Drosophila* (Hanson & Lemaitre, 2020). Expression of AMPs is induced following infection by the Toll and Imd pathways (Tzou, Reichhart & Lemaitre, 2002). In mutant flies deficient for the both immune pathways, infection did not induce expression of any known AMPs, and flies succumbed to the infection rapidly (Tzou, Reichhart & Lemaitre, 2002). Constitutive expression of a single AMP was sufficient to rescue mutants in some cases, indicating the importance of AMPs in host defense and the necessity of the pathways for AMP expression (Tzou, Reichhart & Lemaitre, 2002).

In *Anopheles*, a total of 11 AMPs belonging to four classes have been described so far: one attacin, four CECs, five defensins (DEFs), and one gambicin (Christophides et al, 2002). AMPs are synthesized by cells in the fat body and midgut epithelia, where they are secreted into the hemolymph and midgut lumen respectively (Christophides et al, 2002; Yassine & Osta, 2010). Within the midgut, expression of AMPs has been found to be enriched in the cardia compared to more posterior regions (Vizioli et al, 2000; Dong et al, 2006; Warr et al, 2007; Rodgers et al, 2020; Hixson et al, 2022). It is hypothesized that this expression pattern enables the immediate sterilization of a food source and attack of invading pathogens upon food intake, but also increased tolerance of the microbiota in the posterior (Dong et al, 2006; Rodgers et al, 2020).

Little is known about the sole attacin (ATT1) in *Anopheles,* but in *Hyalophora* (cecropia moths), attacin has been shown to interfere with the synthesis of bacterial outer membrane proteins (Carlsson et al, 1998). Cecropins are α-helical AMPs and exert their antimicrobial activity through a pore-forming mechanism, perhaps mediated by the amphipathicity of the helices (Christensen et al, 1988; Oren & Shai, 1998; Fig. 1.3.4). Three of the four cecropins (CEC1–3) cluster together in the genome on the X chromosome (Zheng & Zheng, 2002). The expression of CEC1 and CEC2 is both constitutive and inducible by immune challenge under the regulation of NF-κB factors (Vizioli et al, 2000; Zheng & Zheng, 2002), and CEC1 is potent against bacteria of both Gram types, fungi, and yeasts (Vizioli et al, 2000). Similarly, three defensins (DEF2–4) cluster together in the genome, and their mode of action has also been described to be pore-forming (Cociancich et al, 1993). Defensins are the most widespread AMPs in insects and are enriched with cysteine residues (Bulet et al, 1999). In *Anopheles*, DEF1

has been found to act primarily against Gram-positive bacteria and fungi, and its expression is induced in the hemolymph after bacterial infection (Vizioli et al, 2001a; Blandin et al, 2002).

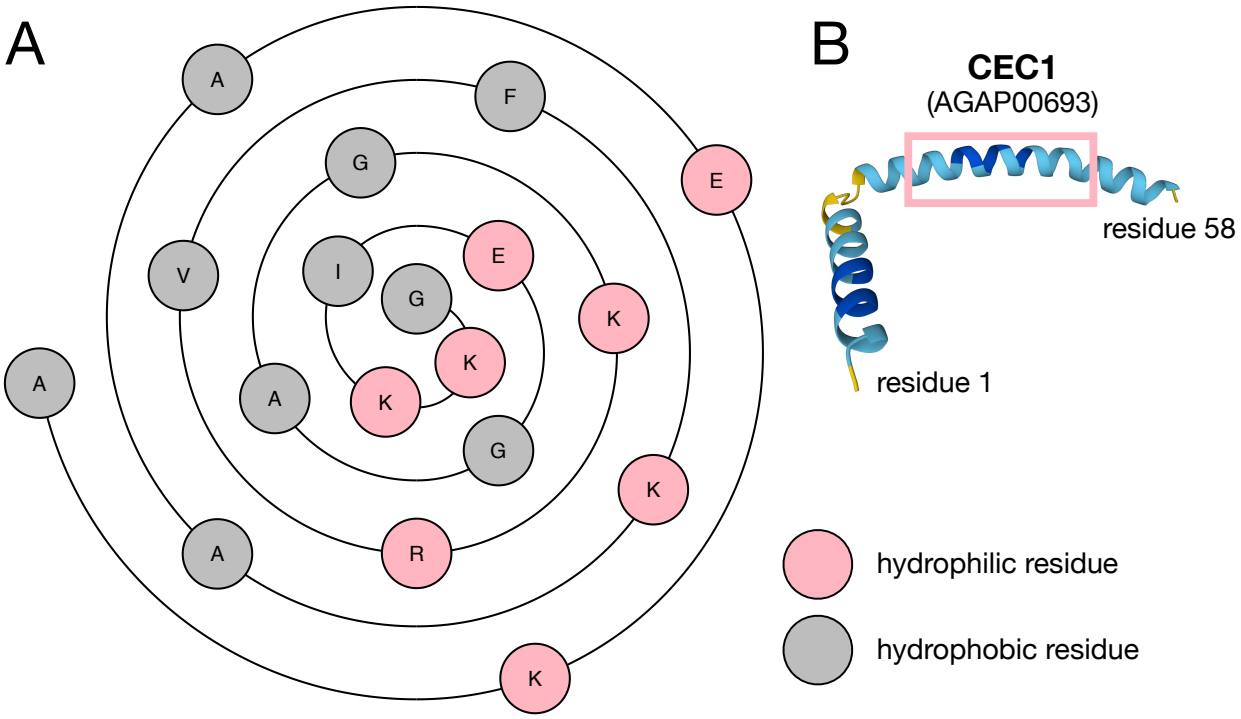


***Figure 1.3.4*. Amphipathicity and folding structure of CEC1**. *A) A wenxiang diagram representing the bird's eye view of an α-helix made up of residues in positions 30–47 of CEC1. Circles represent individual amino acids. The first residue of the sequence (G) is in the middle represented by a grey circle, which is connected to the next residue (K) in the sequence, and so on. Grey circles represent hydrophobic residues, pink circles represent hydrophilic residues. Created with the 'helixvis' package in RStudio (Wadhwa, Subramanian & Stevens-Truss, 2018). B) AlphaFold (Jumper et al, 2021) structural prediction of CEC1 obtained from VectorBase (Amos et al, 2022). Curled twists represent α-helices. Colored regions represent the confidence metric of the prediction, with dark blue, light blue, and yellow representing 'very confident', 'confident', and 'low confidence' regions of the prediction respectively. The region highlighted by the pink rectangle is depicted in part A of this figure.*

Gambicin is a novel AMP identified in *Anopheles* and *Aedes*, and it shares no compelling sequence similarity to any known genes or proteins (Vizioli et al, 2001b). GAM1 contains 8 cysteines forming four disulfide bridges and has been shown to have antimicrobial activity against bacteria of both Gram types, filamentous fungi, and mild lethality to *P. berghei in vitro* (Vizioli et al, 2001b). Similarly, depletion of GAM1 increased mosquito susceptibility to Gram-positive bacteria, *P. berghei*, and to a lower extent *P. falciparum* (Dong et al, 2006). Expression of GAM1 is induced upon bacterial infection, but not following a blood meal (Vizioli et al, 2001b), suggesting that it is also constitutively expressed. Interestingly, expression is induced after an infectious blood meal at 24 hours and 11–21 days post-infection, which corresponds to *P. berghei* ookinete invasion and the release of sporozoites into the hemocoel, followed by salivary gland invasion respectively (Vizioli et al, 2001b). While it has been hypothesized that gambicin evolved as a specific effector against *Plasmodium*, numerous studies have suggested this is unlikely (Morlais et al, 2004; Obbard et al, 2007; Rottschaefer et al, 2011).

1.4 Malaria intervention strategies: the past, present, and future

To date, successful malaria intervention has relied heavily on the use of insecticides to manage vector populations and break the cycle of transmission to humans (Collins & Besansky, 1994; Collins & Paskewitz, 1995; Hammond et al, 2016). Insecticide-treated bed nets and indoor residual spraying have been extremely effective, but their widespread use is thought to have facilitated the spread of insecticide resistance in vector populations (Ranson & Lissenden, 2016; Hancock et al, 2020). Another vector population management strategy is the manipulation of larval habitats, including: the reduction of the availability of suitable aquatic reservoirs, either temporarily or permanently; the reduction of water levels in aquatic habitats *e.g.* dams; and the addition of larvicides into standing water (Kibret et al, 2018; Martello et al, 2022; Peterson & Rolston, 2022).

In 2021, a major milestone was reached when the World Health Organization (WHO) endorsed the first malaria vaccine RTS,S/AS01 for widespread use (WHO, 2021). A significant number of malaria cases are to be prevented, however the vaccine efficacy in infants and children, the most vulnerable group to malaria-caused deaths, remains relatively low at around 40% after administration of three doses and a booster (RTS,S Clinical Trials Partnership, 2015). Another vaccine, R21/Matrix-M^{TM}, was also recently recommended by the WHO and showed an improved vaccine efficacy of up to approximately 80% (Datoo et al, 2022). The effect of this vaccine on reducing malaria incidences remains to be determined over the next years. Currently, deaths and cases of malaria are on the rise; the COVID-19 pandemic caused significant disruption to maintaining essential malaria services and led to an increase in malaria deaths between 2019 and 2022 (WHO, 2021; WHO, 2022; WHO, 2023). These limitations coupled with the spread of drug (*e.g.* artemisinin) resistance amongst parasites has caused malaria control efforts to stall over the last decade (WHO, 2022). Therefore, novel intervention strategies are needed now more than ever to continue curbing the spread of malaria.

A number of alternative intervention strategies are being developed, and if successful, will be implemented in time (Nash et al, 2019). One promising area of research is the generation of transgenic mosquitoes, namely the modulation of the spatiotemporal expression of immune effectors such as AMPs. As the midgut serves as

the initial point of interaction between the mosquito host and pathogen(s) they transmit, it is therefore a tissue of particular interest to study vector control for disease transmission. Here, a major bottleneck also exists for the parasite during its transition from gametocytes to oocysts, *i.e.* midgut stages and during midgut invasion (Sinden, 2010). These massive parasite losses are largely attributed to midgut mosquito immune responses (Cirimotich et al, 2010). By modulating immune mechanisms that already exist within the mosquito midgut, expression of immune effectors can be modified to directly target and kill the invading parasites, or damage them to the extent that their fitness is reduced and are eliminated by other immune responses such as the complement-like system (Povelones et al, 2009). This has already been achieved by a number of studies, such as through the overexpression of immune effectors, spatiotemporal manipulation of their expression (Kim et al, 2004; Corby-Harris et al, 2010; Dong et al, 2011; Dong et al, 2012; Simões et al, 2017b), and alteration of gene haplotypes (Volohonsky et al, 2017), although a limited subset of effectors and loci have been explored.

In a study by Dong et al, REL2 was fused to the zinc carboxypeptidase A1 (CP-REL2) and vitellogenin (VG-REL2) promoters for blood meal-inducible expression in the midgut and fat body respectively (2011). These transgenic lines had increased resistance to *P. falciparum* infection with minimal fitness costs, although the CP-REL2 line was more resistant than VG-REL2 transgenic mosquitoes (Dong et al, 2011; Pike et al, 2017). This effect was achieved independently of the microbiota, and was also prominent when infection intensities were low (Dong et al, 2011). Further testing of the lines involved mixed cage experiments, where transgenic mosquitoes were co-reared with WT mosquitoes in equal starting numbers. After one generation only, the proportion of the CP-REL2 transgene rose from 50% to 90%, which carried over for the next nine generations (Pike et al, 2017). This was only observed in the CP-REL2 line, and it was discovered that the modulation of midgut microbiota through midgut-specific *REL2* expression conferred a mating preference of males for the opposite genotype, *i.e.* WT males preferred to mate with transgenic females, and transgenic males preferred to mate with WT females (Pike et al, 2017). The CP-REL2 mosquitoes also maintain heightened *Plasmodium* resistance even after 50 generations of culturing (Pike et al, 2017). These results suggest stability of the phenotypes and the potential for this transgene to efficiently

self-spread if released into wild populations. This transgenic line further highlights the importance of considering the microbiota, and is also a great candidate line for deployment into the field to reduce malaria transmission.

Another study also generated transgenic mosquitoes, where a kinase essential to insulin signaling was overexpressed in the midgut, which led to the total inhibition of *P. falciparum* development in the mosquito (Corby-Harris et al, 2010). Heterozygous individuals were also able to block up to 99% of parasite development, however strong fitness costs were associated with this transgenic manipulation (Corby-Harris et al, 2010). For use as a malaria control strategy, specifically field deployment and/or coupling with a gene drive system, it is imperative that fitness costs are assessed.

Exogenous immune effectors have also been introduced into *Anopheles* to target *Plasmodium*. For example, exogenous effectors have been sourced from other insects (Dong, Simões & Dimopoulos, 2020; Hoermann et al, 2022) or were synthetically designed and integrated (Isaacs et al, 2011; Dong, Simões & Dimopoulos, 2020). Monoclonal antibodies against parasite proteins have been found to inhibit parasite development *in vitro* or when fed to mosquitoes in an infectious blood meal (Hollingdale et al, 1984; Barr et al, 1991; Burkot et al, 1991; Li et al, 2004). Specifically, single-chain antibody (scFv) 2A10 binds to circumsporozoite protein on the parasite surface and was found to inhibit hepatocyte invasion by sporozoites *in vitro* (Hollingdale et al, 1984; Burkot et al, 1991). In order for successful integration of antibodies into the *Anopheles* genome, Isaacs et al utilized scFvs derived from the monoclonal antibodies, which are smaller and less complex while still retaining specific binding activity (2011). The scFv of 2A10 was coupled with CEC1, a known antimalarial, and regulatory elements of VG to enable fat body expression during and after midgut invasion. Transgenic mosquitoes with scFv-2A10 had reduced infection intensities and prevalence of *P. falciparum* sporozoites, demonstrating the *in vivo* transmission blocking effect of cecropin and 2A10 (Isaacs et al, 2011).

Using a similar strategy, Dong, Simões & Dimopoulos introduced several exogenous effectors into the *An. stephensi* with anti-*Plasmodium* activity (2020). Under the CP promoter, five exogenous effectors enabled targeting of midgut parasites: melittin, an AMP from *Apis mellifera* (honey bee); transportan 10 dimer, an AMP from *Vespula*

lewisii (Korean yellow-jacket wasp); Shiva-1, a cecropin-like synthetic peptide; enolase-plasminogen interaction peptides (EPIPs); and scorpine, an AMP from *Pandinus imperator* (emperor scorpion). Moreover, under the VG promoter, scorpine and scFv 2A10 were expressed in the fat body. This line of transgenic mosquito enabled bombardment of invading parasites from either side of the midgut epithelium. Indeed, the infection intensity and prevalence of *P. falciparum* salivary gland sporozoites were significantly reduced in this transgenic line compared to WT mosquitoes (Dong, Simões & Dimopoulos, 2020). This superior *Plasmodium* resistance came at a cost, as longevity was significantly reduced (Dong, Simões & Dimopoulos, 2020).

Another interesting avenue of research has utilized the relatively quick and simple manipulation of microbes to deliver effector molecules. In 2007, Riehle and colleagues found that expression of exogenous antimalarials phospholipase A2 (PLA2) and salivary gland and midgut peptide 1 (SM1) in *Escherichia coli* reduced *P. berghei* oocyst development in *An. stephensi* mosquitoes by 39%. *E. coli* cells were cleared rapidly from the midgut (~2 days) and this had an effect on parasite development when fed one day post-*Plasmodium* infection. If this were to be used as a malaria control strategy for wild populations, this would mean that mosquitoes only have a short window after parasite ingestion to uptake the *E. coli* for this strategy to work. Introduction of *Enterobacteria agglomerans*, an abundant member of the microbiota in the researchers' mosquito colony, also led to a survival time of two days in the midgut. As a proof of principle, this mosquito-associated bacteria underwent two rounds of selection, which enabled persistence for up to 14 days (Riehle et al, 2007). Taken together, the significant reduction in oocysts and ability to select for midgut persistence demonstrates the feasibility of this approach (Riehle et al, 2007).

A few years later, the same group genetically engineered another natural symbiotic bacterium, *Pantoae agglomerans*, to secrete different anti-*Plasmodium* effectors including scorpine, EPIP, and SM1. Feeding of this species to *An. gambiae* significantly reduced *P. falciparum* infection prevalence without having a significant impact on their lifespan (Wang et al, 2012). Another strategy involved an engineered fungus *Metarhizium anisopliae* to express SM1 and scorpine, where spraying of the transgenic fungal spores reduced salivary gland *P. falciparum* sporozoite density by over 90% in *An. gambiae*

(Fang et al, 2011). Importantly, mosquitoes could be infected with the fungal spores at 11 days post-*Plasmodium* infection, and a reduction in sporozoite density could be observed as early as three days later (Fang et al, 2011). Therefore, the use of this fungus drastically expands the window in which mosquitoes can be infected and subsequently block salivary gland invasion by sporozoites. The introduction of such engineered microbes to wild *Anopheles* can be achieved through the use of sugar bait traps, however the caveat is that these microbes need to be widely accessible and 'released' regularly in order for an increased chance of uptake by mosquitoes in the field. A study in 2017 by the same group identified an endosymbiont that was able overcome this problem. *Serratia* strain AS1, a natural endosymbiont of *Anopheles*, colonizes numerous organs of the mosquito and exhibits superior transmission, where it was observed to be spread in a vertical, horizontal, and transstadial manner (Wang et al, 2017). AS1 also propagates in water which enables ingestion by larvae, and while the larval gut microbiota is purged during metamorphosis, the bacteria persisted in the adults of AS1-infected larvae (Wang et al, 2017). Similarly, while microbiota are purged post-blood meal digestion with the PM, AS1 populations expand and persist post-blood meal (Wang et al, 2017). As before, Wang et al genetically engineered this strain to express and secrete various antimalarials, such as scorpine, EPIP, Shiva1, and PLA2 (2017). Introduction of transgenic AS1 to mosquitoes two days prior to *P. falciparum* infection inhibited oocyst formation by up to 93% (Wang et al, 2017). Therefore, the self-propagation and antimalarial properties of transgenic AS1 makes it a feasible bacterial strain to be applied to wild populations of mosquitoes.

These studies provide valuable insights in the use of transgenic mosquitoes in blocking malaria transmission. Coupled with a gene drive system, release of such mosquitoes into the wild can be an effective tool for population replacement, where wild mosquitoes are eventually replaced with malaria-refractory transgenic populations. Gene drives involve the use of selfish elements to drive the transmission of a trait to the next generation in a super-Mendelian manner (Alphey et al, 2020). This can be achieved through the expression of an endonuclease and a guide RNA in the germline, which enables cleavage of a target site and subsequent homing of a desired transgene to spread to the offspring. Some of the main limitations of using genetically modified

mosquitoes are legislative and regulatory hurdles imposed by local governments, as such technology does not have a long-standing history of use (Roberts et al, 2017; Teem et al, 2019; Connolly et al, 2021; Peterson & Rolston, 2022). Results from controlled field trials need to be performed, and if successful, will catalyze legislative changes and provide peace of mind to stakeholders, locals in the affected areas, and governing bodies. While the use of transgenic mosquitoes is an incredibly promising tool in combating malaria, there is much more to be understood. Currently, there is no available system to experimentally assess transmission rates of human malaria parasites from mosquitoes.

1.5 Thesis aims

The overarching aim of this Ph.D. project was to characterize mosquito immune responses in the adult malaria mosquito, particularly in the midgut. This was further sub-divided into the following aims:

1) To characterize mosquito transcriptomes and immunity in a spatiotemporal manner, particularly in the midgut, including the regulation of tissue identity and in the context of *P. falciparum* infection.
2) To identify novel AMPs and investigate their roles in defense against *P. falciparum* and bacteria.

The first aim is addressed in Chapters 3 and 4 through the RNA-sequencing (RNA-seq) of mosquito tissues at various time points post-blood feeding/infection. The second aim is addressed in Chapter 5, where machine learning predictive algorithms were employed to identify putative AMPs, which were subsequently screened for antibacterial and antiplasmodial activity. The characterization of the mosquito midgut and particularly its immune responses could enable the generation of malaria-refractory transgenic mosquitoes in a two-fold manner: 1) the expression of endogenous antimalarial effectors can be genetically manipulated to improve efficacy against *Plasmodium*, and/or 2) the loci of genes with a desirable expression profile can be modified to include endogenous and/or exogenous effectors to target the parasite in a controlled, spatiotemporal manner. In the future, these transgenic mosquitoes can be assessed as a potential malaria control strategy and coupled with a gene drive.

Chapter 2: Methods

2.1 Mosquito rearing

The lab colony of *An. coluzzii* N'gousso strain mosquitoes (formerly *An. gambiae* M form, and henceforth referred to as *An. gambiae* for simplicity and comparison) was previously established from field-collected mosquitoes in Yaoundé, Cameroon in 2006. The majority of mosquitoes used for experiments were reared and provided by Temesgen Kebede. Mosquitoes were reared and maintained under standard insectary conditions at 80% humidity, 28˚C, and a 12 hour light/dark cycle, including half an hour of dawn and dusk transitions. Eggs were surface-sterilized with a 0.37% bleach solution and hatched in plastic trays with demineralized water. All aquatic stages of larvae were provided with fish food (Nishikoi, UK) on a daily basis. Throughout development, larvae were further sub-divided into additional plastic trays until pupation to prevent overcrowding and competition over resources. Pupae were collected into a small plastic cup with ~100 mL demineralized water and placed in mesh cages (BugDorm, Taiwan) for adult emergence. Adults were provided with *ad libitum* access to a sterile 10% fructose solution (Sigma-Aldrich, UK). At 2–20 days post-emergence, adults were provided human concentrated red blood cells (cRBCs; NHS Blood and Transplant, UK) or cow blood meals (First Link, UK) to stimulate egg production every 3 days or as needed. The human cRBCs were processed within 17 days post-blood draw and were derived from a single donor.

2.2 P. falciparum cultures and mosquito infections

The methods described in this section were adapted from Habtewold et al (2019) and were performed according to the Imperial College London containment level 3 laboratory (CL3) Code of Practice and standard operating procedures. The majority of P. falciparum *parasites used in experiments were cultured and provided by Maria Grazia Inghilterra.*

2.2.1 Asexual cultures

P. falciparum (NF54 strain) obtained from the Malaria Research and Reference Reagent Resource Center were used for all malaria infections of mosquitoes. Asexual parasite cultures were maintained at 37˚C in 25 cm^3 flasks containing 500 µL human cRBCs and 10 mL complete media (Roswell Park Memorial Institute media [Sigma-Aldrich, UK] supplemented with 10% human serum [Interstate Blood Bank, USA], hypoxanthine to a final concentration of 0.05 g/L [Sigma-Aldrich, UK], and L-glutamine to a final concentration of 0.3 mg/L [Sigma-Aldrich, UK]). Media was replaced daily (10 mL) and flasks were fumigated with malaria gas (92% N_2, 5% CO_2, and 3% O_2) each time the culture flasks were opened. Blood smear staining of the asexual cultures was performed with Kwik-Diff™ reagents (Thermo Fisher Scientific, UK) every few days to assess the parasitemia and life stage of the cultures. Asexual parasite cultures were diluted and passaged as needed for maintenance, typically every 2–4 days.

2.2.2 Gametocyte cultures and mosquito infections

Gametocytes were induced from ring-stage asexual cultures at 1% parasitemia with 7 mL complete media and 500 µL human blood, and maintained for at least 14 days as described above for asexual cultures, with the exception that 7 mL instead of 10 mL of media was replaced daily. Gametocyte viability and maturity were checked at 14–17 days post-induction via an exflagellation assay and blood smear staining of the culture using Kwik-Diff™ reagents. Exflagellation assays involved pelleting a ~500 µL aliquot of the gametocyte culture and resuspending the blood pellet with ~30 µL ookinete media (complete media supplemented with 0.1% xanthurenic acid to a final concentration of 100 µM). Samples were incubated for 12–15 minutes at room temperature and loaded into Neubauer chambers. Exflagellating male gametocytes were observed under a light

microscope using the 10x objective in a randomly selected 4×4 square grid. All cultures containing viable and mature parasites were pooled in 15 mL Falcon tubes pre-heated to 38˚C with an additional aliquot of fresh blood (~100 µL for each flask) and centrifuged for 5 min at 0.6 × *g* and 38˚C. The supernatant was discarded and blood pellets were pooled, and an equal volume of heat-inactivated serum pre-heated to 38˚C was added. The serum and blood were mixed gently with a pipette and maintained at 38˚C prior to feeding.

Prior to malaria infection, approximately 70 female mosquitoes were collected in paper cups covered with a double-layered mesh netting and starved for at least 3 hours. Glass feeders were connected to a water bath heated to 38˚C and covered with a stretched piece of Parafilm (Bemis Company, USA) to create a membrane for the mosquitoes to feed on. Using a pre-heated pipette tip, *P. falciparum*-infected blood was transferred into the cavity of the glass feeders, which were then lowered onto the mesh netting over the paper cups containing the mosquitoes. Mosquito were allowed to feed on 0.5–1 mL of *P. falciparum*-infected human blood for 15 minutes in the dark via a standard membrane feeding assay (SMFA). *P. falciparum*-infected mosquitoes were then transferred to a 26˚C, 80% humidity incubator in secondary containment boxes and kept in the dark for 48 hours following infection. During this time, mosquitoes had no access to food or water to starve and kill unfed mosquitoes. After the 48 hour darkness period, the mosquitoes were then exposed to the standard insectary lighting conditions, and provided a 10% sterile fructose solution on a cotton ball which was replaced every 48 hours.

2.2.3 Oocyst enumeration

Malaria-infected mosquitoes at 7–10 days post-infection were killed with CO_2 gas and submerged in 70% ethanol for 1–2 minutes. Mosquitoes were washed thrice with phosphate-buffered saline (PBS) prior to dissections (Sigma-Aldrich, UK). Infected midguts were dissected into ice-cold PBS and stained in 0.3% mercurochrome (Sigma-Aldrich, UK) for 20 minutes. Midguts were washed with PBS and fixed in 4% paraformaldehyde (Sigma-Aldrich, UK) for 30 minutes, followed by another PBS wash prior to mounting. Stained midguts were mounted on a glass microscope slide with filter-sterilized 80% glycerol, and cover slips were sealed with clear nail polish. Oocysts were

counted and imaged via an EVOS^{TM} M5000 (Thermo Fisher Scientific, UK) imaging system with 10x or 20x objectives using the transmission filter. Wilcoxon rank-sum tests and generalized linear models (GLM) with a quasi-Poisson distribution factoring in the effect of condition, replicate, and the interaction between condition and replicate were used to assess differences in oocyst infection intensities. Prior to these analyses, midguts with zero oocysts (*i.e.* uninfected) were excluded. To assess differences in infection intensity between groups, χ^2 tests were employed. χ^2 tests were performed without the default Yates continuity correction as it is unnecessarily restrictive (Camilli & Hopkins, 1979). All statistical analyses were performed in RStudio (RStudio Team, Boston, MA, USA).

2.2.4 P. falciparum *and treatment with artemisinin/MF/KIN1*

Ring-stage sexual cultures were diluted to 1% parasitemia in 10 mL cultures and were smeared to calculate the proportion of each asexual life stage. Artemisinin (Thermo Fisher Scientific, UK) and methyl farnesoate (MF; Echelon Biosciences, USA) were dissolved in dimethyl sulfoxide (DMSO; Sigma-Aldrich, UK) to four stock concentrations of 100 mM, 10 mM, 1 mM, and 0.1 mM. Cultures were treated with the artemisinin or MF stocks in 1:1000 dilutions to a final concentration of 100 µM, 10 µM, 1 µM, and 0.1 µM respectively, or an equal volume of DMSO (10 µL) as a control. Over a period of two days, ~500 µL aliquots of the cultures were smeared daily to calculate the proportion of each asexual life stage. No media changes took place, and flasks were fumigated with malaria gas each time the flasks were opened. Differences in growth were assessed with Wilcoxon signed-rank tests at each time point in RStudio.

Exflagellation assays involving MF were performed as described previously in section 2.2.2, with the addition of the 100 mM MF stock diluted 1:1000 in the ookinete media to a final concentration of 100 µM or an equal volume of DMSO. Exflagellating male gametocytes were counted and differences in the exflagellation rate were assessed with paired Wilcoxon rank-sum tests in RStudio.

Prior to mosquito infections, gametocyte-infected blood was supplemented with the 100 mM MF stock to a final concentration of 100 µM (1:1000 dilution), or an equal volume of DMSO. In separate experiments, the kininogen-1 (KIN1) synthetic peptide

synthesized by GenScript, USA, was dissolved in nuclease-free water (Thermo Fisher Scientific, UK) to a stock concentration of 1 mM. The KIN1 peptide stock was added to infected blood in a 1:20 dilution to a final concentration of 50 µM, or an equal volume of nuclease-free water was added as a control. Mosquitoes were infected with *P. falciparum* and oocysts were counted and analyzed at 7–10 days post-infection as described previously in section 2.2.3.

2.3 Mosquito sample collection

2.3.1 Adult and midgut sample collection

For RNA-seq of midgut tissues, adult female mosquitoes (2–3 days post-emergence) were fed human blood via an SMFA. Mosquitoes were anesthetized with 2–5% CO_2 and dissected at 0 hours (*i.e.* before blood feeding), 5 hours, and 24 hours post-blood feeding. At each time point, the cardia, anterior midgut, and posterior midgut tissues of approximately 50 mosquitoes were collected into 5–10 µL of ice-cold PBS. Each sample was then stored in 500 µL TRIzol™ reagent (Thermo Fisher Scientific, UK) and homogenized prior to RNA extraction and sequencing. Three biological replicates were completed. The experimental workflow is summarized in Figure 2.3.1.

For other experiments, various tissues from sugar-fed mosquitoes were dissected and collected, including: whole adults (with heads removed, also referred to as 'carcasses'), whole midguts, three midgut tissue sections (cardia, anterior midgut, posterior midgut), two midgut tissue sections (cardia and anterior midgut together, posterior midgut separately), and remaining carcasses (midgut removed). For one experiment involving the feeding of exogenous MF, the 100 mM MF stock (as described in section 2.2.4) was diluted into nuclease-free water to a final concentration of 100 µM (1:1000 dilution) and provided to mosquitoes via a cotton ball. Mosquito midguts (cardia/anterior and posterior regions) were dissected 2 days later.

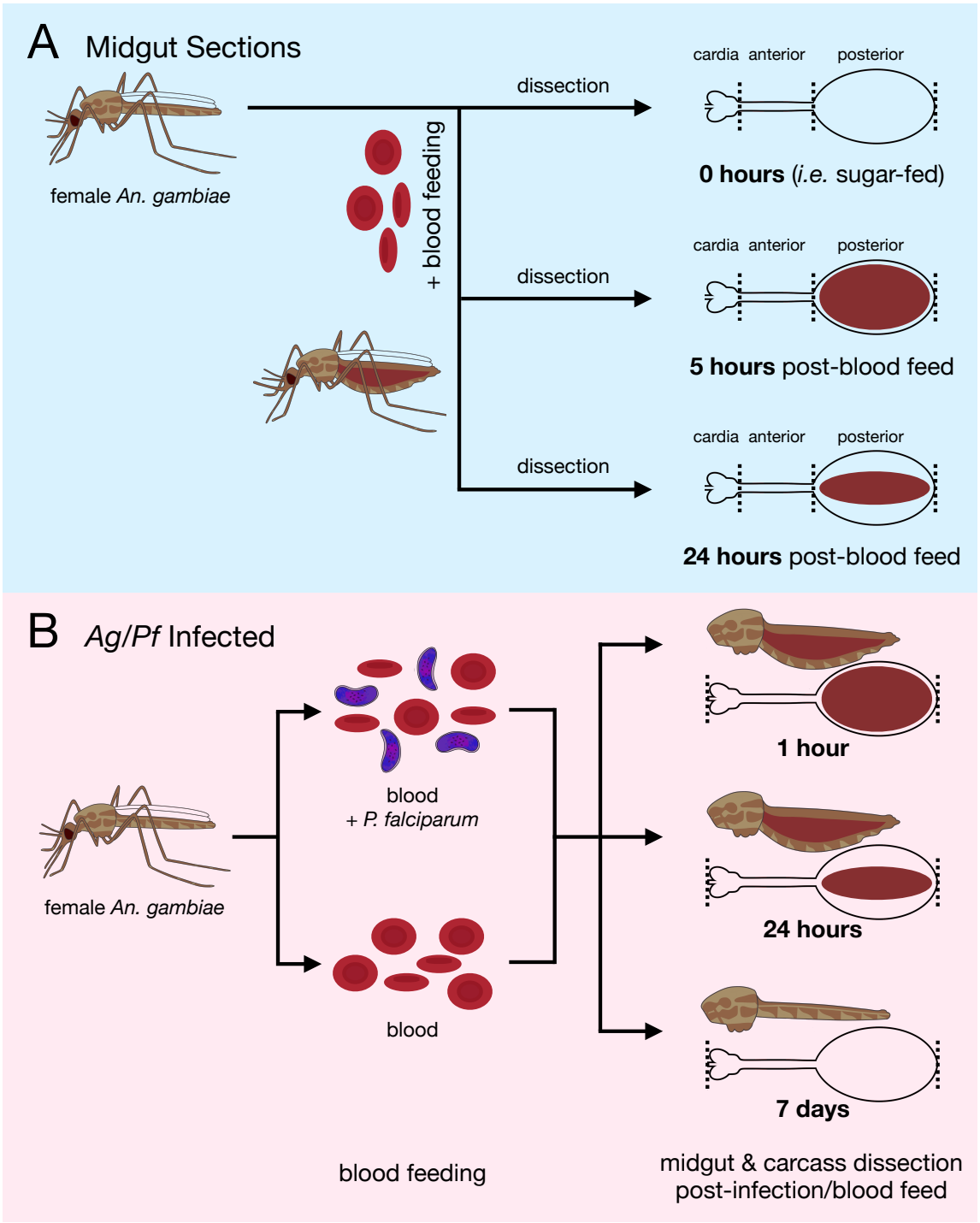


Figure 2.3.1. Mosquito samples collected for RNA-seq. *Experimental workflows for A) the midgut sections RNA-seq data; midgut tissues (cardia, anterior midgut, posterior midgut) were collected from approximately 50 mosquito individuals at various time points post-blood feeding, at 0 hours (*i.e. *prior to blood feeding, sugar-fed), 5 hours, and 24 hours, three replicates were performed in total; and B) the* Ag/Pf *data; whole midguts and whole, decapitated carcasses were collected from approximately 20–30 mosquito individuals at various time points post-blood meal or* P. falciparum *infection. Three replicates were performed for midgut samples and two replicates were performed for carcass samples.*

2.3.2 Collection of life stages

At least 10 mosquito individuals of the following developmental stages were collected in 500 µL TRIzol™ reagent for RNA extraction to assess spatiotemporal expression of specific genes: eggs, L2 larvae, L4 larvae, pupae, sugar-fed adult females (never been blood-fed), adult males, and blood-fed adult females. Adult mosquitoes were anesthetized with CO_2 and beheaded prior to storage into 500 µL TRIzol™ reagent. A total of six biological replicates were completed for this experiment.

2.3.3 P. falciparum*-infected sample collection*

Infected samples were handled according to the CL3 Code of Practice and standard operating procedures. Malaria-infected mosquitoes were killed with CO_2 gas and submerged in 70% ethanol for 1–2 minutes. Mosquitoes were washed thrice with PBS prior to dissections (Sigma-Aldrich, UK). For RNA-sequencing, approximately 30 midguts and carcasses (whole, decapitated mosquitoes) were collected at three time points: 1 hour, 24 hours, and 7 days post-infection. Samples from at least 20 mosquito individuals were pooled in 500 µL TRIzol™ reagent and homogenized with an electronic mortar and pestle. At 7 days post-infection, additional midguts were dissected for oocyst staining and counting as described in section 2.2.3 to assess infection intensity and prevalence. A total of 2 and 3 replicates were performed for the carcass and midguts samples respectively. The final replicate of midgut samples was performed with assistance from Yunyue (Zoe) Zhou (MRes student, project supervised by Dr. Dina Vlachou). This methodology was repeated for uninfected controls (*i.e.* blood-fed without *Plasmodium*) and is outlined in Figure 2.3.1.

2.4 RNA extraction

RNA extraction with TRIzol™ reagent was performed according to the manufacturer's instructions with a starting volume of 500 µL. Samples were first homogenized with an electronic mortar and pestle. RNA pellets were resuspended in 10, 20, or 50 µL nuclease-free water depending on the sample type. The RNA concentration was determined with a NanoDrop spectrophotometer (Thermo Fisher Scientific, UK). Isolated total RNA samples were submitted to GeneWiz Ltd. (UK) for RNA-sequencing, or used as templates for complementary DNA (cDNA) synthesis.

2.5 cDNA synthesis and qRT-PCR

From the isolated RNA, cDNA was synthesized using random primers and SuperScript™ III Reverse Transcriptase following the manufacturer's protocol (Thermo Fisher Scientific, UK) and an overnight incubation at 50˚C. Quantitative reverse-time polymerase chain reactions (qRT-PCR) were performed on the Applied Biosystems 7500 Fast Real-Time PCR system (Thermo Fisher Scientific, UK) with the cDNA as a template, gene-specific primers (Sigma-Aldrich, UK; Table A.1), and Fast SYBR™ Green Master Mix (Thermo Fisher Scientific, UK) in 10 µL reactions according to the manufacturer's instructions. Reactions were performed in triplicate, and each reaction was assessed for specificity by analysis of a melt curve. The relative expression of a gene was calculated as described by Pfaffl (2001), normalizing to mosquito housekeeping gene *S7* before comparing to the expression ratio of a control condition or mosquito cDNA pool:

$$\text{Relative Expression} = \frac{\text{PE}\,(\textit{gene})^{C_T1 - C_T2}}{\text{PE}\,(\textit{S7})^{C_T3 - C_T4}}$$

where PE (*gene*) and PE (*S7*) refer to the primer efficiencies of the gene of interest and internal control *S7* respectively, C_T1 is the C_T value of the gene of interest in the control condition, C_T2 is the C_T value of the gene of interest in the treatment condition, C_T3 is the C_T value of *S7* in the control condition, and C_T4 is the C_T value of *S7* in the treatment condition. In other words, relative expression is equal to the primer efficiency of the gene of interest to the power of C_T1 minus C_T2, divided by the primer efficiency of

the housekeeping gene to the power of C_T3 minus C_T4. Primer efficiencies of a primer pair were determined by qRT-PCR amplification of serially-diluted mosquito cDNA pools with the following formula:

$$\text{Primer Efficiency} = 10^{\frac{-1}{\text{slope}}}$$

where the slope is calculated from C_T values of at least 3 serially-diluted concentrations of mosquito cDNA templates amplified with the primer pair of interest. In other words, the primer efficiency is equal to 10 to the power of -1 divided by the slope. Calculated primer efficiencies are listed in Appendix A.

Kruskal–Wallis tests were used to compare relative expression values between multiple groups. In order to identify which pairs of conditions were significantly different, Dunn's tests were performed *post-hoc* with Benjamini–Hochberg false discovery rate corrections. For individual comparisons, *e.g*. expression of a single gene in a KD condition versus *dsLacZ*-injected controls, Shapiro–Wilk tests were used to determine the normality of the data. Subsequently, Student's *t*-tests or Wilcoxon tests were used depending on the distribution of the data to determine significant differences in relative expression between two sample conditions. All statistical analyses were performed in RStudio.

2.6 *An. gambiae* RNA-sequencing data processing and analysis

2.6.1 RNA-seq data processing

Paired-end RNA-sequencing was performed on total RNA by GeneWiz Ltd. (UK) on the Illumina® NovaSeq™ platform with a configuration of 2 × 150 bp. The data was uploaded and processed on the public Galaxy webserver (Afgan et al, 2016). Reads were filtered for quality (Phred score > 15 or 20) and length (> 20 bp) with Cutadapt (Martin, 2011), and aligned to the *An. gambiae* PEST genome (AGAP4) from VectorBase (Amos et al, 2022) with HISAT2 (Kim et al, 2019). Alignment files were converted into transcript counts using featureCounts (Liao, Smyth & Shi, 2014) and the AGAP4 annotation file from VectorBase (Amos et al, 2022). Using DESeq2, the counts were normalized and differences in gene expression were determined, which were then assessed with Wald tests and Benjamini–Hochberg multiple test corrections (Love, Huber & Anders, 2014). Transcripts per million (TPM) were calculated from aligned reads using Stringtie (Pertea et al, 2015).

2.6.2 Hierarchical clustering

Hierarchical clustering was performed on the normalized counts across conditions, where counts were scaled and k = 11 in RStudio using the tidyverse (Wickham et al, 2019) and cluster packages (Maechler et al, 2021). The number of clusters were determined through the use of the 'elbow' (Within-Cluster-Sum of Squared Errors), average silhouette, and gap statistic methods. Clusters were visualized with the following packages: factoextra (Kassambara & Mundt, 2020), gplots (Warnes et al, 2020), and dendextend (Galili, 2015).

2.6.3 Venn diagrams and enriched GO terms

Venn diagrams of significantly DE genes were generated using Venny 2.1 (Oliveros, 2015) to determine shared and unique DE genes of various sample conditions. Significantly enriched gene ontology (GO) terms of a given list of genes/transcripts were identified with g:Profiler (Raudvere et al, 2019). In the identification of enriched GO terms in the top secreted transcriptome of each midgut tissue at each time point, the statistical domain scope (*i.e.* background) was changed to only contain genes with a signal peptide.

2.7 An. darlingi data

All sample/data collection experiments pertaining to this section were conducted by Dr. Bianca Carlos of São Paolo State University (UNESP), Brazil, in the context of a collaboration between Imperial College London (Prof. George K. Christophides and Dr. Dina Vlachou), UNESP (Dr. Jayme Souza-Neto), and Amazonia International Centers of Excellence for Malaria Research (Dr. Marta Moreno and Prof. Joseph Vinetz).

Laboratory-reared *An. darlingi* mosquitoes were allowed to feed on 6 *P. vivax*-infected patients in Iquitos, Peru, or heat-treated control blood from the same patients. Mosquito midguts from 20 individuals were dissected and pooled at 6 time points post-infection/blood feed (1 hour, 6 hours, 18 hours, 22 hours, 26 hours, and 7 days) for RNA extraction and RNA-sequencing using the Illumina® MiSeq™ platform. At 7 days post-infection, an additional 15 mosquito midguts fed on each patient were dissected to assess the infection intensity and prevalence (Table 2.7.1; unpublished data).

Table 2.7.1**. **Infection intensities of** **An. darlingi** **midguts infected with** **P. vivax** **from gametocyte-carrying patients in the Amazonian rainforest of Peru.

Patient #	Oocyst range	Oocysts (n = 15)		Prevalence (%)
		Mean	Median	
1	0–5	1.9	2	67
2	0–162	102.9	104	87
3	0–4	1.7	2	80
4	0–5	2.6	3	80
5	9–146	72.4	72	100
6	0–62	32.4	40	80

The RNA-seq data were originally analyzed by the investigators using the fragments per kilobase of transcript per million mapped reads method and were reanalyzed in the context of this project as described previously in section 2.6.1 in order to make direct comparisons with the *An. gambiae*/*P. falciparum* data.

2.8 Novel AMP identification

2.8.1 AMP prediction

From the RNA-seq data, gene sequences were filtered by $\log_2$ fold change (≥ 2 or ≤ -2) between any pairwise comparison, significance (adjusted *p*-value ≤ 0.05), length (≤ 800 bp), having a signal peptide, and those that did not have descriptions/annotations. Sequences that fulfilled these criteria were selected for further analysis through 6 AMP prediction algorithms available on the following 3 AMP databases: ADAM (Lee et al, 2015), iAMP-2L (Xiao et al, 2013), and $CAMP_{R3}$ (Waghu et al, 2015). Some genes predicted to be AMPs by at least 4 of the 6 predictive algorithms were tested in subsequent assays for antimicrobial activity.

2.8.2 de novo *transcriptome assembly*

Quality control steps of the RNA-seq data of midgut tissues were performed as described in section 2.6.1, and reads were mapped to the genome with altered splice options using HISAT2 (Kim et al, 2019). In order to generate a transcriptome assembly from the data, Stringtie (Pertea et al, 2015) was used to predict transcripts with a minimum transcript length of 30 bp, which were subsequently merged with the original list of transcripts (annotation file) to combine redundancies. GffCompare (Pertea & Pertea, 2020) was employed to collate annotations of the transcripts from the new and original transcriptomes into a new annotation file. Using this new annotation file, counts were assigned to reads using featureCounts (Liao, Smyth & Shi, 2014) and differentially expressed (DE) genes were determined with DESeq2 (Love, Huber & Anders, 2014). Novel gene sequences (*i.e*. without gene name annotations) were filtered as described in section 2.8.1 with the addition of two new criteria: significantly upregulated in the cardia compared to the posterior of the same time point, and lacking any association with a known gene. Signal peptide prediction was performed on the sequences of these remaining transcripts using SignalP 6.0 (Teufel et al, 2022).

2.9 RNAi

Students Yunyue (Zoe) Zhou (MRes, project supervised by Dr. Dina Vlachou), Fanyun (Coco) Wu (BSc final year project, project supervised by Prof. George K. Christophides), and Sihao Ma (BSc final year project, project supervised by Prof. George K. Christophides) contributed to the collection of data involving the methods detailed in this section.

2.9.1 dsRNA synthesis

Genes were amplified with gene-specific primers designed with E-RNAi (Horn & Boutros, 2010) from mosquito midgut cDNA using CloneAmp HiFi (Takara, Japan) or Phusion high fidelity (NEB, USA) polymerases in 100 µL polymerase chain reactions (PCR; Table A.1). *LacZ*, a bacterial gene encoding β-galactosidase, and a gene encoding green fluorescent protein (GFP) served as controls for the effect of injection as these sequences are not naturally present in the mosquito genome. For *LacZ,* the pCR™-Blunt II-TOPO® vector (Thermo Fisher Scientific, UK) was used as a template. For *GFP*, cDNA from a transgenic line of mosquito expressing GFP as a marker was used as a template (Hoermann et al, 2022). PCR products were purified using the QIAquick™ PCR Purification Kit (Qiagen, Germany) and analyzed by electrophoresis (110 V, 30–40 minutes) on a 1% agarose gel. If a multiple bands were present, the band corresponding to the correct fragment size was extracted and purified from the gel using the QIAquick® Gel Extraction Kit (Qiagen, Germany). Purified/extracted PCR products were used as a template for further amplification with T7-flanked gene-specific primers and CloneAmp HiFi or Phusion high fidelity polymerases in 100 µL reactions (Table A.1). The T7 PCR product was purified with the QIAquick™ PCR Purification Kit and verified for the correct band size via gel electrophoresis. Additionally, purified PCR products were sent for Sanger sequencing using GeneWiz Ltd. services to check for the correct amplicon sequence. The T7 PCR product was then used as a template for dsRNA synthesis with the TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific, UK). The dsRNA was purified with the RNeasy® Mini Kit (Qiagen, Germany), eluted in nuclease-free water, analyzed for the correct size via gel electrophoresis as described above, and adjusted to a concentration of 3,000 ng/µL.

2.9.2 RNAi-mediated silencing of genes

Adult mosquitoes aged 0–2 days post-emergence were anesthetized with CO_2 and injected thoracically with 69 nL of 3,000 ng/μL gene-specific dsRNA using a Nanoject II (Drummond Scientific, USA) to achieve transient, *in vivo* KD of genes. In order to silence two genes, dsRNA constructs of both genes were mixed in a 1:1 ratio and 138 nL was injected into each mosquito (*i.e.* 69 nL of the mixture was injected twice). Injected mosquitoes were divided into paper cups covered with a mesh net secured with a rubber band. Sugar was provided via cotton balls, and mosquitoes were maintained under insectary conditions in secondary containment boxes lined with water-soaked paper towels. After 2–3 days, sample collection (section 2.3.1), *P. falciparum* infection (section 2.2), or bacterial infection (section 2.10) was performed. In the case of *P. falciparum* infections following dsRNA injections, an additional cup of non-injected mosquitoes (~50) was included as a positive control for the infection.

2.9.3 Calculation of knockdown efficiency

The calculation of KD efficiencies was performed on specific mosquito tissues or whole mosquitoes depending on the experiment. For example, if expression of genes in the cardia were to be investigated post-knockdown, the KD efficiencies were calculated from the same samples (*i.e.* the cardia in this example). Following bacterial or *P. falciparum* infections, approximately 10 whole mosquitoes were collected for each gene in each replicate. RNA extraction, cDNA synthesis, and qRT-PCR reactions were performed as described in section 2.4 and section 2.5. The relative expression of each gene was calculated from the C_T values normalized to *S7*, and compared to *dsLacZ*-injected control mosquitoes as described in section 2.5.

2.10 Bacterial infections

2.10.1 Bacterial cultures

Frozen glycerol stocks of *Enterococcus faecalis* (OG1RF strain), *Serratia marcescens* (Db11 strain)*, E. coli* (ATCC 23724), and *Staphylococcus aureus* (RN6390 strain) kindly gifted by Dr. Dominique Ferrandon (IBMC, Strasbourg, France) were cultured in 200–500 mL of lysogeny broth (Lennox) (for *S. marcescens* and *E. coli;* Sigma-Aldrich, UK) or brain heart infusion broth (for *E. faecalis* and *S. aureus;* Sigma-Aldrich) and grown overnight at 37˚C in a shaking incubator at 225 rpm. Bacterial cells were pelleted, washed twice with sterile PBS, and resuspended in sterile PBS or a 10% fructose solution for systemic and oral infections respectively (Sigma-Aldrich).

2.10.2 Bacterial infections

Adult mosquitoes aged 2–5 days were fed with *S. marcescens* (OD_{600} = 0.1) resuspended in a sterile 10% fructose solution and provided in a cotton ball for 24 hours, or injected thoracically with 69 nL *E. coli* (OD_{600} = 0.4) with a Nanoject II (Drummond Scientific). Control mosquitoes were fed with sugar or injected with 69 nL sterile PBS respectively. Mosquitoes were maintained as described in section 2.9.2 and at 2–3 days post-infection, mosquitoes were dissected and collected as described in section 2.3.1. For *S. marcescens*-infected mosquitoes, midguts and the remaining carcasses (heads removed) were collected for RNA extraction whereas headless mosquito carcasses were collected for *E. coli*-infected mosquitoes.

2.10.3 Bacterial survival assays

At 2 days post-dsRNA injection, the surviving mosquitoes were either 1) fed with *E. faecalis* or *S. marcescens* (OD_{600} = 0.1) resuspended in a sterile 10% fructose solution provided via a cotton ball for 24 hours, or 2) injected thoracically with 69 nL *E. coli*, *S. aureus* (OD_{600} = 0.4), or *S. marcescens* (OD_{600} = 0.1) with a Nanoject II. Injected mosquitoes were further maintained in secondary containment boxes lined with water-soaked paper towels for 48 hours. The number of surviving mosquitoes were counted daily until all mosquitoes had died, or until 10 days post-injection. In bacteria-injected mosquitoes, individuals that died within 24 hours post-injection were omitted from

subsequent analyses. Mosquitoes were provided with a sterile 10% fructose solution in a cotton ball throughout the experiment. A minimum of 3 biological replicates were completed for each bacterial species and gene KD.

2.10.4 Survival assay analyses

Survival assays were visualized with Kaplan–Meier survival curves using the survival (Therneau & Grambsch, 2000) and survminer (Kassambara, Kosinski & Biecek, 2020) packages in RStudio. Differences in survival curves were compared with Cox proportional hazard analyses, where the effect of gene KD and replicate were considered as covariates. A hazard ratio (HR) and its associated 95% confidence interval (CI) were calculated for control mosquitoes (*dsLacZ*-injected) versus putative AMP KD (*i.e.* treatment) mosquitoes. HR values greater than 1 indicate an increased risk associated with the control group, and vice versa. Wald test *p*-values were used to determine whether the HR significantly deviates from a HR of 1, *i.e.* no difference in survival between the two groups of mosquitoes.

2.10.5 Incubation of bacterial cultures with MF

Overnight cultures of *E. coli* and *S. marcescens* were prepared according to section 2.10.1. *E. coli* bacterial cultures were diluted to OD_{600} = 0.05 and a final volume of 5 mL into 15 mL tubes with ventilated caps. A total of 6 aliquots were made. MF was added to half of the 5 mL cultures to a final concentration of 100 µM (1:1000 dilution as described in section 2.2.4) and incubated in a 37˚C shaking incubator at 225 rpm for 8 hours. An equal volume of DMSO was added to the remaining 5 mL aliquots of cultures as controls. From each 5 mL culture, a 1 µL aliquot was taken every hour to determine the OD_{600} with a NanoDrop spectrophotometer. One biological replicate was performed for *E. coli* (includes the 3 technical replicates). For *S. marcescens*, cultures were diluted to OD_{600} = 0.1 and three biological replicates were performed. Of these replicates, one was performed with the addition of MF to a final concentration of 1 mM (1:100 dilution). Growth curves were compared using permutation tests (10,000 permutations) from the statmod package (Smyth et al, 2013) in RStudio.

2.11 Immunofluorescence assays

A transgenic line was generated by Hoermann et al, where exogenous AMPs magainin 2 from *X. laevis* and melittin from *A. mellifera* were inserted into the *GAM1* locus to be co-expressed (2022). Ribosomal skip sequences (2A) were inserted between the three AMPs to allow for three separate protein products to be translated. This transgenic line will be referred to as the 'GAM1-Mag-Mel' (GMM) line.

Whole midguts of WT (KIL/G3 strain) and transgenic GMM mosquitoes were dissected in ice-cold PBS and fixed in 4% paraformaldehyde for 45 minutes at room temperature. Midguts were washed thrice with PBS for 10 minutes, followed by a 1 hour incubation in a blocking buffer (PBS, 1% bovine serum albumin, 0.1% Triton X-100; Sigma-Aldrich, UK). Midguts were then incubated with polyclonal rabbit-anti-T2A antibodies provided by Dr. Astrid Hoermann (1:500 dilution, Discovery® Antibodies [Biosynth], UK) overnight at 4˚C, washed in blocking buffer thrice for 10 minutes, and incubated with goat anti-rabbit-IgG-Alexa Fluor™ 488 (1:1000 dilution, Thermo Fisher Scientific, UK) for 1 hour at room temperature. The tissues were then washed thrice with blocking buffer and stained with 1 µg/mL Hoechst in PBS for 10 minutes (Thermo Fisher Scientific, UK). Midguts were mounted in VECTASHIELD® Antifade Mounting Medium with DAPI (Vector Laboratories, UK), and imaged on the EVOS™ M5000 imaging system with 4x, 10x, or 20x objectives.

2.12 Whole-mount midgut RNA FISH

The methods presented in this section have been adapted from Kosman et al, 2004 and Juhn & James, 2012 to apply RNA fluorescence in situ *hybridization (FISH) to whole-mount mosquito midgut tissues for the first time.*

DNA templates for *CEC2* probe synthesis were generated from mosquito cDNA via PCR as described in section 2.9.1. The purified *CEC2* amplicon was subsequently amplified with two different primer pairs to generate template DNA for the synthesis of sense and antisense *CEC2* probes respectively: 1) T7-flanked forward primer and reverse primer (without modifications), and 2) forward primer (without modifications) and T7-flanked reverse primer (Table A.1). In order to confirm the correct amplicon, the correct length and sequence was verified via gel electrophoresis and Sanger sequencing as described in section 2.9.1. The T7-flanked amplicons were then used as templates for RNA probe synthesis using the FISH Tag™ RNA Red Kit according to the manufacturer's instructions (Thermo Fisher Scientific, UK), or the TranscriptAid T7 High Yield Transcription Kit. Using the FISH Tag™ RNA Red Kit, the RNA synthesis and ethanol precipitation steps were increased to overnight incubations. RNA probes were subsequently stained with an amine-reactive dye supplied in the FISH Tag™ RNA Red Kit according to the manufacturer's instructions. RNA probes for AGAP009313 and AGAP013543 were also generated using the same methodology.

Approximately 7 mosquito midguts were dissected from female adults aged 2–6 days post-emergence (N'gousso strain) into ice-cold PBS. Excess PBS was removed and midguts were fixed in 4% paraformaldehyde for 45 minutes. Midguts were washed with PBT (1X PBS, 0.1% Tween® 20 [Sigma-Aldrich, UK]) for 4 × 5 minutes and equilibrated into 1:1 PBT/hybridization buffer (50% formamide, 5X saline-sodium citrate buffer 100 µg/mL salmon testes DNA, 50 µg/mL heparin, 0.1% Tween® 20 [Sigma-Aldrich, UK]) for 30 minutes with gentle rocking. Subsequently, samples were pre-hybridized in hybridization buffer for 30 minutes at 55˚C with rotation. RNA probes were added to ~50 µL hybridization buffer to a final concentration of 2–5 ng/µL and were hybridized to mRNA in the midgut samples overnight at 55˚C.

The hybridization buffer containing RNA probes was removed and samples were incubated twice: first with hybridization buffer (without the probe) for 5 minutes at 55˚C

with rocking, followed by a 30 minute incubation with fresh hybridization buffer at 55˚C with rotation. Samples were equilibrated into 1:1 PBT/hybridization buffer for 10 minutes with gentle rocking, followed PBT washes for by 4 × 5 minutes with gentle rocking. The midguts were then subject to an RNAse treatment (PBT, 20 µg/mL RNAse A; Qiagen, Germany) for 30 minutes at 37˚C, followed by PBT washes for 4 × 5 minutes with rocking. In the final PBT wash, 1 mg/mL of Hoechst was added to PBT at a final concentration of 1 µg/mL (1:1000 dilution). For some replicates, a stock of 66 µM Alexa Fluor™ 488 phalloidin (Thermo Fisher Scientific, UK) suspended in DMSO was also added to the final PBT wash in a 1:400 dilution. Midguts were mounted in VECTASHIELD® Antifade Mounting Medium with DAPI and cover slips were sealed with clear nail polish. Slides were imaged on an EVOS™ M5000 imaging system with 10x or 20x objectives, or a Leica SP8 inverted confocal microscope with the HC PL APO 40x/1.3 oil CS2 objective at the Facility for Imaging by Light Microscopy (FILM) at Imperial College London, with assistance from Matteo Vitale for some images. Training for image acquisition was provided by Dr. David Gaboriau of FILM. Captured images were analyzed and processed with ImageJ (Schneider, Rasband & Eliceiri, 2012) and/or Icy (de Chaumont et al, 2012).

Chapter 3: Results I
Mosquito Transcriptomes

3.1 Introduction

3.1.1 Abstract

Malaria remains a global health issue and the lack of significant progress in reducing its impact in recent decades highlights the need for novel intervention strategies. The malaria parasite experiences a large population bottleneck in the midgut due to local immune responses. Hence, further characterization of mosquito responses to parasite infection can inform novel malaria intervention strategies. The midgut is also highly compartmentalized, with anterior regions more immunocompetent and posterior regions specialized for blood meal digestion, however, this compartmentalization has mostly been examined under the context of sugar feeding. In order to fill this gap of knowledge, RNA-seq was performed on parasite-infected mosquito tissues at three time points, as well as on midgut regions at different time points post-blood meal. Here, few genes were modulated in response to parasite infection, which could be a result of varying infection intensities across the replicates. Following intake of a blood meal, the midgut is transcriptionally reprogrammed to prioritize blood meal digestion. The transcriptomic data confirmed the functional specialization of the different midgut regions and that these functions persist before and after blood feeding. Additionally, the cardia is extremely immunocompetent and constitutively secretes high titers of AMPs. These RNA-seq datasets can be used as a resource for the generation of malaria-refractory mosquitoes, as effectors can be introduced into loci which enable targeting of the parasite in a controlled spatiotemporal manner.

3.1.2 Background

The midgut is highly compartmentalized, with the anterior regions highly immunocompetent, and the posterior regions more immunotolerant to enable the persistence of symbiotic microbiota. Following a blood meal, the influx of nutrients from the blood meal allow the microbiota population to expand drastically (Demaio et al, 1996; Dong et al, 2009; Rodgers et al, 2020). A PM is formed in the posterior midgut following blood feeding, encapsulating the blood bolus and physically separating it from the midgut epithelial cells. The microbiota proliferate within this sac, which prevents the continuous activation of epithelial immune responses.

The mosquito midgut is an important tissue to study in the context of malaria as it marks the first site of interaction between the parasite and vector. While the PM protects the midgut epithelia from microbial dissemination post-blood meal, it also enables the sequestration of *Plasmodium*, which are able to develop in the posterior midgut undetected (Kumar et al, 2010). At the same time, the PM represents another barrier the parasite must cross before traversing the midgut epithelium, which they overcome by secreting chitinases (Shahabuddin et al, 1993; Tsai et al, 2001). Despite protection by the PM, the parasite experiences a large population bottleneck in the midgut, specifically between the ookinete to oocyst transition, which can be attributed to both humoral and cellular immune responses (Sinden et al, 2010; Cirimotich et al, 2010). For example, AMPs are highly and constitutively expressed in the cardia (Rodgers et al, 2020; Hixon et al, 2022), and a variety of potent nitration responses are launched following invasion of epithelial cells by ookinetes (Garver, de Almeida Oliveira, & Barillas-Mury, 2013).

To date, detailed investigation of the compartmentalization of the midgut has mostly been restricted to sugar-fed conditions (Warr et al, 2007; Hixson et al, 2022), whereas RNA-seq of mosquito responses to different stages of the human malaria parasite has not been performed before. In order to fill this knowledge gap, RNA-seq was performed to elucidate the responses of mosquito tissues to parasite infection, as well as different regions of the midgut to blood feeding. Characterization of these transcriptomes can lead to the identification of parasite-responsive effectors and the spatiotemporal expression of all genes in the midgut, which may inform the generation of future malaria-refractory transgenic mosquitoes.

3.1.3 Methods

RNA-seq was employed here to generate two datasets: 1) the *Ag/Pf* data and 2) the midgut sections data. An overview of the samples collected are presented in Figure 2.3.1. For the *Ag/Pf* data, midguts and carcasses (decapitated, whole mosquitoes) were dissected at three time points post-*P. falciparum* infection or post-blood meal (controls) to enable the characterization of mosquito responses to gametocytes (1 hour), ookinetes (24 hours), and oocysts (7 days). Samples were collected according to section 2.3.3, and RNA was extracted from samples as detailed in section 2.4. Additional midguts were

dissected and stained at 7 days post-infection to assess oocyst infection intensity as an indicator of infection success.

For the midgut sections data, the cardia, anterior midgut, and posterior midgut regions were sampled at three time points pre- and post-blood meal: 0 hours (*i.e.* sugar-fed), to enable characterization of the transcriptomes just before blood feeding; 5 hours, to assess early responses to blood feeding/blood meal digestion; and 24 hours, to investigate late responses as the blood bolus is noticeably reduced in size at this time point. The detailed methodology is described in sections 2.3.1 and 2.4.

RNA-sequencing on isolated RNA was performed with GeneWiz Ltd. services, and the raw data files were processed on the Galaxy webserver (Afgan et al, 2016) to obtain normalized counts as detailed in section 2.6.1. Significant differences in gene expression between all pairwise comparisons of conditions were calculated with DESeq2 (Love, Huber & Anders, 2014). To enable the characterization of investment towards the expression of specific genes in the midgut sections data, TPMs were calculated with Stringtie (Pertea et al, 2015).

Hierarchical clustering was performed on the midgut sections data to identify clusters of genes with similar expression profiles. A variety of validation metrics were used to confirm that an appropriate number of clusters were selected (section 2.6.2). In order to associate clusters and/or gene lists (*e.g.* significantly DE genes between two conditions) with biological mechanisms, GO term enrichment analyses were performed using g:Profiler (Raudvere et al, 2019).

3.2 RNA-sequencing data of infected midguts and whole carcasses

Comparing mosquito immune responses of *Plasmodium*-infected and blood-fed mosquitoes can lead to the identification of *Plasmodium*-responsive/specific effectors. In order to investigate the transcriptional responses to *P. falciparum*, RNA-sequencing was performed on *P. falciparum*-infected mosquitoes and uninfected controls. Pools of ~30 midguts and carcasses (decapitated whole mosquitoes) were sampled at 1 hour, 24 hours, and 7 days post-infection/blood feeding to assess responses to gametocytes, invading ookinetes, and oocysts respectively. The intensity and prevalence of the infections was performed at 7 days post-blood feeding by dissecting additional midguts and counting oocysts (Table 3.2.1). Three biological replicates were completed for midgut samples, with MRes student Yunyue (Zoe) Zhou (project supervised by Dr. Dina Vlachou) contributing towards the sample collection of the final replicate, and two replicates were performed for carcass samples. This data will be referred to as the '*Ag/Pf* data'.

Table 3.2.1*. Ag/Pf *data infection statistics. *Intensity and prevalence of* P. falciparum *infection in* An. gambiae *across the three replicates submitted for RNA-sequencing. Midguts were dissected at 7 days post-infection, where oocysts were stained and counted. MRes student Yunyue (Zoe) Zhou (project supervised by Dr. Dina Vlachou) contributed to the sample collection of replicate 3.*

		Oocysts			
Replicate	# of Midguts	Maximum	Mean	Median	Prevalence (%)
1	27	51	12.89	10	74
2	41	32	2.14	0	37
3	40	98	7.25	2	62.5

A total of 30 samples were sequenced, which yielded 1,734,746,113 pairs of reads. Following filtering and removal of low-quality reads, a total of 1,268,590,314 reads corresponding to 73.1% of all reads aligned uniquely to one region of the *An. gambiae* PEST4 genome. The aligned data was converted into counts, and 1,238,855,247 reads (97.7% of aligned reads) corresponded to annotated transcripts in the genome. Normalization and differential expression analysis of the count data was performed with DESeq2 (Love, Huber & Anders, 2014). Principal component analysis (PCA) revealed that samples separated primarily by time point (PC1) and by tissue (PC2) into four groups (Fig. B.1). Infection status did not appear to influence clustering. An infected midgut

sample at 7 days post-infection from replicate 3 (sample IM7C) was an outlier and was removed from subsequent analyses as it did not cluster with any midgut samples (Fig. 3.2.1; Fig. B.1). Upon removal of this sample, the clustering of the other samples remained nearly identical (Fig. 3.2.1).

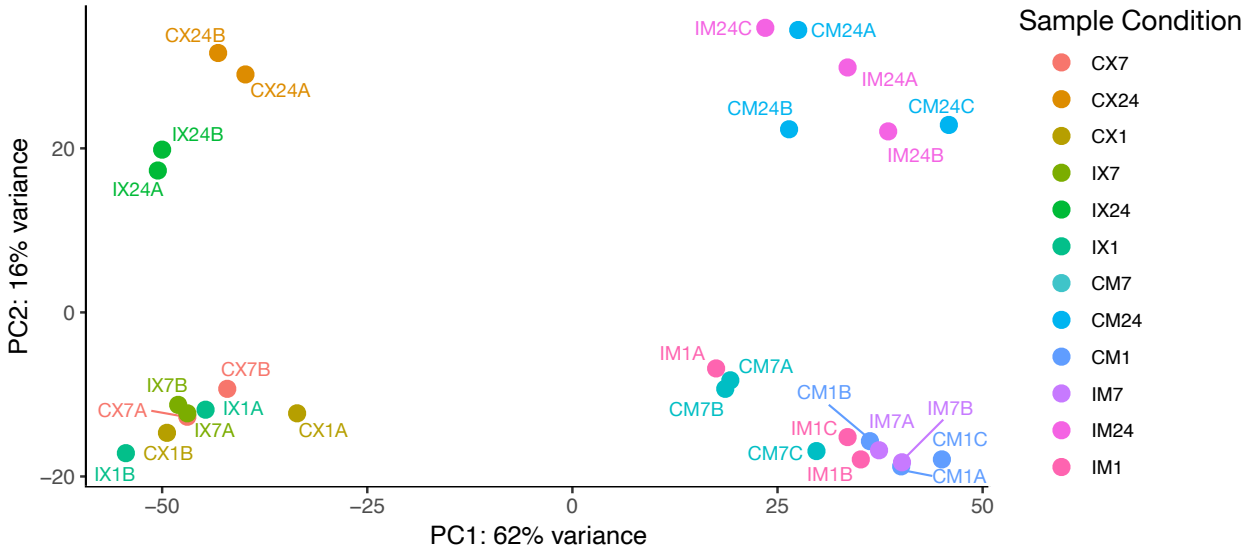


Figure 3.2.1. PCA of the* Ag/Pf *RNA-seq samples. *PCA was performed following the normalization of read counts with DESeq2 (Love, Huber & Anders, 2014). Each dot represents a transcriptome sequenced from a pool of midguts or carcasses. Sample IM7C was removed as it did not cluster with other replicates of the same condition. PC = principal component; I =* P. falciparum*-infected, C = control (blood-fed); M = midgut, X = carcass; 1 = 1 hour post-infection/blood feed, 24 = 24 hours post-infection/blood feed, 7 = 7 days post-infection/blood feed; A/B/C represent replicates 1, 2, and 3 respectively. Three replicates were performed for midgut samples, two replicates were performed for carcass samples.*

3.3 Responses to *Plasmodium* are dissimilar between the midgut and carcass

The significantly DE transcripts between pairs of conditions were identified with DESeq2. In order to have an overview of the transcriptional responses to *P. falciparum* infection, up- and downregulated transcripts in infected and control comparisons of the same time point (*e.g.* infected midguts at 1 hour versus control midguts at 1 hour) were extrapolated and GO term enrichment analysis was performed with g:Profiler (Raudvere et al, 2019). Very few transcripts were significantly DE between infected and control samples (420 in the midgut, 39 in the carcass, representing ~3% of the transcriptome), and no genes were significantly upregulated in any of the infected carcass samples compared to controls. Just 2 and 37 transcripts were downregulated at 1 and 24 hours post-infection in the carcass respectively. This is in concordance with the PCA clustering, where samples did not separate by infection status (Fig. 3.1.1).

Within midgut samples, no overlaps were observed between the upregulated transcripts at each time point (Fig. 3.3.1). Similarly, just 3 downregulated transcripts were shared across the time points, representing ~2.5% of the total number of significantly downregulated transcripts (Fig. 3.3.1). This suggests that different but limited responses are employed to deal with different stages of *Plasmodium*.

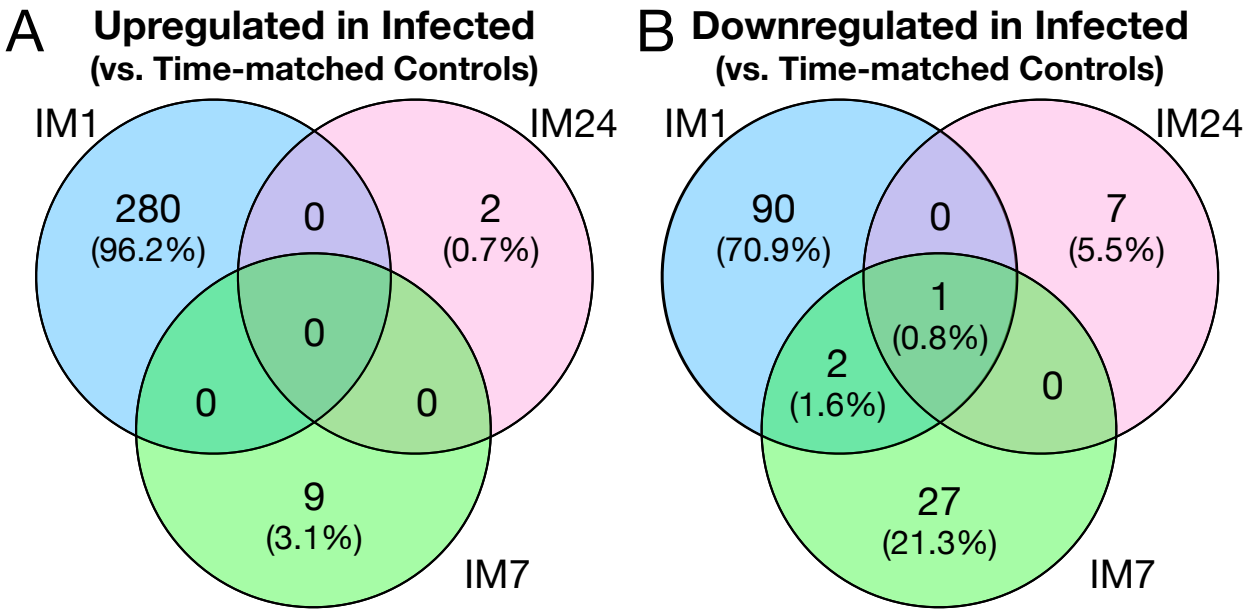


Figure 3.3.1. DE genes in infected and control samples. *Venn diagrams were constructed from the unique and shared significantly (Wald tests, false discovery rate adjusted-*p *< 0.05) A) up- and B) downregulated transcripts in infected midgut samples versus their time-matched controls across three time points post-*P. falciparum *infection or blood meal. Lists of shared and unique DE genes were generated using Venny 2.1 (Oliveros, 2015). IM = infected midgut samples; 1 = 1 hour post-infection, 24 = 24 hours post-infection, 7 = 7 days post-infection.*

Enriched GO terms were identified from 3 infected versus control comparisons: up- and downregulated transcripts at 1 hour in infected midguts, and upregulated transcripts in infected midguts at 7 days post-infection compared to controls of the same time point and tissue. At 1 hour, 17 GO terms were enriched from the upregulated transcripts in the infected midgut versus controls including, 'oogenesis' (GO:0048477) and 'aromatic amino acid metabolic process' (GO:0009072). Of the downregulated transcripts at 1 hour in infected midguts, 9 GO terms were enriched including 'endoplasmic reticulum' (GO:00058783). At 7 days, 15 terms were overrepresented from the upregulated genes, many of which relating to the binding of misfolded/unfolded proteins. The significantly enriched GO terms of each group of DE transcripts are listed in Table B.1.

A previous study reported similar findings at 24 hours post-*P. falciparum* infection in the midgut, where 265 and 62 genes were up- and downregulated in response to ookinete invasion respectively (Dong et al, 2006). The number of DE genes (*i.e.* percentage of the transcriptome) are comparable to those identified in this study, where 282 and 90 transcripts were up- and downregulated in the response to ookinete traversal respectively. Similar findings were also present for carcass samples, where fewer genes were significantly DE compared to the midgut (Dong et al, 2006).

Next, immunity genes which were significantly DE between the infected and control conditions were identified. A total of 12 immunity-related genes were identified, 9 of which were upregulated in infected samples (Table 3.3.1). In the previous study by Dong et al, dozens of immune effectors were DE in response to the presence of the parasite (2006).

Table 3.3.1. Significantly DE transcripts between infected and control samples. *Differentially expressed genes were identified from comparisons between the same tissue type and time point with Wald statistical tests.*

					Wald test	
Transcript ID	Name	lfc	Comparison (Avg. Counts)		*W* Statistic	Adjusted *p*-value
AGAP013506-RA	UPD3A	-6.54	IM24 (15)	UM24 (3965)	1.18	4.39×10^{-4}
AGAP001375-RA	SRPN12	-2.35	IX24 (26)	UX24 (146)	0.58	2.20×10^{-2}
AGAP011950-RA	CASPS2	-2.17	IM1 (19)	CM (93)	0.65	3.29×10^{-2}
AGAP003194-RA	SRPN8	1.24	IM1 (88)	CM1 (36)	0.38	4.20×10^{-2}
AGAP000573-RA	CLIPC4	1.34	IM1 (314)	CM1 (120)	0.42	4.81×10^{-2}
AGAP009221-RA	SRPN5	1.78	IM1 (281)	CM1 (77)	0.52	2.46×10^{-2}
AGAP001212-RA	PGRP-LB	1.81	IM1 (71,771)	CM1 (19,771)	0.48	1.07×10^{-2}
AGAP029769-RA	CLIPB9	2.37	IM1 (68)	UM1 (11)	0.74	4.72×10^{-2}
AGAP005717-RA	LYSC6	2.95	IM1 (276)	UM1 (31)	0.31	2.95×10^{-3}
AGAP004456-RA	GNBP-B5	3.27	IM1 (62)	UM1 (20)	1.02	4.50×10^{-2}
AGAP004036-RA	HPX7	4.78	IM1 (131)	UM1 (2)	0.93	9.05×10^{-5}
AGAP004038-RA	HPX8	5.51	IM1 (1,767)	UM1 (19)	0.90	1.42×10^{-6}

lfc = $\log_2$ fold change comparing counts of a gene between an infected and control sample of the same tissue and time point, negative values refer to an upregulation in the control sample (or a downregulation in the infected), positive values refer to an upregulation in the infected sample (or a downregulation in control); avg. counts = average counts of a transcript in a sample condition; adjusted p*-value = false discovery rate-corrected* p*-values. I = infected, C = control, M = midgut, X = carcass, 1 = 1 hour post-infection/blood feed, 24 = 24 hours post-infection/blood feed.*

Here, expression of two HPXs were upregulated in response to early mosquito stages of *P. falciparum*. Interestingly, these HPXs were also previously observed to be upregulated in response to *P. berghei* ookinetes (de Almeida Oliveira, Lieberman & Barillas-Mury, 2012; Garver, de Almeida Oliveira & Barillas-Mury, 2013). Three serine protease inhibitors (serpins [SRPNs]) were also significantly DE between infected and control conditions. As most immune signaling cascades involve serine proteases, at least 10 SRPNs can inhibit numerous immune mechanisms such as POs involved in the melanization pathway (Christophides et al, 2002; Michel et al, 2005; Michel et al, 2006). SRPNs 5 and 8 are upregulated in the parasite presence, suggesting that the parasite modulates their expression to suppress immune stimulation, whereas SRPN12 expression could be downregulated by the host to boost immune activity. Despite this, expression of SRPNs 8 and 12 is primarily found in larval stages (Suwanchaichinda & Kanost, 2009), and the average counts of these SRPNs in midgut samples were relatively low. This suggests that although significant, changes to their expression are overall very small. PGRP-LB, a negative regulator of the IMD pathway in the midgut, was significantly upregulated in the midgut at 1 hour post-infection compared to controls. While PGRP-LB is upregulated following a blood meal (Gendrin et al, 2017), its further upregulation in the presence of the parasite may suggest tolerance mechanisms in place to enable parasite development.

No AMPs were significantly DE between infected and control samples of the same infection status and time point. Expression of the known AMPs CEC1, DEF1, and GAM1 were further examined within the same sample condition but across the different time points (Fig. 3.3.2). In all cases, differences in average counts were insignificant with the exception of CEC1 and DEF1 in control carcass samples, which were significantly enriched at 1 hour compared to 7 days post-blood feed (Wald tests; $W = 0.64$, false discovery rate-adjusted $p = 0.01$ and $W = 0.55$, false discovery rate-adjusted $p = 0.02$ respectively). This suggests that AMP expression is largely unchanged by blood feeding and *P. falciparum* infection.

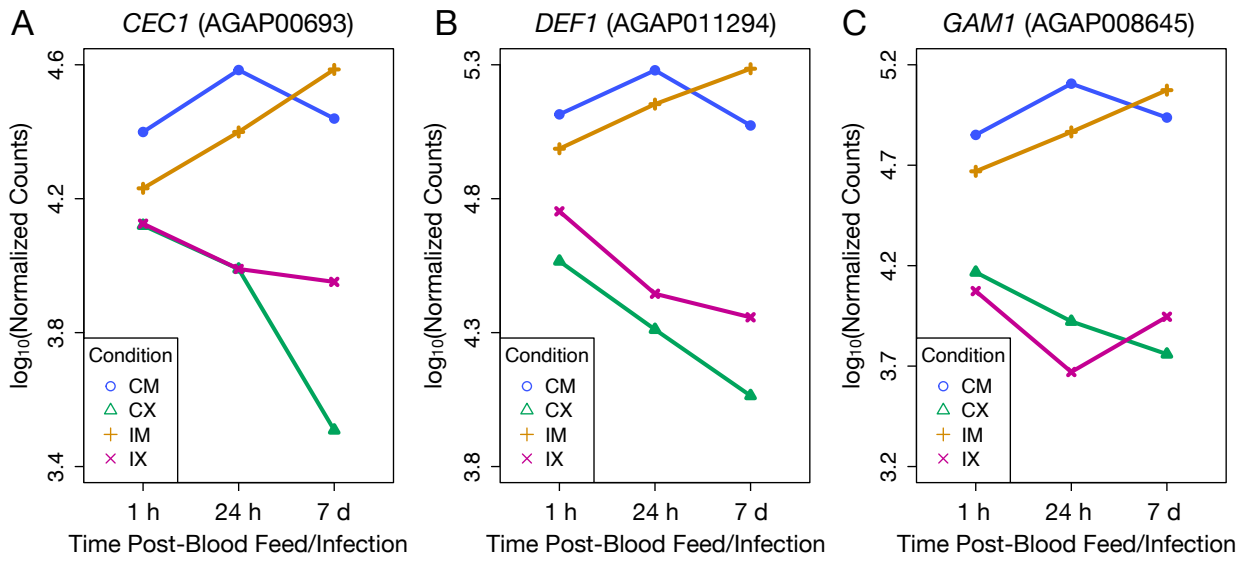


Figure 3.3.2. Average normalized counts of known AMPs in the* Ag/Pf *data. *Mosquito tissues were sampled at three time points in midguts and carcasses post-blood feed (controls) or post-*P. falciparum *infection (infected). CM = control midguts, CX = control carcasses, IM = infected midguts, IX = infected carcasses. Three replicates were performed for midgut samples, two replicates were performed for carcass samples.*

Given the limited number of significantly DE genes between infected and control conditions, additional approaches were taken to further characterize transcriptional responses to *P. falciparum* infection. First, overlaps in the DE transcripts at each time point of the four conditions were compared (Fig. 3.3.3). At 24 hours post-infection, 3,638 and 4,903 transcripts were up- and downregulated respectively compared to at 1 hour in all four sample conditions (infected/control and carcass/midgut). In concordance with the PCA, responses in the different tissues appear to be distinct. Among the upregulated transcripts, 54.4% and 25.9% were unique to midgut and carcass samples respectively, and of the downregulated transcripts, 38.6% and 44.3% were unique to the midgut and carcass samples respectively (Fig. 3.3.3). At 7 days post-infection, the majority of transcripts were unique to either the midgut or carcass samples, suggesting dissimilar responses in the tissues following blood feeding and/or *Plasmodium* infection (Fig. 3.3.3).

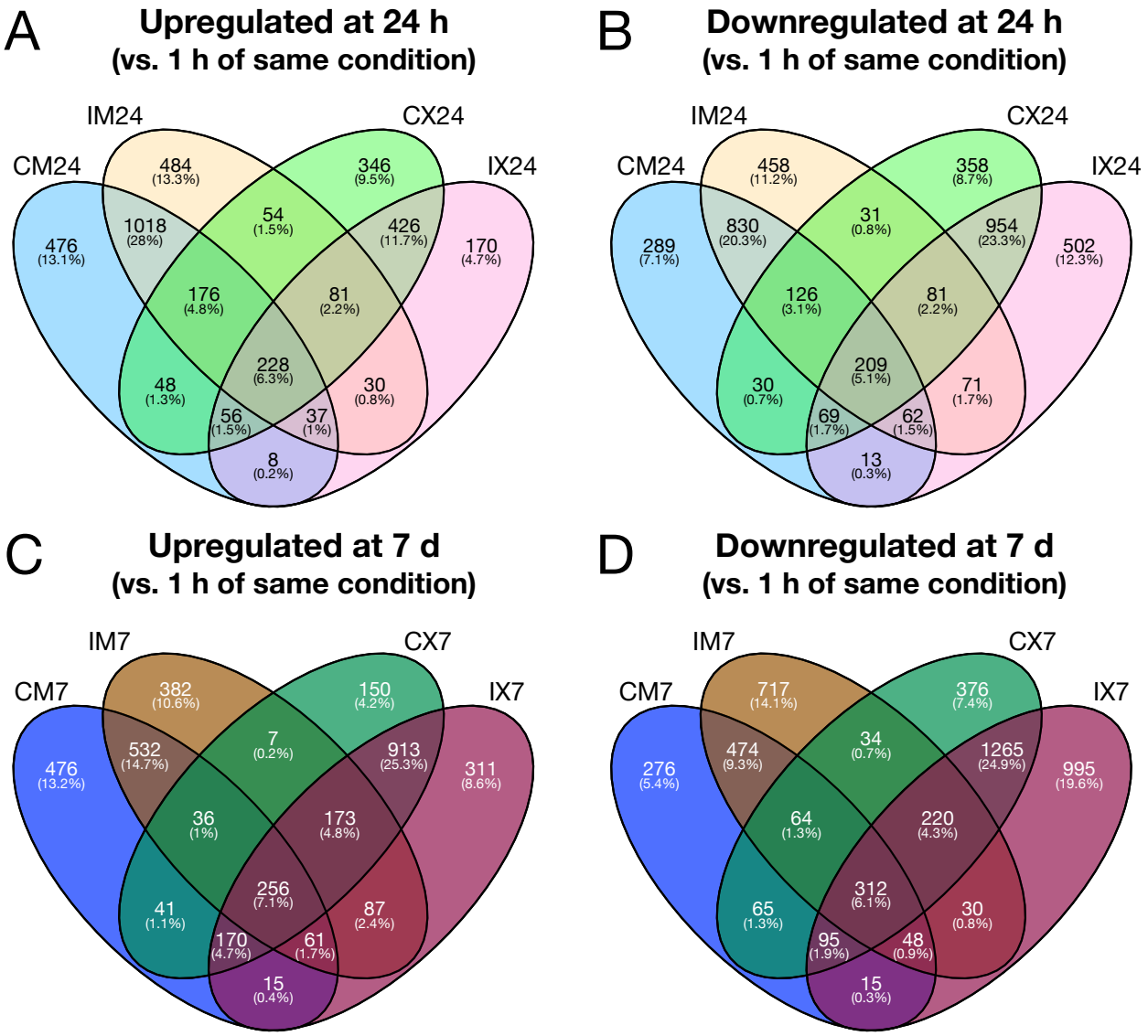


Figure 3.3.3. Venn diagrams of significantly DE transcripts at each time point*. Unique and shared DE genes significantly (Wald tests, false discovery rate adjusted-*p *< 0.05) A) up- and B) downregulated transcripts at 24 hour-post-infection/blood feed versus at 1 hour; and C) up- and D) downregulated transcripts at 7 days post-infection/blood feed versus at 1 hour. Lists of shared and unique DE genes were generated using Venny 2.1 (Oliveros, 2015). I = infected, C = control; M = midgut, X = carcass; 24 = 24 hours post-*P. falciparum *infection/blood feed, 7 = 7 days post-infection/blood feed.*

In order to assess responses to parasite invasion of the midgut epithelium, DE genes at 24 hours compared to 1 hour were extrapolated from infected and control midguts separately (*i.e.* DE genes between 24 hours and 1 hour post-infection, and DE genes between 24 hours and 1 hour post-blood feeding). These transcripts were then used to construct a Venn diagram to compare and contrast the DE transcripts from the infected and control midguts, and GO term enrichment analysis was performed on the genes that were shared between and unique to each condition (Fig. 3.3.4). The majority of DE genes were shared between infected and control midguts (Fig. 3.3.4A–B). Of the shared upregulated transcripts at 24 hours compared to 1 hour, 34 GO terms were significantly enriched, including 'peroxisome' (GO:0005777), 'catalytic activity' (GO:0003824), which includes over 450 different transcripts, and 'lipid metabolic process' (GO:0006629). These terms reflect blood meal digestion-related processes which are necessary in both conditions. Many of the shared downregulated 222 GO terms were related to transcription and translation, *e.g.* 'RNA binding' (GO:0003723) and 'translation factor activity, RNA binding' (GO:0008135).

At 24 hours post-blood feed in control midguts, many of the 25 GO terms representative of the unique upregulated transcripts relate to reproduction, whereas of the 19 enriched GO terms from the downregulated transcripts, actin and ATPase activity-related terms were common (Table B.2). In infected midguts, 16 and 32 GO terms were enriched in the up- and downregulated terms respectively. Chromatin and nuclear GO terms in particular were overrepresented from the upregulated transcripts, whereas ribosomal components and fatty acid elongation GO terms were enriched from the downregulated transcripts (Table B.2). The full list of enriched GO terms are listed in Table B.2.

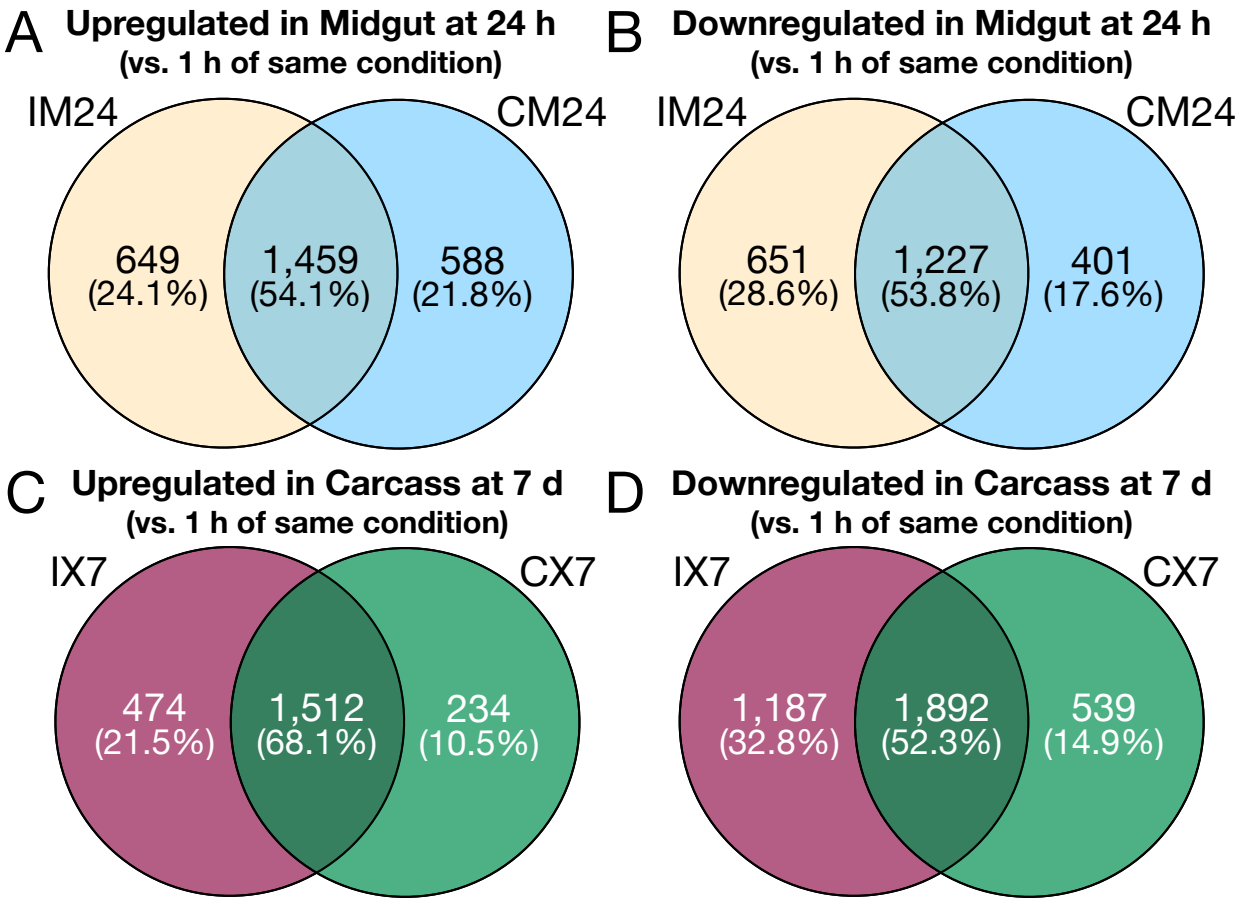


Figure 3.3.4. Venn diagrams of DE transcripts in midgut and carcass samples. *Unique and shared significantly (Wald tests, false discovery rate adjusted-*p *< 0.05) A) up- and B) downregulated transcripts at 24 hour-post-infection/blood feed in the midgut versus at 1 hour; and C) up- and D) downregulated transcripts at 7 days post-infection/blood feed in the carcass versus at 1 hour. Lists of shared and unique DE genes were generated using Venny 2.1 (Oliveros, 2015). I = infected, C = control; M = midgut, X = carcass; 24 = 24 hours post-*P. falciparum *infection/blood feed, 7 = 7 days post-infection/blood feed.*

The above analysis was repeated in carcass samples at 7 days post-infection (compared to 1 hour) to understand systemic responses to late-stage oocysts. Similar to the midgut at 24 hours post-infection/blood feed, the majority of genes were shared between infected and control carcass samples (Fig. 3.3.4C–D). The shared up- and downregulated genes contributed to the enrichment of 288 and 188 GO terms respectively. Nuclear and nucleic acid binding terms were among the 60 enriched terms from the upregulated genes unique to the infected samples (Table B.3). From the unique downregulated transcripts at 7 days in the infected carcass, the 76 enriched GO terms included transmembrane and receptor signaling terms (Table B.3). In the transcripts unique to the control samples, 40 GO terms were enriched in the downregulated transcripts, including translation, ribosomal, and prostaglandin synthesis terms. The upregulated transcripts in the control carcasses yielded 40 enriched terms, specifically binding- and signaling activity-related terms were overrepresented in the upregulated transcripts.

