



UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
F. EDWARD HÉBERT SCHOOL OF MEDICINE
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND 20814-4799



March 17, 2008

GRADUATE PROGRAMS IN THE
BIOMEDICAL SCIENCES AND
PUBLIC HEALTH

APPROVAL SHEET

Ph.D. Degrees

Interdisciplinary
-Emerging Infectious Diseases
-Molecular & Cell Biology
-Neuroscience

Departmental
-Clinical Psychology
-Environmental Health Sciences
-Medical Psychology
-Medical Zoology
-Pathology

Doctor of Public Health (Dr.P.H.)

Physician Scientist (MD/Ph.D.)

Master of Science Degrees

-Public Health

Masters Degrees

-Military Medical History
-Public Health
-Tropical Medicine & Hygiene

Graduate Education Office

Dr. Eleanor S. Metcalf, Associate Dean
Janet Anastasi, Program Coordinator
Tanice Acevedo, Education Technician

Web Site

<http://www.usuhs.mil/graded/>

E-mail Address

graduateprogram@usuhs.mil

Phone Numbers

Commercial: 301-295-9474
Toll Free: 800-772-1747
DSN: 295-9474
FAX: 301-295-6772

Title of Dissertation: "A Functional Transcriptomic Approach to Understanding the Sand Fly Vector Relationships to the Host and *Leshmania* parasites"

Name of Candidate: Ryan Jochim
Doctor of Philosophy Degree
7 April 2008

Dissertation and Abstract Approved:

Richard Andre, Ph.D.
Department of Preventive Medicine & Biometrics
Committee Chairperson

4/7/08

Date

Jesus Valenzuela, Ph.D.
National Institute of Health
Committee Member

04/07/08

Date

Stephen Davies, Ph.D.
Department of Microbiology & Immunology
Committee Member

4/7/08

Date

Ernest Maynard, Ph.D.
Department of Biochemistry & Molecular Biology
Committee Member

4/7/08

Date

Carolina Barillas-Mury, M.D., Ph.D.
National Institutes of Health
Committee Member

4/7/08

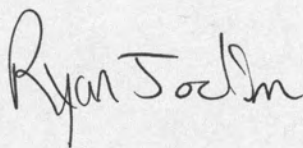
Date

Copyright Statement

The author hereby certifies that the use of any copyrighted material in the thesis manuscript entitled:

“A functional transcriptomic approach to understanding the sand fly vector relationships to the host and *Leishmania* parasites”

is appropriately acknowledged and, beyond brief excerpts, is with the permission of the copyright owner.

A handwritten signature in black ink, reading "Ryan C. Jochim". The signature is written in a cursive, flowing style.

Ryan C. Jochim

Emerging Infectious Diseases Program

Uniformed Services University of the Health Sciences

Abstract

Title of Dissertation:

A functional transcriptomic approach to understanding the sand fly vector relationships to the host and *Leishmania* parasites

Ryan C. Jochim, Doctor of Philosophy, 2008

Thesis Directed by:

Jesus G. Valenzuela, Ph.D.

Principal Investigator, Vector Molecular Biology Unit, NIAID, NIH

Phlebotomine sand flies are the only known biological vectors of *Leishmania*. The maintenance of *Leishmania* can be represented by the epidemiological triad; the relationship between the vertebrate host, the parasite and the insect vector. This research focuses on the relationships between the sand fly and the host and also between the sand fly and *Leishmania*. The feeding success of the sand fly and the transmission of *Leishmania* are linked to the pharmacological cocktail of molecules in the sand fly saliva – the host-sand fly relationship. Adenosine deaminase (ADA) was identified as a salivary constituent of *Lutzomyia longipalpis* and is not present in the saliva of *Phlebotomus papatasi*, *P. argentipes*, *P. perniciosus* and *P. ariasi*, leading to the false presumption that ADA is a *Lutzomyia*-specific enzyme. Two transcripts encoding for ADA were identified by the analysis of cDNA libraries produced from salivary glands of *P. duboscqi*. Our research revealed the presence of ADA activity in the saliva of *P.*

duboscqi and demonstrated that this activity was attributable to the identified transcripts. Our expert use of functional transcriptomics was expanded from sand fly saliva to include sand fly midgut tissue. The successful completion of the *Leishmania* life cycle, from the ingested amastigote to the transmitted metacyclic promastigote, occurs within the sand fly midgut – the sand fly-*Leishmania* relationship. We randomly sequenced a large number of transcripts from high-quality, full-length, female *P. papatasi* and *Lu. longipalpis* midgut-specific cDNA libraries. By means of customized bioinformatics analysis, cDNA libraries were evaluated from sugar-fed and blood-fed *P. papatasi* and *Lu. longipalpis*, in the presence or absence of *L. major* or *L. infantum chagasi*, respectively. Subsequently, we evaluated cDNA libraries generated from *L. infantum chagasi*-infected and uninfected *Lu. longipalpis* midguts after blood meal digestion. Sand fly midgut transcripts modulated by blood-feeding or *Leishmania* colonization were identified using a functional transcriptomic approach. By cataloging the midgut tissue molecular repertoire, we identified novel sand fly midgut-derived molecules, including proteases, microvillar proteins, peritrophins, oxidative stress proteins and antimicrobials. Expression profiling of selected transcripts from *Lu. longipalpis* described the temporal midgut transcript abundance and verified the striking effect of *L. infantum chagasi* colonization on microvillar protein, protease and peritrophin transcription.

**A functional transcriptomic approach to understanding the
sand fly vector relationships to the host and *Leishmania*
parasites**

By

Ryan C. Jochim

Dissertation submitted to the Faculty of the Emerging Infectious Diseases Program of the
Uniformed Services University of the Health Sciences in partial fulfillment of the
requirements for the degree of Doctor of Philosophy 2008

Acknowledgements

Jesus G. Valenzuela, Ph.D., Dissertation Advisor

Richard G. Andre, Ph.D., Dissertation Committee Chair

Carolina V. Barillas-Mury, M.D., Ph.D., Dissertation Committee Member

Stephen J. Davies, BVSc., Ph.D., Dissertation Committee Member

Ernest L. Maynard, Ph.D., Dissertation Committee Member

Members of the Valenzuela Lab

Krystle L. Mohawk, My Better Half

“Science is a wonderful thing if one does not have to earn one's living at it.”

