Transcriptomic analysis of the digestive system of the desert locust, *Schistocerca gregaria*

Michiel HOLTOF

Supervisor:
Prof. Dr. Jozef Vanden Broeck

Examination Committee:
Prof. Dr. J. Billen, *chair*
Prof. Dr. R. Huybrechts
Prof. Dr. D. Bylemans
Dr. M. Peferoen
Dr. J. Spit
Dr. P. Van Wielendaele

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor of Science: Biochemistry and Biotechnology

November 2019

Doctoraatsproefschrift aan de faculteit Wetenschappen van de KU Leuven

Uitgegeven in eigen beheer, Michiel Holtof, Leuven, Belgium

Dankwoord

Het is zover. De langverwachte dag waarop ik mijn dankwoord mag schrijven en bijgevolg de laatste hand aan deze thesis leg, is aangebroken. En wat een rit is dit geweest! Dit gehele proefstuk zou nooit tot stand gekomen zijn zonder de hulp en aanwezigheid van een hele reeks mensen. Ik neem in dit dankwoord dan ook met veel plezier de tijd om iedereen even in de schijnwerpers te plaatsen.

De eerste persoon die ik wil bedanken is mijn promotor, Jef. Bedankt om mij de kans te geven om ongestoord onderzoek te kunnen voeren in uw onderzoeksgroep. De laatste jaren hebben mij enorm veel bijgeleerd over het zelfstandig uitvoeren van wetenschappelijk onderzoek; een onvervangbare leerschool.

Doorheen de jaren heb ik het voorrecht gehad om op regelmatige tijdsstippen de nodige hulp en input te mogen ontvangen van mijn Supervisory Committee. Beste Jef, Roger, Pieter, en Jornt, via deze weg wil ik mijn dankbaarheid uitdrukken voor jullie advies gedurende de afgelopen jaren.

Jornt, bedankt om mij als thesisstudent te introduceren in het labo en de wereld van het wetenschappelijke onderzoek. Jouw enthousiasme heeft ervoor gezorgd dat ook ik gebeten werd door de onderzoeksmicrobe. Bedankt om mij nadien met raad en daad bij te staan tijdens de eerste jaren van mijn doctoraat, en bedankt om mij doorheen de jaren te blijven stimuleren om nooit op te geven. Het is dan ook geheel mijn genoegen om je na al die jaren eindelijk een volledige tabel met *S. gregaria* serine proteasen te kunnen presenteren.

Verder wil ik natuurlijk ook graag al mijn juryleden ontzettend bedanken. Jullie namen de tijd om mijn werk kritisch na te lezen en te beoordelen. Door jullie gezamenlijke kennersblik zijn zowel de helderheid alsook de kwaliteit van dit manuscript toegenomen. Ondanks de nervositeit die een preliminaire verdediging met zich meebrengt, heb ik ervan genoten om samen met jullie mijn werk te overlopen en te bespreken.

Bedankt aan alle collega's van de HSVTBA labo's (wat een mond vol is dit geworden doorheen de jaren) om de afgelopen jaren een open en toffe werksfeer te creëren. Een extra bedankje aan de symbiobar, de talloze recepties en afterwork activiteiten die steeds voor de nodige ontspanning zorgden.

Ik heb het geluk gekend om talloze collega's vandaag mijn vrienden te mogen noemen. Evert, hoe zou ik die laatste jaren zonder jou zijn doorgekomen? Ontelbare keren hebben we samen gelachen, plezier gemaakt, en onze gedachten verzet. Meermaals bleek dit voor mij het ideale recept voor een geslaagde werkdag. Moest onze vriendschap ooit verfilmd worden, dan hoop ik dat Nadezhda Goncharova de alwetende verteller is. Sven, dankzij jou weet ik dat elk gek idee in mijn hoofd werkelijkheid gemaakt kan worden via Photoshop. Het leukste vond ik dat jij deze ideeën meestal niet eens zo gek vond. Timon, bedankt om altijd klaar te staan als vriend en als collega. Jullie drie haalden er meermaals (onbewust) het randje af voor mij, iets waar ik jullie enorm dankbaar voor ben. Ik mis onze Alma-uitstapjes nu al.

Rik, bedankt om mij te introduceren aan de wondere wereld van NGS. Van Leuven, via Leipzig, tot in Porto: we hebben toffe momenten beleefd! Ook een extra bedankje voor al de moeite en tijd die je gestoken hebt in het samenstellen van het *S. gregaria* referentietranscriptoom. Een pionierswerk dat al te vaak te snel als vanzelfsprekend wordt beschouwd. Indien we dit niet hadden, was er alvast geen sprake geweest van mijn onderzoek.

02.10! Zonder deze bende bestond dit proefstuk met zekerheid niet! Charline, Katleen, Els en Jornt, vanuit de grond van mijn hart: bedankt om de afgelopen jaren van ons bureau mijn home

away from home te maken. Naar het werk gaan is heus niet zo erg als je omringt bent door zulke fijne mensen als jullie.

Mijn vrienden van T&P: zijn we nu een voetbal-, bowling-, boks-, of biljartclub? Ik ben het ondertussen even kwijt. Angelo, Benoit, Bert, Dieter (Brasseur), Keaton, Gionatan, Michiel (ik reken jou erbij), Tom en Wim: vriendschappen als deze helpen je doorheen de moeilijkste momenten. Dat is exact wat jullie gedaan hebben.

Ik wil ook heel graag deze kans nemen om mijn fantastische familie en schoonfamilie te bedanken voor alle steun en motivatie die ik gedurende de afgelopen jaren heb mogen voelen. De wetenschap dat zoveel mensen achter mij staan en in mij geloven, heeft mij zeker meermaals geholpen tijdens dit doctoraat.

Mama en papa, bedankt om mij altijd gesteund te hebben tijdens mijn studies. Zonder jullie zou ik nooit aan dit avontuur begonnen zijn. Dankzij jullie sta ik nu waar ik sta, en hier ben ik jullie eindeloos dankbaar voor. Mama, Celien, de laatste jaren waren niet gemakkelijk voor ons. Desondanks heb ik steeds de nodige steun van jullie blijven voelen om dit hoofdstuk in mijn leven tot een goed einde te brengen. Daar ben ik jullie ontzettend dankbaar voor.

Liefste Jana, ik zal onze honderden koosnaampjes even achterwege laten, ik weet niet eens waar te beginnen met jou te bedanken. Ik kan je eindeloos bedanken voor al je steun, al de motivatie, al de leuke momenten van afleiding, al de liefde, *etc*. Laat me je daarom bedanken voor wie je voor mij bent. Ik kijk reikhalzend uit naar al de avonturen die we nog samen gaan beleven!

Ik wil graag nog even de hoofdrolspelers van dit verhaal in de bloemetjes zetten: de sprinkhanen. Liefste sprinkhanen, bedankt voor al jullie opofferingen in de naam van de wetenschap. Jullie leven voor eeuwig voort in mijn herinneringen.

Tot slot wil ik de laatste alinea van dit dankwoord opdragen aan mijn papa. Papa, je was altijd ontzettend trots dat ik een doctoraat deed en loofde mijn doorzettingskracht. Dat is een grote motivatie geweest voor mij. Jij en mama hebben mij steeds de vrijheid gegeven om te doen wat ik zelf wilde, zodat ik zelf kon ontdekken wie ik ben en waar mijn interesses liggen. Je hebt mij altijd en overal gesteund. De laatste jaren van dit doctoraat waren bijzonder zwaar zonder jouw steun. Ik mis je elke dag. Het idee om jouw trotse blik met dit boekje in je handen te kunnen zien, heeft mij meermaals gepusht om dit tot een goed einde te brengen.
Ik draag dit werk op aan jou.

Deze thesis is opgedragen aan mijn papa. Je wordt ontzettend gemist.

Sky of blackness and sorrow (a dream of life)
Sky of love, sky of tears (a dream of life)
Sky of glory and sadness (a dream of life)
Sky of mercy, sky of fear (a dream of life)
Sky of memory and shadow (a dream of life)
Your burnin' wind fills my arms tonight
Sky of longing and emptiness (a dream of life)
Sky of fullness, sky of blessed life (a dream of life)

Come on up for the rising
Come on up, lay your hands in mine

.

The rising – Bruce Springsteen

Table of contents

List of abbreviations

20E:	20-hydroxyecdysone
A:	Alanine
AA:	Amino acid
ABC:	ATP binding cassette
Ago2:	Argonaute-2
BLAST:	Basic local alignment search tool
BT:	*Bacillus thuringiensis*
BWT:	Burrows-Wheeler transform algorithm
CAATCH1:	Cation-anion-activated amino acid transporter/channel
CAT:	Cationic amino acid transporter
CBD:	Chitin binding domain
CBP:	Chitin binding protein
CDS:	Coding sequence
CE:	Carboxylesterase
ChtBD2:	Cysteine-containing type-2 chitin-binding domain
Chy:	Chymotrypsin
CNS:	Central nervous system
COG:	Clusters of Orthologous Groups
CPM:	Count Per Million
CSP:	Chemosensory protein
CV:	Celangulin V
D:	Aspartate
DAG:	Diacylglycerol
Dcr2:	Dicer-2
DDT:	Dichlorodiphenyltrichloroethane
***Dm*OPT1:**	*Drosophila melanogaster* oligopeptide transporter 1
DNA:	Deoxyribonucleic acid
dNTP:	Deoxynucleoside triphosphate
dsRNA:	Double-stranded RNA
EB (Chapter 1):	Enteroblast
EB (Chapter 4):	Empirical Bayes
EC:	Enterocyte cells
EcR:	Ecdysone receptor
EEC:	Enteroendocrine cells
EM:	Expectation Maximization
EPA:	Environmental Protection Agency of the United States
ER:	Endoplasmic reticulum
ERAP:	Endoplasmic reticulum aminopeptidase
EST:	Expressed sequence tag
EU:	European Union
FABP:	Fatty acid binding proteins
FAO:	Food and Agricultural Organization of the United Nations
FATP:	Fatty acid transport protein
FC:	Fold change
FDR:	False discovery rate
G:	Glycine
GFP:	Green fluorescent protein
GH:	Glycosyl hydrolase
GLUT:	Glucose facilitator

GO:	Gene Ontology
GST:	Glutathione S-transferase
H^+ V-ATPase:	Vacuolar-type H^+-ATPase
HAT:	Heterodimeric amino acid transporter
Hex:	Hexamerin
HMM:	Hidden Markov Model
Hsp:	Heat shock protein
HUGO:	Human Genome Organization
ICS:	Intestinal stem cell
IgE:	Immunoglobulin E
iNAT:	Insect nutrient amino acid transporter
Indy:	Protein I'm not dead yet
IPM:	Integrated pest management
JH:	Juvenile hormone
JHBP:	Juvenile hormone binding protein
KAAT1:	K^+-coupled amino acid transporter
KEGG:	Kyoto Encyclopedia of Genes and Genomes
Kr-h1:	Krüppel homolog 1
MAFFT:	Multiple Alignment using Fast Fourier Transform
MAG:	Monoacylglycerol
Met:	Methoprene-tolerant
MFS:	Major facilitator superfamily
miRNA:	Micro RNA
mRNA:	Messenger RNA
MRP:	Multidrug resistance protein
MWW:	Mann-Whitney-Wilcoxon
NAT:	Nutrient amino acid transporter
NB:	Negative binominal
NBD:	Nucleotide-binding domain
NCBI:	National Center for Biotechnology Information
ncRNA:	Non-coding RNA
NGS:	Next-generation sequencing
NPC1:	Niemann-Pick C1
NSP:	Nitrile-specifier protein
ORF:	Open reading frame
P450:	Cytochrome P450 monooxygenase
PA1b:	Pea albumin 1b
PAR-2:	Protease-activated receptor-2
PC:	Principal Component
PCA:	Principal Component Analyses
PG:	Prothoracic gland
P-gp:	P-glycoprotein
PI:	Protease inhibitor
PM:	Peritrophic membrane
PMP:	Peritrophic matrix protein
proPO:	Prophenoloxidase
PSTI:	Pancreatic secretory trypsin inhibitor
PUFA:	Polyunsaturated fatty acid
Q:	Phred score
QL:	Quasi-likelihood
RISC:	RNA induced silencing complex

Rlog:	Regularized log
RNA:	Ribonucleic acid
RNAi:	RNA interference
RNA-Seq:	RNA sequencing
RQ:	Relative quantity
RSEM:	RNA-Seq by Expectation Maximization
RXR/USP:	Retinoid-X-receptor/Ultraspiracle
S:	Serine
SCP:	Sterol carrier protein
SCP2:	Sterol carrier protein-2
SCPx:	Sterol carrier protein-x
SCRT:	Sucrose transporter protein
Serpin:	Serine protease inhibitor
SGLT:	Sodium-driven glucose symporter
SGPP:	Pacifastin-related peptide precursor
siRNA:	Small interfering RNA
SLC:	Solute Carrier
SNS:	Stomatogastric nervous system
Sog:	Suboesophageal ganglion
SSD:	Sterol-sensing domain
SUH:	Sucrose-specific sucrases
TAG:	Triacylglycerol
TMD:	Transmembrane domain (TMD)
TMM:	Trimmed Mean of M-values
TO:	Protein Takeout
Tret1:	Facilitated trehalose transporter 1
Try:	Trypsin
UGT:	Uridine diphosphate-glucuronosyltransferase
Uniprot:	Universal Protein Resource
VNC:	Ventral nerve cord

Abstract

Insects are among the most successful organisms on the planet. One explanation for their success is their extraordinary ability to successfully consume a wide range of foods. Like all heterotrophic organisms, insects need to acquire vital nutrients from their diet. The central organ for food digestion and absorption of nutrients is the gastrointestinal tract. This organ's principal functions are to mediate the efficient digestion of food and to protect the organism against harmful chemicals, microorganisms, and mechanical damage from the food. These functions are achieved through regional differentiation of the alimentary canal, as well as highly flexible adaptations to the consumed diets, both at anatomical and molecular levels. This central role of the digestive tract in the insect's life history has made it an important subject of study. Numerous studies describing the general gut morphology and associated digestive mechanisms of various insects exist. In contrast, the molecular patterns underlying digestion and nutrient uptake in insects are still poorly and only partially characterized.

Insects have a high socio-economic impact. Many are considered to be pests, endangering the health and livelihood of a large proportion of the world's population. These pest insects are generally combated in different ways, mostly including a variety of biological and chemical insecticides. However, as many chemical insecticides pose threats to human health and the environment, development of new and eco-friendly alternatives is essential. Moreover, an increasing number of studies are reporting on resistance of insect populations against widely applied insecticides, both in the laboratory and on the field. A promising alternative would be to disrupt essential molecules within the insect gut, ideally resulting in consequent mortality of the pest insect. However, in order to find such target sites, more research on the gut physiology of insects is pivotal.

In this context, we decided to examine the changes in the midgut transcriptome of the desert locust, *Schistocerca gregaria*, during the digestive process. The relative size of the desert locust together with its polyphagous nature makes it a highly favorable organism for studying gut physiology. We performed RNA sequencing (RNA-Seq) analysis of the messenger RNA (mRNA) content of midguts dissected at various time points following food uptake. The midgut is a key part of the digestive tract of insects, which is typically the site of enzymatic digestion and nutrient uptake, and it was therefore selected for deep sequencing. The RNA-Seq data were used to create a *S. gregaria* midgut reference transcriptome, which was consulted to study the specific transcript profile of this tissue. Moreover, this transcriptome database is a useful resource to further study the digestive process in insects in general, and additionally represents an excellent database for future midgut-associated studies in this insect species in particular. Furthermore, a differential expression analysis was performed to investigate differences in the midgut transcript profiles between two hours after feeding and twenty-four hours after feeding, in order to find genes mediating the digestive process in the desert locust. A total of 569 and 212 transcripts were found to be significantly up- and downregulated in the midgut two hours after feeding, respectively. Briefly, this analysis clearly demonstrated the desert locust's ability to swiftly induce the expression of a large array of genes during the digestive process in response to food availability in the gut.

The list of transcripts upregulated two hours after feeding was further subjected to a detailed analysis in search for putatively lethal candidate targets for future pest management. Consequently, two upregulated transcripts were further investigated *in vivo* by means of RNA interference (RNAi). A vacuolar-type H^+-ATPase (H^+ V-ATPase) subunit a encoding transcript, denoted *Sg-VAHa_1*, and a Niemann-Pick C1 b (NPC1b) encoding transcript, denoted *Sg-NPC1b,* were selected based on their expected pivotal role in the intestine of the desert locust.

In general, insect H^+ V-ATPases are well-known for supporting transepithelial molecular transport by generating favorable membrane potentials, while insect NPC1b is probably responsible for the dietary sterol uptake in the midgut. Silencing each of these transcripts resulted in developmental defects and high mortality rates within two weeks after the first injections with double-stranded (ds) RNA, indicating their vital importance for the desert locust. These RNAi experiments demonstrated the possibility of discovering essential genes by analyzing the *S. gregaria* midgut transcriptome, hence emphasizing their promising potential as candidate targets for combating insect pests.

This doctoral research provides for the first time an insight into the midgut transcriptome of *S. gregaria* during the digestive process. This information was then used to further investigate the regulation of feeding and digestion in this insect species. In addition, this study also generated a broad and promising list of possible novel insecticide targets present in the midgut of the desert locust.

Beknopte samenvatting

Insecten behoren tot de meest succesvolle organismen op aarde. Een belangrijke verklaring voor dit succes is hun uitzonderlijke vermogen om op efficiënte wijze een grote variëteit aan voedingsstoffen te nuttigen. Net zoals andere heterotrofe organismen, moeten ook insecten hun essentiële nutriënten opnemen via het voedsel. Hun darmkanaal vormt hiervoor het centrale orgaan. De belangrijkste functies van dit orgaan zijn de efficiënte enzymatische vertering van het voedsel en de daaropvolgende nutriëntopname in de middendarm, en de bescherming van het organisme tegen chemicaliën, micro-organismen en mechanische schade veroorzaakt door het voedsel. Om deze functies efficiënt te kunnen uitvoeren, beschikt het insectendarmkanaal over meerdere specifieke regionale aanpassingen, en kan het bovendien, zowel op anatomisch als moleculair niveau, zeer flexibel reageren op specifieke voedselsamenstellingen. Deze centrale rol in het leven van het insect maakt van het insectendarmkanaal een veelvuldig onderzocht insectenorgaan. Vooral de algemene morfologie van het darmkanaal alsook de aanwezige spijsverteringsenzymen werden voor meerdere insectensoorten reeds in detail beschreven. Hiertegenover staat echter de relatief beperkte kennis inzake de moleculaire regulatie van de spijsvertering en nutriëntopname bij insecten.

Insecten hebben een belangrijke sociaaleconomische impact. Meerdere insectensoorten vormen een aanzienlijke bedreiging voor de gezondheid en voedselvoorziening van een groot deel van de wereldbevolking, en worden bijgevolg als plaagsoorten beschouwd. Insectenplagen worden op verschillende wijzen, voornamelijk gebaseerd op biologische en chemische middelen, bestreden. Veel van de gebruikte chemische insecticiden zijn echter nefast voor de mens en het milieu, waardoor de vraag naar nieuwe, milieubewuste(re) insecticiden de laatste jaren sterk is toegenomen. Bovendien blijken steeds meer plaaginsecten resistent te worden tegen frequent gebruikte insecticiden. Een veelbelovend alternatief bestrijdingsmiddel omvat het specifiek uitschakelen van essentiële, vitale moleculen in het plaaginsect, idealiter resulterend in zijn vroegtijdige dood, waardoor de negatieve impact op mens en natuur beperkt blijven. Zulke *target* moleculen kunnen onder andere gevonden worden in het insectendarmkanaal. Maar vooraleer deze ontdekt kunnen worden, is een verbeterde kennis van de algemene moleculaire samenstelling van de insectendarm noodzakelijk.

Tijdens dit doctoraatsonderzoek werden de tijdsgebonden transcriptveranderingen in de middendarm van de woestijnsprinkhaan, *Schistocerca gregaria*, gedurende het spijsverteringsproces door middel van RNA sequenering (RNA-Seq) in kaart gebracht. Door zijn relatieve grootte en polyfaag karakter vormt de woestijnsprinkhaan een ideaal onderzoeksorganisme voor studies naar de spijsvertering bij insecten. De RNA-Seq gegevens werden eerst gebruikt om een referentie transcriptoom van de middendarm van *S. gregaria* op te stellen. Dit transcriptoom werd geraadpleegd om het specifieke transcriptenprofiel van dit weefsel in detail te bestuderen. Vervolgens werd een differentiële genexpressie analyse uitgevoerd met als doel de veranderingen in het transcriptenprofiel van de middendarm tussen twee uur en vierentwintig uur na voedselopname te bestuderen. In totaal werden 569 en 212 respectievelijk differentieel op- en neergereguleerde genen twee uur na de voedselopname geïdentificeerd. Dit toonde duidelijk aan dat de woestijnsprinkhaan zeer snel de expressie van verschillende, spijsverterings-gerelateerde, genen kan doen toenemen als reactie op de beschikbaarheid van voedsel in het darmkanaal.

Vervolgens werd de lijst van opgereguleerde transcripten twee uur na voedselopname doorzocht naar mogelijk lethale doelwitten met betrekking tot plaagbestrijding. Twee

opgereguleerde transcripten werden, op basis van hun voorspelde sleutelfunctie in de darm van de woestijnsprinkhaan, onderworpen aan verdere *in vivo* experimenten: een *vacuolar-type H^+-ATPase* (H^+ V-ATPase) *subunit a* coderend transcript, *Sg-VAHa_1*, en een *Niemann-Pick C1 b* (NPC1b) coderend transcript, *Sg-NPC1b.* In het algemeen zijn H^+ V-ATPase complexen in insecten belangrijk voor het genereren van gunstige membraanpotentialen, waarmee het transepitheel transport van moleculen wordt bevorderd, terwijl NPC1b proteïnen voorspeld worden een centrale rol te spelen in de opname van vitale sterolen in de middendarm. Om het belang van beide transcripten voor de vertering en levensvatbaarheid van de woestijnsprinkhaan nader te onderzoeken, werden hun genproducten uitgeschakeld door middel van RNA interferentie (RNAi) gemedieerde *knockdowns*. De RNAi-gemedieerde *knockdowns* van beide transcripten resulteerden binnen de twee weken na initiatie in sterke ontwikkelingsdefecten en hoge sterftecijfers**,** waarmee het vitale belang van beide transcripten voor de woestijnsprinkhaan werd benadrukt. Deze RNAi-experimenten demonstreerden tevens de mogelijkheid om essentiële genen te ontdekken in het *S. gregaria* middendarm-transcriptoom, en beklemtoonden daarmee het veelbelovend potentieel van deze kandidaat-doelwitten voor de bestrijding van plaaginsecten.

Dit doctoraatsonderzoek geeft voor de allereerste keer een inzicht in het middendarm-transcriptoom van *S. gregaria* gedurende het spijsverteringsproces. De bekomen informatie werd gebruikt om de regulatie van voedselopname en vertering in dit insect verder te onderzoeken. Bovendien genereerde dit onderzoek een brede en veelbelovende lijst van potentiele moleculaire doelwitten in de middendarm die getest kunnen worden met het oog op de ontwikkeling van nieuwe insecticiden.

Chapter 1

Introduction

Paragraphs 1.1, 1.2, 1.4 and 1.5 of this chapter are published as:

Holtof M., Lenaerts C., Cullen D., Vanden Broeck J. (2019) Extracellular nutrient digestion and absorption in the insect gut. Cell and Tissue Research. doi: 10.1007/s00441-019-03031-9.

1.1. General anatomy of the insect gastrointestinal tract

The alimentary canal of insects runs through the entire body cavity from mouth to anus, and is divided into three main regions: foregut, midgut and hindgut (Fig. 1.1). The epithelia of both foregut and hindgut are of ectodermal origin and have a cuticular inner lining. The midgut epithelium is the only part of the alimentary duct that is of endodermal origin, and consists of a single layer of cells, lacking a cuticular lining. Therefore, the majority of the enzymatic digestion and the absorption of nutrients largely takes place in the midgut (Chapman, 2013).

The insect mouth, through which food will enter the body, is directly connected to the foregut. The foregut usually consists of four distinct areas: the pharynx, esophagus, crop and proventriculus. The pharynx and esophagus are both involved in the backwards movement (by peristalsis) of the food towards the crop. The food bolus is then stored inside the crop, until it is passed on to the midgut for digestion. Some initial enzymatic digestion occurs in the crop, which is mediated by salivary enzymes and enzymes regurgitated from the midgut. In most insects, the crop is situated in line with the alimentary canal and expands laterally. However, in adult Lepidoptera and adult Diptera, the crop is an associated lobe of the esophagus (Chapman, 2013: Stoffolano and Haselton, 2013). The proventriculus is located at the transition of the foregut to the midgut. This muscular region is an important valve for the passage of food, and is in some insects, mostly Orthoptera, Blattodea and some Coleoptera, even responsible for additional mechanical degradation of the food through the action of modified tooth-like chitinous structures. Next the food bolus enters the midgut: the primary site of enzymatic digestion in insects. Most insects possess two to eight small pouches, called caeca, diverging from the proximal anterior end of the midgut. These gastric caeca provide extra surface area for enzyme secretion and biochemical digestion. Hence they strongly increase digestive efficiency of the midgut. After enzymatic digestion in the midgut, the remaining food bolus passes on to the hindgut. The hindgut's general function is the osmotic regulation of internal fluids. The hindgut can be divided into three distinct regions: the pylorus, ileum and rectum. Malpighian tubules are specialized tubular excretory organs that are attached to the pylorus (Maddrell and O'Donnell, 1992). Inside the pylorus, the food bolus and the secretions from the Malpighian tubules are mixed. To prevent adverse reflux of the food bolus back into the midgut, the pylorus is surrounded by circular muscles to act as a sphincter. The ileum connects the pylorus with the rectum, and can harbor important symbiotic microorganisms. For example in termites and scarab beetles, microbiota inside the ileum contribute largely to the degradation of plant polysaccharides utilized for aerobic metabolism by the insect (Douglas, 2015). The third and final part of the hindgut, the rectum, is the main site for ion and water exchange. In order to support its role in osmosis, the cuticular lining of the rectum is substantially thinner than that of the other regions of the hindgut (Chapman, 2013).

Gut motility results from the presence of muscular layers in the wall of the alimentary canal. The entire alimentary canal is enclosed by striated circular and longitudinal muscles. The circular muscles directly enclose the gut and are surrounded by an external layer of longitudinal muscles. The muscles surrounding the fore- and hindgut are the most developed, since these parts of the gut function as dilators. The gastrointestinal tract of insects generally has three profoundly innervated regions: the foregut/anterior midgut, the midgut/hindgut junction, and the posterior hindgut (Hartenstein, 1997; Cognigni et al., 2011). These innervations largely control the peristaltic regulation of the visceral muscles to promote the intestinal food transit, but might also regulate epithelial functions, such as enzyme production and nutrient absorption (see Ch1|1.3) (Cognigni et al., 2011; Lemaitre and Miguel-Aliaga, 2013).

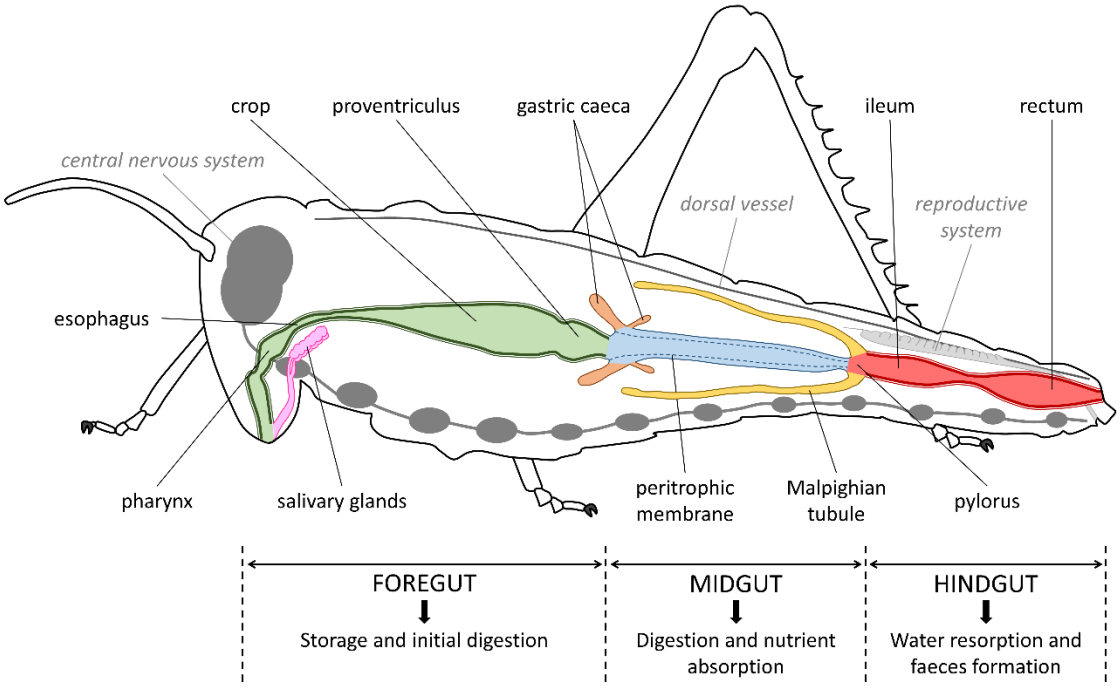


Figure 1.1. General anatomy of the insect gastrointestinal tract. The insect intestine is divided into three main regions: foregut (green), midgut (blue) and hindgut (red). The foregut consists of the pharynx, esophagus, crop and proventriculus. The midgut is divided by the peritrophic membrane into an endoperitrophic and ectoperitrophic space. The hindgut encompasses the pylorus, ileum and rectum. Each area is specifically adapted to its respective function in digestion, with the majority of enzymatic digestion occurring inside the lumen of the midgut and associated caeca (orange).The Malpighian tubules (yellow) are connected to the pylorus of the hindgut. The salivary glands are drawn in pink. The cuticular lining of the foregut and hindgut are indicated by thicker lines.

1.2. Anatomy and functioning of the insect midgut

The general morphology of the midgut epithelium is comparable in all insects. It consists of a single layer of epithelial cells, containing four cell types: the intestinal stem cells (ISCs) and the enteroblasts (EBs), which are immature cells that can further differentiate into enteroendocrine cells (EECs) or into enterocyte cells (ECs) largely depending on the Notch/Delta and Wingless signaling pathways (Tanaka et al., 2007; Takashima et al., 2011; Miguel-Aliaga et al., 2018) (Fig. 1.2). EECs are part of the enteroendocrine system and represent a vast proportion of cells present in the midgut epithelium. These cells act as sensors for the internal intestinal environment, and are able to produce specific gut regulatory hormones and peptides, as well as mediate the communication of the nutritional status to other organs (Sternini et al., 2008). Well-known examples of enteroendocrine peptides in insects are allatostatin A, myosuppresin, neuropeptide F and tachykinin-related peptides. For more in-depth summaries on the presence and functions of enteroendocrine peptides in different insect species, the reader is referred to reviews by Spit *et al.* (2012a) and Wegener and Veenstra (2015). ECs are the most prevalent cells in the midgut epithelium. They are characterized by a columnar shape and bear microvilli on the apical membrane that significantly increase the surface area of the midgut. The enterocytes of the midgut are responsible for the production of digestive enzymes and the absorption of digested products (Huang et al., 2015).

In addition, in some insects, enterocytes are also specialized in controlling the midgut luminal pH. In general, the midgut lumen of most insects has a neutral pH between 6 and 7.5. Exceptions are the more acidic to neutral midgut pH (pH 5–7) of most Coleoptera, the high

alkaline midgut pH of Lepidoptera (pH > 9), and specialized acidic anterior midgut regions of most Diptera and Hemiptera (Terra and Ferreira, 1994; Chapman, 2013). In *D. melanogaster* and other higher dipterans, specialized ECs, named coprophilic or copper cells, are responsible for the acidification of the anterior midgut (Dubreuil et al., 1998). The acidic environment is generated by a membrane associated H^+ V-ATPase pump, a carbonic anhydrase and five transporters or channels mediating K^+, Cl^- and HCO_3^- transport (Overend et al., 2016). The acidic pH is important for the initial breakdown of food, as well as for the protection against pathogens. In contrast to the well described mechanism of acidification, less is known about how the pH in other midgut areas is maintained (Miguel-Aliaga et al., 2018). This pH associated compartmentalization inside the *Drosophila* midgut is remarkable, but not unique in insects. In the midgut of lepidopteran insects, specialized ECs, called goblet cells, are responsible for the global alkalization of the midgut lumen. Goblet cells form cavities inside the midgut epithelium and mediate the active transport of K^+ into the gut lumen via $K^+/2H^+$ antiporters driven by a V-ATPase pump (Flower and Filshie, 1976; Chapman, 2013). The gut pH is one of the most important regulators of digestive enzyme activity in insects. Consequently, different enzyme profiles are found amongst different insect orders, with a clear correlation between the gut pH and the pH optima of the observed digestive enzymes. Maintaining the gut pH is therefore a pivotal role of the midgut epithelium (Terra and Ferreira, 1994).

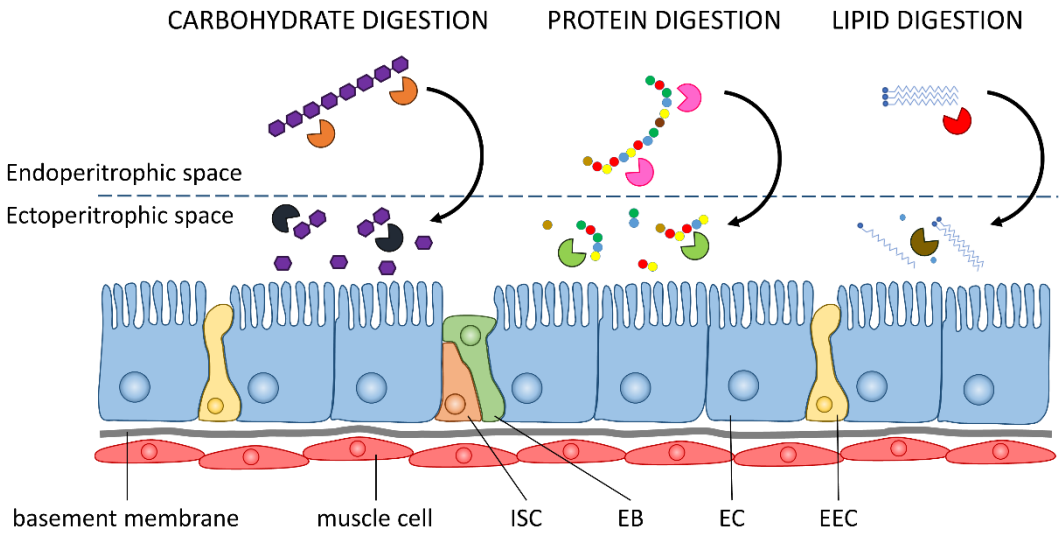


Figure 1.2. Compartmentalized digestion of macronutrients inside the midgut. The midgut epithelium is a single layer of epithelial cells, containing four cell types: the intestinal stem cells (ISCs), enteroblasts (EBs), enteroendocrine cells (EECs), and enterocyte cells (ECs). Gut motility results from the presence of muscular layers in the wall of the alimentary canal. ECs bear a brush border of microvilli on the apical membrane, produce and release digestive enzymes, absorb the digested end products from the midgut lumen and regulate the midgut luminal environment. The midgut also synthesizes the peritrophic membrane (PM, blue dashed line). The PM compartmentalizes the midgut into an endo- and ectoperitrophic space, which gives rise to spatial organization of digestion, with initial macronutrient digestion occurring inside the endoperitrophic space, and the assimilation of digestive end products occurring inside the ectoperitrophic space. Distinct digestive enzymes are active in the separated areas of the midgut.

In most insect species, a membranous, noncellular structure is observed inside the midgut. This semipermeable membrane is the peritrophic matrix (PM) and envelops the food bolus (Fig. 1.3). The PM is composed of chitin microfibrils embedded in a matrix of proteins (peritrophins), glycoproteins and proteoglycan. Two types of PMs exist according to their respective way of synthesis as reviewed by Lehane (Lehane, 1997). Type I PMs are produced by the entire midgut and can be found in the majority of PM producing insects. This type of

PMs can be generated continuously, found in continuous feeders, such as caterpillars and cockroaches, or in response to feeding, found in batch feeders, such as mosquitos. Type II PMs are produced by restricted areas in the anterior midgut and are usually only found in Diptera, Dermaptera, Embiodea, and some families of Lepidoptera (Lehane, 1997). In some insects however, both types of PM can occur at different life stages. For example, in mosquitos, larvae produce a type II PM, while adult females produce a type I PM (Dinglasan et al., 2009). Not all insects produce the PM, and some orders, such as Hemiptera and Thysanoptera, have evolved an analogous extracellular lipoprotein structure called the perimicrovillar membrane (PMM) (Lehane, 1997; Terra, 2001; Silva et al., 2004). Generally, the PM encloses the food bolus to separate the food from the midgut epithelium, thereby protecting the midgut epithelium from mechanical and chemical damage by the food bolus, as well as from harmful pathogens and toxins (Huang et al., 2015). As a semipermeable membrane, the PM also controls the passage of molecules, hence dividing the midgut lumen into two physiologically separate regions: the endoperitrophic space (inside the PM) and the ectoperitrophic space (outside the PM). The compartmentalization of the midgut lumen results in the spatial organization of digestive events. In the endoperitrophic space, food flows from the anterior to the posterior end of the midgut. The pore size at the anterior side of the PM allows the movement of digestive enzymes into the endoperitrophic space. The initial digestion of the food thus takes place inside the endoperitrophic space, where mostly large macromolecules are degraded. The pore size of the PM gradually decreases towards the posterior end, only allowing smaller sized molecules to pass through the membrane to the ectoperitrophic space (Gutiérrez-Cabrera et al., 2016). Inside the ectoperitrophic space, a countercurrent flow pushes the food particles backwards towards the caeca. This flow is created by the excretion of water and ions by the Malpighian tubules and their subsequent uptake by the caeca (Fig. 1.3). In addition, neuropeptides also influence gut motility as well as gut contents, thereby affecting the digestive flux (Spit et al., 2012a). For example, in the migratory locust, *Locusta migratoria*, the neuropeptide sulfakinin was demonstrated to mediate the clearing of food contents from the caeca (Zels et al., 2015). Whenever food particles become small enough to diffuse through the PM, the countercurrent flow in the ectoperitrophic space will push them back towards the caeca where digestion continues. From the caeca, food particles can again enter the endoperitrophic space restarting the digestive cycle. The recycling of food particles and digestive enzymes strongly increases digestive efficiency inside the insect midgut (Lehane, 1997; Terra, 2001; Chapman, 2013; Bolognesi et al., 2008). The eventual end products of digestion will be taken up from the ectoperitrophic space by the midgut epithelial cells (Chapman, 2013).

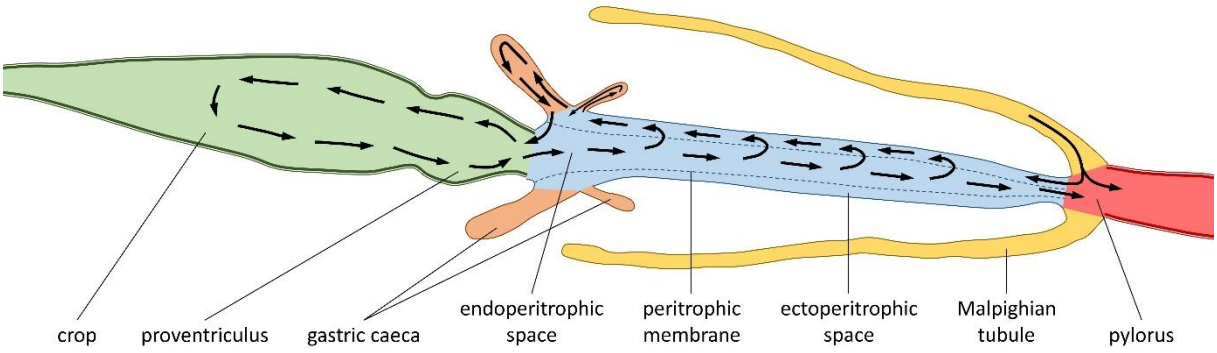


Figure 1.3. Flow of food particles inside the insect gut. The countercurrent flow inside the midgut is created by the secretion of water and ions by the Malpighian tubules (yellow) and their subsequent uptake by the caeca (orange). In the endoperitrophic space, food particles flow towards the distal end of the midgut, while inside the ectoperitrophic space food particles are transported proximally towards the caeca. Some midgut contents may even re-enter the crop. The arrows indicate the path followed by food particles inside the midgut. The countercurrent flow inside the midgut strongly increases digestive efficiency.

1.3. Regulation of feeding and digestion

Insect feeding behavior and digestion are strictly regulated processes mainly influenced by a complex combination of physical, chemical and nutritional cues. In general, positive inputs from food and non-food stimuli are alternated by deterrent stimuli and by feedbacks from amongst others gut stretch receptors, hormones, peptides and haemolymph composition. A crucial role in initiating feeding and food uptake is reserved for the suboesophageal ganglion (SOG). The SOG, as part of the central nervous system (CNS), is a composite of fused ganglia (mandibular, maxillary and labial) located below the brain and esophagus, which coordinate the movements of the insect mouthparts (Griss et al., 1991; Audsley and Weaver, 2009).

The anterior portion of insect gastrointestinal tract is strongly innervated by the CNS and the stomatogastric nervous system (SNS). The insect SNS typically consists of a chain of ganglia connecting the muscles of the mouth cavity, foregut and midgut with the brain and endocrine system (Hartenstein, 1997). It is mainly involved with the regulation of foregut motor activity associated with feeding. Three main partitions are distinguished in the insect SNS (Fig. 1.4). (1) The frontal ganglion (FG) connecting the brain with the anterior foregut via the frontal connectives. The FG has been demonstrated to play a crucial role in feeding behavior of insects, as its peripheral projections innervate the musculature of the pharynx and esophagus, coordinating the peristaltic movements of the foregut. (2) The hypocerebral ganglion (HCG) is connected to the FG and the corpora cardiaca (CC), a pair of neuroendocrine organs which produce and release several different neurohormones and neuropeptides mediating digestive and metabolic processes. (3) The proventricular ganglion (PVG) is located at the proventriculus, near the foregut/midgut junction, and is connected to the FG and HCG via processes running across the lateral surface of the foregut (Hartenstein, 1997; Copenhaver, 2007; Audsley and Weaver, 2009; Spit et al., 2012a).

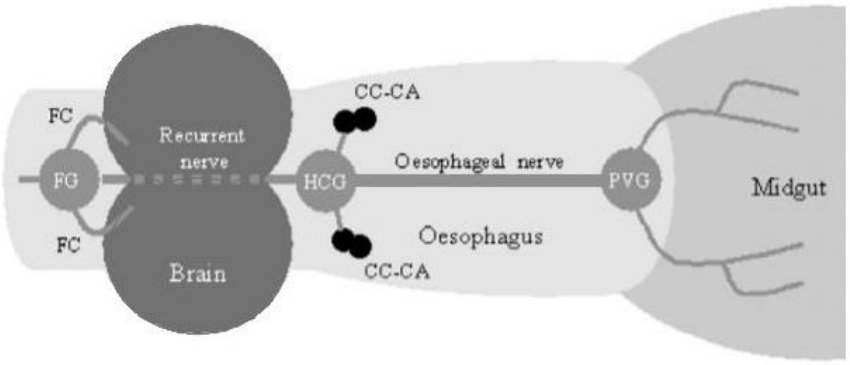


Figure 1.4. Generalized structure showing the different ganglia and interconnecting nerves constituting the SNS in insects. Abbreviations: CC = corpus cardiacum; CA = corpus allatum; FG = frontal ganglion; FC = frontal connectives; HCG = hypocerebral ganglion; PVG = proventricular ganglion. Picture adapted from Audsley and Weaver (2009).

The movement of food through the gut is strongly regulated by the contractions of the foregut, which are orchestrated by complex interactions between CNS and SNS. In contrast, the midgut enteric nervous system (ENS) generally lacks ganglia, but rather consists of diffuse nerve plexuses that extend across the midgut musculature and often contain dispersed groups of enteric neurons. The hindgut is innervated by proctodeal and rectal nerves originating from the terminal abdominal ganglion of the ventral nerve cord (CNS). Branches of the proctodeal nerve also extend onto the posterior midgut. Several peripheral neurosecretory cells are frequently present within these nerves (Copenhaver, 2007; Spit et al., 2012a).

Numerous neuropeptides are found in the insect CNS and SNS demonstrating the importance of brain-gut peptidergic communication in the regulation of the digestive process. For example, several insect peptides, such as proctolin, tachykinins, FMRFamide-related peptides, exhibit excitatory effects on the visceral muscles thereby affecting gut motility and feeding behavior. Other typical insect peptides demonstrated to be involved in the stimulation or inhibition of feeding include amongst others allatostatins, diuretic hormone, neuropeptide F, short neuropeptide F and sulfakinins. For details on the mode of action of different neuropeptides regulating feeding behavior and digestion in insects, the reader is referred to in-depth reviews by Gäde and Goldsworthy (2003), Audsley and Weaver (2009) and Spit *et al*. (2012a). Noteworthy, several of these peptides have also been localized in the midgut EECs indicating the ability of the midgut to regulate its own functioning (Ch1|1.2).

Until now, there exist only a few studies describing the regulation of digestive enzyme production and release in insects. This is generally believed to be regulated by a combination of hormonal, paracrine and prandial mechanisms (Lehane et al., 1995). However, the lack of additional in-depth research explains that today we still do not fully understand how this process is exactly regulated. Interestingly, also some regulatory peptides appear to exhibit stimulatory or inhibitory effects on enzyme activity levels in the gut, indicating that digestive enzyme release in response to food is putatively, to a certain extent, mediated through neuropeptides (Spit et al., 2012a).

1.4. Extracellular enzymatic nutrient digestion

To digest dietary proteins, carbohydrates and lipases, insects rely on the hydrolytic action of digestive enzymes (Fig. 1.2; Table 1.1). The composition of digestive enzymes mediating the degradation of food inside the gut is complex and often species-specific. It is largely determined by feeding habits, ingested food quality and quantity, and specific midgut luminal environments. The evolutionary relationships between insects and their feeding hosts has also stimulated this complexity in great ways. This section will provide a comprehensive overview of the enzymatic digestion and digestive enzyme classes found in different insect orders.

Table 1.1 | General overview of macronutrient digestion in the insect's gastrointestinal tract.

	PROTEIN DIGESTION	**CARBOHYDRATE DIGESTION**	**LIPID DIGESTION**
ENZYMES	exopeptidases ○ carboxypeptidases ○ aminopeptidases endopeptidases ○ serine proteases ○ cysteine proteases ○ aspartic proteases ○ metalloproteases	α-amylases α/β-glucosidases endo/exo-β-1,4-glucanases (cellulases)	lipases phospholipases sterol reductases
EXPRESSION	caeca, midgut	salivary glands, caeca, midgut	caeca, midgut
SUBSTRATES	proteins, polypeptides	polysaccharides, disaccharides	acylglycerols, fatty acids, galactolipids, phospholipids, sterols
PRODUCTS	amino acids, oligopeptides	mono- and disaccharides	free fatty acids, phospholipids, cholesterol

1.4.1. Protein digestion

The adequate uptake of proteins is pivotal for the survival of any insect. Proteins are digested to release their amino acid contents, which are then absorbed across the midgut epithelium to be used in vital processes, such as growth and development, energy storage and reproduction. Some amino acids cannot be synthesized *de novo* by the insect, and need to be acquired from the environment. These essential amino acids for insects are: the aromatic phenylalanine, tryptophan and histidine; the aliphatic leucine, isoleucine, valine and threonine; the sulfur containing methionine; and the basic arginine and lysine (Boudko, 2012).

Different proteolytic enzymes, also called proteases, mediate protein breakdown (Table 1.1). In general, digestive proteases are divided into two groups: the endopeptidases that cleave the proteins internally and the exopeptidases that cleave the terminal amino acids from the proteins. The exopeptidases are further divided into two groups: the carboxypeptidases that cleave at the carboxylic terminus and the aminopeptidases that cleave at the amino terminus of proteins. Endopeptidases are typically classified into four groups, based on the composition of the catalytically active site: (1) the serine proteases, having a serine residue at the active site, (2) cysteine proteases, having a cysteine residue at the active site, (3) the aspartic (acid) proteases, having a aspartyl residue at the active site, (4) the metalloproteases, having a metal ion at the active site (Rao et al., 1998; Berg et al., 2002).

Serine and cysteine proteases rely on the activity of a catalytic triad, respectively Ser-His-Asp and Cys-His-Asp, to degrade peptide bonds. Serine proteases include trypsins, chymotrypsins and elastases. Trypsins preferentially cleave peptide bonds following an arginine or lysine residue, chymotrypsins preferentially cleave after a tyrosine, phenylalanine or tryptophan residue, and elastases preferentially cleave peptide bonds following an alanine or serine residue (Terra and Ferreira, 1994; Berg et al., 2002). Cysteine proteases in insects include the Cathepsin B-like and Cathepsin L-like proteases (Terra and Ferreira, 1994, 2012; Chapman, 2013). Cathepsin B-like proteases are in fact defined as important insect peptidyl dipeptidases, instead of real endopeptidases. The cathepsin L-like proteases, on the other hand, are true endopeptidases that preferably cleave peptide bonds with hydrophobic amino acid residues (Terra and Ferreira, 2012). Metallo- and aspartic proteases differ from serine and cysteine proteases because they depend on an activated water molecule instead of an amino acid as nucleophile (Berg et al., 2002; Zhu-Salzman and Zeng, 2015). The water molecule is activated by a divalent metal cation or an aspartic residue respectively, present in the active site. The most prevalent metalloproteases in insects are the zinc-metalloproteases, representing a major fraction of the exopeptidase activity found in insects (Terra and Ferreira, 2005). The best-known insect digestive aspartic acid proteases are the Cathepsin D-like proteases. These proteases show high sequence similarity with Cathepsin D proteases, a major family of intracellular aspartic proteases in the lysosomes of all animals, and thus are probably derived from the same ancestral gene as the intracellular cathepsin D found in lysosomes (Padilha et al., 2009). The shift towards the use of lysosomal enzymes for digestion presumably evolved from adaptations to dietary challenges, such as the presence of plant-derived defensive compounds in the food (Pimentel et al., 2017).

An important factor determining the digestive enzyme profile and activity in insects is the midgut luminal pH. The pH optimum of serine proteases is neutral to alkaline. Hence, serine proteases are found in all studied insect orders, including Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, and Orthoptera. A large phylogenetic analysis of trypsin and chymotrypsin sequences originating from over 60 different insect species belonging to the above mentioned orders, illustrated that trypsins and chymotrypsins are clearly divided in two

phylogenetic clades and that trypsins or chymotrypsins originating from the same insect class cluster together inside their respective clade (Spit et al., 2014). This indicates that these proteases are specifically adapted to their associated midgut environment. Serine proteases make up for the majority (as much as 95%) of the total proteolytic activity in the gut of larval Lepidoptera (Srinivasan et al., 2006). Serine proteases isolated from Lepidoptera midguts show clear alkaline pH optima, probably as a consequence of adaptations to the alkaline midgut environment (Christeller et al., 1992). Most of the remaining proteolytic activity in Lepidoptera midguts is executed by cathepsin-B like cysteine proteases, and different exopeptidases (Christeller et al., 1992; Terra and Ferreira, 1994; Bown et al., 1997; Patankar et al., 2001; Breugelmans et al., 2009; Tabatabaei et al., 2011). The midguts of Orthoptera are characterized by a neutral pH, and in parallel, the majority of proteolytic activity in the midguts of this order results from serine proteases (Chapman, 2013). Transcriptional and enzymatic studies in *Locusta migratoria* and *Schistocerca gregaria* (Orthoptera) revealed high serine protease activity in the midgut (Spit et al., 2012b, 2014). Three trypsins purified from *L. migratoria* showed maximum activity at pH 8, matching the neutral midgut environment of these insects (Lam et al., 2000). Serine proteases are also the most prevalent digestive proteases in other studied Orthoptera, including the grasshopper *Oedaleous asiaticus* and the cricket *Gryllus bimaculatus* (Huang et al., 2017; Woodring, 2017). Serine proteases also represent the majority of proteases identified in Dipteran midguts. Mosquitos largely depend on trypsin serine proteases for their proteolytic digestion (Borovsky, 2003). Genome analysis of *D. melanogaster* has revealed a substantial amount of putative serine proteases present in the midgut. Other identified proteases were metalloproteases and, probably restricted to the acid midgut region, cysteine and aspartic proteases (Lemaitre and Miguel-Aliaga, 2013). Cysteine and aspartic proteases are described to have slightly acidic pH optima, and are therefore predominantly found in the acidic midguts of Coleoptera, and specific acidic midgut regions of Diptera, Heteroptera and Hemiptera (Terra et al., 2012). The Colorado potato beetle, *Leptinotarsa decemlineata*, depends on the proteolytic actions of both cysteine proteases, with cathepsin L-like, B-like and H-like enzymes, and aspartic proteases, with Cathepsin D-like activity. In these insects, pH ranges from 5 to 7 along the midgut lumen. Hence the dominant protease activity shifts from aspartic proteases mediating initial proteolytic digestion at mildly acidic pH to cysteine and serine proteases for subsequent proteolytic digestion at neutral pH (Brunelle et al., 2004; Srp et al., 2016). Cathepsin D-like proteases in insects have low pH optima and were first discovered in the very acidic anterior midgut region of *Musca domestica* (Diptera) (Greenberg, 1955; Padilha et al., 2009) and were later also identified in other Diptera, Hemiptera, and several Coleoptera (Terra et al., 2012). For example, in *Dysdercus peruvianus* (Hemiptera), cathepsin D-like proteases active in the very acidic anterior midgut portion (pH~3.5) are responsible for the breakdown of specific cysteine protease inhibitors (PIs) present in their cotton seed diet. By degrading these plant defensive compounds, cathepsin D-like proteases both complement the general protein digestion as well as protect the digestive activity of the cathepsin L-like cysteine proteases in the posterior midgut regions (Pimentel et al., 2017). In *M. domestica,* the cathepsin D-like proteases in the acidic anterior midgut region have important anti-bacterial functions (Espinoza-Fuentes and Terra, 1987).

Besides gut pH, adaptations to their specific feeding habits have also largely influenced the digestive enzyme profiles in insects. For example, the interaction between herbivorous insects and plants has massively impacted their digestive enzyme profiles. One common defensive strategy of plants is to inhibit the insect digestion via digestive enzyme inhibitors, such as PIs. Plant PIs are abundant proteins in all tissues susceptible to insect voracity, namely storage, reproductive and vegetative tissues, and their expression can be induced upon wounding. Once present in the insect gut, these inhibitory proteins are able to block the proteolytic enzymes, thereby dramatically decreasing the digestive efficiency. To continue feeding on

these plants, insects evolved adaptive strategies to overcome the detrimental effects of these plant PIs (Jongsma, 1997; Zhu-Salzman and Zeng, 2015). This evolutionary arms race has resulted in several PI induced adaptations, typically characterized by changes in the midgut enzymatic composition. Different counter adaptive strategies include the overproduction of active proteases to overrule the inhibitory effects of PIs, a shift to inhibitor insensitive proteases, and the production of proteases capable of degrading the plant PIs (Vorster et al., 2015; Zhu-Salzman and Zeng, 2015). Consequently, plant-insect interactions and the adaptation to plant defensive compounds have resulted in the positive selection for certain digestive enzyme profiles in insects. The PI induced response has been extensively investigated in many pest insects. For example, the Lepidoptera *Helicoverpa armigera* and *Spodoptera exigua* both respond to the uptake of PIs by producing both more and inhibitor insensitive proteases (Srinivasan et al., 2006). In the locust *S. gregaria*, in addition to induced overall proteolytic activity, increased carboxypeptidase activity was observed when the serine protease activity was inhibited by plant and/or insect derived PIs (Spit et al., 2012b). In *L. migratoria*, a more generalized upregulation of serine protease activity in response to dietary plant serine PIs was observed (Spit et al., 2014, 2016). When fed with PI containing diet, *T. castaneum* responded by shifting from a cysteine to a serine protease-based digestion (Oppert, 2005). The Colorado potato beetle appears to downregulate the expression of cysteine and aspartic proteases targeted by ingested plant PIs, while simultaneously upregulating distinct PI insensitive cysteine proteases and serine proteases (Petek et al., 2012). And as described earlier, *D. peruvianus* (Hemiptera) uses cathepsin D-like proteases in its acidic anterior midgut portion to degrade of specific cysteine PIs present in their diet. More examples of PI induced counter defenses are described in several in-depth reviews (see for example Jongsma, 1997; Mello and Silva-Filho, 2002; Lopes et al., 2004; Srinivasan et al., 2006; Zhu-Salzman and Zeng, 2015).

Endogenous mechanisms regulating the protease activity in insects remain largely elusive. Specific endogenous pancreatic secretory trypsin inhibitor- like (PSTI-like) proteins, according to their analogy with mammalian PSTIs, have been identified in several insect species, including *D. melanogaster*, *L. migratoria* and *Bombyx mori.* These endogenous trypsin inhibitors probably protect the organism from prematurely activated digestive enzymes, as was firstly discovered in *L. migratoria* in 2011 by Van Hoef and colleagues (Van Hoef et al., 2011). This protective mechanism suggests that insects are able to autonomously control protease activity inside the intestine, analogous to vertebrates. Recently, a study in *H. armigera* confirmed the presence of endogenous PIs mediating the gut proteolytic activity in response to food availability (Lomate et al., 2018). Nevertheless, further research is needed to characterize these regulatory mechanisms in insects.

1.4.2. Carbohydrate digestion

Dietary carbohydrates are used as a direct source of energy, or they can be either stored as glycogen or lipids for energy storage in the fat boy or recycled as the carbon skeleton for the synthesis of various amino acids. Moreover, the insect cuticle largely consists of the polysaccharide chitin, which can be generated from simple sugars absorbed from the diet. The majority of carbohydrate digestion in insects is concentrated in the midgut, but initial digestion can already take place in the foregut and mouth. The most common sources of dietary sugars for insects are starch, cellulose, and sucrose. Carbohydrate digestion is typically mediated by digestive enzymes, carbohydrases, that break down the glycosylic bonds of polysaccharides. The end products of carbohydrate digestion are mainly monosaccharides that are absorbed across the midgut epithelium (Table 1.1) (Chapman, 2013; Miguel-Aliaga et al., 2018). These

end products in turn regulate the carbohydrate digestion rate inside the gut. For example, in *D. melanogaster*, the accumulation of sucrose, glucose and fructose stimulates a feedback loop that inhibits the gene expression of carbohydrases and lipases (Chng et al., 2014; Miguel-Aliaga et al., 2018).

The initial digestion of carbohydrates is mediated by amylases. These types of carbohydrases catalyze the breakdown of the α-1,4-glucan chains present inside dietary polysaccharides, such as starch and glycogen. The only type of amylases recognized in insects are α-amylases (Terra and Ferreira, 1994). Alpha-amylases can act on both internal (endo-amylases) and external (exo-amylases) linkages in a random manner, thereby producing smaller polysaccharides and disaccharides that will be further degraded to monosaccharides during subsequent steps of carbohydrate digestion performed by distinct enzymes (Terra and Ferreira, 1994; Chapman, 2013; Da Lage, 2018). A recent review by Da Lage shows that most insects possess multiple gene copies encoding α-amylase digestive enzymes, called the *Amy* genes (Da Lage, 2018). The copy numbers vary from one to as much as twelve *Amy* genes and can vary drastically within the same insect genus. A well-studied genus in this context is that of *Drosophila*, where the *Amy* gene has undergone multiple duplication events. In this insect lineage, *Amy* copy numbers vary from one in *D. virilism* to six or seven in *D. ananassae*, even though these insects grossly share the same diet (Da Lage et al., 2000; Da Lage, 2018). Also in other insect orders, this drastic divergence between *Amy* copy numbers was observed. The presence of multiple *Amy* gene copies might be part of specific counter defenses against plant anti-digestive factors, such as amylase inhibitors (Franco et al., 2002). The vast majority of amylase production and activity is situated in the midgut. However, part of the carbohydrase activity can already be initiated in the mouth and foregut. This can be largely ascribed to amylases originating from the midgut that migrate towards the foregut upon food uptake, as observed in the larvae of Coleoptera, Diptera, Lepidoptera, and Orthoptera (Da Lage, 2018). Interestingly, in some insects, including *H. armigera* and the mosquito *Aedes aegypti,* the salivary glands are also capable of producing some amylases themselves (Grossman et al., 1997; Kotkar et al., 2012; Da Lage, 2018).

After the initial breakdown of large oligosaccharides by amylases, glucosidases mediate the subsequent degradation of oligo- and disaccharides into monosaccharides. Two types of glucosidases are found in insects: α- and β-glucosidases. The nomenclature correlates with the type of linkage the enzymes targets, namely α-1,4-glucose terminal linkages and β-1,4-glucose terminal linkages (Terra and Ferreira, 1994). Glucosidase activity generates monosaccharides, the end products of carbohydrate digestion, to be absorbed by the midgut enterocytes. The substrate preference of glucosidases is largely dependent on the glucosidase size and its active site preference. The presence of different types of glucosidases in the midgut helps regulating the efficient stepwise digestion of both long and short dietary carbohydrate chains. Glucosidases are typically highly expressed by the midgut epithelium, and dominant activity is observed in the ectoperitrophic space, as opposed to amylases, which act mainly in the endoperitrophic space (Terra and Ferreira, 1994; Guzik et al., 2015). In some insects, glucosidase expression is also observed in the salivary glands (Terra and Ferreira, 1994; Juhn et al., 2011). The most abundant α-glucosidases in insects are maltases and sucrases, named after their preferred disaccharide substrates maltose and sucrose (Chapman, 2013). Starch and sucrose represent the dominant carbohydrate reserves in plants and are the principal sources of dietary carbohydrates for herbivorous insects. Starch is a long polysaccharide composed of a large number of glucose units and is successfully degraded inside the insect gut by the subsequent actions of α-amylases and maltases (α-glucosidase). Sucrose is a disaccharide composed of a glucose and a fructose unit, and its degradation is generally mediated by α-glucosidase and β-fructosidase activity (Terra and Ferreira, 1994).

Recently, membrane-bound sucrose-specific sucrases (SUHs) were for the first time detected in the high alkaline intestinal lumen of Lepidoptera larvae (Li et al., 2017). These SUHs presumably diverged from α-glucosidases to enhance the sucrose digestion, the major energy source for Lepidoptera. For Hemiptera, sap-feeding insects, sucrose is one of the most concentrated nutrients available in their diet, and accordingly, high sucrase activity has been observed in these insects (Douglas, 2006). The β-glucosidases are named after the terminal sugar unit they cleave, namely glucose. Other examples are β-galactosidase and β-fructosidase, which cleave the terminal galactose and fructose respectively. Insect β-glucosidases are part of the cellulytic system responsible for the degradation of cellulose. Interestingly, many of these enzymes have important industrial applications because of their high catalytic activity for degrading lactose (Husain, 2010).

Cellulose and hemicellulose are two of the most abundant compounds in plant cell walls. These β-1,4-glucose polymers both represent a significant source of carbohydrates for plant-eating insects as their breakdown releases high amounts of free glucose molecules. Cellulose can only be completely degraded by the combined action of three sets of enzymes, often referred to as the cellulytic system, composed of endo-β-1,4-glucanases, exo-β-1,4-glucanases and β-glucosidases. The endo- and exo-β-1,4-glucanases, sometimes referred to as cellulases, randomly target inner and outer bonds of the cellulose chain (Watanabe and Tokuda, 2010). Various cellulases might be active inside the gut, such as cellulases of microbial endosymbionts, exogenous cellulases ingested via the food and endogenous cellulases. It was long believed that animals were not capable of endogenously producing cellulases, but this has been re-evaluated ever since more genomic and enzymatic evidence of cellulase activity in insects and other animals appeared (Chapman, 2013). The first record of endogenous cellulase activity observed in insects was in the subterranean termite *Reticulitermes speratus* (Yokoe, 1964). Since then, cellulase genes or their homologs were identified in various insect species (Watanabe and Tokuda, 2010). More recently, endogenous cellulase activity has also been detected in the midguts of the well-studied insect models *T. castaneum*, the western corn rootworm, *Diabrotica virgifera virgifera, and M. domestica* (Willis et al., 2011; Valencia et al., 2013; Zhang et al., 2017b). Remarkably, endogenous cellulases appear to be absent in *Anopheles gambiae*, *B. mori* and *D. melanogaster* (Kunieda et al., 2006; Watanabe and Tokuda, 2010). In-depth studies on the identity of these endogenous cellulases in insect midguts demonstrated that important cellulase activity was performed by the endogenously-produced glycoside hydrolase (GH) family of endoglucanases. Members of the GH9 endoglucanases family have been identified in all insect orders with reported cellulase activity. Additionally, members of the GH5 and GH49 families appear to be exclusively active in Coleoptera. Although these GH cellulase families differ in structure, they do share the same substrate specificities and are suggested to have evolved convergently (Watanabe and Tokuda, 2010; Kirsch et al., 2012; Chapman, 2013).

1.4.3. Lipid digestion

All insects use lipids for energy storage in the fat body, which is accessed during subsequent periods of high energy demand or starvation (Horst, 2003; Chapman, 2013). Lipids are also incorporated in the growing oocyte to support oogenesis (Fruttero et al., 2017). Additionally, many insect pheromones present in the insect cuticle are synthesized from dietary lipids (Yew and Chung, 2015). However, in the majority of insects, these dietary lipids seem to be less essential than proteins and carbohydrates, since insects are capable of endogenously synthesizing many fatty acids and phospholipids from dietary carbohydrates. Nevertheless, all insects do require dietary sources of sterol and polyunsaturated fatty acids (PUFAs) as

structural components of the cell membrane, secondary metabolites, and starting material for steroid synthesis (Chapman, 2013; Zibaee et al., 2014).

Acylglycerols, fatty acids, galactolipids, phospholipids and sterols are the most prevalent dietary lipids consumed by insects. The key enzymes involved in lipid digestion are glycerol ester hydrolases, called lipases and phospholipases, the latter being a special family of lipases depending on a different catalytic mechanism that specifically degrade phospholipids (Turunen, 1979). Lipases degrade dietary lipids to generate typical end products, such as free fatty acids, glycerols, partial acylglycerols and phospholipid derivatives, in a process called lipolysis (Table 1.1) (Turunen, 1979; Berg et al., 2012; Chapman, 2013). In most insects, the majority of lipase production and activity takes place in the midgut and associated caeca (Majerowicz and Gondim, 2013). In some insects, including *S. gregaria* and the grain aphid *Sitobion avenae*, lipases were detected in the transcriptome of salivary glands. However, they are most probably involved in insect-host interactions and their possible role in the digestive breakdown of dietary lipids remains to be elucidated (Valenzuela et al., 2003; Shukle et al., 2009; Schafer et al., 2011; Zhang et al., 2017a). Similar to other organisms, the insect's midgut produces many different families of lipases, namely neutral lipases, acid lipases, lipase2, lipase3, GDSL-like lipase, hormone sensitive lipases, and galactolipases (Turunen, 1979; Horne et al., 2009; Christeller et al., 2011; Chapman, 2013; Gondim et al., 2018). To degrade lipid substrates all lipases apply a similar reaction mechanism, which is typical to the α/β hydrolase fold superfamily of proteins they belong to. This mechanism is based on the active site's catalytic triad of residues, usually being Ser-His-Asp/Glu. Most lipases are capable of hydrolyzing a wide range of substrates, albeit with variable specificity. They typically cleave fatty acid residues from triacylglycerol (TAG), diacylglycerol (DAG), monoacylglycerol (MAG) and phospholipids. Moreover, many of these lipases can also hydrolyze carboxylester and thiolester substrates (Terra and Ferreira, 1994; Chapman, 2013).

Total gut lipase content and composition varies among insect species and largely depends on the gut environment (*i.e.* gut pH), diet composition and dietary requirements. Due to divergent feeding habits, insects encounter different lipid compositions in their respective diets and therefore specific repertoires of lipases with different substrate preferences are required for efficient dietary lipid uptake (Chapman, 2013). For example, in the midgut of larval *Epiphyas postvittana* (Lepidoptera), six neutral and three acid lipases were detected. All identified lipases had alkaline pH optima matching the pH of the lepidopteran midgut. Sequence analyses revealed that the neutral lipases were only capable of hydrolyzing phospholipids and galactolipids, but not TAG. This correlates very well with the diet composition of the larvae, since phospholipids and galactolipids are part of the chloroplast thylakoid membranes in the green leaves consumed by the larvae. These neutral lipases probably perform the majority of lipase activity in the lepidopteran midgut, while degradation of TAGs might be an exclusive task of the acid lipases (Christeller et al., 2011). A comparative and functional genomics screen for lipases in four main representatives of holometabolous insect species, namely *D. melanogaster* and *A. gambiae* (Diptera), *Apis mellifera* (Hymenoptera), *B. mori* (Lepidoptera) and *T. castaneum* (Coleoptera), indicated that divergent mixtures of lipases are active inside the midguts of these animals. Low lipase activity was found in *A. mellifera* and might be the result of their specialized sugar-rich, low-fat diet. In contrast, in *T. castaneum*, a fourfold higher lipase concentration was observed, probably related to their high lipid containing diet (Horne et al., 2009).

Since insects are unable to biosynthesize sterols *de novo*, they need to acquire certain sterols from the environment, either from their diet or from gut microbial symbionts (Svoboda, 1999; Chapman, 2013). The dominant sterol in insects is cholesterol. Depending on their diet, insects

will either directly or indirectly obtain cholesterol from their food. Carnivorous insects obtain cholesterol directly from their food, while herbivorous and fungivorous insects synthesize most cholesterol from dietary phytosterols, mainly sitosterol and campesterol, via intermediary dealkylation pathways in their gut (Ciufo et al., 2011; Jing et al., 2012). Degradation of dietary sterols is most probably mediated inside the midgut lumen by the combined action of dehydrogenases, epoxidases and reductases (Svoboda, 1999). This has been demonstrated in several insect species including *A. aegypti, A. mellifera,* and *B. mori* (Svoboda, 1999; Ciufo et al., 2011). Next, the resulting cholesterols are absorbed from the midgut by specific transporter systems. After absorption from the midgut lumen, cholesterol is typically incorporated into (peripheral) cell membranes, or used for the synthesis of ecdysteroids, which are crucial hormonal regulators of insect post-embryonic development (Niwa and Niwa, 2014).

1.5. Intestinal nutrient absorption

Once the food has been processed inside the intestine, the end products of digestion are absorbed by the apical cell membrane of the midgut epithelial cells. Absorbed molecules can then be further processed inside the epithelial cells or be released from the basal membrane into the haemolymph to be transported to their respective site(s) of action or specific storage tissues. This section will describe the most common ways of intestinal nutrient absorption in insects.

Table 1.2 | General overview of putative insect nutrient transporter systems.

AMINO ACID ABSORPTION	CARBOHYDRATE ABSORPTION	LIPID ABSORPTION
amino acid ○ insect nutrient amino acid transporters of the solute carrier family 6 (iNAT-SLC6; Table 3) ○ solute carrier family 7 (SLC7; Table 3)	monosaccharide ○ glucose facilitator 2 (GLUT2) ○ glucose facilitator 5 (GLUT5) ○ sodium-driven glucose symporter (SGLT)	free fatty acids ○ fatty acid binding protein (FABP) ○ fatty acid transport protein (FATP) ○ scavenger receptor class B type I (Sr-BI)
oligopeptide ○ solute carrier family 15 (SLC15; Table 3)	disaccharide ○ sucrose transporter protein (SCRT)* ○ Trehalose transporter 1 (Tret1)	sterol ○ Niemann-Pick C1 (NPC1) ○ sterol carrier protein (SCP)

* Only characterized in *D. melanogaster* (*Dm*SCRT) (Meyer et al., 2011)

1.5.1. Protein absorption

The common end products of protein digestion in the insect midgut are a mixture of small peptides and free amino acids. The lipid bilayer of the midgut epithelial membrane only allows passive diffusion of a negligible amount of end products, while the vast majority need to be absorbed from the midgut lumen by both active or passive transporters located in the apical membranes of the enterocytes (Chakrabarti and Deamer, 1994). Some transporter activity has also been detected in the proventriculus and the hindgut, associated with the action of the Malpighian tubules (Chapman, 2013). Several amino acid transporters, with clear homology to mammalian transporter systems, have been identified in insects. These transporters include

cationic acid transporters, ion-dependent and independent amino acid transporters, and oligopeptide transporters (Miguel-Aliaga et al., 2018). All identified transporters putatively mediating the trafficking of amino acids and oligopeptides across the gut epithelial membrane belong to the Solute Carrier (SLC) transporter family (Table 1.2). This SLC family consists of different types of electrochemical-energy-coupled ‘‘secondary’’ membrane transporters originally enlisted by the Human Genome Organization (HUGO) (Hediger et al., 2004). Generalized functional descriptions of characterized human SLC superfamily members can be found in several mini-reviews (Hediger et al., 2004; Schlessinger et al., 2013).

In insects, the majority of identified nutrient amino acid transporters (NATs) belong to the SLC6 subfamily. Within the SLC6 subfamily, insect NATs are referred to as insect NATs or iNATs (Boudko, 2012). Similar to their vertebrate variants, these iNAT-SLC6 transporters are integral parts of the plasma membrane and typically use Na^+ ions to actively translocate amino acids against their concentration gradient into the cell (Bröer, 2006). Multiple iNAT-SLC6 members have been identified in the genomes of various insects, but to date, only a limited number of transporters have been functionally characterized across just a few insect species (Boudko, 2012; Meleshkevitch et al., 2013). To the best of my knowledge, a total of 9 iNAT-SLC6 transporters have been functionally characterized in insects: 2 in *B. mori*, 1 in *L. decemlineata,* 2 in *A. aegypti*, 3 in *A. gambiae*, and 1 in *D. melanogaster* (Table 1.3). However, additional putative iNAT-SLC6 transporters were identified in the genomes of the Dipteran insects *A. gambiae* (7 members), *A. aegypti* (6 members) and *D. melanogaster* (9 members), suggesting that these insects rely on expanded iNAT-SLC6 transporter networks for nutrient amino acid uptake (Okech et al., 2008). Interestingly, the few characterized iNAT-SLC6 transporters either exhibit a broad substrate spectrum, allowing the trafficking of various amino acids with uniform affinity, or are substrate specific, only allowing the trafficking of specific types of amino acids (Boudko, 2012). Therefore, the observed iNAT-SLC6 gene enrichment and functional differentiation is believed to significantly increase the nutrient amino acid absorption spectrum and efficiency of the midgut epithelium (Boudko, 2012; Chapman, 2013).

The first characterized iNAT-SLC6 transporters were a K^+-coupled amino acid transporter (KAAT1) and a cation-anion-activated amino acid transporter/channel (CAATCH1) both identified in the midgut of the lepidopteran species, *Manduca sexta* (Castagna et al., 1998; Feldman et al., 2000). Remarkably, while transporters belonging to the NAT-SLC6 family are described to couple to Na^+ ions, those identified in *M. sexta* strongly couple to K^+, and less to Na^+ ions. This is an adaptation to the alkaline luminal midgut environment created by the action of H^+ V-type ATPase pumps located in the plasma membrane of the goblet cells, which excrete H^+ into the gut lumen, which on its turn is exchanged for K^+ by the action of $2H^+/K^+$ exchangers. The resulting net K^+ electrochemical gradient is used by the epithelial *M. sexta* transporters to drive amino acid absorption into the cells. Both transporters exhibit broad and overlapping substrate specificity. Additionally, the pH optima of both *Ms*KAAT1 and *Ms*CAATCH1 transporter systems also strongly correlate to the high alkaline midgut pH of Lepidoptera (Castagna et al., 1998; Feldman et al., 2000; Castagna et al., 2009). The transporters *Ae*NAT1 and *Dm*NAT1 were the first Dipteran iNAT-SLC6 transporters characterized in *A. aegypti* and *D. melanogaster* respectively and both represent close phylogenetic relatives of *Ms*KAAT1 and *Ms*CAATCH1 (Boudko et al., 2005; Miller et al., 2008). They exhibit broad substrate specificity, but, unlike *Ms*KAAT1 and *Ms*CAATCH1, their activity depends on a Na^+ electrochemical gradient. *Dm*Nat1 also shows an unusual ability to transport D-isomers of several amino acids. This is a clear adaptation to its diet, since D-amino acids are abundantly present inside the cell walls of bacteria, which represent a considerable part of the *Drosophila* diet (Miller et al., 2008; Genchi, 2017). Recent research has indicated that the lepidopteran NATs *Ms*KAAT1 and *Ms*CAATCH1 are also capable of transporting D-amino acids across the membrane, and that

their isomeric preference probably depends on the coupled cation (Vollero et al., 2016). Other characterized substrate specific iNAT-SLC6 transporters are *Ag*NAT6 and *Ag*NAT8, identified in *A. gambiae*, exhibiting high selectivity for indole- and phenyl-branched aromatic amino acids respectively (Okech et al., 2008), and *Ae*NAT5 and *Ag*NAT5, identified in *A. aegypti* and *A. gambiae*, both exhibiting high methionine selectivity (Boudko, 2012; Meleshkevitch et al., 2013). More recently, the first coleopteran putative iNAT-SLC6 gene (*Ld*NAT1), exhibiting broad substrate specificity, was identified and characterized in *L. decemlineata* (Fu et al., 2015).

Table 1.3 | Functionally characterized transporters of amino acids or oligopeptides.

GENE NAME	TRANSPORTER FAMILY	INSECT SPECIES	REFERENCE
MsKAAT1	iNAT-SLC6	*M. sexta*	Castagna et al., 1998
MsCAATCH1	iNAT-SLC6	*M. sexta*	Feldman et al., 2000
DmNAT1	iNAT-SLC6	*D. melanogaster*	Miller et al., 2008
AeNAT1	iNAT-SLC6	*A. aegypti*	Boudko et al., 2005
AeNAT5	iNAT-SLC6	*A. aegypti*	Meleshkevitch et al., 2013
AgNAT5	iNAT-SLC6	*A. gambiae*	Meleshkevitch et al., 2013; Boudko, 2012
AgNAT6	iNAT-SLC6	*A. gambiae*	Okech et al., 2008
AgNAT8	iNAT-SLC6	*A. gambiae*	Okech et al., 2008
LdNAT1	iNAT-SLC6	*L. decemlineata*	Fu et al., 2015
AaCAT1	SLC7-CAT	*A. aegypti*	Hansen et al., 2011
AaCAT2	SLC7-CAT	*A. aegypti*	Hansen et al., 2011
Dmslif	SLC7-CAT	*D. melanogaster*	Colombiani et al., 2003
DmMnd	SLC7-HAT	*D. melanogaster*	Martin et al., 2000
DmjhI-21	SLC7-HAT	*D. melanogaster*	Reynolds et al., 2009
DmOPT1	SLC15	*D. melanogaster*	Roman et al., 1998

Abbreviations: KAAT = K^+-coupled amino acid transporter; CAATCH = cation-anion-activated amino acid transporter/channel; NAT = nutrient amino acid transporter; CAT = cationic amino acid transporter; slif = slimfast; mnd = minidiscs; OPT= oligopeptide transporter; iNAT = insect nutrient amino acid transporter; SLC = solute carrier; HAT = heterodimeric amino acid transporters; *Ms* = *Manduca sexta*; *Dm* = *Drosophila melanogaster*; *Ae* = *Aedes aegypti*; *Ag* = *Anopheles gambiae*; *Ld* = *Leptinotarsa decemlineata.*

The iNAT-SLC6 transporter family is clearly essential for the dietary uptake of essential amino acids. In addition, other types of transporters associated with the midgut epithelium might also contribute to the uptake of dietary amino acids and oligopeptides. Suggested candidate transporters are members of the SLC7 and SLC15 family (Table 1.2). The SLC7 transporter family is well-characterized in mammals and includes the cationic amino acid transporters (CATs), mediating transport of essential cationic amino acids, and the heterodimeric amino acid transporters (HATs), transporting mainly essential amino acids across the plasma membrane (Verrey et al., 2004). Insect SLC7 transporters mediating amino acid transport into the fat body cells have been identified in *D. melanogaster* and *A. aegypti* (Martin et al., 2000; Colombani et al., 2003; Reynolds et al., 2009; Hansen et al., 2011; Boudko, 2012). Nevertheless, hitherto, their contribution to amino acid uptake across the midgut epithelium membrane has not yet been experimentally confirmed. For example, the CAT-SLC7 transporter, *Aa*CAT1, characterized in *A. aegypti* is highly expressed in the fat body and shows Na^+-independent cationic amino acid transporter activity with high L-histidine selectivity at

neutral pH (Hansen et al., 2011). However, the contribution of *Aa*CAT1 to the uptake of dietary amino acids can only be hypothesized, since its distribution across the alimentary tract of *A. aegypti* has not yet been investigated (Boudko, 2012).

Alongside dietary amino acid uptake, the trafficking of peptides across the midgut epithelium by the action of oligopeptide transporters is also possible. This is based on the predicted parallel to mammals, where the SLC15 transporters have been identified as a family of oligopeptide transporters (Hediger et al., 2004; Schlessinger et al., 2013). However, up to date, the only characterized insect SLC15 transporter with activity in the alimentary canal is the *D. melanogaster* oligopeptide transporter 1 (*Dm*OPT1). The OPT1 transporter exhibits broad substrate specificity, with highest selectivity for L-alanine and alanine derived short peptides, and is predicted to be involved in the initial absorption of oligopeptides in a proton-mediated manner (Roman et al., 1998). Genomic studies have revealed the presence of the SLC15 in other investigated insect genomes, including the Diptera *A. gambiae, A. aegypti*, indicating a putative universal function of SLC15 transporters as dietary oligopeptide transporters in insects. Though, further functional characterizations are needed to fully elucidate their functions, it is believed that both SLC7 as SLC15 transporters could potentially complement the nutrient transporter activity of iNAT-SLC6 to a certain degree. But until further investigation, their contribution to dietary nutrient uptake across the midgut epithelium should not be overestimated (Boudko, 2012; Fu et al., 2015).

1.5.2. Carbohydrate absorption

The digestive end products of carbohydrates are mainly monosaccharides and disaccharides. These simple sugars are absorbed across the midgut epithelium into the ECs for further processing or allocation to peripheral tissues. Carbohydrates can be used as direct energy sources, but can also be stored as energy reserves, predominantly in the form of TAGs, in the fat body. In mammals, the internalization of dietary monosaccharides in intestinal cells is almost exclusively transporter mediated. Similarly, in insects, two major types of glucose transporter systems have been identified, namely the major facilitator superfamily (MFS) glucose facilitators (GLUT2 and GLUT5) and the sodium-driven glucose symporters (SGLTs). Both transporter systems belong to the SLC transporter family (Table 1.2) (Hediger et al., 2004). GLUTs belong to the SLC2 family and SGLTs to the SLC5 family (Scheepers et al., 2004; Kellett and Brot-Laroche, 2005). Few reports on the presence and activity of GLUT and SLGT transporters in the insect alimentary tract exist (see for example Escher and Rasmuson-Lestander, 1999; Caccia et al., 2005, 2007; Price et al., 2007, 2010; Bifano et al., 2010; Price and Gatehouse, 2014; Kikuta et al., 2015; Govindaraj et al., 2016). However, a model for dietary sugar absorption in insects is still lacking, and therefore, most of our understanding about dietary carbohydrate absorption in insects is based on expected parallels with the classical mammalian model (Chapman, 2013). In mammals, the SGLT1 and GLUT5 transporters are present on the apical membrane of the intestinal ECs, while GLUT2 is detected on both the apical and basolateral sides. Functional studies show that SGLT1 is responsible for the Na^+-mediated uptake of dietary glucose across the apical membrane of the gut lumen, GLUT5 is responsible for the facilitated transport of dietary fructose, and GLUT2 can transfer both glucose and fructose. Moreover, GLUT2 is located at both the apical and basolateral membrane and is involved in the movement of glucose and fructose both in and out the gut epithelial cells. Since homologous transporters have been identified in several insect species (see references above), similar dietary sugar uptake mechanisms are expected to be active in insects. This was for the first time experimentally observed in the hymenopteran parasitoid of aphids, *Aphidius ervi* (Caccia et al., 2005, 2007). However, to date, further

experimental data supporting this mechanism of dietary monosaccharide absorption in insects remain scarce (Miguel-Aliaga et al., 2018).

A putative mechanism for the uptake of disaccharides was first discovered in *D. melanogaster*. Meyer and colleagues discovered that a protein, named sucrose transporter protein (SCRT), highly similar to members of the SLC45 transporter family, demonstrated sucrose transporter activity on the apical side of the hindgut and the vesicular membranes of ovarian follicle cells (Table 1.2) (Meyer et al., 2011). The apical position along the alimentary tract suggests a role in the uptake of dietary sucrose. However, until today, no other records on putative sucrose absorbing mechanisms in insects exist. Another suggested dietary disaccharide transporter is the facilitated trehalose transporter 1 (Tret1) identified in the gut of various insect species (Table 1.2) (Kikawada et al., 2007; Kanamori et al., 2010). The disaccharide trehalose (α-D-glucopyranosyl-(1,1)-α-D-glucopyranoside) is known to be the dominant haemolymph sugar in most insects. Trehalose is mainly synthesized from dietary sugars inside the fat body, and plays a key role as mobile energy store (Thompson, 2003). Up to date, the transporter function of Tret1 has mainly been analyzed in the fat body and other peripheral tissues, where it plays a role in the transmembrane trafficking of trehalose. Its putative function in the absorption of dietary disaccharides from the intestinal lumen has largely been neglected and remains to be further investigated (Miguel-Aliaga et al., 2018).

1.5.3. Lipid absorption

The typical end products of lipid digestion in the gastrointestinal tract are free fatty acids, glycerol, phospholipids and non-esterified cholesterol (Majerowicz and Gondim, 2013). The majority of these compounds, together with sterol, can diffuse rather easily across the plasma membrane of the midgut epithelium, whether or not facilitated by emulsification. This emulsification is achieved by the formation of fatty acid, amino acid and glycolipid complexes, as well as fatty acids and lysophospholipid micelles (Chapman, 2013; Miguel-Aliaga et al., 2018). The majority of absorbed dietary lipids are stored as TAG energy reserves in the fat body or transported to the oocytes during female oogenesis (Majerowicz and Gondim, 2013; Chapman, 2013; Fruttero et al., 2017).

The diffusion of hydrophobic ligands across the plasma membranes of different tissues is facilitated by fatty acid binding proteins (FABP). These membrane associated cytosolic proteins have clear lipid binding potencies and were first characterized in mammals. FABPs have also been identified in several insect tissues, but a limited body of research exists on their putative nutritional role in insects (Esteves and Ehrlich, 2006). Some reports clearly demonstrate the presence of FABPs in the midgut epithelium of the investigated insect, hence suggesting a function in dietary lipid uptake (Table 1.2). FABPs have been identified in the midguts of the Lepidoptera *M. sexta* (MFB1, MFB2) (Smith et al., 1992), *B. mori* (*Bm*FATP) (Ohnishi et al., 2009), and *Spodoptera litura* (*Sl*fabp1) (Huang et al., 2012), as well as the Hymenoptera *Aphidius ervi* (*Ae*FABP) (Caccia et al., 2012). Nevertheless, further experimental investigation is needed to better understand their contributions to the dietary lipids absorption. Once absorbed inside the midgut ECs, free fatty acids are further processed to either be oxidized in the mitochondria or used for the synthesis of TAGs, DAGs and phospholipids. Lipids are mainly transported towards the peripheral storage tissues as DAGs bound to carrier proteins, called lipophorins, where they are subsequently stored as TAGs (Gondim et al., 2018). Fast handling of the absorbed dietary lipids helps to ensure low lipid concentrations inside the ECs, which creates a favorable concentration gradient promoting lipid diffusion across the membrane (Chapman, 2013). Next to FABPs, other membrane associated

transporter proteins, namely the fatty acid transport proteins (FATPs) and the scavenger receptors class B type I belonging to the CD36 protein family, might be involved in the active transport of dietary fatty acids across the midgut membrane (Table 1.2). This expectation is mainly based on their observed fatty acid binding activities in peripheral tissues, for example in the silk glands (CD36) and the pheromone glands (FATP) of *B. mori* (Ohnishi et al., 2009; Sakudoh et al., 2010). However, no experimental data supporting their role in dietary lipid uptake across the midgut epithelium exist (Majerowicz and Gondim, 2013; Miguel-Aliaga et al., 2018).

Since insects are unable to synthesize sterols *de novo,* their dietary absorption is of vital importance. Several mechanisms of sterol uptake have been proposed, involving both passive and active migration across the plasma membrane. Unfortunately, the molecular mechanisms behind dietary sterol absorption in insects are not yet fully elucidated. Based on vertebrate studies, two types of sterol-binding proteins have recently gained attention in insects: the Niemann-Pick C1 proteins and the sterol carrier proteins (SCPs) (Table 1.2). The Niemann-Pick C1 proteins are a well-described family of transmembrane sterol-binding proteins in mammals, which are described to be critical for the cholesterol absorption in the intestine. Two members of the Niemann-Pick C1 family have been identified in the gastrointestinal tract of mammals, namely the Niemann-Pick C1 (NPC1), which is important for intracellular cholesterol transport, and the Niemann-Pick C1-like 1 (NPC1L1), which is important in general sterol absorption (Dixit et al., 2007). Two Niemann-Pick C1 genes were first identified in *D. melanogaster*. These closely related Niemann-Pick C1 homologs, called NPC1a and NPC1b, showed higher sequence similarity with NPC1 than with NPC1L1. NPC1b is only expressed in the midgut, while NPC1a has a wide, uniform tissue distribution. Functional characterization showed that NPC1b is responsible for the dietary cholesterol absorption inside the midgut epithelium, while NPC1a plays a role in intracellular sterol trafficking in the midgut and peripheral tissues (Voght et al., 2007; Zheng et al., 2018). More recently, the available genomes and transcriptomes of 39 insect species belonging to 10 different orders were analyzed for the presence of putative Niemann-Pick C1 genes. Similar to *D. melanogaster*, most studied insect genomes contained two *NPC1* genes, but a few had only one *NPC1* gene. Moreover, this research also indicated a gut-specific expression of *NPC1b* among the studied insects (Zheng et al., 2018). Another type of well characterized sterol-binding proteins in vertebrates are the sterol carrier proteins (SCP2). The *SCP2* gene contains two distinct promoter sites, resulting in two different sterol carrier proteins: SCPx (sterol carrier protein-x), and SCP2 (sterol carrier protein-2) (Ohba et al., 1995). Some SCP2 encoding genes have been identified in insects, including *D. melanogaster* (Kitamura et al., 1996), *A. aegypti* (Krebs and Lan, 2003), *S. littoralis* (Takeuchi et al., 2004), and *B. mori* (Gong et al., 2006). Tissue distributions of these identified insect SCPs suggest a mediating role in both the uptake and translocation of cholesterol. Moreover, RNA interference (RNAi) mediated loss-of-function studies in *S. litura* have recently demonstrated the role of SCP-x in the uptake of cholesterol across the midgut epithelium (Guo et al., 2009). Further functional characterizations of the above mentioned putative sterol-binding proteins are pivotal for understanding the precise mechanisms of dietary sterol uptake in insects.

1.6. Insect gut transcriptome studies

From the above described paragraphs, it should be clear that the way food is handled inside the gastrointestinal tract varies dramatically between different insect species. Most of this variation can be attributed to the diverse feeding strategies exhibited by insects, ranging from herbivory to hematophagy. Furthermore, insects have evolved numerous coping strategies to

overcome different host defensive strategies, which are displayed by their overall large repertoire of digestive enzymes and defensive molecules. The ways in which insects internalize the degraded macronutrients from the intestinal lumen and translocate these throughout the body are far less described. Despite a consensus that the general mechanisms are probably analogous to those observed in vertebrates, there are only a limited number of experimental studies supporting this hypothesis. Consequently, more in-depth studies are essential to shed light on the exact mechanisms of nutrient uptake and allocation in insects.

Although transcript profiling has become a standard approach to investigate biological processes, studies specifically focusing on the insect gut transcriptome appear to be limited. Only a restricted number of annotated, or partially annotated, insect gut transcriptomes are publicly available on PubMed (https://www.ncbi.nlm.nih.gov/pubmed). Table 1.4 summarizes the total of 43 research papers describing the gut transcriptome of 42 different insect species belonging to 7 different insect orders that could be found on PubMed. Immediately apparent from this table is the high number of studies performed with insects belonging to the orders of the Diptera, Coleoptera and Lepidoptera. Currently, there are only two Orthoptera with a publicly available annotated midgut transcriptome: *Locusta migratoria* and *Oedaleous asiaticus* (Spit et al., 2016; Huang et al., 2017).