GO term enrichment analysis was employed here to identify differences in functions between the various conditions. While these analyses can provide broad-spectrum and valuable insights about the functions of tissues/conditions, incomplete or absent gene annotations can drastically impact the enriched terms within a list of genes/transcripts. For example, in the *Ag/Pf* data, multiple reproduction-related terms (*e.g.* 'female gamete generation' [GO:0007292]) were enriched in the transcripts upregulated at 1 hour post-*P. falciparum* infection compared to 1 hour post-blood feeding (controls), but further investigation found that same four genes triggered the enrichment of these terms: two unspecified products predicted to have Bicoid (Bcd) functions (AGAP007364 and AGAP007365; Amos et al, 2022), and HPX 7 and 8 (AGAP004036 and AGAP004038 respectively). Bcd is involved anterior/posterior axis patterning in *Drosophila* embryos, but *bcd* is transcribed during oogenesis (Ali-Murthy & Kornberg, 2016). While HPXs have been implicated in the eggshell hardening process (Konstandi et al, 2005; Amenya et al, 2010), HPX7–8 also catalyze prostaglandin biosynthesis in the midgut which enables an enhanced hemocyte response to *P. berghei* infection (Barletta et al, 2019). Moreover, *Anopheles* HPXs are upregulated in response to invading *P. berghei* ookinetes (de Almeida Oliveira, Lieberman & Barillas-Mury, 2012; Garver, de Almeida Oliveira & Barillas-Mury, 2013), control midgut microbiota proliferation (Kakani, Gupta & Kumar, 2020), and are induced by blood feeding (Kumar et al, 2010). Here, while it can be concluded that HPX7–8 are also induced in response to *P. falciparum* and at an earlier time point than *P. berghei*, results from GO term enrichment analyses should be further interpreted. The functions of some genes which caused the enrichment of the reproduction-related terms appear to lie on the periphery of each term, implying that the enrichment of this term may be misleading. Given that HPXs are enriched in infected samples compared to control samples, it could be more indicative of anti-*Plasmodium* responses rather than increased reproductive investment following infection. In contrast, reproductive terms were also enriched from genes uniquely upregulated in control midguts at 24 hours post-blood feed compared to at 1 hour. While the same four genes contributed to the enrichment of these terms, additional, more canonical reproductive genes are present, including AGAP008847 which encodes a γ-tubulin complex component and is implicated in the formation of meiotic spindle fibers; AGAP001522,

predicted to be involved in meiotic recombination checkpoint signaling; AGAP002829, encoding ATP-dependent RNA helicase spindle-E which is involved in the meiotic cell cycle; and AGAP007280, which encodes a peptidase S1 domain-containing protein predicted to be involved in eggshell chorion assembly and oocyte dorsal–ventral axis patterning.

3.4 RNA-sequencing data of mosquito midgut tissues

The mosquito midgut is functionally compartmentalized across its anterior-posterior axis, with each tissue having distinct transcriptomes (Warr et al, 2007; Hixson et al, 2022). Detailed characterization of these transcriptomes has previously been performed in the context of sugar-fed midguts, but not in response to blood feeding (Warr et al, 2007; Hixson et al, 2022). Dissecting the midgut tissue transcriptomes in response to blood feeding can enable the analysis of the spatiotemporal expression of genes in the context of blood-borne infections. As the parasite undergoes a major bottleneck in the midgut, this could also facilitate the identification of effectors and loci which may be beneficial for use in the generation of a transgenic, malaria-refractory mosquito. In order to fill this gap of knowledge, pools of three *An. gambiae* midgut tissues (the cardia, anterior midgut, and posterior midgut) were sampled at three different time points: 0 hours (*i.e.* sugar-fed), 5 hours, and 24 hours post-blood feeding and submitted for RNA-sequencing.

In total, three biological replicates were completed for the RNA-sequencing of midgut tissues, resulting in 1,222,124,235 pairs of reads across the 27 samples. After quality control steps, reads were aligned to the *An. gambiae* PEST4 genome where 809,923,718 reads (66.3% of all reads) aligned uniquely. Conversion of the aligned data into count data led to 797,451,232 reads (98.5% of aligned reads) corresponding to annotated transcripts. The count data was then normalized and analyzed with DESeq2 (Love, Huber & Anders, 2014). PCA revealed that samples clustered by tissue and time point (Fig. 3.4.1). No obvious outliers were detected. This dataset will be referred to as the ‘midgut sections data’.

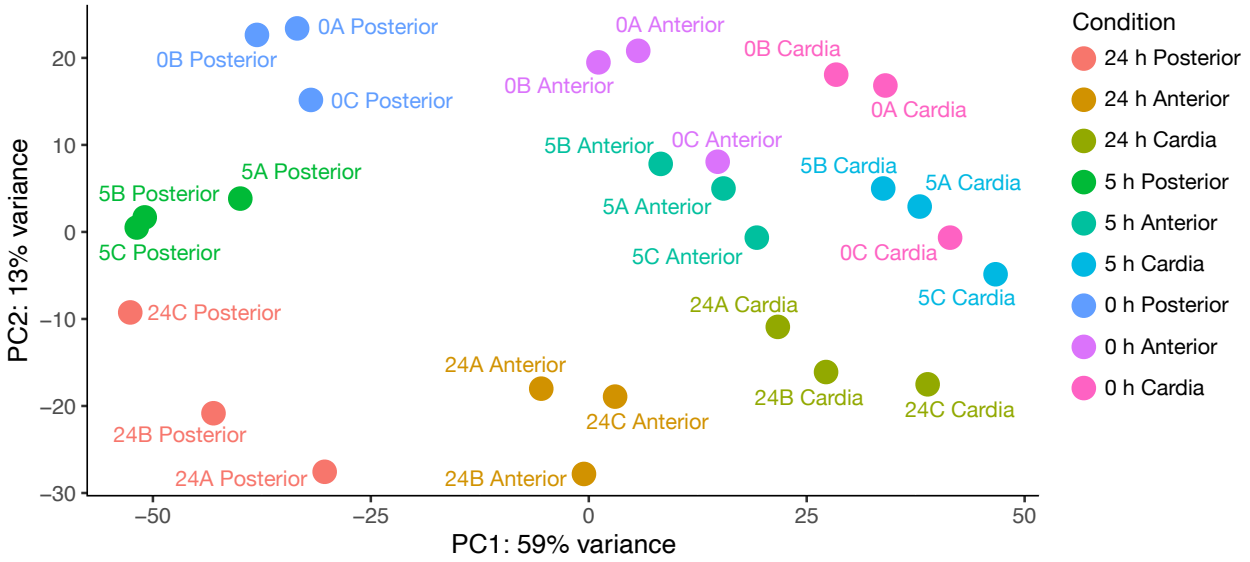


Figure 3.4.1. PCA of the midgut sections RNA-seq samples. *Read counts of each transcript in each condition were normalized with DESeq2 (Love, Huber & Anders, 2014). Each dot represents a transcriptome sequenced from a pool of midgut tissues from approximately 50 individuals. A/B/C represent replicates 1, 2, and 3 respectively. Three replicates were performed.*

3.5 Mosquito midgut tissues have distinct transcriptomes which are impacted by blood feeding

To gain an overview of mosquito midgut tissue responses to blood feeding, hierarchical clustering was performed on the midgut sections data. Counts of each transcript were summed to obtain counts of the entire gene, and genes with average counts below 50 were excluded. These average counts were scaled so that the median expression value of each gene was 0 and the standard deviation was 1. Using the scaled average counts of each tissue and time point, expression profiles across all time points and midguts tissues of 8,436 genes were hierarchically clustered using the tidyverse (Wickham et al, 2019) and cluster packages in RStudio (Maechler et al, 2021).

Clustering of the samples and genes showed that each tissue had a distinct transcriptome signature, where the majority of genes were enriched in the cardia (Fig. 3.5.1). In the anterior and posterior midgut, expression of genes was relatively lower than in the cardia except at 24 hours post-blood feeding, where expression of genes was upregulated in these tissues (Fig. 3.5.1).

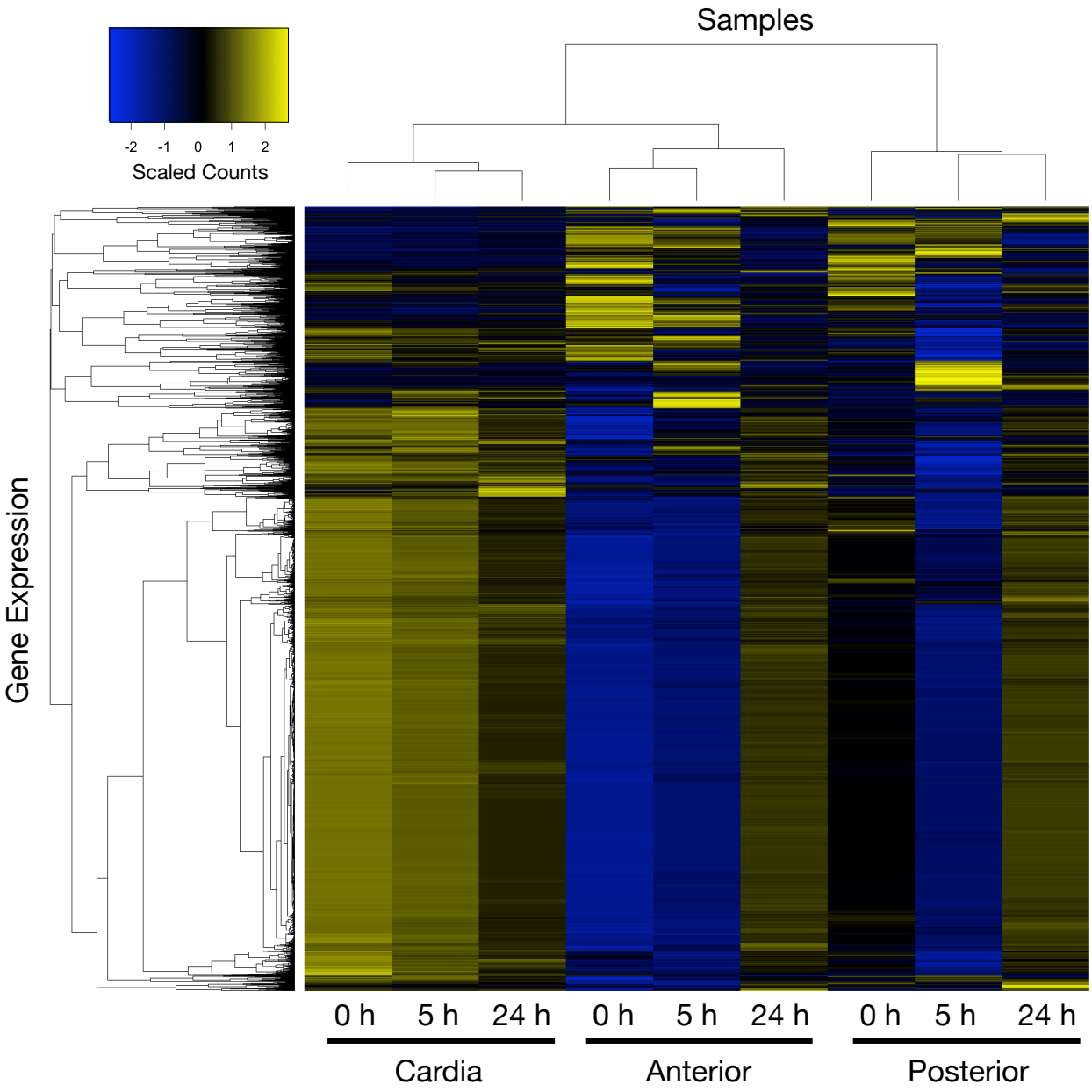


Figure 3.5.1. Hierarchical clustering of midgut sections data. *The heatmap depicts expression profiles of each gene in each sample, where yellow represents increased expression and blue represents decreased. Average normalized counts of each gene from three replicates were scaled prior to hierarchical clustering so that mean expression values were 0 and standard deviations were 1. Time points: 0 h = prior to blood feeding (*i.e. *sugar-fed), 5 h = 5 hours post-blood feeding, 24 h = 24 hours post-blood feeding.*

The expression profiles of the genes were arbitrarily separated into 11 clusters (Fig. 3.5.2). The average expression profiles for each cluster are shown in Figure 3.5.3. Significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) terms of each cluster were subsequently identified with g:Profiler (Raudvere et al, 2019). Enriched KEGG terms were identified in 10 of the 11 clusters (Table 3.5.1). These clusters can be loosely compiled into three groups: 1) enriched in the cardia/anterior regions of the midgut, 2) blood meal-responsive (enriched in the posterior regions of the midgut), and 3) sugar meal-responsive.

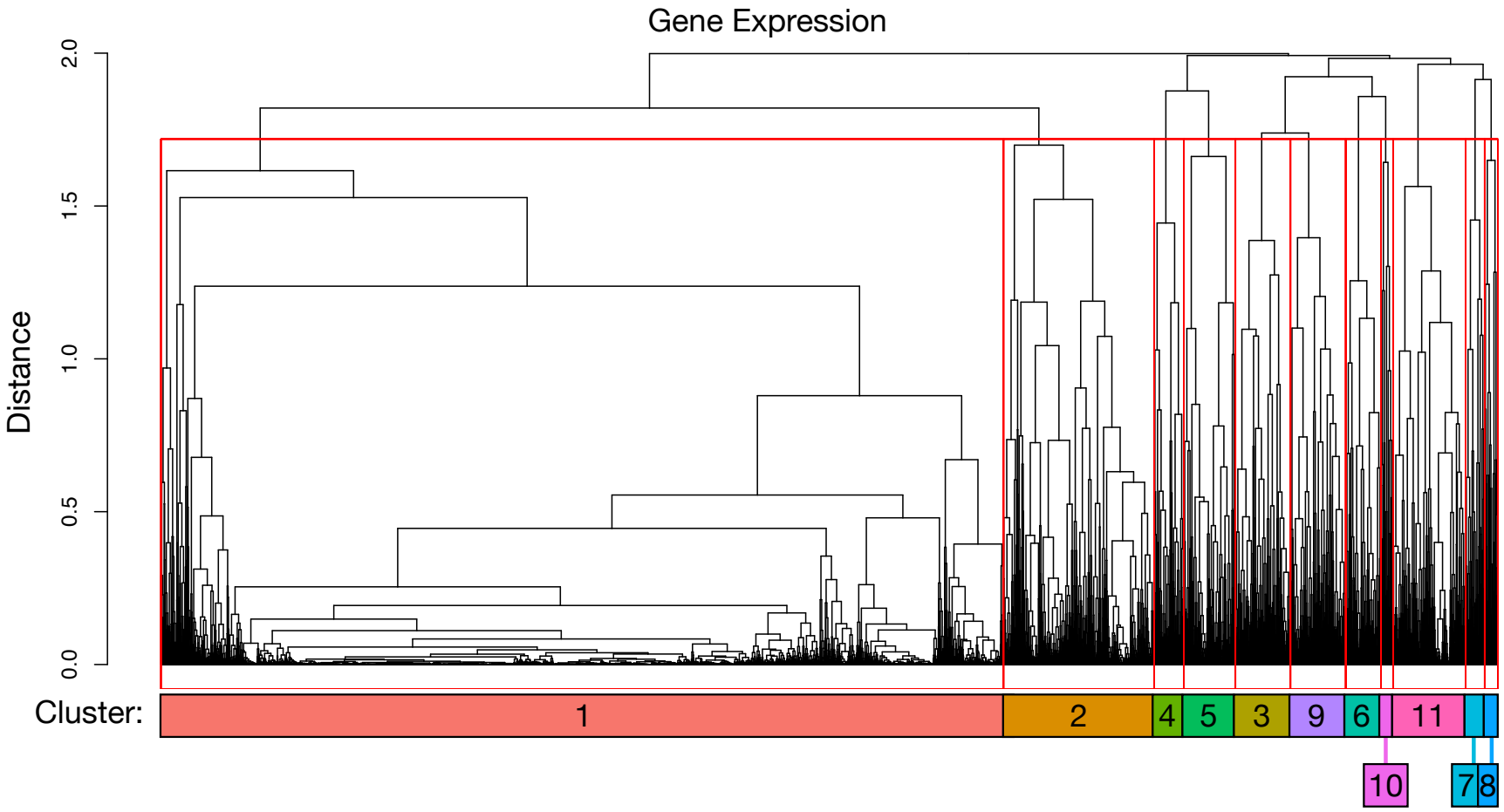


Figure 3.5.2. Clusters of gene expression profiles in the midgut sections data. *Clusters were identified from hierarchical clustering of gene expression profiles.*

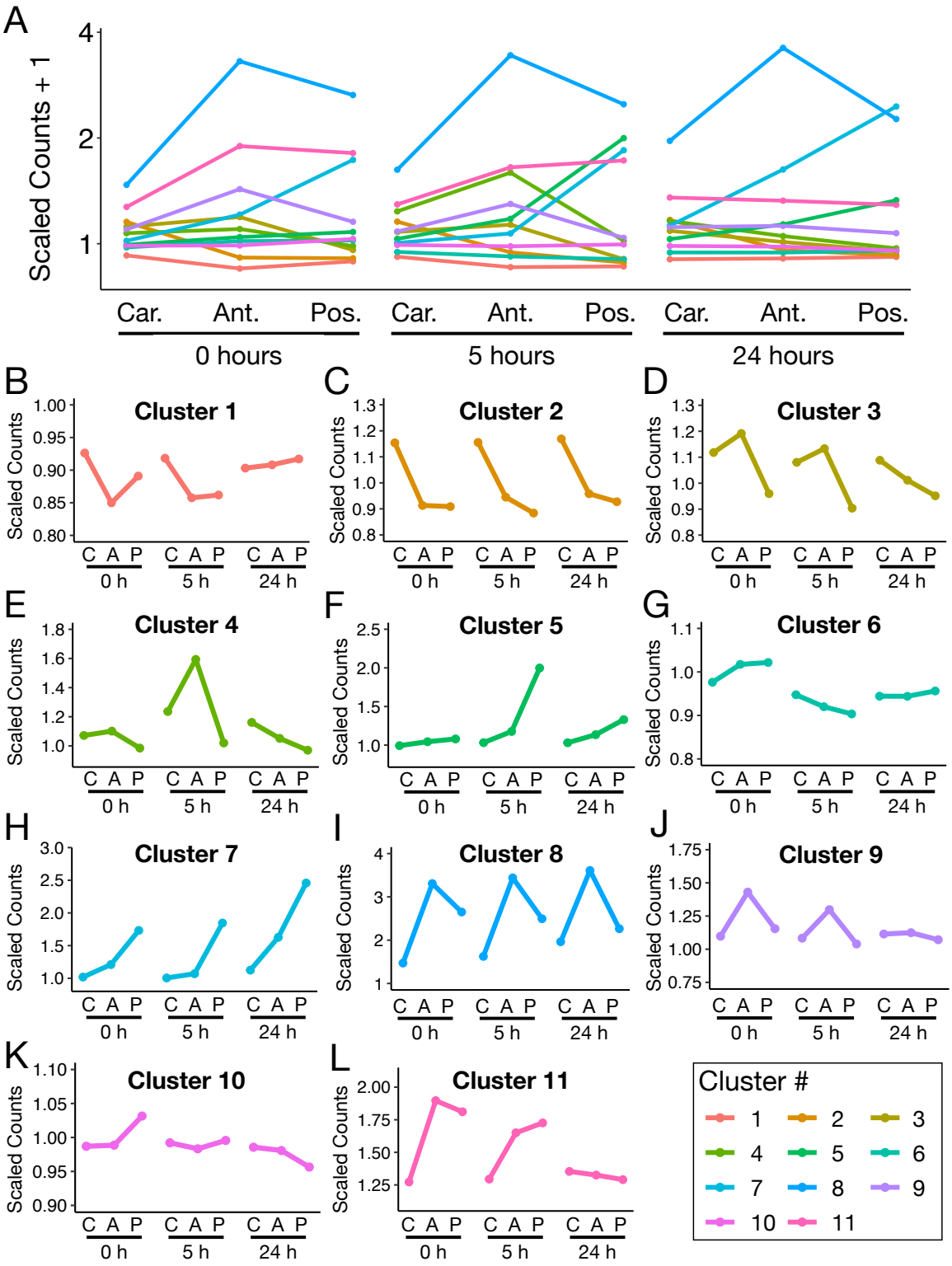


Figure 3.5.3. Average expression profile of each cluster. *Average expression profile of A) all clusters and B-L) each cluster individually. Genes were hierarchically clustered based on their expression profile across the time points and midgut tissues. Normalized counts of each gene were scaled prior to clustering. All panels show the scaled counts + 1. C/Car. = cardia, A/Ant. = anterior, P/Pos. = posterior tissues of the midgut.*

Table 3.5.1. Significantly enriched KEGG terms associated with gene clusters identified in the midgut sections RNA-seq data.

Cluster #	n	Expression Profile	KEGG Term	∩	Term ID	Adjusted *p*
1	5,317	Peaks in cardia, low expression throughout	Spliceosome	89	KEGG:03040	6.34×10^{-13}
			Basal transcription factors	29	KEGG:03022	1.52×10^{-5}
			RNA degradation	41	KEGG:03018	1.50×10^{-4}
			DNA replication	27	KEGG:03030	2.14×10^{-4}
			Fanconi anemia pathway	20	KEGG:03460	1.64×10^{-3}
			Ribosome biogenesis in eukaryotes	47	KEGG:03008	2.05×10^{-3}
			Mismatch repair	15	KEGG:03430	2.48×10^{-3}
			RNA polymerase	20	KEGG:03020	1.63×10^{-2}
			Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	19	KEGG:00563	2.92×10^{-2}
			Homologous recombination	17	KEGG:03440	3.59×10^{-2}
			Nucleotide excision repair	26	KEGG:03420	3.70×10^{-2}
2	950	Cardia >>> Anterior > Posterior	Valine, leucine and isoleucine degradation	7	KEGG:00280	6.35×10^{-9}
			Metabolic pathways	10	KEGG:01100	7.44×10^{-5}
			Fatty acid degradation	11	KEGG:00071	1.40×10^{-4}
			Propanoate metabolism	12	KEGG:00640	1.86×10^{-3}
			One carbon pool by folate	18	KEGG:00670	1.87×10^{-3}
			Glycine, serine and threonine metabolism	139	KEGG:00260	6.20×10^{-3}
3	347	Cardia = Anterior > Posterior	Circadian rhythm - fly	4	KEGG:04711	1.61×10^{-2}
4	189	Peaks at 5 h Anterior	Aminoacyl-tRNA biosynthesis	12	KEGG:00970	1.04×10^{-8}
			Protein processing in endoplasmic reticulum	10	KEGG:04141	8.28×10^{-3}
5	325	Peaks at 5 h Posterior	Protein processing in endoplasmic reticulum	30	KEGG:04141	2.82×10^{-13}
			Phagosome	19	KEGG:04145	1.10×10^{-8}
			Glutathione metabolism	13	KEGG:00480	5.85×10^{-3}
			Protein export	6	KEGG:03060	2.24×10^{-2}
			Metabolic pathways	72	KEGG:01100	4.94×10^{-2}
6	221	Peaks at 0 h (all tissues)	Longevity regulating pathway - multiple species	8	KEGG:04213	7.31×10^{-4}
			Pentose and glucuronate interconversions	6	KEGG:00040	9.32×10^{-3}
			Autophagy - animal	8	KEGG:04140	4.58×10^{-2}
7	120	Posterior >>> Anterior > Cardia	Lysosome	8	KEGG:04142	4.17×10^{-3}
8	83	Anterior >>> Posterior > Cardia	N/A	N/A	N/A	N/A
9	349	Peaks at 0, 5 h Anterior, low expression at 24 h	FoxO signaling pathway	8	KEGG:04068	1.31×10^{-2}
			Autophagy - animal	10	KEGG:04140	2.84×10^{-2}
			Mitophagy - animal	6	KEGG:04137	3.06×10^{-2}
			Glycolysis / Gluconeogenesis	6	KEGG:00010	5.00×10^{-2}

Cluster #	n	Expression Profile	KEGG Term	∩	Term ID	Adjusted *p*
10	77	Peaks at 0 h Posterior	Oxidative phosphorylation	32	KEGG:00190	3.96×10^{-30}
			Metabolic pathways	43	KEGG:01100	6.08×10^{-8}
11	208	0, 5 h: Anterior = Posterior > Cardia, low expression at 24 h	Ribosome	76	KEGG:03010	1.86×10^{-51}
			Citrate cycle (TCA cycle)	18	KEGG:00020	9.23×10^{-12}
			Oxidative phosphorylation	30	KEGG:00190	8.37×10^{-8}
			Carbon metabolism	23	KEGG:01200	8.12×10^{-6}
			2-Oxocarboxylic acid metabolism	8	KEGG:01210	6.70×10^{-5}
			Biosynthesis of amino acids	12	KEGG:01230	2.08×10^{-2}

Cluster number corresponds to the clusters in Figure 3.5.3; n = number of genes belonging to each cluster; ∩ *= intersection,* i.e. *the number of overlapping genes belonging to a cluster and a KEGG term; adjusted* p *= false discovery rate-adjusted* p*-values of Fisher's one-tailed tests, significant if adjusted* p < *0.05. Significantly enriched KEGG terms were identified with g:Profiler (Raudvere et al, 2019).*

3.5.1 Genes enriched in more anterior tissues of the midgut

Genes belonging to Cluster 1 represented approximately 60% of all genes detected in this RNA-seq dataset. Although the average expression profile of genes belonging to this cluster indicates that expression peaks in the cardia (Fig. 3.5.3B), the overall average counts were extremely low (Fig. 3.5.3A). Prior to clustering, the normalized counts of each gene across all tissues and time points were scaled in the same manner. As a result, minute differences (*e.g.* difference of 1 count between samples) can skew the expression profiles drastically and the average expression profile may be misleading. Therefore, the expression of this cluster will be considered as relatively stable expression throughout the tissues and time points and may include the genes which do not strongly cluster to genes in any of the other clusters. Many of the enriched KEGG terms of this cluster are related to transcription. This could be interpreted as transcription factors and other components involved in gene expression (*e.g.* splicing factors, polymerases) are necessary in all cells along the midgut axis, meaning these genes are expressed in a relatively stable manner across the tissues and time points.

Cluster 2 was the second largest cluster identified, representing over 11% of all genes and was characterized by those which are highly expressed in the cardia at all time points (Fig. 3.5.3C). The enriched KEGG terms from this cluster included a series of metabolic pathways such as ‘fatty acid degradation’ (KEGG:00071) and ‘glycine, serine, and threonine metabolism’ (KEGG:00260). The intersection of genes between the ‘metabolic pathway’ (KEGG:01100) term and those belonging to Cluster 2 include α-galactosidase (AGAP005846), ribose-phosphate pyrophosphokinase (AGAP004980), 6 glutathione S-transferases (GSTs), and 2 cytochrome c oxidase components. The former enzyme catalyzes the hydrolysis of raffinose, a sugar found in plant nectars which has been isolated from mosquito crops (Burkett, Carlson & Kline, 1998; Manda et al, 2007). The latter three groups of enzymes have been implicated in conferring insecticide resistance in a variety of insects including mosquitoes (Pridgeon & Liu, 2003; Ranson & Hemingway, 2005; Hong et al, 2013; Wilding et al, 2015). Additionally, GSTs have numerous substrates and are involved in the detoxification of free radicals, such as those generated during hemoglobin digestion, and also the metabolism of dietary molecules (Ranson & Hemingway, 2005). Alongside many other metabolic enzymes, their

enrichment in the cardia at all time points might parallel a hypothesis proposed by Dong et al, where immune effectors are produced in the cardia to be blended into the blood meal for immediate targeting of blood borne pathogens, and to have an effect in the posterior midgut (2006). This strategy might also be useful for the digestion of sugars. In the case of ingested sugars, the anterior region of the midgut is heavily involved in their metabolism and absorption (Hecker et al, 1977; Hixson et al, 2022). Therefore, their immediate breakdown with such enzymes in the cardia may aid or allow for more efficient absorption in the anterior midgut. Two other sugar digestive enzymes were also enriched in the cardia but showed membership to Cluster 1.

Genes belonging to Cluster 8 showed peaks of expression in the anterior at all time points (Fig, 3.5.3I). Just 83 genes belonged to the cluster, and no KEGG terms were enriched. One of the genes in this cluster was sugar transporter early response to dehydration 6-like 7 (AGAP001027). This gene shares homology to glucose and trehalose transporters in *Ae. albopictus* and other anophelines, further implicating the anterior midgut in sugar absorption. REL2 (AGAP006747), the transcription factor of the IMD pathway, was one of the only immunity genes belonging to this cluster. In addition to peaks of expression in anterior regions of the midgut, a peak of expression was also observed in the posterior at 24 hours post-blood meal. This suggests enrichment of REL2 in the cardia and anterior, but post-blood meal, its expression is induced in the posterior by the proliferating microbiota.

Cluster 3 consisted of genes with higher expression in both the cardia and anterior regions of the midgut at all time points (Fig. 3.5.3D). This cluster included over 300 genes, and just one enriched KEGG term was identified in this cluster: ‘circadian rhythm - fly’ (KEGG:04711). The enrichment of this term was driven by 4 genes encoding: Clock (CLK*)*, Vrille (VRI), glycogen synthase kinase-3β, and an unknown protein, while the term itself is made up of 10 genes. CLK is a master regulator of circadian rhythms and exhibits cyclic expression itself (Allada et al, 1998). In the mosquito body (excluding the head/antennae), *CLK* expression peaks during the daytime (Das & Dimopoulos, 2008). *VRI* encodes a transcription factor and is under the control of CLK, which in turn acts as a negative regulator of *CLK* (Cyran et al, 2003; Glossop et al, 2003).

Another gene in this cluster relating to circadian rhythms was *takeout (TO) 2* (AGAP012703). In *Drosophila*, *to* is highly expressed in structures relating to feeding/food detection, including the cardia, crop, and antennae (Sarov-Blat et al, 2000). *Drosophila to* expression is induced upon starvation and is downstream of CLK, meaning circadian rhythmic control and feeding responses are linked (Sarov-Blat et al, 2000; So et al, 2000). Additionally, *to* drives locomotion and food foraging behaviors (Meunier, Belgacem & Martin, 2007). In *Anopheles*, three homologs of *to* were identified (Das & Dimopoulos, 2008). Here, the two other *TO* genes, *TO1* (AGAP004263) and *TO3* (AGAP004262), belonged to Cluster 2 and were also enriched in the more anterior regions of the midgut. *Anopheles TO* expression is also induced upon starvation and is cyclic, corresponding with the peak of *CLK* expression in the body during daytime (Das & Dimopoulos, 2008; Rund et al, 2013). Given that sample collection for this present RNA-seq experiment occurred during the daytime, it was assumed that *CLK* and hence *TO* expression was at its peak. Expression of all TOs peaked at 0 hours, with expression decreasing after blood feeding. In addition to the digestive enzymes in Cluster 2, their abundance in the more anterior regions of the midgut may therefore indicate sugar-seeking behavior. The cardia is connected to the sugar storage organs and acts as valve, where its closure ensures sugars are transported into the crop and dorsal diverticula. The enrichment of circadian and feeding mechanisms may elucidate a role of the cardia in signaling or receiving signals for sugar-seeking behavior.

Strikingly, all canonical AMPs belonged to Clusters 1–3, indicating that they were enriched in the more anterior regions of the midgut. *ATT1* was found in Cluster 1, as its expression showed a slight upregulation in the cardia compared to other regions of the midgut, but overall expression was low. *DEF2* was also found in Cluster 1, but had relatively stable and low expression throughout the tissues and time points. *CEC1–3* and *DEF1* on the other hand were found in Cluster 2 and showed extremely drastic differences in expression between the cardia and posterior midgut. The differences in AMP expression between the cardia and posterior midgut at the same time point ranged between 200 to nearly 900-fold. Many previous studies have identified this expression profile in sugar-fed midguts (Vizioli et al, 2000; Warr et al, 2007; Rodgers et al, 2017; Hixson et al, 2022). Comparisons between AMP expression within the cardia at various

time points were insignificant (Wald tests, false discovery rate-adjusted $p > 0.05$), suggesting that AMP expression is not induced or changed by the blood meal. In the posterior, AMP levels were the lowest, and for *CEC1*, expression significantly decreased at 24 hours post-blood meal compared to pre-blood feed levels (Wald test, $W = 3.67$, false discovery rate-adjusted $p = 0.003$). *GAM1* showed membership to Cluster 3, where expression was enriched in the cardia at both 0 and 5 hours post-blood feeding, but at 24 hours, expression was upregulated in the anterior to a similar level as the cardia. Taken together, this suggests a significant investment into the production of AMPs by the cardia, and reveals its key role in immunity. The remaining AMPs (*CEC4*, *DEF3–5*) were excluded from the clustering as their average counts were too low. Other genes with antimicrobial activity found in Cluster 2 include class C lysozymes 1 (*LYSC1*; AGAP007347) and 4 (*LYSC4*; AGAP0073850). Lysozymes digest bacterial cell walls and can potentiate immune recognition by producing bacterial PAMPs. Previous studies have shown that LYSC1 plays a role in antibacterial responses (Kajla et al, 2010; Habtewold et al, 2022).

A mixture of immune activators and suppressors were also found to be consistently enriched in the cardia. IMD pathway antagonist *PGRP-LB* (AGAP001212) was found in Cluster 2, indicating an enrichment in the cardia at all time points. Inhibitory SPRNs 2, 4, and 9 (AGAP006911, AGAP009670, and AGAP003139 respectively) were also found in Cluster 2. All annotated CLIPs with average counts greater than 50 were found in Clusters 1 and 2, totaling 36. CLIPs are separated into 5 classes (A–E) and contribute towards a variety of functions. For example, CLIPA8 (AGAP010731) promotes parasite melanization, but A2 (AGAP011790) and A5 (AGAP011787) inhibit melanization of *P. berghei* (Volz et al, 2006). CLIPB14 (AGAP010833) and B15 (AGAP009844) were previously shown to be induced by bacteria in the midgut and play a role in antibacterial responses (Christophides et al, 2002; Volz et al, 2005). Although *CLIPB14* was also enriched in the cardia, it was excluded from hierarchical clustering due to its low counts. The enrichment of CLIPs in the anterior regions of the midgut was unexpected, given some of their roles in the melanization of malaria parasites, which occurs in the hemolymph. Indeed, in the *Ag/Pf* data, the expression levels of these CLIPs were generally enriched in the carcass compared to the midgut, indicating that they are more

abundantly expressed in the hemolymph. While CLIPAs and Cs are considered to be inactive proteases, CLIPBs do possess protease activity and their expression in the midgut may be related to digestion rather than immunity.

The majority of LRIMs were also enriched in the cardia. LRIMs are putative PRRs, and several have previously been characterized to antagonize *Plasmodium* development (Dong et al, 2006; Frolet et al, 2006; Garver et al, 2012; Molina-Cruz et al, 2012; Upton, Povelones & Christophides, 2014). Their enrichment in the cardia might indicate pathogen sensing by the cardia and/or secretion into food sources to aid pathogen neutralization or killing.

3.5.2 Blood meal-responsive genes in the more posterior regions

The expression of genes in Clusters 4 and 5 peaked at 5 hours post-blood feeding in the anterior and posterior respectively (Fig. 3.5.3E–F). Both clusters shared the enrichment of the term 'protein processing in endoplasmic reticulum' (KEGG:04141), in addition to each having additional enriched KEGG terms relating to protein expression (Table 3.5.1). Secreted and membrane-associated proteins are synthesized by ribosomes attached to the endoplasmic reticulum (ER). Polypeptides are folded or modified in the ER post-translation, and correctly folded proteins are transported to the Golgi body for further modification/secretion. The other KEGG term enriched in Cluster 4 was 'aminoacyl-tRNA biosynthesis' (KEGG:00970), where 13 tRNA synthetase genes overlapped between the term and the genes within the cluster. These enzymes are involved in attaching the correct amino acid to its corresponding tRNA, which is an essential component of translation. In Cluster 5, membership of SEC61 complex subunits α and β (AGAP009182 and AGAP011504 respectively) contributed to the enrichment of the KEGG term 'protein export' (KEGG:03060). Heterotrimeric SEC61 complexes form a protein translocation channel in the ER, enabling shuttling of proteins into the cytosol and the Golgi body, where they can be secreted outside of the cell (Römisch, 1999). The enrichment of such terms suggests heavy investment of these tissues into the production of proteins in response to the blood meal, which are likely to be secreted into the gut lumen. These proteins include digestive enzymes such as peptidases and trypsins. Other genes relevant to protein production and transport in these clusters include

eukaryotic translation initiation factor 3 subunit J (AGAP006613, Cluster 4), translation initiation factor 4E (AGAP007172; Cluster 5), and 4 tRNA-synthetases in Cluster 5.

Another blood meal-induced group of genes was Cluster 7. While expression peaked in the posterior at all time points, the average expression further increased in the posterior at 24 hours post-blood meal (Fig. 3.5.3H). Cluster 7 was made up of 120 genes and one KEGG term was significantly enriched: 'lysosome' (KEGG:04142). Lysosomes are acidic organelles which contain a wide variety of digestive enzymes for the breakdown of macromolecules. Three of the 8 genes in the intersection between the 'lysosome' term and Cluster 7 encode Niemann-Pick type C (NPC) proteins. NPCs are implicated in lipid transport, where NPC1 proteins are transmembrane proteins on the lysosome surface and NPC2s are soluble (Willenborg et al, 2005; Castellano et al, 2017). During blood meal digestion, lipids are absorbed and transported through the ER, where lipid droplets are formed for lipid storage (Briegel & Hörler, 1990; Beller et al, 2010; Pinch et al, 2021), which in turn has implications on oogenesis. Thus, the enrichment of NPCs in this cluster reflect the digestion and transport of lipids from the blood meal. NPCs have also been experimentally implicated in mosquito immunity against pathogens: 1) NPC expression is induced in response to *P. falciparum* and bacterial infections of both Gram type, and 2) silencing of NPCs increases susceptibility to infection by the aforementioned pathogens (Dong et al, 2006; Cui et al, 2020; Kulkarni et al, 2021). Therefore, upregulation of NPCs observed in this cluster may additionally represent induced responses against the expanding microbiota post-blood meal.

Given that the blood meal is digested in the posterior region of the midgut, Clusters 5 and 7 were investigated in more detail to characterize blood meal responses. These two clusters contain an abundance of genes annotated with peptidase/proteolytic activity. Clusters 5 and 7 include 21 and 31 peptidases respectively, totaling 52 unique genes encoding enzymes with proteolytic activity. These enzymes include 5 carboxypeptidases and 9 trypsins. Digestive enzymes are known to be upregulated in response to a blood meal to aid with its digestion (Marinotti et al, 2006). However, their enrichment in the posterior was observed even at 0 hours, *i.e*. in the absence of a blood meal. These results further confirm the digestive role of the posterior midgut, but also elucidate that

high expression prior to blood meal may be an anticipatory mechanism for more efficient digestion.

A negative regulator of the Imd pathway was a member of Cluster 7. Enrichment of *CAD* (AGAP009646) in the posterior midgut suggests tolerance mechanisms in place to restrict excess immune activation and protect the microbiota. At the same time, CAD expression was downregulated post-blood meal, perhaps reflecting the need for immune activation to control the proliferating microbiota. The IMD pathway is more implicated in controlling the overexpansion of microbiota (Meister et al, 2009). IMD pathway component *IKK2* (AGAP005933) was also a member of Cluster 7 and is involved in pathway activation. *IKK2* was enriched in the posterior at 0 and 24 hours post-blood meal. The other component of the IKK complex, *IKK1* (AGAP009166), belonged to Cluster 1 and had stable expression across the tissues and time points. Following *S. marcescens* infection, a previous study found that *IKK2* but not *IKK1* was upregulated (Kulkarni et al, 2021). Here, the enrichment of *IKK2* in the posterior could be in response to bacterial expansion *(*Bando et al, 2013; Gendrin et al, 2015). Three TOLL pathway components were also identified in Clusters 5 and 7. TOLL pathway negative regulator *CACT* (AGAP007938) was enriched in the posterior midgut at 5 hours post-blood meal. Like *CAD*, this can also be explained as being a tolerance mechanism in the posterior midgut. Membership of two TOLL pathway promoters *TOLL9* (AGAP006974) and tumor necrosis factor receptor-associated factor 6 (*TRAF6*; AGAP000388) were also enriched in Clusters 5 and 7 respectively. At all three time points, these genes were enriched in the posterior compared to other regions of the midgut, which may represent activation by microbiota. The membership of positive and negative regulators of immune pathways highlight the co-existence of immune activation and suppression in a concurrent manner, and/or the presence of self-regulatory mechanisms. IMPer (AGAP013327) protects the microbiota by forming a protein network around the blood bolus and was induced by blood feeding as previously described (Kumar et al, 2010). Congruently, *IMPer* was a member of Cluster 5 and its expression was consistently enriched in the posterior. At 5 and 24 hours post-blood feeding, expression increased up to 200-fold compared to at 0 hours.

Three scavenger receptor (SCRs) were also found in Cluster 5, indicating a blood meal-responsive upregulation in the posterior. Specifically, 3 of the 4 orthologs of the

Drosophila SCR class B protein Croquemort (SCRBQ) displayed membership to Cluster 5. SCRs are transmembrane receptors that have many functions, including binding to lipoproteins, the recognition and phagocytosis of bacteria, as well as the clearance of apoptotic cells. In *Drosophila* embryos, Croquemort was found on the surface of phagocytic macrophages and aid the binding to and phagocytosis of apoptotic cells (Franc et al, 1999). In adults, Croquemort aids the clearance of bacteria and apoptotic cells, in addition to phagosome maturation (Guillou et al, 2016). Very few studies have investigated SCRBQs in *Anopheles*, however *SCRBQ2* (AGAP010133), the closest homolog to Croquemort, was found to be upregulated in the midgut post-blood meal and during *P. berghei* ookinete invasion (González-Lázaro et al, 2009). *SCRBQ2* is also responsive to bacterial infections of both Gram-types, as well as sterile injuries (Christophides et al, 2002). The assignment of the remaining three SCRBQs to Cluster 5 may indicate that they are expressed on the surface of epithelial cells and perhaps are involved in the recognition of bacteria, as the microbiota will start to flourish in the posterior following a blood meal.

3.5.3 Sugar meal-responsive genes at earlier time points

Cluster 6 was defined by genes with a peak of expression in all tissues prior to blood feeding, followed by a subsequent downregulation post-blood meal (Fig. 3.5.3G). Three KEGG terms were enriched, including ‘pentose and glucuronate interconversions (KEGG:00040) and ‘autophagy - animal’ (KEGG:04140; Table 3.5.1). The enrichment of these terms could be explained by the fact that in the absence of a blood meal, cellular maintenance activities are performed, but following blood feeding, resources are invested towards blood digestion rather than these general maintenance mechanisms. The former term is closely related to glycolysis, which produces energy, and the pentose phosphate pathway (PPP). The PPP is an alternative branch of glycolysis and is a crucial part of cellular maintenance and metabolism. This pathway metabolizes sugars to produce precursors to DNA bases and amino acids. Additionally, this pathway also protects against oxidative stress (reviewed in Stincone et al, 2015). The latter term also relates to maintaining cellular homeostasis. Autophagy is the lysosome-mediated degradation of

intracellular material and relates to development, immunity, and waste/damage removal (Mulakkal et al, 2014).

Genes belonging to Clusters 9–11 were also characterized by a similar expression profile to that of Cluster 6. Cluster 10 included genes which peak in the posterior during sugar feeding. Clusters 9 and 11 included genes which peak at both 0 and 5 hours followed by reduced but stable expression across the tissues at 24 hours post-blood meal. Genes in Cluster 9 peaked in the anterior, while genes in Cluster 11 peaked in the both the anterior and posterior regions of the midgut. As with Cluster 6, the enriched GO terms were very relevant to cellular maintenance mechanisms, such as 'citrate cycle (TCA cycle)' (KEGG:00020) and 'carbon metabolism' (KEGG:01200) in Cluster 11, 'oxidative phosphorylation' (KEGG:00190) in Clusters 10 and 11, and 'glycolysis / gluconeogenesis' (KEGG:00010) in Cluster 9, which also had the 'autophagy - animal' KEGG term. The abundance of these cellular respiration-related terms further supports the notion that a large transcriptional reprogramming occurs following blood feeding to digest the blood meal.

The presence of the 'autophagy - animal' term in two different clusters means that different genes triggered the enrichment of the same term. Further investigation of the genes that triggered the enrichment of this term found that Cluster 6 contained some inhibitors of autophagy. Characterization of autophagy in *Anopheles* is limited; much of the knowledge gained so far is derived from yeast, plant, and mammalian models. Kontroller of growth 1 (*KOG1*; AGAP010035; known as Raptor in *Drosophila* and mammals) and target of rapamycin (*TOR*; AGAP007873) encode components of the TOR complex (TORC) 1 and were enriched in Cluster 6. Additionally, Ras homolog enriched in brain (*RHEB*; AGAP001902) was also present in this cluster and encodes an activator of TORC1. TORC1 inhibits autophagy in favor of cell growth through processes such as protein synthesis, transcription, and metabolism (Wullschleger, Loewith & Hall, 2006; Jung et al, 2010). Under starvation conditions (*e.g.* in the absence of glucose) or other forms of stress, autophagy can be triggered through the inhibition of TORC1 (Wullschleger, Loewith & Hall, 2006). RHEB also activates TORC2, which also contains TOR, and is involved in actin organization (Wullschleger, Loewith & Hall, 2006). Inhibition of TORC2 initiates autophagy in skeletal muscle cells of mammals (Zhao et al, 2007).

The enrichment of this term in Cluster 9 was driven by three cathepsins, which are intracellular proteinases that mediate autophagy (Kaminskyy & Zhivotovsky, 2012). Two other terms in Cluster 9 relevant to autophagy are 'mitophagy - animal' (KEGG:04137) and 'FoxO signaling pathway' (KEGG:04068). ROS inhibit mitochondrial function, which is hypothesized to also trigger mitophagy in an TORC1-dependent manner (Desai, Myers & Schreiber, 2002; Kim et al, 2002; Jung et al, 2010). The Forkhead family subgroup O (FOXO) transcription factors are downstream of TORC2 and promote autophagy in mammals (Zhao et al, 2007). It is unclear what the TORC2 pathway controls in *Anopheles*, however more recent research has shown that it mediates heat resistance in *Drosophila* (Jevtov et al, 2015). Therefore, the enrichment of this term in this dataset could reflect a combination of basal biological processes, including autophagy, cellular growth, and metabolism.

Two other genes relevant to apoptosis were also enriched in Cluster 9. *SCRBQ2*, the closest homolog to *Drosophila* Croquemort, was found in Cluster 9. Given the role of Croquemort in *Drosophila* immune responses, its expression being higher in posterior regions could be explained by its putative role in antibacterial responses as the microbiota are harbored there. Its downregulation at 5 hours could indicate a shift in investment into blood meal digestion-related genes. In addition, *IAP6* (AGAP002651), a homolog of *Drosophila* baculovirus IAP repeat-containing ubiquitin-conjugating enzyme (*Bruce*), was also a member of Cluster 9. Closer examination of its expression profile suggests that it is expressed in a relatively stable manner across the tissues and time points. IAPs, as their name suggests, are involved in the inhibition of apoptosis, typically achieved through the regulation of caspases via ubiquitination (Silke & Vaux, 2001; Orme & Meier, 2009). Bruce in *Drosophila* has been found to inhibit cell death and apoptosis in eye cells (Vernooy et al, 2002), and additionally inhibit autophagy under nutrient-rich conditions in the ovary (Hou et al, 2008). Another IAP, IAP2, is involved in Imd pathway activation and signaling via the ubiquitination of various other Imd pathway components (Meinander et al, 2012; Kleino & Silverman, 2014). To date, the roles of the majority of *Anopheles* IAPs remain elusive. In *Ae. aegypti*, the KD of *IAP6* and 3 other IAPs had no effect on apoptosis of cells in *vitro* and mosquito mortality (Pugliese, Estep & Becnel, 2015). Taken

together, these findings point towards most IAPs performing signaling rather than apoptotic roles in *Anopheles*.

Many immune pathway components and effectors were also enriched in these clusters, indicating that they are generally downregulated following a blood meal. *PGRP-LC* (AGAP005203) encodes an activator of the IMD pathway and was found in Cluster 9. *PGRP-LC* expression peaked in the cardia and anterior prior to blood feeding. Post-blood meal, expression generally decreased, perhaps reflecting a need for tolerance of the microbiota. *SRPN10* (AGAP005246), which encodes an inhibitory SRPN, was a member of Cluster 11, and *CASPAR* (AGAP006473), which encodes a negative regulator of the IMD pathway, was a member of Cluster 6. Both genes shared a similar profile, where expression consistently peaked in the posterior, but blood feeding also generally led to their downregulation. Previously, *SRPN10* expression was induced upon *P. berghei* ookinete invasion of the midgut (Danielli, Kafatos & Loukeris, 2003). In the same study, *SRPN10* expression was not altered upon bacterial challenge. Their downregulation at 24 hours post-blood feed observed in this experiment might represent the opposite of PGRP-LC, *i.e.* need for immune resistance.

Three carbohydrate-binding lectins were also found in these clusters. Galectins (GALEs) and C-type lectins (CTLs) are proteins which can bind to glycans in glycoproteins and glycolipids and thus can act as PRRs (Lu et al, 2020). In *Ae. aegypti* and *Culex pipiens* mosquitoes, some CTLs have been described to be microbiota-induced and also defend the microbiota from AMPs in the midgut, acting as a tolerance mechanism (Pang et al, 2016). In *Drosophila*, some CTLs directly bind bacteria and have been hypothesized to mediate melanization and encapsulation (Ao, Ling & Yu, 2007). Few studies have elucidated the functions of CTLs in *Anopheles*, but many appear to be induced in response to infection and have a role in anti-Gram-negative bacterial responses (Schnitger et al, 2009). At the same time, some CTLs also enable *Plasmodium* development (Osta, Christophides & Kafatos, 2004). *CTL6* (AGAP006267) was a member of Cluster 11, meaning expression was highest in the anterior and posterior at 0 and 5 hours post-blood feeding. This lectin was previously shown to be upregulated in midgut following *S. marcescens* oral infection (Dekmak et al, 2021). GALEs on the other hand have roles in *Drosophila* development in addition to immunity (Pace et al, 2002).

GALE5 (AGAP011287) and *GALE8* (AGAP012529) belong to Cluster 9, but the peak of expression was found in the posterior at 24 hours post-blood feeding. *GALE8* was previously shown to be induced upon bacterial and malarial infection (Dimopoulos et al, 1998). Contrastingly, *GALE5* was marginally induced upon *E. coli* infection, and is downregulated during oocyst and salivary gland sporozoite stages of *P. berghei* infection (Christophides et al, 2002). Further research on these lectins will be required to determine whether they are agonists or antagonists of pathogens and/or the microbiota.

3.5.4 The midgut shows a clear dichotomy of functions

The *Ag/Pf* data showed that AMPs are consistently expressed, irrespective of infection status and up to at least 7 days post-infection/blood feed. Congruent to this observation, expression of AMPs was enriched in the cardia at all time points post-blood feed. Taken together, this suggests that AMPs are constitutively expressed in the cardia. In response to blood feeding, the transcription of many digestive enzymes was initiated regardless of infection status. Within the midgut tissues, dozens of proteolytic enzymes were exclusively enriched in the posterior at all time points. These results indicate a constitutive digestive role of the posterior midgut. To date, studies have shown that *An. gambiae* midgut tissues have highly specialized functions, but have only been examined under sugar-fed conditions (Warr et al, 2007; Hixson et al, 2022). Considering these similar findings in the RNA-seq datasets, the secreted transcriptomes of each midgut tissue were compared to further characterize the compartmentalization of the midgut. TPM values were calculated for each sample to enable these comparisons. Transcripts without signal peptide annotations were filtered out, and TPM values were converted into percentages of the secreted transcriptome. Approximately 20% of the transcriptome was annotated to have signal peptides.

In the cardia, the 11 known AMPs make up a staggering ~40% of the secreted transcriptome prior to and post-blood feeding (Fig. 3.5.4). This is a disproportionately small number of genes to contribute towards such a large proportion of the transcriptome. Post-blood feeding, investment into AMPs fell, but only marginally. A stark contrast was observed in the posterior, where AMP expression was negligible (Fig. 3.5.4). Instead, peptidases and trypsins, enzymes which are involved in blood meal digestion, make up a significant proportion of the transcriptome. Expression of these enzymes increased by over 50% at 24 hours post-blood meal compared to earlier time points, indicating a clear additional investment into blood meal digestion following blood feeding. Surprisingly, although expression of these enzymes was relatively low in the cardia, their expression doubled at 24 hours post-blood meal to 5% of the secreted transcriptome. This indicates a role of the cardia in aiding blood meal digestion, despite its distance from the blood meal, or that these enzymes contribute to immune defense and digest microbial proteins. The transcriptome of the anterior was an intermediate between the cardia and posterior,

where expression of both AMPs and digestive enzymes was relatively similar. However, investment into digestive enzymes also increased at 24 hours post-blood meal and was coupled with a reduction in AMP expression. The differential expression of AMPs and digestive enzymes in the cardia and posterior highlight a clear dichotomy of functions across the midgut axis.

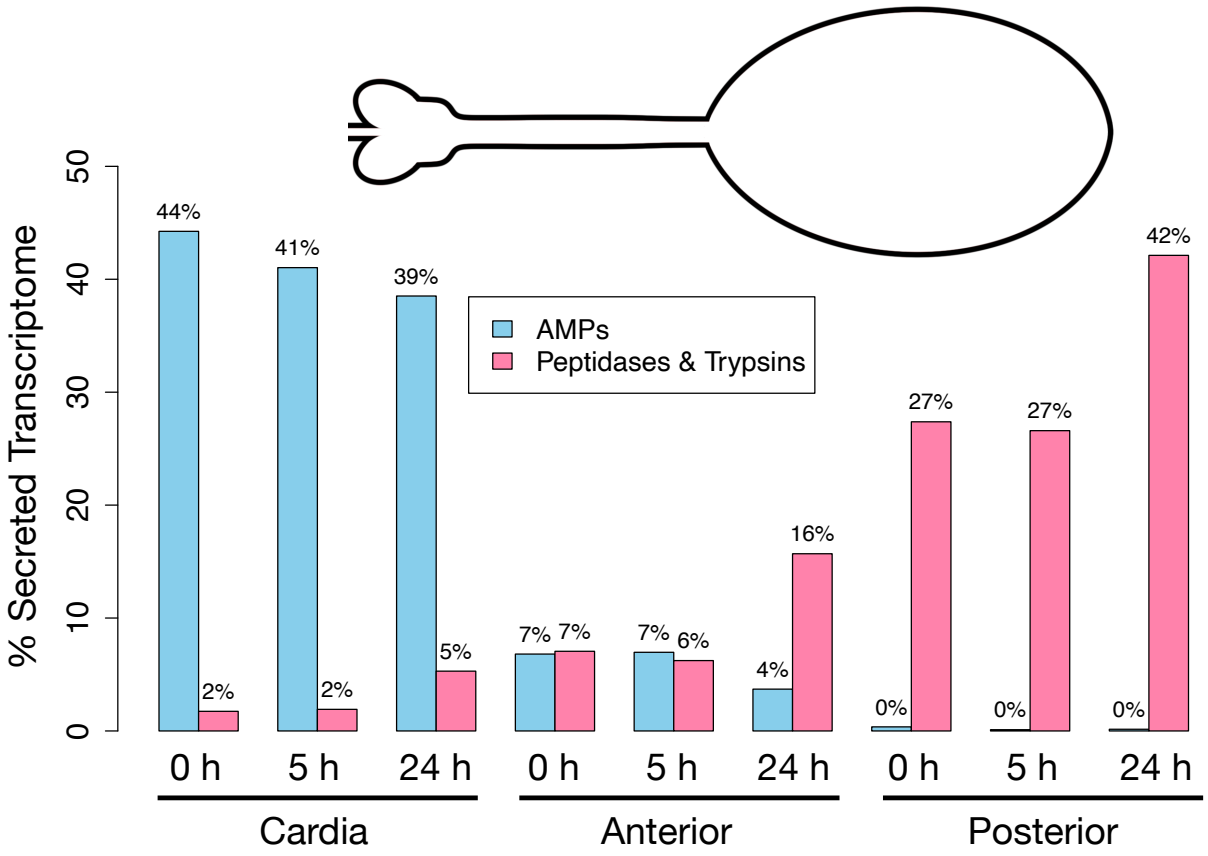


Figure 3.5.4. Midgut investment into expression of AMPs and digestive enzymes. *Percentage of secreted transcriptomes made up of known AMPs, and peptidases and trypsins. TPM values were calculated and averaged from three replicates of the midgut sections RNA-seq data, and converted into percentages of the secreted transcriptome. A total of 11 AMPs and 58 digestive enzymes were included in the secreted transcriptomes.*

3.6 Discussion

Characterization of the mosquito midgut transcriptomes in response to blood feeding and infection can provide important insights about the biology of the midgut and its interaction with blood-borne pathogens such as *Plasmodium*. In this study, very few *Plasmodium*-responsive genes were identified between direct comparisons of infected and control (blood-fed) tissues. A previous study reported similar findings at 24 hours post-*P. falciparum* infection in the midgut, where a similar percentage of the transcriptome was significantly DE in response to infection (Dong et al, 2006). In carcass samples of the same study, fewer genes were also significantly DE compared to the midgut (Dong et al, 2006). However, this previous study utilized DNA microarrays which are less sensitive and specific than RNA-seq. Here, RNA-seq was performed, so it was expected that a greater number of genes would be identified as *Plasmodium*-responsive in this present study. Despite the similarity in the number of DE genes/transcripts, only 12 immunity-related transcripts were identified in this dataset as *Plasmodium*-responsive, which contrasts with the dozens identified in the study by Dong et al (2006). This discrepancy could be explained by a mixture of high and low infection intensities in the mosquitoes sampled across the three replicates. Other studies have characterized infection intensity-dependent responses, where different effectors are launched in response to different infection intensities (Garver et al, 2012), immune responses are suppressed in low-intensity *P. falciparum* infections (Mendes et al, 2011), or responses are directly proportionate to the number of invading parasites (Kumar & Barillas-Mury, 2005; Simões et al, 2017). For example, oxidative responses launched against invading parasites in the midgut are costly, both in terms of energy and damage to host cells, and therefore are expected to only be launched when infection intensities are high (Kumar et al, 2003; Kumar & Barillas-Mury, 2005; Mendes et al, 2011). During low intensity *P. falciparum* infections, immune responses have been observed to be suppressed (Mendes et al, 2011). Given the mixture of high and low infection intensities across the replicates, it is likely that immune responses differed between the replicates, and thus obscured the immune responses against *P. falciparum*.

This explanation is also partially supported by the observation that PGRP-LB, a negative regulator of the IMD pathway, was significantly upregulated in response to early

stages of *P. falciparum* infection. PGRP-LB promotes tolerance to the microbiota post-blood meal by sequestering PGN, but was also previously characterized to play a role in the tolerance of *P. falciparum* (Gendrin et al, 2017). Its upregulation in infected midguts compared to controls therefore support the notion that immune responses were suppressed. Alternatively, PGRP-LB might have direct antiparasitic effects that have yet to be characterized. In the tsetse fly, PGRP-LB promotes tolerance to bacteria but also directly restricts trypanosome development (Wang et al, 2009; Wang & Aksoy, 2012). It would be interesting to test whether *Anopheles* PGRP-LB has direct antiplasmodial activity and if it does, its upregulation in response to parasite can then be characterized as a host response to restrict *Plasmodium* development.

In order to further investigate anti-*Plasmodium* responses, significantly DE genes identified from comparisons between different time points within the same tissue and infection status were compared. This enabled indirect comparisons of the impact of *Plasmodium* infection on mosquito tissues. GO term enrichment analysis was performed on the genes which were shared between and unique to each condition. In all cases, the majority of the significantly DE genes were shared between the infected and control samples, indicating that transcriptional responses to blood feeding are more pronounced than responses to *Plasmodium*. This was reflected in the GO terms enriched in the shared DE genes, and parallels the midgut sections data where the midgut also undergoes transcriptional reprogramming to aid blood meal digestion. Hierarchical clustering also revealed significant changes in the expression of the majority of genes post-blood feeding. Previously, comparisons of whole mosquito transcriptomes before and after blood feeding suggested that expression of approximately half of all genes were modulated in response to the blood meal (Marinotti et al, 2006). This translates to the majority of metabolic processes switching to blood meal digestion at the expense of other activities such as flight (Marinotti et al, 2006). In this study, these transcriptional responses include the upregulation of enzymes with proteolytic activity in the posterior post-blood meal, highlighting the crucial role of this tissue in blood meal digestion.

Additionally, a cluster of genes was identified on the basis that their expression was downregulated post-blood meal, and includes genes involved in housekeeping activities. Among the shared upregulated genes in midguts at 24 hours post-

infection/blood feed (compared to at 1 hour), terms such as ‘peroxisome’ and ‘catabolic process’ were enriched. These terms reflect biological processes upregulated following a blood meal, including the increased metabolic activity to digest the blood meal which in turn leads to increased oxidative stress (Kumar et al, 2003). Detoxification processes occur in the peroxisome. Of the shared downregulated genes at 24 hours versus at 1 hour, many transcription and translation-related terms were enriched. This perhaps indicates a reduction in the investment towards the expression of genes directly involved in blood meal digestion. Expression of many digestive enzymes such as trypsin 1 and CP peak before or at 24 hours post-blood meal (Müller et al, 1993; Müller et al, 1995; Edwards et al, 1997; Marinotti et al, 2006). Within the midgut, a drastic reduction in blood bolus size can be observed at 24 hours post-blood feeding, indicating that digestive activities have taken place. Investment into blood meal digestion can therefore begin to be reduced as the midgut starts to revert back to its metabolic state prior to blood feeding.

In the *Ag/Pf* data, female reproductive terms were enriched from the upregulated genes of control samples at 24 hours post-blood feed in the midgut, but not the infected samples. This reflects the reproductive and immune trade-offs present in many female insects (reviewed in Schwenke, Lazzaro & Wolfner, 2016). For example, after *S. marcescens* infection, the viability of stored sperm in oceanic field crickets (*Teleogryllus oceanicus*) is reduced (McNamara, van Lieshout & Simmons, 2014). Following fungal, bacterial, and parasitoid wasp infections, *D. melanogaster* exhibit reduced fecundity (Fellowes, Kraaijeveld & Godfray, 1999; Lazzaro et al, 2008; Bashir-Tanoli & Tinsley, 2014). In *Anopheles, Plasmodium* infection has been observed to increase the number of apoptotic follicles (Ahmed & Hurd, 2006) and reduce fecundity during infection (Hogg & Hurd, 1995). Here, *P. falciparum* infection likely caused an increase in resource allocation towards immunity at the expense of reproduction in infected *An. gambiae* females compared to uninfected controls. Accordingly, multiple fatty acid-related GO terms were enriched from the downregulated genes of infected mosquitoes at 24 hours post-blood feed. In *Ae. aegypti*, KD of elongase of long-chain fatty acids (ELO) significantly reduces egg length, female adult fertility, and fecundity (Chen et al, 2023). In the case of *AaELO9* depletion, dissection of unhatched eggs showed that no larvae were developing inside (Chen et al, 2023). Lipid metabolism has also been shown to be

essential for oogenesis in *Drosophila* (Reiff et al, 2015; Almeida-Oliveira et al, 2021). Taken together, these results suggest that ELOs are crucial for reproductive processes in mosquitoes. Among the fatty acid-related terms, AGAP010695, which encodes *An. gambiae* ELO4, contributed to the enrichment of these terms and shares homology to *AaELO9*. The downregulation of such terms and *ELO4* in infected mosquitoes further suggest a lack of investment into reproductive processes during infection.

In order to identify potential oocyst-responsive mechanisms, DE genes from carcass samples at 7 days-post infection/blood feed were compared. As before, GO term enrichment analysis was performed on transcripts significantly DE between 7 day and 1 hour time points of the same infection status in carcass samples. Among the genes downregulated specifically in control carcasses, translation-related and prostaglandin biosynthetic GO terms were enriched. Three microsomal GSTs triggered the enrichment of the prostaglandin-related terms. Cytosolic GSTs are involved in insecticide resistance in *Anopheles* (Wilding et al, 2015), but little is known about the microsomal GSTs. In humans, microsomal GSTs are involved in prostaglandin synthesis and protect against oxidative stress (Morgenstern, Zhang & Johansson, 2011). In *Drosophila*, genetic disruption of microsomal GST1 does not impact larval development, but the adult lifespan of mutants is significantly reduced (Toba & Aigaki, 2000). Assuming *Anopheles* microsomal GSTs also catalyze the synthesis of prostaglandins, the upregulation of GSTs at 1 hour (versus at 7 days) in control carcasses might reflect an anticipation of oxidative stress response to blood meal digestion, or the expanding microbiota in the midgut (García Gil de Muñoz et al, 2008). The absence of enriched prostaglandin-related GO terms in the infected carcass can be explained by a constant elicitation of GST expression in response to the parasite at both time points, or that GST expression is being modulated by *P. falciparum* (Ramos, Custódio & Silveira, 2014). Previously, *P. falciparum* has been shown to modulate and suppress host NOS expression (Tahar et al, 2002).

Analysis of transcriptomes typically relies on GO term enrichment analysis to capture key biological information about samples. Here, enriched GO/KEGG terms were identified to characterize differences in transcriptomes of different conditions, as well as provide valuable insight on the biology of the tissues in response to blood feeding and/or *P. falciparum* infection. However, the accuracy and/or comprehensiveness of these

analyses are only as robust as the annotations of genes in a genome. In other words, incomplete assignments of genes to GO terms can have a great impact on the enriched terms obtained. The absence of GO terms representing a specific biological mechanism does not necessarily reflect the absence of that mechanism. Hence, for more robust analyses and conclusions, individual genes within a condition should be considered in addition to the intersections.

One design flaw of this study was that in the *Ag/Pf* data, the carcass samples contained midguts, which likely further contributed to diluting hemolymph/fat body responses. Dissections of mosquito midguts are performed in PBS and typically lead to parts of the hemolymph and fat body falling out of the body cavity. Hence, the initial rationale was to collect and sequence the entire carcass to capture as much of the hemolymph/fat body transcriptome as possible. Future experiments should proceed with the removal of the midgut, but the tissue leakage into the dissection PBS could also be collected as well.

The *Ag/Pf* data provided some additional insights into mosquito responses to *P. falciparum* infection. As defined by its name, the immune responses observed are in response to an infection/behavior/stimulus. Both the *Ag/Pf* and midgut sections datasets have also enabled the observation of another form of immunity. AMPs are constitutively and highly expressed in the cardia, irrespective of infection or feeding status. While AMP expression has been known to be enriched in the cardia for decades (Vizioli et al, 2000; Warr et al, 2007; Rodgers et al, 2020; Hixson et al, 2022), their significant contribution to the secreted transcriptome throughout all time points post-blood feeding is extraordinary. It is hypothesized that these AMPs are secreted into the cardia lumen and remain there, as the cardia act as a valve. This would enable the accumulation of significant titers of AMPs, creating a disinfectant pool for the immediate sterilization of ingested food. This also challenges the traditionally characterized concepts of immune responses, where effectors are launched upon the recognition of a pathogen. Instead, the nature of AMP expression in the cardia appears to be anticipatory of potential pathogens introduced through feeding. Unexpectedly, *PGRP-LB*, which encodes a negative regulator of the IMD pathway, was also enriched in the cardia. Despite this, many antimicrobial effector genes controlled by the IMD pathway were also enriched in the cardia including *CEC1*,

LRIM1, and *LYSC1* (Meister et al, 2005; Luna et al, 2006; Gendrin et al, 2017, Rodgers et al, 2020). This suggests that mechanisms or pathways beyond the traditional insect immunity pathways control this constitutive expression of AMPs/other effectors in the cardia. The concept of anticipatory immunity was previously described by Upton, Povelones & Christophides following the observation that *LRIM9* expression was upregulated in response to the blood meal but not by exogenous or endogenous bacteria (2014). In the midgut sections data, hundreds of additional genes were identified to be blood meal-responsive, and many more were characterized to be consistently enriched in a specific midgut tissue across the three time points. The constant, high expression of AMPs in the cardia at all time points could be also be described as a form of basal immunity, as this pattern of expression appears to be an evolved, innate mechanism rather than a response to pathogens.

It is important to note that the many forms of immunity co-exist. Numerous studies have characterized the induction of AMPs/immune effectors in response to pathogen detection, *i.e.* the REL-dependent, traditionally characterized immune pathways (Vizioli et al, 2001a; Vizioli et al, 2001b; Dimopoulos et al, 2002; Dong et al, 2006; Rodgers et al, 2017). Expression of many AMPs, PGRPs, and CTLs are also microbiota-dependent in the midgut (Rodgers et al, 2017). In the posterior midgut, expression of TOLL pathway inhibitor *CACT* was responsive to the blood meal and/or the microbiota. *IMPer* expression was also blood meal-responsive and contributes towards the tolerance of the microbiota (Kumar et al, 2010). Other genes with immunity-related functions which were blood meal/microbiota-responsive include NPCs, *SCRBQ2*, *IKK2*, *TOLL9*, and *TRAF6*. The IMD pathway inhibitor *CAD* was constitutively enriched in the posterior, indicating that either tolerance of this pathway is a built-in mechanism, or *CAD* is being constantly activated by basal levels of microbiota which are harbored in the posterior. *REL2*, transcription factor of the IMD pathway, had enriched expression in the cardia and anterior throughout the time points, but expression in the posterior gradually increased post-blood meal to peak at 24 hours. This upregulation in the posterior could be a direct response to the expanding microbiota, and hence initiate the expression of other effectors. Subsequent studies could sample more and finer time points post-blood meal to further investigate the coexistence of such resistance and tolerance responses, including

whether these mechanisms are self-regulatory. As with the enrichment of *PGRP-LB* and AMP expression in the cardia, both resistance and tolerance mechanisms are present in the posterior midgut at the same time. Given the greater implication of the IMD pathway in antibacterial responses (Meister et al, 2005; Meister et al, 2009; Dong et al, 2011), modulation of the expression of its pathway components, downstream effectors, and negative regulators has a greater impact on the microbiota than modification of the TOLL pathway.

It is interesting to speculate a form of immune signaling across the midgut axis and also to other tissues. An increasing population of microbiota in the midgut leads an increased risk of disease and an increased detection of PAMPs by the mosquito; it would be beneficial for the host if these detection signals can be spread to the hemolymph so that expression of effectors is initiated in case the midgut barrier is breached. While all AMPs were highly expressed in the cardia, expression of GAM1 was also induced in the anterior at 24 hours post-blood feeding. It would be intriguing to determine whether this expression profile is an innate mechanism of the midgut in response to the blood meal, or if signals from the posterior relating to the proliferating microbiota induced this upregulation to boost immune responses. Previous experiments involving the use of antibiotics to clear the midgut microbiota showed a significant reduction in midgut AMP expression in whole midguts, despite the majority of the microbiota being hosted in the posterior (Rodgers et al, 2017; Rodgers et al, 2020). This suggests some form of signaling at least across the midgut axis.

Expression of half of the 11 AMPs present in the *Anopheles* genome was negligible in the midgut at all time points. These AMPs might be highly specific and their expression is induced in response to specific pathogens and/or in specific tissues or life stages. In the *Ag/Pf* data, attacin was expressed primarily in the carcass. *DEF4* and *CEC4* are enriched the gastric caeca of the final instar larvae (Neira Oviedo et al, 2008), but expression peaks overall in L1–2 and L3–4 larvae respectively (Papa et al, 2017). Expression of *DEF2* was previously found to be approximately 80-fold higher in L2–3 larvae than adults (Papa et al, 2017). The remaining two defensins (3 and 5) also appear to be expressed in a larvae-specific manner. These discrepancies might reflect the

differences in habitat and therefore different pathogens that interact with mosquito individuals at different life stages.

The prominent gradient of AMPs produced by the cardia and digestive enzymes by the posterior midgut creates a stark division of functionality across the midgut axis, highlighting the compartmentalization of the midgut. The more anterior regions show enhanced immunocompetence whereas the posterior regions show tolerance. This is traded off with digestion, where the posterior has a greater ability to digest blood meals. The anterior regions show an intermediate phenotype with similar investment into immunity and blood meal digestion, in addition to being better equipped to digest/absorb sugar meals. These results are not only consistent with findings from previous studies that examined the midgut tissue transcriptomes under sugar-fed conditions (Warr et al, 2007; Hixson et al, 2022), but also indicate that the functions of these tissues remain clearly defined post-blood meal. One key difference between the investment into immunity and digestion by the cardia and posterior midgut respectively is that investment into AMP production is extremely constitutive. Although digestive enzymes are also expressed throughout the time points pre- and post-blood meal, investment is in part reliant on a blood meal as their expression increased post-blood feeding. This might be due to the fact that immune functions are necessary throughout the mosquito lifespan, as all food sources need to be sterilized. Blood meal digestion events are only required upon taking a blood meal.

This dataset enabled the evaluation of the spatiotemporal expression of every gene in the midgut post-blood meal, and will be a valuable tool for characterizing the mosquito midgut. Using this information, new targets for transgenesis can also be selected to enable the specific spatiotemporal modulation or expression of other effectors, which can have major impacts on malaria parasite development and transmission. For example, highly-specific nanobodies targeting gametocyte surface proteins could be introduced into the locus of a gene which is abundantly expressed in the cardia at early time points post-infection. The major bottleneck observed in the gametocyte to oocyst stages of *Plasmodium* are attributed to midgut immune responses (Cirimotich et al, 2010; Sinden, 2010), perhaps more specifically due to the bombardment of AMPs in the anterior regions. Thus, more potent anti-*Plasmodium* immune effectors could be introduced or

modulated to have a similar expression profile as the AMPs through the use of endogenous AMP promoters. Alternatively, given this expression profile, there is high selection pressure exerted on the malaria parasite for adaptations to protect against this. An alternative strategy could be target the parasite in the posterior midgut to subvert these putative adaptations. These two strategies were tested by Hoermann et al, where exogenous and potent anti-*Plasmodium* AMPs were introduced into the *GAM1* (GMM line) and *CP* loci to enable enhanced front- and back-end AMP expression in the midgut respectively (2022). Both strategies proved to be successful, as oocysts were significantly reduced and parasite development was impaired. The enhanced back-end AMP expression was a more successful strategy at least until oocyst development. This impairment of oocyst development translated to a reduced number of sporozoites. These transcriptomic data will enable additional loci and effectors to be identified for the generation of more transgenic mosquito lines, as genes/effectors with a desired expression profile can be selected.

Chapter 4: Results II Midgut Homeostasis and the Cardia

4.1 Introduction

4.1.1 Abstract

The cardia is the anterior-most region of the midgut and is known to secrete AMPs. It has been proposed that these AMPs are blended into food sources to kill ingested pathogens. Transcriptomic data of the cardia has shown that it is highly specialized in function, and few genes contribute to the majority of its secreted transcriptome, of which AMPs make up nearly half. The high, constitutive expression of AMPs pre- and post-blood meal in the cardia is underpinned by a concentration of DVE in a manner reminiscent of embryonic developmental patterning, which marks another example of a developmental transcription factor co-opted by the immune system. These AMPs enable the sterilization of incoming food sources without impacting the microbiota harbored in the posterior midgut. Additionally, the cardia contributes to other biological processes occurring the posterior midgut. Inducible expression of AMPs is conferred by activation of the IMD/REL2 pathway. These findings solidify the immunocompetent role of the cardia and stark dichotomy of functions across the anterior-posterior axis of the mosquito midgut.

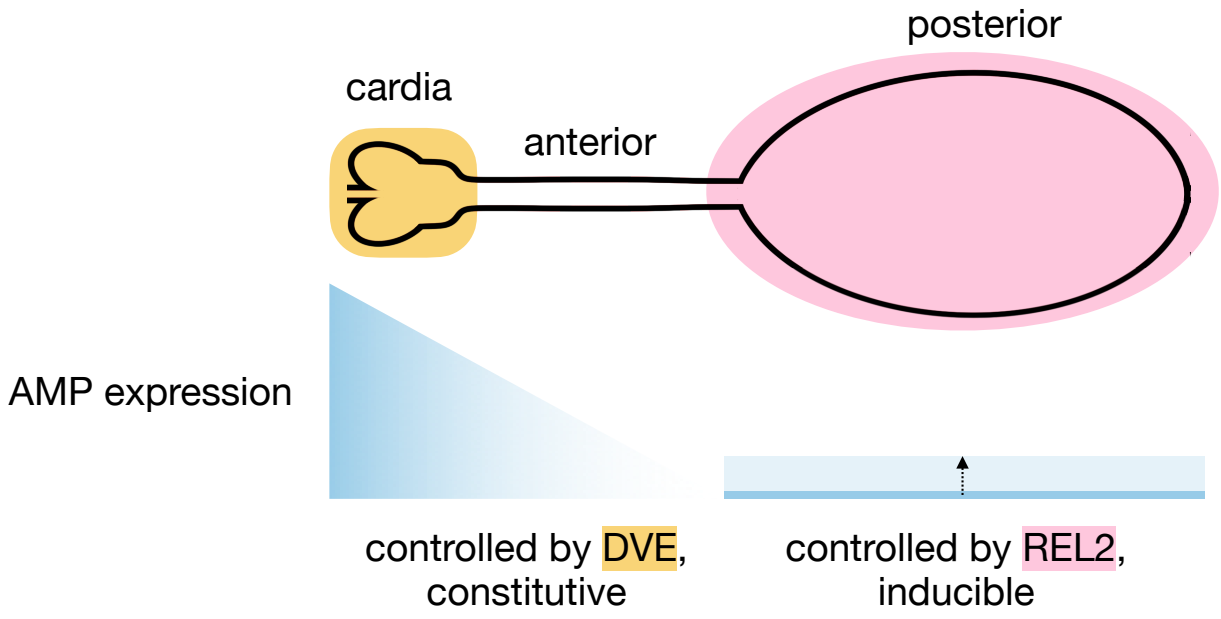


Figure 4.1.1. Graphical abstract: model of AMP expression in the midgut. *High constitutive expression of AMPs is controlled by transcription factor DVE in the cardia. In the posterior midgut, AMP expression is low, but is induced upon infection by transcription factor REL2.*

4.1.2 Background

The cardia, situated at the junction between the foregut and midgut tissues, is a bulbous structure that functions as a valve that controls the diet entering the midgut. Over the last few decades, other functions have also been attributed to insect cardias. In *Drosophila*, a pool of multipotent stem cells are present in the cardia which give rise to progenitor cells that differentiate upon migration (Singh et al, 2011), highlighting the role of the cardia in maintaining the midgut. In *An. gambiae*, the cardia is known to secrete an abundance of antimicrobial peptides (AMPs), which are hypothesized to be blended into a sugar/blood meal to protect against orally ingested pathogens (Dong et al, 2006; Warr et al, 2007; Neira Oviedo et al, 2008; Rodgers et al 2020; Hixson et al, 2022). This parallels tsetse flies, another hematophagous insect which transmits trypanosome parasites, where the cardia is speculated to have an important role in immunity (Hao, Kasumba & Aksoy, 2003).

In *Drosophila*, cardia development begins at the boundary of the ectodermal (foregut) and endodermal (midgut) tissues in embryos. The ectodermal cells evaginate to form a keyhole structure devoid of the mesoderm, where the anterior cells of the keyhole then migrate towards the endodermal tissue to create a triple-layered heart-shape structure (Fig. 4.1.2; Fuss et al, 2004). Cell migration and keyhole formation is reliant on the Notch and JAK/STAT signaling pathways (Fuss et al, 2004; Josten et al, 2004). Transcription factors forkhead (Fkh) and defective proventriculus (Dve) drive the development of ectodermal and endodermal tissues respectively (Weigel et al, 1989; Nakagoshi et al, 1998). In the absence of *fkh*, the anterior domains of mutant embryos lack an esophagus and cardia, but retain a pharynx (Jürgens & Weigel, 1988). Dve is necessary for the morphogenesis and function of the cardia, as *dve* mutants die within a day of hatching and heterozygous mutants have malformed and dysfunctional cardias (Nakagoshi et al, 1998). Additionally, expression of *dve* persists into adulthood in the cardia tissue of endodermal origin (Nakagawa et al, 2011; Singh et al, 2011). Taken together, these results indicate that Dve provides the cardia tissue its identity.

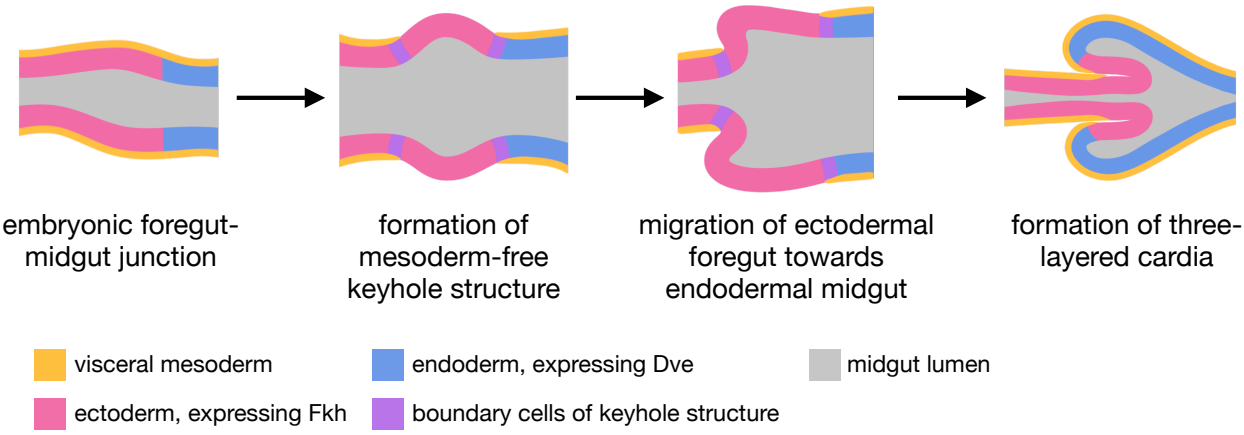


Figure 4.1.2. Cardia morphogenesis in the* Drosophila *embryo. *Images depict cross-sections of the foregut-midgut junction. Adapted from Nakagoshi, 2005.*

Here, the underlying mechanisms and purpose of high AMP expression in the cardia were explored. While AMPs are enriched in the cardia, the transcription factors downstream of TOLL and IMD immune pathway activation did not share the same expression profile. To date, AMP expression has mostly been characterized to be under the control of immune pathways (Meister et al, 2005; Luna et al, 2006). Therefore, it was hypothesized that other factors must be involved in controlling AMP expression. Given the role of Dve in *Drosophila* cardia development, the role of DVE on AMP expression was investigated here through the use of RNAi-mediated silencing. Additionally, the induced expression of AMPs following an immune challenge was characterized in more detail, and various visualization methodologies were developed/implemented to confirm the transcriptomic data.

4.1.3 Methods

In order to profile the secreted transcriptomes of midgut tissues and to enable comparisons between genes and samples, RNA-seq reads of the midgut tissues data (sections 2.3.1 and 3.4) were converted into TPMs using Stringtie (Pertea et al, 2015) on the Galaxy webserver (Afgan et al, 2016). Transcripts which made up the top two-thirds of the secreted transcriptomes were considered, and GO term enrichment analysis was employed on these genes to further characterize the functions of tissues at different time points.

Given the high expression of AMPs in the cardia, experiments were performed to determine the molecular mechanism behind this expression profile. RNAi-mediated silencing was employed here to elucidate and assess the roles of specific genes. The synthesis and intra-thoracic injection of dsRNA molecules into adults enable the systemic and temporary silencing of genes over the course of a couple of days (section 2.9). The silencing efficiency and relative expression of other genes of interest (*e.g.* AMPs, bacterial *16S*) was calculated through the dissection of tissues, extraction of RNA, conversion into cDNA, and quantification with qRT-PCR as detailed in sections 2.3.1, 2.4, 2.5, and 2.9.3.

Bacteria species *E. coli* and *S. marcescens* were utilized as different models of bacterial infection. *S. marcescens* was introduced via sugar feeding into the midgut as this species is competent in colonizing the midgut and has been characterized as part of the natural microbiota (Bando et al, 2013; Bahia et al, 2014; Stathopoulos et al, 2014). Contrastingly, *E. coli* was injected intrathoracically as this species is pathogenic and unable to establish itself within the midgut (Riehle et al, 2007). The culturing of bacteria and infection of mosquitoes is outlined in sections 2.10.1 and 2.10.2.

Investigation into the role of an exogenous metabolite (MF) on midgut AMP expression and bacterial growth was performed as described in sections 2.3.1 and 2.10.5 respectively. Downstream quantification of AMP expression was performed as above. MF was mixed into water instead of sugar as water is believed to directly enter the midgut.

Results obtained by molecular means were visualized through the use of fluorescently-labelled molecules to confirm the transcriptomic data. A whole-mount RNA FISH protocol was developed and implemented on *Anopheles* midguts for the first time to observe the expression of cardia-enriched genes (sections 2.12 and 4.3), and IFAs were performed in a transgenic mosquito line post-*DVE* silencing to observe the change in AMP expression (section 2.11).

4.2 Few genes make up the majority of the secreted transcriptome of the cardia

Considering the significant investment into AMP expression in the cardia and proteolytic enzymes in the posterior midgut, the secreted transcriptomes of each tissue at each time point were assessed. In order to enable comparisons between samples, the transcriptomic data were converted into TPMs. From the TPMs, genes that were annotated to be secreted were extrapolated, and TPMs of secreted genes were converted into percentages of the secreted transcriptome. Genes contributing to the top two-thirds (≥ 66.7%) of the predicted secreted transcriptome were considered. In the cardia at 0 hours, the top two-thirds of the secreted transcriptome comprised of just 5 genes (Table C.1). At 5 and 24 hours, this increased to 6 and 11 genes respectively and included the original 5 genes from 0 hours (Table C.1). The contribution of each of the 11 genes towards the secreted transcriptomes of the cardia at all time points are portrayed in Figure 4.2.1.

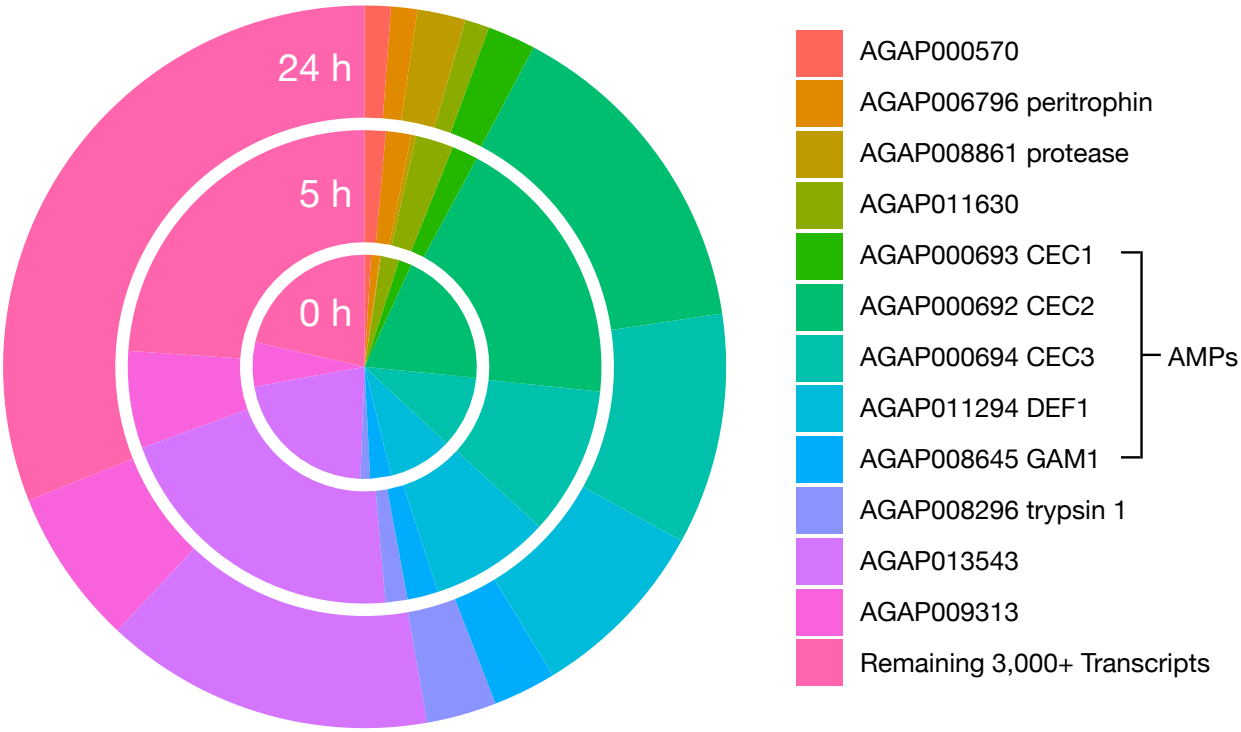


Figure 4.2.1. Secreted transcriptomes of the cardia. *Transcripts which make up the top two-thirds of the secreted transcriptome at any time point were included in all time points. Inner circle = 0 hours, middle ring = 5 hours, and outer ring = 24 hours post-blood feeding. Average TPMs were calculated from three replicates of the midgut sections transcriptomic data and converted into percentages.*

A total of 5 AMPs contributed to the top two-thirds of the cardia transcriptome. The overall investment into these 5 AMPs was nearly identical to that of all AMPs, indicating the remaining 6 AMPs are hardly expressed in the midgut. Additionally, two unspecified products also contributed significantly to the secreted transcriptome of the cardia. AGAP009313 represented nearly 7% of the transcriptome at all time points. This uncharacterized gene is annotated as sharing homology with tsetseEP, a protein from the tsetse fly *Glossina morsitans* containing a series of glutamic acid (E) and proline (P) repeats at the C-terminus. Additionally, AGAP009313 was the sixth most abundant protein in the PM proteome (Dinglasan et al, 2009). AGAP013543 was the most abundant transcript at 0 and 5 hours post-blood feeding, representing over 20% of the entire secreted transcriptome. At 24 hours, expression was downregulated to 14% of the transcriptome. The presence of 3 blood meal digestion-related proteins were also observed in the cardia. At all of the time points, peritrophin (AGAP006796), a component of the PM, represented over 1% of the secreted transcriptome. Similarly, trypsin 1 (AGAP008296) and a protease (AGAP008861) contributed towards ~1% and 0.2% of the secreted transcriptome respectively at 0 and 5 hours. At 24 hours, investment towards the expression of these two genes increased to 3% and 2% respectively, indicating a role of the cardia in the formation of the peritrophic matrix that is restricted to the posterior midgut and aiding blood meal digestion. This parallels the observation that investment towards peptidases and trypsins doubles in the posterior at 24 hours post-blood meal, although at a much larger scale than in the cardia.

In the anterior, 24, 23, and 19 genes made up the top two-thirds of the secreted transcriptome at 0, 5, and 24 hours post-blood feeding respectively (Table C.1). Three proteins related to sugar digestion were identified in the anterior across the three time points, in addition to 6 digestive enzymes. Similarly, 22, 26, and 15 genes were identified in the posterior at each time point (Table C.1). Many peptidases and trypsins were also identified, and the reduction in the number of genes making up the top two-thirds at 24 hours was a result of increased investment into the expression of these enzymes. For example, trypsin 1 and chymotrypsin 1 (AGAP006709) made up ~5% and < 2% of the transcriptome at 0 and 5 hours post-blood meal respectively. At 24 hours, expression increased to ~8% and ~4% of the secreted transcriptome respectively.