— *Albert Einstein* —

Table of Contents

APPROVAL SHEET	I
COPYRIGHT STATEMENT	II
ABSTRACT	III
ACKNOWLEDGEMENTS.....	VI
TABLE OF FIGURES	XIII
TABLE OF TABLES	XVI
INTRODUCTION.....	1
EPIDEMIOLOGY	2
CLINICAL ASPECTS OF LEISHMANIASIS.....	2
SAND FLIES	5
LEISHMANIA LIFE CYCLE	10
SAND FLY SALIVA	16
PREMISE.....	18
REFERENCES	19
IDENTIFICATION AND CHARACTERIZATION OF A SALIVARY ADENOSINE DEAMINASE FROM THE SAND FLY <i>PHLEBOTOMUS DUBOSCQI</i>, THE VECTOR OF <i>LEISHMANIA MAJOR</i> IN SUB-SAHARAN AFRICA	28
ABSTRACT	29
BACKGROUND	30
RESULTS	31
<i>Salivary gland ADA activity</i>	<i>32</i>
<i>Lack of adenosine and AMP in the saliva of P. duboscqi.....</i>	<i>37</i>
<i>Expression and activity of recombinant P. duboscqi salivary ADA</i>	<i>40</i>
DISCUSSION	42

MATERIALS AND METHODS	46
<i>Sand flies and preparation of salivary gland homogenate (SGH)</i>	46
<i>Salivary gland cDNA library</i>	46
<i>Phylogenetic analysis</i>	47
<i>Enzymatic assays</i>	47
<i>Molecular sieving-high-performance liquid chromatography</i>	48
<i>Expression of P. duboscqi ADA</i>	48
<i>Sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis (PAGE) and western blotting</i>	49
REFERENCES	50
 EXPLORING THE MIDGUT TRANSCRIPTOME OF <i>PHLEBOTOMUS PAPATASI</i>:	
COMPARATIVE ANALYSIS OF EXPRESSION PROFILES OF SUGAR-FED, BLOOD-FED AND	
<i>LEISHMANIA MAJOR</i>-INFECTED SAND FLIES.....	53
 ABSTRACT	54
BACKGROUND	55
RESULTS AND DISCUSSION	56
<i>Microvilli-associated like proteins</i>	62
<i>Peritrophin-like proteins</i>	65
<i>Trypsin</i>	67
<i>Chymotrypsin</i>	67
<i>Carboxypeptidase</i>	69
<i>Astacin-like zinc metalloprotease</i>	72
<i>Kazal-type serine protease inhibitor</i>	72
<i>Ferritin</i>	76
<i>Glutathione S-transferase (GST)</i>	78
<i>Unknown proteins</i>	78
<i>Functionally characterized proteins</i>	79

<i>Comparative analysis of transcripts that significantly differ from the sugar-fed and blood-fed midgut cDNA libraries.....</i>	79
<i>Validation of transcript abundance of selected sequences by real-time PCR.....</i>	81
<i>Comparative analysis of transcripts significantly differs from the blood-fed and L. major-infected midgut cDNA libraries</i>	83
CONCLUSION	87
METHODS	89
<i>Sand flies.....</i>	89
<i>Messenger RNA extraction and cDNA library construction.....</i>	89
<i>Random sequencing</i>	90
<i>Bioinformatic analysis</i>	91
<i>Quantitative PCR.....</i>	92
<i>Semi-quantitative PCR.....</i>	93
REFERENCES.....	93

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 99

ABSTRACT	100
BACKGROUND	101
RESULTS AND DISCUSSION.....	103
<i>Proteases</i>	107
<i>Trypsin.....</i>	109
<i>Chymotrypsin.....</i>	114
<i>Carboxypeptidases.....</i>	115
<i>Astacin</i>	118
<i>Peritrophin-like proteins.....</i>	123
<i>Microvillar proteins.....</i>	128
<i>Oxidative stress molecules.....</i>	132

<i>Serine protease inhibitors</i>	136
<i>Anti-bacterial molecules</i>	140
<i>Transcripts differentially expressed by blood-feeding and digestion</i>	144
<i>Transcripts differentially expressed by the presence of Leishmania infantum chagasi</i> ..	146
CONCLUSION	149
METHODS	151
<i>Sand flies</i>	151
<i>Leishmania and sand fly infection</i>	151
<i>cDNA library construction</i>	153
<i>DNA Sequencing</i>	154
<i>Bioinformatics</i>	154
REFERENCES	156

**TEMPORAL PROFILING OF TRANSCRIPTS IN THE MIDGUT TISSUE OF THE SAND FLY
LUTZOMYIA LONGIPALPIS: EFFECTS OF BLOOD-FEEDING AND *LEISHMANIA INFANTUM*
CHAGASI INFECTION.....160**

ABSTRACT	161
BACKGROUND	162
RESULTS AND DISCUSSION	164
<i>L. infantum chagasi is most abundant prior to the defecation of digested blood.</i>	167
<i>Microvillar protein transcripts are affected by the presence of L. infantum chagasi.</i>	167
<i>L. infantum chagasi accelerates and upregulates peritrophin transcription.</i>	174
<i>Carboxypeptidases are increased in response to L. infantum chagasi colonization.</i>	182
<i>Late-stage L. infantum chagasi infection increases chymotrypsin abundance.</i>	184
<i>Astacin transcripts increase but trypsin is relatively unaffected by L. infantum chagasi.</i>	187
<i>Blood meal digestion induces actin transcription.</i>	189
CONCLUSION	190
METHODS	191

<i>Sand flies</i>	191
<i>Leishmania and sand fly infection</i>	192
<i>RNA extraction</i>	193
<i>Multiplex RT-PCR</i>	193
<i>Real-time RT-PCR</i>	195
<i>Data analysis</i>	195
REFERENCES	196
DISCUSSION AND FUTURE DIRECTIONS	199
OVERVIEW	200
PHLEBOTOMUS DUBOSQI ADENOSINE DEAMINASE	201
FUNCTIONAL TRANSCRIPTOMICS OF THE SAND FLY MIDGUT	202
TEMPORAL PROFILING OF LUTZOMYIA LONGIPALPIS MIDGUT TRANSCRIPTS.....	205
REFERENCES	206