Table 1.4 | Summary of all the (partially) annotated insect gut transcriptomes available on PubMed.

INSECT ORDER	INSECT SPECIES	REFERENCE
Diptera	*Aedes aegypti*	Sanders et al., 2003
	Culicoides sonorensis	Campbell et al., 2005
	Phlebotomus papatasi	Ramalho-Ortigão et al., 2007
	Lutzomyia longipalpis	Oliveira et al., 2008; Pitaluga et al., 2009
	Mayetiola destructor	Zhang et al., 2010
	Leishmania infantum	Dostálová et al., 2011
	Musca domestica	Gazara et al., 2017
	Drosophila melanogaster	Bost et al., 2018
Coleoptera	*Callosobruchus maculatus*	Pedra et al., 2003; Chi et al., 2009
	Tribolium castaneum	Oppert et al., 2009; Oppert et al., 2018
	Chrysomela tremulae	Pauchet et al., 2009
	Phaedon cochleariae	Kirsch et al., 2012
	Anoplophora parallela	Scully et al., 2013
	Leptinotarsa decemlineata	Smagghe et al., 2013
	Holotrichia parallela	Shu et al., 2015
	Dendroctonus ponderosae	Keeling et al., 2016
	Cosmopolites sordidus	Valencia et al., 2016
	Tenebrio molitor	Gazara et al., 2017; Oppert et al., 2018
	Holotrichia oblita	Jiang et al., 2017
Lepidoptera	*Chilo suppressalis*	Ma et al., 2012
	Lymantria dispar	Sparks et al., 2013
	Helicoverpa zea	Hum-Musser et al., 2013
	Plutella xylostella	Lei et al., 2014
	Trichoplusia ni	Herde and Howe, 2014
	Heliothis virescens	Popham et al., 2015
	Mythimna separata	Liu et al., 2016
	Bombyx mori	Qian et al., 2016; Wang et al., 2016b; Yang et al., 2016
	Antheraea pernyi	Wang et al., 2016a
	Pectinophora gossypiella	Fabrick et al., 2016

	Spodoptera frugiperda	Gazara et al., 2017
	Procecidochares utilis	Li et al., 2018
	Pieris rapae	Xiang et al., 2018
	Achaea janata	Dhania et al., 2019
Hemiptera	*Rhodnius prolixus*	Da Silva et al., 2014
	Bemisia tabaci	Ye et al., 2014; Shao et al., 2018
	Dysdercus peruvianus	Gazara et al., 2017
Orthoptera	*Locusta migratoria*	Spit et al., 2016
	Oedaleous asiaticus	Huang et al., 2017
Blattidae	*Periplaneta americana*	Zhang et al., 2016
Hymenoptera	*Apis mellifera ligustica*	Xiong et al., 2017

So far, only few research has actually focused on the changes in the midgut transcriptome during food uptake, digestion and nutrient absorption. Most insect gut transcriptomic studies in Table 1.4 are either exploratory studies, investigating overall transcript enrichment, or more applied studies, for example investigating immunity or xenobiotic metabolism in the digestive tract. With different studies focusing on different aspects of the digestive tract, it is difficult to compare these data. Nevertheless, these studies did allow us to notice some obvious parallels among annotated gut transcriptomes of insects belonging to different insect orders.

In general, most investigated insect gut transcriptomes appear to contain broad collections of digestive enzymes, putative nutrient transporters and enzymes mediating immunity, stress response and xenobiotic metabolism. Especially herbivorous insects, feeding on leaves and seeds, harbor a remarkably high fraction of transcripts encoding proteolytic enzymes in their gut. Examples of such herbivorous insects include amongst others the Coleoptera, *C. tremulae*, *T. castaneum* and *C. sordidus* (Pauchet et al., 2009; Oppert et al., 2009; Valencia et al., 2016), the Lepidoptera, *L. dispar* and *T. ni* (Sparks et al., 2013; Herde and Howe, 2014), the Diptera, *M. destructor* (Zhang et al., 2010), and the Orthoptera, *L. migratoria* (Spit et al., 2016). But also the blood-sucking insects, *A. aegypti, C. sonorensis* and *R. prolixus* have an extensive set of transcripts encoding proteolytic enzymes in their gut (Sander et al, 2003; Campell et al., 2005; Da Silva et al., 2014). Transcripts encoding carbohydrases and lipases are also present in these insects. However, in some insects, these appear to be less represented than transcripts encoding proteolytic enzymes. This was observed in *C. tremulae, M. destructor*, *C. sordidus* and *L. migratoria* (Pauchet et al., 2009; Zhang et al., 2010; Valencia et al., 2016; Spit et al., 2016). These varying gut enzymatic profiles probably reflect the dietary preferences of the insect host, but this needs to be further investigated (Zhu-Salzman and Zeng, 2015).

All transcriptomic studies in Table 1.4 that investigate the xenobiotic metabolism in the insect intestinal tract clearly revealed a strong ability to metabolize different toxins introduced by the food, *i.e.* plant-derived xenobiotics, bacterial entomotoxins or synthetic insecticides. Consistently high numbers of transcripts encoding cytochrome P450 monooxygenases (P450s), carboxylesterases (CEs), glutathione S-transferases (GSTs) and uridine diphosphate-glucuronosyltransferases (UGTs), typically described to be involved in distinct phases of xenobiotic metabolism in insects, were observed in a variety of insect species (Puchet et al., 2009; Zhang et al., 2010; Hum-Musser et al., 2013; Scully et al., 2013; Sparks et al., 2013; Popham et al., 2015; Zhang et al., 2016; Spit et al., 2016; Li et al., 2018). This is indicative for a general ability and flexibility of insects to cope with different challenges posed by their diet and is especially very common among herbivorous insects, which naturally encounter a vast amount of plant-derived defensive molecules in their diet.

In contrast to our generally profound knowledge about transcript identities in the insect gut, far less is known about how the gut responds at the transcript level to the passage of food. However, this knowledge is a crucial missing link in our understanding of the complexity of this tissue. It will help to further elucidate the blind spots of insect digestion research, such as the way nutrients are absorbed and processed by the midgut epithelial cells or how food availability in the gut lumen is perceived. Until now, this has only been experimentally addressed in a few insect species, while most of our knowledge so far is largely based on putative parallels with vertebrates. Moreover, understanding insect gut physiology might be crucial for the development of new alternative pest management strategies, as will be explained in the next paragraphs.

1.7. Combating insect pests for crop protection

1.7.1. Future challenges for sustainable agriculture

In a report by the Food and Agricultural Organization of the United Nations (FAO) published in 2017, it was estimated that the world's population will grow to reach almost 10 billion by 2050, consequently boosting agricultural demands by an expected fifty percent compared to 2013. This projected growth estimate of the world's population is expected to be concentrated in Africa and South Asia and in the world's cities, resulting in the expectation that two-thirds of the global population will be living in urban areas by 2050. To meet the associated ever increasing global demand for food, farmers worldwide will have to produce more food per unit land than they do today. Moreover, increasing urbanization is expected to result in a 1.8-2.4% loss of global croplands by 2030, of which around 80% will occur in Asia and Africa (Bren d'Amour et al., 2017). Expansion of agricultural land is hence prospected to be mostly at the expense of forests and other valuable ecosystems (Erb et al., 2016). Given these limitations, sustainable production and increasing productivity on the available agricultural land are corner stones to secure food production, as well as precious natural land, in the future. However, increase of production is strongly hampered by different challenges across regions, including climate change, and crop loss due to pests and diseases (Calicioglu et al., 2019).

Around 10-16% of the yearly global crop yield is lost to a variety of pests, including insects, fungi and microorganisms (Oerke, 2006). It is for example estimated that around 15% of some of the world's most important crops, such as wheat, rice and maize, are lost every year due to various insect pests (FAO). Thousands of insects are known to threaten crop production around the world. Some of the major pest insects belong to the insect classes of the Coleoptera, Hemiptera, Lepidoptera, and Orthoptera (Brezeanu et al., 2014). Examples of relevant pest insects include the western corn root worm, currently costing US farmers about 1 billion U.S. dollars yearly due to crop losses, and the desert locust, which feeds on a wide variety of crops, and poses great threats to local farmers in Saharan African countries and the Middle-East (Cumming et al., 2008; Flagel et al., 2014). Furthermore, climate change is expected to result in the rise of insect pest activity. A recent study by Deutsch and colleagues demonstrated that pest-related yield losses of wheat, rice and maize are predicted to increase by as much as 46%, 19% and 31%, respectively, when global temperature would rise 2 °C above pre-industrial levels, which is the limit set by the Paris Agreement of the United Nations formulated back in 2015 (Deutsch et al., 2018) (https://unfccc.int/process-and-meetings/the-paris-agreement/d2hhdC1pcy).

Overall, it is clear that if we want to continue feeding the rapidly growing world population in ever increasing difficult circumstances, effective global insect pest management strategies are crucial.

1.7.2. Insect pest management

Protection of food supplies already starts in the field, where numerous insect pests form serious threats for crops. Over the past decades, several strategies have been developed to increase food availability by improving crop protection from these insect pests. Especially chemical insecticides, such as organochlorine, organophosphorus and carbamate insecticides, have been playing indispensable roles in agriculture since the 1960s (Popp et al., 2013; Nicolopoulou-Stamati et al., 2016). Because of their relatively low costs and high effectiveness, many synthetic insecticides are nowadays still widely used for crop protection, especially in developing countries (FAO). It is estimated that the annual global chemical-pesticide market, including insecticides amongst others, is about 9 million tons associated with a total cost of *ca.* 50 billion U.S. dollars (Phillips McDougall, 2018).

However, the use of chemical insecticides is becoming more and more ecologically unacceptable, and the production of several of these, such as dichlorodiphenyltrichloroethane (DDT), has already been banned (Aktar et al., 2009). As broad-spectrum insecticides, these chemical compounds are known to cause a large number of negative health and environmental effects. Many of these have direct effects on human health through contact with the skin, ingestion, or inhalation, as well as indirect effects through food commodities. The numerous hazardous effects associated with these chemicals include, amongst others, carcinogenic, dermatological, endocrine, gastrointestinal, neurological, respiratory, and reproductive effects (reviewed by Aktar et al., 2009; Nicolopoulou-Stamati et al., 2016). Not only humans, but also other non-target organisms, such as plants, microorganisms, animals, and insects, can be affected by insecticides. Most chemical insecticides are persistent and remain in the ecosystem for a long period of time. Moreover, they can easily spread in the environment through, for example, the wind, surface runoff, nearby streams of water, *etc.* (Aktar et al., 2009). Recent studies have demonstrated that the extensive usage of insecticides has caused a strong decline of birds in grasslands, and aquatic organisms in streams in the U.S. over the past decades (Mineau and Whiteside, 2013; Beketov et al., 2013). Additionally, chemical insecticides are a major driver behind the observed global decline in insect numbers, including beneficial insects, such as bumblebees and ladybugs (reviewed by Desneux et al., 2007; Sánchez-Bayo and Wyckhuys, 2019).

In addition to their detrimental effects on the environment, many chemical insecticides are rapidly losing their effectiveness as a result of the growing insecticide resistance of pest insects. It is estimated that over 500 species of pest insects are already resistant to one or more insecticides, and these numbers are predicted to grow exponentially (Bass and Field, 2011; Gould et al., 2018; South and Hastings, 2018). Insecticide resistance is generally defined as the decreased sensitivity of a pest population to an insecticide that was previously effective at controlling that population. Natural selection boosts insecticide resistance: insects with resistance traits survive the insecticide and pass on these traits to their offspring. When the same insecticide, or insecticides with a similar mode of action, are continued to be applied to combat pest populations on the same field, then resistant individuals will increase in the population, while susceptible individuals will be eliminated by the insecticide. Eventually, only those insects with resistant traits will remain, making the insecticide no longer effective in combating the pest population (Hawkins et al., 2019). There are several documented ways

insects can become resistant to insecticides, of which metabolic resistance, the ability of an insect to detoxify or destroy the toxin using endogenous enzymes, is predominantly observed. Another important mechanism of resistance is target-site resistance resulting from standing genetic variations or *de novo* mutations in the insecticide target site, making the insect not or no longer susceptible (Chandler et al., 2011; Hawkins et al., 2019).

It is indisputable that the many disadvantages of heavy chemical insecticides, sometimes with similar modes of actions, are gradually outweighing their advantages, and novel strategies for successful pest management are stringently needed.

1.7.3. Search for eco-friendly alternatives

One simple way of reducing the amount of chemical insecticides is by using them more considerately, as part of integrated pest management (IPM) techniques. Briefly, IPM is described as the combined approach of all available biological and chemical control strategies to combat plant pests, with minimal applications of chemical pesticides (Knipling, 1972). However, IPM techniques are sometimes so costly, time consuming, and complicated that local farmers still mostly rely on chemical insecticides to combat insect pests (Hokkanen, 2015). Therefore, better alternatives than chemical insecticides are pivotal if we want to effectively reduce their use in the field (Popp et al., 2013).

In this context, the search for efficient eco-friendly alternatives, such as biopesticides, has become a priority for the crop protection industry. Biopesticides are natural materials derived from animals, plants, and bacteria that are used for pest control. Today, they comprise only a small share of the global crop protection market, with an estimated value of around 3 billion U.S. dollars, equal to 5% of the total global crop protection market. However, biopesticide usage at a global scale is increasing by almost 10% every year (Kumar and Singh, 2015; Damalas and Koutroubas, 2018). In the United States, more than 200 biopesticide products are commercially available, compared to 60 analogous products in the European Union (EU), mostly due to stricter regulations (Olson, 2015).

The vast majority of all microbial biopesticides currently available on the market are derived from one entomopathogenic bacterium, *Bacillus thuringiensis* (Bt) (Damalas and Koutroubas, 2018). During the stationary phase of its growth cycle, Bt produces spores containing protein crystals, typically consisting of one or more Cry or Cyt proteins (δ-endotoxins). Once ingested by the insect, these δ-endotoxins bind to specific receptors located in the insect midgut epithelial membrane, resulting in cell disruption and premature insect death (reviewed by Hofte and Whiteley, 1989; Broderick et al., 2006; Sanahuja et al., 2011). Interestingly, Bt toxins act host-specifically, and different strains of Bt produce different types of δ-endotoxins, each affecting a narrow taxonomic group of insects. Up to now, at least 952 toxin genes have already been identified and characterized in different Bt strains around the world (Jouzani et al., 2017). Because of their specific character, Bt toxins are considered to be safe to people, non-target organisms, and the environment. For these reasons, Bt toxins have been used for crop protection against insect pests for almost a century now (Sporeine was the first commercial topical Bt insecticide released in 1938) (Sanahuja et al., 2011). Nowadays, they are mainly used as topical insecticides, or incorporated into transgenic Bt crop varieties (Chandler et al., 2011). Nevertheless, despite their obvious overall success, there are also some factors limiting their use in crop protection. For example, Bt toxins have a limited host range targeting only those insects with specific binding sites present in their gut (Palma et al., 2014). Moreover,

increasing numbers of insect resistance against Bt toxins are being reported in the laboratory and on the field (Tabashnik et al., 2013).

The advantages as well as disadvantages of Bt toxins clearly highlight the need for additional eco-friendly insecticides. Moreover, it is known that insecticide resistance to both chemical and biological insecticides can be hampered by deploying different insecticides with varying modes of action. This way, the selection pressure directed towards insecticides with a similar mode of action is reduced, and the survival of resistant individuals is significantly decreased (Sparks and Nauen, 2015). Therefore, the development of additional eco-friendly insecticides with novel modes of action is one of the biggest future challenges for the crop protection industry.

1.7.4. Targeting endogenous insect molecules

Newly emerging and promising alternative pest management strategies comprise the specific targeting of non-conserved gene products in pest insects, for example with specific inhibitors or double-stranded (ds) RNA molecules, resulting in rapid lethality. The use of dsRNA in crop protection by downregulating essential gene functions in herbivorous insects through RNAi has been recognized for years (Price and Gatehouse, 2008). Briefly, RNAi is a post-transcriptional gene-silencing mechanism that acts as the main antiviral immune response in insects. This process can also be triggered when applying synthetic dsRNA molecules, specifically targeting endogenous gene products, making it a highly interesting mechanism for the development of specific alternative insecticides (Burand and Hunter, 2013). The molecular mechanism of RNAi in insects is explained in chapter 5 (Ch5|5.1.4). However, despite many ongoing research, only very recently the first RNAi-based insecticide, developed by Monsanto and Dow Agrosciences, and which will be known as SmartStax Pro®, was approved by the United States Environmental Protection Agency (EPA) for commercial use (Vogel et al., 2019). There are several factors still limiting the widespread application of dsRNA in crop protection, including its mode of delivery, the target insect's RNAi response, and finding appropriate target mRNA. This is further discussed in chapter 7 (Ch7|7.4).

The insect alimentary canal, as the barrier between the inner and outer environment, is considered to be an interesting site for application of pest control strategies (Linser and Dinglasan, 2014). It is generally accepted that interfering with the proper functioning of the gut, for example by inhibiting PM formation, or the digestion and uptake of nutrients, will cause severe detrimental effects in the affected insects, weakening them and possibly even resulting in their premature death. Moreover, in several insect species the digestive tract has been demonstrated to be a principal site for the detoxification of both naturally occurring as well as synthetic insecticidal molecules (see, for example Xu et al., 2015; Zhu et al., 2016; Tian et al., 2018; Sun et al., 2019). Inhibiting detoxification pathways in the gut might therefore re-establish the toxicity of insecticides that have previously lost their effectiveness due to insect resistance.

The power of ever developing high-resolution technologies, such as next-generation sequencing (NGS), has already improved the general knowledge of the insect gut functioning, as well as the different detoxification pathways present. As more genomic databases of insect guts are being constructed, more information is assembled for comparative studies between insect species and the identification of genes associated with, for example, the regulation of digestion, host preference, or insecticide resistance. This will ultimately aid in the identification of new molecular targets for novel insecticides.

1.8. *Schistocerca gregaria* as a research organism

The organism used throughout this study is the desert locust, *Schistocerca gregaria*. This insect belongs to the insect order of the Orthoptera and is a member of the family of the short-horned grasshoppers (Acrididae). It usually inhabits the semi-arid and arid deserts of Africa, the Middle-East and South-West Asia. This hemimetabolous insect has a life cycle comprising three distinct stages: the egg stage (ovum and embryo), the larval stage (nymphs or hoppers), and the adult stage (Fig. 1.4). The entire larval stage, typically consisting of five separate stages in which the locust increases in size after every molt, takes about 30 to 40 days. In the adult stage, the locust has functional wings and it becomes sexually mature. Adult desert locusts generally live around two to four months, during which the females can lay at least three times *ca.* 80-150 eggs (Uvarov, 1966).

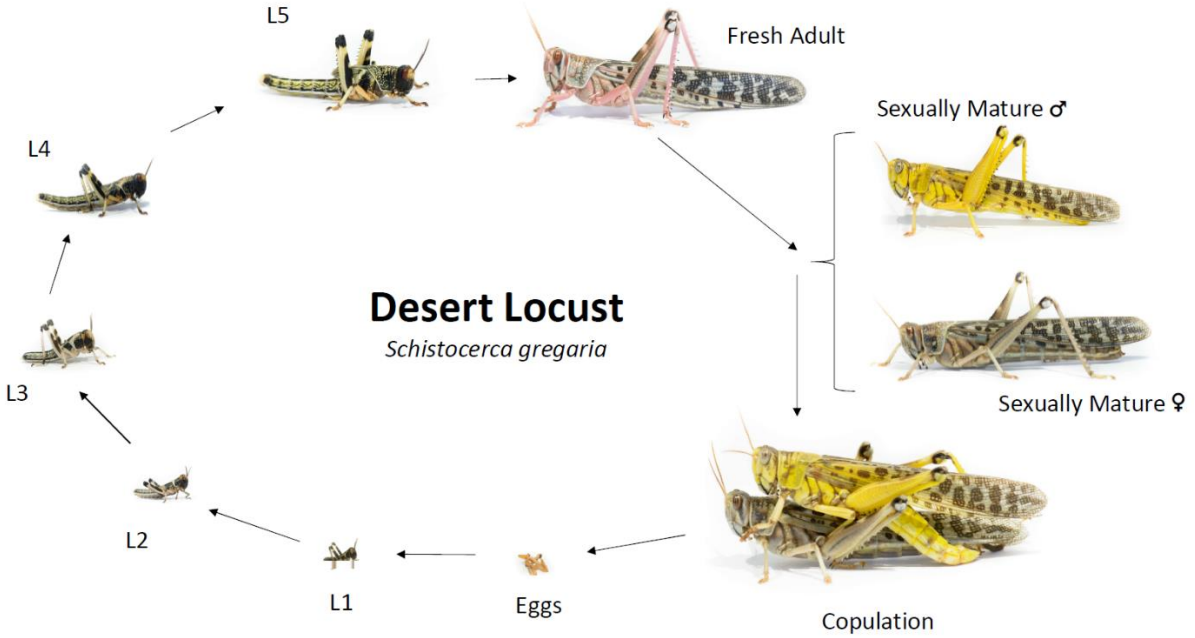


Figure 2.4. The life cycle of gregarious *S. gregaria*. The desert locust's life cycle comprises three distinct stages: the egg stage, the larval stage (L1 – L5), and the adult stage. Picture made by Timon Smeets.

An important feature distinguishing locusts from other grasshoppers in the family of Acrididae is their phase polyphenism, a remarkable type of phenotypic plasticity (reviewed by Pener and Simpson, 2009). Polyphenism is broadly described as the existence of two or more distinct phenotypes which can be induced in individuals of the same genotype by extrinsic factors. The desert locust has two extreme phenotypes: a solitarious phase (*solitaria*) and a gregarious, swarming phase (*gregaria*). Transition between the two phases is a continuous process called phase transition. Solitarious locusts live more isolated and tend to avoid contact with conspecifics. During this phase, locust populations are generally small and dispersed, which represents no economic threat to agriculture. However, when population density increases and locusts are forced together, typically after periods of rainfall, again followed by drought, transition to the gregarious phase is stimulated (Sword et al., 2010). Following this transformation, sometimes occurring over two or three generations, locusts can form dense bands of flightless hoppers and swarms of flying adults, often referred to as 'plagues'. Such locust swarms are sometimes composed of billions of individuals, spanning over a massive area of more than 1,000 km^2, forming devastating threats for (local) agriculture as they decimate every crop they encounter (Cumming et al., 2008). The desert locust, for example, is a highly polyphagous pest insect able to consume a wide range of host plants, including

important economic crops, such as rice, cotton, barley, wheat, sugarcane, and maize. Because of their large range, the voracious locust swarms have high socio-economic impacts in multiple regions in Africa, the Middle-East and South-West Asia (FAO).

Apart from their obvious behavioral differences, solitarious and gregarious locusts also clearly differ in morphology and physiology. For example in *S. gregaria*, striking color differences are observed between the different phases. The color of solitarious nymphs ranges from green to brown, whereas gregarious ones are yellow with black patterning. Solitarious adult *S. gregaria* are beige/brown, whereas gregarious adults are bright yellow (males) or beige/brown-yellowish (females) (pictures of gregarious *S. gregaria* are displayed in figure 1.4). Other differences are observed in body size (solitarious locusts are generally larger than gregarious locusts) and in reproductive processes (gregarious female desert locusts usually mature faster but lay less eggs than solitarious locusts) (Uvarov, 1966; Pener and Simpson, 2009).

Because of its polyphagous, herbivorous lifestyle and its relatively large size, *S. gregaria* is a highly favorable model insect for gastrointestinal studies. Moreover, it is described to exhibit a very potent systemic RNAi response when injected with dsRNA, which allows the straightforward *in vivo* analysis of the function of genes in specific physiological processes (Wynant et al., 2012). For these reasons, and the fact that it is a highly relevant pest insect, *S. gregaria* was used as research organism in the context of this doctoral research.

1.9. Objectives of this study

The main objective of this study was to gain insight in the transcript profile of the *S. gregaria* midgut. Therefore, the messenger RNA (mRNA) content of midguts at ten minutes, two hours and twenty-four hours after food uptake was investigated by means of RNA sequencing (RNA-Seq), which allowed us to investigate the midgut transcriptome during key moments of digestion. As the introduction of this thesis already clearly demonstrated, there is a pressing need for more in-depth studies investigating the mechanisms by which digestion and food uptake are controlled in insects. Additionally, understanding how the insect midgut operates during the digestive process is expected to lead to the identification of novel molecular target sites for insecticides. Therefore, by investigating the transcript profile of the midgut of the desert locust, a useful model organism for herbivorous polyphagous insects, this knowledge can be further expanded and compared to the situation in other insect species. The objectives of this doctoral research can thus be summarized as follows:

- Investigating the general transcript profile of the *S. gregaria* midgut by constructing a midgut reference transcriptome. This will further complement the general knowledge of the insect digestive tract.
- Investigating the transcriptomic changes in the *S. gregaria* midgut during the digestive process by means of RNA-Seq differential expression analysis. This will expand the knowledge of how the insect midgut responds at the transcript level to the presence or absence of food.
- Investigating those genes whose expression changes in response to the presence or absence of food. This will help to identify putative lethal target genes present in the *S. gregaria* midgut.

1.10. References

Aktar W, Sengupta D, Chowdhury A (2009) Impact of pesticides use in agriculture: their benefits and hazards. Interdiscip Toxicol 2:1–12. doi: 10.2478/v10102-009-0001-7

Audsley N, Weaver RJ (2009) Neuropeptides associated with the regulation of feeding in insects. Gen Comp Endocrinol 162:93–104. doi: 10.1016/j.ygcen.2008.08.003

Bass C, Field LM (2011) Gene amplification and insecticide resistance. Pest Manag Sci 67:886–890. doi: 10.1002/ps.2189

Beketov MA, Kefford BJ, Schafer RB, Liess M (2013) Pesticides reduce regional biodiversity of stream invertebrates. Proc Natl Acad Sci 110:11039–11043. doi: 10.1073/pnas.1305618110

Berg JM, Tymoczko JL, Stryver L (2012) Biochemistry, 6th edn W H Freeman, New York, pp 251-252

Bifano TD, Alegria TGP, Terra WR (2010) Transporters involved in glucose and water absorption in the *Dysdercus peruvianus* (Hemiptera: Pyrrhocoridae) anterior midgut. Comp Biochem Physiol Part B Biochem Mol Biol 157:1–9. doi: 10.1016/j.cbpb.2010.05.014

Bolognesi R, Terra WR, Ferreira C (2008) Peritrophic membrane role in enhancing digestive efficiency. Theoretical and experimental models. J Insect Physiol 54:1413–1422. https://doi.org/10.1016/j.jinsphys.2008.08.002

Borovsky D (2003) Biosynthesis and Control of Mosquito Gut Proteases. IUBMB Life 55:435–441. doi: 10.1080/1521654031000159772 1

Bost A, Franzenburg S, Adair KL, Martinson VG, Loeb G, Douglas AE (2018) How gut transcriptional function of Drosophila melanogaster varies with the presence and composition of the gut microbiota. Mol Ecol 27:1848–1859. doi: 10.1111/mec.14413

Boudko DY (2012) Molecular basis of essential amino acid transport from studies of insect nutrient amino acid transporters of the SLC6 family (NAT-SLC6). J Insect Physiol 58:433–449. doi: 10.1016/j.jinsphys.2011.12.018

Boudko DY, Kohn AB, Meleshkevitch EA, Dasher MK, Seron TJ, Stevens BR, Harvey WR (2005) Ancestry and progeny of nutrient amino acid transporters. Proc Natl Acad Sci 102:1360–1365. doi: 10.1073/pnas.0405183101

Bown DP, Wilkinson HS, Gatehouse JA (1997) Differentially regulated inhibitor-sensitive and insensitive protease genes from the phytophagous insect pest, *Helicoverpa armigera*, are members of complex multigene families. Insect Biochem Mol Biol 27:625–638. doi: 10.1016/S0965-1748(97)00043-X

Bren d'Amour C, Reitsma F, Baiocchi G, Barthel S, Güneralp B, Erb K-H, Haberl H, Creutzig F, Seto KC (2017) Future urban land expansion and implications for global croplands. Proc Natl Acad Sci 114:8939–8944. doi: 10.1073/pnas.1606036114

Breugelmans B, Simonet G, van Hoef V, Van Soest S, Smagghe G, Vanden Broeck J (2009) A lepidopteran pacifastin member: cloning, gene structure, recombinant production, transcript profiling and in vitro activity. Insect Biochem Mol Biol 39:430–439. https://doi.org/10.1016/j.ibmb.2009.03.005

Brezeanu PM, Brezeanu C, Ambărus S, Călin M, Cristea TO (2014) a Review of the Most Important Pest Insects and Its Influence on Tomato Culture. Sci Stud Res Ser Biol / Stud si Cercet Stiint Ser Biol 23:68–73

Broderick NA, Raffa KF, Handelsman J (2006) Midgut bacteria required for *Bacillus thuringiensis* insecticidal activity. Proc Natl Acad Sci 103:15196–15199. doi: 10.1073/pnas.0604865103

Bröer S (2006) The SLC6 orphans are forming a family of amino acid transporters. Neurochem Int 48:559–567. doi: 10.1016/j.neuint.2005.11.021

Brunelle F, Cloutier C, Michaud D (2004) Colorado potato beetles compensate for tomato cathepsin D inhibitor expressed in transgenic potato. Arch Insect Biochem Physiol 55:103–113. doi: 10.1002/arch.10135

Burand JP, Hunter WB (2013) RNAi: Future in insect management. J Invertebr Pathol 112:S68–S74. doi: 10.1016/j.jip.2012.07.012

Caccia S, Leonardi MG, Casartelli M, Grimaldi A, de Eguileor M, Pennacchio F, Giordana B (2005) Nutrient absorption by *Aphidius ervi* larvae. J Insect Physiol 51:1183–1192. doi: 10.1016/j.jinsphys.2005.06.010

Caccia S, Casartelli M, Grimaldi A, Losa E, de Eguileor M, Pennacchio F, Giordana B (2007)

Unexpected similarity of intestinal sugar absorption by SGLT1 and apical GLUT2 in an insect (*Aphidius ervi*, Hymenoptera) and mammals. Am J Physiol - Regul Integr Comp Physiol 292:R2284–R2291. doi: 10.1152/ajpregu.00847.2006

Caccia S, Grimaldi A, Casartelli M, Falabella P, Eguileor M de, Pennacchio F, Giordana B (2012) Functional analysis of a fatty acid binding protein produced by *Aphidius ervi* teratocytes. J Insect Physiol 58:621–627. doi: 10.1016/j.jinsphys.2011.12.019

Calicioglu O, Flammini A, Bracco S, Bellù L, Sims R (2019) The Future Challenges of Food and Agriculture: An Integrated Analysis of Trends and Solutions. Sustainability 11:222. doi: 10.3390/su11010222

Capelson M, Hetzer MW (2009) The role of nuclear pores in gene regulation, development and disease. EMBO Rep. 2009 Jul;10(7):697-705. doi: 10.1038/embor.2009.147.

Campbell CL, Vandyke KA, Letchworth GJ, Drolet BS, Hanekamp T, Wilson WC (2005) Midgut and salivary gland transcriptomas of the arbovirus vector *Culicoides sonorensis* (Diptera: Ceratopogonidae). Insect Mol Biol 14:121–136. doi: 10.1111/j.1365-2583.2004.00537.x

Castagna M, Shayakul C, Trotti D, Sacchi F V., Harvey WR, Hediger MA (1998) Cloning and characterization of a potassium-coupled amino acid transporter. Proc Natl Acad Sci 95:5395–5400

Castagna M, Bossi E, Sacchi VF (2009) Molecular physiology of the insect K-activated amino acid transporter 1 (KAAT1) and cation-anion activated amino acid transporter/channel 1 (CAATCH1) in the light of the structure of the homologous protein LeuT. Insect Mol Biol 18:265–279. doi: 10.1111/j.1365-2583.2009.00881.x

Chakrabarti AC, Deamer DW (1994) Permeation of membranes by the neutral form of amino acids and peptides: relevance to the origin of peptide translocation. J Mol Evol 39(1):1–5

Chandler D, Bailey AS, Tatchell GM, Davidson G, Greaves J, Grant WP (2011) The development, regulation and use of biopesticides for integrated pest management. Philos Trans R Soc B Biol Sci 366:1987–1998. doi: 10.1098/rstb.2010.0390

Chapman RF (2013) The insects: structure and function. Cambridge University Press, New York

Chi YH, Salzman RA, Balfe S, Ahn JE, Sun W, Moon J, Yun DJ, Lee SY, Higgins TJV, Pittendrigh B (2009) Cowpea bruchid midgut transcriptome response to a soybean cystatin - Costs and benefits of counter-defence. Insect Mol Biol 18:97–110. doi: 10.1111/j.1365-2583.2008.00854.x

Ciufo LF, Murray PA, Thompson A, Rigden DJ, Rees HH (2011) Characterisation of a Desmosterol Reductase Involved in Phytosterol Dealkylation in the Silkworm, *Bombyx mori*. PLoS One 6:e21316. doi: 10.1371/journal.pone.0021316

Chng W bin A, Sleiman MSB, Schüpfer F, Lemaitre B (2014) Transforming growth factor β/activin signaling functions as a sugar-sensing feedback loop to regulate digestive enzyme expression. Cell Rep 9:336–348. doi: 10.1016/j.celrep.2014.08.064

Christeller JT, Laing WA, Markwick NP, Burgess EPJ (1992) Midgut protease activities in 12 phytophagous lepidopteran larvae: Dietary and protease inhibitor interactions. Insect Biochem Mol Biol 22:735–746. doi: 10.1016/0965-1748(92)90052-G

Christeller JT, Amara S, Carrière F (2011) Galactolipase , phospholipase and triacylglycerol lipase activities in the midgut of six species of lepidopteran larvae feeding on different lipid diets. 57:1232–1239. doi: 10.1016/j.jinsphys.2011.05.012

Cognigni P, Bailey AP, Miguel-Aliaga I (2011) Enteric Neurons and Systemic Signals Couple Nutritional and Reproductive Status with Intestinal Homeostasis. Cell Metab 13:92–104. doi: 10.1016/j.cmet.2010.12.010

Colombani J, Raisin S, Pantalacci S, Radimerski T, Montagne J, Léopold P (2003) A Nutrient Sensor Mechanism Controls *Drosophila* Growth. Cell 114:739–749. doi: 10.1016/S0092-8674(03)00713-X

Copenhaver PF (2007) How to innervate a simple gut: Familiar themes and unique aspects in the formation of the insect enteric nervous system. Dev Dyn 236:1841–1864. doi: 10.1002/dvdy.21138

Cui F, Li M-X, Chang H-J, Mao Y, Zhang H-Y, Lu L-X, Yan S-G, Lang M-L, Liu L, Qiao C-L (2015) Carboxylesterase-mediated insecticide resistance: Quantitative increase induces broader metabolic resistance than qualitative change. Pestic Biochem Physiol 121:88–96. doi: 10.1016/j.pestbp.2014.12.016

Cumming JM, Sinclair BJ, Triplehorn CA, Aldryhim Y, Galante E, Marcos-Garcia MA, Edmunds M, Edmunds M, Lounibos LP, Frank JH, et al (2008) Desert Locust, *Schistocerca gregaria* Forskål (Orthoptera: Acrididae) Plagues. In: Encyclopedia of Entomology. Springer Netherlands,

Dordrecht, pp 1181–1186
Da Lage J-L (2018) The Amylases of Insects. Int J Insect Sci 10:117954331880478. doi: 10.1177/1179543318804783
Da Lage JL, Maczkowiak F, Cariou ML (2000) Molecular characterization and evolution of the amylase multigene family of *Drosophila ananassae*. J Mol Evol 51(4):391–403
Damalas CA, Koutroubas SD (2018) Current Status and Recent Developments in Biopesticide Use. Agriculture 8:13. doi: 10.3390/agriculture8010013
Da Silva ML, Melo ACA, Araujo H, Logullo R, Koerich L, Mesquita RD, Balczun C, Genta FA, Silber AM, Salmon D, et al (2014) An Insight into the Transcriptome of the Digestive Tract of the Bloodsucking Bug, *Rhodnius prolixus*. PLoS Negl Trop Dis 8:e2594. doi: 10.1371/journal.pntd.0002594
Desneux N, Decourtye A, Delpuech J-M (2007) The Sublethal Effects of Pesticides on Beneficial Arthropods. Annu Rev Entomol 52:81–106. doi: 10.1146/annurev.ento.52.110405.091440
Deutsch CA, Naylor RL, Battisti DS, Tewksbury JJ, Tigchelaar M, Merrill SC, Huey RB (2018) Increase in crop losses to insect pests in a warming climate. Science 361:916–919. doi: 10.1126/science.aat3466
Dhania NK, Chauhan VK, Chaitanya RK, Dutta-Gupta A (2019) Midgut de novo transcriptome analysis and gene expression profiling of *Achaea janata* larvae exposed with *Bacillus thuringiensis* (Bt)-based biopesticide formulation. Comp Biochem Physiol Part D Genomics Proteomics 30:81–90. doi: 10.1016/j.cbd.2019.02.005
Dinglasan RR, Devenport M, Florens L, Johnson JR, McHugh CA, Donnelly-Doman M, Carucci DJ, Yates JR, Jacobs-Lorena M (2009) The *Anopheles gambiae* adult midgut peritrophic matrix proteome. Insect Biochem Mol Biol 39:125–134. https://doi.org/ 10.1016/j.ibmb.2008.10.010
Dixit SS, Sleat DE, Stock AM, Lobel P (2007) Do mammalian NPC1 and NPC2 play a role in intestinal cholesterol absorption? Biochem J 408:1–5. doi: 10.1042/BJ20071167
Dostálová A, Votýpka J, Favreau AJ, Barbian KD, Volf P, Valenzuela JG, Jochim RC (2011) The midgut transcriptome of *Phlebotomus (Larroussius) perniciosus*, a vector of *Leishmania infantum*: Comparison of sugar fed and blood fed sand flies. BMC Genomics 12:223. doi: 10.1186/1471-2164-12-223
Douglas AE (2006) Phloem-sap feeding by animals: problems and solutions. J Exp Bot 57:747–754. doi: 10.1093/jxb/erj067
Douglas AE (2015) Multiorganismal Insects: Diversity and Function of Resident Microorganisms. Annu Rev Entomol 60:17–34. doi: doi.org/10.1146/annurev-ento-010814-020822
Dubreuil RR, Frankel J, Wang P, Howrylak J, Kappil M, Grushko TA (1998) Mutations of α spectrin and labial block cuprophilic cell differentiation and acid secretion in the middle midgut of *Drosophila* larvae. Dev Biol 194:1–11. doi: 10.1006/dbio.1997.8821
Erb K-H, Lauk C, Kastner T, Mayer A, Theurl MC, Haberl H (2016) Exploring the biophysical option space for feeding the world without deforestation. Nat Commun 7:11382. doi: 10.1038/ncomms11382
Escher SA, Rasmuson-Lestander Å (2004) The *Drosophila* Glucose Transporter Gene: cDNA Sequence, Phylogenetic Comparisons, Analysis of Functional Sites and Secondary Structures. Hereditas 130:95–103. doi: 10.1111/j.1601-5223.1999.00095.x
Espinoza-Fuentes FP, Terra WR (1987) Physiological adaptations for digesting bacteria. Water fluxes and distribution of digestive enzymes in *Musca domestica* larval midgut. Insect Biochem 17:809–817. doi: 10.1016/0020-1790(87)90015-1
Esteves A, Ehrlich R (2006) Invertebrate intracellular fatty acid binding proteins. Comp Biochem Physiol - C Toxicol Pharmacol 142:262–274. doi: 10.1016/j.cbpc.2005.11.006
Fabrick JA, Flexner JL, Nelson ME, Wu G, Tabashnik BE, Mathis J, Zastrow-Hayes G, Carrière Y, Tassone EE (2016) Sequencing, de novo assembly and annotation of a pink bollworm larval midgut transcriptome. Gigascience 5:1–5. doi: 10.1186/s13742-016-0130-9
Feldman DH, Harvey WR, Stevens BR (2000) A Novel Electrogenic Amino Acid Transporter Is Activated by K + or Na + , Is Alkaline pH-dependent, and Is Cl − -independent. J Biol Chem 275:24518–24526. doi: 10.1074/jbc.M907582199
Flagel LE, Bansal R, Kerstetter RA, Chen M, Carroll M, Flannagan R, Clark T, Goldman BS, Michel AP (2014) Western corn rootworm (*Diabrotica virgifera virgifera*) transcriptome assembly and genomic analysis of population structure. BMC Genomics 15:195. doi: 10.1186/1471-2164-15-195
Flower NE, Filshie BK (1976) Goblet cell membrane differentiations in the midgut of a lepidopteran larva.

J Cell Sci 20:357–375
Franco OL, Rigden DJ, Melo FR, Grossi-de-Sá MF (2002) Plant α-amylase inhibitors and their interaction with insect α-amylases. Eur J Biochem 269:397–412. doi: 10.1046/j.0014-2956.2001.02656.x
Fruttero LL, Leyria J, Canavoso LE (2017) Lipids in insect oocytes: from the storage pathways to their multiple functions. Results Probl Cell Differ 63:403–434. https://doi.org/10.1007/978-3-319-60855-6_18
Fu K-Y, Guo W-C, Ahmat T, Li G-Q (2015) Knockdown of a nutrient amino acid transporter gene LdNAT1 reduces free neutral amino acid contents and impairs *Leptinotarsa decemlineata* pupation. Sci Rep 5:18124. doi: 10.1038/srep18124
Gäde G, Goldsworthy GJ (2003) Insect peptide hormones: A selective review of their physiology and potential application for pest control. Pest Manag Sci 59:1063–1075. doi: 10.1002/ps.755
Gazara RK, Cardoso C, Bellieny-Rabelo D, Ferreira C, Terra WR, Venancio TM (2017) De novo transcriptome sequencing and comparative analysis of midgut tissues of four non-model insects pertaining to Hemiptera, Coleoptera, Diptera and Lepidoptera. Gene 627:85–93. doi: 10.1016/j.gene.2017.06.008
Genchi G (2017) An overview on d-amino acids. Amino Acids 49:1521–1533. doi: 10.1007/s00726-017-2459-5
Gondim KC, Atella GC, Pontes EG, Majerowicz D (2018) Lipid metabolism in insect disease vectors. Insect Biochem Mol Biol 101:108–123. doi: 10.1016/j.ibmb.2018.08.005
Gong J, Hou Y, Zha X-F, Lu C, Zhu Y, Xia Q-Y (2006) Molecular cloning and characterization of *Bombyx mori* sterol carrier protein x/sterol carrier protein 2 (SCPx/SCP2) gene. DNA Seq 17:326–333. doi: 10.1080/10425170600886706
Gould F, Brown ZS, Kuzma J (2018) Wicked evolution: Can we address the sociobiological dilemma of pesticide resistance? Science (80-) 360:728–732. doi: 10.1126/science.aar3780
Govindaraj L, Gupta T, Esvaran VG, Awasthi AK, Ponnuvel KM (2016) Genome-wide identification, characterization of sugar transporter genes in the silkworm *Bombyx mori* and role in *Bombyx mori* nucleopolyhedrovirus (BmNPV) infection. Gene 579:162–171. doi: 10.1016/j.gene.2015.12.057
Greenberg B, Paretsky D (1955) Proteolytic enzymes in the house fly, *Musca domestica* (L.). Ann Entomol SocAm48:46–50. https://doi. org/10.1093/aesa/48.1-2.46
Griss C, Simpson SJ, Rohrbacher J, Rowell CHF (1991) Localization in the central nervous system of larval *Manduca sexta* (Lepidoptera: Sphingidae) of areas responsible for aspects of feeding behaviour. J Insect Physiol 37:477–482. doi: 10.1016/0022-1910(91)90023-S
Grossman GL, Campos Y, Severson DW, James AA (1997) Evidence for two distinct members of the amylase gene family in the yellow fever mosquito, *Aedes aegypti*. Insect Biochem Mol Biol 27:769–781. doi: 10.1016/S0965-1748(97)00063-5
Guo X-R, Zheng S-C, Liu L, Feng Q-L (2009) The sterol carrier protein 2/3-oxoacyl-CoA thiolase (SCPx) is involved in cholesterol uptake in the midgut of *Spodoptera litura*: gene cloning, expression, localization and functional analyses. BMC Mol Biol 10:102. doi: 10.1186/1471-2199-10-102
Gutiérrez-Cabrera AE, Córdoba-Aguilar A, Zenteno E, Lowenberger C, Espinoza B (2016) Origin, evolution and function of the hemipteran perimicrovillar membrane with emphasis on Reduviidae that transmit Chagas disease. Bull Entomol Res 106:279–291. doi: 10.1017/S0007485315000929
Guzik J, Nakonieczny M, Tarnawska M, Bereś PK, Drzewiecki S, Migula P (2015) Glycolytic enzymes activity in the midgut of *Diabrotica virgifera virgifera* (Coleoptera: Chrysomelidae) adult and their seasonal changes. J Insect Sci 15:1–8. doi: 10.1093/jisesa/iev036
Hansen IA, Boudko DY, Shiao S-H, Voronov DA, Meleshkevitch EA, Drake LL, Aguirre SE, Fox JM, Attardo GM, Raikhel AS (2011) AaCAT1 of the Yellow Fever Mosquito, *Aedes aegypti*. J Biol Chem 286:10803–10813. doi: 10.1074/jbc.M110.179739
Hartenstein V (1997) Development of the insect stomatogastric nervous system. Trends Neurosci 20:421–427. doi: http://doi.org/10.1016/S0166-2236(97)01066-7
Hawkins NJ, Bass C, Dixon A, Neve P (2019) The evolutionary origins of pesticide resistance. Biol Rev 94:135–155. doi: 10.1111/brv.12440
Hediger MA, Romero MF, Peng J-B, Rolfs A, Takanaga H, Bruford EA (2004) The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. Pflugers Arch Eur J Physiol 447:465–468. doi: 10.1007/s00424-003-1192-y
Herde M, Howe GA (2014) Host plant-specific remodeling of midgut physiology in the generalist insect

herbivore *Trichoplusia ni*. Insect Biochem Mol Biol 50:58–67. doi: 10.1016/j.ibmb.2014.03.013
Hofte H, Whiteley HR (1989) Insecticidal Crystal Proteins of *Bacillus thuringiensis*. Microbiol Rev 53:242–255
Hokkanen HMT (2015) Integrated pest management at the crossroads: Science, politics, or business (as usual)? Arthropod Plant Interact 9:543–545. doi: 10.1007/s11829-015-9403-y
Horne I, Haritos VS, Oakeshott JG (2009) Comparative and functional genomics of lipases in holometabolous insects. Insect Biochem Mol Biol 39:547–567. doi: 10.1016/j.ibmb.2009.06.002
Horst DJ Van Der (2003) Insect adipokinetic hormones : release and integration of flight energy metabolism. 136:217–226. doi: 10.1016/S1096-4959(03)00151-9
Huang J-H, Jing X, Douglas AE (2015) The Multi-Tasking Gut Epithelium of Insects. Insect Biochem Mol Biol 67:15–20. doi: 10.1016/j.ibmb.2015.05.004
Huang X, McNeill MR, Ma J, Qin X, Tu X, Cao G, Wang G, Nong X, Zhang Z (2017) Gut Transcriptome Analysis Shows Different Food Utilization Efficiency by the Grasshopper *Oedaleous asiaticus* (Orthoptera: Acrididae). J Econ Entomol 110:1831–1840. doi: 10.1093/jee/tox128
Huang Z, Zhou D, Gao G, Zheng S, Feng Q, Liu L (2012) Cloning and characterization of midgut-specific fatty acid binding protein in *Spodoptera litura*. Arch Insect Biochem Physiol 79:1–17. doi: 10.1002/arch.21001
Hum-Musser S, Musser R, Popham H, Vogel H, Breitenbach J, Noland J (2013) Gut Transcription in *Helicoverpa zea* is Dynamically Altered in Response to Baculovirus Infection. Insects 4:506–520. doi: 10.3390/insects4030506
Husain Q (2010) β Galactosidases and their potential applications: A review. Crit Rev Biotechnol 30:41–62. doi: 10.3109/07388550903330497
Jing X, Vogel H, Grebenok RJ, Zhu-Salzman K, Behmer ST (2012) Dietary sterols/steroids and the generalist caterpillar *Helicoverpa zea*: Physiology, biochemistry and midgut gene expression. Insect Biochem Mol Biol 42:835–845. doi: 10.1016/j.ibmb.2012.07.009
Jiang J, Huang Y, Shu C, Soberón M, Bravo A, Liu C, Song F, Lai J, Zhang J (2017) *Holotrichia oblita* Midgut Proteins That Bind to *Bacillus thuringiensis* Cry8-Like Toxin and Assembly of the *H. oblita* Midgut Tissue Transcriptome. Appl Environ Microbiol 83:1–11. doi: 10.1128/AEM.00541-17
Jongsma MA (1997) The Adaptation of Insects to Plant Protease Inhibitors. 43:885–895
Jouzani GS, Valijanian E, Sharafi R (2017) *Bacillus thuringiensis*: a successful insecticide with new environmental features and tidings. Appl Microbiol Biotechnol 101:2691–2711. doi: 10.1007/s00253-017-8175-y
Juhn J, Naeem-Ullah U, Maciel Guedes B, Majid A, Coleman J, Paolucci Pimenta P, Akram W, James A, Marinotti O (2011) Spatial mapping of gene expression in the salivary glands of the dengue vector mosquito, *Aedes aegypti*. Parasit Vectors 4:1. doi: 10.1186/1756-3305-4-1
Kanamori Y, Saito A, Hagiwara-Komoda Y, Tanaka D, Mitsumasu K, Kikuta S, Watanabe M, Cornette R, Kikawada T, Okuda T (2010) The trehalose transporter 1 gene sequence is conserved in insects and encodes proteins with different kinetic properties involved in trehalose import into peripheral tissues. Insect Biochem Mol Biol 40:30–37. doi: 10.1016/j.ibmb.2009.12.006
Kellett GL, Brot-Laroche E (2005) Apical GLUT2: A major pathway of intestinal sugar absorption. Diabetes 54:3056–3062. doi: 10.2337/diabetes.54.10.3056
Keeling CI, Li M, Sandhu HK, Henderson H, Yuen MM Saint, Bohlmann J (2016) Quantitative metabolome, proteome and transcriptome analysis of midgut and fat body tissues in the mountain pine beetle, Dendroctonus ponderosae Hopkins, and insights into pheromone biosynthesis. Insect Biochem Mol Biol 70:170–183. doi: 10.1016/j.ibmb.2016.01.002
Kikawada T, Saito A, Kanamori Y, Nakahara Y, Iwata K -i., Tanaka D, Watanabe M, Okuda T (2007) Trehalose transporter 1, a facilitated and high-capacity trehalose transporter, allows exogenous trehalose uptake into cells. Proc Natl Acad Sci 104:11585–11590. doi: 10.1073/pnas.0702538104
Kikuta S, Nakamura Y, Hattori M, Sato R, Kikawada T, Noda H (2015) Herbivory-induced glucose transporter gene expression in the brown planthopper, *Nilaparvata lugens*. Insect Biochem Mol Biol 64:60–67. doi: 10.1016/j.ibmb.2015.07.015
Kirsch R, Wielsch N, Vogel H, Svatoš A, Heckel DG, Pauchet Y (2012) Combining proteomics and transcriptome sequencing to identify active plant-cell-wall-degrading enzymes in a leaf beetle. BMC Genomics 13:587. doi: 10.1186/1471-2164-13-587
Kitamura T, Kobayashi S, Okada M (1996) Regional expression of the transcript encoding sterol carrier protein x-related thiolase and its regulation by homeotic genes in the midgut of *Drosophila*

embryos. Dev Growth Differ 38:373–381. doi: 10.1046/j.1440-169X.1996.t01-3-00005.x
Knipling EF (1972) Entomology and the management of man's environment. Aust J Entomol 11:153–167. doi: 10.1111/j.1440-6055.1972.tb01618.x
Kotkar HM, Bhide AJ, Gupta VS, Giri AP (2012) Amylase gene expression patterns in *Helicoverpa armigera* upon feeding on a range of host plants. Gene 501:1–7. doi: 10.1016/j.gene.2012.04.010
Krebs KC, Lan Q (2003) Isolation and expression of a sterol carrier protein-2 gene from the yellow fever mosquito, *Aedes aegypti*. Insect Mol Biol 12:51–60. doi: 10.1046/j.1365-2583.2003.00386.x
Kumar S, Singh A (2015) Biopesticides : Present Status and the Future Prospects. J Fertil Pestic 6:. doi: 10.4172/2471-2728.1000e129
Kunieda T, Fujiyuki T, Kucharski R, Foret S, Ament SA, Toth AL, Ohashi K, Takeuchi H, Kamikouchi A, Kage E, et al (2006) Carbohydrate metabolism genes and pathways in insects: Insights from the honey bee genome. Insect Mol Biol 15:563–576. doi: 10.1111/j.1365-2583.2006.00677.x
Lam W, Coast GM, C. Rayne R (2000) Characterisation of multiple trypsins from the midgut of *Locusta migratoria*. Insect Biochem Mol Biol 30:85–94. doi: 10.1016/S0965-1748(99)00103-4
Lei Y, Zhu X, Xie W, Wu Q, Wang S, Guo Z, Xu B, Li X, Zhou X, Zhang Y (2014) Midgut transcriptome response to a Cry toxin in the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). Gene 533:180–187. doi: 10.1016/j.gene.2013.09.091
Lehane MJ (1997) Peritrophic Matrix Structure and Function. Annu Rev Entomol 42:525–550. doi: 10.1146/annurev.ento.42.1.525
Lemaitre B, Miguel-Aliaga I (2013) The Digestive Tract of *Drosophila melanogaster*. Annu Rev Genet 47:377–404. doi: 10.1146/annurev-genet-111212-133343
Li X, Shi L, Zhou Y, Xie H, Dai X, Li R, Chen Y, Wang H (2017) Molecular evolutionary mechanisms driving functional diversification of α-glucosidase in Lepidoptera. Sci Rep 7:1–13. doi: 10.1038/srep45787
Li L, Lan M, Lu W, Li Z, Xia T, Zhu J, Ye M, Gao X, Wu G (2018) De novo transcriptomic analysis of the alimentary tract of the tephritid gall fly, *Procecidochares utilis*. PLoS One 13:1–28. doi: 10.1371/journal.pone.0201679
Liu Y, Qi M, Chi Y, Wuriyanghan H (2016) De novo assembly of the transcriptome for oriental armyworm *Mythimna separata* (Lepidoptera: Noctuidae) and analysis on insecticide resistance-related genes. J Insect Sci 16:0–7. doi: 10.1093/jisesa/iew079
Linser PJ, Dinglasan RR (2014) Insect Gut Structure, Function, Development and Target of Biological Toxins, 1st edn. Elsevier Ltd.
Lomate PR, Dewangan V, Mahajan NS, Kumar Y, Kulkarni A, Wang L, Saxena S, Gupta VS, Giri AP (2018) Integrated transcriptomic and proteomic analyses suggest the participation of endogenous protease inhibitors in the regulation of protease gene expression in . Molecular & Cellular Proteomics 17 (7):1324-1336
Lopes AR, Juliano MA, Juliano L, Terra WR (2004) Coevolution of insect trypsins and inhibitors. Arch Insect Biochem Physiol 55:140–152. doi: 10.1002/arch.10134
Ma W, Zhang Z, Peng C, Wang X, Li F, Lin Y (2012) Exploring the midgut transcriptome and brush border membrane vesicle proteome of the rice stem borer, *Chilo suppressalis* (walker). PLoS One 7:1–7. doi: 10.1371/journal.pone.0038151
Maddrell SHP, O'Donnell MJ (1992) Insect Malpighian tubules: V-ATPase action in ion and fluid transport. J Exp Biol 172:417–429. doi: 10.1016/bs.apha.2014.10.006
Majerowicz D, Gondim KC (2013) Insect Lipid Metabolism: Insights into Gene Expression Regulation. In book: Recent Trends in Gene Expression, Edition 1, Chapter 7
Martin JF, Hersperger E, Simcox A, Shearn A (2000) minidiscs encodes a putative amino acid transporter subunit required non-autonomously for imaginal cell proliferation. Mech Dev 92:155–167. doi: 10.1016/S0925-4773(99)00338-X
Meleshkevitch EA, Voronov DA, Miller MM, Penneda M, Fox JM, Metzler R, Boudko DY (2013) A novel eukaryotic Na+ methionine selective symporter is essential for mosquito development. Insect Biochem Mol Biol 43:755–767. doi: 10.1016/j.ibmb.2013.05.008
Mello MO, Silva-Filho MC (2002) Plant-insect interactions: An evolutionary arms race between two distinct defense mechanisms. Brazilian J Plant Physiol 14:71–81. doi: 10.1590/S1677-04202002000200001
Meyer H, Vitavska O, Wieczorek H (2011) Identification of an animal sucrose transporter. J Cell Sci 124:1984–1991. doi: 10.1242/jcs.082024

Miguel-Aliaga I, Jasper H, Lemaitre B (2018) Anatomy and Physiology of the Digestive Tract of *Drosophila melanogaster*. Genetics 210:357–396. doi: 10.1534/genetics.118.300224

Miller MM, Popova LB, Meleshkevitch EA, Tran P V., Boudko DY (2008) The invertebrate B0 system transporter, *D. melanogaster* NAT1, has unique d-amino acid affinity and mediates gut and brain functions. Insect Biochem Mol Biol 38:923–931. doi: 10.1016/j.ibmb.2008.07.005

Mineau P, Whiteside M (2013) Pesticide Acute Toxicity Is a Better Correlate of U.S. Grassland Bird Declines than Agricultural Intensification. PLoS One 8:e57457. doi: 10.1371/journal.pone.0057457

Nicolopoulou-Stamati P, Maipas S, Kotampasi C, Stamatis P, Hens L (2016) Chemical Pesticides and Human Health: The Urgent Need for a New Concept in Agriculture. Front Public Heal 4:148. doi: 10.3389/fpubh.2016.00148

Niwa R, Niwa YS (2014) Enzymes for ecdysteroid biosynthesis: their biological functions in insects and beyond. Biosci Biotechnol Biochem 78:1283–1292. doi: 10.1080/09168451.2014.942250

Oerke E-C (2006) Crop losses to pests. J Agric Sci 144:31–43. doi: 10.1017/S0021859605005708

Ohba T, Holt JA, Billheimer JT, Strauss JF 3rd (1995) Human sterol carrier protein x/sterol carrier protein 2 gene has two promoters. Biochemistry. 34:10660–10668

Ohnishi A, Hashimoto K, Imai K, Matsumoto S (2009) Functional characterization of the *Bombyx mori* fatty acid transport protein (BmFATP) within the silkmoth pheromone gland. J Biol Chem 284:5128–5136. doi: 10.1074/jbc.M806072200

Okech BA, Meleshkevitch EA, Miller MM, Popova LB, Harvey WR, Boudko DY (2008) Synergy and specificity of two Na+-aromatic amino acid symporters in the model alimentary canal of mosquito larvae. J Exp Biol 211:1594–1602. doi: 10.1242/jeb.017244

Oliveira F, Laughinghouse A, Teixeira CR, Gomes RB, Mu J, Valenzuela JG, Elnaiem D-E, Jochim RC (2008) The midgut transcriptome of Lutzomyia longipalpis: comparative analysis of cDNA libraries from sugar-fed, blood-fed, post-digested and Leishmania infantum chagasi-infected sand flies. BMC Genomics 9:15. doi: 10.1186/1471-2164-9-15

Olson S (2015) An Analysis of the Biopesticide Market Now and Where it is Going. Outlooks Pest Manag 26:203–206. doi: 10.1564/v26_oct_04

Oppert B, Morgan TD, Hartzer K, Kramer KJ (2005) Compensatory proteolytic responses to dietary proteinase inhibitors in the red flour beetle, *Tribolium castaneum* (Coleoptera: Tenebrionidae). Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology 140 (1):53-58

Oppert B, Perkin L, Martynov AG, Elpidina EN (2018) Cross-species comparison of the gut: Differential gene expression sheds light on biological differences in closely related tenebrionids. J Insect Physiol 106:114–124. doi: 10.1016/j.jinsphys.2017.03.010

Oppert C, Oppert B, Morris K, Lorenzen MD, Jurat-Fuentes JL, Vinokurov K, Elpidina EN, Hiromasa Y, Fabrick J, Tomich JM (2009) *Tribolium castaneum* Larval Gut Transcriptome and Proteome: A Resource for the Study of the Coleopteran Gut. J Proteome Res 8:3889–3898. doi: 10.1021/pr900168z

Overend G, Luo Y, Henderson L, Douglas AE, Davies SA, Dow JAT (2016) Molecular mechanism and functional significance of acid generation in the Drosophila midgut. Sci Rep 6:1–11. doi: 10.1038/srep27242

Padilha MHP, Pimentel AC, Ribeiro AF, Terra WR (2009) Sequence and function of lysosomal and digestive cathepsin D-like proteinases of *Musca domestica* midgut. Insect Biochem Mol Biol 39:782–791. doi: 10.1016/j.ibmb.2009.09.003

Palma L, Muñoz D, Berry C, Murillo J, Caballero P (2014) *Bacillus thuringiensis* Toxins: An Overview of Their Biocidal Activity. Toxins (Basel) 6:3296–3325. doi: 10.3390/toxins6123296

Patankar AG, Giri AP, Harsulkar AM, Sainani MN, Deshpande V V, Ranjekar PK, Gupta VS (2001) Complexity in specificities and expression of *Helicoverpa armigera*. Insect Biochem Mol Biol 31:453–464

Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, Ffrench-Constant RH (2009) Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera. Insect Biochem Mol Biol 39:403–413. doi: 10.1016/j.ibmb.2009.04.001

Pedra JHF, Brandt A, Westerman R, Lobo N, Li HM, Romero-Severson J, Murdock LL, Pittendrigh BR (2003) Transcriptome analysis of the cowpea weevil bruchid: Identification of putative proteinases and α-amylases associated with food breakdown. Insect Mol Biol 12:405–412. doi: 10.1046/j.1365-2583.2003.00425.x

Pener MP, Simpson SJ (2009) Locust Phase Polyphenism: An Update. In: Advances in Insect Physiology. Elsevier, pp 1–272

Petek M, Turnšek N, Gašparič MB, Novak MP, Gruden K, Slapar N, Popovič T, Štrukelj B, Gruden K, Štrukelj B, et al (2012) A complex of genes involved in adaptation of *leptinotarsa decemlineat*a larvae to induced potato defense. Arch Insect Biochem Physiol 79:153–181. doi: 10.1002/arch.21017

Pimentel AC, Fuzita FJ, Palmisano G, Ferreira C, Terra WR (2017) Role of cathepsins D in the midgut of *Dysdercus peruvianus*. Comp Biochem Physiol Part - B Biochem Mol Biol 204:45–52. doi: 10.1016/j.cbpb.2016.11.004

Pitaluga AN, Beteille V, Lobo AR, Ortigão-Farias JR, Dávila AMR, Souza AA, Ramalho-Ortigão JM, Traub-Cseko YM (2009) EST sequencing of blood-fed and Leishmania-infected midgut of *Lutzomyia longipalpis*, the principal visceral leishmaniasis vector in the Americas. Mol Genet Genomics 282:307–317. doi: 10.1007/s00438-009-0466-2

Phillips McDougall (2018) Evolution of the Crop Protection Industry since 1960. Informa, November 20, 2018

Popp J, Pető K, Nagy J (2013) Pesticide productivity and food security. A review. Agron Sustain Dev 33:243–255. doi: 10.1007/s13593-012-0105-x

Price DR, Gatehouse JA (2014) Genome-wide annotation and functional identification of aphid GLUT-like sugar transporters. BMC Genomics 15:647. doi: 10.1186/1471-2164-15-647

Price DRG, Gatehouse JA (2008) RNAi-mediated crop protection against insects. Trends Biotechnol 26:393–400. doi: 10.1016/j.tibtech.2008.04.004

Price DRG, Tibbles K, Shigenobu S, Smertenko A, Russell CW, Douglas AE, Fitches E, Gatehouse AMR, Gatehous JA (2010) Sugar transporters of the major facilitator superfamily in aphids; From gene prediction to functional characterization. Insect Mol Biol 19:97–112. doi: 10.1111/j.1365-2583.2009.00918.x

Price DRG, Wilkinson HS, Gatehouse JA (2007) Functional expression and characterisation of a gut facilitative glucose transporter, NlHT1, from the phloem-feeding insect *Nilaparvata lugens* (rice brown planthopper). Insect Biochem Mol Biol 37:1138–1148. doi: 10.1016/j.ibmb.2007.07.001

Popham HJR, Gould F, Adang MJ, Jurat-Fuentes JL, Perera OP, Shelby KS (2015) Generation of a Transcriptome in a Model Lepidopteran Pest, Heliothis virescens, Using Multiple Sequencing Strategies for Profiling Midgut Gene Expression. PLoS One 10:e0128563. doi: 10.1371/journal.pone.0128563

Qian H, Li G, He Q, Zhang H, Xu A (2016) Analysis of differentially expressed genes between fluoride-sensitive and fluoride-endurable individuals in midgut of silkworm, *Bombyx mori*. Gene 588:47–53. doi: 10.1016/j.gene.2016.04.033

Ramalho-Ortigão M, Jochim RC, Anderson JM, Lawyer PG, Pham V-M, Kamhawi S, Valenzuela JG (2007) Exploring the midgut transcriptome of *Phlebotomus papatasi*: comparative analysis of expression profiles of sugar-fed, blood-fed and Leishmania major-infected sandflies. BMC Genomics 8:300. doi: 10.1186/1471-2164-8-300

Rao MB, Tanksale AM, Ghatge MS, Deshpande V V. (1998) Molecular and Biotechnological Aspects of Microbial Proteases. Microbiol Mol Biol Rev 62:597–635

Reynolds B, Roversi P, Laynes R, Kazi S, Boyd CAR, Goberdhan DCI (2009) *Drosophila* expresses a CD98 transporter with an evolutionarily conserved structure and amino acid-transport properties. Biochem J 420:363–372. doi: 10.1042/BJ20082198

Roman G, Meller V, Wu KH, Davis RL (1998) The opt1 gene of *Drosophila melanogaster* encodes a proton-dependent dipeptide transporter. Am J Phys 275(3):857–869

Sakudoh T, Iizuka T, Narukawa J, Sezutsu H, Kobayashi I, Kuwazaki S, Banno Y, Kitamura A, Sugiyama H, Takada N, et al (2010) A CD36-related Transmembrane Protein Is Coordinated with an Intracellular Lipid-binding Protein in Selective Carotenoid Transport for Cocoon Coloration. J Biol Chem 285:7739–7751. doi: 10.1074/jbc.M109.074435

Sanahuja G, Banakar R, Twyman RM, Capell T, Christou P (2011) *Bacillus thuringiensis*: a century of research, development and commercial applications. Plant Biotechnol J 9:283–300. doi: 10.1111/j.1467-7652.2011.00595.x

Sánchez-Bayo F, Wyckhuys KAG (2019) Worldwide decline of the entomofauna: A review of its drivers. Biol Conserv 232:8–27. doi: 10.1016/j.biocon.2019.01.020

Sanders HR, Evans AM, Ross LS, Gill SS (2003) Blood meal induces global changes in midgut gene

expression in the disease vector, *Aedes aegypti*. Insect Biochem Mol Biol 33:1105–1122. doi: 10.1016/S0965-1748(03)00124-3
Schafer M, Fischer C, Meldau S, Seebald E, Oelmuller R, Baldwin IT (2011) Lipase Activity in Insect Oral Secretions Mediates Defense Responses in *Arabidopsis*. PLANT Physiol 156:1520–1534. doi: 10.1104/pp.111.173567
Scheepers A, Joost H, Schurmann A (2004) The glucose transporter families SGLT and GLUT: molecular basis of normal and aberrant function. J Parenter Enter Nutr 28:364–371. doi: 10.1177/0148607104028005364
Schlessinger A, Yee SW, Sali A, Giacomini KM (2013) SLC Classification: An Update. Clin Pharmacol Ther 94:19–23. doi: 10.1038/clpt.2013.73
Scully ED, Hoover K, Carlson JE, Tien M, Geib SM (2013) Midgut transcriptome profiling of Anoplophora glabripennis, a lignocellulose degrading cerambycid beetle. BMC Genomics 14:. doi: 10.1186/1471-2164-14-850
Shao R-X, Geng L, Liu S-S, Wang X-W, Liu Y-Q, Qian L-X (2018) Transcriptome profiling of whitefly guts in response to Tomato yellow leaf curl virus infection. Virol J 15:1–12. doi: 10.1186/s12985-018-0926-6
Shu C, Tan S, Yin J, Soberón M, Bravo A, Liu C, Geng L, Song F, Li K, Zhang J (2015) Assembling of Holotrichia parallela (dark black chafer) midgut tissue transcriptome and identification of midgut proteins that bind to Cry8Ea toxin from *Bacillus thuringiensis.* Appl Microbiol Biotechnol 99:7209–7218. doi: 10.1007/s00253-015-6755-2
Shukle RH, Mittapalli O, Morton PK, Chen M-S (2009) Characterization and expression analysis of a gene encoding a secreted lipase-like protein expressed in the salivary glands of the larval Hessian fly, *Mayetiola destructor* (Say). J Insect Physiol 55:105–112. doi: 10.1016/j.jinsphys.2008.10.008
Silva CP, Silva JR, Vasconcelos FF, Petretski MDA, DaMatta RA, Ribeiro AF, Terra WR (2004) Occurrence of midgut perimicrovillar membranes in paraneopteran insect orders with comments on their function and evolutionary significance. Arthropod Struct Dev 33:139–148. doi: 10.1016/j.asd.2003.12.002
Smagghe G, Huvenne H, ffrench-Constant R, Swevers L, Kontogiannatos D, Menschaert G, Kourti A, Pauchet Y (2013) Colorado potato beetle (Coleoptera) gut transcriptome analysis: expression of RNA interference-related genes. Insect Mol Biol 22:668–684. doi: 10.1111/imb.12054
Smith AF, Tsuchida K, Hanneman E, Suzuki TC, Wells MA (1992) Isolation, characterization, and cDNA sequence of two fatty acid-binding proteins from the midgut of *Manduca sexta* larvae. J Biol Chem 267:380–384
South A, Hastings IM (2018) Insecticide resistance evolution with mixtures and sequences: a model-based explanation. Malar J 17:80. doi: 10.1186/s12936-018-2203-y
Sparks ME, Blackburn MB, Kuhar D, Gundersen-Rindal DE (2013) Transcriptome of the *Lymantria dispar* (Gypsy Moth) Larval Midgut in Response to Infection by *Bacillus thuringiensis*. PLoS One 8:. doi: 10.1371/journal.pone.0061190
Sparks TC, Nauen R (2015) IRAC: Mode of action classification and insecticide resistance management. Pestic Biochem Physiol 121:122–128. doi: 10.1016/j.pestbp.2014.11.014
Spit J, Badisco L, Verlinden H, Van Wielendaele P, Zels S, Dillen S, Vanden Broeck J (2012a) Peptidergic control of food intake and digestion in insects. Can J Zool 90:489–506. https://doi.org/10. 1139/z2012-014
Spit J, Breugelmans B, Van Hoef V, Simonet G, Zels S, Vanden Broeck J (2012b) Growth-inhibition effects of pacifastin-like peptides on a pest insect: The desert locust, *Schistocerca gregaria*. Peptides 34:251–257. doi: 10.1016/j.peptides.2011.06.019
Spit J, Zels S, Dillen S, Holtof M, Wynant N, Vanden Broeck J (2014) Effects of different dietary conditions on the expression oftrypsin-and chymotrypsin-like protease genes in the digestive system of the migratory locust, *Locusta migratoria*. Insect Biochem Mol Biol 48:100–109. doi: 10.1016/j.ibmb.2014.03.002
Spit J, Badisco L, Vergauwen L, Knapen D, Vanden Broeck J (2016) Microarray-based annotation of the gut transcriptome of the migratory locust, *Locusta migratoria*. Insect Mol Biol 25:745–756. doi: 10.1111/imb.12258
Srinivasan A, Giri A, Gupta V (2006) Structural and functional diversities in lepidopteran serine proteases. Cell Mol Biol Lett 11:132–154. doi: 10.2478/s11658-006-0012-8
Srp J, Nussbaumerová M, Horn M, Mareš M (2016) Digestive proteolysis in the Colorado potato beetle,

Leptinotarsa decemlineata: Activity-based profiling and imaging of a multipeptidase network. Insect Biochem Mol Biol 78:1–11. doi: 10.1016/j.ibmb.2016.08.004

Sternini C, Anselmi L, Rozengurt E (2008) Enteroendocrine cells-a site of taste in gastrointestinal chemosensing. Curr Opin Endocrinol Diabetes Obes 15:73–78. doi: 10.1097/MED.0b013e3282f43a73.Enteroendocrine

Stoffolano JG, Haselton AT (2013) The Adult Dipteran Crop: A Unique and Overlooked Organ. Annu Rev Entomol 58:205–225. doi: 10.1146/annurev-ento-120811-153653

Sun Z, Xu C, Chen S, Shi Q, Wang H, Wang R, Song Y, Zeng R (2019) Exposure to Herbicides Prime P450-Mediated Detoxification of *Helicoverpa armigera* against Insecticide and Fungal Toxin. Insects 10:28. doi: 10.3390/insects10010028

Svoboda JA (1999) Variability of Metabolism and Function of Sterols in Insects. Crit Rev Biochem Mol Biol 34:49–57. doi: 10.1080/10409239991209183

Sword GA, Lecoq M, Simpson SJ (2010) Phase polyphenism and preventative locust management. J Insect Physiol 56:949–957. doi: 10.1016/j.jinsphys.2010.05.005

Tabashnik BE, Brévault T, Carrière Y (2013) Insect resistance to Bt crops: lessons from the first billion acres. Nat Biotechnol 31:510–521. doi: 10.1038/nbt.2597

Tabatabaei PR, Hosseininaveh V, Goldansaz SH, Talebi K (2011) Biochemical characterization of digestive proteases and carbohydrases of the carob moth, *Ectomyelois ceratoniae* (Zeller) (Lepidoptera: Pyralidae). J Asia Pac Entomol 14:187–194. doi: 10.1016/j.aspen.2010.12.010

Takashima S, Adams KL, Ortiz P a., Ying CT, Moridzadeh R, Younossi-Hartenstein A, Hartenstein V (2011) Development of the *Drosophila* entero-endocrine lineage and its specification by the Notch signaling pathway. Dev Biol 353:161–172. doi: 10.1016/j.ydbio.2011.01.039

Takeuchi H, Chen J-H, Jenkins JR, Bun-Ya M, Turner PC, Rees HH (2004) Characterization of a sterol carrier protein 2/3-oxoacyl-CoA thiolase from the cotton leafworm (*Spodoptera littoralis*): a lepidopteran mechanism closer to that in mammals than that in dipterans. Biochem J 382:93–100. doi: 10.1042/BJ20040717

Tanaka R, Takase Y, Kanachi M, Enomoto-Katayama R, Shirai T, Nakagoshi H (2007) Notch-, Wingless-, and Dpp-mediated signaling pathways are required for functional specification of *Drosophila* midgut cells. Dev Biol 304:53–61. doi: 10.1016/j.ydbio.2006.12.018

Terra WR (2001) The origin and functions of the insect peritrophic membrane and peritrophic gel. Arch Insect Biochem Physiol 47:47–61. doi: 10.1002/arch.1036

Terra WR, Ferreira C (1994) Insect digestive enzymes : properties , compartmentalization and function. Comp Biochem Physiol 109B:1–62

Terra WR, Ferreira C (2005) Biochemistry of digestion. Compr Mol Insect Sci 4:171–224 Elsevier

Terra WR, Ferreira C (2012) Biochemistry and molecular biology of digestion. Insect Mol Biol Biochem 365–418. Elsevier

Tian F, Mo X, Rizvi SAH, Li C, Zeng X (2018) Detection and biochemical characterization of insecticide resistance in field populations of Asian citrus psyllid in Guangdong of China. Sci Rep 8:12587. doi: 10.1038/s41598-018-30674-5

Turunen S (1979) Digestion and absorption of lipids in insects. Comp Biochem Physiol Part A Physiol 63:455–460. doi: 10.1016/0300-9629(79)90171-3

Valencia A, Alves AP, Siegfried BD (2013) Molecular cloning and functional characterization of an endogenous endoglucanase belonging to GHF45 from the western corn rootworm, *Diabrotica virgifera virgifera*. Gene 513:260–267. doi: 10.1016/j.gene.2012.10.046

Valencia A, Wang H, Soto A, Aristizabal M, Arboleda JW, Eyun S Il, Noriega DD, Siegfried B (2016) Pyrosequencing the midgut transcriptome of the banana weevil *Cosmopolites sordidus* (germar) (coleoptera: Curculionidae) reveals multiple protease-like transcripts. PLoS One 11:. doi: 10.1371/journal.pone.0151001

Valenzuela JG, Francischetti IMB, Pham VM, Garfield MK, Ribeiro JMC (2003) Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. Insect Biochem Mol Biol 33:717–732. doi: 10.1016/S0965-1748(03)00067-5

Van Der Horst DJ (2003) Insect adipokinetic hormones : release and integration of flight energy metabolism. Comp Biochem Physiol B Biochem Mol Biol 136:217–226. https://doi.org/10.1016/S1096- 4959(03)00151-9

Van Hoef V, Breugelmans B, Spit J, Simonet G, Zels S, Billen J, Vanden Broeck J (2011) Functional analysis of a pancreatic secretory trypsin inhibitor-like protein in insects: silencing effects resemble

the human pancreatic autodigestion phenotype. Insect Biochem Mol Biol 41:688–695. https://doi.org/10.1016/j.ibmb.2011.04.012
Verrey F, Closs EI, Wagner CA, Palacin M, Endou H, Kanai Y (2004) CATs and HATs: the SLC7 family of amino acid transporters. Pflugers Arch Eur J Physiol 447:532–542. doi: 10.1007/s00424-003-1086-z
Vogel E, Santos D, Mingels L, Verdonckt T-W, Vanden Broeck J (2019) RNA Interference in Insects: Protecting Beneficials and Controlling Pests. Front Physiol 9:1912. doi: 10.3389/fphys.2018.01912
Voght SP, Fluegel ML, Andrews LA, Pallanck LJ (2007) *Drosophila* NPC1b Promotes an Early Step in Sterol Absorption from the Midgut Epithelium. Cell Metab 5:195–205. doi: 10.1016/j.cmet.2007.01.011
Vollero A, Imperiali FG, Cinquetti R, Margheritis E, Peres A, Bossi E (2016) The D-amino acid transport by the invertebrate SLC6 transporters KAAT1 and CAATCH1 from *Manduca sexta*. Physiol Rep 4:e12691. doi: 10.14814/phy2.12691
Vorster J, Rasoolizadeh A, Goulet MC, Cloutier C, Sainsbury F, Michaud D (2015) Positive selection of digestive Cys proteases in herbivorous Coleoptera. Insect Biochem Mol Biol 65:10–19. doi: 10.1016/j.ibmb.2015.07.017
Wang F-C, Li X-S, Qin L, Jiang Y-R, Wang G-B, Liu W, Sun Y, He Y-Z (2016a) Transcriptome Analysis of the Midgut of the Chinese Oak Silkworm *Antheraea pernyi* Infected with *Antheraea pernyi* Nucleopolyhedrovirus. PLoS One 11:e0165959. doi: 10.1371/journal.pone.0165959
Wang XY, Yu HZ, Geng L, Xu JP, Yu D, Zhang SZ, Ma Y, Fei DQ (2016b) Comparative transcriptome analysis of *Bombyx mori* (Lepidoptera) larval midgut response to BmNPV in susceptible and near-isogenic resistant strains. PLoS One 11:1–22. doi: 10.1371/journal.pone.0155341
Watanabe H, Tokuda G (2010) Cellulolytic Systems in Insects. Annu Rev Entomol 55:609–632. doi: 10.1146/annurev-ento-112408-085319
Wegener C, Veenstra JA (2015) Chemical identity, function and regulation of enteroendocrine peptides in insects. Curr Opin Insect Sci 11:8–13. doi: 10.1016/j.cois.2015.07.003
Willis JD, Oppert B, Oppert C, Klingeman WE, Jurat-Fuentes JL (2011) Identification, cloning, and expression of a GHF9 cellulase from *Tribolium castaneum* (Coleoptera: Tenebrionidae). J Insect Physiol 57:300–306. doi: 10.1016/j.jinsphys.2010.11.019
Woodring J (2017) The flow and fate of digestive enzymes in the field cricket, *Gryllus bimaculatus*. Arch Insect Biochem Physiol 95:1–12. doi: 10.1002/arch.21398
Woodring J, Weidlich S, Hoffmann KH, Woodring J, Huang X, McNeill MR, Ma J, Qin X, Tu X, Cao G et al (2017) Effects of different dietary conditions on the expression oftrypsin-and chymotrypsinlike protease genes in the digestivesystem of the migratory locust, *Locusta migratoria*. Arch Insect Biochem Physiol 110:100–109. https://doi.org/10.1002/arch.21303
Wynant N, Verlinden H, Breugelmans B, Simonet G, Vanden Broeck J (2012) Tissue-dependence and sensitivity of the systemic RNA interference response in the desert locust, *Schistocerca gregaria*. Insect Biochem Mol Biol 42:911–917. doi: 10.1016/j.ibmb.2012.09.004
Xiang M, Zhang X, Deng Y, Li Y, Yu J, Zhu J, Huang X, Zhou J, Liao H (2018) Comparative transcriptome analysis provides insights of anti-insect molecular mechanism of *Cassia obtusifolia* trypsin inhibitor against Pieris rapae. Arch Insect Biochem Physiol 97:1–21. doi: 10.1002/arch.21427
Xiong C, Xu X, Kumar D, Zheng Y, Xi W, Liang Q, Chen D, Zhang Z, Huang Z, Zou X, et al (2017) Uncovering the immune responses of Apis *Mellifera ligustica* larval gut to *Ascosphaera apis* infection utilizing transcriptome sequencing. Gene 621:40–50. doi: 10.1016/j.gene.2017.04.022
Xu Z-B, Zou X-P, Zhang N, Feng Q-L, Zheng S-C (2015) Detoxification of insecticides, allechemicals and heavy metals by glutathione S -transferase SlGSTE1 in the gut of *Spodoptera litura*. Insect Sci 22:503–511. doi: 10.1111/1744-7917.12142
Yang B, Huang W, Zhang J, Xu Q, Zhu S, Zhang Q, Beerntsen BT, Song H, Ling E (2016) Analysis of gene expression in the midgut of *Bombyx mori* during the larval molting stage. BMC Genomics 17:1–17. doi: 10.1186/s12864-016-3162-8
Ye XD, Su YL, Zhao QY, Xia WQ, Liu SS, Wang XW (2014) Transcriptomic analyses reveal the adaptive features and biological differences of guts from two invasive whitefly species. BMC Genomics 15:1–12. doi: 10.1186/1471-2164-15-370
Yew JY, Chung H (2015) Insect pheromones: An overview of function, form, and discovery. Prog Lipid Res 59:88–105. doi: 10.1016/j.plipres.2015.06.001
Yokoe Y (1964) Cellulase activity in the termite, *Leucotermes speratus*, with new evidence in support of

a cellulase produced by the termite itself. Sci Pap Coll Gen Educ 14:115–120
Zels S, Dillen S, Crabbé K, Spit J, Nachman RJ, Vanden Broeck J (2015) Sulfakinin is an important regulator of digestive processes in the migratory locust, *Locusta migratoria*. Insect Biochem Mol Biol 61: 8–16. https://doi.org/10.1016/j.ibmb.2015.03.008
Zhang J, Zhang Y, Li J, Liu M, Liu Z (2016) Midgut transcriptome of the cockroach *Periplaneta americana* and its microbiota: Digestion, detoxification and oxidative stress response. PLoS One 11:1–20. doi: 10.1371/journal.pone.0155254
Zhang S, Shukle R, Mittapalli O, Zhu YC, Reese JC, Wang H, Hua BZ, Chen MS (2010) The gut transcriptome of a gall midge, *Mayetiola destructor.* J Insect Physiol 56:1198–1206. doi: 10.1016/j.jinsphys.2010.03.021
Zhang S, Huang J, Hu R, Guo G, Shang X, Wu J (2017b) Characterization of a new multifunctional beta-glucosidase from *Musca domestica*. Biotechnol Lett 39:1219–1227. https://doi.org/10.1007/s10529-017-2351-0
Zhang Y, Fan J, Sun J, Francis F, Chen J (2017a) Transcriptome analysis of the salivary glands of the grain aphid, *Sitobion avenae*. Sci Rep 7: 15911. https://doi.org/10.1038/s41598-017-16092-z
Zheng J-C, Sun S-L, Yue X-R, Liu T-X, Jing X (2018) Phylogeny and evolution of the cholesterol transporter NPC1 in insects. J Insect Physiol 107:157–166. doi: 10.1016/j.jinsphys.2018.04.007
Zhu-Salzman K, Zeng R (2015) Insect Response to Plant Defensive Protease Inhibitors. Annu Rev Entomol 60:233–252. doi: 10.1146/annurev-ento-010814-020816
Zhu F, Moural TW, Nelson DR, Palli SR (2016) A specialist herbivore pest adaptation to xenobiotics through up-regulation of multiple Cytochrome P450s. Sci Rep 6:20421. doi: 10.1038/srep20421
Zibaee A, Alborzi Z, Hoda H (2014) A review on digestive TAG-lipases of insects. Arch Phytopathol Plant Prot 47:2187–2192. doi: 10.1080/03235408.2013.870112

Chapter 2

Set-up of the RNA-Seq study

2.1. Introduction

The primary aim of this doctoral research was to investigate the transcript regulation of digestion in the desert locust. Therefore, we decided to perform an RNA-Seq profiling of the midgut at three well-selected timepoints after food uptake. The selection of these timepoints was largely based on prior literature studies, discussions and pilot experiments analyzing the general feeding behavior of *S. gregaria*. We eventually decided to study the midgut transcriptome in the following three temporal conditions: ten minutes, two hours and twenty-four hours after food uptake. These conditions would allow us to analyze transcriptomic changes associated with the initiation and regulation of digestion in response to food intake. Moreover, this study would also allow us to create a reference transcriptome for *S. gregaria* midgut, containing all relevant transcripts required for the proper functioning of this organ. This transcriptomic database will be an excellent resource for future gut-related studies in the desert locust.

Because the majority of this thesis is based on the outcome of this RNA-Seq experiment, I will first explain the experimental set-up as well as all the general and specialized methods that were applied in preparing this large-scale experiment.

2.2. Design of the experiment

The desert locust is described to be a continuous feeder. This implies that the insect's feeding pattern is not restricted to discrete periods of feeding, as is observed in discontinuous feeders, such as mosquitoes. However, the desert locust demonstrates a more irregular feeding pattern than for example most lepidopteran larvae that continuously take up food (Chapman, 2013). The desert locust will ingest food until its crop is completely filled. At that point, the crop will be maximally expanded, resulting in the activation of stretch receptors present on the foregut. Active stretch receptors will lead to the cessation of feeding (Finlay and Lowenstein, 1958). Next, the food bolus will be passed on to the midgut lumen via the proventriculus where it will be further degraded enzymatically. Only when the majority of the food bolus has been passed on to the midgut, new food can be ingested. The new food bolus will then clear the remains of the previous meal from the midgut. This entire process is expected to be accompanied with distinct alterations in the transcriptome of tissues involved in the digestive process. However, to the best of my knowledge, such specific large-scale studies investigating gene expression at different stages of digestion have not yet been reported in insects.

We observed that the initial feeding stage of *S. gregaria* typically took about ten minutes. On average, one hour later, the insects continued to feed, indicating that the majority of food had passed on to the midgut. A similar feeding pattern has also been described in literature for another locust species, *L. migratoria* (Chapman, 2013). Interestingly, if no extra food could be consumed after the initial feeding stage, the food bolus remained inside the midgut for an extended period of time. Based on these observations, we decided to analyze the transcript profile of the midgut in *S. gregaria* during the initial phase of digestion (10 minutes after food uptake), and the core phase of digestion, when the majority of food was located in the midgut lumen (2 hours after food uptake), and compare both to a reference stage when no fresh food was available in the gut (24 hours after food uptake). This would allow us to investigate transcriptomic changes that are directly associated with the onset and maintenance of the digestive process in the desert locust.

We decided to primarily focus on the transcript profile of the midgut at these stages, since this is the principal site of enzymatic digestion and nutrient uptake (Chapter 1). Additionally, next to the midgut, also the foregut, caeca, hindgut, Malpighian tubules and the head of locusts were collected for putative future studies of digestion-associated processes in these tissues.

2.3. General methods

2.3.1. Customized rearing conditions prior to RNA-Seq

2.3.1.1. Removing possible gut contaminants: theory

In next-generation sequencing (NGS), which will be explained in detail later in this chapter (Ch2|2.3.3), every mRNA molecule present in a sample in theory has a chance of being sequenced. Therefore, it was absolutely critical to only introduce pure *S. gregaria* tissue extracts for NGS, without any external contaminations. This is extremely challenging for the midgut. Next to endogenous genetic material, many other potential sources of genetic contamination can be present, including food particles and (parasitic, symbiotic or commensal) microorganisms. These contaminants had to be removed (or at least minimized) prior to sequencing. Any exogenous genetic material left in the samples could potentially bias the sequencing results.

Therefore, midguts were always thoroughly cleaned after dissection. This is typically sufficient to remove food particles and other visible sources of exogenous genetic material. However, the most problematic source of contamination in *S. gregaria* midgut samples are gregarines (Apicomplexa). Gregarines are parasitic microorganisms that, at the time of this experiment, were persistently present in our *S. gregaria* colony. Their presence in the gastrointestinal tract of laboratory-reared, as well as free-living grasshoppers and locusts has been frequently reported (Valigurová and Koudela, 2008). They exhibit a typical fecal-oral lifecycle, with a trophozoite feeding stage inside the host's intestine, and are transferred to a new host organism as oocysts in feces via oral ingestion. Once inside their new host, these oocysts rapidly develop into sporozoites and subsequently into trophozoites exhibiting an active metabolism inside the host's gastrointestinal tract (Dillon and Charnley, 2002; Takahashi et al., 2003; Valigurová and Koudela, 2008; Lantova and Volf, 2014).

Figure 2.1. Light microscope picture of a gregarine infested food bolus dissected from the midgut of a *S. gregaria* 5th instar from our colony. Yellow rods are gregarines, brown material is the food bolus.

In our *S. gregaria* colony, large amounts of gregarines could be found in the midgut of larval and adult locusts at the time of this research. For most experiments, this does not need to be alarming, since the locusts seem to develop normally and many reports of gregarine infested laboratory-reared insects exist. However, because this research focused specifically on the intrinsic digestive system of the desert locust, extreme care to avoid the introduction of gregarines in the samples designated for RNA-Seq was required. If introduced to these samples, the mRNA of these microorganisms would mix with the endogenous mRNA library of *S. gregaria,* and would end up competing with the *S. gregaria* mRNA for sequencing. Therefore, to avoid introducing contaminating gregarines into the RNA-Seq samples, we decided to (1) disinfect the eggs and (2) rear the animals in a disinfect, closed environment.

2.3.1.2. Removing possible gut contaminants: practice

From six different egg pots, six to seven days old eggs were carefully removed from the egg tubes and rinsed with distilled water to remove excess sand particles. Next, the eggs were disinfected by first washing them in ethanol (70%) for 30 seconds and subsequently rinsing them with distilled water (few seconds). The disinfected eggs were then transferred to cleaned egg pots (10% bleach) containing an autoclaved sand/soil mixture (1:3 ratio). On average, around 300 *S. gregaria* eggs were disinfected and divided over six new egg pots, generating a large enough population for the later RNA-Seq sampling. The egg pots were placed in a disinfected cage (10% bleach) located inside a closed incubator (J.P. Selecta®) representing a clean environment. The incubator was set at the same conditions as in the breeding facility (14h/10h day/night cycle, 32 °C, 60% relative humidity). An extra lamp was placed inside the cage to produce extra heat and light, analogous to the cages in the regular breeding facility. The egg pots were removed from the cage once all animals hatched from the eggs. After hatching, the animals were fed daily *ad libitum* with freshly washed cabbage, *Brassica oleracea*, until they reached the 5th larval stage. During the 5th larval stage, the insects were subjected to a restricted diet (Ch2|2.3.1.3) and were later sacrificed (Ch2|2.3.2.1).