There were two genes which were present in the top two-thirds of the secreted transcriptome of all three tissues in at least one time point. Trypsin 1 was present in the anterior and posterior at all time points, compared to the cardia at just one time point (24 hours). Investment into trypsin 1 expression by the cardia increased over time and peaked at 24 hours. An uncharacterized, MBF2 transcription factor domain-containing gene AGAP000570 followed a similar pattern, however its expression in the cardia was relatively unchanged throughout the time points. Investment into this gene peaked in the anterior, where it made up approximately 6–9% of the secreted transcriptome. In the posterior, expression of this gene decreased from 6% to 1% of the secreted transcriptome across the time points. The product of AGAP000570 was putatively named secreted glycoconjugate of unknown function (SGU) (Mathias et al, 2014), and was found to be the second most abundant protein in the PM (Dinglasan et al, 2009). Delivery of antibodies targeting SGU blocks *P. falciparum* oocyst development, but has no effect on field isolates of *P. vivax* (Mathias et al, 2014). Interestingly, SGU also interacts with P47Rec (AGAP006398), which interacts with *Plasmodium* P47 and is thought to mediate the suppression of immune responses that target the parasite (Molina-Cruz et al, 2020; Singh, Kirar & Sehrawat, 2022). In the midgut sections data, expression of SGU was abundant and peaked in the anterior at all time points, while P47Rec peaked in the posterior and was generally downregulated in the entire midgut post-blood meal.

GO term enrichment analysis was performed on the genes making up the top two-thirds of each transcriptome. Genes annotated with a signal peptide were used as a background to more accurately identify enriched GO terms. The enriched GO terms confirmed the results so far, where immunity-related terms were exclusively enriched in the cardia at all time points, and a single term (‘digestion’ [GO:0007586]) was enriched from the posterior (Table C.2). No terms were identified for the anterior midgut at any time point. The CECs and DEF1 led to the enrichment of immunity terms in all cases. The lack of GAM1 also being present in the intersections of genes belonging to the top two-thirds of the transcriptome and the immunity GO term highlights the incomplete nature of the GO terms. Similarly, 4–5 digestive enzymes (trypsins, chymotrypsin, and CP) led to the enrichment of the term ‘digestion’, despite not being the only enzymes related to digestion.

4.3 Developing a whole-mount midgut RNA FISH protocol

Considering the cardia-specific high expression of some genes, RNA FISH was performed to visualize the spatial expression of these mRNA molecules. CEC2, AGAP0013543, and AGAP009313 were selected. CEC2 was selected because it was the most abundantly expressed AMP in the cardia. The latter two were selected as they are uncharacterized, but also extremely abundant in the cardia. To my knowledge, RNA FISH has never been performed on whole-mount mosquito midgut samples before. Thus, a protocol needed to be developed to enable the visualization of the expression of these genes.

Existing protocols for whole-mount RNA FISH in *Drosophila* embryos (Kosman et al, 2004) and *Anopheles* ovaries and salivary glands (Juhn & James, 2012) were used as a foundation to adapt a protocol to *Anopheles* midguts. The main steps of RNA FISH include: 1) probe generation, 2) pre-hybridization, 3) hybridization, 4) post-hybridization, and 5) sample imaging. In many cases, RNA probes are synthesized with modified UTPs, and secondary antibodies conjugated with a fluorophore which target this modification are used to enable the indirect detection of mRNA molecules. Using the FISH Tag™ RNA Red kit, the need for secondary antibodies with fluorophores is removed as the uracil bases have amine modifications and the dye provided is amine-reactive. Therefore, the probes are directly labelled and can be detected by fluorescence microscopy.

Sense and antisense probe templates were generated via PCR amplification of cDNA derived from mosquito midguts. These templates were sequenced using GeneWiz Ltd. services to ensure the correct sequence. Using the FISH Tag™ RNA Red kit, RNA probes were synthesized from the DNA templates and labelled with a dye according to the manufacturer's instructions. As the designed probe lengths were relatively short at approximately 100–300 bp for the three genes, some modifications were made to the manufacturer's protocol. All ethanol precipitation steps were performed overnight instead of half an hour. Additionally, RNA synthesis reactions were also performed overnight to maximize the RNA yield. Following probe synthesis, a total of three attempts were made to optimize the FISH protocol.

For the first attempt, 50 midguts were dissected in ice-cold PBS and split into two groups for the sense and antisense probes of *CEC2*. Samples were fixed in 4%

paraformaldehyde for 45 minutes as this is standard practice for preparation of tissues for IFAs. Following fixation, samples were washed with PBT. Samples were then treated with proteinase K (Thermo Fisher Scientific, UK) to increase cell permeability, and then equilibrated into the hybridization buffer. Hybridization buffers were made as described in Kosman et al (2004). Probes were added to the hybridization buffer at a concentration of 1 ng/µL as previously described (Kosman et al, 2004). Hybridization took place overnight, followed by equilibration back into PBT. Nuclei were stained with Hoechst and mounted on slides with VECTASHIELD®. Apart from stained nuclei, no signal was observed whatsoever from these samples. This was hypothesized to be a result of low probe concentration and too many midguts within one reaction tube.

For the second attempt, each reaction was reduced to just 5 midguts, and the probe concentration was tripled. Upon imaging, nuclei (blue) were observed throughout the entire midgut of all samples, and probe signal (red) was observed in the cardias of antisense probe-treated samples. However, a strong red signal in the distal half of the posterior midgut was also observed. This was hypothesized to be a result of residual probes not washed out of the midgut, as the RNA-seq data did not indicate any significant expression of *CEC2* in the posterior midgut. To address this, the next attempt was performed with RNAse A, an RNAse which specifically cleaves single-stranded RNA, *i.e.* probes that have not hybridized to midgut mRNA. In the final optimization attempt, 6–7 midguts were used in each reaction, and probe concentrations were kept between 2 to 3 ng/µL. The proteinase K treatment step was removed, as Tween® 20 in the PBT was deemed sufficient enough to enable permeabilization of the midgut, particularly as the probe used was very small (~100 bp). In this attempt, imaging of the midguts showed a phenotype which was identical to that of molecular data. The fluorescent probe signal was strictly concentrated in the cardias of antisense probe-treated samples, whilst nuclei were observed throughout the entire midgut (Fig. 4.3.1). Therefore, a whole-mount RNA FISH protocol was successfully developed for mosquito midguts.

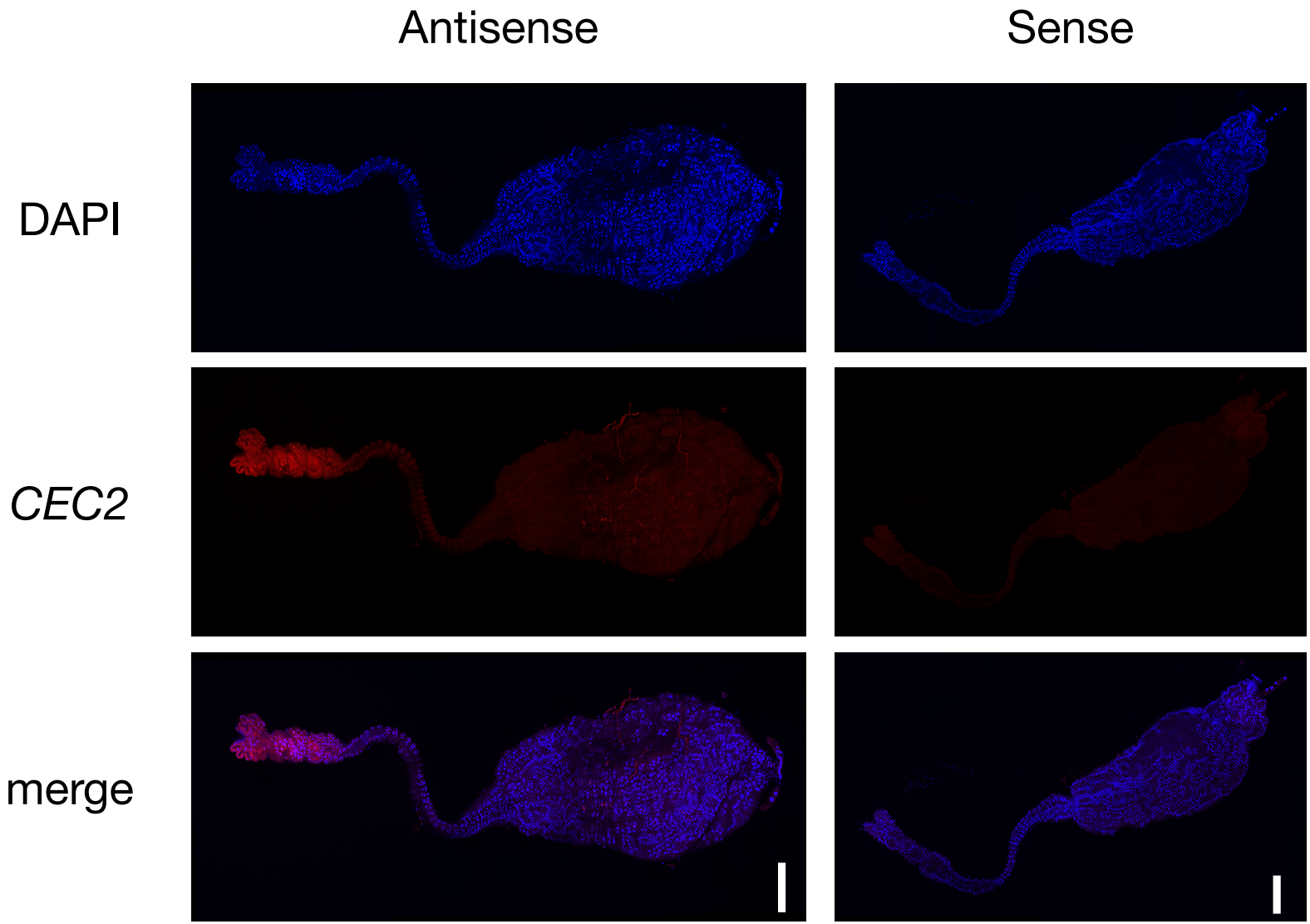


Figure 4.3.1. Confocal imaging of* CEC2 *and nuclei in the mosquito midgut. *Nuclei were stained with DAPI and are represented in blue. Alexa Fluor™ 594-labeled RNA probes (sense and antisense strands separately) were hybridized to* CEC2 *mRNA and are represented in red. Scale bar represents 200 µm.*

This optimized protocol was repeated for the two uncharacterized genes abundant in the cardia, AGAP009313 and AGAP013543. For both genes, nuclei and the antisense probe signals were detected throughout the midgut and in the cardia respectively (Fig. C.1; Fig. C.2). No probe signal was detected with the sense probes (Fig. C.1; Fig. C.2). Interestingly, the fluorescence signal in the cardia was not identical to that of *CEC2*; expression of these genes appeared to be more restricted to the proximal, bulbous regions of the cardia (Fig. 4.3.2).

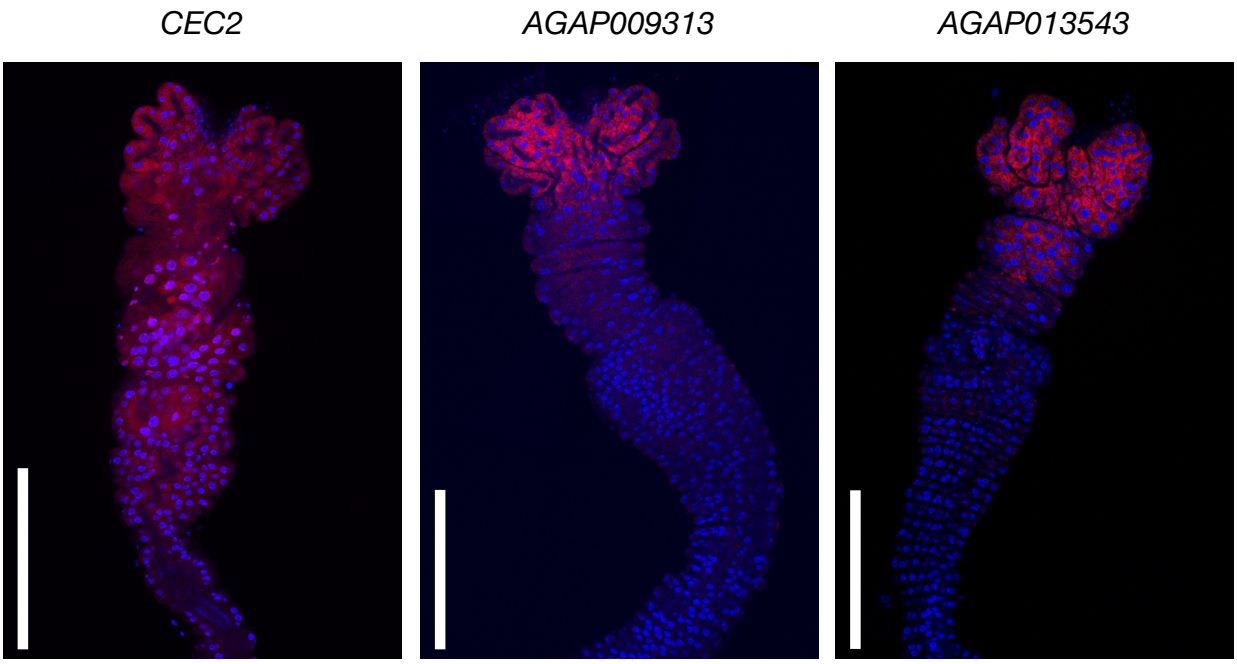


Figure 4.3.2. Spatial visualization of highly expressed genes in the cardia. *Confocal imaging of Alexa Fluor™-labeled antisense RNA probes and nuclei in the cardia. Nuclei were stained with DAPI and are represented in blue. Alexa Fluor™ 594-labeled antisense RNA probes were hybridized to mRNA and are represented in red. Scale bar represents 200 µm.*

Confocal imaging of the cardia with these probes also enabled the observation of many crypts present within the tissue (Fig. 4.3.3). These crypts resemble those reported in the small intestine of other animals and perhaps represents a structural architecture to increase the surface area to volume ratio. This organization of the cardia tissue might then enable an increased number of cells to produce an increased amount of transcripts/proteins such as the AMPs. Using the Alexa Fluor™ 488 phalloidin stain, the actin molecules on the surface of the midguts were able to be visualized. This revealed that the surface of the cardia is not smooth, but rather lined by invaginations.

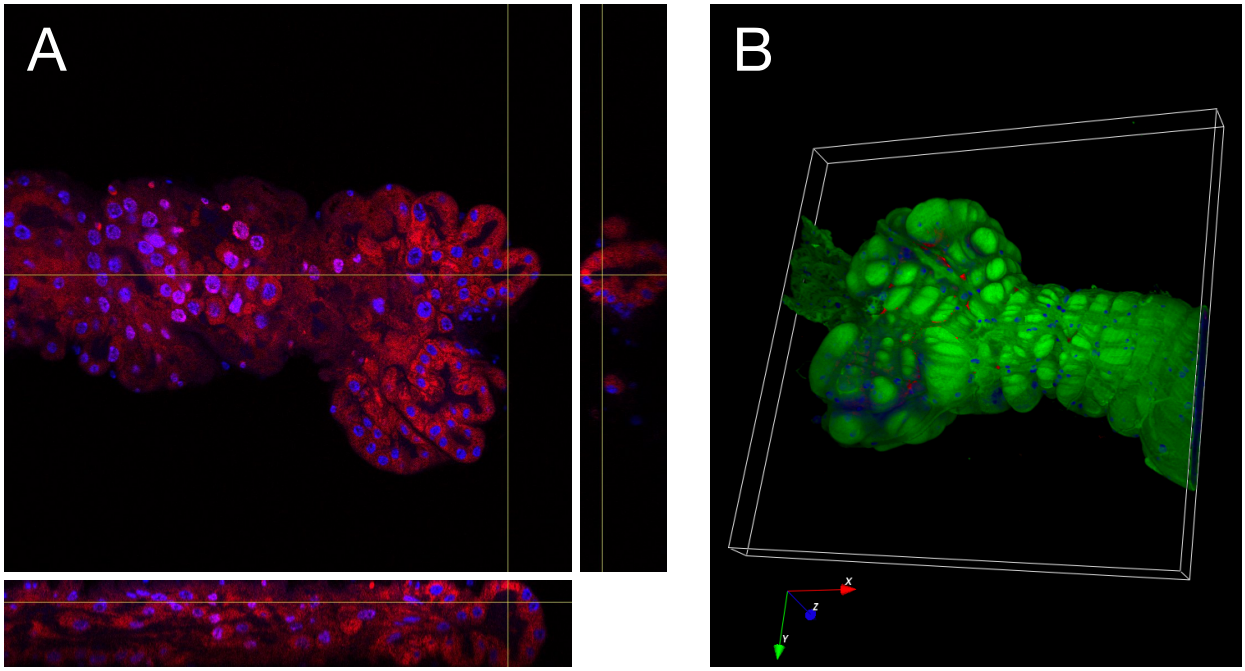


Figure 4.3.3. 3D image stacks of the cardia. *Confocal imaging of Alexa Fluor™-labeled antisense* CEC2 *RNA probes, Alexa Fluor™-labelled actin, and nuclei in the cardia. A) Cross section of the cardia. Bottom and right panels represent orthogonal views of the cardia cross-section. B) 3D-rendering of the cardia. Nuclei were stained with DAPI and are represented in blue. Alexa Fluor™ 594-labeled antisense* CEC2 *RNA probes were hybridized to mRNA and are represented in red. Actin was stained with Alexa Fluor™ 488 phalloidin and is represented in green.*

4.4 DVE controls the high AMP expression in the cardia

AMPs have so far been characterized to be under the control of the TOLL and IMD pathways. The high and constitutive expression of AMPs in the cardia, coupled with the relatively stable yet low expression of REL1/REL2 in the midgut suggest a basal, housekeeping mechanism that maintains high AMP expression in the cardia and anterior-most part of the anterior midgut, in addition to inducible immune expression. This AMP expression dichotomy could be a result of two possible explanations: 1) a default high and constitutive mechanism of AMP expression throughout the midgut and concomitant suppression in the posterior midgut, or 2) the expression of a transcription factor specific to the cardia and less so in the anterior midgut causing high AMP expression in these tissues.

To address the former possibility, the KD of putative REL2 suppressors was performed in a REL2-overexpressed line (CP-REL2) generated by Dr. Astrid Hoermann (unpublished data). In this transgenic mosquito line, *REL2* was fused with the *CP* promoter, leading to increased expression of *REL2* in the posterior midgut post-blood meal. RNA-seq experiments confirmed this increased *REL2* expression, however AMP expression was not impacted in the posterior midgut post-blood meal (A. Hoermann, unpublished data). This suggests a mechanism or factor suppressing AMP expression in the posterior. Considering REL2 is part of the IMD pathway, negative regulator CAD and two uncharacterized CACT-like genes were silenced to test this hypothesis. The CACT-like genes (AGAP004935 and AGAP005326) contain an ankyrin-rich repeat domain, which is reminiscent of that of *CACT*, the IκB inhibitor of the TOLL pathway NF-κB transcription factor REL1. Both *CACT*-like genes also do not contain signal peptides or transmembrane domains, indicating intracellular localization. Based on this shared domain, these two genes were tested here as putative pathway regulators. The hypothesis was that CACT-like proteins may inhibit the short form of REL2, which lacks an IκB domain (Meister et al, 2005). CACT is a known negative regulator of the TOLL (REL1) pathway that was never reported to regulate the IMD pathway, and hence was not considered in this assay. In the midgut sections data, AGAP004935 was part of Cluster 4, meaning expression peaked in the anterior at 5 hours post-blood feed. Prior to blood feeding, expression was enriched in the posterior. At 24 hours post-blood

feeding, expression was downregulated but relatively stable across the midgut tissues. AGAP005326 expression peaked in the cardia prior to blood feeding, but was downregulated to a similar expression level as the other tissues at 5 hours. At 24 hours post-blood feeding, expression in all tissues was upregulated.

Each of these genes were individually silenced in the CP-REL2 transgenic mosquitoes, totaling 3 biological replicates. As AMP expression was not increased in the posterior midgut of this transgenic line despite the enhanced expression of *REL2*, AMP expression (*CEC1–2*, *DEF1*, and *GAM1*) was assayed in the posterior midgut post-KD as a reporter. Overall, the effect of KD on AMP expression was mixed for these genes (Fig. 4.4.1). The average *CAD* KD efficiency was approximately 44% in the midgut, and *DEF1* was the only AMP where expression was consistently affected in the same manner (*i.e.* expression was reduced post-KD in all replicates). Expression of the remaining AMPs varied across the replicates compared to the posterior midgut of d*sLacZ*-injected controls. Silencing of the two CACT-like genes also yielded mixed results, with the exception of reduced *GAM1* expression post-AGAP005326 silencing (Fig. 4.4.1). The average KD efficiency of each *CACT-like* gene in the midgut was approximately 70%. Expression of all AMPs in each condition were not significantly different compared to that in controls (paired Student's *t*-tests and Wilcoxon signed-rank tests, $p > 0.05$). Therefore, these genes may play some role in contributing to the suppression of AMPs in the posterior midgut of the CP-REL2 line, but it is unlikely to be major.

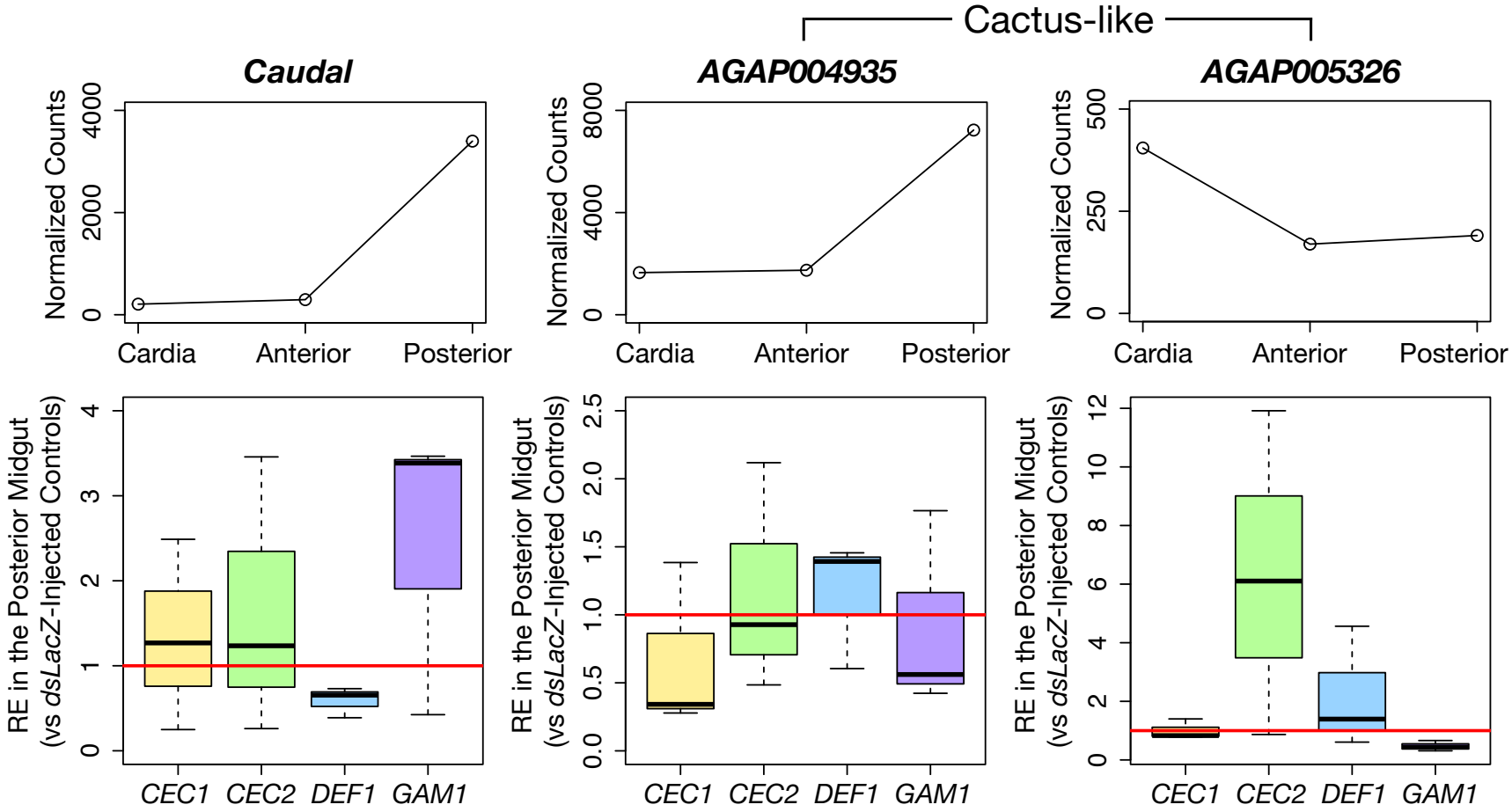


Figure 4.4.1. Midgut expression profiles of (putative) IMD pathway suppressors (WT) and relative expression of AMPs post-KD in the posterior midguts of CP-REL2 transgenic mosquitoes. *Normalized counts were calculated and averaged from three replicates of the midgut sections transcriptomic data under sugar feeding conditions (0 hours) with DESeq2. Relative expression values were normalized to that of* dsLacZ*-injected controls, represented by the red line. Three replicates were performed. RE = relative expression. No significant differences were observed between expression post-KD and in* dsLacZ*-injected controls (Wilcoxon signed-rank tests,* p > *0.05).*

In order to address the latter possibility, potential transcription factors controlling the high expression of AMPs in the cardia were first identified. Expression of AMPs exists as a gradient, which is reminiscent of embryo developmental networks, where concentration gradients of various transcription factors control the identity and differentiation of the developing tissues. Therefore, some developmental transcription factors were selected to test this hypothesis. Considering the role of Dve in *Drosophila* cardia development (Weigel et al, 1989; Nakagoshi et al, 1998; Nakagoshi, 2005), expression of *DVE* (AGAP029578) in the *Anopheles* midgut was examined. Indeed, expression of this homeobox domain-containing transcription factor was enriched in the cardia at all time points. Two other genes involved in patterning of specific structures during development were also considered. In *Drosophila*, *araucan* (*ara*), *mirror* (*mirr*), and *caupolican* belong to the Iroquois (Iro) family of genes and form the Iroquois complex (Iro-C) (Gómez-Skarmeta et al, 1996; McNeill et al, 1997). These transcription factors also contain homeobox domains and are heavily implicated in the development of the *Drosophila* eye, head, and mesothorax (Cavodeassi, Modolelle & Gómez-Skarmeta, 2001). *ARA* (AGAP005010) and *MIRR* (AGAP005011) have been identified in *Anopheles*, and both also share homology to *Drosophila caupolican*. A 1:1 homolog of *caupolican* does not exist in *Anopheles*. In the midgut, expression of *ARA* and *MIRR* peak in the cardia and anterior respectively. Finally, a fourth gene was also selected: AGAP009771. This gene contains a zinc finger domain, indicating a putative DNA binding role. Expression of this gene was at least 500-fold higher in the cardia compared to the posterior. These four transcription factors were silenced separately in WT (N'gousso strain) mosquitoes, and AMP expression was measured in the cardia/anterior as these regions contribute to the majority of AMP expression. A total of 4 replicates were performed for all genes except AGAP009771, with 9 replicates.

Silencing of the Iro-C genes had nearly opposite effects on AMP expression in the cardia/anterior. The *ARA* depletion rate was 60%, which led to the reduction of *CEC1–2* expression compared to *dsLacZ*-injected controls, although this was significant only for *CEC2* (paired Student's *t*-test, $t = 3.5$, degrees of freedom [df] = 3, $p = 0.04$; Fig. 4.4.2). The median expression of *DEF1* was also reduced, but no clear effect was observed for *GAM1*. Following *MIRR* silencing, expression of the AMPs generally increased, with

consistent results except for *CEC2* (Fig. 4.4.2). *CEC1* expression significantly increased in the cardia/anterior compared to *dsLacZ*-injected controls (paired Student's *t*-test, $t = -3.2$, df = 3, $p = 0.05$; Fig. 4.4.2). The KD efficiency of *MIRR* in the midgut was relatively poor at approximately 10% across the replicates; improving the KD efficiency could lead to a more profound effect on AMP expression. Overall, the results suggest that *ARA* may play a role in the high expression of cecropins and *DEF1* observed in the cardia.

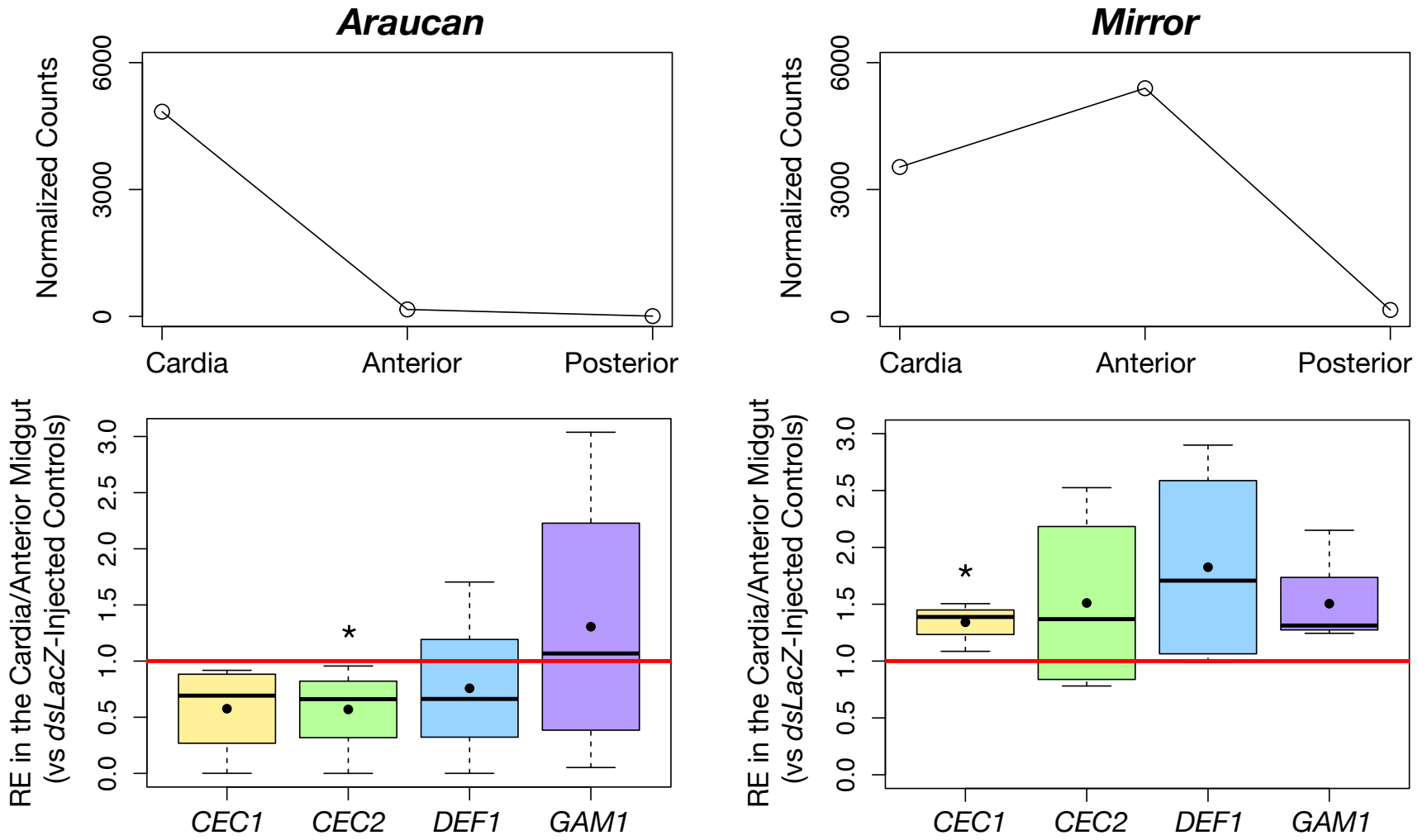


Figure 4.4.2. Midgut expression profiles of Iro-C members and relative expression of AMPs post-KD in the cardia/anterior midgut. *Normalized counts were calculated and averaged from three replicates of the midgut sections transcriptomic data under sugar feeding conditions (0 hours) with DESeq2. Relative expression values were normalized to that of* dsLacZ*-injected controls, represented by the red line. Black bars represent medians, dots represent means. Four replicates were performed. RE = relative expression. Paired student's* t*-tests between AMP expression in KD mosquitoes versus controls, * p < 0.05.*

For the remaining two gene KD conditions (AGAP009771 and *DVE*), expression of two additional genes were assayed alongside the AMPs post-KD. AGAP009313 and AGAP013543 made up a significant portion of the secreted transcriptome in the cardia and were therefore also included as reporter genes. The depletion of AGAP009771 was achieved at an average rate of 42% and had no apparent effect on any of the genes assayed (Wilcoxon signed-rank tests, $p > 0.05$; Fig. 4.4.3). Therefore, the genes under the control of AGAP009771 remain elusive.

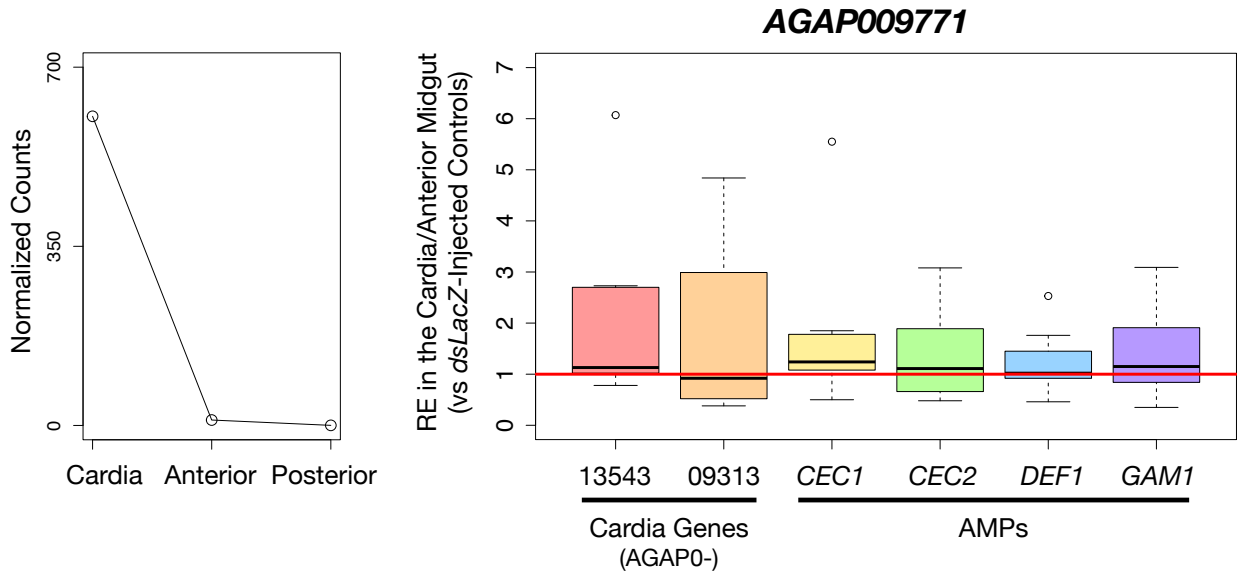


Figure 4.4.3. Midgut expression profile of transcription factor AGAP009771 and relative expression of AMPs and cardia genes post-KD in the cardia/anterior midgut. *Normalized counts were calculated and averaged from three replicates of the midgut sections transcriptomic data under sugar feeding conditions (0 hours) with DESeq2. Relative expression values were normalized to that of* dsLacZ*-injected controls, represented by the red line. Black bars represent medians. Nine replicates were performed. RE = relative expression. No significant differences were observed between expression post-KD and in* dsLacZ*-injected controls (Wilcoxon signed-rank tests,* p > *0.05).*

Remarkably, KD of *DVE* suppressed AMP expression in the cardia/anterior compared to controls in every replicate, although this reduction was significant only for *CEC1* (paired Student's *t*-test, $t = 6.2$, df = 3, $p = 0.01$). On average, *CEC1* was reduced 44%, *CEC2* 50%, *DEF1* 37%, and *GAM1* 22% in the cardia/anterior compared to controls. This reduction in AMP expression was not a result of reduced *REL2* levels, as *REL2* expression was similar between *DVE* KD and control mosquitoes (paired Student's *t*-test, $t = -0.1$, df = 3, $p = 0.9$; Fig. 4.4.4). Furthermore, the reduction in expression was comparable to the degree of *DVE* silencing, where KD efficiencies in the midgut were on average ~44% across the replicates. Expression of the other two uncharacterized genes enriched in the cardia were not affected, suggesting a different transcription factor is involved in the expression of these genes. In the posterior, AMP expression was not affected (Wilcoxon signed-rank tests, $p > 0.05$; Fig. D.1). Therefore, this data serves as preliminary evidence that DVE is specifically involved in the enrichment of AMPs in the cardia.

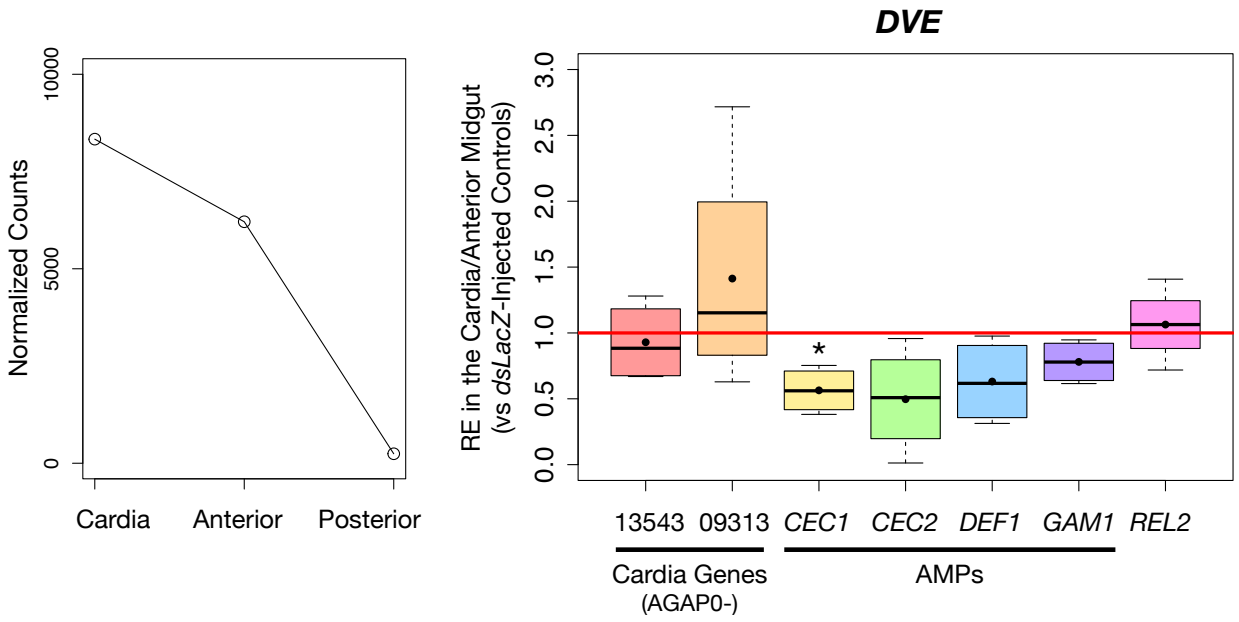


Figure 4.4.4. Midgut expression profile of* DVE *and relative expression of AMPs post-KD in the cardia/anterior midgut. *Normalized counts were calculated and averaged from three replicates of the midgut sections transcriptomic data under sugar feeding conditions (0 hours) with DESeq2. Relative expression values were normalized to that of* dsLacZ*-injected controls, represented by the red line. Black bars represent medians, dots represent means. Four replicates were performed. RE = relative expression. Paired Student's* t*-tests between AMP expression in KD mosquitoes versus controls,* * p < *0.05.*

To further validate the role of DVE in AMP expression, IFAs were performed in GMM transgenic mosquitoes post-*DVE* KD. The GMM line was generated by Hoermann et al, where two exogenous AMPs were inserted into the *GAM1* locus (2022). In between *GAM1* and the two exogenous AMPs are two 2A ribosomal skip signals to enable the generation of three separate polypeptides. Antibodies which target a 2A signal were used in this experiment as a proxy for *GAM1* expression. IFAs were also performed in WT mosquitoes (KIL/G3 strain) as the negative control. In WT mosquitoes, no fluorescent signal was observed, whereas the 2A skip signal was visible throughout the cardia in untreated GMM mosquitoes (Fig. 4.4.5; Fig. D.2). As a control for the injection process, *dsGFP* was injected into GMM mosquitoes, which produced a similar phenotype to the GMM mosquitoes without injection treatment (Fig. 4.4.5). Strikingly, the antibody signal was almost completely abolished in the *DVE* KD midguts with a phenotype reminiscent of WT *An. gambiae* (Fig. 4.4.5; Fig. D.2). Despite only achieving partial KDs when measuring number of transcripts via qRT-PCR, microscopy of these midguts suggest a much greater KD efficiency at the protein level. Therefore, the calculated KD efficiencies at the transcript level do not necessarily represent the same reduction at the protein level and must be interpreted with caution. These results indicate a key role of DVE in controlling the constitutive expression of GAM1 and, likely, other AMPs in the cardia.

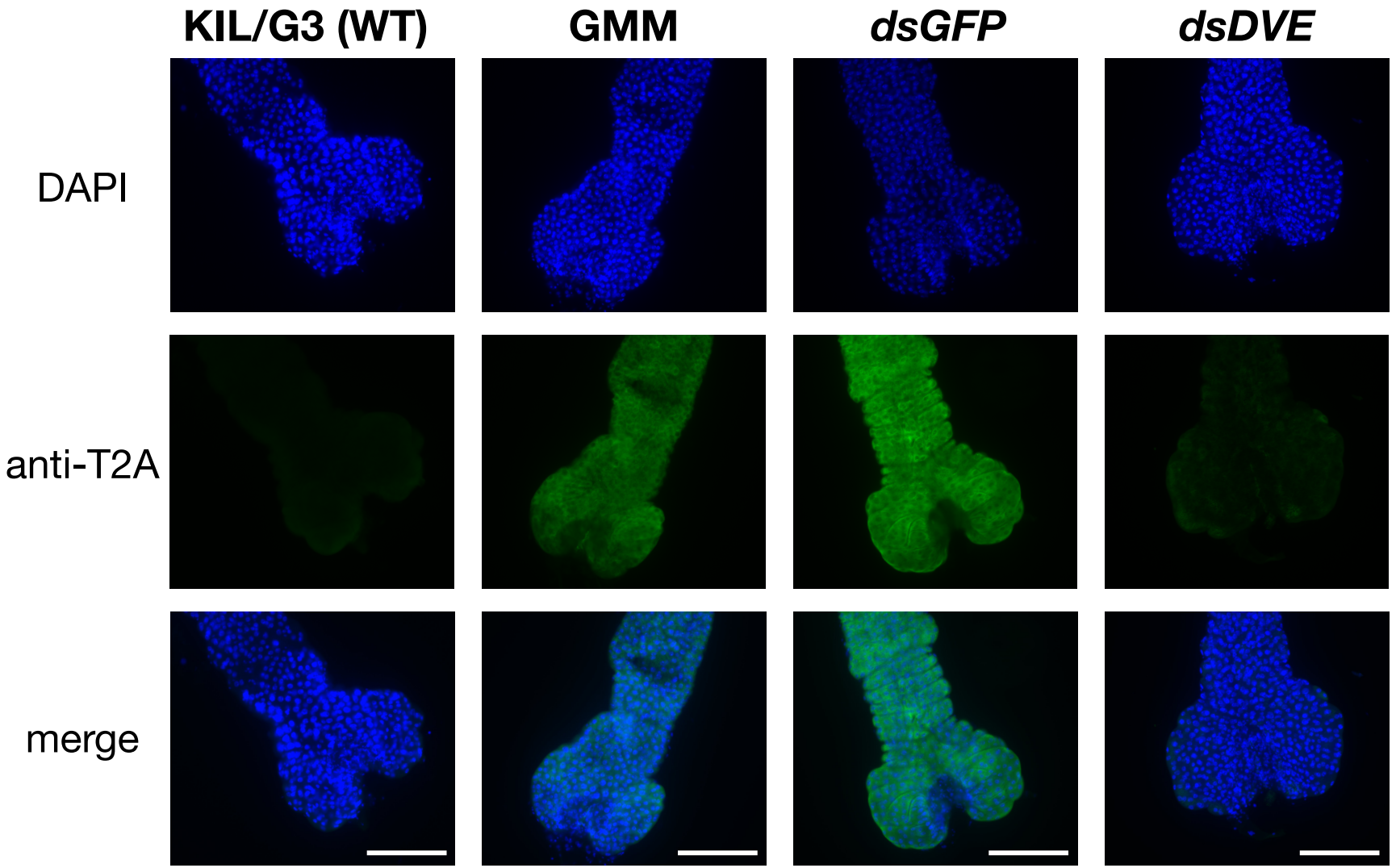


Figure 4.4.5. Fluorescent imaging of* GAM1 *expression in the cardia of GMM transgenic and WT mosquitoes. *Antibodies raised against a ribosomal skip sequence in the transgenic mosquito line were conjugated to GFP-secondary antibodies and are represented in green. Nuclei were stained with DAPI and are represented in blue. Scale bars represent 125 μm.*

4.5 AMPs launched in response to bacterial infection are *DVE*-independent

Given the role of DVE in controlling AMP expression in the cardia, further experiments were performed to identify whether *DVE* expression is modulated in response to bacterial infections. The induction of AMPs post-immune challenge has been described previously (Vizioli et al, 2000; Vizioli et al, 2001a; Vizioli et al, 2001b), but it is unclear whether this is also a result of *DVE* upregulation in the posterior in conjunction with TOLL/IMD pathway activation. Mosquitoes were infected with *S. marcescens* Db11 strain (orally) to induce a local infection in the midgut, or *E. coli* ATCC 23724 (injected thoracically) to induce a systemic infection of the hemolymph. Sugar-fed and PBS-injected mosquitoes were used as controls respectively. *S. marcescens* is commonly found in the mosquito midgut microbiota, and has the ability to colonize the midgut effectively (Bando et al, 2013; Bahia et al, 2014; Stathopoulos et al, 2014). The systemic infection of *E. coli* was included in this assay to act as a control; if *DVE* expression plays a role in the induced expression of AMPs following bacterial infection, it was expected that *DVE* expression is also inducible and can be observed in tissues where it is normally not expressed. The expression of AMPs and *DVE* was measured with qRT-PCR in whole carcasses (*E. coli*) or in midguts and the remaining carcasses (*S. marcescens*).

Following *E. coli* infection, *CEC1* and *DEF1* expression significantly increased compared to PBS-injected controls, indicating an induction of AMPs in response to infection (Student's *t*-tests, $t = 10.3$, df = 3.4, $p = 0.001$, and $t = 13.0$, df = 3.6, $p = 0.0004$ respectively; Fig. 4.5.1). The median and mean expression of *CEC2* and *GAM1* also increased in *E. coli*-infected mosquitoes, although insignificantly (Student's *t*-tests, $p > 0.05$; Fig. 4.5.1). This induction of AMP expression was independent of *DVE* expression, which was elevated albeit insignificantly in PBS-injected controls (Fig. 4.5.1). Therefore, the observed induction of AMPs was likely a result of REL2-dependent immune pathway activation rather than an upregulation of *DVE*.

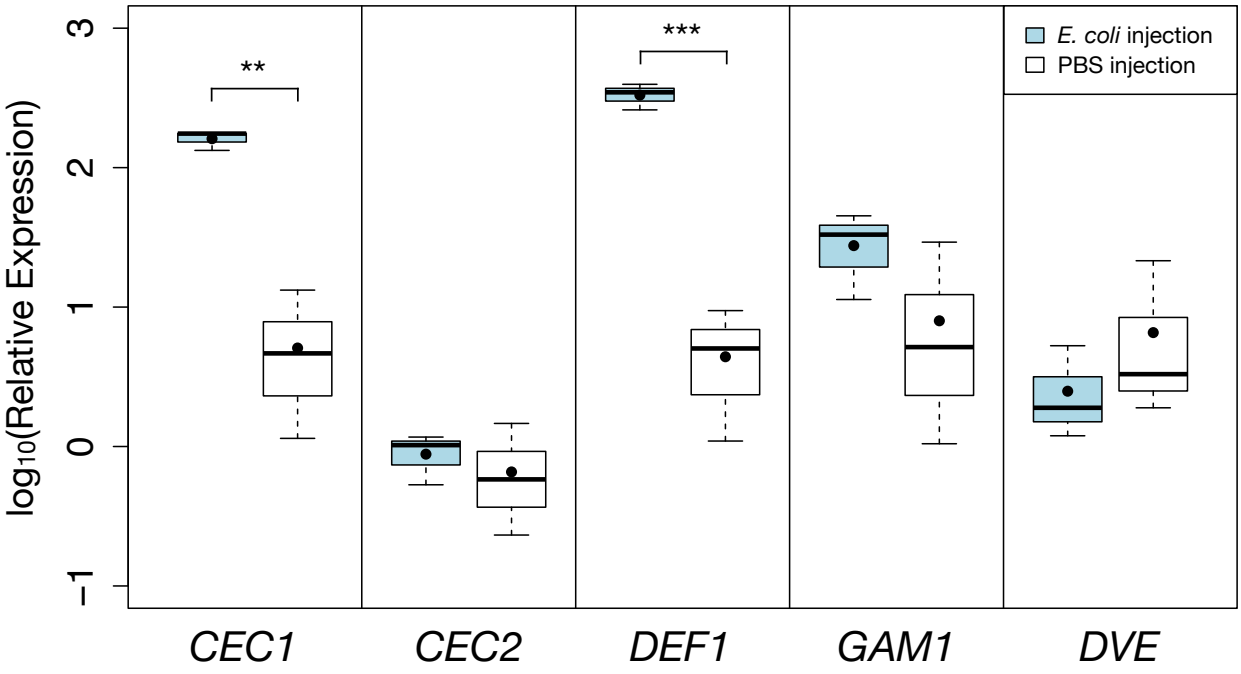


Figure 4.5.1. Relative expression of AMPs and* DVE *post-*E. coli *infection in the whole carcass. *Control mosquitoes were injected with an equal volume of sterile PBS. Relative expression values were log_{10}-transformed for presentation purposes and are relative to a pool of mosquito cDNA. Three replicates were performed using 10–15 mosquito individuals per condition for each replicate. Black bars represent medians, dots represent means. Differences in the original relative expression values were compared with Student's* t*-tests, ** $p < 0.01$, *** $p < 0.001$.*

S. marcescens infection had opposing effects on AMP expression in different tissues (Fig. 4.5.2). In the midgut, median and average AMP expression were reduced in *S. marcescens*-infected mosquitoes compared to sugar-fed controls, with *CEC2* expression reduced significantly (Student's *t*-test, $t = -3.5$, df = 7.4, $p = 0.01$). This downregulation was independent of *DVE*, as *DVE* expression was comparable between *S. marcescens*-infected and control mosquitoes. In the carcass, median expression of most AMPs increased post-infection, with *GAM1* increasing significantly (Wilcoxon rank-sum test, $W = 12$, $p = 0.05$). While median expression of *DVE* also increased in the carcass post-infection, this was insignificant (Wilcoxon rank-sum test, $W = 11$, $p = 0.37$), and *DVE* expression was found to be primarily in the midgut.

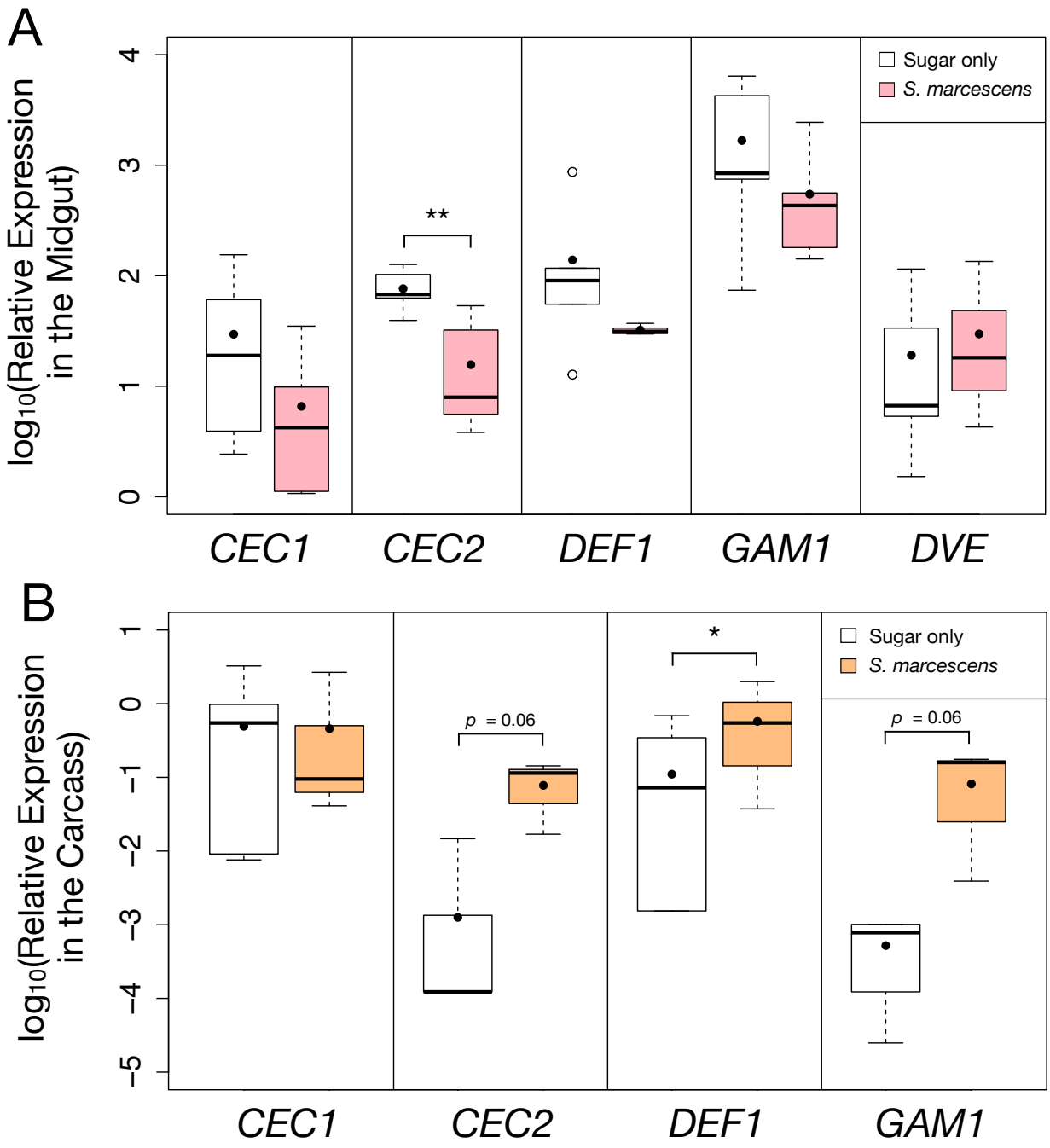


Figure 4.5.2. Relative expression of AMPs and* DVE *post-*S. marcescens *infection *in the A) midgut and B) remaining carcass. Control mosquitoes were fed a sterile sugar solution without bacteria. Relative expression values were log_{10}-transformed for presentation purposes and are relative to a pool of mosquito cDNA. Five replicates were completed for all midgut samples and control carcasses, three replicates were completed for bacteria-infected midguts. A total of 10–15 mosquito individuals were pooled for each condition. Black bars represent medians, dots represent means. Differences in the original relative expression values were compared with Student's* t*-tests or Wilcoxon rank-sum tests,* * p < *0.05,* ** p < *0.01.*

These discrepancies suggest tolerance mechanisms within the midgut so that immune responses are suppressed/unchanged in response to *S. marcescens*. As *S. marcescens* are commonly found to be associated with *Anopheles* midguts, unknown mechanisms could be in place to recognize endosymbionts to prevent overactivation of immune responses. If *S. marcescens* breach the midgut epithelium and enter the hemolymph/fat body, they are pathogenic. Therefore, the induction of AMPs in the carcass is likely in anticipation of *S. marcescens* infection, or an indication that they have spread into the hemolymph. Similarly, as *E. coli* is an exogenous/pathogenic species of bacteria to *Anopheles*, immune responses are immediately launched upon their detection. These responses are DVE-independent and likely due to IMD and/or TOLL pathway signaling.

4.6 High AMP expression maintains midgut sterility

The enrichment of AMPs in the cardia has been proposed to sterilize incoming food sources and the midgut (Dong et al, 2006; Warr et al, 2007). In order to experimentally validate this hypothesis, *DVE* was silenced in mosquitoes, and were subsequently orally infected with *S. marcescens*. Considering DVE plays a role in controlling AMP expression in the cardia, its KD will lead to the simultaneous suppression of many AMPs. Furthermore, *S. marcescens* was selected as it did not elicit AMP expression in the midgut, meaning the effect of *DVE* KD would not be cancelled out.

DVE KD and subsequent *S. marcescens* infection led to a significant increase in midgut colonization by *S. marcescens* compared to *dsLacZ*-injected controls (Wilcoxon signed-rank test, $V = 21$, $p = 0.03$; Fig. 4.6.1). In concordance with this finding, the overall microbiota load in *DVE*-silenced midguts was significantly increased (Wilcoxon signed-rank test, $V = 35$, $p = 0.02$; Fig. 4.6.1). In both cases, relative microbiota load was calculated in comparison to a pool of non-injected, uninfected mosquito midguts. Among the *dsLacZ*-injected mosquitoes, relative expression of *Serratia 16S* was at least 3-fold higher than in the unmanipulated control mosquitoes, indicating successful infection (Fig. 4.6.1). The relative loads of *S. marcescens* were therefore further increased in the *dsDVE*-injected mosquitoes. Thus, the high expression of AMPs plays a role in the sterilizing food sources.

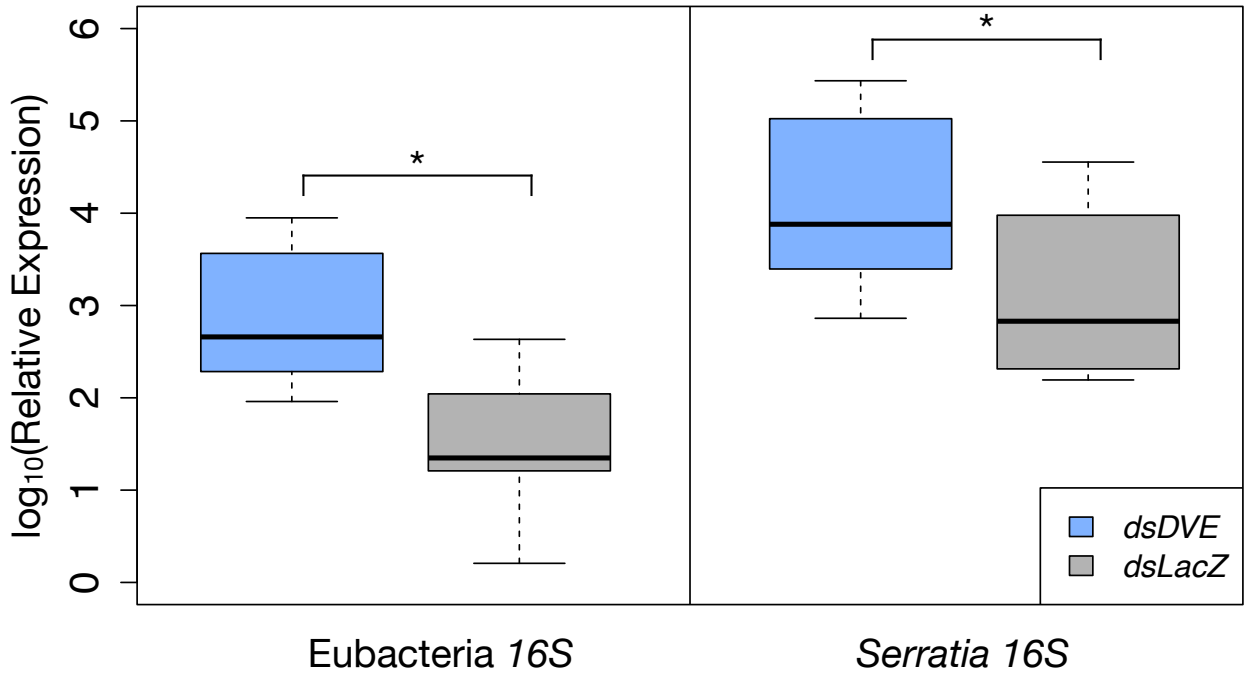


Figure 4.6.1. Relative expression of 16S in the posterior midgut post-*DVE *KD. *Control mosquitoes were injected with* dsLacZ*. Relative expression values were* log_{10}*-transformed for presentation purposes and are relative to a pool of mosquito cDNA. Four replicates were performed using 15–30 mosquito individuals per condition. Black bars represent medians. Differences in Eubacteria* 16S *(microbiota load) and* Serratia 16S *were compared with Wilcoxon signed-rank tests,* * p < *0.05.*

Further experiments were performed to identify whether the AMPs expressed in the cardia have a major effect on the microbiota harbored in the posterior. Here, *DVE* was silenced, and the relative loads of 3 major bacterial families (Acetobacteraceae, Enterobacteriaceae, and Flavobacteriaceae) in the midgut was measured with qRT-PCR. These three families of bacteria have been previously identified to be part of the mosquito microbiota (Boissière et al, 2012; Gendrin et al, 2015; Akorli et al, 2016; Tapanelli et al, 2021). Depletion of *DVE* had no effect on overall microbiota load or any of the three bacterial families (Wilcoxon signed-rank tests, $p > 0.05$; Fig. 4.6.2). This elucidates the possibility that AMPs in the cardia have little effect on the microbiota communities already established in the posterior midgut under sugar feeding conditions. Instead, the cardia-expressed AMPs appear to be important in sterilizing the diet, preventing infections in the posterior midgut or changes in the microbiota community. In this case, the high expression of AMPs in the cardia would also enable tolerance in the posterior midgut due to the distance between the tissues. This lack of impact could also be a result of a number of other explanations. First, as the expression of many AMPs are impacted following *DVE* KD, the microbiota could be affected in a collective manner as opposed to specific clades being impacted by the modulation of a single AMP. Alternatively, as the midguts were collected 2 days post-*DVE* KD, this may not have been enough time to significantly alter the microbiota composition.

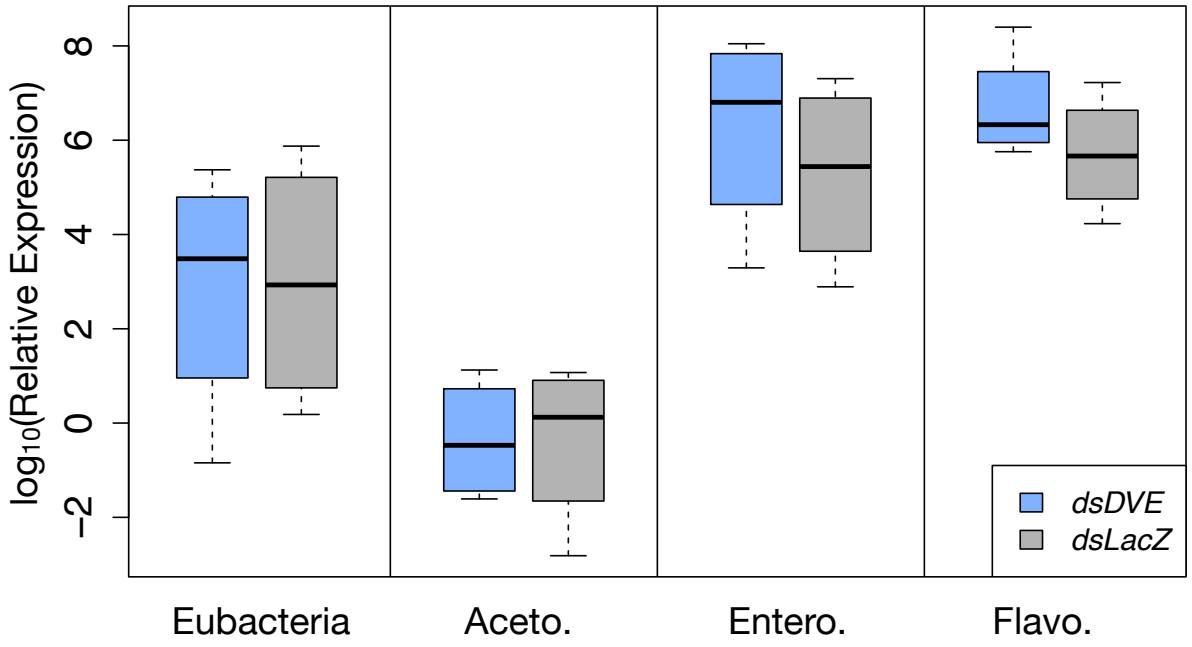


Figure 4.6.2. Relative expression of major bacterial families in the posterior midgut post-DVE KD. *Control mosquitoes were injected with* dsLacZ*. Relative expression values were log$_{10}$-transformed for presentation purposes and are relative to a pool of mosquito cDNA. Four replicates were performed using 15–30 mosquito individuals per condition. Aceto. = Acetobacteraceae; Entero. = Enterobacteriaceae; Flavo. = Flavobacteriaceae. Black bars represent medians. No differences were observed in bacterial loads post-KD (Wilcoxon signed-rank tests, * p < 0.05).*

4.7 KD of *DVE* and *REL2* reduces AMP expression throughout the midgut

Activation of AMP expression in the posterior midgut has been characterized to be microbiota- and REL2-dependent (Meister et al, 2005; Luna et al, 2006; Meister et al, 2009; Rodgers et al, 2017). Considering the role of DVE in AMP expression in the cardia but not the posterior, co-silencing of *DVE* and *REL2* was therefore hypothesized to reduce AMP expression across the entire midgut axis. Mosquitoes were injected with dsRNA targetting *DVE* and *REL2* KDs, and AMP expression in the cardia/anterior and posterior midgut was analyzed.

Depletion of both *DVE* and *REL2* led to significant reductions in AMP expression along the entire midgut axis (Wilcoxon signed-rank tests, $p < 0.05$; Fig. 4.7.1). On average, *CEC1*, *CEC2*, and *DEF1* expression was reduced ~25% in the cardia/anterior and ~50% in the posterior. *GAM1* expression was reduced an average of 36% in the cardia/anterior and 46% in the posterior. In the posterior, microbiota loads were also significantly increased, with an average increase of over 2-fold (Wilcoxon signed-rank test, $V = 55$, $p = 0.002$). The reduction of AMPs and increase in microbiota load in the posterior was due to *REL2* depletion, as *DVE* KD alone previously had no effect. These results indicate that midgut is organized so that different transcription factors and pathways control AMP expression in different regions of the midgut.

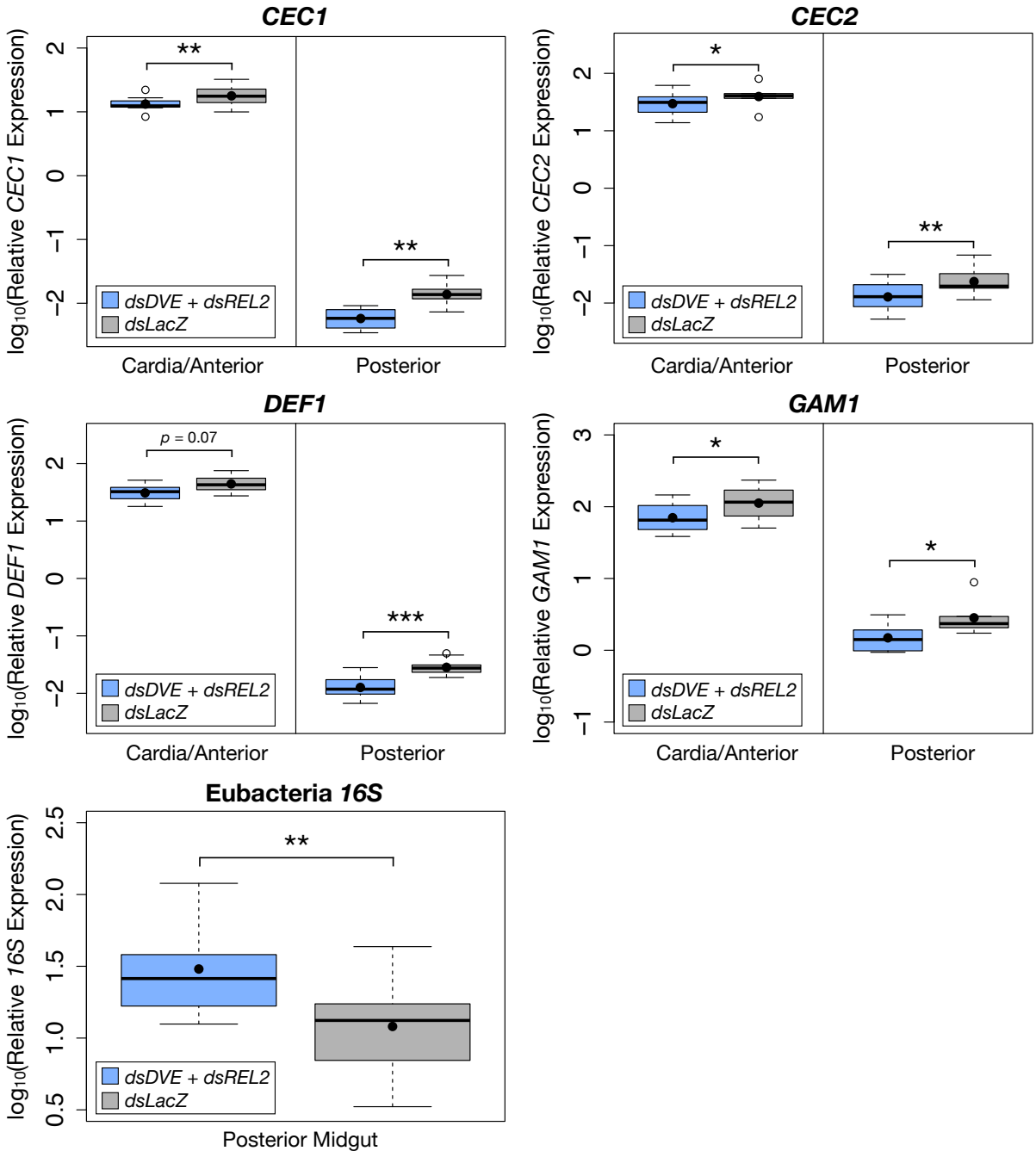


Figure 4.7.1. Relative expression of AMPs and Eubacteria* 16S *post-*DVE *and* REL2 *co-silencing in the midgut. *Control mosquitoes were injected with* dsLacZ*. Relative expression values were log_{10}-transformed for presentation purposes and are relative to a pool of mosquito cDNA. Between 7 and 11 replicates were performed using approximately 30 mosquito individuals for each condition. Black bars represent medians, dots represent means. Differences in relative expression of AMPs/*16S *between the KD and control mosquitoes were compared with Wilcoxon signed-rank tests and paired Student's* t*-tests,* * p < *0.05,* ** p < *0.01,* *** p < *0.001.*

4.8 FAMeTs in the cardia contribute to immunotolerance in the posterior midgut

DVE and the cardia play a pivotal role in shaping the immune landscape of the mosquito midgut. In the midgut sections transcriptomic data, farnesoic acid O-methyl transferases (FAMeTs) 1 (AGAP006194) and 2 (AGAP006195) were found to be enriched in the cardia and anterior midgut sections (Fig. D.3). Importantly, a previous study has shown that the expression of these FAMeTs in the midgut was largely microbiota-dependent, as antibiotic treatment of mosquitoes significantly reduced their expression (Rodgers et al, 2017). FAMeTs convert farnesoic acid (FA) to MF, which is a precursor of juvenile hormone III (JH III). JH III is converted from MF by an epoxidase, and is typically produced in the brain (corpora allata) and secreted into the hemolymph (Hiruma & Kaneko, 2013). While FAMeT 1 was previously found to be the tenth most abundant protein in the PM (Dinglasan et al, 2009), its enrichment in the cardia was unexpected. In *Aedes* and *Drosophila*, JH III signals through Methoprene-tolerant (Met) and Taiman (Tai), which form a receptor complex (Ashok, Turner & Wilson, 1998; Zhu, Busche & Zhang, 2010; Li, Mead & Zhu, 2011; Jindra, Bellés & Shinoda, 2015). This signal is transduced to Krüppel homolog 1, a transcription factor which represses metamorphosis and promotes investment into reproduction (Zhu, Busche & Zhang, 2010). Another hormone, 20-hydroxy-ecdysone (20E) has an opposing role where it initiates metamorphosis. The interaction between JH III and 20E is antagonistic and represents an immunity–reproduction trade-off. JH III suppresses AMP induction in favor of investment into reproduction, whereas 20E promotes AMP expression (Flatt et al, 2008).

In *Anopheles*, homologs for Met and Tai are AGAP006022 and AGAP009664 respectively. Both of these components were enriched in the anterior/posterior at 0 and 5 hours post-blood feeding (Fig. D.4). JH III in mosquitoes also has similar roles as in *Drosophila*, where it regulates metamorphosis, reproductive maturation, and suppresses AMPs in the fat body (Zhu & Noriega, 2016; Chang et al, 2021). Interestingly, the role of JH III is not necessary for larvae to development into adults; MF can in part replace the function of JH III in its absence (Nouzova et al, 2021). Taken together, it was hypothesized that FA is converted to MF in the cardia or along the midgut axis, which then has immunosuppressive roles in more posterior regions of the midgut.

Both FAMeTs were highly expressed in the midgut compared to the remaining carcass of mosquitoes, indicating enrichment in the midgut and cardia (Fig. D.3). In order to elucidate the roles of MF in the mosquito midgut, RNAi-mediated silencing of the FAMeTs was performed to imitate MF depletion, and *dsLacZ*-injected mosquitoes were used as a control. Midguts were dissected and separated into cardia/anterior and posterior regions 48 hours later, and expression of four AMPs was measured via qRT-PCR as previously performed. Fanyun (Coco) Wu, an undergraduate student carrying out their final year project in the lab (project supervised by Prof. George K. Christophides), contributed towards the collection of data from all of the experiments outlined in this section. On average, the KD efficiencies for *FAMeT1* and *FAMeT2* were 43% and 50% respectively in the midgut. *FAMeT1*-silencing increased the expression of cecropins and *DEF1* in both regions of the midgut, although this was only significant for *CEC2* in the posterior midgut (Wilcoxon signed-rank test, $V = 21$, $p = 0.03$; Fig. 4.8.1). The depletion of *FAMeT2* led to similar observations, where expression of the same 3 AMPs increased, with *CEC1* and *DEF1* expression increasing significantly (Fig. 4.8.1; Wilcoxon signed-rank tests, $p < 0.05$). *GAM1* expression was not significantly affected following the silencing of either *FAMeT*. Taken together, MF suppresses immune responses in the posterior midgut, with *FAMeT2* having a greater effect. Despite an increase in AMPs, the microbiota load was unaffected post-*FAMeT1* silencing, but increased by an average of 8-fold post-*FAMeT2* silencing although insignificantly (Student's *t*-test, $t = 2.4$, df = 5, $p = 0.06$).

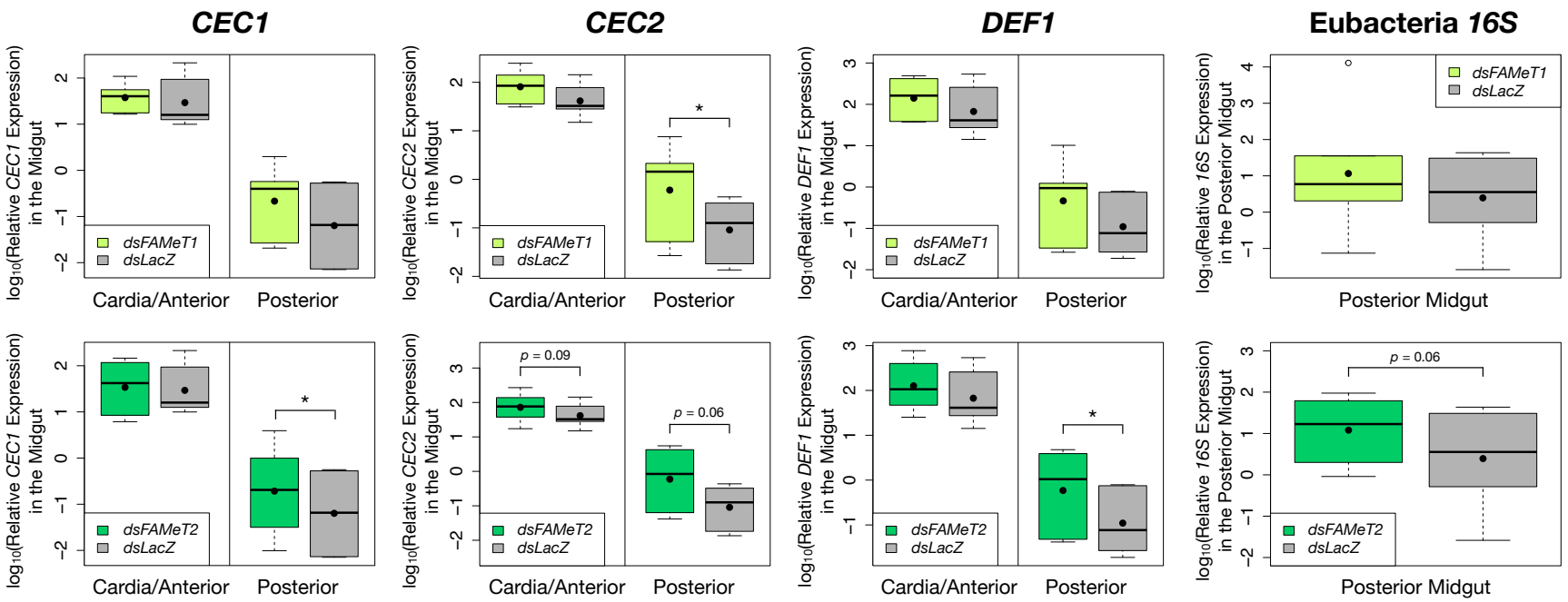


Figure 4.8.1. Relative expression of AMPs and* 16S *post-*FAMeT *KD in the midgut. *Control mosquitoes were injected with* dsLacZ. *Relative expression values were log$_{10}$-transformed for presentation purposes and are relative to a pool of mosquito cDNA. A total of 4–6 replicates were completed using approximately 15–30 mosquito individuals per condition, with Fanyun (Coco) Wu (BSc final year project student, project supervised by Prof. George K. Christophides) contributing towards 2–3 replicates. Black bars represent medians, dots represent means. Differences in the original relative expression values were compared with paired Student's* t*-tests or Wilcoxon signed-rank tests,* * p < *0.05,* ** p < *0.01.*

Given the suggestive immunosuppressive effect of MF in the posterior midgut and increase in microbiota load post-*FAMeT2* depletion, the next objective was to establish whether MF has a direct effect on the microbiota, or if its effects are indirect (*e.g.* acting through the MET–TAI receptor complex). Exogenous MF was added to cultures of *E. coli* and *S. marcescens*, and the OD_{600} of these cultures was measured every hour to assess whether growth rates were inhibited. An equal volume of DMSO was added to separate cultures as controls. The growth rates of either species of bacteria were not affected with 100 µM and/or 1 mM treatment of MF (Fig. 4.8.2, permutation tests, 10,000 permutations, $p > 0.05$). Therefore, these initial experiments indicate that MF does not have a direct impact on bacteria and may exert its effect through immunosuppressive signaling, although additional species need to be tested to further validate this finding.

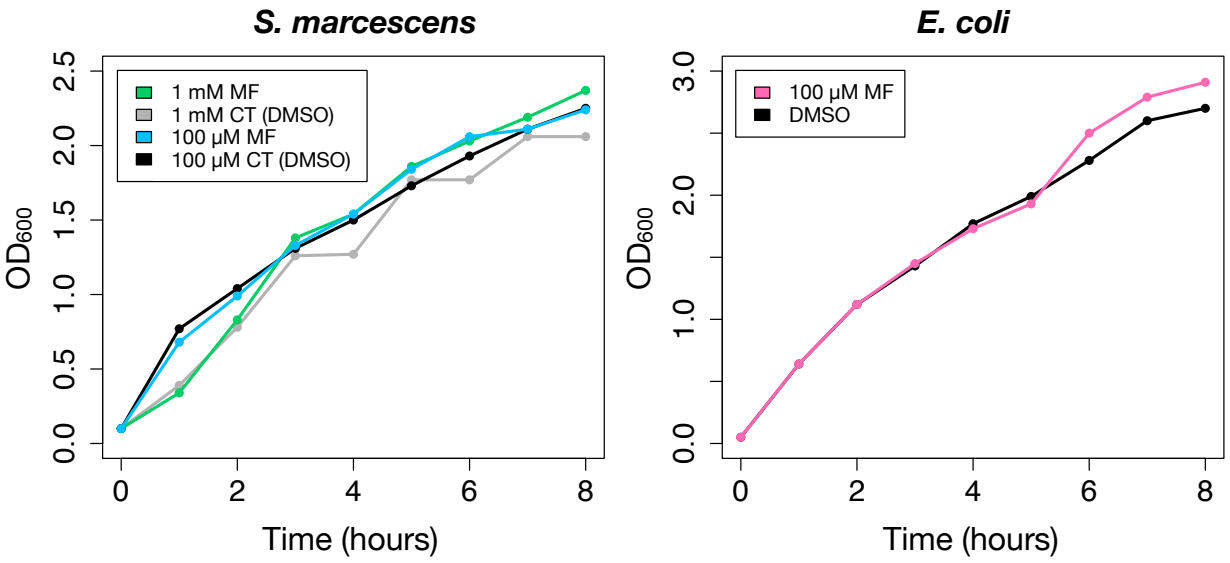


Figure 4.8.2. Average growth of* S. marcescens *and* E. coli *with MF or DMSO treatment. *OD_{600} was measured hourly for 9 time points as an indicator of bacterial growth. Fanyun (Coco) Wu (BSc final year project student, project supervised by Prof. George K. Christophides) contributed to the experiments. Three biological replicates (inclusive of three technical replicates each) were performed for* S. marcescens *treated with 100 µM MF, and one biological replicate (inclusive of three technical replicates) was performed for* S. marcescens *treated with 1 mM MF and* E. coli. *CT = control condition for the corresponding MF concentration. MF treatment had no effect on the growth of either species of bacteria (permutation tests, 10,000 permutations,* $p > 0.05$*).*

To further investigate how FAMeTs (through MF) exert their effects, exogenous MF was mixed into water or sugar and fed to mosquitoes. Feeding MF to mosquitoes through water is thought to enable MF to be transported immediately into the midgut, whereas sugar is stored in the crop and gradually released into the midgut. Two days later, the cardia/anterior and posterior regions of the midgut were collected and qRT-PCR was performed to measure the expression of AMPs, FAMeTs, and *16S*. Undergraduate student Sihao Ma, carrying out their final year project in the lab (project supervised by Prof. George K. Christophides), assisted with data collection for some of the replicates of this experiment. In the posterior midgut, median expression of all four AMPs was reduced following MF feeding (Fig. 4.8.3). Despite a relatively lower expression of AMPs in the posterior regions compared to the anterior regions of the midgut, this reduction was consistently observed although it did not reach statistical significance. In the cardia/anterior region, MF feeding had no effect on the expression of AMPs (Fig. D.5), corroborating the hypothesis that MF exerts its immunosuppressive effects in the posterior midgut. Similarly, the microbiota load was reduced (Fig. 4.8.3). These results are the opposite of what was observed following *FAMeT* KD (*i.e.* MF depletion).

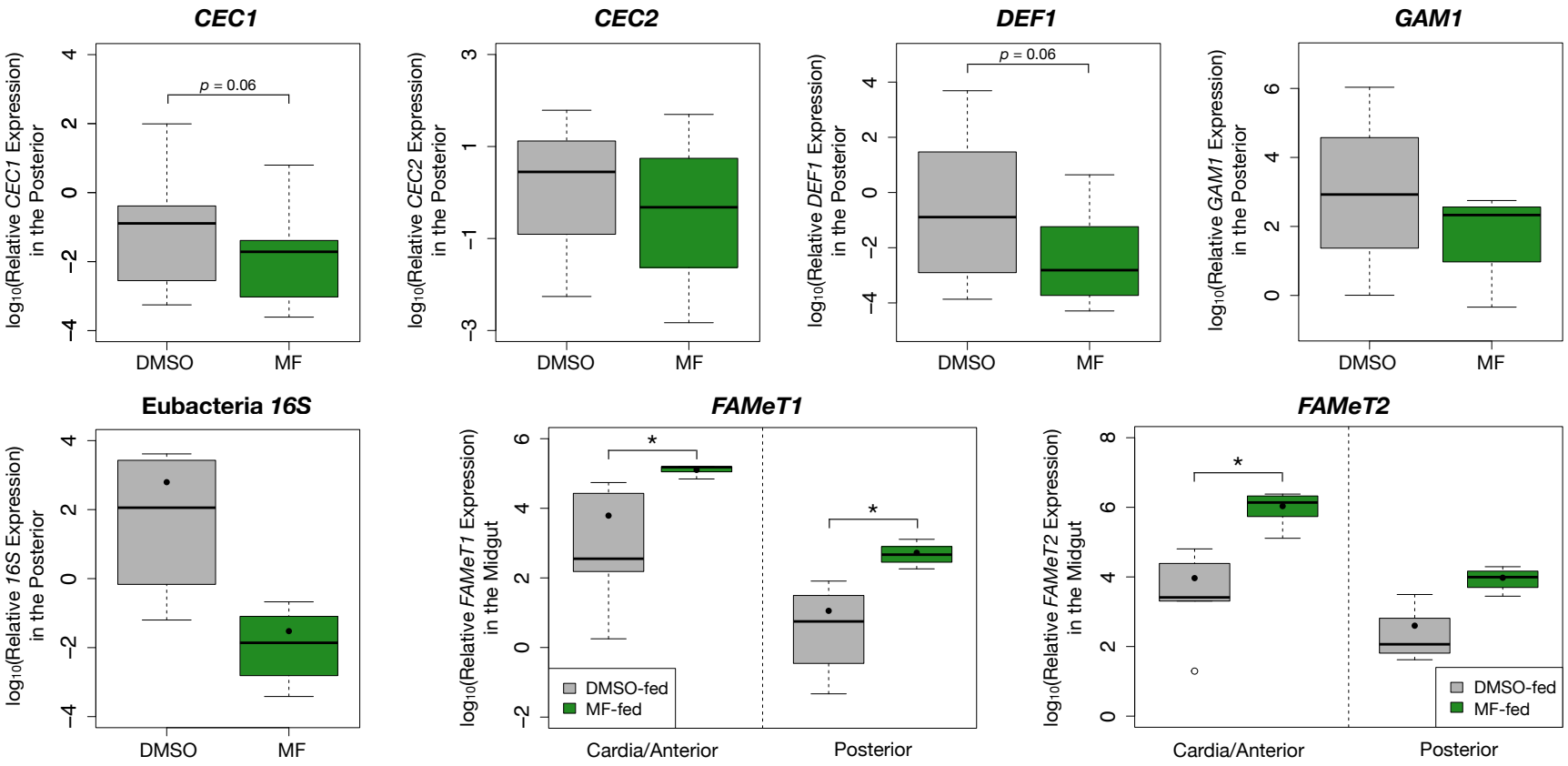


Figure 4.8.3. Relative expression of AMPs,* 16S*, and FAMeTs post-MF feeding in the midgut. *Relative expression values were log_{10}-transformed for presentation purposes and are relative to a pool of mosquito cDNA. Black bars represent medians, dots represent means. A total of 4–6 replicates were performed using approximately 15–30 mosquito individuals per condition, with BSc final year project students Fanyun (Coco) Wu and Sihao Ma (both projects supervised by Prof. George K. Christophides) contributing towards the data collection of 2–3 replicates each. Student's* t*-tests and Wilcoxon signed-rank tests,* * $p < 0.05$.

Following the feeding of exogenous MF, *FAMeT1* expression was significantly upregulated in both the anterior and posterior regions of the midgut (Fig. 4.8.3; paired Student's *t*-tests, $p < 0.05$). *FAMeT2* expression was also upregulated in both regions of the midgut, although this was only significant for the cardia/anterior region (paired Student's *t*-test, $t = 3.9$, df = 4, $p = 0.02$). This indicates that the downregulation of AMPs post-MF feeding could be a result of upregulated *FAMeT* expression and hence increased MF levels, as well as the possibility that exogenous MF induces *FAMeT* expression.

4.9 Discussion

The mosquito midgut is a tissue of particular significance as it marks the first site of major interactions with food, foreign molecules, and microbes/pathogens. A clear dichotomy in function exists across the midgut axis, and further analysis of secreted transcriptomes of each tissue at each time point enabled in-depth characterization of these functions. The cardia invests a significant amount of resources into the production of a disproportionately few number of genes (of which many are uncharacterized), suggesting it is highly specialized in function. One of these key functions is immune defense, as over a third of the secreted transcriptome is made up of just four known AMPs. This high, constitutive expression of AMPs is at least partially controlled by the developmental transcription factor DVE, which is also constitutively expressed in the cardia. In *Drosophila*, Dve controls cardia morphogenesis and its expression is maintained in the cardia throughout adulthood (Nakagoshi et al, 1998; Nakagawa et al, 2011). Throughout larval development, *DVE* expression peaks in L1 larvae, suggesting a similar role in cardia morphogenesis in *Anopheles* (Papa et al, 2017). In midguts of L4 larvae, AMP expression is reminiscent of that in adults, where AMPs peak in the gastric caeca, the anterior-most tissue of the midgut that is analogous to the adult cardia (Neira Oviedo et al, 2008).

Control of AMP expression in the cardia by DVE, a developmental transcription factor, is another example of developmental pathways being co-opted for the regulation of immune responses in *Anopheles*. The Toll pathway is also believed to be secondarily co-opted by the immune system for immune signaling in *Drosophila*, however two transcription factors exist downstream of Toll: Dorsal, which regulates dorsal–ventral patterning of the embryo, and Dif, which regulates the transcription of immune genes (Anderson & Nüsslein-Volhard, 1984; Wu & Anderson, 1998; Towb, Bergmann & Wasserman, 2001). In *Anopheles*, REL1 is the sole transcription factor downstream of TOLL signaling, implicating its role in both developmental and immune signaling, further highlighting the synergy between development and immunity. Other developmental transcription factors such as Cad have also been co-opted as immune regulators in both genera (Ryu et al, 2008; Clayton et al, 2013). Depletion of *DVE* leads to reduced AMP expression in the cardia/anterior midgut, which in turn led to increased loads of *S.*

marcescens in the midgut following oral infection. These oral infections were performed by resuspending bacteria in sugar. Thus, the high expression of AMPs in the cardia also plays a role in the sterilization of the midgut, whereby any ingested food passing through the cardia are rid of pathogens. These results also provide evidence against the notion that these highly expressed cardia genes may remain as transcripts/peptides in the cell for immediate translation, activation, and/or secretion post-blood meal (Felix et al, 1991; Müller et al, 1995). Taken together, DVE provides the cardia its transcriptional identity of high AMP expression and drives its ability to function as an immunocompetent region of the midgut.

A few uncharacterized genes also made up a significant portion of the secreted transcriptome. Two of these genes, AGAP009313 and AGAP013543 made up approximately a quarter of the secreted transcriptome of the cardia at all time points. AGAP013543 appears to be *Anopheles*-specific, warranting further examination into the function of this gene. Expression of both genes is microbiota-dependent, which is also a feature of the four AMPs that are highly expressed in the cardia (Rodgers et al, 2017; Rodgers et al, 2020). AGAP009313 shares some homology with the tsetse fly protein tsetseEP, a putative lectin that is abundantly and constitutively expressed in the midgut (Chandra et al, 2004). This tsetse protein is involved in anti-trypanosome responses and expression is induced following bacterial challenge (Haines et al, 2005; Haines et al, 2010). Thus, AGAP009313 may have antimicrobial activity as well. AGAP009313 is also a major component of the PM (Dinglasan et al, 2009), and could also have a role in PM structure.