Table of Figures

FIGURE 1. WORLD MAPS INDICATING ENDEMIC AREAS OF VISCERAL AND CUTANEOUS LEISHMANIASIS.....	3
FIGURE 2. TIMELINE OF EVENTS LEADING TO THE INCRIMINATION OF NEW WORLD AND OLD WORLD SAND FLY VECTORS OF LEISHMANIASIS.	8
FIGURE 3. LIFE CYCLE OF <i>LEISHMANIA</i> PARASITES WITHIN THE SAND FLY.	11
FIGURE 4. AMINO ACID ALIGNMENT OF THE TWO ADENOSINE DEAMINASE (ADA) MOLECULES DERIVED FROM TRANSCRIPTS FOUND IN <i>P. DUBOSCQI</i> SALIVARY GLANDS.	33
FIGURE 5. CLUSTAL ALIGNMENTS OF INVERTEBRATE PUTATIVE SALIVARY ADENOSINE DEAMINASE (ADA) OF <i>P. DUBOSCQI</i> (PduM73 AND PduM74), <i>LUTZOMYIA LONGIPALPIS</i> , <i>AEDES AEGYPTI</i> , <i>A. ALBOPICTUS</i> , <i>CULEX PIPIENS</i> AND MAMMALIAN ADA (MOUSE, RAT AND HUMAN).	34
FIGURE 6. PHYLOGENETIC TREE ANALYSIS OF PUTATIVE ADENOSINE DEAMINASE (ADA).	36
FIGURE 7. ADA ACTIVITY OF SALIVARY HOMOGENATES OF <i>P. DUBOSCQI</i>	38
FIGURE 8. SALIVARY ADENOSINE DEAMINASE (ADA) ACTIVITY FROM SALIVARY GLAND HOMOGENATE (SGH) OF UNFED AND BLOOD-FED SAND FLIES AND FROM PUNCTURED SALIVARY GLANDS.	39
FIGURE 9. THREE-DIMENSIONAL CHROMATOGRAPHIC DISPLAY OF PHOTODIODE ARRAY DATA OBTAINED FROM MS-HPLC OF <i>P. PAPATASI</i> AND <i>P. DUBOSCQI</i> SALIVARY GLAND HOMOGENATE (SGH).	41
FIGURE 10. EXPRESSION OF RECOMBINANT <i>P. DUBOSCQI</i> SALIVARY ADENOSINE DEAMINASE (ADA).	43
FIGURE 11. ENZYMATIC ACTIVITY OF RECOMBINANT <i>P. DUBOSCQI</i> SALIVARY ADENOSINE DEAMINASE (ADA).	44
FIGURE 12. DISTRIBUTION OF SEQUENCES ANALYZED FROM EACH CDNA LIBRARY SEPARATED BY PUTATIVE BIOLOGIC FUNCTION.	60
FIGURE 13. MULTIPLE SEQUENCE ALIGNMENT OF THE FOUR PUTATIVE MICROVILLI ASSOCIATED-LIKE PROTEINS FOUND IN THE MIDGUT OF <i>PHLEBOTOMUS PAPATASI</i>	64
FIGURE 14. CHARACTERIZATION OF PERITROPHIN SEQUENCES.	66
FIGURE 15. PHYLOGENETIC ANALYSIS OF TRYPSINS.	68
FIGURE 16. CHYMOTRYPSIN SEQUENCE ANALYSIS.	70
FIGURE 17. <i>PHLEBOTOMUS PAPATASI</i> MIDGUT CARBOXYPEPTIDASE-LIKE PROTEINS.	71
FIGURE 18. MULTIPLE SEQUENCE ANALYSIS OF ASTACIN-LIKE PROTEINS.	73

FIGURE 19. SEQUENCE ANALYSIS OF KAZAL-TYPE PROTEINS.....	75
FIGURE 20. SEQUENCE ANALYSIS OF FERRITIN HEAVY AND LIGHT CHAIN MOLECULES.	77
FIGURE 21. COMPARATIVE ABUNDANCE OF PERITROPHIN TRANSCRIPTS IN SUGAR-FED OR BLOOD-FED SAND FLIES.....	82
FIGURE 22. TRANSCRIPT ABUNDANCE OF MICROVILLI ASSOCIATED-LIKE PROTEINS COMPARED BETWEEN UNFED AND BLOOD-FED SAND FLIES.	84
FIGURE 23. HISTOGRAPH OF THE NUMBER OF SEQUENCES GROUPED INTO FUNCTIONAL CLASSES FROM THE SUGAR-FED, BLOOD-FED AND POST-BLOOD MEAL DIGESTION CDNA LIBRARIES.	106
FIGURE 24. SEQUENCE ANALYSIS OF TRYPSIN-LIKE SERINE PROTEASES.	112
FIGURE 25. CHYMOTRYPSIN SEQUENCE ANALYSIS.....	116
FIGURE 26. ANALYSIS OF PUTATIVE CARBOXYPEPTIDASE MOLECULES.	119
FIGURE 27. ASTACIN-LIKE METALLOPROTEASE SEQUENCE COMPARISON AND ANALYSIS.	121
FIGURE 28. CHARACTERIZATION OF PERITROPHIN SEQUENCES.	125
FIGURE 29. SEQUENCE ANALYSIS OF MICROVILLAR PROTEINS.	130
FIGURE 30. PHYLOGENETIC ANALYSIS OF GLUTATHIONE S-TRANSFERASE MOLECULES.	134
FIGURE 31. SEQUENCE ANALYSIS OF PEPTIDOGLYCAN RECOGNITION PROTEINS.....	141
FIGURE 32. MULTIPLE SEQUENCE ALIGNMENT OF PUTATIVE DEFENSIN SEQUENCES.....	143
FIGURE 33. TEMPORAL PROFILE OF <i>L. INFANTUM CHAGASI</i> INFECTION IN THE MIDGUT OF <i>LU. LONGIPALPIS</i> . 168	
FIGURE 34. COMPARATIVE ABUNDANCE OF MICROVILLAR PROTEIN <i>MVP3</i> IN UNINFECTED AND <i>L. INFANTUM</i> <i>CHAGASI</i> -INFECTED <i>LU. LONGIPALPIS</i> MIDGUTS.	170
FIGURE 35. COMPARATIVE ABUNDANCE OF MICROVILLAR PROTEIN <i>MVP4</i> IN UNINFECTED AND <i>L. INFANTUM</i> <i>CHAGASI</i> -INFECTED <i>LU. LONGIPALPIS</i> MIDGUTS.	172
FIGURE 36. COMPARATIVE ABUNDANCE OF PERITROPHIN <i>PER1</i> IN UNINFECTED AND <i>L. INFANTUM CHAGASI</i> - INFECTED <i>LU. LONGIPALPIS</i> MIDGUTS.....	175
FIGURE 37. COMPARATIVE ABUNDANCE OF PERITROPHIN <i>PER2</i> IN UNINFECTED AND <i>L. INFANTUM CHAGASI</i> - INFECTED <i>LU. LONGIPALPIS</i> MIDGUTS.....	177
FIGURE 38. COMPARATIVE ABUNDANCE OF CARBOXYPEPTIDASE <i>CPEPA1</i> IN UNINFECTED AND <i>L. INFANTUM</i> <i>CHAGASI</i> -INFECTED <i>LU. LONGIPALPIS</i> MIDGUTS.	180