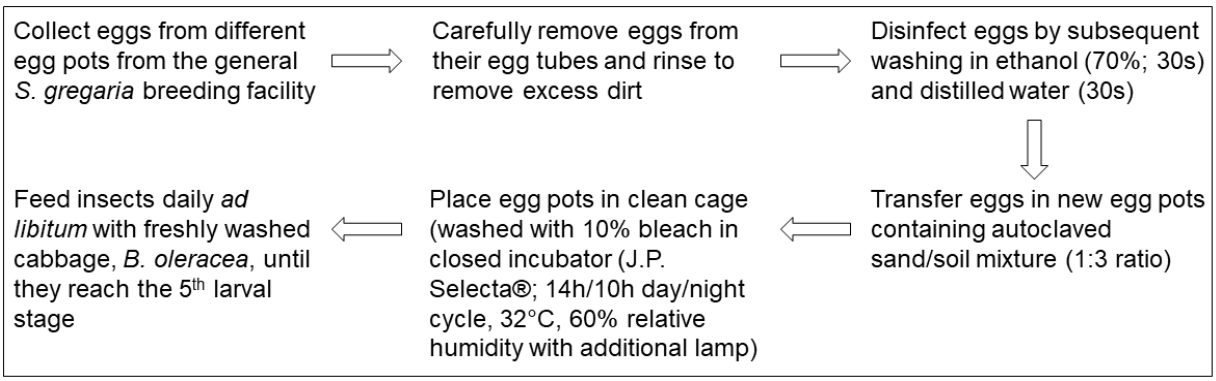


Figure 2.2. Summary of the applied method for disinfecting eggs and customized rearing conditions to avoid gregarine contaminants in the *S. gregaria* gut.

2.3.1.3. Synchronizing feeding behavior

The *S. gregaria* larvae were fed daily *ad libitum* with freshly washed cabbage until they reached the 5th larval stage. A total of 108 nymphs were developmentally synchronized on the day of the molt to the 5th larval stage (marked as day 0). During the fifth larval stage, animals were put on a routine diet in order to synchronize their feeding behavior. The animals were only allowed to feed from freshly washed cabbage twice a day: one hour in the morning and one

hour in the afternoon (Fig 2.3). At day three of the fifth larval stage, all animals displayed the same feeding behavior and the final group selection for tissue collection could occur. Because we were interested in the transcriptomic profile of the midgut at precise timepoints after food uptake, the synchronization of their feeding behavior was essential. This way, potential variation introduced by divergent food uptake between animals was reduced. It was also interesting to observe that a routine behavior was easily implemented in a large group of locusts.

2.3.2. Tissue collection and mRNA isolation

2.3.2.1. Dissections and tissue collection

On the morning of day 4 of the 5th larval stage, all 108 animals (equal male/female ratio) were fed with freshly washed cabbage for one hour. After this meal, three groups were established. This is graphically summarized in figure 2.3. One group of 36 animals was deprived from food for the next 24 hours and sacrificed the next day, day 5 (Group C). The other animals were allowed to eat for 1 hour in the afternoon of day 4, and received a final meal in the morning of day 5. They were allowed to eat for 10 minutes from this meal, which was equal to their typically observed initial feeding stage as explained in paragraph 2.2 of this chapter. Within these 10 minutes, all animals finished eating and were freely moving away from the food. A group of 36 animals was sacrificed at ten minutes (Group A) and another group of 36 animals was sacrificed at 2 hours (Group B) after this initial feeding period. The dissections of all three groups were executed on the same day around the same time to exclude any potential day-related effects (Fig. 2.3).

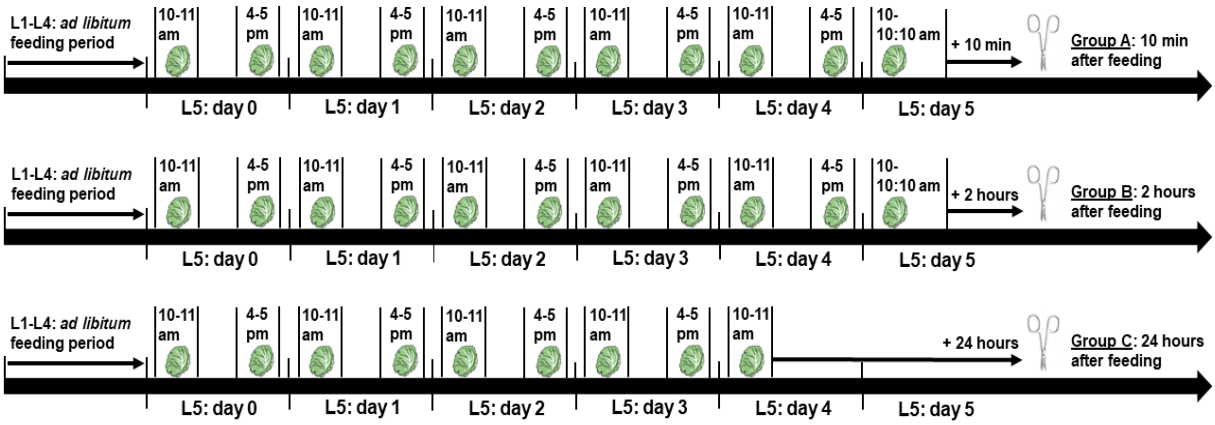


Figure 2.3. Graphical summary of the experimental set-up for the collection of the RNA-Seq samples for three different conditions. The different conditions were: sacrificed 10 minutes after feeding (Group A), sacrificed 2 hours after feeding (Group B), and sacrificed 24 hours after feeding (Group C).

On day 5 of the 5th larval stage, a total of 108 *S. gregaria* larvae were dissected for their head, foregut, caeca, midgut and hindgut with Malpighian tubules. Each timepoint after feeding was represented by 36 *S. gregaria* larvae (equal male/female ratio). Their tissues were equally divided over 6 different samples per tissue (biological replicates), with each sample thus containing pooled tissues from 6 individual locusts (Fig. 2.4). Only heads were stored separately. The different timepoints were: 10 minutes after feeding (Group A), 2 hours after feeding (Group B) and 24 hours after feeding (Group C). In order to minimize time-dependent variation between biological replicates, all insects belonging to the same condition were simultaneously dissected. Dissected tissues were immediately transferred to screwcapped 2

mL Eppendorf vials containing 0.5 mL RNA*later* (Thermo Scientific™) to avoid RNA degradation. Further sampling of the tissues was continued once all tissues of that condition were collected. For further sampling, tissues were removed from RNA*later* and were further dissected and cleaned from exogenous material in *S. gregaria* Ringer's solution (1 L: 8.766 g NaCl; 0.188 g $CaCl_2$; 0.746 g KCl; 0.407 g $MgCl_2$; 0.336 g $NaHCO_3$; 30.807 g sucrose; 1.892 g trehalose; pH 7.2). Midgut tissues were stored in 2 mL MagNA Lyser Green Bead vials (Roche) for subsequent total RNA extraction, while other tissues were stored in regular 2 mL Eppendorf screwcap vials. All samples were immediately transferred to liquid nitrogen to prevent RNA degradation and then stored at -80 °C until further processing.

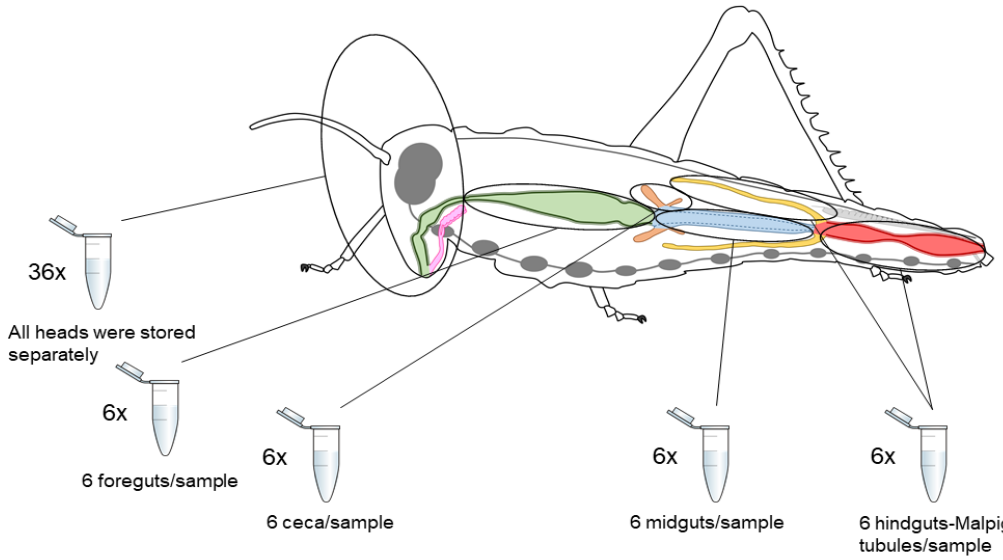


Figure 2.4. Graphical summary of the sampling for each condition. For each condition, 6 samples containing tissues of 6 different *S. gregaria* L5 nymphs were collected. The different conditions were: sacrificed 10 minutes after feeding (Group A), sacrificed 2 hours after feeding (Group B), and sacrificed 24 hours after feeding (Group C).

2.3.2.2. Total RNA extraction

To each MagNA Lyser Green Beads vial (Roche) containing midgut tissues, 1 mL QIAzol lysis reagent buffer (Thermo Scientific™) was added. Next, the tissues were homogenized for 30 seconds at 6000 x g using a MagNA Lyser instrument (Roche). Subsequent total RNA extraction from the midgut tissues was performed using the RNeasy Lipid Tissue extraction kit (Qiagen) according to manufacturer's instructions. An additional DNaseI treatment (Qiagen) was performed to remove contaminating genomic DNA. In the final step, total RNA was extracted in 80 µL nuclease-free water.

2.3.2.3. Nucleic acid concentration and quality control

Concentration of the extracted total RNA was assessed using a NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Scientific™). Quality of the extracted total RNA was further assessed using the Agilent® 2100 Bioanalyzer. The in-house obtained NanoDrop data output is available in the supplementary data of this thesis (Supp. Table S2.1). Additional concentration and quality controls were performed at the sequencing facility NXTGNT (Ghent University, Ottergemsesteenweg 460, B-9000 Ghent, Belgium) by using the 'Quant-it ribogreen RNA assay' (Life Technologies™) and the RNA 6000 Nano chip (Agilent Technologies®), respectively.

2.3.3. Next generation sequencing: Illumina®

2.3.3.1. TruSeq® library preparation[1]

Our study was based on the Illumina® deep sequencing of mRNA molecules. Therefore, mRNA was first purified from the total RNA mixture in every sample. From each sample, one µg of RNA was used to perform an Illumina® sequencing library preparation using the TruSeq® Stranded mRNA Library prep kit (Illumina®) according to the manufacturer's protocol. The goal of the library preparation was to first purify the poly-A containing mRNA molecules from the total RNA mixture in every sample using poly-T oligo attached magnetic beads. Next, these purified mRNA molecules were fragmented into smaller fragments (100–1000 nt) using divalent metal cations under elevated temperature in a proprietary Illumina® fragmentation buffer. Fragmentation was necessary because sequencing can only be performed on shorter strands. The main reason for this is that the quality, and therefore also the confidence level, of sequenced bases decreases with the length of a fragment. Sequencing shorter reads massively improves the correctness of the sequenced fragment across its entire length. Our sequencing was performed on a high throughput Illumina® NextSeq 500 flow cell generating 75 bp single reads. The smaller RNA fragments resulting from the fragmentation step were copied into first strand cDNA using reverse transcriptase and random primers. This was done because NGS is designed for sequencing DNA material, and is not suited for direct sequencing of RNA. These cDNA fragments were further prepared for sequencing. Next, a single 'A' base was added to these cDNA fragments followed by the ligation of the Illumina® adapters. Such adapters contain the sequencing primer promoter site, indices for multiplexing and terminal sequences for hybridization to the oligo flow cell. All cDNA fragments originating from the same sample were identified by an identical index sequence. This provided fragments from each sample with a unique barcode and allowed combining of samples on the flow cell. The resulting cDNA products were then purified and enriched with PCR to create the final cDNA library. During library preparation, 13 PCR cycles were used. Libraries were quantified by qPCR, according to Illumina's protocol 'Sequencing Library qPCR Quantification protocol guide', version February 2011. A High Sensitivity DNA chip (Agilent Technologies®) was used to control the library's size distribution and quality.

2.3.3.2. The Illumina® sequencing principle[1]

Illumina® sequencing is a universally used NGS technology based on the sequencing by synthesis principle. It is composed of three basic steps: library preparation, cluster generation and sequencing. Here I will provide a general explanation on NGS on the Illumina® platform based on the technical information provided by Illumina®.

The constructed cDNA library (as explained in Ch2|2.3.3.1) is denatured into single-stranded DNA (ssDNA) fragments and loaded onto the flow cell, a glass slide containing lanes coated with two types of oligos complementary to the adapter sequences ligated to the cDNA fragments. The 3’ terminals of the fragmented cDNA strands will randomly hybridize to the oligos on the surface of the flow cell. Now, a polymerase creates a complement of the hybridized ssDNA fragment. The resulting double stranded molecule is denatured and the original template is washed away. The remaining strands are clonally amplified through a process called bridge amplification (Fig. 2.5). During this process, the remaining ssDNA strand

[1] TruSeq® library preparation and Illumina® sequencing were performed at the sequencing facility NXTGNT (Ghent University, Ottergemsesteenweg 460, B-9000 Ghent, Belgium).

bends over and the free adapter region will hybridize to the second type of oligo on the flow cell, creating a bridge. Polymerases then generate the complementary strand forming a double stranded bridge. This bridge is then denatured, resulting in two single stranded copies of the original molecule attached to the flow cell. These resulting neighboring complementary strands serve as templates for a next cycle of bridge amplification. This process is repeated over and over and occurs simultaneously for millions of clusters. Each cluster consists of thousands of identical copies of the original template ssDNA molecules. After bridge amplification, the reverse strands are cleaved and washed off, leaving only clonal colonies of the forward strands ready for sequencing.

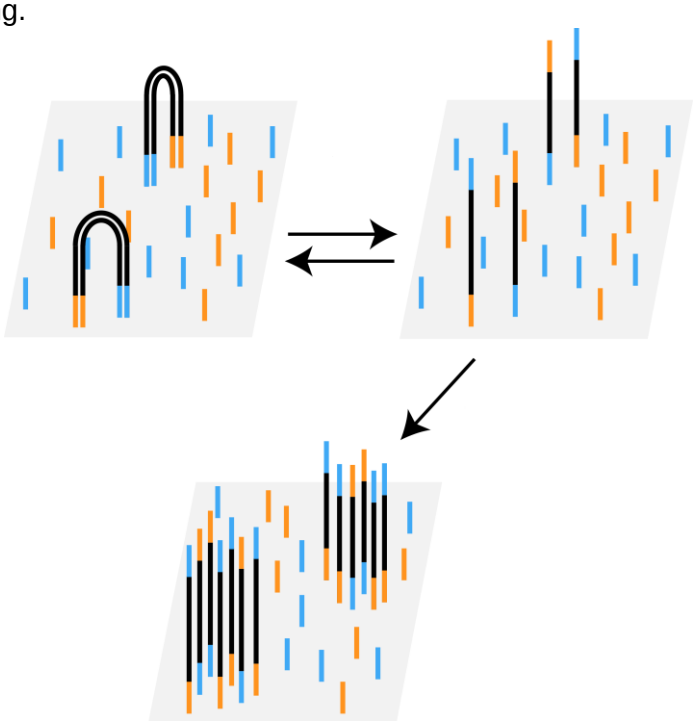


Figure 2.5. Graphical representation of cluster generation through bridge amplification in an Illumina® flow cell. ssDNA fragments containing Illumina® adapters are hybridized to complimentary oligos on the flow cell surface. Polymerases create complements of the hybridized ssDNA fragments and the original templates are washed away. The remaining ssDNA strands bend over and the free adapter regions will hybridize to the second type of oligo on the flow cell, creating a bridge. During repetitive cycles of amplification and denaturation, polymerases generate millions of clusters representing the original ssDNA strand.

Sequencing begins with the hybridization of the sequencing primers to the adapter sequences at the free 3' ends of the ssDNA strands. Sequencing by synthesis occurs in a cyclic process. Each cycle, four fluorescently tagged deoxynucleoside triphosphates (dNTPs) compete for addition to the growing chain, complementary to the attached template strand. Only one nucleotide can be incorporated at a time based on the sequence of the template. Once added to the chain, the fluorescent tag serves as a terminator for polymerization, inhibiting the incorporation of a new dNTP at the next open spot. The sequencing reaction is conducted simultaneously to the tens of millions of clusters present on the surface of the flow cell. Once every growing fragment has incorporated exactly one dsNTP, the fluorescent tags are excited by a light source and a nucleotide-specific fluorescent signal is emitted. Following this, the fluorescent tags are removed and a new sequencing cycle can begin (Fig. 2.6). This base-by-base sequencing is repeated until the desired length is attained; 75 nucleotides for our RNA-Seq experiment. The signal wave-length along with the signal intensity determines the identity and reliability of each incorporated nucleotide. This process is referred to as base calling. Sequence identities and base calling qualities of all the generated reads are provided by the sequencing platform in typical FASTQ records (Chapter 3).

In this doctoral research, a total of 18 midgut samples were sequenced on a high throughput Illumina® NextSeq 500 flow cell generating 75 bp single reads. The samples were multiplexed, allowing to sequence all samples on one Illumina® flow cell. Moreover, to increase sequencing depth, which is the amount of distinct reads produced during sequencing, sequencing was performed *in duplo* on two distinct flow cells. Per sample, on average 55 ± 10 million reads were generated.

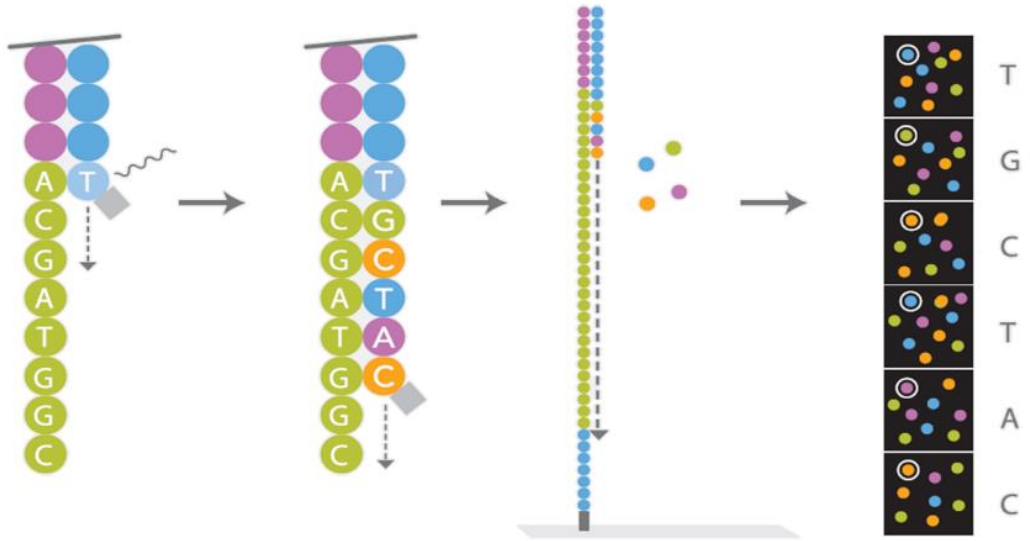


Figure 2.6. Graphical representation of the sequencing by synthesis principle on the Illumina® NGS platform. First, sequencing primers hybridize to the adapter sequences at the free 3' ends of the ssDNA strands. Then, four fluorescently tagged bases compete for addition to the growing chain, complementary to the attached template strand. Only one nucleotide can be incorporated at a time. Once every growing fragment has incorporated exactly one complementary base, the fluorescent tags are excited by a light source and a nucleotide-specific fluorescent signal is detected. Next, the label is removed and a new sequencing cycle is initiated.

2.4. Discussion

2.4.1. Effectiveness of the optimized rearing conditions

We were able to successfully raise gregarine-free insects following the proposed rearing method. All animals reared in the clean, closed environment were healthy and showed normal development. But foremost, these animals did not harbor any gregarines in their intestine. In contrast, during the same period, all conventionally reared animals inside the breeding facility harbored many gregarines in their gut.

These results clearly prove the possibility of completely removing a gregarine contamination from a colony in only one generation. Allowing disinfected eggs to hatch in disinfected soil was probably the key to obtaining healthy, gregarine-free insects. It has already been described that the oocysts of gregarines are passed on orally via feces. In breeding cages, insects are constantly exposed to feces and will indiscriminately end up orally ingesting gregarine oocysts. Moreover, feces also end up on top of the soil in the egg pots. When hatchlings crawl out of the soil, they will encounter gregarine infected feces. It is assumable that hence the gregarines are passed on to the next generation of locusts. But when larvae hatched from disinfected eggs in a disinfected environment, the chances of them encountering gregarine infected feces were removed, consequently eliminating gregarine infection from the locust population.

Remarkably, the animals reared the incubators molted, on average, one day earlier to a subsequent larval stage compared to the conventionally reared animals. An indication that the incubator reared insects possibly acquired their critical weights faster, suggesting that the conventionally reared insects' acquisition of nutrients might be inhibited by the parasites in their gut. This demonstrates the importance and effectiveness of this treatment in generating healthy and gregarine-free *S. gregaria* larvae for RNA-Seq.

It was not evaluated if the treatment resulted in the removal of other, possibly beneficial, gut microbiota. Nevertheless, since all animals were healthy and showed no developmental defects, there was no reason to believe that this treatment negatively affected the intrinsic digestive system of the animals. Interestingly, earlier research in the desert locust suggested that its gut microbiome does not significantly contribute to its nutrition, although it does seem to play a substantial role in host defense against pathogens (Dillon and Charnley, 2002). Moreover, a recent study demonstrated the absence of resident microbiomes in several Lepidoptera insect species, illustrating that some insect species have a lifestyle independent from symbionts (Hammer et al., 2017). Further research is thus needed to truly investigate the contribution of *S. gregaria* symbionts to its digestive process. Furthermore, additional tests to compare the gut microbiota of incubator reared with conventionally reared insects would be an interesting addition to this research.

2.4.2. Midgut appearance at the different timepoints

The appearance of all dissected midguts for RNA-Seq sampling perfectly met our prior expectations. The observed midgut physiology at the three different timepoints after feeding is described in detail below, and is summarized in Table 2.1.

Ten minutes after food uptake (Group A):
Ten minutes after feeding, the crop was completely filled with fresh food, while the midgut only contained highly digested remainders of a previous meal. At this point, signals from the foregut had resulted in the termination of feeding, and food slowly migrated towards the midgut for enzymatic digestion. This timepoint was selected to investigate the initial stage of food uptake and digestion in the midgut.

Two hours after food uptake (Group B):
Two hours after feeding, the entire intestine, from foregut to hindgut, was filled with food. The majority of the food bolus was present in the midgut and the associated gastric caeca, where enzymatic processing is occurring. This timepoint was selected to investigate the active digestion of the food bolus and nutrient absorption in the midgut. This group was expected to give the highest amount of information on the regulation of genes playing a role in digestive processes in *S. gregaria*.

Twenty-four hours after food uptake (Group C):
Twenty-four hours after feeding, only highly digested food remains were still present in the midgut and hindgut, while the foregut was empty of food. At this point, the midgut was probably in a preliminary state of starvation. This timepoint was included as a reference, a point zero, during which the midgut was expected to perform the least amount of digestive activity. This condition could then be used to investigate digestion-related transcript enrichment in the *S. gregaria* midgut in the other two conditions.

Table 2.1 | Overview of the physical appearance of the intestinal tract of *S. gregaria* during the three different timepoints after feeding.

GROUP	TIMEPOINT	INTESTINAL TRACT APPEARANCE
A	10 minutes	Foregut full of fresh food, midgut and hindgut contain only highly digested food remains
B	2 hours	Food spread across gut, food bolus abundantly present in midgut lumen
C	24 hours	Foregut empty of food, midgut and hindgut contain only highly digested food remains

Per sample, midguts originating from six individuals were pooled. Each condition consisted of six samples. A total of eighteen samples representing three different conditions were sent for sequencing. All 18 samples met the sequencing facility's criteria for quality control and were subsequently sequenced on a high throughput Illumina® NextSeq 500 flow cell generating 75 bp single reads.

Next, the generated reads were mapped to a *S. gregaria* reference transcriptome, and the number of reads mapped to each transcript were quantified. These data were then used for constructing a *S. gregaria* midgut reference transcriptome and for analyzing differential expression between the three conditions. These procedures will be explained in more detail in the next chapters.

2.4.3. Challenges of using cabbage as a food source

B. oleracea, as a member of the Brassicaceae family of plants, is known to contain numerous poisonous allelochemicals, especially glucosinolates, which are non-volatile sulphur-containing secondary plant metabolites. It is therefore considered to be an extremely challenging food source for many insects. Nevertheless, our desert locust colony is fed daily with this type of food, supplemented with rolled oats, and shows normal developmental rates and fecundity. This suggests that *S. gregaria,* which is considered as a broad-spectrum herbivore (polyphagous species), is able to tolerate the detrimental factors present in their daily diet, possibly as a result of specific adaptations. This RNA-Seq study was also used to investigate if the desert locust indeed possesses distinct coping mechanisms dealing with this high concentration of allelochemicals ingested on a daily basis. Interestingly, wild *S. gregaria* in the Sahara desert sometimes feed on *Schouwia purpurea* (Brassicaceae), and is able to actively cope with the toxins released by these plants, demonstrating its intrinsic capacity to detoxify allelochemicals of Brassicaceae (Mainguet et al., 2000). We will look deeper into these coping mechanisms in chapter 3 (Ch3| 3.4.2.2.5) and chapter 4 (Ch4|4.4.3.1).

2.5. Conclusion

To investigate the transcript profile of the *S. gregaria* midgut during the digestive process, midgut samples representing three different timepoints after food uptake were sequenced by means of RNA-Seq. The three temporal conditions were: ten minutes, two hours, and twenty-four hours after feeding, representing the onset and maintenance of the digestive process, and a short period of food deprivation, respectively. The current chapter explained in detail the sample preparation and the tissue collection for this RNA-Seq study. The next chapters will explain the processing and subsequent analyses of the generated RNA-Seq reads.

Remarkably, it was possible to remove all gregarine contamination from the digestive tract of the desert locust when following customized rearing conditions. In general, these specifically reared animals were gregarine-free, healthy and appeared to develop faster than their conspecifics reared in conventional conditions. It is therefore advisable to rely on the customized way of insect rearing explained in this chapter for any future studies involving the *S. gregaria* intestine.

2.6. References

Chapman R. F. (2013) The Insects: Structure and Function. Cambridge University Press, New York

Dillon R, Charnley K (2002) Mutualism between the desert locust *Schistocerca gregaria* and its gut microbiota. Res Microbiol 153:503–509. doi: 10.1016/S0923-2508(02)01361-X

Finlay LH, Lowenstein O (1958) The structure and function of abdominal stretch receptors in insects. Proc R Soc London Ser B - Biol Sci 148:433–449. doi: 10.1098/rspb.1958.0037

Hammer TJ, Janzen DH, Hallwachs W, Jaffe SP, Fierer N (2017) Caterpillars lack a resident gut microbiome. Proc Natl Acad Sci 114:9641–9646. doi: 10.1073/pnas.1707186114

Lantova L, Volf P (2014) Mosquito and sand fly gregarines of the genus *Ascogregarina* and *Psychodiella* (Apicomplexa: Eugregarinorida, Aseptatorina) - Overview of their taxonomy, life cycle, host specificity and pathogenicity. Infect Genet Evol 28:616–627. doi: 10.1016/j.meegid.2014.04.021

Mainguet AM, Louveaux A, Sayed G, Rollin P (2000) Ability of a generalist insect, *Schistocerca gregaria*, to overcome thioglucoside defense in desert plants: tolerance or adaptation? Entomol Exp Appl 94:309–317. doi: 10.1046/j.1570-7458.2000.00632.x

Takahashi KT, Kawaguchi S, Kobayashi M, Toda T (2003) Parasitic eugregarines change their spatial distribution within the host digestive tract of Antarctic krill, *Euphausia superba*. Polar Biol 26:468–473. doi: 10.1007/s00300-003-0511-2

Valigurová A, Koudela B (2008) Morphological analysis of the cellular interactions between the eugregarine *Gregarina garnhami* (Apicomplexa) and the epithelium of its host, the desert locust *Schistocerca gregaria*. Eur J Protistol 44:197–207. doi: 10.1016/j.ejop.2007.11.006

Chapter 3

A *Schistocerca gregaria* midgut reference transcriptome for RNA-Seq

3.1. Introduction

Even today, the amount of studies describing the transcriptomic changes in the intestine of herbivorous insects in relation to feeding and digestion is still remarkably limited (Chapter 1). However, such transcriptome profiles would provide crucial information to help us better understand the divergent aspects of the digestive system of these insects, ranging from the general enzymatic digestion and nutrient uptake to the specialized detoxication of the numerous (plant) toxins they encounter. Within this context, we decided to create an annotated reference transcriptome of the midgut of *S. gregaria*, a polyphagous herbivorous insect which therefore is considered to be a good representative for this class of insects.

In the previous chapter, I have explained how three different timepoints after feeding were selected for differential gene expression analysis during different stages of digestion. This will be further described in chapter 4. Moreover, RNA-Seq of these samples provided us with the almost entire transcriptomic profile of the *S. gregaria* midgut during these three phases after food uptake. By combining this information, we were able to create an annotated midgut specific transcriptome that could be used as a reference for future studies, including the differential expression analysis in chapter 4.

3.2. Experimental procedure

3.2.1. Aims of the experiment

The millions of raw sequencing reads generated by the Illumina® sequencer were further processed and analyzed. The current chapter will explain the entire workflow from the processing of these short sequencing reads until the construction and the analysis of a *de novo* annotated midgut reference transcriptome for RNA-Seq. This reference transcriptome was later used to study differential gene expression in the midgut at different timepoints after food uptake (Chapter 4).

3.2.2. Preprocessing and quality control

After obtaining the short sequencing reads files for each sample from the Illumina® machine, these need to be preprocessed and assessed for quality prior to any downstream analyses. The standard output file of an Illumina® machine is a so-called FASTQ record. This file type typically comprises four lines containing the FASTA-like sequence output with associated base calling qualities, allowing the user to determine the quality of each sequenced base (Cock et al., 2010). A typical FASTQ record generally consists of four sections:

1. FASTA-like header, but the > symbol replaced is by the @ symbol. This is followed by an ID and an optional text.
2. The generated nucleotide sequence (typically one line).
3. + sign connecting section 2 and section 4. Sometimes + is followed by the same string of information as in section 1.
4. String of ASCII characters representing the quality score for each nucleotide in section 2.

An example of a FASTQ record generated in this RNA-Seq experiment:

```
@NS500361:286:HGLKMBGX2:1:11101:3435:10351:N:0:TAGCTT
AACAANATTACTTACGTTTATGAAGTTATTCTCAGATAACTCTACCCTCAGCAACTTGCTCATCTGGTAAAGACTG
+
AAAAA#EEEEEE6EEEEEEEEEEEEEEEEAEEAEEEEEEEEEEEEEEEEAEEE/EEEEEEEEEEEEEEEEEEEAAA
```

A first preprocessing step includes the removal of all contaminating Illumina® adapter sequences. It is indeed possible that the Illumina® machine continues sequencing into the adapter region at the 3' terminal of a molecule. If so, the output sequence will also contain (parts of) that adapter sequence. Therefore, it is crucial to find such reads and remove the adapter sequences where they occur, in a process called adapter trimming. This way, adapter sequences are removed from the FASTA-like sequence of the reads and only the biologically relevant features of the reads are retained for further downstream analyses. In this RNA-Seq study, a widely-applied software tool called Cutadapt (version 1.11) was used to trim the Illumina® adapter sequences from the reads (Martin, 2011).

Next, the FASTQ records are investigated to determine the overall base calling quality of the reads. This is typically done by calculating the Phred score (Q) of each base in every read, representing its error possibility. This is encoded as ASCII characters in the fourth section of the FASTQ record, parallel to the nucleotide sequence in section two. The current standard ASCII encoding is the so-called Sanger (+33) format, in which each character in section four of the FASTQ record has a numerical value according to the following scale:

!	**“**	**#**	**$**	**%**	**&**	**‘**	**(**	**)**	*****	**+**	**,**	**-**	**.**	**/**	**0**	**1**	**2**	**3**	**4**	**5**
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
6	**7**	**8**	**9**	**:**	**;**	**<**	**=**	**>**	**?**	**@**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	

There is a well-defined logic behind this alphabet, but this is beyond the scope of this work. It is important to know that Q is specified by a value between 0 and 40, and indicates the probability of a base call being correct by the formula P=10^(-Q/10). The following table illustrates this principle:

Q	ERROR	ACCURACY
0	1 in 1	**0%**
10	1 in 10	**90%**
20	1 in 100	**99%**
30	1 in 1000	**99.9%**
40	1 in 10 000	**99.99%**

For example, a Q of 0 for a base indicates that the base at this position is predicted to be 0% accurate, signifying a very bad base calling quality. A Q of 30 indicates that the base calling is predicted to be 99.9% correct, signifying a high quality. Phred scores are calculated for each base in every read of every sample and thus provide a global impression of the overall quality of the sequencing. A widely used tool for assessing and visualizing this information from FASTQ records is FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

3.2.3. Mapping of the sequencing reads

At the end of chapter 2, we ended up with millions of short sequencing reads (75 bp) for each sample representing one of the three conditions to be compared. The general goal of any NGS experiment is to investigate from which transcripts these sequencing reads originate and to calculate these transcripts' abundances in each sample. Differential gene expression can then be studied by comparing the transcript abundances among conditions.

After quality assessment and preprocessing, the next step in detecting and analyzing gene expression in any RNA-Seq experiment is aligning the short sequencing reads from each sample to a reference genome/transcriptome. This task is called mapping and poses significant computational challenges. Millions of short reads need to be assigned to the correct location in the reference genome/transcriptome, sometimes consisting of billions of bases. This becomes even more challenging if we take into account possible sequencing errors and all the repeated areas that can be present in the reference. Many software tools, called mappers, have been developed to tackle these substantial challenges and to accurately align short reads to the reference (Evans et al., 2009).

One of the most conventionally used mappers is called Bowtie2 (Langmead and Slazberg, 2013). Because Bowtie2 also performs well with reference transcriptomes, this mapper was utilized in the current RNA-Seq experiment. Bowtie2 is a fast aligner that relies on the prior transformation of the reference genome/transcriptome according to the Burrows-Wheeler transform (BWT) algorithm, before aligning the short sequencing reads to this reference. BWT is a data compression algorithm of which the details are beyond the scope of this study. Briefly, the BWT algorithm transforms and comprises the reference genome/transcriptome to a, for Bowtie2, easier-to-navigate representation of this reference. Bowtie2 then aligns the short reads character by character from left to right to the BWT-transformed reference. Each successful character alignment reduces the list of potential target positions for the subsequent read character in the BWT-transformed reference. If Bowtie2 is unable to find a location where a character aligns perfectly, it makes a substitution for this character, which reduces the overall similarity score of the read. Then Bowtie2 continues the search for the next character in the read, and so on. If Bowtie2 is unable to find an appropriate location for a read, this read remains unmapped. Bowtie2 thus subdivides the mapping of a read into smaller search jobs, by aligning one character at a time until the entire read has been aligned. This massively reduces the alignment speed and required computational memory (Trapnell and Salzberg, 2009; Langmead and Slazberg, 2013).

3.2.4. Quantifying gene expression

The following step after mapping the short reads to the reference transcriptome is quantifying the mappings per reference transcript. This way, a transcript profile of all samples is established and comparative analyses can be conducted. Importantly, major complications during quantification are imposed by sequencing reads that do not map uniquely to a single transcript. This happens because genomes and transcriptomes generally contain a certain degree of duplicate sequences, like for example splice variants or repetitive elements. When a read maps to multiple reference sequences, it is impossible to distinguish to which reference transcript this read actually belongs.

How such multireads are processed largely depends on the selected quantification software. The simplest solution is to ignore all multireads, and only take into account the uniquely

mapping reads. However, this way, lots of useful data are discarded and biases are inevitably introduced. Some software packages are able to estimate the abundance of all available reads, including the multireads, by evenly distributing multireads over all their target transcripts. This way, estimated count values for each transcript, instead of true count values, are produced. The software used in this RNA-Seq project, RSEM (RNA-Seq by Expectation Maximization), improves upon this approach, by utilizing an expectation-maximization (EM) algorithm to estimate maximum likelihood expression levels of all available isoforms (Li and Dewey, 2011). Moreover, this software package has the great advantage of being compatible with *de novo* assembled transcriptomes and has the Bowtie2 mapping software built-in.

Furthermore, RSEM also takes into account other variables that are inherently present in RNA-Seq datasets. Especially transcript length distribution and the total amount of sequenced reads per sample (library size) massively influence the read distribution in a sample (Li and Dewey, 2011; Li et al., 2015). As explained in chapter 2, the mRNA content in each sample is fragmented prior to sequencing. Consequently, the amount of reads originating from a transcript not only depends on its expression level, but also on its length. If expression levels are equal, longer transcripts will have relatively more reads mapping back to them than shorter transcripts, positively discriminating longer transcripts (Oshlack and Wakefield, 2009). Another important source of variation in every NGS dataset are differences in sample library sizes. The higher the total amount of sequenced reads in a sample, the more reads are available to map to the reference transcriptome, and hence the higher the estimated transcript abundances. RSEM takes both variables into account when estimating the abundance of transcripts in a sample.

3.2.5. Reference transcriptome preprocessing for mapping and quantification

At the time this experiment was initiated, only an in-house *S. gregaria* whole-body transcriptome was available (Verdonck 2017, unpublished data). Since *S. gregaria* is a non-model organism and no genomic information was yet available, many conventional tools for NGS analyses were unfortunately not applicable. Moreover, the reference transcriptome appeared to contain large amounts of duplicate transcript sequences, which are (quasi) identical sequences originating from *i.a.* splice variants, repetitive elements or redundant sequences. In any NGS experiment, redundant sequences in the reference transcriptome can result in ambiguously mapping sequencing reads, decreasing the power of any downstream analytical test. We decided to utilize RSEM with Bowtie2 to perform the mappings and subsequent quantification of the mapped reads, because these software packages are described to perform well with multireads (Li and Dewey, 2011). Nevertheless, mapping multireads to the appropriate transcripts still remains an ‘estimated guess’, and when the ratio of multireads in a dataset increases, these estimated count values will tend to further deviate from the true count values. This implies that the reliability of the estimated counts decreases as the amount of multireads increases, simultaneously increasing the error rate of any downstream statistical analysis. Therefore, our in-house transcriptome was first preprocessed in order to reduce redundancy, prior to using it as a reference for the mapping of the short RNA-Seq reads. This preprocessed reference transcriptome will be further denoted as the “adjusted whole-body transcriptome”.

The construction of the adjusted whole-body transcriptome was performed in two steps:

1. Generating SuperTranscripts (Trinity)
2. Removing remaining redundancy (CD-HIT-EST)

The tool SuperTranscripts is part of the Trinity RNA-Seq *de novo* assembly software, that was used to assemble the in-house *S. gregaria* whole-body transcriptome (Verdonck 2017, unpublished data). A SuperTranscript is a single linear sequence constructed from collapsing unique and common sequence regions of all available splicing isoforms of a transcript (Davidson et al., 2017). This way, all isoforms of a transcript are combined to one assembled contiguous sequence creating a genome-like version of the transcriptome, as illustrated in figure 3.1 from Davidson *et al.* (2017). Although the newly generated SuperTranscripts do not necessarily represent true biological molecules, no actual transcript data get lost. It is a reversible technique in the sense that the original transcripts that make up the SuperTranscripts can easily be retraced in the original reference transcriptome. Therefore, the SuperTranscripts version of the transcriptome provides a very potent alternative reference for the mapping of RNA-Seq reads.

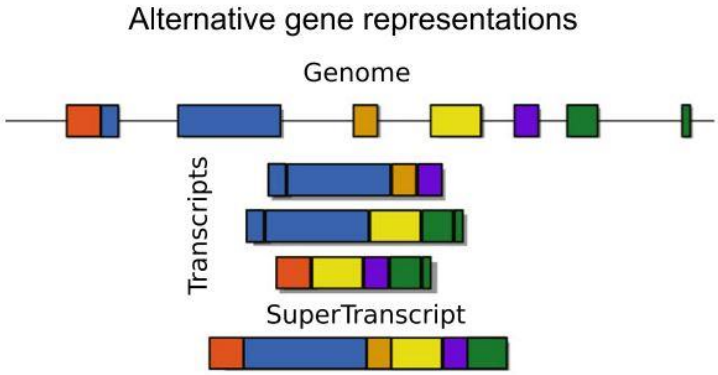


Figure 3.1. Graphic representation of the SuperTranscript synthesis (Davidson et al., 2017). Unique and common sequence regions among different splicing isoforms are collapsed into one SuperTranscipt.

The tool SuperTranscripts uses the typical Trinity identifiers (IDs) to recognize transcript isoforms in a transcriptome assembled by Trinity. An example of such a Trinity ID is 'TR1|c1_g1 _i1', representing contig 'TR1', cluster 'c1', gene 'g1' and gene isoform 'i1'. If this gene or transcript has two predicted isoforms, then another version 'TR1|c1_g1_i2' exists in the transcriptome. The SuperTranscripts tool recognizes these isoforms by their IDs and creates the combined SuperTranscript 'TR1|c1_g1', containing all unique and common sequence information of the original isoform sequences.

After synthesizing SuperTranscripts, the transcriptome was subjected to a final filtering to remove remaining redundant transcripts. Redundant transcripts share such high sequence similarity that they probably have the same origin, resulting in redundant information. CD-HIT-EST is an excellent tool for reducing redundancy in a nucleotide dataset. This program clusters nucleotide sequences based on an *a priori* set similarity threshold, and therefrom selects the best representative sequence (Li and Godzik, 2006; Fu et al., 2012). The similarity threshold was set at 95%, implying that sequences with more than 95% similarity were clustered together into one representative sequence.

3.2.6. Normalization of the estimated count data

The short RNA-Seq reads were mapped to the adjusted whole-body transcriptome using Bowtie2. The subsequent transcript quantification was performed by RSEM. This generated a count table containing all the estimated counts of all the transcripts sequenced in the adjusted

whole-body transcriptome. Next, the estimated counts were normalized. This was especially necessary for the analysis of differential gene expression, presented in chapter 4. However, since the *de novo* annotated midgut reference transcriptome, constructed in the current chapter, would primarily serve as a database for the identification of differentially expressed genes, only those transcripts passing the normalization criteria were retained for further *de novo* annotation. This also massively decreased the computational demands for *de novo* annotation.

Transcripts with very low counts across all the libraries were removed prior to downstream analyses. This has both biological and statistical motives. First, it is reasonable to assume that a minimal concentration of a transcript is needed before it is likely to be translated into a protein or to be biologically relevant. Second, allowing low abundant transcripts to stay as 'passengers' during further downstream analyses, might interfere with some of the statistical approximations. Due to their low abundance, it is also impossible to identify their differential expression between samples. Rare transcripts can therefore be removed from the dataset without losing essential information. One of the most commonly used thresholds for removal of low abundant transcripts is based on their count-per-million (CPM). For the construction of the midgut reference transcriptome, any transcript that did not have a CPM of at least 1 in at least 6 out of the total of 18 samples was considered to be irrelevant and was removed from the dataset.

Subsequently, for normalization, the RSEM count table was imported into the edgeR Bioconductor package, a statistical R package developed for identifying differential expression (Robinson et al., 2009). In edgeR, the estimated count data were normalized according to the package's standard normalization method, based on the Trimmed Mean of M-values (TMM) principle, which takes into account both sequencing depth and library compositions. The general principle of the TMM normalization method is to balance the overall expression levels of transcripts across samples. This is especially important when a small set of transcripts are highly expressed in some, but not all samples. These highly expressed transcripts will consume a large proportion of the total library size, resulting in the under-representation of other transcripts in these samples, which might lead to falsely reported downregulated transcripts. TMM normalization for library composition is achieved by generating a normalization factor for each sample, which rescales its original library size to an effective library size, minimizing differential expression between samples (Robinson and Oshlack, 2010). For example, a normalization factor smaller than one indicates that a small proportion of highly expressed transcripts are biasing the library, resulting in the underrepresentation of the remaining transcripts present in that library. Multiplying the library content by this normalization factor downscales the original library size, resulting in reduced sample variation. The effective library size replaces the original library size for further downstream analyses, such as exploring differential expression (Chapter 4).

The remaining TMM-normalized transcripts were used to calculate differential expression among the three different conditions. This will be elaborated in more detail in chapter 4. Simultaneously, these transcripts were also used to create an annotated *S. gregaria* midgut reference transcriptome for RNA-Seq, which served as a database for identifying the differentially expressed genes in chapter 4.

3.3. Materials and methods

3.3.1. Animal rearing

The optimized rearing conditions are described in detail in chapter 2 (Ch2|2.3.1.2).

3.3.2. Tissue collection

Tissue collection is described in detail in chapter 2 (Ch2|2.3.2.1).

3.3.3. Total RNA extraction

Total RNA extraction of the midgut samples is explained in detail in chapter 2 (Ch2|2.3.2.2). Determination of nucleic acid concentrations and quality control of the RNA samples are described in chapter 2 (Ch2|2.3.2.3).

3.3.4. Illumina® sequencing library preparation and sequencing[1]

Library preparation using the TruSeq® Stranded mRNA Library prep kit (Illumina®) was performed according to the manufacturer's protocol and is explained in detail in chapter 2 (Ch2|2.3.3.1). Libraries were quantified by qPCR, according to Illumina's protocol 'Sequencing Library qPCR Quantification protocol guide', version February 2011. A High Sensitivity DNA chip (Agilent Technologies®) was used to control the library's size distribution and quality. Sequencing was performed on the Illumina® NextSeq 500 platform generating 75 bp single reads and is explained in detail in chapter 2 (Ch2|2.3.3.2).

3.3.5. Data acquisition/quality control/mapping[2]

Per sample, on average 55 ± 10 million reads were generated. First, these reads were trimmed using Cutadapt, version 1.11, to remove the Illumina® adaptor sequence (Martin, 2011). The trimmed reads were mapped against the adjusted *S. gregaria* whole-body transcriptome using bowtie2, version 2.2.5 (Langmead and Salzberg, 2012). The RSEM software, version 1.2.31, was used to generate the count tables (Li and Dewey, 2011).

3.3.6. Construction of the midgut reference transcriptome

After mapping to the adjusted whole-body reference transcriptome, following steps were performed for all samples. (1) Only transcripts with a counts per million (CPM) above 1 in at least six samples were retained. (2) A normalization using Bioconductor software edgeR's standard TMM normalization method (Robinson et al., 2009). Transcripts passing the cut-off

[1] Illumina® sequencing was performed at the sequencing facility NXTGNT (Ghent University, Ottergemsesteenweg 460, B-9000 Ghent, Belgium).

[2] Data acquisition/quality control/mapping were performed at the sequencing facility NXTGNT (Ghent University, Ottergemsesteenweg 460, B-9000 Ghent, Belgium).

criteria were used to filter the adjusted whole-body reference transcriptome to only retain those transcripts present in the midgut.

3.3.7. Annotation of the midgut reference transcriptome

Transcripts composing the midgut reference transcriptome were annotated using Trinotate (http://trinotate.github.io), a protocol and toolkit for *de novo* assembled transcriptomes designed by Bryant *et al.* (2017). Trinotate assists in matching the unknown transcriptome sequences to databases of known sequences, retrieving functional information. In this study, *de novo* annotation was performed in following steps:

1. Nucleotide sequences were translated to the best candidate coding sequences (CDS) by Transdecoder.
2. Transcripts and their predicted CDS were used as query for BLASTx and BLASTp in UniProtKB/Swiss-Prot. These BLASTs also generated GO, eggNOG and KEGG terms.
3. Transcripts and their predicted CDS were used as query for BLASTx and BLASTp in the non-redundant arthropod protein database (NCBI).
4. Coding regions were scanned (HMMER) in the Pfam database.
5. Coding regions were scanned for potential signal peptide sequences with SignalP.
6. Coding regions were scanned for potential transmembrane domain sequences with TMHMM.
7. Data were organized in a tab-delimited file using Trinotate.

Transdecoder (http://transdecoder.github.io) was first used to translate the nucleotide sequences to their predicted best CDS. Transdecoder incorporates BLAST hits information to help identify the best candidate open reading frame (ORF). Next, all transcript sequences and their predicted CDS were used as query for BLAST in the annotated UniProtKB/Swiss-Prot databases. Standard parameters for BLAST were used, with BLAST e-value cut-off set at 1e-5. The hits from these homology searches were used to provide the functional annotations of Gene Ontology (GO), Clusters of Orthologous Groups (COG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) terms. The predicted coding regions were further used to annotate domain content (HMMER scan in Pfam database), signal peptides (SignalP), and transmembrane domains (TMHMM). Additionally, transcripts and their predicted CDS were used as query for BLAST in the non-redundant arthropod protein database downloaded from NCBI, with BLAST e-value cut-off set at 1e-5. Finally, Trinotate organized all data in a SQLite database, which was extracted to a tab-delimited file. In this file, each row represented a transcript of the midgut reference transcriptome together with all the high-throughput annotation information obtained from the above mentioned searches, such as expected homologs (BLAST and Pfam), GO terms and putative transmembrane regions.

3.3.8. *In silico* analysis of the midgut reference transcriptome

Gene transcript enrichment analyses were performed using GO and COG terms obtained by Trinotate. Transcripts putatively involved in various aspects of digestion were searched by name and Pfam accession number in the midgut reference transcriptome. For completeness, additional BLASTn/tBLASTn searches in the midgut reference transcriptome were performed using known arthropod sequences downloaded from the NCBI taxonomy browser as query. Individual sequence analyses were performed in Jalview. Multiple sequence alignments were performed using MAFFT (Multiple Alignment using Fast Fourier Transform). BOXSHADE

version 3.21 was used for shading multiple sequence alignments. A maximum likelihood phylogenetic tree of the putative *S. gregaria* serine proteases was constructed using a WAG substitution model and tested by the bootstrap method, using 500 replications. All graphs were designed in R, version 3.4.2. (https://cran.r-project.org/).

3.4. Results and discussion

3.4.1. Designing the *S. gregaria* midgut reference transcriptome

3.4.1.1. Trimming and quality control of the sequenced reads

A total of 18 samples, 6 per timepoint, was sequenced on a high throughput Illumina® NextSeq 500 flow cell generating 75 bp single reads. Sequencing was performed twice, generating on average 55 ± 10 million (M) reads per sample (Table 3.2). The generated reads were first subjected to a preprocessing step by Cutadapt, version 1.11, to remove the Illumina® adapter sequences. Next, the overall quality of the trimmed reads was assessed by calculating the Phred scores (Q) of a small subset of the reads (200,000 per sample) of all 18 RNA-seq samples with FastQC. The FASTQC report demonstrated an overall high base calling quality in all RNA-Seq samples. On average, all identified bases in all samples had a Q of 35, indicating a consistent very high base calling quality in every sample (Fig. 3.2).

Figure 3.2. Mean quality scores of clean sequencing reads of all samples. The x-axis represents the position of the base. The y-axis represents the associated Phred score (Q). The collapsed green lines represent all the reads sequenced in all 18 samples.

3.4.1.2. Mapping on an adjusted whole-body transcriptome

In a first attempt, the trimmed sequencing reads were mapped to our in-house whole-body transcriptome using Bowtie2, version 2.2.5. On average, 94-95% of the reads mapped at least once, among which 40-42% mapped uniquely, and 50-53% mapped multiple times (multireads). This was not surprising, since this whole-body reference transcriptome was not *a priori* filtered for *i.a.* splice variants and redundant contigs. The in-house whole-body transcriptome contained 175,066 contigs, of which 140,093 were distinct transcripts (80%) and 34,973 (20%) were transcript isoforms (Verdonck 2017, unpublished data) (Supplementary disk S1). Although the utilized quantification software RSEM is described to perform well for multiread handling and quantification, such high amount of multireads could potentially bias the downstream analysis of the count data (Li and Dewey, 2011). Therefore, the whole-body

reference transcriptome was preprocessed to reduce redundancy and the therefrom resulting amount of multireads.

Two preprocessing steps were conducted to obtain the adjusted whole-body transcriptome. First, SuperTranscripts were created with the SuperTranscript tool of Trinity. Next, the transcriptome was filtered for remaining redundant transcripts using CD-HIT-EST, with the similarity threshold set at 95%. SuperTranscripts re-assembled the whole-body reference transcriptome to ultimately retain the 140,093 distinct transcripts, by collapsing all distinct transcript isoforms into generalized representative SuperTranscripts (Supplementary disk S2). Nevertheless, no sequence information got lost in this step. Subsequently, a total of 16,379 redundant contigs were removed by CD-HIT-EST. The adjusted whole-body transcriptome had a total of 123,714 contigs (Supplementary disk S3). The table below demonstrates the effect of the different preprocessing steps on the reference transcriptome dataset.

Table 3.1 | Summary of the effect of the different transcriptome preprocessing steps on the total number of contigs.

Transcriptome version	# contigs
Reference transcriptome - original	175,066
Reference transcriptome - SuperTranscripts	140,093
Reference transcriptome - SuperTranscripts + CD-HIT-EST	123,714

Next, this adjusted whole-body transcriptome was used as a reference for mapping the clean RNA-Seq reads. On average, slightly less (88-90%) reads mapped at least once to the adjusted whole-body transcriptome. However, this time, 70-73% mapped uniquely to the adjusted whole-body transcriptome, while only 15-20% were multireads. This large reduction of multireads compared to the previous mapping to the original reference whole-body reference transcriptome (50-53% multireads) was expected to significantly enhance the power of the subsequent statistical tests. Detailed statistics of the mappings to the adjusted whole-body transcriptome by Bowtie2 are presented in Table 3.2.

Table 3.2 | General mapping statistics of the 18 RNA-Seq samples on the adjusted whole-body transcriptome by Bowtie2.

Sample	% mapped uniquely	% multimapped	% unmapped	Total amount of reads
1	72.5	16.5	10.9	47.4 M
2	70.8	18.2	11	55.8 M
3	71.5	17.4	11.1	45.9 M
4	71	18	11	45.9 M
5	72.4	16.4	11.2	50.5 M
6	69.8	19.6	10.6	43.6 M
7	71	18	11	51.8 M
8	69.8	19.2	11	61.7 M
9	71.8	17.2	11	78 M
10	70.8	18.2	11	65.4 M
11	71.1	17.7	11.2	72.2 M
12	70.8	18.3	11	48.7 M
13	72.6	16.2	11.2	57.1 M
14	73.1	15.4	11.6	52 M
15	73	15.8	11.2	55.9 M
16	71.8	16.9	11.4	48.7 M
17	72.9	15.7	11.4	42.9 M
18	73.4	15.3	11.4	57.9 M

Additional mappings were performed on the available genome assemblies of a gregarine species (*Gregarina niphandrodes*), *B. oleracea* and *Escherichia coli* to check for any genomic contamination. None of the sequencing reads mapped significantly to these databases, demonstrating the overall purity of the sequenced mRNA pool.

RSEM was used to estimate the transcript abundances in all 18 RNA-Seq samples. The generated output file was a table of estimated counts, with rows corresponding to transcripts and columns to samples (Supplementary disk S4). This count table was used to (1) design the annotated *S. gregaria* midgut reference transcriptome, and to (2) identify differential expression across the different conditions (Chapter 4). The design of the midgut reference transcriptome will be further described below.

3.4.1.3. Sequencing coverage

The sequencing coverage was calculated before and after mapping to the adjusted whole-body transcriptome. None of the sample libraries were saturated when the amount of distinct reads was plotted in function of the total amount of sequenced reads (Fig. 3.3 A). This indicated that extra sequencing could lead to the sequencing of more distinct reads. However, this clearly changed after mapping the sequenced reads to the adjusted whole-body transcriptome. All sample libraries were slowly saturating when the amount of distinct transcripts (genes) was plotted in function of the amount of sequenced reads (Fig. 3.3 B). This indicated that most transcripts present in the midgut were sequenced at least once during the two rounds of Illumina® sequencing. These observations supported the idea of designing an annotated midgut reference transcriptome from these data, because this was expected to contain the majority of transcripts present in the *S. gregaria* midgut.

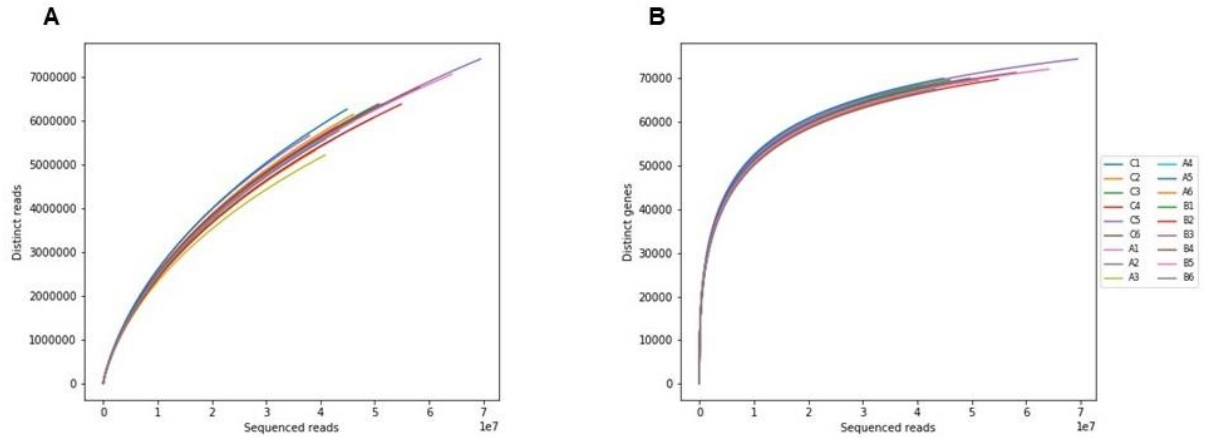


Figure 3.3. (A) The amount of distinct sequenced reads (y-axis) plotted in function of the total amount of sequenced reads (x-axis). **(B)** The amount of distinct transcripts (genes) (y-axis) plotted in function of the total amount of sequenced reads (x-axis). Colored lines in A and B represent all sequenced samples of the RNA-Seq experiment: A1-A6 is 10 minutes after feeding, B1-B6 is 2 hours after feeding and C1-C6 is 24 hours after feeding.

3.4.1.4. Construction of the midgut reference transcriptome

First, all transcripts that did not have a CPM of at least 1 in at least 6 out of 18 samples were removed. Second, the estimated count data were normalized according to edgeR's standard TMM normalization method (Supplementary disk S5). The transcripts retained after filtering

were utilized to design an annotated *S. gregaria* midgut reference transcriptome for RNA-Seq analyses. This final selection of transcripts contained 19,345 midgut-associated transcripts (Table 3.3) (Supplementary disk S6). Among these transcripts; 5,732 were SuperTranscripts constructed from at least two transcript isoforms, while the remaining 13,713 were unique transcripts.

Table 3.3 | General statistics of the *S. gregaria* midgut reference transcriptome.

Total number of contigs	19,345
Total contig length (nt)	33,359,723
Max contig length (nt)	20,472
Mean contig length (nt)	1,724
Contig N50 length (nt)	2,375
Contig N90 length (nt)	904
GC content (%)	38.9

3.4.1.5. Annotation of the midgut reference transcriptome

Transdecoder (http://transdecoder.github.io) was utilized to translate transcript nucleotide sequences in the midgut reference transcriptome to their predicted best CDS. A total of 10,536 transcripts had at least one predicted ORF (Supplementary disk S7). The midgut reference transcriptome was then annotated by matching the unknown transcriptome nucleotide sequences and their predicted best CDS to known databases. A total of 8,659 contigs retrieved a BLASTx hit, while 7,990 coding regions retrieved a BLASTp hit in the UniProtKB/Swiss-Prot database, with e-value cut-off set at 1e-5. Additionally, 10,660 contigs retrieved a BLASTx hit in the non-redundant arthropod protein database (NCBI), with e-value cut-off set at 1e-5. Searching the UniProtKB/Swiss-Prot database also provided GO, COG and KEGG functional annotations. The transcriptome annotation is summarized in Table 3.4. Altogether, a total of 10,939 transcripts had at least one functional annotation, representing 56.5 % of the total midgut reference transcriptome. The *de novo* annotated *S. gregaria* midgut reference transcriptome is available in supplementary disk S9.

Table 3.4 | Number of transcripts in the *S. gregaria* midgut reference transcriptome with at least one hit in the associated databases.

Annotation	**# Transcripts**
BLASTx UniProtKB/Swiss-Prot	8,659
BLASTp UniProtKB/Swiss-Prot	7,990
BLASTx non-redundant arthropod protein database (NCBI)	10,660
Pfam	7,656
GO	8,645
COG	7,142
KEGG	8,448

Consequently, a total of 8,406 transcripts (43.5%) of the midgut reference transcriptome did not retrieve any functional annotation. In this study, *de novo* annotation of the transcripts was primarily based on BLASTs, which searches for regions of similarity between sequences of numerous species. Identification of sequences was therefore strongly based on assumed functional and evolutionary relationships. It is possible that large amounts of transcripts in the *S. gregaria* midgut are too divergent from sequences in other, even closely related, species. Moreover, it is also putative that certain transcripts were specific to the desert locust and their

specialized functions need to be discovered first. Interestingly, high numbers of genes without a functional annotation record were also observed in *de novo* annotations of gut transcriptomes of other insect species, possibly indicating high evolutionary diversification of insect midgut-associated genes (Pedra et al., 2003; Ma et al., 2012; Ye et al., 2014; Liu et al., 2016; Qian et al., 2016; Spit et al., 2016a; Wang et al., 2016; Gazara et al., 2017; Huang et al., 2017; Dhania et al., 2019). These findings clearly demonstrate the need for more functional research of gene products in insects or other less investigated organisms. Additionally, for completeness, other databases including those of non-coding (nc) RNAs were searched, but no significant hits were retrieved.

The length distribution of all contigs and their predicted CDS in the midgut reference transcriptome are presented in figure 3.4. The mean contig length was 1,724 nucleotides, while the mean CDS length was 315 amino acids.

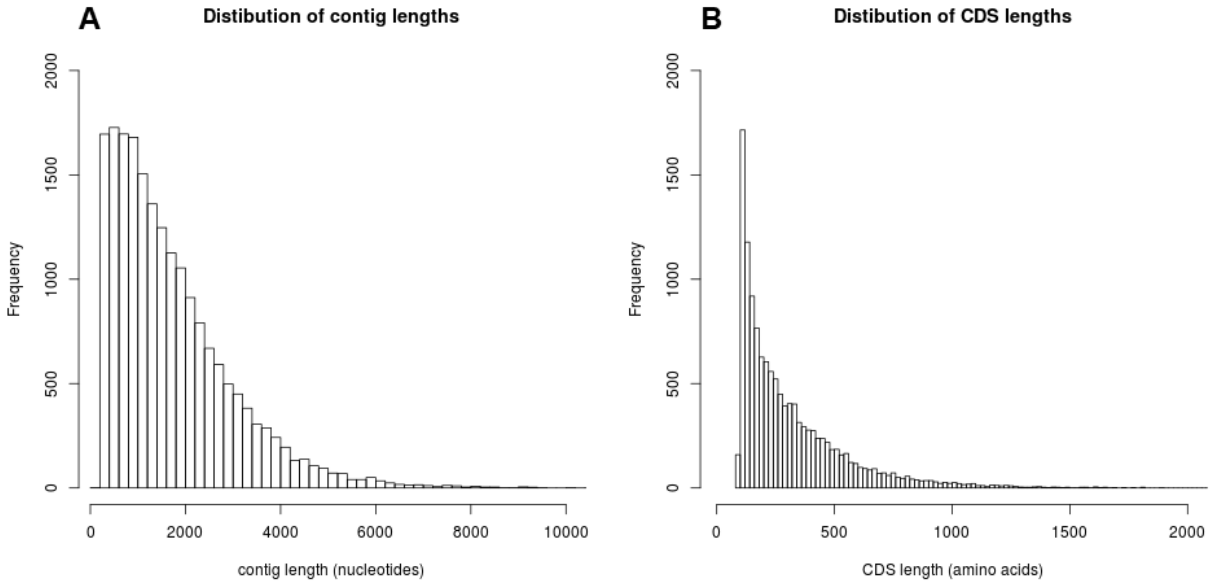


Figure 3.4. (A) Contig length distribution in midgut reference transcriptome. Mean contig length is 1,724 nucleotides. **(B)** Length distribution of predicted coding sequences (CDS) in midgut reference transcriptome. Mean CDS length is 315 amino acids.

3.4.2. Exploring the *S. gregaria* midgut reference transcriptome

3.4.2.1. General statistics of the midgut reference transcriptome

Trinotate retrieved Gene Ontology (GO) and Clusters of Orthologous Groups (COG) based functional annotations of the transcripts from the Swiss-Prot database. These are two widely used controlled vocabularies connecting transcripts to predicted functions, which are typically used to discover common characteristics in omics datasets. This information can be displayed into several graphs or tables to facilitate and summarize the large amount of information present in such datasets.

3.4.2.1.1. Gene Ontology functional classification

The Gene Ontology database is world's largest database containing generalized information on functions of genes among species (http://geneontology.org/) (Ashburner et al., 2000). This

publicly available database contains so-called GO terms, which are species-independent functional annotations of gene products based on three categories:

1. Cellular component: location in which the gene product performs its function.
2. Molecular function: activities of the gene product at a molecular level.
3. Biological process: involvement of the gene product in specific operations or molecular events with a defined beginning and end.

In this study, GO assignments retrieved from the Swiss-Prot database were used to categorize annotated transcripts according to predicted functions based on the three GO categories. A total of 22,961 GO terms could be assigned to 8,645 distinct transcripts (44.6%) of the midgut reference transcriptome. These GO terms were further divided into 7,443; 7,837 and 7,681 terms representing biological process, cellular component, and molecular function, respectively (Fig. 3.5).

All three GO categories displayed predictable gene enrichment patterns. The GO category ‘biological process’ indicated that the annotated transcripts are predicted to mediate a wide range of biological processes, including cellular, metabolic and regulatory processes. In the GO category ‘cellular component’, the most abundant subcategories were ‘cell’ and ‘cell part’ (7,098 and 7,079 transcripts, respectively). Additionally, also a high number of transcripts were assigned to the subcategories ‘membrane’ (3,274 transcripts) and ‘extracellular region’ (1,022 transcripts). Altogether, these functions can include intra- and extracellular enzymatic activity, nutrient absorption across the epithelial membrane and intracellular processing of molecules. In the GO category ‘molecular function’ the majority of transcripts, 5,947 and 4,090, respectively, were assigned to the subcategories ‘binding’ and ‘catalytic activity’, which can be clearly associated with the digestive activity of the midgut.

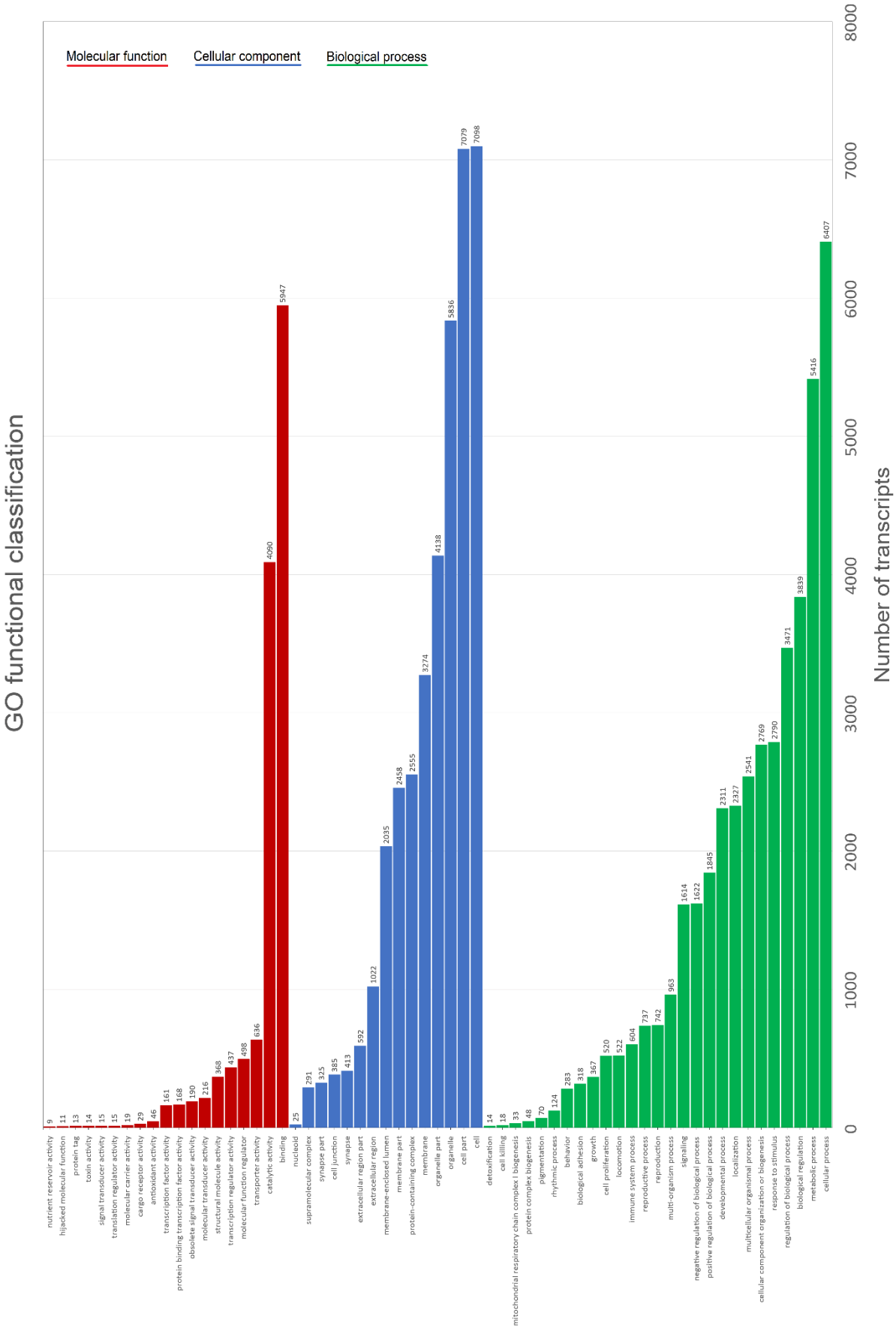


Figure 4.5. Gene Ontology (GO) classification of nucleotide sequences in the *S. gregaria* midgut reference transcriptome.

3.4.2.1.2. COG functional classification

Clusters of orthologous groups (COG) is a publicly available database containing information about functions of genes based on orthologous relationships across prokaryotes and eukaryotes (https://www.ncbi.nlm.nih.gov/COG/). COG provides an alternative way of clustering transcripts in a dataset based on predicted orthologous relationships. Functional groups within the COG database are classified according to the letters of the alphabet (Tatusov, 2000; Tatusov et al., 2003; Galperin et al., 2015). In this study, COG assignments were retrieved from the Swiss-Prot database and were used to cluster transcripts according to predicted orthology. A total of 7,142 transcripts (36.9%) of the midgut reference transcriptome could be attributed to at least one COG category (Fig. 3.6).

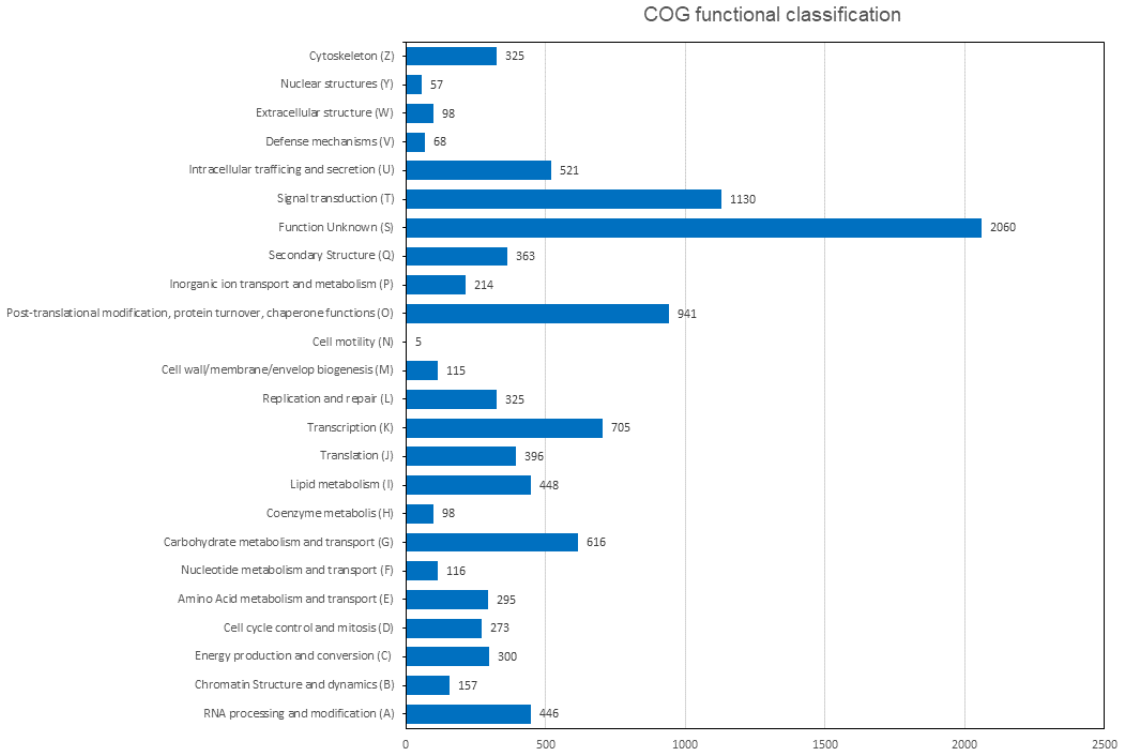


Figure 3.6. COG functional classification of coding sequences in the *S. gregaria* midgut reference transcriptome based on the COG alphabet.

Function unknown (S) was by far the largest represented COG group (2,060; 20.5%), followed by signal transduction (T) (1,130; 11.2%), and post-translational modification, protein turnover and chaperone functions, respectively (O) (941; 9.3%). High numbers of orthologs with unknown function were also observed in the gut transcriptomes of other insect species (Gazara et al., 2017). Other categories contained less than 1,000 transcripts. Categories with clear functions in digestion and nutrient uptake were also well-represented: carbohydrate metabolism and transport (G) (616; 6.1%), lipid metabolism and transport (I) (448; 4.5%), and amino acid metabolism and transport (E) (295; 3%).

3.4.2.2. Exploring the midgut reference transcriptome

The *S. gregaria* midgut reference transcriptome for RNA-Seq was initially designed as a reference for studying differential expression in the midgut during the digestive process, which will be explained in detail in chapter 4. However, this transcriptome is at the same time a broad

database containing midgut-derived transcripts, which could be used as a tool for investigating transcript enrichment in the midgut of *S. gregaria*. The following paragraphs will provide an overview of transcripts identified in the midgut reference transcriptome that are putatively involved in important intestinal processes, such as proteolytic digestion, nutrient absorption and detoxification of xenobiotics.

3.4.2.2.1. Proteolytic digestion

Proteins comprise a substantial part of the daily diet of the desert locust. Numerous protease encoding transcripts were identified in the midgut reference transcriptome, of which the majority code for serine proteases. This is in line with previous research conducted in our lab by Spit *et al.* (2012), who demonstrated that the majority of proteolytic activity inside the *S. gregaria* midgut could be attributed to serine protease activity, namely from trypsins and chymotrypsins (Spit et al., 2012). Serine proteases have neutral pH optima, and their activity is therefore highly favored in the neutral midgut luminal environment of *S. gregaria*. Additionally, this study by Spit *et al.* also revealed minor cysteine and carboxypeptidase activity in the midgut of this insect. In parallel, several transcripts coding for cysteine and aspartic proteases, as well as for amino- and carboxypeptidases are present in the midgut reference transcriptome (Table 3.5).

Table 3.5 | Predicted proteases with annotation in the *S. gregaria* midgut reference transcriptome.

Predicted identity	**# Transcripts**
Endopeptidase activity	
Serine protease	
Trypsin	84
Chymotrypsin	22
Elastase	6
Cysteine protease	9
Aspartic protease	3
Metalloprotease	8
Exopeptidase activity	
Aminopeptidase	47
Carboxypeptidase	32
Dipeptidase	3

The high numbers of proteolytic enzymes in Table 3.5 were not surprising. Comparable high amounts of protease transcripts were also discovered in the gut transcriptomes of other insect species, including *T. castaneum, B. mori, P. americana, C. suppressalis, H. virescens, C. tremulae* and *P. xylostella* (Oppert et al., 2009; Pauchet et al., 2009; Ma et al., 2012; Lin et al., 2015; Popham et al., 2015; Yang et al., 2016; Zhang et al., 2016). These broad sets of protease transcripts present in the intestinal tracts of these insects have already been attributed to their specific life history and associated interactions with different plant defensive mechanisms (Zhu-Salzman and Zeng, 2015).

In a previous research in the host lab, a total of 95 putative serine protease sequences were identified in the LocustDB expressed sequence tag (EST) database (for *L. migratoria*) (Ma et al., 2006; Spit et al., 2014, 2016a). After careful inspection, this number was reduced to 20 unique serine protease encoding sequences, among which 6 and 13 were predicted to encode putatively active trypsins and chymotrypsins, respectively. These were later extensively

investigated by *in vivo* tissue distribution profiling and differential expression analysis as part of the protease inhibitor (PI) induced response (Spit et al., 2014, 2016b).

The *S. gregaria* midgut reference transcriptome also contains a large number of trypsins and chymotrypsins (106 in total). Analogous to the research by Spit et *al.*, their predicted amino acid sequences were carefully investigated *in silico*. This *in silico* examination was largely based on Perona and Craiks' description of the general structure of serine proteases, as well as on both *in silico* and *in vivo* studies performed in several insect species (Perona and Craik, 1995; Hedstrom, 2002; Lopes et al., 2004; Ge et al., 2012; Spit et al., 2014; Cao et al., 2015). Digestive serine proteases in insects are members of the S1 protease family, which are characterized by a catalytic triad consisting of a histidine, aspartate and serine (His57- Asp102- Ser195; bovine chymotrypsin numbering system). The catalytic triad is responsible for the correct proteolytic functioning of a serine protease. The members of the catalytic triad are generally part of more or less conserved amino acid motifs: TAA**H**$_{57}$C, **D**$_{102}$IA and GD**S**$_{195}$GG. Whether a serine protease exhibits trypsin or chymotrypsin activity largely depends on the identity of the amino acid residue at position 189, located at the base of the substrate binding pocket. In trypsin, an aspartate at position 189 exhibits high specificity towards positively charged substrate residues, while in chymotrypsin, a serine, glycine or alanine residue at position 189 exhibits high specificity towards large, neutral substrate residues (Perona and Craik, 1995). Most serine proteases are synthesized as inactive zymogens to be activated upon cleavage of a specific peptide bond. In insects, a strongly conserved RIVGG amino acid motif near the N-terminal site of the protein represents the activation cleavage site of serine proteases (Muhlia-Almazán et al., 2008; Cao et al., 2015).

According to the above described typical trypsin and chymotrypsin features, a total of 8 complete putative trypsin and 21 complete putative chymotrypsin encoding transcripts were retained from a total of 106 putative trypsin and chymotrypsin sequences identified in the midgut reference transcriptome for RNA-Seq[3]. Only those transcripts with a complete catalytic triad and a correct substrate determining amino acid at position 189 were included. Nevertheless, this list of complete trypsin and chymotrypsin coding transcripts does not exclude proteolytic activity of any of the other identified serine proteases in the midgut reference transcriptome. These 29 proteases were solely selected based on sequence predictions. The conserved motifs of the retained complete trypsin and chymotrypsin coding transcripts are presented in table 3.6. A MAFFT multiple sequence alignment of the trimmed amino acid sequences of these putative proteases and an unrooted maximum likelihood tree (500 bootstrap repeats) can be found in the supplementary data (Supp. Fig. S3.1 and Supp. Fig. S3.2).

[3] Among these 29 transcripts, 11 are SuperTranscripts: TR42553|c0_g1 (3), TR45971|c0_g1 (3), TR55871|c1_g3 (7), TR55943|c3_g1 (4), TR56617|c1_g6 (5), TR62847|c0_g1 (2), TR64159|c0_g1 (2), TR68072|c0_g6 (6), TR69019|c0_g1 (3), TR82999|c0_g1 (4), TR86989|c0_g1 (2). Between parentheses is their number of originating transcript isoforms. Interestingly, when analyzing all the original transcript isoforms of these trypsins and chymotrypsins, the exact same 29 unique amino acid sequences as retrieved from the midgut reference transcriptome are obtained. This can be explained by the high overall nucleotide sequence similarity (> 85%) of the transcript isoforms making up the different SuperTranscripts. Since SuperTranscripts are constructed by collapsing unique and common sequence regions of all available isoforms of a transcript, most of these SuperTranscripts indeed strongly resemble their originating transcript isoforms. Consequently, the CDS of these SuperTranscripts are similar to the CDS of their originating transcript isoforms. The SuperTranscript IDs were further used to denote the different trypsin and chymotrypsin transcripts of the *S. gregaria* midgut.

Table 3.6 | Conserved motifs of predicted trypsin and chymotrypsin sequences in the *S. gregaria* midgut reference transcriptome.

Transcript	Sequence length (AA)	Trypsinogen activation site	TAAHC motif	DIA motif	GDSGG motif	AA189
			Predicted trypsins			
TR49755|c0_g1	288	RIVGG	TAGHC	DIG	GDSGG	D
TR50818|c0_g1	282	RIVNG	TAAHC	DIG	SDSGG	D
TR55943|c3_g1	296	RIVGG	SAAHC	DIC	GDSGG	D
TR60258|c0_g1	518	LIVGG	TAAHC	DIA	GDSGG	D
TR62847|c0_g1	747	RIVGG	SAGHC	DIT	GDSGG	D
TR68072|c0_g6	255	RIVGG	TAAHC	DIG	GDSGG	D
TR79895|c0_g1	227	RIVGG	SAAHC	DIC	GDSGG	D
TR98054|c0_g1	254	RIVGG	TAAHC	DIG	GDSGG	D
			Predicted chymotrypsins			
TR45971|c0_g1	234	RIVGG	TAAHC	DIG	GDSGG	G
TR53843|c0_g1	212	RIVGG	SAAHC	DVA	GDSGG	G
TR60303|c0_g2	261	RIIGG	TAGHC	DIA	GDSGG	G
TR63225|c0_g1	254	RIVGG	TVTHC	DIA	MDSGS	G
TR65593|c0_g1	259	RIVGG	TAAHC	DVA	GDSGS	G
TR91215|c0_g1	284	RIIGG	TAAHC	DVA	GDSGG	G
TR60658|c0_g1	252	RIVNG	TAAHC	DIA	GDSGG	G
TR86989|c0_g1	376	RIVNG	TAAHC	DIA	GDSGG	G
TR106489|c0_g1	230	RIIGG	TAAHC	DIA	GDSGS	G
TR42553|c0_g1	295	RIISG	TAAHC	DVA	GDSGG	S
TR54977|c0_g1	210	RIVGG	TAAHC	DIT	GDSGS	S
TR55871|c1_g3	429	RIYGG	TAGHC	DIA	GDSGG	S
TR56617|c1_g6	353	RIIGG	TAAHC	DVA	GDSGG	S
TR64159|c0_g1	248	RIING	TAAHC	DVA	GDSGG	S
TR67531|c0_g1	277	RIYGG	TAGHC	DIA	GDSGG	S
TR69019|c0_g1	273	RITGG	TAGHC	DIA	GDSGG	S
TR82999|c0_g1	273	RIING	TAAHC	DVA	GDSGG	S
TR89528|c0_g1	288	RITNG	TAAHC	DIA	GDSGG	S
TR91381|c0_g1	259	RVIGG	TAAHC	DIA	GDSGG	S
TR96765|c0_g1	257	RIIGG	TAAHC	DIA	GDSGS	S
TR67399|c0_g1	295	RIVGG	TAGHC	DVA	GDSGG	A

RIVGG is the standard activation cleavage site of insect serine proteases. The catalytic triad H_{57}, D_{102}, S_{195} is mostly found in the conserved motifs $TAAH_{57}C$, $D_{102}IA$ and $GDS_{195}GG$. Serine proteases were predicted to exhibit trypsin activity when amino acid 189 was an aspartate (D), or chymotrypsin activity when amino acid 189 was a serine (S), glycine (G) or alanine (A).

Small variations in the conserved motifs were allowed, as long as the essential amino acids of the catalytic triad were present. Similar deviating conserved motifs were also observed in the serine protease amino acid sequences of other insect species (Muhlia-Almazán et al., 2008; Spit et al., 2014; Cao et al., 2015). It is however remarkable that only 29 out of 106 predicted trypsin and chymotrypsin coding transcripts contained all the conserved motifs for molecular recognition and activity (Lopes et al., 2004). Moreover, the current study demonstrated that the other genes encoding serine proteases predicted to be enzymatically inactive (77 in total) were also expressed in the *S. gregaria* midgut; some even at high levels (Supp. Table S3.1). It cannot be fully excluded that a (small) proportion of these sequences originate from assembly artifacts. Therefore, their presence in the midgut as well as their nucleotide sequence should first be confirmed. It is possible that some of these serine proteases predicted to be

proteolytically inactive have acquired other, physiologically relevant functions. Interestingly, one transcript (TR91530|c0_g1) coding for a Ser-protease-related protein (*Scg*-SPRP), has already been cloned from the *S. gregaria* midgut in the host lab in 1998 by Chiou and colleagues (Chiou et al., 1998). This trypsin-like molecule was predicted to be enzymatically inactive, since its derived amino acid sequence lacked two out of three residues indispensable for catalytic activity of serine proteases. *Scg*-SPRP was nevertheless suggested to play a distinct role in the life history of the desert locust, as it could be demonstrated that its occurrence fluctuated during the molting cycle and the female reproductive cycle, and its expression appeared to be stimulated by juvenile hormone (JH). In addition, it has already been hypothesized that herbivorous insects probably produce sets of inactive proteases as a protective mechanism against ingested plant PIs (Christeller, 2005). It is therefore possible that some of the predicted inactive proteases in the *S. gregaria* midgut reference transcriptome are part of an elaborate PI counter defensive mechanism. In conclusion, further *in vivo* analyses are required to determine whether the proteins encoded by these transcripts have proteolytic activity or perform deviating physiological functions.

The RSEM generated estimated count file of midgut transcripts allowed us to analyze the overall abundances of serine proteases at different timepoints after feeding. Table 3.7 displays the mean log2-transformed TMM-normalized CPM of all 29 predicted enzymatically active serine proteases at ten minutes (group A), two hours (group B) and twenty-four hours (group C) after feeding. Three validated *S. gregaria* housekeeping genes are added as a reference (Van Hiel et al., 2009). This table reveals that most serine protease transcripts were ubiquitously present during the different physiological stages of the midgut, even in the absence of food, demonstrating a continuous expression of serine proteases in the *S. gregaria* midgut. Some of these serine proteases however significantly increased in transcript numbers when food was present in the gut. These are indicated in bold in Table 3.7, and will be further elaborated in chapter 4.

Table 3.7 | Log2-transformed mean TMM-normalized CPM of predicted enzymatically active trypsin and chymotrypsin proteases in the *S. gregaria* midgut reference transcriptome at different timepoints after feeding.

Transcript	Predicted activity	10 minutes after feeding	2 hours after feeding	24 hours after feeding
		Predicted trypsins		
TR49755|c0_g1	Trypsin	11.80191245	11.91217595	11.39694373
TR50818|c0_g1	Trypsin	4.498075035	4.426129507	4.770609442
TR55943|c3_g1	Trypsin	7.710081024	7.686827057	6.053431797
TR60258|c0_g1	Trypsin	0.336838301	0.064576848	1.200581373
TR62847|c0_g1	Trypsin	0.296317749	0.228784405	0.759066844
TR68072|c0_g6	Trypsin	2.804149609	2.843115938	2.198689429
TR79895|c0_g1	Trypsin	2.727771444	2.894997684	2.032219884
TR98054|c0_g1	Trypsin	8.885575789	9.28288423	8.580451783
		Predicted chymotrypsins		
TR45971|c0_g1	Chymotrypsin	6.188148895	6.52933214	6.271865761
TR53843|c0_g1	Chymotrypsin	3.78703924	3.770277339	3.160964007
TR60303|c0_g2	Chymotrypsin	3.208410077	3.51902224	2.423411298
TR63225|c0_g1	Chymotrypsin	5.594360049	5.765397282	5.184430013
TR65593|c0_g1	Chymotrypsin	6.446141517	6.724022373	6.027411798

TR91215|c0_g1	Chymotrypsin	10.06581181	10.37970205	9.287313577
TR60658|c0_g1	Chymotrypsin	2.791458488	3.010757813	1.689789089
TR86989|c0_g1	Chymotrypsin	3.550949094	3.753902179	2.632360389
TR106489|c0_g1	Chymotrypsin	2.799227178	3.098074753	1.899243659
TR42553|c0_g1	Chymotrypsin	8.12345437	8.137015175	6.151351628
TR54977|c0_g1	Chymotrypsin	3.850974152	3.833255649	3.180912107
TR55871|c1_g3	Chymotrypsin	7.941876571	7.988247085	7.263511542
TR56617|c1_g6	Chymotrypsin	11.66055628	11.92825072	11.50605986
TR64159|c0_g1	Chymotrypsin	10.9181831	11.16712988	10.79077311
TR67531|c0_g1	Chymotrypsin	6.903773513	6.781295149	6.047236402
TR69019|c0_g1	Chymotrypsin	9.324101777	9.132102634	7.739269606
TR82999|c0_g1	Chymotrypsin	5.545239205	5.596212707	5.427693268
TR89528|c0_g1	Chymotrypsin	8.217567818	8.269140774	6.357195296
TR91381|c0_g1	Chymotrypsin	11.04311203	11.37790491	10.62128004
TR96765|c0_g1	Chymotrypsin	3.855135974	4.017078204	2.779896814
TR67399|c0_g1	Chymotrypsin	4.72356393	4.504248334	3.179283838
		Schistocerca gregaria housekeeping genes		
TR69021|c0_g6	Actin	11.50379434	11.65481405	11.12516027
TR68374|c0_g1	EF-1α	12.07511839	11.99933504	12.20399421
TR64652|c0_g1	GAPDH	9.076841672	9.046443973	8.97755426

In bold are the serine proteases that were significantly upregulated two hours after feeding compared to twenty-four hours after feeding (see chapter 4).

Very high numbers of amino- and carboxypeptidase encoding transcripts are observed in the *S. gregaria* midgut reference transcriptome. Exopeptidases are generally described to cleave amino acid residues from the amino- or carboxyl-terminus of proteins, respectively, and are categorized as either digestive or regulatory peptidases. As digestive enzymes, they are mostly described to be active in the ectoperitrophic space, where they mediate the final steps of the proteolytic digestion generating free peptides to be absorbed by the midgut ECs. A total of 47 aminopeptidases were identified in the *S. gregaria* midgut reference transcriptome, of which 29 are predicted to have intracellular activity, 13 are predicted to be plasma membrane-bound, and 11 are predicted to have extracellular activity. The majority of identified aminopeptidases (29 in total) in the *S. gregaria* midgut belongs to the M1 peptidase family. This family of peptidases is present in all insect species and is mainly expressed in the gut, where they play important roles in protein digestion (Terra and Ferreira, 1994). Most insect digestive carboxypeptidases have been classified as carboxypeptidases A or B, belonging to the M14 peptidase family (Terra and Ferreira, 1994). A total of 32 putative carboxypeptidases were identified in the *S. gregaria* midgut reference transcriptome, of which 13 belong to the M14 peptidase family. These are all predicted to be active extracellularly, probably mediating protein digestion in the midgut lumen. Amino- and carboxypeptidases with predicted intracellular activity could be performing essential cellular functions, such as regulating protein turnover or the immune response. Additionally, three transcripts encoding putative membrane-bound dipeptidases (Pfam01244) were identified. These enzymes are described to mediate the final cleavage of dipeptides to amino acids, which can then be absorbed by the midgut epithelial cells.

3.4.2.2.2. Other digestive enzymes

The plant-based diet of the desert locust is also rich in sugars and, to a lesser extent, fats. The main sources of dietary carbohydrates for leaf-eating herbivorous insects are starch, cellulose

and sucrose, while the majority of dietary fats originate from phospholipids, fatty acids and sterols. Numerous transcripts encoding specialized digestive enzymes able to degrade these nutrients could be identified in the *S. gregaria* midgut reference transcriptome (Table 3.8).

Table 3.8 | Predicted digestive enzymes mediating carbohydrate and lipid digestion with annotation in the *S. gregaria* midgut reference transcriptome.