Fluorescently-labelled molecules enabled the visualization of AMPs in the midgut. Using the GMM transgenic line, antibodies raised against a ribosomal skip signal were utilized as a proxy for GAM1 expression. GAM1 was detected throughout the cardia and some of the anterior midgut, confirming the transcriptomic data. Furthermore, following the silencing of *DVE*, GAM1 signal was fully depleted; no obvious differences could be observed between the phenotype of *DVE* KD and WT mosquitoes. This was unexpected, as the calculated KD efficiency of *DVE* was never complete in the midgut, averaging 44% per replicate using molecular techniques. This highlights the fact that transcription rates do not necessarily reflect translation rates. Indeed, a previous study found that mRNA

and protein levels of the same gene could vary up to 30-fold within a sample in yeast (Gygi et al, 1999).

For the first time, whole-mount RNA FISH was successfully performed on *Anopheles* midguts. RNA FISH of *CEC2* and the two uncharacterized cardia genes AGAP009313 and AGAP013543 showed that they were indeed expressed exclusively in the cardia. While all transcripts were expressed in the cardia, the spatial expression still varied. CEC2 was expressed throughout the entire cardia, including at the boundary of the anterior midgut. On the other hand, expression of the other two genes was restricted to the bulbous, more anterior regions of the cardia. These experiments also enabled the visualization of a series of crypts within the cardia. It is hypothesized that these crypts enable an increased surface area to volume ratio, enabling many cells in the same region to produce, secrete, and store high titers of AMPs in the crypts.

Post-blood feeding, the cardia and anterior midgut also upregulate the expression of genes encoding digestion-related proteins such as peritrophins and trypsins. This indicates that these regions of the midgut also contribute to blood meal digestion, despite the blood meal being harbored in the posterior midgut. Notably, the secreted transcriptome of the posterior midgut was consistently made up of a plethora of digestive enzymes, even in the absence of a blood meal. Many of these transcripts or proteins are likely stored intracellularly and immediately translated or released following a blood meal (Felix et al, 1991; Müller et al, 1995). This high investment into digestive enzymes is coupled with a lack of investment into immunity, creating an immune quiescent region in the posterior midgut and enables persistence of the microbiota.

The transcriptomes of each midgut tissue were previously investigated by Hixson et al and are congruent with the findings presented here, but these tissues were previously only examined under sugar feeding conditions (2022). This present study extends these findings to the time that blood-borne pathogens invade the midgut and indicates that the investment into immunity and digestion by the cardia and posterior midgut regions respectively are not impacted by blood feeding. Many digestive enzymes already make up a large proportion of the transcriptome in the posterior under sugar feeding conditions, and further investment is dedicated towards these enzymes after blood meal uptake. Given the role of DVE in the high AMP expression of the cardia, the

arrangement of the midgut functions across the tissues is likely to be mediated by concentration gradients of transcription factors.

During embryogenesis, the patterning, regionalization, and specialization of body regions relies on these gradients of transcription factors. In *Drosophila*, regionalized and 'striped' expression of multiple transcription factors enable the segmentation of the embryo (reviewed in Clark, Peel & Akam 2019). For example, *paired* (*prd*) is expressed in the anterior while *cad* is expressed in the trunk of the *Drosophila* embryo (Clark, Peel & Akam, 2019). *Prd* marks parasegment boundaries, and *cad* enables early-stage segment patterning of the trunk (Clark, Peel & Akam, 2019). Expression of *cad* retracts in a dorsal and posterior manner, which is concurrent with the formation of multiple bands of *prd* expression moving from the anterior to the posterior regions of the embryo (Clark & Peel, 2018). Many more of these developmental transcription factors could have been co-opted for the controlling the expression of many of these genes investigated here in adulthood. For example, despite sharing an expression profile with AMPs, DVE does not control the high expression of the two uncharacterized cardia genes (AGAP009313 and AGAP013543). Instead, another cardia-enriched transcription factor is hypothesized to confer this phenotype. In the posterior, high transcription of many digestive enzymes would also be controlled by one or multiple transcription factors. The middle ground in terms of immune and digestive investment in the anterior midgut could therefore be a result of 'spillage' of the waning transcription factor concentration gradients into this region.

The observation that FAMeTs are highly expressed in the anterior midgut sections was unexpected as JH III is produced in the corpora allata (brain) and acts in the hemolymph of the mosquito. Given the immunosuppressive effects of JH III (Flatt et al, 2008), in addition to the ability of MF to replace the functions of JH III (Nouzova et al, 2021), it was hypothesized that MF in *Anopheles* also exerts immunosuppressive effects. The role of MF was investigated through its depletion and augmentation *in vivo*. The former was achieved through RNAi-mediated KD of FAMeTs, where AMP expression increased in the posterior. The latter was achieved through the introduction of exogenous MF into the midgut to mimic *FAMeT* overexpression, where AMP expression was found to be suppressed in the posterior midgut. This data provides evidence of MF having

immunosuppressive effects in the posterior midgut, which perhaps contributes to the low expression of AMPs in this region. It is unclear whether these changes in AMP expression were a result of the microbiota. The microbiota load was modulated in a counterintuitive manner, where increased AMP expression post-*FAMeT* KD was coupled with an increased microbiota load, and the reduced AMP expression following MF feeding led was observed with reduced microbiota loads. This suggests that MF is able to directly kill the microbiota which in turn affects the level of basal immune activation. Although bacterial co-cultures with MF did not identify a bactericidal effect, only one species used in these experiments represented an endogenous microbiota member in the mosquito midgut. Therefore, other species/strains of bacteria, particularly those isolated directly from mosquito midguts, could be tested in the future to confirm whether MF has direct bactericidal effects.

Surprisingly, the expression of FAMeTs was significantly affected by MF feeding. Introduction of exogenous MF led to the increase of *FAMeT* expression in both regions of the midgut. FAMeTs catalyze the conversion of FA to MF; their upregulation following MF feeding therefore invites the speculation of many biological explanations. A positive feedback loop could exist, where MF induces the expression of FAMeTs, which then leads to the production of more MF. Alternatively, FAMeTs may also catalyze the reverse reaction, *i.e.* the demethylation of MF into FA, or some other modification. Enzymes such as aconitase, succinate dehydrogenase, and malate dehydrogenase from the TCA cycle catalyze reversible reactions (*i.e.* the forward and reverse reactions). The upregulation of FAMeTs could therefore represent the need for the reduction of MF levels to equilibrate the midgut.

It is hypothesized that a source of the FAMeT substrate FA in the midgut is likely to come from the nectar of the flowers mosquitoes feed on. Then, the conversion of FA to MF occurs either in the cardia/anterior midgut sections and is transported to the posterior, or FAMeTs are secreted and act in the posterior where they suppress immune responses. A previous study found that FAMeTs 1 and 2 were the tenth and eleventh most abundant proteins in the midgut PM respectively (Dinglasan et al, 2009), indicating that they are secreted from the anterior to posterior regions of the midgut. Their embedment/attachment to the PM could further promote tolerance mechanisms during

the blood meal to protect the microbiota in the ectoperitrophic space. Taken together, the enrichment of FAMeTs in the cardia further strengthen its role in midgut immunity and contributing to the gradient of immune effectors expressed across the midgut axis, as these enzymes not only have a role in regulating infection but also microbial colonization of the posterior midgut.

Throughout the entire mosquito, induction of AMP expression in response to infection is not homogeneous across the tissues but rather dependent on the bacterial species and route of infection. Oral infections of *S. marcescens*, a known endosymbiont of mosquito midguts (Bando et al, 2013; Bahia et al, 2014; Stathopoulos et al, 2014), led to the downregulation of AMPs in the midgut, but an upregulation in the hemolymph/fat body. This suggests a mechanism which recognizes microbiota in the midgut and promotes tolerance rather than the induction of immune responses. In *Drosophila*, this mechanism is controlled by Imd pathway activator PGRP-LE (Bosco-Drayon et al, 2012). PGRP-LE recognizes PAMPs in the midgut and induces the expression of AMPs (Bosco-Drayon et al, 2012). In response to microbiota, Pirk and PGRP-LB, negative regulators of the Imd pathway, are upregulated instead of AMPs (Bosco-Drayon et al, 2012). The proposed model behind the mechanism involves the duration of interaction between the bacteria and fruit fly midgut, and also the bacterial load (Bosco-Drayon et al, 2012). Microbiota form long-lasting interactions at relatively lower bacterial loads, which therefore elicit moderate Imd activation and the induction of tolerance responses. Larger loads of pathogenic bacteria that are unable to persist in the midgut have brief interactions with the midgut and elicit a strong immune response. No homologs exist for PGRP-LE and Pirk in *Anopheles* (Christophides et al, 2002), so it is unclear whether such a mechanism could explain the results observed here. Other molecules may perform these roles, or an entirely different mechanism may exist in *Anopheles*. The upregulation of AMPs in the carcass following *S. marcescens* infection might reflect a variety of mechanisms, including anticipatory responses, immune priming, inter-tissue signaling between the midgut and hemolymph, and/or that the microbiota have breached the midgut barrier and have become pathogenic (Dekmak et al, 2021). A previous RNA-seq study found that the responses to *S. marcescens* (Db11 strain, the same strain used here) depended on route of infection, where oral infections led to suppressed immune

responses compared to systemic infections (Dekmak et al, 2021). Additionally, a previous study found that mosquito hemocoel infections with *Enterobacter* and *Serratia*, bacterial species naturally associated with mosquito midguts, leads to rapid mosquito death (Kulkarni et al, 2021). Taken together, these findings indicate that microbiota species are pathogenic in tissues where they should not be present (*i.e.* most tissues except the midgut), and immune responses are launched to tackle this infection.

Systemic infections of *E. coli*, a non-microbiota bacterial species, elicited AMP expression. As this species is pathogenic regardless of localization of infection, it is likely that immune responses were launched immediately following their detection, similar to systemic *S. marcescens* infections. Interestingly, comparisons of systemic infections between endogenous species (*Enterobacter* and *Serratia*) and pathogenic (*E. coli*) found that the thoracic injection of endogenous species was much more virulent than *E. coli* (Powers et al, 2020; Dekmak et al, 2021). In the case of *E. coli*, the bacterial loads remained high but the survival time was much longer, indicating tolerance mechanisms in place for such infections (Powers et al, 2020; Dekmak et al, 2021). Although all bacterial species discussed so far are Gram-negative, the differences in immune responses and survival outcomes varies greatly, highlighting a need for more nuanced models of infection/immune mechanisms.

Ultimately, the cardia, in its strategic position at the forefront of the midgut, plays a major defensive role to protect the midgut against ingested pathogens. The investment towards the constitutive expression of an arsenal of AMPs is largely conferred by DVE, contrasting with REL2 in the posterior, which induces AMP expression upon pathogen recognition. In the midgut, two main opposing gradients exist across the entire axis. Immunocompetence is enhanced in the cardia and gradually reduces towards more posterior regions. This prevents overactivation of immune responses against the microbiota, which provide many benefits to the mosquito host. Contrastingly, in addition to immune quiescence, the posterior midgut invests significant resources into the production of digestive enzymes to enable rapid blood meal digestion upon blood feeding. This biological process is extremely demanding, and the cardia and anterior also contribute a small percentage of their transcriptomic investment into the transcription of digestive enzymes post-blood feeding. The cardia also contributes towards immune

tolerance in the posterior through the secretion of FAMeTs, which have an immunosuppressive effect. This further enhances the stark contrast of the anterior versus posterior regions of the midgut in terms of immunity. Clearly, the midgut is highly compartmentalized, with each region having dedicated biological functions.

Chapter 5: Results III
Novel Antimicrobial Effectors

5.1 Introduction

5.1.1 Abstract

AMPs are important components of immune responses in all domains of life and include a broad range of molecules. Despite diversity across AMPs, the majority are thought to be broad-spectrum and exert antimicrobial effects through membrane disruption. In *Anopheles*, four families of AMPs have been identified, some of which possess anti-malarial activity. AMPs from other organisms have also been examined in the context of malaria infection and have also been introduced into the *Anopheles* genome, which successfully restricted parasite development. Given that the genomes of many other insects encode dozens of AMPs across multiple families, it was hypothesized that many more were to be discovered in *Anopheles*. Utilizing RNA-seq datasets and machine learning algorithms, a variety of *in silico* analyses were performed to identify novel, putative AMPs in the malaria mosquito. Multiple candidates AMPs were discovered and were screened for activity against bacteria and the human malaria parasite. Of these candidates, many possessed antimicrobial activity and deserve to be further characterized. Following further experimentation to validate their antimicrobial activity, the expression of these novel, putative AMPs could be modulated to generate malaria-refractory mosquitoes.

5.1.2 Background

AMPs represent a diverse group of defense molecules. The majority are thought to share some characteristics which enable their antimicrobial activity, such as having a positive charge and exhibiting amphipathicity, both of which contribute to their affinity to negatively-charged microbial membranes (Ladram & Nicholas, 2016). AMPs are also typically short in length, and have the propensity to form secondary structures such as α-helices (Wang, Li & Wang, 2009; Ladram & Nicholas, 2016). To date, four families have been identified in *Anopheles*, of which GAM1 appears to be mosquito-specific (Vizioli et al, 2001b).

In the last few decades, AMP research has flourished. Numerous databases exist and contain thousands of AMP sequences (Thomas et al, 2010; Lee et al, 2015, Waghu et al, 2016). Additionally, the advent of various machine learning algorithms has enabled

accurate prediction of protein sequences, such as determining whether a sequence is an AMP and predicting protein folding structures (Xiao et al, 2013; Khosravian et al, 2013; Lee et al, 2015; Waghu et al, 2016; Jumper et al, 2021). These algorithms include support vector machines (SVM), random forests (RF), discriminant analysis (DA), and artificial neural networks (ANN) (Thomas et al, 2010). For example, the predictive algorithms of AMP database $CAMP_{R3}$ (Waghu et al, 2016) were trained on two generated datasets: a positive (*i.e.* a list of experimentally validated 2,500+ AMP sequences) and negative (~5,000 sequences). The negative dataset was made up of a combination of experimentally validated non-AMPs, randomly generated sequences, non-secreted known proteins, and randomly retrieved sequences lacking an antimicrobial annotation (Thomas et al, 2010; Waghu et al, 2016). During the testing process, the various algorithms were able to achieve up to 86% accuracy (Thomas et al, 2010). Using these technologies, novel AMPs can be identified from sequences originating from any organism (Khosravian et al, 2013; Waghu et al, 2016).

The use of AMPs in generating malaria-refractory transgenic mosquitoes is highly promising, as demonstrated through the modulation/modification of endogenous AMPs, or introduction of exogenous AMPs into the *Anopheles* genome (Isaacs et al, 2011; Dong, Simões & Dimopoulos, 2020; Hoermann et al, 2022). A recent study generated two transgenic lines, where exogenous AMPs were introduced into the *GAM1* and *CP* loci to enable enriched expression in the cardia and posterior midgut respectively (Hoermann et al, 2022). Melittin, an AMP from a honey bee (*A. mellifera*), and magainin 2, an AMP from a frog (*X. laevis*) were genetically engineered into the mosquito genome as they previously showed promising transmission blocking activity following experiments involving their supplementation into infectious blood meals (Habtewold et al, 2019). While both transgenic lines displayed significantly reduced incidences of malaria infection, the expression of AMPs in the posterior midgut yielded a stronger anti-*Plasmodium* phenotype, and further characterization of this line showed that oocyst development was attenuated or delayed compared to that of infected WT mosquitoes, which in turn reduced the number of sporozoites (Hoermann et al, 2022).

Many other transgenic lines have been characterized to possess enhanced antimalarial activity, however the majority utilize the same subset of effectors, while a

plethora of effectors are available and have yet to be characterized in the context of developing transgenic, malaria-refractory mosquitoes. Similarly, a limited number of promoters/loci have been used in these strategies, whereas many other sites have yet to be considered, and may be more effective.

It was hypothesized that there are putative AMPs/effectors present in the *Anopheles* genome, as many genes/proteins are uncharacterized, and also because other insect species encode many more AMPs in their genomes. The various AMP prediction algorithms and other machine learning platforms, in addition to the RNA-seq datasets generated in this thesis were leveraged to identify novel and putative AMPs. The discovery of novel effectors can be used to generate the next transgenic mosquito lines that are malaria-refractory.

5.1.3 Methods

In order to identify putative AMP candidates for functional characterization, a list of selection criteria were applied to transcripts in the previously generated RNA-seq datasets (described in sections 3.2 and 3.4), including possession of a signal peptide and significantly DE between any pair of sample conditions. From this initial candidate list, transcript peptide sequences were run through various AMP predictive algorithms, and their expression profiles in the RNA-seq datasets were compared to that of known AMPs. Beyond the generated RNA-seq datasets, a *de novo* assembly was built on the midgut sections RNA-seq data and a *P. vivax*-infected *An. darlingi* dataset was reanalyzed to identify additional AMP candidates. These methodologies are outlined in sections 2.7 and 2.8.

Following the candidate selection process, the structural characteristics of putative AMP candidates were investigated with AlphaFold (Jumper et al, 2021) and αCharges (Schindler et al, 2023), which were also compared to that of known AMPs. RNAi-mediated silencing (detailed in section 2.9) was performed to investigate whether these uncharacterized peptides possess antibacterial and antimalarial activity through the analysis of survival curves (section 2.10) and oocyst enumeration respectively (section 2.2). Survival curves were compared with Cox proportional hazard models, where HRs

and their associated *p*-values were calculated with respect to control mosquitoes (*i.e, dsLacZ*-injected).

Given that AMPs are responsive to immune challenge, the relative expression of the putative AMP candidates were assayed following bacterial infection. Local and systemic responses were investigated through the use of *S. marcescens* in the midgut and *E. coli* in the hemolymph respectively (section 2.10). Expression of putative AMPs was measured via qRT-PCR according to sections 2.4 and 2.5. Additionally, the expression of some putative AMPs was measured in various tissues and throughout development as described above.

Finally, the effect of exogenous molecules were investigated for antimalarial activity. These molecules were utilized in spike-in experiments to identify their impact on asexual stage development *in vitro*, male gametocyte exflagellation *in vitro*, and midgut stage parasite development in the mosquito (section 2.2.4).

5.2 Identification of novel, putative AMPs from RNA-seq datasets

In order to identify novel AMPs in *An. gambiae*, 4 selection strategies were performed and 6 different predictive algorithms from three AMP databases were utilized: 1) SVM from ADAM (Lee et al, 2015); 2) fuzzy K-nearest neighbor from iAMP-2L (Xiao et al, 2013); and 3–6) SVM, RF, ANN, and DA from $CAMP_{R3}$ (Waghu et al, 2015). Transcripts which fulfilled the following initial selection criteria were selected from the RNA-seq datasets for AMP prediction: significantly DE (log_2 fold change ≥ 2 or ≤ -2) between any pair of sample conditions (adjusted *p*-value < 0.05), less than 800 bp in length, possess a signal peptide, average normalized counts greater than 50, and lacking an annotation (*i.e.* annotated as 'unspecified product' in VectorBase).

For the first selection strategy, both *An. gambiae* RNA-seq datasets were utilized, where 98 unique transcripts (93 and 62 transcripts from the *Ag/Pf* and midgut sections datasets respectively) fulfilled the initial selection criteria. These sequences were assessed by the predictive algorithms and continued to the next round of selection if they passed the threshold of being predicted to be an AMP by at least 5 of 6 algorithms. Here, a total of 12 and 8 genes were identified in the *Ag/Pf* and midgut sections datasets respectively, with all of the putative AMPs in the midgut sections data overlapping with those identified in the *Ag/Pf* data (Table 5.2.1). The 4 genes that were unique to the *Ag/Pf* data were excluded from further analyses as their average counts in the midgut sections data were low. Similarly, AGAP000152 and AGAP005612 were excluded because their average counts were amongst the lowest in both datasets. From this selection strategy, a total of 6 AMP candidates were identified.

Three of the putative AMPs are located in the same locus and are predicted to have antibacterial activity, which potentially reflect the possibility of being novel types of AMPs. These genes, AGAP005888, AGAP005889, and AGAP005890, are enriched in the cardia and are upregulated in response to *P. falciparum* infection in the first 24 hours of infection. Additionally, they were previously identified by the host laboratory in microarray experiments involving *Plasmodium*-infected midguts, where they were upregulated in the presence of the parasite at 24 hours post-infection (unpublished data). AGAP005888, known as kininogen 1 (KIN1), was previously shown to be under the control of REL2 (Meister et al, 2005), but had no impact on the development of *P. berghei*

and *P. falciparum* oocysts (Vlachou et al, 2005; Dong et al, 2006; Mendes et al, 2008). KIN1 is made up of 93 amino acids, 25.8% and 17.2% of which are histidine and glycine respectively. Despite an abundance of positively-charged histidines, the overall charge of the peptide is slightly negative. AGAP005889 is positively-charged and 114 amino acids long, 42% of which are glycine residues and will be referred to as 'glycine-rich' (GLYR). AGAP005890 is a positively-charged, alanine- and glycine-rich (21% and 13% composition respectively) peptide and will be referred to as 'alanine-rich' (ALAR). Collectively, they will be referred to as members of the putative AMP locus.

Next, the midgut expression profiles of each of the 6 putative AMPs identified so far were compared to that of known AMPs, which show enrichment in the cardia at all time points post-blood feeding. All genes selected showed a peak of expression in the cardia, with the exception of AGAP003713 which peaked in the anterior (Fig. 5.2.1).

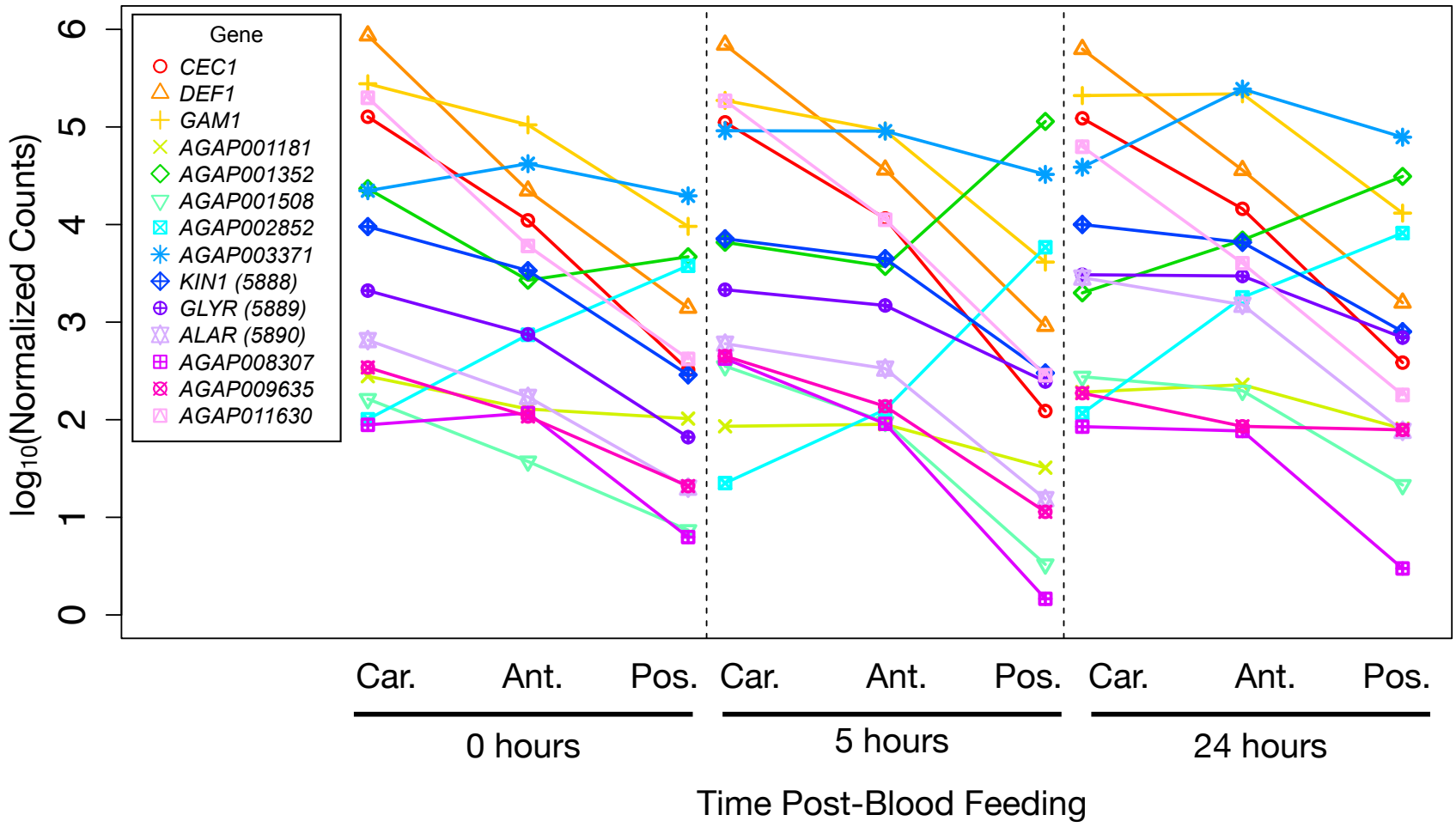


Figure 5.2.1. Expression profiles of putative and known AMPs in the midgut post-blood feeding. *Normalized counts were averaged and calculated from three replicates of the midgut sections transcriptomic data. Car. = cardia, Ant. = anterior, Pos. = posterior midgut tissues.*

In order to identify potential *P. falciparum*-responsive/induced genes, an additional search strategy was employed. First, genes that were predicted to be AMPs by 4 of 6 algorithms in the *Ag/Pf* data were identified, yielding 24 sequences. From these sequences, their expression profiles were assessed in both RNA-seq datasets and genes were selected for further characterization on a case-by-case basis. Four genes were selected in total: AGAP001181, which was the only gene that was significantly upregulated in an infected sample compared to its time- and tissue-matched control; AGAP002852, which was expressed approximately 40% higher in infected midguts compared to control midguts at 24 hours post-infection; AGAP008307, which showed a peak of expression in the cardia at 5 hours post-blood feed; and AGAP009635, which also had a peak of expression in the cardia at 5 hours post-blood feed and is glycine-rich (32 of 125 residues are G) (Fig. 5.2.1; Fig. 5.2.2; Table 5.2.1).

In addition, as just five AMPs made up over a third of the cardia's secreted transcriptome (section 4.2), two uncharacterized genes which also made up a significant portion of the cardia's transcriptome were also assessed as putative AMPs. These two genes were highly enriched in the cardia at all time points post-blood meal, making up nearly a quarter of the secreted transcriptome combined. AGAP009313 is made up of 247 amino acids and is 970 bp long, meaning it was filtered out using the initial selection strategies. However, the majority of the algorithms predicted it to be an AMP. AGAP013543 fulfilled the initial selection criteria as it is 170 amino acids and 677 bp long, but was filtered out as it was only predicted to be an AMP by half of the algorithms. Both genes were upregulated in the presence of the parasite (Fig. 5.2.2). Given their significant enrichment in the cardia, they were included as putative AMPs. These two genes will be collectively referred to as the 'cardia genes'.

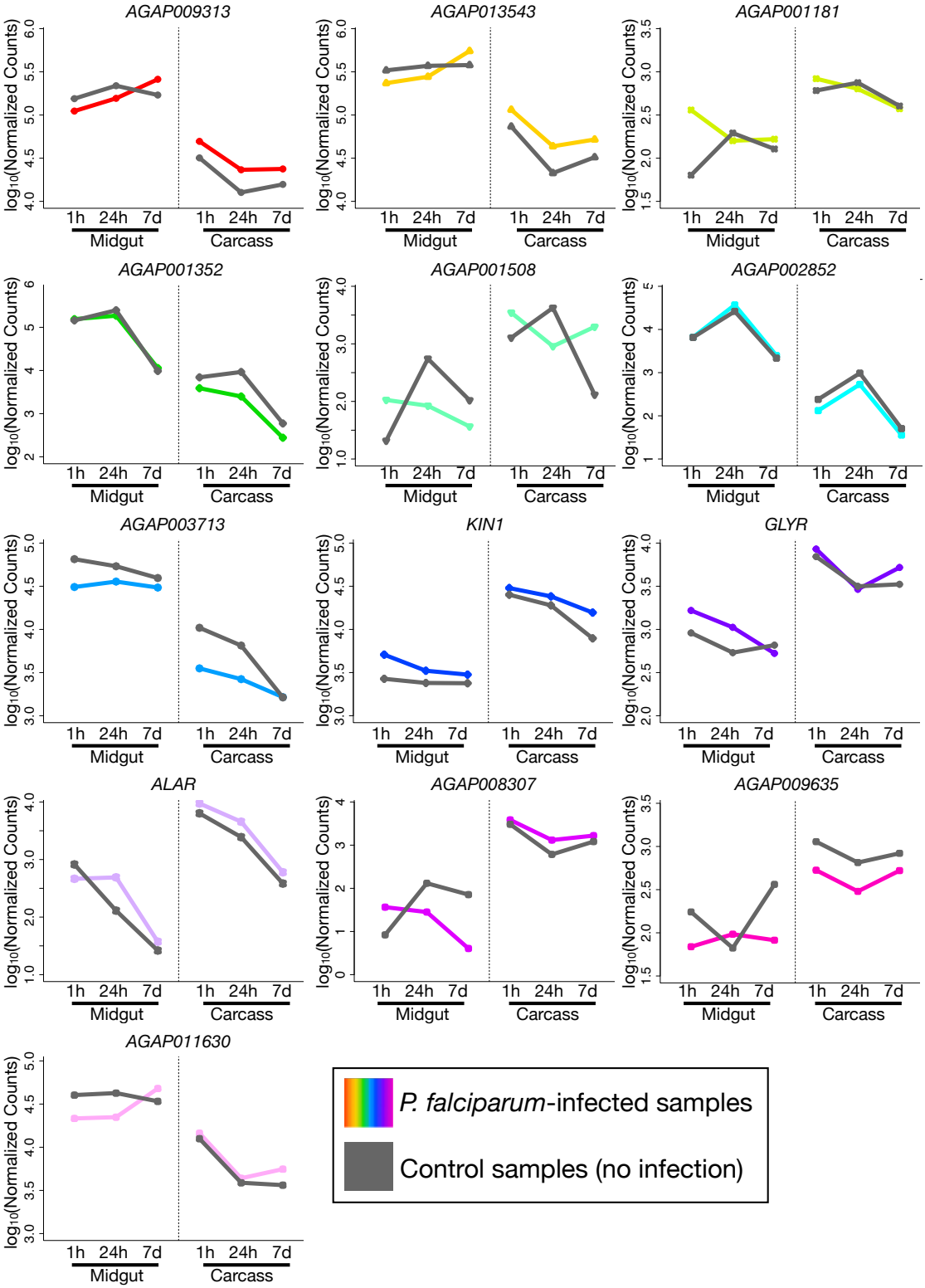


Figure 5.2.2. Average normalized counts of putative AMPs in the Ag/Pf RNA-seq data. *Normalized counts were log_{10}-transformed for presentation purposes. Colored lines represent expression in infected samples, dark grey lines represent expression in control samples. A total of three and two biological replicates were completed for midgut and carcass samples respectively.*

The third selection strategy utilized RNA-seq data of *An. darlingi* midguts infected with *P. vivax* generated by Dr. Bianca Carlos of UNESP, Brazil, in the context of a collaboration between Imperial College London (Prof. George K. Christophides and Dr. Dina Vlachou), UNESP (Dr. Jayme Souza-Neto), and ICEMER-Amazonia (Dr. Marta Moreno and Prof. Joseph Vinetz). In this present study, the raw sequencing data was reprocessed and analyzed using the same pipeline as the other two *An. gambiae* RNA-seq datasets. The reprocessed data yielded 1,812,028,976 pairs of reads, where an average of 88.3% aligned uniquely to the *An. darlingi* Coari genome. From these aligned reads, an average of 77.2% were associated with annotated transcripts. This dataset will be referred to as the '*Ad/Pv* data'.

The same initial selection criteria was applied, with the following exceptions: the adjusted *p*-value cut-off was increased to ≤ 0.1 (instead of ≤ 0.05), and no restrictions on $\log_2$ fold change between conditions were used. This analysis identified 30 sequences. The 1:1 orthologs in *An. gambiae* of these sequences were identified and these orthologs were assessed by the six algorithms. Approximately half of the genes passed the 4 of 6 algorithms threshold. Two genes were ultimately selected for subsequent analyses: AGAP001508, the ortholog of ADAC008083, which was also identified in the first selection strategy *i.e.* predicted to be an AMP in both *An. gambiae* datasets; and AGAP001352, the ortholog of ADAC009053, which was highly expressed in the cardia and posterior prior to blood feeding and at 5 hours post-blood meal respectively (Fig. 5.2.1). ADAC009053 was also predicted to have hemolymph juvenile hormone binding properties (InterPro ID: IPR010562).

The fourth and final strategy was to generate *de novo* transcriptome assemblies from the midgut sections RNA-seq data. Other insect species have dozens of AMPs comprising multiple classes, whereas only 11 AMPs have been identified in *Anopheles*. Thus, it was hypothesized that some novel AMP genes may exist, which are too short to be predicted by automatic gene prediction algorithms. Over 32,000 transcripts were predicted, compared to the ~15,000 transcripts currently annotated in VectorBase. Sequences were filtered with the same initial selection criteria as before, with the additional criteria of lacking of an association with a known gene, *i.e*. not an alternative transcript of an annotated gene, and significantly upregulated in the cardia compared to

the posterior midgut at any time point. In total, 12 transcripts fulfilled these criteria, 9 of which did not map to any known chromosome. The highest average counts value out of all of the novel transcripts was just 615.69. All transcripts were run through SignalP 6.0 (Teufel et al, 2022) to predict whether each sequence contained a signal peptide. No signal peptides were identified in any of the sequences (Table E.1). Therefore, no putative AMPs were identified under this strategy. All selected putative AMPs and a summary of the selection strategies are presented in Table 5.2.1 and Figure 5.2.3 respectively.

Table 5.2.1*. *Expression profile and average counts of selected putative AMPs. *Putative AMPs were identified from three RNA-seq datasets: two* An. gambiae *and one* An. darlingi*.*

Gene ID (AGAP-)	Source (RNA-seq)	Domains & Predictions	Expression Profile in *An. gambiae* RNA-seq Datasets		Average Counts		
					Infected		
			Infected (Midgut/Carcass)	Midgut	Midgut	Carcass	Midgut
001181	*Ag/Pf*: 4/6		1 h parasite, 24 h blood response	C/A > P	179	600	136
001352	*Ad/Pv*		24 h blood response, suppression in parasite presence	P > C/A	125,774	3,908	21,658
001508	All		1 h parasite, 24 h blood response	C > A > P	151	2,002	128
002852	*Ag/Pf*: 4/6	NPC2-like sterol transport	M: 24 h blood & parasite response; X: 24 h blood response, suppression in parasite presence	P > A > C	13,510	328	2,303
003713	Both *Ag*	MBF2 TF	1/24 h blood response, suppression in parasite presence	A > C/P	42,606	4,402	73,384
005888	Both *Ag*		1 h parasite response, upregulation in parasite presence	C > A > P	3,145	20,330	4,721
005889	Both *Ag*		M: 1/24 h parasite response; X: 1 h/7 d parasite response	C > A > P	891	5,017	1,503
005890	Both *Ag*		M: 1 h blood, 24 h parasite response; X: 1/24 h parasite response, upregulation in parasite presence	C > A > P	328	3,958	694
008307	*Ag/Pf*: 4/6		M: 1 h parasite response, 24 h blood response; X: 1 h blood and parasite response	C > A > P	47	1,947	99
009313	*Ag* Midgut	TsetseEP-like; DUF725	M: 7 d parasite response, 24 h blood response; X: 1 h parasite response, upregulation in parasite presence	C > A > P	178,293	26,124	277,368
009635	*Ag/Pf*: 4/6		M: 7 d blood response, suppression in parasite presence; X: 1h blood response, suppression in parasite presence	C > A > P	142	663	158
011630	Both *Ag*	MBF2 TF	M: 7 d parasite response, 1/24 h blood response; X: 1 h parasite response, upregulation in parasite presence	C > A > P	34,731	7,432	52,251
013543	*Ag* Midgut		M: 7 d parasite response; X: 1 h parasite response, upregulation in parasite presence	C > A > P	357,061	56,329	676,634

Ag/Pf: *4/6 refers to AMPs identified in the* Ag/Pf *RNA-seq data and were predicted to be AMPs by 4 of the 6 algorithms; '*Ad/Pv*' refers to the* An. darlingi/P. vivax *RNA-seq data; '*Ag *Midgut' refers to the* An. gambiae *midgut sections RNA-seq data; TF = transcription factor; h = hours post-infection/blood feeding, d = days post-infection/blood feeding, M = midgut, X = carcass of the* Ag/Pf *data; C = cardia, A = anterior, P = posterior of the midgut sections data. Average counts were calculated from all conditions and replicates of the respective RNA-seq data.*

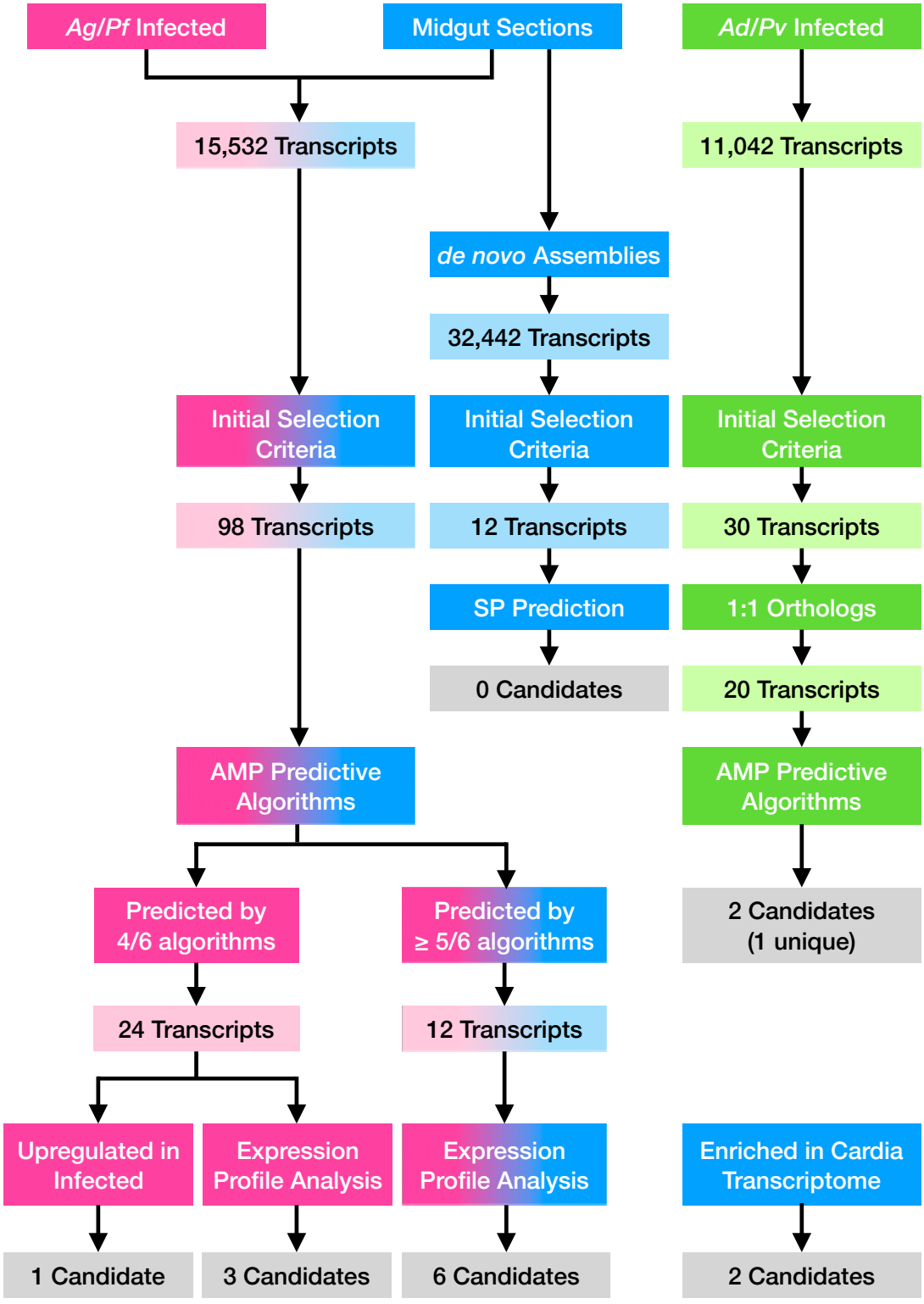


Figure 5.2.3. Putative AMP selection strategies. *Transcripts from RNA-Seq datasets were filtered using the initial selection criteria and processed through six AMP predictive algorithms to identify putative AMP candidates for further characterization. Additionally, two genes which were highly enriched in the cardia were included as candidates, and a* de novo *transcriptome assembly was generated to identify putative AMPs but did not yield any candidates. SP = signal peptide.*

Positive and negative control genes were used to validate the predictive algorithms. AMPs CEC1 (AGAP000693) and GAM1 (AGAP008645) were used as positive controls, and α-tubulin (AGAP010971) was used as a negative control. CEC1 was predicted to be an AMP by all 6 algorithms and GAM1 was predicted by 5 of 6 algorithms. Tubulin was predicted to be an AMP by half of the 6 algorithms.

AlphaFold utilizes a machine learning approach to predict the folding structures of proteins (Jumper et al, 2021) and was employed here to identify if any of the putative AMPs have secondary structures which are commonly associated with AMPs (*i.e.* α-helices and β-sheets). Structural predictions of the putative AMP locus genes typically had low confidence levels (Fig. 5.2.4). The confidence metric (pLDDT) of GLYR and KIN1 was very low (< 50), and no structures were predicted except an α-helix in the signal peptide region (Fig. 5.2.4F-G). ALAR on the other hand was predicted to have around four regions of α-helices (Fig. 5.2.4H). Structural predictions of AGAP001508 and AGAP009635 were also of low confidence and without any secondary structures predicted (Fig. 5.2.4C, K). AGAP001181, AGAP002852, AGAP003713, AGAP011630 were predicted to form multiple β-sheets with high confidence, whereas the majority of AGAP008307 and the two cardia genes form α-helices (Fig. 5.2.4). AGAP001352 was predicted to have both secondary structures (Fig. 5.2.4B). Almost the entirety of the sequence of the known AMPs CEC1 and DEF1 were predicted to be α-helical, whereas GAM1 has a series of β-sheets (Fig. 5.2.4N–P).

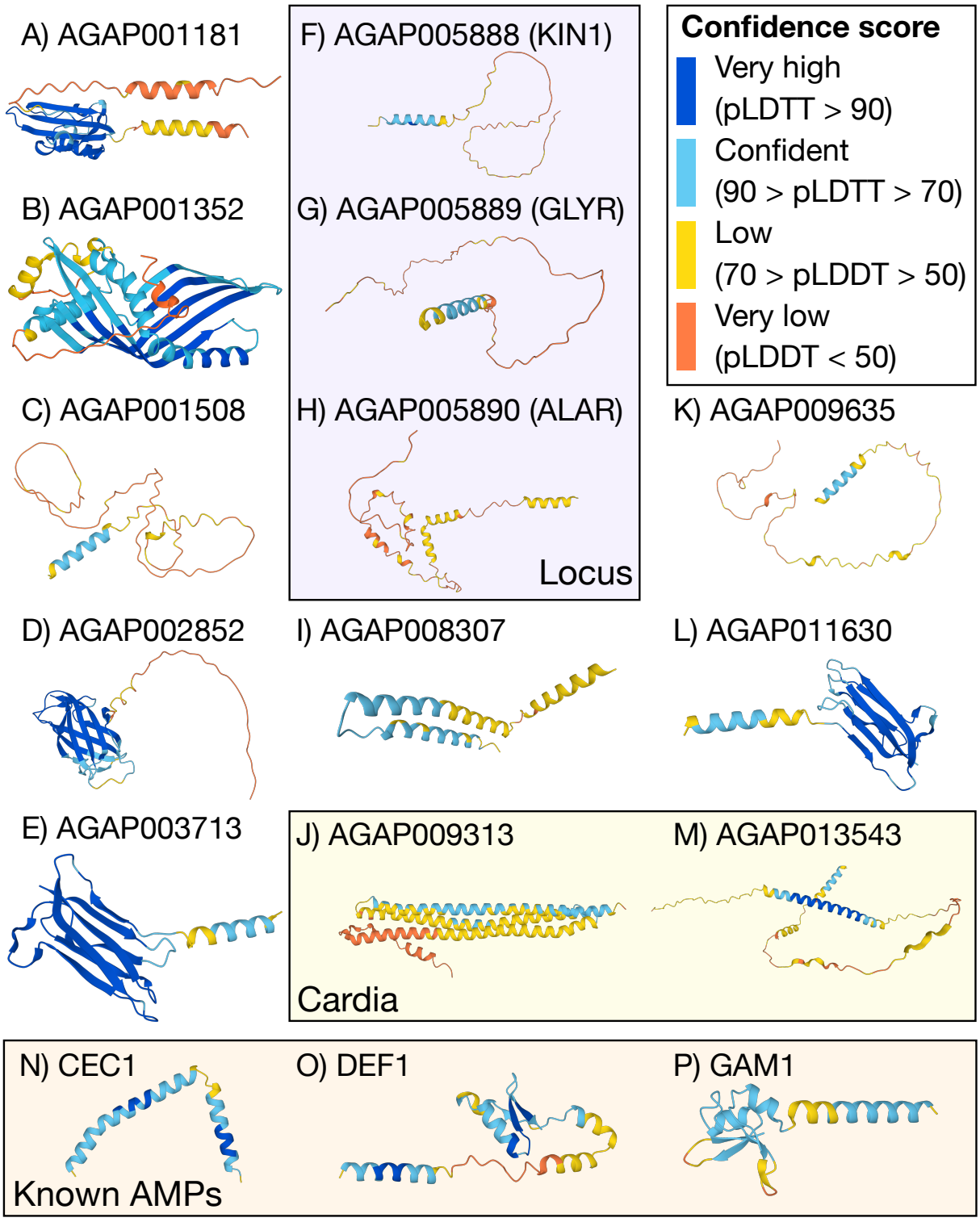


Figure 5.2.4. AlphaFold structural predictions of putative and known AMP proteins. *'Locus' putative AMPs were identified in both* An. gambiae *RNA-seq datasets and are located in the same locus. 'Cardia' putative AMPs were found to be highly expressed in the cardia. Residues are colored based on the confidence metric of their prediction. Rectangles with arrowheads represent β-sheets and curled twists represent α-helices. Structural predictions created with AlphaFold (Jumper et al, 2021) were obtained from Vectorbase (Amos et al, 2022).*

Using αCharges (Schindler et al, 2023), the AlphaFold predictions of the putative AMPs were colored to represent the relative charges of the residues in each polypeptide sequence, enabling the identification of regions with amphipathicity (Fig. 5.2.5). Known AMP CEC1 exhibits amphipathicity on some regions of its α-helical structure, where one side of the α-helix is positively-charged, which contributes to its antimicrobial effects. The α-helices of many putative AMPs contain regions made up of charged residues, including the cardia genes (Fig. 5.2.5J, M), AGAP008307 (Fig. 5.2.5I), and AGAP001352 (Fig. 5.2.5B). Many regions of AGAP009313 appear to possess small regions of amphipathicity, in addition to a mixture of positive- and negatively-charged residues in close proximity (Fig. 5.2.5J). However, none of these peptides exhibit discrete amphipathicity as that of CEC1. As the predicted folding structures of many of these putative AMPs were not of high confidence (Fig. 5.2.4), the projection of charges onto these structural predictions may not fully represent an accurate model.

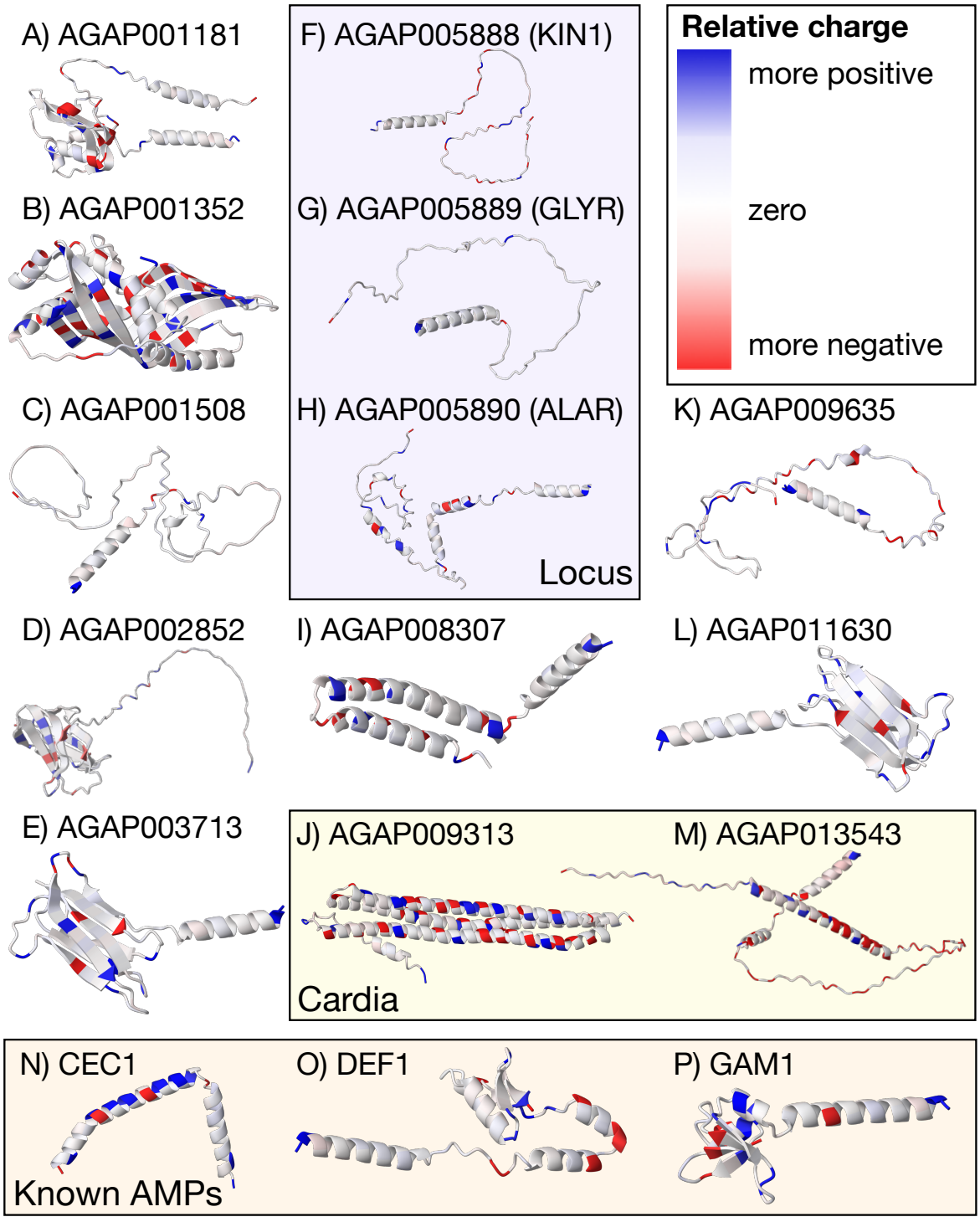


Figure 5.2.5. Charges and predicted structures of putative and known AMP proteins. *'Locus' putative AMPs were identified in both* An. gambiae *RNA-seq datasets and are located in the same locus. 'Cardia' putative AMPs were found to be highly expressed in the cardia. Residues are colored based on their relative charge, with bluer colors being more positively-charged, and redder colors being more negatively-charged. Rectangles with arrowheads represent β-sheets and curled twists represent α-helices. Structures and relative charges were computed using AlphaCharge (Schindler et al, 2023).*

5.3 Putative AMP screen: antibacterial activity

In order to screen the putative AMPs for antimicrobial activity, various infection assays were carried out. First, a bacterial survival assay was performed, where the putative AMPs belonging to the same locus (*ALAR*, *GLYR*, *KIN1*) were silenced via RNAi and the mosquitoes were subsequently infected with various species of bacteria. The number of mosquito deaths were recorded daily until all mosquitoes had died. This was performed alongside the known AMPs *LYSC1* and *GAM1*, and *LacZ* was used as a control.

Mosquitoes were infected with one of four species of bacteria after (putative) AMP-silencing: *S. marcescens* and *E. faecalis* were introduced orally through sugar feeding to initiate a midgut (local) infection, and *E. coli* and *S. aureus* were introduced through thoracic injection to initiate a systemic infection. Silencing of each gene was successful and the average KD efficiencies are listed in Table F.1. Survival curves of putative AMP KD and *dsLacZ*-injected control mosquitoes were scored as HRs with respect to the control, *i.e.* HR > 1 indicates reduced survival of the dsLacZ-injected controls, or conversely increased survival in putative AMP KD mosquitoes and vice versa.

The KD of *ALAR* and *KIN1* significantly increased mosquito susceptibility to *S. marcescens* infection in the midgut (Cox proportional hazard analyses, HR [with respect to control mosquitoes] < 1, $p < 0.05$; Fig. 5.3.1; Table F.1). Survival following *E. coli* infection was significantly reduced with *KIN1* silencing (Cox proportional hazard analyses, HR [with respect to control mosquitoes] = 0.79, 95% CIs = 0.65–0.97, $p = 0.03$; Fig. 5.3.1; Table F.1). Contrastingly, *GAM1* KD led to increased survival following *S. aureus* infection, although this was insignificant (Cox proportional hazard analysis, HR [with respect to control mosquitoes] = 1.21, 95% CIs = 0.96–1.51, $p = 0.1$; Fig. 5.3.1; Table F.1).

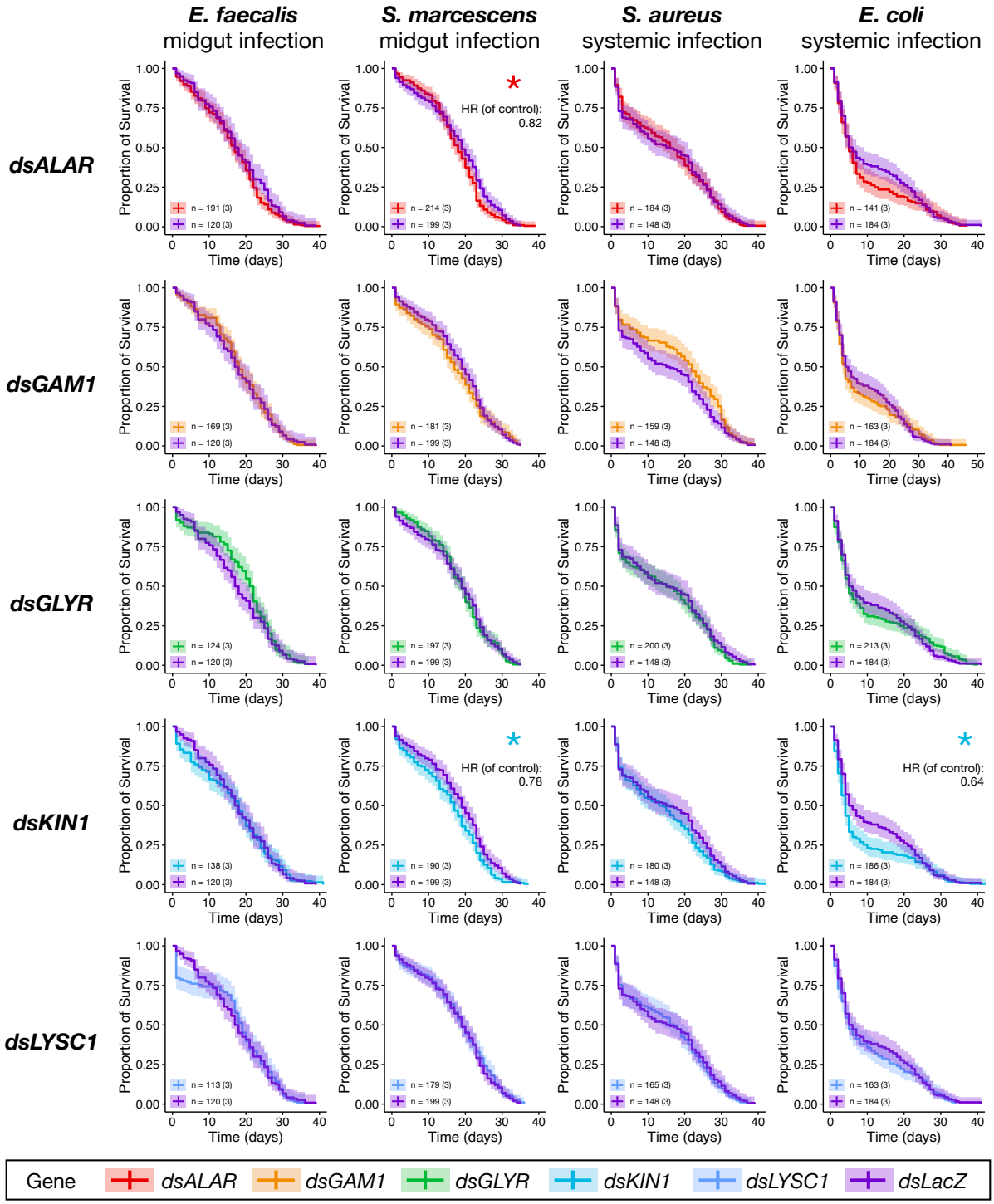


Figure 5.3.1. Kaplan–Meier survival curves of mosquitoes post-locus AMP KD*. Mosquitoes were subjected to (putative) AMP KD via RNAi, and infected with one species of bacteria two days later. The left two species of bacteria were introduced to the mosquito through a sugar meal. The right two species of bacteria were introduced via thoracic injection. Mosquito deaths were recorded daily until all mosquitoes had died. Shaded regions represent 95% CIs. HR = hazard ratio (with respect to controls), n = number of mosquitoes, parentheses indicate number of replicates. Cox proportional hazard analyses, * p < 0.05. Full model results are listed in Table F.1.*

Over the course of the survival assay, it was expected that some mosquitoes eventually will have recovered from infection. Thus, in order to assess the effect of each putative AMP on the shorter-term resistance against infection, mosquito survival over a period of ten days was analyzed (Fig. 5.3.2; Fig. F.1).

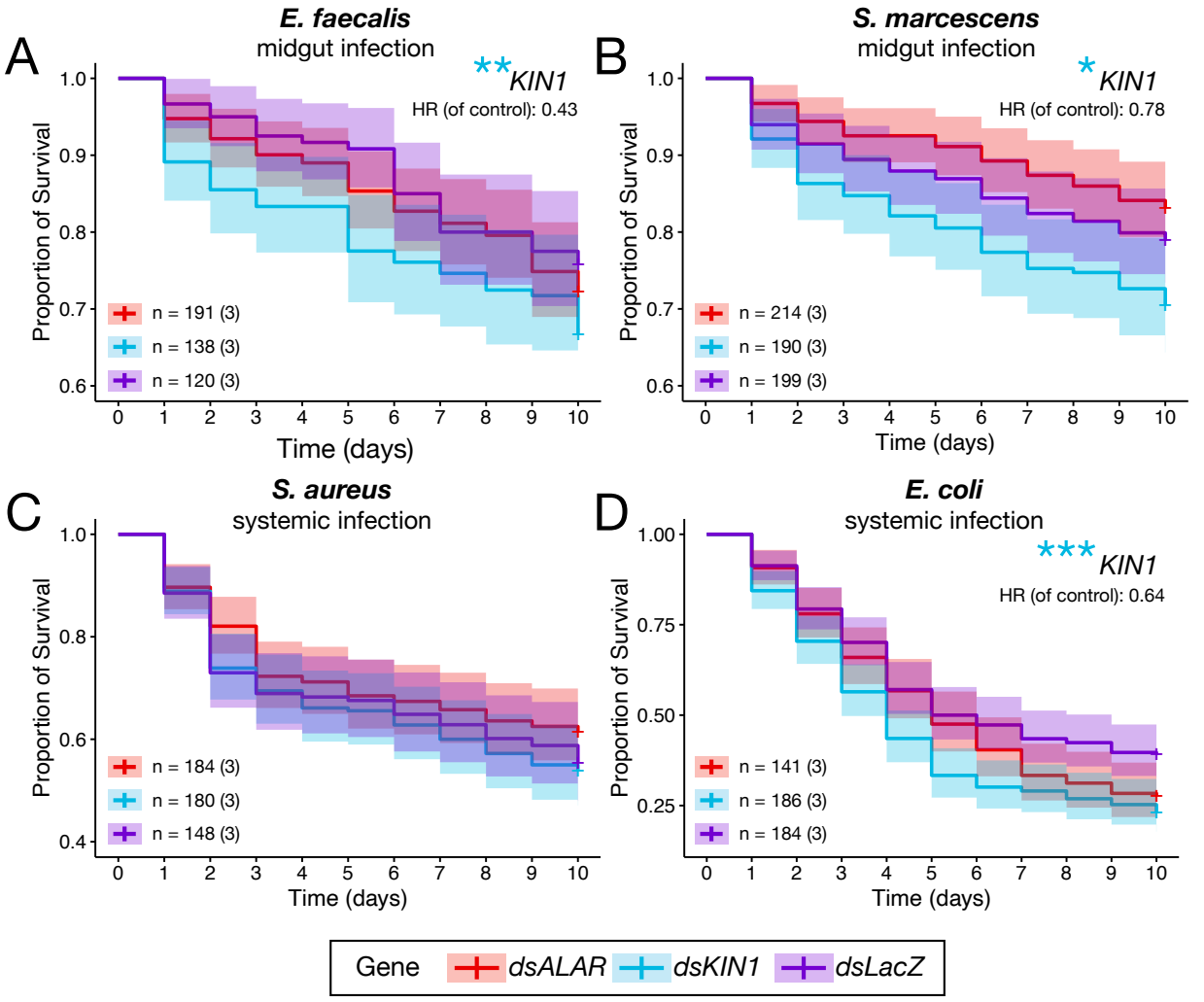


Figure 5.3.2. Kaplan–Meier survival curves (10 days) of mosquitoes post-locus AMP KD. *Mosquitoes were subjected to (putative) AMP via RNAi, and infected with one species of bacteria two days later. The top two species of bacteria were introduced to the mosquito through a sugar meal. The bottom two species of bacteria were introduced via thoracic injection. Mosquito deaths were recorded daily until 10 days post-infection. Shaded regions represent 95% CIs. HR = hazard ratio (with respect to controls), n = number of mosquitoes, parentheses indicate number of replicates. Cox proportional hazard analyses, * p < 0.05, ** p < 0.01, *** p < 0.001. Full model results are listed in Table F.1.*

In all infections, *KIN1* KD reduced survival although this was only significant for infections with 3 of the 4 bacterial species (Cox proportional hazard analyses, HR [with respect to control mosquitoes] < 1, $p > 0.05$; Fig. 5.3.2; Table F.1). Depletion of *GAM1* reduced survival against Gram-negative bacteria, whereas *GLYR* silencing improved survival following oral infections but had the opposite effect on systemic infections (Fig. F.1). *ALAR* and *LYSC1* KD showed similar effects, where survival was reduced in *E. faecalis* and *E. coli* infection, but improved survival following *S. marcescens* and *S. aureus* infection (Fig. 5.3.2; Fig. F.1; Table F.1). The improvement of survival outcomes in *ALAR* KD mosquitoes contradicts the significant reduction in survival over the long term in *S. marcescens* infection. Taken together, *KIN1* plays a role in limiting bacterial infections of both Gram types, whereas the other genes have mixed effects on different types of bacterial infections.

Next, the two novel cardia genes were examined under the same assay. KD of the two cardia genes were successful, averaging 53% for AGAP009313 and 60% for AGAP013543 (Table F.1). Over the course of mosquito adulthood, depletion of AGAP0013543 improved survival in *S. marcescens*-infected mosquitoes for the majority of mosquitoes, but survival rates did not differ from controls in mosquitoes living over 22 days post-infection. Overall, KD of AGAP0013543 had no significant effect on mosquito survival following infection with any bacterial species tested (Cox proportional hazard analyses, HR ≈ 1, $p > 0.05$; Fig. 5.3.3; Table F.1). Interestingly, KD of AGAP009313 improved survival after infection with three of the four bacterial species, although this was only significant for *S. marcescens* and *E. coli* infection (Cox proportional hazard analyses, HR [with respect to control mosquitoes] > 1, $p < 0.05$; Fig. 5.3.3; Table F.1).

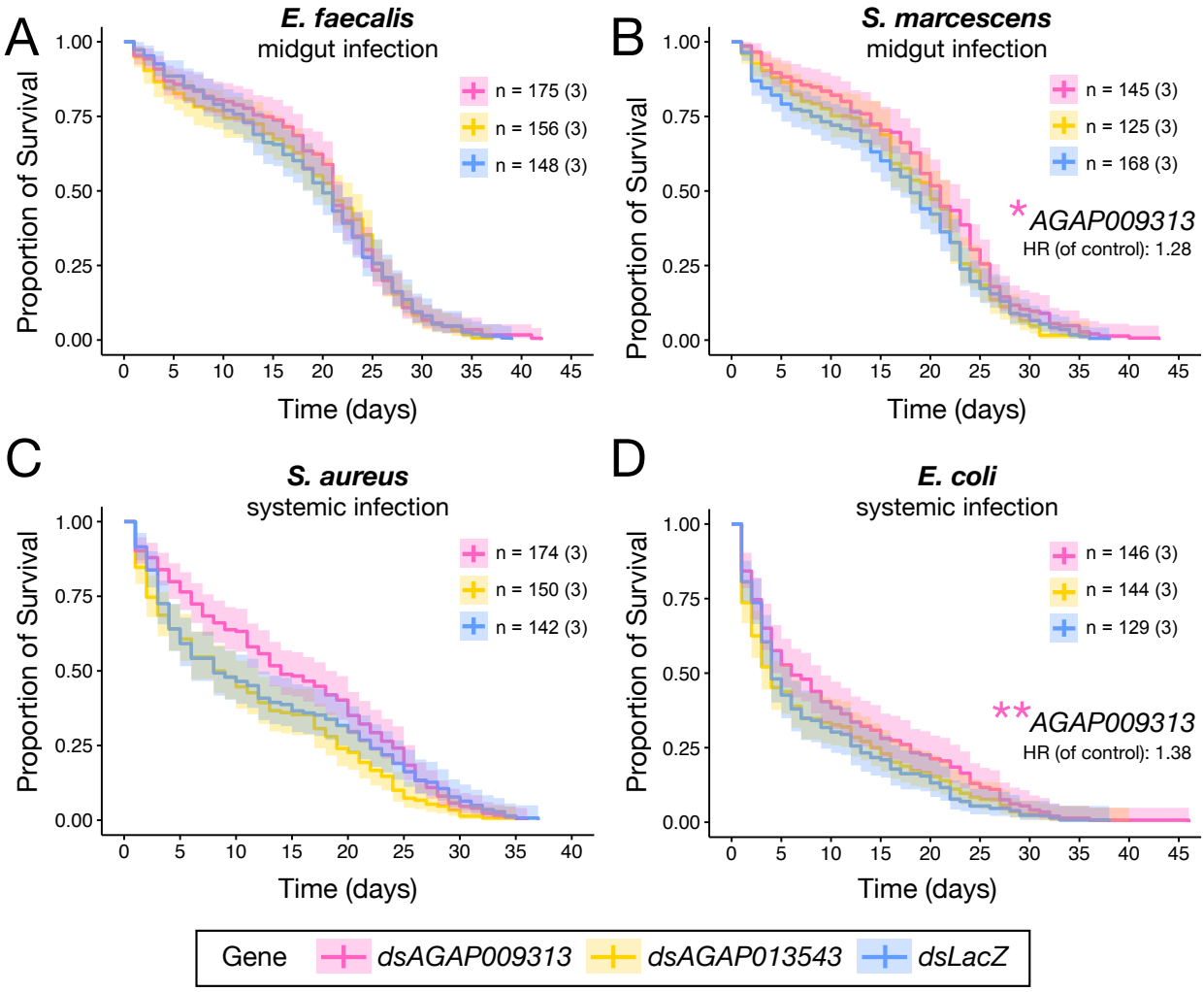


Figure 5.3.3. Kaplan–Meier survival curves of mosquitoes post-cardia gene KD. *Putative AMPs enriched in the cardia were knocked down in mosquitoes via RNAi and infected with one species of bacteria two days later. The top two species of bacteria were introduced to the mosquitoes through a sugar meal. The bottom two species of bacteria were introduced via thoracic injection. Mosquito deaths were recorded daily until all mosquitoes had died. Shaded regions represent 95% CIs. HR = hazard ratio (with respect to controls), n = number of mosquitoes, parentheses indicate number of replicates. Cox proportional hazard analyses, * p < 0.05, ** p < 0.01. Full model results are listed in Table F.1.*

In the short-term, silencing of AGAP009313 significantly improved survival following *S. marcescens* and *S. aureus* infection (Cox proportional hazard analyses, HR [with respect to control mosquitoes] > 1, $p < 0.05$; Table F.1; Fig. F.2). These results suggest a role of AGAP009313 in bacterial infections of both Gram-types, however a batch effect was also present, as replicate had a significant impact on survival post-infection (Cox proportional hazard analyses, $p < 0.05$; Table F.1).

Next, the antibacterial activity of the remaining selected putative AMPs were tested. MRes student Yunyue (Zoe) Zhou (project supervised by Dr. Dina Vlachou) contributed towards the data collection of the subsequent experiments outlined in this section. Here, putative AMP-silenced mosquitoes were injected with either *S. marcescens* or *E. coli*, and survival was measured over a period of 10 days. Injection of *S. marcescens* drastically increased mortality over the course of 10 days, with survival rates reaching 25% at around 10 days compared to around 25 days using oral infections. Effects of putative AMP KD were mixed across the different genes and bacteria species (Fig. 5.3.4). Only AGAP001508 KD had a significant effect, where its depletion improved survival following *S. marcescens* infection compared to controls (Cox proportional hazard analysis, HR [with respect to control mosquitoes] = 1.63, 95% CIs = 1.10–2.41, $p = 0.02$; Table F.1).

Interestingly, replicate was a significant predictor of survival for many of the survival curves, but particularly with regard to those involving bacterial infections via injection (Cox proportional hazard analyses, $p < 0.05$; Table F.1). This could be attributed to variability between colonies of mosquitoes to tolerate injection.

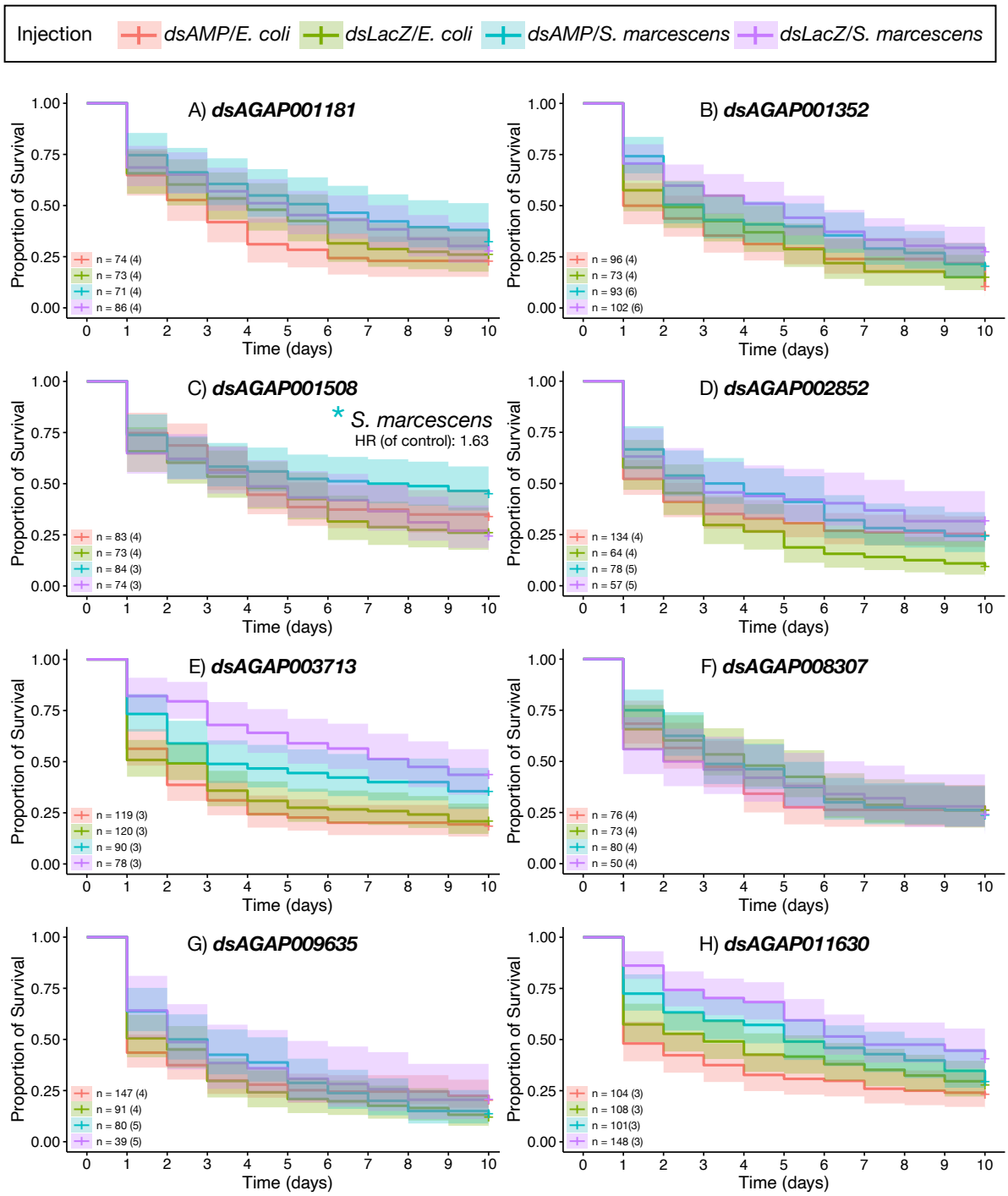


Figure 5.3.4. Kaplan–Meier survival curves of mosquitoes post-putative AMP KD. *All remaining putative AMP candidates were silenced in the mosquito via RNAi and infected with one species of bacteria two days later. Both species of bacteria were introduced to the mosquito via thoracic injection. Mosquito deaths were recorded daily for 10 days post-infection. Shaded regions represent 95% CIs. MRes student Yunyue (Zoe) Zhou contributed towards the data collection of some replicates (project supervised by Dr. Dina Vlachou). HR = hazard ratios (with respect to controls), n = number of mosquitoes, parentheses indicate number of replicates. Cox proportional hazard analyses, * p < 0.05. Full model results are listed in Table F.1.*

5.4 Putative AMP screen: induction post-bacterial challenge

Many putative AMPs exhibited antibacterial effects in the previous assays. Additionally, the expression levels of known AMPs were also previously shown to be modulated in response to bacterial challenge (section 4.5). Thus, expression levels of the putative AMPs were assessed post-bacterial challenge to identify whether they are bacteria-responsive. Mosquitoes were subjected to *E. coli* challenge (via thoracic injection) or midgut *S. marcescens* challenge (via sugar feeding), and AMP expression was measured from samples collected 48 hours later via qRT-PCR.

Among the three putative AMP locus genes, depletion of *ALAR* and *KIN1* led to significantly reduced survival rates following *E. coli* and *S. marcescens* infections. However, following bacterial challenge of these species, *ALAR* and *KIN1* expression was mostly unchanged (Fig. 5.4.1). The exception lied within the carcass of *S. marcescens*-infected mosquitoes, where expression was consistently downregulated by an average of 40–60% following infection (Fig. 5.4.1). The same observation was made for *GLYR*, but none of these differences were significant (Student's *t*-tests and Wilcoxon rank-sum tests, $p > 0.05$; Fig. 5.4.1). Nevertheless, relative expression values of all three putative AMPs were enriched in the carcass compared to whole midguts, which is in agreement with the *Ag/Pf* transcriptomic data.

The expression of the two cardia genes was also measured. The near opposite effect was observed, where expression was on average increased up to 20-fold in the carcass following infection of *S. marcescens*. Despite this, these differences were insignificant, and as expression of these two genes is concentrated in the cardia, such drastic changes could reflect relatively minute changes in low expression levels. Of the remaining 8 putative AMPs, two showed significant changes in expression following infection (Fig. 5.4.1; Fig. G.1). *AGAP001508* expression was significantly reduced following *E. coli* infection (Student's *t*-test, $t = 7.5$, df = 2.4, $p = 0.01$). Previously, depletion of this gene improved survival outcomes following systemic *S. marcescens* infection after its KD. Expression of *AGAP008307*, which is abundant in the carcass, was significantly reduced following *S. marcescens* infection in the carcass (Wilcoxon rank-sum test, $W = 0$, $p = 0.04$).

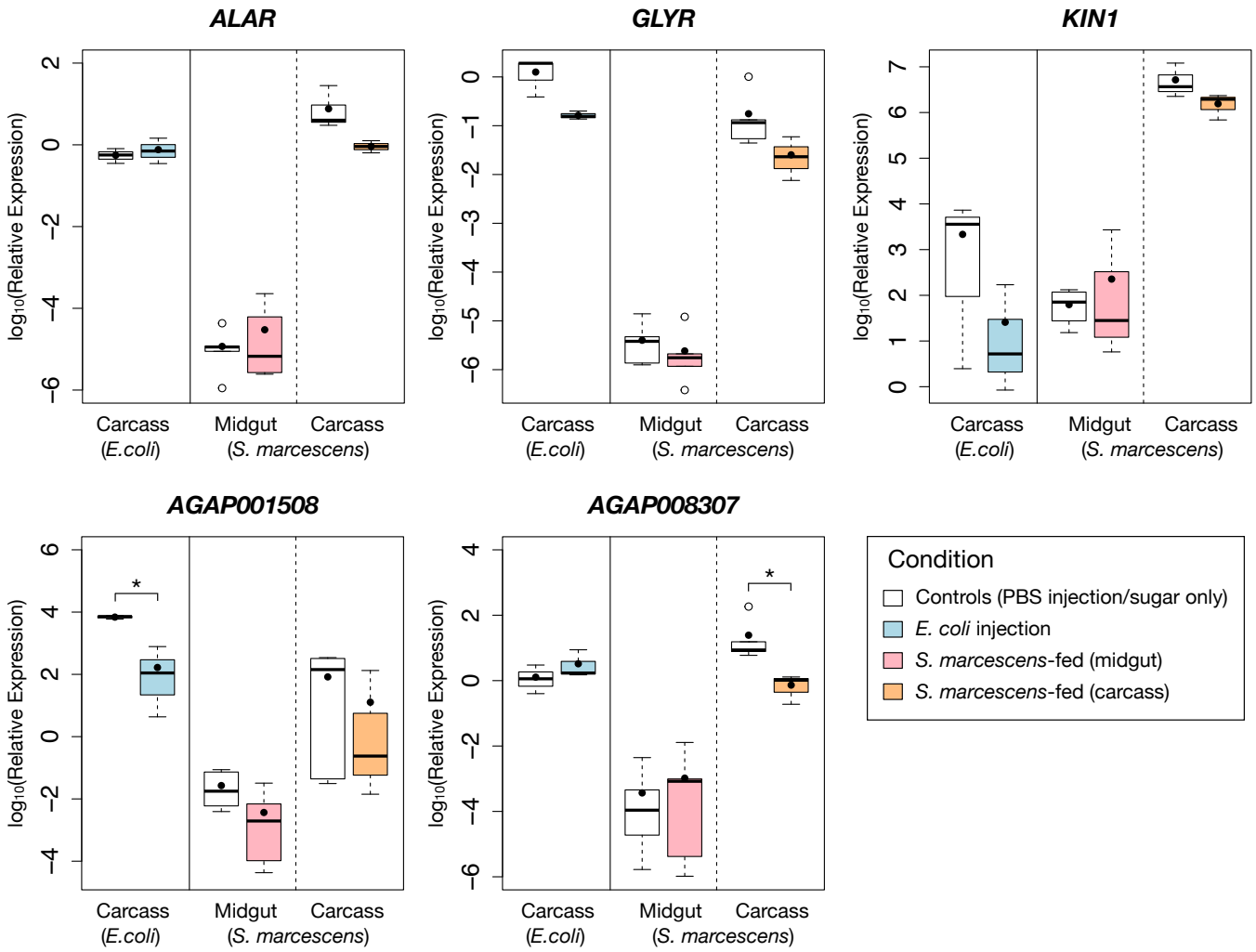


Figure 5.4.1. Relative expression of putative AMPs post-bacterial challenge. *The three genes at the top belong to the same locus. Controls for* E. coli*-injected mosquitoes were PBS-injected mosquitoes. Controls for* S. marcescens*-fed mosquitoes were fed with sugar only. Relative expression values were* log_{10}*-transformed for presentation purposes and are relative to a pool of mosquito cDNA. A total of 3–5 replicates were performed with 10–15 mosquito individuals per condition for each replicate. Black bars represent medians, dots represent means. Differences in the original relative expression values were compared with Student's* t*-tests or Wilcoxon rank-sum tests,* * $p < 0.05$.

5.5 Putative AMP screen: anti-*Plasmodium* activity

The antimalarial activity of the putative AMPs were investigated next. Similar to the bacterial infections, putative AMPs and the cardia genes were silenced in mosquitoes, and mosquitoes were infected with *P. falciparum* three days later. Here, *LacZ* was also used as a control. Midgut oocysts were counted at 8 to 9 days post-infection to assess the antimalarial effects of the putative AMPs. The infection statistics and KD efficiencies are summarized in Table 5.5.1 and Figure 5.5.1. Wilcoxon rank-sum tests were employed to identify whether differences in oocyst infection intensity in infected midguts (*i.e.* midguts without oocysts were excluded) were significant, and χ^2 tests were used to assess differences in infection prevalence. MRes student Yunyue (Zoe) Zhou (project supervised by Dr. Dina Vlachou) contributed towards the collection of data for some of the genes in this section.

Table 5.5.1.* P. falciparum *infection statistics and average KD efficiencies of putative AMP KD mosquitoes. *At least three biological replicates were completed for each gene. Statistics refer to the infection intensity (Wilcoxon rank-sum tests) or prevalence (χ^2 tests) of the putative AMP KD mosquitoes compared to that of* dsLacZ*-injected control mosquitoes. GLMs were modelled on the effect of gene KD and replicate on infection intensity, model statistics are listed in Table H.1. MRes student Yunyue (Zoe) Zhou (project supervised by Dr. Dina Vlachou) contributed towards data collection for some of the replicates.*

Gene ID/Name	KD eff. (%)	n	Oocysts				Wilcoxon Rank-Sum Tests		χ^2 Tests		
			Max.	Average	Median	Prev. (%)	*W*	*p*-value	χ^2	df	*p*-value
AGAP001181	39	118	151	14.57	1.5	68	4063	0.19	6.14	1	0.01 *
LacZ	N/A	171	182	6.09	1	53	-	-	-	-	-
AGAP001352	64	94	83	5.00	0	37	572	0.17	0.75	1	0.39
LacZ	N/A	92	182	8.26	0	43	-	-	-	-	-
AGAP001508	74	54	51	4.91	1	54	842.5	0.78	0.11	1	0.74
LacZ	N/A	110	85	3.02	1	51	-	-	-	-	-
AGAP002852	60	123	128	4.98	0	43	1789	0.55	1.68	1	0.20
LacZ	N/A	141	182	5.76	1	51	-	-	-	-	-
AGAP003713	65	77	212	10.45	1	53	1422	0.04 *	0.10	1	0.75
LacZ	N/A	110	85	3.02	1	51	-	-	-	-	-
KIN1	85	71	46	5.37	0	48	969	0.10	2.89	1	0.09
LacZ	N/A	76	85	4.17	1.5	62	-	-	-	-	-
GLYR	90	132	130	6.10	1	58	2747	0.19	5.73	1	0.02 *
ALAR	65	116	48	4.10	0	39	1783	0.06	0.53	1	0.47
LacZ	N/A	148	85	2.90	0	43	-	-	-	-	-
AGAP008307	80	71	76	6.58	1	56	1329.5	0.36	0.11	1	0.74
LacZ	N/A	102	85	5.02	1	59	-	-	-	-	-
AGAP009635	59	88	131	9.77	1	57	1948	0.44	0.72	1	0.40
LacZ	N/A	141	182	5.76	1	51	-	-	-	-	-
AGAP011630	52	118	50	2.62	0	41	1435	0.98	0.31	1	0.58
LacZ	N/A	136	85	2.29	0	44	-	-	-	-	-
AGAP009313	56	109	51	3.85	1	59	1318	0.71	13.9	1	1.96×10^{-4} ***
AGAP013543	59	93	289	7.56	1	53	910	0.26 *	7.31	1	0.01 **
LacZ	N/A	125	44	2.59	0	34	-	-		-	-

KD eff. = knockdown efficiency calculated from whole mosquitoes, R = number of replicates performed, n = number of midguts dissected, max. = maximum # of oocysts, prev. = infection prevalence, df = degrees of freedom; * $p < 0.05$, *** $p < 0.001$.

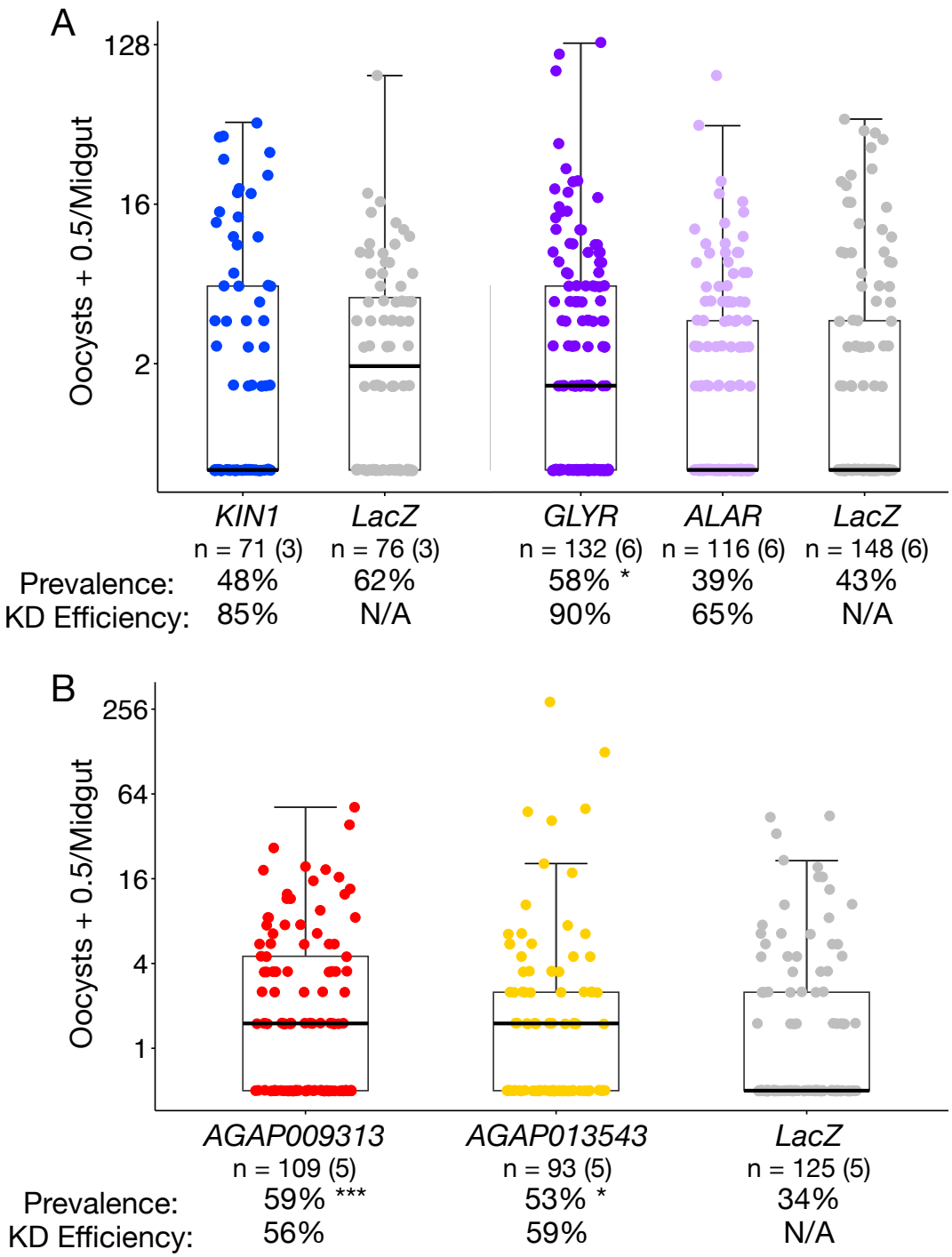


Figure 5.5.1. The effect of putative AMP (locus and cardia genes) KD on* P. falciparum *oocyst development. *A) Locus genes, B) cardia genes. Infections were performed three days post-KD, and oocysts were counted at eight days post-infection. Each dot represents the total number of oocysts in one midgut. Prior to plotting against a $\log_2$-transformed y-axis, 0.5 was added to each oocyst count value. Thick black lines within box plots indicate the median number of oocysts. The number of replicates performed are listed in parentheses. No significant differences were observed in infection intensity (Wilcoxon rank-sum tests performed on midguts with non-zero counts). Infection prevalence: χ^2 tests, * $p < 0.05$, *** $p < 0.001$.*

The KD of *GLYR*, AGAP001181, and both cardia genes significantly increased the prevalence of *P. falciparum*, indicating a potential antimalarial role of the peptides (Table 5.5.1; Fig. 5.5.1; Fig. 5.5.2). Additionally, only silencing of AGAP003713 had a significant effect on oocyst infection intensity amongst all of the putative AMPs (Table 5.5.1; Fig. 5.5.1; Fig. 5.5.2).

Considering the great variation in infection statistics between replicates, GLMs with quasi-Poisson distributions were built to examine the effect of replicate and gene KD on infection intensities. Indeed, replicate was a significant explanatory variable for a staggering 11 of the 13 putative AMP genes assayed, highlighting large differences between replicates (*F*-tests, $p < 0.05$). The GLMs built for AGAP001181 and *ALAR* also revealed a significant effect of gene KD on oocyst infection intensities (*F*-tests, $p < 0.05$; Table H.1). Mean and average oocyst infection intensities of infected midguts (*i.e.* excluding midguts with zero oocysts) in these two gene KD conditions were greater than that of *dsLacZ*-injected control groups. These results contribute to the fact that these genes may have antimalarial activity (Table 5.5.1; Fig 5.5.2; Table H.1). GLMs also indicated that AGAP003713 KD also had a significant impact on infection intensity, however replicate and the interaction between gene KD and replicate were also significant predictors (*F*-tests, $p < 0.05$; Table H.1). This suggests that AGAP003713 may not have antimalarial activity, however additional replicates are needed to confirm this result.