FIGURE 39. COMPARATIVE ABUNDANCE OF CARBOXYPEPTIDASE <i>CPEP</i> B IN UNINFECTED AND <i>L. INFANTUM</i> <i>CHAGASI</i> -INFECTED <i>LU. LONGIPALPIS</i> MIDGUTS.	181
FIGURE 40. COMPARATIVE ABUNDANCE OF CHYMOTRYPSIN, <i>CHYM</i> 3, IN UNINFECTED AND <i>L. INFANTUM</i> <i>CHAGASI</i> -INFECTED <i>LU. LONGIPALPIS</i> MIDGUTS.	183
FIGURE 41. COMPARATIVE ABUNDANCE OF ASTACIN IN UNINFECTED AND <i>L. INFANTUM CHAGASI</i> -INFECTED <i>LU. LONGIPALPIS</i> MIDGUTS.	185
FIGURE 42. COMPARATIVE ABUNDANCE OF TRYPSIN <i>TRYP</i> 3 IN UNINFECTED AND <i>L. INFANTUM CHAGASI</i> - INFECTED <i>LU. LONGIPALPIS</i> MIDGUTS.	186
FIGURE 43. COMPARATIVE ABUNDANCE OF ACTIN IN UNINFECTED AND <i>L. INFANTUM CHAGASI</i> -INFECTED <i>LU.</i> <i>LONGIPALPIS</i> MIDGUTS.	188

Table of Tables

TABLE 1: OLD WORLD DISTRIBUTION OF <i>LEISHMANIA</i> AND THE PROVEN OR SUSPECTED VECTORS.....	6
TABLE 2: NEW WORLD DISTRIBUTION OF <i>LEISHMANIA</i> AND THE PROVEN OR SUSPECTED VECTORS	7
TABLE 3: LIST OF <i>PHLEBOTOMUS PAPATASI</i> MIDGUT-SPECIFIC SEQUENCES, CLUSTERS, AND SEQUENCES PER CLUSTER OF cDNA LIBRARIES MADE FROM FLIES SUGAR-FED, BLOOD-FED, AND BLOOD-FED WITH <i>LEISHMANIA MAJOR</i> PARASITES	58
TABLE 4: CLUSTERS OF COMBINED <i>P. PAPATASI</i> MIDGUT cDNA LIBRARIES (SUGAR-FED, BLOOD-FED AND <i>LEISHMANIA MAJOR</i> -INFECTED) OF TRANSCRIPTS WITH HIGH QUALITY SEQUENCES	61
TABLE 5: CLUSTERS OVERREPRESENTED IN THE SUGAR-FED AND BLOOD-FED MIDGUT cDNA LIBRARIES AS DETERMINED BY χ^2 STATISTICAL ANALYSIS.....	80
TABLE 6: CLUSTERS OVERREPRESENTED IN THE BLOOD-FED AND <i>LEISHMANIA MAJOR</i> -INFECTED SAND FLY MIDGUT cDNA LIBRARIES AS DETERMINED BY χ^2 STATISTICAL ANALYSIS	85
TABLE 7: OVERALL EXAMINATION OF THE 5 INDIVIDUAL cDNA LIBRARIES AND THE COMBINED ANALYSIS	105
TABLE 8: PUTATIVE MIDGUT-ASSOCIATED PROTEASES; BEST MATCHED RESULTS AND CORRESPONDING E- VALUES FROM BLAST INQUIRIES OF A GENBANK-DERIVED NON-REDUNDANT PROTEIN DATABASE AND <i>LUTZOMYIA LONGIPALPIS</i> EST DATABASE	108
TABLE 9: PUTATIVE MIDGUT-ASSOCIATED PROTEASES; PUTATIVE FUNCTION AND SEQUENCE DISTRIBUTION CONTRIBUTED FROM EACH cDNA LIBRARY	108
TABLE 10: PUTATIVE MIDGUT-ASSOCIATED PROTEASES; LOCALIZATION, MOLECULAR WEIGHT AND ISOELECTRIC POINT OF PUTATIVE MIDGUT PROTEINS	110
TABLE 11: PUTATIVE MIDGUT-ASSOCIATED PERITROPHIN PROTEINS; BEST MATCHED RESULTS AND CORRESPONDING E-VALUES FROM BLAST INQUIRIES OF A GENBANK-DERIVED NON-REDUNDANT PROTEIN DATABASE AND <i>LUTZOMYIA LONGIPALPIS</i> EST DATABASE.....	124
TABLE 12: PUTATIVE MIDGUT-ASSOCIATED PERITROPHIN PROTEINS; PUTATIVE FUNCTION AND SEQUENCE DISTRIBUTION CONTRIBUTED FROM EACH cDNA LIBRARY	124
TABLE 13: PUTATIVE MIDGUT-ASSOCIATED PERITROPHIN PROTEINS; LOCALIZATION, MOLECULAR WEIGHT AND ISOELECTRIC POINT OF PUTATIVE MIDGUT PROTEINS	124