Predicted identity	**# Transcripts**
Sugar digestion	
α-Amylase	6
α-Glucosidase	20
β-Glucosidase	44
Cellulase	3
Lipid digestion	
Lipase	22
Phospholipase	13
Sterol dehydrogenase/reductase	29

Table 3.8 shows that the *S. gregaria* midgut contains large quantities of transcripts encoding putative carbohydrate and lipid degrading enzymes. This possibly originates from its generalist herbivorous lifestyle: in order to more efficiently digest its diverse diet, the desert locust is equipped with a broad arsenal of digestive enzymes. A plant-based diet also implies encountering different plant defensive molecules, such as carbohydrase inhibitors. Therefore, a proportion of these digestive enzymes is probably involved in specific counter defensive mechanisms (Zhu-Salzman and Zeng, 2015; Holtof et al., 2019). Broad repertoires of digestive enzymes are also frequently observed in other insect species, analogous to the earlier described high enrichment of proteolytic enzymes (Oppert et al., 2009; Pauchet et al., 2009; Ma et al., 2012; Popham et al., 2015; Zhu-Salzman and Zeng 2015; Spit et al., 2016a; Yang et al., 2016; Zhang et al., 2016; Da Lage, 2018). Our data are in line with the general observation that insects deploy divergent sets of enzymes to mediate enzymatic digestion of food.

Large numbers of the most commonly found types of carbohydrases were present in the *S. gregaria* midgut transcriptome: 6 α-Amylase transcripts (Pfam00128), 20 α-Glucosidase transcripts (Pfam01055), and 44 β-Glucosidase transcripts (Pfam00232). Interestingly, three transcripts coding for enzymes putatively involved in the digestion of cellulose were identified in the midgut reference transcriptome. This is in accordance with the more recent findings that suggest that some insects exhibit an endogenous cellulase activity (Holtof et al., 2019). Several in-depth studies have demonstrated that this cellulase activity could be assigned to proteins belonging to the glycoside hydrolase (GH) family of endoglucanases (Kunieda et al., 2006; Watanabe and Tokuda, 2010; Shelomi et al., 2014). Members of the GH9 endoglucanases family (Pfam00759) have been identified in all insect orders with reported cellulase activity. The three cellulase encoding transcripts identified in the *S. gregaria* midgut reference transcriptome all belong to the GH9 family.

The dietary uptake of lipids generally seems to be less essential for insects compared to the uptake of proteins and carbohydrates, mostly because insects are capable of endogenously synthesizing many fatty acids and phospholipids from dietary carbohydrates. Nevertheless, all insects do require some dietary absorption of PUFAs and sterols to be applied as structural components of the cell membrane, secondary metabolites or as starting materials for steroid synthesis, respectively (Zibaee et al., 2014). Moreover, in *S. gregaria,* adequate dietary supplementation of fatty acid energy reserves is essential, as these are the primary energizers

for long distance flights (Haunerl and Chisholm, 1990). Several lipases (22 in total) and phospholipases (13 in total) were identified in the *S. gregaria* midgut reference transcriptome. The availability of a broad set of lipid hydrolases possibly contributes to the efficient breakdown of dietary lipids, and is in line with findings from other insect species, including *D. melanogaster, A. gambiae, A. mellifera, B. mori* and *T. castaneum* (Horne et al., 2009).

Insects are unable to biosynthesize sterols *de novo*, making their dietary uptake of sterols vital (Chapman, 2013). The dominant sterol in insects is cholesterol. Herbivorous insects acquire cholesterol from plant phytosterols via intermediate dealkylation pathways in their gut (Ciufo et al., 2011). A total of 29 putative sterol dehydrogenases/reductases (Pfam00106) were identified in the *S. gregaria* midgut. The use of sterol dehydrogenases and reductases converting dietary phytosterols to cholesterol has already been observed in several other phytophagous insect species. it is possible that some of the sterol dehydrogenases/reductases identified in the *S. gregaria* midgut serve a similar function (Svoboda, 1999; Ciufo et al., 2011; Chapman, 2013).

3.4.2.2.3. Nucleic acid digestion

The diet of most insects also contains large quantities of nucleic acids. Therefore, different nucleases are present in the gut to mediate nucleic acid breakdown (Katoch and Thakur, 2012). These enzymes are able to degrade nucleic acids by breaking down their phosphodiester linkages, and are called deoxyribonuclease (DNase) or ribonuclease (RNase) according to their substrate preference, typically being DNA or RNA, respectively (Yang, 2011). A total of 4 putative DNase and 38 putative RNase encoding transcripts were present in the *S. gregaria* midgut reference transcriptome. In addition, 14 putative nucleotidase and 1 putative nucleosidase encoding transcripts could be identified. These enzymes generally catalyze the further breakdown of free nucleotides following the catalytic action of nucleases. It is expected that most of these enzymes play critical roles in cellular nucleic acid metabolism. However, it is possible that some of these also contribute to the dietary breakdown of plant nucleic acids inside the midgut lumen. Further research is needed to determine which of these enzymes are indeed active inside the midgut lumen to assist the degradation of consumed nucleic acids.

3.4.2.2.4. Nutrient absorption

Research on the different mechanisms of nutrient absorption in insects is limited. Most of our knowledge is based on predicted parallels with mammalian systems (Holtof et al., 2019). Based on this information, the *S. gregaria* midgut reference transcriptome was scanned for putative nutrient transporter encoding transcripts belonging to specific Pfam protein families (Table 3.9).

Table 3.9 | Summary of Pfam annotation of predicted nutrient transporters and their substrates present in the *S. gregaria* midgut reference transcriptome.

Substrate	Pfam	# Transcripts	Description
Amino acids	PF01490	17	Solute carrier protein 36
	PF00324	12	Cationic amino acid transporter (SLC7)
	PF00209	7	Sodium coupled amino acid transporter (iNAT-SLC6)
	PF10149	2	Solute carrier protein 38

	PF00854	2	Oligopeptide transporter (SLC15)
Carbohydrates	PF00083	53	Facilitated hexose and polyol transporters (GLUT)
	PF00474	3	Sodium-driven glucose transporters (SGLT)
Lipids	PF00501	24	Fatty acid transport protein (FATP)
	PF00061	15	Fatty acid binding protein (FABP)
	PF01130	7	Scavenger receptors class B type I (CD36)
	PF00108	7	Sterol carrier protein-x (SCPx)
	PF02036	4	Sterol carrier protein-2 (SCP2)
	PF02221	4	Niemann-Pick C2 (NPC2)
	PF16414/ PF02460/ PF12349	1	Niemann-Pick C1b (NPC1b)

Abbreviations: iNAT = insect nutrient amino acid transporter; SLC = solute carrier protein.

Dietary protein absorption

Protein digestion in the midgut lumen results in the accumulation of amino acids that, together with remaining dipeptides and small oligopeptides, are subsequently internalized by the midgut ECs. This absorption is accomplished by specific nutrient transporters present in the midgut epithelial membrane (Chapter 1). The midgut reference transcriptome of *S. gregaria* contains a total of 39 transcripts encoding amino acid transporters, of which 7 and 12 are predicted to belong to the SLC6 (Pfam00209) and SLC7 (Pfam00324) transporter family, respectively. However, the majority of predicted amino acid transporters present in the *S. gregaria* midgut transcriptome belong to the phylogenetically related SLC32, 36 and 38 protein families (Pfam01490). High numbers of this clade of transmembrane amino acid transporters have also been reported in some other insect species (Boudko, 2012). Nevertheless, to the best of my knowledge, there exists only one study reporting midgut nutrient amino acid absorption mediated by a protein of this family. In *A. aegypti*, a SLC36 proton coupled amino acid transporter (*Aea*PAT1) localized in the apical membrane of the midgut and caeca was demonstrated to transport a broad range of amino acids across the midgut epithelium. Moreover, *Aea*PAT1 expression was significantly induced after a blood meal, illustrating its role in the transport of nutrient amino acids during digestion (Evans et al., 2009). The observed high number of SLC36 protein family members (17 in total) in the midgut transcriptome of *S. gregaria* strongly suggests that they might have an analogous function in this insect species.

The few functionally characterized insect nutrient amino acid transporters (iNAT) almost all belong to the solute carrier 6 transporter family (SLC6) (Boudko, 2012; Holtof et al., 2019). Other putative iNATs are believed to belong to the SLC7 transporter family, which includes the cationic amino acid transporters (CATs) and the heterodimeric amino acid transporters (HATs) (Verrey et al., 2004). However, until now, their contribution to amino acid transport across the midgut epithelium in insects has not been experimentally confirmed (Holtof et al., 2019). The identification of several SLC7 transporter encoding transcripts in the midgut transcriptome, strongly suggests their contribution to the uptake of dietary amino acids. Until now, amino acid transporter activity by SLC7 transporters has only been functionally characterized in fat body tissue of *A. aegypti* and *D. melanogaster* (Martin et al., 2000; Colombiani et al., 2003; Reynolds et al., 2009; Hansen et al., 2011). Further *in vivo* analyses are required to better comprehend their function in nutrient amino acid absorption.

Additionally, two putative oligopeptide transporters were identified in the midgut reference transcriptome, both predicted to belong to the SLC15 transporter family (Pfam00324). The

expectation that SLC15 transporters contribute to the absorption of oligopeptides in insects is also based on predicted parallels with vertebrates. Members of the SLC15 transporter family have already been identified in the genomes of several insect species. However, their contribution to the dietary oligopeptide absorption has only been characterized in *D. melanogaster* (Roman et al., 1998; Boudko, 2012).

Dietary carbohydrate absorption

Following the breakdown of dietary carbohydrates into mono- and disaccharides, specific transporters facilitate their absorption into the midgut ECs. It is expected that insects, similarly to mammals, rely on the MFS of glucose facilitators (GLUT2 and GLUT5 of the SLC2 family) and SGLTs (SLC5 family) to mediate this transport of dietary sugars (Chapter 1). A total of 53 transcripts encoding proteins that belong to the SLC2 family (Pfam00083) were identified in the midgut transcriptome, among which 32 transcripts are predicted to encode facilitated glucose transporters (GLUTs) and 21 transcripts are predicted to encode facilitated trehalose transporter 1 (Tret1). Additionally, three transcripts putatively encoding sodium-driven glucose transporters (SGLT) (Pfam00474) were also identified (Table 3.9).

Based on predicted analogies with the mammalian system of dietary carbohydrate absorption, the majority of these GLUTs are expected to be responsible for the uptake of dietary monosaccharides. Furthermore, it has been suggested that Tret1, a known transporter of the disaccharide trehalose in peripheral insect tissues, might be responsible for the uptake of dietary disaccharides from the intestinal lumen (Kikawada et al., 2007; Kanamori et al., 2010). Tret1 transporters have been identified in the guts of various insects, but until now, their putative role in this tissue has not yet been characterized (Holtof et al., 2019). Nevertheless, the observed high numbers of Tret1 encoding transcripts in the *S. gregaria* midgut reference transcriptome further supports the hypothesis that these transporters are involved in the cellular uptake of dietary disaccharides in the insect midgut.

Dietary lipid Absorption

Our knowledge on the receptor mediated transport of dietary lipids across the midgut epithelium in insects is still rudimentary (Chapter 1). The main end products of lipid digestion in the midgut lumen include free fatty acids, glycerol, mono- and diacylglycerol and phospholipids. The majority of these molecules are able to passively diffuse across the membrane. However, their movement might be facilitated by specific binding proteins, such as the fatty acid binding proteins (FABPs). Several transcripts (15 in total) encoding putative FABPs (Pfam00061) were identified in the *S. gregaria* midgut reference transcriptome (Table 3.9). Next to FABPs, 24 transcripts encoding fatty acid transport proteins (FATPs) (Pfam00501) and 7 transcripts encoding the scavenger receptors class B type I belonging to the CD36 protein family (Pfam01330) were identified in the midgut reference transcriptome. Both proteins have already been suggested to interact in the active transport of dietary fatty acids across the midgut membrane, and their high abundance in the midgut transcriptome further supports this (Holtof et al., 2019).

Dietary uptake of sterol in insects is expected to be mediated by Niemann-Pick C proteins (NPC) and sterol carrier proteins (SCP) (Holtof et al., 2019). A total of 4 NPC1 encoding transcripts were identified in the *S. gregaria* midgut reference transcriptome. However, from these 4 transcripts annotated as NPC1, only 1 contained all the conserved domains

(Pfam16414/Pfam02460/Pfam12349) characteristic to NPC1, and was annotated as NPC1b. The remaining 3 NPC1 transcripts contained only one or two NPC1 protein domains. Their sequences need to be further evaluated. Until now, the role of NPC1 in insect sterol homeostasis has only been functionally studied in *D. melanogaster*. In these insects, 2 NPC1 homologs were identified: NPC1b being responsible for the absorption of dietary cholesterol across the midgut epithelium, and NPC1a playing a role in intracellular sterol trafficking in the midgut and peripheral tissues (Huang, 2005; Fluegel et al., 2006; Voght et al., 2007). Recently, the presence of NPC1 homologs was investigated in the genomes of 39 insect species belonging to 10 different orders. It could be demonstrated that most insects species possessed two NPC1 homologs. However, in some, NPC1b was absent, while NPC1a was present in all investigated genomes (Zheng et al., 2018). Further investigation is needed to classify the identified *S. gregaria* NPC1 sequences. In addition, 4 NPC2 (Pfam02221) encoding transcripts were identified in the *S. gregaria* midgut reference transcriptome. As of yet, NPC2 has only been studied in *D. melanogaster*. In this insect species, eight NPC2 proteins (NPC2a-h) were identified, but only NPC2a is highly expressed in the midgut and seems to play a minor role in sterol homeostasis (Huang et al., 2007; Shi et al., 2012). Altogether, more research is needed to elucidate the exact roles of the identified NPC1 and NPC2 encoding transcripts in the *S. gregaria* midgut, especially regarding their function in the uptake of dietary sterols.

Additionally, 7 putative SCPx (Pfam00108) and 4 putative SCP2 (Pfam02036) encoding transcripts were identified in the midgut reference transcriptome. SCPs are suggested to mediate sterol uptake in the insect intestinal tract. In general, SCP genes give rise to two different, functionally active sterol carrier proteins: SCPx (sterol carrier protein-x), and SCP2 (sterol carrier protein-2) (Ohba et al., 1994). In vertebrates, SCPx is characterized by a 3-oxoacyl-CoA thiolase domain (Pfam00108), and is expected to be involved in the oxidation of branched chain fatty acids, while SCP2 is involved in the transfer of lipids (Kim and Lan, 2010). Both gene products have been identified in the guts of several insect species, including *B. mori, S. littoralis* and *D. melanogaster* (Kitamura et al., 1996; Takeuchi et al., 2004; Gong et al., 2006). However, further *in vivo* studies are required to fully elucidate their role in the dietary sterol uptake in insects.

3.4.2.2.5. Other digestive process related transcripts

Finally, the midgut reference transcriptome was scanned for transcripts generally known to contribute to the digestive physiology of herbivorous insects. Table 3.10 provides an overview of these transcripts and their predicted functions. Transcripts of interest included those that were predicted to contribute to PM establishment, transmembrane trafficking and xenobiotic metabolism.

Table 3.10 | Pfam annotation of other important midgut associated proteins present in the *S. gregaria* midgut reference transcriptome.

Category	Pfam	# Transcripts	Description
Peritrophic membrane	PF01607	59	Chitin binding peritrophin A domain
	PF00379	11	Chitin binding domain
	PF00704	13	Endochitinase
Transmembrane proton pump	PF00005	55	ABC transporter NBD
	PF00664	36	ABC transporter TMD
	PF02874	4	H^+ V-ATPase V_1 subunit A
	PF00137	4	H^+ V-ATPase V_1 subunit C

	PF01813	2	H^+ V-ATPase V_1 subunit D
	PF01991	1	H^+ V-ATPase V_1 subunit E
	PF01990	1	H^+ V-ATPase V_1 subunit F
	PF03179	1	H^+ V-ATPase V_1 subunit G
	PF11698	1	H^+ V-ATPase V_1 subunit H
	PF01496	5	H^+ V-ATPase V_0 subunit a
Detoxification	PF00067	132	CYP450
	PF00201	63	Uridine diphosphate-glucuronosyltransferase (UGT)
	PF02798/13417	59	Glutathione S-transferase (GST)
	PF00135	123	Carboxylesterase (CE)
Others	PF00230	4	Aquaporin
	PF00999	4	Sodium/hydrogen exchanger
	PF00689/00690	1	Sodium/potassium V-ATPase alpha
	PF00287	3	Sodium/potassium V-ATPase beta
	PF00955	1	Anion exchanger

Abbreviations: ABC = ATP binding cassette; NBD = nuclease binding domain; TMD = transmembrane domain; CYP450 = cytochrome P450 monooxygenases.

Peritrophic membrane proteins

The majority of insects produce a PM lining the intestinal lumen (Chapter 1). This semipermeable barrier has a dual function: it compartmentalizes the midgut lumen to increase digestive efficiency, and it protects the midgut epithelium from mechanical and chemical damage by the food bolus, as well as from harmful pathogens and toxins (Lehane, 1997). In the desert locust, the PM is produced by all midgut ECs (type I) (Chapman, 2013). The PM mainly consists of chitin microfibrils embedded in a matrix of chitin binding proteins (CBPs), called peritrophins. Two major types of CBPs belonging to Pfam00379 and Pfam01607 protein families have been described in insects (Tetreau et al., 2015). Most peritrophic matrix proteins (PMPs) are described to belong to the Pfam01607 protein family and are characterized by the presence of one or multiple cysteine-containing type-2 chitin-binding domains (ChtBD2), also called peritrophin A-type domains. Several (59 in total) transcripts encoding Pfam01607 proteins were identified in the *S. gregaria* midgut reference transcriptome, and might be involved in establishing the PM. Also three transcripts putatively encoding chitin synthase (CHS) were present in the midgut reference transcriptome and are possibly involved in the synthesis of chitin incorporated in the PM. Additionally, several transcripts (13 in total) encoding putative glycoside hydrolases containing chitin binding domains (CBDs) were identified in the midgut reference transcriptome. These proteins, endochitinases, belong to the Pfam00704 protein family and are involved in the degradation of chitinous materials. Their presence in the midgut might also suggest the (partial) recycling of the PM. However, until now, this has not been experimentally observed in *S. gregaria*.

Other transmembrane protein complexes

Transmembrane proton pumps are important mediators of the nutrient transport across the epithelial barrier in insects. They can either actively transport nutrients, or they can cooperate with other proton coupled transporters in a two-component transporter system. The majority of identified proton pumps in insects belong to the Vacuolar-type H^+-ATPase (H^+ V-ATPase) proton pumps and ATP binding cassette (ABC) transporters.

Multiple H^+ V-ATPase subunit encoding transcripts were identified in the *S. gregaria* midgut reference transcriptome, indicating their significance in this tissue (Table 3.10). H^+ V-ATPases are highly conserved transmembrane proton pump complexes located at the endomembrane system and the apical plasma membrane of cells. In the latter, they typically utilize the energy released from hydrolyzing ATP to pump protons into the extracellular space. This creates an electrochemical gradient across the epithelial membrane which, in synch with, for example, Na^+/H^+ exchangers, drives Na^+ coupled transporters in stimulating the inward movement of nutrients into the cell (Wieczorek et al., 2009). This two-component transporter mechanism is expected to be present in the gastrointestinal tract of all insect species, and will be explained in more detail in chapter 5 (Ch5|5.1.2). H^+ V-ATPases consist of two major subunits: a cytosolic V_1 domain and a transmembrane V_0 domain, with both domains containing multiple structural subunits. Briefly, the cytosolic V_1 domain is composed of eight different subunits (A-H), while the transmembrane V_0 domain is generally composed of six subunits (a, c, c'', d, e, and Ac45 in mammals and a, c, c',c'',d, and e in yeast) (Toei et al., 2010). Moreover, in mammalian cells, most of the V-ATPase subunits are described to exist in multiple isoforms which are often expressed in a tissue specific manner (Toei et al., 2010). Also in *D. melanogaster*, such isoform specific expression of H^+ V-ATPase subunits has also been identified (Chintapalli et al., 2013). Remarkably, not all functional H^+ V-ATPase subunits could be retrieved from the *S. gregaria* reference midgut transcriptome, and more in-depth research is needed to examine this.

ABC transporters are another important class of transmembrane proteins. They are generally responsible for the ATP-powered active transport of a wide range of molecules, including all types of nutrients and xenobiotics, across plasma membranes (Higgins, 1992). ABC transporters generally consist of four core domains: two cytosolic nucleotide-binding domains (NBDs) and two transmembrane domains (TMDs). Several transcripts encoding both the NBD (Pfam00005) as well as the TMD (Pfam00664) domains of ABC transporters were identified in the *S. gregaria* midgut reference transcriptome. The NBDs are typically responsible for the hydrolysis of ATP, which induces conformational changes in the TMDs allowing unidirectional transport of molecules across the membrane (Locher, 2004). The majority of ABC transporters described in eukaryotes appear to solely function as exporters, whereas import by ABC transporters seems to be restricted to prokaryotes (Rees et al., 2009). ABC transporters in insects are typically classified into eight different subfamilies based on sequence analyses (ABCA-H) (Dermauw and Van Leeuwen, 2014). However, until now, detailed information about the different ABC transporter subfamilies and their functionalities in insects is still limited. Members of several ABC subfamilies are represented in the *S. gregaria* midgut reference transcriptome: 5 ABCA, 9 ABCB, 5 ABCC, 3 ABCD, 1 ABCE, 3 ABCF and 14 ABCG. The subfamilies ABCC and ABCG are most represented in the transcriptome, which was analogous to what is described in other insect species (Broehan et al., 2013; Qi et al., 2016; Tian et al., 2017). Remarkably, no annotated representatives for ABCH were found in the transcriptome.

Detoxification proteins

Herbivorous insects encounter many naturally occurring plant toxins in their diet. The adequate metabolization of these toxins is critical for their survival. In general, the detoxification process in insects can be divided into three phases, phase I-III. In phase I, enzymes such as cytochrome P450 monooxygenases (P450s) and carboxylesterases (CEs) directly modify the toxic molecules by introducing hydroxyl, carboxyl or amino groups through oxidation, reduction, and/or hydrolysis reactions, making them more reactive and water soluble. In phase II, xenobiotic metabolites are conjugated with *e.g.* glutathione or glycosyl groups by glutathione

S-transferases (GSTs) or uridine diphosphate-glucuronosyltransferases (UGTs), respectively, generating excretable water-soluble products (Rowland et al., 2013; Dasari, 2017). Finally, in phase III, the generated polar compounds or conjugates are transported out of the cells by transmembrane transporters, typically ABC-type transporters (Dermauw and Van Leeuwen, 2014).

CYP450 monooxygenases comprise a well-described family of proteins present in all eukaryotes. They exhibit distinct functions in insects, including the biosynthesis of steroid hormones, and phase I xenobiotic metabolism (Scott, 1999; Iga and Kataoka, 2012). In general, large quantities of CYP450 enzymes are present in insects, with highest observed activity in midgut and fat body tissues (Hlavica, 2011; Iga and Kataoka, 2012). Genomic analyses revealed that insect CYP450 proteins typically fall into four clades: the CYP2, CYP3, CYP4 clade showing highest homology with vertebrate microsomal CYPs, and the mitochondrial CYP clade (Feyereisen, 2006). The detoxifying action of CYP450 enzymes is mainly linked to the action of the CYP6 subfamily of the CYP3 insect clade. Additionally, some members of the CYP4 insect clade are assumed to also contribute to the metabolism of xenobiotics (Hlavica, 2011). Members of all four CYP450 insect clans are present in the *S. gregaria* midgut reference transcriptome, of which the majority are CYP6 (80 in total) and CYP4 transcripts (19 in total). These data indicate that *S. gregaria* can rely on a wide range of CYP450 proteins in its midgut, which is in agreement with earlier observations in other insect species (Hlavica, 2011).

A total of 123 transcripts putatively encoding CEs (Pfam00135) were also identified in the *S. gregaria* midgut reference transcriptome. These enzymes are involved in phase I xenobiotic metabolism in insects by directly hydrolyzing the ester bonds of various toxic carboxylic ester substrates, such as organophosphorus insecticides, resulting in their detoxification. It has already been observed that insect species can contain large numbers of CE encoding genes. For example *B. mori* and *L. migratoria* contain 76 and 27 CE encoding genes, respectively (Cui et al., 2015; Spit et al., 2016a).

High numbers of UGT encoding transcripts (63 in total) were also identified in the *S. gregaria* midgut reference transcriptome. Together with the GSTs (59 in total), they contribute to phase II xenobiotic metabolism, generating excretable water-soluble products (Rowland et al., 2013; Dasari, 2017). Interestingly, as described in chapter 2, our *S. gregaria* colony is fed daily with cabbage, *B. oleracea*, which is known to produce high amounts of glucosinolates as defensive compounds (Ahuja et al., 2010). Insects feeding on these plants are described to heavily rely on UGTs and GSTs to effectively detoxify these xenobiotic compounds (Li et al., 2018). It is possible that this high enrichment of UGT and GST encoding transcripts in its midgut makes it possible for the desert locust to efficiently survive on cabbage.

ABC transporters described to mediate phase III xenobiotic metabolism typically belong to the multidrug resistance protein 1 (MRP1) family. This family includes the ABC transporter subfamilies ABCB, ABCC and ABCG, which have also been demonstrated to be involved in xenobiotic metabolism in insects (Labbé et al., 2011; Sodani et al., 2012; Dermauw and Van Leeuwen, 2014; Qi et al., 2016). All these ABC subfamilies were also represented in the *S. gregaria* midgut reference transcriptome (see above). In general, MRP1s are organic anion transporters that excrete compounds conjugated to glutathione or glycosyl groups by GSTs or UGTs, respectively, from the cell (Borst et al., 2000).

This paragraph clearly illustrates that herbivorous insects rely on a wide range of detoxification mechanisms as defense against host plant toxins. Transcripts encoding proteins that are part

of distinct phases of xenobiotic metabolism are abundantly present in the *S. gregaria* midgut reference transcriptome. Interestingly, insects are demonstrated to rely on similar mechanisms for their protection against pesticides (Hawkins et al., 2019). How they succeed in doing this is still an important subject of study (Zhu et al., 2016). The identification and characterization of the underlying patterns orchestrating metabolic resistance against pesticides will be crucial in combatting insect pesticide resistance in the future. Within this context, studying the insect midgut genome/transcriptome/proteome is of pivotal importance (Chapter 1).

3.5. Conclusion

This chapter explained the construction and annotation of an *S. gregaria* midgut transcriptome as a reference for differential expression analysis. Briefly, this reference transcriptome was created by filtering an adjusted *S. gregaria* whole-body transcriptome with transcripts identified from sequencing the mRNA content of midgut samples as part of a differential expression analysis of the digestive process.

Because *S. gregaria* is a non-model organism, and no genomic information was available, several preprocessing steps were performed on the in-house whole-body transcriptome in order to make it more accessible for subsequent transcriptional studies. An important preprocessing step involved generating so-called SuperTranscripts. These long contiguous sequences were created by collapsing unique and common sequence regions of all available splicing isoforms of a transcript present in the whole-body transcriptome. Although SuperTranscripts did not represent actual 'true' biological sequences, they represented excellent alternative sequences for subsequent mapping of RNA-Seq reads using Bowtie2. This way, the number of ambiguously mapping multireads was drastically reduced, inducing the reliability of all performed downstream analyses. It is however important to note that before conducting any downstream *in vivo* experiment based on sequence information obtained from this midgut reference transcriptome, one must always first check if one is analyzing a SuperTranscript or an 'original' transcript. In the case of a SuperTranscript, the multiple originating sequences that comprise the SuperTranscript need to be investigated separately, instead of the corresponding SuperTranscript.

The midgut reference transcriptome was *de novo* annotated, and investigated for functional enrichment focusing on the general midgut physiology of *S. gregaria*. A high number of sequences remained without annotation, possibly indicating strong evolutionary diversification of the *S. gregaria* midgut-associated genes. Nevertheless, among the annotated transcripts, large numbers of digestive enzymes and nutrient transporters, mediating the digestive breakdown and subsequent absorption of food particles in the midgut, respectively, could be identified. Additionally, several other transcripts encoding proteins described to play distinct roles in the insect midgut physiology were identified. Overall, the *S. gregaria* midgut reference transcriptome was demonstrated to be a resourceful database for the investigation of midgut associated processes.

In the next chapter, this annotated midgut reference transcriptome is used to identify differentially expressed genes in the RNA-Seq investigation of the digestive process in *S. gregaria*.

3.6. References

Ahuja I, Rohloff J, Bones AM, Magnar A, Defence B (2010) Defence mechanisms of *Brassicaceae*: implications for plant-insect interactions and potential for integrated pest management. A review. Agron Sustain Dev 30:311–348

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25:25–29. doi: 10.1038/75556

Borst P, Zelcer N, Van Helvoort A (2000) ABC transporters in lipid transport. Biochim Biophys Acta - Mol Cell Biol Lipids 1486:128–144. doi: 10.1016/S1388-1981(00)00053-6

Boudko DY (2012) Molecular basis of essential amino acid transport from studies of insect nutrient amino acid transporters of the SLC6 family (NAT-SLC6). J Insect Physiol 58:433–449. doi: 10.1016/j.jinsphys.2011.12.018

Broehan G, Kroeger T, Lorenzen M, Merzendorfer H (2013) Functional analysis of the ATP-binding cassette (ABC) transporter gene family of *Tribolium castaneum*. BMC Genomics 14:6. doi: 10.1186/1471-2164-14-6

Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, Lee TJ, Leigh ND, Kuo TH, Davis FG, et al (2017) A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. Cell Rep 18:762–776. doi: 10.1016/j.celrep.2016.12.063

Cao X, He Y, Hu Y, Zhang X, Wang Y, Zou Z, Chen Y, Blissard GW, Kanost MR, Jiang H (2015) Sequence conservation, phylogenetic relationships, and expression profiles of nondigestive serine proteases and serine protease homologs in *Manduca sexta*. Insect Biochem Mol Biol 62:51–63. doi: 10.1016/j.ibmb.2014.10.006

Chintapalli VR, Wang J, Herzyk P, Davies SA, Dow JA (2013) Data-mining the FlyAtlas online resource to identify core functional motifs across transporting epithelia. BMC Genomics 14:518. doi: 10.1186/1471-2164-14-518

Chiou S-J, Vanden Broeck J, Janssen I, Borovsky D, Vandenbussche F, Simonet G, De Loof A (1998) Cloning of the cDNA encoding Scg-SPRP, an unusual Ser-protease-related protein from vitellogenic female desert locusts (*Schistocerca gregaria*). Insect Biochem Mol Biol 28:801–808. doi: 10.1016/S0965-1748(98)00054-X

Christeller JT (2005) Evolutionary mechanisms acting on proteinase inhibitor variability. FEBS J 272:5710–5722. doi: 10.1111/j.1742-4658.2005.04975.x

Ciufo LF, Murray PA, Thompson A, Rigden DJ, Rees HH (2011) Characterisation of a Desmosterol Reductase Involved in Phytosterol Dealkylation in the Silkworm, *Bombyx mori*. PLoS One 6:e21316. doi: 10.1371/journal.pone.0021316

Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM (2010) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res 38:1767–1771. doi: 10.1093/nar/gkp1137

Colombani J, Raisin S, Pantalacci S, Radimerski T, Montagne J, Léopold P (2003) A Nutrient Sensor Mechanism Controls Drosophila Growth. Cell 114:739–749. doi: 10.1016/S0092-8674(03)00713-X

Cui F, Li M-X, Chang H-J, Mao Y, Zhang H-Y, Lu L-X, Yan S-G, Lang M-L, Liu L, Qiao C-L (2015) Carboxylesterase-mediated insecticide resistance: Quantitative increase induces broader metabolic resistance than qualitative change. Pestic Biochem Physiol 121:88–96. doi: 10.1016/j.pestbp.2014.12.016

Da Lage J-L (2018) The Amylases of Insects. Int J Insect Sci 10:117954331880478. doi: 10.1177/1179543318804783

Dasari S (2017) Glutathione S-transferases Detoxify Endogenous and Exogenous Toxic Agents-Minireview. J Dairy, Vet Anim Res 5:157–159. doi: 10.15406/jdvar.2017.05.00154

Davidson NM, Hawkins ADK, Oshlack A (2017) SuperTranscripts: A data driven reference for analysis and visualisation of transcriptomes. Genome Biol 18:1–10. doi: 10.1186/s13059-017-1284-1

Dermauw W, Van Leeuwen T (2014) The ABC gene family in arthropods: Comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol 45:89–110. doi: 10.1016/j.ibmb.2013.11.001

Dhania NK, Chauhan VK, Chaitanya RK, Dutta-Gupta A (2019) Midgut de novo transcriptome analysis and gene expression profiling of *Achaea janata* larvae exposed with *Bacillus thuringiensis* (Bt)-based biopesticide formulation. Comp Biochem Physiol Part D Genomics Proteomics 30:81–90. doi: 10.1016/j.cbd.2019.02.005

Evans AM, Aimanova KG, Gill SS (2009) Characterization of a blood-meal-responsive proton-dependent amino acid transporter in the disease vector, *Aedes aegypti*. J Exp Biol 212:3263–

3271. doi: 10.1242/jeb.029553
Feyereisen R (2006) Evolution of insect P450 Introduction: insect genomes. Biochem Soc Trans 34:1252–1255
Fluegel ML, Parker TJ, Pallanck LJ (2006) Mutations of a *Drosophila* NPC1 Gene Confer Sterol and Ecdysone Metabolic Defects. Genetics 172:185–196. doi: 10.1534/genetics.105.046565
Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. doi: 10.1093/bioinformatics/bts565
Galperin MY, Makarova KS, Wolf YI, Koonin E V. (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261–D269. doi: 10.1093/nar/gku1223
Gazara RK, Cardoso C, Bellieny-Rabelo D, Ferreira C, Terra WR, Venancio TM (2017) De novo transcriptome sequencing and comparative analysis of midgut tissues of four non-model insects pertaining to Hemiptera, Coleoptera, Diptera and Lepidoptera. Gene 627:85–93. doi: 10.1016/j.gene.2017.06.008
Ge Z-Y, Wan P-J, Han Z-J (2012) Cloning and characterization of trypsin- and chymotrypsin-like genes in the striped rice stem borer, *Chilo suppressalis*. Genome 55:281–288. doi: 10.1139/g2012-015
Gong J, Hou Y, Zha X-F, Lu C, Zhu Y, Xia Q-Y (2006) Molecular cloning and characterization of *Bombyx mori* sterol carrier protein x/sterol carrier protein 2 (SCPx/SCP2) gene. DNA Seq 17:326–333. doi: 10.1080/10425170600886706
Hansen IA, Boudko DY, Shiao S-H, Voronov DA, Meleshkevitch EA, Drake LL, Aguirre SE, Fox JM, Attardo GM, Raikhel AS (2011) AaCAT1 of the Yellow Fever Mosquito, *Aedes aegypti*. J Biol Chem 286:10803–10813. doi: 10.1074/jbc.M110.179739
Haunerl NH, Chisholm JM (1990) Fatty acid binding protein in flight muscle of the locust, Schistocerca gregaria. Biochim Biophys Acta - Lipids Lipid Metab 1047:233–238. doi: 10.1016/0005-2760(90)90521-X
Hawkins NJ, Bass C, Dixon A, Neve P (2019) The evolutionary origins of pesticide resistance. Biol Rev 94:135–155. doi: 10.1111/brv.12440
Hedstrom L (2002) Serine Protease Mechanism and Specificity. Chem Rev 102:4501–4524. doi: 10.1021/cr000033x
Higgins CF (1992) ABC transporters e from microorganisms to man. Annu Rev Cell Biol 8:67–113
Hlavica P (2011) Insect cytochromes P450: Topology of structural elements predicted to govern catalytic versatility. J Inorg Biochem 105:1354–1364. doi: 10.1016/j.jinorgbio.2011.05.003
Holtof M, Lenaerts C, Cullen D, Vanden Broeck J (2019) Extracellular nutrient digestion and absorption in the insect gut. Cell Tissue Res. doi: 10.1007/s00441-019-03031-9
Horne I, Haritos VS, Oakeshott JG (2009) Comparative and functional genomics of lipases in holometabolous insects. Insect Biochem Mol Biol 39:547–567. doi: 10.1016/j.ibmb.2009.06.002
Huang X (2005) A *Drosophila* model of the Niemann-Pick type C lysosome storage disease: dnpc1a is required for molting and sterol homeostasis. Development 132:5115–5124. doi: 10.1242/dev.02079
Huang X, McNeill MR, Ma J, Qin X, Tu X, Cao G, Wang G, Nong X, Zhang Z (2017) Gut Transcriptome Analysis Shows Different Food Utilization Efficiency by the Grasshopper *Oedaleous asiaticus* (Orthoptera: Acrididae). J Econ Entomol 110:1831–1840. doi: 10.1093/jee/tox128
Huang X, Warren JT, Buchanan J, Gilbert LI, Scott MP (2007) *Drosophila* Niemann-Pick Type C-2 genes control sterol homeostasis and steroid biosynthesis: a model of human neurodegenerative disease. Development 134:3733–3742. doi: 10.1242/dev.004572
Iga M, Kataoka H (2012) Recent Studies on Insect Hormone Metabolic Pathways Mediated by Cytochrome P450 Enzymes. Biol Pharm Bull 35:838–843. doi: 10.1248/bpb.35.838
Kanamori Y, Saito A, Hagiwara-Komoda Y, Tanaka D, Mitsumasu K, Kikuta S, Watanabe M, Cornette R, Kikawada T, Okuda T (2010) The trehalose transporter 1 gene sequence is conserved in insects and encodes proteins with different kinetic properties involved in trehalose import into peripheral tissues. Insect Biochem Mol Biol 40:30–37. doi: 10.1016/j.ibmb.2009.12.006
Katoch R, Thakur N (2012) Insect gut nucleases: a challenge for RNA interference mediated insect control strategies. Int J Biochem Biotechnol 1:198–203
Kikawada T, Saito A, Kanamori Y, Nakahara Y, Iwata K -i., Tanaka D, Watanabe M, Okuda T (2007) Trehalose transporter 1, a facilitated and high-capacity trehalose transporter, allows exogenous trehalose uptake into cells. Proc Natl Acad Sci 104:11585–11590. doi: 10.1073/pnas.0702538104
Kim M-S, Lan Q (2010) Sterol carrier protein-x gene and effects of sterol carrier protein-2 inhibitors on lipid uptake in *Manduca sexta*. BMC Physiol 10:9. doi: 10.1186/1472-6793-10-9
Kitamura T, Kobayashi S, Okada M (1996) Regional expression of the transcript encoding sterol carrier protein x-related thiolase and its regulation by homeotic genes in the midgut of *Drosophila*

embryos. Dev Growth Differ 38:373–381. doi: 10.1046/j.1440-169X.1996.t01-3-00005.x
Kunieda T, Fujiyuki T, Kucharski R, Foret S, Ament SA, Toth AL, Ohashi K, Takeuchi H, Kamikouchi A, Kage E, et al (2006) Carbohydrate metabolism genes and pathways in insects: Insights from the honey bee genome. Insect Mol Biol 15:563–576. doi: 10.1111/j.1365-2583.2006.00677.x
Labbé R, Caveney S, Donly C (2011) Genetic analysis of the xenobiotic resistance-associated ABC gene subfamilies of the Lepidoptera. Insect Mol Biol 20:243–256. doi: 10.1111/j.1365-2583.2010.01064.x
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. doi: 10.1038/nmeth.1923
Langmead B, Slazberg SL (2013) Fast gapped-read alignmnet with Bowtie 2. Nat Methods 9:357–359. doi: 10.1038/nmeth.1923.Fast
Lehane MJ (1997) Peritrophic matrix structure and function. Annu Rev Entomol 42:525–550. doi: 10.1146/annurev.ento.42.1.525
Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323. doi: 10.1186/1471-2105-12-323
Li P, Piao Y, Shon HS, Ryu KH (2015) Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. BMC Bioinformatics 16:347. doi: 10.1186/s12859-015-0778-7
Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659. doi: 10.1093/bioinformatics/btl158
Li X, Shi H, Gao X, Liang P (2018) Characterization of UDP-glucuronosyltransferase genes and their possible roles in multi-insecticide resistance in *Plutella xylostella* (L.). Pest Manag Sci 74:695–704. doi: 10.1002/ps.4765
Lin H, Xia X, Yu L, Vasseur L, Gurr GM, Yao F, Yang G, You M (2015) Genome-wide identification and expression profiling of serine proteases and homologs in the diamondback moth, *Plutella xylostella* (L.). BMC Genomics 16:1–13. doi: 10.1186/s12864-015-2243-4
Liu Y, Qi M, Chi Y, Wuriyanghan H (2016) De novo assembly of the transcriptome for oriental armyworm *Mythimna separata* (Lepidoptera: Noctuidae) and analysis on insecticide resistance-related genes. J Insect Sci 16:0–7. doi: 10.1093/jisesa/iew079
Locher KP (2004) Structure and mechanism of ABC transporters. Curr Opin Struct Biol 14:426–431. doi: 10.1016/j.sbi.2004.06.005
Lopes AR, Juliano MA, Juliano L, Terra WR (2004) Coevolution of insect trypsins and inhibitors. Arch Insect Biochem Physiol 55:140–152. doi: 10.1002/arch.10134
Ma W, Zhang Z, Peng C, Wang X, Li F, Lin Y (2012) Exploring the midgut transcriptome and brush border membrane vesicle proteome of the rice stem borer, *Chilo suppressalis* (walker). PLoS One 7:1–7. doi: 10.1371/journal.pone.0038151
Ma Z, Yu J, Kang L (2006) LocustDB: A relational database for the transcriptome and biology of the migratory locust (*Locusta migratoria*). BMC Genomics 7:1–3. doi: 10.1186/1471-2164-7-11
Martin JF, Hersperger E, Simcox A, Shearn A (2000) minidiscs encodes a putative amino acid transporter subunit required non-autonomously for imaginal cell proliferation. Mech Dev 92:155–167. doi: 10.1016/S0925-4773(99)00338-X
Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10. doi: 10.14806/ej.17.1.200
Muhlia-Almazán A, Sánchez-Paz A, García-Carreño FL (2008) Invertebrate trypsins: a review. J Comp Physiol B 178:655–672. doi: 10.1007/s00360-008-0263-y
Ohba T, Rennert H, Pfeifer SM, He Z, Yamamoto R, Holt JA, Billheimer JT, Strauss JF (1994) The Structure of the Human Sterol Carrier Protein X/Sterol Carrier Protein 2 Gene (SCP2). Genomics 24:370–374. doi: 10.1006/geno.1994.1630
Oppert C, Oppert B, Morris K, Lorenzen MD, Jurat-Fuentes JL, Vinokurov K, Elpidina EN, Hiromasa Y, Fabrick J, Tomich JM (2009) *Tribolium castaneum* Larval Gut Transcriptome and Proteome: A Resource for the Study of the Coleopteran Gut. J Proteome Res 8:3889–3898. doi: 10.1021/pr900168z
Oshlack A, Wakefield MJ (2009) Transcript length bias in RNA-seq data confounds systems biology. Biol Direct 4:14. doi: 10.1186/1745-6150-4-14
Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, Ffrench-Constant RH (2009) Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera. Insect Biochem Mol Biol 39:403–413. doi: 10.1016/j.ibmb.2009.04.001
Pedra JHF, Brandt A, Westerman R, Lobo N, Li HM, Romero-Severson J, Murdock LL, Pittendrigh BR (2003) Transcriptome analysis of the cowpea weevil bruchid: Identification of putative proteinases

and α-amylases associated with food breakdown. Insect Mol Biol 12:405–412. doi: 10.1046/j.1365-2583.2003.00425.x
Perona JJ, Craik CS (1995) Structural basis of substrate specificity in the serine proteases. Protein Sci 4:337–360
Popham HJR, Gould F, Adang MJ, Jurat-Fuentes JL, Perera OP, Shelby KS (2015) Generation of a Transcriptome in a Model Lepidopteran Pest, *Heliothis virescens*, Using Multiple Sequencing Strategies for Profiling Midgut Gene Expression. PLoS One 10:e0128563. doi: 10.1371/journal.pone.0128563
Qi W, Ma X, He W, Chen W, Zou M, Gurr GM, Vasseur L, You M (2016) Characterization and expression profiling of ATP-binding cassette transporter genes in the diamondback moth, *Plutella xylostella* (L.). BMC Genomics 17:760. doi: 10.1186/s12864-016-3096-1
Qian H, Li G, He Q, Zhang H, Xu A (2016) Analysis of differentially expressed genes between fluoride-sensitive and fluoride-endurable individuals in midgut of silkworm, *Bombyx mori*. Gene 588:47–53. doi: 10.1016/j.gene.2016.04.033
Rees DC, Johnson E, Lewinson O (2009) ABC transporters: the power to change. Nat Rev Mol Cell Biol 10:218–227. doi: 10.1038/nrm2646
Reynolds B, Roversi P, Laynes R, Kazi S, Boyd CAR, Goberdhan DCI (2009) Drosophila expresses a CD98 transporter with an evolutionarily conserved structure and amino acid-transport properties. Biochem J 420:363–372. doi: 10.1042/BJ20082198
Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. doi: 10.1093/bioinformatics/btp616
Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25. doi: 10.1186/gb-2010-11-3-r25
Roman G, Meller V, Wu KH, Davis RL (1998) The opt1 gene of *Drosophila melanogaster* encodes a proton-dependent dipeptide transporter. Am J Physiol Physiol 275:C857–C869. doi: 10.1152/ajpcell.1998.275.3.C857
Rowland A, Miners JO, Mackenzie PI (2013) The UDP-glucuronosyltransferases: Their role in drug metabolism and detoxification. Int J Biochem Cell Biol 45:1121–1132. doi: 10.1016/j.biocel.2013.02.019
Scott JG (1999) Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol 29:757–777. doi: 10.1016/S0965-1748(99)00038-7
Shelomi M, Watanabe H, Arakawa G (2014) Endogenous cellulase enzymes in the stick insect (Phasmatodea) gut. J Insect Physiol 60:25–30. doi: 10.1016/j.jinsphys.2013.10.007
Shi X-Z, Zhong X, Yu X-Q (2012) *Drosophila melanogaster* NPC2 proteins bind bacterial cell wall components and may function in immune signal pathways. Insect Biochem Mol Biol 42:545–556. doi: 10.1016/j.ibmb.2012.04.002
Sodani K, Patel A, Kathawala RJ, Chen Z-S (2012) Multidrug resistance associated proteins in multidrug resistance. Chin J Cancer 31:58–72. doi: 10.5732/cjc.011.10329
Spit J, Badisco L, Vergauwen L, Knapen D, Vanden Broeck J (2016a) Microarray-based annotation of the gut transcriptome of the migratory locust, *Locusta migratoria*. Insect Mol Biol 25:745–756. doi: 10.1111/imb.12258
Spit J, Breugelmans B, Van Hoef V, Simonet G, Zels S, Vanden Broeck J (2012) Growth-inhibition effects of pacifastin-like peptides on a pest insect: The desert locust, *Schistocerca gregaria*. Peptides 34:251–257. doi: 10.1016/j.peptides.2011.06.019
Spit J, Holtof M, Badisco L, Vergauwen L, Vogel E, Knapen D, Vanden Broeck J (2016b) Transcriptional Analysis of The Adaptive Digestive System of The Migratory Locust in Response to Plant Defensive Protease Inhibitors. Sci Rep 6:32460. doi: 10.1038/srep32460
Spit J, Zels S, Dillen S, Holtof M, Wynant N, Vanden Broeck J (2014) Effects of different dietary conditions on the expression of trypsin- and chymotrypsin-like protease genes in the digestive system of the migratory locust, *Locusta migratoria*. Insect Biochem Mol Biol 48:100–109. doi: 10.1016/j.ibmb.2014.03.002
Svoboda JA (1999) Variability of Metabolism and Function of Sterols in Insects. Crit Rev Biochem Mol Biol 34:49–57. doi: 10.1080/10409239991209183
Takeuchi H, Chen J-H, Jenkins JR, Bun-Ya M, Turner PC, Rees HH (2004) Characterization of a sterol carrier protein 2/3-oxoacyl-CoA thiolase from the cotton leafworm (*Spodoptera littoralis*): a lepidopteran mechanism closer to that in mammals than that in dipterans. Biochem J 382:93–100. doi: 10.1042/BJ20040717
Tatusov RL (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36. doi: 10.1093/nar/28.1.33

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin E V, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:1–14. doi: 10.1186/1471-2105-4-41

Terra WR, Ferreira C (1994) Insect digestive enzymes : properties , compartmentalization and function. Comp Biochem Physiol 109B:1–62

Tetreau G, Dittmer NT, Cao X, Agrawal S, Chen Y-R, Muthukrishnan S, Haobo J, Blissard GW, Kanost MR, Wang P (2015) Analysis of chitin-binding proteins from *Manduca sexta* provides new insights into evolution of peritrophin A-type chitin-binding domains in insects. Insect Biochem Mol Biol 62:127–141. doi: 10.1016/j.ibmb.2014.12.002

Tian L, Song T, He R, Zeng Y, Xie W, Wu Q, Wang S, Zhou X, Zhang Y (2017) Genome-wide analysis of ATP-binding cassette (ABC) transporters in the sweetpotato whitefly, *Bemisia tabaci*. BMC Genomics 18:330. doi: 10.1186/s12864-017-3706-6

Toei M, Saum R, Forgac M (2010) Regulation and Isoform Function of the V-ATPases. Biochemistry 49:4715–4723. doi: 10.1021/bi100397s

Trapnell C, Salzberg SL (2009) How to map billions of short reads onto genomes. Nat Biotechnol 27:455–457. doi: 10.1038/nbt0509-455

Van Hiel MB, Van Wielendaele P, Temmerman L, Van Soest S, Vuerinckx K, Huybrechts R, Vanden Broeck J, Simonet G (2009) Identification and validation of housekeeping genes in brains of the desert locust *Schistocerca gregaria* under different developmental conditions. BMC Mol Biol 10:56. doi: 10.1186/1471-2199-10-56

Verrey F, Closs EI, Wagner CA, Palacin M, Endou H, Kanai Y (2004) CATs and HATs: the SLC7 family of amino acid transporters. Pflugers Arch Eur J Physiol 447:532–542. doi: 10.1007/s00424-003-1086-z

Voght SP, Fluegel ML, Andrews LA, Pallanck LJ (2007) *Drosophila* NPC1b Promotes an Early Step in Sterol Absorption from the Midgut Epithelium. Cell Metab 5:195–205. doi: 10.1016/j.cmet.2007.01.011

Wang XY, Yu HZ, Geng L, Xu JP, Yu D, Zhang SZ, Ma Y, Fei DQ (2016) Comparative transcriptome analysis of *Bombyx mori* (Lepidoptera) larval midgut response to BmNPV in susceptible and near-isogenic resistant strains. PLoS One 11:1–22. doi: 10.1371/journal.pone.0155341

Watanabe H, Tokuda G (2010) Cellulolytic Systems in Insects. Annu Rev Entomol 55:609–632. doi: 10.1146/annurev-ento-112408-085319

Wieczorek H, Beyenbach KW, Huss M, Vitavska O (2009) Vacuolar-type proton pumps in insect epithelia. J Exp Biol 212:1611–1619. doi: 10.1242/jeb.030007

Yang B, Huang W, Zhang J, Xu Q, Zhu S, Zhang Q, Beerntsen BT, Song H, Ling E (2016) Analysis of gene expression in the midgut of *Bombyx mori* during the larval molting stage. BMC Genomics 17:1–17. doi: 10.1186/s12864-016-3162-8

Yang W (2011) Nucleases: diversity of structure, function and mechanism. Q Rev Biophys 44:1–93. doi: 10.1017/S0033583510000181

Ye XD, Su YL, Zhao QY, Xia WQ, Liu SS, Wang XW (2014) Transcriptomic analyses reveal the adaptive features and biological differences of guts from two invasive whitefly species. BMC Genomics 15:1–12. doi: 10.1186/1471-2164-15-370

Zhang J, Zhang Y, Li J, Liu M, Liu Z (2016) Midgut transcriptome of the cockroach *Periplaneta americana* and its microbiota: Digestion, detoxification and oxidative stress response. PLoS One 11:1–20. doi: 10.1371/journal.pone.0155254

Zheng J-C, Sun S-L, Yue X-R, Liu T-X, Jing X (2018) Phylogeny and evolution of the cholesterol transporter NPC1 in insects. J Insect Physiol 107:157–166. doi: 10.1016/j.jinsphys.2018.04.007

Zhu-Salzman K, Zeng R (2015) Insect Response to Plant Defensive Protease Inhibitors. Annu Rev Entomol 60:233–252. doi: 10.1146/annurev-ento-010814-020816

Zhu F, Moural TW, Nelson DR, Palli SR (2016) A specialist herbivore pest adaptation to xenobiotics through up-regulation of multiple Cytochrome P450s. Sci Rep 6:20421. doi: 10.1038/srep20421

Zibaee A, Alborzi Z, Hoda H (2014) A review on digestive TAG-lipases of insects. Arch Phytopathol Plant Prot 47:2187–2192. doi: 10.1080/03235408.2013.870112

Chapter 4

RNA-Seq analysis of the *Schistocerca gregaria* midgut during digestion

4.1. Introduction

The central theme of this doctoral research was the investigation of the regulation at transcript level of the digestive process in the desert locust, a well-known polyphagous herbivorous pest species. Therefore, an RNA-Seq analysis of the midgut of *S. gregaria* larvae at three well-considered timepoints after feeding was conducted. These three timepoints were: (A) ten minutes, (B) two hours, and (C) twenty-four hours after feeding, representing the initial stage of digestion, the active digestion, and a reference point, respectively. This was already elaborated in detail in chapter 2.

Until now, only limited research on the transcript regulation of digestion in insects exists. With this study we wanted to gain a better insight in how major events during the digestive process, such as enzymatic digestion and nutrient uptake, are mediated inside the insect midgut. In addition, we also dissected other parts of the alimentary canal, as well as the heads. The transcript profiles of these tissues might also be explored in the future. For now, we decided to focus on the midgut, since this is the primary site of digestion and nutrient uptake in insects.

In chapter 3, we already investigated the global transcript enrichment in the midgut reference transcriptome of the desert locust. Next, we decided to analyze how this transcript profile alters in response to feeding, providing us more information on the regulation of digestion at the gene transcript level, a field of research that has been largely neglected up to now (Chapter 1). Moreover, because of the many predicted parallels between insect and mammalian digestion, insects are considered to be relevant small model organisms for studying gastrointestinal functions. In this context, our RNA-Seq study might contribute to the further unravelling of the regulation of digestion in general. Additionally, gaining a better understanding of how insects orchestrate their digestion, might for example generate fundamental information for the development of new insecticides specifically targeting vital gene transcripts or proteins in the midgut. The gastrointestinal tract of insects, as the barrier between the inner and outer environment, has long been considered to be an interesting target site for insecticides. The best known examples of such insecticides are the different toxins produced by the entomopathogenic bacterium, *Bacillus thuringiensis* (Bt), as described in chapter 1. About 952 toxin genes have already been identified and characterized in different Bt strains around the world (Jouzani et al., 2017). Briefly, after ingestion by the insect, Bt toxins bind to specific receptors at the luminal side of the midgut epithelium, disrupting the midgut wall, resulting in the preliminary death of the insect (Hofte and Whiteley, 1989; Broderick et al., 2006). Although reports on insect resistance against Bt toxins are rapidly increasing, these molecules have clearly demonstrated the power of biopesticides targeting gut function and integrity (reviewed by Tabashnik et al., 2013; Tabashnik, 2015; Coates, 2016). Recently, several other midgut-associated protein complexes received growing attention as putative insecticidal targets, such as for example V-ATPases (Eyraud et al., 2017). We believe that many other, perhaps more interesting, targets are available inside the insect midgut. For example, proteins (or the corresponding transcripts) essential for nutrient absorption or protection of the midgut epithelial barrier. With this RNA-Seq study, we provide a first step towards identifying novel potential targets.

In this chapter, differential expression between each timepoint is analyzed. Transcripts with significantly higher abundances in a specific condition, when compared to the other, are expected to play distinct roles in the midgut during this phase of digestion. The *S. gregaria* midgut reference transcriptome constructed in chapter 3 was used to identify these transcripts and predict their putative functions. The most noteworthy findings with an expected link to the extracellular digestion in insects are further described.

4.2. Experimental procedure

4.2.1. Aims of the experiment

In this chapter, we will continue to utilize the estimated count table obtained in previous chapter after subsequently mapping (Bowtie2, version 2.2.5) and counting (RSEM) the sequencing reads of every sample of the RNA-Seq study onto the adjusted *S. gregaria* whole-body transcriptome. For detailed explanations about mapping and counting of sequencing reeds, see chapter 3 paragraphs Ch3|3.2.3 and Ch3|3.2.4, respectively. These data were then used to investigate the transcript changes possibly mediating the onset and maintenance of the digestive process in *S. gregaria*.

To investigate differential expression between samples from the three different timepoints after feeding, pairwise differential expression analyses were performed. Hence, three separate comparative analyses were executed:

1. Transcript profile of midguts at ten minutes and twenty-four hours after feeding.
2. Transcript profile of midguts at two hours and twenty-four hours after feeding.
3. Transcript profile of midguts at ten minutes and two hours after feeding.

Several software packages and accessory statistical models exist for differential expression analysis from RNA-Seq data. For more information, the reader is referred to the several in-depth reviews on this topic (Rapaport et al., 2013; Soneson and Delorenzi, 2013; Huang et al., 2015; Seyednasrollah et al., 2015; Finotello and Di Camillo, 2015; Costa-Silva et al., 2017). We decided to investigate differential expression using the edgeR Bioconductor software in R (Robinson et al., 2009).

4.2.2. Differential expression analysis using edgeR

A typical RNA-seq experiment consists of the sequencing of RNA samples representing different conditions, mapping short reads to the appropriate genome/transcriptome, and recording the number of reads mapped to each gene/transcript. After normalization of the count data, significant differences in gene expression between the different conditions are calculated. However, since we mostly observe only a fraction of the true amounts of mRNA molecules in each sample, appropriate statistical modeling of transcript counts is required to investigate differential expression among different conditions.

The distribution of every transcript in an RNA-Seq sample might be described by a Poisson distribution. This distribution describes processes in which discrete events (here: transcripts) are sampled out of a large pool (here: sequenced mRNA pool) with low probability, and is therefore frequently used to model count data. A fundamental property of the Poisson distribution states that the mean of the transcript counts should be equal to the variance across all samples. However, in the case of biological replicates in an RNA-Seq experiment, the observed variance of the counts of a transcript over different samples within a given condition is usually higher than the average count for that same transcript. In other words, RNA-Seq datasets generally contain more variability than Poisson can account for, called overdispersion. Hence, RNA-Seq data are commonly described to follow a negative binominal (NB) distribution, in which a dispersion parameter is added to the Poisson distribution (Fig. 4.1) (Robinson and Smyth, 2007; Marioni et al., 2008; Robinson et al., 2009).

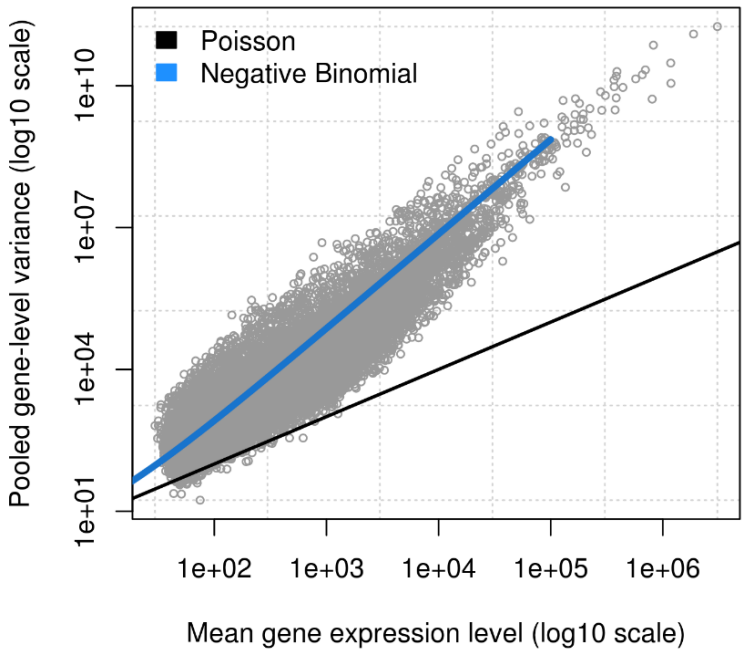


Figure 4.1. Scatter plot of the log10-transformed variances against the log10-transformed mean expression level of all transcripts in all samples of the RNA-Seq dataset. This plot clearly illustrates that the variance of the transcript counts is larger than their mean, and that the data therefore follow a NB distribution rather than a Poisson distribution.

Appropriate modelling of the transcript variability between biological replicates is essential when calculating differential expression among different conditions. Two levels of variation can be distinguished in any RNA-Seq experiment: first a biological variance, representing the 'true' biological differences in transcript abundances between samples, and second a technical variance, representing differences in transcript abundances mainly due to technical or measurement errors. The technical variance decreases as the size of the counts, thus the sequencing depth, increases. The biological variance on the other hand does not, and is expected to be the dominant source of variation for high-count reads. Therefore, reliable estimation of this biological variance is essential for differential expression analysis in RNA-Seq studies.

EdgeR uses the NB distribution to model the read counts for each transcript in each sample. The dispersion parameter of the NB distribution accounts for the overall transcript variability between biological replicates (Robinson and Smyth, 2007; Robinson et al., 2009). This general NB model is extended with quasi-likelihood (QL) methods to account for the transcript-specific variability from both biological and technical sources. The NB dispersion trend accounts for the global transcript variation across all transcripts, while the additional QL dispersions account for transcript-specific variation differing from this overall level of variation. Employing these transcript-specific QL dispersions is necessary to avoid that differential expression is driven by outliers. Estimation of the transcript-specific QL dispersions is often challenging because of the limited number of replicates in an RNA-Seq dataset, generating too little information to confidently estimate the dispersion of each transcript. To overcome this, an empirical Bayes (EB) strategy is applied to moderate the degree of transcript-specific dispersion, in which information is shared among transcripts. Briefly, a mean-dependent trend is fitted to the raw QL dispersions of each transcript. The raw QL dispersions are then squeezed towards this trend generating moderated EB estimates, that replace the raw QL dispersions for downstream analyses (Fig. 4.2). Hence, the EB strategy reduces the ambiguity of the transcript-specific QL dispersions and improves testing power (Robinson et al., 2009; McCarthy et al., 2012).

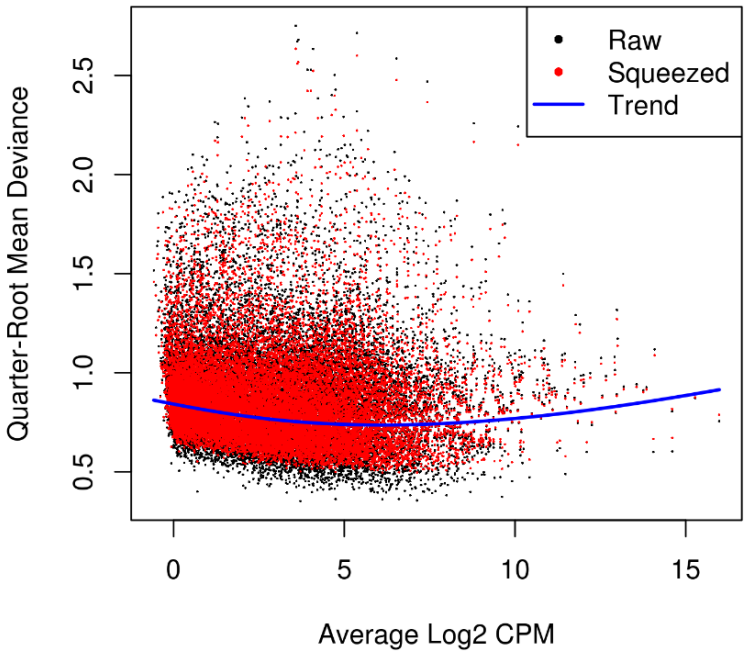


Figure 4.2. Scatterplot of the quarter-root mean deviance against the average abundance (log2-transformed CPM) of each transcript in all samples of the RNA-Seq dataset. The quarter-root transformation was applied to improve visibility for dispersions. The trended dispersion displays the global level of dispersion across the samples. Raw transcript-specific QL dispersions are squeezed towards the trended estimate dispersions by EB moderation to obtain squeezed transcript-specific QL dispersions. The relatively small shrinkage indicates that the transcript-specific QL dispersions were more variable, and less squeezing was performed.

Once the negative binomial model is fitted and the dispersion estimates are obtained, edgeR calculates if transcript quantities differ significantly between conditions. This is specified by the log2-ratio of the fold change (FC), which is the ratio by which the transcript counts vary between conditions. A positive log_2FC above a certain threshold indicates the transcript is quantitively upregulated in one condition compared to the control condition, whereas a negative log_2FC below a certain threshold indicates the transcript is quantitatively downregulated. For example, a transcript with a log_2FC of 1 has doubled in quantity compared to the control condition; *vice versa* for a transcript with a log_2FC of -1. To determine the statistical significance of differential expression, edgeR relies on quasi-likelihood F-tests to test the null hypothesis that the transcript abundance is the same in the tested conditions (Wu and Wu, 2016). Because an RNA-Seq dataset consists of a large number of values, proper correction for multiple comparisons is crucial, *i.e.* the higher the number of tests, the higher the chance of false positive observations to occur. Therefore, the false discovery rate (FDR) is typically applied as an extra *post-hoc* method to determine the statistical significance of observations in RNA-Seq datasets (Li et al., 2012b). For each transcript, the obtained p-value, describing the overall probability of any observation to be a false positive, is adjusted to a q-value, or FDR, using the Benjamini-Hochberg FDR procedure (Benjamini and Hochberg, 1995). This FDR eventually reports the chance of a significant result being false positive. For example, an FDR smaller than 0.05 indicates that less than 5% of the total amount of predicted differentially expressed genes will be false positives.

4.3. Materials and methods

4.3.1. Animal rearing

The optimized rearing conditions are described in detail in chapter 2 (Ch2|2.3.1.2).

4.3.2. Tissue collection

Tissue collection is described in detail in chapter 2 (Ch2|2.3.2.1).

4.3.3. Total RNA extraction

Total RNA extraction of the midgut samples is explained in detail in chapter 2 (Ch2|2.3.2.2). Determination of nucleic acid concentrations and quality control of the RNA samples are described in chapter 2 (Ch2|2.3.2.3).

4.3.4. Illumina® sequencing library preparation and sequencing[1]

Library preparation using the TruSeq® Stranded mRNA Library prep kit (Illumina®) was performed according to the manufacturer's protocol and is explained in detail in chapter 2 (Ch2|2.3.3.1). Libraries were quantified by qPCR, according to Illumina's protocol 'Sequencing Library qPCR Quantification protocol guide', version February 2011. A High Sensitivity DNA chip (Agilent Technologies®) was used to control the library's size distribution and quality. Sequencing was performed on the Illumina® NextSeq 500 platform generating 75 bp single reads and is explained in detail in chapter 2 (Ch2|2.3.3.2).

4.3.5. Data acquisition/quality control/mapping[2]

Per sample, on average 55 ± 10 million reads were generated. First, these reads were trimmed using cutadapt, version 1.11, to remove the Illumina adaptor sequence (Martin 2011). The trimmed reads were mapped against the adjusted *S. gregaria* whole-body transcriptome using bowtie2, version 2.2.5 (Langmead and Salzberg, 2012). The RSEM software, version 1.2.31, was used to generate the count tables (Li and Dewey, 2011).

4.3.6. Differential expression analysis

To explore if the samples from different groups clustered together and to detect outlier samples, Principal Component Analyses (PCAs) on regularized log (rlog) transformed counts, Deseq2 Bioconductor software, were performed using the R statistical computing software (Love et al., 2014). Differential expression analysis between groups of samples was performed

[1] Illumina® sequencing was performed at the sequencing facility NXTGNT (Ghent University, Ottergemsesteenweg 460, B-9000 Ghent, Belgium).

[2] Data acquisition, quality control and mapping were performed at the sequencing facility NXTGNT (Ghent University, Ottergemsesteenweg 460, B-9000 Ghent, Belgium).

using the edgeR Bioconductor software in R (Robinson et al., 2009). Three separate differential expression analyses were performed. The three different pairwise comparisons were: group A vs group B, group B vs group C and group A vs group C. For each separate analysis, the following steps were performed. (1) Only transcripts with a counts per million (CPM) above 1 in at least six samples were retained. (2) A normalization using Bioconductor software edgeR's standard TMM normalization method (Robinson et al., 2009). (3) A general linear model was built, and statistical testing was done using the empirical Bayes quasi-likelihood F-test. Transcripts having a FDR < 0.05 and a $\log_2$FC > 1 or < -1 were considered to be significantly differential. Differential transcripts were annotated using the *de novo* annotated *S. gregaria* midgut reference transcriptome generated in chapter 3.

4.4. Results and discussion

4.4.1. First exploratory analysis of the data

To explore if samples clustered condition-dependently and to detect any outlier samples, a principal component analysis (PCA) was performed. First, all TMM-normalized transcript counts in all 18 samples were log-transformed in order to minimize any internal variation putatively dominating the PCA. This transformation was solely performed with the purpose of clustering information in the dataset, and was not used for later differential expression analyses. PCA is a frequently used data compression method allowing to get insights into the general structure of large datasets. In general, PCA creates new uncorrelated variables, principal components (PCs), that together describe all the variability in the dataset. These PCs are displayed as new orthogonal axes representing the directions where there is most variation. The first PC accounts for as much of the variability in the dataset as possible, and each succeeding component accounts for as much of the remaining variability as possible (Jackson, 1991; Jolliffe, 2011; Jolliffe and Cadima, 2016). In our dataset, the majority of the variability, 48% and 17.7% respectively, could be explained by PC1 and PC2. Therefore, a two-dimensional PCA plot of all samples was generated (Fig. 4.3).

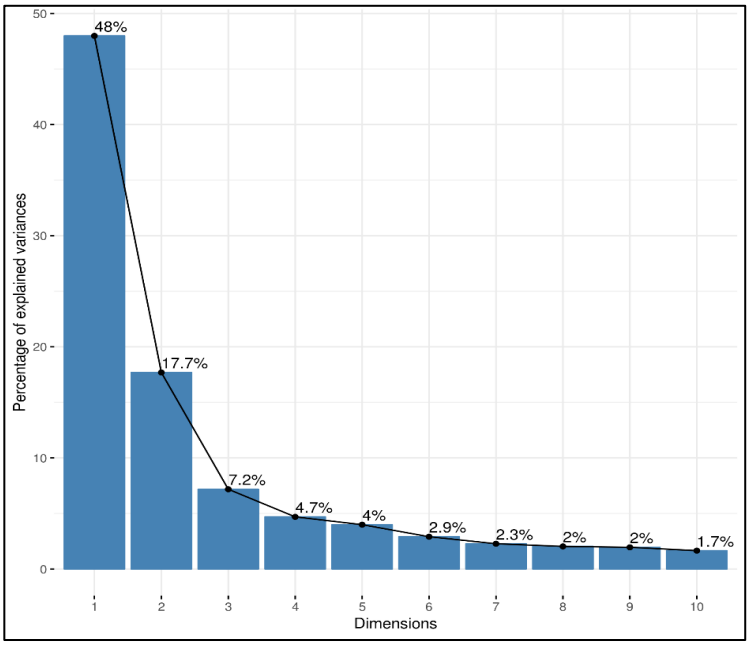


Figure 4.3. Scree plot displaying the percentages of variance between all samples explained by the different PCs (dimensions).

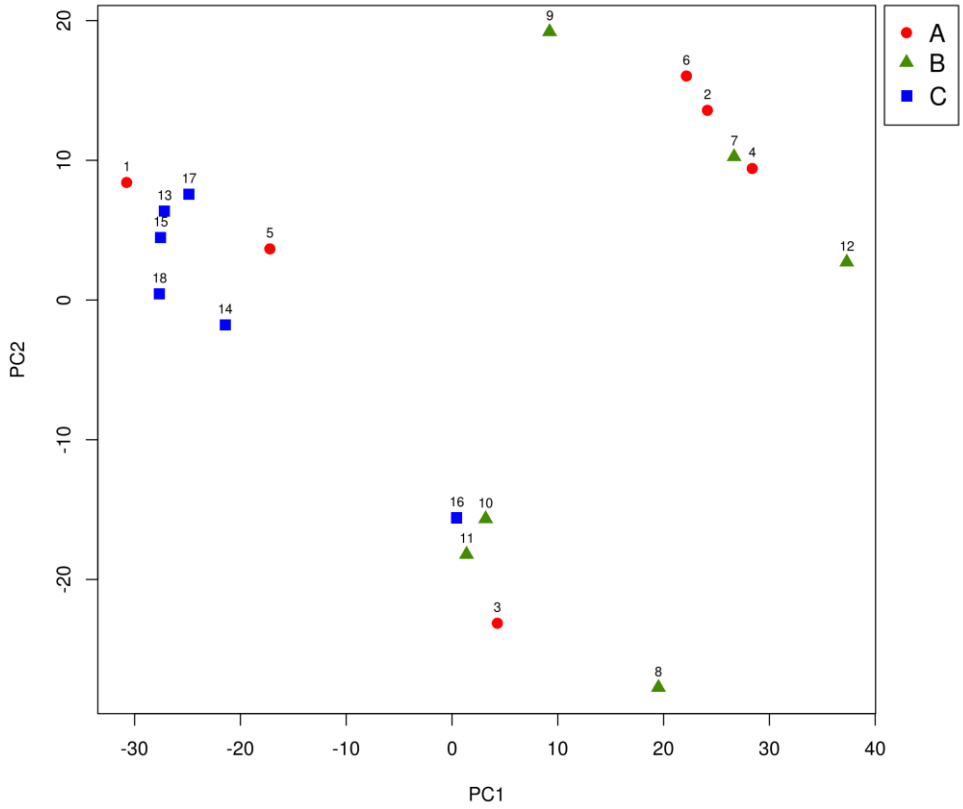


Figure 4.4. Principal component analysis (PCA) on log2 transformed counts of all samples. PC1 accounts for 48% and PC2 for 17.7% of all variation in the dataset. Group A: 10 minutes after feeding; sample 1-6 (Red dots). Group B: 2 hours after feeding; sample 7-12 (Green triangles). Group C: 24 hours after feedings; sample 13-18 (Blue squares).

The PCA in figure 4.4 illustrates that none of the samples clustered in a unique condition-dependent group. Only the samples belonging to group C (blue circles; 24 hours after feeding) represented a more-or-less cohesive group. This condition was possibly least subjected to variation, because the global activity of the midguts was reduced when no food was available.

This drastically changed in response to food entering the alimentary canal, and different biological processes, and hence altering gene expression, were induced. Samples belonging to group A (red circles; 10 minutes after feeding) were the most dispersed in the PCA. Samples 1 and 5, belonging to group A, clustered together with the majority of the samples of group C. This suggested that their transcript profiles were more similar to samples of group C than to the others of group A. All the other samples of group A clustered together with samples of group B, indicating that their transcript profiles were comparable. These observations suggest that at ten minutes after feeding, the midgut transcriptomes can exhibit characteristics of both twenty-four hours and two hours after feeding.

Ten minutes after feeding, food was present in the foregut, while the midgut only contained highly digested remains of a previous meal (Ch2|2.4.2). Consequently, the midgut contents at ten minutes and twenty-four hours after feeding, when no fresh food was introduced to the alimentary tract, were very alike. However, it is possible that at ten minutes after feeding, small amounts of food already entered the midgut, or that the midgut already anticipated at the transcript level to the entrance of the food from the foregut. In those cases, the midgut physiology was probably more similar to that of two hours after feeding. This could explain why some samples of group A clustered together with samples of group B, and others with samples of group C.

Group B and Group C, representing midguts at two and twenty-four hours after feeding, respectively, could be separated on PC1 of the PCA (Fig. 4.5). This is probably the consequence of condition-dependent transcript profiles in both groups. Moreover, the visible differences between midguts of group B and group C were also the most obvious. Two hours after feeding, the majority of the food bolus was present in the midgut lumen, while twenty-four hours after feeding, only highly digested food remains were left (Ch2|2.4.2).

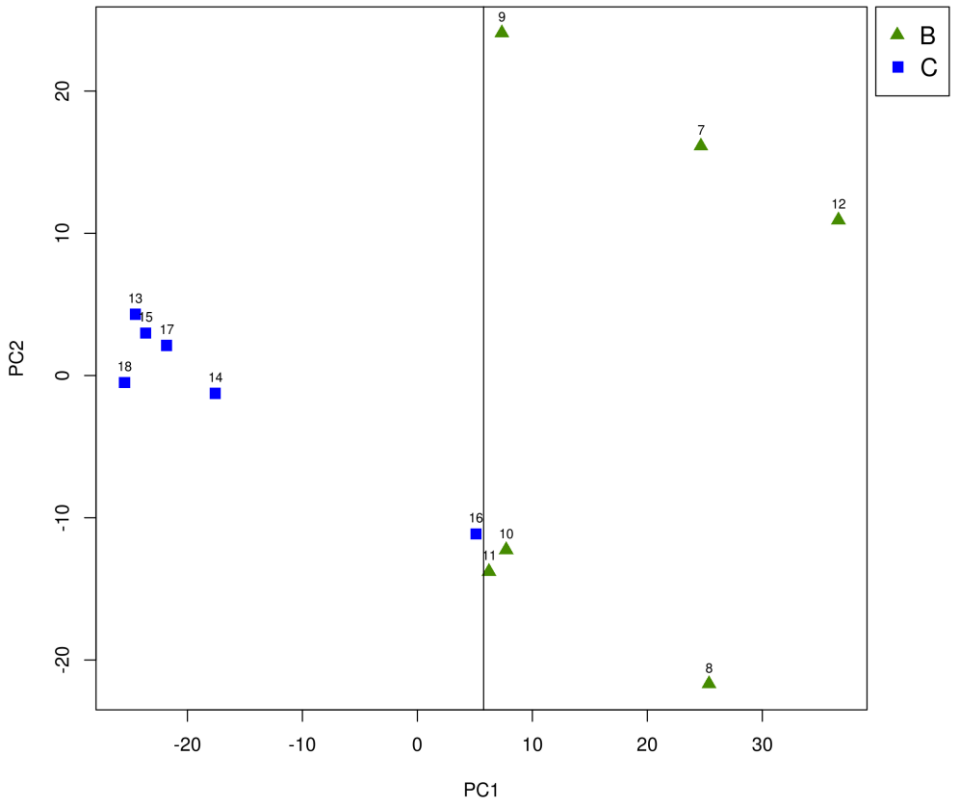


Figure 4.5. Principal component analysis (PCA) on log2 transformed counts of samples belonging to group B and group C. PC1 and PC2 explain 50.4% and 18.1% of all variance between the samples. Group B: 2 hours after feeding; sample 7-12 (Green triangles). Group C: 24 hours after feedings; sample 13-18 (Blue squares). Straight vertical line separates group B and Group C on PC1.

4.4.2. Differential expression analysis

Differential expression was investigated using the Bioconductor package edgeR (Robinson et al., 2009). The analyses were performed in a pairwise fashion, resulting in a total of three different differential expression analyses:

1. Group A vs Group C: 10 minutes after feeding vs 24 hours after feeding.
2. Group B vs Group C: 2 hours after feeding vs 24 hours after feeding.
3. Group A vs Group B: 10 minutes after feeding vs 2 hours after feeding.

For each pairwise analysis, only transcripts with a counts per million (CPM) above 1 in at least 6 samples were retained. Subsequently, the estimated count data of each sample were normalized using edgeR's standard TMM normalization method. The theory behind this was explained in detail in chapter 3 (Ch3|3.2.6). Next, differential expression between the tested conditions was assessed using EdgeR. The statistical model utilized by edgeR to calculate

differential expression is explained in detail paragraph Ch4|4.2.2 of this chapter. Transcripts having an FDR < 0.05 and a log_2FC > 1 were considered to be significantly differential.

EdgeR was unable to detect significantly differentially expressed genes when comparing group A to group C and group A to group B. This was already assumed after the PCA, in which samples of group A grouped together with samples of both group B, being samples 2, 3, 4 and 6, as well as group C, being samples 1 and 5 (Fig. 4.4). The transcript profiles of the midguts at ten minutes after feeding thus appeared to be at intermediary levels between those of midguts at two and twenty-four hours after feeding. Moreover, strong internal variation between biological replicates of group A was observed, possibly resulting from divergent physiological states of the dissected midguts. Where some insects already initiated digestion in their midgut at ten minutes after feeding, others did not, resulting in different transcript profiles within the same condition. Although it is unfortunate that we were unable to detect any significantly differential expression ten minutes after feeding, it is very interesting to notice that the *S. gregaria* midgut appears to be able to very swiftly respond to the presence of food in the midgut; already around ten minutes after feeding.

Nevertheless, a total of 569 and 212 transcripts were found to be significantly up- and downregulated when comparing group B to group C, *i.e.* two hours after feeding to twenty-four hours after feeding. This is graphically summarized in the volcano plot of figure 4.6, in which, for each transcript, the negative log of the FDR was plotted against the log2 of the FC. In this graph, the upper right red dots represent the transcripts that were quantitively more abundant two hours after feeding compared to twenty-four hours after feeding. Upper left, *vice versa*.

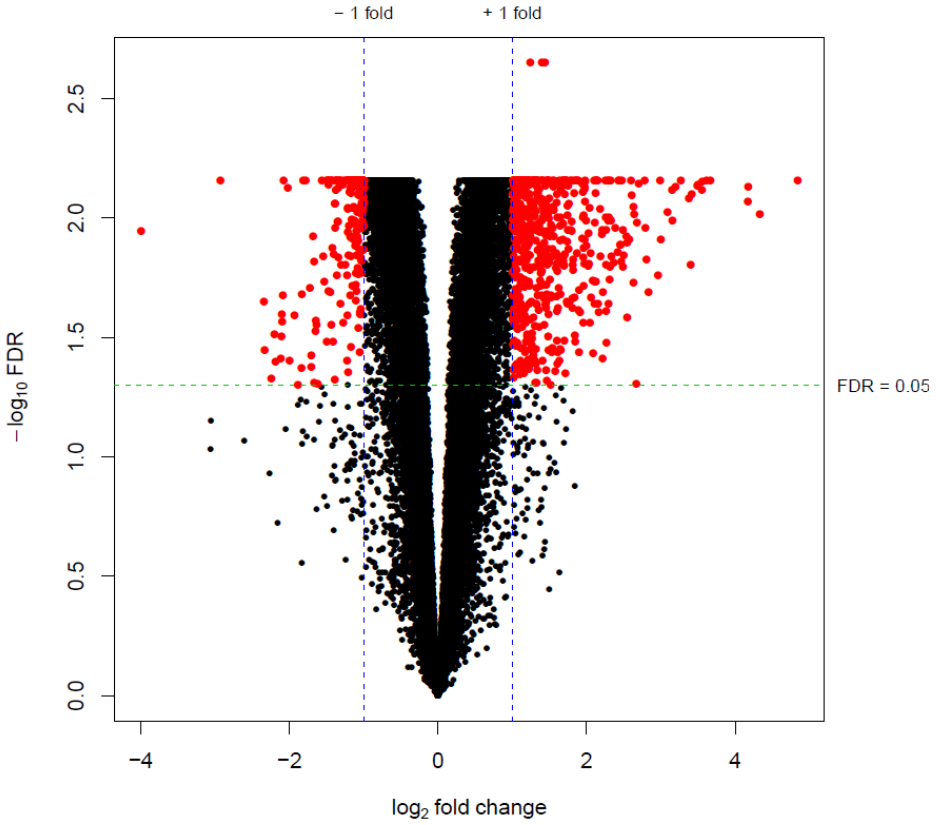


Figure 4.6. Volcano plot of the comparison between group B (2 hours after feeding) and group C (24 hours after feeding). Dots represent the transcripts. For each transcript, the negative log FDR is plotted against the log_2 fold change. Red dots represent differential transcripts. 569 transcripts with log_2 FC > 1 and FDR < 0.05 were considered upregulated 2 hours after feeding (upper right). 212 transcripts with log_2 FC < -1 and FDR < 0.05 were considered downregulated 2 hours after feeding (upper left).

4.4.3. Differential expression analysis between two hours and twenty-four hours after feeding

From the 569 and 212 differentially up- and downregulated transcripts, 283 and 84 had at least one GO term annotation, respectively. These functional annotations were used to analyze the enrichment of up- and downregulated processes in the midgut after feeding (Fig. 4.7). The GO graphs displayed in figure 4.7 can be used as a graphical summary that provides an overview of the general functionalities of differential transcripts two hours after feeding. In general, more processes were predicted to be upregulated than downregulated, which was already expected from the relative difference in differential transcript abundance. Especially the molecular function GO terms associated with catalytic and binding activity were abundantly upregulated two hours after feeding (Fig. 4.7 B). The biological process GO terms associated with biological regulation, metabolic process, and cellular process were also abundantly upregulated two hours after feeding (Fig. 4.7 C). From these graphs, it appears that the majority of the downregulated transcripts also belonged to the above mentioned GO terms. It is to be expected that most of these transcripts were active in other processes than the upregulated transcripts, however this cannot be deduced from these GO graphs and further investigations of the annotation records are thus required.

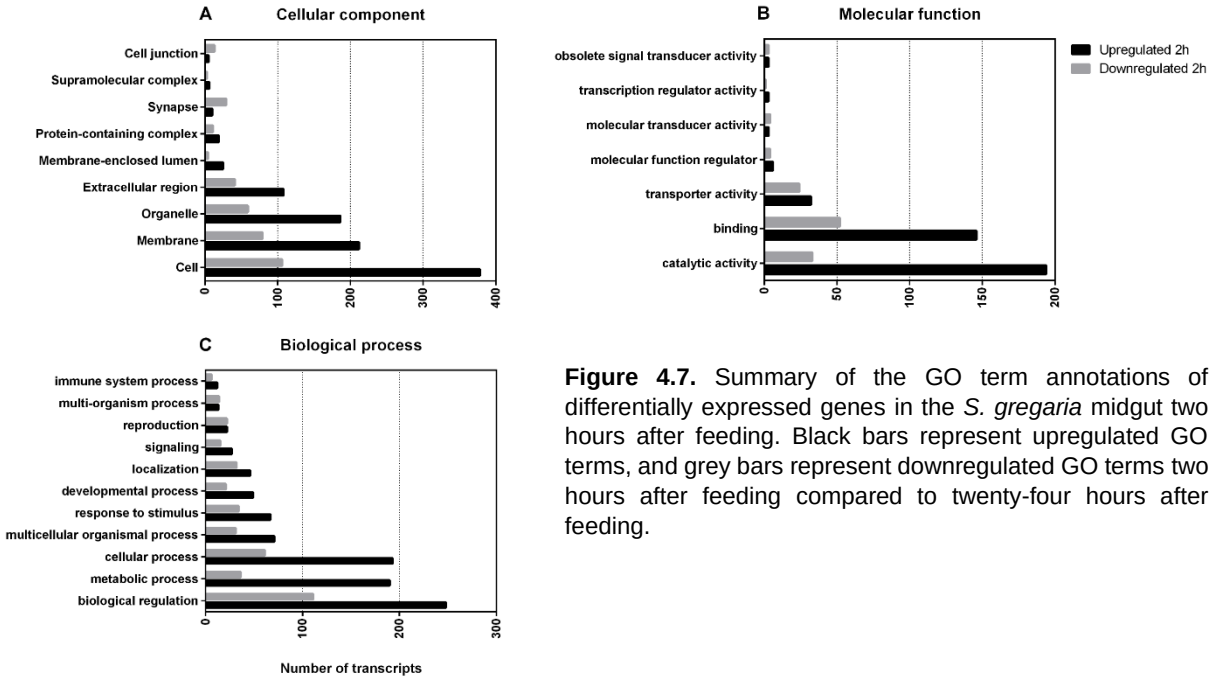


Figure 4.7. Summary of the GO term annotations of differentially expressed genes in the *S. gregaria* midgut two hours after feeding. Black bars represent upregulated GO terms, and grey bars represent downregulated GO terms two hours after feeding compared to twenty-four hours after feeding.

4.4.3.1. Differentially upregulated transcripts two hours after feeding

A total of 569 transcripts had increased in abundance in response to the presence of food inside the midgut, of which many are expected to be involved in digestion-related processes. The *S. gregaria* midgut reference transcriptome (Chapter 3) was used to annotate these transcripts. From the 569 transcripts, a total of 305 transcripts had at least one high-throughput annotation. Unfortunately, this indicates that almost half of the differential transcripts did not retain a high-throughput annotation. The entire list of upregulated transcripts with associated FC, FDR and, if available, high-throughput annotation is available in the supplementary data of this thesis, and can be used by the reader to self-explore the data (Supplementary disk S10).

As already mentioned in chapter 3, high numbers of un-annotated transcripts are fairly common among *de novo* insect transcriptome assemblies. Here we also observed a high number of un-annotated transcripts with induced transcript levels in response to feeding, suggesting that these might perform important functions in the midgut of *S. gregaria* following food uptake. The absence of an annotation record might even indicate a species-specific character, and if so, these transcripts are extremely interesting for further identification and characterization. Although I fully appreciate the potential value of these transcripts, for the remainder of this thesis, I will focus on the annotated transcripts and their putative roles in digestion. Furthermore, I believe that describing and later characterizing any of the annotated transcripts in this list of differential transcripts will only improve our confidence of investigating any un-annotated transcript in the future.

First, differentially upregulated transcripts two hours after feeding were classified according to the Pfam protein families they were predicted to belong to. A total of 239 transcripts could be ascribed to at least one Pfam family (Table 4.1).

Table 4.1 | Summary of Pfam families of upregulated transcripts two hours after feeding.