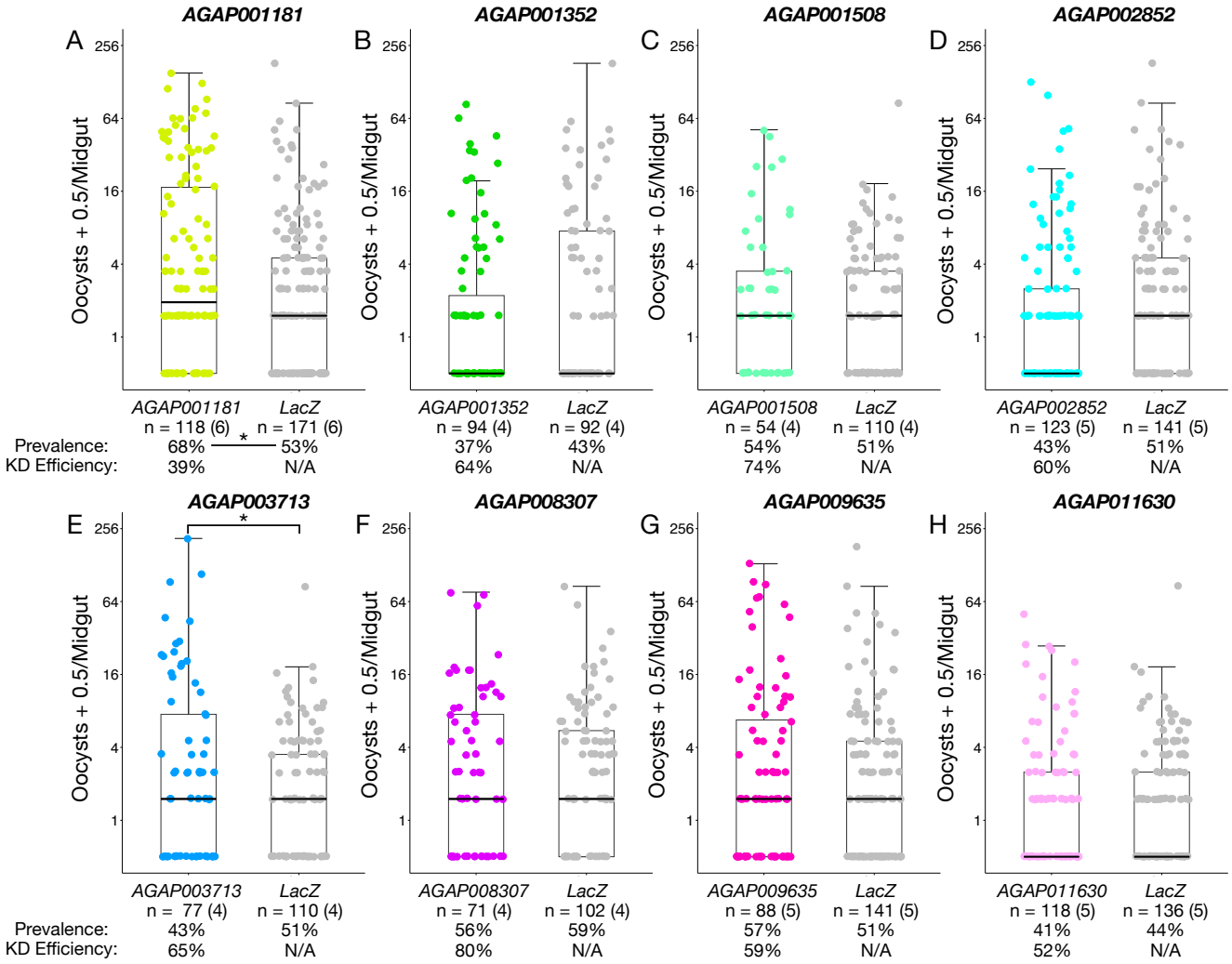


Figure 5.5.2. The effect of putative AMP KD on* P. falciparum *oocyst development. *Infections were performed three days post-KD, and oocysts were counted at eight days post-infection. Each dot represents the total number of oocysts in one midgut. Prior to plotting against a $\log_2$-transformed y-axis, 0.5 was added to each oocyst count value. Thick black lines within box plots represent the median number of oocysts. MRes student Yunyue (Zoe) Zhou contributed towards the data collection of some replicates (project supervised by Dr. Dina Vlachou). The number of replicates performed are listed in parentheses. Infection intensity: Wilcoxon rank-sum (performed on midguts with non-zero counts), infection prevalence: χ^2 tests, * $p < 0.05$.*

5.6 Putative AMP locus genes throughout mosquito development and tissues

Given that the members of the putative AMP locus (*ALAR*, *GLYR*, and *KIN1*) were found to have antimicrobial effects, their expression throughout mosquito development and tissues were further characterized. If they are to be used as effectors or a locus for the generation of transgenic mosquitoes, it is important that their expression throughout development is understood, as modifying their expression or introducing additional effectors into the locus may impact mosquito development. The relative expression of these genes was measured using qRT-PCR and compared to that of the known AMP *GAM1* (Fig. 5.6.1).

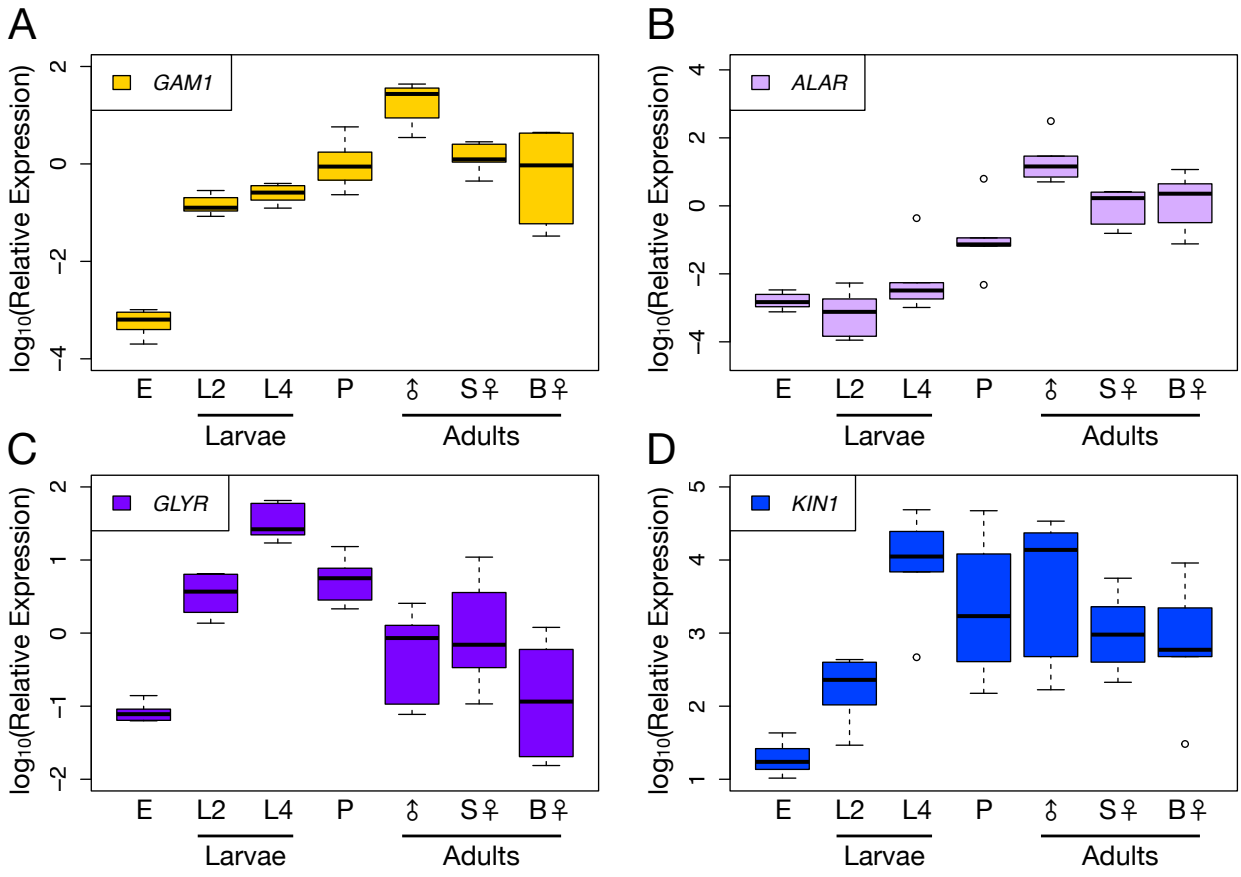


Figure 5.6.1. Relative expression of* GAM1 *and putative locus AMPs throughout mosquito development. *Expression values were log_{10}-transformed for presentation purposes and are relative to expression of a pool of mosquito cDNA. A total of 6 biological replicates were performed using at least 10 mosquito individuals for each life stage. E = eggs, L2 = second instar larvae, L4 = fourth instar larvae, P = pupae, S = sugar-fed adults, B = blood-fed adults.*

The expression of *GAM1* and the putative AMPs changed significantly throughout development (Kruskal–Wallis tests, $p < 0.05$; Table I.1). *GAM1* and *ALAR* expression increased significantly throughout development (Table I.1; Fig. 5.6.1A–B). Amongst adults, median expression levels of these genes were higher in males than females of both feeding conditions, although significantly only for *ALAR* between male and sugar-fed females (Dunn's test with Benjamini–Hochberg false discovery rate corrections, $Z = 2.99$, $p = 0.01$). Similarly, *KIN1* expression also significantly increased throughout mosquito development and was higher in males than females (Table I.1; Fig. 5.6.1D). *GLYR* expression peaked in the final larval stage, and median expression was marginally higher in males than females (Table I.1; Fig. 5.6.1C). All pairwise comparisons and the associated statistical test results are listed in Table I.1.

The expression of these genes was analyzed via qRT-PCR in additional midgut tissue samples to confirm the RNA-seq results, and also in the remaining mosquito carcass to assess expression in the fat body/hemolymph. The qRT-PCR results confirm the spatial expression profiles observed in the RNA-seq data, where expression was higher in more anterior tissues compared to the posterior midgut (Table I.1; Fig. 5.2.1).

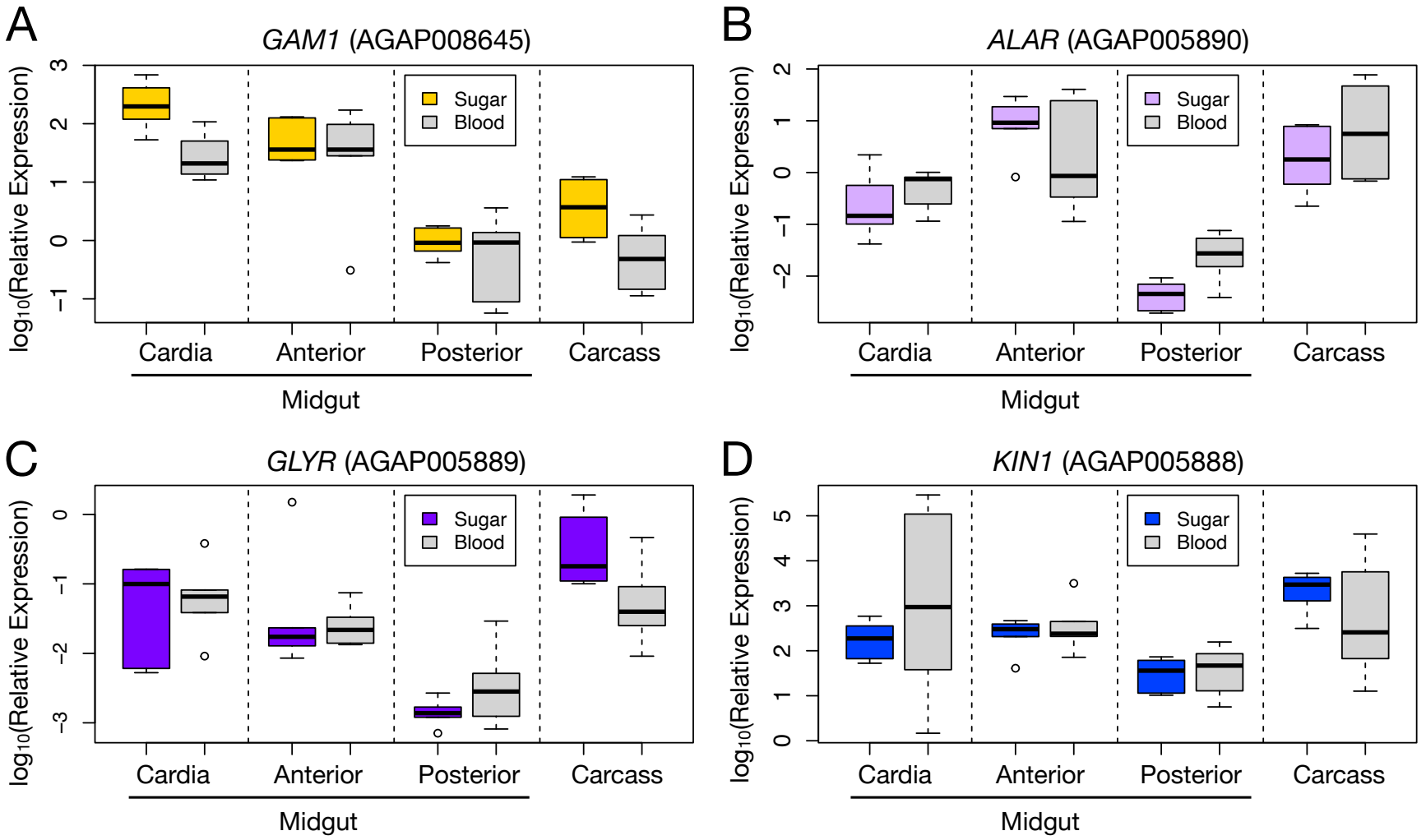


Figure 5.6.2. Relative expression of* GAM1 *and putative locus AMPs in mosquito midgut tissues and the carcass before and after blood feeding. *Sugar refers to sugar-fed tissues,* i.e. *tissues collected prior to blood feeding, blood refers to tissues collected 24 hours post-blood meal. A total of 6 biological replicates were performed using 15-30 mosquito individuals for each condition. Expression values were log_{10}-transformed for presentation purposes and are relative to expression of a pool of mosquito cDNA.*

Relative expression values of all genes were significantly different across the four different tissues sampled post-blood feeding (Kruskal–Wallis tests, $p < 0.05$; Table I.2). Along the midgut axis, *GAM1* expression was significantly enriched in the cardia compared to the anterior and posterior regions (Dunn's tests with Benjamini–Hochberg false discovery rate corrections, $p < 0.05$; Table I.2). Expression in the carcass was comparable to that in the posterior and significantly reduced compared to the anterior and cardia (Table I.2). In the cardia, *GAM1* expression was significantly reduced following blood feeding (Wilcoxon rank-sum test, $W = 1$, $p = 0.004$; Fig. 5.6.2). In the carcass, *GAM1* expression was also downregulated post-blood meal (Wilcoxon rank-sum test, $W = 5$, $p = 0.04$; Fig. 5.6.2). Expression in the anterior and posterior was not impacted by blood feeding (Wilcoxon rank-sum tests, $p > 0.05$; Fig. 5.6.2).

ALAR expression peaked in the anterior under sugar- and blood-fed conditions (Fig. 5.6.2), which differs from what was observed in the midgut sections RNA-seq data, where *ALAR* expression peaked in the cardia (Fig. 5.2.1). However, *ALAR* expression in the cardia was comparable to that of the anterior at 24 hours post-blood meal, which was observed with in both the RNA-seq and qRT-PCR results (Fig. 5.6.2; Fig. 5.2.1). *ALAR* was significantly upregulated in the posterior following a blood meal (Wilcoxon rank-sum test, $W = 33$, $p = 0.015$). In the carcass, expression was comparable to that of the anterior regions of the midgut (Table I.2).

Similarly, *GLYR* expression peaked in sugar-fed carcass samples compared to the midgut tissues and was significantly downregulated following blood feeding (Wilcoxon rank-sum test, $W = 4$, $p = 0.026$; Fig. 5.6.2). Within the midgut, *GLYR* expression was consistent with known AMPs and the RNA-seq data, where expression peaked in the cardia (Table 5.2.1). Additionally, blood feeding had no impact on expression levels within the midgut (Wilcoxon rank-sum tests, $p > 0.05$; Table 5.6.1).

KIN1 was also enriched in carcass samples, and blood feeding did not significantly impact expression levels in any tissue (Wilcoxon rank-sum tests, $p > 0.05$; Table 5.6.1). Under sugar feeding conditions, expression across the tissues did not differ significantly from each other (Kruskal–Wallis test, $\chi^2 = 4.93$, df = 3, $p = 0.18$). This is similar to what was observed in the RNA-seq data, where expression was relatively stable within the midgut compared to other putative and known AMPs.

Table 5.6.1. Wilcoxon rank-sum test statistics of locus AMP expression comparisons between blood- and sugar-fed tissues.

Comparison		Gene	Wilcoxon Rank-Sum Test Statistics	
			W	*p*-value
Cardia (BF)	Cardia (SF)	*ALAR*	27	0.18
Cardia (BF)	Cardia (SF)	*GAM1*	1	4.33×10^{-3} **
Cardia (BF)	Cardia (SF)	*GLYR*	18	1.00
Cardia (BF)	Cardia (SF)	*KIN1*	20	0.82
Anterior (BF)	Anterior (SF)	*ALAR*	12	0.39
Anterior (BF)	Anterior (SF)	*GAM1*	18	1.00
Anterior (BF)	Anterior (SF)	*GLYR*	22	0.59
Anterior (BF)	Anterior (SF)	*KIN1*	18	1.00
Posterior (BF)	Posterior (SF)	*ALAR*	33	0.02 *
Posterior (BF)	Posterior (SF)	*GAM1*	16	0.82
Posterior (BF)	Posterior (SF)	*GLYR*	27	0.18
Posterior (BF)	Posterior (SF)	*KIN1*	22	0.59
Carcass (BF)	Carcass (SF)	*ALAR*	25	0.31
Carcass (BF)	Carcass (SF)	*GAM1*	5	0.04 *
Carcass (BF)	Carcass (SF)	*GLYR*	4	0.03 *
Carcass (BF)	Carcass (SF)	*KIN1*	13	0.48

BF = blood-fed, SF = sugar-fed; * $p < 0.05$, ** $p < 0.01$.

5.7 Synthetic KIN1 shows anti-*Plasmodium* activity *in vivo*

Given that the depletion of putative AMPs from the locus had antimicrobial effects, the next step was to investigate the effect of overexpression. As a proxy for overexpression, the proteins of these putative AMPs were requested to be synthesized *in vitro* by an external company (GenScript). Unfortunately, only KIN1 was able to be synthesized successfully. KIN1 was initially planned to be used as a negative control, as its KD did not significantly affect *P. falciparum* infection intensity and prevalence.

Synthetic KIN1 was resuspended in nuclease-free water and supplemented into infectious *P. falciparum* blood meals for mosquito infections to a final concentration of 50 μM. Control mosquitoes were fed infectious blood meals containing an equal volume of nuclease-free water, although the use of another synthetic peptide would have acted as a more appropriate control. After eight days, mosquito midguts were dissected and oocysts were counted. Three biological replicates were completed. Feeding of synthetic KIN1 consistently had a significant impact on the parasite, where prevalence was reduced when fed to mosquitoes (Fig. 5.7.1; Table 5.7.1). Infection intensities were significantly reduced in all but one of the replicates following treatment with KIN1 (Table 5.7.1); indeed, a GLM revealed that replicate had a significant impact on the oocyst infection intensities ($F_{1,121}$, $p = 1.11 \times 10^{-10}$; Table H.1), warranting further study.

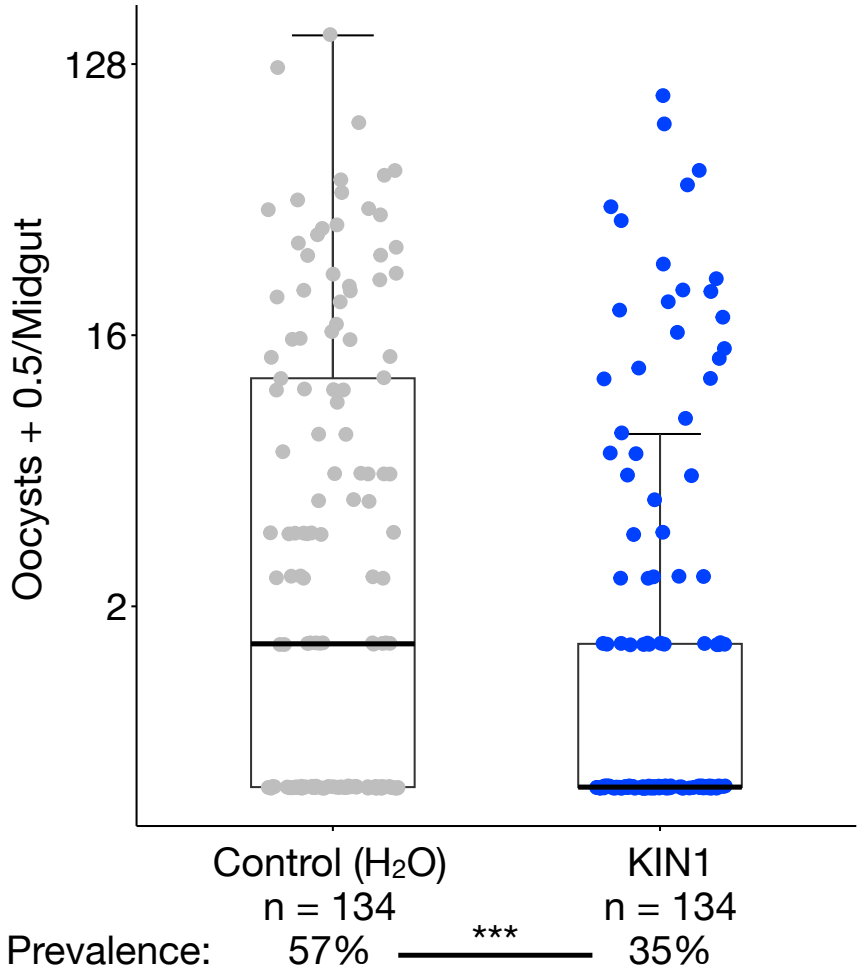


Figure 5.7.1. The effect of exogenous, synthetic KIN1 on* P. falciparum *oocyst development. *Synthetic KIN1 peptides were supplemented into infectious blood to a final concentration of 50 μM. Infectious blood meals with an equal volume of water were fed to other mosquitoes as a control. Oocysts were counted at eight days post-infection. Each dot represents the total number of oocysts in one midgut. Prior to plotting against a $\log_2$-transformed y-axis, 0.5 was added to each oocyst count value. Black lines indicate the median number of oocysts. Three biological replicates were completed in total. No significant difference was observed in oocyst infection intensity between the conditions (Wilcoxon rank-sum test performed on midguts with non-zero counts, p > 0.05). GLMs were modelled on the effect of treatment condition and replicate on infection intensity, model statistics are listed in Table H.1. Infection prevalence: χ^2 test, *** p < 0.001.*

Table 5.7.1. Summary of* P. falciparum *infection statistics following the addition of KIN1 or water (control) to an infectious blood meal. *Wilcoxon rank-sum tests were used to assess differences in oocyst infection intensity (midguts with 0 oocysts were excluded), χ^2 tests were used to assess differences in infection prevalence.*

	Replicate 1		Replicate 2		Replicate 3		Combined	
	KIN1	Control	KIN1	Control	KIN1	Control	KIN1	Control
# of Midguts	49	46	49	55	36	33	134	134
Oocysts								
Maximum	100	159	1	15	6	17	100	159
Mean	13.2	27.4	0.06	0.94	0.42	3.56	4.9	10.7
Median	4	22	0	0	0	2	0	1
Prev. (%)	71	87	6	29	25	67	35	57
Wilcoxon Rank-Sum Tests								
W	954.5		36		167.5		2107.5	
p-value	3.27×10^{-3} **		0.15		2.49×10^{-3} **		0.12	
χ^2 Tests								
χ^2	2.46		9.16		12.08		13.51	
df	1		1		1		1	
p-value	0.12		2.48×10^{-4} **		5.09×10^{-4} ***		2.38×10^{-4} ***	

*Prev. = infection prevalence, df = degrees of freedom; ** p < 0.01, *** p < 0.001.*

5.8 MF significantly impacts asexual *Plasmodium* development *in vitro*

Previously, it was found that FAMeTs, enzymes which convert FA to MF, the precursor of JH III, were enriched in the cardia (section 4.8). Like JH III, MF was found to have immunosuppressive effects in the posterior, although it is unclear whether if this is due to MF having antibacterial effects and directly killing the microbiota. Other precursors of JH III are synthesized by the mevalonate pathway from acetyl-CoA, resulting in the metabolite farnesyl diphosphate. Interestingly, farnesyl diphosphate is also a precursor shared by the squalene/cholesterol biosynthetic pathway in insects (Bellés, Martín & Piulachs, 2005), as well as the artemisinin pathway in *Artemisia annua* (Schramek et al, 2010; Chen et al, 2017; Fig. 5.8.1), which has been widely used as an antimalarial drug.

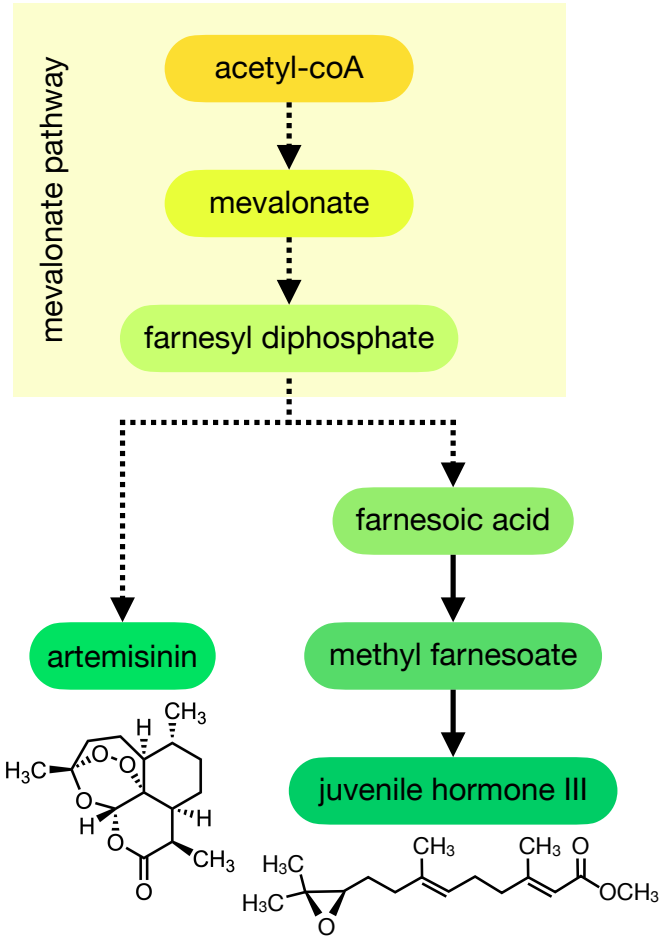


Figure 5.8.1. The artemisinin and JH III biosynthetic pathways. *Both pathways share a common precursor metabolite: farnesyl diphosphate. Dotted lines represent other metabolites and biochemical reactions take place between the production of two metabolites, solid lines represent direct biochemical reactions.*

Given that MF has immunosuppressive effects in the posterior midgut and shares precursor metabolites with artemisinin, exploratory experiments were conducted to assess whether MF has antimalarial activity. Here, various life stages of the human malaria parasite were co-cultured or treated with MF.

First, asexual stages of *P. falciparum* were co-cultured with various concentrations of MF, artemisinin, or an equal volume of DMSO over the course of 48 hours *in vitro*. The parasitemia of each culture was assessed every 24 hours for a total of three time points (0, 24, and 48 hours). Four replicates were conducted in total. On average, all 4 concentrations of artemisinin tested significantly reduced growth rates of *P. falciparum* (Wilcoxon signed-rank tests, $p < 0.05$; Table J.1), whereas only the highest concentration of MF (100 μM) significantly impacted asexual parasite growth (Wilcoxon signed-rank test, $V = 4$, $p = 0.01$; Fig. 5.8.2; Table J.1). The three highest concentrations of artemisinin killed the majority if not all parasites after 24 hours, whereas the 100 μM MF treatment appeared to maintain the initial parasitemia levels over the course of the experiment (Fig. 5.8.2).

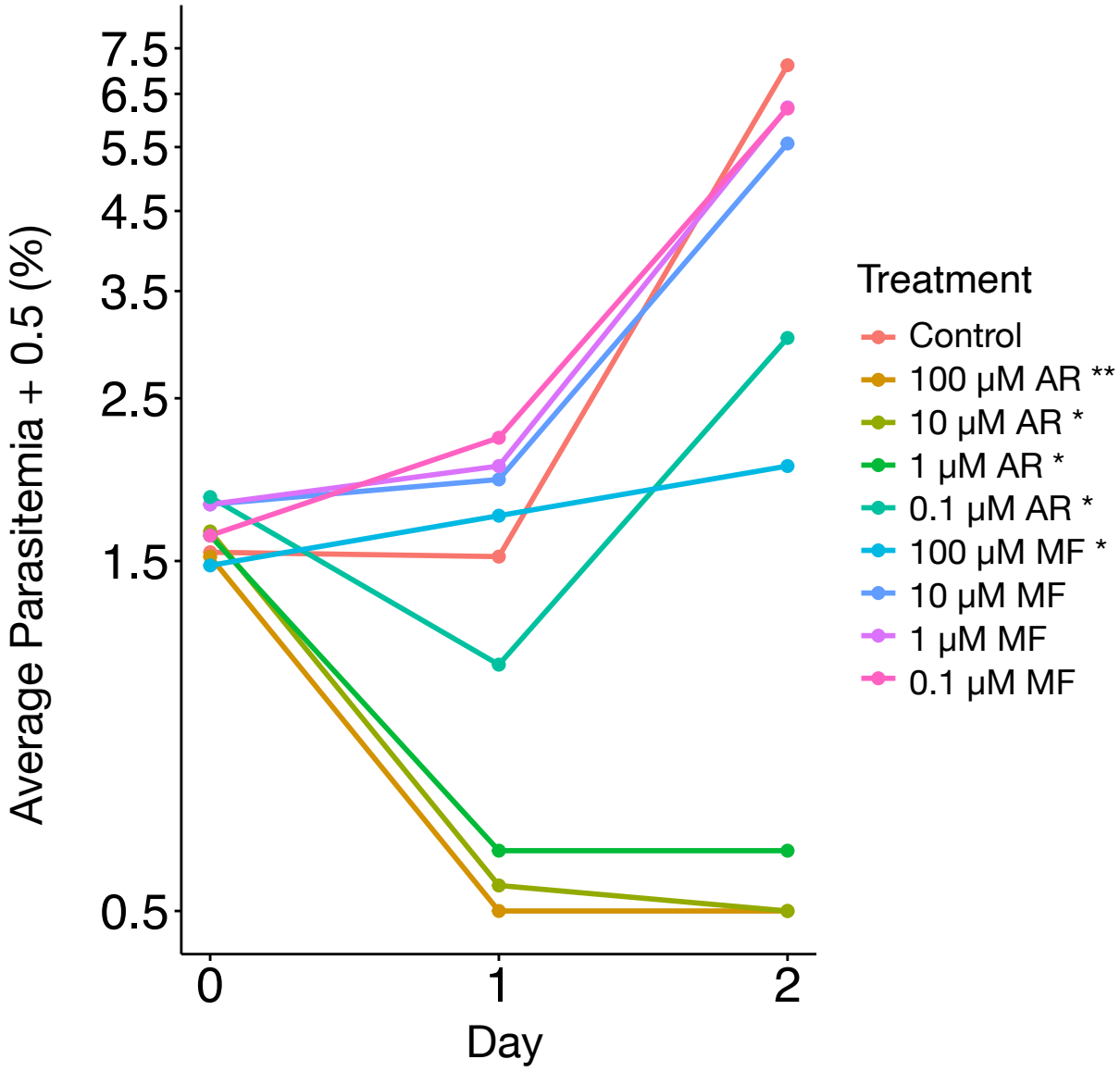


Figure 5.8.2. Average parasitemia of* in vitro P. falciparum *asexual cultures following co-incubation with various concentrations of artemisinin (AR), MF, or DMSO (control). *Four replicates were completed in total. Parasitemia was assessed daily from an aliquot of each culture, and 0.5 was added to each average parasitemia value in order to utilize a log_2-transformed y-axis scale for presentation purposes. Wilcoxon signed-rank tests, * $p < 0.05$, ** $p < 0.01$.*

Next, the effect of MF on male gametocyte exflagellation was tested. Two aliquots were collected from each *in vitro* culture containing mature gametocytes, which were pelleted and mixed with ookinete media containing 100 μM MF or an equal volume of DMSO prior to the exflagellation assay. After exflagellation initiated, exflagellation centers in Neubauer chambers were counted under a light microscope using the 10x objective. No significant difference was observed between MF- and DMSO-treated gametocytes (Fig. 5.8.3A; Wilcoxon signed-rank test, $V = 104.5$, $p = 0.31$).

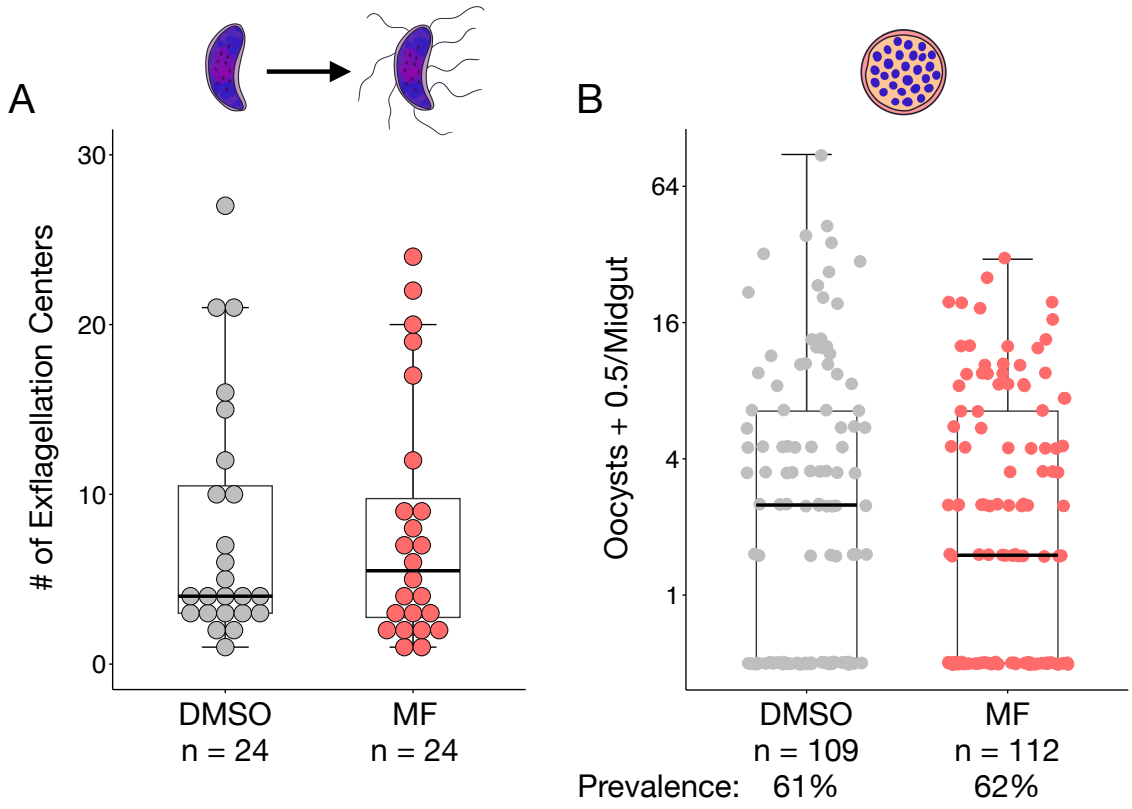


***Figure 5.8.3. The effect of MF on mosquito midgut stages of* P. falciparum.** *A) The number of exflagellation centers observed* in vitro *following MF or DMSO (control) treatment. B) The number of oocysts per mosquito midgut observed at eight days post-infection following blending of MF or DMSO into the infectious blood meal. Each dot represents the total number of oocysts in one midgut. Prior to plotting against a $\log_2$-transformed y-axis, 0.5 was added to each oocyst count value. Black lines indicate the median number of oocysts. Three biological replicates were completed in total. No significant differences were observed in the number of exflagellation centers and oocysts between the DMSO and MF-treated parasites and mosquitoes respectively (Wilcoxon rank-sum tests, $p > 0.05$; for oocyst infection intensity, only midguts with non-zero counts were considered). GLMs were modelled on the effect of treatment condition and replicate on infection intensity, model statistics are listed in Table H.1.*

In order to assess the impact on mosquito midgut stages of the parasite, MF was added to infectious blood meals to a final concentration of 100 μM immediately before mosquito infections. As a control, an equal volume of DMSO was used. Midguts were dissected at eight days post-infection for oocyst staining and counting. A total of three biological replicates were performed, and MF had no effect on *P. falciparum* infection prevalence (χ^2 test, $\chi^2 = 4.53 \times 10^{-4}$, df = 1, $p = 0.98$; Fig. 5.8.3B). Given that oocyst infection intensities can vary greatly between replicates, a GLM model was constructed to assess the impact of treatment (MF/DMSO) and replicate on infection intensity. While condition was a significant explanatory variable of infection intensity, this was also confounded by a significant effect of replicate number (*F*-tests, $p < 0.05$; Table H.1). Additional replicates are required to elucidate whether MF has a true impact on midgut parasite development. Thus, MF appears to only have a clear impact on asexual stages of *Plasmodium*.

5.9 Discussion

In the midgut, the parasite faces a major bottleneck between the transition from gametocytes to oocysts. This is attributed to midgut immune responses, such as AMPs which are enriched in the more anterior regions of the midgut (Cirimotich et al, 2010). As malaria continues to burden hundreds of millions of people annually (WHO, 2023), further characterization of mosquito immunity could lead to the generation of malaria-refractory mosquitoes which could form the foundation for a novel intervention strategy. The main aim of this research was to identify and characterize novel AMPs.

A number of putative AMPs were identified using six different AMP prediction algorithms. These algorithms were trained on different datasets and were generated by different mathematical models. Multiple algorithms for AMP prediction were utilized to select the best candidates for functional characterization. Additionally, a *de novo* transcriptome assembly was generated in efforts to identify novel AMPs. The dozens of AMPs in other insects dwarf the 11 encoded by *Anopheles*; the possibility that more exist but have yet to be identified was thought to explain this difference. While tens of thousands of transcripts were identified, the majority were associated with known genes. Of the remaining transcript predictions, few were short in length and none were predicted to have signal peptide sequences. Nevertheless, the other candidate selection strategies identified a total of 13 putative AMP candidates for further characterization.

Interestingly, three of the identified putative AMPs belong to the same locus and include *KIN1*. These genes have a number of characteristics which suggest they are antimicrobial. First, they share a similar spatiotemporal expression pattern in the midgut as known AMPs. Second, they are abundant in glycine. A class of AMPs has been defined by a richness in glycine (Wang, Li & Wang, 2009). Glycine and alanine, which are abundant in ALAR, are the two smallest amino acids, and their abundance in a sequence aids the folding of secondary structures (Wang, Li & Wang, 2009). Despite this, AlphaFold structural predictions of GLYR and KIN1 did not yield any secondary structures, but the confidence level of each model was relatively low and is perhaps inaccurate. ALAR on the other hand was predicted to have a series of α-helices. Additionally, approximately one-fifth of KIN1 is made up of positively-charged histidine residues. While the overall peptide charge is slightly negative, clusters of histidine residues exist within

the sequence which could form regions of the peptide which have a high affinity for microbial membranes. Lastly, their proximity to each other in the genome perhaps reflect that they are novel types of AMPs, similar to the cecropins.

Throughout mosquito development, *ALAR* and *KIN1* share a similar expression profile to *GAM1*, where expression increases into adulthood. *GLYR* peaked in the final larval instar. These results are in agreement with a previous study which performed RNA-sequencing of whole larvae of each instar (Papa et al, 2017). Expression of all four genes was also higher in males than in sugar- and blood-fed females. The lower expression of AMPs in females could be attributed to a trade-off between immunity and reproduction (Sheldon & Verhulst, 1996; Schmid-Hempel, 2003; Schwenke, Lazzaro & Wolfner, 2016). Following blood feeding, female mosquitoes undergo major transcriptional changes, including the expression of digestive enzymes and investment into ovary maturation (Ribeiro, 2003; Rogers et al, 2008). These biological processes are exclusive to females, as males feed on sugar throughout their entire lifetime and do not produce eggs. Thus, males can invest more resources into immunity. Between the carcass and midgut, expression data of these putative AMPs indicated that they were enriched in the carcass, which contrasts with the expression profiles of known AMPs.

RNAi-mediated KDs of the putative AMPs and subsequent immune challenge showed an indication that many of the putative AMPs have antimicrobial activity. In bacterial infection assays, *E. faecalis* and *S. marcescens* were introduced through the sugar meal to induce a local infection in the midgut as they have been previously characterized to be able to colonize insect midguts (Cox & Gilmore, 2007; Stathopoulos et al, 2014). *E. coli* and *S. aureus* were thoracically-injected into mosquitoes to induce a systemic infection. KD of the known AMP *GAM1* had mixed effects on bacterial infections, and interestingly, the silencing of *GAM1* improved long-term survival after *S. aureus* infection. In this case, as *GAM1* expression is abundant in the midgut, it is likely that the depletion of *GAM1* enabled the proliferation of midgut microbiota. This in turn allowed for immune priming in the midgut and hemolymph/fat body, which increased survival rates following infection. On the contrary, a previous study found that *GAM1* depletion reduced survival rates following *S. aureus* infection in the short-term, however the bacterial density used for infections was approximately five times greater than what was used in this study

(Dong et al, 2006; Chang et al, 2013). As GAM1 has been characterized to exert antimicrobial activity against Gram-positive bacteria (Vizioli et al, 2001b), its depletion could severely compromise antibacterial defenses when mosquitoes are subjected to a more potent immune challenge.

Similarly, silencing of cardia gene AGAP009313 consistently and significantly improved survival rates following infection with three of the four species of bacteria tested. This observation may suggest antibacterial activity, as it could be explained by a similar mechanism as *GAM1* where its depletion led to immune priming and better survival outcomes following a systemic infection. In the context of *S. marcescens* infection, better survival outcomes could also be explained by AGAP009313 KD not only promoting immune priming, but also tolerance in the midgut. In insects, the dampening of immune responses or enduring fewer infections have been shown to increase life span through the reduction of immunopathological damage and increased tolerance (Pursall & Rolff, 2011; Fabian et al, 2018). Silencing of the cardia genes AGAP009313 and AGAP013543 also led to an increased prevalence of *P. falciparum* infection, suggesting a role of both in antimalarial responses. AGAP009313 was previously found to be a major constituent of the PM, suggesting it is either is a functional component of the matrix, or as it is highly expressed by the cardia, these proteins are being 'caught' during PM formation (Dinglasan et al, 2009). In any case, depletion of AGAP009313 could impact PM integrity and thus allow for immune priming by the proliferating microbiota, or allow for parasites to be detected by components of the immune system instead of being sequestered by the PM. This mechanism might be similar to that described following IMPer/DUOX depletion, which increased *Plasmodium* interactions with mosquito immune responses and led to their reduced survival (Kumar et al, 2010). As AGAP013543 lacked any significant effect on bacterial infections, it is convenient to speculate that it may have *Plasmodium*-specific activity. However, multiple previous studies have outlined the unlikelihood of the mosquito evolving *Plasmodium*-specific effectors (Morlais et al, 2004; Obbard et al, 2007; Rottschaefer et al, 2011). Nevertheless, their effect on *Plasmodium* could be a result of broad-spectrum antimicrobial activity, or indirect mechanisms. For example, these genes may be involved in another important biological process in the mosquito midgut, and their depletion may reduce the suitability of the midgut for parasite survival.

Given that oral infections were less lethal to mosquitoes, an alternative approach was taken to assess antibacterial activity in the remaining putative AMPs. Instead of oral infections, *S. marcescens* was thoracically-injected into mosquitoes like in experiments performed with *E. coli*, and mosquito survival was observed over a period of ten days. Compared to oral infections, systemic infections of *S. marcescens* led to more dramatic reductions in mosquito survival as described previously (Bahia et al, 2014; Dekmak et al, 2021). This parallels a previous study in *Drosophila*, where host adaptations to pathogens were described as route-specific; adaptation of a host to a pathogen from one infection route does not lead to increased resistance against the same pathogen from a different infection route (Martins et al, 2013). Additionally, this parallels previous findings presented here, where oral infection of *S. marcescens* did not elicit AMP expression in the midgut, but AMP expression was upregulated in the remaining carcass. As *S. marcescens* is a natural midgut endosymbiont of *Anopheles*, it is tolerated in the midgut but not in the hemolymph/fat body.

Of the remaining putative AMPs, only AGAP001508 significantly impacted survival rates over a period of ten days following bacterial infection. While AGAP001508 depletion significantly improved survival rates post-*S. marcescens* infection, expression of this putative AMP was downregulated following systemic *E. coli* infection, and replicate played a significant role in survival outcomes. In the context of its depletion followed by *E. coli* infection, survival rates were increased compared to controls after the 10 day period, although this increase was insignificant. This putative AMP was also enriched in the carcass compared to the midgut. Taken together, this gene could be involved in immune tolerance/suppression against bacteria in the hemolymph/fat body. Although KD of AGAP001508 had no effect on *P. falciparum* in this study, it was previously implicated in the melanization of *P. berghei* ookinetes in *An. gambiae* (Pinto et al, 2009).

While injection of *S. marcescens* drastically reduced the experimentation time required, one of the costs was the inability to observe the effect of putative AMP depletion on midgut infections. In order to minimize time and also assess the effect of depletion in tissues where these putative AMPs are expressed, subsequent experiments could use a higher concentration of bacteria in oral infections, or the use of a more pathogenic species with the ability to colonize the midgut could be investigated.

Antimicrobial activity was observed for all three genes in the putative AMP locus. *ALAR* plays a role in Gram-negative defense, where its depletion significantly reduced survival following *S. marcescens* infection in the long-term. Contrastingly, in the short-term, *ALAR* depletion improved survival outcomes. This could be explained by the fact that *S. marcescens* is able to effectively colonize the mosquito midgut (Stathopoulos et al, 2014), and depletion of *ALAR* could have promoted tolerance during the initial infection stages. The effect of KD typically lasts a few days, but during this time the *S. marcescens* may have established itself in the midgut and eventually mosquitoes succumbed to the infection. ALAR also influenced infection intensities of *P. falciparum*. GLYR is involved in anti-*Plasmodium* responses, and KIN1 appears to have antibacterial and antimalarial activity. Interestingly, previous experiments performed in the host laboratory found that in the context of *P. berghei* infections, *ALAR* but not *GLYR* depletion significantly increased oocyst infection intensities (unpublished data). Therefore, there is evidence for the involvement of all three of the genes in anti-*Plasmodium* responses.

KIN1 appears to exert its antibacterial activity in the hemolymph and fat body, as its expression peaks in the carcass, and its depletion consistently reduced survival following systemic *E. coli* infection. Here, its expression was not induced following *E. coli* infection despite being recently characterized as one of the most abundant genes post-*E. coli* challenge in *An. gambiae* (Hixson et al, 2023). This might be attributed to differences in the time points assayed as well as dosage, as this previous study examined *KIN1* expression at 12 hours post-challenge instead of 48 hours, and infected mosquitoes with half of the dosage employed here. In all other species of bacteria, KD of *KIN1* consistently reduced survival rates in the long-term. Therefore, KIN1 plays a clear role in protecting against bacterial infections.

Surprisingly, depletion of *KIN1* had no effect on oocyst development as previously described (Vlachou et al, 2005; Dong et al, 2006), whereas feeding of synthetic KIN1 peptides to replicate its overexpression significantly reduced the infection prevalence. This could be a result of a few possible explanations. Firstly, while *KIN1* expression is enriched in the carcass, it may be more potent against midgut stages of the parasite. Therefore, upon supplementing the synthetic peptide with the infectious blood meal, it is able to impact midgut stages of parasite development when transported into the midgut.

Additionally, the silencing of *KIN1* would then have little effect on the parasite. Secondly, the measured silencing efficiency may not accurately reflect the reduction in protein turnover. In other words, while the number of *KIN1* transcripts were reduced, it is unclear whether KIN1 proteins were also reduced. Subsequent studies could include the quantification of proteins in addition to transcripts when measuring KD efficiency. Lastly, if KIN1 is an AMP with antibacterial activity, altering KIN1 levels may impact the microbiota, which in turn impacts the parasite. This mechanism could involve synthetic KIN1 altering the microbiota community in favor of species with anti-*Plasmodium* activity. A similar mechanism has been previously described, where the overexpression of a single AMP in *Drosophila* significantly altered the population size of one species of bacteria in favor of another (Ryu et al, 2008). To test this hypothesis, midgut microbiota (*16S*) sequencing of mosquitoes treated with or without exogenous KIN1 could be completed to compare whether the microbiota community structure is affected. Given the significant reduction of infection prevalence following KIN1 feeding, future experiments could entail the generation of a transgenic mosquito line, with *KIN1* introduced into the locus of a midgut-specific and blood meal-responsive gene such as *CP*. Assessing parasite development in this transgenic line would enable a better understanding of whether KIN1 has antimalarial activity. It is also important to note that the use of water as a control may not have been the most appropriate. The initial aim was to test the effect of synthetic ALAR and GLYR on parasite development *in vivo,* and KIN1 was to be used as a negative control given the lack of impact on *P. falciparum* infection following its KD. Only KIN1 was able to be synthesized successfully, and therefore the planned experiments were carried out as exploratory research. The synthetic peptide was resuspended into water, and hence an equal volume of water was added to infectious blood meals as a control condition. Future experiments could utilize another synthetic peptide with no impact on *Plasmodium* development as a control. Despite this, a previous study utilized the addition of PBS as a control and found that supplementation of synthetic exogenous AMPs magainin 2 and melittin significantly reduced parasite development (Habtewold et al, 2019). In this study, other peptides were also tested but had no or little effect. These AMPs were then introduced into the *An. gambiae* genome to generate two

transgenic lines, both of which were able to significantly reduce *P. falciparum* development in the midgut (Hoermann et al, 2022).

AGAP001181 and AGAP003713 were the only other putative AMPs with antimalarial activity, where their silencing significantly reduced the prevalence and infection intensity of *P. falciparum* oocysts respectively. The AGAP003713 gene is abundant in the midgut compared to the rest of the body and is downregulated post-infection, potentially indicative of the parasite suppressing its expression. It is annotated to be have a transcription factor domain and hence may control the expression of a variety of effectors that may impact midgut stages of parasite development. However, other factors were also significant predictors of infection intensity for AGAP003713; further study is required to confirm its putative antimalarial role. The AGAP00181 gene was identified to be enriched in infected midguts compared to controls, and might be a novel *Plasmodium*-responsive peptide that also possesses antimalarial activity. Once again, it is unlikely that this putative effector evolved to specifically target the parasite. As expression of AGAP001181 was found to be relatively low in both the midgut and carcass, future experiments could blend synthetic AGAP001181 peptides into infectious blood meals to see if *Plasmodium* development is further impacted. Alternatively, increasing the expression of this peptide in a transgenic mosquito could also be explored. When assessing the malaria refractoriness of future transgenic lines, it would also be important to measure the number and prevalence of salivary gland sporozoites to assess whether transmission has been reduced. Counting the number of oocysts only provides insight on how parasite development is impacted and not their transmission; in other words, if no significant reduction of oocysts is observed, this may not reflect a lack of reduction in sporozoite numbers, as the oocysts may be damaged and unable to produce sporozoites.

Many of the *P. falciparum* infection and bacterial survival assays were severely impacted by replicate, where replicate was a significant contributor to infection and survival outcomes respectively. This has the potential to obscure effects of putative AMP KD on *P. falciparum* infection or mosquito survival following bacterial infection. In the bacterial survival assays, this effect was more prevalent in bacterial infections performed via injection, which parallels the fact that the vast majority of the infection intensities observed in *P. falciparum* infection assays, where all gene KD conditions were subjected

to intrathoracic injection, were affected by replicate. The differences between replicates could be attributed to variation in number of factors, including differences in: 1) fitness, 2) tolerance/resistance responses to malaria/bacterial infection, 3) response to the injection process (*i.e.* wounding and healing), and/or 4) microbiota community structure both within and between mosquito colonies (replicates) and/or individuals. Performing additional replicates could perhaps alleviate these effects. Alternatively, other biological factors can be more tightly controlled between replicates, such as microbiota composition, which are known to impact infection outcomes (Beier et al, 1994; Dong, Manfredini & Dimopoulos, 2009; Weiss & Aksoy, 2011; Engel & Moran, 2013; Clay, 2014; Kulkarni et al, 2021). Despite a significant impact of replicate on survival, if gene KD is also a significant predictor of survival/infection intensity, this indicates that the overall trend in survivability/malaria susceptibility following putative AMP KD and infection is still consistent across replicates, and hence provides an indication that these putative AMPs contribute to antimicrobial responses.

The *P. falciparum* infection assays were severely impacted by the effect of the injections, as it was difficult to achieve successful *P. falciparum* infections under laboratory conditions and more so post-injection. The median number of oocysts in all replicates was between 0 and 1.5, indicating that a large proportion of mosquitoes were not infected. Approximately half of the replicates failed to produce a successful *P. falciparum* infection, with one gene requiring eight replicates in order to achieve three successful replicates of infection. Compared to positive control mosquitoes (*i.e.* not injected), putative AMP-silenced mosquitoes typically had reduced oocyst infection intensities. The same observation was found in a study by Nsango et al, where *dsLacZ*-injected mosquitoes had reduced infection intensities compared to non-injected controls (2012). Nsango and colleagues proposed that the wound healing response, which is initiated upon injection, provides protection against parasites as well (2012). An alternative explanation is immune priming, as the injection process was not completed under sterile conditions. Microbes in the environment and/or on the mosquito cuticle can be introduced into the thorax during the injection process. A previous study also observed a similar effect, where injured mosquitoes (*i.e.* pierced with a needle) had better survival outcomes upon facing an *E. coli* infection compared to bacteria-primed mosquitoes and

naïve controls (Powers et al, 2020). Ideally, the characterization of these putative AMP genes should be performed via the generation of transgenic knockout or conditional knockout lines, but this would be require much more resource investment, and the latter tool has yet to be developed in mosquitoes.

The lack of antimicrobial activity in some of the putative AMPs does not preclude the possibility of being an AMP or an antimicrobial effector. Functional redundancies among AMPs exist, and so the effect of knocking down one AMP could be masked by another AMP which has similar activity. While AMPs are typically described as having broad-spectrum activity against microbes, some possess activity against a narrow range or even individual species of microbes. In *Drosophila*, specific resistance against *Providencia rettgeri* bacteria is conferred by a single, non-synonymous mutation of one amino acid in diptericin A (Unckless, Rottschaefer & Lazzaro, 2015; Unckless, Howick & Lazzaro, 2016). This mutation has independently evolved multiple times in different *Drosophila* species (Unckless & Lazzaro, 2016). Additionally, this might also explain the lack of expression of some known AMPs in the RNA-seq datasets *e.g*. DEF3 and DEF5. These (putative) AMPs may have activity against specific microbes, or may be expressed within specific tissues/life stages such as the salivary glands which were not examined. Furthermore, whether the putative AMPs have antifungal or antiviral effects was not tested and could be the focus of subsequent studies. As approximately half of the putative AMPs were only tested against Gram-negative species of bacteria, it would also be important to test their impact during Gram-positive infections.

Finally, preliminary experiments have shown that MF negatively impacts the development of asexual stages of *Plasmodium*. Other life stages were not impacted. While the mechanism of this antimalarial action was not explored, it is unlikely that the impact of MF is related to the developmental arrest of the parasites or the commitment to gametocytogenesis, as cultures continued to alternate between ring and trophozoite asexual stages over the course of the experiments. Compared to artemisinin, the antimalarial activity of MF was observed to be much weaker and only potent at the highest concentration tested. Artemisinin is known to kill human blood stages of *P. falciparum in vivo* (ter Kuile et al, 1993; Chen et al, 1994). Although structurally dissimilar to artemisinin, MF also appears to exhibit similar antimalarial activity against human blood stages of *P.*

falciparum in vitro. While parasites were not completely killed by MF unlike artemisinin, it is likely that the induction of gametocytes cannot occur as parasitemia levels appear to remain unchanged over the course of two days. Longer term experiments with MF could confirm this hypothesis. It could also be possible that MF possesses antimalarial activity, but the metabolite itself degraded within the timeframe of the experiment as parasite cultures were maintained at human body temperatures. In any case, this study identified MF as a metabolite which targets *Plasmodium* in an asexual stage-specific manner.

Chapter 6: Conclusions and Perspectives

Malaria continues to cause harm and death to a quarter of a billion people each year, despite being a preventable and curable disease (WHO, 2023). A myriad of strategies targeting different aspects of malaria transmission exist, but the lack of reduction in malaria cases in recent years highlights a need for novel strategies to be developed and implemented (WHO, 2023). One promising approach is the use of transgenic mosquitoes, particularly the manipulation of mosquito immunity to impede or block parasite development and subsequent transmission to humans. Such engineered genetic features can be spread in wild mosquito populations using gene drives. Along these lines, multiple transgenic lines have been generated over the last decade, with some lines exhibiting near-complete malaria refractoriness. Most of these transgenic mosquitoes utilize the same subset of effectors and promoter sequences. Further study into the spatiotemporal expression of mosquito immune system components would greatly benefit the understanding of responses to infection, in addition to identifying novel elements of the mosquito genome that can be used in such approaches. Thus, the focus of this thesis revolved around the characterization of mosquito immune responses in the midgut, which is highly compartmentalized, and to identify novel, putative AMPs.

The mosquito midgut is a particularly relevant tissue for studying immunity, as the parasite undergoes many developmental stages and are subjected to a major bottleneck as a result of midgut immune responses (Cirimotich et al, 2010, Sinden, 2010). The midgut is also highly compartmentalized in terms of function. The cardia is incredibly immunocompetent, as it is constitutively expressing an arsenal of AMPs to bombard pathogens and sterilize incoming food sources (Warr et al, 2007; Hixson et al, 2022). The posterior midgut is immune quiescent, enabling the colonization of beneficial microbiota which also influence host vectorial capacity. Rather than investment into immunity, resources in the posterior midgut are heavily put towards the expression of genes relating to digestive processes (Warr et al, 2007; Hixson et al, 2022). Blood meal digestion, itself a crucial biological process of female mosquitoes, underpins other important biological aspects such as reproduction. Indeed, following blood feeding, up to half of the metabolism is invested towards blood meal digestion and egg production at the detriment to other processes such as flight (Marinotti et al, 2006). The process of blood meal digestion is also associated with an increase in oxidative stress (Graça-Souza et al, 2006),

as well as the potential ingestion of pathogens such as malaria parasites. Therefore, further characterization of immune responses after a blood meal could inform and enable the development of malaria-refractory transgenic mosquitoes.

In this study, spatiotemporal responses to the blood meal were characterized in the midgut tissues for the first time and compared to those after sugar feeding. Very few genes, including 5 AMPs, represented the majority of the secreted transcriptome of the cardia across all time points, indicating a highly specialized role of this tissue. This was confirmed microscopically through the development and implementation of the first whole-mount mosquito midgut RNA FISH method. It has been demonstrated that the immunocompetence of the cardia persists throughout blood feeding and is not dependent on malaria infection status, indicating that other mechanisms beyond pathogen recognition and the traditional immune pathways are involved in enabling this constitutive expression of AMPs. Importantly, it was found that DVE, a developmental homeobox transcription factor, plays a major role in this pattern of AMP expression. Dve is known in *Drosophila* to be involved in anterior-posterior midgut axis development and in cardia morphogenesis, and its expression in the adult midgut mosquito is shown to exist as a gradient across the midgut, which is paralleled by the expression of AMPs. This represents yet another example of a hard-wired developmental pathway in insects co-opted by the immune system, in this case to prevent rather than respond to infection.

The constant expression of immune effectors plays the role of sterilization, whereby ingested food sources are rid of pathogens prior to transportation into posterior regions of the midgut for digestion. One main hypothesis is that as the cardia acts as an immune filter, with high titers of AMPs accumulated in the cardia lumen. The crypt-like structures within the cardia are thought to function as storage compartments of a cocktail of highly concentrated AMPs that are blended into the food as this passes through the cardia. Sugar feeding does not immediately open the cardia valve as it happens with the blood, and the sugar is directed to the crop. As sugar drops are gradually released to enter the midgut, passing through the cardia, they are thought to mix with the AMPs stored in the cardia lumen. The antimicrobial cocktail is likely to continue its activity as it travels through the anterior midgut and after it reaches the posterior midgut (Dong et al, 2006). Rather than be a general sterilizer, the antimicrobial cocktail of the cardia is thought to

act as a selectivity filter, allowing only specific sets of microbes to enter the midgut. This is supported by the small number of different AMPs found to be highly expressed in the cardia, opening a new avenue for future investigations into specificity of mosquito AMPs.

In the posterior midgut, AMP expression is largely under the control of typical immune signaling pathways, especially the IMD/REL2, triggered upon pathogen recognition. Despite the expression of IMD pathway receptors such as PGRP-LC and PGRP-LA, as well as all known pathway components, the posterior midgut is highly immunotolerant due to increased expression of negative pathway regulators, such as PGRP-LB (Gendrin et al, 2017). Instead, investment of resources is poured into the expression of digestive enzymes, both prior and after a blood meal. Presumably, the upregulation of these genes in the posterior is the result of another concentration gradient of a transcription factor in the opposite direction to DVE. Many previous transgenic mosquitoes were generated with the expression of antimalarial effectors driven by the *CP* promoter (Corby-Harris et al, 2010; Dong et al, 2011; Dong et al, 2012; Dong, Simões & Dimopoulos, 2020; Hoermann et al, 2022). The data presented in this thesis provide a multitude of alternative promoters and host gene loci for high constitutive or blood meal-inducible expression of effectors. For example, it would be interesting to examine whether high expression of effectors specifically in the posterior midgut at 5 hours post-blood meal rather than consistently throughout all time points, could reduce fitness costs often associated with transgenes. The resources required from the mosquito to express transgenes would be reduced, as it would only be expressed at a specific time point rather than constitutively.

The compartmentalization of functions in the midgut is by no means absolute; following blood meal uptake, more anterior regions of the midgut also contribute to aid blood meal digestion by increasing investment into digestive enzyme expression. This reflects a transcriptomic reprograming of the midgut post-blood meal, when digestion becomes the most important biological function that needs to be carried out. As a result, many cellular maintenance activities are downregulated in favor of blood meal digestion. It is demonstrated in this thesis that effector molecules that promote tolerance in the posterior midgut are also produced in the cardia. These results provide the foundation of the cardia being an immune organ, indeed one that resembles an immune gland

considering its crypt-rich architecture. Its contribution to immunity and strategic position at the forefront of the midgut enables it to be the gatekeeper of the midgut. This would mark the first time such a structure is identified in insects.

A second important contribution of the work conducted for this Ph.D. thesis is the identification of several new, putative AMPs in the *An. gambiae* genome. This includes a new genomic locus consisting of three putative AMPs, *ALAR*, *GLYR*, and *KIN1*, which merit further characterization to validate their possession of antimicrobial activity. Two additional genes which were abundant in the cardia (AGAP009313 and AGAP013543) were also found to have AMP characteristics and possess significant antimalarial activity. A previous RNA-seq experiment of the entire mosquito found that expression levels of both uncharacterized cardia genes were in the 99th percentile in terms of abundance of expression (Vannini et al, 2014; Amos et al, 2022), and appear to be restricted to the adult mosquito (Papa et al 2017). Their unique expression profile and very high abundance in the entire mosquito suggests they have an important role in *Anopheles*. AGAP013543 is specific for the *Anopheles* genus. The generation of knockout mosquitoes will be crucial in better understanding the roles of these uncharacterized genes and validate their antimicrobial activity. Additionally, future work could aim to manipulate the spatiotemporal expression of these putative AMPs, or utilize their loci for the expression of known antimalarial effectors to assess whether this could further increase the bottleneck experienced by parasites in the mosquito and fully block parasite transmission. AGAP001181 was found to be a novel, parasite-responsive putative AMP that possesses antimalarial activity. Further characterization of AGAP001181 is warranted and like the other putative AMPs, could be used in engineering malaria-resistant mosquitoes. It would also be important to determine whether these putative AMPs are expressed in other malaria-relevant tissues of the mosquito such as the salivary glands, as they may have an impact on later stages of the malaria parasite which have not been considered here.

In conclusion, the results presented in this thesis provide novel insights into the mosquito immunity landscape and suggest a novel model of the regulation of the midgut immune transcriptome. Immunity and digestion, two biological functions associated with the cardia and posterior midgut respectively, are innate mechanisms of these tissues

which occur even in the absence of a stimulus (*e.g.* pathogenic insults and blood feeding). It is proposed that concentration gradients of developmental transcription factors drive the compartmentalization of midgut functions in a similar manner to embryo patterning (Fig. 6.1). This represents another example of developmental transcription factors and pathways being co-opted by the immune system. Many more transcription factors involved in embryo development could be co-opted to have additional immune roles in adulthood. Gradients of different transcription factors therefore contribute to the expression of different repertoires of genes, such as DVE driving the expression of AMPs in the cardia, and CAD contributing to tolerance in the posterior midgut. Multiple transcription factors could occupy the same location, such as the transcription factor(s) behind the AMP-like expression of AGAP009313 and AGAP013543 in the cardia. Subsequent studies could investigate other transcription factors in the midgut and the genes under their control.

As a follow-up from the work reported in this thesis, many interesting avenues of research are worthwhile to be explored. The midgut sections RNA-seq data has uncovered the spatiotemporal expression of all genes in the midgut. This will be extremely beneficial in informing the generation of the new malaria-refractory transgenic mosquitoes. Endogenous effectors and/or loci with desired expression profiles could be exploited to include the expression of powerful antimalarial effectors, enabling the targeting of the parasite in a controlled time- and tissue-specific manner. Ultimately, this knowledge, coupled with genetic engineering techniques, could form the foundation of novel malaria control strategies.

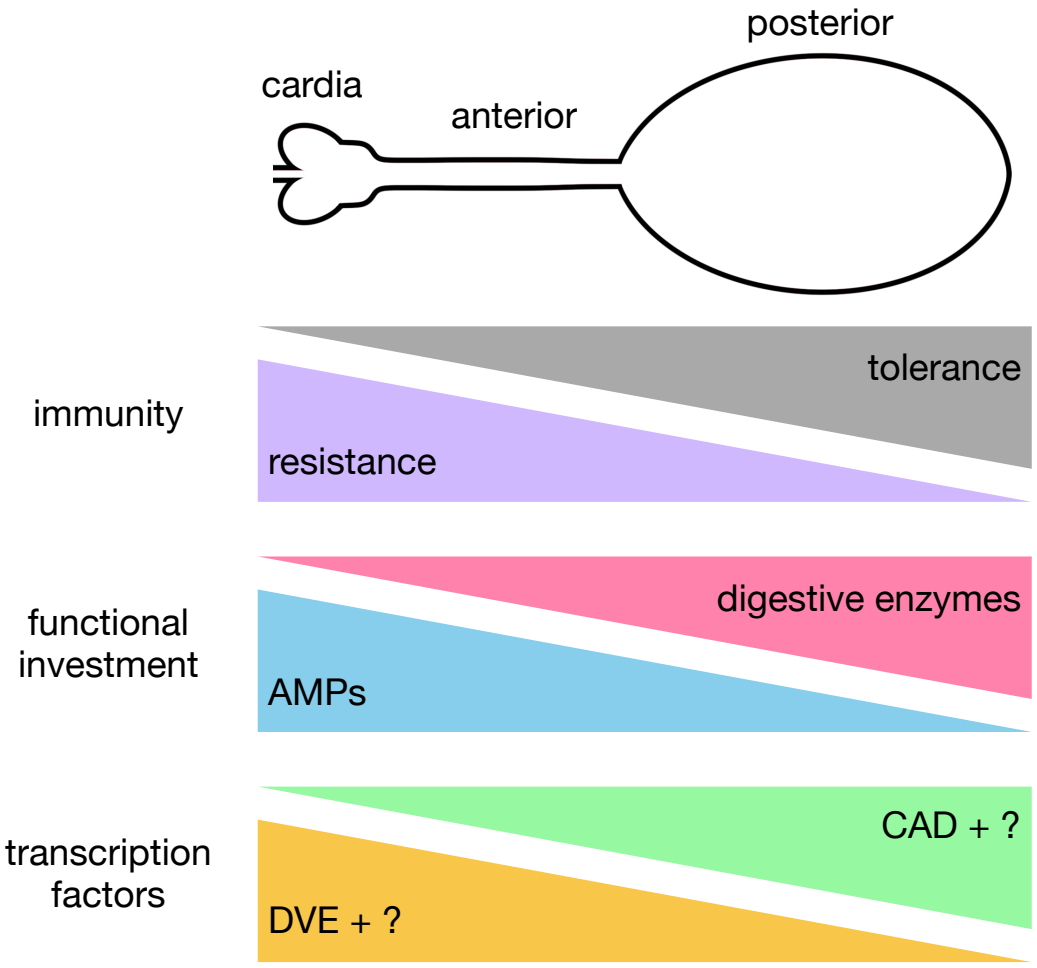


Figure 6.1. Graphical summary of the proposed model of midgut compartmentalization*. The midgut is compartmentalized in function, which is underpinned by concentration gradients of transcription factors driving the expression of different genes.*

Chapter 7: References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N. & Grüning, B. A. (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research.* 46 (W1), W537–W544.

Agrawal, N., Dasaradhi, P. V. N., Mohmmed, A., Malhotra, P., Bhatnagar, R. K. & Mukherjee, S. K. (2003) RNA interference: biology, mechanism, and applications. *Microbiology and Molecular Biology Reviews.* 67 (4), 657–470.

Ahmed, A. M. & Hurd, H. (2006) Immune stimulation and malaria infection impose reproductive costs in *Anopheles gambiae* via follicular apoptosis. *Microbes and Infection.* 8 (2), 308–315.

Akman-Anderson, L., Olivier, M. & Luckhart, S. (2007) Induction of nitric oxide synthase and activation of signaling proteins in *Anopheles mosquitoes* by the malaria pigment, hemozoin. *Infection and Immunity.* 75 (8), 4012–4019.

Akorli, J., Gendrin, M., Pels, N. A. P., Yeboah-Manu, D., Christophides, G. K. & Wilson, M. D. (2016) Seasonality and locality affect the diversity of *Anopheles gambiae* and *Anopheles coluzzii* midgut microbiota from Ghana. *PLoS ONE.* 11 (6), e0157529.

Ali-Murthy, Z. & Kornberg, T. B. (2016) *Bicoid* gradient formation and function in the *Drosophila* pre-syncytial blastoderm. e*Life.* 5, e13222.

Allada, R., White, N. E., So, W. V., Hall, J. C. & Rosbash, M. (1998) A mutant *Drosophila* homolog of mammalian *Clock* disrupts circadian rhythms and transcription of *period* and *timeless*. *Cell.* 93 (5), 791–804.

Almeida-Oliveira, F., Tuthill II, B. F., Gondim, K. C., Majerowicz, D. & Musselman, L. P. (2021) dHNF4 regulates lipid homeostasis and oogenesis in *Drosophila melanogaster*. *Insect Biochemistry and Molecular Biology.* 133, 103569.

Alphey, L. S., Crisanti, A., Randazzo, F. & Akbari, O. S. (2020) Standardizing the definition of gene drive. *Proceedings of the National Academy of Sciences.* 117 (49), 30864–30867.

Alves e Silva, T. L., Canepa, G. E., Sweeney, B., Hessab Alvarenga, P., Zhao, M., Vega-Rodríguez, J., Molina-Cruz, A. & Barillas-Mury, C. (2023) The heat shock protein Hsc70-3 mediates an anti-apoptotic response critical for *Plasmodium* evasion of *Anopheles gambiae* immunity. *Microbiology Spectrum.* 11 (6), 940.

Amenya, D. A., Chou, W., Li, J., Yan, G., Gershon, P. D., James, A. A. & Marinotti, O. (2010) Proteomics reveals novel components of the *Anopheles gambiae* eggshell. *Journal of Insect Physiology.* 56 (10), 1414–1419.

Amos, B., Aurrecoechea, C., Barba, M., Barreto, A., Basenko, E. Y., Bażant, W., Belnap, R., Blevins, A. S., Böhme, U., Brestelli, J., Brunk, B. P., Caddick, M., Callan, D., Campbell, L., Christensen, M. B., Christophides, G. K., Crouch, K., Davis, K., DeBarry, J., Doherty, R., Duan, Y., Dunn, M., Falke, D., Fisher, S., Flicek, P., Fox, B., Gajria, B., Giraldo-Calderón, G. I., Harb, O. S., Harper, E., Hertz-Fowler, C., Hickman, M. J., Howington, C., Hu, S., Humphrey, J., Iodice, J., Jones, A., Judkins, J., Kelly, S. A., Kissinger, J. C., Kwon, D. K., Lamoureux, K., Lawson, D., Li, W., Lies, K., Lodha, D., Long, J., MacCallum, R. M., Maslen, G., McDowell, M. A., Nabrzyski, J., Roos, D. S., Rund, S. S. C., Schulman, S. W., Shanmugasundram, A., Sitnik, V., Spruill, D., Starns, D., Stoeckert Jr, C. J., Tomko, S. S., Wang, H., Warrenfeltz, S., Wieck, R., Wilkinson, P. A., Xu, L. & Zheng, J. (2022) VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Research.*

Anderson, K. V. & Nüsslein-Volhard, C. (1984) Information for the dorsal-ventral pattern of the *Drosophila* embryo is stored as maternal mRNA. *Nature.* 311 (5983), 223–227.

Angleró-Rodríguez, Y. I., Blumberg, B. J., Dong, Y., Sandiford, S. L., Pike, A., Clayton, A. M. & Dimopoulos, G. (2016) A natural *Anopheles*-associated *Penicillium chrysogenum* enhances mosquito susceptibility to *Plasmodium* infection. *Scientific Reports.* 6 (1), 34084.

Angleró-Rodríguez, Y. I., MacLeod, H. J., Kang, S., Carlson, J. S., Jupatanakul, N. & Dimopoulos, G. (2017) *Aedes aegypti* molecular responses to Zika virus: modulation of infection by the Toll and Jak/Stat immune pathways and virus host factors. *Frontiers in Microbiology.* 8, 2050.

Ao, J., Ling, E. & Yu, X. (2007) *Drosophila* C-type lectins enhance cellular encapsulation. *Molecular Immunology.* 44 (10), 2541–2548.

Arbouzova, N. I., Bach, E. A. & Zeidler, M. P. (2006) Ken & Barbie selectively regulates the expression of a subset of JAK/STAT pathway target genes. *Current Biology.* 16 (1), 80–88.

Arbouzova, N. I. & Zeidler, M. P. (2006) JAK/STAT signalling in *Drosophila*: insights into conserved regulatory and cellular functions. *Development*. 133 (14), 2605–2616.

Arnot, C. J., Gay, N. J. & Gangloff, M. (2010) Molecular mechanism that induces activation of Spätzle, the ligand for the *Drosophila* Toll receptor. *The Journal of Biological Chemistry.* 285 (25), 19502–19509.

Ashok, M., Turner, C. & Wilson, T. G. (1998) Insect juvenile hormone resistance gene homology with the bHLH-PAS family of transcriptional regulators. *Proceedings of the National Academy of Sciences.* 95 (6), 2761–2766.

Bahia, A. C., Dong, Y., Blumberg, B. J., Mlambo, G., Tripathi, A., BenMarzouk-Hidalgo, O. J., Chandra, R. & Dimopoulos, G. (2014) Exploring *Anopheles* gut bacteria for *Plasmodium* blocking activity. *Environmental Microbiology.* 16 (9), 2980–2994.

Bahia, A., Kubota, M., Tempone, A., Araújo, H., Guedes, B., Orfanó, A., Tadei, W., Ríos-Velásquez, C., Han, Y., Secundino, N., Barillas-Mury, C., Pimenta, P. & Traub-Csekö, Y. (2011) The JAK-STAT pathway controls *Plasmodium vivax* load in early stages of *Anopheles aquasalis* infection. *PLoS Neglected Tropical Diseases.* 5 (11), e1317.

Bai, L., Wang, L., Vega-Rodríguez, J., Wang, G. & Wang, S. (2019) A gut symbiotic bacterium *Serratia marcescens* renders mosquito resistance to *Plasmodium* infection through activation of mosquito immune responses. *Frontiers in Microbiology.* 10, 1580.

Bando, H., Okado, K., Guelbeogo, W. M., Badolo, A., Aonuma, H., Nelson, B., Fukumoto, S., Xuan, X., Sagnon, N. & Kanuka, H. (2013) Intra-specific diversity of *Serratia marcescens* in *Anopheles* mosquito midgut defines *Plasmodium* transmission capacity. *Scientific Reports.* 3 (1), 1–9.

Barillas-Mury, C., Charlesworth, A., Gross, I., Richman, A., Hoffmann, J. A. & Kafatos, F. C. (1996) Immune factor Gambif1, a new rel family member from the human malaria vector, *Anopheles gambiae*. *The EMBO Journal.* 15 (17), 4691–4701.

Barillas-Mury, C., Han, Y., Seeley, D. & Kafatos, F. C. (1999) *Anopheles gambiae* Ag-STAT, a new insect member of the STAT family, is activated in response to bacterial infection. *The EMBO Journal.* 18 (4), 959–967.

Barletta, A. B. F., Trisnadi, N., Ramirez, J. L. & Barillas-Mury, C. (2019) Mosquito midgut prostaglandin release establishes systemic immune priming. *iScience.* 19, 54–62.

Barr, P. J., Green, K. M., Gibson, H. L., Bathurst, I. C., Quakyi, I. A. & Kaslow, D. C. (1991) Recombinant Pfs25 protein of *Plasmodium falciparum* elicits malaria transmission-blocking immunity in experimental animals. *The Journal of Experimental Medicine.* 174 (5), 1203–1208.

Bashir-Tanoli, S. & Tinsley, M. C. (2014) Immune response costs are associated with changes in resource acquisition and not resource reallocation. *Functional Ecology.* 28 (4), 1011–1019.

Baton, L. A., Robertson, A., Warr, E., Strand, M. R. & Dimopoulos, G. (2009) Genome-wide transcriptomic profiling of *Anopheles gambiae* hemocytes reveals pathogen-specific signatures upon bacterial challenge and *Plasmodium berghei* infection. *BMC Genomics.* 10 (1), 257.

Baxter, R. H., Steinert, S., Chelliah, Y., Volohonsky, G., Levashina, E. A. & Deisenhofer, J. (2010) A heterodimeric complex of the LRR proteins LRIM1 and APL1C regulates complement-like immunity in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences.* 107 (39), 16817–16822.

Beier, M. S., Pumpuni, C. B., Beier, J. C. & Davis, J. R. (1994) Effects of para-aminobenzoic acid, insulin, and gentamicin on *Plasmodium falciparum* development in anopheline mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology.* 31 (4), 561–565.

Beller, M., Thiel, K., Thul, P. J. & Jäckle, H. (2010) Lipid droplets: a dynamic organelle moves into focus. *FEBS Letters.* 584 (11), 2176–2182.

Bellés, X., Martín, D. & Piulachs, M. (2005) The mevalonate pathway and the synthesis of juvenile hormone in insects. *Annual Review of Entomology.* 50 (1), 181–199.

Betz, A., Lampen, N., Martinek, S., Young, M. W. & Darnell Jr, J. E. (2001) A *Drosophila* PIAS homologue negatively regulates *stat92E*. *Proceedings of the National Academy of Sciences.* 98 (17), 9563–9568.

Billingsley, P. F. & Hecker, H. (1991) Blood digestion in the mosquito, *Anopheles stephensi* Liston (Diptera: Culicidae): activity and distribution of trypsin, aminopeptidase, and α-glucosidase in the midgut. *Journal of Medical Entomology.* 28 (6), 865–871.

Blandin, S., Moita, L. F., Köcher, T., Wilm, M., Kafatos, F. C. & Levashina, E. A. (2002) Reverse genetics in the mosquito *Anopheles gambiae*: targeted disruption of the *Defensin* gene. *EMBO Reports.* 3 (9), 852–856.

Blandin, S., Shiao, S., Moita, L. F., Janse, C. J., Waters, A. P., Kafatos, F. C. & Levashina, E. A. (2004) Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell.* 116 (5), 661–670.

Boissière, A., Tchioffo, M. T., Bachar, D., Abate, L., Marie, A., Nsango, S. E., Shahbazkia, H. R., Awono-Ambene, P. H., Levashina, E. A. & Christen, R. (2012) Midgut microbiota of the malaria mosquito vector *Anopheles gambiae* and interactions with *Plasmodium falciparum* infection. *PLoS Pathogens.* 8 (5), e1002742.

Bosco-Drayon, V., Poidevin, M., Boneca, I. G., Narbonne-Reveau, K., Royet, J. & Charroux, B. (2012) Peptidoglycan sensing by the receptor PGRP-LE in the *Drosophila* gut induces immune responses to infectious bacteria and tolerance to microbiota. *Cell Host & Microbe.* 12 (2), 153–165.

Boutros, M., Agaisse, H. & Perrimon, N. (2002) Sequential activation of signaling pathways during innate immune responses in *Drosophila*. *Developmental Cell.* 3 (5), 711–722.

Briegel, H. & Hörler, E. (1993) Multiple blood meals as a reproductive strategy in *Anopheles* (Diptera: Culicidae). *Journal of Medical Entomology.* 30 (6), 975–985.

Brogden, K. A. (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Reviews Microbiology.* 3 (3), 238–250.

Brown, L. D., Shapiro, L. L., Thompson, G. A., Estévez-Lao, T. Y. & Hillyer, J. F. (2019) Transstadial immune activation in a mosquito: Adults that emerge from infected larvae have stronger antibacterial activity in their hemocoel yet increased susceptibility to malaria infection. *Ecology and Evolution.* 9 (10), 6082–6095.

Brown, S., Hu, N. & Hombría, J. C. (2001) Identification of the first invertebrate interleukin JAK/STAT receptor, the *Drosophila* gene *domeless*. *Current Biology.* 11 (21), 1700–1705.

Buchon, N., Broderick, N. A., Poidevin, M., Pradervand, S. & Lemaitre, B. (2009) *Drosophila* intestinal response to bacterial infection: activation of host defense and stem cell proliferation. *Cell Host & Microbe.* 5 (2), 200–211.
Bukhari, T., Aimanianda, V., Bischoff, E., Brito-Fravallo, E., Eiglmeier, K., Riehle, M. M., Vernick, K. D. & Mitri, C. (2022) Genetics and immunity of *Anopheles* response to the entomopathogenic fungus *Metarhizium anisopliae* overlap with immunity to *Plasmodium*. *Scientific Reports. 1*2 (1), 6315.
Bulet, P., Hetru, C., Dimarcq, J. & Hoffmann, D. (1999) Antimicrobial peptides in insects; structure and function. *Developmental & Comparative Immunology.* 23 (4–5), 329–344.
Burkett, D. A., Carlson, D. A. & Kline, D. L. (1998) Analysis of composition of sugar meals of wild mosquitoes by gas chromatography. *Journal of the American Mosquito Control Association.* 14 (4), 373–379.
Burkot, T. R., Zhou, W. D., Geysen, H. M., Wirtz, R. A. & Saul, A. (1991) Fine specificities of monoclonal antibodies against the *Plasmodium falciparum* circumsporozoite protein: recognition of both repetitive and non-repetitive regions. *Parasite Immunology.* 13 (2), 161–170.
Camilli, G. & Hopkins, K. D. (1979) Testing for association in 2 × 2 contingency tables with very small sample sizes. *Psychological Bulletin.* 86 (5), 1011.
Canepa, G. E., Molina-Cruz, A. & Barillas-Mury, C. (2016) Molecular analysis of Pfs47-mediated *Plasmodium* evasion of mosquito immunity. *PLoS ONE.* 11 (12), e0168279.
Cappelli, A., Damiani, C., Mancini, M. V., Valzano, M., Rossi, P., Serrao, A., Ricci, I. & Favia, G. (2019) *Asaia* activates immune genes in mosquito eliciting an anti-*Plasmodium* response: implications in malaria control. *Frontiers in Genetics.* 10, 836.
Carissimo, G., Pondeville, E., McFarlane, M., Dietrich, I., Mitri, C., Bischoff, E., Antoniewski, C., Bourgouin, C., Failloux, A. & Kohl, A. (2015) Antiviral immunity of *Anopheles gambiae* is highly compartmentalized, with distinct roles for RNA interference and gut microbiota. *Proceedings of the National Academy of Sciences.* 112 (2), E176–E185.
Carlsson, A., Nyström, T., de Cock, H. & Bennich, H. (1998) Attacin-an insect immune protein-binds LPS and triggers the specific inhibition of bacterial outer-membrane protein synthesis. *Microbiology.* 144 (8), 2179–2188.
Castellano, B. M., Thelen, A. M., Moldavski, O., Feltes, M., Van Der Welle, R. E., Mydock-McGrane, L., Jiang, X., Van Eijkeren, R. J., Davis, O. B. & Louie, S. M. (2017) Lysosomal cholesterol activates mTORC1 via an SLC38A9–Niemann-Pick C1 signaling complex. *Science.* 355 (6331), 1306–1311.
Castillo, J. C., Robertson, A. E. & Strand, M. R. (2006) Characterization of hemocytes from the mosquitoes *Anopheles gambiae* and *Aedes aegypti*. *Insect Biochemistry and Molecular Biology.* 36 (12), 891–903.
Cavodeassi, F., Modolell, J. & Gómez-Skarmeta, J. L. (2001) The Iroquois family of genes: from body building to neural patterning. *Development.* 128 (15), 2847–2855.
Chabanol, E., Behrends, V., Prévot, G., Christophides, G. K. & Gendrin, M. (2020) Antibiotic treatment in *Anopheles coluzzii* affects carbon and nitrogen metabolism. *Pathogens.* 9 (9), 679.
Champion, C. J. & Xu, J. (2018) Redox state affects fecundity and insecticide susceptibility in *Anopheles gambiae*. *Scientific Reports.* 8 (1), 13054.
Chamy, L. E., Leclerc, V., Caldelari, I. & Reichhart, J. (2008) Sensing of 'danger signals' and pathogen-associated molecular patterns defines binary signaling pathways 'upstream' of Toll. *Nature Immunology.* 9 (10), 1165–1170.
Chandra, M., Liniger, M., Tetley, L., Roditi, I. & Barry, J. D. (2004) TsetseEP, a gut protein from the tsetse *Glossina morsitans*, is related to a major surface glycoprotein of trypanosomes transmitted by the fly and to the products of a *Drosophila* gene family. *Insect Biochemistry and Molecular Biology.* 34 (11), 1163–1173.
Chang, M., Wang, Y., Yang, Q., Wang, X., Wang, M., Raikhel, A. S. & Zou, Z. (2021) Regulation of antimicrobial peptides by juvenile hormone and its receptor, Methoprene-tolerant, in the mosquito *Aedes aegypti*. *Insect Biochemistry and Molecular Biology.* 128, 103509.
Chang, Y., Yang, C., Sun, R., Cheng, Y., Kao, W. & Yang, P. (2013) Rapid single cell detection of *Staphylococcus aureus* by aptamer-conjugated gold nanoparticles. *Scientific Reports.* 3 (1), 1863.
Chen, J., Wu, Y., Chen, J., Zhu, X., Merkler, D., Liao, C. & Han, Q. (2023) Elongases of long-chain fatty acids ELO2 and ELO9 are involved in cuticle formation and function in fecundity in the yellow fever mosquito, *Aedes aegypti*. *Insects.* 14 (2), 189.
Chen, M., Yan, T., Shen, Q., Lu, X., Pan, Q., Huang, Y., Tang, Y., Fu, X., Liu, M. & Jiang, W. (2017) GLANDULAR TRICHOME-SPECIFIC WRKY 1 promotes artemisinin biosynthesis in *Artemisia annua*. *New Phytologist.* 214 (1), 304–316.
Chen, P., Li, G., Guo, X., He, K., Fu, Y., Fu, L. & Song, Y. (1994) The infectivity of gametocytes of *Plasmodium falciparum* from patients treated with artemisinin. *Chinese Medical Journal.* 107 (09), 709–711.
Choe, K. M., Werner, T., Stöven, S., Hultmark, D. & Anderson, K. V. (2002) Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in *Drosophila*. *Science.* 296 (5566), 359–362.
Choe, K., Lee, H. & 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.* 102 (4), 1122–1126.

Chouaia, B., Rossi, P., Epis, S., Mosca, M., Ricci, I., Damiani, C., Ulissi, U., Crotti, E., Daffonchio, D. & Bandi, C. (2012) Delayed larval development in *Anopheles* mosquitoes deprived of *Asaia* bacterial symbionts. *BMC Microbiology.* 12 (1), S2.
Christensen, B., Fink, J., Merrifield, R. B. & Mauzerall, D. (1988) Channel-forming properties of cecropins and related model compounds incorporated into planar lipid membranes. *Proceedings of the National Academy of Sciences.* 85 (14), 5072–5076.
Christensen, B. M., Li, J., Chen, C. & Nappi, A. J. (2005) Melanization immune responses in mosquito vectors. *Trends in Parasitology.* 21 (4), 192–199.
Christophides, G. K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P. T., Collins, F. H., Danielli, A., Dimopoulos, G., Hetru, C., Hoa, N. T., Hoffmann, J. A., Kanzok, S. M., Letunic, I., Levashina, E. A., Loukeris, T. G., Lycett, G., Meister, S., Michel, K., Moita, L. F., Muller, H. M., Osta, M. A., Paskewitz, S. M., Reichhart, J. M., Rzhetsky, A., Troxler, L., Vernick, K. D., Vlachou, D., Volz, J., von Mering, C., Xu, J., Zheng, L., Bork, P. & Kafatos, F. C. (2002) Immunity-related genes and gene families in *Anopheles gambiae*. *Science.* 298 (5591), 159–165.
Cirimotich, C. M., Dong, Y., Garver, L. S., Sim, S. & Dimopoulos, G. (2010) Mosquito immune defenses against *Plasmodium* infection. *Developmental and Comparative Immunology.* 34 (4), 387–395.
Cirimotich, C. M., Dong, Y., Clayton, A. M., Sandiford, S. L., Souza-Neto, J. A., Mulenga, M. & Dimopoulos, G. (2011) Natural microbe-mediated refractoriness to *Plasmodium* infection in *Anopheles gambiae*. *Science.* 332 (6031), 855–858.
Clark, E. & Peel, A. D. (2018) Evidence for the temporal regulation of insect segmentation by a conserved sequence of transcription factors. *Development.* 145 (10), dev155580.
Clark, E., Peel, A. D. & Akam, M. (2019) Arthropod segmentation. *Development.* 146 (18), dev170480.
Clay, K. (2014) Defensive symbiosis: a microbial perspective. *Functional Ecology.* 28 (2), 293–298.
Clayton, A. M., Cirimotich, C. M., Dong, Y. & Dimopoulos, G. (2013) Caudal is a negative regulator of the *Anopheles* IMD pathway that controls resistance to *Plasmodium falciparum* infection. *Developmental & Comparative Immunology.* 39 (4), 323–332.
Clements, A. N. (1992) *The biology of mosquitoes: development, nutrition and reproduction.* London, Chapman & Hall.
Cociancich, S., Ghazi, A., Hetru, C., Hoffmann, J. A. & Letellier, L. (1993) Insect defensin, an inducible antibacterial peptide, forms voltage-dependent channels in *Micrococcus luteus*. *Journal of Biological Chemistry.* 268 (26), 19239–19245.
Coetzee, M., Hunt, R. H., Wilkerson, R., Della Torre, A., Coulibaly, M. B. & Besansky, N. J. (2013) *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa.* 3619 (3), 246–274.
Collins, F. H., Sakai, R. K., Vernick, K. D., Paskewitz, S., Seeley, D. C., Miller, L. H., Collins, W. E., Campbell, C. C. & Gwadz, R. W. (1986) Genetic selection of a *Plasmodium*-refractory strain of the malaria vector *Anopheles gambiae*. *Science.* 234 (4776), 607–610.
Collins, F. H. & Besansky, N. J. (1994) Vector biology and the control of malaria in Africa. *Science.* 264 (5167), 1874.
Collins, F. H. & Paskewitz, S. M. (1995) Malaria: current and future prospects for control. *Annual Review of Entomology.* 40, 195–219.
Connolly, J. B., Mumford, J. D., Fuchs, S., Turner, G., Beech, C., North, A. R. & Burt, A. (2021) Systematic identification of plausible pathways to potential harm via problem formulation for investigational releases of a population suppression gene drive to control the human malaria vector *Anopheles gambiae* in West Africa. *Malaria Journal.* 20 (1), 1–69.
Coon, K. L., Vogel, K. J., Brown, M. R. & Strand, M. R. (2014) Mosquitoes rely on their gut microbiota for development. *Molecular Ecology.* 23 (11), 2727–2739.
Corby-Harris, V., Drexler, A., Watkins de Jong, L., Antonova, Y., Pakpour, N., Ziegler, R., Ramberg, F., Lewis, E. E., Brown, J. M. & Luckhart, S. (2010) Activation of *Akt* signaling reduces the prevalence and intensity of malaria parasite infection and lifespan in *Anopheles stephensi* mosquitoes. *PLoS Pathogens.* 6 (7), e1001003.
Cox, C. R. & Gilmore, M. S. (2007) Native microbial colonization of *Drosophila melanogaster* and its use as a model of *Enterococcus faecalis* pathogenesis. *Infection and Immunity.* 75 (4), 1565–1576.
Cox, F. E. (2010) History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors.* 3 (1), 1–9.
Cronin, S. J., Nehme, N. T., Limmer, S., Liegeois, S., Pospisilik, J. A., Schramek, D., Leibbrandt, A., Simoes, R. d. M., Gruber, S. & Puc, U. (2009) Genome-wide RNAi screen identifies genes involved in intestinal pathogenic bacterial infection. *Science.* 325 (5938), 340–343.
Cui, Y., Niu, G., Li, V. L., Wang, X. & Li, J. (2020) Analysis of blood-induced *Anopheles gambiae* midgut proteins and sexual stage *Plasmodium falciparum* interaction reveals mosquito genes important for malaria transmission. *Scientific Reports.* 10 (1), 14316.
Cullen, B. R. (2014) Viruses and RNA Interference: Issues and Controversies. *Journal of Virology.* 88 (22), 12934–12936.
Cyran, S. A., Buchsbaum, A. M., Reddy, K. L., Lin, M., Glossop, N. R., Hardin, P. E., Young, M. W., Storti, R. V. & Blau, J. (2003) *vrille, Pdp1*, and *dClock* form a second feedback loop in the *Drosophila* circadian clock. *Cell.* 112 (3), 329–341.