TABLE 14: PUTATIVE MIDGUT-ASSOCIATED MICROVILLAR PROTEINS; BEST MATCHED RESULTS AND CORRESPONDING E-VALUES FROM BLAST INQUIRIES OF A GENBANK-DERIVED NON-REDUNDANT PROTEIN DATABASE AND <i>LUTZOMYIA LONGIPALPIS</i> EST DATABASE.....	129
TABLE 15: PUTATIVE MIDGUT-ASSOCIATED MICROVILLAR PROTEINS; PUTATIVE FUNCTION AND SEQUENCE DISTRIBUTION CONTRIBUTED FROM EACH cDNA LIBRARY	129
TABLE 16: PUTATIVE MIDGUT-ASSOCIATED MICROVILLAR PROTEINS; LOCALIZATION, MOLECULAR WEIGHT AND ISOELECTRIC POINT OF PUTATIVE MIDGUT PROTEINS	129
TABLE 17: PUTATIVE MIDGUT-ASSOCIATED OXIDATIVE STRESS MOLECULES; BEST MATCHED RESULTS AND CORRESPONDING E-VALUES FROM BLAST INQUIRIES OF A GENBANK-DERIVED NON-REDUNDANT PROTEIN DATABASE AND <i>LUTZOMYIA LONGIPALPIS</i> EST DATABASE.....	133
TABLE 18: PUTATIVE MIDGUT-ASSOCIATED OXIDATIVE STRESS MOLECULES; PUTATIVE FUNCTION AND SEQUENCE DISTRIBUTION CONTRIBUTED FROM EACH CDNA LIBRARY	133
TABLE 19: PUTATIVE MIDGUT-ASSOCIATED OXIDATIVE STRESS MOLECULES; LOCALIZATION, MOLECULAR WEIGHT AND ISOELECTRIC POINT OF PUTATIVE MIDGUT PROTEINS	133
TABLE 20: HOUSE KEEPING AND LOW ABUNDANT TRANSCRIPTS FROM THE MIDGUT OF <i>LUTZOMYIA</i> <i>LONGIPALPIS</i> ; BEST MATCHED RESULTS AND CORRESPONDING E-VALUES FROM BLAST INQUIRIES OF A GENBANK-DERIVED NON-REDUNDANT PROTEIN DATABASE AND <i>LU. LONGIPALPIS</i> EST DATABASE...	137
TABLE 21: HOUSE KEEPING AND LOW ABUNDANT TRANSCRIPTS FROM THE MIDGUT OF <i>LUTZOMYIA</i> <i>LONGIPALPIS</i> ; PUTATIVE FUNCTION AND SEQUENCE DISTRIBUTION CONTRIBUTED FROM EACH CDNA LIBRARY	138
TABLE 22: HOUSE KEEPING AND LOW ABUNDANT TRANSCRIPTS FROM THE MIDGUT OF <i>LUTZOMYIA</i> <i>LONGIPALPIS</i> ; LOCALIZATION, MOLECULAR WEIGHT AND ISOELECTRIC POINT OF PUTATIVE MIDGUT PROTEINS	139
TABLE 23: SEQUENCE DISTRIBUTION ALTERED DURING SUGAR-FEEDING AND BLOOD MEAL DIGESTION; CLUSTERS OVERREPRESENTED IN THE SUGAR-FED, BLOOD-FED AND POST-BLOOD MEAL DIGESTION MIDGUT CDNA LIBRARIES AS DETERMINED BY χ^2 STATISTICAL ANALYSIS	145

TABLE 24: SEQUENCE DISTRIBUTION ALTERED DURING SUGAR-FEEDING AND BLOOD MEAL DIGESTION; CLUSTERS THAT APPEAR OVERABUNDANT, BUT ARE NOT STATISTICALLY SIGNIFICANT BY χ^2 ANALYSIS, IN THE SUGAR-FED, BLOOD-FED AND POST-BLOOD MEAL DIGESTION MIDGUT cDNA LIBRARIES	145
TABLE 25: SEQUENCE DISTRIBUTION ALTERED BY <i>LEISHMANIA INFANTUM CHAGASI</i> ; CLUSTERS OVERREPRESENTED IN THE BLOOD-FED AND BLOOD-FED <i>L. INFANTUM CHAGASI</i> -INFECTED MIDGUT cDNA LIBRARIES AS DETERMINED BY χ^2 STATISTICAL ANALYSIS	147
TABLE 26: SEQUENCE DISTRIBUTION ALTERED BY <i>LEISHMANIA INFANTUM CHAGASI</i> ; CLUSTERS OVERREPRESENTED IN THE POST-BLOOD MEAL DIGESTION AND POST-BLOOD MEAL DIGESTION <i>L.</i> <i>INFANTUM CHAGASI</i> -INFECTED MIDGUT cDNA LIBRARIES AS DETERMINED BY χ^2 STATISTICAL ANALYSIS	147
TABLE 27: SEQUENCE DISTRIBUTION ALTERED BY <i>LEISHMANIA INFANTUM CHAGASI</i> ; LULO ^{TRYP3} APPEARS UNDERREPRESENTED IN THE POST-BLOOD MEAL DIGESTION <i>L. INFANTUM CHAGASI</i> -INFECTED MIDGUT cDNA LIBRARY	147
TABLE 28: TRANSCRIPT ABUNDANCE IN THE cDNA LIBRARIES OF <i>LUTZOMYIA LONGIPALPIS</i>	165

Chapter 1

Introduction

Epidemiology

Leishmaniasis is a spectrum of diseases that are caused by the protozoan parasite belonging to the genus *Leishmania* and are transmitted by various species of sand flies. Specific species of the parasite commonly are associated with different clinical manifestations. Leishmaniasis afflicts 88 countries and affects approximately 12 million people with two million new cases each year; thus, leishmaniasis is becoming a worldwide re-emerging public-health problem [1]. Current epidemiological assessments report 500,000 cases and 80,000 deaths annually due to visceral leishmaniasis [1]. Although globally distributed, most all of the cases of visceral leishmaniasis occur in Bangladesh, India, Nepal, Sudan and Brazil. Cases of cutaneous leishmaniasis occur in Afghanistan, Algeria, Brazil, Iran, Peru, Saudi Arabia and Syria (Figure 1) [2-4].

Clinical aspects of leishmaniasis

The clinical forms of the disease include visceral, cutaneous, mucocutaneous, and diffuse cutaneous leishmaniasis.

Visceral disease occurs after the parasite infects cells of the reticuloendothelial system and there is suppression of specific cell-mediated immunity. Multiplying unchecked by the immune system, both the parasites-and infected cells replicate mainly in spleen, liver and bone marrow, forming granulomas and resulting in serious complications. Visceral leishmaniasis is characterized by weight loss, fever, lymphadenopathy and hepatosplenomegaly, among other numerous conditions. The visceral form of the disease can be caused by *L. donovani*, *L. infantum* and *L. tropica* in the Old World and by *L. infantum chagasi* and *L. amazonensis* in the New World.