CLASSIFICATION	PFAM	# DEGS	DESCRIPTION
Peritrophins	PF01607	17	Chitin binding Peritrophin-A domain
Enzymatic digestion	PF00089	20	Trypsin
	PF00232	15	Glycosyl hydrolase family 1
	PF01055	8	Glycosyl hydrolases family 31
	PF00151	5	Lipase
	PF01433	3	Peptidase family M1 domain
	PF11838	3	ERAP1-like C-terminal domain - Aminopeptidase family
	PF05577	1	Serine carboxypeptidase S28
	PF00128	1	Alpha amylase, catalytic domain
	PF02449	1	Beta-galactosidase
	PF00686	1	Starch binding molecule
	PF02837	1	Glycosyl hydrolase family 2
	PF00759	1	Glycosyl hydrolase family 9
	PF02055	1	Glycosyl hydrolase family 30
	PF01301	1	Glycosyl hydrolase family 35
	PF01074	1	Glycosyl hydrolase family 38, N-terminal domain
	PF04116	1	Fatty acid hydroxylase superfamily
	PF16884	1	N-terminal domain of oxidoreductase
Detoxification	PF00135	15	Carboxylesterase family
	PF00201	6	UDP-glucoronosyl and UDP-glucosyl transferase
	PF06757	4	Insect-allergen-repeat protein, nitrile-specifier detoxification
	PF00106	4	Short chain dehydrogenase
	PF00248	3	Aldo/keto reductase family
	PF02798	3	Glutathione S-transferase, N-terminal domain
	PF04488	2	Glycosyltransferase sugar-binding region
	PF01762	2	Galactosyltransferase
Nutrient absorption	PF00083	6	Sugar (and other) transporter
	PF01490	4	Transmembrane amino acid transporter protein
	PF00061	2	Cytosolic fatty-acid binding protein family
	PF16414	2	Niemann-Pick C1 N terminus
	PF00061	1	Fatty acid binding protein family
	PF00854	1	Proton-dependent oligopeptide transporter family
	PF13906	1	C-terminus of amino acid permease

Membrane pumps	PF07690	3	Major Facilitator Superfamily
	PF00664	1	ABC transporter transmembrane region
	PF00005	1	ABC transporter
	PF01061	1	ABC-2 type transporter
	PF00999	1	Sodium/hydrogen exchanger family
	PF03619	1	Organic solute transporter Ostalpha
	PF01496	1	V-type ATPase a subunit family
Other	PF00011	4	Heat shock protein 20 (Hsp20)
	PF06585	3	Haemolymph juvenile hormone binding protein (JHBP)
	PF07714	2	Protein tyrosine kinase
	PF01370	2	NAD dependent epimerase/dehydratase family
	PF13561	2	Enoyl-(Acyl carrier protein) reductase
	PF01683	2	EB domain – Function unknown
	PF00226	2	Chaperone DnaJ domain
	PF02518	2	Histidine kinase-, and HSP90-like ATPase
	PF16984	1	Group 7 allergen
	PF00012	1	Heat shock protein 70 (Hsp70)
	PF00183	1	Heat shock protein 83 (Hsp83)
	PF13855	1	Leucine rich repeat
	PF00063	1	Myosin head (motor domain)
	PF00076	1	RNA recognition motif
	PF00642	1	Zinc finger C-x8-C-x5-C-x3-H type / RNA-binding
	PF00096	1	Zinc finger, C2H2 type
	PF00390	1	Malic enzyme
	PF16177	1	Acetyl-coenzyme A synthetase N-terminus
	PF00109	1	Beta-ketoacyl synthase
	PF17297	1	Phosphoenolpyruvate carboxykinase, N-terminal domain
	PF03953	1	Tubulin C-terminal domain
	PF01553	1	Acyltransferase
	PF00085	1	Thioredoxin
	PF08211	1	Cytidine and deoxycytidylate deaminase
	PF00330	1	Aconitase family (aconitate hydratase)
	PF01869	1	BadF/BadG/BcrA/BcrD ATPase family
	PF00352	1	Transcription factor TFIID (or TATA-binding protein, TBP)
	PF02460	1	Patched family / membrane receptor for Sonic Hedgehog
	PF13637	1	Ankyrin repeats (many copies)
	PF01569	1	Phosphatidic acid phosphatase superfamily
	PF04193	1	PQ loop repeat
	PF13306	1	BspA type Leucine rich repeat region (6copies)domain
	PF01564	1	Spermine/spermidine synthase domain
	PF01510	1	N-acetylmuramoyl-L-alanine amidase
	PF07084	1	Thyroid hormone-inducible hepatic protein Spot 14
	PF06463	1	Molybdenum Cofactor Synthesis C
	PF06293	1	Lipopolysaccharide kinase (Kdo/WaaP) family
	PF00560	1	Leucine Rich Repeat
	PF00501	1	AMP-binding enzyme
	PF04055	1	Radical SAM superfamily
	PF05368	1	NmrA-like family
	PF04922	1	DIE2/ALG10 family
	PF02668	1	Taurine catabolism dioxygenase TauD, TfdA family
	PF03009	1	Glycerophosphoryl diester phosphodiesterase family
	PF17284	1	Spermidine synthase tetramerisation domain
	PF13606	1	Ankyrin repeat
	PF00389	1	D-isomer specific 2-hydroxyacid dehydrogenase
	PF07991	1	Acetohydroxy acid isomeroreductase
	PF13353	1	4Fe-4S single cluster domain
	PF12796	1	Ankyrin repeats(3 copies)
	PF12349	1	Sterol-sensing domain of SREBP cleavage-activation
	PF09335	1	SNARE associated Golgi protein

	PF08659	1	KR domain part of polyketide synthases
	PF13664	1	Domain of unknown function (DUF4149)
	PF16012	1	Domain of unknown function (DUF4780)
	PF06155	1	Protein of unknown function (DUF971)
	PF05705	1	Eukaryotic protein of unknown function (DUF829)

This table clearly demonstrates that transcripts predicted to contribute to various aspects of the digestive process were upregulated two hours after food uptake. Amongst others, transcripts encoding proteins contributing to the enzymatic digestion, nutrient uptake and detoxification of xenobiotics are highly represented in Table 4.1. In the next few paragraphs, upregulated transcripts with an expected or described link to extracellular digestion in insects will be defined in more detail based on their available high-throughput annotations.

Peritrophins

A total of seventeen transcripts encoding putative peritrophins (Pfam01607) were upregulated two hours after feeding (Table 4.2, Supp. Fig. S4.1). Peritrophins are chitin binding proteins (CBPs) crucial for the establishment of the peritrophic membrane (PM) (Chapter 1). The PM in insects envelops the food bolus during its passage through the midgut to increase digestive efficiency and to protect the midgut epithelial membrane. The upregulation of high numbers of peritrophins in response to the presence of food in the gut is a widely described phenomenon in insects (Lehane, 1997). This result demonstrated that the *S. gregaria* larvae were able to swiftly induce the expression of a large array of peritrophin encoding genes upon food uptake.

Table 4.2 | Summary of differentially upregulated peritrophin encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR98067|c0_g1	4.172450418	4.325603906	Peritrophin 1
TR102591|c1_g1	2.641130923	6.058407569	Peritrophin 1
TR52696|c1_g15 *	2.536348628	6.693028962	Peritrophin 1
TR98067|c0_g2	2.460845637	4.704511174	/
TR100696|c1_g4 *	2.299878139	5.661780536	Peritrophin 1
TR68297|c0_g1 *	2.296897555	3.536703918	/
TR52696|c1_g14	2.220537718	2.852976215	Probable CBP 10
TR56571|c0_g3 *	2.119278746	4.466490304	Peritrophin 1
TR52696|c1_g2	2.085477212	7.14872373	Peritrophin 1
TR100696|c0_g4 *	1.941643596	2.512669307	Peritrophin 1
TR52696|c1_g5	1.604635994	6.901222363	Peritrophin 1
TR52696|c1_g7	1.457532697	6.748202496	Probable CBP 10
TR70703|c0_g1	1.452524771	9.891773792	Peritrophin 1
TR56571|c0_g1 *	1.37956144	11.63805384	Peritrophin 1
TR56571|c0_g2 *	1.304014937	7.54942339	Peritrophin 1
TR52696|c1_g9	1.265406215	7.508129722	Probable CBP 10
TR68798|c0_g1	1.175101603	4.384178836	Peritrophin 1

* TR52696|c1_g15 (8), TR100696|c1_g4 (5), TR68297|c0_g1 (2), TR56571|c0_g3 (2), TR100696|c0_g4 (2), TR56571|c0_g1 (11), TR56571|c0_g2 (2) are SuperTranscripts. Between parentheses is the number of originating transcript isoforms.

Endopeptidases

Twenty transcripts encoding putative digestive endopeptidases were differentially upregulated two hours after food uptake. All were predicted to exhibit serine protease-like activity (Pfam00089) (Table 4.3, Supp. Fig. S4.2). Interestingly, only eight transcripts were predicted to code for active serine proteases based on their amino acid sequence, as explained in chapter 3 (Ch3|3.4.2.2.1). Consequently, the other twelve transcripts did not possess all the conserved domains necessary for serine protease activity. Nevertheless, their induced abundance in response to the presence of food in the midgut, suggested that they might, in some way, contribute to the digestive process. Further investigations are required to fully understand their putative roles in the midgut of *S. gregaria.*

Table 4.3 | Summary of differentially upregulated serine protease encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION	PREDICTED ACTIVITY IN CHAPTER 3
TR69767|c0_g1 *	2.572320643	1.646216911	Trypsin-1	/
TR68951|c0_g1 *	2.215386701	1.734311351	Trypsin-1	/
TR70075|c0_g1	2.184449564	1.102668285	Trypsin eta	/
TR89726|c0_g1 *	2.040133643	6.737878953	Trypsin	/
TR64781|c0_g2	1.957306308	2.340724662	Trypsin-1	/
TR42553|c0_g1 *	1.900141648	7.766546072	Collagenase	Chymotrypsin
TR82403|c2_g5	1.709560967	3.86615692	Chymotrypsin BII	/
TR55943|c1_g1 *	1.648940764	5.082040557	Trypsin delta/gamma-like protein	/
TR64781|c0_g1	1.60779006	3.942606549	Trypsin-7	/
TR55943|c3_g1 *	1.516158942	7.410049326	Trypsin-1	Trypsin
TR67399|c0_g1	1.332986509	4.179482903	Trypsin-7	/
TR68929|c0_g1	1.305019796	6.031463222	Vitellin-degrading protease	/
TR60658|c0_g1	1.239178822	2.65927194	Trypsin-1	Chymotrypsin
TR55871|c1_g2 *	1.238779242	3.081712554	Trypsin-7	Chymotrypsin
TR52278|c1_g2	1.21258586	5.513350151	Elastase-1	/
TR106489|c0_g1	1.160863658	2.705844256	Trypsin-3	Chymotrypsin
TR86989|c0_g1 *	1.114663766	3.368465806	Mite allergen protein	Chymotrypsin
TR82403|c2_g3 *	1.110382787	7.281529546	Chymotrypsin BI	/
TR91215|c0_g1	1.064024529	9.977839223	Trypsin-1	Chymotrypsin
TR60303|c0_g2	1.042017693	3.141113401	Trypsin-1	Chymotrypsin

Transcripts in bold: enzymes encoded by these transcripts were predicted to be enzymatically active according to *in silico* investigation of the amino acid sequence (Chapter 3). Last column indicates their predicted activity based on their amino acid sequence. * TR69767|c0_g1 (5), TR68951|c0_g1 (8), TR89726|c0_g1 (7), TR42553|c0_g1 (3), TR55943|c1_g1 (2), TR55943|c3_g1 (4), TR55871|c1_g2 (2), TR86989|c0_g1 (2), TR82403|c2_g3 (7) are SuperTranscripts. Between parentheses is the number of originating transcript isoforms.

One transcript (TR86989|c0_g1) with predicted chymotrypsin activity was annotated as a mite allergen protein, because its amino acid sequence shows resemblance to a serine protease extracted from house dust mites. It has been demonstrated that the stimulation of the protease-activated receptor-2 (PAR-2) by this type of serine proteases can significantly contribute to the onset and progression of airway inflammatory diseases, such as asthma and allergies (Knight et al., 2001; Ock et al., 2005). Also in the cockroach, *B. germanica*, a similar serine protease was extracted from the midgut. It was demonstrated that this cockroach serine protease could also activate PAR-2, similarly to the house dust mite serine protease (Ock et al., 2005). Interestingly, cockroach and locust allergies are frequently reported types of allergies,

especially among workers regularly exposed to these insects under laboratory conditions (Lopata et al., 2005). It is possible that serine proteases originating from the gut and faeces actively contribute to the development of these allergies by activating PAR-2 in the recipient. Therefore, the appropriate measures, such as wearing a dust mask, are advisable when working with these insect species. From my own experience of working with locusts through the course of this PhD research, I have noticed that allergy-related symptoms appear to be most induced when dissecting the gut.

Exopeptidases

Most digestive exopeptidases are expected to be active in the ectoperitrophic space, mediating the final steps of proteolytic digestion (Chapter 3). A total of eight transcripts encoding exopeptidases was upregulated two hours after feeding (Table 4.5, Supp. Fig. S4.3).

Table 4.5 | Summary of differentially upregulated exopeptidase encoding transcripts different two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION	PREDICTED ACTIVITY
TR81225|c0_g1	2.148656507	3.1528777686	Membrane alanyl aminopeptidase	Aminopeptidase (M1)
TR63230|c0_g1	1.80520091	4.6050501382	Aminopeptidase	Aminopeptidase (M1)
TR94178|c0_g1 *	1,4079175935	6,5784171695	Zinc carboxypeptidase	Carboxypeptidase (M14)
TR26668|c0_g1	1.323032538	1.8035068308	Puromycin-sensitive aminopeptidase	Aminopeptidase (M1)
TR79410|c0_g2	1.279659082	5.343188059	Aminopeptidase N	Aminopeptidase (ERAP1)
TR67875|c0_g1	1.219585375	4.6178361811	Aminopeptidase N	Aminopeptidase (ERAP1)
TR79607|c0_g1	1.133965349	3.370565173	Carboxypeptidase N	Carboxypeptidase
TR74016|c0_g1	1.12735452	3.7129230917	Aminopeptidase N	Aminopeptidase (ERAP1)
TR53666|c0_g1 *	1.07651299	6.148797271	Putative serine protease	Carboxypeptidase (S28)

* TR94178|c0_g1 (8) and TR53666|c0_g1 (2) are SuperTranscripts. Between parentheses is the number of originating transcript isoforms.

The abundance of six putative aminopeptidase encoding transcripts was induced two hours after feeding. They were all predicted to be part of the M1 peptidase family, among which three were predicted to belong to the family of endoplasmic reticulum aminopeptidase 1 (ERAP1; Pfam11838). In mammals, these aminopeptidases are described to be active in the endoplasmic reticulum (ER) and are involved in the processing and transport of proteins (Nguyen et al., 2011). Knowledge about ERAP1 structure and function in insects is limited. In the Lepidopteran insect pest, *A. Janata*, however, a membrane-bound proteolytic enzyme with high sequence similarity to ERAP1-like proteins was demonstrated to be active in dietary protein digestion (Ningshen et al., 2013). According to the *de novo* annotation record, two ERAP1 coding transcripts upregulated two hours after feeding were predicted to be plasma-membrane bound, and possibly contribute to the proteolytic digestion in the *S. gregaria* lumen. The other one was predicted to be active in the ER. The other three upregulated aminopeptidases (Pfam01433) were all predicted to exhibit extracellular activity, of which two were described to be plasma membrane-bound.

Three putative carboxypeptidases were identified, of which one was annotated to putatively belong to the S28 peptidase family (Pfam05577). This type of endopeptidase is predicted to have intracellular activity, and is believed to play a role in mediating protein turnover (Ferreira et al., 2015). However, in *T. molitor*, a S28 carboxypeptidase could be extracted from the midgut lumen suggesting a role in dietary protein digestion (Goptar et al., 2008). Whether the carboxypeptidases identified in this RNA-Seq analysis exhibit extracellular digestive activity needs to be further investigated *in vivo*.

Carbohydrate digestion

Many transcripts encoding enzymes mediating carbohydrate digestion were more abundant two hours after feeding compared to twenty-four hours after feeding, indicating that the desert locust was able to induce the expression of a large array of carbohydrases in response to the presence of food in the midgut. In total, 29 such transcripts were identified: 12 myrosinase1 transcripts (Pfam00232), 8 myogenesis regulating glycosidase transcripts (Pfam01055), 4 lactase-like transcripts (Pfam00232), 1 maltase transcript (Pfam00128), 1 β-galactosidase transcript (Pfam02449), 1 glucosylceramidase (Pfam02055), 1 β-mannosidase (Pfam02837) and 1 endoglucanase-like enzyme (Pfam00759). The entire list can be found in the supplementary data of this thesis (Supp. Table S4.1).

Twelve upregulated transcripts predicted to belong to the GH1 family (Pfam00232), typically containing the β-glucosidases, were annotated as myrosinase1 proteins. Endogenous insect myrosinases have already been identified in the intestinal tract of several insect species. For example, in *T. castaneum*, high expression levels of different myrosinase1 genes were observed in the midgut (Oppert et al., 2018). Structural characterization of a myrosinase1 purified from the gut of the cabbage aphid, *Brevicoryne brassicae*, confirmed its strong similarity to animal β-glucosidase (Jones et al., 2002). Interestingly, myrosinase activity has already been observed in the midgut of *S. gregaria* in response to feeding from the cabbage *S. purpurea* (Mainguet et al., 2000). In that study, it was demonstrated that the synthesis and release of endogenous myrosinase proteins was induced as part of a counter-defense mechanism against the glucosinolate-myrosinase anti-herbivory defense system of the host plant. A similar defensive mechanism might possibly explain the increased myrosinase1 transcript levels observed in our RNA-Seq study. It is however also possible that myrosinases, as structurally active β-glucosidases, assist in the digestion of dietary carbohydrates in the midgut, although this needs to be further investigated. Another large group of enzymes with induced transcript levels two hours after feeding were annotated as myogenesis-regulating glycosidases, belonging to the GH31 family (Pfam01055). These enzymes appear to display all necessary structural domains for α-glucosidase activity (Datta et al., 2009). However, until today, their catalytic activity has not yet been experimentally characterized. Their presence in the *S. gregaria* midgut, as well as their induced expression after food uptake, suggests a role in the dietary carbohydrate digestion.

Glucosylceramidases are described to catalyze the hydrolysis of glucosylceramide, important glycolipids present in the lipid bilayers of eukaryotic cells, into free ceramide and glucose (Sugawara et al., 2010). Their observed upregulation upon feeding as well as their ability to generate free glucose molecules inside the midgut lumen suggest that these enzymes possibly contribute to the carbohydrate metabolism in *S. gregaria*. However, this has not yet been described in any insect species and needs to be further investigated.

One transcript was annotated to encode a glucanase-like enzyme belonging to the GH9 family (Pfam00759). As already described in chapter 1 and chapter 3 of this thesis, these proteins, together with β-glucosidases, are demonstrated to exhibit cellulase activity. The observation that the transcript levels of a GH9 encoding transcript were induced two hours after feeding, suggests that the *S. gregaria* larvae are also able to utilize endogenous enzymes to assist in dietary cellulose breakdown. Furthermore, the upregulated β-mannosidase possibly also contributes to the breakdown of hemicellulose originating from the plant diet, since these enzymes are described to hydrolyze mannose bonds, which are ubiquitously present in certain types of hemicelluloses (Scrivener et al., 1997; Scully et al., 2013).

Lipid digestion

In contrast to the high amount of induced protease and carbohydrase encoding transcripts, significantly less lipid degrading enzymes were induced in transcript levels two hours after feeding. Five transcripts encoding lipases, two transcripts encoding cholesterol desaturases and one transcript with a weak resemblance to a sterol desaturase were more abundant after food uptake (Supp. Table S4.2).

Nutrient transmembrane transporters

A total of twenty putative nutrient transmembrane transporters were identified in the list of annotated transcripts upregulated two hours after feeding. These included transporters for all major types of nutrients present in the food (Table 4.6, Supp. Fig. S4.4).

Table 4.6 | Summary of differentially upregulated nutrient transporter protein encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
Amino acid transport			
TR67832|c3_g1	1.979377877	0.951318169	SLC36 AA transporter
TR59587|c0_g1 *	1.783930186	5.226061179	SLC36 AA transporter
TR42645|c0_g2	1.344884395	4.726081161	SLC36 AA transporter
TR81531|c0_g1	1.030927378	2.612912024	SLC36 AA transporter
TR37582|c1_g2	1.325054799	7.290867034	SLC6 AA transporter
TR39911|c0_g2 *	1.166471802	5.944414531	SLC7 AA transporter
Oligopeptide transport			
TR55312|c0_g1 *	1.030003063	7.493914670	SLC15 peptide transporter
Cholesterol transport			
TR102686|c0_g1	1.71736284	4.781254114	NPC1b
TR48631|c0_g1	1.099632521	0.201753377	NPC1
Fatty acid transport			
TR25769|c0_g1	1.832404323	4.141285854	FABP
TR36332|c0_g3 *	1.329795641	8.372391773	FABP
TR77389|c0_g2	1.294318438	3.239098648	FABP
Monosaccharide transport			
TR63477|c0_g1	1.724551824	1.818073696	GLUT

Disaccharide transport			
TR83533|c0_g1	2.308748488	1.408528522	Tret1
TR93769|c0_g1	2.108001281	2.827259152	Tret1
TR65138|c0_g1	2.054040162	1.131879024	Tret1
TR55075|c0_g1	1.95785285	3.903906663	Tret1
TR34420|c0_g1	1.729000332	4.635280077	Tret1
TR69991|c0_g1 *	1.705327515	3.917422773	Tret1
TR42518|c0_g1 *	1.019374187	3.331240928	Tret1

Abbreviations: SLC = solute carrier; NPC = Niemann-Pick C; FABP = Fatty acid binding protein; GLUT = Facilitated glucose transporter; Tret = Facilitated trehalose transporter. * TR59587|c0_g1 (13), TR39911|c0_g2 (4), TR55312|c0_g1 (2), TR36332|c0_g3 (3), TR69991|c0_g1 (3), TR42518|c0_g1 (8) are SuperTranscripts. Between parentheses is the number of originating transcript isoforms.

Seven transcripts putatively encoding the facilitated trehalose transporter (Tret1) were upregulated two hours after feeding. Tret1 has already been suggested to mediate the absorption of dietary disaccharides from the midgut lumen in insects (Holtof et al., 2019). More research is needed to identify and characterize the different Tret1 proteins *in vivo*. However, for now, their observed induced expression in response to food uptake is a first indication that they might play a role in the uptake of disaccharides from the midgut lumen in *S. gregaria*.

The transcript levels of four transcripts encoding proton-coupled amino acid transporter-like proteins belonging to the SLC36 transporter family were also strongly induced after feeding. Until now, absorption of dietary amino acids from the midgut lumen in insects has mainly been attributed to the action of SLC6 and SLC7 transporters. Especially members of the SLC6 transporter family have been characterized in several insect species (Holtof et al., 2019). In chapter 3, the remarkably high abundance of SLC36 transporters in the *S. gregaria* midgut reference transcriptome was already highlighted (Ch3|3.4.2.2.4). The observed strong upregulation of multiple SLC36 transporter encoding transcripts in response to the presence of food in the midgut further suggested their contribution to the dietary amino acid uptake in *S. gregaria*. In addition, one transcript encoding a Na^+-dependent nutrient amino acid transporter (SLC6) and one transcript encoding a cationic amino acid transporter (CAT) (SLC7) were also upregulated two hours after food uptake. The amino acid transporter activity of members of the SLC7 family has only been characterized in the fat bodies of *A. aegypti* and *D. melanogaster* (Colombani et al., 2003; Hansen et al., 2011). The upregulation of a CAT in response to feeding in the midgut of *S. gregaria* also suggested an analogous function in this tissue.

Increased levels of transcripts encoding the Niemann-Pick C1 homologue b (NPC1b) transporter in response to feeding were observed. Interestingly, a recent study on a wide range of insects has demonstrated that NPC1b appears to be exclusively expressed in the midgut epithelium (Zheng et al., 2018). Since insects are unable to biosynthesize sterols *de novo*, the adequate uptake of cholesterol from the food is essential for the growth and development of any insect. This absorption is probably mediated by NPC1b. Furthermore, also three transcripts encoding fatty acid binding proteins (FABPs) were more abundantly expressed two hours after feeding. These proteins have already been described to be involved in the cellular uptake of free fatty acids released from the digestion of dietary lipids in insects (Majerowicz and Gondim, 2013; Holtof et al., 2019).

Other transmembrane proteins

A total of eight ATP binding cassette (ABC) transporter encoding transcripts were upregulated two hours after feeding (Table 4.7, Supp. Fig. S4.5). Since eukaryotic ABC transporters are

described to only mediate the unidirectional export of substrates, they are not expected to be involved in the absorption of any nutrients from the midgut lumen (Rees et al., 2009). Six upregulated ABC transporter transcripts were further described to belong to the multidrug resistance protein (MRP) family, including the ABC transporter subfamilies ABCB, ABCC and ABCG, generally described to have important functions in xenobiotic metabolism in insects (Ch3|3.4.2.2.5) (Labbé et al., 2011; Sodani et al., 2012; Dermauw and Van Leeuwen, 2014; Qi et al., 2016). Among these MRPs, one was further categorized as a P-glycoprotein 1 (P-gp, ABCB1), which is a type of MRP that preferentially transfers neutral or basic organic compounds out of the cell. An earlier study performed in the host lab already demonstrated the presence of a Pgp efflux transporter in the brain barrier of *S. gregaria* (Andersson et al., 2014). A few other studies on the *in vivo* function of P-gp, conducted in *D. melanogaster*, *Spodoptera exigua*, and *P. xylostella,* could demonstrate the ability of P-gp to excrete certain types of pesticides, especially abamectin and its derivates (Luo et al., 2013; Tian et al., 2013; Zuo et al., 2018). Another upregulated MRP was annotated as MRP49 (ABCB1). Up to now, this type of MRP has only been characterized in *D. melanogaster,* in which specific knockout studies demonstrated its role in the resistance against certain pesticides (Denecke et al., 2017). Three other upregulated MRPs were annotated as MRP1s (ABCC1), which are generally described to transport organic anions across membranes (Borst et al., 2000). To the best of my knowledge, only one study describing the role of MRP1 in insects exists. In *D. melanogaster*, it was demonstrated that MRP1 stimulates the resistance against the pesticide DTT, indicating that these proteins also play a role in insecticide detoxification (Gellatly et al., 2015). Finally, two other upregulated ABC transporter transcripts were predicted to belong to the ABCG subfamily. Only recently, it has been demonstrated that an ABCG encoding gene is responsible for Bt toxin resistance in the Asian corn borer, *Ostrinia furnacalis* (Zhang et al., 2017).

Table 4.7 | Summary of differentially upregulated ABC transporter protein encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR12891|c0_g1	1.747893046	1.677950574	ATP-binding cassette sub-family G member 1 (ABCG1)
TR55774|c1_g1	1.493754215	1.322649418	Multidrug resistance protein homolog 49 (MRP49)
TR16343|c0_g1	1.47414058	1.523526518	Multidrug resistance protein; P-glycoprotein 1 (P-gp1)
TR89712|c0_g1	1.421776022	5.381743792	ATP-binding cassette sub-family G member 4 (ABCG4)
TR78861|c0_g1	1.41452713	1.31063235	Multidrug resistance protein 1 (MRP1)
TR23105|c0_g1	1.237846056	1.592478221	Multidrug resistance protein 1B (MRP1B)
TR67617|c0_g1	1.078209937	2.517320331	Multidrug resistance-associated protein 9
TR80358|c0_g1	1.070611366	1.85518941	Multidrug resistance protein 1 (MRP1)

One H^+ V-ATPase subunit a (Pfam01496) encoding transcript was upregulated two hours after feeding (Table 4.8, Supp. Fig. S4.6). The H^+ V-ATPase subunit a is part of the V_0 transmembrane region and is crucial for the V_1/V_0 subunit interaction and hence the proper formation of this transmembrane pump system (Cotter et al., 2015). In general, H^+ V-ATPase complexes are responsible for the establishment of specific membrane potentials stimulating the trafficking of substances across the epithelial membrane (Ch3|3.4.2.2.5 and Ch5|5.1.2). The swift upregulation of a transcript encoding a crucial component of the H^+ V-ATPase complex after feeding suggests its increased importance during digestion and dietary nutrient uptake. Additionally, a transcript belonging to the sodium/hydrogen exchanger family (Pfam00999) was also observed to be upregulated two hours after feeding, albeit with very low

CPM (Table 4.1). These transmembrane proteins putatively work together with the H^+ V-ATPase pumps to accumulate Na^+ molecules in the midgut lumen, consequently supporting the absorption of nutrients via sodium driven nutrient transporters. A similar mechanism is observed in the goblet cells of *M. sexta*, where collaboration between H^+ V-ATPase pumps and $2H^+/K^+$ exchangers create net K^+ electrochemical gradients which are used by the epithelial transporters to drive amino acid absorption into the cells (Castagna et al., 1998).

Table 4.8 | Summary of a differentially upregulated H^+ V-ATPase subunit a encoding transcript two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR39983|c0_g2 *	1.229345517	7.4025206198	V-type proton ATPase 116 kDa subunit a
TR44784|c0_g1	1.090715908	0.058940432	Sodium/hydrogen exchanger 8

* TR39983|c0_g2 (2) is a SuperTranscript. Between parentheses is the number of originating transcript isoforms.

Detoxification

The above described ABC transporters are predicted to be involved in phase III xenobiotic metabolism. Additionally, various transcripts encoding proteins active in phase I and II xenobiotic metabolism were identified among the upregulated transcripts (Table 4.9, Supp. Fig. S4.7). This clearly demonstrates that the desert locust actively detoxifies its ingested food in the midgut by swiftly increasing the transcript levels of several detoxification enzymes. This is well within the line of expectations, since herbivorous insects have already been described to possess numerous mechanisms to overcome plant defenses (Mello and Silva-Filho, 2002).

Most remarkable are the four upregulated transcripts predicted to belong to a protein family of insect-allergen-repeat proteins (Pfam06757). This family of proteins is expected to have resulted from co-evolution between herbivorous insects and their host plants. Almost all plants of the order Brassicales, including the cabbage fed to our locust colony, heavily rely on a glucosinolate-myrosinase based defensive mechanism against insect herbivory, by storing inactive glucosylated toxins in their tissues, separately from β-glucosidases, such as myrosinases. Upon insect herbivory and the resulting tissue damage, these myrosinases come into contact with the glucosylated toxins, resulting in the formation of a β-d-glucose and an unstable aglycone; thiohydroximate-O-sulfonate. Spontaneous reorganization of this aglycone intermediate results in the release of a sulfate ion and the formation of toxic metabolites, such as isothiocyanate (Barba et al., 2016). However, insects have developed several counter-defensive mechanisms against this plant anti-herbivory system. One such mechanism is based on the nitrile-specifier proteins (NSPs), also referred to as insect-allergen-repeat proteins. NSPs are structurally different from any other known detoxifying enzyme and are described to promote the formation of simple nitriles from the ingested aglycone intermediates, which can be further metabolized or excreted with the faeces (Burow et al., 2006). High levels of NSPs have been identified in the midguts of several insect species, including *B. mori*, *B. germanica* and *T. castaneum* (Fischer et al., 2008). Also in *L. migratoria*, different copies of insect-allergen related proteins were identified in the midgut (Spit et al., 2016a). The *S. gregaria* midgut reference transcriptome also contains many insect-allergen-repeat protein coding transcripts (11 unique in total). The observed upregulation of several of these transcripts after eating cabbage may suggest a detoxifying function in the *S. gregaria* midgut. Interestingly, members of the NSP family isolated from other insect species, including *A. gambiae* and *B. germanica*, have demonstrated binding capacity with human immunoglobulin E (IgE), provoking an allergic reaction in atopic persons (Wang et al., 1999). Locust NSPs are expected to provoke allergies in a similar manner.

In addition, a total of six UGT (Pfam00201) and three GST (Pfam02798) encoding transcripts were more abundantly present in the midgut two hours after feeding. These enzymes are probably also utilized by the locust to detoxify the high amounts of ingested glucosinolate-derived toxins, typically produced by Brassicaceae plants. Together with the upregulated NSP encoding transcripts, these data clearly indicate that our locust colony exhibits a strong ability to detoxify these major toxic components present inside their daily diet.

Table 4.9 | Summary of differentially upregulated transcripts encoding the most general types of detoxification enzymes two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
Cytochrome P450			
TR84160|c0_g2	1.556645372	3.542774805	CYP6J
TR104907|c0_g1	1.456239279	3.29252191	CYP6K
TR78286|c0_g1	1.32325274	4.073988125	CYP4C
TR99929|c0_g1	1.258700027	4.32464478	CYP304A
UDP-glycosyltransferase			
TR46182|c3_g1	3.414567702	3.851581464	UGT
TR90718|c2_g1	2.022542841	2.195526846	UGT
TR41238|c0_g1 *	1.755162415	1.722723701	UGT
TR31766|c0_g1	1.527876817	1.06122072	UGT
TR49921|c0_g3 *	1.109835877	4.172896427	UGT
TR75988|c2_g2 *	1.013761684	5.715463858	UGT
Glutathione S-transferase			
TR87683|c0_g1	1.485343197	1.381226037	GST
TR39022|c3_g3 *	1.160199897	5.848493212	GST
TR31999|c0_g1	1.081588133	4.080286125	GST
Carboxylesterase			
TR95462|c0_g1	2.366501838	2.411421598	Esterase SG1
TR92205|c0_g1	1.768416297	0.489375859	Putative carboxylesterase
TR81487|c0_g2	1.41230202	2.879257669	Esterase FE4
TR98141|c0_g1	1.397026431	1.845612629	Venom carboxylesterase-6
TR107208|c0_g1	1.338826951	2.321089084	/
TR59908|c0_g1 *	1.334934541	3.490548007	Venom carboxylesterase-6
TR63470|c0_g1 *	1.334696799	0.659991544	Carboxylesterase 3
TR3368|c0_g1	1.323205712	1.10004019	Venom carboxylesterase-6
TR82404|c2_g2	1.241824714	1.006732959	Venom carboxylesterase-6
TR91367|c0_g1	1.207160058	4.291832871	Venom carboxylesterase-6
TR48351|c1_g2	1.2068	3.328499237	Esterase E4
TR75066|c1_g1 *	1.164202058	4.689336513	Esterase B1
TR77194|c0_g1	1.112164533	1.147704257	Putative carboxylesterase
TR56539|c1_g1 *	1.107379431	5.977031809	Venom carboxylesterase-6
TR73959|c0_g1	1.007339247	2.421449955	Esterase B1
Insect allergen			
TR39781|c0_g1	1.780627124	4.491097972	/
TR56550|c0_g9	1.546699994	8.779883048	Protein G12
TR56550|c0_g6 *	1.146880374	7.80717078	Protein G12
TR56550|c0_g2 *	1.021782458	10.45747763	Protein G12

Abbreviations: CYP = Cytochrome P450; UGT = UDP-glycosyltransferase; GST = Glutathione S-transferase.
* TR41238|c0_g1 (4), TR49921|c0_g3 (4), TR75988|c2_g2 (3), TR39022|c3_g3 (3), TR59908|c0_g1 (6), TR63470|c0_g1 (2), TR75066|c1_g1 (13), TR56539|c1_g1 (4), TR56550|c0_g6 (5), TR56550|c0_g2 (2) are SuperTranscripts. Between parentheses is the number of originating transcript isoforms.

Many transcripts encoding proteins of the carboxylesterase (CE) family (Pfam00135) were also upregulated upon feeding, indicating that these are probably also exhibiting important roles in detoxifying ingested plant secondary metabolites. The induced expression of CEs upon ingestion of plant toxins has also been observed in other insect species. For example, in the aphid, *S. avenae*, CEs are upregulated in the midgut after eating wheat in order to detoxify ingested indole alkaloids (Cai et al., 2004). Furthermore, different studies have also demonstrated that CE activity is strongly associated with insect resistance against ester-containing pesticides, such as pyrethroids (Sogorb and Vilanova, 2002; Wheelock et al., 2005). This demonstrates that insects are able to rely on similar mechanisms for the detoxification of a large array of natural, as well as synthetic, toxins.

Haemolymph juvenile hormone binding proteins

Another interesting observation was the upregulation of three different transcripts encoding putative haemolymph juvenile hormone binding proteins (JHBPs) two hours after feeding (Table 4.10, Supp. Fig. S4.8). In general, JHBPs in insects are described to facilitate the movement of the lipophilic juvenile hormone (JH) through the aqueous haemolymph, and are important factors mediating the uptake of JH in target tissues (Gilbert et al., 2000). Juvenile hormone (which can occur in different molecular forms) is one of the most ubiquitous and most studied insect hormones, with several well-described functions throughout the lifespan of all insects. However, it is best known for its role as a regulator of larval molting and reproductive maturation (for in-depth reviews, see Jindra et al., 2013, 2015; Cheong et al., 2015; Li et al., 2019). Interestingly, JH has also been demonstrated to be involved in the regulation of digestive enzyme expression in certain insect species (Noriega and Wells, 1999; Bian et al., 2008; Sui et al., 2009; Zhang et al., 2010; Cornette et al., 2013). Earlier work in our lab has also highlighted the JH induced expression of three chymotrypsin-like transcripts in the midgut of *L. migratoria* (Spit et al., 2016b). This will be further elaborated in chapter 6 of this thesis. The same study by Spit and colleagues also demonstrated that several JHBPs, as well as hexamerins, which are specific types of JHBPs, were upregulated as part of the PI induced response in the midgut of *L. migratoria*. It was suggested that the hexamerins were responsible for the increased supply of JH towards the midgut, thereby inducing the expression of specific types of chymotrypsins (Spit et al., 2016b). In this RNA-Seq experiment, the observed upregulated JHBPs were not predicted to be related to hexamerins. However, the transcript *TR89612|c1_g1* showed 83% sequence similarity with the *L. migratoria* transcript *LMC_004216-LM_GM5_00321*, encoding a JHBP that was upregulated in response to PI ingestion (Supplementary data of Spit et al., 2016b). Therefore, it is possible that the JHBPs in *L. migratoria* and *S. gregaria* might also be involved in supporting the JH supply towards the midgut in an analogous manner as the hexamerins, thereby mediating the expression of certain digestive proteases during digestion. More *in vivo* research is required to validate these assumptions.

Table 4.10 | Summary of differentially upregulated JHBP encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR69345|c0_g1 *	1.353678254	0.997625665	Haemolymph juvenile hormone binding protein
TR89612|c1_g1	1.156251949	7.739925381	Haemolymph juvenile hormone binding protein
TR59915|c1_g1 *	1.004894045	7.152291231	Haemolymph juvenile hormone binding protein

* TR69345|c0_g1 (2) and TR59915|c1_g1 (2) are SuperTranscripts. Between parentheses is the number of originating transcript isoforms.

Heat shock proteins

Five transcripts encoding different types of heat shock proteins (HSPs) were upregulated two hours after feeding (Table 4.11, Supp. Fig. S4.9). Moreover, all were characterized by high logFC changes, indicating strongly increased abundance upon feeding. This possibly indicates that the insects were responding to stress stimuli, for example from metabolic stress. It could however also suggest that HSPs are involved in digestion-associated events in the midgut.

Table 4.11 | Summary of differentially upregulated heat shock protein encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR107741|c0_g2	4.330894668	3.38896233	Protein lethal(2)essential for life; Hsp20
TR73100|c0_g4	3.402323232	4.311458315	Protein lethal(2)essential for life; Hsp20
TR36797|c0_g2	2.804129255	2.330079128	Protein lethal(2)essential for life; Hsp20
TR75043|c0_g1 *	2.679951942	8.111635384	Endoplasmic reticulum chaperone BiP; Hsp70
TR45928|c0_g2	2.126765776	9.532854038	Hsp83
TR96056|c0_g1	1.740660823	6.393041175	Protein lethal(2)essential for life; Hsp20

* TR75043|c0_g1 (10) is a SuperTranscript. Between parentheses is the number of originating transcript isoforms.

Most HSP encoding transcripts were annotated as protein “lethal(2) essential for life”. This type of proteins belong to the HSP20 family, also known as small heat shock proteins (sHSPs). These proteins are present in both prokaryotic and eukaryotic organisms and are induced upon heat shock or other environmental stress. The HSP20s have an average molecular weight of 20 kDa, and are described to act as protein chaperones protecting other proteins against stress-induced denaturation and aggregation (Li et al., 2012a). However, until now, no reports on the function of HSP20 in the digestive process of insects exists.

Another highly abundant and strongly upregulated transcript after feeding was predicted to encode HSP70. Interestingly, in *R. prolixus*, the expression of a HSP70 encoding gene was demonstrated to be rapidly upregulated in the midgut during the first hours after feeding on blood. When knocking down the *HSP70* gene, *R. prolixus* insects failed to activate essential signaling pathways involved in the processing of nutrients, resulting in a significantly induced mortality rate (Paim et al., 2016). A similar impairment of digestion was also observed in *A. aegypti* when knocking down *HSP70* (Benoit et al., 2010). Our RNA-Seq study illustrates a similar expression pattern for a HSP70 encoding gene in *S. gregaria* two hours after food uptake, suggesting an important role of HSP70 during digestion in this herbivorous insect.

HSP83 transcript levels were also strongly induced two hours after feeding. A recent study has demonstrated that HSP83 physically interacts with the JH receptor Methoprene-tolerant (Met) in *D. melanogaster* (He et al., 2014). Loss-of-function of HSP83 caused the reduction of JH binding to Met, resulting in the reduced downstream JH signaling via Krüppel homolog 1 (Kr-h1), a conserved downstream element of Met (Minakuchi et al., 2008; Lozano and Belles, 2011). Hence, these data clearly indicated the crucial role of HSP83 for the Met mediated JH signaling in *D. melanogaster* (He et al., 2014). Remarkably, an earlier study in our lab has demonstrated the JH dependent expression of three chymotrypsin-like transcripts in the midgut of *L. migratoria*, as already mentioned above. In that study, it was shown that knocking down Met in this locust species resulted in the significantly reduced expression of these chymotrypsin-like genes (Spit et al., 2016b). This will be elaborated in more detail in chapter 6. In this RNA-Seq study, the observed induced transcript levels of HSP83 encoding transcripts, putatively interacting with Met, together with the previously described upregulated

JHBP encoding transcripts, suggest a possible role for JH in the stimulation of expression of specific digestive proteases the *S. gregaria* midgut.

4.4.3.2. Differentially downregulated transcripts two hours after feeding

A total of 212 transcripts were less abundant two hours after feeding compared to twenty-four hours after feeding in the *S. gregaria* midgut ($Log_2FC < -1$ and $FDR < 0.05$). From this, a total of 88 transcripts had a high-throughput annotation. The entire list of downregulated transcripts two hours after feeding with associated FC, FDR and, if available, high-throughput annotation is available in the supplementary data of this thesis (Supplementary disk S11). Their lower relative abundance in the midgut two hours after feeding compared to twenty-four hours after feeding, suggests that they are less crucial in the response to food uptake in *S. gregaria*. Moreover, some of these transcripts with a higher abundance twenty-four hours after feeding are possibly involved in the midgut's response to (short-term) food deprivation.

First, differentially downregulated transcripts two hours after feeding were classified according to the Pfam protein families they were predicted to belong to. A total of 68 transcripts could be ascribed to at least one Pfam family (Table 4.12).

Table 4.12 | Summary of Pfam families of downregulated transcripts two hours after feeding.

Classification	Pfam	# DEGs	Description
Peritrophins	PF01607	5	Chitin binding Peritrophin-A domain
Enzymes	PF00089	2	Trypsin
	PF00351	1	Biopterin-dependent aromatic amino acid hydroxylase
	PF01019	1	Gamma-glutamyltranspeptidase
	PF00930	1	Dipeptidyl peptidase IV (DPP IV) N-terminal region
	PF01619	1	Proline dehydrogenase
Inhibitors	PF00079	1	Serpin (serine protease inhibitor)
Detoxification	PF00067	6	Cytochrome P450
	PF00106	1	Short chain dehydrogenase
	PF00135	1	Carboxylesterase family
	PF00171	1	Aldehyde dehydrogenase family
	PF16884	1	N-terminal domain of oxidoreductase
Membrane associated	PF00664	3	ABC transporter transmembrane region
	PF00083	2	Sugar (and other) transporter
	PF03522	1	Solute carrier family 12
	PF00860	1	Permease family
	PF00909	1	Ammonium Transporter Family
	PF00916	1	Sulfate permease family
	PF01061	1	ABC-2 type transporter
	PF00155	1	Aminotransferase class I and II
	PF04547	1	Calcium-activated chloride channel
	PF03600	1	Citrate transporter
	PF00061	1	Fatty acid binding protein
	PF00335	1	Tetraspanin family
Other	PF00595	2	PDZ domain
	PF01347	2	Lipoprotein amino terminal region
	PF03392	2	Insect pheromone-binding family
	PF03722	2	Hemocyanin, all-alpha domain
	PF12796	2	Ankyrin repeats (3 copies)

	PF06585	1	Haemolymph juvenile hormone binding protein (JHBP)
	PF00059	1	Lectin C-type domain
	PF00046	1	Homeodomain
	PF00078	1	Reverse transcriptase (RNA-dependent DNA polymerase)
	PF00096	1	Zinc finger, C2H2 type
	PF00041	1	Fibronectin type III domain
	PF00169	1	PH domain
	PF00560	1	Leucine Rich Repeat
	PF00875	1	DNA photolyase
	PF02014	1	Reeler domain
	PF02931	1	Neurotransmitter-gated ion-channel ligand binding domain
	PF04321	1	RmlD substrate binding domain
	PF04500	1	FLYWCH zinc finger domain
	PF05978	1	Ion channel regulatory protein UNC-93
	PF07177	1	Neuralized
	PF09298	1	Fumarylacetoacetase N-terminal
	PF15993	1	Fuseless
	PF12017	1	Transposase protein
	PF12803	1	mRNA (guanine-7-)methyltransferase (G-7-MTase)
	PF13516	1	Leucine Rich repeat
	PF16984	1	Group 7 allergen
	PF00650	1	CRAL/TRIO domain
	PF03729	1	Short repeat of unknown function (DUF308)

Not surprisingly, according to their Pfam classification, none of the downregulated transcripts was predicted to exhibit extracellular digestive or nutrient absorption activity. This is in contrast to the high number of upregulated transcripts with digestive process related functions two hours after feeding (Table 4.1). Next, the downregulated transcripts were defined in more detail by further investigating and describing their available high-throughput annotations. In the next few paragraphs, I will focus on explaining some of the most noteworthy findings.

Endogenous protease inhibitors

Among the downregulated transcripts two hours after feeding, two were predicted to encode for serine protease inhibitors (Table 4.13, Supp. Fig. S4.10). Endogenous serine protease inhibitors are generally part of specific regulatory mechanisms designated to avoid the unwanted activation of intracellular serine protease-dependent cascades, and possible tissue damage that could result from the loose activity of these proteases. In addition, some protease inhibitors are also able to directly protect the organism against harmful pathogens.

Table 4.13 | Summary of differentially downregulated serine protease inhibitor encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR68956|c0_g2 *	-2.929609462	1.684026062	*Schistocerca gregaria* protease inhibitor
TR78634|c0_g1 *	-1.029331438	4.388620717	Leukocyte elastase inhibitor (Serpin)

* TR68956|c0_g2 (8) and TR78634|c0_g1 (2) are SuperTranscripts. Between parentheses is the number of originating transcript isoforms.

The second strongest downregulated transcript after feeding was annotated as a *Schistocerca gregaria* serine protease inhibitor. A detailed *in silico* investigation of its nucleotide sequence revealed that the *S. gregaria* pacifastin-related peptide precursor 5 (SGPP-5) had 91% sequence similarity with the downregulated transcript, TR68956|c0_g2. Pacifastins are canonical serine protease inhibitors found ubiquitously and exclusively among arthropods, and are characterized by a conserved pattern of six cysteine residues Cys-Xaa$_{9–12}$-Cys-Asn-Xaa-

Cys-Xaa-CysXaa$_{2–3}$-Gly-Xaa$_{3–6}$-Cys-Thr-Xaa$_3$-Cys. In the desert locust, seven different pacifastin-related precursors have been identified (Breugelmans et al., 2009b). It was demonstrated that SGPP-5 has high expression levels in the fore- and hindgut, as well as detectable levels in the haemolymph of adult desert locust. A dual function of SGPP-5 was therefore proposed: the protection against cuticle-degrading proteases of invading pathogens ingested via the food and the protection against leaking enzymes from the midgut. Remarkably, SGPP transcripts have not been detected in the midgut of *S. gregaria* (Simonet et al., 2004). However, in a different locust species *L. migratoria*, three pacifastin-related precursor peptides could be detected in the midgut, suggesting their activity in this tissue as well (Clynen et al., 2006; Breugelmans et al., 2009b). Also in the Lepidoptera *B. mori*, the expression of a pacifastin-like protease inhibitor could be identified in the midgut (Breugelmans et al., 2009a). Interestingly, purified pacifastin-related peptides from *S. gregaria* were demonstrated to exhibit strong inhibitory capacity towards the proteolytic activity of *S. gregaria* midgut secretions (Spit et al., 2012). This illustrated that pacifastin-related peptides are strong inhibitors of endogenous proteases, supporting a possible role as protective agents against leaking midgut digestive enzymes.

In this RNA-Seq study, the observed downregulation of SGPP-5 expression in the midgut during digestion suggests that the insects depend less on the inhibitory action of SGPP-5 when food was available. At this point, the digestive proteases in the midgut lumen are targeting the macronutrient peptide bonds of the food, and inhibiting their action would reduce the proteolytic capacity of the midgut juice. Moreover, higher transcript levels of SGPP-5 in the absence of food substrates might also be part of a protective mechanism against serine proteases leaking from the midgut, or to limit the residual serine protease activity in the midgut lumen. In chapter 3, it was already illustrated that most serine proteases are expressed at a constant rate within the timeframe of twenty-four hours after feeding (Table 3.7). However, in the absence of food substrates, these serine proteases need to be inactivated or removed to avoid uncontrolled autocatalysis and potential tissue damage. Additionally, higher abundance of SGPP-5 twenty-four hours after feeding could also be part of an immune response protecting the organism against pathogens left in the midgut. Further *in vivo* investigations are required to address these interesting open questions.

In addition to a pacifastin-like protease inhibitor, the expression of another serine protease inhibitor (serpin) (Pfam00079) was reduced in the midgut two hours after feeding. Serpins constitute a large group of functionally diverse serine protease inhibitors that are able to inhibit a large array of enzymes, including serine proteases, cysteine proteases and cathepsins (Danielli et al., 2003). Most insect serpins are characterized to modulate different innate immune responses, including the Toll or prophenoloxidase (proPO) pathways in the haemolymph, by negatively regulating their protease-dependent activation cascades (Meekins et al., 2017). The expression of serpins in the midgut has been demonstrated in several insect species, but as of yet, their putative functions during digestion have not been broadly described (Meekins et al., 2017). In *M. sexta,* however, complexes of endogenous serpins (serpin-1K) with digestive chymotrypsins were identified in the haemolymph, suggesting that serpin-1K could be involved in the protection of *B. mori* against digestive enzyme leakage from the midgut (Ragan et al., 2010). Interestingly, in *A. gambiae*, specific serpins were demonstrated to be upregulated in the midgut epithelial cells in response to parasite passage through the midgut, suggesting a direct function in parasite tolerance which needs to be further investigated (Danielli et al., 2003; Meekins et al., 2017). Such upregulation of serpin expression in the midgut epithelial cells upon parasite infection was also observed in *A. aegypti* and *Glossina morsitans morsitans* (Bian et al., 2005; Telleria et al., 2014). Lastly, in *D. melanogaster*, two

serpin proteins were found to be upregulated in the haemolymph during starvation (Handke et al., 2013).

Because of their broad substrate spectrum, it is possible that the serpins expressed in midgut were part of a protective mechanism against autocatalysis by secreted digestive enzymes when no food substrates were present. Then, when food was available in the midgut, their expression levels were reduced to avoid unwanted inhibition of digestive proteases. However, the observed downregulation of serpins in the *S. gregaria* midgut two hours after feeding might also suggest that the locust reduces the inhibitory control of protease inhibitors on specific innate immunity pathways in response to food passing through the midgut, possibly to protect itself against harmful pathogens. However, as observed in *A. gambiae*, *A. aegypti* and *G. morsitans morsitan,* some serpins expressed in the midgut might be directly involved in parasite tolerance (Danielli et al., 2003; Abraham et al., 2005; Telleria et al., 2014). Hence, it is also putative that certain serpins protect the insect against residual pathogens possibly left in the midgut after the food has left the body, which would explain their higher transcript levels twenty-four hours after feeding. Altogether, further *in vivo* investigations are needed to fully elucidate the role of serpins during digestion.

Trypsins

Interestingly, two transcripts encoding trypsin-like enzymes were downregulated two hours after feeding (Pfam00089) (Table 4.14, Supp. Fig. S4.11). These transcripts are predicted to be involved in other physiological processes than digestion. One transcript was annotated as “serine protease snake”, the other as “modular serine protease”. Both types of proteases have already been associated with innate immunity by activating the Toll pathway in response to infection with Gram-positive bacteria and fungi (Jiang and Kanost, 2000; Meekins et al., 2017). Our results suggests that the locust midgut invests less in innate immunity during digestion by downregulating the transcript levels of serine proteases that initiate specific immune responses. Contrastingly, the higher abundance of both transcripts twenty-four hours after feeding might, on the other hand, also suggest that these immune responses in the midgut are more important during periods of food deprivation.

Table 4.14 | Summary of differentially downregulated serine protease encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR60258|c0_g1	-1.3473358901	0.8723890192	Serine protease snake
TR84957|c0_g1	-1.0807379435	1.2968873303	Modular serine protease

Protein Takeout

Another interesting transcript that was downregulated two hours after feeding was annotated to encode the protein Takeout (TO) (Table 4.15, Supp. Fig. S4.12). TO is a haemolymph JHBP (Pfam06585) that has already been linked to the circadian control of feeding behavior in *D. melanogaste*r (So et al., 2000). Its gene expression in the head was demonstrated to be downregulated during starvation and was suggested to play a role in energy homeostasis. In an earlier study performed in the host lab, three different TO encoding sequences were identified in the EST database of brain tissue of *S. gregaria,* named *Sg*-TO1, *sg*-TO2 and *Sg*TO3 (Vuerinckx 2012, unpublished data). Interestingly, it was demonstrated that, in contrast to *D. melanogaster*, expression levels of *Sg-TO2* increased in the heads of adult locusts during

starvation. This is analogous to the observation in this RNA-Seq of higher levels of a TO encoding transcript in the midgut of *S. gregaria* larvae twenty-four hours compared to two hours after feeding. Remarkably, the study of Vuerinckx indicated that transcript levels of *Sg-TO2* were negligible in the midgut tissues of adult locusts. Nevertheless, in analogy to earlier findings in both *D. melanogaster* and *S. gregaria* adults, these data suggest that TO might also play a role in energy homeostasis in the midgut of *S. gregaria* larvae. However, further *in vivo* testing is necessary to fully elucidate its function in the insect midgut.

Table 4.15 | Summary of a differentially downregulated protein Takeout encoding transcript two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR42461|c0_g1 *	-1.3723027639	2.9477862799	Protein Takeout

* TR42461|c0_g1 (2) is a SuperTranscript. Between parentheses is the number of originating transcript isoforms.

Protein I'm not dead yet

One transcript annotated as "protein I'm not dead yet" (Indy) was strongly downregulated two hours after feeding (Table 4.16, Supp. Fig. S4.13). This non-electrogenic solute transporter (Pfam03600) has been demonstrated to mediate the cellular uptake of citrate across plasma membranes in *D. melanogaster*. Interestingly, silencing *Indy* in this insect species extended their life span by a mechanism resembling caloric restriction (Wang et al., 2009; Willmes et al., 2018). In the butterfly, *Bicyclus anynana*, expression of *Indy* was induced in response to nutrient deprivation (Pijpe et al., 2011). It was hypothesized by Pijpe and colleagues that the upregulation of *Indy* under starvation in *B. anynana* was essential for the uptake of nutrients mobilized from reserves, crucial for survival (Pijpe et al., 2011). It is possible that INDY plays a similar role in the midgut of *S. gregaria.*

Table 4.16 | Summary of a differentially downregulated "protein I'm not dead yet" encoding transcript two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR67978|c0_g3	-2.0778405027	3.8739728442	Protein I'm not dead yet

Chitin binding Peritrophin-A domain

The transcript levels of as many as five transcripts encoding peritrophin-like proteins (Pfam01607) were downregulated after food uptake (Table 4.17, Supp. Fig. S4.14). In theory, one would foremost expect an increased production of these proteins when food is present in the midgut, as was clearly demonstrated in Table 4.2. The downregulation of peritrophin encoding transcripts upon feeding has not yet been demonstrated in insects. Remarkably, three transcripts showed weak sequence similarity to a putative obstructor-E protein. This type of CBP has only been described in *D. melanogaster*, where it appears to be a pivotal constituent of the larval cuticle (Tajiri et al., 2017). Further research is therefore needed to elucidate the function of these CBPs in the *S. gregaria* midgut and to identify any differences between these proteins and the peritrophins that were induced two hours after feeding.

Table 4.17 | Summary of differentially downregulated peritrophin-like protein encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR69775|c0_g1	-1.814971968	2.991666704	Protein obstructor-E
TR25864|c0_g1	-1.340184597	0.343243118	Protein obstructor-E
TR16735|c0_g1	-1.1799032	1.917589887	Protein obstructor-E
TR60276|c0_g7	-1.172169252	4.555330591	/
TR86896|c0_g1	-1.104186936	2.213622181	/

Insect pheromone binding proteins

Two transcripts encoding insect chemosensory proteins (CSPs) belonging to the insect pheromone-binding family (Pfam03392) were downregulated after feeding (Table 4.18, Supp. Fig. S4.15). Recently, it has been demonstrated in *P. americana* and *S. litura* that gut CSPs play a role in food perception and preference in these insects (Li et al., 2017; Yi et al., 2017). Moreover, high numbers of CSPs have already been identified in both *S. gregaria* and *L. migratoria* and some of these were even predicted to play a role in locust phase transition (Martín-Blázquez et al., 2017). The downregulation of these CSP encoding transcripts two hours after feeding, as observed in the current experiment, might be the consequence of the presence of food in the midgut. In addition, it could be possible that certain CSPs play distinct roles during periods of food deprivation in *S. gregaria*, hence their higher abundance twenty-four hours compared to two hours after feeding.

Table 4.18 | Summary of differentially downregulated insect chemosensory protein encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR36295|c0_g1	-2.019973003	3.1496272097	Ejaculatory bulb-specific protein 3
TR48408|c0_g1	-1.0213656835	2.2464490822	Ejaculatory bulb-specific protein 3

ABC transporter proteins

The transcript levels of five ABC protein encoding transcripts, of which the majority were predicted to be involved in phase III xenobiotic metabolism, were reduced upon feeding (Table 4.19, Supp. Fig. S4.16). It is possible that these ABC transporter proteins are less important during digestion, as compared to the ABC transporters of which the transcript levels were upregulated upon feeding.

Table 4.19 | Summary of differentially downregulated ABC transporter protein encoding transcripts two hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR106329|c0_g1 *	-1.725569242	3.737879906	Multidrug resistance protein 49 (MRP49)
TR79577|c0_g2	-1.105411473	1.299373209	Multidrug resistance protein 4 (MRP4)
TR46161|c0_g2	-1.097492701	0.895089396	Multidrug resistance protein 49(MRP49)
TR102683|c0_g1	-1.09585907	2.7332134	Multidrug resistance protein 4 (MRP4)
TR41001|c0_g1 *	-1.026869839	3.437706595	ATP-binding cassette sub-family G member 20 (ABCG20)

* TR106329|c0_g1 (2) and TR102683|c0_g1 (2) are SuperTranscripts. Between parentheses is the number of originating transcript isoforms.

4.4.3.3. Observations ten minutes after feeding

Supplementary figures S4.1-S4.17 display boxplots of the log2-transformed TMM-normalized CPM of all the above described transcripts at ten minutes, two hours and twenty-four hours after feeding. Remarkably, these figures illustrate comparable expression patterns of these transcripts at ten minutes and two hours after feeding when compared to twenty-four hours after feeding. However, no significant differential expression was detected by edgeR between the conditions ten minutes and twenty-four hours after feeding (Ch4|4.4.2).

EdgeR was probably unable to detect these differential transcripts because of the overall low level of diversity between samples of ten minutes and twenty-four hours after feeding. This especially resulted from samples 1 and 5 of group A being highly similar to the majority of the samples of group C, representing the midguts at twenty-four hours after feeding (Fig. 4.4). Consequently, group A is dominated by strong within-group variation. The consequences of this were dual. Firstly, this large proportion of within-group variation in group A led to less squeezing of the estimated QL transcript-specific dispersions, resulting in the increase of p-values. Secondly, when performing the *post-hoc* FDR calculations, all positive discoveries were rejected. However, when the count data of specific transcripts are observed separately across samples before statistical modelling, as displayed in supplementary figures S4.1-S4.17, differences among conditions do become visible, because this does not take into account the batch effects of the entire mRNA-pool as calculated by edgeR.

As described in paragraph 4.4.2 of this chapter, the transcript profile of midguts dissected at ten minutes after feeding could have been in an intermediate state between the transcript profiles of two hours and twenty-four hours after feeding. Nevertheless, midguts dissected at this timepoint were never filled with fresh food, and were in a clearly distinct physiological state compared to midguts two hours after feeding. The fact that the same transcripts up- and downregulated at two hours after feeding were already displaying a similar trend at ten minutes after feeding, probably illustrates the extraordinary speed by which the midgut operates at the transcript level in response to the uptake of food.

4.5. Conclusions

4.5.1. Experimental design

The general idea of this RNA-Seq study was to investigate several crucial timepoints during digestion in *S. gregaria* larvae. Therefore, three different timepoints were carefully selected based on differences in midgut physiology. Midguts were dissected at ten minutes, two hours and twenty-four hours after food uptake. In theory, each timepoint represented a decisive moment during digestion in the midgut, namely the initial (preparative) phase of digestion, ongoing enzymatic digestion and a (short) period of food deprivation as a reference, respectively, as explained in detail in chapter 2.

A first exploratory analysis of the data using PCA clearly illustrated that the samples originating from the ten minutes after feeding condition (group A) were very dispersed and clustered with samples originating from the two hours (group B), as well as the twenty-four hours (group C), after feeding condition (Fig. 4.4). Because of this strong within-group variation, edgeR was also unable to detect any significantly differentially expressed genes when comparing group A to the other two groups, unfortunately eliminating two of the three possible comparisons.

In this RNA-Seq study, each sample represented midguts originating from six different insects that were subjected to a tightly regulated feeding routine. Based on their physical appearances, the dissected midguts could be clearly distinguished per condition. Furthermore, the dissections were performed in a highly organized manner to avoid technical variation as much as possible. Nevertheless, high proportions of variation could still be observed between biological replicates, demonstrating that the digestive process in the midgut of the desert locust is a highly irregular process, probably strongly subjected to biological variation.

The least internal variation was observed in the samples of group C, at twenty-four hours after feeding. It is assumable that at twenty-four hours after feeding, the midguts were less active at a transcript level, resulting in less variation. In addition, the chances of introducing any variation that would potentially bias the samples were significantly lower, because no food was present to stimulate any digestion-associated process in the midgut. This was clearly different for the other two conditions, in which variation could be more easily introduced. For example, minuscule food particles and fluids could have already been passed on to the midgut in one insect at ten minutes after feeding, while another insect's midgut still only contained the highly/fully digested remains of a previous meal. In that case, both insects belonging to the same condition were expected to have different transcript profiles, and extra variation in the RNA-Seq dataset was introduced. Moreover, in general, the midguts of insects belonging to group A and B were expected to be more active at these timepoints and consequently biological variation might be more inherent.

In retrospect, it could be possible that the expected differences at transcript level in the midgut at ten minutes after feeding on the one hand and at two and twenty-four hours after feeding on the other hand were overestimated in advance. In fact, one could argue that at ten minutes after feeding, the physiological state of the midgut may have intermediate characteristics, corresponding to both other conditions. This RNA-Seq study has demonstrated that also the transcript profiles of the midguts at these timepoints have shared properties. Moreover, interesting transcript changes might be occurring in other tissues during this short period after feeding, and are worth investigating.

4.5.2. Differential expression analyses

The most visible differences in midgut physiology were observed between group B and C, namely at two hours and twenty-four hours after feeding. The midguts of the animals of group B were completely filled with food, while those of group C were almost empty, containing only highly/fully digested remains of a previous meal. This was also the only comparison in which we were able to detect differential gene expression. Using the Bioconductor package edgeR, a total of 569 and 212 transcripts were found to be significantly more abundant two hours and twenty-four hours after feeding, respectively (Fig. 4.6). From these transcripts, respectively 305 and 88 had a high-throughput annotation in the *S. gregaria* midgut reference transcriptome.

Several transcripts associated with the digestive process were upregulated two hours after feeding, and were selected for more detailed *in silico* investigations. Transcript levels of transcripts involved in PM construction, enzymatic digestion, nutrient absorption and detoxification were strongly induced when food was present inside the midgut lumen. These data clearly demonstrated that the desert locust was able to swiftly induce the expression of a large array of digestion-related genes in response to food availability. Remarkably, when analyzing the normalized count data of the investigated upregulated transcripts across all

samples, it could be observed that these transcripts were generally also more abundant in the midgut ten minutes after feeding compared to twenty-four hours after feeding, despite not being called by edgeR. This suggests that the midgut is able to respond very swiftly at the transcript level to the uptake of food.

In contrast, no genes predicted to be involved in enzymatic digestion or nutrient uptake were downregulated two hours after feeding, indicating that the organism fully invests in the digestive (and associated) process(es). The most noteworthy findings at this timepoint were the reduced transcript levels of several serine protease inhibitors, trypsins putatively involved in immunity, a TO protein, and Indy. Moreover, it is not excluded that the proteins encoded by these transcripts play important roles during periods of food deprivation. They might, for example, be involved in mediating stress tolerance, tissue maintenance/protection, or immune defenses during starvation. Again, separate analyses of the normalized count data of these transcripts across all samples also demonstrated their reduced abundances ten minutes after feeding, albeit at a more intermediate level between group B and C.

While this chapter mostly focused on summarizing so-called digestion-related transcripts, which are expected to actively mediate different aspects of digestion, Table 4.1 and Table 4.12 also clearly demonstrate that several other transcripts, with distinct predicted functions, were up- and downregulated upon feeding, and might be interesting for future studies involving the *S. gregaria* midgut. The complete annotation records of annotated up- and downregulated transcripts are available in supplementary disk S10 and S11.

Finally, this RNA-Seq study clearly illustrated the contrasting gene expression in the *S. gregaria* midgut at two hours *versus* twenty-four hours after feeding. When food was available in the midgut, the insects heavily invested in digestion-related processes, such as enzymatic digestion, nutrient uptake and xenobiotic metabolism. This chapter summarized several interesting differentially expressed genes two hours after feeding. These can now be further investigated *in vivo* to better understand their role in the digestive physiology of the desert locust, and eventually to validate their potential as insecticidal targets.

4.6. References

Abraham EG, Pinto SB, Ghosh A, Vanlandingham DL, Budd A, Higgs S, Kafatos FC, Jacobs-Lorena M, Michel K (2005) An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. Proc Natl Acad Sci 102:16327–16332. doi: 10.1073/pnas.0508335102

Andersson O, Badisco L, Hansen AH, Hansen SH, Hellman K, Nielsen PA, Olsen LR, Verdonck R, Abbott NJ, Vanden Broeck J, et al (2014) Characterization of a novel brain barrier ex vivo insect-based P-glycoprotein screening model. Pharmacol Res Perspect 2:e00050. doi: 10.1002/prp2.50

Barba F, Nikmaram N, Roohinejad S, Khelfa A, Zhu Z, Koubaa M (2016) ioavailability of Glucosinolates and Their Breakdown Products: Impact of Processing. Front Nutr. 2016 3: 24. doi: 10.3389/fnut.2016.00024 PMCID: PMC4985713 PMID: 27579302 B

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc 57:289–300

Benoit JB, Lopez-Martinez G, Phillips ZP, Patrick KR, Denlinger DL (2010) Heat shock proteins contribute to mosquito dehydration tolerance. J Insect Physiol 56:151–156. doi: 10.1016/j.jinsphys.2009.09.012

Bian G, Raikhel AS, Zhu J (2008) Characterization of a juvenile hormone-regulated chymotrypsin-like serine protease gene in *Aedes aegypti* mosquito. Insect Biochem Mol Biol 38:190–200. doi: 10.1016/j.ibmb.2007.10.008

Bian G, Shin SW, Cheon H-M, Kokoza V, Raikhel AS (2005) Transgenic alteration of Toll immune pathway in the female mosquito *Aedes aegypti*. Proc Natl Acad Sci 102:13568–13573. doi: 10.1073/pnas.0502815102

Borst P, Zelcer N, Van Helvoort A (2000) ABC transporters in lipid transport. Biochim Biophys Acta - Mol Cell Biol Lipids 1486:128–144. doi: 10.1016/S1388-1981(00)00053-6

Breugelmans B, Simonet G, van Hoef V, Van Soest S, Smagghe G, Vanden Broeck J (2009a) A lepidopteran pacifastin member: Cloning, gene structure, recombinant production, transcript profiling and in vitro activity. Insect Biochem Mol Biol 39:430–439. doi: 10.1016/j.ibmb.2009.03.005

Breugelmans B, Simonet G, van Hoef V, Van Soest S, Vanden Broeck J (2009b) Pacifastin-related peptides: Structural and functional characteristics of a family of serine peptidase inhibitors. Peptides 30:622–632. doi: 10.1016/j.peptides.2008.07.026

Broderick NA, Raffa KF, Handelsman J (2006) Midgut bacteria required for *btus thuringiensis* insecticidal activity. Proc Natl Acad Sci 103:15196–15199. doi: 10.1073/pnas.0604865103

Burow M, Markert J, Gershenzon J, Wittstock U (2006) Comparative biochemical characterization of nitrile-forming proteins from plants and insects that alter myrosinase-catalysed hydrolysis of glucosinolates. FEBS J 273:2432–2446. doi: 10.1111/j.1742-4658.2006.05252.x

Cai QN, Zhang QW, Cheo M (2004) Contribution of indole alkaloids to *Sitobion avenae* (F.) resistance in wheat. J Appl Entomol 128:517–521. doi: 10.1111/j.1439-0418.2004.00770.x

Castagna M, Shayakul C, Trotti D, Sacchi F V., Harvey WR, Hediger MA (1998) Cloning and characterization of a potassium-coupled amino acid transporter. Proc Natl Acad Sci 95:5395–5400

Cheong SPS, Huang J, Bendena WG, Tobe SS, Hui JHL (2015) Evolution of ecdysis and metamorphosis in arthropods: The rise of regulation of juvenile hormone. Integr Comp Biol 55:878–890. doi: 10.1093/icb/icv066

Clynen E, Huybrechts J, Verleyen P, De Loof A, Schoofs L (2006) Annotation of novel neuropeptide precursors in the migratory locust based on transcript screening of a public EST database and mass spectrometry. BMC Genomics 7:1–15. doi: 10.1186/1471-2164-7-201

Coates BS (2016) Bacillus thuringiensis toxin resistance mechanisms among Lepidoptera: Progress on genomic approaches to uncover causal mutations in the European corn borer, *Ostrinia nubilalis*. Curr Opin Insect Sci 15:70–77. doi: 10.1016/j.cois.2016.04.003

Colombani J, Raisin S, Pantalacci S, Radimerski T, Montagne J, Léopold P (2003) A Nutrient Sensor Mechanism Controls *Drosophila* Growth. Cell 114:739–749. doi: 10.1016/S0092-8674(03)00713-X

Cornette R, Hayashi Y, Koshikawa S, Miura T (2013) Differential gene expression in response to juvenile hormone analog treatment in the damp-wood termite *Hodotermopsis sjostedti* (Isoptera, Archotermopsidae). J Insect Physiol 59:509–518. doi: 10.1016/j.jinsphys.2013.02.002

Costa-Silva J, Domingues D, Lopes FM (2017) RNA-Seq differential expression analysis: An extended review and a software tool. PLoS One 12:e0190152. doi: 10.1371/journal.pone.0190152

Cotter K, Stransky L, McGuire C, Forgac M (2015) Recent Insights into the Structure, Regulation, and Function of the V-ATPases. Trends Biochem Sci 40:611–622. doi: 10.1016/j.tibs.2015.08.005

Danielli A, Kafatos FC, Loukeris TG (2003) Cloning and Characterization of Four *Anopheles gambiae* Serpin Isoforms, Differentially Induced in the Midgut by *Plasmodium berghei* Invasion. J Biol Chem

278:4184–4193. doi: 10.1074/jbc.M208187200
Datta K, Guan T, Gerace L (2009) NET37, a Nuclear Envelope Transmembrane Protein with Glycosidase Homology, Is Involved in Myoblast Differentiation. J Biol Chem 284:29666–29676. doi: 10.1074/jbc.M109.034041
Denecke S, Fusetto R, Batterham P (2017) Describing the role of *Drosophila melanogaster* ABC transporters in insecticide biology using CRISPR-Cas9 knockouts. Insect Biochem Mol Biol 91:1–9. doi: 10.1016/j.ibmb.2017.09.017
Dermauw W, Van Leeuwen T (2014) The ABC gene family in arthropods: Comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol 45:89–110. doi: 10.1016/j.ibmb.2013.11.001
Eyraud V, Balmand S, Karaki L, Rahioui I, Sivignon C, Delmas AF, Royer C, Rahbé Y, Da Silva P, Gressent F (2017) The interaction of the bioinsecticide PA1b (Pea Albumin 1 subunit b) with the insect V-ATPase triggers apoptosis. Sci Rep 7:4902. doi: 10.1038/s41598-017-05315-y
Ferreira C, Rebola KGO, Cardoso C, Bragatto I, Ribeiro AF, Terra WR (2015) Insect midgut carboxypeptidases with emphasis on S10 hemipteran and M14 lepidopteran carboxypeptidases. Insect Mol Biol 24:222–229. doi: 10.1111/imb.12151
Finotello F, Di Camillo B (2015) Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. Brief Funct Genomics 14:130–142. doi: 10.1093/bfgp/elu035
Fischer HM, Wheat CW, Heckel DG, Vogel H (2008) Evolutionary Origins of a Novel Host Plant Detoxification Gene in Butterflies. Mol Biol Evol 25:809–820. doi: 10.1093/molbev/msn014
Gellatly KJ, Yoon KS, Doherty JJ, Sun W, Pittendrigh BR, Clark JM (2015) RNAi validation of resistance genes and their interactions in the highly DDT-resistant 91-R strain of *Drosophila melanogaster*. Pestic Biochem Physiol 121:107–115. doi: 10.1016/j.pestbp.2015.01.001
Gilbert LI, A. Granger N, Roe RM (2000) The juvenile hormones: historical facts and speculations on future research directions. Insect Biochem Mol Biol 30:617–644. doi: 10.1016/S0965-1748(00)00034-5
Goodman BE (2010) Insights into digestion and absorption of major nutrients in humans. Adv Physiol Educ 34:44–53. doi: 10.1152/advan.00094.2009
Goptar IA, Filippova IY, Lysogorskaya EN, Oksenoit ES, Vinokurov KS, Zhuzhikov DP, Bulushova N V., Zalunin IA, Dunaevsky YE, Belozersky MA, et al (2008) Localization of post-proline cleaving peptidases in *Tenebrio molitor* larval midgut. Biochimie 90:508–514. doi: 10.1016/j.biochi.2007.11.002
Handke B, Poernbacher I, Goetze S, Ahrens CH, Omasits U, Marty F, Simigdala N, Meyer I, Wollscheid B, Brunner E, et al (2013) The Hemolymph Proteome of Fed and Starved *Drosophila* Larvae. PLoS One 8:e67208. doi: 10.1371/journal.pone.0067208
Hansen IA, Boudko DY, Shiao S-H, Voronov DA, Meleshkevitch EA, Drake LL, Aguirre SE, Fox JM, Attardo GM, Raikhel AS (2011) AaCAT1 of the Yellow Fever Mosquito, *Aedes aegypti*. J Biol Chem 286:10803–10813. doi: 10.1074/jbc.M110.179739
He Q, Wen D, Jia Q, Cui C, Wang J, Palli SR, Li S (2014) Heat Shock Protein 83 (Hsp83) Facilitates Methoprene-tolerant (Met) Nuclear Import to Modulate Juvenile Hormone Signaling. J Biol Chem 289:27874–27885. doi: 10.1074/jbc.M114.582825
Hofte H, Whiteley HR (1989) Insecticidal Crystal Proteins of *Bacillus thuringiensis*. Microbiol Rev 53:242–255
Holtof M, Lenaerts C, Cullen D, Vanden Broeck J (2019) Extracellular nutrient digestion and absorption in the insect gut. Cell Tissue Res. doi: 10.1007/s00441-019-03031-9
Huang H-C, Niu Y, Qin L-X (2015) Differential Expression Analysis for RNA-Seq: An Overview of Statistical Methods and Computational Software. Cancer Inform 14:57–67. doi: 10.4137/CIN.S21631
Jackson EJ (1991) A User's Guide To Principal Components
Jiang H, Kanost MR (2000) The clip-domain family of serine proteinases in arthropods. Insect Biochem Mol Biol 30:95–105. doi: 10.1016/S0965-1748(99)00113-7
Jindra M, Bellés X, Shinoda T (2015) Molecular basis of juvenile hormone signaling. Curr Opin Insect Sci 11:39–46. doi: 10.1016/j.cois.2015.08.004
Jindra M, Palli SR, Riddiford LM (2013) The Juvenile Hormone Signaling Pathway in Insect Development. Annu Rev Entomol 58:181–204. doi: 10.1146/annurev-ento-120811-153700
Jolliffe I (2011) Principal component analysis. In International encyclopedia of statistical science. 1094–1096
Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philos Trans R Soc A Math Phys Eng Sci 374:20150202. doi: 10.1098/rsta.2015.0202
Jones AME, Winge P, Bones AM, Cole R, Rossiter JT (2002) Characterization and evolution of a

myrosinase from the cabbage aphid *Brevicoryne brassicae*. Insect Biochem Mol Biol 32:275–284. doi: 10.1016/S0965-1748(01)00088-1
Jouzani GS, Valijanian E, Sharafi R (2017) *Bacillus thuringiensis*: a successful insecticide with new environmental features and tidings. Appl Microbiol Biotechnol 101:2691–2711. doi: 10.1007/s00253-017-8175-y
Knight DA, Lim S, Scaffidi AK, Roche N, Chung KF, Stewart GA, Thompson PJ (2001) Protease-activated receptors in human airways: Upregulation of PAR-2 in respiratory epithelium from patients with asthma. J Allergy Clin Immunol 108:797–803. doi: 10.1067/mai.2001.119025
Labbé R, Caveney S, Donly C (2011) Genetic analysis of the xenobiotic resistance-associated ABC gene subfamilies of the Lepidoptera. Insect Mol Biol 20:243–256. doi: 10.1111/j.1365-2583.2010.01064.x
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. doi: 10.1038/nmeth.1923
Lehane MJ (1997) Peritrophic matrix structure and function. Annu Rev Entomol 42:525–550. doi: 10.1146/annurev.ento.42.1.525
Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323. doi: 10.1186/1471-2105-12-323
Li D-C, Yang F, Lu B, Chen D-F, Yang W-J (2012a) Thermotolerance and molecular chaperone function of the small heat shock protein HSP20 from hyperthermophilic archaeon, *Sulfolobus solfataricus* P2. Cell Stress Chaperones 17:103–108. doi: 10.1007/s12192-011-0289-z
Li J, Witten DM, Johnstone IM, Tibshirani R (2012b) Normalization, testing, and false discovery rate estimation for RNA-sequencing data. Biostatistics 13:523–538. doi: 10.1093/biostatistics/kxr031
Li K, Jia Q, Li S (2019) Juvenile hormone signaling – a mini review. Insect Sci 26:600–606. doi: 10.1111/1744-7917.12614
Li Z-Q, He P, Zhang Y-N, Dong S-L (2017) Molecular and Functional Characterization of Three Odorant-Binding Protein from *Periplaneta americana*. PLoS One 12:e0170072. doi: 10.1371/journal.pone.0170072
Lopata AL, Fenemore B, Jeebhay MF, Gäde G, Potter PC (2005) Occupational allergy in laboratory workers caused by the African migratory grasshopper *Locusta migratoria*. Allergy 60:200–205. doi: 10.1111/j.1398-9995.2005.00661.x
Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. doi: 10.1186/s13059-014-0550-8
Lozano J, Belles X (2011) Conserved repressive function of Krüppel homolog 1 on insect metamorphosis in hemimetabolous and holometabolous species. Sci Rep 1:1–7. doi: 10.1038/srep00163
Luo L, Sun YJ, Wu YJ (2013) Abamectin resistance in *Drosophila* is related to increased expression of P-glycoprotein via the dEGFR and dAkt pathways. Insect Biochem Mol Biol 43:627–634. doi: 10.1016/j.ibmb.2013.04.006
Mainguet AM, Louveaux A, Sayed G, Rollin P (2000) Ability of a generalist insect, *Schistocerca gregaria*, to overcome thioglucoside defense in desert plants: tolerance or adaptation? Entomol Exp Appl 94:309–317. doi: 10.1046/j.1570-7458.2000.00632.x
Majerowicz D, Gondim KC (2013) Insect Lipid Metabolism: Insights into Gene Expression Regulation. In: Recent Trends in Gene Expression. Nova Science Publishers, Inc., pp 147–190
Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18:1509–1517. doi: 10.1101/gr.079558.108
Martín-Blázquez R, Chen B, Kang L, Bakkali M (2017) Evolution, expression and association of the chemosensory protein genes with the outbreak phase of the two main pest locusts. Sci Rep 7:6653. doi: 10.1038/s41598-017-07068-0
Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10. doi: 10.14806/ej.17.1.200
McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40:4288–4297. doi: 10.1093/nar/gks042
Meekins DA, Kanost MR, Michel K (2017) Serpins in arthropod biology. Semin Cell Dev Biol 62:105–119. doi: 10.1016/j.semcdb.2016.09.001
Mello MO, Silva-Filho MC (2002) Plant-insect interactions: an evolutionary arms race between two distinct defense mechanisms. Brazilian J Plant Physiol 14:71–81. doi: 10.1590/S1677-04202002000200001
Minakuchi C, Zhou X, Riddiford LM (2008) Krüppel homolog 1 (Kr-h1) mediates juvenile hormone action

during metamorphosis of *Drosophila melanogaster*. Mech Dev 125:91–105. doi: 10.1016/j.mod.2007.10.002
Nguyen TT, Chang S-C, Evnouchidou I, York IA, Zikos C, Rock KL, Goldberg AL, Stratikos E, Stern LJ (2011) Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. Nat Struct Mol Biol 18:604–613. doi: 10.1038/nsmb.2021
Ningshen TJ, Aparoy P, Ventaku VR, Dutta-Gupta A (2013) Functional interpretation of a non-gut hemocoelic tissue aminopeptidase n (APN) in a lepidopteran insect pest *Achaea janata*. PLoS One 8:. doi: 10.1371/journal.pone.0079468
Noriega FG, Wells MA (1999) A molecular view of trypsin synthesis in the midgut of *Aedes aegypti*. J Insect Physiol 45:613–620. doi: 10.1016/S0022-1910(99)00052-9
Ock MS, Kim BJ, Kim SM, Byun KH (2005) Cloning and expression of trypsin-encoding cDNA from *Blattella germanica* and its possibility as an allergen. Korean J Parasitol 43:101. doi: 10.3347/kjp.2005.43.3.101
Oppert B, Perkin L, Martynov AG, Elpidina EN (2018) Cross-species comparison of the gut: Differential gene expression sheds light on biological differences in closely related tenebrionids. J Insect Physiol 106:114–124. doi: 10.1016/j.jinsphys.2017.03.010
Paim RMM, Araujo RN, Leis M, Sant'anna MRV, Gontijo NF, Lazzari CR, Pereira MH (2016) Functional evaluation of Heat Shock Proteins 70 (HSP70/HSC70) on *Rhodnius prolixus* (Hemiptera, Reduviidae) physiological responses associated with feeding and starvation. Insect Biochem Mol Biol 77:10–20. doi: 10.1016/j.ibmb.2016.07.011
Pijpe J, Pul N, van Duijn S, Brakefield PM, Zwaan BJ (2011) Changed gene expression for candidate ageing genes in long-lived *Bicyclus anynana* butterflies. Exp Gerontol 46:426–434. doi: 10.1016/j.exger.2010.11.033
Qi W, Ma X, He W, Chen W, Zou M, Gurr GM, Vasseur L, You M (2016) Characterization and expression profiling of ATP-binding cassette transporter genes in the diamondback moth, *Plutella xylostella* (L.). BMC Genomics 17:760. doi: 10.1186/s12864-016-3096-1
Ragan EJ, An C, Yang CT, Kanost MR (2010) Analysis of Mutually Exclusive Alternatively Spliced Serpin-1 Isoforms and Identification of Serpin-1 Proteinase Complexes in *Manduca sexta* Hemolymph. J Biol Chem 285:29642–29650. doi: 10.1074/jbc.M110.125419
Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND, Betel D (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol 14:R95. doi: 10.1186/gb-2013-14-9-r95
Rees DC, Johnson E, Lewinson O (2009) ABC transporters: the power to change. Nat Rev Mol Cell Biol 10:218–227. doi: 10.1038/nrm2646
Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. doi: 10.1093/bioinformatics/btp616
Robinson MD, Smyth GK (2007) Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics 9:321–332. doi: 10.1093/biostatistics/kxm030
Scrivener A., Watanabe H, Noda H (1997) Diet and carbohydrate digestion in the yellow-spotted longicorn beetle *Psacothea hilaris*. J Insect Physiol 43:1039–1052. doi: 10.1016/S0022-1910(97)00063-2
Scully ED, Hoover K, Carlson JE, Tien M, Geib SM (2013) Midgut transcriptome profiling of *Anoplophora glabripennis*, a lignocellulose degrading cerambycid beetle. BMC Genomics 14:850. doi: 10.1186/1471-2164-14-850
Seyednasrollah F, Laiho A, Elo LL (2015) Comparison of software packages for detecting differential expression in RNA-seq studies. Brief Bioinform 16:59–70. doi: 10.1093/bib/bbt086
Simonet G, Claeys I, Van Soest S, Breugelmans B, Franssens V, De Loof A, Vanden Broeck J (2004) Molecular identification of SGPP-5, a novel pacifastin-like peptide precursor in the desert locust. Peptides 25:941–950. doi: 10.1016/j.peptides.2004.03.005
So W V., Sarov-Blat L, Kotarski CK, McDonald MJ, Allada R, Rosbash M (2000) takeout, a Novel *Drosophila* Gene under Circadian Clock Transcriptional Regulation. Mol Cell Biol 20:6935–6944. doi: 10.1128/MCB.20.18.6935-6944.2000
Sodani K, Patel A, Kathawala RJ, Chen Z-S (2012) Multidrug resistance associated proteins in multidrug resistance. Chin J Cancer 31:58–72. doi: 10.5732/cjc.011.10329
Sogorb MA, Vilanova E (2002) Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. Toxicol Lett 128:215–228. doi: 10.1016/S0378-4274(01)00543-4
Soneson C, Delorenzi M (2013) A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 14:91. doi: 10.1186/1471-2105-14-91

Spit J, Badisco L, Vergauwen L, Knapen D, Vanden Broeck J (2016a) Microarray-based annotation of the gut transcriptome of the migratory locust, *Locusta migratoria*. Insect Mol Biol 25:745–756. doi: 10.1111/imb.12258

Spit J, Breugelmans B, Van Hoef V, Simonet G, Zels S, Vanden Broeck J (2012) Growth-inhibition effects of pacifastin-like peptides on a pest insect: The desert locust, *Schistocerca gregaria*. Peptides 34:251–257. doi: 10.1016/j.peptides.2011.06.019

Spit J, Holtof M, Badisco L, Vergauwen L, Vogel E, Knapen D, Vanden Broeck J (2016b) Transcriptional Analysis of The Adaptive Digestive System of The Migratory Locust in Response to Plant Defensive Protease Inhibitors. Sci Rep 6:32460. doi: 10.1038/srep32460

Sugawara T, Tsuduki T, Yano S, Hirose M, Duan J, Aida K, Ikeda I, Hirata T (2010) Intestinal absorption of dietary maize glucosylceramide in lymphatic duct cannulated rats. J Lipid Res 51:1761–1769. doi: 10.1194/jlr.M002204

Sui Y-P, Wang J-X, Zhao X-F (2009) The impacts of classical insect hormones on the expression profiles of a new digestive trypsin-like protease (TLP) from the cotton bollworm, *Helicoverpa armigera*. Insect Mol Biol 18:443–452. doi: 10.1111/j.1365-2583.2009.00884.x

Tabashnik BE (2015) ABCs of Insect Resistance to Bt. PLOS Genet 11:e1005646. doi: 10.1371/journal.pgen.1005646

Tabashnik BE, Brévault T, Carrière Y (2013) Insect resistance to Bt crops: lessons from the first billion acres. Nat Biotechnol 31:510–521. doi: 10.1038/nbt.2597

Tajiri R, Ogawa N, Fujiwara H, Kojima T (2017) Mechanical Control of Whole Body Shape by a Single Cuticular Protein Obstructor-E in *Drosophila melanogaster*. PLOS Genet 13:e1006548. doi: 10.1371/journal.pgen.1006548

Telleria EL, Benoit JB, Zhao X, Savage AF, Regmi S, e Silva TLA, O'Neill M, Aksoy S (2014) Insights into the Trypanosome-Host Interactions Revealed through Transcriptomic Analysis of Parasitized Tsetse Fly Salivary Glands. PLoS Negl Trop Dis 8:e2649. doi: 10.1371/journal.pntd.0002649

Tian L, Yang J, Hou W, Xu B, Xie W, Wang S, Zhang Y, Zhou X, Wu Q (2013) Molecular Cloning and Characterization of a P-Glycoprotein from the Diamondback Moth, *Plutella xylostella* (Lepidoptera: Plutellidae). Int J Mol Sci 14:22891–22905. doi: 10.3390/ijms141122891

Wang NM, Lee MF, Wu CH (1999) Immunologic characterization of a recombinant American cockroach (*Periplaneta americana*) Per a 1 (Cr-PII) allergen. Allergy Eur J Allergy Clin Immunol 54:119–127. doi: 10.1034/j.1398-9995.1999.00902.x

Wang P-Y, Neretti N, Whitaker R, Hosier S, Chang C, Lu D, Rogina B, Helfand SL (2009) Long-lived Indy and calorie restriction interact to extend life span. Proc Natl Acad Sci 106:9262–9267. doi: 10.1073/pnas.0904115106

Wheelock CE, Shan G, Ottea J (2005) Overview of Carboxylesterases and Their Role in the Metabolism of Insecticides. J Pestic Sci 30:75–83. doi: 10.1584/jpestics.30.75

Willmes DM, Kurzbach A, Henke C, Schumann T, Zahn G, Heifetz A, Jordan J, Helfand SL, Birkenfeld AL (2018) The longevity gene INDY (I'm Not Dead Yet) in metabolic control: Potential as pharmacological target. Pharmacol Ther 185:1–11. doi: 10.1016/j.pharmthera.2017.10.003

Wu Z, Wu H (2016) Experimental design and power calculation for RNA-seq experiments

Yi X, Qi J, Zhou X, Hu MY, Zhong GH (2017) Differential expression of chemosensory-protein genes in midguts in response to diet of *Spodoptera litura*. Sci Rep 7:296. doi: 10.1038/s41598-017-00403-5

Zhang C, Zhou D, Zheng S, Liu L, Tao S, Yang L, Hu S, Feng Q (2010) A chymotrypsin-like serine protease cDNA involved in food protein digestion in the common cutworm, *Spodoptera litura*: Cloning, characterization, developmental and induced expression patterns, and localization. J Insect Physiol 56:788–799. doi: 10.1016/j.jinsphys.2010.02.001

Zhang T, Coates BS, Wang Y, Wang Y, Bai S, Wang Z, He K (2017) Down-regulation of aminopeptidase N and ABC transporter subfamily G transcripts in Cry1Ab and Cry1Ac resistant Asian corn borer, *Ostrinia furnacalis* (Lepidoptera: Crambidae). Int J Biol Sci 13:835–851. doi: 10.7150/ijbs.18868

Zheng J-C, Sun S-L, Yue X-R, Liu T-X, Jing X (2018) Phylogeny and evolution of the cholesterol transporter NPC1 in insects. J Insect Physiol 107:157–166. doi: 10.1016/j.jinsphys.2018.04.007

Zuo Y-Y, Huang J-L, Wang J, Feng Y, Han T-T, Wu Y-D, Yang Y-H (2018) Knockout of a P-glycoprotein gene increases susceptibility to abamectin and emamectin benzoate in *Spodoptera exigua*. Insect Mol Biol 27:36–45. doi: 10.1111/imb.12338

Chapter 5

In vivo analyses of RNA-Seq findings

5.1. Introduction

5.1.1. General introduction

The RNA-Seq differential expression analysis has highlighted numerous interesting up- and downregulated transcripts in the midgut of the desert locust two hours after feeding (Chapter 4). Further experiments to investigate the exact *in vivo* functions of these transcripts can now be performed. For this doctoral research, it was decided to analyze a few transcripts in more detail. These were selected primarily based on their high-throughput *de novo* annotations, which included their predicted mode(s) and site(s) of action. Among all differential transcripts, an H^+ V-ATPase subunit a (TR39983|c0_g2) encoding transcript and a NPC1b (TR102686|c0_g1) encoding transcript, were further evaluated, based on their predicted key functions in the digestive process. Briefly, H^+ V-ATPases are well-described regulators of membrane potentials, facilitating molecular trafficking across epithelial membranes, while NPC1b proteins are expected to mediate the dietary sterol absorption in the insect midgut (Toei et al., 2010; Zheng et al., 2018).

Before performing *in vivo* analyses from these RNA-Seq data, we first validated our findings by analyzing the expression of several randomly selected upregulated transcripts two hours after feeding by means of real-time PCR (RT-qPCR). Only after confirming these expression profiles, experiments starting from results obtained from the RNA-Seq study were performed.

The current chapter will explain a number of pilot experiments that were conducted to evaluate the importance of the above mentioned transcripts for the viability of the desert locust. This was mainly done by suppressing their gene products by means of RNAi mediated knockdowns and subsequently evaluating the resulting effects on the overall survival and fitness of the treated animals.