Damiani, C., Ricci, I., Crotti, E., Rossi, P., Rizzi, A., Scuppa, P., Capone, A., Ulissi, U., Epis, S. & Genchi, M. (2010) Mosquito-bacteria symbiosis: the case of *Anopheles gambiae* and *Asaia*. *Microbial Ecology.* 60, 644–654.
Danielli, A., Kafatos, F. C. & Loukeris, T. G. (2003) Cloning and characterization of four *Anopheles gambiae* serpin isoforms, differentially induced in the midgut by *Plasmodium berghei* invasion. *Journal of Biological Chemistry.* 278 (6), 4184–4193.
Das, S. & Dimopoulos, G. (2008) Molecular analysis of photic inhibition of blood-feeding in *Anopheles gambiae*. *BMC Physiology.* 8 (1), 1–19.
Datoo, M. S., Natama, H. M., Somé, A., Bellamy, D., Traoré, O., Rouamba, T., Tahita, M. C., Ido, N. F. A., Yameogo, P. & Valia, D. (2022) Efficacy and immunogenicity of R21/Matrix-M vaccine against clinical malaria after 2 years' follow-up in children in Burkina Faso: a phase 1/2b randomised controlled trial. *The Lancet Infectious Diseases.* 22 (12), 1728–1736.
de Almeida Oliveira, G., Lieberman, J. & Barillas-Mury, C. (2012) Epithelial nitration by a peroxidase/NOX5 system mediates mosquito antiplasmodial immunity. *Science.* 335 (6070), 856–859.
de Chaumont, F., Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., Meas-Yedid, V., Pankajakshan, P., Lecomte, T. & Le Montagner, Y. (2012) Icy: an open bioimage informatics platform for extended reproducible research. *Nature Methods.* 9 (7), 690–696.
DeJong, R. J., Miller, L. M., Molina-Cruz, A., Gupta, L., Kumar, S. & Barillas-Mury, C. (2007) Reactive oxygen species detoxification by catalase is a major determinant of fecundity in the mosquito *Anopheles gambiae*. *Proceedings of the National Academy of Sciences.* 104 (7), 2121–2126.
Dekmak, A. S., Yang, X., Zu Dohna, H., Buchon, N. & Osta, M. A. (2021) The route of infection influences the contribution of key immunity genes to antibacterial defense in *Anopheles gambiae*. *Journal of Innate Immunity.* 13 (2), 107–126.
Demaio, J., Pumpuni, C. B., Kent, M. & Beier, J. C. (1996) The midgut bacterial flora of wild *Aedes triseriatus, Culex pipiens*, and *Psorophora columbiae* mosquitoes. *The American Journal of Tropical Medicine and Hygiene.* 54 (2), 219–223.
Desai, B. N., Myers, B. R. & Schreiber, S. L. (2002) FKBP12-rapamycin-associated protein associates with mitochondria and senses osmotic stress via mitochondrial dysfunction. *Proceedings of the National Academy of Sciences.* 99 (7), 4319–4324.
Dimopoulos, G., Christophides, G. K., Meister, S., Schultz, J., White, K. P., Barillas-Mury, C. & Kafatos, F. C. (2002) Genome expression analysis of *Anopheles gambiae*: responses to injury, bacterial challenge, and malaria infection. *Proceedings of the National Academy of Sciences.* 99 (13), 8814–8819.
Dimopoulos, G., Seeley, D., Wolf, A. & Kafatos, F. C. (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), 6115–6123.
Ding, S. & Voinnet, O. (2007) Antiviral immunity directed by small RNAs. *Cell.* 130 (3), 413–426.
Dinglasan, R. R., Devenport, M., Florens, L., Johnson, J. R., McHugh, C. A., Donnelly-Doman, M., Carucci, D. J., Yates III, J. R. & Jacobs-Lorena, M. (2009) The *Anopheles gambiae* adult midgut peritrophic matrix proteome. *Insect Biochemistry and Molecular Biology.* 39 (2), 125–134.
Dong, Y., Aguilar, R., Xi, Z., Warr, E., Mongin, E. & Dimopoulos, G. (2006) *Anopheles gambiae* immune responses to human and rodent *Plasmodium* parasite species. *PLoS Pathogens.* 2 (6), e52.
Dong, Y., Cirimotich, C. M., Pike, A., Chandra, R. & Dimopoulos, G. (2012) *Anopheles* NF-κB-regulated splicing factors direct pathogen-specific repertoires of the hypervariable pattern recognition receptor AgDscam. *Cell Host & Microbe.* 12 (4), 521–530.
Dong, Y., Das, S., Cirimotich, C., Souza-Neto, J. A., McLean, K. J. & Dimopoulos, G. (2011) Engineered A*nopheles* immunity to *Plasmodium* infection. *PLoS Pathogens.* 7 (12), e1002458.
Dong, Y. & Dimopoulos, G. (2009) *Anopheles* fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. *Journal of Biological Chemistry.* 284 (15), 9835–9844.
Dong, Y., Manfredini, F. & Dimopoulos, G. (2009) Implication of the mosquito midgut microbiota in the defense against malaria parasites. *PLoS Pathogens.* 5 (5), e1000423.
Dong, Y., Simões, M. L. & Dimopoulos, G. (2020) Versatile transgenic multistage effector-gene combinations for *Plasmodium falciparum* suppression in *Anopheles*. *Science Advances.* 6 (20), eaay5898.
Dong, Y., Taylor, H. E. & Dimopoulos, G. (2006) AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biology.* 4 (7), e229.
Dostert, C., Jouanguy, E., Irving, P., Troxler, L., Galiana-Arnoux, D., Hetru, C., Hoffmann, J. A. & Imler, J. (2005) The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of *Drosophila*. *Nature Immunology.* 6 (9), 946–953.
Edwards, M. J., Lemos, F. J., Donnelly-Doman, M. & Jacobs-Lorena, M. (1997) Rapid induction by a blood meal of a carboxypeptidase gene in the gut of the mosquito *Anopheles gambiae*. *Insect Biochemistry and Molecular Biology.* 27 (12), 1063–1072.
Engel, P. & Moran, N. A. (2013) The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews.* 37 (5), 699–735.

Epand, R. M. & Vogel, H. J. (1999) Diversity of antimicrobial peptides and their mechanisms of action. *Biochimica et Biophysica Acta (BBA) - Biomembranes.* 1462 (1–2), 11–28.
Fabian, D. K., Garschall, K., Klepsatel, P., Santos-Matos, G., Sucena, E., Kapun, M., Lemaitre, B., Schlötterer, C., Arking, R. & Flatt, T. (2018) Evolution of longevity improves immunity in *Drosophila*. *Evolution Letters.* 2 (6), 567–579.
Fang, W., Vega-Rodriguez, J., Ghosh, A. K., Jacobs-Lorena, M., Kang, A. & St Leger, R. J. (2011) Development of transgenic fungi that kill human malaria parasites in mosquitoes. *Science.* 331 (6020), 1074–1077.
Favia, G., Ricci, I., Damiani, C., Raddadi, N., Crotti, E., Marzorati, M., Rizzi, A., Urso, R., Brusetti, L. & Borin, S. (2007) Bacteria of the genus *Asaia* stably associate with *Anopheles stephensi*, an Asian malarial mosquito vector. *Proceedings of the National Academy of Sciences.* 104 (21), 9047–9051.
Felix, C. R., Betschart, B., Billingsley, P. F. & Freyvogel, T. A. (1991) Post-feeding induction of trypsin in the midgut of *Aedes aegypti* L. (Diptera: Culicidae) is separable into two cellular phases. *Insect Biochemistry.* 21 (2), 197–203.
Fellowes, M., Kraaijeveld, A. R. & Godfray, H. (1999) The relative fitness of *Drosophila melanogaster* (Diptera, Drosophilidae) that have successfully defended themselves against the parasitoid *Asobara tabida* (Hymenoptera, Braconidae). *Journal of Evolutionary Biology.* 12 (1), 123–128.
Flatt, T., Heyland, A., Rus, F., Porpiglia, E., Sherlock, C., Yamamoto, R., Garbuzov, A., Palli, S. R., Tatar, M. & Silverman, N. (2008) Hormonal regulation of the humoral innate immune response in *Drosophila melanogaster*. *Journal of Experimental Biology.* 211 (16), 2712–2724.
Foster, W. A. (1995) Mosquito sugar feeding and reproductive energetics. *Annual Review of Entomology.* 40 (1), 443–474.
Fraiture, M., Baxter, R. H., Steinert, S., Chelliah, Y., Frolet, C., Quispe-Tintaya, W., Hoffmann, J. A., Blandin, S. A. & Levashina, E. A. (2009) Two mosquito LRR proteins function as complement control factors in the TEP1-mediated killing of *Plasmodium*. *Cell Host & Microbe.* 5 (3), 273–284.
Franc, N. C., Heitzler, P., Alan B, R., Ezekowitz & White, K. (1999) Requirement for Croquemort in phagocytosis of apoptotic cells in *Drosophila*. *Science.* 284 (5422), 1991–1994.
Frolet, C., Thoma, M., Blandin, S., Hoffmann, J. A. & Levashina, E. A. (2006) Boosting NF-κB-dependent basal immunity of *Anopheles gambiae* aborts development of *Plasmodium berghei*. *Immunity.* 25 (4), 677–685.
Fuss, B., Josten, F., Feix, M. & Hoch, M. (2004) Cell movements controlled by the Notch signalling cascade during foregut development in *Drosophila*. *Development.* 131 (7), 1587–1595.
Galili, T. (2015) dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics.* 31 (22), 3718–3720.
García Gil de Muñoz, F. L., Martínez-Barnetche, J., Lanz-Mendoza, H., Rodríguez, M. H. & Hernández-Hernández, F. C. (2008) Prostaglandin E2 modulates the expression of antimicrobial peptides in the fat body and midgut of *Anopheles albimanus*. *Archives of Insect Biochemistry and Physiology.* 68 (1), 14–25.
Garver, L. S., Bahia, A. C., Das, S., Souza-Neto, J. A., Shiao, J., Dong, Y. & Dimopoulos, G. (2012) *Anopheles* Imd pathway factors and effectors in infection intensity-dependent anti-*Plasmodium* action. *PLoS Pathogens.* 8 (6), e1002737.
Garver, L. S., de Almeida Oliveira, G. & Barillas-Mury, C. (2013) The JNK pathway is a key mediator of *Anopheles gambiae* antiplasmodial immunity. *PLoS Pathogens.* 9 (9), e1003622.
Garver, L. S., Dong, Y. & Dimopoulos, G. (2009) Caspar controls resistance to *Plasmodium falciparum* in diverse anopheline species. *PLoS Pathogens.* 5 (3), e1000335.
Gendrin, M., Turlure, F., Rodgers, F. H., Cohuet, A., Morlais, I. & Christophides, G. K. (2017) The peptidoglycan recognition proteins PGRPLA and PGRPLB regulate *Anopheles* immunity to bacteria and affect infection by *Plasmodium*. *Journal of Innate Immunity.* 9 (4), 333–342.
Gendrin, M., Rodgers, F. H., Yerbanga, R. S., Ouédraogo, J. B., Basáñez, M., Cohuet, A. & Christophides, G. K. (2015) Antibiotics in ingested human blood affect the mosquito microbiota and capacity to transmit malaria. *Nature Communications.* 6, 5921.
Gendrin, M., Zaidman-Rémy, A., Broderick, N. A., Paredes, J., Poidevin, M., Roussel, A. & Lemaitre, B. (2013) Functional analysis of PGRP-LA in *Drosophila* immunity. *PLoS ONE.* 8 (7), e69742.
Gillies, M. T. (1953) The duration of the gonotrophic cycle in *Anopheles gambiae* and *Anopheles funestus*, with a note on the efficiency of hand catching. *East African Medical Journal.* 30 (4), 129–135.
Glossop, N. R., Houl, J. H., Zheng, H., Ng, F. S., Dudek, S. M. & Hardin, P. E. (2003) VRILLE feeds back to control circadian transcription of *Clock* in the *Drosophila* circadian oscillator. *Neuron.* 37 (2), 249–261.
Gobert, V., Gottar, M., Matskevich, A. A., Rutschmann, S., Royet, J., Belvin, M., Hoffmann, J. A. & Ferrandon, D. (2003) Dual activation of the *Drosophila* Toll pathway by two pattern recognition receptors. *Science.* 302 (5653), 2126–2130.
Gokudan, S., Muta, T., Tsuda, R., Koori, K., Kawahara, T., Seki, N., Mizunoe, Y., Wai, S. N., Iwanaga, S. & Kawabata, S. (1999) Horseshoe crab acetyl group-recognizing lectins involved in innate immunity are structurally related to fibrinogen. *Proceedings of the National Academy of Sciences.* 96 (18), 10086–10091.

Gómez-Skarmeta, J., del Corral, R. D., de la Calle-Mustienes, E., Ferrés-Marcó, D. & Modolell, J. (1996) *araucan* and *caupolican*, two members of the novel Iroquois complex, encode homeoproteins that control proneural and vein-forming genes. *Cell.* 85 (1), 95–105.
Gonzalez-Ceron, L., Santillan, F., Rodriguez, M. H., Mendez, D. & Hernandez-Avila, J. E. (2003) Bacteria in midguts of field-collected *Anopheles albimanus* block *Plasmodium vivax* sporogonic development. *Journal of Medical Entomology.* 40 (3), 371–374.
González-Lázaro, M., Dinglasan, R. R., de la Cruz Hernández-Hernández, F., Rodríguez, M. H., Laclaustra, M., Jacobs-Lorena, M. & Flores-Romo, L. (2009) *Anopheles gambiae* Croquemort SCRBQ2, expression profile in the mosquito and its potential interaction with the malaria parasite *Plasmodium berghei*. *Insect Biochemistry and Molecular Biology.* 39 (5–6), 395–402.
Gottar, M., Gobert, V., Matskevich, A. A., Reichhart, J., Wang, C., Butt, T. M., Belvin, M., Hoffmann, J. A. & Ferrandon, D. (2006) Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors. *Cell.* 127 (7), 1425–1437.
Graça-Souza, A. V., Maya-Monteiro, C., Paiva-Silva, G. O., Braz, G. R., Paes, M. C., Sorgine, M. H., Oliveira, M. F. & Oliveira, P. L. (2006) Adaptations against heme toxicity in blood-feeding arthropods. *Insect Biochemistry and Molecular Biology.* 36 (4), 322–335.
Guillou, A., Troha, K., Wang, H., Franc, N. C. & Buchon, N. (2016) The *Drosophila* CD36 homologue *croquemort* is required to maintain immune and gut homeostasis during development and aging. *PLoS Pathogens.* 12 (10), e1005961.
Gupta, L., Molina-Cruz, A., Kumar, S., Rodrigues, J., Dixit, R., Zamora, R. E. & Barillas-Mury, C. (2009) The STAT pathway mediates late-phase immunity against *Plasmodium* in the mosquito *Anopheles gambiae*. *Cell Host & Microbe.* 5 (5), 498–507.
Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology.* 19 (3), 1720–1730.
Habtewold, T., Tapanelli, S., Masters, E. K., Hoermann, A., Windbichler, N. & Christophides, G. K. (2019) Streamlined SMFA and mosquito dark-feeding regime significantly improve malaria transmission-blocking assay robustness and sensitivity. *Malaria Journal.* 18 (1), 1–11.
Habtewold, T., Tapanelli, S., Masters, E. K., Windbichler, N. & Christophides, G. K. (2022) The circadian clock modulates *Anopheles gambiae* infection with *Plasmodium falciparum*. *PLoS ONE.* 17 (12), e0278484.
Haines, L. R., Jackson, A. M., Lehane, M. J., Thomas, J. M., Yamaguchi, A. Y., Haddow, J. D. & Pearson, T. W. (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), 413–423.
Haines, L. R., Lehane, S. M., Pearson, T. W. & Lehane, M. J. (2010) Tsetse EP protein protects the fly midgut from trypanosome establishment. *PLoS Pathogens.* 6 (3), e1000793.
Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E. & Russell, S. (2016) A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nature Biotechnology.* 34 (1), 78–83.
Hammond, S. M. (2005) Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS Letters.* 579 (26), 5822–5829.
Han, Y. S., Thompson, J., Kafatos, F. C. & Barillas-Mury, C. (2000) Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes. *The EMBO Journal.* 19 (22), 6030–6040.
Hancock, P. A., Heniks, C. J. M., Tangena, J. A., Gibson, H., Hemingway, J., Coleman, M., Gething, P. W., Cameron, E., Bhatt, S. & Moyes, C. L. (2020) Mapping trends in insecticide resistance phenotypes in African malaria vectors. *PLoS Biology.* 18 (6), e3000633.
Hanson, M. A. & Lemaitre, B. (2020) New insights on *Drosophila* antimicrobial peptide function in host defense and beyond. *Current Opinion in Immunology.* 62, 22–30.
Hao, Z., Kasumba, I. & Aksoy, S. (2003) Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera: Glossinidiae). *Insect Biochemistry and Molecular Biology.* 33 (11), 1155–1164.
Hecker, H. (1977) Structure and function of midgut epithelial cells in Culicidae mosquitoes (Insecta, Diptera). *Cell and Tissue Research.* 184, 321–341.
Hegedus, D., Erlandson, M., Gillott, C. & Toprak, U. (2009) New insights into peritrophic matrix synthesis, architecture, and function. *Annual Review of Entomology.* 54, 285–302.
Hillyer, J. F., Barreau, C. & Vernick, K. D. (2007) Efficiency of salivary gland invasion by malaria sporozoites is controlled by rapid sporozoite destruction in the mosquito haemocoel. *International Journal for Parasitology.* 37 (6), 673–681.
Hillyer, J. F. & Estévez-Lao, T. Y. (2010) Nitric oxide is an essential component of the hemocyte-mediated mosquito immune response against bacteria. *Developmental and Comparative Immunology.* 34 (2), 141–149.
Hillyer, J. F., Schmidt, S. L. & Christensen, B. M. (2003) Rapid phagocytosis and melanization of bacteria and *Plasmodium* sporozoites by hemocytes of the mosquito *Aedes aegypti*. *The Journal of Parasitology.* 89 (1), 62–69.

Hillyer, J. F. & Strand, M. R. (2014) Mosquito hemocyte-mediated immune responses. *Current Opinion in Insect Science.* 3, 14–21.
Hiruma, K. & Kaneko, Y. (2013) Hormonal regulation of insect metamorphosis with special reference to juvenile hormone biosynthesis. *Current Topics in Developmental Biology.* 103, 73–100.
Hixson, B., Bing, X., Yang, X., Bonfini, A., Nagy, P. & Buchon, N. (2022) A transcriptomic atlas of *Aedes aegypti* reveals detailed functional organization of major body parts and gut regional specializations in sugar-fed and blood-fed adult females. *eLife.* 11, e76132.
Hixson, B., Huot, L., Morejon, B., Yang, X., Nagy, P., Michel, K. & Buchon, N. (2023) The transcriptional response in mosquitoes distinguishes between fungi and bacteria but not Gram types. *bioRxiv.* 2023–07.
Hoermann, A., Habtewold, T., Selvaraj, P., Del Corsano, G., Capriotti, P., Inghilterra, M. G., Kebede, T. M., Christophides, G. K. & Windbichler, N. (2022) Gene drive mosquitoes can aid malaria elimination by retarding *Plasmodium* sporogonic development. *Science Advances.* 8 (38), eabo1733.
Hogg, J. C. & Hurd, H. (1995) *Plasmodium yoelii nigeriensis*: the effect of high and low intensity of infection upon the egg production and bloodmeal size of *Anopheles stephensi* during three gonotrophic cycles. *Parasitology.* 111 (5), 555–562.
Hollingdale, M. R., Nardin, E. H., Tharavanij, S., Schwartz, A. L. & Nussenzweig, R. S. (1984) Inhibition of entry of *Plasmodium falciparum* and *P. vivax* sporozoites into cultured cells; an *in vitro* assay of protective antibodies. *The Journal of Immunology.* 132 (2), 909–913.
Hong, S., Zhou, D., Chen, C., Wang, W., Lv, Y., Ye, Y., Zou, P., Yv, J., Chang, X. & Yv, X. (2013) Ribose-phosphate pyrophosphokinase 1 (PRPS1) associated with deltamethrin resistance in *Culex pipiens pallens*. *Parasitology Research.* 112, 847–854.
Horn, T. & Boutros, M. (2010) E-RNAi: a web application for the multi-species design of RNAi reagents—2010 update. *Nucleic Acids Research.* 38 (Web Server issue), W332–W339.
Horng, T. & Medzhitov, R. (2001) *Drosophila* MyD88 is an adapter in the Toll signaling pathway. *Proceedings of the National Academy of Sciences.* 98 (22), 12654–12658.
Horton, A. A., Wang, B., Camp, L., Price, M. S., Arshi, A., Nagy, M., Nadler, S. A., Faeder, J. R. & Luckhart, S. (2011) The mitogen-activated protein kinome from *Anopheles gambiae*: identification, phylogeny and functional characterization of the ERK, JNK and p38 MAP kinases. *BMC Genomics.* 12 (1), 574.
Hou, Y. C., Chittaranjan, S., Barbosa, S. G., McCall, K. & Gorski, S. M. (2008) Effector caspase Dcp-1 and IAP protein Bruce regulate starvation-induced autophagy during *Drosophila melanogaster* oogenesis. *The Journal of Cell Biology.* 182 (6), 1127–1139.
Huang, H. W., Chen, F. & Lee, M. (2004) Molecular mechanism of peptide-induced pores in membranes. *Physical Review Letters.* 92 (19), 198304.
Hughes, G. L., Dodson, B. L., Johnson, R. M., Murdock, C. C., Tsujimoto, H., Suzuki, Y., Patt, A. A., Cui, L., Nossa, C. W. & Barry, R. M. (2014) Native microbiome impedes vertical transmission of *Wolbachia* in *Anopheles* mosquitoes. *Proceedings of the National Academy of Sciences.* 111 (34), 12498–12503.
Iatsenko, I., Kondo, S., Mengin-Lecreulx, D. & Lemaitre, B. (2016) PGRP-SD, an extracellular pattern-recognition receptor, enhances peptidoglycan-mediated activation of the *Drosophila* Imd pathway. *Immunity.* 45 (5), 1013–1023.
Ip, Y. T., Kraut, R., Levine, M. & Rushlow, C. A. (1991) The *dorsal* morphogen is a sequence-specific DNA-binding protein that interacts with a long-range repression element in *Drosophila*. *Cell.* 64 (2), 439–446.
Isaacs, A. T., Li, F., Jasinskiene, N., Chen, X., Nirmala, X., Marinotti, O., Vinetz, J. M. & James, A. A. (2011) Engineered resistance to *Plasmodium falciparum* development in transgenic *Anopheles stephensi*. *PLoS Pathogens.* 7 (4), e1002017.
Janeway Jr., C. A. & Medzhitov, R. (2002) Innate immune recognition. *Annual Review of Immunology.* 20 (1), 197–216.
Jaramillo-Gutierrez, G., Molina-Cruz, A., Kumar, S. & Barillas-Mury, C. (2010) The *Anopheles gambiae* oxidation resistance 1 (OXR1) gene regulates expression of enzymes that detoxify reactive oxygen species. *PLoS ONE.* 5 (6), e11168.
Jevtov, I., Zacharogianni, M., van Oorschot, M. M., van Zadelhoff, G., Aguilera-Gomez, A., Vuillez, I., Braakman, I., Hafen, E., Stocker, H. & Rabouille, C. (2015) TORC2 mediates the heat stress response in *Drosophila* by promoting the formation of stress granules. *Journal of Cell Science.* 128 (14), 2497–2508.
Jindra, M., Bellés, X. & Shinoda, T. (2015) Molecular basis of juvenile hormone signaling. *Current Opinion in Insect Science.* 11, 39–46.
Josten, F., Fuss, B., Feix, M., Meissner, T. & Hoch, M. (2004) Cooperation of JAK/STAT and Notch signaling in the *Drosophila* foregut. *Developmental Biology.* 267 (1), 181–189.
Juhn, J. & James, A. A. (2012) Hybridization *in situ* of salivary glands, ovaries, and embryos of vector mosquitoes. *Journal of Visualized Experiments: JoVE.* (64), e3709.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A. & Potapenko, A. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature.* 596 (7873), 583–589.

Jung, C. H., Ro, S., Cao, J., Otto, N. M. & Kim, D. (2010) mTOR regulation of autophagy. *FEBS Letters.* 584 (7), 1287–1295.
Jürgens, G. & Weigel, D. (1988) Terminal versus segmental development in the *Drosophila* embryo: the role of the homeotic gene *fork head*. *Roux's Archives of Developmental Biology.* 197, 345–354.
Kajla, M. K., Andreeva, O., Gilbreath III, T. M. & Paskewitz, S. M. (2010) Characterization of expression, activity and role in antibacterial immunity of *Anopheles gambiae* lysozyme c-1. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology.* 155 (2), 201–209.
Kakani, P., Gupta, L. & Kumar, S. (2020) Heme-peroxidase 2, a peroxinectin-like gene, regulates bacterial homeostasis in *Anopheles stephensi* midgut. *Frontiers in Physiology.* 11, 572340.
Kaminskyy, V. & Zhivotovsky, B. (2012) Proteases in autophagy. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics.* 1824 (1), 44–50.
Kaneko, T., Yano, T., Aggarwal, K., Lim, J., Ueda, K., Oshima, Y., Peach, C., Erturk-Hasdemir, D., Goldman, W. E. & Oh, B. (2006) PGRP-LC and PGRP-LE have essential yet distinct functions in the *Drosophila* immune response to monomeric DAP-type peptidoglycan. *Nature Immunology.* 7 (7), 715–723.
Kassambara, A., Kosinski, M. & Biecek, P. (2020) *survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.8.* www.sthda.com/english/rpkgs/survminer.
Kassambara, A. & Mundt, F. (2020) *factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0.7.* http://cran.r-project.org/package=factoextra.
Keene, K. M., Foy, B. D., Sanchez-Vargas, I., Beaty, B. J., Blair, C. D. & Olson, K. E. (2004) RNA interference acts as a natural antiviral response to O'nyong-nyong virus (*Alphavirus*, Togaviridae) infection of *Anopheles gambiae*. *Proceedings of the National Academy of Sciences.* 101 (49), 17240–17245.
Khosravian, M., Faramarzi, F. K., Beigi, M. M., Behbahani, M. & Mohabatkar, H. (2013) Predicting antibacterial peptides by the concept of Chou's pseudo-amino acid composition and machine learning methods. *Protein and Peptide Letters.* 20 (2), 180–186.
Kibret, S., Wilson, G. G., Ryder, D., Tekie, H. & Petros, B. (2018) Can water-level management reduce malaria mosquito abundance around large dams in sub-Saharan Africa? *PLoS ONE.* 13 (4), e0196064.
Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology.* 37 (8), 907–915.
Kim, D., Sarbassov, D. D., Ali, S. M., King, J. E., Latek, R. R., Erdjument-Bromage, H., Tempst, P. & Sabatini, D. M. (2002) mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. *Cell.* 110 (2), 163–175.
Kim, M., Lee, J. H., Lee, S. Y., Kim, E. & Chung, J. (2006) Caspar, a suppressor of antibacterial immunity in *Drosophila*. *Proceedings of the National Academy of Sciences.* 103 (44), 16358–16363.
Kim, W., Koo, H., Richman, A. M., Seeley, D., Vizioli, J., Klocko, A. D. & O'brochta, D. A. (2004) Ectopic expression of a cecropin transgene in the human malaria vector mosquito *Anopheles gambiae* (Diptera: Culicidae): effects on susceptibility to *Plasmodium*. *Journal of Medical Entomology.* 41 (3), 447–455.
Kim, Y., Kim, T., Yoon, J., Cho, H., Lee, W., Kim, J., Song, Y., Kim, S. N., Yoon, J. H. & Kim-Ha, J. (2005) Downregulation of lipopolysaccharide response in *Drosophila* by negative crosstalk between the AP1 and NF-κB signaling modules. *Nature Immunology.* 6 (2), 211–218.
King, J. G. & Hillyer, J. F. (2013) Spatial and temporal *in vivo* analysis of circulating and sessile immune cells in mosquitoes: hemocyte mitosis following infection. *BMC Biology.* 11 (1), 55.
Kiu, H. & Nicholson, S. E. (2012) Biology and significance of the JAK/STAT signalling pathways. *Growth Factors.* 30 (2), 88–106.
Kleino, A., Myllymäki, H., Kallio, J., Vanha-aho, L., Oksanen, K., Ulvila, J., Hultmark, D., Valanne, S. & Rämet, M. (2008) Pirk is a negative regulator of the *Drosophila* Imd pathway. *The Journal of Immunology.* 180 (8), 5413–5422.
Kleino, A. & Silverman, N. (2014) The *Drosophila* IMD pathway in the activation of the humoral immune response. *Developmental & Comparative Immunology.* 42 (1), 25–35.
Konstandi, O. A., Papassideri, I. S., Stravopodis, D. J., Kenoutis, C. A., Hasan, Z., Katsorchis, T., Wever, R. & Margaritis, L. H. (2005) The enzymatic component of *Drosophila melanogaster* chorion is the Pxd peroxidase. *Insect Biochemistry and Molecular Biology.* 35 (9), 1043–1057.
Kosman, D., Mizutani, C. M., Lemons, D., Cox, W. G., McGinnis, W. & Bier, E. (2004) Multiplex detection of RNA expression in *Drosophila* embryos. *Science.* 305 (5685), 846.
Kriventseva, E. V., Rahman, N., Espinosa, O. & Zdobnov, E. M. (2007) OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Research.* 36 (Database issue), D271–D275.
Kulkarni, A., Pandey, A., Trainor, P., Carlisle, S., Chhilar, J. S., Yu, W., Moon, A. & Xu, J. (2021) Trained immunity in *Anopheles gambiae*: Antibacterial immunity is enhanced by priming via sugar meal supplemented with a single gut symbiotic bacterial strain. *Frontiers in Microbiology.* 12, 649213.
Kukutla, P., Lindberg, B. G., Pei, D., Rayl, M., Yu, W., Steritz, M., Faye, I. & Xu, J. (2014) Insights from the genome annotation of *Elizabethkingia anophelis* from the malaria vector *Anopheles gambiae*. *PLoS ONE.* 9 (5), e97715.

Kumar, S. & Barillas-Mury, C. (2005) Ookinete-induced midgut peroxidases detonate the time bomb in anopheline mosquitoes. *Insect Biochemistry and Molecular Biology.* 35 (7), 721–727.
Kumar, S., Molina-Cruz, A., Gupta, L., Rodrigues, J. & Barillas-Mury, C. (2010) A peroxidase/dual oxidase system modulates midgut epithelial immunity in *Anopheles gambiae*. *Science.* 327 (5973), 1644–1648.
Kumar, S., Christophides, G. K., Cantera, R., Charles, B., Han, Y. S., Meister, S., Dimopoulos, G., Kafatos, F. C. & Barillas-Mury, C. (2003) The role of reactive oxygen species on *Plasmodium* melanotic encapsulation in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences.* 100 (24), 14139.
Kumar, S., Gupta, L., Han, Y. S. & Barillas-Mury, C. (2004) Inducible peroxidases mediate nitration of *Anopheles* midgut cells undergoing apoptosis in response to *Plasmodium* invasion. *The Journal of Biological Chemistry.* 279 (51), 53475–53482.
Ladram, A. & Nicolas, P. (2016) Antimicrobial peptides from frog skin: biodiversity and therapeutic promises. *Frontiers in Bioscience.* 21, 1341–1371.
Lazzaro, B. P., Flores, H. A., Lorigan, J. G. & Yourth, C. P. (2008) Genotype-by-environment interactions and adaptation to local temperature affect immunity and fecundity in *Drosophila melanogaster*. *PLoS Pathogens.* 4 (3), e1000025.
Lee, H., Lee, C., Yang, J., Lai, J. Z. & Chang, K. Y. (2015) A large-scale structural classification of antimicrobial peptides. *BioMed Research International.* 2015.
Lemaitre, B., Reichhart, J. & Hoffmann, J. A. (1997) *Drosophila* host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. *Proceedings of the National Academy of Sciences.* 94 (26), 14614–14619.
Leone, P., Bischoff, V., Kellenberger, C., Hetru, C., Royet, J. & Roussel, A. (2008) Crystal structure of *Drosophila* PGRP-SD suggests binding to DAP-type but not lysine-type peptidoglycan. *Molecular Immunology.* 45 (9), 2521–2530.
Leulier, F., Rodriguez, A., Khush, R. S., Abrams, J. M. & Lemaitre, B. (2000) The *Drosophila* caspase Dredd is required to resist Gram-negative bacterial infection. *EMBO Reports.* 1 (4), 353–358.
Leulier, F., Vidal, S., Saigo, K., Ueda, R. & Lemaitre, B. (2002) Inducible expression of double-stranded RNA reveals a role for dFADD in the regulation of the antibacterial response in *Drosophila* adults. *Current Biology.* 12 (12), 996–1000.
Levashina, E. A., Moita, L. F., Blandin, S., Vriend, G., Lagueux, M. & Kafatos, F. C. (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell.* 104 (5), 709–718.
Li, F., Templeton, T. J., Popov, V., Comer, J. E., Tsuboi, T., Torii, M. & Vinetz, J. M. (2004) *Plasmodium* ookinete-secreted proteins secreted through a common micronemal pathway are targets of blocking malaria transmission. *The Journal of Biological Chemistry.* 279 (25), 26635–26644.
Li, J., Wang, X., Zhang, G., Githure, J. I., Yan, G. & James, A. A. (2013) Genome-block expression-assisted association studies discover malaria resistance genes in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences.* 110 (51), 20675–20680.
Li, M., Mead, E. A. & Zhu, J. (2011) Heterodimer of two bHLH-PAS proteins mediates juvenile hormone-induced gene expression. *Proceedings of the National Academy of Sciences.* 108 (2), 638–643.
Liao, Y., Smyth, G. K. & Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 30 (7), 923–930.
Ligoxygakis, P., Pelte, N., Hoffmann, J. A. & Reichhart, J. (2002) Activation of *Drosophila* Toll during fungal infection by a blood serine protease. *Science.* 297 (5578), 114–116.
Lim, J., Kim, M., Kim, H., Yano, T., Oshima, Y., Aggarwal, K., Goldman, W. E., Silverman, N., Kurata, S. & Oh, B. (2006) Structural basis for preferential recognition of diaminopimelic acid-type peptidoglycan by a subset of peptidoglycan recognition proteins. *Journal of Biological Chemistry.* 281 (12), 8286–8295.
Lim, J., Gowda, D. C., Krishnegowda, G. & Luckhart, S. (2005) Induction of nitric oxide synthase in *Anopheles stephensi* by *Plasmodium falciparum*: mechanism of signaling and the role of parasite glycosylphosphatidylinositols. *Infection and Immunity.* 73 (5), 2778–2789.
Lindsay, S. A. & Wasserman, S. A. (2014) Conventional and non-conventional *Drosophila* Toll signaling. *Developmental & Comparative Immunology.* 42 (1), 16–24.
Love, M. I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology.* 15 (12), 1–21.
Lu, Y., Su, F., Li, Q., Zhang, J., Li, Y., Tang, T., Hu, Q. & Yu, X. (2020) Pattern recognition receptors in *Drosophila* immune responses. *Developmental & Comparative Immunology.* 102, 103468.
Luckhart, S., Vodovotz, Y., Cui, L. W. & Rosenberg, R. (1998) The mosquito *Anopheles stephensi* limits malaria parasite development with inducible synthesis of nitric oxide. *Proceedings of the National Academy of Sciences.* 95 (10), 5700–5705.
Luckhart, S., Crampton, A. L., Zamora, R., Lieber, M. J., Dos Santos, P. C., Peterson, T. M., Emmith, N., Lim, J., Wink, D. A. & Vodovotz, Y. (2003) Mammalian transforming growth factor β1 activated after ingestion by *Anopheles stephensi* modulates mosquito immunity. *Infection and Immunity.* 71 (6), 3000–3009.

Ludtke, S. J., He, K., Heller, W. T., Harroun, T. A., Yang, L. & Huang, H. W. (1996) Membrane pores induced by magainin. *Biochemistry.* 35 (43), 13723–13728.
Luna, C., Hoa, N. T., Lin, H., Zhang, L., Nguyen, H., Kanzok, S. M. & Zheng, L. (2006) Expression of immune responsive genes in cell lines from two different anopheline species. *Insect Molecular Biology.* 15 (6), 721–729.
Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2021) *cluster: cluster analysis basics and extensions. R package version 2.1.2.* https://cran.r-project.org/package=cluster.
Maillet, F., Bischoff, V., Vignal, C., Hoffmann, J. & Royet, J. (2008) The *Drosophila* peptidoglycan recognition protein PGRP-LF blocks PGRP-LC and IMD/JNK pathway activation. *Cell Host & Microbe.* 3 (5), 293–303.
Manda, H., Gouagna, L., Nyandat, E., Kabiru, E. W., Jackson, R. R., Foster, W. A., Githure, J. I., Beier, J. C. & Hassanali, A. (2007) Discriminative feeding behaviour of *Anopheles gambiae s.s.* on endemic plants in western Kenya. *Medical and Veterinary Entomology.* 21 (1), 103–111.
Marinotti, O., Calvo, E., Nguyen, Q. K., Dissanayake, S., Ribeiro, J. & James, A. A. (2006) Genome-wide analysis of gene expression in adult *Anopheles gambiae*. *Insect Molecular Biology.* 15 (1), 1–12.
Martello, E., Yogeswaran, G., Reithinger, R. & Leonardi-Bee, J. (2022) Mosquito aquatic habitat modification and manipulation interventions to control malaria. *Cochrane Database of Systematic Reviews.* 2022 (11).
Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 17 (1), 10–12.
Martins, N. E., Faria, V. G., Teixeira, L., Magalhães, S. & Sucena, É. (2013) Host adaptation is contingent upon the infection route taken by pathogens. *PLoS Pathogens.* 9 (9), e1003601.
Mathias, D. K., Jardim, J. G., Parish, L. A., Armistead, J. S., Trinh, H. V., Kumpitak, C., Sattabongkot, J. & Dinglasan, R. R. (2014) Differential roles of an anopheline midgut GPI-anchored protein in mediating *Plasmodium falciparum* and *Plasmodium vivax* ookinete invasion. *Infection, Genetics and Evolution.* 28, 635–647.
Matsuzaki, K., Murase, O., Fujii, N. & Miyajima, K. (1996) An antimicrobial peptide, magainin 2, induced rapid flip-flop of phospholipids coupled with pore formation and peptide translocation. *Biochemistry.* 35 (35), 11361–11368.
McFarlane, M., Arias-Goeta, C., Martin, E., O'Hara, Z., Lulla, A., Mousson, L., Rainey, S. M., Misbah, S., Schnettler, E., Donald, C. L., Merits, A., Kohl, A. & Failloux, A. (2014) Characterization of *Aedes aegypti* innate-immune pathways that limit chikungunya virus replication. *PLoS Neglected Tropical Diseases.* 8 (7), e2994.
McNamara, K. B., van Lieshout, E. & Simmons, L. W. (2014) Females suffer a reduction in the viability of stored sperm following an immune challenge. *Journal of Evolutionary Biology.* 27 (1), 133–140.
McNeill, H., Yang, C., Brodsky, M., Ungos, J. & Simon, M. A. (1997) *mirror* encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the *Drosophila* eye. *Genes & Development.* 11 (8), 1073–1082.
Medzhitov, R. & Janeway Jr., C. A. (1997) Innate immunity: the virtues of a nonclonal system of recognition. *Cell.* 91 (3), 295–298.
Meinander, A., Runchel, C., Tenev, T., Chen, L., Kim, C., Ribeiro, P. S., Broemer, M., Leulier, F., Zvelebil, M. & Silverman, N. (2012) Ubiquitylation of the initiator caspase DREDD is required for innate immune signalling. *The EMBO Journal.* 31 (12), 2770–2783.
Meister, S., Kanzok, S. M., Zheng, X. L., Luna, C., Li, T. R., Hoa, N. T., Clayton, J. R., White, K. P., Kafatos, F. C., Christophides, G. K. & Zheng, L. (2005) Immune signaling pathways regulating bacterial and malaria parasite infection of the mosquito *Anopheles gambiae*. *Proceedings of the National Academy of Sciences.* 102 (32), 11420–11425.
Meister, S., Agianian, B., Turlure, F., Relógio, A., Morlais, I., Kafatos, F. C. & Christophides, G. K. (2009) *Anopheles gambiae* PGRPLC-mediated defense against bacteria modulates infections with malaria parasites. *PLoS Pathogens.* 5 (8), e1000542.
Mellroth, P., Karlsson, J., Håkansson, J., Schultz, N., Goldman, W. E. & Steiner, H. (2005) Ligand-induced dimerization of *Drosophila* peptidoglycan recognition proteins in vitro. *Proceedings of the National Academy of Sciences.* 102 (18), 6455–6460.
Mendes, A. M., Awono-Ambene, P. H., Nsango, S. E., Cohuet, A., Fontenille, D., Kafatos, F. C., Christophides, G. K., Morlais, I. & Vlachou, D. (2011) Infection intensity-dependent responses of *Anopheles gambiae* to the African malaria parasite *Plasmodium falciparum*. *Infection and Immunity.* 79 (11), 4708–4715.
Mendes, A. M., Schlegelmilch, T., Cohuet, A., Awono-Ambene, P., De Iorio, M., Fontenille, D., Morlais, I., Christophides, G. K., Kafatos, F. C. & Vlachou, D. (2008) Conserved mosquito/parasite interactions affect development of *Plasmodium falciparum* in Africa. *PLoS Pathogens.* 4 (5), e1000069.
Meunier, N., Belgacem, Y. H. & Martin, J. (2007) Regulation of feeding behaviour and locomotor activity by *takeout* in *Drosophila*. *Journal of Experimental Biology.* 210 (8), 1424–1434.
Michel, K., Budd, A., Pinto, S., Gibson, T. J. & Kafatos, F. C. (2005) *Anopheles gambiae* SRPN2 facilitates midgut invasion by the malaria parasite *Plasmodium berghei*. *EMBO Reports.* 6 (9), 891–897.
Michel, K., Suwanchaichinda, C., Morlais, I., Lambrechts, L., Cohuet, A., Awono-Ambene, P. H., Simard, F., Fontenille, D., Kanost, M. R. & Kafatos, F. C. (2006) Increased melanizing activity in *Anopheles gambiae* does not affect development of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences.* 103 (45), 16858–16863.

Mishra, A. K., Sharma, V., Mutsuddi, M. & Mukherjee, A. (2021) Signaling cross-talk during development: Context-specific networking of Notch, NF-κB and JNK signaling pathways in *Drosophila*. *Cellular Signalling.* 82, 109937.
Mitraka, E., Stathopoulos, S., Siden-Kiamos, I., Christophides, G. K. & Louis, C. (2013) *Asaia* accelerates larval development of *Anopheles gambiae*. *Pathogens and Global Health.* 107 (6), 305–311.
Mitri, C., Jacques, J., Thiery, I., Riehle, M. M., Xu, J., Bischoff, E., Morlais, I., Nsango, S. E., Vernick, K. D. & Bourgouin, C. (2009) Fine pathogen discrimination within the APL1 gene family protects *Anopheles gambiae* against human and rodent malaria species. *PLoS Pathogens.* 5 (9), e1000576.
Moita, L. F., Wang-Sattler, R., Michel, K., Zimmermann, T., Blandin, S. A., Levashina, E. A. & Kafatos, F. C. (2005) *In vivo* identification of novel regulators and conserved pathways of phagocytosis in *A. gambiae*. *Immunity.* 23 (1), 65–73.
Molina-Cruz, A., DeJong, R. J., Ortega, C., Haile, A., Abban, E., Rodrigues, J., Jaramillo-Gutierrez, G. & Barillas-Mury, C. (2012) Some strains of *Plasmodium falciparum*, a human malaria parasite, evade the complement-like system of *Anopheles gambiae* mosquitoes. *Proceedings of the National Academy of Sciences.* 109 (28), 1957.
Molina-Cruz, A., Canepa, G. E., Kamath, N., Pavlovic, N. V., Mu, J., Ramphul, U. N., Ramirez, J. L. & Barillas-Mury, C. (2015) *Plasmodium* evasion of mosquito immunity and global malaria transmission: The lock-and-key theory. *Proceedings of the National Academy of Sciences.* 112 (49), 15178–15183.
Molina-Cruz, A., Canepa, G. E., Alves e Silva, T. L., Williams, A. E., Nagyal, S., Yenkoidiok-Douti, L., Nagata, B. M., Calvo, E., Andersen, J. & Boulanger, M. J. (2020) *Plasmodium falciparum* evades immunity of anopheline mosquitoes by interacting with a Pfs47 midgut receptor. *Proceedings of the National Academy of Sciences.* 117 (5), 2597–2605.
Molina-Cruz, A., Canepa, G. E. & Barillas-Mury, C. (2017) *Plasmodium P47*: a key gene for malaria transmission by mosquito vectors. *Current Opinion in Microbiology.* 40, 168–174.
Molina-Cruz, A., DeJong, R. J., Charles, B., Gupta, L., Kumar, S., Jaramillo-Gutierrez, G. & Barillas-Mury, C. (2008) Reactive oxygen species modulate *Anopheles gambiae* immunity against bacteria and *Plasmodium*. *The Journal of Biological Chemistry.* 283 (6), 3217–3223.
Molina-Cruz, A., Garver, L. S., Alabaster, A., Bangiolo, L., Haile, A., Winikor, J., Ortega, C., van Schaijk, B. C., Sauerwein, R. W. & Taylor-Salmon, E. (2013) The human malaria parasite *Pfs47* gene mediates evasion of the mosquito immune system. *Science.* 340 (6135), 984–987.
Moll, R. M., Romoser, W. S., Modrakowski, M. C., Moncayo, A. C. & Lerdthusnee, K. (2001) Meconial peritrophic membranes and the fate of midgut bacteria during mosquito (Diptera: Culicidae) metamorphosis. *Journal of Medical Entomology.* 38 (1), 29–32.
Mongelli, V. & Saleh, M. (2016) Bugs are not to be silenced: small RNA pathways and antiviral responses in insects. *Annual Review of Virology.* 3 (1), 573–589.
Morgenstern, R., Zhang, J. & Johansson, K. (2011) Microsomal glutathione transferase 1: mechanism and functional roles. *Drug Metabolism Reviews.* 43 (2), 300–306.
Morlais, I., Ponçon, N., Simard, F., Cohuet, A. & Fontenille, D. (2004) Intraspecific nucleotide variation in *Anopheles gambiae*: new insights into the biology of malaria vectors. *The American Journal of Tropical Medicine and Hygiene.* 71 (6), 795–802.
Mosquito Taxonomic Inventory. (2008) *Anopheles.* https://mosquito-taxonomic-inventory.myspecies.info/simpletaxonomy/term/6047 [Accessed Mar 16, 2023].
Mulakkal, N. C., Nagy, P., Takats, S., Tusco, R., Juhász, G. & Nezis, I. P. (2014) Autophagy in *Drosophila*: from historical studies to current knowledge. *BioMed Research International.* 2014.
Müller, H. M., Crampton, J. M., Della Torre, A., Sinden, R. & Crisanti, A. (1993) Members of a trypsin gene family in *Anopheles gambiae* are induced in the gut by blood meal. *The EMBO Journal.* 12 (7), 2891–2900.
Müller, H., Catteruccia, F., Vizioli, J., della Torre, A. & Crisanti, A. (1995) Constitutive and blood meal-induced trypsin genes in *Anopheles gambiae*. *Experimental Parasitology*. 81 (3), 371–385.
Myllymäki, H. & Rämet, M. (2014) JAK/STAT pathway in *Drosophila* immunity. *Scandinavian Journal of Immunology.* 79 (6), 377–385.
Naitza, S., Rossé, C., Kappler, C., Georgel, P., Belvin, M., Gubb, D., Camonis, J., Hoffmann, J. A. & Reichhart, J. (2002) The *Drosophila* immune defense against Gram-negative infection requires the death protein dFADD. *Immunity.* 17 (5), 575–581.
Nakagawa, Y., Fujiwara-Fukuta, S., Yorimitsu, T., Tanaka, S., Minami, R., Shimooka, L. & Nakagoshi, H. (2011) Spatial and temporal requirement of defective proventriculus activity during *Drosophila* midgut development. *Mechanisms of Development.* 128 (5–6), 258–267.
Nakagoshi, H., Hoshi, M., Nabeshima, Y. & Matsuzaki, F. (1998) A novel homeobox gene mediates the Dpp signal to establish functional specificity within target cells. *Genes & Development.* 12 (17), 2724–2734.
Nakagoshi, H. (2005) Functional specification in the *Drosophila* endoderm. *Development, Growth & Differentiation.* 47 (6), 383–392.

Nanfack-Minkeu, F., Mitri, C., Bischoff, E., Belda, E., Casademont, I. & Vernick, K. D. (2019) Interaction of RNA viruses of the natural virome with the African malaria vector, *Anopheles coluzzii*. *Scientific Reports.* 9 (1), 6319.
Nappi, A. J. & Christensen, B. M. (2005) Melanogenesis and associated cytotoxic reactions: applications to insect innate immunity. *Insect Biochemistry and Molecular Biology.* 35 (5), 443–459.
Nash, A., Urdaneta, G. M., Beaghton, A. K., Hoermann, A., Papathanos, P. A., Christophides, G. K. & Windbichler, N. (2019) Integral gene drives for population replacement. *Biology Open.* 8 (1).
Neira Oviedo, M., Vanekeris, L., Corena-Mcleod, M. & Linser, P. J. (2008) A microarray-based analysis of transcriptional compartmentalization in the alimentary canal of *Anopheles gambiae* (Diptera: Culicidae) larvae. *Insect Molecular Biology.* 17 (1), 61–72.
Nouzova, M., Edwards, M. J., Michalkova, V., Ramirez, C. E., Ruiz, M., Areiza, M., DeGennaro, M., Fernandez-Lima, F., Feyereisen, R. & Jindra, M. (2021) Epoxidation of juvenile hormone was a key innovation improving insect reproductive fitness. *Proceedings of the National Academy of Sciences.* 118 (45), e2109381118.
Nsango, S. E., Abate, L., Thoma, M., Pompon, J., Fraiture, M., Rademacher, A., Berry, A., Awono-Ambene, P. H., Levashina, E. A. & Morlais, I. (2012) Genetic clonality of *Plasmodium falciparum* affects the outcome of infection in *Anopheles gambiae*. *International Journal for Parasitology.* 42 (6), 589–595.
O'Loughlin, S. (2020) The expanding *Anopheles gambiae* species complex. *Pathogens and Global Health.* 114 (1), 1.
Obbard, D. J., Linton, Y. M., Jiggins, F. M., Yan, G. & Little, T. J. (2007) Population genetics of *Plasmodium* resistance genes in *Anopheles gambiae*: no evidence for strong selection. *Molecular Ecology.* 16 (16), 3497–3510.
Ogienko, A. A., Omelina, E. S., Bylino, O. V., Batin, M. A., Georgiev, P. G. & Pindyurin, A. V. (2022) *Drosophila* as a model organism to study basic mechanisms of longevity. *International Journal of Molecular Sciences.* 23 (19), 11244.
Oliveira, J. H. M., Gonçalves, R. L., Oliveira, G. A., Oliveira, P. L., Oliveira, M. F. & Barillas-Mury, C. (2011) Energy metabolism affects susceptibility of *Anopheles gambiae* mosquitoes to *Plasmodium* infection. *Insect Biochemistry and Molecular Biology.* 41 (6), 349–355.
Oliveros, J. C. (2015) *Venny 2.1.0.* https://bioinfogp.cnb.csic.es/tools/venny/ [Accessed Jan 23, 2022].
Oren, Z. & Shai, Y. (1998) Mode of action of linear amphipathic α-helical antimicrobial peptides. *Peptide Science.* 47 (6), 451–463.
Orme, M. & Meier, P. (2009) Inhibitor of apoptosis proteins in *Drosophila*: gatekeepers of death. *Apoptosis.* 14, 950–960.
Osei-Poku, J., Mbogo, C. M., Palmer, W. J. & Jiggins, F. M. (2012) Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. *Molecular Ecology.* 21 (20), 5138–5150.
Osta, M. A., Christophides, G. K. & Kafatos, F. C. (2004) Effects of mosquito genes on *Plasmodium* development. *Science.* 303 (5666), 2030–2032.
Pace, K. E., Lebestky, T., Hummel, T., Arnoux, P., Kwan, K. & Baum, L. G. (2002) Characterization of a novel *Drosophila melanogaster* galectin: expression in developing immune, neural, and muscle tissues. *Journal of Biological Chemistry.* 277 (15), 13091–13098.
Pang, X., Xiao, X., Liu, Y., Zhang, R., Liu, J., Liu, Q., Wang, P. & Cheng, G. (2016) Mosquito C-type lectins maintain gut microbiome homeostasis. *Nature Microbiology.* 1 (5), 1–11.
Papa, F., Windbichler, N., Waterhouse, R. M., Cagnetti, A., D'Amato, R., Persampieri, T., Lawniczak, M. K., Nolan, T. & Papathanos, P. A. (2017) Rapid evolution of female-biased genes among four species of *Anopheles* malaria mosquitoes. *Genome Research.* 27 (9), 1536–1548.
Park, J. M., Brady, H., Ruocco, M. G., Sun, H., Williams, D., Lee, S. J., Kato, J., Tomohisa, Richards, N., Chan, K., Mercurio, F., Karin, M. & Wasserman, S. A. (2004) Targeting of TAK1 by the NF-κB protein Relish regulates the JNK-mediated immune response in *Drosophila*. *Genes & Development.* 18 (5), 584–594.
Pastor-Pareja, J. C., Wu, M. & Xu, T. (2008) An innate immune response of blood cells to tumors and tissue damage in *Drosophila*. *Disease Models & Mechanisms.* 1 (2–3), 144–154.
Peixoto, R. S., Harkins, D. M. & Nelson, K. E. (2021) Advances in microbiome research for animal health. *Annual Review of Animal Biosciences.* 9, 289–311.
Pertea, G. & Pertea, M. (2020) GFF utilities: GffRead and GffCompare. *F1000Research.* 9, 304.
Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T., Mendell, J. T. & Salzberg, S. L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology.* 33 (3), 290–295.
Peterson, R. K. & Rolston, M. G. (2022) Larval mosquito management and risk to aquatic ecosystems: A comparative approach including current tactics and gene-drive *Anopheles* techniques. *Transgenic Research.* 31 (4–5), 489–504.
Peterson, T. M., Gow, A. J. & Luckhart, S. (2007) Nitric oxide metabolites induced in *Anopheles stephensi* control malaria parasite infection. *Free Radical Biology and Medicine.* 42 (1), 132–142.
Pfaffl, M. W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research.* 29 (9), e45.
Pike, A., Dong, Y., Dizaji, N. B., Gacita, A., Mongodin, E. F. & Dimopoulos, G. (2017) Changes in the microbiota cause genetically modified *Anopheles* to spread in a population. *Science.* 357 (6358), 1396–1399.

Pinch, M., Mitra, S., Rodriguez, S. D., Li, Y., Kandel, Y., Dungan, B., Holguin, F. O., Attardo, G. M. & Hansen, I. A. (2021) Fat and happy: profiling mosquito fat body lipid storage and composition post-blood meal. *Frontiers in Insect Science.* 1, 693168.
Pinto, S. B., Lombardo, F., Koutsos, A. C., Waterhouse, R. M., McKay, K., An, C., Ramakrishnan, C., Kafatos, F. C. & Michel, K. (2009) Discovery of *Plasmodium* modulators by genome-wide analysis of circulating hemocytes in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences.* 106 (50), 21270–21275.
Povelones, M., Waterhouse, R. M., Kafatos, F. C. & Christophides, G. K. (2009) Leucine-rich repeat protein complex activates mosquito complement in defense against *Plasmodium* parasites. *Science.* 324 (5924), 258–261.
Povelones, M., Bhagavatula, L., Yassine, H., Tan, L., Upton, L., Osta, M. & Christophides, G. (2013) The Clip-domain serine protease homolog SPCLIP1 regulates complement recruitment to microbial surfaces in the malaria mosquito *Anopheles gambiae*. *PLoS Pathogens.* 9 (9), e1003623.
Powers, J. C., Turangan, R., Joosse, B. A. & Hillyer, J. F. (2020) Adult mosquitoes infected with bacteria early in life have stronger antimicrobial responses and more hemocytes after reinfection later in life. *Insects.* 11 (6), 331.
Pridgeon, J. W. & Liu, N. (2003) Overexpression of the cytochrome *c* oxidase subunit I gene associated with a pyrethroid resistant strain of German cockroaches, *Blattella germanica* (L.). *Insect Biochemistry and Molecular Biology.* 33 (10), 1043–1048.
Puglise, J. M., Estep, A. S. & Becnel, J. J. (2016) Expression profiles and RNAi silencing of inhibitor of apoptosis transcripts in *Aedes, Anopheles,* and *Culex* mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology.* 53 (2), 304–314.
Pumpuni, C. B., Beier, M. S., Nataro, J. P., Guers, L. D. & Davis, J. R. (1993) *Plasmodium falciparum*: inhibition of sporogonic development in *Anopheles stephensi* by Gram-negative bacteria. *Experimental Parasitology.* 77 (2), 195–199.
Pursall, E. R. & Rolff, J. (2011) Immune responses accelerate ageing: proof-of-principle in an insect model. *PLoS ONE.* 6 (5), e19972.
Ramirez, J. L., Short, S. M., Bahia, A. C., Saraiva, R. G., Dong, Y., Kang, S., Tripathi, A., Mlambo, G. & Dimopoulos, G. (2014) *Chromobacterium Csp_P* reduces malaria and dengue infection in vector mosquitoes and has entomopathogenic and *in vitro* anti-pathogen activities. *PLoS Pathogens.* 10 (10), e1004398.
Ramos, S., Custódio, A. & Silveira, H. (2014) *Anopheles gambiae* eicosanoids modulate *Plasmodium berghei* survival from oocyst to salivary gland invasion. *Memórias do Instituto Oswaldo Cruz.* 109, 668–671.
Ramphul, U. N., Garver, L. S., Molina-Cruz, A., Canepa, G. E. & Barillas-Mury, C. (2015) *Plasmodium falciparum* evades mosquito immunity by disrupting JNK-mediated apoptosis of invaded midgut cells. *Proceedings of the National Academy of Sciences.* 112 (5), 1273–1280.
Ranson, H. & Hemingway, J. (2005) Mosquito glutathione transferases. *Methods in Enzymology.* 401, 226–241.
Ranson, H. & Lissenden, N. (2016) Insecticide resistance in African *Anopheles* mosquitoes: a worsening situation that needs urgent action to maintain malaria control. *Trends in Parasitology.* 32 (3), 187–196.
Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H. & Vilo, J. (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research.* 47 (W1), W191–W198.
Reiff, T., Jacobson, J., Cognigni, P., Antonello, Z., Ballesta, E., Tan, K. J., Yew, J. Y., Dominguez, M. & Miguel-Aliaga, I. (2015) Endocrine remodelling of the adult intestine sustains reproduction in *Drosophila*. *eLife.* 4, e06930.
Reinert, J. F. (1975) Mosquito generic and subgeneric abbreviations (Diptera: Culicidae). *Mosquito Systematics.* 7 (2), 105–110.
Rhodes, V. L., Thomas, M. B. & Michel, K. (2018) The interplay between dose and immune system activation determines fungal infection outcome in the African malaria mosquito, *Anopheles gambiae*. *Developmental & Comparative Immunology.* 85, 125–133.
Ribeiro, J. M. C. (2003) A catalogue of *Anopheles gambiae* transcripts significantly more or less expressed following a blood meal. *Insect Biochemistry and Molecular Biology.* 33 (9), 865–882.
Riehle, M. A., Moreira, C. K., Lampe, D., Lauzon, C. & Jacobs-Lorena, M. (2007) Using bacteria to express and display anti-*Plasmodium* molecules in the mosquito midgut. *International Journal for Parasitology.* 37 (6), 595–603.
Riehle, M. M., Xu, J., Lazzaro, B. P., Rottschaefer, S. M., Coulibaly, B., Sacko, M., Niare, O., Morlais, I., Traore, S. F. & Vernick, K. D. (2008) *Anopheles gambiae* APL1 is a family of variable LRR proteins required for Rel1-mediated protection from the malaria parasite, *Plasmodium berghei*. *PLoS ONE.* 3 (11), e3672.
Ríos-Barrera, L. D. & Riesgo-Escovar, J. R. (2013) Regulating cell morphogenesis: The *Drosophila* Jun N-terminal kinase pathway. *Genesis.* 51 (3), 147–162.
Roberts, A., De Andrade, P. P., Okumu, F., Quemada, H., Savadogo, M., Singh, J. A. & James, S. (2017) Results from the workshop “problem formulation for the use of gene drive in mosquitoes”. *The American Journal of Tropical Medicine and Hygiene.* 96 (3), 530–533.
Rodgers, F. H., Gendrin, M., Wyer, C. A. & Christophides, G. K. (2017) Microbiota-induced peritrophic matrix regulates midgut homeostasis and prevents systemic infection of malaria vector mosquitoes. *PLoS Pathogens.* 13 (5), e1006391.

Rodgers, F. H., Cai, J. A., Pitaluga, A. N., Mengin-Lecreulx, D., Gendrin, M. & Christophides, G. K. (2020) Functional analysis of the three major PGRPLC isoforms in the midgut of the malaria mosquito *Anopheles coluzzii*. *Insect Biochemistry and Molecular Biology.* 118, 103288.
Rodrigues, J., Brayner, F. A., Alves, L. C., Dixit, R. & Barillas-Mury, C. (2010) Hemocyte differentiation mediates innate immune memory in *Anopheles gambiae* mosquitoes. *Science.* 329 (5997), 1353–1355.
Rodriguez-Andres, J., Rani, S., Varjak, M., Chase-Topping, M. E., Beck, M. H., Ferguson, M. C., Schnettler, E., Fragkoudis, R., Barry, G., Merits, A., Fazakerley, J. K., Strand, M. R. & Kohl, A. (2012) Phenoloxidase activity acts as a mosquito innate immune response against infection with Semliki Forest virus. *PLoS Pathogens.* 8 (11), e1002977.
Rogers, D. W., Whitten, M. M., Thailayil, J., Soichot, J., Levashina, E. A. & Catteruccia, F. (2008) Molecular and cellular components of the mating machinery in *Anopheles gambiae* females. *Proceedings of the National Academy of Sciences.* 105 (49), 19390–19395.
Romoli, O., Schönbeck, J. C., Hapfelmeier, S. & Gendrin, M. (2021) Production of germ-free mosquitoes via transient colonisation allows stage-specific investigation of host–microbiota interactions. *Nature Communications.* 12 (1), 942.
Römisch, K. (1999) Surfing the Sec61 channel: bidirectional protein translocation across the ER membrane. *Journal of Cell Science.* 112 (23), 4185–4191.
Rottschaefer, S. M., Riehle, M. M., Coulibaly, B., Sacko, M., Niare, O., Morlais, I., Traore, S. F., Vernick, K. D. & Lazzaro, B. P. (2011) Exceptional diversity, maintenance of polymorphism, and recent directional selection on the *APL1* malaria resistance genes of *Anopheles gambiae*. *PLoS Biology.* 9 (3), e1000600.
RTS,S Clinical Trials Partnership. (2015) Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *The Lancet.* 386 (9988), 31–45.
Rund, S. S., Bonar, N. A., Champion, M. M., Ghazi, J. P., Houk, C. M., Leming, M. T., Syed, Z. & Duffield, G. E. (2013) Daily rhythms in antennal protein and olfactory sensitivity in the malaria mosquito *Anopheles gambiae*. *Scientific Reports.* 3 (1), 1–9.
Rutschmann, S., Jung, A. C., Zhou, R., Silverman, N., Hoffmann, J. A. & Ferrandon, D. (2000) Role of *Drosophila* IKKγ in a Toll-independent antibacterial immune response. *Nature Immunology.* 1 (4), 342–347.
Ryu, J., Kim, S., Lee, H., Bai, J. Y., Nam, Y., Bae, J., Lee, D. G., Shin, S. C., Ha, E. & Lee, W. (2008) Innate immune homeostasis by the homeobox gene *caudal* and commensal-gut mutualism in *Drosophila*. *Science.* 319 (5864), 777–782.
Sarov-Blat, L., So, W. V., Liu, L. & Rosbash, M. (2000) The *Drosophila takeout* gene is a novel molecular link between circadian rhythms and feeding behavior. *Cell.* 101 (6), 647–656.
Schindler, O., Berka, K., Cantara, A., Křenek, A., Tichý, D., Raček, T. & Svobodová, R. (2023) αCharges: partial atomic charges for AlphaFold structures in high quality. *Nucleic Acids Research.* 51 (W1), W11–W16.
Schmid-Hempel, P. (2003) Variation in immune defence as a question of evolutionary ecology. *Proceedings of the Royal Society of London. Series B: Biological Sciences.* 270 (1513), 357–366.
Schmid-Hempel, P. (2005) Evolutionary ecology of insect immune defenses. *Annual Review of Entomology.* 50, 529–551.
Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods.* 9 (7), 671–675.
Schnitger, A. K., Yassine, H., Kafatos, F. C. & Osta, M. A. (2009) Two C-type lectins cooperate to defend *Anopheles gambiae* against Gram-negative bacteria. *Journal of Biological Chemistry.* 284 (26), 17616–17624.
Schramek, N., Wang, H., Römisch-Margl, W., Keil, B., Radykewicz, T., Winzenhörlein, B., Beerhues, L., Bacher, A., Rohdich, F. & Gershenzon, J. (2010) Artemisinin biosynthesis in growing plants of *Artemisia annua*. A $^{13}CO_2$ study. *Phytochemistry.* 71 (2–3), 179–187.
Schwartz, A. & Koella, J. C. (2002) Melanization of *Plasmodium falciparum* and C-25 Sephadex beads by field-caught *Anopheles gambiae* (Diptera: Culicidae) from Southern Tanzania. *Journal of Medical Entomology.* 39 (1), 84–88.
Schwenke, R. A., Lazzaro, B. P. & Wolfner, M. F. (2016) Reproduction–immunity trade-offs in insects. *Annual Review of Entomology.* 61, 239–256.
Shahabuddin, M., Toyoshima, T., Aikawa, M. & Kaslow, D. C. (1993) Transmission-blocking activity of a chitinase inhibitor and activation of malarial parasite chitinase by mosquito protease. *Proceedings of the National Academy of Sciences.* 90 (9), 4266–4270.
Sheldon, B. C. & Verhulst, S. (1996) Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology. *Trends in Ecology & Evolution.* 11 (8), 317–321.
Silke, J. & Vaux, D. L. (2001) Two kinds of BIR-containing protein-inhibitors of apoptosis, or required for mitosis. *Journal of Cell Science.* 114 (10), 1821–1827.
Silverman, N., Zhou, R., Erlich, R. L., Hunter, M., Bernstein, E., Schneider, D. & Maniatis, T. (2003) Immune activation of NF-κB and JNK requires *Drosophila* TAK1. *Journal of Biological Chemistry.* 278 (49), 48928–48934.
Simões, M. L., Mlambo, G., Tripathi, A., Dong, Y. & Dimopoulos, G. (2017a) Immune regulation of *Plasmodium* is *Anopheles* species specific and infection intensity dependent. *MBio.* 8 (5), e01631–17.

Simões, M. L., Dong, Y., Hammond, A., Hall, A., Crisanti, A., Nolan, T. & Dimopoulos, G. (2017b) The *Anopheles* FBN9 immune factor mediates *Plasmodium* species-specific defense through transgenic fat body expression. *Developmental & Comparative Immunology.* 67, 257–265.
Sinden, R. (2010) A biologist's perspective on malaria vaccine development. *Human Vaccines.* 6 (1), 3–11.
Singh, S. R., Zeng, X., Zheng, Z. & Hou, S. X. (2011) The adult *Drosophila* gastric and stomach organs are maintained by a multipotent stem cell pool at the foregut/midgut junction in the cardia (proventriculus). *Cell Cycle.* 10 (7), 1109–1120.
Singh, H., Kirar, M. & Sehrawat, N. (2022) *In silico* characterization and evolutionary analysis of conserved GPI-anchored protein SGU (Secretory Glycoconjugate of Unknown function) in *Anopheles gambiae*. *International Journal of Mosquito Research.* 9 (6), 1–14.
Sluss, H. K., Han, Z., Barrett, T., Goberdhan, D. C., Wilson, C., Davis, R. J. & Ip, Y. T. (1996) A JNK signal transduction pathway that mediates morphogenesis and an immune response in *Drosophila*. *Genes & Development.* 10 (21), 2745–2758.
Smyth, G., Hu, Y., Dunn, P., Phipson, B. & Chen, Y. (2013) *Package 'statmod'. R package version 1.5.0.* https://cran.r-project.org/web/packages/statmod/index.html.
So, W. V., Sarov-Blat, L., Kotarski, C. K., McDonald, M. J., Allada, R. & Rosbash, M. (2000) *Takeout*, a novel *Drosophila* gene under circadian clock transcriptional regulation. *Molecular and Cellular Biology.* 20 (18), 6935–6944.
Souza-Neto, J. A., Sim, S. & Dimopoulos, G. (2009) An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. *Proceedings of the National Academy of Sciences.* 106 (42), 17841–17846.
Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M. & Levine, M. (2002) Whole-genome analysis of dorsal-ventral patterning in the *Drosophila* embryo. *Cell.* 111 (5), 687–701.
Stathopoulos, S., Neafsey, D. E., Lawniczak, M. K., Muskavitch, M. A. & Christophides, G. K. (2014) Genetic dissection of *Anopheles gambiae* gut epithelial responses to *Serratia marcescens*. *PLoS Pathogens.* 10 (3), e1003897.
Stincone, A., Prigione, A., Cramer, T., Wamelink, M. M., Campbell, K., Cheung, E., Olin-Sandoval, V., Grüning, N., Krüger, A. & Tauqeer Alam, M. (2015) The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biological Reviews.* 90 (3), 927–963.
Stöven, S., Silverman, N., Junell, A., Hedengren-Olcott, M., Erturk, D., Engström, Y., Maniatis, T. & Hultmark, D. (2003) Caspase-mediated processing of the *Drosophila* NF-κB factor Relish. *Proceedings of the National Academy of Sciences.* 100 (10), 5991–5996.
Sun, H., Bristow, B. N., Qu, G. & Wasserman, S. A. (2002) A heterotrimeric death domain complex in Toll signaling. *Proceedings of the National Academy of Sciences.* 99 (20), 12871–12876.
Suwanchaichinda, C. & Kanost, M. R. (2009) The serpin gene family in *Anopheles gambiae*. *Gene.* 442 (1–2), 47–54.
Tafesh-Edwards, G. & Eleftherianos, I. (2020) JNK signaling in *Drosophila* immunity and homeostasis. *Immunology Letters.* 226, 7–11.
Tahar, R., Boudin, C., Thiery, I. & Bourgouin, C. (2002) Immune response of *Anopheles gambiae* to the early sporogonic stages of the human malaria parasite *Plasmodium falciparum*. *The EMBO Journal.* 21 (24), 6673–6680.
Takehana, A., Katsuyama, T., Yano, T., Oshima, Y., Takada, H., Aigaki, T. & Kurata, S. (2002) Overexpression of a pattern-recognition receptor, peptidoglycan-recognition protein-LE, activates imd/relish-mediated antibacterial defense and the prophenoloxidase cascade in *Drosophila* larvae. *Proceedings of the National Academy of Sciences.* 99 (21), 13705–13710.
Tapanelli, S., Inghilterra, M. G., Cai, J. A., Philpott, J., Capriotti, P., Windbichler, N. & Christophides, G. K. (2021) Assessment of *Plasmodium falciparum* infection and fitness of genetically modified *Anopheles gambiae* aimed at mosquito population replacement. *Frontiers in Tropical Diseases.* 2, 806880.
Tauszig-Delamasure, S., Bilak, H., Capovilla, M., Hoffmann, J. A. & Imler, J. (2002) *Drosophila* MyD88 is required for the response to fungal and Gram-positive bacterial infections. *Nature Immunology.* 3 (1), 91–97.
Tchioffo, M. T., Boissière, A., Abate, L., Nsango, S. E., Bayibeki, A. N., Awono-Ambene, P. H., Christen, R., Gimonneau, G. & Morlais, I. (2016) Dynamics of bacterial community composition in the malaria mosquito's epithelia. *Frontiers in Microbiology.* 6, 1500.
Tchioffo, M. T., Boissière, A., Churcher, T. S., Abate, L., Gimonneau, G., Nsango, S. E., Awono-Ambéné, P. H., Christen, R., Berry, A. & Morlais, I. (2013) Modulation of malaria infection in *Anopheles gambiae* mosquitoes exposed to natural midgut bacteria. *PLoS ONE.* 8 (12), e81663.
Teem, J. L., Ambali, A., Glover, B., Ouedraogo, J., Makinde, D. & Roberts, A. (2019) Problem formulation for gene drive mosquitoes designed to reduce malaria transmission in Africa: results from four regional consultations 2016–2018. *Malaria Journal.* 18, 1–13.
ter Kuile, F., White, N. J., Holloway, P., Pasvol, G. & Krishna, S. (1993) *Plasmodium falciparum*: in vitro studies of the pharmacodynamic properties of drugs used for the treatment of severe malaria. *Experimental Parasitology.* 76 (1), 85–95.

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G. & Nielsen, H. (2022) SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology.* 40 (7), 1023–1025.
Therneau, T. M. & Grambsch, P. M. (2000) *Modeling survival data: extending the Cox model.* New York, Springer.
Thomas, S., Karnik, S., Barai, R. S., Jayaraman, V. K. & Idicula-Thomas, S. (2010) CAMP: a useful resource for research on antimicrobial peptides. *Nucleic Acids Research.* 38 (Database issue), D774–D780.
Toba, G. & Aigaki, T. (2000) Disruption of the *microsomal glutathione S-transferase-like* gene reduces life span of *Drosophila melanogaster*. *Gene.* 253 (2), 179–187.
Tossi, A., Sandri, L. & Giangaspero, A. (2000) Amphipathic, α-helical antimicrobial peptides. *Peptide Science.* 55 (1), 4–30.
Towb, P., Bergmann, A. & Wasserman, S. A. (2001) The protein kinase Pelle mediates feedback regulation in the *Drosophila* Toll signaling pathway. *Development.* 128 (23), 4729–4736.
Tsai, Y., Hayward, R. E., Langer, R. C., Fidock, D. A. & Vinetz, J. M. (2001) Disruption of *Plasmodium falciparum* chitinase markedly impairs parasite invasion of mosquito midgut. *Infection and Immunity.* 69 (6), 4048–4054.
Tzou, P., Reichhart, J. & Lemaitre, B. (2002) Constitutive expression of a single antimicrobial peptide can restore wild-type resistance to infection in immunodeficient *Drosophila* mutants. *Proceedings of the National Academy of Sciences.* 99 (4), 2152–2157.
Ukegbu, C. V., Giorgalli, M., Tapanelli, S., Rona, L. D., Jaye, A., Wyer, C., Angrisano, F., Blagborough, A. M., Christophides, G. K. & Vlachou, D. (2020) PIMMS43 is required for malaria parasite immune evasion and sporogonic development in the mosquito vector. *Proceedings of the National Academy of Sciences.* 117 (13), 7363–7373.
Ukegbu, C. V., Giorgalli, M., Yassine, H., Ramirez, J. L., Taxiarchi, C., Barillas-Mury, C., Christophides, G. K. & Vlachou, D. (2017) *Plasmodium berghei* P47 is essential for ookinete protection from the *Anopheles gambiae* complement-like response. *Scientific Reports.* 7 (1), 6026.
Unckless, R. L., Howick, V. M. & Lazzaro, B. P. (2016) Convergent balancing selection on an antimicrobial peptide in *Drosophila*. *Current Biology.* 26 (2), 257–262.
Unckless, R. L. & Lazzaro, B. P. (2016) The potential for adaptive maintenance of diversity in insect antimicrobial peptides. *Philosophical Transactions of the Royal Society B: Biological Sciences.* 371 (1695), 20150291.
Unckless, R. L., Rottschaefer, S. M. & Lazzaro, B. P. (2015) The complex contributions of genetics and nutrition to immunity in *Drosophila melanogaster*. *PLoS Genetics.* 11 (3), e1005030.
Upton, L. M., Povelones, M. & Christophides, G. K. (2014) *Anopheles gambiae* blood feeding initiates an anticipatory defense response to *Plasmodium berghei*. *Journal of Innate Immunity.* 7 (1), 74–86.
Valzano, M., Cecarini, V., Cappelli, A., Capone, A., Bozic, J., Cuccioloni, M., Epis, S., Petrelli, D., Angeletti, M., Eleuteri, A. M., Favia, G. & Ricci, I. (2016) A yeast strain associated to *Anopheles* mosquitoes produces a toxin able to kill malaria parasites. *Malaria Journal.* 15 (20), 21.
Vannini, L., Dunn, W. A., Reed, T. W. & Willis, J. H. (2014) Changes in transcript abundance for cuticular proteins and other genes three hours after a blood meal in *Anopheles gambiae*. *Insect Biochemistry and Molecular Biology.* 44, 33–43.
Vernooy, S. Y., Chow, V., Su, J., Verbrugghe, K., Yang, J., Cole, S., Olson, M. R. & Hay, B. A. (2002) *Drosophila* Bruce can potently suppress Rpr-and Grim-dependent but not Hid-dependent cell death. *Current Biology.* 12 (13), 1164–1168.
Vizioli, J., Bulet, P., Hoffmann, J. A., Kafatos, F. C., Muller, H. M. & Dimopoulos, G. (2001b) Gambicin: a novel immune responsive antimicrobial peptide from the malaria vector *Anopheles gambiae*. *Proceedings of the National Academy of Sciences.* 98 (22), 12630–12635.
Vizioli, J., Bulet, P., Charlet, M., Lowenberger, C., Blass, C., Müller, H., Dimopoulos, G., Hoffmann, J., Kafatos, F. C. & Richman, A. (2000) Cloning and analysis of a cecropin gene from the malaria vector mosquito, *Anopheles gambiae*. *Insect Molecular Biology.* 9 (1), 75–84.
Vizioli, J., Richman, A. M., Uttenweiler-Joseph, S., Blass, C. & Bulet, P. (2001a) The defensin peptide of the malaria vector mosquito *Anopheles gambiae*: antimicrobial activities and expression in adult mosquitoes. *Insect Biochemistry and Molecular Biology.* 31 (3), 241–248.
Vlachou, D., Schlegelmilch, T., Christophides, G. K. & Kafatos, F. C. (2005) Functional genomic analysis of midgut epithelial responses in *Anopheles* during *Plasmodium* invasion. *Current Biology.* 15 (13), 1185–1195.
Volohonsky, G., Hopp, A., Saenger, M., Soichot, J., Scholze, H., Boch, J., Blandin, S. A. & Marois, E. (2017) Transgenic expression of the anti-parasitic factor TEP1 in the malaria mosquito *Anopheles gambiae*. *PLoS Pathogens.* 13 (1), e1006113.
Volz, J., Müller, H., Zdanowicz, A., Kafatos, F. C. & Osta, M. A. (2006) A genetic module regulates the melanization response of *Anopheles* to *Plasmodium*. *Cellular Microbiology.* 8 (9), 1392–1405.
Volz, J., Osta, M. A., Kafatos, F. C. & Müller, H. (2005) The roles of two Clip domain serine proteases in innate immune responses of the malaria vector *Anopheles gambiae*. *Journal of Biological Chemistry.* 280 (48), 40161–40168.
Wadhwa, R. R., Subramanian, V. & Stevens-Truss, R. (2018) Visualizing alpha-helical peptides in R with helixvis. *Journal of Open Source Software.* 3 (31), 1008.

Waghu, F. H., Barai, R. S., Gurung, P. & Idicula-Thomas, S. (2016) $CAMP_{R3}$: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Research.* 44 (Database issue), D1094–D1097.
Wang, G., Li, X. & Wang, Z. (2009) APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Research.* 37 (Database issue), D933–D937.
Wang, J. & 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), 10552–10557.
Wang, J., Wu, Y., Yang, G. & Aksoy, S. (2009) Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. *Proceedings of the National Academy of Sciences.* 106 (29), 12133–12138.
Wang, M. C., Bohmann, D. & Jasper, H. (2005) JNK extends life span and limits growth by antagonizing cellular and organism-wide responses to insulin signaling. *Cell.* 121 (1), 115–125.
Wang, S., Ghosh, A. K., Bongio, N., Stebbings, K. A., Lampe, D. J. & Jacobs-Lorena, M. (2012) Fighting malaria with engineered symbiotic bacteria from vector mosquitoes. *Proceedings of the National Academy of Sciences.* 109 (31), 12734–12739.
Wang, X., Rocheleau, T. A., Fuchs, J. F., Hillyer, J. F., Chen, C. C. & Christensen, B. M. (2004) A novel lectin with a fibrinogen-like domain and its potential involvement in the innate immune response of *Armigeres subalbatus* against bacteria. *Insect Molecular Biology.* 13 (3), 273–282.
Wang, X., Zhao, Q. & Christensen, B. M. (2005) Identification and characterization of the fibrinogen-like domain of fibrinogen-related proteins in the mosquito, *Anopheles gambiae*, and the fruitfly, *Drosophila melanogaster*, genomes. *BMC Genomics.* 6 (1), 114.
Wang, Y., Gilbreath III, T. M., Kukutla, P., Yan, G. & Xu, J. (2011) Dynamic gut microbiome across life history of the malaria mosquito *Anopheles gambiae* in Kenya. *PLoS ONE.* 6 (9), e24767.
Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M. & Venables, B. (2020) *gplots: Various R Programming Tools for Plotting Data.*
Warr, E., Das, S., Dong, Y. & Dimopoulos, G. (2008) The *Gram-Negative Bacteria-Binding Protein* gene family: its role in the innate immune system of *Anopheles gambiae* and in anti-*Plasmodium* defence. *Insect Molecular Biology.* 17 (1), 39–51.
Warr, E., Aguilar, R., Dong, Y., Mahairaki, V. & Dimopoulos, G. (2007) Spatial and sex-specific dissection of the *Anopheles gambiae* midgut transcriptome. *BMC Genomics.* 8 (1), 1–11.
Waterhouse, R. M., Kriventseva, E. V., Meister, S., Xi, Z., Alvarez, K. S., Bartholomay, L. C., Barillas-Mury, C., Bian, G., Blandin, S. & Christensen, B. M. (2007) Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science.* 316 (5832), 1738–1743.
Weber, A. N., Tauszig-Delamasure, S., Hoffmann, J. A., Lelièvre, E., Gascan, H., Ray, K. P., Morse, M. A., Imler, J. & Gay, N. J. (2003) Binding of the *Drosophila* cytokine Spätzle to Toll is direct and establishes signaling. *Nature Immunology.* 4 (8), 794–800.
Weigel, D., Jürgens, G., Küttner, F., Seifert, E. & Jäckle, H. (1989) The homeotic gene *fork head* encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. *Cell.* 57 (4), 645–658.
Weiss, B. & Aksoy, S. (2011) Microbiome influences on insect host vector competence. *Trends in Parasitology.* 27 (11), 514–522.
White, G. B. (1974) *Anopheles gambiae* complex and disease transmission in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene.* 68 (4), 278–298.
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L. & Hester, J. (2019) Welcome to the Tidyverse. *Journal of Open Source Software.* 4 (43), 1686.
Wilding, C. S., Weetman, D., Rippon, E. J., Steen, K., Mawejje, H. D., Barsukov, I. & Donnelly, M. J. (2015) Parallel evolution or purifying selection, not introgression, explains similarity in the pyrethroid detoxification linked GSTE4 of *Anopheles gambiae* and *An. arabiensis*. *Molecular Genetics and Genomics.* 290, 201–215.
Willenborg, M., Schmidt, C. K., Braun, P., Landgrebe, J., von Figura, K., Saftig, P. & Eskelinen, E. (2005) Mannose 6-phosphate receptors, Niemann-Pick C2 protein, and lysosomal cholesterol accumulation. *Journal of Lipid Research.* 46 (12), 2559–2569.
Williams, M., Contet, A., Hou, C. D., Levashina, E. A. & Baxter, R. H. G. (2019) *Anopheles gambiae* TEP1 forms a complex with the coiled-coil domain of LRIM1/APL1C following a conformational change in the thioester domain. *PLoS ONE.* 14 (6), e0218203.
World Health Organization. (2021) *World Malaria Report 2021.*
World Health Organization. (2022) *World Malaria Report 2022.*
World Health Organization. (2023) *World Malaria Report 2023.*
Wu, L. P. & Anderson, K. V. (1998) Regulated nuclear import of Rel proteins in the *Drosophila* immune response. *Nature.* 392 (6671), 93–97.
Wullschleger, S., Loewith, R. & Hall, M. N. (2006) TOR signaling in growth and metabolism. *Cell.* 124 (3), 471–484.
Xiao, X., Wang, P., Lin, W., Jia, J. & Chou, K. (2013) iAMP-2L: a two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Analytical Biochemistry.* 436 (2), 168–177.
Yassine, H., Kamareddine, L. & Osta, M. A. (2012) The mosquito melanization response is implicated in defense against the entomopathogenic fungus *Beauveria bassiana*. *PLoS Pathogens.* 8 (11), e1003029.

Yassine, H. & Osta, M. A. (2010) *Anopheles gambiae* innate immunity. *Cellular Microbiology.* 12 (1), 1–9.
Yu, S., Luo, F., Xu, Y., Zhang, Y. & Jin, L. H. (2022) *Drosophila* innate immunity involves multiple signaling pathways and coordinated communication between different tissues. *Frontiers in Immunology.* 13, 905370.
Zaidman-Rémy, A., Hervé, M., Poidevin, M., Pili-Floury, S., Kim, M., Blanot, D., Oh, B., Ueda, R., Mengin-Lecreulx, D. & Lemaitre, B. (2006) The *Drosophila* amidase PGRP-LB modulates the immune response to bacterial infection. *Immunity.* 24 (4), 463–473.
Zdobnov, E. M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R. R., Christophides, G. K., Thomasova, D., Holt, R. A., Subramanian, G. M., Mueller, H., Dimopoulos, G., Law, J. H., Wells, M. A., Birney, E., Charlab, R., Halpern, A. L., Kokoza, E., Kraft, C. L., Lai, Z., Lewis, S., Louis, C., Barillas-Mury, C., Nusskern, D., Rubin, G. M., Salzberg, S. L., Sutton, G. G., Topalis, P., Wides, R., Wincker, P., Yandell, M., Collins, F. H., Ribeiro, J., Gelbart, W. M., Kafatos, F. C. & Bork, P. (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science.* 298 (5591), 149–159.
Zhang, G., Niu, G., Franca, C. M., Dong, Y., Wang, X., Butler, N. S., Dimopoulos, G. & Li, J. (2015) *Anopheles* midgut FREP1 mediates *Plasmodium* invasion. *Journal of Biological Chemistry.* 290 (27), 16490–16501.
Zhang, L. & Gallo, R. L. (2016) Antimicrobial peptides. *Current Biology.* 26 (1), R14–R19.
Zhang, S., Zeng, Y. & Loker, E. S. (2008) Expression profiling and binding properties of fibrinogen-related proteins (FREPs), plasma proteins from the schistosome snail host *Biomphalaria glabrata*. *Innate Immunity.* 14 (3), 175–189.
Zhang, W., Tettamanti, G., Bassal, T., Heryanto, C., Eleftherianos, I. & Mohamed, A. (2021) Regulators and signalling in insect antimicrobial innate immunity: functional molecules and cellular pathways. *Cellular Signalling.* 83, 110003.
Zhao, J., Brault, J. J., Schild, A., Cao, P., Sandri, M., Schiaffino, S., Lecker, S. H. & Goldberg, A. L. (2007) FoxO3 coordinately activates protein degradation by the autophagic/lysosomal and proteasomal pathways in atrophying muscle cells. *Cell Metabolism.* 6 (6), 472–483.
Zhao, P., Lu, Z., Strand, M. R. & Jiang, H. (2011) Antiviral, antiparasitic, and cytotoxic effects of 5,6-dihydroxyindole (DHI), a reactive compound generated by phenoloxidase during insect immune response. *Insect Biochemistry and Molecular Biology.* 41 (9), 645–652.
Zheng, X. & Zheng, L. (2002) Genomic organization and regulation of three cecropin genes in *Anopheles gambiae*. *Insect Molecular Biology.* 11 (6), 517–525.
Zhu, J. & Noriega, F. G. (2016) The role of juvenile hormone in mosquito development and reproduction. In: *Advances in Insect Physiology.* Elsevier. pp. 93–113.
Zhu, J., Busche, J. M. & Zhang, X. (2010) Identification of juvenile hormone target genes in the adult female mosquitoes. *Insect Biochemistry and Molecular Biology.* 40 (1), 23–29.