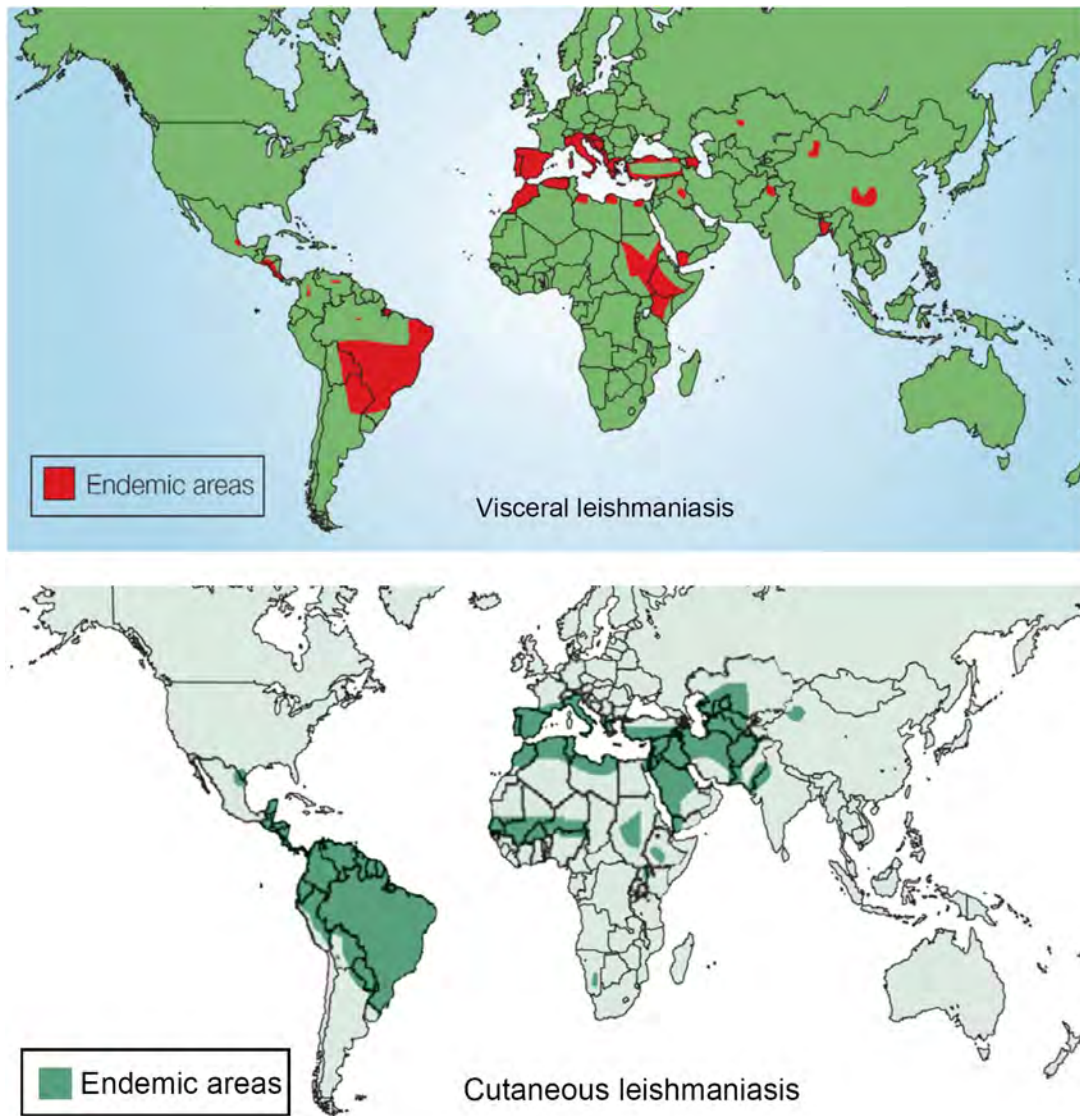


Figure 1. World maps indicating endemic areas of visceral and cutaneous leishmaniasis.

Interestingly, *L. infantum chagasi* is a New World variant of *L. infantum* [5]. Visceral leishmaniasis is a fatal disease if left untreated, and treatment is a lengthy procedure commonly requiring at least 3-4 weeks of drug administration and possibly inpatient monitoring depending upon the drug used [6]. .

Cutaneous leishmaniasis begins as a papule at the sand fly bite site, as the parasite continues to recruit and replicate in macrophages. During this period the cellular response is being mounted or a T-helper 2 cellular response is initiated, both of which are ineffective at activating the leishmanicidal activity of infected macrophages. The infection can develop to an ulcerated lesion if the T-helper 1 cellular response is vigorous enough to cause adjacent tissue damage once recruited macrophages are activated to kill the *Leishmania* amastigotes. Diffuse cutaneous leishmaniasis is a variation of common cutaneous leishmaniasis, which results in a numerous skin lesions due to an inefficient cell-mediated immune response to the parasite. Cutaneous leishmaniasis can be caused by numerous species of *Leishmania* parasite and includes *L. major*, *L. tropica*, and *L. aethiopica* in the Old World and *L. mexicana*, *L. braziliensis*, and *L. peruviana* in the New World [6]. Diffuse cutaneous leishmaniasis is more commonly associated with *L. aethiopica* and *L. mexicana* parasites. Mucosal leishmaniasis is a South American disease most commonly caused by *L. braziliensis* and is characterized by mucosal ulcerations and disfigurement [4]. Most cutaneous lesions will resolve over time, leaving a scar; however, treatment for persistent lesions and to reduce scarring can consist of paromomycin ointment, antimony, or heat therapy [4]. Diffuse cutaneous and mucosal leishmaniases are more difficult to treat and immunotherapy may be applied in conjunction with amphotericin or antimony treatment. These are examples of the more

commonly diagnosed leishmaniasis pathogens; however, the number of parasite species and the disease afflictions that they can cause are far more numerous. The number of species of *Leishmania* capable of causing human disease is greater than any other parasitic infection, and a complementary complexity exists in the number of sand fly species that are proven or possible insect vectors of *Leishmania* parasites.

Sand flies

Sand flies are small Dipteran insects of the Family Psychodidae and Subfamily Phlebotominae, which require a blood meal from a vertebrate host for egg production. It is during the acquisition of the blood meal that there is the potential for disease transmission. *Phlebotomus* and *Lutzomyia* are the most important genera of anthropophilic sand flies, as they are vectors of human and animal diseases. Of the approximately 700 species of sand flies, about 80 have been implicated as disease vectors. Some of the sand flies implicated in transmitting *Leishmania* have not been fully incriminated as definitive vectors (Table 1,2) [7-11]. A large research effort at the turn of the 19th century was instrumental to incriminate sand fly transmission of *L. tropica* (1941) and *L. infantum chagasi* (1977) [8, 12]. Figure 2 is a historical timeline of events leading to the eventual definitive incrimination of sand flies as *Leishmania* vectors in both the New World and the Old World [13-27]. The work of those pioneering individuals, spanning nearly 100 years, set the foundation of our current understanding of the inherent complexity in the epidemiology and basic biology of the transmission of *Leishmania* parasites.