5.1.2. Insect H^+ V-ATPases

The general structure and function of insect H^+ V-ATPases was already briefly explained in chapter 3 (Ch3|3.4.2.2.5). These highly conserved transmembrane proton pump complexes are located in endomembrane system and in plasma membranes of eukaryotic cells. They typically consist of two major multi-subunit complexes: a cytosolic V_1 domain and a transmembrane V_0 domain. Both domains consist of multiple structural subunits as is clearly demonstrated in figure 5.1, according to Wieczorek *et al.* (2009). The cytosolic V_1 domain is composed of eight different subunits (A-H). The catalytic sites located at subunits A and B, which are arranged in a ring of three copies per subunit in alternating way, function to hydrolyze ATP (Fig. 5.1). The transmembrane V_0 domain is generally composed of six subunits (a, c, c'', d, e, and Ac45 in mammals and a, c, c',c'',d, and e in yeast) and mediates the transport of H^+ ions out of the cell (Toei et al., 2010). In insects, the number of V_0 subunits appears to differ across species. For example, homologues of all mammalian V_0 subunits have been identified in the genome of *D. melanogaster,* while in *M. sexta,* only four V_0 subunits were identified (a, c, d, and e) (Merzendorfer et al., 2000; Chintapalli et al., 2013). Subunit a is essential for the assembly and proton transporting activity of the H^+ V-ATPase complex. Each V_0 domain contains multiple copies of subunit c that are arranged into a proteolipid ring spanning the plasma membrane (Fig. 5.1). Several V_1 and V_0 subunits form stalks connecting both domains. A central stalk consisting of subunits D, F, and d extends from the proteolipid ring of c subunits through the center of the A_3B_3 ring of the V_1 domain. Additionally, there are three peripheral

stalks consisting of subunits C, E, G and H connecting the A_3B_3 ring to subunit a of the V_0 domain (Fig. 5.1). ATP hydrolysis at the V_1 subunit drives the rotation of the central stalk and the connected transmembrane proteolipid ring of c subunits. This movement stimulates the translocation of protons across the membrane through subunit a of the V_0 domain (Fig. 5.1 A) (Wieczorek et al., 2009; Toei et al., 2010).

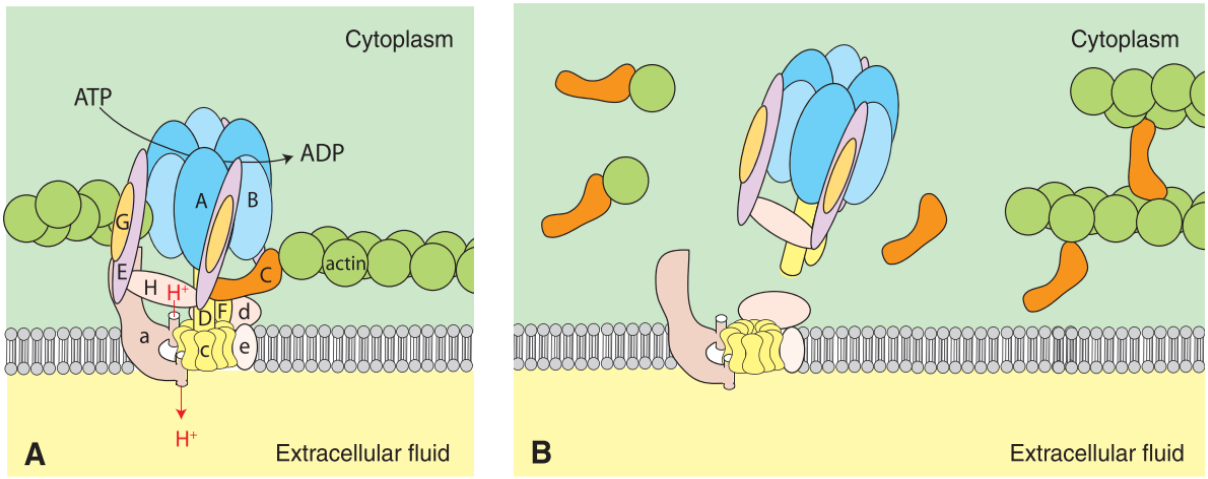


Figure 5.1. (A) (B) These pictures display the reversible dissociation of the V_1 domain from the V_0 domain of an insect H^+ V-ATPase. The V_0 remains in the plasma membrane. The transmembrane V_0 domain of insects typically consists of six subunits (a, c, d, e). The cytosolic V_1 domain is composed of eight different subunits (A-H). Structure and function of the different H^+ V-ATPase subunits are explained in the text. Picture from Wieczorek *et al.* (2009).

H^+ V-ATPase complexes use the energy released from hydrolyzing ATP to translocate protons across the membrane, generating favorable membrane potentials which are essential for the majority of transmembrane trafficking of molecules (Wieczorek et al., 1999). H^+ V-ATPases were demonstrated to be crucial for numerous biological processes, such as regulating the internal pH of intracellular compartments, the energization of secondary transport across membranes, and the loading of neurotransmitters into secretory vesicles (Marshansky and Futai, 2008; Toei et al., 2010). Also in insects, H^+ V-ATPases have been shown to play a key role in the molecular trafficking across the epithelial membranes of several tissues including the salivary glands, the intestine, the Malpighian tubules and the central nervous system (Merzendorfer et al., 2000; Weng 2003; Hiesinger et al., 2005; Wieczorek et al., 2009a; Baumann and Walz, 2012; Overend et al., 2016; Nepomuceno et al., 2017). In midgut epithelial cells, H^+ V-ATPases assist in the establishment of electrochemical gradients stimulating Na^+ or K^+ coupled nutrient transport across the midgut epithelium. This two-component transporter mechanism is expected to be present in the gastrointestinal tract of every insect species (Wieczorek et al., 2009). In addition, in the Diptera *D. melanogaster* and *A. aegypti*, H^+ V-ATPases of specialized midgut cells were demonstrated to generate specific acidic midgut regions protecting the insects from harmful pathogens that may be present in their gut lumen (Overend et al., 2016; Nepomuceno et al., 2017).

The crucial role of H^+ V-ATPases in the physiology of major insect tissues has also marked them as promising insecticidal targets. Especially the subunits positioned at the luminal side of the gut wall might constitute highly interesting candidate targets, as in theory these are easy to reach through the gastrointestinal tract (Wolfersberger, 1992). There already exist a number studies evaluating the insecticidal potential of targeting H^+ V-ATPases in several insect species, including a variety of insect pests, *D. melanogaster*, *L. decemlineata*, *M. sexta*, *T. castaneum*, *A. pisum*, and *N. lugens* (Davies et al., 1996; Baum et al., 2007; Whyard et al., 2009; Zhu et al., 2011; Li et al., 2011; Li and Xia, 2012; Ahmed, 2016; Rebijith et al., 2016;

Pan et al., 2017; Sato et al., 2018). So far, the majority of these studies focused on targeting the different subunits of the V_1 domain of these proton pumps. To the best of my knowledge, only one study describing the targeting of subunit a of the H^+ V-ATPase V_0 domain in insects exists (Ahmed, 2016). Remarkably, all above mentioned studies demonstrated strong negative impacts on insect growth and survival, clearly illustrating the strong insecticidal potential of targeting insect H^+ V-ATPases.

Interestingly, also several plant defensive molecules interacting with insect H^+ V-ATPases exist. The pea albumin 1b (PA1b), a small protein extracted from legume seeds, was the first of such molecules proven to selectively inhibit H^+ V-ATPases in the insect midgut by binding on the proteolipid ring of c subunits of the V_0 membrane domain. PA1b then forces the c ring to attach to the static subunit e instead of subunit a, resulting in the blockage of the motive force of the H^+ V-ATPase (Chouabe et al., 2011; Muench et al., 2014). Later, it was demonstrated that in the cereal weevil, *Sitophilus oryzae*, this interaction of PA1b with the c and e subunits of midgut H^+ V-ATPases triggers apoptosis in the midgut cells, resulting in premature death of the treated animals (Eyraud et al., 2017). Interestingly, the host range of PA1b is restricted to only certain insect species, and therefore, this entomotoxin is considered to be a highly attractive alternative biopesticide (Rahioui et al., 2014). More recently, another molecule with known insecticidal capacity, Celangulin V (CV) extracted from Chinese bittersweet, was demonstrated to interact with the A_3B_3 ring of midgut H^+ V-ATPases in *Mythimna separata* (Wei et al., 2017; Ding et al., 2019).

The current RNA-Seq study has highlighted the significant increase in abundance of a transcript encoding an H^+ V-ATPase subunit a two hours after feeding (Ch4|4.4.3.1). This subunit is described to be essential for the proper establishment of the entire H^+ V-ATPase proton pump. Its upregulation therefore suggests an increased formation of these crucial complexes when food is present in the midgut lumen, possibly stimulating the absorption of nutrients across the midgut epithelial barrier. Until now, such rapid upregulation of this subunit in the midgut in response to feeding has not been described in insects. Moreover, as membrane-bound essential structural elements of H^+ V-ATPases, these subunits might be excellent candidate insecticidal target sites. Therefore, the effects of RNAi-mediated silencing of H^+ V-ATPase subunit a on the fitness and survival of *S. gregaria* nymphs were further evaluated.

5.1.3. Insect Niemann-Pick C1 proteins

The role of NPC1 proteins in the digestive tract of insects was discussed in detail in chapter 1 (Ch1|1.5.3). In general, NPC1 proteins are polytropic transmembrane proteins typically containing three conserved sites: a sterol-sensing domain (SSD), a cysteine-rich-N-terminal sterol binding domain (“NPC_N” domain) and a patched domain, all described to have sterol binding capacity (Wang and Song, 2012; Zheng et al., 2018). Mammalian NPC1 has been shown to be highly expressed in the epithelial cells of the small intestine, where they play a critical role in cholesterol absorption. In insects, two NPC1 resembling genes, denoted NPC1a and NPC1b, were first identified in *D. melanogaster* (Fluegel et al., 2006). Later, it was demonstrated that the majority of insects possess these two functional NPC1 homologues. Furthermore, it could be demonstrated that NPC1b has a gut-specific expression profile, while NPC1a has a wider, more uniform expression profile (Zheng et al., 2018). Functional studies in *D. melanogaster* have illustrated that NPC1b homologue is primarily involved in the early dietary sterol uptake across the midgut epithelium (Voght et al., 2007). However, until now, this is the only published study, in which the role of NPC1b in the insect gastrointestinal tract

has been investigated (Holtof et al., 2019). The observed rapid upregulation of NPC1b encoding transcripts in the *S. gregaria* midgut two hours after feeding suggests that also in this insect species this transporter plays an important role during the digestive process.

Since insects are obligate sterol auxotrophs, they must acquire sterols from their diet. Upon absorption, the vast majority of sterols are utilized as structural components in cell membranes, while others are utilized for the synthesis of key steroid hormones, such as ecdysteroids. In insects, ecdysteroids, and especially their biologically most active form, 20-hydroxyecdysone (20E), are important regulators of the molting process by activating a neuropeptide signaling cascade that orchestrates the ecdysis at the end of each molting cycle (Roller et al., 2010). For detailed descriptions on this neuroendocrine regulation of ecdysis in insects, the reader is referred to Ewer *et al.* (1997); Zitnan (2003); Clark (2004); Žitňan *et al.* (2007); Roller *et al.* (2010); Lenaerts *et al.* (2016). Additionally, in adult insects, ecdysteroids are described to be involved in the control of a variety of other physiological processes, including reproduction, diapause and innate immunity (Uryu et al., 2015). In larval stages, ecdysteroid biosynthesis mainly takes place in the prothoracic glands (PG). Only the final step, the production of 20E, takes place in peripheral tissues, such as the fat body (Nakagawa and Henrich 2009). The PG, however, degenerate during the adult molt, and in adults, ecdysteroid biosynthesis takes place in the reproductive tissues (Hentze et al., 2013). It has been demonstrated in several insect species that disrupting the ecdysteroid signaling cascade results in molting deficiencies and lethality (Hackney et al., 2012; Luan et al., 2013; Song et al., 2017; Gouveia et al., 2018). Also in *S. gregaria* nymphs, the knockdown of the two nuclear receptors, ecdysone receptor (EcR) and retinoid-X-receptor/ultraspiracle (RXR/USP), that together form the heterodimeric nuclear 20E receptor complex in insects, resulted in an arrested development at the final molt (Lenaerts et al., 2016). For additional information on the structure and function of EcR and RXR/USP, the reader is referred to Nakagawa and Henrich (2009). Based on the obvious vital role of this signaling pathway, a variety of chemicals targeting EcR and thus disrupting the molting process in insects have already been developed as part of insect pest management strategies (Song et al., 2017).

The adequate supply of dietary sterols to the PG or to the reproductive tissues in insects is crucial for the production of ecdysteroids, which control key events in an insect's lifetime. Thus far, the effect of disrupting the NPC1b transporter on the viability of insects has only been evaluated in *D. melanogaster*. It was demonstrated that the midgut epithelial cells of NPC1b mutants were deficient in sterol and sterol intermediates, which resulted in lethal phenotypes (Voght et al., 2007). Apart from this research, the role of NPC1b in insect digestion and molting has never been evaluated. Furthermore, NPC1b might be an interesting candidate insecticidal target, since it functions as a crucial component on which are depending many key events in the life of insects. Therefore, when targeting NPC1b, severe developmental defects leading to lethal phenotypes may be expected due to an inadequate nutritional uptake of essential sterols. Therefore, we further analyzed the possible effects of silencing NPC1b on both the fitness and the molting success of *S. gregaria* nymphs.

5.1.4. RNA interference

RNAi is a posttranscriptional gene silencing mechanism that is present in nearly all eukaryotic organisms (Shabalina and Koonin, 2008). In insects, RNAi relies on the small interfering RNA (siRNA) pathway, as part of their innate anti-viral immune response. In this pathway, double stranded (ds) RNA molecules are digested by an RNase III-like enzyme, Dicer-2 (Dcr2), into smaller siRNA helices (18-24 nucleotides long) with two-nucleotide 3′ overhangs. These siRNA

helices consist of a guide and a passenger strand. A sequence specific nuclease, Argonaute-2 (Ago2), as part of the 'RNA induced silencing complex' (RISC), then cleaves and removes the siRNA passenger strand, to only retain the guide strand in the RISC. This siRNA guide strand directs the RISC to cleave RNA molecules with complementary sequence (Siomi and Siomi, 2009).

The RNAi machinery is also widely addressed as a research tool for studying gene functionalities in insects. For example, by injecting *S. gregaria* with long dsRNA molecules derived from an endogenous gene of interest, the insect's siRNA pathway will target the transcripts of this gene. As such, RNAi can be used to suppress gene products of interest, providing an excellent technique for analyzing gene functionalities *in vivo* (Vogel et al., 2019). Moreover, the desert locust exhibits a robust systemic RNAi response upon injecting dsRNA in its body cavity, resulting in a potent RNAi response throughout the entire body (Wynant et al., 2012).

In the current chapter, we relied on the systemic RNAi response in *S. gregaria* to further investigate the physiological roles of the H^+ V-ATPase subunit a and NPC1b proteins identified by RNA-Seq in a temporal transcript profiling analysis of the digestive tract. Therefore, dsRNA molecules with sequences homologous to both transcripts were constructed and injected into the body cavity of *S. gregaria* instars. These were then further monitored to examine the possible phenotypic effects of these treatments on overall fitness and survival.

5.2. Materials and methods

5.2.1. Animal rearing

Desert locusts were reared under crowded conditions in large cages in which temperature (32 °C), ambient relative humidity (40-60%) and light (14 h photoperiod) were kept constant. The animals were fed daily with dry oat flakes and fresh cabbage.

5.2.2. Production of dsRNA for RNAi experiments

The dsRNA constructs for *GFP* (589 bp), *Sg-VAHa_1* (542 bp), and *Sg-NPC1b* (795 bp) were synthesized using the MEGAscript® RNAi Kit (Ambion), according to the manufacturer's instructions. The procedure is based on the high-yield transcription reaction of a user-provided linear template transcript with a T7 promoter sequence attached. This is done by synthesizing both sense and antisense strands from a PCR-generated DNA template containing T7 RNA Polymerase promoters on both 5' ends of the primers. Primer sequences are presented in Table 5.1. cDNA derived from midguts of fifth larval *S. gregaria* instars was used as a template for the PCR reactions. A PCR with REDTaq® DNA polymerase (Sigma-Aldrich©) was performed to synthesize the linear template transcript for dsRNA construction. For *dsSg-VAHa_1* production, the following thermocycling profile was used: 5 min at 95 °C followed by 10 cycles of 30 s at 95 °C, 1 min at 55 °C and 1 min at 68 °C and 30 cycles of 30 s at 95 °C, 1 min at 59 °C and 1 min at 68 °C. For *dsSg-NPC1b* production, the following thermocycling profile was used: 5 min at 95 °C followed by 10 cycles of 30 s at 95 °C, 1 min at 55 °C and 1 min at 68 °C and 30 cycles of 30 s at 95 °C, 1 min at 61 °C and 1 min at 68 °C. For the production of *GFP* dsRNA, an amplified *GFP* cDNA fragment was cloned in both sense and antisense direction in a TOPO 4.1 sequencing vector (Life technologies), containing a T7

promotor site, and was subsequently used as template for dsRNA production. The length of the generated PCR products was checked before dsRNA synthesis by performing 1% gel electrophoresis. The final concentration of the produced dsRNA was estimated using a NanoDrop ND-1000 UV-VIS Spectrophotometer, and 1% agarose gel electrophoresis was performed to assess the integrity of the dsRNA.

Table 5.1. | Oligonucleotide sequences for primers used for dsRNA construct design.

TARGET GENE	FORWARD PRIMER	REVERSE PRIMER
Sg-VAHa_1	TAATACGACTCACTATAGGGCCGACG AATGATAGGTACTAAAC	TAATACGACTCACTATAGGGGCAGAAG CGGCAGTATTT
Sg-NPC1b	TAATACGACTCACTATAGGGTCGGAG CTGGTGAAAGCTA	TAATACGACTCACTATAGGGGTGCAAC ATGTTCCTCATGG
GFP	TAATACGACTCACTATAGGGAAGGTGA TGCTACATACGGAA	TAATACGACTCACTATAGGGATCCCAG CAGCAGTTACAAAC

Underlined sequences are the T7 promoter sequences.

5.2.3. RNAi experiments

5.2.3.1. Investigating the role of *Sg-VAHa_1*

On the day of the 5th larval molt, 91 *S. gregaria* nymphs were developmentally synchronized. To investigate the physiological role of *Sg-VAHa_1*, 46 experimental animals were subjected to repeated injections with 800 ng dsRNA (in 4 µL Ringer solution) targeting *Sg-VAHa_1*. Injections were initiated at the first day after the day of the 5th larval molt (day 1). To ensure a potent and persistent knockdown, injections were repeated on day 3 and day 6 following the day of the 5th larval molt. In parallel, 45 control animals were subjected to *dsGFP* injections. On day 7, a total 15 animals per condition were sacrificed to check the efficiency of the knockdown in the midgut. Survival and fitness of all other animals were monitored for the remainder of the experiment.

5.2.3.2. Investigating the role of *Sg-NPC1b*

Start 5th larval stage. On the day of the 5th larval molt, 39 *S. gregaria* nymphs were developmentally synchronized. To investigate the physiological role of *Sg-NPC1b*, 20 experimental animals were subjected to repeated injections with 800 ng dsRNA (in 4 µL Ringer solution) targeting *Sg-NPC1b.* Injections were initiated at the first day after the 5th larval molt (day 1). To ensure a potent and persistent knockdown, injections were repeated three and six days after the 5th larval molt. In parallel, 19 control animals were subjected to *dsGFP* injections. All animals were monitored for their survival and fitness during the experiment. Animals in both conditions were weighted on day 1 and day 8 after the 5th larval molt.

Start 4th larval stage. Six days after the 4th larval molt, 21 developmentally synchronized *S. gregaria* nymphs were injected with 800 ng dsRNA (in 4 µL Ringer solution) targeting *Sg-NPC1b*. Simultaneously, 20 *S. gregaria* were injected with 800ng dsRNA (in 4 µL Ringer solution) targeting *GFP.* Two days later, 11 experimental and 16 control animals that molted to the 5th larval stage were subjected to a follow-up injection of 800 ng dsRNA targeting either *Sg-NPC1b* or *GFP*, respectively (day 1). To ensure a potent and persistent knockdown, injections were repeated three and six days after the 5th larval molt. All animals were monitored

for their survival and fitness during the experiment. Animals in both conditions were weighted on day 1 and day 6 of the 5th larval stage.

5.2.4. Sample collection for *Sg-VAHa_1* knockdown experiment

To check the knockdown efficiency of *Sg-VAHa_1*, midguts of 15 *dsGFP* and 15 *dsSg-VAHa_1* treated animals were dissected on day 7 following the initiation of the 5th larval stage. Midguts were dissected in *S. gregaria* ringer solution (1 L: 8.766 g NaCl; 0.188 g $CaCl_2$; 0.746 g KCl; 0.407 g $MgCl_2$; 0.336 g $NaHCO_3$; 30.807 g sucrose; 1.892 g trehalose; pH 7.2) and pooled in 2 mL MagNA Lyser Green Bead vials (Roche) for subsequent total RNA extraction. Each pool consisted of three midguts each. These vials were immediately transferred to liquid nitrogen to prevent RNA degradation. Tissues were stored at -80 °C prior to RNA extraction.

5.2.5. RNA extraction for *Sg-VAHa_1* knockdown experiment

To each MagNA Lyser Green Beads vial (Roche) containing midgut tissues, 1 mL QIAzol lysis reagent buffer (Thermo Scientific™) was added. Next, the tissues were homogenized for 30 seconds at 6000 x g using a MagNA Lyser instrument (Roche). Subsequent total RNA extraction from the midgut tissues was performed using the RNeasy Lipid Tissue extraction kit (Qiagen) according to manufacturer's instructions. An additional DNaseI treatment (Qiagen) was performed to remove contaminating genomic DNA. Concentration of extracted RNA was assessed using a NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Scientific™). Prior to RT-qPCR analyses, for each sample, equal amounts of RNA (1 µg) were reverse transcribed to complementary DNA (cDNA) using the PrimeScript™ RT Reagent Kit (Takara) following the manufacturer's protocol.

5.2.6. RT-qPCR analyses

RT-qPCR analyses were performed to (1) validate the RNA-Seq differential expression analysis and (2) check knockdown efficiency of *Sg-VAHa_1*. Prior to RT-qPCR transcript profiling, several previously described housekeeping genes were tested for their stability in the designed experiment (Van Hiel et al., 2009). Optimal housekeeping genes were selected using geNorm software (Vandesompele et al., 2002). RT-qPCR primers for reference genes and target genes were designed using the online tool primer3plus (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi). Primer sets were validated by designing relative standard curves for gene transcripts with a serial ten-fold dilution of a calibrator cDNA sample. Efficiency of RT-qPCR and correlation coefficient (R^2) were measured for each primer pair. Primers for RT-qPCR are given in table 5.2. All PCR reactions were performed in duplicate in 96-well plates on a StepOne System (ABI Prism, Applied Biosystems). Each reaction contained 5 µl fast Sybr Green, 0.5 µl Forward and Reverse primer (10 µM), 1.5 µl MQ water, and 2.5 µl of diluted cDNA. For all RT-qPCR reactions, the following thermal cycling profile was used: 50 °C for 2 min, 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. Finally, a melt curve analysis was performed to check for primer dimers. For all transcripts, only a single melting peak was found during the dissociation protocol. Relative transcript levels were calculated using the ΔΔCt method according to (Livak and Schmittgen, 2001). To correct for sample to sample variation, expression was normalized against two housekeeping genes, *Sg-RP49* and *Sg-GAPDH*, as earlier determined with the geNorm software.

Table 5.2 | Oligonucleotide sequences for primers used for RT-qPCR.

TARGET GENE	FORWARD PRIMER	REVERSE PRIMER
TR51342|c0_g1	GTTTGGAGAGCAACATGAAAG	TGTTGCACTCATCACAGTC
TR57768|c0_g1	GTACAGGAGCTGTGTCATTC	GATTGGTCGCTGGAAGTATC
TR55774|c1_g1	CGTATGTTCTCTGCAATGGT	CGATATCCGGAAGCTGAATG
TR82403|c2_g5	GCACAAGTGGCTTTGTCTTC	CAGAATGAACGTGTTGGCAAT
Sg-VAHa_1	TATTACTGTCGCGTCGCATC	TCTTCTCGCTCCGAAACATC
Sg-NPC1b	CGGTGAATGTTACGATGACG	TTCTTGGTCCACCAGAAAGG
Sg-RP49	CGCTACAAGAAGCTTAAGAGGTCAT	CCTACGGCGCACTCTGTTG
Sg-GAPDH	GTCTGATGACAACAGTGCAT	GTCCATCACGCCACAACTTTC

5.2.7. PCR isoform expression analysis

To investigate the expression of TR39983|c0_g2_i1 and TR39983|c0_g2_i2, two transcript isoforms present in the *S. gregaria* whole-body transcriptome, in the midgut of *S. gregaria*, a PCR with REDTaq® DNA polymerase (Sigma-Aldrich©) was performed with template cDNA originating from midgut tissues of 5th larval stage *S. gregaria* instars. The primers used for this PCR were, forward primer: TATTACTGTCGCGTCGCATC, reverse primer: GCAGAAGCGGCAGTATTT. For this PCR, following thermocycling profile was used: 5 min at 95 °C followed by 32 cycles of 30 s at 95 °C, 1 min at 58 °C and 1 min at 68 °C. The length of the generated PCR products was checked by performing 1% gel electrophoresis.

5.2.8. Statistical analyses

All expression data and weight data were always tested for normality using the D'Agostino-Pearson omnibus normality test prior to analyzing statistical differences. If the data were normally distributed, significant differences between the control group and the treatment group were analyzed using the student's T-test. If the data were not normally distributed, significant differences between the control group and the treatment group were analyzed using the nonparametric Mann-Whitney-Wilcoxon (MWW) test. Differences were considered statistically significant if $p < 0.05$. All statistical analyses were performed using GraphPad Prism software (Version 6).

5.3. Results

5.3.1. RT-qPCR validation of RNA-Seq results

In order to validate the findings of the RNA-Seq differential expression analysis, the expression of several randomly selected upregulated genes two hours after feeding were analyzed using RT-qPCR. For this RT-qPCR analysis, samples of a duplicate feeding experiment, performed analogously to the feeding experiment for RNA-Seq sample collection, were used. The reader is referred to chapter 2 for detailed information about the conducted feeding experiment. This RT-qPCR analysis confirmed the induced expression of all randomly selected genes in the midgut samples two hours after feeding compared to twenty-four hours after feeding ($log_2FC > 1$ and $FDR < 0.05$). Moreover, the FCs observed in this RT-qPCR analysis did not differ drastically from the FCs observed in the RNA-Seq analysis. The log2-transformed FC for the different examined genes from both the RNA-Seq study and the RT-qPCR analysis are displayed in Table 5.3.

Table 5.3 | Log2-transformed FC of transcripts two hours after feeding compared to twenty-four hours after feeding calculated from RNA-Seq and RT-qPCR analysis.

TRANSCRIPT	LOG FC RNA-SEQ	LOG FC RT-QPCR	ANNOTATION
TR51342|c0_g1	4.8391234511	6.365083126	Endochitinase
TR57768|c0_g1	1.1357227476	1.879275567	Organic solute transporter alpha-like protein
TR55774|c1_g1	1.493754215	1.093913583	Multidrug resistance protein homolog 49 (MRP49)
TR82403|c2_g5	1.709560967	1.59412368	Chymotrypsin BII
Sg-VAHa_1	1.229345517	1.067133766	V-type proton ATPase 116 kDa subunit a
Sg-NPC1b	1.71736284	2.175706406	Niemann-Pick C1 b (NPC1b)

5.3.2. H^+ V-ATPase subunit a

5.3.2.1. Identification and expression analysis of *Sg-VAHa_1*

The RNA-Seq study of the midgut demonstrated the significant upregulation of an H^+ V-ATPase (116 kDa) subunit a two hours after food uptake (Chapter 4). This transcript, TR39983|c0_g2, was a SuperTranscript composed of two transcript isoforms in the in-house whole-body *S. gregaria* transcriptome, namely TR39983|c0_g2_i1 and TR39983|c0_g2_i2. Further analysis of these two transcript isoforms revealed 89.3% similarity between their respective nucleotide sequences. The transcript TR39983|c0_g2_i2 had an additional unique nucleotide sequence region between position 473 and 692 compared to TR39983|c0_g2_i1. The MAFFT pairwise sequence alignment of these nucleotide sequences is displayed in supplementary figure S5.1. When examining the putative coding regions of both isoforms, the longest predicted ORF of isoform TR39983|c0_g2_i1 and TR39983|c0_g2_i2 encoded 430 and 619 amino acids, respectively. However, both ORFs appeared to be incomplete at the C-terminal end, missing a stop codon.

The expression of both transcript isoforms in the midgut of *S. gregaria* was investigated *in vivo* by performing a PCR with 5th instar midgut cDNA as template and specific primers flanking both sides of the unique sequence region of TR39983|c0_g2_i2. This was achieved by synthesizing a forward primer with its hybridization site located between nucleotide 81 and 101 on both isoforms, and a reverse primer with its hybridization site located between nucleotide 962 and 981 for TR39983|c0_g2_i1, and nucleotide 1181 and 1200 for TR39983|c0_g2_i2 (Supplementary figure S5.1). Therefore, if two isoforms existed in the midgut, then two different sequences with a total length of 900 bp and 1119 bp, representing both TR39983|c0_g2_i1 as well as TR39983|c0_g2_i2 respectively, would be amplified and two bands would be detected on a 1% agarose gel. However, upon 32 cycles of PCR, only one product with a length clearly larger than 1000 bp, thus representing TR39983|c0_g2_i2, was detected (Supplementary figure S5.2). This suggests that only TR39983|c0_g2_i2 is expressed in the midgut, while the isoform TR39983|c0_g2_i1 is probably not expressed in the midgut or not expressed at all. Therefore, only TR39983|c0_g2_i2 was used for further analysis.

Interestingly, in addition to TR39983|c0_g2, four other transcripts annotated as H^+ V-ATPase subunit a were present in the *S. gregaria* midgut reference transcriptome (see Ch3|3.4.2.2.5). After careful investigation of their original nucleotide sequences in the in-house whole-body reference transcriptome, only two unique transcripts with a full ORF were retained: TR70116|c0_g1_i2 (843 amino acids) and TR91547|c0_g1_i1 (846 amino acids). Altogether, the *S. gregaria* midgut appears to contain three different H^+ V-ATPase subunit a isoforms,

namely TR39983|c0_g2_i2, TR70116|c0_g1_i2, and TR91547|c0_g1_i1. The MAFFT multiple sequence alignment of their amino acid sequences is displayed in supplementary figure S5.3. The remainder of this chapter will focus on isoform TR39983|c0_g2_i2, which will be further referred to as *Sg-VAHa_1*. This was the only *S. gregaria* H^+ V-ATPase subunit for which the transcript levels significantly increased in the midgut two hours after feeding (Ch4|4.4.3.1).

The relative tissue-specific expression profile of *Sg-VAHa_1* was examined in mature adult male *S. gregaria* (in-house RNA library) using RT-qPCR. *Sg-VAHa_1* expression was quasi limited to the midgut and associated caeca. In other parts of the alimentary tract, expression was far lower. Relative levels of the *Sg-VAHa_1* transcript were negligible in the fat body, the brain, ventral nerve cord (VNC), suboesophageal ganglion, epidermis and the gonads (Fig. 5.2).

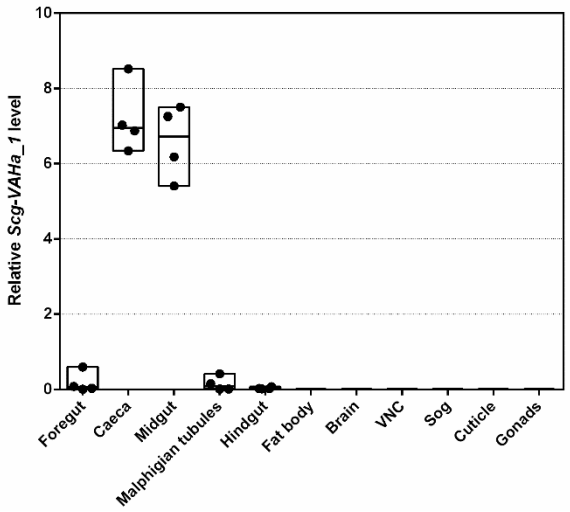


Figure 5.2. Tissue distribution of *Sg-VAHa_1* in mature male adult *S. gregaria* tissues. The relative *Sg-VAHa_1* transcript quantities in the foregut, caeca, midgut, Malpighian tubules, hindgut, fat body, brain, ventral nerve cord (VNC), suboesophageal ganglion (sog), epidermis and gonads were measured using RT-qPCR. Box plots are based on four pools of four individuals. Transcript levels were normalized against two reference genes, *Sg-GAPDH* and *Sg-RP49*.

5.3.2.2. *Sg-VAHa_1* knockdown experiment

To investigate how silencing *Sg-VAHa_1* would affect the overall fitness of *S. gregaria* instars, an experimental group of 31 insects was subjected to an injection of 800 ng *dsSg-VAHa_1* on day 1, day 3, and day 6 of the 5th larval stage. These repeated injections were performed to ensure a strong and persistent knockdown. Simultaneously, a control group of 30 animals was treated in a similar manner with *dsGFP*. Six days after the first injection of *dsSg-VAHa_1*, the average transcript levels of *Sg-VAHa_1* were demonstrated to be significantly reduced with 88.5% in the midguts of the experimentally treated locusts (Fig. 5.3 C).

Between day 6 and day 9, all 27 surviving control animals (90%) molted to the adult stage, while only 3 control animals died before day 10. All control animals reaching the adult stage were in a healthy condition and stayed alive throughout the experiment. In contrast, only 3 insects (9.67%) eventually survived the *dsSg-VAHa_1* treatment. The highest mortality rates were observed on day 9 and day 10, with 11 and 7 casualties on these days, respectively. Between day 7 and day 10, 76.7% of all experimentally treated animals died (Fig. 5.3 A). In

contrast to the control group, a total of only 18 *dsSg-VAHa_1* treated animals could reach the adult stage during the experimental timeframe. However, these insects were clearly not healthy, and between day 8 and day 14, as much as 15 adults died while only 3 survived (Fig. 5.3 B). Different lethal phenotypes were observed as a result of silencing *Sg-VAHa_1.* These could be categorized into three different groups: 5 insects died in the 5th larval stage, 8 insects died during ecdysis to the adult stage, and 15 insects reached the adult stage but died soon after. Representative pictures of the different lethal phenotypes are displayed in figure 5.3 D.

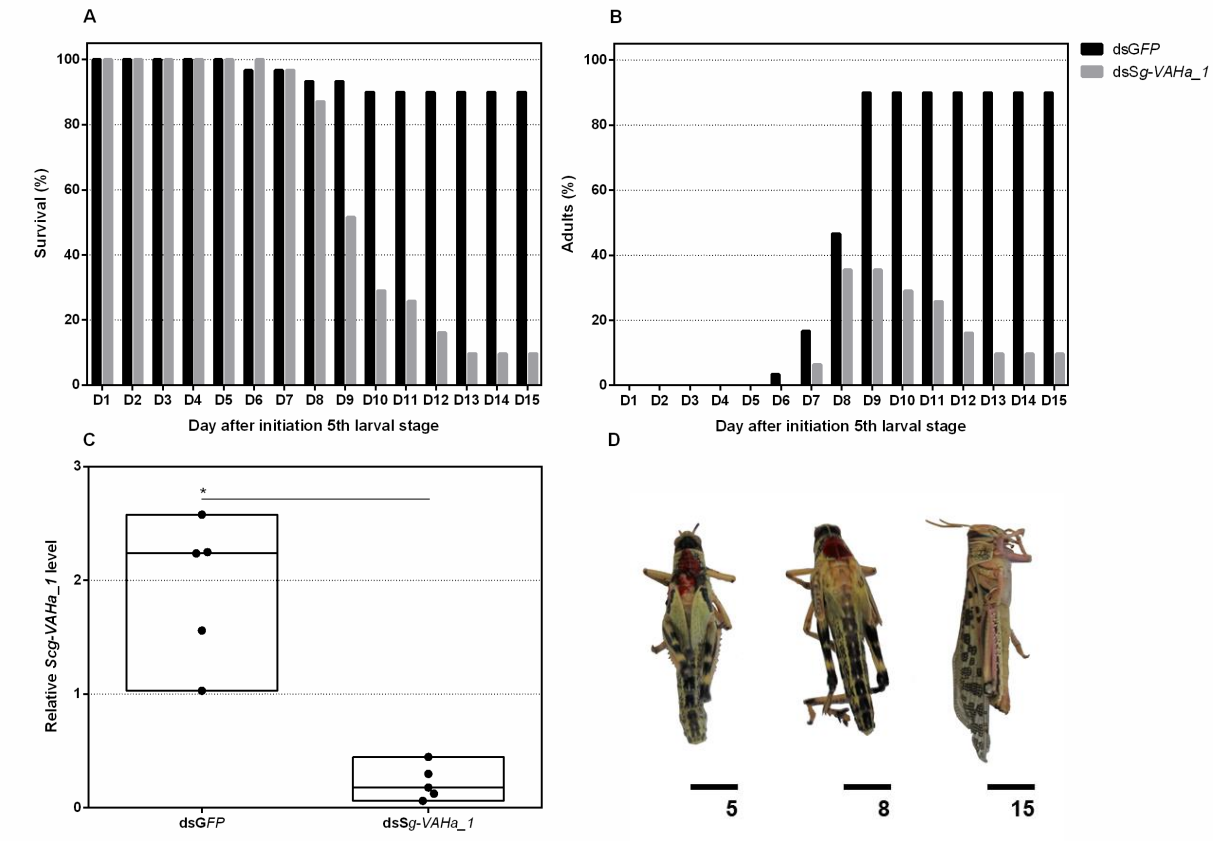


Figure 5.3. (A) Survival rate in percent of *dsGFP* (black bars) and *dsSg-VAHa_1* (grey bars) treatment groups. Daily percentages were calculated by dividing the number of animals alive on each day with the number of initial animals on D1. (B) Percentage of adult insects in *dsGFP* (black bars) and *dsSg-VAHa_1* (grey bars) treatment groups. Daily percentages were calculated by dividing the number of alive adults on each day with the number of initial animals on D1. (C) Relative mRNA levels of *Sg-VAHa_1* in the midguts of *dsGFP* and *dsSg-VAHa_1* treated animals measured with RT-qPCR. Box plots are based on five pools of three individuals. Transcript levels were normalized against two reference genes, *Sg-GAPDH* and *Sg-RP49*. The overall *Sg-VAHa_1* transcript levels were significantly reduced by 88.5% in the *dsSg-VAHa_1* treated animals, indicated with a * ($p < 0.05$, Student's T-test). (D) Pictures of the different lethal phenotypes and their numbers observed upon *dsSg-VAHa_1* treatment.

5.3.3. Niemann-Pick C1 b

5.3.3.1. Identification and expression analysis of *Sg-VNPC1b*

Transcript TR102686|c0_g1, encoding a NPC1b protein, was significantly upregulated in the *S. gregaria* midgut two hours after feeding (Chapter 4). This transcript originated from a single unique sequence in the in-house whole-body *S. gregaria* transcriptome, and will be further referred to as *Sg-NPC1b.* Moreover, *Sg-NPC1b* was the only NPC1b encoding transcript present in the transcriptome with a complete ORF encoding 1,257 amino acids containing the "NPC_N" domain (Pfam16414) between amino acids 179-270, the SSD domain (Pfam12349) between amino acids 628-780, and the patched domain (Pfam02460) between amino acids 507-1238, described to be characteristic to NPC1 proteins (Supp. Fig. S5.5). A MAFFT multiple sequence alignment of *Sg-NPC1b* with other publicly available insect NPC1b amino acid sequences is displayed in supplementary figure S5.6.

The relative tissue-specific expression profile of *Sg-NPC1b* in male *S. gregaria* 5th larval instars was examined via RT-qPCR. The different tissues selected for expression analysis were either part of the alimentary tract, or known to traffic sterols across their epithelial barrier. High expression levels of *Sg-NPC1b* were measured in the midgut and associated caeca, while very low or negligible expression was measured in the other parts of the alimentary tract. No detectable expression of *Sg-NPC1b* was observed in the fat body, gonads and prothoracic glands (Fig. 5.4).

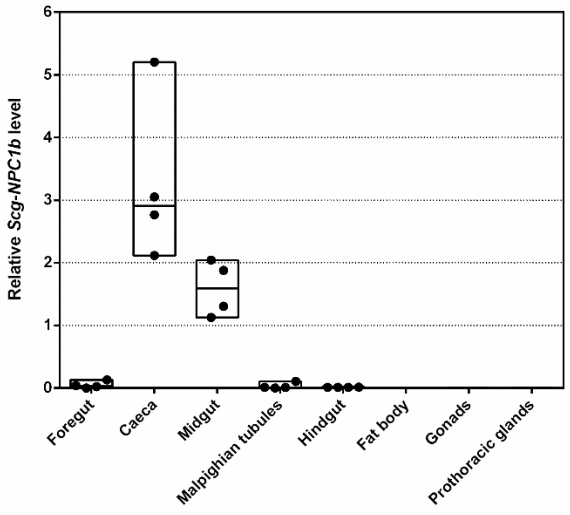


Figure 5.4. Tissue distribution of *Sg-NPC1b* in male *S. gregaria* 5th larval instar tissues. The relative *Sg-NPC1b* transcript quantities in the foregut, caeca, midgut, Malpighian tubules, hindgut, fat body, gonads and prothoracic glands were measured using RT-qPCR. Box plots are based on four pools of four individuals. Transcript levels were normalized against two reference genes, *Sg-GAPDH* and *Sg-RP49*.

5.3.3.2. *Sg-NPC1b* knockdown experiments

Two separate RNAi experiments were performed to investigate how silencing *Sg-NPC1b* would affect both molting and overall fitness in *S. gregaria* instars. In a first experiment, a total of 19 control and 20 experimental animals were subjected to repeated injections of 800 ng *dsGFP* or *dsSg-NPC1b*, respectively, on day 1, day 3, and day 6 of the 5th larval stage. Starting from day 10 of the 5th larval stage, mortality rates increased rapidly in the *dsSg-NPC1b* injected

condition. In total, 95% of all locusts in the experimental group died between day 10 and day 16 after initiating the 5th larval stage (Fig 5.5 A). At that point, all surviving control animals (16) already molted to the adult stage. Only 5 experimental animals reached the adult stage on day 10 after initiating the 5th larval stage. However, none of these animals survived long and all died soon after the adult molt (Fig. 5.5 B). On day 16, the only surviving insect subjected to the experimental treatment was arrested in the 5th larval stage. To check for any developmental effects during the experiment, all animals were weighted on day 1 and day 8 (surviving animals) of the 5th larval stage. There was no statistical difference in weight between both groups on day 1 of the 5th larval stage. Both groups gained weight between day 1 and day 8 of the 5th larval stage, but on average, insects injected with *dsSg-NPC1b* weighed significantly less than insects injected with *dsGFP* on day 8 of the 5th larval stage. (Fig. 5.5 C).

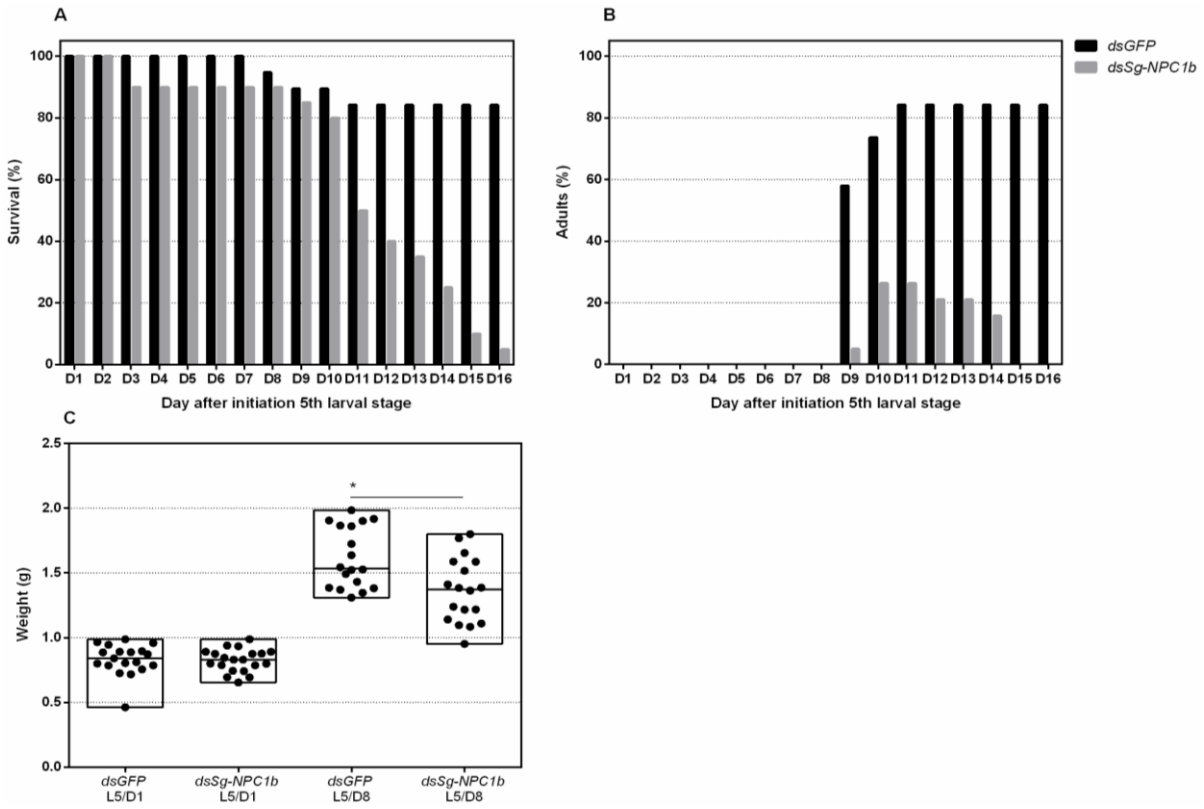


Figure 5.5. (A) Survival rate in percent of *dsGFP* (black bars) and *dsSg-NPC1b* (grey bars) treatment groups. Daily percentages were calculated by dividing the number of animals alive on each day with the number of initial animals on D1. (B) Percentage of adult insects in *dsGFP* (black bars) and *dsSg-NPC1b* (grey bars) treatment groups. Daily percentages were calculated by dividing the number of adults alive on each day with the number of initial animals on D1. (C) Total body weight in grams of *dsGFP* and *dsSg-NPC1b* treated animals on day1 and 8 of the 5th larval stage. There was no statistical difference in weight between both groups on day 1 of the 5th larval stage The *dsSg-NPC1b* treated group weighed significantly less than the *dsGFP* treated group on day 8 of the 5th larval stage, indicated with a * ($p < 0.05$, MWW).

To examine if molting to the adult stage could be completely inhibited when the *dsSg-NPC1b* injections were initiated earlier, a group of insects was already injected during the late phase of the 4th larval stage with 800 ng *dsSg-NPC1b*. In parallel, a control group was injected with an equal dosage of *dsGFP*. A total of 16 control and 11 experimental animals that simultaneously molted to the 5th larval stage 2 days after this first dsRNA injection were selected for further analysis. Depending on the condition, these animals were subsequently subjected to repeated *dsGFP* or *dsSg-NPC1b* injections (800 ng) on day 1, day 3, and day 6 of the 5th larval stage.

When *dsSg-NPC1b* injections were initiated earlier, the effect on the survival rate during the 5th larval stage was more pronounced. Within 15 days after the initiation of the 5th larval stage all 11 experimental animals died before initiating the next molt (Fig 5.6 A). Consequently, none of these animals reached the adult stage. Simultaneously, all 16 control animals survived and molted to the adult stage within 11 days after the initiation of the 5th larval stage (Fig 5.6 B). On day 1 of the 5th larval stage, no statistical difference in weight was observed between both groups. However, on day 6 of the 5th larval stage, all animals of the experimental condition (9 in total) weighed significantly less than the control animals (16 in total). Moreover, the total body weights of animals treated with *dsSg-NPC1b* did not alter significantly between day 1 and day 6 of the 5th larval stage (Fig. 5.6 C).

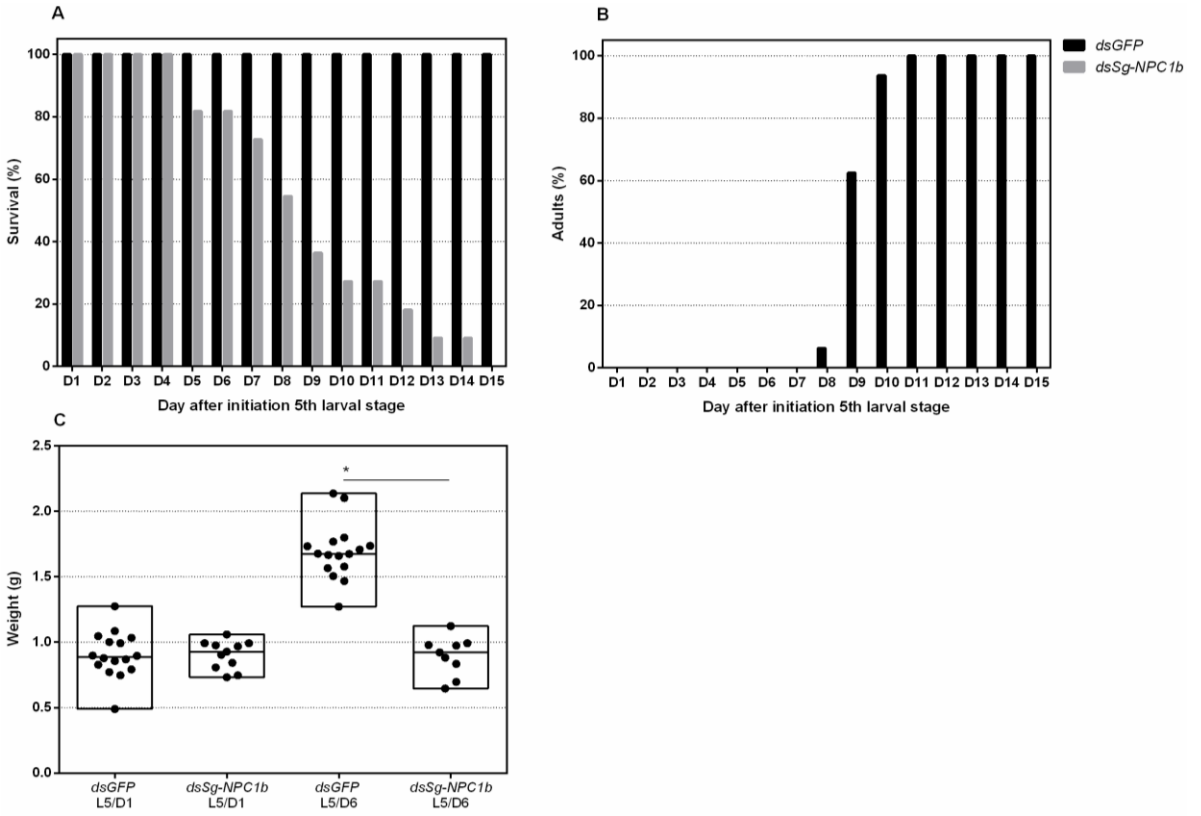


Figure 5.6. (A) Survival rate in percent of *dsGFP* (black bars) and *dsSg-NPC1b* (grey bars) treatment groups. Daily percentages were calculated by dividing the number of animals alive on each day with the number of initial animals on D1. (B) Percentage of adult insects in *dsGFP* (black bars) and *dsSg-NPC1b* (grey bars) treatment groups. Daily percentages were calculated by dividing the number of adults alive on each day with the number of initial animals on D1. (C) Total body weight in grams of *dsGFP* and *dsSg-NPC1b* treated animals on day1 and 6 of the 5th larval stage. There was no statistical difference in weight between both groups on day 1 of the 5th larval stage. The *dsSg-NPC1b* treated group weighed significantly less than the *dsGFP* treated group on day 6 of the 5th larval stage, indicated with a * ($p < 0.05$, Student's T-test).

5.4. Discussion

5.4.1. Physiological effects of silencing *Sg-VAHa_1*

The H^+ V-ATPase subunit a is known to be a key structural element of the H^+ V-ATPase transporter complex, as it is essential for complex assembly, as well as proton translocation. RNA-Seq analysis of the *S. gregaria* midgut has demonstrated the significant upregulation of an H^+ V-ATPase subunit a isoform, called *Sg-VAHa_1*, two hours after food uptake. This

induced expression possibly resulted in the increased H^+ V-ATPase complex assembly in the midgut cells upon feeding, which on its turn would boost the proton transport towards the midgut lumen. This way, a favorable membrane potential would be established to be subsequently utilized by specific nutrient transporters for driving the nutrient translocation across the apical membrane of the midgut epithelial cells. Such two-component transporter mechanism has been proposed to be present in all insect species (Wieczorek et al., 2009). Simultaneously, we detected a significant upregulation of several nutrient transporter encoding transcripts two hours after feeding. The activity of these transporters is possibly energized by an induced H^+ V-ATPase activity (Ch4|4.4.3.1). Such swift upregulation of a member of the H^+ V-ATPase transporter complex in the midgut in response to feeding has not been described previously in any insect species. In this chapter, the possible effects of RNAi-mediated knockdowns of *Sg-VAHa_1* on the viability of *S. gregaria* nymphs were evaluated as a first step towards characterizing its potential as a pesticidal target.

The tissue distribution of *Sg-VAHa_1* in *S. gregaria* highlighted extreme differences in relative expression levels. The expression of *Sg-VAHa_1* was very high in the midgut and caeca, and very low or negligible in the foregut and Malpighian tubules. Remarkably, almost no expression of *Sg-VAHa_1* was detectable in the fat body, the brain, the ventral nerve cord (VNC), the suboesophageal ganglion, the epidermis, and the gonads (Fig. 5.2). These results suggest a midgut-specific expression of *Sg-VAHa_1,* possibly implying a specialized function in this organ. The tissue-specific expression of several H^+ V-ATPase subunit isoforms has already been demonstrated in mammalian cells (Toei et al., 2010). In insects, such tissue-specific expression of H^+ V-ATPase subunit isoforms has only been analyzed in *D. melanogaster*. In this insect species, a midgut-specific H^+ V-ATPase subunit a isoform, called vha100-4, was identified. Other *D. melanogaster* H^+ V-ATPase subunit a isoforms, four in total, are more uniformly expressed in various tissues (Chintapalli et al., 2013). Interestingly, the *Sg*-VAHa_1 amino acid sequence appeared to be slightly more similar to vha100-2 (57.6%), than to the midgut-specific vha100-4 (53.3%) of *D. melanogaster*.

In addition to *Sg-VAHa_1*, two other unique H^+ V-ATPase subunit a isoform sequences were identified in the *S. gregaria* midgut. However, their tissue-specific expression profile has not yet been evaluated. More research is therefore needed to functionally characterize the different *S. gregaria* H^+ V-ATPase subunit a isoforms. Furthermore, a MAFFT multiple sequence alignment of the three identified *S. gregaria* H^+ V-ATPase subunit a isoform amino acid sequences with other publicly available insect H^+ V-ATPase subunit a sequences showed overall high sequence similarity among all sequences (Supp. Fig S5.4). The *Sg*-VAHa_1 amino acid sequence however appears to be incomplete at the C-terminal end, and is therefore shorter than the other insect H^+ V-ATPase subunit a isoform sequences. The complete *Sg-VAHa_1* sequence, including a full ORF, can hopefully be identified in the near future.

To investigate the importance of *Sg*-VAHa_1 on the viability of *S. gregaria,* an RNAi-mediated knockdown of *Sg-VAHa_1* was performed. Significant mortality was observed between days seven and ten after the first *dsSg-VAHa_1* injection. As much as 76.7% of all experimental animals died within this timeframe. At the end of the experiment, only three out of thirty-one insects survived the experimental treatment (Fig. 5.3 A-B). Three different lethal phenotypes were observed: death in 5^{th} larval stage, death during ecdysis and death in the adult stage (Fig 5.3 D). The majority of the experimental animals died in the adult stage, shortly after molting. In general, the rate of mortality appeared to be slow, which suggests a slow manifestation of the deathly effect. It might also be the consequence of a low turnover rate of these transmembrane V_0 subunits, resulting in a delayed effect of the dsRNA treatment.

Nevertheless, a mortality rate of 90% within thirteen days after the first injection clearly illustrates the crucial importance of *Sg*-VAHa_1 for the viability of the locusts.

How silencing *Sg-VAHa_1* resulted in lethality needs to be further addressed. The tissue distribution already demonstrated the midgut-specific expression of *Sg-VAHa_1*, and therefore, the lethal effects are expected to be induced at the cells of the midgut and its associated caeca. It is possible that lethality was induced by apoptosis in the midgut cells, similarly to the mode of action of the entomotoxin PA1b that inhibits H^+ V-ATPase activity (Eyraud et al., 2017). It would therefore be interesting to investigate the integrity of the midgut epithelium of *dsSg-VAHa_1* treated animals with, for example, transmission electron microscopy. Another suggested way to evaluate apoptosis in the midgut cells is to investigate the caspase-3 expression levels, which are key enzymes of the caspase family and known to play a major role in the activation of apoptosis in insect cells (Eyraud et al., 2017). Lethality might have also resulted from a defective uptake of essential nutrients due to the silencing of the H^+ V-ATPase complex, since the formation of the H^+ V-ATPase complex, and hence the favorable membrane potential it induces, may have been inhibited upon knocking down *Sg-VAHa_1*. This can be evaluated by measuring the nutrient uptake at the midgut cells or nutrient levels in the haemolymph upon silencing *Sg-VAHa_1.*

5.4.2. Physiological effects of silencing *Sg-NPC1b*

The RNA-Seq analysis highlighted the significant upregulation of the NPC1b encoding transcript, *Sg-NPC1b*, two hours after feeding. This *Sg-NPC1b* was the only NPC1b transcript present in the *S. gregaria* transcriptome with a complete ORF encoding 1,257 amino acids containing the conserved SSD, “NPC” and patched domains, characteristic to NPC1 proteins (Wang and Song 2012; Zheng et al., 2018). NPC1b gene expression is described to be midgut-specific in insects, mediating sterol absorption across the midgut epithelial membrane (Zheng et al., 2018). This was also investigated in *S. gregaria* by examining the relative expression levels of *Sg-NPC1b* in various tissues known to have traffic of sterols across their plasma membrane. We observed that the expression of *Sg-NPC1b* was restricted to the midgut and its associated caeca, demonstrating a midgut-specific profile in the desert locust (Fig. 5.4). A MAFFT multiple sequence alignment of the amino acid sequence of *Sg-NPC1b* with other publicly available insect NPC1b sequences revealed overall high sequence similarity between the different insect NPC1b sequences (Supp. Fig. S5.6).

RNAi-mediated knockdowns of *Sg-NPC1b* were induced in *S. gregaria* 4th and 5th larval instars. These experiments clearly indicate the detrimental effects of *dsSg-NPC1b* injections on survival and molting. When initiating the knockdowns in the 5th larval stage, only 26% of the experimental animals were able to reach the adult stage but died soon after, while simultaneously, all surviving control animals were healthy and molted to the adult stage. Moreover, as much as 95% of the *dsSg-NPC1b* treated animals died between days ten and sixteen after initiating the experimental treatment (Fig. 5.5 A-B). When *dsSg-NPC1b* injections were initiated in the late 4th larval stage, the effects were even more pronounced. None of the experimentally treated insects reached the adult stage, and all died within fifteen days after initiating the final larval stage. However, all control insects were healthy and reached the adult stage (Fig. 5.6 A-B). To determine if growth defects manifested during the experiment, individual locusts were weighted towards the end of the 5th larval stage. In both experiments, the *dsSg-NPC1b* treated animals weighed significantly less than the control animals. However, this effect was most pronounced when the injections were already initiated in the 4th larval stage (Fig 5.5 C and Fig. 5.6 C). These results indicate that the experimentally treated animals

were unable to gain as much weight as the control animals during the same period of time. Therefore, this observation could indeed be in accordance with an impaired absorption of essential nutrients.

Targeting *Sg-NPC1b* in *S. gregaria* instars not only resulted in high mortality, moreover, almost none of the *dsSg-NPC1b* treated animals reached the adult stage. One putative explanation is that knocking down *Sg-NPC1b* resulted in the reduced absorption of vital dietary sterols in the midgut. This can be further investigated by determining the total sterol contents in the midgut cells, similar to the experiments described by Voght *et al.* (2007). The expected drop in dietary sterol absorption caused by *Sg-NPC1b* knockdown might on its turn have resulted in reduced ecdysteroid production, crucial for insect development, resulting in the observed molting defects and premature death. This could explain why none of the animals was able to molt to the adult stage when the injections were initiated in the late 4th larval stage, instead of the early 5th larval stage. In this case, the ecdysteroid production may have been inhibited earlier, because of the earlier defective dietary sterol absorption, so that no insect was able to initiate the adult molt. The link between *Sg-NPC1b* knockdown and ecdysteroid synthesis can be further investigated in several ways. One straightforward way would be to determine the haemolymph ecdysteroid titers using anti-ecdysone antiserum (O–6) as described in Sakurai *et al.* (1998). However, this procedure has proven to be difficult to reproduce. Another possibility would be to examine the expression levels of the so-called Halloween genes, described to be involved in the ecdysteroid biosynthesis pathway in *S. gregaria* (Niwa and Niwa, 2014; Lenaerts et al., 2016). An earlier study in our lab demonstrated that silencing EcR/RXR, the receptor complex of ecdysteroids, significantly reduced the expression of the Halloween genes *Spook*, *Shade*, and *Shadow* in *S. gregaria* (Lenaerts et al., 2016). We observed phenotypes that were similar to these observed after silencing the *Sg-EcR*/*Sg-RXR* receptor components during the last larval stage of *S. gregaria* (Lenaerts et al., 2016). When silencing this receptor complex, ecdysteroid signaling was reduced, and nymphs were arrested in the 5th larval stage for a prolonged time, up to 25 days, and eventually died. Interestingly, the reproductive organs of some of these *dsSg-EcR/Sg-RXR* treated insects already initiated adult characteristics during their prolonged 5th larval stage. Moreover, the majority of these insects displayed clear locomotive defects (Lenaerts et al., 2016; Supplementary data). Similar locomotive defects were observed in several *dsSg-NPC1b* treated animals, albeit sooner after initiating dsRNA injections. Additionally, the lethal effects of the *dsSg-NPC1b* treatment manifested much faster and stronger compared to *Sg-EcR*/*Sg-RXR* knockdown. In this short timeframe between treatment and death, the *dsSg-NPC1b* treated animals did not initiate adult-like maturation of the reproductive organs as observed by Lenaerts *et al.* (2016).

In addition to the possible defective ecdysteroid production resulting in reduced molting and premature death, it is also plausible that lethality upon *dsSg-NPC1b* injection was induced in a different way. It is of course possible that the *dsSg-NPC1b* treatment also affected other vital physiological processes. It is generally accepted that in insects most dietary sterols are utilized for cell membrane structure, while only a small pool of sterols is necessary as precursors for steroidogenesis. However, the specificity of the sterols required for ecdysteroid production is very strict, since these can only be synthesized from cholesterol. The specificity for sterols incorporated in the cell membranes on the other hand is less strict (Behmer, 2017). Sterols are indispensable structural elements of cell membranes where they influence membrane fluidity and rigidity. Reduction of the amount of dietary acquired sterols as a result of injecting *dsSg-NPC1b* might have caused cell membrane abnormalities resulting in mortality. Remarkably, in *D. melanogaster,* sterols appeared to be dispensable for maintaining the basic membrane biophysical properties during periods of extreme dietary sterol restriction. During such periods, some biophysical properties of the membrane, such as impermeability, were performed by

non-sterol lipids, such as sphingolipids (Carvalho et al., 2010). It therefore needs to be investigated whether *S. gregaria* is also able to rely on similar compensatory mechanisms when the uptake of dietary sterols is restricted by *dsSg-NPC1b* treatments. Moreover, investigating the substrate specificity of NPC1b in *S. gregaria* as well as the exact sterol depletions caused by *dsSg-NPC1b* injections would help us to better understand the role of NPC1b in the digestive process of this insect species. Altogether, more research is definitely needed to further elucidate the cause(s) of mortality induced by *dsSg-NPC1b* injections in *S. gregaria* nymphs.

Overall, these experiments were the first to investigate the role of NPC1b in an insect species distinct from *D. melanogaster*. We clearly demonstrated that injecting *dsSg-NPC1b* resulted in growth and developmental (molting) defects. This may be due to a reduced uptake of dietary sterols. Until now, insect NPC1b has only been characterized in *D. melanogaster,* where it was demonstrated to mediate dietary sterol absorption in the midgut (Voght et al., 2007). Our results suggest a similar role for NPC1b in the desert locust. Nevertheless, this needs to be further characterized.

5.5. Conclusion

The RNA-Seq analysis of the digestive process in *S. gregaria* has highlighted the upregulation of an H^+-V-ATPase subunit a (*Sg-VAHa_1*), and a NPC1b (*Sg-NPC1b*) encoding transcript two hours after food uptake. Both gene products were expected to play a crucial role during digestion, and consequently be of vital importance for the insect. The current chapter described several pilot experiments that were conducted to evaluate for the first time their importance for the viability of the desert locust. It could be demonstrated that targeting both transcripts with dsRNA resulted in high rates of mortality, clearly confirming their vital importance. However, more research is necessary to further characterize their exact roles in the digestive process as well as to elucidate what exactly is causing the premature death of the insects upon dsRNA treatments.

5.6. References

Ahmed MAM (2016) RNAi-based silencing of genes encoding the vacuolar-ATPase subunits a and c in pink bollworm (*Pectinophora gossypiella*). African J Biotechnol 15:2547–2557. doi: 10.5897/AJB2016.15611

Baum JA, Bogaert T, Clinton W, Heck GR, Feldmann P, Ilagan O, Johnson S, Plaetinck G, Munyikwa T, Pleau M, et al (2007) Control of coleopteran insect pests through RNA interference. Nat Biotechnol 25:1322–1326. doi: 10.1038/nbt1359

Baumann O, Walz B (2012) The blowfly salivary gland – A model system for analyzing the regulation of plasma membrane V-ATPase. J Insect Physiol 58:450–458. doi: 10.1016/j.jinsphys.2011.11.015

Behmer ST (2017) Overturning dogma: tolerance of insects to mixed-sterol diets is not universal. Current Opinion in Insect Science 23;89-95. doi: 10.1016/j.cois.2017.08.001

Carvalho M, Schwudke D, Sampaio JL, Palm W, Riezman I, Dey G, Gupta GD, Mayor S, Riezman H, Shevchenko A, Kurzchalia TV, Eaton S (2010) Survival strategies of a sterol auxotroph.Development 137(21): 3675–3685. doi: 10.1242/dev.044560

Chintapalli VR, Wang J, Herzyk P, Davies SA, Dow JA (2013) Data-mining the FlyAtlas online resource to identify core functional motifs across transporting epithelia. BMC Genomics 14:518. doi: 10.1186/1471-2164-14-518

Chouabe C, Eyraud V, Da Silva P, Rahioui I, Royer C, Soulage C, Bonvallet R, Huss M, Gressent F (2011) New Mode of Action for a Knottin Protein Bioinsecticide. J Biol Chem 286:36291–36296. doi: 10.1074/jbc.m111.281055

Clark AC (2004) Neuroendocrine Control of Larval Ecdysis Behavior in *Drosophila*: Complex Regulation by Partially Redundant Neuropeptides. J Neurosci 24:4283–4292. doi: 10.1523/JNEUROSCI.4938-03.2004

Davies SA, Goodwin SF, Kelly DC, Wang Z, Sözen MA, Kaiser K, Dow JAT (1996) Analysis and Inactivation of vha55 , the Gene Encoding the Vacuolar ATPase B-subunit in *Drosophila melanogaster* Reveals a Larval Lethal Phenotype. J Biol Chem 271:30677–30684. doi: 10.1074/jbc.271.48.30677

Ding L, Guo Z, Xu H, Li T, Wang Y, Tao H (2019) The Inhibitory Effect of Celangulin V on the ATP Hydrolytic Activity of the Complex of V-ATPase Subunits A and B in the Midgut of *Mythimna separata*. Toxins (Basel) 11:130. doi: 10.3390/toxins11020130

Ewer J, Gammie SC, Truman JW (1997) Control of insect ecdysis by a positive-feedback endocrine system: roles of eclosion hormone and ecdysis triggering hormone. J Exp Biol 200:869–81

Eyraud V, Balmand S, Karaki L, Rahioui I, Sivignon C, Delmas AF, Royer C, Rahbé Y, Da Silva P, Gressent F (2017) The interaction of the bioinsecticide PA1b (Pea Albumin 1 subunit b) with the insect V-ATPase triggers apoptosis. Sci Rep 7:4902. doi: 10.1038/s41598-017-05315-y

Fluegel ML, Parker TJ, Pallanck LJ (2006) Mutations of a *Drosophila* NPC1 Gene Confer Sterol and Ecdysone Metabolic Defects. Genetics 172:185–196. doi: 10.1534/genetics.105.046565

Gouveia D, Bonneton F, Almunia C, Armengaud J, Quéau H, Degli-Esposti D, Geffard O, Chaumot A (2018) Identification, expression, and endocrine-disruption of three ecdysone-responsive genes in the sentinel species *Gammarus fossarum*. Sci Rep 8:3793. doi: 10.1038/s41598-018-22235-7

Hackney JF, Zolali-Meybodi O, Cherbas P (2012) Tissue Damage Disrupts Developmental Progression and Ecdysteroid Biosynthesis in *Drosophila*. PLoS One 7:e49105. doi: 10.1371/journal.pone.0049105

Hentze JL, Moeller ME, Jørgensen AF, Bengtsson MS, Bordoy AM, Warren JT, Gilbert LI, Andersen O, Rewitz KF (2013) Accessory Gland as a Site for Prothoracicotropic Hormone Controlled Ecdysone Synthesis in Adult Male Insects. PLoS One 8:e55131. doi: 10.1371/journal.pone.0055131

Hiesinger PR, Fayyazuddin A, Mehta SQ, Rosenmund T, Schulze KL, Zhai RG, Verstreken P, Cao Y, Zhou Y, Kunz J, et al (2005) The v-ATPase V0 subunit a1 is required for a late step in synaptic vesicle exocytosis in *Drosophila*. Cell 121:607–620. doi: 10.1016/j.cell.2005.03.012

Holtof M, Lenaerts C, Cullen D, Vanden Broeck J (2019) Extracellular nutrient digestion and absorption in the insect gut. Cell Tissue Res. doi: 10.1007/s00441-019-03031-9

Lenaerts C, Van Wielendaele P, Peeters P, Vanden Broeck J, Marchal E (2016) Ecdysteroid signalling components in metamorphosis and development of the desert locust, *Schistocerca gregaria*. Insect Biochem Mol Biol 75:10–23. doi: 10.1016/j.ibmb.2016.05.003

Li C, Xia Y (2012) Vacuolar ATPase subunit H is essential for the survival and moulting of *Locusta*

migratoria manilensis. Insect Mol Biol 21:405–413. doi: 10.1111/j.1365-2583.2012.01147.x
Li J, Chen Q, Lin Y, Jiang T, Wu G, Hua H (2011) RNA interference in *Nilaparvata lugens* (Homoptera: Delphacidae) based on dsRNA ingestion. Pest Manag Sci 67:852–859. doi: 10.1002/ps.2124
Livak KJ, Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods 25:402–408. doi: 10.1006/meth.2001.1262
Luan J-B, Ghanim M, Liu S-S, Czosnek H (2013) Silencing the ecdysone synthesis and signaling pathway genes disrupts nymphal development in the whitefly. Insect Biochem Mol Biol 43:740–746. doi: 10.1016/j.ibmb.2013.05.012
Marshansky V, Futai M (2008) The V-type H+-ATPase in vesicular trafficking: targeting, regulation and function. Curr Opin Cell Biol 20:415–426. doi: 10.1016/j.ceb.2008.03.015
Merzendorfer H, Reineke S, Zhao X-F, Jacobmeier B, Harvey WR, Wieczorek H (2000) The multigene family of the tobacco hornworm V-ATPase: novel subunits a, C, D, H, and putative isoforms1The nucleotide sequences reported in this paper have been submitted to the GenBank/EMBL data bank with accession numbers AJ249388, AJ249389, AJ249390 a. Biochim Biophys Acta - Biomembr 1467:369–379. doi: 10.1016/S0005-2736(00)00233-9
Muench SP, Rawson S, Eyraud V, Delmas AF, Da Silva P, Phillips C, Trinick J, Harrison MA, Gressent F, Huss M (2014) PA1b Inhibitor Binding to Subunits c and e of the Vacuolar ATPase Reveals Its Insecticidal Mechanism. J Biol Chem 289:16399–16408. doi: 10.1074/jbc.M113.541250
Nakagawa Y, Henrich VC (2009) Arthropod nuclear receptors and their role in molting. FEBS J 276:6128–6157. doi: 10.1111/j.1742-4658.2009.07347.x
Nepomuceno DB, Santos VC, Araújo RN, Pereira MH, Sant'Anna MR, Moreira LA, Gontijo NF (2017) pH control in the midgut of *Aedes aegypti* under different nutritional conditions. J Exp Biol 220:3355–3362. doi: 10.1242/jeb.158956
Overend G, Luo Y, Henderson L, Douglas AE, Davies SA, Dow JAT (2016) Molecular mechanism and functional significance of acid generation in the *Drosophila* midgut. Sci Rep 6:27242. doi: 10.1038/srep27242
Pan H, Yang X, Bidne K, Hellmich RL, Siegfried BD, Zhou X (2017) Dietary Risk Assessment of v-ATPase A dsRNAs on Monarch Butterfly Larvae. Front Plant Sci 8:1–9. doi: 10.3389/fpls.2017.00242
Rahioui I, Eyraud V, Karaki L, Sasse F, Carre-Pierrat M, Qin A, Zheng MH, Toepfer S, Sivignon C, Royer C, et al (2014) Host range of the potential biopesticide Pea Albumin 1b (PA1b) is limited to insects. Toxicon 89:67–76. doi: 10.1016/j.toxicon.2014.07.004
Rebijith KB, Asokan R, Ranjitha HH, Rajendra Prasad BS, Krishna V, Krishna Kumar NK (2016) Diet-Delivered dsRNAs for juvenile hormone-binding protein and vacuolar atpase-h implied their potential in the management of the melon aphid (Hemiptera: Aphididae). Environ Entomol 45:268–275. doi: 10.1093/ee/nvv178
Roller L, Žitňanová I, Dai L, Šimo L, Park Y, Satake H, Tanaka Y, Adams ME, Žitňan D (2010) Ecdysis triggering hormone signaling in arthropods. Peptides 31:429–441. doi: 10.1016/j.peptides.2009.11.022
Sato K, Miyata K, Ozawa S, Hasegawa K (2018) Systemic RNAi of V-ATPase subunit B causes molting defect and developmental abnormalities in *Periplaneta fuliginosa*. Insect Sci 721–731. doi: 10.1111/1744-7917.12565
Shabalina SA, Koonin E V. (2008) Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol 23:578–587. doi: 10.1016/j.tree.2008.06.005
Siomi H, Siomi MC (2009) On the road to reading the RNA-interference code. Nature 457:396–404. doi: 10.1038/nature07754
Song Y, Villeneuve DL, Toyota K, Iguchi T, Tollefsen KE (2017) Ecdysone Receptor Agonism Leading to Lethal Molting Disruption in Arthropods: Review and Adverse Outcome Pathway Development. Environ Sci Technol 51:4142–4157. doi: 10.1021/acs.est.7b00480
Toei M, Saum R, Forgac M (2010a) Regulation and Isoform Function of the V-ATPases. Biochemistry 49:4715–4723. doi: 10.1021/bi100397s
Uryu O, Ameku T, Niwa R (2015) Recent progress in understanding the role of ecdysteroids in adult insects: Germline development and circadian clock in the fruit fly *Drosophila melanogaster*. Zool Lett 1:32. doi: 10.1186/s40851-015-0031-2
Van Hiel MB, Van Wielendaele P, Temmerman L, Van Soest S, Vuerinckx K, Huybrechts R, Vanden Broeck J, Simonet G (2009) Identification and validation of housekeeping genes in brains of the

desert locust *Schistocerca gregaria* under different developmental conditions. BMC Mol Biol 10:56. doi: 10.1186/1471-2199-10-56

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:1–30. doi: Artn 0034.1Doi 10.1186/Gb-2002-3-7-Research0034

Vogel E, Santos D, Mingels L, Verdonckt T-W, Vanden Broeck J (2019) RNA Interference in Insects: Protecting Beneficials and Controlling Pests. Front Physiol 9:1912. doi: 10.3389/fphys.2018.01912

Voght SP, Fluegel ML, Andrews LA, Pallanck LJ (2007) *Drosophila* NPC1b Promotes an Early Step in Sterol Absorption from the Midgut Epithelium. Cell Metab 5:195–205. doi: 10.1016/j.cmet.2007.01.011

Wang L-J, Song B-L (2012) Niemann–Pick C1-Like 1 and cholesterol uptake. Biochim Biophys Acta - Mol Cell Biol Lipids 1821:964–972. doi: 10.1016/j.bbalip.2012.03.004

Wei J, Li D, Xi X, Liu L, Zhao X, Wu W, Zhang J (2017) Molecular insights into the potential insecticidal interaction of β-Dihydroagarofuran derivatives with the H subunit of V-ATPase. Molecules 22:. doi: 10.3390/molecules22101701

Weng X-H (2003) The V-type H+-ATPase in Malpighian tubules of *Aedes aegypti*: localization and activity. J Exp Biol 206:2211–2219. doi: 10.1242/jeb.00385

Whyard S, Singh AD, Wong S (2009) Ingested double-stranded RNAs can act as species-specific insecticides. Insect Biochem Mol Biol 39:824–832. doi: 10.1016/j.ibmb.2009.09.007

Wieczorek H, Beyenbach KW, Huss M, Vitavska O (2009a) Vacuolar-type proton pumps in insect epithelia. J Exp Biol 212:1611–1619. doi: 10.1242/jeb.030007

Wieczorek H, Brown D, Grinstein S, Ehrenfeld J, Harvey WR (1999) Animal plasma membrane energization by proton-motive V-ATPases. BioEssays 21:637–648. doi: 10.1002/(SICI)1521-1878(199908)21:8<637::AID-BIES3>3.0.CO;2-W

Wolfersberger MG (1992) V-ATPase-energized epithelia and biological insect control. J Exp Biol 172:377–386

Wynant N, Verlinden H, Breugelmans B, Simonet G, Vanden Broeck J (2012) Tissue-dependence and sensitivity of the systemic RNA interference response in the desert locust, *Schistocerca gregaria*. Insect Biochem Mol Biol 42:911–917. doi: 10.1016/j.ibmb.2012.09.004

Zheng J-C, Sun S-L, Yue X-R, Liu T-X, Jing X (2018a) Phylogeny and evolution of the cholesterol transporter NPC1 in insects. J Insect Physiol 107:157–166. doi: 10.1016/j.jinsphys.2018.04.007

Zhu F, Xu J, Palli R, Ferguson J, Palli SR (2011) Ingested RNA interference for managing the populations of the Colorado potato beetle, *Leptinotarsa decemlineata*. Pest Manag Sci 67:175–182. doi: 10.1002/ps.2048

Zitnan D (2003) Conservation of ecdysis-triggering hormone signalling in insects. J Exp Biol 206:1275–1289. doi: 10.1242/jeb.00261

Žitňan D, Kim Y-J, Žitňanová I, Roller L, Adams ME (2007) Complex steroid–peptide–receptor cascade controls insect ecdysis. Gen Comp Endocrinol 153:88–96. doi: 10.1016/j.ygcen.2007.04.002

Chapter 6

The role of juvenile hormone in the digestive process of locusts

This chapter is partially published as:

Spit J., **Holtof M.**, Badisco L., Vergauwen L., Vogel E., Knapen D., Vanden Broeck J. (2016) Transcriptional Analysis of The Adaptive Digestive System of The Migratory Locust in Response to Plant Defensive Protease Inhibitors. Scientific Reports 6:32460. doi: 10.1038/srep32460.

6.1. Introduction

6.1.1 General introduction

The RNA-Seq study of the digestive process of *S. gregaria* has highlighted the differential upregulation of several transcripts predicted to be involved in the JH signaling cascade. Three JHBP and one HSP83 encoding transcripts were found to be upregulated in the midgut two hours after feeding (Ch4|4.4.3.1). This is not the first report of the possible involvement of JH in the regulation of insect digestion. There are several other articles of insect digestive serine proteases seemingly under the regulatory control of JH (Bian et al., 2008; Sui et al., 2009; Zhang et al., 2010; Cornette et al., 2013). The best-studied case of JH involvement in the digestive system of an insect species comes from the mosquito, *A. aegypti*. It is well established that the ingestion of a blood meal in the female mosquito takes place in two distinct phases. A first stage when a small amount of early trypsins initiates the proteolytic digestion is always followed by a second stage when much larger amounts of late trypsins continue this digestion. Studies have shown that the transcription of the early trypsins is under strict regulatory control of JH (Noriega and Wells, 1999).

A recent microarray study performed in our lab investigating the protease inhibitor (PI) induced response in the midgut of another locust species, *L. migratoria*, demonstrated that JH putatively also regulates the expression of several digestive enzymes in this insect species (Spit et al., 2016). This chapter will explain the experiments that were conducted to evaluate the possible role of JH on the expression of serine proteases in the midgut of *L. migratoria*.

6.1.2. The PI induced response in *L. migratoria*

Plants have evolved a wide range of specific protective strategies against herbivorous insects. Among these defensive mechanisms, the secondary metabolites, such as PIs, are the most studied (Mello and Silva-Filho, 2002). These PIs primarily act in the digestive tract of the insect by targeting proteolytic digestive enzymes, or interfering with the structural properties or protective functions of the gut (Jongsma, 1997). Resulting amino acid deficiencies will eventually manifest themselves in a limitation of the insect's growth, development, fecundity, and can ultimately lead to an increased mortality. However, several important pest insects have evolved mechanisms to compensate for the antimetabolic effects of PIs, with attenuation, or a complete lack of developmental defects as a result. This co-evolutionary relationship and the therefrom resulting PI induced compensation mechanisms observed in major pest insects are described in more detail in chapter one (Ch1|1.4.1).

Over the past decades, the number of published studies describing these different counter-adaptive strategies in insects has grown (See, for example: Jongsma, 1997; Mello and Silva-Filho, 2002; Lopes et al., 2004; Srinivasan et al., 2006; Zhu-Salzman and Zeng, 2015). One such study performed in our lab investigated the PI induced response in *L. migratoria*. It was demonstrated that this locust species copes with the detrimental effects of ingested PIs by swiftly inducing the overall expression and release of active serine proteases, hence outnumbering the PIs in the midgut lumen (Spit et al., 2014). Especially serine proteases belonging to cluster I in the phylogenetic tree of serine proteases in *L. migratoria*: *Lm*Chy1 (GenBank: BK008825), *Lm*Chy2 (GenBank: BK008826), *Lm*Chy3 (GenBank: BK008827), *Lm*Chy4 (GenBank: BK008828), *Lm*Try1B (GenBank: BK008820), *Lm*Try2A (GenBank: BK008822), and *Lm*Try2B (GenBank: BK008823), shown in figure 6.1, were found to be

consistently upregulated for a prolonged time after PI uptake. Moreover, members of this cluster were previously predicted to constitute the majority of protease activity inside the *L. migratoria* gut (Lam et al., 1999, 2000; Spit et al., 2014).

Additionally, we conducted a microarray analysis to investigate the regulation of this PI induced response at the transcript level (Spit et al., 2016). This study revealed that the presence of PIs in the midgut lumen resulted in the up- and downregulation of a wide range of transcripts. In general, these data showed that fewer resources were invested in immunity and structural integrity of the gut, while numerous digestive enzymes were upregulated to compensate for the presence of the PIs in the intestinal tract. Moreover, the microarray analysis revealed the strong upregulation of a group of hexamerin and JHBP encoding transcripts in response to the dietary uptake of PIs. Hexamerins are best known as hemolymph storage proteins providing amino acids and energy during non-feeding periods in insects (Burmester, 1999). Apart from their function as storage proteins, in *L. migratoria*, hexamerins were also reported as JHBPs that function in the regulation of JH levels (Braun and Wyatt, 1996). Such JH binding capacity of hexamerins was later also demonstrated in the termites of the genus *Reticulitermes* (Zhou et al., 2006). Interestingly, the putative involvement of JH in the PI induced response was also suggested in several other insect species (Chi et al., 2009; Govind et al., 2010; Petek et al., 2012). Because of this specific function of hexamerins as JHBPs, the possible role of hexamerins and JH in the regulation of proteolytic digestion in the migratory locust was further investigated.

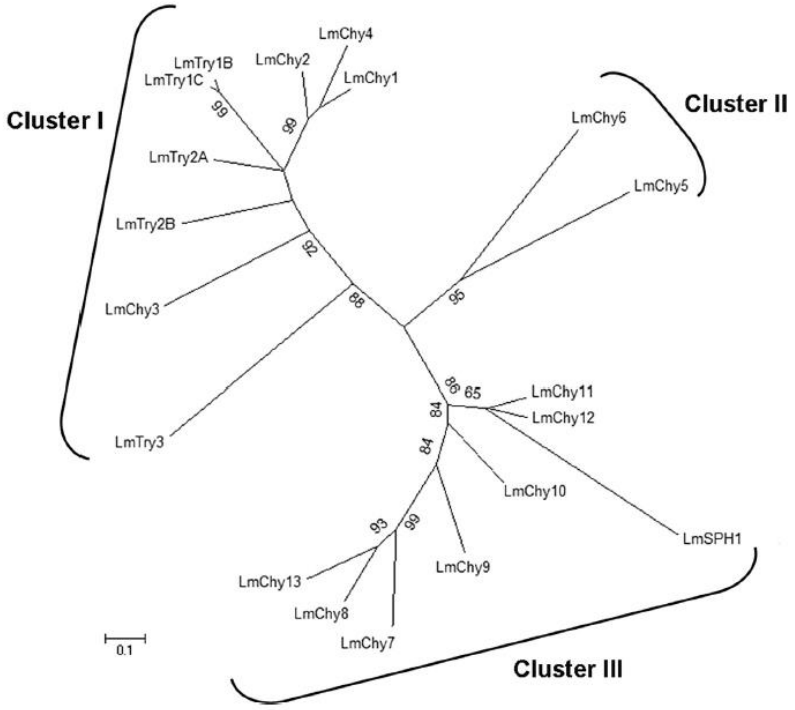


Figure 6.1. Unrooted maximum likelihood phylogenetic tree of *L. migratoria* serine proteases. The consensus tree based on 500 bootstrap replicates is presented, showing three different clusters. Only bootstrap values higher than 60 are shown. The chymotrypsins *Lm*Chy1, *Lm*Chy2 and *Lm*Chy4 cluster together in cluster I. Figure according to Spit *et al.* (2014).

After careful sequence investigations, the group of upregulated hexamerins was reduced to a total of three unique hexamerin encoding transcripts, namely *LmHex2* (GenBank: FJ609739.1), *LmHex3* (GenBank: FJ609740.1), and *LmHex6* (accession number: BK009413). To examine the role of these hexamerins in the PI induced response in *L. migratoria*,

simultaneous RNAi mediated knockdowns of their encoding transcripts were performed, while feeding the insects with either control or PI-containing diet. Simultaneously silencing *LmHex2*, *LmHex3*, and *LmHex6* strongly reduced the PI induced upregulation of three chymotrypsins: *LmChy1*, *LmChy2*, and *LmChy4*. Reduced compensatory upregulation of this cluster of chymotrypsins resulted in significantly reduced weight gain observed in RNAi animals fed with the PI-containing diet, illustrating that their PI induced response was compromised by the treatment. Interestingly, these chymotrypsins tightly clustered in a phylogenetic analysis of *L. migratoria* serine proteases (Fig. 6.1) (Spit et al., 2014). These observations suggest that the expression of these genes might be under the same regulatory control mechanism. The transcript levels of other serine proteases known to be part of the PI induced response in *L. migratoria*, namely *LmChy3*, *LmTry1B*, *LmTry2A*, and *LmTry2B*, remained unaffected by the knockdowns, and their usual upregulation after PI ingestion was still observed (Spit et al., 2016).

In conclusion, the following mechanism was suggested. When the diet of *L. migratoria* contains elevated levels of PIs, more proteases are needed in the midgut lumen as part of a compensatory mechanism. One plausible strategy to meet these demands relies on a JH-stimulated overexpression of digestive proteases. In order to capture more JH in the digestive tract, hexamerins (or other JHBPs) may act as specific carriers of the hormone. Hence, it was hypothesized that after PI ingestion, transcript levels of hexamerins are swiftly upregulated in the gut to assist the JH-stimulated compensatory expression of certain proteases (Spit et al., 2016).

The ability of *L. migratoria* hexamerins to bind and transport JH in the hemolymph opened the question whether JH might also be mediating the expression of this cluster of three chymotrypsins under regular dietary condition. This chapter will describe the experiments that were conducted to validate this hypothesis.

6.1.3. Aims of the experiment

In this chapter, the role of JH on the expression of serine proteases in the midgut of *L. migratoria* under regular dietary conditions was evaluated in two distinct ways.

(1) First, the expression levels of important serine proteases in the midgut were analyzed upon interfering with JH signaling through the knockdown of the sole known JH receptor, methoprene tolerant (*Lm*Met) (Huang et al., 2015).

(2) Second, the expression levels of important serine proteases in the midgut were analyzed upon stimulation of JH signaling through topical application of methoprene, a widely-used JH analogue in insects (Kamita et al., 2011).

6.2. Materials and methods

6.2.1. Rearing of animals and sample collection

Locusts (*Locusta migratoria*) were reared under crowded conditions in large cages in which temperature (32 °C), ambient relative humidity (40-60%) and light (14 h photoperiod) were kept constant. Locusts were fed daily with dry oat flakes and grass. For experiments, larvae were developmentally synchronized on the day of the 5th larval molt (Day 0). Midguts were dissected in *L. migratoria* Ringer's solution (composition: 9.82 g/l NaCl; 0.48 g/l KCl; 0.19 g/l NaH2PO4; 0.25 g/l NaHCO3; 0.73 g/l MgCl2; 0.32 g/l CaCl2; pH 6.5) and pooled in 2 mL MagNA Lyser Green Bead vials (Roche) for subsequent total RNA extraction. Each pool consisted of five midguts each. These vials were immediately transferred to liquid nitrogen to prevent RNA degradation. Tissues were stored at -80 °C until prior to RNA extraction.

6.2.2. Synthesis of dsRNA for RNAi studies

The dsRNA constructs for *GFP* (589 bp) and *LmMet* (421 bp) were synthesized using the MEGAscript® RNAi Kit (Ambion), according to the manufacturer's instructions. The procedure is based on the high-yield transcription reaction of a user-provided linear template transcript with a T7 promoter sequence attached. This is done by synthesizing both sense and antisense strands from a PCR-generated DNA template containing T7 RNA Polymerase promoters on both 5' ends of the primers. Primer sequences are presented in Table 6.1. cDNA derived from midguts of fifth larval *S. gregaria* was used as a template for the PCR reactions. A PCR reaction with REDTaq® DNA polymerase (Sigma-Aldrich©) was used to synthesize the linear template transcript for dsRNA construction. For the production of GFP dsRNA, an amplified GFP cDNA fragment was cloned in both sense and antisense direction in a TOPO 4.1 sequencing vector (Life technologies), containing a T7 promotor site, and was subsequently used as template for dsRNA production. The length of the generated PCR products was checked before dsRNA synthesis by performing 1% gel electrophoresis. The final concentration of the produced dsRNA was assessed using a NanoDrop ND-1000 UV-VIS Specro-photometer (Thermo Scientific™), and 1% agarose gel electrophoresis was performed to assess the integrity of the dsRNA.

Table 6.1. | Oligonucleotide sequences for primers used for dsRNA construct design.

TARGET GENE	FORWARD PRIMER	REVERSE PRIMER
LmMet	TAATACGACTCACTATAGGGTTAGGGC AGCATCAGAAAG	TAATACGACTCACTATAGGGTCGTCGGG AGGAAGTGTAT
GFP	TAATACGACTCACTATAGGGAAGGTGA TGCTACATACGGAA	TAATACGACTCACTATAGGGATCCCAGC AGCAGTTACAAAC

Underlined sequences are the T7 promoter sequences.

6.2.3. Met knockdown experiment

At the first day after ecdysis (day 0), a total of 85 5th instar larvae were divided in two groups. One group was injected with 600 ng of *Lm*Met dsRNA, while the other group was injected with 600 ng of dsRNA for GFP as a control. Since initial knockdowns of the receptor proved difficult, locusts received a boost injection at day 2 to maximize the effect. During the experiment, locusts were fed ad libitum with grass. On day 4, midguts were collected 3 hours post feeding

for total RNA extraction. Tissues for RNA extraction were dissected in 6 pools of at least 5 individuals.

6.2.4. Methoprene treatments

At the first day after ecdysis (day 0), 60 male larvae were divided in two groups. An experimental group of 30 larvae was topically treated with the JH III hormone mimic, methoprene (Sigma-Aldrich©), dissolved in acetone, while a control group received only acetone. The locusts were treated with a daily dose of 100 µg for four consecutive days. During the experiment, locusts were fed with grass ad libitum. On day 4, midguts were collected 3 hours post feed for total RNA extraction. Tissues for RNA extraction were dissected in 5 pools of at least 4 individuals.