Chapter 8: Appendices

Appendix A

Table A.1. List of primers. *Primer pair sequences used for dsRNA generation, RNA probe generation, and in qRT-PCR reactions to quantify expression levels.*

Gene	Type	Sequence	PE
AGAP000692 *CEC2*	probe F	taatacgactcactatagggGTACTGATGGCTGCGTTCCT	N/A
	probe R	taatacgactcactatagggAAACCCTTCTGGGCTGCATT	
	qPCR F	GTACTGATGGCTGCGTTCCT	1.82
	qPCR R	AAACCCTTCTGGGCTGCATT	
AGAP000693 *CEC1*	qPCR F	ACCAACCAACCACCAAACAAC	2.02
	qPCR R	CTTCTCTGCTGCCTTGAACACT	
AGAP001181	dsRNA F	taatacgactcactatagggGCGGAGTAACTTACCGGTTG	N/A
	dsRNA R	taatacgactcactatagggCCGTTTGCATTAAACTTTACACC	
	qPCR F	CGAGGCTGTGTCCATCAGTT	1.96
	qPCR R	AGTAACGAACAGACGGCGAG	
AGAP001352	dsRNA F	taatacgactcactatagggGTAGCGACACTCCGTACCGT	N/A
	dsRNA R	taatacgactcactatagggTCACGAAGCTAAAACTGCGA	
	qPCR F	CCAGGACGAGGGGTTTCTTC	1.99
	qPCR R	GAGAGCCGGATTGGGACAAA	
AGAP001508	dsRNA F	taatacgactcactatagggAGTCGTTCAACCAGGGCTTC	N/A
	dsRNA R	taatacgactcactatagggTATCCTCCGAACGACTGGG	
	qPCR F	CGTTCTTTCCCACACCACCT	2.05
	qPCR R	AAGCCAACCTGTCGCACTA	
AGAP002852	dsRNA F	taatacgactcactatagggGGTAGCTCTGCGTCGTCTGT	N/A
	dsRNA R	taatacgactcactatagggCGTGCTCGAGAGGAACATC	
	qPCR F	CAGCGTAAAGACGAAAGCCG	1.97
	qPCR R	AGCACATCGCCCCACTTATT	
AGAP003713	dsRNA F	taatacgactcactatagggATGAAAGGACTAGCCGTCGT	N/A
	dsRNA R	taatacgactcactatagggCCTTGAGCGGATACACGAGA	
	qPCR F	CCTGTATGCCGGTGGTATCG	2.11
	qPCR R	CGCCTTCACTCGTCTATCGG	
AGAP004935 *CACT-like*	dsRNA F	taatacgactcactatagggGAAAAGCGAACCGACGTAAT	N/A
	dsRNA R	taatacgactcactatagggCAGACAGCTGTGAGCCGTG	
	qPCR F	ATAAACCATCCGACGCACGA	1.97
	qPCR R	GTAAAACCCAAGCCGCACAG	
AGAP005010 *ARA*	dsRNA F	taatacgactcactatagggGGCGATATGGACCATTCAAA	N/A
	dsRNA R	taatacgactcactatagggCTGATGGTGGTGGTGCTG	
	qPCR F	ATGGCCGCATACGCTCAAT	2.03
	qPCR R	TGTCGTCTGTCCTCCGCTAA	
AGAP005011 *MIRR*	dsRNA F	taatacgactcactatagggACCGACAGTTCCGTCTGC	N/A
	dsRNA R	taatacgactcactatagggGTATCGCAACAGCGTCCAC	
	qPCR F	AGACAGTGATTGGAGAACGGC	1.97
	qPCR R	GCCTCGGAAGGGAAAATGTG	
AGAP005326 *CACT-like*	dsRNA F	taatacgactcactatagggGGCGGAAAGTGTTGTGGAT	N/A
	dsRNA R	taatacgactcactatagggGTCTGCTCGCAGTCCTCCT	
	qPCR F	TGGACGTGTTGGGGGAAAAA	1.95
	qPCR R	GTCCTCCCATTGCTGTCCAA	

Gene	Type	Sequence	PE
AGAP005888 *KIN1*	qPCR F	GAACGGCAAGAAATGCGATCA	2.21
	qPCR R	GATCGTGATTGTGATGTCCCG	
	dsRNA F	taatacgactcactatagggCGGAGACGCTGCACAGTAAA	N/A
	dsRNA R	taatacgactcactatagggGCCACGATGACAAAGATCGC	
AGAP005889 *GLYR*	dsRNA F	taatacgactcactatagggGATGGTTAGCGCACAAGGT	N/A
	dsRNA R	taatacgactcactatagggACCTCCGTGCACGTTCTTCT	
	qPCR F	AACAAGCGAACGCGTCTGTTA	2.21
	qPCR R	ACGATGGCGATCACCAGAAA	
AGAP005890 *ALAR*	dsRNA F	taatacgactcactatagggAGTCGGCCTGGAACAACA	N/A
	dsRNA R	taatacgactcactatagggGGCCATGTCATTACGGTTG	
	qPCR F	CCAGTGGAACCGCTATGGAG	2.17
	qPCR R	CTACATCGGCCGCTTTACCA	
AGAP006194 *FAMeT1*	dsRNA F	taatacgactcactatagggGTTCGCTGAAAGGTGTACCG	N/A
	dsRNA R	taatacgactcactatagggGCGTACGTGTTGCAGGTTG	
	qPCR F	ATTTGCGGTTTGGAAGGTCG	2.19
	qPCR R	GCTGTTTCGCCATCCTCCTA	
AGAP006195 *FAMeT2*	dsRNA F	taatacgactcactatagggATCGTCTACAACGGCACGAC	N/A
	dsRNA R	taatacgactcactatagggCCGTACTGCTTACAGCCTTT	
	qPCR F	TGTCCGCTAATCGTCCAACC	2.16
	qPCR R	ACGGGTAATCCTGTCCTTCG	
AGAP007347 *LYSC1*	dsRNA F	taatacgactcactatagggGAAGCTAAAACGTTCGGCAA	N/A
	dsRNA R	taatacgactcactatagggTCGCACCAGTACTTGTTGTTG	
	qPCR F	TGCAAGATCGCGTGCAAAA	2.04
	qPCR R	CGTTTGTGGATCAGCTTGGC	
AGAP007938 *CACT*	dsRNA F	taatacgactcactatagggGGTTTCTTTCCGGGCATAAC	N/A
	dsRNA R	taatacgactcactatagggCGGAGTCTAGGTCGGTTTCC	
	qPCR F	TGTCGTAGAGAGCCCACTGA	1.9
	qPCR R	CGCTGCCAACAGGATGAATG	
AGAP008307	dsRNA F	taatacgactcactatagggAGCACTTAGCCGACAGCG	N/A
	dsRNA R	taatacgactcactatagggACCTATCAGCTCGGTACGGG	
	qPCR F	TCGGTTGTCCCTTCAGCTTG	2.15
	qPCR R	TAAGGTCAAACAGCCCCTGC	
AGAP008645 *GAM1*	dsRNA F	taatacgactcactatagggCTATCTCAACCGGAAGG	N/A
	dsRNA R	taatacgactcactatagggCCAACGTCTGGCACTGATT	
	qPCR F	GTGCTGCCAGGGTTTGATTC	2.1
	qPCR R	AACACCATTGCATCCGCAAC	
AGAP009313	dsRNA F	taatacgactcactatagggTTAGCGTACGTCCAAACGAT	N/A
	dsRNA R	taatacgactcactatagggGTCGAGCAGATGGGTGAACT	
	probe F	taatacgactcactatagggCAAATTGTGGCGATTGGGCA	N/A
	probe R	taatacgactcactatagggATTGCATGGCACACTACGAC	
	qPCR F	AGATGGTGCGGAAAAGGTGA	2.14
	qPCR R	GCCGATCCACGAGTTTTCCT	
AGAP009635	dsRNA F	taatacgactcactatagggAGAGCTCGCTAGCGGAAAG	N/A
	dsRNA R	taatacgactcactatagggCTGCATTCCGTCGTCCAT	
	qPCR F	GGATTCGATGGGCGAAATGG	2.11
	qPCR R	CCTCTGTATCCTCCGAAGCC	

Gene	Type	Sequence	PE
AGAP009646 *CAD*	dsRNA F	taatacgactcactatagggCCATGTACTATCCCCATCCG	N/A
	dsRNA R	taatacgactcactatagggTGTCTTTCGTGCGTGTCTTC	
	qPCR F	ACCAGAGTCAACCCAATCCA	2.11
	qPCR R	ATCGTGATGTAGCGCGTGTA	
AGAP009771	dsRNA F	taatacgactcactatagggTTGTGTGTGATCTGTGCGG	N/A
	dsRNA R	taatacgactcactatagggGATGCGTCAGGACGAAGTG	
	qPCR F	ATACGGCAGGCAGAGGAAAC	2.05
	qPCR R	CCGTAGCTTGGTGGAGTAGG	
AGAP010592 *S7*	qPCR F	GTGCGCGAGTTGGAGAAGA	2.1
	qPCR R	ATCGGTTTGGGCAGAATGC	
AGAP011294 *DEF1*	qPCR F	CATGCCGCGCTGGAGAACTA	2.06
	qPCR R	GATAGCGGCGAGCGATACAGTGA	
AGAP011630	dsRNA F	taatacgactcactatagggGGTGGCATCGGAAAGAACTA	N/A
	dsRNA R	taatacgactcactatagggTGCACAATGAAAGAGCATGT	
	qPCR F	TCCCGAGCAACGAGATGAAA	2.04
	qPCR R	TAACCATACAGGCACCGAGC	
AGAP013543	dsRNA F	taatacgactcactatagggGACGAGGACCACCCAGAG	N/A
	dsRNA R	taatacgactcactatagggAGCGTTTCAGGTAGTCCAGC	
	probe F	taatacgactcactatagggCCTTTGCTCTCGTGCTATGC	N/A
	probe R	taatacgactcactatagggCCAGTTCCTCCGCTACCAAG	
	qPCR F	TTTGCTCTCGTGCTATGCGT	2
	qPCR R	TTGTTCGGCGTCATCCAGAG	
AGAP029578 *DVE*	dsRNA F	taatacgactcactatagggGCTCGGCTACTCCACAGAGA	N/A
	dsRNA R	taatacgactcactatagggGTTGCGCAGGATCGTTATCT	
	qPCR F	CCATCGCACAACCTGATCCT	2.15
	qPCR R	CGGATCGTACAGCGACATCT	
16S	qPCR F	TCCTACGGGAGGCAGCAGT	1.87
	qPCR R	GGACTACCAGGGTATCTAATCCTGTT	
Acetobacteraceae *16S*	qPCR F	GTGCCGATCTCTAAAAGCCGTCTCA	2.03
	qPCR R	TTCGCTCACCGGCTTCGGGT	
E. faecalis 16S	qPCR F	CCGAGTGCTTGCACTCAATTGG	1.97
	qPCR R	CTCTTATGCCATGCGGCATAAAC	
Enterobacteria-ceae *16S*	qPCR F	CGTGCTACAATGGCATATACAAAGAGAAG	2.15
	qPCR R	AGCATTCTGATCTACGATTACTAGCGATTC	
Flavobacteriaceae *16S*	qPCR F	TAAGGTTGAAGTGGCTGGAATAA	2.1
	qPCR R	GTCCATCAGCGTCAGTTAAGACT	
GFP	dsRNA F	taatacgactcactatagggGACGTAAACGGCCACAAGTT	N/A
	dsRNA R	taatacgactcactatagggTGTTCTGCTGGTAGTGGTCG	
LacZ	dsRNA F	taatacgactcactatagggAGAATCCGACGGGTTGTTACT	N/A
	dsRNA R	taatacgactcactatagggCACCACGCTCATCGATAATTT	
S. marcescens 16S	qPCR F	ACGTTCATCAATTGACGTTACTCGCA	1.97
	qPCR R	AACCGCCTGCGTGCGCTTTA	

Primer types: dsRNA = used to amplify a gene from cDNA to generate dsRNA, T7 promoter sequence is in lowercase; qPCR = verification primers for assessment of KD efficiency and/or relative expression via qRT-PCR, probe = used for RNA probe synthesis in RNA FISH, F = forward primer, R = reverse prime, PE = primer efficiency.

Appendix B

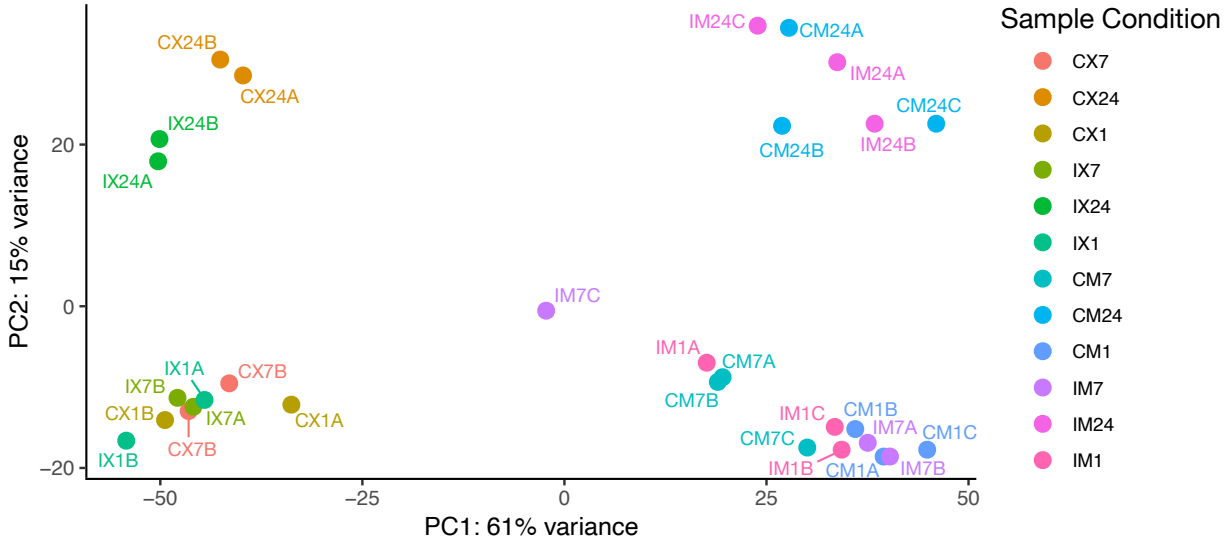


Figure B.1. PCA of the* Ag/Pf *data prior to outlier removal. *PCA was performed following the normalization of read counts with DESeq2 (Love, Huber & Anders, 2014). Each dot represents a transcriptome sequenced from a pool of midguts. Sample IM7C was removed from subsequent analyses because it did not cluster with other replicates of the same sample type. I =* P. falciparum*-infected, C = control (blood-fed); M = midgut, X = carcass; 1 = 1 hour post-infection/blood feed, 24 = 24 hours post-infection/blood feed, 7 = 7 days post-infection/blood feed; A/B/C represent replicates 1, 2, and 3 respectively. Three replicates were performed for midgut samples, two replicates were performed for carcass samples.*

Table B.1. Significantly enriched GO terms from DE transcripts between infected and control samples. *Significantly DE transcripts were identified from comparisons of* P. falciparum*-infected samples and blood-fed controls of the same tissue type and time point (Wald tests, false discovery rate adjusted-*p *< 0.05). Identified using g:Profiler (Raudvere et al, 2019).*

Source	GO Term	Term ID	Adjusted p	∩
IM1 – upregulated versus CM1				
MF	phosphatidylinositol binding	GO:0035091	1.86×10^{-2}	8
BP	oogenesis	GO:0048477	3.82×10^{-3}	4
BP	female gamete generation	GO:0007292	3.82×10^{-3}	4
BP	phenol-containing compound metabolic process	GO:0018958	1.18×10^{-2}	3
BP	germ cell development	GO:0007281	1.72×10^{-2}	4
BP	cellular process involved in reproduction in multicellular organism	GO:0022412	2.54×10^{-2}	4
BP	aromatic amino acid metabolic process	GO:0009072	3.61×10^{-2}	4
CC	host intracellular organelle	GO:0033647	2.34×10^{-2}	8
CC	host intracellular membrane-bounded organelle	GO:0033648	2.34×10^{-2}	8
CC	host cell nucleus	GO:0042025	2.34×10^{-2}	8
CC	host cell	GO:0043657	2.34×10^{-2}	8
CC	host cell part	GO:0033643	2.34×10^{-2}	8
CC	host cellular component	GO:0018995	2.34×10^{-2}	8
CC	host intracellular part	GO:0033646	2.34×10^{-2}	8
CC	host intracellular region	GO:0043656	2.34×10^{-2}	8
CC	extracellular region	GO:0005576	2.97×10^{-2}	20
CC	cluster of actin-based cell projections	GO:0098862	4.99×10^{-2}	2
IM1 – downregulated versus CM1				
MF	oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor	GO:0016671	4.53×10^{-3}	3
MF	oxidoreductase activity, acting on a sulfur group of donors	GO:0016667	5.49×10^{-3}	4
MF	peptide-methionine (S)-S-oxide reductase activity	GO:0008113	8.62×10^{-3}	2
MF	L-methionine:thioredoxin-disulfide S-oxidoreductase activity	GO:0033744	8.62×10^{-3}	2
MF	L-methionine-(S)-S-oxide reductase activity	GO:0036456	8.62×10^{-3}	2
MF	structural constituent of chromatin	GO:0030527	3.90×10^{-2}	4
BP	cellular response to chemical stimulus	GO:0070887	4.28×10^{-2}	5
CC	endoplasmic reticulum	GO:0005783	2.00×10^{-3}	9
CC	nucleosome	GO:0000786	2.88×10^{-2}	4
IM7 – upregulated versus CM7				
MF	ATP-dependent protein folding chaperone	GO:0140662	1.42×10^{-4}	3
MF	protein folding chaperone	GO:0044183	1.77×10^{-4}	3
MF	misfolded protein binding	GO:0051787	1.44×10^{-3}	2
MF	heat shock protein binding	GO:0031072	2.20×10^{-2}	2
BP	protein refolding	GO:0042026	7.41×10^{-3}	2
BP	'de novo' post-translational protein folding	GO:0051084	7.41×10^{-3}	2
BP	chaperone cofactor-dependent protein refolding	GO:0051085	7.41×10^{-3}	2
BP	'de novo' protein folding	GO:0006458	9.97×10^{-3}	2
BP	cellular response to unfolded protein	GO:0034620	1.14×10^{-2}	2
BP	chaperone-mediated protein folding	GO:0061077	1.29×10^{-2}	2
BP	response to unfolded protein	GO:0006986	1.45×10^{-2}	2
BP	cellular response to topologically incorrect protein	GO:0035967	2.84×10^{-2}	2
BP	response to topologically incorrect protein	GO:0035966	3.57×10^{-2}	2
CC	plasma membrane	GO:0005886	3.60×10^{-3}	4
CC	cell periphery	GO:0071944	1.19×10^{-2}	4

IM1 = P. falciparum*-infected midguts at 1 hour post-infection, CM1 = control midguts at 1 hour post-blood feed, IM7 = infected midguts at 7 days post-infection, CM7 = control midguts at 7 days post-blood feed; MF = molecular function, BP = biological process, CC = cellular component;* ∩ *= intersection,* i.e*. the number of overlapping transcripts belonging to a list of transcripts from a comparison (of a Venn diagram) and to a GO/KEGG term; adjusted* p *= false discovery rate-adjusted* p*-values of Fisher's one-tailed tests, significant if adjusted* p *< 0.05.*

Table B.2. Significantly enriched GO terms from DE transcripts unique to or shared between infected and control midgut samples at 24 hours. *Significantly DE transcripts were identified from comparisons of* P. falciparum*-infected/control midgut samples at 24 hours compared to at 1 hour of the same infection status (Wald tests, false discovery rate adjusted-*p *< 0.05). GO terms were identified using g:Profiler (Raudvere et al, 2019).*

Source	GO Term	Term ID	Adjusted *p*	∩
Unique to CM24 - downregulated				
MF	actin monomer binding	GO:0003785	7.83×10^{-4}	4
MF	ATPase-coupled ion transmembrane transporter activity	GO:0042625	3.85×10^{-3}	6
MF	ATPase activity, coupled to transmembrane movement of ions, rotational mechanism	GO:0044769	3.85×10^{-3}	6
MF	proton-transporting ATPase activity, rotational mechanism	GO:0046961	3.85×10^{-3}	6
MF	pyrophosphate hydrolysis-driven proton transmembrane transporter activity	GO:0009678	5.25×10^{-3}	6
MF	proton transmembrane transporter activity	GO:0015078	8.10×10^{-3}	11
MF	ATPase-coupled monoatomic cation transmembrane transporter activity	GO:0019829	2.46×10^{-2}	6
BP	actin filament organization	GO:0007015	2.20×10^{-4}	11
BP	protein-containing complex organization	GO:0043933	2.92×10^{-3}	28
BP	supramolecular fiber organization	GO:0097435	5.56×10^{-3}	11
BP	protein-containing complex assembly	GO:0065003	1.33×10^{-2}	25
BP	proton transmembrane transport	GO:1902600	1.63×10^{-2}	8
BP	cellular component assembly	GO:0022607	2.37×10^{-2}	32
BP	actin filament-based process	GO:0030029	4.79×10^{-2}	11
BP	actin cytoskeleton organization	GO:0030036	4.79×10^{-2}	11
CC	cytoplasm	GO:0005737	1.64×10^{-3}	106
CC	membrane protein complex	GO:0098796	1.26×10^{-2}	25
CC	endoplasmic reticulum	GO:0005783	2.13×10^{-2}	20
CC	proton-transporting V-type ATPase complex	GO:0033176	4.37×10^{-2}	5
Unique to CM24 - upregulated at 24 hours				
MF	protein binding	GO:0005515	8.15×10^{-4}	131
MF	binding	GO:0005488	3.51×10^{-2}	281
BP	female gamete generation	GO:0007292	8.29×10^{-5}	6
BP	oogenesis	GO:0048477	8.29×10^{-5}	6
BP	germ cell development	GO:0007281	1.23×10^{-3}	6
BP	reproductive process	GO:0022414	1.63×10^{-3}	11
BP	cellular process involved in reproduction in multicellular organism	GO:0022412	2.37×10^{-3}	6
BP	sexual reproduction	GO:0019953	3.97×10^{-3}	10
BP	developmental process involved in reproduction	GO:0003006	4.26×10^{-3}	7
BP	gamete generation	GO:0007276	4.26×10^{-3}	7
BP	multicellular organismal reproductive process	GO:0048609	4.26×10^{-3}	7
BP	reproduction	GO:0000003	1.97×10^{-2}	11
BP	positive regulation of catalytic activity	GO:0043085	1.97×10^{-2}	11
BP	multicellular organism reproduction	GO:0032504	2.33×10^{-2}	8
BP	positive regulation of molecular function	GO:0044093	4.07×10^{-2}	12
BP	cell cycle	GO:0007049	4.13×10^{-2}	22
CC	host cell	GO:0043657	6.07×10^{-3}	12
CC	host cell nucleus	GO:0042025	6.07×10^{-3}	12
CC	host intracellular membrane-bounded organelle	GO:0033648	6.07×10^{-3}	12
CC	host intracellular organelle	GO:0033647	6.07×10^{-3}	12
CC	host intracellular part	GO:0033646	6.07×10^{-3}	12
CC	host cell part	GO:0033643	6.07×10^{-3}	12
CC	host cellular component	GO:0018995	6.07×10^{-3}	12
CC	host intracellular region	GO:0043656	6.07×10^{-3}	12
CC	nucleosome	GO:0000786	2.04×10^{-2}	8
Unique to IM24 - downregulated at 24 hours				
MF	structural constituent of ribosome	GO:0003735	2.19×10^{-3}	22
MF	fatty acid synthase activity	GO:0004312	7.58×10^{-3}	8
MF	fatty acid elongase activity	GO:0009922	9.44×10^{-3}	7
MF	3-oxo-lignoceronyl-CoA synthase activity	GO:0102338	9.44×10^{-3}	7
MF	3-oxo-arachidoyl-CoA synthase activity	GO:0102336	9.44×10^{-3}	7

MF	3-oxo-cerotoyl-CoA synthase activity	GO:0102337	9.44×10^{-3}	7
MF	very-long-chain 3-ketoacyl-CoA synthase activity	GO:0102756	9.44×10^{-3}	7
MF	pyridoxal phosphate binding	GO:0030170	1.51×10^{-2}	9
MF	vitamin B6 binding	GO:0070279	1.51×10^{-2}	9
MF	vitamin binding	GO:0019842	2.15×10^{-2}	11
BP	aromatic amino acid family catabolic process	GO:0009074	1.79×10^{-3}	6
BP	amide metabolic process	GO:0043603	2.06×10^{-3}	44
BP	translation	GO:0006412	2.36×10^{-3}	36
BP	peptide biosynthetic process	GO:0043043	2.96×10^{-3}	36
BP	carboxylic acid metabolic process	GO:0019752	4.16×10^{-3}	33
BP	amide biosynthetic process	GO:0043604	4.68×10^{-3}	37
BP	oxoacid metabolic process	GO:0043436	4.86×10^{-3}	33
BP	organic acid metabolic process	GO:0006082	5.67×10^{-3}	33
BP	aromatic amino acid metabolic process	GO:0009072	1.27×10^{-2}	6
BP	peptide metabolic process	GO:0006518	1.47×10^{-2}	37
BP	erythrose 4-phosphate/phosphoenolpyruvate family amino acid catabolic process	GO:1902222	4.96×10^{-2}	4
BP	L-phenylalanine catabolic process	GO:0006559	4.96×10^{-2}	4
BP	L-phenylalanine metabolic process	GO:0006558	4.96×10^{-2}	4
BP	erythrose 4-phosphate/phosphoenolpyruvate family amino acid metabolic process	GO:1902221	4.96×10^{-2}	4
BP	tyrosine metabolic process	GO:0006570	4.96×10^{-2}	4
CC	cytosolic ribosome	GO:0022626	1.01×10^{-7}	20
CC	cytosolic large ribosomal subunit	GO:0022625	1.25×10^{-5}	13
CC	ribosome	GO:0005840	2.07×10^{-5}	26
CC	cytosol	GO:0005829	6.84×10^{-5}	40
CC	ribosomal subunit	GO:0044391	1.58×10^{-4}	21
CC	large ribosomal subunit	GO:0015934	2.24×10^{-3}	14
CC	nuclear pore outer ring	GO:0031080	4.99×10^{-2}	4
Unique to IM24 - upregulated at 24 hours				
MF	zinc ion binding	GO:0008270	3.26×10^{-6}	54
MF	transition metal ion binding	GO:0046914	4.15×10^{-5}	66
MF	structural constituent of chromatin	GO:0030527	3.28×10^{-4}	12
MF	nucleic acid binding	GO:0003676	3.81×10^{-2}	111
BP	chromatin organization	GO:0006325	4.23×10^{-6}	19
BP	chromatin remodeling	GO:0006338	8.40×10^{-4}	14
CC	chromatin	GO:0000785	2.78×10^{-5}	22
CC	nucleus	GO:0005634	1.62×10^{-4}	123
CC	nucleosome	GO:0000786	3.65×10^{-4}	12
CC	intracellular membrane-bounded organelle	GO:0043231	1.48×10^{-3}	177
CC	protein-DNA complex	GO:0032993	2.49×10^{-3}	13
CC	membrane-bounded organelle	GO:0043227	2.88×10^{-3}	179
CC	intracellular organelle	GO:0043229	4.52×10^{-3}	195
CC	intracellular anatomical structure	GO:0005622	8.69×10^{-3}	224
CC	organelle	GO:0043226	1.06×10^{-2}	195
CC	chromosome	GO:0005694	1.08×10^{-2}	25
Shared between M24 samples - downregulated				
MF	RNA binding	GO:0003723	4.95×10^{-37}	158
MF	catalytic activity, acting on RNA	GO:0140098	6.23×10^{-33}	91
MF	catalytic activity, acting on a nucleic acid	GO:0140640	3.05×10^{-21}	99
MF	structural constituent of ribosome	GO:0003735	4.56×10^{-20}	62
MF	protein folding chaperone	GO:0044183	1.38×10^{-18}	25
MF	ATP-dependent protein folding chaperone	GO:0140662	1.82×10^{-18}	24
MF	catalytic activity, acting on a tRNA	GO:0140101	9.26×10^{-18}	42
MF	heterocyclic compound binding	GO:1901363	1.62×10^{-15}	403
MF	unfolded protein binding	GO:0051082	3.96×10^{-15}	32
MF	organic cyclic compound binding	GO:0097159	7.85×10^{-15}	403
MF	snoRNA binding	GO:0030515	5.66×10^{-13}	15
MF	translation factor activity, RNA binding	GO:0008135	1.24×10^{-12}	29
MF	translation regulator activity, nucleic acid binding	GO:0090079	1.62×10^{-12}	30
MF	translation regulator activity	GO:0045182	6.05×10^{-12}	31
MF	rRNA binding	GO:0019843	1.78×10^{-10}	21

MF	translation initiation factor activity	GO:0003743	4.61×10^{-10}	20
MF	nucleic acid binding	GO:0003676	9.52×10^{-10}	248
MF	ATP-dependent activity	GO:0140657	1.07×10^{-9}	80
MF	purine nucleotide binding	GO:0017076	2.48×10^{-9}	164
MF	transferase activity, transferring one-carbon groups	GO:0016741	2.99×10^{-9}	36
MF	small molecule binding	GO:0036094	3.21×10^{-9}	190
MF	methyltransferase activity	GO:0008168	6.81×10^{-9}	34
MF	RNA methyltransferase activity	GO:0008173	9.81×10^{-9}	20
MF	ribonucleotide binding	GO:0032553	1.14×10^{-8}	156
MF	purine ribonucleotide binding	GO:0032555	1.79×10^{-8}	154
MF	S-adenosylmethionine-dependent methyltransferase activity	GO:0008757	1.80×10^{-8}	26
MF	purine ribonucleoside triphosphate binding	GO:0035639	2.33×10^{-8}	152
MF	nucleotide binding	GO:0000166	2.95×10^{-8}	177
MF	nucleoside phosphate binding	GO:1901265	2.95×10^{-8}	177
MF	anion binding	GO:0043168	1.51×10^{-7}	172
MF	ligase activity, forming carbon-oxygen bonds	GO:0016875	2.14×10^{-7}	20
MF	aminoacyl-tRNA ligase activity	GO:0004812	2.14×10^{-7}	20
MF	adenyl nucleotide binding	GO:0030554	5.92×10^{-7}	136
MF	tRNA methyltransferase activity	GO:0008175	4.17×10^{-6}	11
MF	adenyl ribonucleotide binding	GO:0032559	4.20×10^{-6}	126
MF	ATP binding	GO:0005524	5.88×10^{-6}	124
MF	ATP-dependent activity, acting on RNA	GO:0008186	1.11×10^{-5}	19
MF	RNA helicase activity	GO:0003724	1.11×10^{-5}	19
MF	structural molecule activity	GO:0005198	2.09×10^{-5}	81
MF	ligase activity	GO:0016874	3.20×10^{-5}	30
MF	carbohydrate derivative binding	GO:0097367	7.63×10^{-5}	162
MF	tRNA binding	GO:0000049	8.89×10^{-5}	14
MF	ribonucleoside triphosphate phosphatase activity	GO:0017111	1.12×10^{-4}	60
MF	heat shock protein binding	GO:0031072	1.40×10^{-4}	13
MF	pyrophosphatase activity	GO:0016462	5.10×10^{-4}	61
MF	hydrolase activity, acting on acid anhydrides	GO:0016817	5.70×10^{-4}	62
MF	5'-3' RNA polymerase activity	GO:0034062	5.88×10^{-4}	11
MF	RNA polymerase activity	GO:0097747	5.88×10^{-4}	11
MF	DNA-directed 5'-3' RNA polymerase activity	GO:0003899	5.88×10^{-4}	11
MF	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	GO:0016818	7.84×10^{-4}	61
MF	ATP hydrolysis activity	GO:0016887	1.60×10^{-3}	37
MF	signal sequence binding	GO:0005048	1.62×10^{-3}	10
MF	tRNA (guanine) methyltransferase activity	GO:0016423	2.20×10^{-3}	6
MF	amide binding	GO:0033218	2.39×10^{-3}	20
MF	RNA endonuclease activity	GO:0004521	2.53×10^{-3}	13
MF	helicase activity	GO:0004386	2.78×10^{-3}	23
MF	catalytic activity, acting on a rRNA	GO:0140102	3.06×10^{-3}	8
MF	endonuclease activity	GO:0004519	4.81×10^{-3}	17
MF	N-methyltransferase activity	GO:0008170	5.52×10^{-3}	13
MF	nuclear localization sequence binding	GO:0008139	8.01×10^{-3}	6
MF	ribonucleoprotein complex binding	GO:0043021	1.12×10^{-2}	13
MF	ATPase-coupled monoatomic cation transmembrane transporter activity	GO:0019829	1.19×10^{-2}	11
MF	RNA endonuclease activity, producing 5'-phosphomonoesters	GO:0016891	1.26×10^{-2}	8
MF	rRNA methyltransferase activity	GO:0008649	1.79×10^{-2}	7
MF	ATPase-coupled ion transmembrane transporter activity	GO:0042625	2.31×10^{-2}	9
MF	ATPase activity, coupled to transmembrane movement of ions, rotational mechanism	GO:0044769	2.31×10^{-2}	9
MF	proton-transporting ATPase activity, rotational mechanism	GO:0046961	2.31×10^{-2}	9
MF	RNA nuclease activity	GO:0004540	2.84×10^{-2}	14
MF	U3 snoRNA binding	GO:0034511	3.07×10^{-2}	4
MF	pyrophosphate hydrolysis-driven proton transmembrane transporter activity	GO:0009678	3.67×10^{-2}	9
BP	ribonucleoprotein complex biogenesis	GO:0022613	7.53×10^{-77}	141
BP	ribosome biogenesis	GO:0042254	1.41×10^{-73}	123

BP	ncRNA metabolic process	GO:0034660	2.95×10^{-70}	147
BP	rRNA metabolic process	GO:0016072	5.52×10^{-63}	96
BP	ncRNA processing	GO:0034470	3.09×10^{-62}	124
BP	rRNA processing	GO:0006364	1.11×10^{-60}	92
BP	RNA processing	GO:0006396	3.46×10^{-38}	143
BP	peptide biosynthetic process	GO:0043043	3.06×10^{-35}	121
BP	translation	GO:0006412	5.36×10^{-35}	120
BP	amide biosynthetic process	GO:0043604	7.10×10^{-33}	122
BP	gene expression	GO:0010467	1.21×10^{-32}	303
BP	peptide metabolic process	GO:0006518	2.12×10^{-32}	125
BP	cellular component biogenesis	GO:0044085	6.42×10^{-32}	169
BP	cellular nitrogen compound metabolic process	GO:0034641	1.29×10^{-29}	348
BP	ribosomal large subunit biogenesis	GO:0042273	8.12×10^{-29}	43
BP	amide metabolic process	GO:0043603	4.01×10^{-26}	128
BP	organonitrogen compound biosynthetic process	GO:1901566	5.90×10^{-20}	160
BP	tRNA metabolic process	GO:0006399	4.19×10^{-18}	54
BP	cellular process	GO:0009987	1.08×10^{-17}	658
BP	cellular component organization or biogenesis	GO:0071840	1.09×10^{-16}	217
BP	ribosomal small subunit biogenesis	GO:0042274	1.10×10^{-16}	31
BP	maturation of LSU-rRNA	GO:0000470	2.32×10^{-16}	22
BP	cellular metabolic process	GO:0044237	1.18×10^{-15}	419
BP	ribonucleoprotein complex assembly	GO:0022618	5.30×10^{-14}	40
BP	ribonucleoprotein complex subunit organization	GO:0071826	1.52×10^{-13}	40
BP	cytoplasmic translation	GO:0002181	3.33×10^{-13}	29
BP	protein folding	GO:0006457	3.39×10^{-13}	38
BP	maturation of SSU-rRNA	GO:0030490	1.20×10^{-12}	23
BP	maturation of 5.8S rRNA	GO:0000460	3.94×10^{-12}	19
BP	RNA metabolic process	GO:0016070	4.04×10^{-12}	208
BP	tRNA processing	GO:0008033	6.14×10^{-10}	31
BP	ribosome assembly	GO:0042255	1.50×10^{-9}	23
BP	cytoplasmic translational initiation	GO:0002183	2.34×10^{-9}	15
BP	RNA modification	GO:0009451	2.52×10^{-9}	33
BP	cellular aromatic compound metabolic process	GO:0006725	3.30×10^{-9}	256
BP	heterocycle metabolic process	GO:0046483	3.57×10^{-9}	256
BP	nucleobase-containing compound metabolic process	GO:0006139	3.98×10^{-9}	250
BP	organic cyclic compound metabolic process	GO:1901360	1.35×10^{-8}	257
BP	nucleic acid metabolic process	GO:0090304	2.98×10^{-8}	220
BP	establishment of protein localization to organelle	GO:0072594	2.31×10^{-7}	35
BP	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	GO:0000463	2.43×10^{-7}	11
BP	translational initiation	GO:0006413	4.54×10^{-7}	19
BP	rRNA modification	GO:0000154	4.59×10^{-7}	13
BP	formation of cytoplasmic translation initiation complex	GO:0001732	8.52×10^{-7}	12
BP	cellular nitrogen compound biosynthetic process	GO:0044271	8.58×10^{-7}	189
BP	tRNA aminoacylation for protein translation	GO:0006418	1.11×10^{-6}	20
BP	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	GO:0000462	1.84×10^{-6}	14
BP	cellular biosynthetic process	GO:0044249	3.16×10^{-6}	213
BP	amino acid activation	GO:0043038	8.97×10^{-6}	20
BP	tRNA aminoacylation	GO:0043039	8.97×10^{-6}	20
BP	cleavage involved in rRNA processing	GO:0000469	9.25×10^{-6}	12
BP	non-membrane-bounded organelle assembly	GO:0140694	1.72×10^{-5}	26
BP	maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	GO:0000466	1.98×10^{-5}	11
BP	RNA phosphodiester bond hydrolysis	GO:0090501	2.52×10^{-5}	16
BP	tRNA modification	GO:0006400	3.57×10^{-5}	19
BP	establishment of protein localization to mitochondrion	GO:0072655	5.22×10^{-5}	17
BP	protein localization to mitochondrion	GO:0070585	5.22×10^{-5}	17
BP	protein targeting to mitochondrion	GO:0006626	7.78×10^{-5}	16
BP	rRNA methylation	GO:0031167	1.17×10^{-4}	8
BP	protein localization to organelle	GO:0033365	1.58×10^{-4}	35
BP	biosynthetic process	GO:0009058	2.32×10^{-4}	226

BP	RNA phosphodiester bond hydrolysis, endonucleolytic	GO:0090502	2.67×10^{-4}	9
BP	tricarboxylic acid cycle	GO:0006099	2.80×10^{-4}	12
BP	RNA methylation	GO:0001510	3.30×10^{-4}	15
BP	ribosomal large subunit assembly	GO:0000027	3.30×10^{-4}	11
BP	methylation	GO:0032259	4.54×10^{-4}	26
BP	nucleic acid phosphodiester bond hydrolysis	GO:0090305	5.44×10^{-4}	16
BP	endonucleolytic cleavage of tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	GO:0000479	8.98×10^{-4}	7
BP	endonucleolytic cleavage involved in rRNA processing	GO:0000478	8.98×10^{-4}	7
BP	organic substance biosynthetic process	GO:1901576	9.75×10^{-4}	220
BP	protein insertion into mitochondrial membrane	GO:0051204	1.07×10^{-3}	6
BP	nucleocytoplasmic transport	GO:0006913	1.51×10^{-3}	21
BP	protein import into nucleus	GO:0006606	1.84×10^{-3}	12
BP	protein-containing complex organization	GO:0043933	1.88×10^{-3}	65
BP	nuclear transport	GO:0051169	2.03×10^{-3}	21
BP	macromolecule methylation	GO:0043414	2.03×10^{-3}	21
BP	mitochondrial transport	GO:0006839	2.03×10^{-3}	18
BP	macromolecule biosynthetic process	GO:0009059	2.13×10^{-3}	170
BP	protein-containing complex assembly	GO:0065003	2.76×10^{-3}	60
BP	import into nucleus	GO:0051170	3.17×10^{-3}	12
BP	establishment of protein localization to mitochondrial membrane	GO:0090151	3.63×10^{-3}	7
BP	nitrogen compound metabolic process	GO:0006807	4.42×10^{-3}	445
BP	tRNA methylation	GO:0030488	4.66×10^{-3}	8
BP	mitochondrial transmembrane transport	GO:1990542	4.68×10^{-3}	11
BP	protein localization to nucleus	GO:0034504	5.26×10^{-3}	12
BP	response to heat	GO:0009408	6.76×10^{-3}	6
BP	protein insertion into mitochondrial inner membrane	GO:0045039	9.21×10^{-3}	5
BP	protein targeting	GO:0006605	1.53×10^{-2}	22
BP	intracellular protein transport	GO:0006886	1.63×10^{-2}	47
BP	intracellular protein transmembrane transport	GO:0065002	2.02×10^{-2}	12
BP	protein transmembrane transport	GO:0071806	2.02×10^{-2}	12
BP	response to temperature stimulus	GO:0009266	2.44×10^{-2}	6
BP	rRNA catabolic process	GO:0016075	2.69×10^{-2}	7
BP	ncRNA catabolic process	GO:0034661	2.69×10^{-2}	7
BP	mitochondrial membrane organization	GO:0007006	3.06×10^{-2}	9
CC	nucleolus	GO:0005730	1.45×10^{-56}	88
CC	ribonucleoprotein complex	GO:1990904	4.64×10^{-46}	145
CC	preribosome	GO:0030684	2.35×10^{-40}	53
CC	organelle lumen	GO:0043233	7.27×10^{-39}	154
CC	intracellular organelle lumen	GO:0070013	7.27×10^{-39}	154
CC	membrane-enclosed lumen	GO:0031974	7.27×10^{-39}	154
CC	nuclear lumen	GO:0031981	3.24×10^{-27}	112
CC	small-subunit processome	GO:0032040	1.70×10^{-19}	26
CC	intracellular anatomical structure	GO:0005622	1.90×10^{-19}	562
CC	intracellular non-membrane-bounded organelle	GO:0043232	2.31×10^{-19}	184
CC	non-membrane-bounded organelle	GO:0043228	3.22×10^{-19}	184
CC	ribosome	GO:0005840	1.28×10^{-17}	67
CC	protein-containing complex	GO:0032991	2.77×10^{-14}	291
CC	cytoplasm	GO:0005737	9.43×10^{-14}	366
CC	90S preribosome	GO:0030686	2.13×10^{-13}	19
CC	preribosome, large subunit precursor	GO:0030687	1.12×10^{-11}	17
CC	ribosomal subunit	GO:0044391	1.77×10^{-11}	49
CC	mitochondrion	GO:0005739	5.79×10^{-11}	116
CC	cytosol	GO:0005829	6.94×10^{-9}	88
CC	mitochondrial ribosome	GO:0005761	1.03×10^{-7}	24
CC	organellar ribosome	GO:0000313	1.03×10^{-7}	24
CC	sno(s)RNA-containing ribonucleoprotein complex	GO:0005732	1.33×10^{-7}	13
CC	eukaryotic translation initiation factor 3 complex	GO:0005852	1.33×10^{-6}	11
CC	eukaryotic 48S preinitiation complex	GO:0033290	1.33×10^{-6}	11
CC	large ribosomal subunit	GO:0015934	1.65×10^{-6}	29
CC	mitochondrial matrix	GO:0005759	3.15×10^{-6}	32

CC	eukaryotic 43S preinitiation complex	GO:0016282	5.49×10^{-6}	11
CC	translation preinitiation complex	GO:0070993	5.49×10^{-6}	11
CC	chaperonin-containing T-complex	GO:0005832	1.09×10^{-5}	8
CC	protein folding chaperone complex	GO:0101031	1.09×10^{-5}	8
CC	mitochondrial protein-containing complex	GO:0098798	5.84×10^{-5}	44
CC	intracellular organelle	GO:0043229	1.17×10^{-4}	443
CC	organelle	GO:0043226	2.54×10^{-4}	445
CC	small ribosomal subunit	GO:0015935	5.21×10^{-4}	20
CC	exoribonuclease complex	GO:1905354	6.08×10^{-4}	9
CC	preribosome, small subunit precursor	GO:0030688	6.14×10^{-4}	7
CC	mitochondrial small ribosomal subunit	GO:0005763	6.70×10^{-4}	10
CC	organellar small ribosomal subunit	GO:0000314	6.70×10^{-4}	10
CC	aminoacyl-tRNA synthetase multienzyme complex	GO:0017101	6.82×10^{-4}	6
CC	organellar large ribosomal subunit	GO:0000315	8.64×10^{-4}	14
CC	mitochondrial large ribosomal subunit	GO:0005762	8.64×10^{-4}	14
CC	exosome (RNase complex)	GO:0000178	3.35×10^{-3}	8
CC	RNA polymerase III complex	GO:0005666	3.35×10^{-3}	8
CC	intracellular membrane-bounded organelle	GO:0043231	3.88×10^{-3}	386
CC	ribonuclease P complex	GO:0030677	4.25×10^{-3}	6
CC	nuclear exosome (RNase complex)	GO:0000176	7.29×10^{-3}	7
CC	envelope	GO:0031975	1.02×10^{-2}	51
CC	organelle envelope	GO:0031967	1.02×10^{-2}	51
CC	cytosolic ribosome	GO:0022626	1.06×10^{-2}	23
CC	RNA polymerase I complex	GO:0005736	1.52×10^{-2}	6
CC	endoribonuclease complex	GO:1902555	1.78×10^{-2}	7
CC	membrane-bounded organelle	GO:0043227	1.97×10^{-2}	389
CC	endonuclease complex	GO:1905348	3.81×10^{-2}	7
CC	cytoplasmic exosome (RNase complex)	GO:0000177	4.06×10^{-2}	6
CC	box H/ACA RNP complex	GO:0072588	4.24×10^{-2}	4
CC	box H/ACA snoRNP complex	GO:0031429	4.24×10^{-2}	4
CC	box C/D RNP complex	GO:0031428	4.24×10^{-2}	4
Shared between M24 samples - upregulated				
MF	transition metal ion binding	GO:0046914	6.06×10^{-5}	122
MF	carbohydrate binding	GO:0030246	3.50×10^{-4}	25
MF	hydrolase activity, hydrolyzing O-glycosyl compounds	GO:0004553	2.00×10^{-3}	25
MF	zinc ion binding	GO:0008270	2.70×10^{-3}	86
MF	cation binding	GO:0043169	2.88×10^{-3}	208
MF	metal ion binding	GO:0046872	3.92×10^{-3}	205
MF	galactokinase activity	GO:0004335	5.80×10^{-3}	5
MF	catalytic activity	GO:0003824	1.19×10^{-2}	484
MF	transmembrane receptor protein serine/threonine kinase activity	GO:0004675	3.14×10^{-2}	5
MF	hydrolase activity, acting on glycosyl bonds	GO:0016798	3.34×10^{-2}	25
MF	ligand-activated transcription factor activity	GO:0098531	4.13×10^{-2}	6
MF	phosphatidylinositol-3-phosphate binding	GO:0032266	4.13×10^{-2}	6
MF	ion binding	GO:0043167	4.30×10^{-2}	315
MF	transferase activity	GO:0016740	4.85×10^{-2}	160
BP	catabolic process	GO:0009056	3.26×10^{-6}	105
BP	process utilizing autophagic mechanism	GO:0061919	1.02×10^{-4}	19
BP	autophagy	GO:0006914	1.02×10^{-4}	19
BP	carbohydrate metabolic process	GO:0005975	8.46×10^{-4}	45
BP	hexose metabolic process	GO:0019318	5.33×10^{-3}	14
BP	monosaccharide metabolic process	GO:0005996	6.20×10^{-3}	15
BP	peroxisome organization	GO:0007031	7.18×10^{-3}	8
BP	organic substance catabolic process	GO:1901575	7.76×10^{-3}	86
BP	autophagy of mitochondrion	GO:0000422	2.94×10^{-2}	8
BP	mitochondrion disassembly	GO:0061726	2.94×10^{-2}	8
BP	lipid metabolic process	GO:0006629	4.67×10^{-2}	58
CC	microbody	GO:0042579	1.73×10^{-4}	17
CC	peroxisome	GO:0005777	1.73×10^{-4}	17
CC	vacuole	GO:0005773	1.51×10^{-3}	23
CC	lysosome	GO:0005764	1.80×10^{-3}	17

CC	lytic vacuole	GO:0000323	1.80×10^{-3}	17
CC	Golgi apparatus	GO:0005794	2.61×10^{-3}	39
CC	endomembrane system	GO:0012505	8.23×10^{-3}	83
CC	Golgi membrane	GO:0000139	1.77×10^{-2}	19
CC	bounding membrane of organelle	GO:0098588	2.04×10^{-2}	38

IM24 = P. falciparum*-infected midguts at 24 hours post-infection, CM24 = control midguts at 24 hours post-blood meal; M24 = midguts (infected and control) at 24 hours; MF = molecular function, BP = biological process, CC = cellular component;* ∩ *= intersection,* i.e. *the number of overlapping transcripts belonging to a list of transcripts from a comparison (of a Venn diagram) and to a GO/KEGG term; adjusted* p *= false discovery rate-adjusted* p*-values of Fisher's one-tailed tests, significant if adjusted* p < *0.05.*

Table B.3. Significantly enriched GO terms from DE transcripts unique to or shared between infected and control carcass samples at 7 days. *Significantly DE transcripts were identified from comparisons of* P. falciparum*-infected/control midgut samples at 7 days compared to at 1 hour of the same infection status. GO terms were identified using g:Profiler (Raudvere et al, 2019).*

Source	GO Term	Term ID	Adjusted *p*	∩
Unique to CX7 - downregulated				
MF	structural constituent of ribosome	GO:0003735	4.49×10^{-22}	43
MF	structural molecule activity	GO:0005198	1.41×10^{-10}	53
MF	prostaglandin-E synthase activity	GO:0050220	7.32×10^{-4}	4
MF	rRNA binding	GO:0019843	1.89×10^{-3}	9
BP	translation	GO:0006412	1.69×10^{-17}	56
BP	peptide biosynthetic process	GO:0043043	2.76×10^{-17}	56
BP	amide biosynthetic process	GO:0043604	1.02×10^{-16}	57
BP	peptide metabolic process	GO:0006518	2.50×10^{-16}	58
BP	amide metabolic process	GO:0043603	1.85×10^{-13}	59
BP	organonitrogen compound biosynthetic process	GO:1901566	1.51×10^{-9}	70
BP	ribosomal small subunit assembly	GO:0000028	1.63×10^{-8}	10
BP	organonitrogen compound metabolic process	GO:1901564	1.45×10^{-6}	148
BP	ribosome assembly	GO:0042255	6.54×10^{-6}	13
BP	protein metabolic process	GO:0019538	1.07×10^{-5}	126
BP	ribosomal small subunit biogenesis	GO:0042274	3.22×10^{-5}	13
BP	ribosome biogenesis	GO:0042254	1.95×10^{-4}	25
BP	ribonucleoprotein complex biogenesis	GO:0022613	4.26×10^{-4}	28
BP	ribonucleoprotein complex assembly	GO:0022618	2.30×10^{-3}	15
BP	ribonucleoprotein complex subunit organization	GO:0071826	3.12×10^{-3}	15
BP	non-membrane-bounded organelle assembly	GO:0140694	8.69×10^{-3}	13
BP	cytoplasmic translation	GO:0002181	2.62×10^{-2}	10
BP	prostaglandin metabolic process	GO:0006693	4.99×10^{-2}	3
BP	prostanoid biosynthetic process	GO:0046457	4.99×10^{-2}	3
BP	prostanoid metabolic process	GO:0006692	4.99×10^{-2}	3
BP	prostaglandin biosynthetic process	GO:0001516	4.99×10^{-2}	3
CC	ribosome	GO:0005840	2.78×10^{-24}	48
CC	cytosolic ribosome	GO:0022626	1.35×10^{-18}	30
CC	ribosomal subunit	GO:0044391	2.81×10^{-18}	37
CC	cytoplasm	GO:0005737	2.53×10^{-16}	165
CC	cytosolic large ribosomal subunit	GO:0022625	2.61×10^{-10}	17
CC	small ribosomal subunit	GO:0015935	1.20×10^{-9}	18
CC	cytosol	GO:0005829	1.79×10^{-9}	47
CC	ribonucleoprotein complex	GO:1990904	1.57×10^{-8}	44
CC	large ribosomal subunit	GO:0015934	9.58×10^{-8}	19
CC	cytosolic small ribosomal subunit	GO:0022627	2.85×10^{-7}	13
CC	mitochondrion	GO:0005739	3.98×10^{-4}	45
CC	intracellular non-membrane-bounded organelle	GO:0043232	1.97×10^{-3}	60
CC	non-membrane-bounded organelle	GO:0043228	2.14×10^{-3}	60
CC	endoplasmic reticulum	GO:0005783	1.43×10^{-2}	24
CC	mitochondrial matrix	GO:0005759	3.16×10^{-2}	13
Unique to CX7 - upregulated				
MF	protein binding	GO:0005515	9.35×10^{-7}	75
MF	binding	GO:0005488	1.58×10^{-4}	138
MF	kinase activity	GO:0016301	9.52×10^{-4}	20
MF	protein serine/threonine kinase activity	GO:0004674	1.64×10^{-3}	12
MF	histone modifying activity	GO:0140993	1.92×10^{-3}	7
MF	purine nucleotide binding	GO:0017076	2.29×10^{-3}	38
MF	transferase activity, transferring phosphorus-containing groups	GO:0016772	2.46×10^{-3}	22
MF	phosphotransferase activity, alcohol group as acceptor	GO:0016773	2.92×10^{-3}	18
MF	nucleotide binding	GO:0000166	7.82×10^{-3}	40
MF	nucleoside phosphate binding	GO:1901265	7.82×10^{-3}	40
MF	ribonucleotide binding	GO:0032553	1.15×10^{-2}	35
MF	protein kinase activity	GO:0004672	1.21×10^{-2}	15
MF	purine ribonucleotide binding	GO:0032555	2.14×10^{-2}	34
MF	adenyl nucleotide binding	GO:0030554	3.10×10^{-2}	31
MF	small molecule binding	GO:0036094	3.44×10^{-2}	40
MF	purine ribonucleoside triphosphate binding	GO:0035639	3.75×10^{-2}	33
BP	protein modification process	GO:0036211	2.43×10^{-4}	36
BP	macromolecule modification	GO:0043412	7.94×10^{-4}	37
BP	phosphorylation	GO:0016310	9.41×10^{-4}	19
BP	intracellular signal transduction	GO:0035556	1.80×10^{-3}	19
BP	phosphate-containing compound metabolic process	GO:0006796	2.42×10^{-3}	30
BP	phosphorus metabolic process	GO:0006793	2.95×10^{-3}	30
BP	biological regulation	GO:0065007	5.10×10^{-3}	62

BP	protein phosphorylation	GO:0006468	6.34×10^{-3}	17
BP	signal transduction	GO:0007165	1.03×10^{-2}	34
BP	regulation of biological process	GO:0050789	1.37×10^{-2}	59
BP	cellular response to stimulus	GO:0051716	1.94×10^{-2}	39
BP	signaling	GO:0023052	2.97×10^{-2}	34
BP	cell communication	GO:0007154	3.60×10^{-2}	34
BP	negative regulation of canonical Wnt signaling pathway	GO:0090090	3.82×10^{-2}	3
BP	Rho protein signal transduction	GO:0007266	4.88×10^{-2}	4
Unique to IX7 - downregulated				
MF	molecular transducer activity	GO:0060089	3.80×10^{-9}	67
MF	signaling receptor activity	GO:0038023	3.80×10^{-9}	67
MF	G protein-coupled receptor activity	GO:0004930	1.45×10^{-8}	34
MF	transmembrane transporter activity	GO:0022857	2.30×10^{-7}	99
MF	monoatomic ion channel activity	GO:0005216	2.61×10^{-7}	43
MF	transmembrane signaling receptor activity	GO:0004888	2.91×10^{-7}	58
MF	monoatomic ion transmembrane transporter activity	GO:0015075	3.05×10^{-7}	61
MF	transporter activity	GO:0005215	5.88×10^{-7}	101
MF	channel activity	GO:0015267	8.94×10^{-7}	44
MF	passive transmembrane transporter activity	GO:0022803	8.94×10^{-7}	44
MF	gated channel activity	GO:0022836	2.07×10^{-6}	28
MF	monoatomic ion gated channel activity	GO:0022839	2.07×10^{-6}	28
MF	neurotransmitter receptor activity	GO:0030594	3.45×10^{-6}	15
MF	monoatomic cation transmembrane transporter activity	GO:0008324	6.99×10^{-5}	46
MF	monoatomic cation channel activity	GO:0005261	1.27×10^{-4}	28
MF	salt transmembrane transporter activity	GO:1901702	3.60×10^{-4}	36
MF	transmitter-gated monoatomic ion channel activity	GO:0022824	3.99×10^{-4}	11
MF	transmitter-gated channel activity	GO:0022835	3.99×10^{-4}	11
MF	DNA-binding transcription factor activity, RNA polymerase II-specific	GO:0000981	1.16×10^{-3}	29
MF	calcium ion binding	GO:0005509	1.86×10^{-3}	35
MF	transmembrane receptor protein tyrosine kinase activity	GO:0004714	2.01×10^{-3}	7
MF	ligand-gated channel activity	GO:0022834	2.42×10^{-3}	18
MF	ligand-gated monoatomic ion channel activity	GO:0015276	2.42×10^{-3}	18
MF	extracellular ligand-gated monoatomic ion channel activity	GO:0005230	2.80×10^{-3}	14
MF	DNA-binding transcription factor activity	GO:0003700	3.77×10^{-3}	47
MF	molecular function regulator activity	GO:0098772	4.40×10^{-3}	50
MF	inorganic molecular entity transmembrane transporter activity	GO:0015318	5.62×10^{-3}	43
MF	metal ion transmembrane transporter activity	GO:0046873	5.90×10^{-3}	28
MF	acetylcholine-gated monoatomic cation-selective channel activity	GO:0022848	6.42×10^{-3}	6
MF	excitatory extracellular ligand-gated monoatomic ion channel activity	GO:0005231	6.42×10^{-3}	6
MF	guanyl-nucleotide exchange factor activity	GO:0005085	6.54×10^{-3}	16
MF	peptide receptor activity	GO:0001653	7.65×10^{-3}	11
MF	G protein-coupled peptide receptor activity	GO:0008528	1.64×10^{-2}	10
MF	ligand-gated monoatomic cation channel activity	GO:0099094	2.27×10^{-2}	8
MF	inorganic cation transmembrane transporter activity	GO:0022890	2.52×10^{-2}	37
MF	protein tyrosine kinase activity	GO:0004713	3.19×10^{-2}	10
MF	transcription regulator activity	GO:0140110	4.17×10^{-2}	53
BP	biological regulation	GO:0065007	1.01×10^{-9}	226
BP	regulation of cellular process	GO:0050794	2.54×10^{-9}	206
BP	regulation of biological process	GO:0050789	4.22×10^{-9}	217
BP	G protein-coupled receptor signaling pathway	GO:0007186	2.77×10^{-8}	39
BP	transmembrane transport	GO:0055085	7.61×10^{-8}	81
BP	monoatomic ion transport	GO:0006811	1.15×10^{-7}	55
BP	cell communication	GO:0007154	1.38×10^{-7}	120
BP	signaling	GO:0023052	1.57×10^{-7}	119
BP	signal transduction	GO:0007165	1.14×10^{-6}	112
BP	monoatomic ion transmembrane transport	GO:0034220	1.30×10^{-5}	34
BP	cell adhesion	GO:0007155	4.99×10^{-5}	26
BP	cell-cell signaling	GO:0007267	7.62×10^{-5}	26
BP	anatomical structure development	GO:0048856	2.13×10^{-3}	39
BP	inorganic ion transmembrane transport	GO:0098660	2.17×10^{-3}	25
BP	chemical synaptic transmission	GO:0007268	2.33×10^{-3}	17
BP	anterograde trans-synaptic signaling	GO:0098916	2.33×10^{-3}	17
BP	synaptic signaling	GO:0099536	2.97×10^{-3}	17
BP	trans-synaptic signaling	GO:0099537	2.97×10^{-3}	17
BP	multicellular organism development	GO:0007275	3.52×10^{-3}	31
BP	metal ion transport	GO:0030001	9.72×10^{-3}	28
BP	monoatomic cation transmembrane transport	GO:0098655	1.10×10^{-2}	22
BP	inorganic cation transmembrane transport	GO:0098662	1.52×10^{-2}	22
BP	monoatomic cation transport	GO:0006812	1.60×10^{-2}	33
BP	cell junction organization	GO:0034330	3.08×10^{-2}	10

BP	cellular response to stimulus	GO:0051716	3.94×10^{-2}	115
CC	cell junction	GO:0030054	1.03×10^{-11}	35
CC	synapse	GO:0045202	2.06×10^{-10}	28
CC	plasma membrane	GO:0005886	4.67×10^{-8}	68
CC	membrane	GO:0016020	3.82×10^{-7}	145
CC	cell periphery	GO:0071944	1.81×10^{-6}	79
CC	plasma membrane protein complex	GO:0098797	6.24×10^{-6}	19
CC	monoatomic ion channel complex	GO:0034702	2.06×10^{-4}	10
CC	cation channel complex	GO:0034703	2.91×10^{-4}	9
CC	synaptic membrane	GO:0097060	6.66×10^{-4}	11
CC	neuron projection	GO:0043005	8.22×10^{-4}	22
CC	plasma membrane region	GO:0098590	1.38×10^{-3}	14
CC	postsynaptic membrane	GO:0045211	1.71×10^{-3}	10
CC	postsynapse	GO:0098794	4.75×10^{-3}	10
CC	collagen-containing extracellular matrix	GO:0062023	4.71×10^{-2}	4
Unique to IX7 - upregulated				
MF	nucleic acid binding	GO:0003676	2.57×10^{-9}	106
MF	RNA polymerase core enzyme binding	GO:0043175	4.44×10^{-4}	6
MF	RNA polymerase binding	GO:0070063	8.00×10^{-4}	6
MF	catalytic activity, acting on a nucleic acid	GO:0140640	1.54×10^{-3}	29
MF	transcription coregulator activity	GO:0003712	3.05×10^{-3}	13
MF	RNA polymerase II complex binding	GO:0000993	6.28×10^{-3}	5
MF	heterocyclic compound binding	GO:1901363	1.01×10^{-2}	130
MF	organic cyclic compound binding	GO:0097159	1.54×10^{-2}	130
MF	basal transcription machinery binding	GO:0001098	1.65×10^{-2}	5
MF	basal RNA polymerase II transcription machinery binding	GO:0001099	1.65×10^{-2}	5
MF	RNA binding	GO:0003723	1.79×10^{-2}	36
BP	nucleic acid metabolic process	GO:0090304	3.92×10^{-12}	108
BP	RNA metabolic process	GO:0016070	2.39×10^{-10}	94
BP	cellular aromatic compound metabolic process	GO:0006725	9.70×10^{-10}	115
BP	nucleobase-containing compound metabolic process	GO:0006139	2.11×10^{-9}	112
BP	heterocycle metabolic process	GO:0046483	2.57×10^{-9}	114
BP	organic cyclic compound metabolic process	GO:1901360	3.23×10^{-9}	115
BP	cellular nitrogen compound metabolic process	GO:0034641	3.14×10^{-6}	117
BP	mRNA processing	GO:0006397	5.45×10^{-6}	25
BP	mRNA metabolic process	GO:0016071	2.38×10^{-5}	28
BP	RNA processing	GO:0006396	4.24×10^{-5}	40
BP	macromolecule metabolic process	GO:0043170	9.61×10^{-4}	163
BP	gene expression	GO:0010467	1.04×10^{-3}	91
BP	mRNA 3'-end processing	GO:0031124	2.54×10^{-3}	7
BP	nitrogen compound metabolic process	GO:0006807	4.80×10^{-3}	174
BP	RNA splicing	GO:0008380	5.01×10^{-3}	17
BP	DNA damage response	GO:0006974	1.80×10^{-2}	19
BP	cellular response to stress	GO:0033554	2.47×10^{-2}	25
BP	RNA localization	GO:0006403	4.25×10^{-2}	8
CC	nuclear protein-containing complex	GO:0140513	7.09×10^{-17}	65
CC	nucleus	GO:0005634	1.18×10^{-13}	134
CC	intracellular anatomical structure	GO:0005622	6.91×10^{-9}	213
CC	membrane-bounded organelle	GO:0043227	1.23×10^{-8}	175
CC	protein-containing complex	GO:0032991	2.55×10^{-8}	119
CC	intracellular membrane-bounded organelle	GO:0043231	9.54×10^{-8}	170
CC	intracellular organelle	GO:0043229	5.10×10^{-6}	182
CC	organelle	GO:0043226	6.55×10^{-6}	183
CC	intracellular protein-containing complex	GO:0140535	8.54×10^{-6}	38
CC	transferase complex, transferring phosphorus-containing groups	GO:0061695	4.89×10^{-4}	16
CC	mediator complex	GO:0016592	2.03×10^{-3}	8
CC	nucleoplasm	GO:0005654	2.22×10^{-3}	20
CC	transferase complex	GO:1990234	2.70×10^{-3}	28
CC	nuclear DNA-directed RNA polymerase complex	GO:0055029	3.32×10^{-3}	11
CC	DNA-directed RNA polymerase complex	GO:0000428	3.32×10^{-3}	11
CC	RNA polymerase II transcription regulator complex	GO:0090575	3.37×10^{-3}	10
CC	RNA polymerase complex	GO:0030880	4.80×10^{-3}	11
CC	spliceosomal complex	GO:0005681	4.95×10^{-3}	12
CC	U5 snRNP	GO:0005682	5.82×10^{-3}	5
CC	catalytic complex	GO:1902494	6.80×10^{-3}	49
CC	peptidase complex	GO:1905368	8.06×10^{-3}	11
CC	U1 snRNP	GO:0005685	1.69×10^{-2}	5
CC	transcription regulator complex	GO:0005667	2.05×10^{-2}	13
CC	small nuclear ribonucleoprotein complex	GO:0030532	3.19×10^{-2}	8
CC	spliceosomal snRNP complex	GO:0097525	3.19×10^{-2}	8

CC	proteasome regulatory particle	GO:0005838	3.86×10^{-2}	6
CC	spliceosomal tri-snRNP complex	GO:0097526	3.86×10^{-2}	6
CC	U2 snRNP	GO:0005686	3.96×10^{-2}	5
CC	nuclear lumen	GO:0031981	4.16×10^{-2}	28
CC	RNA polymerase II, holoenzyme	GO:0016591	4.69×10^{-2}	8
CC	Sm-like protein family complex	GO:0120114	4.69×10^{-2}	8
Shared between X7 samples - downregulated				
MF	transmembrane transporter activity	GO:0022857	2.61×10^{-40}	221
MF	transporter activity	GO:0005215	8.69×10^{-37}	221
MF	active transmembrane transporter activity	GO:0022804	7.16×10^{-28}	100
MF	primary active transmembrane transporter activity	GO:0015399	2.68×10^{-27}	72
MF	oxidoreduction-driven active transmembrane transporter activity	GO:0015453	5.08×10^{-24}	36
MF	inorganic molecular entity transmembrane transporter activity	GO:0015318	4.94×10^{-23}	105
MF	electron transfer activity	GO:0009055	4.90×10^{-21}	46
MF	inorganic cation transmembrane transporter activity	GO:0022890	6.07×10^{-21}	93
MF	monoatomic ion transmembrane transporter activity	GO:0015075	3.23×10^{-20}	114
MF	proton transmembrane transporter activity	GO:0015078	2.99×10^{-19}	50
MF	monoatomic cation transmembrane transporter activity	GO:0008324	3.89×10^{-18}	93
MF	NADH dehydrogenase activity	GO:0003954	1.04×10^{-13}	22
MF	oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor	GO:0016655	6.41×10^{-13}	21
MF	NAD(P)H dehydrogenase (quinone) activity	GO:0003955	6.41×10^{-13}	21
MF	NADH dehydrogenase (ubiquinone) activity	GO:0008137	6.41×10^{-13}	21
MF	NADH dehydrogenase (quinone) activity	GO:0050136	6.41×10^{-13}	21
MF	oxidoreductase activity	GO:0016491	1.64×10^{-12}	158
MF	active monoatomic ion transmembrane transporter activity	GO:0022853	4.87×10^{-12}	46
MF	oxidoreductase activity, acting on NAD(P)H	GO:0016651	1.36×10^{-11}	26
MF	ATPase-coupled transmembrane transporter activity	GO:0042626	5.84×10^{-8}	36
MF	monoatomic ion channel activity	GO:0005216	6.82×10^{-8}	60
MF	salt transmembrane transporter activity	GO:1901702	7.64×10^{-8}	59
MF	channel activity	GO:0015267	7.68×10^{-8}	63
MF	passive transmembrane transporter activity	GO:0022803	7.68×10^{-8}	63
MF	pyrophosphate hydrolysis-driven proton transmembrane transporter activity	GO:0009678	7.94×10^{-8}	16
MF	monoatomic cation channel activity	GO:0005261	3.85×10^{-7}	43
MF	proton-transporting ATPase activity, rotational mechanism	GO:0046961	4.22×10^{-7}	15
MF	ATPase activity, coupled to transmembrane movement of ions, rotational mechanism	GO:0044769	4.22×10^{-7}	15
MF	ATPase-coupled ion transmembrane transporter activity	GO:0042625	4.22×10^{-7}	15
MF	proton-transporting ATP synthase activity, rotational mechanism	GO:0046933	1.28×10^{-6}	12
MF	oxidoreductase activity, acting on a heme group of donors	GO:0016675	1.35×10^{-6}	11
MF	cytochrome-c oxidase activity	GO:0004129	1.35×10^{-6}	11
MF	ATPase-coupled monoatomic cation transmembrane transporter activity	GO:0019829	2.28×10^{-6}	17
MF	potassium ion transmembrane transporter activity	GO:0015079	2.06×10^{-5}	21
MF	proton channel activity	GO:0015252	3.10×10^{-5}	13
MF	potassium channel activity	GO:0005267	7.41×10^{-5}	17
MF	voltage-gated monoatomic ion channel activity	GO:0005244	2.29×10^{-4}	17
MF	voltage-gated channel activity	GO:0022832	2.29×10^{-4}	17
MF	metal ion transmembrane transporter activity	GO:0046873	4.89×10^{-4}	42
MF	voltage-gated monoatomic cation channel activity	GO:0022843	1.21×10^{-3}	14
MF	organic acid transmembrane transporter activity	GO:0005342	1.31×10^{-3}	16
MF	carboxylic acid transmembrane transporter activity	GO:0046943	1.31×10^{-3}	16
MF	catalytic activity	GO:0003824	2.26×10^{-3}	601
MF	monoatomic ion gated channel activity	GO:0022839	2.73×10^{-3}	31
MF	gated channel activity	GO:0022836	2.73×10^{-3}	31
MF	vitamin binding	GO:0019842	2.86×10^{-3}	21
MF	secondary active transmembrane transporter activity	GO:0015291	7.94×10^{-3}	28
MF	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	GO:0016624	1.26×10^{-2}	6
MF	vitamin B6 binding	GO:0070279	1.52×10^{-2}	15
MF	oxidoreductase activity, acting on the aldehyde or oxo group of donors	GO:0016903	1.52×10^{-2}	15
MF	pyridoxal phosphate binding	GO:0030170	1.52×10^{-2}	15
MF	organic anion transmembrane transporter activity	GO:0008514	3.25×10^{-2}	20
MF	ABC-type transporter activity	GO:0140359	3.56×10^{-2}	19
MF	acyl-CoA dehydrogenase activity	GO:0003995	4.41×10^{-2}	6
MF	oxidoreductase activity, acting on the CH-CH group of donors, with a flavin as acceptor	GO:0052890	4.41×10^{-2}	6
BP	energy derivation by oxidation of organic compounds	GO:0015980	2.37×10^{-40}	66
BP	cellular respiration	GO:0045333	1.62×10^{-37}	59

BP	aerobic respiration	GO:0009060	3.25×10^{-34}	54
BP	generation of precursor metabolites and energy	GO:0006091	3.72×10^{-32}	71
BP	electron transport chain	GO:0022900	3.34×10^{-29}	45
BP	respiratory electron transport chain	GO:0022904	1.41×10^{-27}	41
BP	oxidative phosphorylation	GO:0006119	4.08×10^{-25}	38
BP	transmembrane transport	GO:0055085	2.30×10^{-24}	157
BP	ATP synthesis coupled electron transport	GO:0042773	1.75×10^{-23}	36
BP	mitochondrial ATP synthesis coupled electron transport	GO:0042775	8.98×10^{-23}	33
BP	aerobic electron transport chain	GO:0019646	8.98×10^{-23}	33
BP	small molecule metabolic process	GO:0044281	3.91×10^{-18}	154
BP	monoatomic ion transmembrane transport	GO:0034220	3.33×10^{-16}	65
BP	monoatomic ion transport	GO:0006811	8.75×10^{-15}	93
BP	inorganic ion transmembrane transport	GO:0098660	2.46×10^{-14}	53
BP	organic acid metabolic process	GO:0006082	3.83×10^{-14}	92
BP	oxoacid metabolic process	GO:0043436	7.51×10^{-14}	91
BP	carboxylic acid metabolic process	GO:0019752	1.47×10^{-13}	90
BP	inorganic cation transmembrane transport	GO:0098662	2.07×10^{-13}	49
BP	monoatomic cation transmembrane transport	GO:0098655	4.48×10^{-13}	48
BP	ATP metabolic process	GO:0046034	1.06×10^{-11}	22
BP	ribose phosphate biosynthetic process	GO:0046390	6.12×10^{-11}	37
BP	ribose phosphate metabolic process	GO:0019693	8.10×10^{-11}	47
BP	ATP biosynthetic process	GO:0006754	1.29×10^{-10}	19
BP	proton motive force-driven ATP synthesis	GO:0015986	1.29×10^{-10}	19
BP	ribonucleotide biosynthetic process	GO:0009260	1.32×10^{-10}	36
BP	ribonucleotide metabolic process	GO:0009259	1.71×10^{-10}	46
BP	purine ribonucleotide biosynthetic process	GO:0009152	2.02×10^{-10}	34
BP	purine ribonucleotide metabolic process	GO:0009150	3.39×10^{-10}	44
BP	purine nucleoside triphosphate metabolic process	GO:0009144	3.73×10^{-10}	22
BP	purine ribonucleoside triphosphate metabolic process	GO:0009205	3.73×10^{-10}	22
BP	ribonucleoside triphosphate metabolic process	GO:0009199	5.47×10^{-10}	23
BP	nucleoside triphosphate biosynthetic process	GO:0009142	6.70×10^{-10}	21
BP	nucleoside triphosphate metabolic process	GO:0009141	7.01×10^{-10}	24
BP	proton transmembrane transport	GO:1902600	7.01×10^{-10}	24
BP	purine nucleotide biosynthetic process	GO:0006164	1.06×10^{-9}	36
BP	purine-containing compound biosynthetic process	GO:0072522	1.53×10^{-9}	38
BP	purine nucleoside triphosphate biosynthetic process	GO:0009145	1.95×10^{-9}	19
BP	purine ribonucleoside triphosphate biosynthetic process	GO:0009206	1.95×10^{-9}	19
BP	ribonucleoside triphosphate biosynthetic process	GO:0009201	3.55×10^{-9}	20
BP	mitochondrial electron transport, NADH to ubiquinone	GO:0006120	3.76×10^{-9}	13
BP	tricarboxylic acid cycle	GO:0006099	4.64×10^{-9}	17
BP	nucleoside phosphate biosynthetic process	GO:1901293	6.79×10^{-9}	40
BP	monoatomic cation transport	GO:0006812	7.67×10^{-9}	63
BP	nucleotide biosynthetic process	GO:0009165	2.12×10^{-8}	39
BP	purine-containing compound metabolic process	GO:0072521	4.09×10^{-8}	52
BP	nucleobase-containing small molecule metabolic process	GO:0055086	5.83×10^{-8}	61
BP	purine nucleotide metabolic process	GO:0006163	1.10×10^{-7}	48
BP	pyruvate metabolic process	GO:0006090	5.55×10^{-7}	15
BP	nucleoside phosphate metabolic process	GO:0006753	5.89×10^{-7}	52
BP	nucleotide metabolic process	GO:0009117	5.89×10^{-7}	52
BP	amino acid catabolic process	GO:0009063	2.41×10^{-6}	21
BP	potassium ion transport	GO:0006813	4.79×10^{-6}	25
BP	alpha-amino acid metabolic process	GO:1901605	7.26×10^{-6}	33
BP	small molecule catabolic process	GO:0044282	1.33×10^{-5}	34
BP	mitochondrial electron transport, cytochrome c to oxygen	GO:0006123	1.37×10^{-5}	10
BP	organophosphate metabolic process	GO:0019637	2.06×10^{-5}	75
BP	carboxylic acid catabolic process	GO:0046395	2.15×10^{-5}	27
BP	organic acid catabolic process	GO:0016054	2.15×10^{-5}	27
BP	dicarboxylic acid metabolic process	GO:0043648	2.80×10^{-5}	16
BP	glycolytic process	GO:0006096	7.18×10^{-5}	10
BP	alpha-amino acid catabolic process	GO:1901606	7.93×10^{-5}	18
BP	transport	GO:0006810	9.07×10^{-5}	212
BP	monocarboxylic acid metabolic process	GO:0032787	1.25×10^{-4}	35
BP	amino acid metabolic process	GO:0006520	1.37×10^{-4}	44
BP	establishment of localization	GO:0051234	2.18×10^{-4}	212
BP	organophosphate biosynthetic process	GO:0090407	4.11×10^{-4}	51
BP	mitochondrial electron transport, ubiquinol to cytochrome c	GO:0006122	5.85×10^{-4}	8
BP	mitochondrion organization	GO:0007005	2.79×10^{-3}	35
BP	mitochondrial respiratory chain complex assembly	GO:0033108	3.89×10^{-3}	15
BP	localization	GO:0051179	4.77×10^{-3}	213
BP	carbohydrate derivative metabolic process	GO:1901135	5.42×10^{-3}	74

BP	carbohydrate metabolic process	GO:0005975	6.53×10^{-3}	51
BP	metal ion transport	GO:0030001	9.32×10^{-3}	40
BP	potassium ion transmembrane transport	GO:0071805	9.71×10^{-3}	15
BP	purine nucleoside bisphosphate biosynthetic process	GO:0034033	2.17×10^{-2}	8
BP	ribonucleoside bisphosphate biosynthetic process	GO:0034030	2.17×10^{-2}	8
BP	nucleoside bisphosphate biosynthetic process	GO:0033866	2.17×10^{-2}	8
BP	acyl-CoA biosynthetic process	GO:0071616	2.38×10^{-2}	6
BP	thioester biosynthetic process	GO:0035384	2.38×10^{-2}	6
BP	acetyl-CoA biosynthetic process	GO:0006085	2.38×10^{-2}	6
BP	carboxylic acid transport	GO:0046942	4.44×10^{-2}	13
BP	organic acid transport	GO:0015849	4.44×10^{-2}	13
CC	mitochondrial inner membrane	GO:0005743	2.04×10^{-44}	93
CC	organelle inner membrane	GO:0019866	2.72×10^{-43}	93
CC	inner mitochondrial membrane protein complex	GO:0098800	5.08×10^{-43}	71
CC	mitochondrion	GO:0005739	2.62×10^{-42}	179
CC	respirasome	GO:0070469	1.75×10^{-41}	58
CC	oxidoreductase complex	GO:1990204	6.59×10^{-40}	54
CC	respiratory chain complex	GO:0098803	8.06×10^{-40}	55
CC	mitochondrial membrane	GO:0031966	3.32×10^{-39}	98
CC	mitochondrial respirasome	GO:0005746	4.79×10^{-39}	53
CC	mitochondrial envelope	GO:0005740	8.43×10^{-38}	101
CC	mitochondrial protein-containing complex	GO:0098798	1.73×10^{-32}	86
CC	envelope	GO:0031975	6.70×10^{-29}	102
CC	organelle envelope	GO:0031967	6.70×10^{-29}	102
CC	transmembrane transporter complex	GO:1902495	3.55×10^{-28}	52
CC	membrane protein complex	GO:0098796	5.51×10^{-26}	115
CC	transporter complex	GO:1990351	6.81×10^{-26}	52
CC	mitochondrial respiratory chain complex I	GO:0005747	5.14×10^{-23}	32
CC	respiratory chain complex I	GO:0045271	5.14×10^{-23}	32
CC	NADH dehydrogenase complex	GO:0030964	5.14×10^{-23}	32
CC	membrane	GO:0016020	7.51×10^{-21}	291
CC	proton-transporting two-sector ATPase complex	GO:0016469	4.97×10^{-20}	33
CC	organelle membrane	GO:0031090	3.85×10^{-15}	130
CC	cytoplasm	GO:0005737	9.04×10^{-14}	384
CC	cytochrome complex	GO:0070069	5.65×10^{-13}	20
CC	proton-transporting ATP synthase complex	GO:0045259	1.18×10^{-10}	18
CC	proton-transporting two-sector ATPase complex, catalytic domain	GO:0033178	1.24×10^{-9}	15
CC	proton-transporting V-type ATPase complex	GO:0033176	6.51×10^{-9}	15
CC	ATPase complex	GO:1904949	1.10×10^{-8}	33
CC	respiratory chain complex IV	GO:0045277	5.45×10^{-8}	13
CC	mitochondrial proton-transporting ATP synthase complex	GO:0005753	2.55×10^{-7}	13
CC	mitochondrial respiratory chain complex IV	GO:0005751	4.02×10^{-7}	11
CC	proton-transporting two-sector ATPase complex, proton-transporting domain	GO:0033177	1.57×10^{-6}	14
CC	catalytic complex	GO:1902494	1.89×10^{-5}	124
CC	proton-transporting V-type ATPase, V1 domain	GO:0033180	1.20×10^{-4}	7
CC	proton-transporting ATP synthase complex, catalytic core F(1)	GO:0045261	1.27×10^{-4}	8
CC	tricarboxylic acid cycle enzyme complex	GO:0045239	1.27×10^{-4}	8
CC	respiratory chain complex III	GO:0045275	8.49×10^{-4}	7
CC	mitochondrial respiratory chain complex III	GO:0005750	8.49×10^{-4}	7
CC	dihydrolipoyl dehydrogenase complex	GO:0045240	8.97×10^{-4}	6
CC	proton-transporting ATP synthase complex, coupling factor F(o)	GO:0045263	1.82×10^{-3}	8
CC	plasma membrane protein complex	GO:0098797	2.41×10^{-3}	22
CC	mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	GO:0000276	3.37×10^{-3}	7
CC	vacuolar proton-transporting V-type ATPase complex	GO:0016471	9.94×10^{-3}	7
CC	contractile fiber	GO:0043292	1.01×10^{-2}	9
CC	myofibril	GO:0030016	2.27×10^{-2}	8
CC	sarcomere	GO:0030017	2.41×10^{-2}	7
CC	mitochondrial matrix	GO:0005759	3.58×10^{-2}	25
CC	oxoglutarate dehydrogenase complex	GO:0045252	4.99×10^{-2}	4
KEGG	Metabolic pathways	KEGG:01100	2.86×10^{-2}	12
KEGG	Oxidative phosphorylation	KEGG:00190	2.86×10^{-2}	12
Shared between X7 samples - upregulated				
MF	nucleic acid binding	GO:0003676	4.18×10^{-79}	437
MF	heterocyclic compound binding	GO:1901363	2.13×10^{-47}	541
MF	organic cyclic compound binding	GO:0097159	4.22×10^{-46}	541
MF	binding	GO:0005488	9.61×10^{-43}	860
MF	DNA binding	GO:0003677	8.26×10^{-36}	190
MF	zinc ion binding	GO:0008270	5.45×10^{-31}	153

MF	catalytic activity, acting on DNA	GO:0140097	1.13×10^{-24}	59
MF	ATP-dependent activity, acting on DNA	GO:0008094	1.04×10^{-17}	36
MF	transition metal ion binding	GO:0046914	3.24×10^{-15}	158
MF	single-stranded DNA binding	GO:0003697	2.10×10^{-14}	26
MF	ion binding	GO:0043167	1.06×10^{-11}	381
MF	catalytic activity, acting on a nucleic acid	GO:0140640	1.65×10^{-11}	88
MF	DNA replication origin binding	GO:0003688	4.36×10^{-10}	14
MF	metal ion binding	GO:0046872	5.79×10^{-9}	241
MF	cation binding	GO:0043169	2.18×10^{-8}	241
MF	protein binding	GO:0005515	2.83×10^{-7}	329
MF	ATP-dependent activity	GO:0140657	2.95×10^{-7}	81
MF	ATP-dependent chromatin remodeler activity	GO:0140658	6.22×10^{-7}	14
MF	double-stranded DNA binding	GO:0003690	6.38×10^{-6}	43
MF	DNA helicase activity	GO:0003678	9.19×10^{-6}	14
MF	ATP binding	GO:0005524	1.18×10^{-5}	136
MF	adenyl ribonucleotide binding	GO:0032559	3.02×10^{-5}	136
MF	damaged DNA binding	GO:0003684	3.96×10^{-5}	14
MF	nuclease activity	GO:0004518	5.44×10^{-5}	29
MF	adenyl nucleotide binding	GO:0030554	3.55×10^{-4}	139
MF	endonuclease activity	GO:0004519	3.70×10^{-4}	20
MF	helicase activity	GO:0004386	6.72×10^{-4}	26
MF	histone modifying activity	GO:0140993	1.15×10^{-3}	16
MF	sequence-specific double-stranded DNA binding	GO:1990837	3.18×10^{-3}	33
MF	chromatin binding	GO:0003682	4.27×10^{-3}	14
MF	purine ribonucleoside triphosphate binding	GO:0035639	5.42×10^{-3}	145
MF	enzyme binding	GO:0019899	5.95×10^{-3}	30
MF	histone binding	GO:0042393	1.03×10^{-2}	11
MF	purine ribonucleotide binding	GO:0032555	1.11×10^{-2}	145
MF	structural constituent of nuclear pore	GO:0017056	1.26×10^{-2}	9
MF	DNA nuclease activity	GO:0004536	1.60×10^{-2}	12
MF	ATP-dependent DNA damage sensor activity	GO:0140664	1.87×10^{-2}	7
MF	DNA damage sensor activity	GO:0140612	1.87×10^{-2}	7
MF	ribonucleotide binding	GO:0032553	1.96×10^{-2}	145
MF	sequence-specific DNA binding	GO:0043565	2.51×10^{-2}	49
MF	transcription corepressor activity	GO:0003714	3.98×10^{-2}	10
MF	small molecule sensor activity	GO:0140299	4.02×10^{-2}	7
MF	purine nucleotide binding	GO:0017076	4.14×10^{-2}	149
MF	DNA polymerase activity	GO:0034061	4.43×10^{-2}	6
MF	DNA-directed DNA polymerase activity	GO:0003887	4.43×10^{-2}	6
MF	telomeric DNA binding	GO:0042162	4.99×10^{-2}	4
BP	DNA metabolic process	GO:0006259	1.87×10^{-62}	124
BP	nucleic acid metabolic process	GO:0090304	1.20×10^{-59}	335
BP	nucleobase-containing compound metabolic process	GO:0006139	3.56×10^{-48}	345
BP	heterocycle metabolic process	GO:0046483	2.22×10^{-46}	348
BP	cellular aromatic compound metabolic process	GO:0006725	6.31×10^{-46}	347
BP	organic cyclic compound metabolic process	GO:1901360	5.00×10^{-45}	349
BP	DNA repair	GO:0006281	1.12×10^{-43}	85
BP	DNA damage response	GO:0006974	1.15×10^{-41}	90
BP	chromosome organization	GO:0051276	2.35×10^{-36}	65
BP	macromolecule metabolic process	GO:0043170	8.42×10^{-34}	506
BP	cellular nitrogen compound metabolic process	GO:0034641	5.56×10^{-33}	353
BP	cell cycle	GO:0007049	3.27×10^{-32}	94
BP	DNA replication	GO:0006260	5.33×10^{-31}	55
BP	cell cycle process	GO:0022402	1.39×10^{-30}	79
BP	cellular response to stress	GO:0033554	1.87×10^{-26}	96
BP	DNA-templated DNA replication	GO:0006261	7.09×10^{-20}	36
BP	cellular metabolic process	GO:0044237	1.00×10^{-19}	428
BP	nitrogen compound metabolic process	GO:0006807	1.52×10^{-19}	507
BP	response to stress	GO:0006950	9.87×10^{-18}	100
BP	RNA metabolic process	GO:0016070	1.68×10^{-17}	222
BP	DNA recombination	GO:0006310	9.88×10^{-17}	36
BP	regulation of macromolecule metabolic process	GO:0060255	1.15×10^{-16}	181
BP	regulation of metabolic process	GO:0019222	1.25×10^{-16}	185
BP	regulation of gene expression	GO:0010468	2.24×10^{-16}	163
BP	double-strand break repair	GO:0006302	1.51×10^{-15}	31
BP	mitotic cell cycle process	GO:1903047	3.35×10^{-15}	42
BP	chromosome segregation	GO:0007059	3.55×10^{-15}	32
BP	regulation of nucleobase-containing compound metabolic process	GO:0019219	7.40×10^{-15}	144
BP	mitotic cell cycle	GO:0000278	1.74×10^{-14}	49
BP	regulation of cellular metabolic process	GO:0031323	4.36×10^{-14}	160

BP	regulation of primary metabolic process	GO:0080090	5.51×10^{-14}	164
BP	regulation of nitrogen compound metabolic process	GO:0051171	6.78×10^{-14}	163
BP	cellular process	GO:0009987	2.08×10^{-13}	641
BP	regulation of cell cycle	GO:0051726	7.26×10^{-13}	44
BP	regulation of RNA metabolic process	GO:0051252	7.26×10^{-13}	137
BP	regulation of cell cycle process	GO:0010564	2.01×10^{-12}	34
BP	protein modification process	GO:0036211	2.55×10^{-12}	156
BP	recombinational repair	GO:0000725	2.91×10^{-12}	23
BP	double-strand break repair via homologous recombination	GO:0000724	2.91×10^{-12}	23
BP	nuclear division	GO:0000280	3.01×10^{-12}	28
BP	nuclear chromosome segregation	GO:0098813	9.05×10^{-12}	25
BP	RNA biosynthetic process	GO:0032774	9.81×10^{-12}	135
BP	mRNA metabolic process	GO:0016071	3.26×10^{-11}	64
BP	primary metabolic process	GO:0044238	4.00×10^{-11}	514
BP	nucleic acid-templated transcription	GO:0097659	4.19×10^{-11}	133
BP	DNA-templated transcription	GO:0006351	4.19×10^{-11}	133
BP	macromolecule modification	GO:0043412	4.91×10^{-11}	163
BP	chromatin organization	GO:0006325	5.05×10^{-11}	37
BP	DNA conformation change	GO:0071103	5.39×10^{-11}	19
BP	regulation of cell cycle phase transition	GO:1901987	8.01×10^{-11}	26
BP	DNA replication initiation	GO:0006270	1.01×10^{-10}	18
BP	regulation of biological process	GO:0050789	1.63×10^{-10}	296
BP	organic substance metabolic process	GO:0071704	2.88×10^{-10}	529
BP	nucleobase-containing compound biosynthetic process	GO:0034654	3.02×10^{-10}	149
BP	regulation of cellular biosynthetic process	GO:0031326	3.51×10^{-10}	130
BP	cell cycle phase transition	GO:0044770	4.32×10^{-10}	27
BP	gene expression	GO:0010467	5.29×10^{-10}	239
BP	organelle fission	GO:0048285	5.50×10^{-10}	28
BP	regulation of biosynthetic process	GO:0009889	5.79×10^{-10}	130
BP	regulation of macromolecule biosynthetic process	GO:0010556	6.20×10^{-10}	129
BP	biological regulation	GO:0065007	7.85×10^{-10}	303
BP	aromatic compound biosynthetic process	GO:0019438	2.09×10^{-9}	151
BP	heterocycle biosynthetic process	GO:0018130	3.25×10^{-9}	152
BP	telomere organization	GO:0032200	3.47×10^{-9}	13
BP	regulation of DNA-templated transcription	GO:0006355	4.46×10^{-9}	116
BP	regulation of nucleic acid-templated transcription	GO:1903506	4.46×10^{-9}	116
BP	regulation of RNA biosynthetic process	GO:2001141	4.46×10^{-9}	116
BP	DNA geometric change	GO:0032392	4.75×10^{-9}	16
BP	negative regulation of biological process	GO:0048519	7.44×10^{-9}	73
BP	organic cyclic compound biosynthetic process	GO:1901362	9.29×10^{-9}	153
BP	sister chromatid segregation	GO:0000819	1.93×10^{-8}	20
BP	chromatin remodeling	GO:0006338	2.02×10^{-8}	29
BP	meiotic cell cycle	GO:0051321	2.04×10^{-8}	17
BP	regulation of mitotic cell cycle phase transition	GO:1901990	2.24×10^{-8}	18
BP	mRNA processing	GO:0006397	2.76×10^{-8}	49
BP	telomere maintenance	GO:0000723	2.82×10^{-8}	12
BP	DNA duplex unwinding	GO:0032508	3.18×10^{-8}	15
BP	regulation of cellular process	GO:0050794	3.97×10^{-8}	270
BP	cellular component organization	GO:0016043	5.34×10^{-8}	170
BP	negative regulation of cellular process	GO:0048523	5.53×10^{-8}	58
BP	RNA splicing	GO:0008380	7.86×10^{-8}	40
BP	regulation of mitotic cell cycle	GO:0007346	1.02×10^{-7}	23
BP	mitotic sister chromatid segregation	GO:0000070	1.18×10^{-7}	18
BP	mitotic nuclear division	GO:0140014	2.03×10^{-7}	19
BP	mitotic cell cycle phase transition	GO:0044772	2.03×10^{-7}	19
BP	histone modification	GO:0016570	3.60×10^{-7}	29
BP	cell cycle checkpoint signaling	GO:0000075	6.22×10^{-7}	17
BP	metabolic process	GO:0008152	7.39×10^{-7}	533
BP	transcription by RNA polymerase II	GO:0006366	7.69×10^{-7}	68
BP	negative regulation of cell cycle phase transition	GO:1901988	1.28×10^{-6}	17
BP	meiotic cell cycle process	GO:1903046	1.68×10^{-6}	14
BP	negative regulation of cell cycle	GO:0045786	1.90×10^{-6}	18
BP	nuclear DNA replication	GO:0033260	2.83×10^{-6}	12
BP	cell cycle DNA replication	GO:0044786	2.83×10^{-6}	12
BP	protein-DNA complex subunit organization	GO:0071824	4.41×10^{-6}	24
BP	peptidyl-lysine modification	GO:0018205	4.41×10^{-6}	24
BP	protein-DNA complex assembly	GO:0065004	7.21×10^{-6}	23
BP	negative regulation of cell cycle process	GO:0010948	8.90×10^{-6}	17
BP	RNA splicing, via transesterification reactions	GO:0000375	9.79×10^{-6}	34

BP	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	GO:0000377	9.79×10^{-6}	34
BP	mRNA splicing, via spliceosome	GO:0000398	9.79×10^{-6}	34
BP	peptidyl-amino acid modification	GO:0018193	1.32×10^{-5}	39
BP	nucleocytoplasmic transport	GO:0006913	1.40×10^{-5}	24
BP	nucleotide-excision repair	GO:0006289	1.62×10^{-5}	15
BP	organelle organization	GO:0006996	1.79×10^{-5}	113
BP	sexual reproduction	GO:0019953	1.97×10^{-5}	19
BP	nuclear transport	GO:0051169	2.02×10^{-5}	24
BP	negative regulation of mitotic cell cycle	GO:0045930	2.30×10^{-5}	13
BP	mitotic cell cycle checkpoint signaling	GO:0007093	2.30×10^{-5}	13
BP	regulation of chromosome organization	GO:0033044	2.30×10^{-5}	13
BP	negative regulation of macromolecule metabolic process	GO:0010605	2.38×10^{-5}	50
BP	negative regulation of metabolic process	GO:0009892	2.47×10^{-5}	51
BP	cellular response to stimulus	GO:0051716	2.99×10^{-5}	170
BP	double-strand break repair via break-induced replication	GO:0000727	7.13×10^{-5}	9
BP	protein modification by small protein conjugation or removal	GO:0070647	7.20×10^{-5}	41
BP	regulation of transcription by RNA polymerase II	GO:0006357	1.05×10^{-4}	57
BP	nuclear export	GO:0051168	1.23×10^{-4}	18
BP	reproductive process	GO:0022414	1.75×10^{-4}	19
BP	cellular component organization or biogenesis	GO:0071840	2.29×10^{-4}	173
BP	regulation of RNA splicing	GO:0043484	2.64×10^{-4}	14
BP	cell division	GO:0051301	3.07×10^{-4}	17
BP	meiotic nuclear division	GO:0140013	3.21×10^{-4}	10
BP	negative regulation of mitotic cell cycle phase transition	GO:1901991	3.21×10^{-4}	10
BP	negative regulation of cellular metabolic process	GO:0031324	3.97×10^{-4}	34
BP	macromolecule biosynthetic process	GO:0009059	4.24×10^{-4}	172
BP	regulation of G2/M transition of mitotic cell cycle	GO:0010389	5.09×10^{-4}	8
BP	regulation of mRNA metabolic process	GO:1903311	5.85×10^{-4}	22
BP	G2/M transition of mitotic cell cycle	GO:0000086	7.45×10^{-4}	9
BP	poly(A)+ mRNA export from nucleus	GO:0016973	1.06×10^{-3}	6
BP	RNA localization	GO:0006403	1.23×10^{-3}	15
BP	regulation of cell cycle G2/M phase transition	GO:1902749	1.68×10^{-3}	8
BP	metaphase/anaphase transition of mitotic cell cycle	GO:0007091	1.87×10^{-3}	9
BP	chromosome separation	GO:0051304	1.87×10^{-3}	9
BP	cell cycle G2/M phase transition	GO:0044839	1.87×10^{-3}	9
BP	regulation of mitotic metaphase/anaphase transition	GO:0030071	1.87×10^{-3}	9
BP	protein acetylation	GO:0006473	1.89×10^{-3}	15
BP	protein acylation	GO:0043543	2.17×10^{-3}	17
BP	negative regulation of gene expression	GO:0010629	2.35×10^{-3}	28
BP	reproduction	GO:0000003	3.03×10^{-3}	20
BP	DNA integrity checkpoint signaling	GO:0031570	3.36×10^{-3}	10
BP	regulation of alternative mRNA splicing, via spliceosome	GO:0000381	3.36×10^{-3}	10
BP	alternative mRNA splicing, via spliceosome	GO:0000380	3.36×10^{-3}	10
BP	protein modification by small protein conjugation	GO:0032446	3.56×10^{-3}	30
BP	response to stimulus	GO:0050896	3.99×10^{-3}	176
BP	regulation of sister chromatid segregation	GO:0033045	4.20×10^{-3}	9
BP	regulation of chromosome segregation	GO:0051983	4.20×10^{-3}	9
BP	metaphase/anaphase transition of cell cycle	GO:0044784	4.20×10^{-3}	9
BP	regulation of metaphase/anaphase transition of cell cycle	GO:1902099	4.20×10^{-3}	9
BP	meiosis I cell cycle process	GO:0061982	4.52×10^{-3}	8
BP	regulation of nuclear division	GO:0051783	4.52×10^{-3}	8
BP	regulation of mRNA splicing, via spliceosome	GO:0048024	4.95×10^{-3}	12
BP	RNA export from nucleus	GO:0006405	4.95×10^{-3}	12
BP	histone acetylation	GO:0016573	4.95×10^{-3}	12
BP	pre-replicative complex assembly involved in nuclear cell cycle DNA replication	GO:0006267	6.68×10^{-3}	6
BP	pre-replicative complex assembly involved in cell cycle DNA replication	GO:1902299	6.68×10^{-3}	6
BP	pre-replicative complex assembly	GO:0036388	6.68×10^{-3}	6
BP	mitotic G2/M transition checkpoint	GO:0044818	6.68×10^{-3}	6
BP	DNA replication checkpoint signaling	GO:0000076	6.68×10^{-3}	6
BP	regulation of mRNA processing	GO:0050684	6.98×10^{-3}	14
BP	nucleic acid transport	GO:0050657	7.67×10^{-3}	13
BP	RNA transport	GO:0050658	7.67×10^{-3}	13
BP	establishment of RNA localization	GO:0051236	7.67×10^{-3}	13
BP	internal peptidyl-lysine acetylation	GO:0018393	7.97×10^{-3}	12
BP	peptidyl-lysine acetylation	GO:0018394	7.97×10^{-3}	12
BP	internal protein amino acid acetylation	GO:0006475	7.97×10^{-3}	12
BP	negative regulation of nucleobase-containing compound metabolic process	GO:0045934	8.10×10^{-3}	23

BP	sister chromatid cohesion	GO:0007062	1.05×10^{-2}	8
BP	protein deacetylation	GO:0006476	1.05×10^{-2}	8
BP	histone deacetylation	GO:0016575	1.05×10^{-2}	8
BP	meiosis I	GO:0007127	1.06×10^{-2}	7
BP	reciprocal meiotic recombination	GO:0007131	1.06×10^{-2}	7
BP	reciprocal homologous recombination	GO:0140527	1.06×10^{-2}	7
BP	homologous recombination	GO:0035825	1.06×10^{-2}	7
BP	chromosome condensation	GO:0030261	1.06×10^{-2}	7
BP	RNA processing	GO:0006396	1.11×10^{-2}	72
BP	regulation of organelle organization	GO:0033043	1.26×10^{-2}	23
BP	protein ubiquitination	GO:0016567	1.31×10^{-2}	25
BP	negative regulation of cellular biosynthetic process	GO:0031327	1.38×10^{-2}	28
BP	negative regulation of nitrogen compound metabolic process	GO:0051172	1.39×10^{-2}	34
BP	mRNA catabolic process	GO:0006402	1.53×10^{-2}	18
BP	negative regulation of biosynthetic process	GO:0009890	1.97×10^{-2}	28
BP	negative regulation of cell cycle G2/M phase transition	GO:1902750	2.41×10^{-2}	6
BP	double-strand break repair via nonhomologous end joining	GO:0006303	2.41×10^{-2}	6
BP	meiotic chromosome segregation	GO:0045132	2.41×10^{-2}	6
BP	DNA-templated DNA replication maintenance of fidelity	GO:0045005	2.41×10^{-2}	6
BP	negative regulation of G2/M transition of mitotic cell cycle	GO:0010972	2.41×10^{-2}	6
BP	mitotic sister chromatid separation	GO:0051306	2.64×10^{-2}	7
BP	regulation of mitotic sister chromatid separation	GO:0010965	2.64×10^{-2}	7
BP	mitotic DNA integrity checkpoint signaling	GO:0044774	2.64×10^{-2}	7
BP	DNA biosynthetic process	GO:0071897	2.64×10^{-2}	7
BP	regulation of mitotic nuclear division	GO:0007088	2.64×10^{-2}	7
BP	negative regulation of macromolecule biosynthetic process	GO:0010558	2.67×10^{-2}	27
BP	spindle assembly	GO:0051225	4.26×10^{-2}	8
BP	protein export from nucleus	GO:0006611	4.98×10^{-2}	5
BP	interstrand cross-link repair	GO:0036297	4.98×10^{-2}	5
BP	protein sumoylation	GO:0016925	4.98×10^{-2}	5
CC	nucleus	GO:0005634	4.70×10^{-131}	511
CC	intracellular organelle	GO:0043229	3.15×10^{-62}	608
CC	organelle	GO:0043226	5.09×10^{-62}	611
CC	intracellular membrane-bounded organelle	GO:0043231	3.19×10^{-60}	561
CC	membrane-bounded organelle	GO:0043227	8.21×10^{-57}	563
CC	chromosome	GO:0005694	5.83×10^{-46}	114
CC	intracellular anatomical structure	GO:0005622	2.50×10^{-43}	643
CC	nuclear protein-containing complex	GO:0140513	7.81×10^{-32}	152
CC	nuclear chromosome	GO:0000228	1.70×10^{-19}	40
CC	replication fork	GO:0005657	6.06×10^{-13}	21
CC	chromatin	GO:0000785	1.34×10^{-12}	50
CC	nucleoplasm	GO:0005654	1.58×10^{-11}	56
CC	nuclear lumen	GO:0031981	3.40×10^{-11}	88
CC	protein-containing complex	GO:0032991	1.51×10^{-7}	283
CC	nuclear replication fork	GO:0043596	2.93×10^{-7}	13
CC	condensed chromosome	GO:0000793	9.43×10^{-7}	15
CC	protein-DNA complex	GO:0032993	1.08×10^{-6}	27
CC	transferase complex	GO:1990234	1.12×10^{-6}	69
CC	intracellular non-membrane-bounded organelle	GO:0043232	2.51×10^{-6}	155
CC	non-membrane-bounded organelle	GO:0043228	3.04×10^{-6}	155
CC	nuclear envelope	GO:0005635	4.35×10^{-6}	21
CC	SWI/SNF superfamily-type complex	GO:0070603	8.68×10^{-6}	18
CC	transcription regulator complex	GO:0005667	6.48×10^{-5}	30
CC	nuclear body	GO:0016604	4.92×10^{-4}	10
CC	organelle lumen	GO:0043233	4.93×10^{-4}	89
CC	membrane-enclosed lumen	GO:0031974	4.93×10^{-4}	89
CC	intracellular organelle lumen	GO:0070013	4.93×10^{-4}	89
CC	MCM complex	GO:0042555	8.90×10^{-4}	7
CC	origin recognition complex	GO:0000808	9.28×10^{-4}	6
CC	DNA replication preinitiation complex	GO:0031261	9.28×10^{-4}	6
CC	replication fork protection complex	GO:0031298	9.28×10^{-4}	6
CC	nuclear origin of replication recognition complex	GO:0005664	9.28×10^{-4}	6
CC	nuclear pore	GO:0005643	1.31×10^{-3}	15
CC	condensed nuclear chromosome	GO:0000794	5.74×10^{-3}	6
CC	histone acetyltransferase complex	GO:0000123	6.80×10^{-3}	14
CC	acetyltransferase complex	GO:1902493	6.81×10^{-3}	16
CC	protein acetyltransferase complex	GO:0031248	6.81×10^{-3}	16
CC	DNA replication factor C complex	GO:0005663	6.81×10^{-3}	5
CC	nuclear speck	GO:0016607	1.04×10^{-2}	7
CC	nuclear replisome	GO:0043601	2.03×10^{-2}	6

CC	replisome	GO:0030894	2.03×10^{-2}	6
CC	RNA polymerase III transcription regulator complex	GO:0090576	3.62×10^{-2}	5
CC	histone deacetylase complex	GO:0000118	4.51×10^{-2}	8
CC	alpha DNA polymerase:primase complex	GO:0005658	4.99×10^{-2}	4
CC	DNA packaging complex	GO:0044815	4.99×10^{-2}	4
CC	GINS complex	GO:0000811	4.99×10^{-2}	4

CX7 = control carcasses at 7 days post-blood meal, IX7 = P. falciparum-*infected carcasses at 7 days post-infection, X7 = carcasses (infected and control) at 7 days; MF = molecular function, BP = biological process, CC = cellular component;* ∩ = *intersection,* i.e. *the number of overlapping transcripts belonging to a list of transcripts from a comparison (of a Venn diagram) and to a GO/KEGG term; adjusted* p = *false discovery rate-adjusted* p-*values of Fisher's one-tailed tests, significant if adjusted* $p < 0.05$.