Table 1: Old World distribution of *Leishmania* and the proven or suspected vectors

Parasite (Disease) ¹	Sand fly species ²	Sand fly distribution ³
<i>Leishmania aethiopica</i> (LCL, DCL)	<i>Phlebotomus longipes</i> <i>P. pedifer</i> <i>P. sergenti</i>	Kenya, Ethiopia Kenya, Ethiopia Ethiopia
<i>L. donovani</i> (LCL, PKDL, VL)	<i>P. alexandri</i> <i>P. argentipes</i> <i>P. celiae</i> <i>P. chinensis</i> <i>P. longiductus</i> <i>P. martini</i> <i>P. mongolensis</i> <i>P. orientalis</i>	N. Africa to W. China Bangladesh, Nepal, India Kenya, S. Ethiopia Northern, Central China India E. Africa, Ethiopia Central Asia Sudan, Ethiopia, Saudi Arabia, Yemen
<i>L. infantum</i> (LCL, VL)	<i>P. ariasi</i> <i>P. brevis</i> <i>P. chinensis</i> <i>P. halepensis</i> <i>P. kandelakii</i> <i>P. langeroni</i> <i>P. longicuspis</i> <i>P. neglectus</i> <i>P. perfilliewi</i> <i>P. perniciosus</i> <i>P. sichuanensis</i> <i>P. smirnovi</i> <i>P. tobbi</i> <i>P. transcaucasicus</i>	Western Mediterranean Northern Iran to Caucasus Northern, Central China Jordan, Lebanon, Iraq Iran, Afghanistan Egypt-Tunisia North Africa North Africa, Central Asia Mediterranean Basin, Algeria Western Mediterranean China Central Asia Eastern Mediterranean Caucasus
<i>L. killicki</i> (LCL)	<i>P. alexandri</i> <i>P. chaubaudi</i> <i>P. papatasi</i>	Central Tunisia Central Tunisia Central Tunisia
<i>L. major</i> (LCL)	<i>P. alexandri</i> <i>P. ansarii</i> <i>P. caucasicus</i> <i>P. duboscqi</i> <i>P. papatasi</i> <i>P. salehi</i>	North Africa to Western China Iran Iran Sahelian Africa, Kenya N. Africa, Middle East Iran, Pakistan
<i>L. tropica</i> (LCL, VL)	<i>P. aculeatus</i> <i>P. guggisbergi</i> <i>P. halepensis</i> <i>P. longiductus</i> <i>P. sergenti</i>	Kenya Kenya South Caucasus Eurasia India Middle East, North Africa

¹Abbreviations: LCL, localized cutaneous leishmaniasis; DCL, diffuse cutaneous leishmaniasis; PKDL, post kala-azar dermal leishmaniasis; VL, visceral leishmaniasis

²Species names which are bolded are considered proven, although not necessarily fully incriminated, vectors. Other species are implicated as vectors of leishmaniasis based on fewer incriminating factors.

³Sand fly distribution listed here is in correlation with the distribution of the respective leishmaniasis.

Table 2: New World distribution of *Leishmania* and the proven or suspected vectors

Parasite (Disease) ¹	Sand fly species ²	Sand fly distribution ³
<i>Leishmania</i> (L.) <i>amazonensis</i> (LCL, DCL, VL)	<i>Lutzomyia flaviscutellata</i> <i>Lu. olmeca nociva</i>	Northern South America Amazon basin
<i>L. (V.) braziliensis</i> (LCL, MCL)	<i>Lu. amazonensis</i> <i>Lu. ayrozai</i> <i>Lu. carrerai</i> <i>Lu. complexa</i> <i>Lu. intermedia</i> <i>Lu. llanomartinsi</i> <i>Lu. migonei</i> <i>Lu. ovallesi</i> <i>Lu. panamensis</i> <i>Lu. paraensis</i> <i>Lu. pessoai</i> <i>Lu. spinicrassa</i> <i>Lu. trinidadensis</i> <i>Lu. wellcomei</i> <i>Lu. whitmani</i> <i>Lu. yucumensis</i>	Northern Amazon basin Southeastern Brazil Western Amazon basin Para, Brazil Southern Brazil Brazil Brazil, Venezuela Guatemala, Venezuela Central, N. South America N. South America Southern Brazil Colombia Venezuela Para, Brazil Eastern Brazil Bolivia
<i>L. (L.) chagasi</i> (LCL, VL)	<i>Lu. evansi</i> <i>Lu. longipalpis</i>	Colombia Central and South America
<i>L. (V.) garnhami</i> (LCL)	<i>Lu. youngi</i>	Venezuela
<i>L. (V.) guayanensis</i> (LCL, MCL)	<i>Lu. anduzei</i> <i>Lu. umbratilis</i>	N. South America Amazon basin
<i>L. (V.) lainsoni</i> (LCL MCL)	<i>Lu. ubiquitatis</i>	Amazon basin
<i>L. (L.) mexicana</i> (LCL, DCL)	<i>Lu. anthophora</i> <i>Lu. ayacuchensis</i> <i>Lu. diabolica</i> <i>Lu. olmeca olmeca</i> <i>Lu. ylephiletor</i>	Southern Texas, USA Ecuador Southern Texas, USA Central America Guatemala
<i>L. (V.) lindenbergi</i> (CL)	<i>Lu. antunesi</i>	Para, Brazil
<i>L. (V.) naiffi</i> (CL)	<i>Lu. squamiventris</i>	Brazil
<i>L. (V.) panamensis</i> (LCL, MCL)	<i>Lu. gomezi</i> <i>Lu. panamensis</i> <i>Lu. trapidoi</i> <i>Lu. ylephiletor</i>	Central, N. South America Central, N. South America Central America Central America
<i>L. (V.) peruviana</i> (LCL)	<i>Lu. ayacuchensis</i> <i>Lu. peruensis</i> <i>Lu. tejadai</i> <i>Lu. verrucarum</i>	Peru Peru Peru Peru
<i>L. (L.) pifanoi</i> (CL)	<i>Lu. flaviscutellata</i>	N. South America
<i>L. (V.) shawi</i> (CL)	<i>Lu. whitmani</i>	Brazil
<i>L. (L.) venezuelensis</i> (CL)	<i>Lu. olmeca</i>	N. South America