6.2.5. RNA extraction and cDNA synthesis

To each MagNA Lyser Green Beads vial (Roche) containing midgut tissues, 1 mL QIAzol lysis reagent buffer (Thermo Scientific™) was added. Next, the tissues were homogenized for 30 seconds at 6000 x g using a MagNA Lyser instrument (Roche). Subsequent total RNA extraction from the midgut tissues was performed using the RNeasy Lipid Tissue extraction kit (Qiagen) according to manufacturer's instructions. An additional DNaseI treatment (Qiagen) was performed to remove contaminating genomic DNA. Concentration of extracted RNA was assessed using a NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Scientific™). Prior to RT-qPCR analyses, for each sample, equal amounts of RNA (1 µg) was reverse transcribed to complementary DNA (cDNA) using the PrimeScript™ RT Reagent Kit (Takara) following the manufacturer's protocol.

6.2.6. RT-qPCR analyses

RT-qPCR primers for reference genes and target genes were designed using the online tool primer3plus (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi). Primer sets were validated by designing relative standard curves for gene transcripts with a serial ten-fold dilution of a calibrator cDNA sample. Efficiency of RT-qPCR and correlation coefficient (R^2) were measured for each primer pair. Primers for RT-qPCR are given in table 6.2. All PCR reactions were performed in duplicate in 96-well plates on a StepOne System (ABI Prism, Applied Biosystems). Each reaction contained 5 µl fast Sybr Green, 0.5 µl Forward and Reverse primer (10 µM), 1.5 µl MQ water, and 2.5 µl of diluted cDNA. For all RT-qPCR reactions, the following thermal cycling profile was used: 50 °C for 2 min, 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. Finally, a melt curve analysis was performed to check for primer dimers. For all transcripts, only a single melting peak was found during the dissociation protocol. Relative transcript levels were calculated using the ΔΔCt method according to (Livak and Schmittgen, 2001). To correct for sample to sample variation, expression was normalized against two housekeeping genes, *LmRP49* and *LmRPS13,* as determined with the geNorm software (Vandesompele et al., 2002).

Table 6.2 | Oligonucleotide sequences for primers used for RT-qPCR.

TARGET GENE	FORWARD PRIMER	REVERSE PRIMER
LmChy1	GCCAGCTCCGGCTACAAG	CACCTGTACGACGGCAATGT
LmChy2	TGGCAGCTGTCCTTCCAGTA	ACCCAGTCGGAGCTGATGAT
LmChy3	ACGGCGACTTCGATACCTACTC	CCACGGCCTGAATATTGTCA
LmChy4	ACCCAAACCGACATAGCAGAGT	ACGCCCCGCACATGTG
LmTry1B	CAGTGGCAACGACTACGACATC	CACGTTGGTACCGAAGCTGAA
LmTry2A	TCGAGGGCTCCTACATCAACTAC	TGCATGACACCGATGTCGTA
LmTry2B	GACGTGGCCACCATCATTG	CCGTCTATCTCAAGGAGTGCAA
LmMet	TTAGGGCAGCATCAGAAAG	TCGTCGGGAGGAAGTGTAT
LmHex2	CCTATTTGGTCACATTGTCGATCC	GCTGGGTTTCGGGCTGTT
LmHex3	CATCTCTCTCTTCGGCCACATT	GCTGTTCGAGAACGCTAGCA
LmHex6	GCTCCATCTACCGGTACCTCATCT	GGCAGGTGCGCCGTACT
LmRP49	CGCTACAAGAAGCTTAAGAGATCATG	CTTAAACCTACGGCGAACTCTGTT
LmRPS13	CGTGAAGGGTGATATTTTGCA	GCTGACTGGGATATACCCTTACCAG

6.2.7. Statistical analyses

All expression data were always tested for normality using the D'Agostino-Pearson omnibus normality test prior to analyzing statistical differences. If the data were normally distributed, significant differences between the control group and the treatment group were analyzed using student t-tests. If the data were not normally distributed, significant differences between the control group and the treatment group were analyzed using the nonparametric Mann-Whitney-Wilcoxon (MWW) test. Differences were considered statistically significant if $p < 0.05$. All statistical analyses were performed using GraphPad Prism software (Version 6).

6.3. Results

An RNAi-mediated knockdown of the *L. migratoria* Met gene (*LmMet*) [GenBank: KF471131.1] was performed using an identical dsRNA construct as reported by Guo *et al.* (2014). An average knockdown efficiency of 50% was observed in the brain (Supp. Fig. S6.1 A). However, reaching similar knockdown values in the midgut proved difficult, and, remarkably, almost no effect of the *dsLmMet* treatment was observed on the average *LmMet* transcript levels in the midgut (Supp. Fig. S6.1 B). Nevertheless, effects on protease expression in the midgut were still clearly detectable. Figure 6.2 A illustrates that the treatment with *dsLmMet* resulted in significantly decreased transcript levels of the chymotrypsin genes, *LmChy1*, *LmChy2* (MWW; $p < 0.05$), while transcript levels of *LmChy4* showed a strong trend in the same direction. Interestingly, this is the same group of proteases that was affected by the hexamerin knockdown, as explained in the introduction of this chapter. Treatment of locust nymphs with methoprene resulted in the opposite effect. As shown in figure 6.2 B, the transcript levels of *LmChy1*, *LmChy2* and *LmChy4* were significantly induced in the methoprene treated locusts compared to the control group (MWW; $p < 0.05$). None of the other serine proteases that are known to exhibit PI inducible expression were affected by the two different treatments (Fig. 6.3 A-B). In addition, the expression of the hexamerin genes did not alter significantly upon knocking down *LmMet*; however, strong upwards trends were visible (Fig. 6.3 A). Such putative upregulation of hexamerin gene expression could possibly be part of a compensatory mechanism trying to restore the JH signaling. A significant stimulation in the expression of the hexamerin genes, *LmHex2*, *LmHex3*, and a strong trend in the same direction of *LmHex6*, in

the midgut were observed after topical methoprene application. It is possible that hexamerin production was induced for transporting the administered methoprene through the haemolymph, which would confirm the successful absorption of methoprene, as well as the subsequent stimulation of JH signaling in the insects (Fig. 6.3 B).

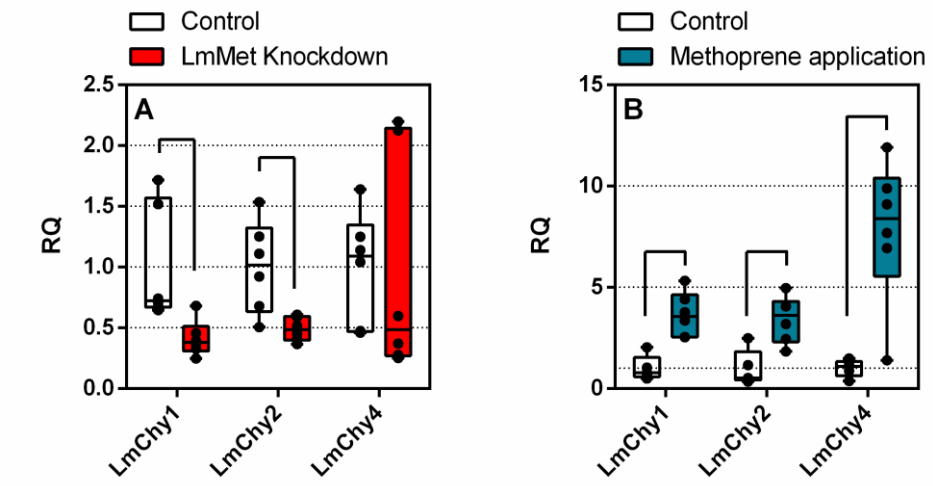


Figure 6.2. Relative transcript levels (RQ) in the midgut of *LmChy1*, *LmChy2* and *LmChy4* following *LmMet* knockdown (A) and following topical application of methoprene (B). Transcript levels in the midgut were normalized against two reference genes, *LmRP49* and *LmRPS13*. Boxplots are presented based on six pools of at least four individuals per pool. Significant differences are visualized by connecting black lines ($p < 0.05$, MWW). Figure according to Spit *et al.* (2016).

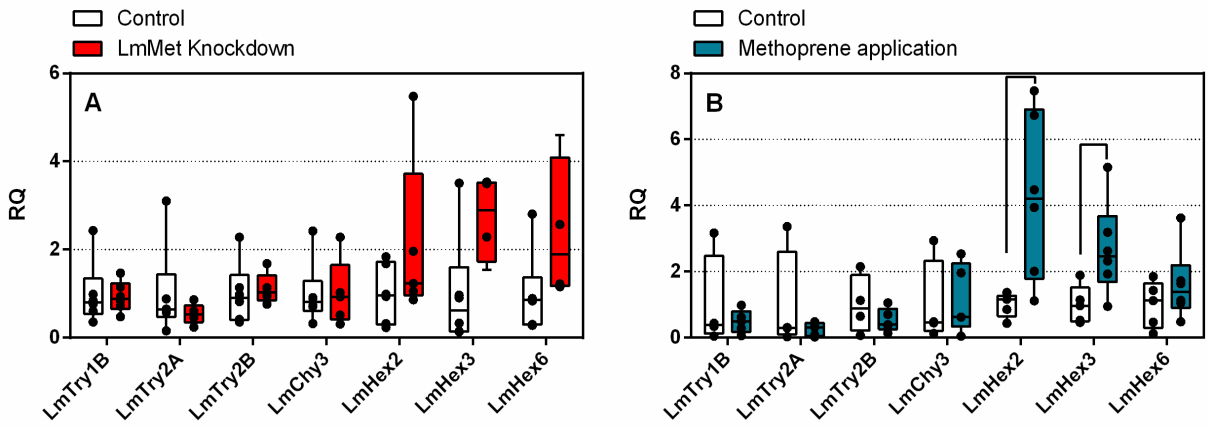


Figure 6.3. Relative transcript levels (RQ) in the midgut of *LmTry1A*, *LmTry1b*, *LmTry2A*, *LmHex2*, *LmHex3*, and *LmHex6* after knockdown of *LmMet* (A) or after topical application of methoprene (B). Transcript levels in the midgut were normalized against two reference genes, *LmRP49* and *LmRPS13*. Boxplots are presented based on six pools of at least four individuals per pool. Significant differences are visualized by connecting black lines ($p < 0.05$, MWW). Figure according to Spit *et al.* (2016).

6.4. Discussion

Juvenile hormone is an ubiquitous insect hormone involved in key biological processes in all insects. As already described earlier in this thesis, it is best known for its role during larval molting and reproductive maturation (For in-depth reviews, see: Jindra et al., 2013; Cheong et al., 2015; Li et al., 2019). We could demonstrate that JH is also involved in the regulation of the expression of several digestive enzymes in the midgut of *L. migratoria*. Silencing *LmMet*,

a known JH receptor, by means of RNAi resulted in the reduced expression of three related chymotrypsin genes, *LmChy1*, *LmChy2*, and *LmChy4*, in the midgut of *L. migratoria* (Fig. 6.2 A). In contrast, the topical application of methoprene, a widely used JH analogue, resulted in the opposite effects, and *LmChy1*, *LmChy2*, and *LmChy4* transcript levels significantly increased (Fig. 6.2 B). Moreover, both treatments had no effect on the expression of other important serine proteases in the *L. migratoria* midgut. Remarkably, the transcript levels of *LmMet* in the midgut were not significantly reduced upon *dsLmMet* treatment as opposed to the successful knockdown of *LmMet* obtained in the brain. Therefore, further evaluations are necessary to elucidate whether or not the midgut is the target organ of JH when regulating digestive enzyme levels, or if other tissues play a significant role in this regulation.

These results demonstrated a stimulatory effect of JH signaling on this specific group of serine proteases, further hinting towards a conserved role for this hormone in the regulation of digestion in insects. JH has already been described to regulate the expression of digestive serine proteases in different insect species (Noriega and Wells, 1999; Bian et al., 2008; Sui et al., 2009; Zhang et al., 2010; Cornette et al., 2013). Based on their feeding behavior, insects can roughly be subdivided into two major categories: the non-continuous feeders (predators and hematophagous insects, such as mosquitoes), and the (more) continuous feeders (e.g. many Coleoptera, Lepidoptera, Orthoptera) (Ch2|2.2). In general, it is believed that the digestive process in continuous feeders is more or less constitutive, whereas in discontinuous feeders, the digestive system is more tightly regulated (Lehane et al., 1995). For example, JH has been demonstrated to play a distinct role in the regulation of the digestive system of the mosquito, *A. aegypti,* which is an obvious discontinuous feeder (Noriega and Wells 1999). Interestingly, phylogenetic analyses have shown a clear divergence of insect trypsins and chymotrypsins. Dipteran early trypsins appear to be more related to an ancient trypsin molecule, while late trypsins show much more sequence resemblance with insect chymotrypsins (Marshall et al., 2008; Spit et al., 2014). Furthermore, *Lm*Chy1, *Lm*Chy2 and *Lm*Chy4 also appear to have evolved from a trypsin ancestor (Spit et al., 2014). The present data therefore indicate that the digestive system of continuous feeders, such as locusts, could also be regulated up to a certain level by JH, and that this activity of JH on the expression of several digestive enzymes could be present in a much broader group of insects. This leads us to hypothesize an evolutionary conserved role of JH in the digestive system in insects, which will have to be further investigated in more species.

Our RNA-Seq study also suggested a possible role for JH in the regulation of digestion in *S. gregaria*. This analysis revealed the significant upregulation of three transcripts encoding JHBPs and one transcript encoding a HSP83 in the midgut two hours after feeding (Ch4|4.4.3.1). Interestingly, one of the upregulated JHBP encoding transcripts, *TR89612|c1_g1*, showed high sequence similarity (83%) with a JHBP that was also found to be upregulated as part of the PI induced response in the midgut of *L. migratoria* (Spit et al., 2016). The current chapter could clearly demonstrate the role of JH in the regulation of digestive enzyme expression in *L. migratoria*. In addition, in *D. melanogaster*, HSP83 is described to facilitate the JH-induced nuclear import of Met necessary for downstream JH signaling via *Kr-h1* (He et al., 2014). Therefore, the following signaling cascade in response to food uptake in *S. gregaria* can be hypothesized. Once food is present in the midgut, specific JHBPs, acting as specialized carriers of JH, are upregulated in order to attract more JH towards the midgut. Once arrived at the midgut, JH interacts with its known receptor Met, inducing the downstream cellular signaling of JH via *Kr-h1*. This was further suggested by the observed upregulation of an HSP83 encoding transcript after feeding. This JH signaling in the midgut cells could then, on its turn, stimulate the expression of certain digestive enzymes, as observed in *L. migratoria*. To further investigate if certain *S. gregaria* digestive enzymes are

indeed under the regulatory control of JH, similar experiments as explained in this chapter for *L. migratoria* can be performed in the future.

6.5. Conclusion

In this chapter, it was demonstrated that under regular dietary conditions, JH signaling (via the Met receptor) elicits a significant stimulatory effect on the expression of a specific group of chymotrypsin-like serine proteases in the midgut of *L. migratoria*. Furthermore, the RNA-Seq study of the *S. gregaria* midgut also suggested a possible involvement of JH in the digestive process of this locust species. In the future, similar studies as presented in this chapter, may help further elucidating this proposed role of JH in the desert locust.

6.6. References

Bian G, Raikhel AS, Zhu J (2008) Characterization of a juvenile hormone-regulated chymotrypsin-like serine protease gene in *Aedes aegypti* mosquito. Insect Biochem Mol Biol 38:190–200. doi: 10.1016/j.ibmb.2007.10.008

Braun RP, Wyatt GR (1996) Sequence of the hexameric juvenile hormone-binding protein from the hemolymph of *Locusta migratoria*. J Biol Chem 271:31756–31762. doi: 10.1074/jbc.271.49.31756

Burmester T (1999) Evolution and function of the insect hexamerins. Eur J Entomol 96:213–225

Cheong SPS, Huang J, Bendena WG, Tobe SS, Hui JHL (2015) Evolution of ecdysis and metamorphosis in arthropods: The rise of regulation of juvenile hormone. Integr Comp Biol 55:878–890. doi: 10.1093/icb/icv066

Chi YH, Salzman RA, Balfe S, Ahn J-E, Sun W, Moon J, Yun D-J, Lee SY, Higgins TJ V., Pittendrigh B, et al (2009) Cowpea bruchid midgut transcriptome response to a soybean cystatin - costs and benefits of counter-defence. Insect Mol Biol 18:97–110. doi: 10.1111/j.1365-2583.2008.00854.x

Cornette R, Hayashi Y, Koshikawa S, Miura T (2013) Differential gene expression in response to juvenile hormone analog treatment in the damp-wood termite *Hodotermopsis sjostedti* (Isoptera, Archotermopsidae). J Insect Physiol 59:509–518. doi: 10.1016/j.jinsphys.2013.02.002

Govind G, Mittapalli O, Griebel T, Allmann S, Böcker S, Baldwin IT (2010) Unbiased Transcriptional Comparisons of Generalist and Specialist Herbivores Feeding on Progressively Defenseless *Nicotiana attenuata* Plants. PLoS One 5:e8735. doi: 10.1371/journal.pone.0008735

Guo W, Wu Z, Song J, Jiang F, Wang Z, Deng S, Walker VK, Zhou S (2014) Juvenile Hormone-Receptor Complex Acts on Mcm4 and Mcm7 to Promote Polyploidy and Vitellogenesis in the Migratory Locust. PLoS Genet 10:. doi: 10.1371/journal.pgen.1004702

He Q, Wen D, Jia Q, Cui C, Wang J, Palli SR, Li S (2014) Heat Shock Protein 83 (Hsp83) Facilitates Methoprene-tolerant (Met) Nuclear Import to Modulate Juvenile Hormone Signaling. J Biol Chem 289:27874–27885. doi: 10.1074/jbc.M114.582825

Huang J-H, Jing X, Douglas AE (2015) The multi-tasking gut epithelium of insects. Insect Biochem Mol Biol 67:15–20. doi: 10.1016/j.ibmb.2015.05.004

Jindra M, Bellés X, Shinoda T (2015) Molecular basis of juvenile hormone signaling. Curr Opin Insect Sci 11:39–46. doi: 10.1016/j.cois.2015.08.004

Jindra M, Palli SR, Riddiford LM (2013) The Juvenile Hormone Signaling Pathway in Insect Development. Annu Rev Entomol 58:181–204. doi: 10.1146/annurev-ento-120811-153700

Jongsma MA (1997) The Adaptation of Insects to Plant Protease Inhibitors. 43:885–895

Kamita SG, Samra AI, Liu J-Y, Cornel AJ, Hammock BD (2011) Juvenile Hormone (JH) Esterase of the Mosquito *Culex quinquefasciatus* Is Not a Target of the JH Analog Insecticide Methoprene. PLoS One 6:e28392. doi: 10.1371/journal.pone.0028392

Lam W, Coast GM, C. Rayne R (1999) Isolation and characterisation of two chymotrypsins from the midgut of *Locusta migratoria*. Insect Biochem Mol Biol 29:653–660. doi: 10.1016/S0965-1748(99)00049-1

Lam W, Coast GM, C. Rayne R (2000) Characterisation of multiple trypsins from the midgut of *Locusta migratoria*. Insect Biochem Mol Biol 30:85–94. doi: 10.1016/S0965-1748(99)00103-4

Lehane MJ, Blakemore D, Williams S, Moffatt MR (1995) Regulation of digestive enzyme levels in insects. Comp Biochem Physiol Part B Biochem Mol Biol 110:285–289. doi: 10.1016/0305-0491(94)00157-P

Li K, Jia Q, Li S (2019) Juvenile hormone signaling – a mini review. Insect Sci 26:600–606. doi: 10.1111/1744-7917.12614

Livak KJ, Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods 25:402–408. doi: 10.1006/meth.2001.1262

Lopes AR, Juliano MA, Juliano L, Terra WR (2004) Coevolution of insect trypsins and inhibitors. Arch Insect Biochem Physiol 55:140–152. doi: 10.1002/arch.10134

Marshall SDG, Gatehouse LN, Becher SA, Christeller JT, Gatehouse HS, Hurst MRH, Boucias DG, Jackson TA (2008) Serine proteases identified from a *Costelytra zealandica* (White) (Coleoptera: Scarabaeidae) midgut EST library and their expression through insect development. Insect Mol Biol 17:247–259. doi: 10.1111/j.1365-2583.2008.00798.x

Mello MO, Silva-Filho MC (2002) Plant-insect interactions: an evolutionary arms race between two

distinct defense mechanisms. Brazilian J Plant Physiol 14:71–81. doi: 10.1590/S1677-04202002000200001

Noriega FG, Wells MA (1999) A molecular view of trypsin synthesis in the midgut of *Aedes aegypti*. J Insect Physiol 45:613–620. doi: 10.1016/S0022-1910(99)00052-9

Petek M, Turnšek N, Buh M, Maruša G, Novak P, Gruden K, Slapar N, Popovič T, Štrukelj B, Gruden K, Štrukelj B, Jongsma MA (2012) A complex of genes involved in adaptation of *Leptinotarsa decemlineata* larvae to induced potato defense. Arch Insect Biochem Physiol 79:153–181. doi: 10.1002/arch.21017

Spit J, Holtof M, Badisco L, Vergauwen L, Vogel E, Knapen D, Vanden Broeck J (2016) Transcriptional Analysis of The Adaptive Digestive System of The Migratory Locust in Response to Plant Defensive Protease Inhibitors. Sci Rep 6:32460. doi: 10.1038/srep32460

Spit J, Zels S, Dillen S, Holtof M, Wynant N, Vanden Broeck J (2014) Effects of different dietary conditions on the expression of trypsin- and chymotrypsin-like protease genes in the digestive system of the migratory locust, *Locusta migratoria*. Insect Biochem Mol Biol 48:100–109. doi: 10.1016/j.ibmb.2014.03.002

Srinivasan A, Giri A, Gupta V (2006) Structural and functional diversities in lepidopteran serine proteases. Cell Mol Biol Lett 11:132–154. doi: 10.2478/s11658-006-0012-8

Sui Y-P, Wang J-X, Zhao X-F (2009) The impacts of classical insect hormones on the expression profiles of a new digestive trypsin-like protease (TLP) from the cotton bollworm, *Helicoverpa armigera*. Insect Mol Biol 18:443–452. doi: 10.1111/j.1365-2583.2009.00884.x

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:1–30. doi: Artn 0034.1Doi 10.1186/Gb-2002-3-7-Research0034

Zhang C, Zhou D, Zheng S, Liu L, Tao S, Yang L, Hu S, Feng Q (2010) A chymotrypsin-like serine protease cDNA involved in food protein digestion in the common cutworm, *Spodoptera litura*: Cloning, characterization, developmental and induced expression patterns, and localization. J Insect Physiol 56:788–799. doi: 10.1016/j.jinsphys.2010.02.001

Zhou X, Tarver MR, Scharf ME (2006) Hexamerin-based regulation of juvenile hormone-dependent gene expression underlies phenotypic plasticity in a social insect. Development 134:601–610. doi: 10.1242/dev.02755

Zhu-Salzman K, Zeng R (2015) Insect Response to Plant Defensive Protease Inhibitors. Annu Rev Entomol 60:233–252. doi: 10.1146/annurev-ento-010814-020816

Chapter 7

General discussion, conclusions and future prospects

7.1. *S. gregaria* midgut transcriptome

In chapter 3, an adjusted version of our in-house *S. gregaria* whole-body transcriptome (123,714 contigs) was constructed to be used as a reference for mapping the short sequencing reads generated from the RNA-Seq study of the midgut. The adjusted whole-body transcriptome was designed by firstly concatenating all available transcript isoforms of the original transcriptome into linear sequences, called SuperTranscripts, and secondly by removing all leftover redundant sequences according to a sequence similarity cut-off (Ch3|3.2.5). These modifications significantly reduced the large proportion of putative transcript isoforms, sometimes with strong sequence similarity, or putative redundant sequences that were present in the original transcriptome (Verdonck 2017, unpublished data). Subsequently mapping the RNA-Seq reads to the adjusted transcriptome instead of the original transcriptome, massively reduced the amount of multimapping sequencing reads, hence increasing the power of all downstream analyses (Ch3|3.4.1.2). This demonstrated the advantage of performing these modification steps prior to mapping. It also highlighted some of the typical challenges of performing an RNA-Seq study in a non-model organism, such as *S. gregaria,* for which no genome database is available. For now, the adjusted transcriptome represents a solid substitute for the original in-house *S. gregaria* whole-body transcriptome for RNA-Seq studies. However, in the future, the availability of a genome sequence database of *S. gregaria* will probably replace the reference transcriptome for RNA-Seq studies. In general, reference genomes are preferred over reference transcriptomes when performing RNA-Seq. A well-assembled genome allows to study not only differential gene expression, but, for example, also differential expression of transcript isoforms, and will therefore aid in identifying authentic transcript isoforms present in the original whole-body transcriptome.

The final midgut reference transcriptome for RNA-Seq was constructed by extracting all relevant (based on their CPM) midgut transcripts identified via RNA-Seq from the adjusted whole-body transcriptome. This transcriptome contained 19,345 contigs covering the vast majority of unique mRNA (cDNA) sequences present in the mRNA (cDNA) library of the *S. gregaria* midgut (Ch3|3.4.1.3). A total of 10,939 contigs (56.6%) could be annotated by performing BLASTs of the unknown sequences in publicly available annotated databases, while 8,406 (43.4%) were left without a high-throughput annotation. Such high numbers of transcripts without a functional annotation were also observed in several other *de novo* gut transcriptome assemblies of other insect species (Ch3|3.4.1.5). Since the annotation of the transcripts in this study was based on BLASTs in publicly available annotated databases, it is possible that some of these *S. gregaria* sequences were too divergent from sequences available in any of the inspected databases, and were therefore not retrieved by BLAST. Such transcripts might possibly be species-specific or specific to a small subgroup of arthropods, and consequently not yet available in any annotated database of sequences. It is reasonable to assume that a vast proportion of transcripts has not yet been functionally investigated and are therefore missing in the searched annotated databases. These results clearly indicate the high necessity of more functional research on gene products in insects or other less investigated organisms. Only then will it be possible to further characterize the as yet unknown elements in the transcriptome of non-model organisms, such as *S. gregaria*. In addition, it is possible that a small number of transcripts might still be annotated when investigating other available databases on NCBI, such as the Nucleotide collection (nt/nr) database containing GenBank records. These BLASTs can be performed in the future to further complement the *S. gregaria* midgut reference transcriptome.

7.2. Differential gene expression

7.2.1. Identification of differential gene expression

The mRNA profile of the *S. gregaria* midgut was examined at three different timepoints, being ten minutes, two hours, and twenty-four hours after feeding, in order to study gene expression during the initial phase of digestion, the ongoing digestion, and a short period of food deprivation, respectively (Chapter 4). However, it appeared that the transcript profiles of the samples representing the midgut during the initial phase of digestion, ten minutes after feeding, had characteristics of both other timepoints, resulting in large within-group variation of transcript quantities. This observation probably demonstrates the high rate at which the *S. gregaria* midgut responds to the presence of food in the alimentary canal. Already at ten minutes after feeding, when the food was only present in the foregut, the midgut of (most) locusts appeared to display the transcript changes that were also observed two hours after feeding, when the food bolus was mainly located in the midgut. To the best of my knowledge, this is the first time that such a high rate of transcriptomic changes has been demonstrated in insect midgut. It is however possible that the insects responded faster than usual to the uptake of food after a non-feeding period of twenty-four hours, which possibly approximated a preliminary state of starvation. Moreover, because of the high within-group variation ten minutes after feeding, it was not possible for edgeR to detect significant differential expression when statistically comparing the transcript quantities at this timepoint to the other timepoints (Ch4|4.4.1). Therefore, conclusions about the transcript profile at this timepoint are rather preliminary and need to be taken with caution. Nevertheless, these findings are still highly interesting and should be further evaluated.

It might be possible that significant differences in gene expression ten minutes after feeding are detectable in tissues other than the midgut, which are possibly less susceptible to biological variation. Therefore, it should be considered to also investigate the mRNA profiles in the other tissues that were dissected during this experiment: the head, the brain, the foregut, the hindgut, or the Malpighian tubules. For example the brain, as a major site for neuropeptide production and release, has already been described to regulate digestive processes in the insect midgut, and investigating its transcript profile will further expand our scientific knowledge of the regulation of feeding and digestion in *S. gregaria* (Nässel and Homberg, 2006; Spit et al., 2012). The rhythmic contractions of the foregut are important for the movement of the food bolus through the alimentary canal and are therefore pivotal for the onset, and maintenance, of digestion. Investigating the transcript profile of the foregut might therefore also result in the identification of transcripts mediating this process. Moreover, in this context, it should be considered to also dissect the SNS for further transcriptomic analysis.

We were eventually able to discover 781 significantly differential transcripts, 569 upregulated and 212 downregulated, in the midgut two hours after feeding compared to twenty-four hours after feeding (Ch4|4.4.2). Among these, 305 upregulated and 88 downregulated transcripts could be annotated, still leaving a large proportion of transcripts without annotation. It was already mentioned earlier that such high amounts of transcripts without an annotation are frequently observed in insect transcriptomic studies, foremost indicating a lack of functional research of gene products in insects and putative strong sequence divergence among insect species (Ch3|3.4.1.5). Such non-conserved or even species-specific transcripts are potentially highly interesting for future applications, such as the development of novel insecticidal products and strategies, and should be examined in more detail.

This doctoral research has mainly focused on the already annotated differential transcripts two hours after food uptake. Among the upregulated transcripts, the majority was predicted to mediate midgut physiology, enzymatic digestion, nutrient uptake, and detoxification of detrimental food compounds (Ch4|4.4.3.1). This study also provides an insight into the *S. gregaria* midgut transcriptome during a short period of food deprivation. In the absence of food, the insects invest significantly less in enzymatic digestion and nutrient absorption. In contrast, transcripts involved in mediating stress tolerance, tissue maintenance/protection, or immune defenses are more abundant during starvation. In summary, there appears to be a trade-off between digestion on the one hand, and immunity and integrity on the other hand in the presence and absence of food, respectively.

7.2.2. Investigation of two differentially expressed genes

Many of the upregulated transcripts in the midgut two hours after feeding were expected to be essential for the desert locust, probably mediating key processes in the midgut during digestion. In chapter 5, two of these upregulated transcripts were further investigated *in vivo*: an H^+ V-ATPase subunit a encoding transcript, denoted *Sg-VAHa_1*, and a NPC1b encoding transcript, denoted *Sg-NPC1b.* Both transcripts were deliberately chosen as they were expected to play a pivotal role in the digestive process: H^+ V-ATPases are well-known for generating favorable membrane potentials essential for transepithelial molecular transport, while NPC1b probably mediates the dietary sterol uptake in the insect midgut (Wieczorek et al., 2009; Holtof et al., 2019).

The expression of both *Sg-VAHa_1* and *Sg-NPC1b* was demonstrated to be midgut-specific, suggesting a specialized role in the midgut of the desert locust. Next, the viability of the locusts upon injection of dsRNA targeting either *Sg-VAHa_1* or *Sg-NPC1b* was examined. Silencing these transcripts resulted in high mortality rates within two weeks after the first injections, indicating the overall high efficiency of the treatments, as well as the vital importance of both transcripts (Ch5|5.3.2 and Ch5|5.3.3). However, the exact mechanisms resulting in the lethal effects of both injections need to be further addressed. Several hypotheses explaining the observed phenotypes upon knocking down both transcripts, as well as interesting follow-up experiments were proposed in the discussion section of chapter 5 (Ch5|5.4.1 and Ch5|5.4.2).

In summary, the results of the RNAi experiments in chapter 5 clearly demonstrated the possibility of discovering lethal targets among the upregulated transcripts in the midgut after feeding. These results not only pave the path for further investigating H^+ V-ATPase and NPC1b proteins as potentially lethal target sites in *S. gregaria* or other insect species, they also demonstrated that the list of differentially expressed genes from this RNA-Seq study could be an extremely valuable database for the future quest for candidate pesticide target genes. In conclusion, we believe that by demonstrating the role of these two annotated transcripts on the viability of the desert locust, we consequently also increased the confidence to investigate in the future other less known or hitherto unknown genes that showed a differential expression after feeding.

7.3. Improved knowledge of the *S. gregaria* digestive system

The generated midgut reference transcriptome as well as the temporal differential expression analysis of the midgut after food uptake have massively increased our general knowledge of the digestive process in the desert locust. In this paragraph, I will summarize some of the major

findings of this RNA-Seq study within the context of extracellular nutrient digestion as described in detail in chapter 1.

7.3.1. Enzymatic digestion

When food enters the body, the midgut responds by inducing the expression of several peritrophin encoding genes, establishing the PM that will envelop the food bolus. This observation is in line with earlier literature (Lehane, 1997). Inside the endoperitrophic space, the food is firstly degraded by the simultaneous action of numerous digestive enzymes. The RNA-Seq data demonstrated that the desert locust relies on a large array of digestive enzymes to mediate the enzymatic digestion of the food bolus. An overview of all the identified digestive enzyme encoding transcripts in the *S. gregaria* midgut reference transcriptome is presented in chapter 3 (Ch3|3.4.2.1 and Ch3|3.4.2.2). As a polyphagous insect, high amounts of digestive enzymes are necessary to efficiently digest the variable diet compositions (Patankar et al., 2001). Moreover, using a large set of digestive enzymes is also a common strategy to overcome plant defensive mechanisms specifically targeting insect digestive enzyme activity (Holtof et al., 2019). Interestingly, some of the identified digestive enzymes clearly exhibit an inducible expression pattern, with increasing expression when food is present in the midgut (Ch4|4.4.3.1). However, many other digestive enzymes appear to have similar transcript levels in the presence of food, ten minutes and two hours after feeding, as well as in the absence of food, twenty-four hours after feeding, indicating a constitutive expression profile. Our data thus demonstrated that the *S. gregaria* midgut harbors a complex mix of inducible and constitutively expressed digestive enzymes.

A large number of identified serine proteases in the midgut of *S. gregaria* (77 out of 106) were predicted to be enzymatically inactive, since they were lacking one or more conserved motifs indispensable for substrate recognition and catalytic activity of serine proteases (Ch3| 3.4.2.2.1). Although this has already been observed in other insect species, their physiological function still remains unclear. We could observe that the majority of these putative serine proteases had modest to high CPM in our RNA-Seq data, indicating that they were ubiquitously present in the midgut transcriptome of *S. gregaria* (Supp. Table S3.1). Moreover, we found that as much as twelve of these enzymatically inactive serine proteases were induced upon feeding, indicating that they might play a distinct role during the digestive process (Ch4|4.4.3.1). First, the presence of these transcripts, as well as their sequence, should be confirmed by for example PCR followed by Sanger sequencing. Next, their putative function and site(s) of action, *e.g.* extracellular or intracellular, should be further investigated to determine their role in the midgut physiology of *S. gregaria*. It would, for example, be interesting to purify all serine proteases from the midgut lumen and midgut tissue of *S. gregaria*. This way, it would be possible to identify which serine proteases are effectively translated, and to verify their possible activity. This would allow us to investigate the digestive capacity and substrate specificity of serine proteases present in the midgut of the desert locust and compare this to the serine proteases present in its midgut transcriptome.

This study illustrated that the midgut transcriptome of the desert locust contains many transcripts encoding protein and carbohydrate degrading enzymes, while less transcripts encoding lipid degrading enzymes appeared to be present in the midgut transcriptome. It is well-established that insect herbivores require a broad set of nutrients for their growth and development, typically including amino acids (mostly obtained from dietary protein), carbohydrates (sugars and starch), lipids (fatty acids, phospholipids and sterols), vitamins and

minerals (Chapman, 2013; Holtof et al., 2019). However, lipids, vitamins and minerals are considered micronutrients because they occur in plants at low levels, and generally insect herbivores require only small amounts (Deans et al., 2016). Except vital sterols, micronutrients are not considered limiting for insect herbivores. Dietary proteins and carbohydrates, on the other hand, are considered limiting nutrients for insect herbivores, and are therefore needed in larger amounts. These macronutrients provide the essential amino acids and energy, respectively, pivotal for their development, growth, and fecundity (Raubenheimer, 2003; Chapman, 2013; Le Gall and Behmer, 2014; Deans et al., 2016; Holtof et al., 2019). In general, the diet of herbivorous insects reflects these strong dietary needs as it is characterized by high carbohydrate (69%), high protein (20%), and low fat (2%) contents (Nguyen et al., 2015). The *S. gregaria* nymphs in this experiment were fed with raw cabbage, which is also characterized by high carbohydrate, high protein, and low fat contents. The specific nutritional contents of their diet might explain why their digestive enzyme transcript profile for protein and carbohydrate digestion appears to be more complex than that for lipid digestion.

Altogether, these findings give a very clear overview of the digestive enzyme encoding transcripts present in the midgut transcriptome of *S. gregaria.* Moreover, it provides, for the first time, an insight into the temporal transcript changes of digestive enzymes in the midgut of this insect species during the digestive process. Furthermore, as an important representative of polyphagous, herbivorous insects, this study also improves our general knowledge of the digestive enzyme production at the transcript level in herbivorous insects.

7.3.2. Dietary nutrient absorption

After the enzymatic breakdown of macro- and micronutrients present in the food, the released molcules are absorbed from the midgut lumen by specific nutrient transporters located at the apical side of the midgut ECs. This RNA-Seq study has illustrated that the *S. gregaria* midgut transcriptome contains all the described nutrient transporters of chapter 1 (Ch3|3.4.2.2.4). Their presence as well as the fluctuating transcript levels of some in response to food uptake seem to confirm the predicted models of nutrient uptake in insects. This information was used to construct a hypothetical model for dietary nutrient uptake in the *S. gregaria* midgut, as illustrated in figure 7.1.

This study has also highlighted some interesting novelties. It was demonstrated that many SLC36 encoding genes were present in the midgut reference transcriptome, among which four increased in expression in response to food availability, for the first time suggesting an active role for SLC36 proteins as nutrient amino acid transporters in insects. Additionally, several SLC6 and SLC7 encoding transcripts were also identified. These proteins, and especially SLC6, have already been linked to the uptake of dietary amino acids in several insect species (Boudko, 2012; Holtof et al., 2019). Moreover, the expression of one SLC6 and one SLC7 encoding gene also increased upon feeding. Furthermore, the upregulation of a SLC15 encoding transcript in the midgut after feeding indicates a distinct role of this transporter in the digestive process. Until now, only one insect SLC15 transporter studied in *D. melanogaster*, *DmOPT1*, has been demonstrated to mediate oligopeptide transport (Roman et al., 1998). The upregulation of a SLC15 encoding gene in the desert locust after food uptake suggests a similar function, but needs to be further examined. From the current study, we propose that the combined transporter action of SLC6, SLC7, SLC36, and SLC15 mediates the absorption of dietary amino acids, and the latter oligopeptides, in *S. gregaria* (Fig 7.1). Additional research is however needed to fully characterize the substrate preference of these transporters.

Far less is known and described about the uptake of carbohydrates and lipids in insects. Our RNA-Seq study however seems to confirm the general expectations as reviewed by Holtof *et al.* (2019). For example, the dietary monosaccharide absorption in the insect midgut is expected to follow the same pathways as described for mammals. In accordance, different putative monosaccharide transporters, mainly GLUT transporters, were identified in the *S. gregaria* midgut reference transcriptome. Disaccharide uptake in insects is expected to be mediated by Tret1 proteins (Kikawada et al., 2007; Holtof et al., 2019). The current study seems to confirm this hypothesis, since large amounts of Tret1 encoding transcripts were present in the midgut transcriptome. Furthermore, many of these increased in transcript levels after food uptake, suggesting a boosted activity when food is present in the midgut. Several FABP, FATP and Scavenger receptor class B type I (CD36) encoding transcripts were detected in the midgut reference transcriptome, which, according to literature, might be responsible for the uptake of free fatty acids from the diet (Majerowicz and Gondim, 2013; Holtof et al., 2019). However, only the transcript levels of three FABP encoding transcripts were induced upon feeding, hinting towards a putative crucial role of these proteins in the uptake of dietary free fatty acids. The experiments in chapter 5 also clearly highlighted the probable involvement of NPC1b in the uptake of vital dietary sterols, however, additional research is needed to fully confirm this. In addition, several SCPx and SCP2 encoding transcripts were identified in the *S. gregaria* midgut reference transcriptome. None of these were upregulated upon feeding, and therefore need to be further investigated in order to validate their function in dietary sterol uptake.

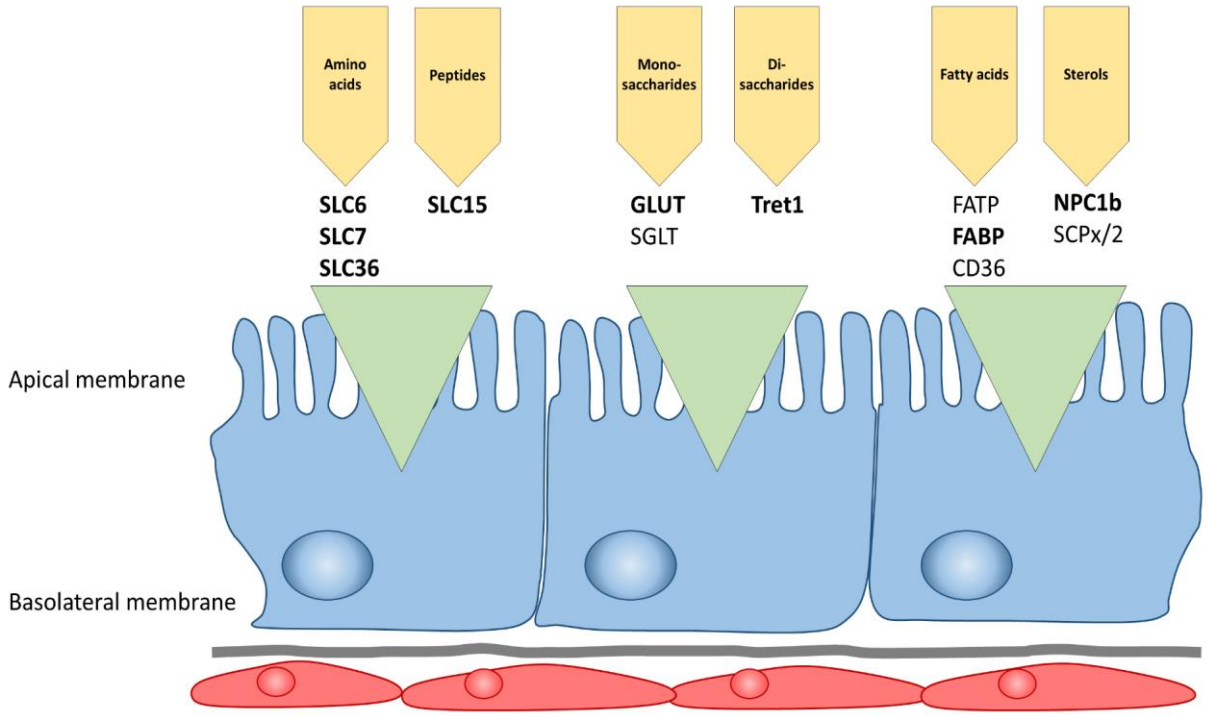


Figure 7.1. Proposed mediators of nutrient absorption in the midgut of *S. gregaria*. After protein breakdown, dietary amino acids are probably absorbed by midgut ECs through SLC6, SLC7, and SLC36 transporters, while dietary oligopeptides are probably absorbed through SLC15 transporters. Monosaccharides originating from dietary carbohydrate digestion are probably absorbed through GLUT and SGLT transporters, while disaccharides are probably absorbed through Tret1 transporters. Free fatty acids originating from dietary lipid digestion are probably absorbed through FATP, FABP, and CD36 transporters. The majority of dietary sterols is probably absorbed through NPC1b, and putatively through SCPx and SCP2 transporters. Cells in blue represent the midgut ECs that absorb nutrient substrates (orange arrows) from the midgut lumen through the action of nutrient transporters (green arrows) located at the apical cell membrane. In bold are the types of receptors of which one or more members are upregulated in the midgut of *S. gregaria* two hours after feeding.

Remarkably, our temporal differential expression analysis illustrated that many of the identified potential nutrient transporters exhibit an inducible expression pattern. Their expression in the midgut is swiftly upregulated in response to the presence of food. These data suggest that the desert locust can adequately respond to the increased availability of nutrients in the midgut lumen, and is the first study to report such swift and flexible upregulation of multiple nutrient transporters during the digestive process in insects.

Most of the amino acid and carbohydrate transporters identified in the *S. gregaria* midgut are secondary active transporters typically requiring the presence of Na^+ ions to translocate nutrients into the midgut ECs. In addition, the observed upregulation of the midgut specific *Sg-VAHa_1*, as part of the H^+ V-ATPase pump system, and a sodium/hydrogen exchanger two hours after feeding (Ch4|4.4.3.1) might indicate that more Na^+ ions are being excreted out of the cell during digestion. Consequently, this accumulation of Na^+ ions in the midgut lumen might on its turn be powering the nutrient transporters located at the apical site of the midgut epithelium, resulting in the uptake of free nutrients into the epithelial cells. This hypothesis can be further evaluated by measuring the transport of ions and nutrients across the epithelial membrane of the midgut in a so-called Ussing chamber. Briefly, an Ussing chamber is able to detect (small) changes in the membrane potential by measuring the changes in voltage between both sides (apical and basolateral) of an epithelium. This way, movement of ions or nutrients in response to a feeding stimulus can be easily analyzed.

7.3.3. Detoxification of food particles

Polyphagous herbivorous insects typically encounter diverse sets of plant defensive molecules in their diet and consequently possess a large arsenal of detoxification enzymes to efficiently sequester these detrimental dietary factors (Mello and Silva-Filho, 2002). The *S. gregaria* midgut transcriptome also contains large amounts of transcripts encoding detoxification enzymes. Many of these are upregulated soon after food uptake, demonstrating the desert locust's ability to quickly respond to harmful elements in its diet. Moreover, in this RNA-Seq study, it could be demonstrated that many of these transcripts encoded proteins that are able to specifically sequester poisonous substances typically present in cabbage, the standard diet of our locust colony. Whether or not our locust colony has developed specific adaptations to their standard diet can be further addressed by comparing the midgut transcriptome of our locust colony with other locust colonies or wild specimens. However, it was already demonstrated that *S. gregaria* in the Sahara desert are able to tolerate *S. purpurea* toxins, demonstrating the intrinsic capacity of this insect species to cope with Brassicaceae allelochemicals (Mainguet et al., 2000).

The identified detoxification enzymes can be further addressed to investigate the metabolic resistance of *S. gregaria* against plant allelochemicals. Furthermore, many of these enzymes are also expected to be responsible for the metabolic resistance of pest insects to insecticides (Hawkins et al., 2019). One promising way to combat insecticide resistance is by inhibiting the establishment of metabolic resistance in the targeted pest insect. Therefore, it is crucial to identify available insect detoxification enzymes and to understand their mode of action. This RNA-Seq research can be considered as a starting point for future studies investigating the various pathways of metabolic resistance present in *S. gregaria*.

7.3.4. The putative role of JH in digestion

One HSP83 and three JHBP encoding transcripts were strongly upregulated after feeding. These transcripts are predicted to be involved in the JH signaling, since JHBPs are known to mediate JH transport through the haemolymph and HSP83 is described to positively affect JH signaling in *D. melanogaster* through the JH receptor Met (Gilbert et al., 2000; He et al., 2014). In chapter 6, the stimulatory effect of JH on the expression of a group of related chymotrypsin-like transcripts was demonstrated in the midgut of another locust species*, L. migratoria* (Ch6|6.3). The current RNA-Seq study also suggest the involvement of JH in the digestive process of *S. gregaria*, hinting towards a conserved function of this hormone in the regulation of the digestive process in locusts. Nevertheless, this hypothesis needs to be further investigated. Experiments analogous to those conducted in chapter 6 can provide a first indication whether JH actively alters the transcript levels of certain serine proteases in the *S. gregaria* midgut. In the future, when a genomic database of *S. gregaria* is available, the promotor regions of gene sequences of enzymes putatively under the regulatory control of JH, can be examined for the presence of specific JH response elements.

7.4. Applicability for future pest management

The current study provides for the first time an insight into the midgut transcriptome of *S. gregaria* during the digestive process. This information can now be used to further investigate the regulation of feeding and digestion in this insect species. In addition, this study also offers a long and promising list of genes expressed in the midgut of the desert locust and regulated during the digestive process. This list might contain potentially interesting candidate targets for novel insecticides. This was illustrated in chapter 5, by demonstrating high levels of lethality following the RNAi-mediated knockdowns of two selected transcripts that were differentially upregulated after feeding. These were still preliminary tests, and more research is needed to further investigate their potential efficacy and selectivity, as well as the safety of targeting these genes or gene products for crop protection.

In theory, target gene(s) (products) for crop protection should ideally be non-conserved and cause rapid lethality. When such an ideal target has been validated, the next step is to develop/discover a (bio)pesticide that can specifically and efficiently disturb its activity. This might, for example, be a biologically available molecule, such as a plant allelochemical, already targeting the gene (product) in nature, or by synthesizing specific pesticidal molecules. Another possible and elegant strategy for pest control, that recently has received growing attention, is the use of RNAi to suppress vital genes, resulting in reduced insect fitness and/or mortality (Price and Gatehouse, 2008). An ideal target gene for RNAi should yield mRNA with a high turnover rate coding for proteins with a short half-life, because this would result in faster acting effects. Another important prerequisite for using dsRNA in pest control, is the sequence match between the siRNAs derived from the dsRNA and the target mRNA in the insect pest. At the same time, the sequences of these siRNAs need to be sufficiently divergent from mRNA sequences present in non-target organisms (Scott et al., 2013). Briefly, in theory, any non-conserved, lethal gene can be targeted by dsRNA resulting in the post-transcriptional silencing of this gene, which would ideally result in the rapid death of the targeted insect pest. The lethal dsRNA molecules can then, for example, be incorporated into genetically modified crops to be taken up by the pest insects upon feeding on these crops. This was for the first time performed by Baum *et al*. in 2007 by genetically engineering transgenic corn expressing dsRNA against the H^+ V-ATPase A encoding transcript of the Western corn rootworm *Diabrotica virgifera virgifera*. Insects eating from these transgenic crops displayed impeded growth and premature

death, resulting in the significant reduction of feeding damage to the transgenic crops (Baum et al., 2007). For more information on the applications of dsRNA in fighting pests or protecting beneficials, the reader is referred to an in-depth review by Vogel et *al.* (2019).

Important factors determining the overall success of RNAi in crop protection are: the appropriate target choice, the successful uptake of the dsRNA by the organism, and its cost-effectiveness for commercial production. The dose of dsRNA necessary to provoke lethality largely depends on several factors, including the targeted insect species, its life stage, the target gene concentration, its expression profile, and the method of delivery to the organism. Especially the latter has posed researchers for significant challenges. It is widely known that the efficiency of RNAi varies between insect species, and is largely dependent on species- and tissue-specific factors, such as presence of dsRNA degrading enzymes or the activity of (different components of) the RNAi machinery (Price and Gatehouse, 2008; Scott et al., 2013; Joga et al., 2016; Vogel et al., 2019).

For RNAi-based pest control, the efficiency of intestinal dsRNA uptake is crucial, since the proposed mechanisms for delivery are primarily based on dsRNA ingestion via the oral route by the insect pests (Abrieux and Chiu, 2016; Joga et al., 2016). Up to now, it has been illustrated in several insect species, including *D. melanogaster*, *T. castaneum*, and *M. sexta,* that oral administration of dsRNA can induce a successful systemic RNAi response (Whyard et al., 2009). However, in many other insects, oral delivery of dsRNA provokes little to no RNAi response (Vélez and Fishilevich, 2018). It is believed that the activity of specific nucleases in the gut lumen is, to a certain extent, responsible for the failure of inducing an RNAi response through oral dsRNA administration in these insect species. A large body of literature exists evidencing dsRNase activity in extracellular fluids of insects, including the haemolymph, saliva, and the gut juice (Garbutt et al., 2013; Gu and Knipple 2013; Wynant et al., 2014b; Christiaens and Smagghe, 2014; Lomate and Bonning, 2016; Spit et al., 2017; Guan et al., 2018; Vélez and Fishilevich, 2018). It was recently proven in our lab that *S. gregaria* is also insensitive to the oral delivery of dsRNA, in contrast to the very potent systemic RNAi response upon dsRNA injection in its body cavity (Wynant et al., 2012, 2014b). Interestingly, Wynant and colleagues could identify four different *dsRNase* sequences (*Sg-dsRNase1-4*) in our in-house whole-body *S. gregaria* transcriptome. Moreover, these *Sg-dsRNases* are primarily expressed in the midgut and thus are expected to strongly contribute to the nuclease activity exhibited by extracted midgut juice (Wynant et al., 2014b). Remarkably, in a later study performed in our lab, it could be demonstrated that simultaneously knocking down all four endonucleases did not enhance the RNAi efficiency upon oral dsRNA delivery in *S. gregaria* (Spit et al., 2017).

The authors suggested several possible explanations for why dsRNA delivery in the desert locust still failed after silencing the four identified dsRNases. It was, for example, possible that residual nuclease activity, or nuclease activity originating from yet unidentified nucleases, could still degrade the dsRNA molecules inside the midgut lumen. In this context, the *S. gregaria* midgut reference transcriptome again profiles itself as a very useful database. Exploring the midgut reference transcriptome revealed numerous other nuclease encoding transcripts, in addition to *Sg-dsRNase1-4* (Ch3|3.4.2.2.3). Some of these nucleases might possibly contribute to the dietary breakdown of plant DNA/RNA (Katoch and Thakur, 2012). Further research is however needed to determine if these nucleases also contribute to dsRNA degradation in the midgut lumen of *S. gregaria*. Another possible explanation for the failure of oral dsRNA delivery in *S. gregaria,* suggested by Spit *et al.* (2017), might be the absence of specific dsRNA uptake mechanisms at the luminal side of the midgut epithelium. Several studies have indicated the role of scavenger receptor-mediated endocytosis in the cellular uptake of dsRNA in insects (Saleh et al., 2006; Ulvila et al., 2006; Wynant et al., 2014a; Xiao

et al., 2015; Ivashuta et al., 2015; Li et al., 2015; Cappelle et al., 2016; Yoon et al., 2016). In line with this, it was demonstrated in *S. gregaria* that inhibiting scavenger receptor function indeed resulted in a decreased sensitivity towards injected dsRNA (Wynant et al., 2014a). In *D. melanogaster* S2 cells, the cellular uptake of dsRNA was demonstrated to be mediated by class C scavenger receptors (SR-C) (Ulvila et al., 2006). However, despite the presence of one SR-C encoding transcript in the whole-body transcriptome of *S. gregaria*, this transcript is not present in its midgut reference transcriptome, suggesting that this receptor is not, or only marginally, expressed in the midgut of the desert locust. This could possibly explain why *S. gregaria* does not show a significant RNAi response to orally ingested dsRNA molecules.

In conclusion, further research to efficiently deliver dsRNA to insect pests continues and important breakthroughs are to be expected in the near future (Vogel et al., 2019). Then, potentially lethal genes identified in this doctoral research can be consulted for the research and development of specific dsRNA (or other insecticidal) molecules.

7.5. More future prospects

Overall, the *S. gregaria* midgut reference transcriptome has proven to be a useful database for midgut-associated studies in the desert locust. The data presented in this thesis have largely contributed to the expansion of our general knowledge of the *S. gregaria* digestive system. Moreover, the acquired information can now be used to further improve our knowledge of the digestive process in insects in general. Throughout this thesis, several interesting research questions were left open for the future. For example: what are the functions of the different identified serine proteases in the midgut, which transporters are mediating the uptake of amino acids in the midgut, is the trehalose transporter Tret1 involved in the dietary disaccharide absorption through the midgut, is JH involved in the regulation of the digestive protease expression, *etc*.? Furthermore, the midgut reference transcriptome can be used to explore the gene enrichment in the midgut of *S. gregaria,* for example to identify different types of detoxification molecules, receptors, *etc*.

The differential expression analysis of chapter 4 has highlighted several key transcripts mediating digestion. These are considered to be interesting starting points for future studies, as was demonstrated in chapter 5. These small-scale pilot experiments clearly indicated the lethal effects of targeting important transcripts in the midgut. In the future, high-throughput RNAi screens of the in this RNA-Seq identified differential transcripts can be executed to rapidly validate the insecticidal potential of many transcripts at once. Interesting lethal targets identified during these screens can then be further evaluated for the development of locust-specific or even wider-spectrum insecticides.

In this doctoral research, we have only investigated the annotated transcripts in the midgut transcriptome. However, many transcripts were left without an annotated record. Even within the list of differential transcripts after feeding, the function of many transcripts is still unknown. Remarkably, also other insect gut transcriptomic studies resulted in high numbers of un-annotated transcripts (Ch3|3.4.1.5). Surprisingly, up to now, none of these transcripts have ever been further investigated, and most studies were largely focused on those transcripts with annotated records. However, transcripts with unknown function might prove to be very interesting, since they could be insect- or species-specific. One simple approach that would allow us to obtain more information about these transcripts would be to investigate putative sequence similarities between un-annotated transcripts of different insect species. Such analyses might lead to the identification of conserved transcripts with as yet unknown

functions. One suggested strategy would be to create a database containing the sequences of all un-annotated transcripts identified in other insect gut transcriptomes. This database could then be used to BLAST the un-annotated *S. gregaria* midgut transcripts in order to find similar sequences. A hit of a similar sequence in other insect species might indicate a common origin and function. Finally, *in vivo* studies can be performed to analyze the function of these, putatively conserved, un-annotated sequences. This way, blind spots in insect gut research can be addressed in a straight-forward approach.

This RNA-Seq study focused on the midgut transcriptome of *S. gregaria*. However, other tissues, such as the brain and the foregut, are also known to influence feeding and digestion. Since several other tissues were dissected for this RNA-Seq experiment (Ch2|2.3.2.1), it would be interesting to also investigate the transcript profiles of these tissues. This way, possible functional interactions between tissues could be identified and an overarching regulatory model of the digestive process in *S. gregaria* could be constructed. Moreover, this would allow us to construct reference transcriptomes of several *S. gregaria* tissues in a similar way as performed for the midgut (Chapter 3). These databases can then be used to study transcript enrichment in these tissues.

Finally, RNA-Seq of the midgut of the desert locust during digestion has helped us improving our knowledge of how the digestive process in insects is regulated. Nevertheless, not all changes in the midgut during digestion are expected to occur at the transcript level. The midgut might, for example, also respond to the presence of food by stimulating protein/peptide release rather than synthesis, or by activating previously inactive proteins/peptides. This might be further investigated by performing proteomic (or peptidomic) based studies of the *S. gregaria* midgut by mass spectrometry. Other molecules that play a major regulatory role in key biological processes, but were not evaluated during this RNA-Seq study, are small ncRNAs. Over the past few years, many micro RNAs (miRNAs), small ncRNAs of about 22 nucleotides, have been demonstrated to mediate gene expression during various aspects of insect development (Asgari, 2013; Wu et al., 2017; He et al., 2017; Ylla et al., 2017). Moreover, in mosquitos, several miRNAs have been demonstrated to mediate gene expression in the midgut during a blood meal (Feng et al., 2018). It is expected that miRNAs also play a role in the gene expression during the digestive process in other insects. For this PhD research however, we solely focused on the mRNA profile of the midgut. Therefore, it would be interesting to also investigate miRNA expression in the midgut, or other relevant tissues, during digestion. That way, possible regulatory miRNAs might be identified and possibly be linked to the differential gene expression identified in the present study.

Ultimately, by implementing all the different studies and techniques suggested throughout this paragraph, we will be able to construct a more holistic view of how the digestive process in the desert locust is regulated at different levels.

7.6. References

Abrieux A, Chiu JC (2016) Oral delivery of dsRNA by microbes: Beyond pest control. Commun Integr Biol 9:e1236163. doi: 10.1080/19420889.2016.1236163

Asgari S (2013) MicroRNA functions in insects. Insect Biochem Mol Biol 43:388–397. doi: 10.1016/j.ibmb.2012.10.005

Baum JA, Bogaert T, Clinton W, Heck GR, Feldmann P, Ilagan O, Johnson S, Plaetinck G, Munyikwa T, Pleau M, Vaughn T, Roberts J (2007) Control of coleopteran insect pests through RNA interference. Nat Biotechnol 25:1322–1326. doi: 10.1038/nbt1359

Boudko DY (2012) Molecular basis of essential amino acid transport from studies of insect nutrient amino acid transporters of the SLC6 family (NAT-SLC6). J Insect Physiol 58:433–449. doi: 10.1016/j.jinsphys.2011.12.018

Christiaens O, Smagghe G (2014) The challenge of RNAi-mediated control of hemipterans. Curr Opin Insect Sci 6:15–21. doi: 10.1016/j.cois.2014.09.012

Deans C, Sword G, Behmer S (2016) Nutrition as a neglected factor in insect herbivore susceptibility to Bt toxins. Curr Opin Insect Sci 15:97–103. doi: 10.1016/j.cois.2016.04.005

Feng X, Zhou S, Wang J, Hu W (2018) microRNA profiles and functions in mosquitoes. PLoS Negl Trop Dis 12:e0006463. doi: 10.1371/journal.pntd.0006463

Garbutt JS, Bellés X, Richards EH, Reynolds SE (2013) Persistence of double-stranded RNA in insect hemolymph as a potential determiner of RNA interference success: Evidence from *Manduca sexta* and *Blattella germanica*. J Insect Physiol 59:171–178. doi: 10.1016/j.jinsphys.2012.05.013

Gilbert LI, A. Granger N, Roe RM (2000) The juvenile hormones: historical facts and speculations on future research directions. Insect Biochem Mol Biol 30:617–644. doi: 10.1016/S0965-1748(00)00034-5

Gu L, Knipple DC (2013) Recent advances in RNA interference research in insects: Implications for future insect pest management strategies. Crop Prot 45:36–40. doi: 10.1016/j.cropro.2012.10.004

Guan RB, Li HC, Fan YJ, Hu SR, Christiaens O, Smagghe G, Miao XX (2018) A nuclease specific to lepidopteran insects suppresses RNAi. J Biol Chem 293:6011–6021. doi: 10.1074/jbc.RA117.001553

Hawkins NJ, Bass C, Dixon A, Neve P (2019) The evolutionary origins of pesticide resistance. Biol Rev 94:135–155. doi: 10.1111/brv.12440

He K, Sun Y, Xiao H, Chang G, Fei L, Zhaojun H (2017) Multiple miRNAs jointly regulate the biosynthesis of ecdysteroid in the holometabolous insects, *Chilo suppressalis*. RNA 23:1817–1833. doi: 10.1261/rna.061408.117

He Q, Wen D, Jia Q, Cui C, Wang J, Palli SR, Li S (2014) Heat Shock Protein 83 (Hsp83) Facilitates Methoprene-tolerant (Met) Nuclear Import to Modulate Juvenile Hormone Signaling. J Biol Chem 289:27874–27885. doi: 10.1074/jbc.M114.582825

Holtof M, Lenaerts C, Cullen D, Vanden Broeck J (2019) Extracellular nutrient digestion and absorption in the insect gut. Cell Tissue Res. doi: 10.1007/s00441-019-03031-9

Joga MR, Zotti MJ, Smagghe G, Christiaens O (2016) RNAi Efficiency, Systemic Properties, and Novel Delivery Methods for Pest Insect Control: What We Know So Far. Front Physiol 7:1–14. doi: 10.3389/fphys.2016.00553

Katoch R, Thakur N (2012) Insect gut nucleases: a challenge for RNA interference mediated insect control strategies. Int J Biochem Biotechnol 1:198–203

Kikawada T, Saito A, Kanamori Y, Nakahara Y, Iwata K, Tanaka D, Watanabe M, Okuda T (2007) Trehalose transporter 1, a facilitated and high-capacity trehalose transporter, allows exogenous trehalose uptake into cells. Proc Natl Acad Sci 104:11585–11590. doi: 10.1073/pnas.0702538104

Le Gall M, Behmer ST (2014) Effects of Protein and Carbohydrate on an Insect Herbivore: The Vista from a Fitness Landscape. Integr Comp Biol 54:942–954. doi: 10.1093/icb/icu102

Lehane MJ (1997) Peritrophic matrix structure and function. Annu Rev Entomol 42:525–550. doi: 10.1146/annurev.ento.42.1.525

Lomate PR, Bonning BC (2016) Distinct properties of proteases and nucleases in the gut, salivary gland and saliva of southern green stink bug, *Nezara viridula*. Sci Rep 6:27587. doi: 10.1038/srep27587

Mainguet AM, Louveaux A, Sayed G, Rollin P (2000) Ability of a generalist insect, *Schistocerca gregaria*, to overcome thioglucoside defense in desert plants: tolerance or adaptation? Entomol Exp Appl 94:309–317. doi: 10.1046/j.1570-7458.2000.00632.x

Majerowicz D, Gondim KC (2013) Insect Lipid Metabolism: Insights into Gene Expression Regulation. In: Recent Trends in Gene Expression. Nova Science Publishers, Inc., pp 147–190

Mello MO, Silva-Filho MC (2002) Plant-insect interactions: an evolutionary arms race between two distinct defense mechanisms. Brazilian J Plant Physiol 14:71–81. doi: 10.1590/S1677-

04202002000200001
Nässel DR, Homberg U (2006) Neuropeptides in interneurons of the insect brain. Cell Tissue Res 326:1–24. doi: 10.1007/s00441-006-0210-8
Nguyen TTX, Tomberlin JK, Vanlaerhoven S (2015) Ability of Black Soldier Fly (Diptera: Stratiomyidae) Larvae to Recycle Food Waste. Environ Entomol 44:406–410. doi: 10.1093/ee/nvv002
Patankar AG, Giri AP, Harsulkar AM, Sainani MN, Deshpande VV, Ranjekar PK, Gupta VS (2001) Complexity in specificities and expression of *Helicoverpa armigera* gut proteinases explains polyphagous nature of the insect pest. Insect Biochem Mol Biol 31:453–464
Price DRG, Gatehouse JA (2008) RNAi-mediated crop protection against insects. Trends Biotechnol 26:393–400. doi: 10.1016/j.tibtech.2008.04.004
Raubenheimer D (2003) Nutrient balancing in grasshoppers: behavioural and physiological correlates of dietary breadth. J Exp Biol 206:1669–1681. doi: 10.1242/jeb.00336
Roman G, Meller V, Wu KH, Davis RL (1998) The opt1 gene of *Drosophila melanogaster* encodes a proton-dependent dipeptide transporter. Am J Physiol Physiol 275:C857–C869. doi: 10.1152/ajpcell.1998.275.3.C857
Scott JG, Michel K, Bartholomay LC, Siegfried BD, Hunter WB, Smagghe G, Zhu KY, Douglas AE (2013) Towards the elements of successful insect RNAi. J Insect Physiol 59:1212–1221. doi: 10.1016/j.jinsphys.2013.08.014
Spit J, Badisco L, Verlinden H, Van Wielendaele P, Zels S, Dillen S, Vanden Broeck J (2012) Peptidergic control of food intake and digestion in insects. Can J Zool 90:489–506. doi: 10.1139/z2012-014
Spit J, Philips A, Wynant N, Santos D, Plaetinck G, Vanden Broeck J (2017) Knockdown of nuclease activity in the gut enhances RNAi efficiency in the Colorado potato beetle, *Leptinotarsa decemlineata* , but not in the desert locust, *Schistocerca gregaria*. Insect Biochem Mol Biol 81:103–116. doi: 10.1016/j.ibmb.2017.01.004
Ulvila J, Parikka M, Kleino A, Sormunen R, Ezekowitz RA, Kocks C, Rämet M (2006) Double-stranded RNA is internalized by scavenger receptor-mediated endocytosis in *Drosophila* S2 cells. J Biol Chem 281:14370–14375. doi: 10.1074/jbc.M513868200
Vélez AM, Fishilevich E (2018) The mysteries of insect RNAi: A focus on dsRNA uptake and transport. Pestic Biochem Physiol 151:25–31. doi: 10.1016/j.pestbp.2018.08.005
Vogel E, Santos D, Mingels L, Verdonckt TW, Vanden Broeck J (2019) RNA Interference in Insects: Protecting Beneficials and Controlling Pests. Front Physiol 9:1912. doi: 10.3389/fphys.2018.01912
Whyard S, Singh AD, Wong S (2009) Ingested double-stranded RNAs can act as species-specific insecticides. Insect Biochem Mol Biol 39:824–832. doi: 10.1016/j.ibmb.2009.09.007
Wieczorek H, Beyenbach KW, Huss M, Vitavska O (2009) Vacuolar-type proton pumps in insect epithelia. J Exp Biol 212:1611–1619. doi: 10.1242/jeb.030007
Wu W, Xiong W, Li C, Zhai M, Li Y, Ma F, Li B (2017) MicroRNA-dependent regulation of metamorphosis and identification of microRNAs in the red flour beetle, *Tribolium castaneum*. Genomics 109:362–373. doi: 10.1016/j.ygeno.2017.06.001
Wynant N, Santos D, Van Wielendaele P, Vanden Broeck J (2014a) Scavenger receptor-mediated endocytosis facilitates RNA interference in the desert locust, *Schistocerca gregaria*. Insect Mol Biol 23:n/a-n/a. doi: 10.1111/imb.12083
Wynant N, Santos D, Verdonck R, Spit J, Van Wielendaele P, Vanden Broeck J (2014b) Identification, functional characterization and phylogenetic analysis of double stranded RNA degrading enzymes present in the gut of the desert locust, *Schistocerca gregaria*. Insect Biochem Mol Biol 46:1–8. doi: 10.1016/j.ibmb.2013.12.008
Wynant N, Verlinden H, Breugelmans B, Simonet G, Vanden Broeck J (2012) Tissue-dependence and sensitivity of the systemic RNA interference response in the desert locust, *Schistocerca gregaria*. Insect Biochem Mol Biol 42:911–917. doi: 10.1016/j.ibmb.2012.09.004
Ylla G, Piulachs M-D, Belles X (2017) Comparative analysis of miRNA expression during the development of insects of different metamorphosis modes and germ-band types. BMC Genomics 18:774. doi: 10.1186/s12864-017-4177-5

Appendix

Table of contents

A.1. Supplementary data of chapter 2

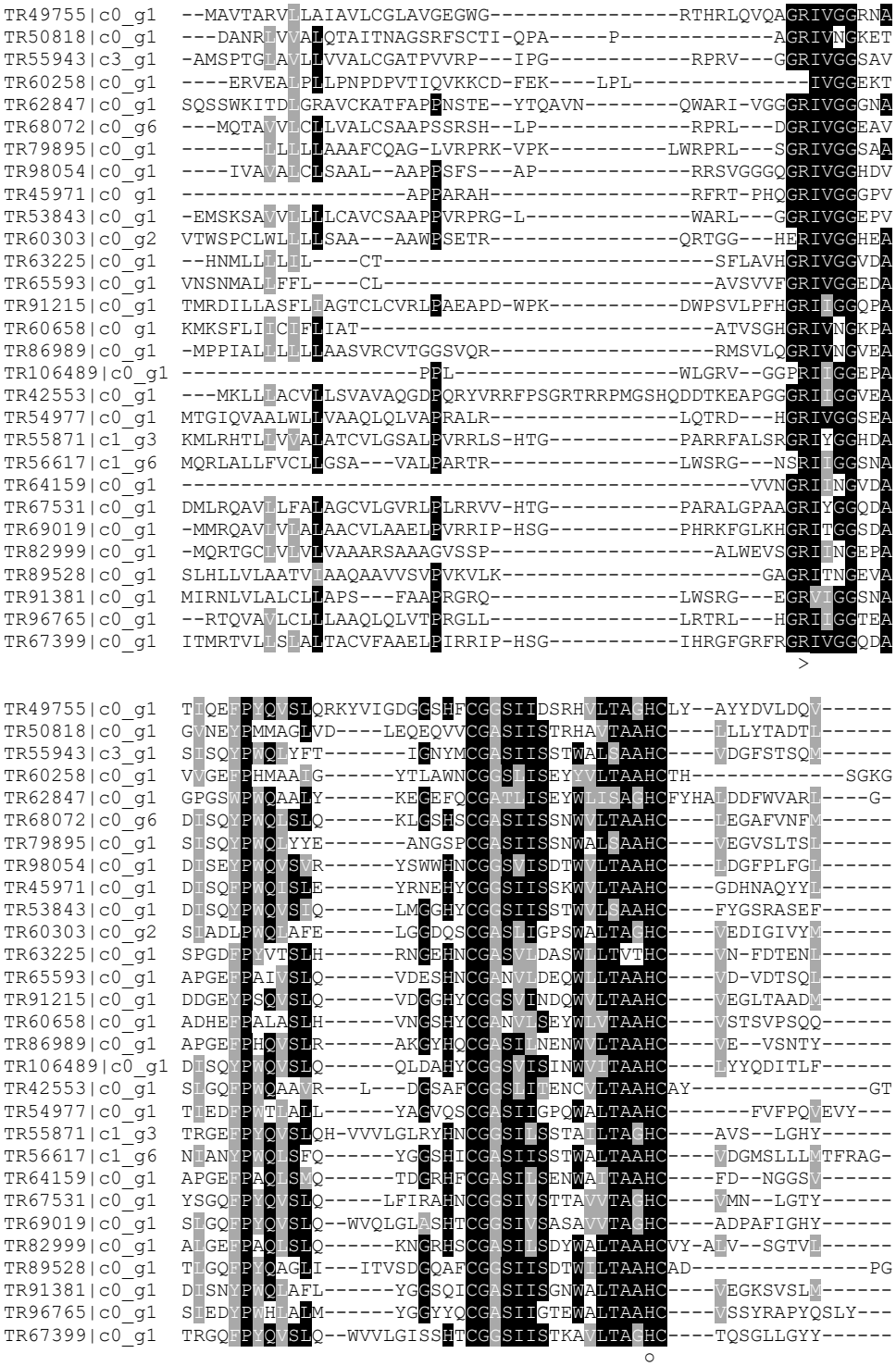
Supplementary table S2.1. RNA concentrations of the 18 samples for RNA-Seq after total RNA extraction using the RNeasy Lipid Tissue extraction kit (Qiagen). Measurements were performed on the NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Scientific™) with nuclease-free water as a blank.

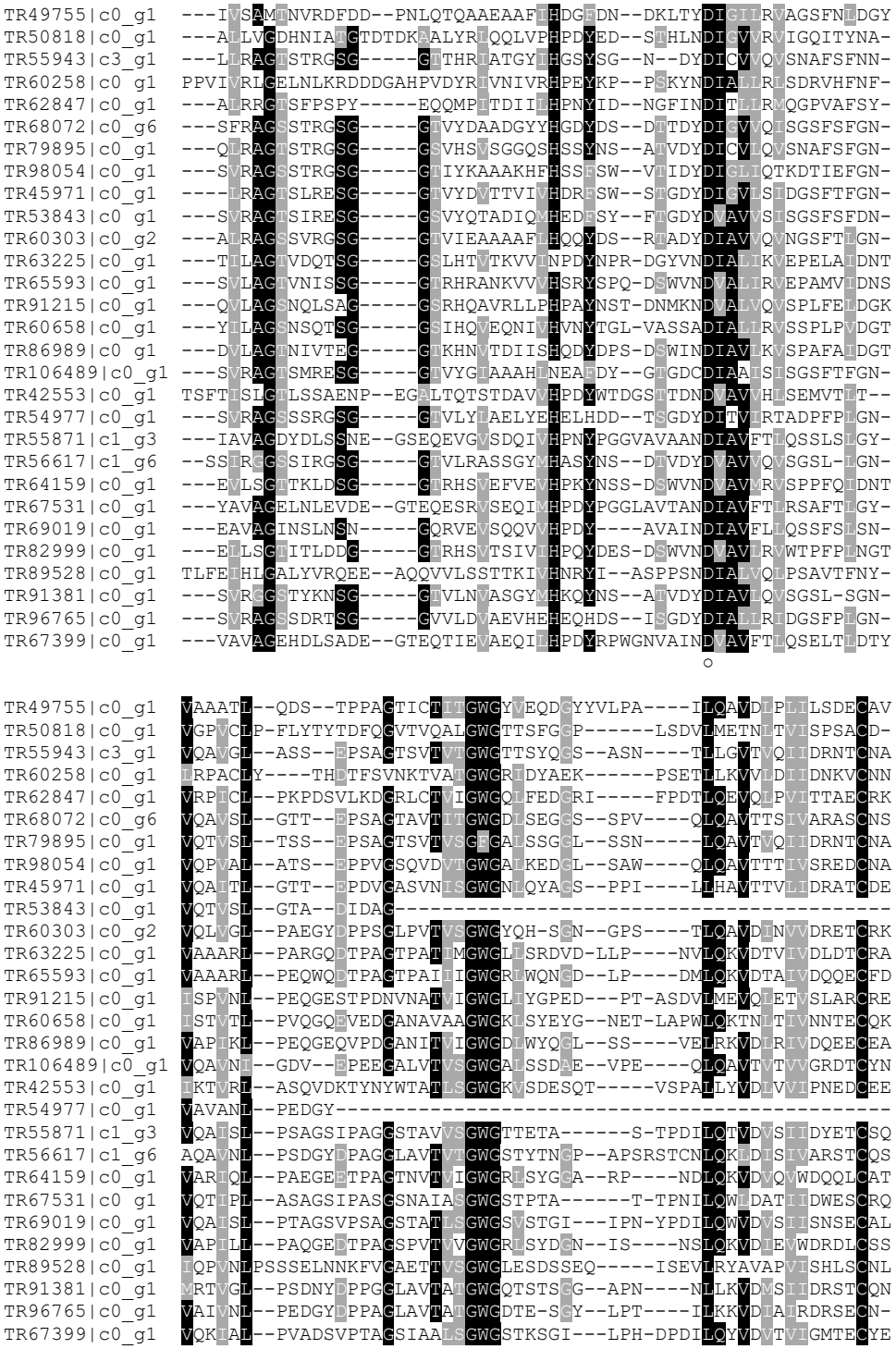
Group	Sample	A260/280	A260/230	Concentration (ng/µL)
Group A (10 minutes)	S1	2.16	1.39	791.47
	S2	2.15	1.70	1098.97
	S3	2.15	2.18	837.12
	S4	2.15	2.22	1152.33
	S5	2.10	1.45	513.62
	S6	2.15	2.34	1212.75
Group B (2 hours)	S7	2.16	1.87	1158.47
	S8	2.15	2.28	830.87
	S9	2.16	1.75	886.88
	S10	2.14	2.20	1033.77
	S11	2.18	2.15	959.83
	S12	2.15	2.13	1057.79
Group C (24 hours)	S13	2.13	1.83	452.44
	S14	2.09	1.81	587.28
	S15	2.14	2.03	229.94
	S16	2.07	1.80	231.56
	S17	2.11	1.71	518.89
	S18	2.11	1.85	503.03

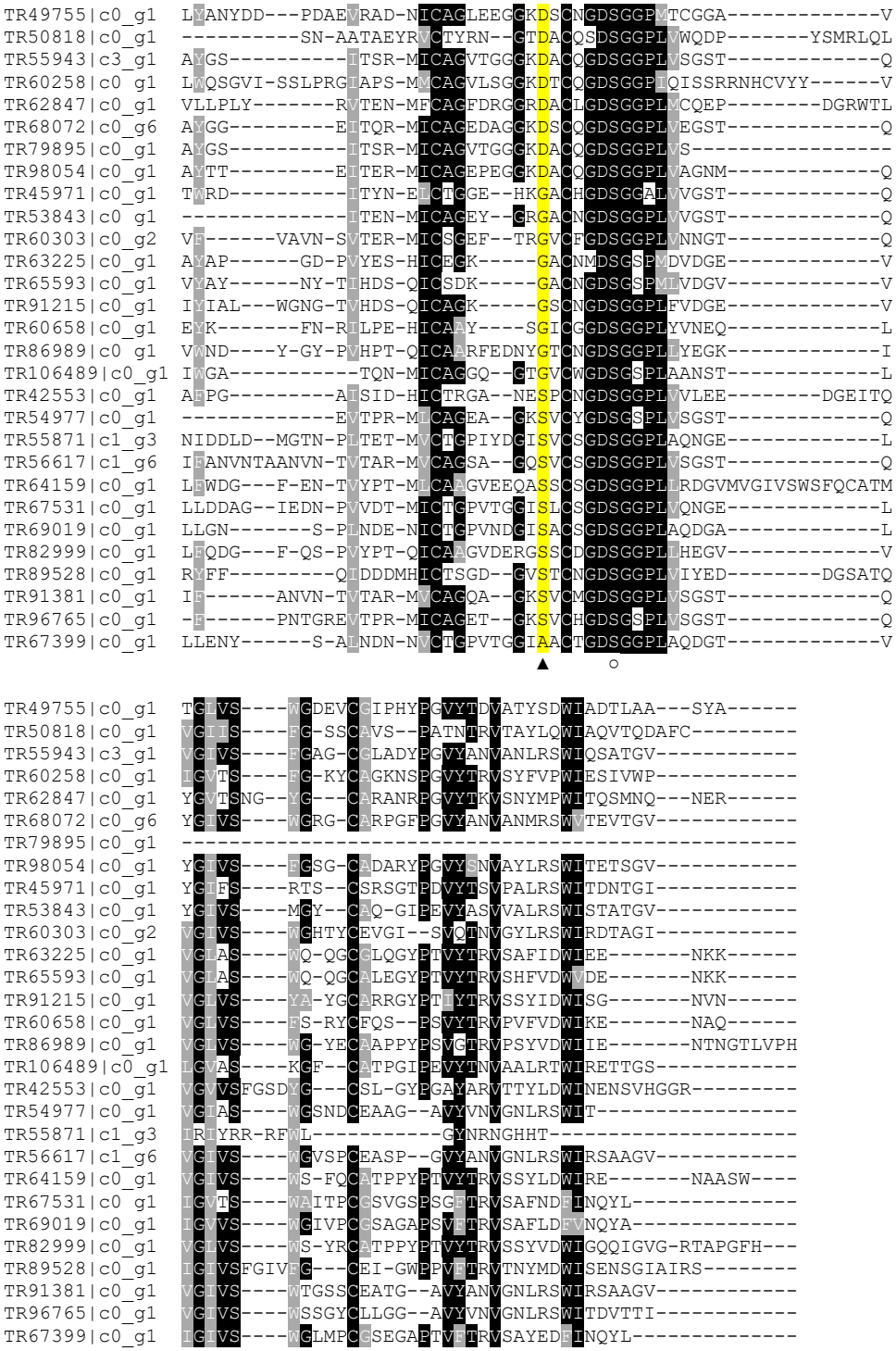
* RNA absorbs light of the wavelength 260 nm, proteins of the wavelength 280 nm, and other contaminants of the wavelength 230.
* A260/280 = Ratio of absorbance at 260 nm over ratio of absorbance at 280 nm; a ratio of ~2.0 is generally accepted as “pure” for RNA.
* A260/230 = Ratio of absorbance at 260 nm over ratio of absorbance at 230 nm; expected values are commonly in the range of 2.0-2.2.

A.2. Supplementary data of chapter 3

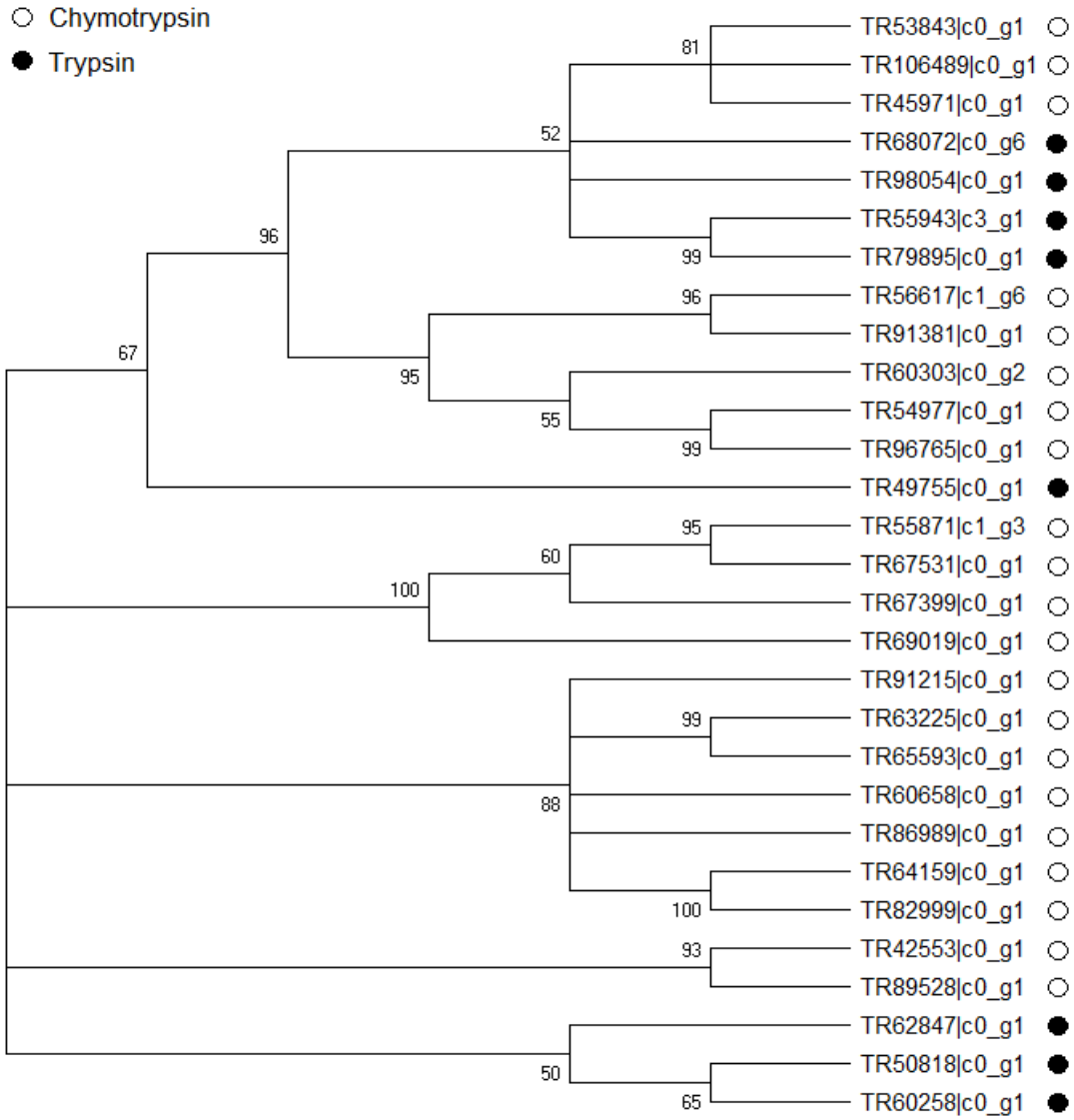
Supplementary figure S3.1. MAFFT multiple sequence alignment of the trimmed amino acid sequences of the predicted enzymatically active trypsin and chymotrypsin sequences in the *S. gregaria* midgut reference transcriptome. Conservation threshold for shading is 60% identity. The standard activation cleavage site of insect serine proteases (RIVGG) is denoted with an arrow (>) below. Amino acids of the catalytic triad are denoted with a circle (○) below. Amino acid 189 determining the substrate specificity of serine proteases is colored yellow and denoted with a triangle (▲) below.




TR49755|c0_g1
TR50818|c0_g1
TR55943|c3_g1
TR60258|c0 g1
TR62847|c0_g1
TR68072|c0_g6
TR79895|c0_g1
TR98054|c0_g1
TR45971|c0_g1
TR53843|c0 g1
TR60303|c0_g2
TR63225|c0_g1
TR65593|c0_g1
TR91215|c0_g1
TR60658|c0_g1
TR86989|c0_g1
TR106489|c0_g1
TR42553|c0_g1
TR54977|c0_g1
TR55871|c1_g3
TR56617|c1_g6
TR64159|c0_g1
TR67531|c0_g1
TR69019|c0_g1
TR82999|c0_g1
TR89528|c0_g1
TR91381|c0_g1
TR96765|c0_g1
TR67399|c0_g1


TR49755|c0_g1
TR50818|c0_g1
TR55943|c3_g1
TR60258|c0 g1
TR62847|c0_g1
TR68072|c0_g6
TR79895|c0_g1
TR98054|c0_g1
TR45971|c0_g1
TR53843|c0_g1
TR60303|c0_g2
TR63225|c0_g1
TR65593|c0_g1
TR91215|c0_g1
TR60658|c0_g1
TR86989|c0_g1
TR106489|c0_g1
TR42553|c0_g1
TR54977|c0_g1
TR55871|c1_g3
TR56617|c1_g6
TR64159|c0_g1
TR67531|c0 g1
TR69019|c0_g1
TR82999|c0_g1
TR89528|c0_g1
TR91381|c0_g1
TR96765|c0_g1
TR67399|c0_g1

Supplementary figure S3.2. Maximum likelihood phylogenetic tree of the trimmed amino acid sequences of the predicted enzymatically active trypsin and chymotrypsin sequences in the *S. gregaria* midgut reference transcriptome. The phylogenetic tree is based on 500 bootstrap replicates with only bootstrap values higher than 50 being displayed. Sequences with predicted chymotrypsin activity are highlighted with an open circle and sequences with predicted trypsin activity are highlighted with a black circle.



Supplementary table S3.1. Mean Log2-transformed TMM-normalized CPM of all predicted trypsin and chymotrypsin proteases in the *S. gregaria* midgut reference transcriptome at different timepoints after feeding.