Appendix C

Table C.1. Transcripts making up the top two-thirds of the secreted transcriptome of each sample.

		Proportion of Secreted Transcriptome (%)								
		Cardia			Anterior			Posterior		
Transcript ID	Name/Annotation	0 h	5 h	24 h	0 h	5 h	24 h	0 h	5 h	24 h
AGAP000692-RA	CEC2	19.7	18.8	14.9	1.7	2.8				
AGAP013543-RA	unspecified product	21.5	20.9	14.8	2.3	3.0				
AGAP000694-RA	CEC3	10.4	10.1	10.4						
AGAP011294-RA	DEF1	9.2	8.2	8.3						
AGAP009313-RA	unspecified product	6.4	6.6	7.0	1.3	1.7				
AGAP008296-RA	trypsin 1			3.1	3.1	3.1	3.3	5.5	4.9	7.6
AGAP008645-RA	GAM1			2.8	3.3	2.1	2.2			
AGAP008861-RA	female digestive tract protease			2.2	7.7	12.1	11.7	2.6	1.9	
AGAP000693-RA	CEC1			2.2						
AGAP006796-RA	peritrophin			1.2	3.0		2.0	1.7		
AGAP000570-RA	unspecified product			1.2	9.1	8.8	5.9	5.7	2.3	
AGAP011630-RA	unspecified product		2.6							
AGAP007745-RA	unspecified product				5.4	4.9	3.6	3.2	2.4	
AGAP006432-RA	unspecified product				3.8	2.6	2.2	2.5		
AGAP007747-RA	unspecified product				3.4	2.4	3.5	2.4	1.9	
AGAP001193-RA	unspecified product				2.9	3.2	2.2	3.2	2.0	
AGAP012401-RA	α-amylase domain-containing				2.6					
AGAP006442-RA	unspecified product				2.5	3.5	2.0	1.3	1.4	1.7
AGAP003995-RA	unspecified product				2.4	1.5				
AGAP006795-RA	peritrophin 1				2.1		2.3	2.9	1.6	2.0
AGAP001192-RA	unspecified product				1.7		2.2	3.6	3.3	2.6
AGAP006711-RA	chymotrypsin 2				1.6		2.6	7.7	1.7	7.5
AGAP006414-RA	chitinase				1.4	1.3		1.3		
AGAP009527-RA	unspecified product				1.2			1.7		
AGAP001199-RA	serine protease				1.2			8.8	6.0	
AGAP012400-RA	α-glucosidase				1.2					
AGAP011939-RA	α-amylase domain-containing				1.1			1.2		
AGAP003713-RA	unspecified product				0.9	1.7	2.5			
AGAP013106-RA	NPC2					2.1				
AGAP003926-RA	aminopeptidase					1.7				
AGAP006434-RA	unspecified product					1.7				
AGAP004883-RA	unspecified product					1.6			1.1	
AGAP007612-RA	unspecified product					1.6				
AGAP003066-RA	cytochrome P450					1.5				
AGAP009830-RA	unspecified product					1.5				
AGAP004918-RA	FBN					1.3				
AGAP006385-RA	unspecified product						7.9		5.1	8.0
AGAP006187-RA	protein G12						3.6		4.1	5.5
AGAP007165-RA	late trypsin						3.3			5.7
AGAP005310-RA	unspecified product						2.6			7.6
AGAP008295-RA	trypsin 2						1.9		2.2	6.7
AGAP006710-RA	chymotrypsin							3.4		3.4
AGAP008292-RA	trypsin 4							2.6	2.8	
AGAP006709-RA	chymotrypsin 1							1.9		4.4
AGAP008294-RA	trypsin 3							1.5	2.1	
AGAP007906-RA	Spink6							1.3		
AGAP002643-RA	unspecified product							1.2		
AGAP009593-RA	zinc carboxypeptidase A1								6.3	1.6
AGAP005065-RA	unspecified product								4.4	
AGAP029692-RA	unspecified product								2.3	
AGAP006707-RA	serine endopeptidase								1.8	
AGAP008487-RA	phosphodiesterase								1.7	
AGAP001198-RA	chymotrypsin								1.5	
AGAP002464-RA	secreted ferritin G subunit								1.2	
AGAP001182-RA	unspecified product								1.1	
AGAP002851-RA	NPC2									1.7
AGAP007142-RA	endopeptidase									1.4

Table C.2. Significantly enriched GO terms of genes making up the top two-thirds of the secreted transcriptome. *All GO terms listed are biological processes (BP) and were identified using g:Profiler (Raudvere et al, 2019).*

GO term	Term ID	Adjusted *p*	∩ (Genes)
Cardia - 0 h			
humoral immune response	GO:0006959	1.28×10^{-3}	3 (*CEC1–3*)
response to bacterium	GO:0009617	1.53×10^{-2}	3 (*CEC1–3*)
defense response to bacterium	GO:0042742	1.53×10^{-2}	3 (*CEC1–3*)
Cardia - 5 h			
humoral immune response	GO:0006959	2.56×10^{-3}	3 (*CEC1–3*)
response to bacterium	GO:0009617	3.06×10^{-2}	3 (*CEC1–3*)
defense response to bacterium	GO:0042742	3.06×10^{-2}	3 (*CEC1–3*)
Cardia - 24 h			
humoral immune response	GO:0006959	2.96×10^{-5}	4 (*CEC1–3*; *DEF1*)
response to bacterium	GO:0009617	1.23×10^{-3}	4 (*CEC1–3*; *DEF1*)
defense response to bacterium	GO:0042742	1.23×10^{-3}	4 (*CEC1–3*; *DEF1*)
antimicrobial humoral response	GO:0019730	8.45×10^{-3}	3 (*CEC1–3*)
antibacterial humoral response	GO:0019731	8.45×10^{-3}	3 (*CEC1–3*)
defense response to Gram-positive bacterium	GO:0050830	8.45×10^{-3}	3 (*CEC1–3*)
innate immune response	GO:0045087	1.38×10^{-2}	4 (*CEC1–3*; *DEF1*)
immune system process	GO:0002376	1.77×10^{-2}	4 (*CEC1–3*; *DEF1*)
immune response	GO:0006955	1.77×10^{-2}	4 (*CEC1–3*; *DEF1*)
response to biotic stimulus	GO:0009607	3.43×10^{-2}	4 (*CEC1–3*; *DEF1*)
response to external biotic stimulus	GO:0043207	3.43×10^{-2}	4 (*CEC1–3*; *DEF1*)
biological process involved in interspecies interaction between organisms	GO:0044419	3.43×10^{-2}	4 (*CEC1–3*; *DEF1*)
response to other organism	GO:0051707	3.43×10^{-2}	4 (*CEC1–3*; *DEF1*)
defense response to other organism	GO:0098542	3.43×10^{-2}	4 (*CEC1–3*; *DEF1*)
defense response to Gram-negative bacterium	GO:0050829	4.21×10^{-2}	3 (*CEC1–3*)
Posterior - 0 h			
digestion	GO:0007586	2.68×10^{-2}	4 (trypsins 1, 4; chymotrypsin 2; CP)
Posterior - 5 h			
digestion	GO:0007586	5.06×10^{-4}	5 (trypsins 1, 2, 4; chymotrypsin 2; CP)
Posterior - 24 h			
digestion	GO:0007586	2.35×10^{-5}	5 (trypsins 1, 2, 4; chymotrypsin 2; CP)

∩ *(genes) = intersection,* i.e. *the overlapping genes making up the top two-thirds of a transcriptome and to a KEGG term; adjusted* p *= false discovery rate-adjusted* p*-values of Fisher's one-tailed tests, significant if adjusted* p *< 0.05. CEC = cecropin, DEF1 = defensin 1, CP = zinc carboxypeptidase A1.*

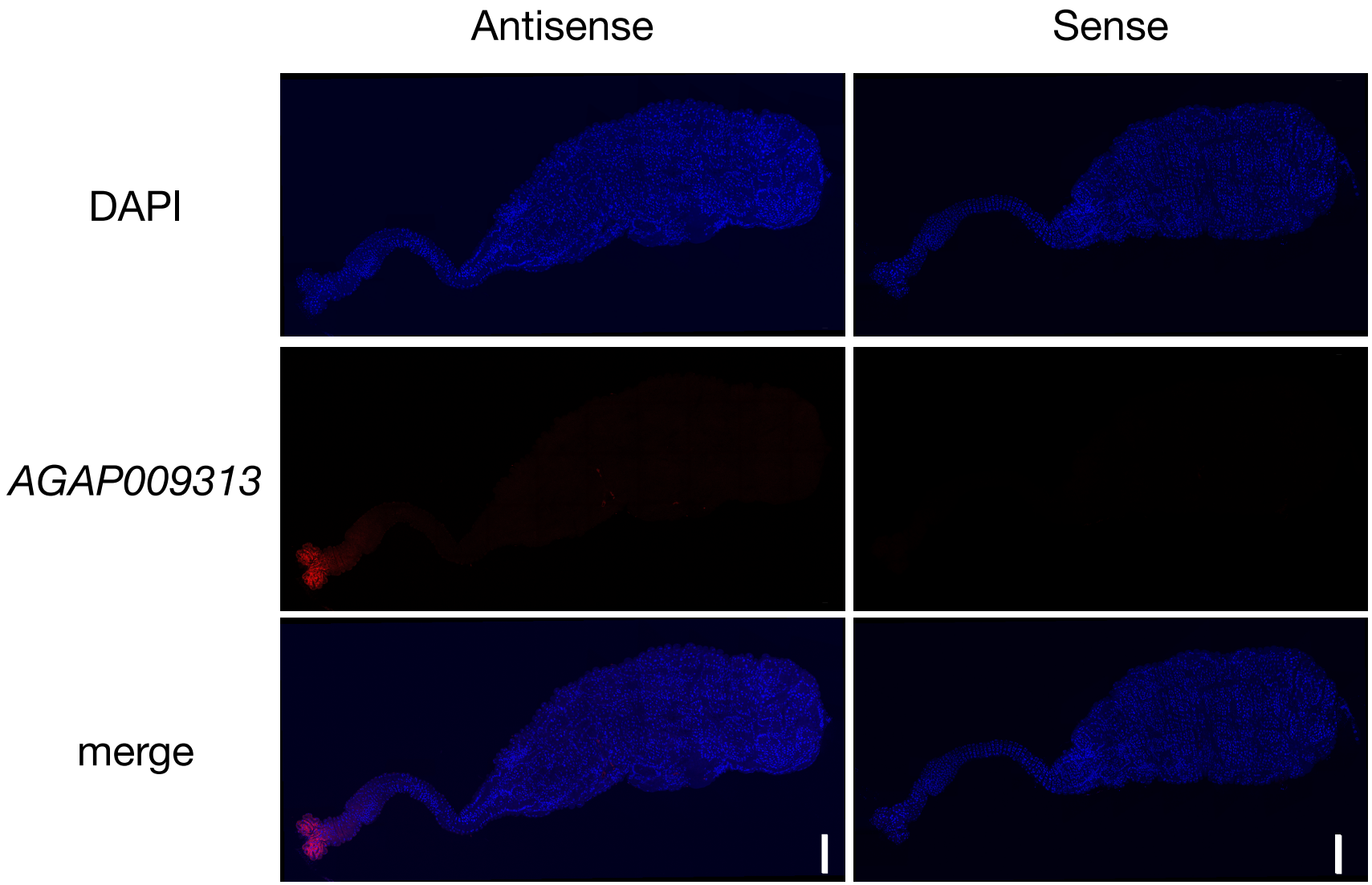


Figure C.1. Confocal imaging of AGAP009313 mRNA and nuclei in the mosquito midgut. *Nuclei were stained with DAPI and are represented in blue. Alexa Fluor™ 594-labeled RNA probes (sense and antisense strands separately) were hybridized to AGAP009313 mRNA and are represented in red. Scale bar represents 200 µm.*

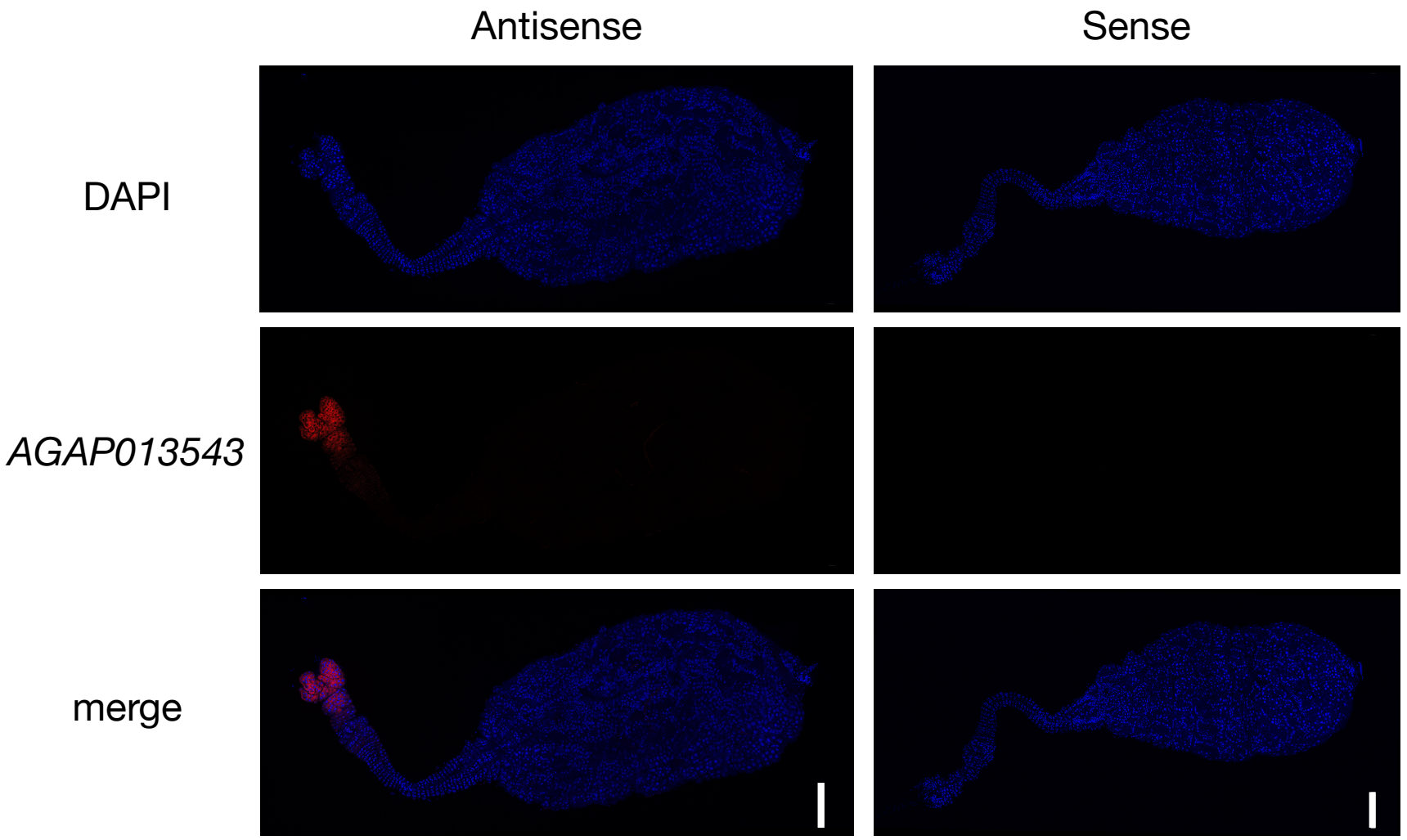


Figure C.2. Confocal imaging of AGAP013543 mRNA and nuclei in the mosquito midgut. *Nuclei were stained with DAPI and are represented in blue. Alexa Fluor™ 594-labeled RNA probes (sense and antisense strands separately) were hybridized to AGAP013543 mRNA and are represented in red. Scale bar represents 200 µm.*

Appendix D

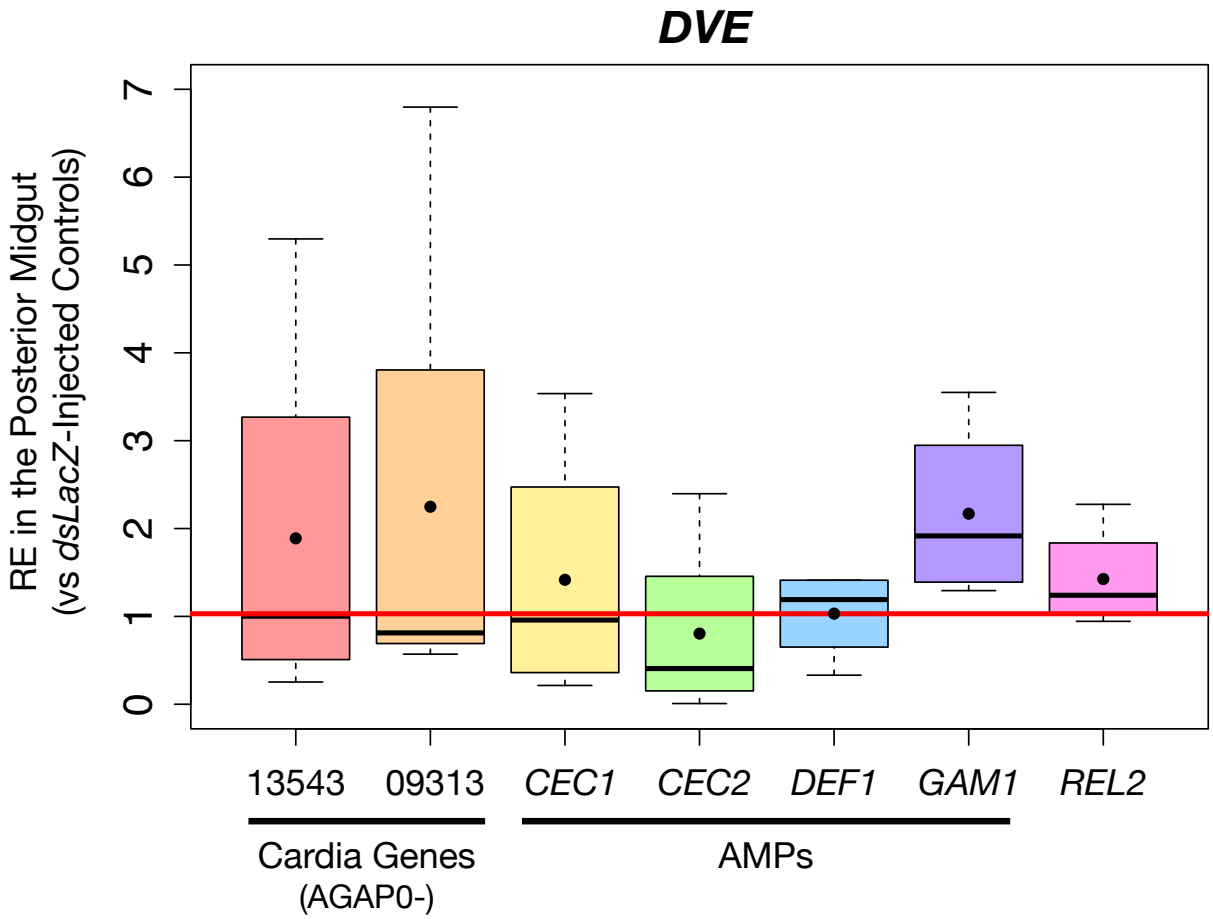


Figure D.1. Relative expression of AMPs post-*DVE *KD in the posterior midgut. *Expression values were normalized to that of* dsLacZ*-injected controls, represented by the red line. Black bars represent medians, dots represent means. Four replicates were performed. RE = relative expression. No significant differences in expression were observed between* DVE*-KD and* dsLacZ*-injected posterior midguts (paired Student's* t*-tests,* p *> 0.05).*

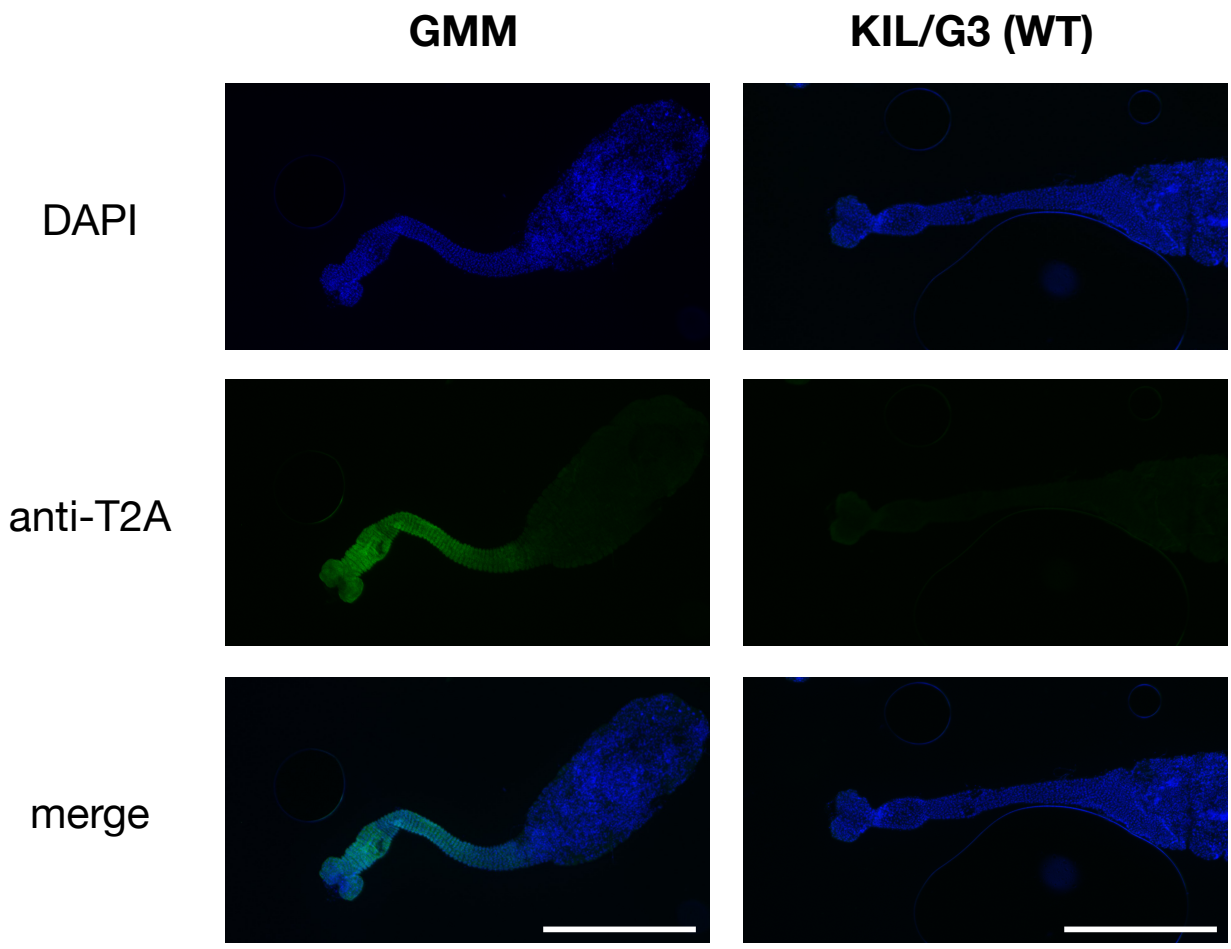


Figure D.2. Fluorescent imaging of* GAM1 *expression in the midgut of GMM transgenic and WT mosquitoes. *Antibodies raised against a ribosomal skip sequence in the transgenic mosquito line were conjugated to GFP secondary antibodies and are represented in green. Nuclei were stained with DAPI and are represented in blue. Scale bars represent 750 μm.*

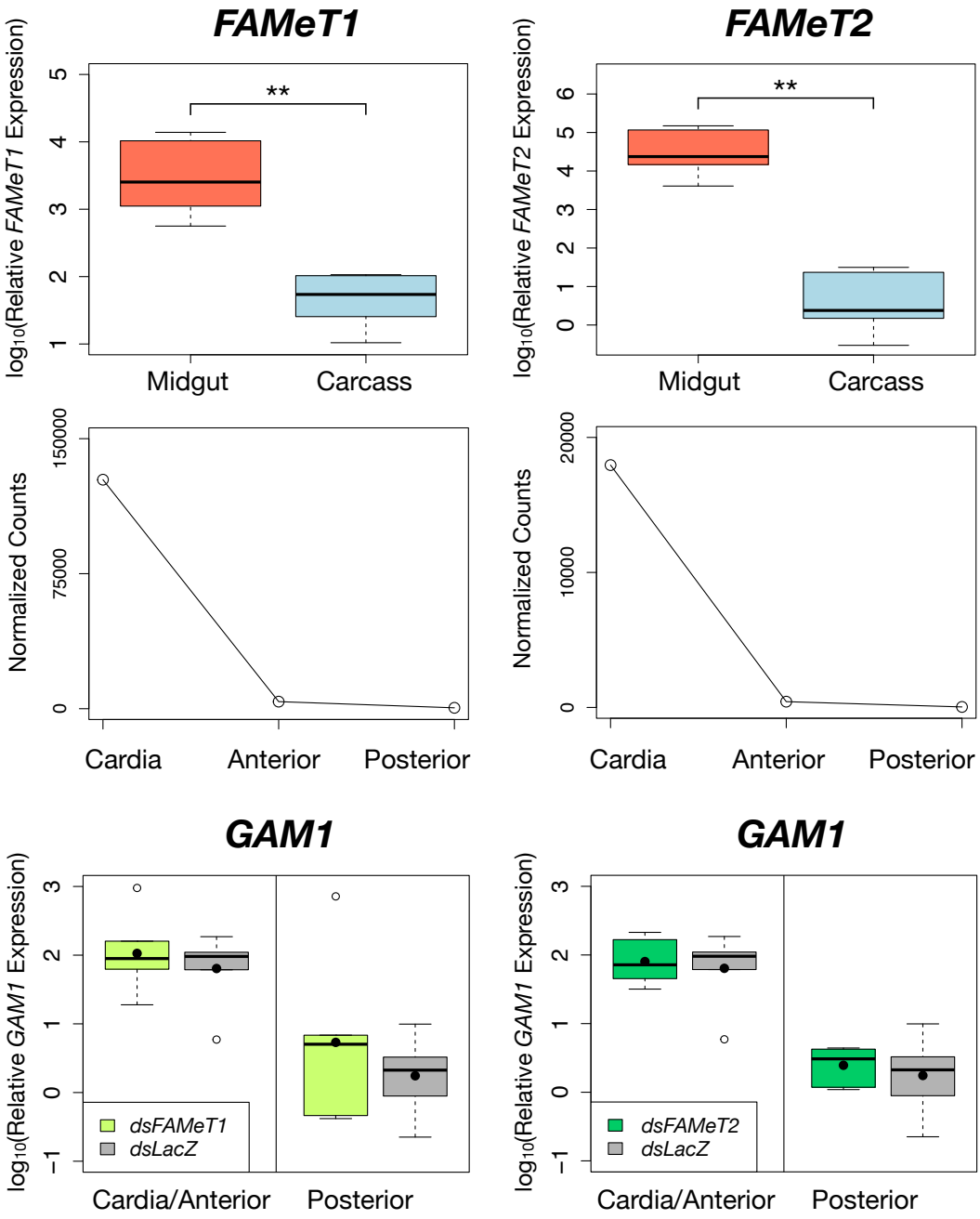


Figure D.3. Expression profile of FAMeTs and expression of* GAM1 *post-*FAMeT *KD in the midgut. *First row: relative expression values were log_{10}-transformed for presentation purposes and are relative to a pool of mosquito cDNA. Five replicates were completed. Wilcoxon rank-sum tests, ** $p < 0.01$. Second row: normalized counts were calculated and averaged from three replicates of the midgut sections transcriptomic data under sugar feeding conditions (0 hours) with DESeq2. Last row: relative expression values were log_{10}-transformed for presentation purposes and are relative to a pool of mosquito cDNA. Black bars represent medians, dots represent means. Six replicates were performed, with Fanyun (Coco) Wu (BSc final year project student, project supervised by Prof. George K. Christophides) contributing towards three replicates. RE = relative expression. No differences were observed in* GAM1 *expression between KD mosquitoes versus controls, Student's* t*-tests,* $p > 0.05$.

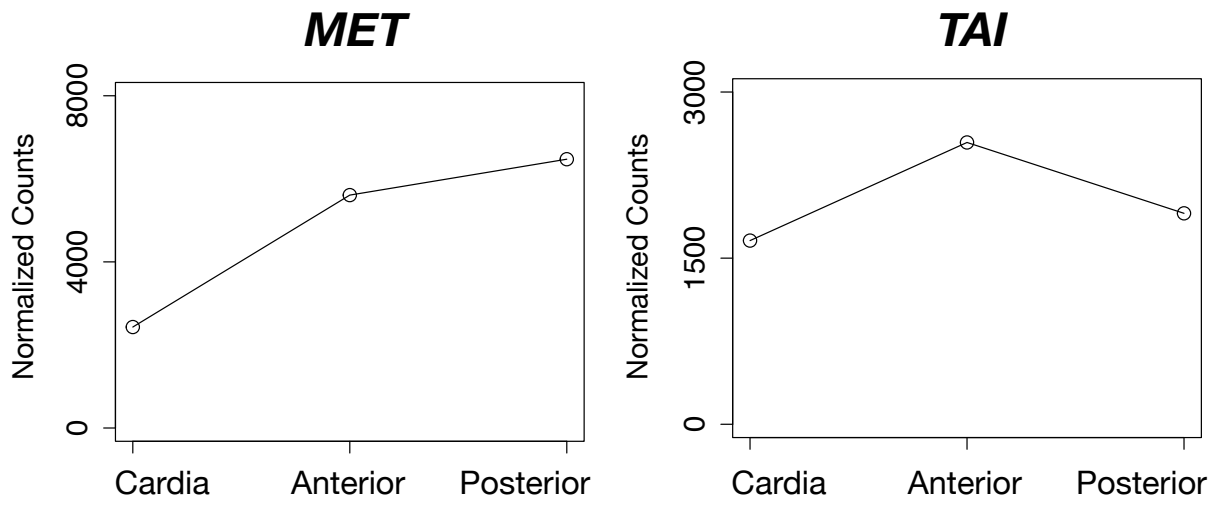


Figure D.4. Expression profiles of* MET *and* TAI *in the midgut. *Normalized counts were calculated and averaged from three replicates of the midgut sections transcriptomic data under sugar feeding conditions (0 hours) with DESeq2.*

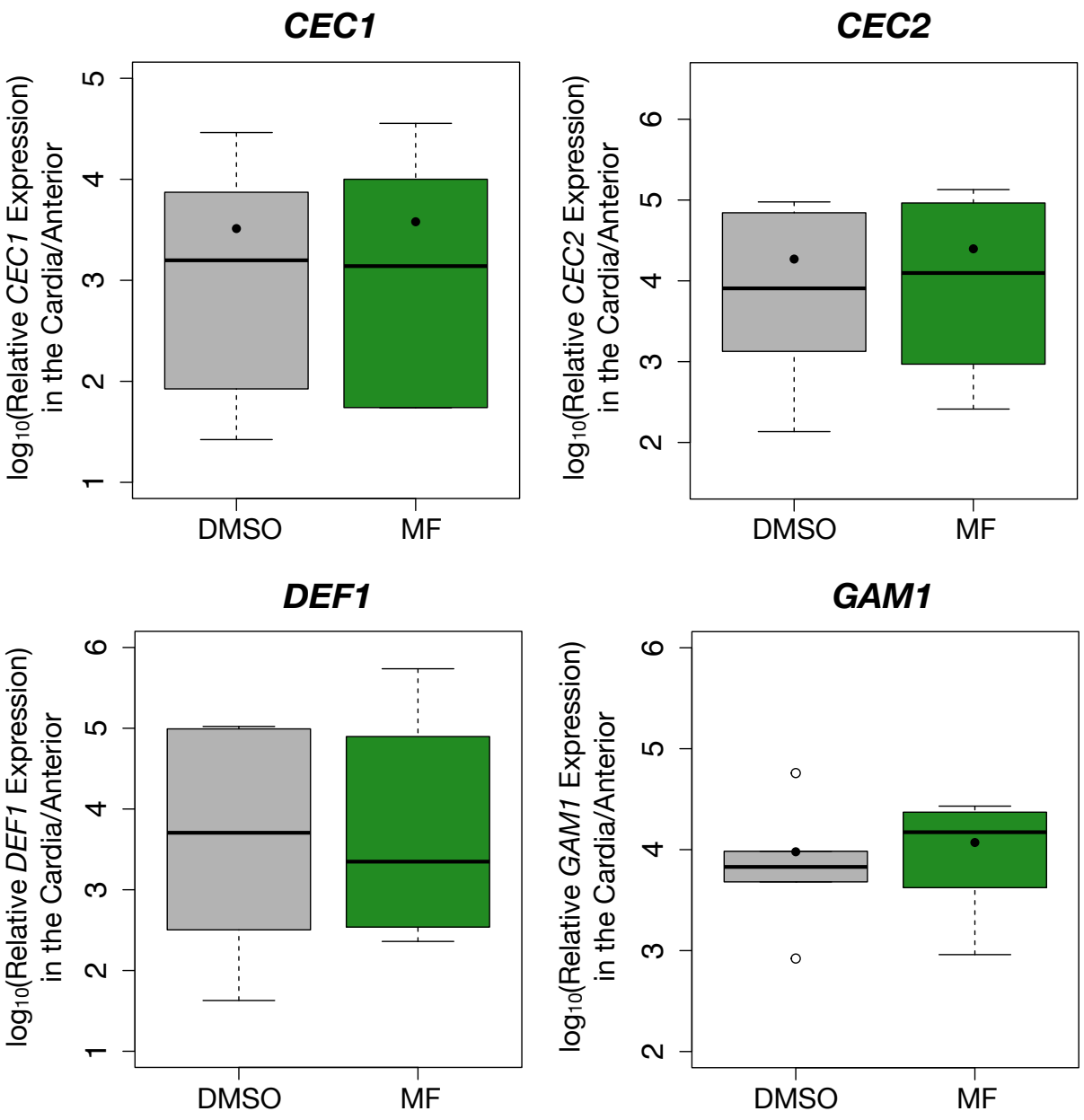


Figure D.5. Relative expression of AMPs post-MF feeding in the cardia and anterior midgut. *Relative expression values were log_{10}-transformed for presentation purposes and are relative to a pool of mosquito cDNA. Black bars represent medians, dots represent means. Six replicates were performed, with Fanyun (Coco) Wu and Sihao Ma (BSc final year project students, project supervised by Prof. George K. Christophides) contributing towards three replicates each. Paired Student's* t*-tests and Wilcoxon signed-rank tests were performed to assess differences in expression between DMSO- and MF-fed mosquitoes. No significant differences were observed* ($p > 0.05$).

Appendix E

Table E.1. Novel transcripts which fulfilled the initial putative AMP selection criteria identified from* de novo *transcriptome assemblies of the midgut sections transcriptomic data.

	DESeq2 Output						Genomic Location				Signal Peptide Likelihood (%)
			Wald Test								
Transcript ID	Avg. Norm. Counts	LFC	*W* Stat.	Adjusted *p*	Comparison		Chromosome/ Scaffold	Start	End	Length (bp)	
MSTRG.60.1	145.53	2.12	2.46	4.90×10^{-2}	5C	5P	AAAB01004737	35678	36130	452	0
MSTRG.223.1	474.04	3.13	3.91	1.22×10^{-3}	24C	24P	AAAB01008449	11082	11795	713	0.01
MSTRG.220.1	80.88	4.33	3.70	1.50×10^{-3}	5C	5P	AAAB01008449	1	547	546	0.03
MSTRG.11095.1	251.85	5.03	8.49	2.05×10^{-15}	24C	24P	3R	50884336	50884761	425	0.02
MSTRG.10129.1	63.36	4.50	4.72	5.20×10^{-5}	24C	24P	3R	29388230	29388431	201	0.12
MSTRG.9824.1	51.56	4.14	3.84	1.57×10^{-3}	24C	24P	3R	20834117	20834733	616	0.01
MSTRG.13602.1	174.07	3.52	5.46	1.49×10^{-6}	24C	24P	Unknown	40373712	40374207	495	0.05
MSTRG.12310.1	65.64	4.40	5.24	4.70×10^{-6}	24C	24P	Unknown	31913415	31914106	691	0.01
MSTRG.11279.1	615.69	2.55	4.64	3.96×10^{-5}	5C	5P	Unknown	1250250	1250723	473	0.01
MSTRG.12669.1	67.14	3.91	5.54	5.79×10^{-7}	5C	5P	Unknown	36083227	36083844	617	0.01
MSTRG.13542.1	265.13	3.91	5.04	6.56×10^{-6}	5C	5P	Unknown	40100585	40101076	491	0
MSTRG.13951.1	321.14	2.28	4.50	7.35×10^{-5}	5C	5P	Unknown	42026166	42026717	551	0.12

Avg. norm. counts = average normalized counts from DESeq2, lfc = log_2 fold change between a cardia and posterior sample, Wald stat = Wald statistic, adjusted p *= false discovery rate-corrected* p*-values, comp. = comparison groups; signal peptide likelihood calculated with SignalP 6.0 (Teufel et al, 2022).*

Appendix F

Table F.2. Average KD efficiencies and Cox proportional hazard models of bacterial survival assays following putative AMP KD. *Survival curves were compared between mosquitoes injected with putative AMP dsRNA and control mosquitoes injected with* dsLacZ*, followed by bacterial infection via oral administration or thoracic injection.*

							Cox Proportional Hazard Model					
Gene	KD Eff.	Bacteria	R	n	n (controls)	Time	Covariates	Coefficient	SE	HR	95% CI	*p*-value
AGAP001181	48%	*E. coli*	4	74	73	10 days	Gene KD	-0.22	0.19	0.81	0.55–1.17	0.26
							Replicate	0.26	0.09	1.29	1.08–1.54	0.004 **
		S. marcescens	4	71	86	10 days	Gene KD	0.15	0.19	1.17	0.80–1.70	0.42
							Replicate	-0.08	0.02	0.92	0.88–0.97	0.001 **
AGAP001352	64%	*E. coli*	4	96	73	10 days	Gene KD	-0.13	0.19	0.88	0.60–1.28	0.51
							Replicate	-0.01	0.03	0.99	0.93–1.06	0.78
		S. marcescens	6	93	102	10 days	Gene KD	-0.13	0.16	0.88	0.64–1.21	0.43
							Replicate	-0.18	0.04	0.84	0.77–0.91	6.68×10^{-5} ***
AGAP001508	57%	*E. coli*	4	83	73	10 days	Gene KD	0.17	0.19	1.19	0.82–1.73	0.37
							Replicate	0.14	0.09	1.15	0.96–1.39	0.12
		S. marcescens	3	84	74	10 days	Gene KD	0.49	0.20	1.63	1.10–2.41	0.02 *
							Replicate	-0.05	0.02	0.95	0.91–1.00	0.04 *
AGAP002852	66%	*E. coli*	4	134	64	10 days	Gene KD	0.22	0.17	1.25	0.89–1.75	0.20
							Replicate	0.004	0.03	1.00	0.94–1.08	0.90
		S. marcescens	5	78	57	10 days	Gene KD	0.01	0.21	1.01	0.67–1.53	0.96
							Replicate	-0.14	0.04	0.87	0.80–0.94	5.84×10^{-4} ***
AGAP003713	68%	*E. coli*	3	119	120	10 days	Gene KD	-0.07	0.14	0.93	0.71–1.24	0.65
							Replicate	0.06	0.05	1.06	0.97–1.12	0.22
		S. marcescens	3	90	78	10 days	Gene KD	-0.33	0.20	0.72	0.48–1.07	0.11
							Replicate	0.02	0.07	1.02	0.90–1.17	0.73
KIN1	63%	*E. faecalis*	3	138	120	All	Gene KD	-0.05	0.13	0.95	0.74–1.23	0.72
							Replicate	-0.17	0.08	0.84	0.72–0.99	0.03 *
						10 days	Gene KD	-0.83	0.26	0.43	0.26–0.72	0.001 **
							Replicate	-0.71	0.16	0.49	0.36–0.67	8.55×10^{-6} ***
		S. marcescens	3	190	199	All	Gene KD	-0.24	0.10	0.78	0.64–0.96	0.02 *
							Replicate	0.08	0.07	1.08	0.93–1.25	0.30
						10 days	Gene KD	-0.43	0.21	0.65	0.43–0.98	0.04 *
							Replicate	-0.16	0.15	0.85	0.63–1.13	0.27
		E. coli	3	186	184	All	Gene KD	-0.23	0.10	0.79	0.65–0.97	0.03 *
							Replicate	-0.28	0.08	0.76	0.65–0.89	6.74×10^{-4} ***
						10 days	Gene KD	-0.44	0.13	0.64	0.50–0.82	4.91×10^{-4} ***
							Replicate	-0.33	0.09	0.71	0.60–0.86	2.92×10^{-4} ***
		S. aureus	3	180	148	All	Gene KD	-0.08	0.11	0.93	0.74–1.15	0.49
							Replicate	-0.64	0.07	0.53	0.46–0.60	2.00×10^{-16} ****
						10 days	Gene KD	-0.05	0.17	0.96	0.69–1.32	0.78
							Replicate	-0.90	0.12	0.41	0.32–0.51	1.12×10^{-14} ****

Gene	KD Eff.	Bacteria	R	n	n (controls)	Time	Cox Proportional Hazard Model					
							Covariates	Coefficient	SE	HR	95% CI	*p*-value
GLYR	49%	*E. faecalis*	3	124	120	All	Gene KD	0.13	0.13	1.13	0.88–1.46	0.33
							Replicate	-0.18	0.08	0.84	0.72–0.98	0.02 *
						10 days	Gene KD	0.29	0.29	1.34	0.76–2.35	0.32
							Replicate	-0.29	0.17	0.75	0.54–1.04	0.09
		S. marcescens	3	197	199	All	Gene KD	-0.07	0.10	0.93	0.76–1.14	0.49
							Replicate	-0.01	0.07	0.99	0.86–1.14	0.90
						10 days	Gene KD	0.19	0.23	1.21	0.78–1.89	0.40
							Replicate	0.17	0.15	1.18	0.88–1.60	0.27
		E. coli	3	213	184	All	Gene KD	-0.05	0.10	0.95	0.78–1.17	0.64
							Replicate	-0.30	0.07	0.74	0.64–0.85	2.80×10^{-5} ****
						10 days	Gene KD	-0.24	0.13	0.79	0.62–1.01	0.06
							Replicate	-0.37	0.09	0.69	0.58–0.82	1.61×10^{-5} ****
		S. aureus	3	200	148	All	Gene KD	-0.09	0.11	0.91	0.74–1.13	0.41
							Replicate	-0.60	0.07	0.55	0.48–0.63	2.16×10^{-12} ****
						10 days	Gene KD	-0.03	0.16	0.97	0.71–1.34	0.87
							Replicate	-0.74	0.10	0.48	0.39–0.59	2.16×10^{-12} ****
ALAR	64%	*E. faecalis*	3	191	120	All	Gene KD	-0.14	0.12	0.87	0.69–1.10	0.24
							Replicate	-0.19	0.07	0.83	0.72–0.95	0.01 **
						10 days	Gene KD	-0.17	0.23	0.85	0.54–1.33	0.47
							Replicate	-0.15	0.13	0.86	0.66–1.12	0.26
		S. marcescens	3	214	199	All	Gene KD	-0.20	0.10	0.82	0.67–1.00	0.04 *
							Replicate	-0.02	0.06	0.98	0.98–0.87	0.79
						10 days	Gene KD	0.23	0.23	1.26	0.80–1.97	0.31
							Replicate	0.23	0.15	1.26	0.95–1.68	0.11
		E. coli	3	141	184	All	Gene KD	-0.13	0.11	0.88	0.70–1.09	0.24
							Replicate	-0.18	0.09	0.84	0.80–1.00	0.04 *
						10 days	Gene KD	-0.24	0.14	0.79	0.60–1.03	0.07
							Replicate	-0.23	0.11	0.79	0.64–0.97	0.03 *
		S. aureus	3	184	148	All	Gene KD	0.04	0.11	1.04	0.84–1.30	0.71
							Replicate	-0.63	0.07	0.54	0.47–0.60	2.00×10^{-16} ****
						10 days	Gene KD	0.18	0.17	1.20	0.86–1.67	0.30
							Replicate	-0.90	0.11	0.45	0.36–0.56	2.3×10^{-12} ****

Gene	KD Eff.	Bacteria	R	n	n (controls)	Time	Cox Proportional Hazard Model					
							Covariates	Coefficient	SE	HR	95% CI	*p*-value
LYSC1	62%	*E. faecalis*	3	113	120	All	Gene KD	0.06	0.13	1.06	0.82–1.38	0.65
							Replicate	-0.27	0.08	0.76	0.65–0.90	0.001 **
						10 days	Gene KD	0.14	0.26	1.15	0.69–1.93	0.59
							Replicate	-0.15	0.16	0.86	0.63–1.19	0.37
		S. marcescens	3	179	199	All	Gene KD	0.07	0.10	1.07	0.87–1.31	0.52
							Replicate	0.04	0.07	1.04	0.90–1.20	0.57
						10 days	Gene KD	0.06	0.23	0.78	0.68–1.67	0.78
							Replicate	0.12	0.15	1.13	0.84–1.53	0.41
		E. coli	3	163	184	All	Gene KD	-0.10	0.11	0.90	0.73–1.12	0.34
							Replicate	-0.02	0.08	0.98	0.83–1.15	0.77
						10 days	Gene KD	-0.11	0.14	0.89	0.69–1.17	0.41
							Replicate	-0.09	0.10	0.91	0.75–1.11	0.34
		S. aureus	3	165	148	All	Gene KD	0.05	0.11	1.06	0.84–1.32	0.63
							Replicate	-0.59	0.07	0.55	0.48–0.64	2.00×10^{-16} ****
						10 days	Gene KD	0.14	0.18	1.15	0.82–1.62	0.42
							Replicate	-0.82	0.12	0.44	0.35–0.56	8.89×10^{-12} ****
AGAP008307	60%	*E. coli*	4	76	73	10 days	Gene KD	-0.001	0.21	1.00	0.66–1.51	1.00
							Replicate	0.03	0.03	1.03	0.98–1.09	0.27
		S. marcescens	4	80	50	10 days	Gene KD	0.14	0.21	1.15	0.76–1.72	0.51
							Replicate	-0.10	0.05	0.90	0.81–1.00	0.04 *
GAM1	37%	*E. faecalis*	3	169	120	All	Gene KD	0.05	0.12	1.05	0.83–1.33	0.07
							Replicate	-0.20	0.07	0.82	0.71–0.94	0.01 **
						10 days	Gene KD	0.26	0.26	1.29	0.78–2.14	0.32
							Replicate	-0.10	0.15	0.90	0.67–1.22	0.52
		S. marcescens	3	181	199	All	Gene KD	-0.06	0.10	0.94	0.77–1.15	0.53
							Replicate	-0.01	0.07	0.99	0.87–1.13	0.90
						10 days	Gene KD	-0.26	0.21	0.77	0.51–1.16	0.21
							Replicate	0.21	0.14	1.23	0.94–1.61	0.14
		E. coli	3	163	184	All	Gene KD	-0.07	0.11	0.93	0.75–1.16	0.52
							Replicate	-0.21	0.08	0.81	0.69–0.96	0.01 *
						10 days	Gene KD	-0.20	0.13	0.81	0.63–1.06	0.13
							Replicate	-0.29	0.10	0.75	0.61–0.91	0.004 **
		S. aureus	3	159	148	All	Gene KD	0.19	0.12	1.21	0.96–1.51	0.10
							Replicate	-0.77	0.07	0.46	0.40–0.53	2.00×10^{-16} ****
						10 days	Gene KD	0.17	0.19	1.19	0.82–1.71	0.36
							Replicate	-1.03	0.13	0.36	0.28–0.46	1.61×10^{-15} ****

Gene	KD Eff.	Bacteria	R	n	n (controls)	Time	Cox Proportional Hazard Model					
							Covariates	Coefficient	SE	HR	95% CI	*p*-value
AGAP009313	53%	*E. faecalis*	3	175	148	All	Gene KD	0.07	0.11	1.07	0.86–1.33	0.56
							Replicate	0.21	0.07	1.23	1.07–1.43	0.005 **
						10 days	Gene KD	0.06	0.24	1.06	0.66–1.70	0.81
							Replicate	-0.47	0.15	0.63	0.46–0.85	0.002 **
		S. marcescens	3	145	168	All	Gene KD	0.25	0.11	1.28	1.02–1.62	0.03 *
							Replicate	0.01	0.07	1.01	0.88–1.16	0.85
						10 days	Gene KD	0.58	0.24	1.78	1.10–2.87	0.02 *
							Replicate	-0.60	0.15	0.55	0.41–0.73	5.14×10^{-5} ****
		E. coli	3	146	129	All	Gene KD	0.32	0.12	1.38	1.09–1.76	0.01 **
							Replicate	0.47	0.08	1.59	1.37–1.86	2.96×10^{-9} ****
						10 days	Gene KD	0.25	0.15	1.29	0.96–1.73	0.09
							Replicate	0.49	0.09	1.64	1.36–1.97	1.72×10^{-7} ****
		S. aureus	3	174	142	All	Gene KD	0.15	0.11	1.16	0.93–1.45	0.2
							Replicate	0.64	0.08	1.89	1.61–2.22	7.74×10^{-15} ***
						10 days	Gene KD	0.51	0.17	1.67	1.20–2.24	0.003 **
							Replicate	0.74	0.12	2.09	1.64–2.65	1.37×10^{-9} ****
AGAP009635	66%	*E. coli*	4	147	91	10 days	Gene KD	0.09	0.15	1.09	0.82–1.45	0.55
							Replicate	0.04	0.03	1.05	1.00–1.10	0.08
		S. marcescens	5	80	39	10 days	Gene KD	-0.10	0.22	0.91	0.59–1.39	0.65
							Replicate	-0.03	0.06	0.97	0.86–1.10	0.68
AGAP011630	51%	*E. coli*	3	104	108	10 days	Gene KD	-0.21	0.16	0.81	0.59–1.12	0.21
							Replicate	-0.03	0.12	0.97	0.76–1.23	0.80
		S. marcescens	3	98	101	10 days	Gene KD	-0.32	0.18	0.73	0.52–1.03	0.07
							Replicate	-0.17	0.13	0.85	0.65–1.10	0.21
AGAP013543	60%	*E. faecalis*	3	156	148	All	Gene KD	-0.01	0.12	0.99	0.79–1.24	0.90
							Replicate	0.13	0.08	1.14	0.98–1.33	0.09
						10 days	Gene KD	-0.20	0.23	0.82	0.52–1.23	0.4
							Replicate	-0.58	0.15	0.56	0.42–0.76	1.30×10^{-4} ***
		S. marcescens	3	125	168	All	Gene KD	0.02	0.12	1.02	0.81–1.29	0.87
							Replicate	0.01	0.07	1.00	0.88–1.15	0.94
						10 days	Gene KD	0.17	0.23	1.19	0.75–1.87	0.46
							Replicate	-0.46	0.13	0.63	0.48–0.82	0.001 ***
		E. coli	3	144	129	All	Gene KD	0.08	0.12	1.08	0.85–1.38	0.53
							Replicate	0.46	0.08	1.59	1.36–1.85	5.38×10^{-9} ****
						10 days	Gene KD	-0.02	0.15	0.98	0.74–1.31	0.9
							Replicate	0.47	0.09	1.60	1.33–1.92	3.62×10^{-7} ****
		S. aureus	3	150	142	All	Gene KD	-0.14	0.12	0.87	0.69–1.10	0.24
							Replicate	0.46	0.09	1.58	1.33–1.87	2.45×10^{-7} ****
						10 days	Gene KD	-0.08	0.16	0.92	0.68–1.26	0.61
							Replicate	0.59	0.11	1.80	1.44–2.25	2.49×10^{-7} ****

KD Eff. = KD efficiency calculated from whole mosquitoes; *R = number of replicates, n = number of mosquitoes, controls refer to* dsLacZ*-injected control mosquitoes; SE = standard error, HR = hazard ratio with respect to control mosquitoes, 95% CI = 95% confidence intervals,* * p < *0.05,* ** p < *0.01,* *** p < *0.001,* **** p < *0.0001.*

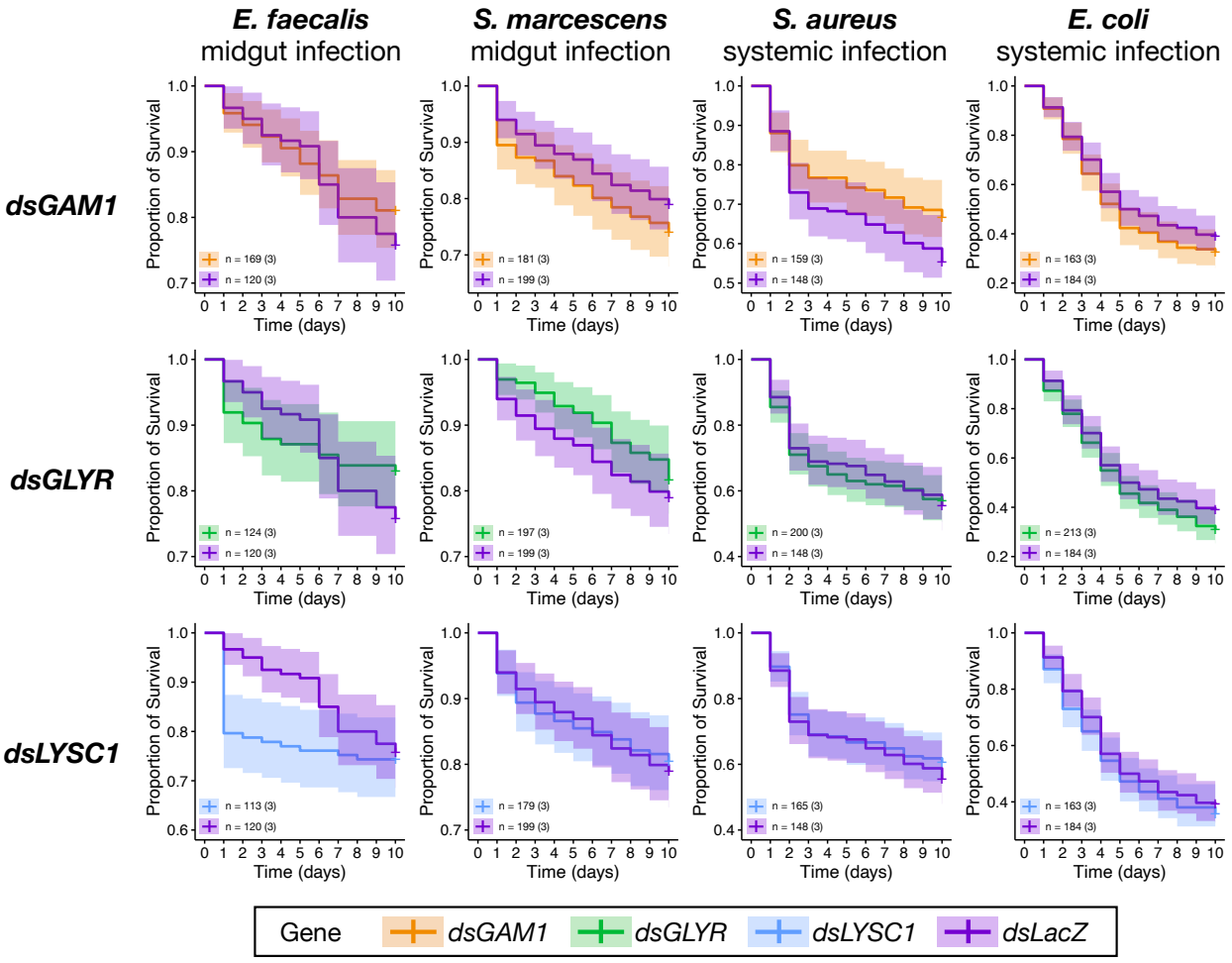


Figure F.1. Kaplan–Meier survival curves (10 days) of mosquitoes post-putative and known AMP KD. *Putative and known AMPs were knocked down via RNAi, and mosquitoes were infected with one species of bacteria two days later. The left two species of bacteria were introduced to the mosquito through a sugar meal. The right two species of bacteria were introduced via thoracic injection. Mosquito deaths were recorded daily until 10 days post-infection. Shaded regions represent 95% CIs. n = number of mosquitoes, parentheses indicate number of replicates. Gene KD did not significantly impact survival outcomes (Cox proportional hazard analyses,* p > *0.05). Full model results are listed in Table F.1.*

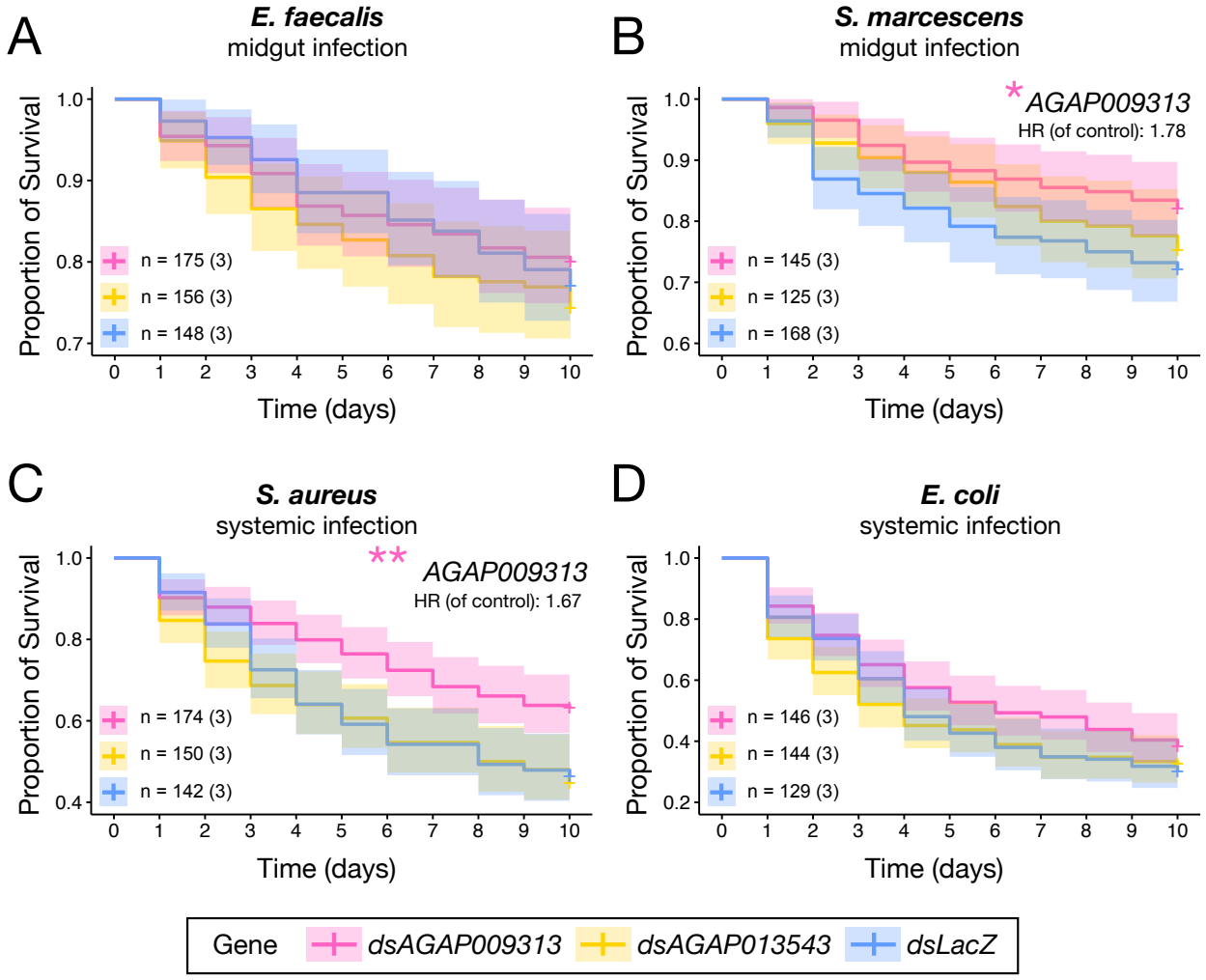


Figure F.2. Kaplan–Meier survival curves (10 days) of mosquitoes post- cardia gene KD. *Putative cardia AMPs were knocked down via RNAi, and mosquitoes were infected with one species of bacteria two days later. The top two species of bacteria were introduced to the mosquito through a sugar meal. The bottom two species of bacteria were via thoracic injection. Mosquito deaths were recorded daily until 10 days post-infection. Shaded regions represent 95% CIs. n = number of mosquitoes, parentheses indicate number of replicates. Cox proportional hazard analyses, * $p < 0.05$, ** $p < 0.01$. Full model results are listed in Table F.1.*

Appendix G

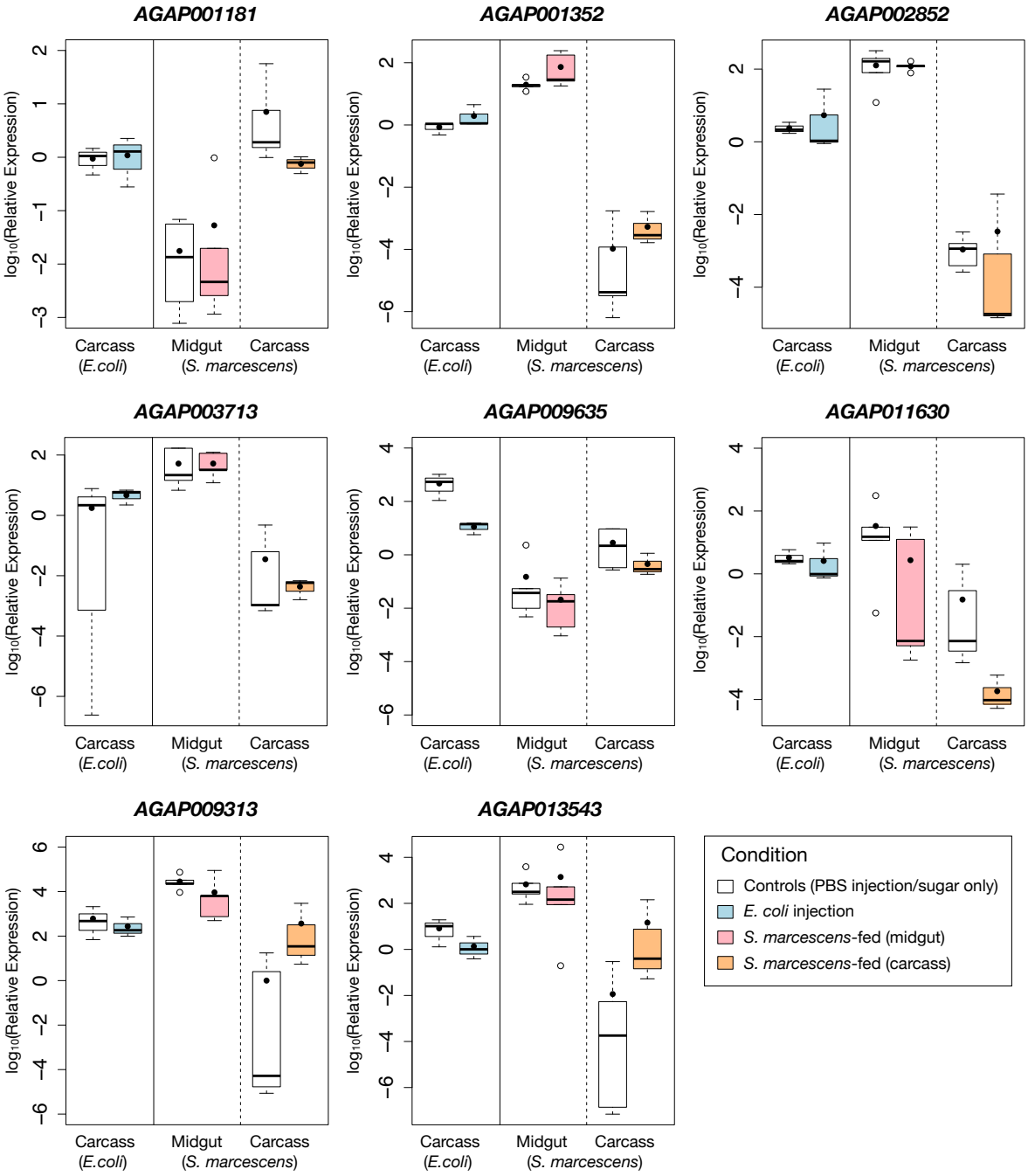


Figure G.1. Relative expression of remaining putative AMPs post-bacterial challenge*. The two genes at the bottom were included as putative AMPs are they are constitutively and highly expressed in the cardia. Controls for* E. coli*-injected mosquitoes were PBS-injected. Controls for* S. marcescens*-fed mosquitoes were fed with sugar only. Relative expression values were* log_{10}*-transformed for presentation purposes and are relative to a pool of mosquito cDNA. A total of 3–5 replicates were performed with 10–15 mosquito individuals per condition for each replicate. Black bars represent medians, dots represent means. No significant differences in the original relative expression values were observed between infected and control mosquitoes (Student's* t*-tests or Wilcoxon rank-sum tests,* $p > 0.05$*).*

Appendix H

Table H.1.* F-*test statistics of GLMs with a quasi-Poisson distribution examining the effect of treatment condition and replicate on* P. falciparum *infection intensity.

Experiment	Gene	Variable	df	Deviance	Residual df	Residual Deviance	*F*	*p*-value
Putative AMP KD	*AGAP001181*	Gene KD	1	267.06	169	4693.9	10.76	0.001 *
		Replicate	5	1306.03	164	3387.9	10.53	9.60×10^{-9} ****
		Gene KD:Replicate	5	156.03	159	3231.8	1.26	0.28
	AGAP001352	Gene KD	1	35.78	73	1838.2	1.23	0.27
		Replicate	3	405.03	70	1433.2	4.65	0.01 **
		Gene KD:Replicate	3	5.99	67	1437.2	0.07	0.98
	AGAP001508	Gene KD	1	26.94	83	877.1	2.36	0.13
		Replicate	3	161.15	80	715.9	4.71	0.004 **
		Gene KD:Replicate	3	100.61	77	615.3	2.94	0.04 *
	AGAP002852	Gene KD	1	0.19	123	2985.2	0.01	0.93
		Replicate	4	821.43	119	2163.8	7.48	2.15×10^{-5} ****
		Gene KD:Replicate	4	103.97	115	2059.8	0.95	0.44
	AGAP003713	Gene KD	1	377.87	95	2086.2	18.50	4.32×10^{-5} ****
		Replicate	3	457.39	92	1628.8	7.47	1.62×10^{-4} ***
		Gene KD:Replicate	3	373.95	89	1254.9	6.11	7.94×10^{-4} ***
	KIN1	Gene KD	1	44.81	79	825.2	3.20	0.08
		Replicate	2	43.56	77	781.7	1.56	0.22
		Gene KD:Replicate	2	67.12	75	714.5	2.40	0.10
	GLYR	Gene KD	1	63.48	138	2097.8	3.40	0.06
		Replicate	5	254.96	133	1842.8	2.73	0.02 *
		Gene KD:Replicate	5	224.23	128	1618.6	2.40	0.04 *
	ALAR	Gene KD	1	49.53	107	1125.1	4.54	0.04 *
		Replicate	5	296.32	102	828.8	5.43	1.91×10^{-4} ***
		Gene KD:Replicate	5	17.53	97	811.3	0.32	0.90
	AGAP008307	Gene KD	1	23.79	98	1376.3	1.29	0.26
		Replicate	3	23.96	95	1352.4	0.43	0.73
		Gene KD:Replicate	3	169.77	93	1182.6	4.61	0.01 *
	AGAP009635	Gene KD	1	74.16	120	3314.8	2.82	0.10
		Replicate	4	1159.85	116	2155.0	11.03	1.40×10^{-7} ****
		Gene KD:Replicate	4	88.82	112	2066.2	0.85	0.50
	AGAP011630	Gene KD	1	7.26	106	929.3	0.68	0.41
		Replicate	4	136.34	102	792.9	3.20	0.02 *
		Gene KD:Replicate	4	66.15	98	726.8	1.56	0.19
	AGAP009313	Gene KD	1	3.45	105	913.2	0.39	0.54
		Replicate	4	181.38	101	731.8	5.05	0.001 ***
		Gene KD:Replicate	4	9.51	97	722.3	0.26	0.90
	AGAP013543	Gene KD	1	69.21	95	2845.0	2.40	0.13
		Replicate	4	943.20	91	1901.8	8.17	1.22×10^{-5} ****
		Gene KD:Replicate	4	89.25	87	1812.5	0.77	0.55

Experiment	Gene	Variable	df	Deviance	Residual df	Residual Deviance	*F*	*p*-value
Synthetic KIN1 Spike-In		Condition	1	34.10	122	2978.8	1.63	0.21
		Replicate	1	1044.80	121	1934.0	49.99	1.11×10^{-10} ****
		Condition:Replicate	1	3.10	120	1930.9	0.15	0.70
Exogenous MF Spike-In		Condition	1	45.84	134	1143.7	4.73	0.03 *
		Replicate	1	210.65	133	933.0	21.75	7.5×10^{-6} ****
		Condition:Replicate	1	9.14	132	923.9	0.94	0.33

MF = methyl farnesoate, df = degrees of freedom; * p < *0.05,* ** p < *0.01,* *** p < *0.001,* **** p < *0.0001.*

Appendix I

Table I.1. Kruskal–Wallis and Dunn's test statistics with Benjamini–Hochberg false discovery rate corrections comparing relative gene expression between various mosquito life stages.

Comparison		Kruskal–Wallis Test Statistics			Dunn's Test Statistics	
		χ^2	df	*p*-value	*Z*	*p*-value
ALAR		34	6	6.69×10^{-6} ****		
Adult Females (BF)	L2 Larvae				2.21	0.06
Adult Females (BF)	Eggs				3.08	0.01 **
L2 Larvae	Eggs				0.87	0.45
Adult Females (BF)	Adult Females (SF)				0.28	0.82
L2 Larvae	Adult Females (SF)				-1.93	0.09
Eggs	Adult Females (SF)				-2.80	0.02 *
Adult Females (BF)	L2 Larvae				3.27	0.01 **
L2 Larvae	L2 Larvae				1.06	0.36
Eggs	L2 Larvae				0.19	0.85
Adult Females (SF)	Adult Males				2.99	0.01 **
Adult Females (BF)	Adult Males				-1.22	0.33
L2 Larvae	Adult Males				-3.44	0.01 **
Eggs	Adult Males				-4.31	1.74×10^{-4} ***
Adult Females (SF)	Adult Males				-1.51	0.21
L2 Larvae	Adult Males				-4.49	1.46×10^{-4} ***
Adult Females (BF)	Pupae				1.11	0.35
L2 Larvae	Pupae				-1.11	0.38
Eggs	Pupae				-1.98	0.09
Adult Females (SF)	Pupae				0.82	0.45
L2 Larvae	Pupae				-2.16	0.06
Adult Males	Pupae				2.33	0.05
GAM1		32.2	6	1.51×10^{-5} ****		
Adult Females (BF)	L2 Larvae				0.87	0.47
Adult Females (BF)	Eggs				2.78	0.02 *
L2 Larvae	Eggs				1.91	0.12
Adult Females (BF)	Adult Females (SF)				-0.75	0.53
L2 Larvae	Adult Females (SF)				-1.62	0.18
Eggs	Adult Females (SF)				-3.53	0.003 **
Adult Females (BF)	L2 Larvae				1.48	0.22
L2 Larvae	L2 Larvae				0.61	0.60
Eggs	L2 Larvae				-1.29	0.26
Adult Females (SF)	Adult Males				2.24	0.07
Adult Females (BF)	Adult Males				-2.24	0.08
L2 Larvae	Adult Males				-3.11	0.01 **
Eggs	Adult Males				-5.01	1.1×10^{-5} ****
Adult Females (SF)	Adult Males				-1.48	0.21
L2 Larvae	Adult Males				-3.72	0.002 **
Adult Females (BF)	Pupae				-0.49	0.65
L2 Larvae	Pupae				-1.36	0.24
Eggs	Pupae				-3.27	0.01 **
Adult Females (SF)	Pupae				0.26	0.80
L2 Larvae	Pupae				-1.98	0.11
Adult Males	Pupae				1.74	0.16

Comparison		Kruskal–Wallis Test Statistics			Dunn's Test Statistics	
		χ^2	df	*p*-value	*Z*	*p*-value
GLYR		32.7	6	1.18×10^{-5} ****		
Adult Females (BF)	L2 Larvae				-4.28	1.94×10^{-4} ***
Adult Females (BF)	Eggs				0.26	0.80
L2 Larvae	Eggs				4.54	1.17×10^{-4} ***
Adult Females (BF)	Adult Females (SF)				-1.53	0.20
L2 Larvae	Adult Females (SF)				2.75	0.02 *
Eggs	Adult Females (SF)				-1.79	0.15
Adult Females (BF)	L2 Larvae				-2.59	0.03 *
L2 Larvae	L2 Larvae				1.69	0.17
Eggs	L2 Larvae				-2.85	0.02 *
Adult Females (SF)	Adult Males				-1.06	0.36
Adult Females (BF)	Adult Males				-1.04	0.35
L2 Larvae	Adult Males				3.25	0.01 **
Eggs	Adult Males				-1.29	0.27
Adult Females (SF)	Adult Males				0.49	0.69
L2 Larvae	Adult Males				1.55	0.21
Adult Females (BF)	Pupae				-3.01	0.01 *
L2 Larvae	Pupae				1.27	0.27
Eggs	Pupae				-3.27	0.01 **
Adult Females (SF)	Pupae				-1.48	0.21
L2 Larvae	Pupae				-0.42	0.71
Adult Males	Pupae				-1.98	0.11
KIN1		25.2	6	3.08×10^{-4} ****		
Adult Females (BF)	L2 Larvae				-1.55	0.28
Adult Females (BF)	Eggs				2.61	0.03 *
L2 Larvae	Eggs				4.16	6.54×10^{-4} ***
Adult Females (BF)	Adult Females (SF)				-0.09	0.93
L2 Larvae	Adult Females (SF)				1.46	0.28
Eggs	Adult Females (SF)				-2.71	0.03 *
Adult Females (BF)	L2 Larvae				1.44	0.26
L2 Larvae	L2 Larvae				2.99	0.01 *
Eggs	L2 Larvae				-1.18	0.39
Adult Females (SF)	Adult Males				1.53	0.26
Adult Females (BF)	Adult Males				-1.08	0.42
L2 Larvae	Adult Males				0.47	0.71
Eggs	Adult Males				-3.69	0.002 **
Adult Females (SF)	Adult Males				-0.99	0.42
L2 Larvae	Adult Males				-2.52	0.04 *
Adult Females (BF)	Pupae				-0.49	0.72
L2 Larvae	Pupae				1.06	0.41
Eggs	Pupae				-3.11	0.01 *
Adult Females (SF)	Pupae				-0.40	0.72
L2 Larvae	Pupae				-1.93	0.14
Adult Males	Pupae				0.59	0.69

*BF = blood-fed, SF = sugar-fed; df = degrees of freedom; * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001.*

Table I.2. Kruskal–Wallis and Dunn's test statistics with Benjamini–Hochberg false discovery rate corrections comparing relative gene expression in mosquito tissues of the same feeding condition.

Comparison		Kruskal–Wallis Test Statistics			Dunn's Test Statistics	
		χ^2	df	*p*-value	*Z*	*p*-value
ALAR (SF)		14.1	3	2.78×10^{-3} **		
Anterior (SF)	Cardia (SF)				2.4	0.03 *
Anterior (SF)	Posterior (SF)				4.1	2.24×10^{-4} ***
Cardia (SF)	Posterior (SF)				1.7	0.14
Anterior (SF)	Carcass (SF)				1.1	0.27
Cardia (SF)	Carcass (SF)				-1	0.21
Posterior (SF)	Carcass (SF)				-3	0.01 **
ALAR (BF)		18.9	3	2.76×10^{-4} ***		
Anterior (BF)	Cardia (BF)				0.4	0.65
Anterior (BF)	Posterior (BF)				2.8	0.02 *
Cardia (BF)	Posterior (BF)				2.3	0.04 *
Anterior (BF)	Carcass (BF)				-1	0.42
Cardia (BF)	Carcass (BF)				-1	0.25
Posterior (BF)	Carcass (BF)				-4	0.001 **
GAM1 (SF)		19.7	3	1.95×10^{-4} ***		
Anterior (SF)	Cardia (SF)				-1	0.30
Anterior (SF)	Posterior (SF)				2.9	0.01 **
Cardia (SF)	Posterior (SF)				4	0.00
Anterior (SF)	Carcass (SF)				1.8	0.10
Cardia (SF)	Carcass (SF)				3	0.01 **
Posterior (SF)	Carcass (SF)				-1	0.29
GAM1 (BF)		15.0	3	0.002 **		
Anterior (BF)	Cardia (BF)				0	0.97
Anterior (BF)	Posterior (BF)				2.6	0.02 *
Cardia (BF)	Posterior (BF)				2.6	0.02 *
Anterior (BF)	Carcass (BF)				2.7	0.04 *
Cardia (BF)	Carcass (BF)				2.7	0.02 *
Posterior (BF)	Carcass (BF)				0.1	1.00
GLYR (SF)		15.5	3	0.001 **		
Anterior (SF)	Cardia (SF)				-0	0.74
Anterior (SF)	Posterior (SF)				2.3	0.04 *
Cardia (SF)	Posterior (SF)				2.7	0.02 *
Anterior (SF)	Carcass (SF)				-2	0.20
Cardia (SF)	Carcass (SF)				-1	0.28
Posterior (SF)	Carcass (SF)				-4	7.46×10^{-4} ***
GLYR (BF)		11.1	3	0.01 *		
Anterior (BF)	Cardia (BF)				-1	0.31
Anterior (BF)	Posterior (BF)				1.8	0.14
Cardia (BF)	Posterior (BF)				3.1	0..01 *
Anterior (BF)	Carcass (BF)				-1	0.47
Cardia (BF)	Carcass (BF)				0.4	0.68
Posterior (BF)	Carcass (BF)				-3	0.02 *
KIN1 (SF)		4.93	3	0.18		
Anterior (SF)	Cardia (SF)				0.3	0.74
Anterior (SF)	Posterior (SF)				2.2	0.06
Cardia (SF)	Posterior (SF)				1.8	0.08
Anterior (SF)	Carcass (SF)				-2	0.10
Cardia (SF)	Carcass (SF)				-2	0.09
Posterior (SF)	Carcass (SF)				-4	2.79×10^{-4} ***
KIN1 (BF)		11.07	3	0.01 *		
Anterior (BF)	Cardia (BF)				0.2	1.00
Anterior (BF)	Posterior (BF)				2	0.30
Cardia (BF)	Posterior (BF)				1.8	0.24
Anterior (BF)	Carcass (BF)				0.3	1.00
Cardia (BF)	Carcass (BF)				0.1	0.93
Posterior (BF)	Carcass (BF)				-2	0.19

BF = blood-fed, SF = sugar-fed; df = degrees of freedom; * p < *0.05,* ** p < *0.01,* *** p < *0.001.*

Appendix J

Table J.1. Wilcoxon signed-rank test statistics of the* P. falciparum *asexual stage parasitemia over time following treatment with artemisinin or MF* in vitro *compared to that of controls (DMSO-treated).

	Wilcoxon Signed-Rank Test Statistics	
Treatment	*V*	*p*-value
100 μM artemisinin	3	0.005 **
10 μM artemisinin	10	0.025 *
1 μM artemisinin	2	0.011 *
0.1 μM artemisinin	9	0.037 *
100 μM methyl farnesoate	4	0.011 *
10 μM methyl farnesoate	11	0.103
1 μM methyl farnesoate	25.5	0.308
0.1 μM methyl farnesoate	30.5	0.859

*Wilcoxon signed-rank tests, * $p < 0.05$, ** $p < 0.01$.*