Transcript	Group A (10 min)	Group B (2 hours)	Group C (24 hours)
TR100336|c0_g1	2.550687877	2.921300214	2.464755237
TR101143|c0_g2	2.938958809	2.946614998	2.189914877
TR102645|c0_g1	4.83715304	5.174708421	4.705522355
TR103876|c0_g1	4.689151063	4.486435945	5.009935659
TR103929|c0_g1	11.27835185	11.60612414	11.72248281
TR104095|c0_g1	0.861405281	0.997556589	0.851690922
TR105511|c0_g1	2.803381069	3.315750868	2.639834944
TR105789|c0_g1	5.596436991	5.684051653	4.887705825
TR10625|c0_g1	1.392318828	1.472586661	0.613525097
TR106489|c0_g1	2.915141603	3.152882528	1.995349612
TR30913|c0_g1	8.941727642	9.355176461	8.644960087
TR34413|c0_g3	6.19207374	5.79318627	6.495772959
TR35640|c0_g2	7.482338913	7.907034832	8.047056534
TR36299|c0_g1	9.4110089	9.541941841	9.088601312
TR37818|c0_g1	0.952945305	0.75527728	0.981597129
TR39518|c0_g1	2.958588631	2.481612536	2.520241537
TR39840|c0_g2	0.634107849	0.323967397	0.673232209
TR42472|c0_g1	10.61562824	10.95750095	11.11559591
TR42553|c0_g1	8.67842931	8.411506608	6.516423798
TR45971|c0_g1	6.246788581	6.589684701	6.388465249
TR48365|c3_g11	9.394524157	9.610386668	9.699227525
TR48428|c1_g2	0.148691347	0.78014764	0.31211548
TR49226|c0_g1	3.580181496	3.209398485	3.864781343
TR49691|c0_g1	5.660556804	5.847788294	5.40798503
TR49728|c0_g1	1.702282358	1.384100827	1.066234463
TR49755|c0_g1	11.82112136	11.94810311	11.41136632
TR50264|c0_g1	3.96932321	3.582368229	4.41735586
TR50818|c0_g1	4.528280546	4.455228532	4.892301532
TR51415|c0_g1	3.903087534	4.23723752	3.456049077
TR52259|c0_g1	0.716251057	0.44414691	0.962152054
TR52278|c1_g2	6.544957078	5.983008418	4.775664417
TR53765|c0_g1	5.471389909	5.632761118	4.871749173
TR53843|c0_g1	3.828017401	3.8941994	3.182310675
TR54977|c0_g1	3.885411139	3.896704863	3.221867578
TR55418|c0_g1	0.605393873	-0.000741635	-2.012564366
TR55871|c1_g2	3.086543299	3.553765933	2.319548611
TR55871|c1_g3	7.986074012	8.047852227	7.300567673
TR55871|c1_g4	7.346150464	7.719347531	7.056496718
TR55943|c1_g1	6.13283854	5.668686369	4.025879216
TR55943|c2_g1	11.04833539	11.25373362	10.66761476
TR55943|c2_g2	11.16945281	11.32051994	10.7533071
TR55943|c3_g1	8.126900474	7.965035302	6.45418429
TR55943|c3_g2	7.139217325	6.819451667	5.508667115
TR56246|c0_g1	2.131730736	1.969909932	2.131138154
TR56617|c1_g1	7.748252672	8.070116244	7.273089926

TR56617|c1_g2	9.148137798	9.295911577	8.71759675
TR56617|c1_g4	8.010461308	7.983026101	6.896365141
TR56617|c1_g5	6.334344437	7.321859358	6.703764235
TR56617|c1_g6	11.681374	11.94697869	11.57114712
TR56617|c1_g7	8.647985108	8.888277859	8.257773105
TR56950|c1_g1	4.205565471	4.104877064	4.090928881
TR60258|c0_g1	0.639928414	-0.005896836	1.362354117
TR60303|c0_g2	3.269953596	3.551558542	2.513315388
TR60658|c0_g1	2.995153469	3.131344931	1.897036092
TR62847|c0_g1	0.361726622	-0.147611511	0.798378519
TR63225|c0_g1	5.625997476	5.776518425	5.2169359
TR64159|c0_g1	10.96360442	11.17914396	10.82984735
TR64731|c0_g1	4.931448663	4.405959095	5.27438207
TR64781|c0_g1	4.82255123	4.518183789	2.914792067
TR64781|c0_g2	2.996160774	2.990875391	1.038883868
TR65593|c0_g1	6.482575363	6.75649295	6.07419657
TR67399|c0_g1	4.956690242	4.682732558	3.35390453
TR67531|c0_g1	7.158407371	6.873885011	6.167978375
TR679|c0_g1	4.331486238	4.578504486	4.025910786
TR68072|c0_g6	3.160577084	2.948071578	2.701415781
TR68669|c0_g1	2.850837756	2.102049798	1.704718779
TR68885|c0_g1	1.399876688	1.501278067	0.646865776
TR68929|c0_g1	6.359005051	6.529474776	5.228096751
TR68951|c0_g1	2.614708023	2.431001359	0.216693775
TR69019|c0_g1	9.665729931	9.254875358	8.067859548
TR69767|c0_g1	1.975258062	2.395700766	-0.154906946
TR69823|c0_g1	5.317861773	5.271670148	5.279489716
TR70075|c0_g1	1.878191677	1.784868469	-0.382471174
TR71811|c0_g1	3.993636557	4.322642115	3.550698533
TR72997|c1_g2	2.617067655	2.806789572	1.709740649
TR75124|c0_g1	2.883429015	2.531514618	3.347424623
TR79895|c0_g1	2.922071464	3.121203809	2.401492562
TR81358|c0_g1	2.017339781	1.198430412	1.311017457
TR81931|c0_g1	0.310610638	-0.280311865	0.151963733
TR82388|c3_g5	4.005593339	4.307007693	3.63310015
TR82403|c2_g3	7.66816856	7.719950277	6.614075533
TR82403|c2_g5	4.428589396	4.463802324	2.761213426
TR82999|c0_g1	5.571215216	5.609045232	5.488091369
TR84957|c0_g1	1.209994019	0.623602619	1.714958966
TR86892|c0_g1	7.168867274	7.416634721	7.158938345
TR86989|c0_g1	3.796726472	3.804029958	2.692945367
TR87238|c0_g1	3.756997295	3.792388118	3.655261117
TR88939|c0_g1	3.535084047	3.895777691	2.907852955
TR89110|c0_g1	4.449617912	4.139066359	4.592834553
TR89409|c0_g1	9.540688889	9.776241487	9.619611726
TR89528|c0_g1	8.6957353	8.549580263	6.896440943
TR89726|c0_g1	7.656785218	7.411309876	5.376903876
TR89872|c0_g1	2.440204186	2.591122825	1.840987552
TR91215|c0_g1	10.09341188	10.40207401	9.342307781
TR91381|c0_g1	11.05571095	11.43151713	10.64242081
TR91530|c0_g1	3.045314595	3.571126345	1.88486648

TR91582|c0_g1	10.57065734	10.64341855	10.73616367
TR93252|c1_g1	7.925374212	8.243642256	7.498069782
TR94178|c0_g1	7.088246158	7.104536505	5.701294254
TR96062|c0_g1	0.364506143	0.898244994	1.042907128
TR96062|c0_g2	3.46206179	3.474987693	2.652598946
TR96184|c0_g1	5.841876635	6.134449589	5.867326368
TR96765|c0_g1	4.044557793	4.331096707	3.033334635
TR98054|c0_g1	9.016013624	9.297574143	8.672112541
TR98295|c0_g1	3.564394278	3.593664456	3.517913805
TR99612|c0_g1	1.130541206	0.685726805	-0.146350265

A.3. Supplementary data of chapter 4

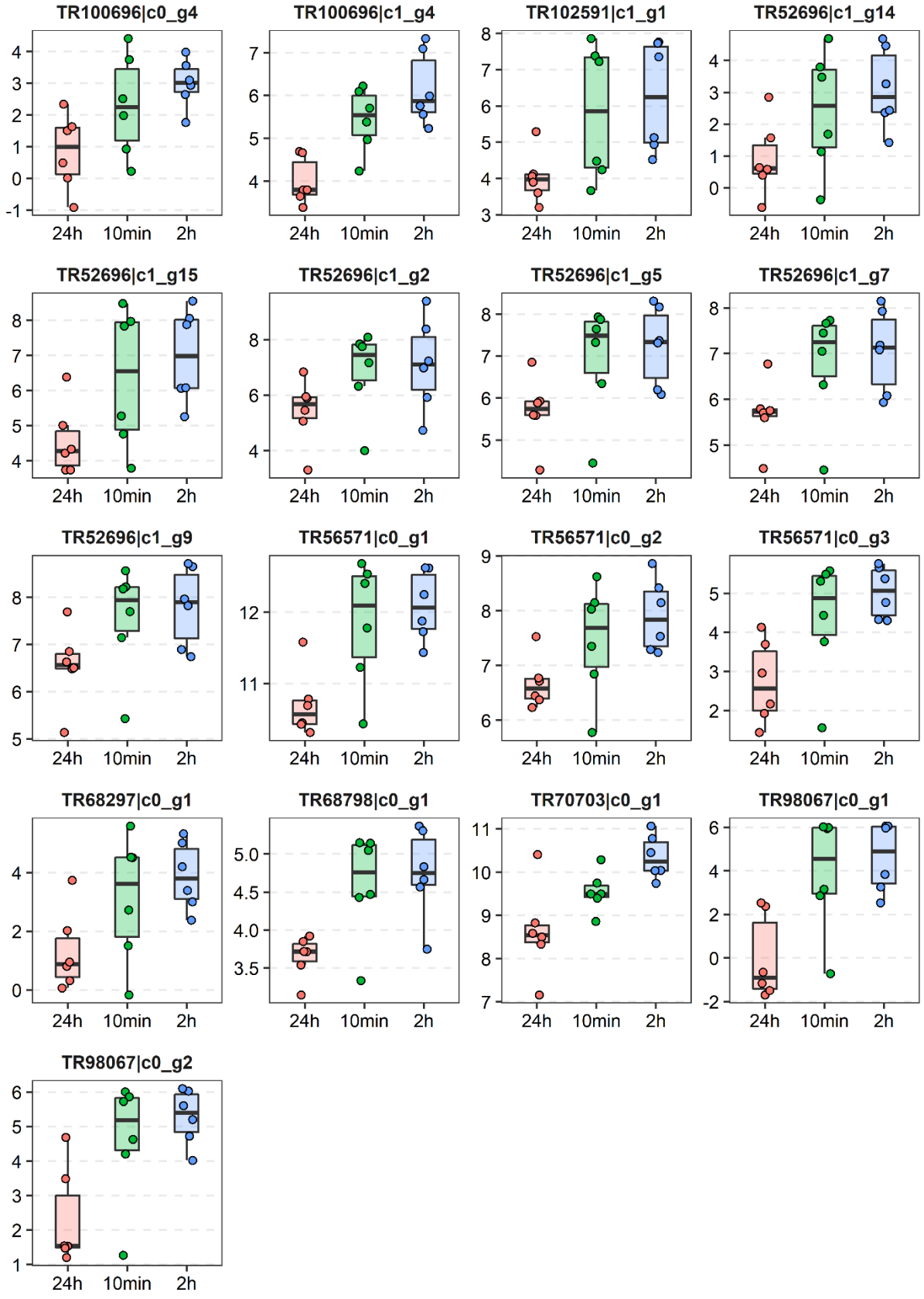
Supplementary table S4.1. Summary of all differentially upregulated transcripts predicted to encode carbohydrate degrading enzymes in the *S. gregaria* midgut two hours after feeding compared to twenty-four hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR52701|c4_g5	2.631604801	1.960616342	Myrosinase 1
TR86870|c1_g1	2.544971466	3.61323586	Myrosinase 1
TR34381|c1_g1	2.703264602	5.651835654	Myogenesis-regulating glycosidase
TR52701|c3_g1	2.499127706	1.618891728	Myrosinase 1
TR60226|c1_g1	2.285039443	5.747366231	Myogenesis-regulating glycosidase
TR52701|c0_g1	2.262198924	1.178479703	Lactase-phlorizin hydrolase
TR52701|c4_g1	2.151692636	0.730438498	Myrosinase 1
TR86870|c0_g3	2.045319172	4.345581619	Lactase-like protein
TR59608|c0_g1	1.979939028	3.118652684	Myogenesis-regulating glycosidase
TR91535|c3_g1	1.809272769	5.596551126	Myrosinase 1
TR33586|c0_g1	1.754027516	1.127114934	Myrosinase 1
TR92287|c0_g1	1.706058852	4.445532187	Beta-galactosidase-1-like protein 2
TR84919|c0_g1	1.699899396	3.181428734	Glucosylceramidase
TR59919|c1_g2	1.680604573	2.471001868	Myrosinase 1
TR59919|c1_g3	1.652571911	3.115864642	Myrosinase 1
TR42015|c0_g1	1.587147731	2.526726157	Myogenesis-regulating glycosidase
TR66519|c0_g1	1.520587863	4.96944381	Beta-mannosidase
TR42387|c1_g2	1.504507875	6.917720096	Lactase-like protein
TR64699|c0_g2	1.47208951	7.39222312	Endoglucanase B
TR99906|c1_g1	1.446616319	1.145407455	Myrosinase 1
TR46352|c0_g1	1.427118032	0.831217127	Myrosinase 1
TR89638|c0_g1	1.407982606	1.11316301	Lactase-phlorizin hydrolase
TR81787|c0_g1	1.330285709	7.471972114	Maltase 1
TR86870|c0_g2	1.224696993	2.579229014	Myrosinase 1
TR33522|c0_g2	1.183007591	6.162850754	Myogenesis-regulating glycosidase
TR50029|c0_g1	1.17392239	1.575575287	Myogenesis-regulating glycosidase
TR52701|c4_g3	1.086996069	3.130656179	Myrosinase 1
TR15214|c0_g1	1.073090487	3.092393632	Myogenesis-regulating glycosidase
TR59608|c0_g4	1.060506271	2.53003911	Myogenesis-regulating glycosidase

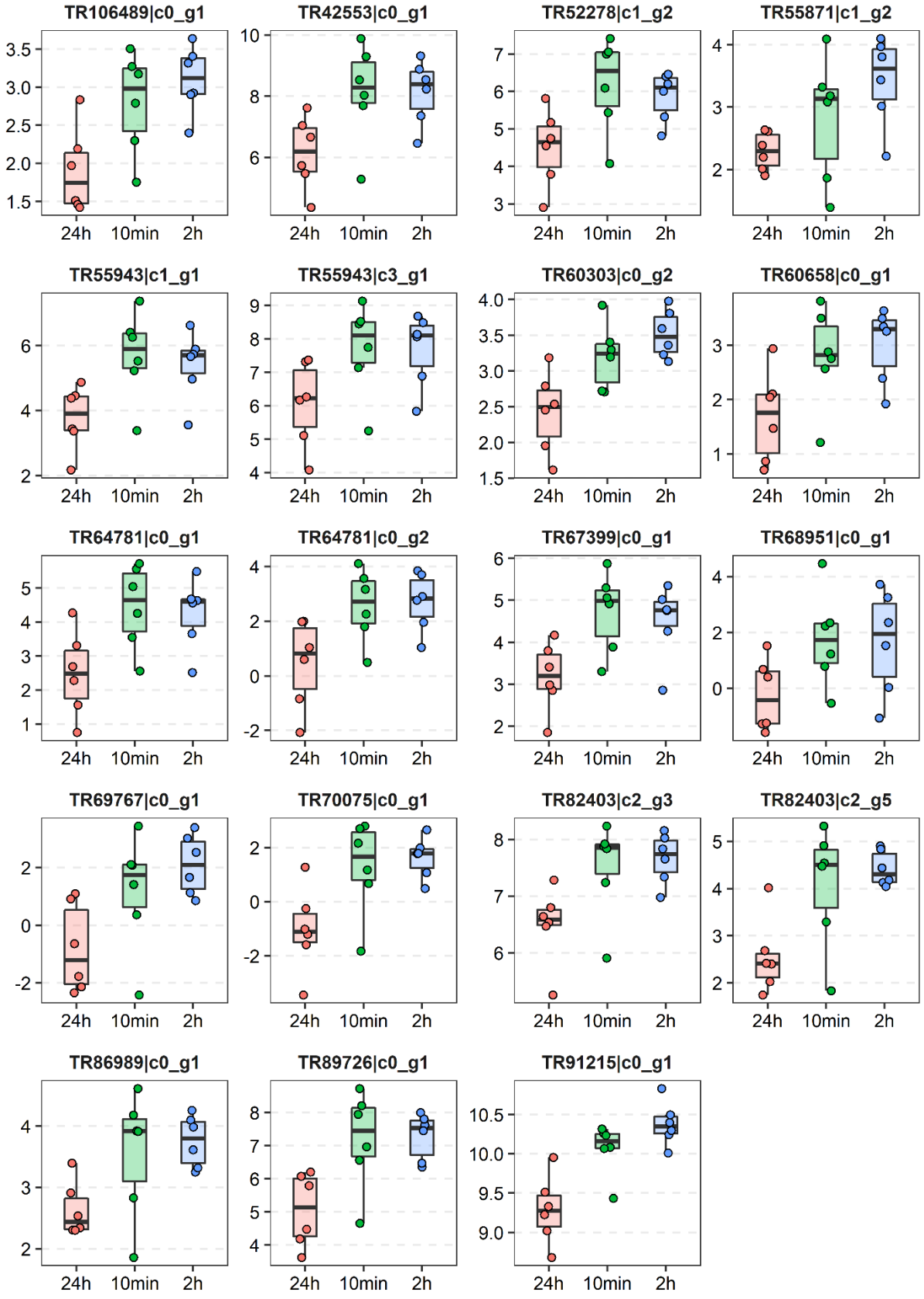
Supplementary table S4.2. Summary of all differentially upregulated transcripts predicted to encode lipid degrading enzymes in the *S. gregaria* midgut two hours after feeding compared to twenty-four hours after feeding.

TRANSCRIPT	LOGFC	LOGCPM	ANNOTATION
TR76143|c0_g1	1.651968865	5.912143791	Lipase member H
TR56546|c0_g2	1.633674409	3.059084846	Lipase member H
TR79714|c0_g1	1.602970019	6.491843564	Probable Delta(7)-sterol 5(6)-desaturase
TR65119|c0_g1	1.592395687	4.093320365	Lipase member H
TR55865|c1_g1	1.489263691	5.472905535	Lipase 3
TR7554|c0_g1	1.467067389	1.869674743	Cholesterol 7-desaturase
TR13589|c0_g1	1.219323591	5.932784041	Cholesterol 7-desaturase
TR55865|c0_g1	1.172528466	6.500432075	Lipase 3

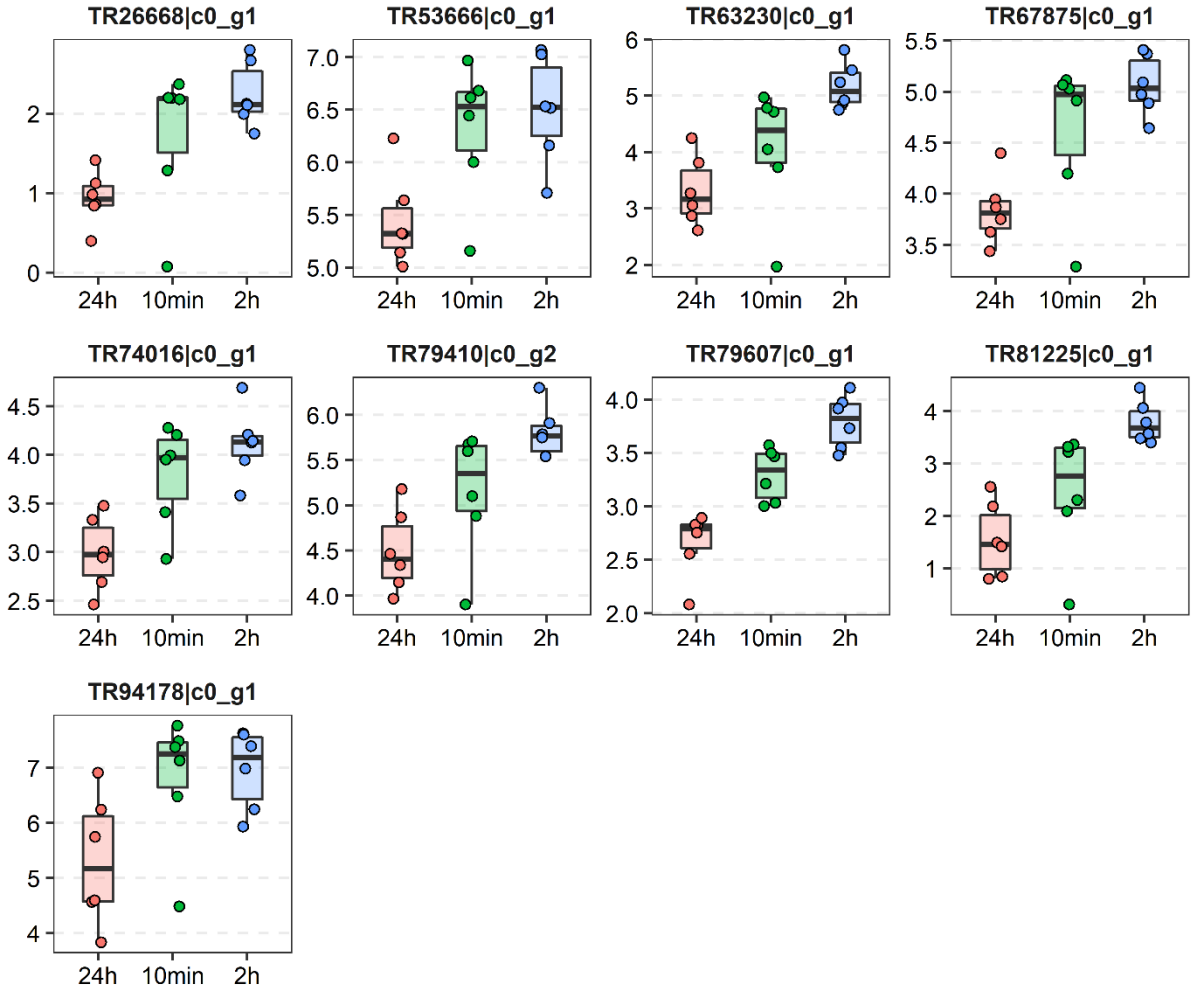
Supplementary figure S4.1. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding peritrophin-like proteins that were significantly upregulated two hours after feeding compared to twenty-four hours after feeding.



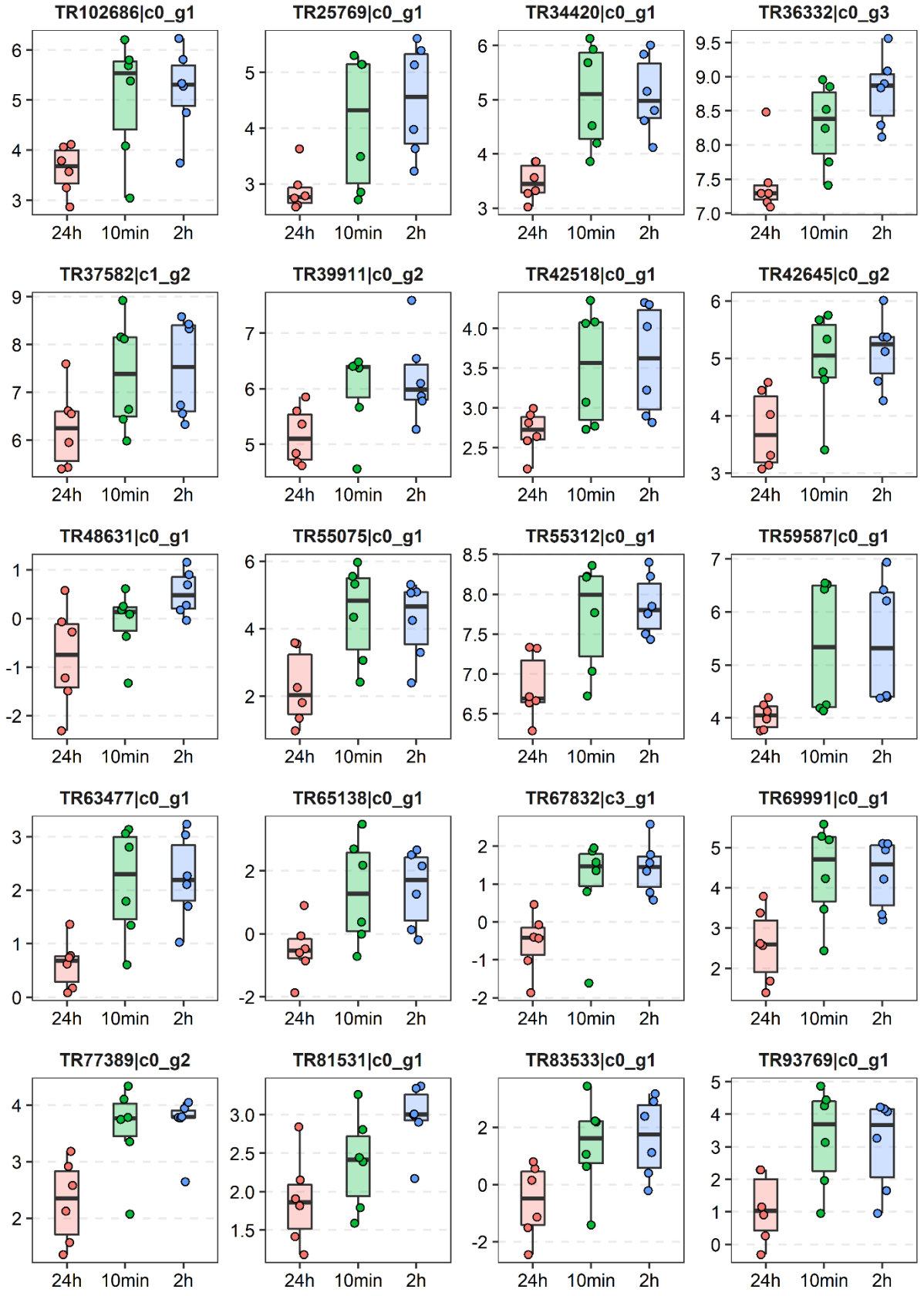
Supplementary figure S4.2. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding endopeptidases that were significantly upregulated two hours after feeding compared to twenty-four hours after feeding.



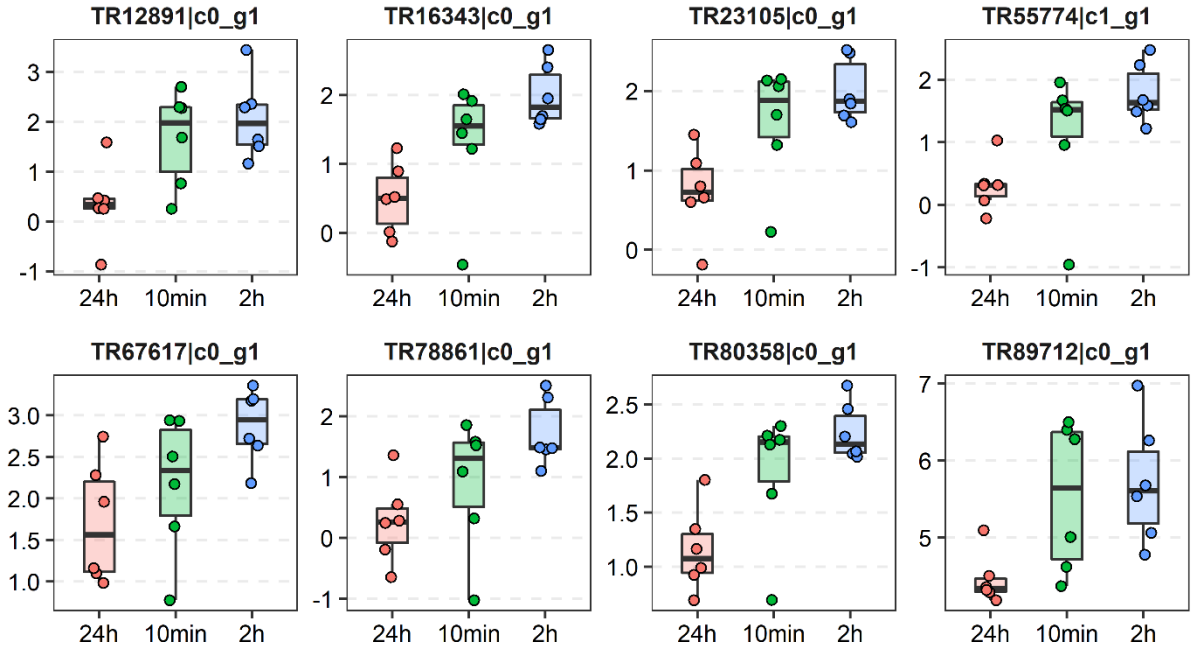
Supplementary figure S4.3. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding exopeptidases that were significantly upregulated two hours after feeding compared to twenty-four hours after feeding.



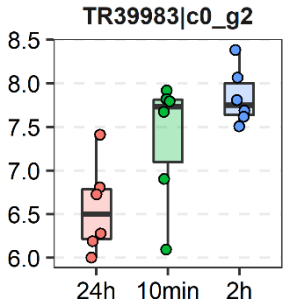
Supplementary figure S4.4. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding putative nutrient transcporter proteins that were significantly upregulated two hours after feeding compared to twenty-four hours after feeding.



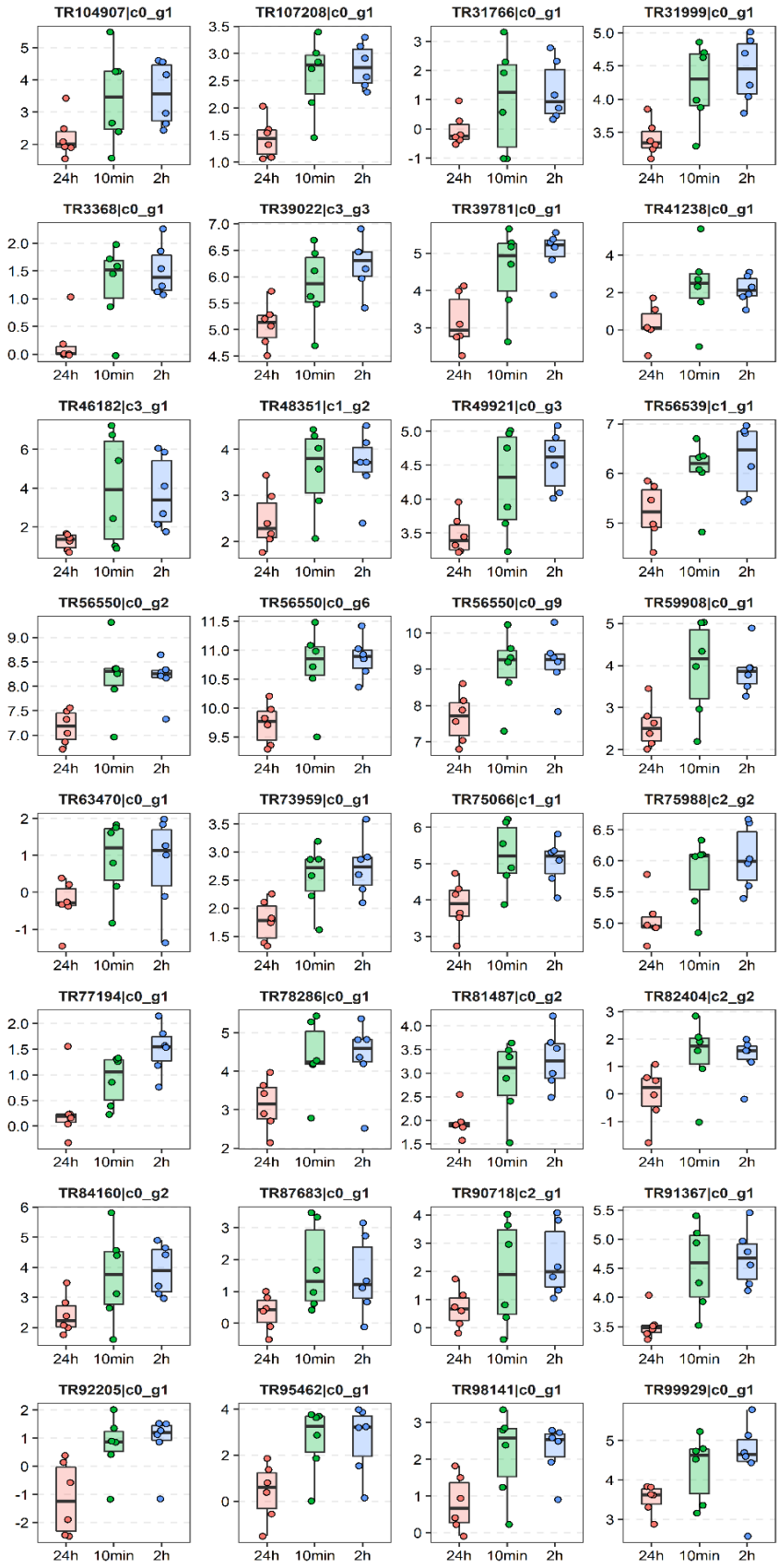
Supplementary figure S4.5. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding ABC transporter proteins that were significantly upregulated two hours after feeding compared to twenty-four hours after feeding.



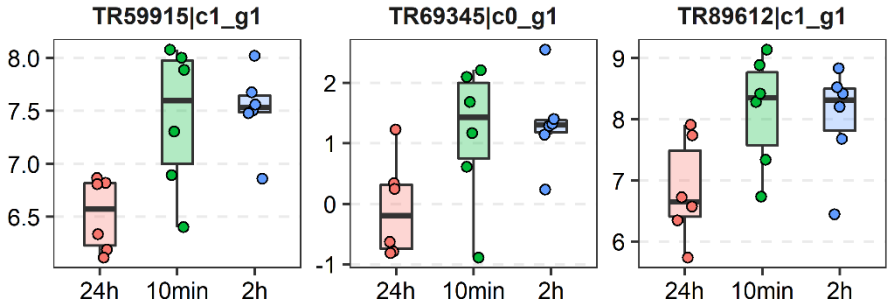
Supplementary figure S4.6. Boxplots of log2-transformed TMM-normalized CPM of a transcript encoding a H^+ V-ATPase subunit that was significantly upregulated two hours after feeding compared to twenty-four hours after feeding.



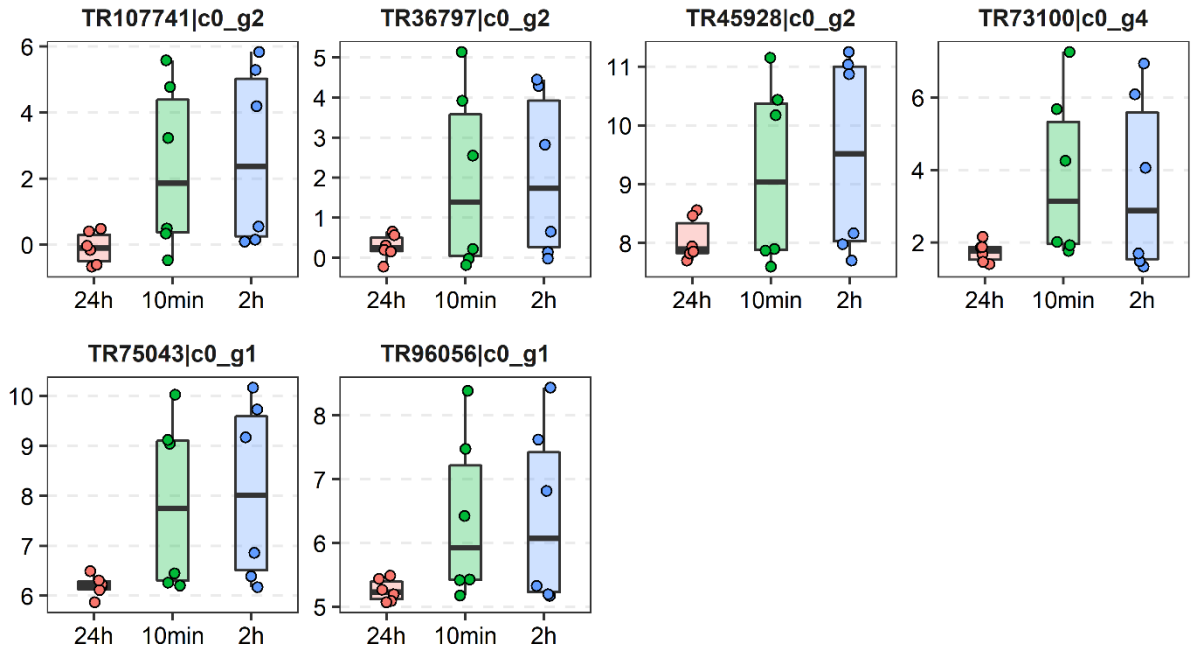
Supplementary figure S4.7. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding the most general types of detoxification proteins that were significantly upregulated two hours after feeding compared to twenty-four hours after feeding.



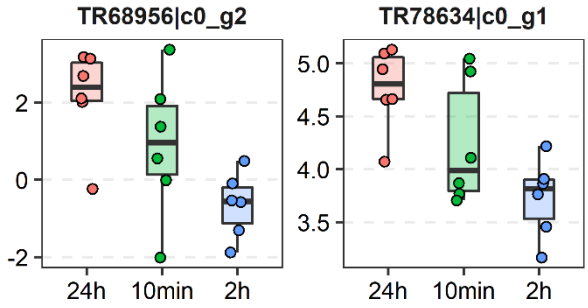
Supplementary figure S4.8. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding JHBPs that were significantly upregulated two hours after feeding compared to twenty-four hours after feeding.



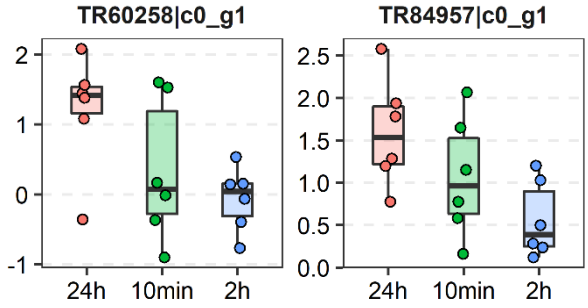
Supplementary figure S4.9. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding several heat shock proteins that were significantly upregulated two hours after feeding compared to twenty-four hours after feeding.



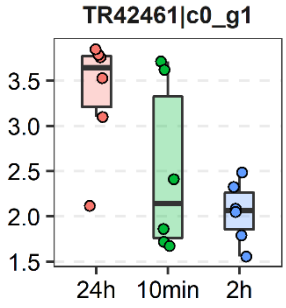
Supplementary figure S4.10. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding protease inhibitors that were significantly downregulated two hours after feeding compared to twenty-four hours after feeding.



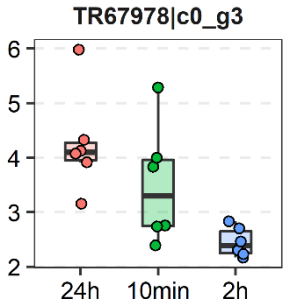
Supplementary figure S4.11. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding trypsin-like proteins that were significantly downregulated two hours after feeding compared to twenty-four hours after feeding.



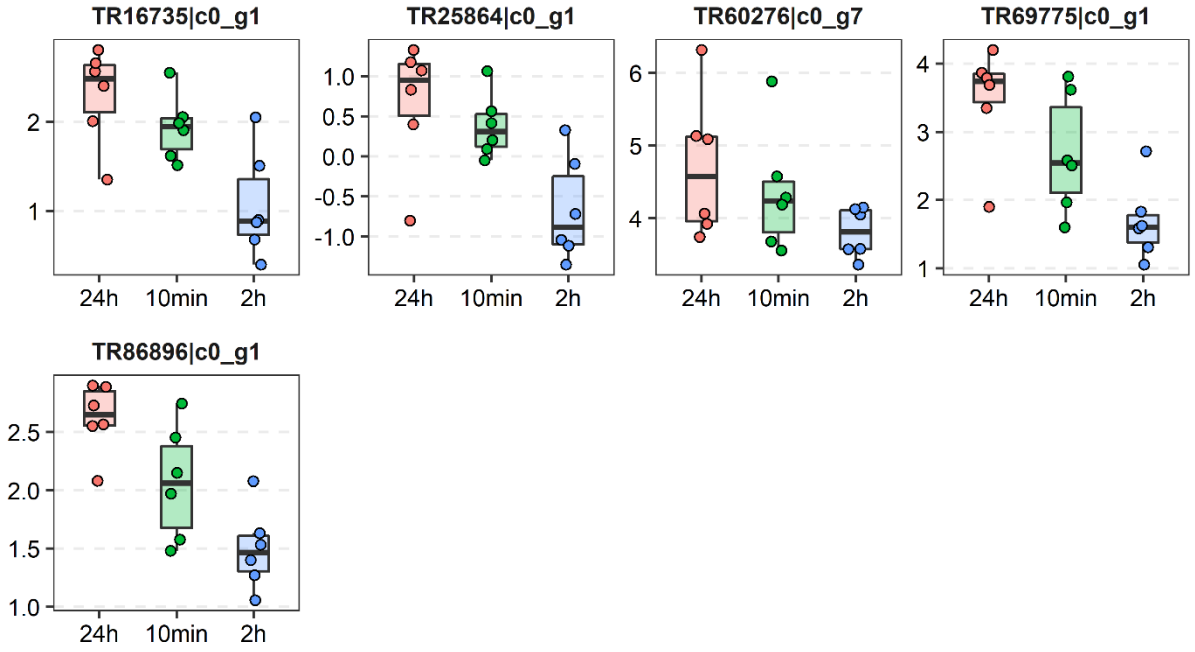
Supplementary figure S4.12. Boxplots of log2-transformed TMM-normalized CPM of a transcript encoding a protein Takeout that was significantly downregulated two hours after feeding compared to twenty-four hours after feeding.



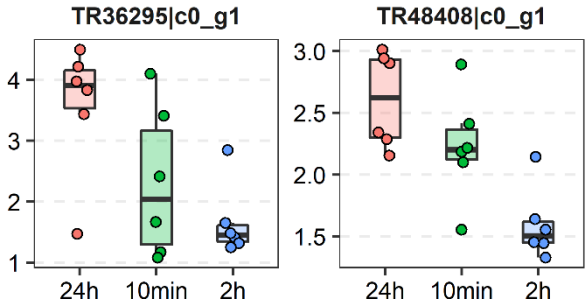
Supplementary figure S4.13. Boxplots of log2-transformed TMM-normalized CPM of a transcripts encoding “protein I’m not dead yet” that was significantly downregulated two hours after feeding compared to twenty-four hours after feeding.



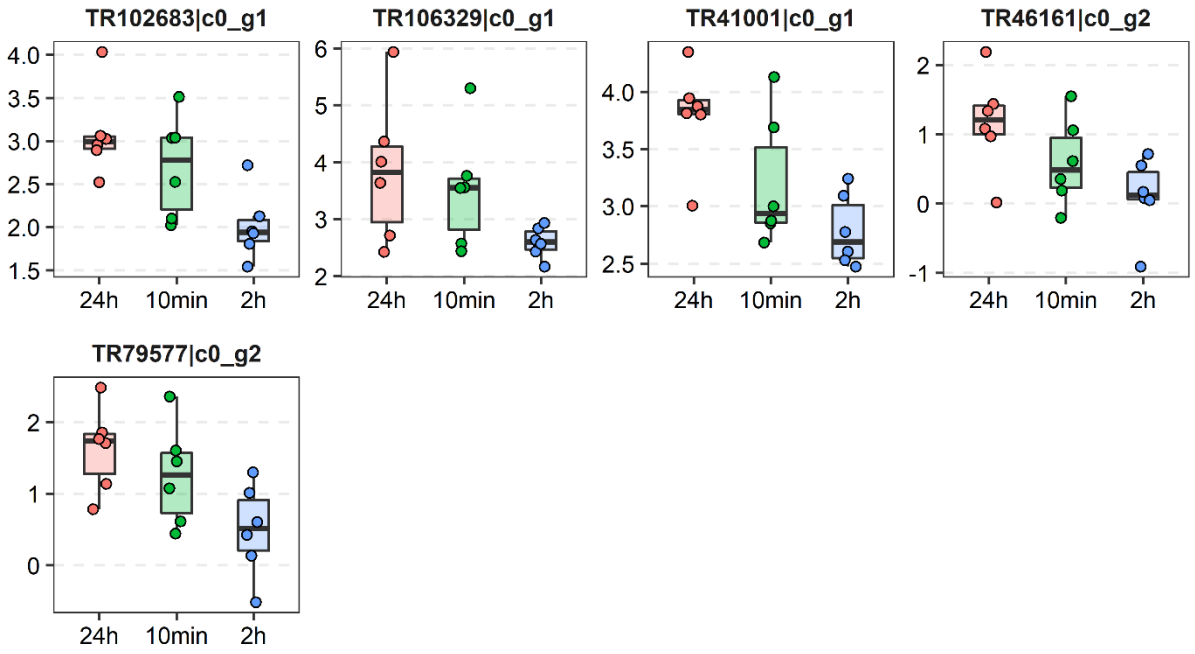
Supplementary figure S4.14. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding peritrophin-like proteins that were significantly downregulated two hours after feeding compared to twenty-four hours after feeding.



Supplementary figure S4.15. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding insect chemosensory proteins that were significantly downregulated two hours after feeding compared to twenty-four hours after feeding.



Supplementary figure S4.16. Boxplots of log2-transformed TMM-normalized CPM of all transcripts encoding ABC transporter proteins that were significantly downregulated two hours after feeding compared to twenty-four hours after feeding.



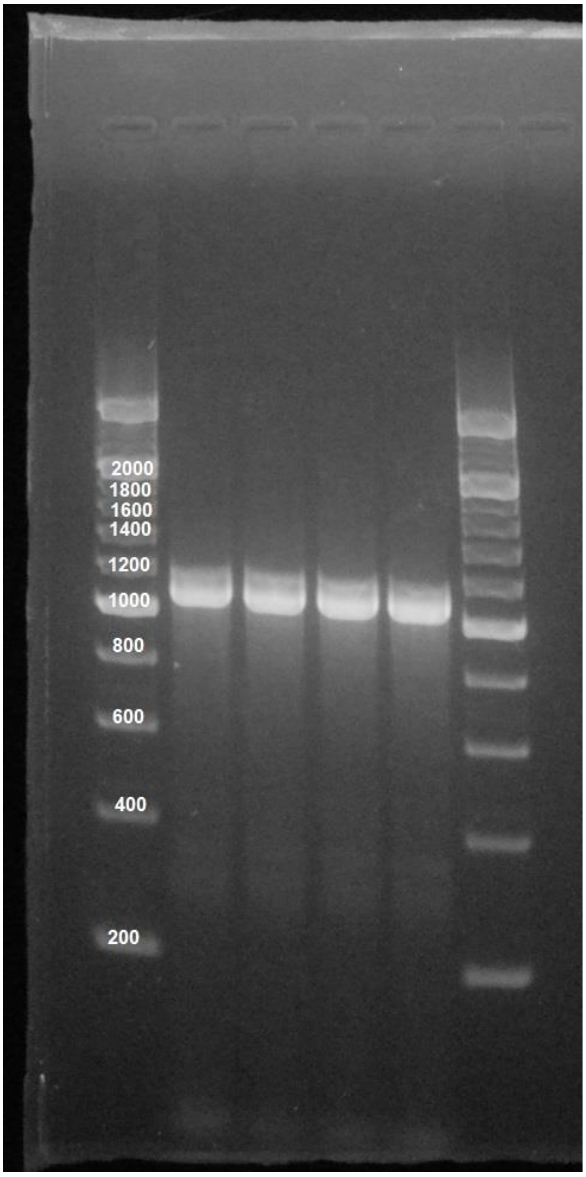
A.4. Supplementary data for chapter 5

Supplementary figure S5.1. MAFFT pairwise alignment of the nucleotide sequences of TR39983|c0_g2_i1 and TR39983|c0_g2_i2. Conservation threshold for shading is 80% identity. Areas in green represent the forward and reverse primer sites used to investigate isoform expression in the *S. gregaria* midgut. The forward primer hybridization site is located between nucleotide 81 and 101 on both isoforms, while the reverse primer hybridization site is located between nucleotide 962 and 981 for TR39983|c0_g2_i1, and nucleotide 1181 and 1200 for TR39983|c0_g2_i2.

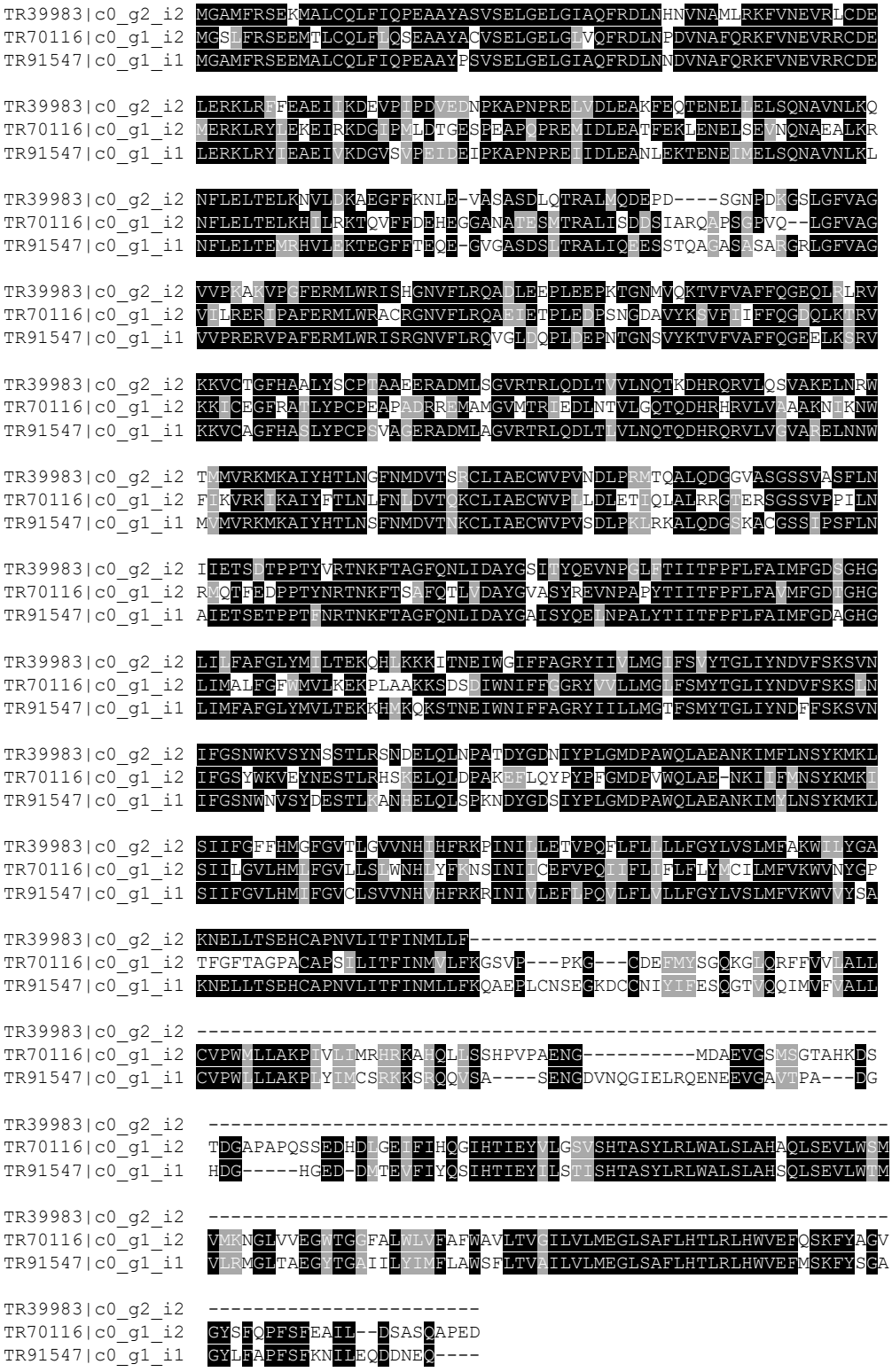
```
TR39983|c0_g2_i1    1  TTCAGCAATCACGTTGGTAATCTCATATTGCCGACGAATGATAGGTACTAAACGCTAAAA
TR39983|c0_g2_i2    1  TTCAGCAATCACGTTGGTAATCTCATATTGCCGACGAATGATAGGTACTAAACGCTAAAA

TR39983|c0_g2_i1   61  GATCTGTGAACAGCCAGCAGCTATTACTGTCGCGTCGCATCTAAATACTGCCGCTTCTGC
TR39983|c0_g2_i2   61  GATCTGTGAACAGCCAGCAGCTATTACTGTCGCGTCGCATCTAAATACTGCCGCTTCTGC

TR39983|c0_g2_i1  121  ATCGGACGGTTGTGGAAGAAATCATCGCTCAGTGAAGCGCACGGAAGAAAAACAGCAGCC
TR39983|c0_g2_i2  121  ATCGGACGGTTGTGGAAGAAATCATCGCTCAGTGAAGCGCACGGAAGAAAAACAGCAGCC

TR39983|c0_g2_i1  181  ATGGGGGCGATGTTTCGGAGCGAGAAGATGGCCTTGTGCCAGCTCTTCATCCAGCCTGAA
TR39983|c0_g2_i2  181  ATGGGGGCGATGTTTCGGAGCGAGAAGATGGCCTTGTGCCAGCTCTTCATCCAGCCTGAA

TR39983|c0_g2_i1  241  GCCGCCTACGCTTCTGTGTCCGAGCTGGGAGAATTGGGCATCGCTCAGTTTCGGGATCTA
TR39983|c0_g2_i2  241  GCCGCCTACGCTTCTGTGTCCGAGCTGGGAGAATTGGGCATCGCTCAGTTTCGGGATCTA

TR39983|c0_g2_i1  301  AATCATAATGTGAATGCCATGTTGCGCAAGTTTGTTAACGAAGTTCGACTATGTGATGAG
TR39983|c0_g2_i2  301  AATCATAATGTGAATGCCATGTTGCGCAAGTTTGTTAACGAAGTTCGACTATGTGATGAG

TR39983|c0_g2_i1  361  CTTGAACGCAAGCTACGATTTTTTGAGGCTGAGATAATCAAGGATGAAGTACCAATTCCT
TR39983|c0_g2_i2  361  CTTGAACGCAAGCTACGATTTTTTGAGGCTGAGATAATCAAGGATGAAGTACCAATTCCT

TR39983|c0_g2_i1  421  GATGTAGAGGACAACCCAAAGGCACCTAACCCAAGAGAGCTTGTCGATCTGGA-------
TR39983|c0_g2_i2  421  GATGTAGAGGACAACCCAAAGGCACCTAACCCAAGAGAGCTTGTCGATCTGGAGGCAAAA

TR39983|c0_g2_i1  474  ------------------------------------------------------------
TR39983|c0_g2_i2  481  TTTGAGCAAACAGAGAATGAATTACTGGAACTGAGTCAGAATGCAGTGAACCTGAAGCAG

TR39983|c0_g2_i1  474  ------------------------------------------------------------
TR39983|c0_g2_i2  541  AACTTCCTAGAATTGACAGAACTGAAGAATGTGCTGGATAAGGCAGAGGGATTTTTTAAG

TR39983|c0_g2_i1  474  ------------------------------------------------------------
TR39983|c0_g2_i2  601  AATCTGGAAGTTGCAAGTGCCTCTGACCTGCAGACTAGGGCACTCATGCAGGATGAGCCT

TR39983|c0_g2_i1  474  --------------------------------GATTTGTTGCTGGAGTTGTACCTAAAGCG
TR39983|c0_g2_i2  661  GACAGTGGGAACCCTGATAAGGGTAGCCTGGGATTTGTTGCTGGAGTTGTACCTAAAGCG

TR39983|c0_g2_i1  503  AAGGTTCCAGGATTTGAGAGAATGCTATGGAG-ATATCACATGGAAATGTGTTCTTGCGA
TR39983|c0_g2_i2  721  AAGGTTCCAGGATTTGAGAGAATGCTATGGAGAATATCACATGGAAATGTGTTCTTGCGA

TR39983|c0_g2_i1  562  CAGGCTGACTTAGAAGAACCCTTAGAAGAGCCAAAGACAGGCAACATGGTACAGAAGACA
TR39983|c0_g2_i2  781  CAGGCTGACTTAGAAGAACCCTTAGAAGAGCCAAAGACAGGCAACATGGTACAGAAGACA

TR39983|c0_g2_i1  622  GTGTTTGTGGCCTTCTTTCAAGGTGAACAGTTGAGACTCCGGGTGAAAAAAGTGTGTACA
TR39983|c0_g2_i2  841  GTGTTTGTGGCCTTCTTTCAAGGTGAACAGTTGAGACTCCGGGTGAAAAAAGTGTGTACA

TR39983|c0_g2_i1  682  GGTTTCCATGCAGCACTGTACTCATGTCCAACTGCTGCAGAAGAGCGTGCTGATATGTTA
TR39983|c0_g2_i2  901  GGTTTCCATGCAGCACTGTACTCATGTCCAACTGCTGCAGAAGAGCGTGCTGATATGTTA

TR39983|c0_g2_i1  742  TCTGGTGTCCGCACCCGCCTTCAAGACCTAACAGTTGTTCTAAATCAAACTAAAGATCAC
TR39983|c0_g2_i2  961  TCTGGTGTCCGCACCCGCCTTCAAGACCTAACAGTTGTTCTAAATCAAACTAAAGATCAC

TR39983|c0_g2_i1  802  AGGCAGAGGGTACTACAGTCTGTTGCTAAAGAGCTGAACAGATGGACAATGATGGTACGC
TR39983|c0_g2_i2 1021  AGGCAGAGGGTACTACAGTCTGTTGCTAAAGAGCTGAACAGATGGACAATGATGGTACGC

TR39983|c0_g2_i1  862  AAGATGAAGGCCATTTACCATACTCTCAATGGTTTCAATATGGATGTCACAAGCAGATGC
TR39983|c0_g2_i2 1081  AAGATGAAGGCCATTTACCATACTCTCAATGGTTTCAATATGGATGTCACAAGCAGATGC

TR39983|c0_g2_i1  922  CTTATTGCAGAGTGCTGGGTGCCTGTGAATGACCTACCAAGGATGACTCAAGCACTGCAA
TR39983|c0_g2_i2 1141  CTTATTGCAGAGTGCTGGGTGCCTGTGAATGACCTACCAAGGATGACTCAAGCACTGCAA
```

```
TR39983|c0_g2_i1  982 GATGGTGGGGTTGCTTCTGGAAGTTCTGTAGCTTCATTTCTTAATATAATTGAAACGTCA
TR39983|c0_g2_i2 1201 GATGGTGGGGTTGCTTCTGGAAGTTCTGTAGCTTCATTTCTTAATATAATTGAAACGTCA

TR39983|c0_g2_i1 1042 GATACTCCACCAACATACGTTAGAACAAATAAGTTCACAGCAGGATTTCAAAACCTTATT
TR39983|c0_g2_i2 1261 GATACTCCACCAACATACGTTAGAACAAATAAGTTCACAGCAGGATTTCAAAACCTTATT

TR39983|c0_g2_i1 1102 GATGCATATGGATCCATAACATACCAAGAAGTTAATCCAGGGTTGTTTACAATTATTACA
TR39983|c0_g2_i2 1321 GATGCATATGGATCCATAACATACCAAGAAGTTAATCCAGGGTTGTTTACAATTATTACA

TR39983|c0_g2_i1 1162 TTCCCGTTCTTGTTTGCAATAATGTTTGGAGATTCTGGACATGGACTCATTCTGTTTGCT
TR39983|c0_g2_i2 1381 TTCCCGTTCTTGTTTGCAATAATGTTTGGAGATTCTGGACATGGACTCATTCTGTTTGCT

TR39983|c0_g2_i1 1222 TTTGGATTGTATATGATTCTCACTGAAAAGCAGCACTTGAAAAAGAAAATTACAAATGAG
TR39983|c0_g2_i2 1441 TTTGGATTGTATATGATTCTCACTGAAAAGCAGCACTTGAAAAAGAAAATTACAAATGAG

TR39983|c0_g2_i1 1282 ATTTGGGGCATCTTCTTCGCTGGACGTTACATTATAGTTTTAATGGGAATATTTTCTGTT
TR39983|c0_g2_i2 1501 ATTTGGGGCATCTTCTTCGCTGGACGTTACATTATAGTTTTAATGGGAATATTTTCTGTT

TR39983|c0_g2_i1 1342 TATACTGGACTAATATACAATGATGTTTTCTCAAAGTCAGTGAACATCTTTGGTAGCAAT
TR39983|c0_g2_i2 1561 TATACTGGACTAATATACAATGATGTTTTCTCAAAGTCAGTGAACATCTTTGGTAGCAAT

TR39983|c0_g2_i1 1402 TGGAAAGTCTCCTATAATAGTTCAACTCTCAGATCAAATGATGAACTTCAGCTCAATCCA
TR39983|c0_g2_i2 1621 TGGAAAGTCTCCTATAATAGTTCAACTCTCAGATCAAATGATGAACTTCAGCTCAATCCA

TR39983|c0_g2_i1 1462 GCAACAGATTATGGTGATAACATATATCCCTTAGGCATGGATCCAGCTTGGCAGTTGGCT
TR39983|c0_g2_i2 1681 GCAACAGATTATGGTGATAACATATATCCCTTAGGCATGGATCCAGCTTGGCAGTTGGCT

TR39983|c0_g2_i1 1522 GAAGCAAACAAGATTATGTTCCTGAACTCATACAAAATGAAGTTATCCATAATATTTGGT
TR39983|c0_g2_i2 1741 GAAGCAAACAAGATTATGTTCCTGAACTCATACAAAATGAAGTTATCCATAATATTTGGT

TR39983|c0_g2_i1 1582 TTCTTTCATATGGGATTTGGAGTGACTCTTGGTGTTGTGAACCATATCCACTTTCGTAAG
TR39983|c0_g2_i2 1801 TTCTTTCATATGGGATTTGGAGTGACTCTTGGTGTTGTGAACCATATCCACTTTCGTAAG

TR39983|c0_g2_i1 1642 CCAATCAACATCCTCCTTGAAACTGTACCCCAATTTCTTTTCCTACTGTTGCTGTTTGGA
TR39983|c0_g2_i2 1861 CCAATCAACATCCTCCTTGAAACTGTACCCCAATTTCTTTTCCTACTGTTGCTGTTTGGA

TR39983|c0_g2_i1 1702 TACTTGGTCTCACTTATGTTTGCTAAATGGATTCTATATGGAGCAAAAAATGAACTGTTG
TR39983|c0_g2_i2 1921 TACTTGGTCTCACTTATGTTTGCTAAATGGATTCTATATGGAGCAAAAAATGAACTGTTG

TR39983|c0_g2_i1 1762 ACAAGTGAACATTGTGCCCCAAATGTGCTCATCACGTTCATCAATATGCTCCTTTTCAA
TR39983|c0_g2_i2 1981 ACAAGTGAACATTGTGCCCCAAATGTGCTCATCACGTTCATCAATATGCTCCTTTTCAA
```

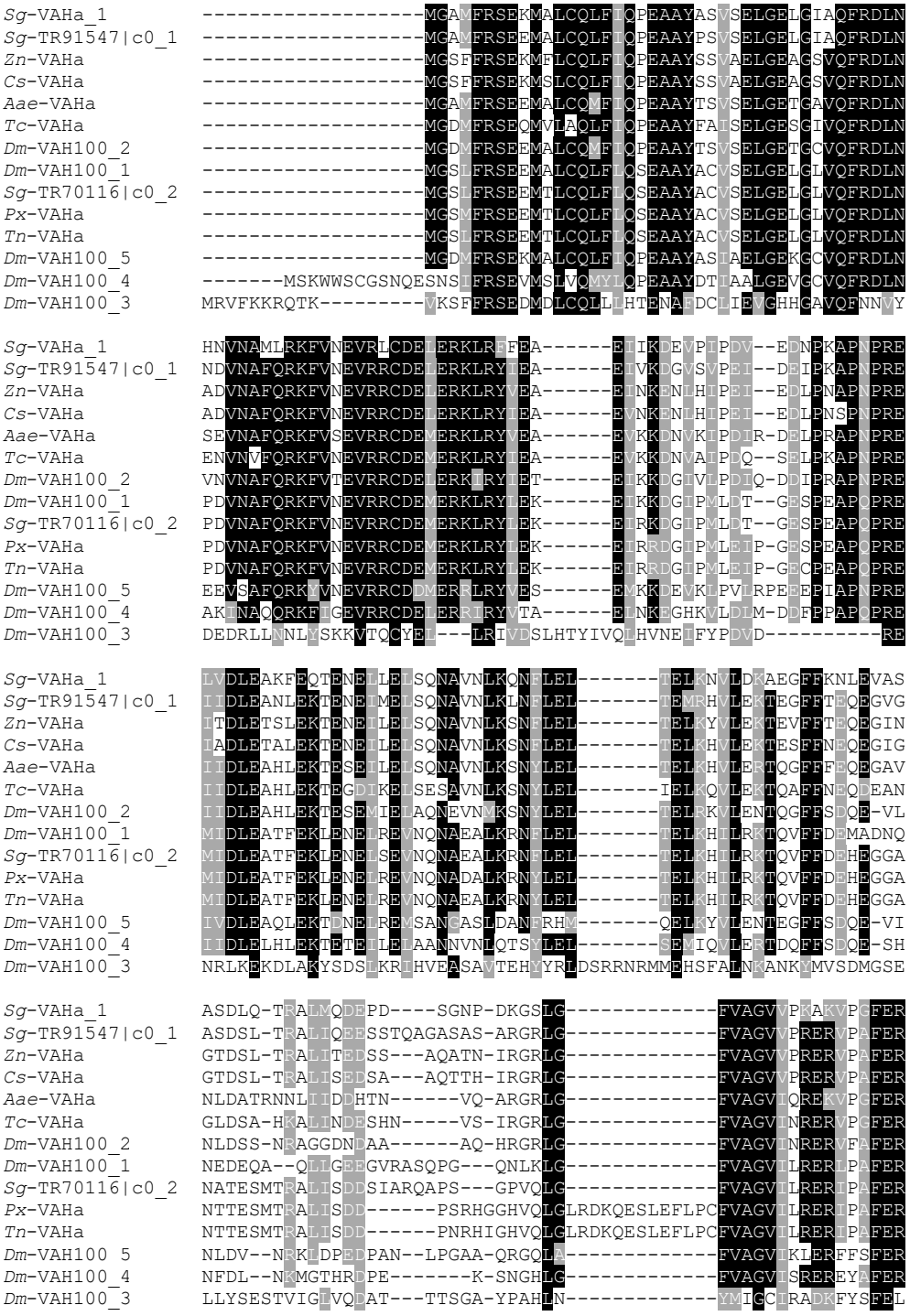
Supplementary figure S5.2. Agarose gel image of the products obtained after PCR to investigate TR39983|c0_g2_i1 and TR39983|c0_g2_i2 expression in the *S. gregaria* midgut (1% agarose gel; 1x TAE buffer; 120 V; 180 A; 1 hour). PCR was performed in quadruplicate. A 200 bp DNA ladder Thermo Scientific™ was used to check the length of the fragments (left lane). The other four lanes display the PCR products obtained from following the thermocycling program: 5 min at 95 °C followed by 32 cycles of 30 s at 95 °C, 1 min at 58 °C and 1 min at 68 °C. Upon 32 cycles of PCR, in all reactions only one product with a length clearly larger than 1000 bp, representing TR39983|c0_g2_i2, was detected.

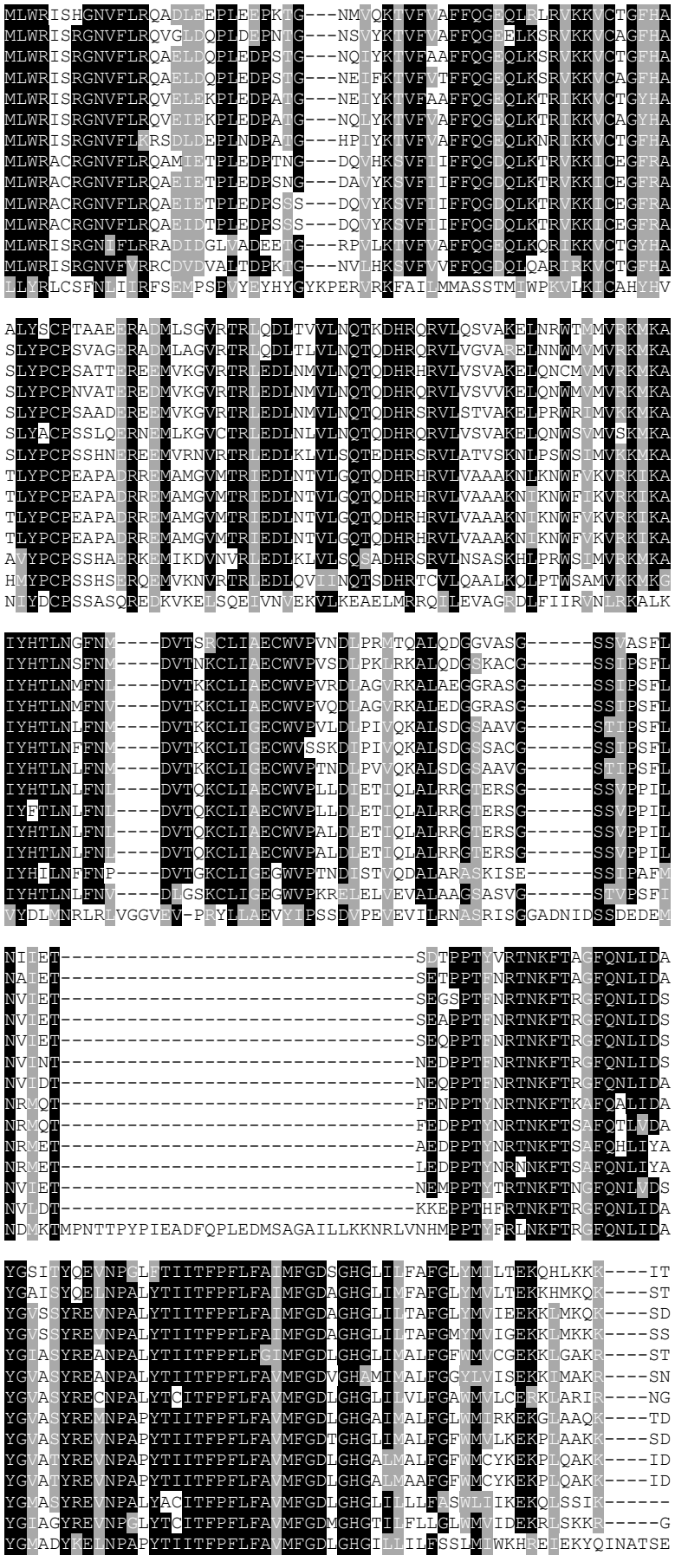


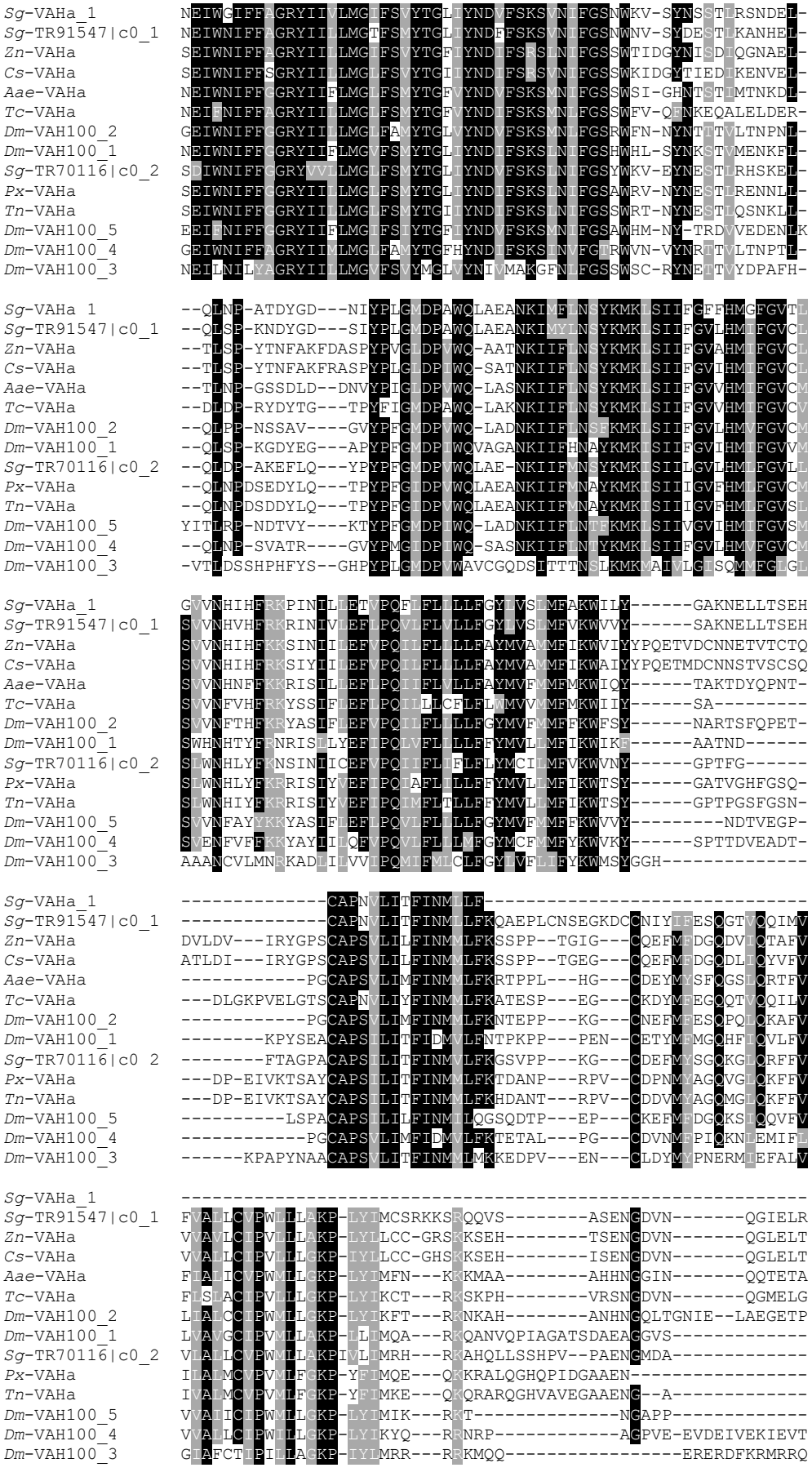
Supplementary figure S5.3. MAFFT multiple sequence alignment of the amino acid sequences of the three different H^+ V-ATPase subunit a isoforms with an ORF present in the *S. gregaria* midgut reference transcriptome, namely TR39983|c0_g2_i2 (*Sg*-VAHa_1), TR70116|c0_g1_i2, and TR91547|c0_g1_i1. Conservation threshold for shading is 80% identity.

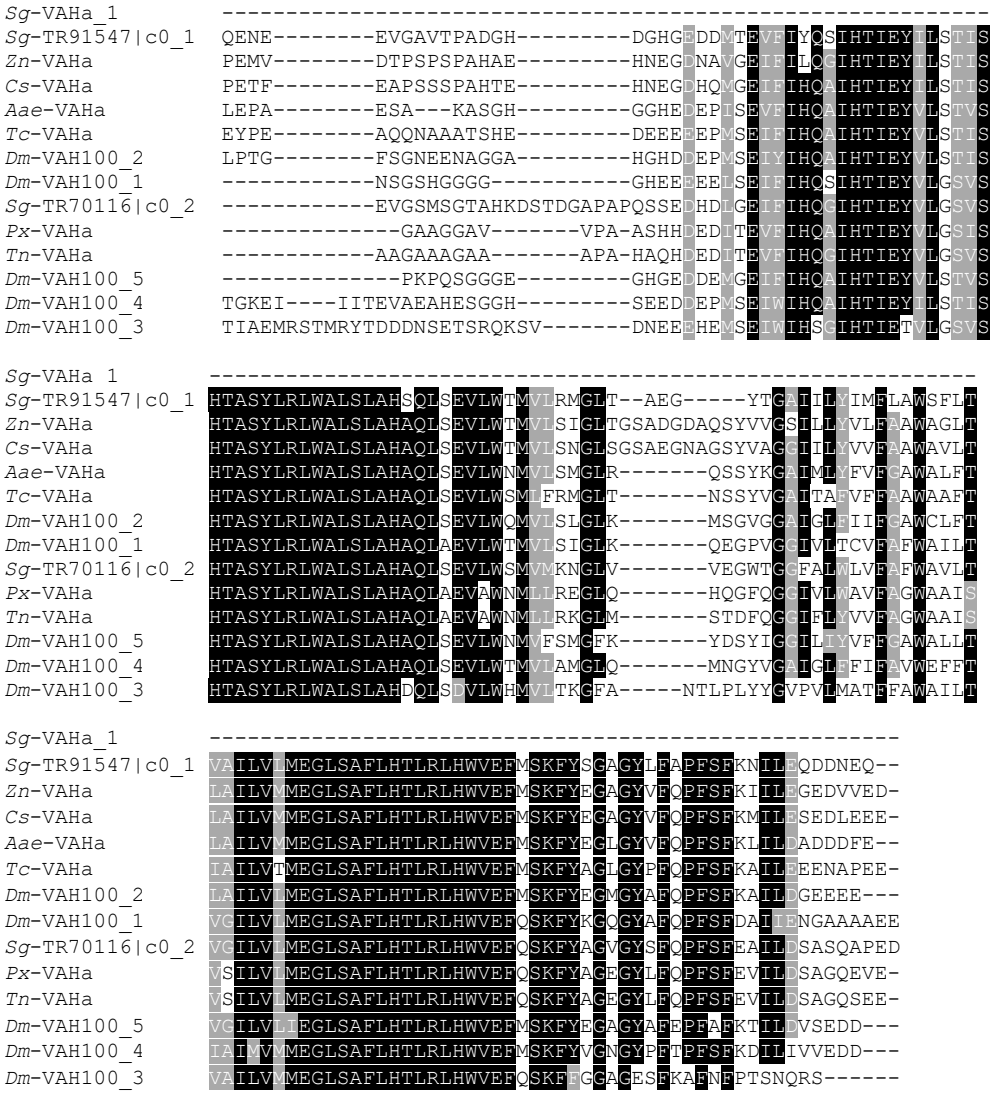


Supplementary figure S5.4. MAFFT multiple sequence alignment of the amino acid sequences of the three *S. gregaria* midgut H^+ V-ATPase subunit a isoforms, namely TR39983|c0_g2_i2 (*Sg*-VAHa_1), TR70116|c0_g1_i2, and TR91547|c0_g1_i1, and other publicly available insect H^+ V-ATPase subunit a sequences. Conservation threshold for shading is 80% identity. Abbreviations: VAHa = H^+ V-ATPase subunit a, *Sg = Schistocerca gregaria*, *Dm = Drosophila melanogaster*, *Zn = Zootermopsis nevadensis*, *Cs = Chilo suppressalis*, *Tc = Tribolium castaneum*, *Aae = Aedes aegypti*, *Px = plutella xylostella.*

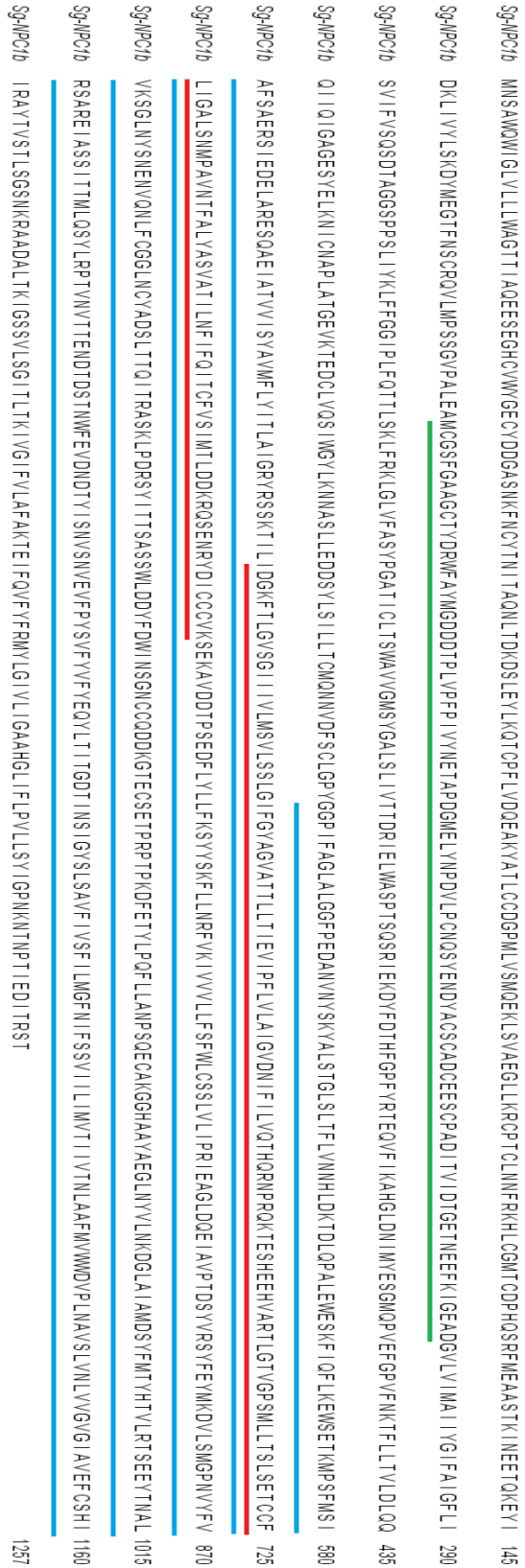



Sg-VAHa_1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0_2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3
Sg-VAHa 1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0_2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3
Sg-VAHa_1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0_2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3
Sg-VAHa_1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0 2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3
Sg-VAHa_1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0_2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3

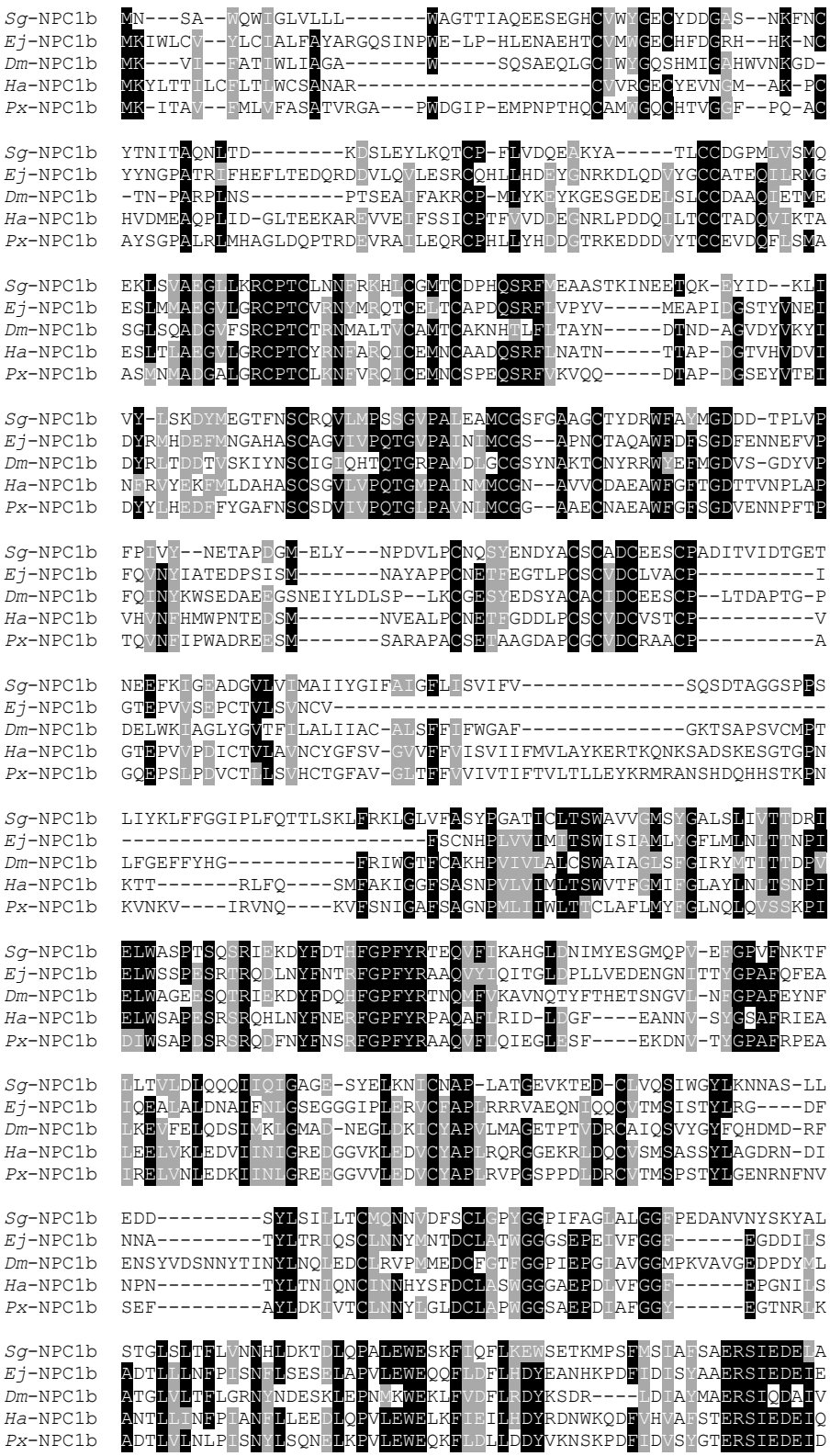

Sg-VAHa_1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0_2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3
Sg-VAHa 1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0_2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3
Sg-VAHa_1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0_2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3
Sg-VAHa_1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0 2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3
Sg-VAHa_1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0_2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3

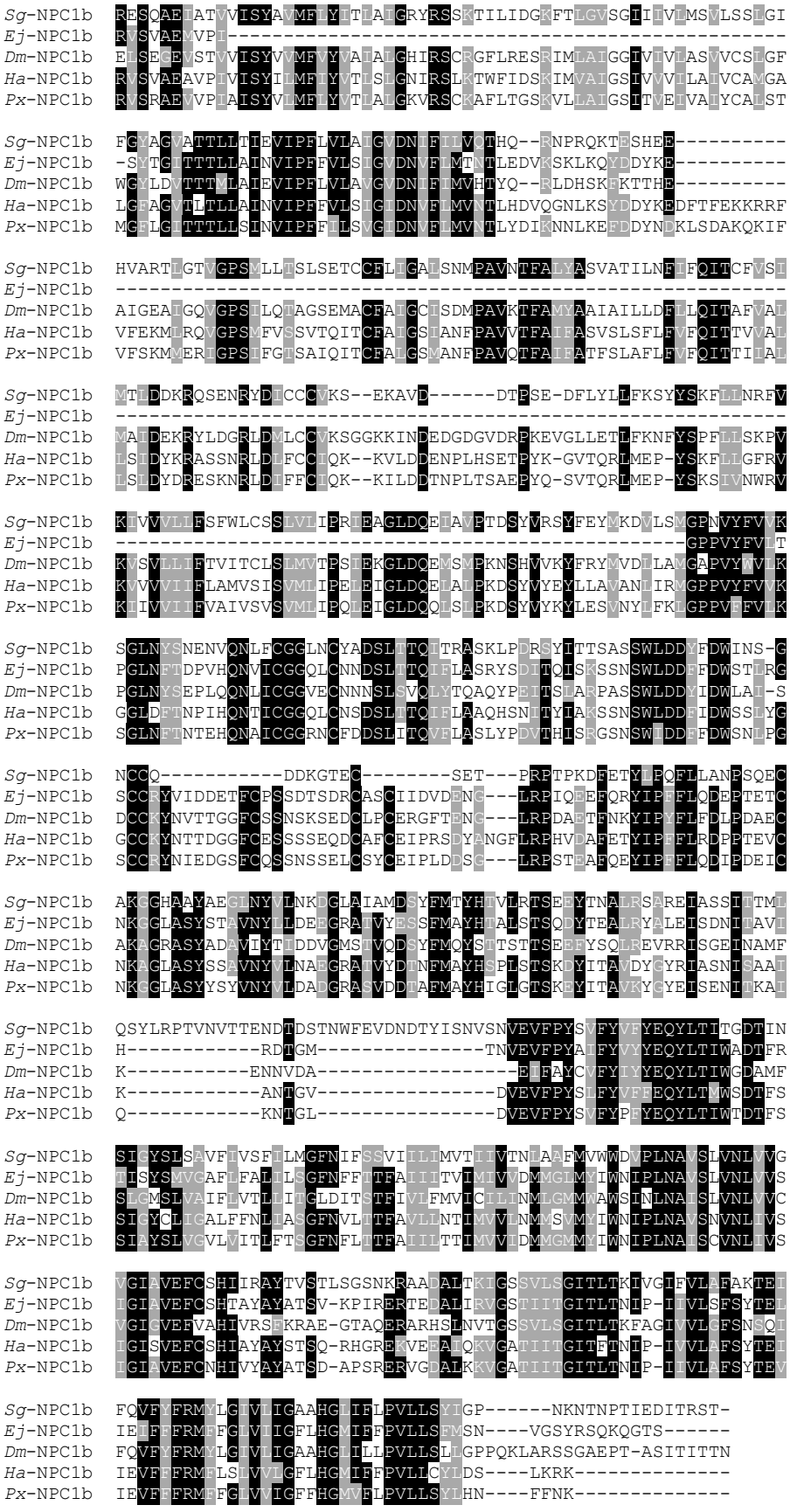

Sg-VAHa_1
Sg-TR91547|c0_1
Zn-VAHa
Cs-VAHa
Aae-VAHa
Tc-VAHa
Dm-VAH100_2
Dm-VAH100_1
Sg-TR70116|c0_2
Px-VAHa
Tn-VAHa
Dm-VAH100_5
Dm-VAH100_4
Dm-VAH100_3

Supplementary figure S5.5. *Sg*-NPC1b amino acid sequence (1,257 amino acids) containing the “NPC_N” domain between amino acids 179-270 (underlined in green), the SSD domain between amino acids 628-780 (underlined in red), and the patched domain between amino acids 507-1238 (underlined in blue), described to be characteristic to NPC1 proteins.



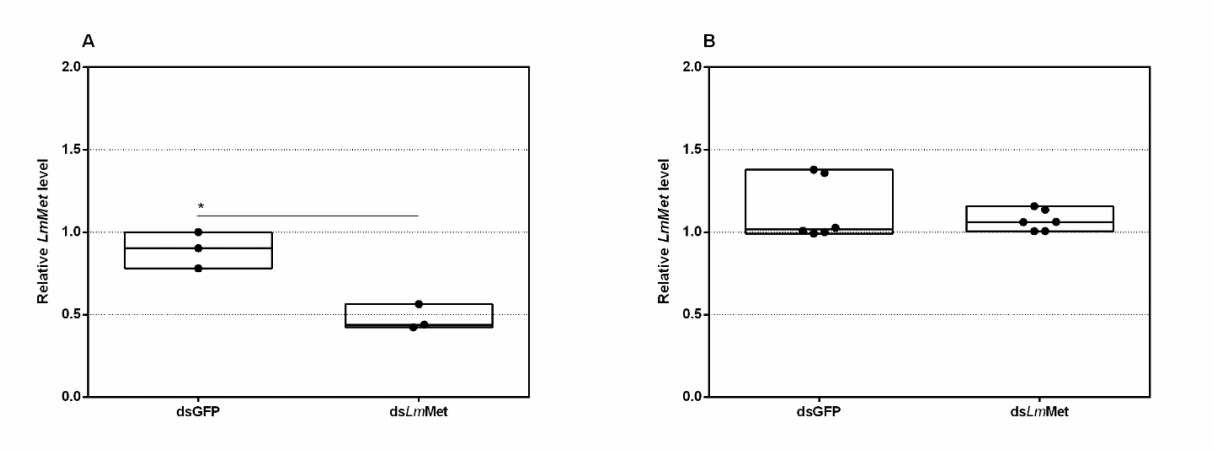
Supplementary figure S5.6. MAFFT multiple sequence alignment of *Sg*-NPC1b and other publicly available insect NPC1b amino acid sequences. Conservation threshold for shading is 80% identity. Abbreviations: NPC1b = Niemann-Pick C1b, *Sg = Schistocerca gregaria*, *Ej = Eumeta japonica*, *Dm = Drosophila melanogaster, Px = Plutella xylostella*, *Ha = Helicoverpa armigera*.




Sg-NPC1b
Ej-NPC1b
Dm-NPC1b
Ha-NPC1b
Px-NPC1b

A.5. Supplementary data for chapter 6

Supplementary figure S6.1. Relative mRNA levels of *LmMet* in brain (A) and midgut (B) of larvae treated with either dsGFP or ds*Lm*Met measured with RT-qPCR. Box plots are based on six pools of at least four individuals. Transcript levels were normalized against two reference genes, *LmRPS13* and *LmRP49*. Significant differences are indicated with an asterisk ($p < 0.05$, Student's T-test).



List of publications

Scientific publications in internationally peer-reviewed academic journals:

Holtof M., Lenaerts C., Cullen D., Vanden Broeck J. (2019) Extracellular nutrient digestion and absorption in the insect gut. Cell and Tissue Research. doi: 10.1007/s00441-019-03031-9.

Wulff JP., Sierra I., Sterkel M., **Holtof M.**, Van Wielendaele P., Francini F., Vanden Broeck J., Ons S. (2017). Orcokinin neuropeptides regulate ecdysis in the hemimetabolous insect, *Rhodnius prolixus*. Insect Biochemistry and Molecular Biology, 81, 91-102. doi: 10.1016/j.ibmb.2017.01.003.

[*] Spit, J., [*] **Holtof, M.**, Badisco, L., Vergauwen, L., Vogel, E., Knapen, D., Vanden Broeck, J. (2016) Transcriptional Analysis of The Adaptive Digestive System of The Migratory Locust in Response to Plant Defensive Protease Inhibitors. Scientific Reports, 6:32460. doi: 10.1038/srep32460.
[*] *Shared first authorship*

Spit J., Zels S., Dillen S., **Holtof M.**, Wynant N., Vanden Broeck J. (2014). Effects of different dietary conditions on the expression of trypsin- and chymotrypsin-like protease genes in the digestive system of the migratory locust, *Locusta migratoria*. Insect Biochemistry and Molecular Biology, 48, 100-109. doi: 10.1016/j.ibmb.2014.03.002.

Presentations and posters at national and international scientific conferences:

Holtof M., Marchal E., Vanden Broeck J. (2018). The role of juvenile hormone in the reproductive development of the male desert locust, *Schistocerca gregaria*. Poster presentation at the 2018 ESA, ESC, and ESBC Joint Annual Meeting (Entomology 2018). Vancouver BC, Canada, 11-14 November 2018.

Holtof M., Spit J., Badisco L., Vogel E., Vanden Broeck J. (2016). Juvenile hormone as a regulator of digestive enzyme synthesis in the African migratory locust, *Locusta migratoria*. Poster presentation at the 2016 CECE conference. Leuven, Belgium, 21 August-25 August 2016

Spit J., **Holtof M.**, Badisco L., Vogel E., Vanden Broeck J. (2015). Insect resistance to dietary protease inhibitors. Poster presentation at the Bioforum 2015 conference. Liege, Belgium, 13 May 2015-13 May 2015.

Holtof M., Jozef Vanden Broeck (2015). Transcriptional analysis of the protease inhibitor induced regulation of digestion in the desert locust, *Schistocerca gregaria*. Oral presentation at the 2015 Programming for Evolutionary Biology Conference. Vairao, Portugal, 27 April - 1 may 2015.

Holtof M., Spit J., Vanden Broeck J. (2014). Compensatory responses to dietary protease inhibitors in the migratory locust, *Locusta migratoria*. Poster presentation at the Zoology 2014 conference. Liège, Belgium, 12-13 December 2014.