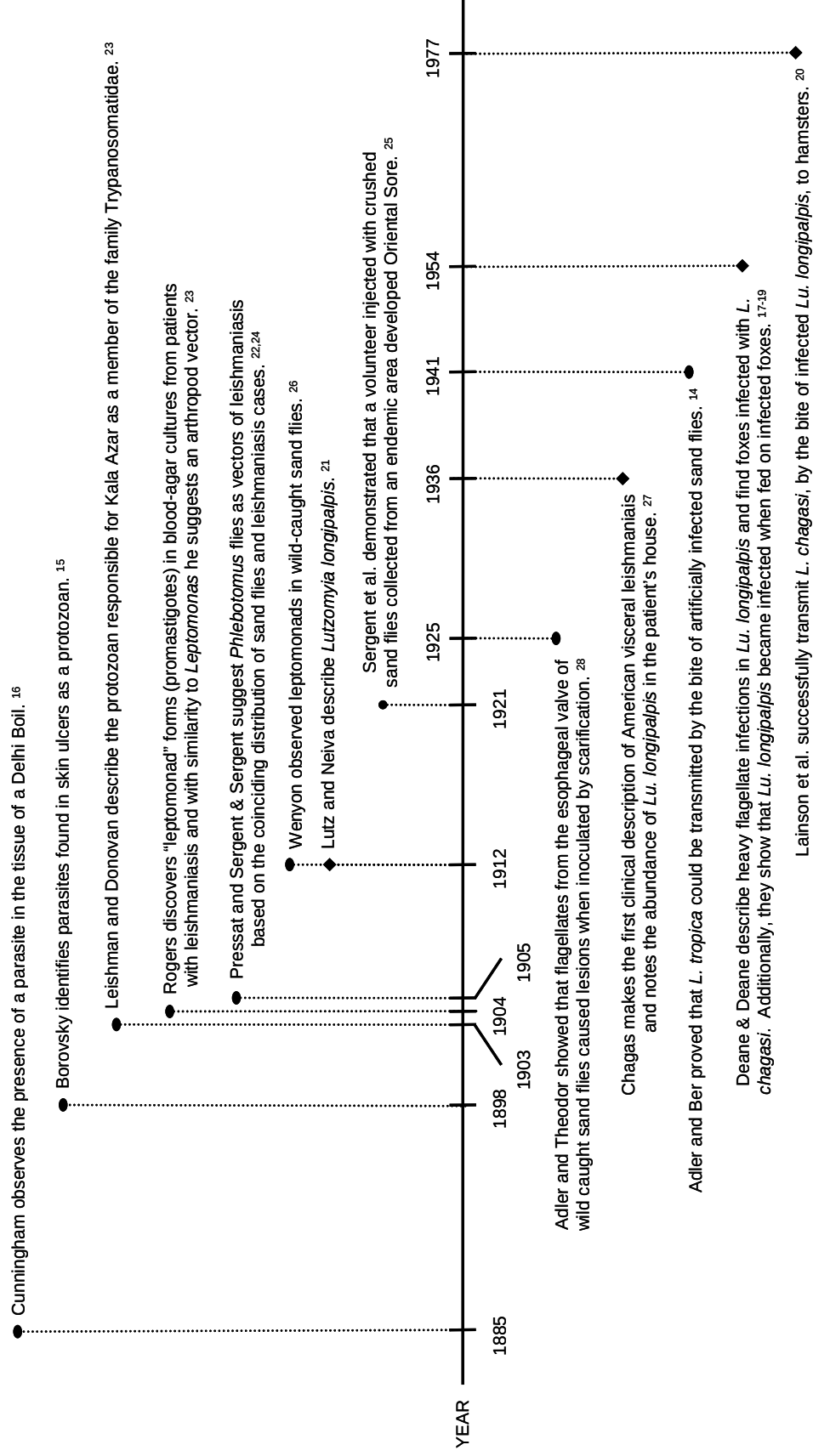
¹Abbreviations: (L.), subgenus *Leishmania*; (V.), subgenus *Viannia*; LCL, localized cutaneous leishmaniasis; DCL, diffuse cutaneous leishmaniasis; MCL, mucocutaneous leishmaniasis; VL, visceral leishmaniasis

²Species names in bolded are considered proven, although not necessarily fully incriminated, vectors. Other species are implicated as vectors of leishmaniasis based on fewer incriminating factors.

³Sand fly distribution listed here is in correlation with the distribution of the respective leishmaniasis.

Figure 2. Timeline of events leading to the incrimination of New World and Old World sand fly vectors of leishmaniasis.

Circles denote Old World sand fly findings and diamonds indicate New World sand fly research.



Leishmania life cycle

Leishmania has a two-stage life cycle; one in the mammalian host as an intracellular amastigote, and one in the sand fly vector as an extracellular motile promastigote. After a vertebrate host receives a bite from an infectious sand fly, the inoculated metacyclic promastigote parasite invades host macrophages and multiplies within the phagolysosome. If an uninfected sand fly feeds on an infectious host, it ingests a blood meal containing amastigote-infected macrophages; thus, beginning the life cycle in the invertebrate host (Figure 3.1). Amastigotes are released after rupture of the macrophage, and the parasite begins the developmental cycle to the first flagellated form of the parasite within the sand fly, the procyclic promastigote (Figure 3.2).

The proliferation and differentiation of the first parasite stages occurs within the peritrophic matrix (PM), a proteo-chitin structure formed to encapsulate the blood meal immediately after feeding. The PM offers a relatively protected environment for the *Leishmania* during the first hours after the blood meal, as the amastigote is susceptible to killing by digestive enzymes [28]. The procyclic form of the promastigote is the first multiplying stage of the parasite within the sand fly and forms rosettes of replicating parasites within the PM as the blood meal is digested [29]. Parasites of the subgenus *Leishmania* develop only in the midgut and foregut (suprapylarian); whereas, parasites of the subgenus *Viannia* have a developmental stage occurring in the hindgut (peripylarian) [30, 31]. The outer surface of procyclic promastigotes is covered in a dense layer of lipophosphoglycans (LPG), a glycoconjugate that has multiple functions [32]. LPG has been shown to determine sand fly species restriction and is the ligand necessary for parasite attachment to the midgut epithelium in what are usually

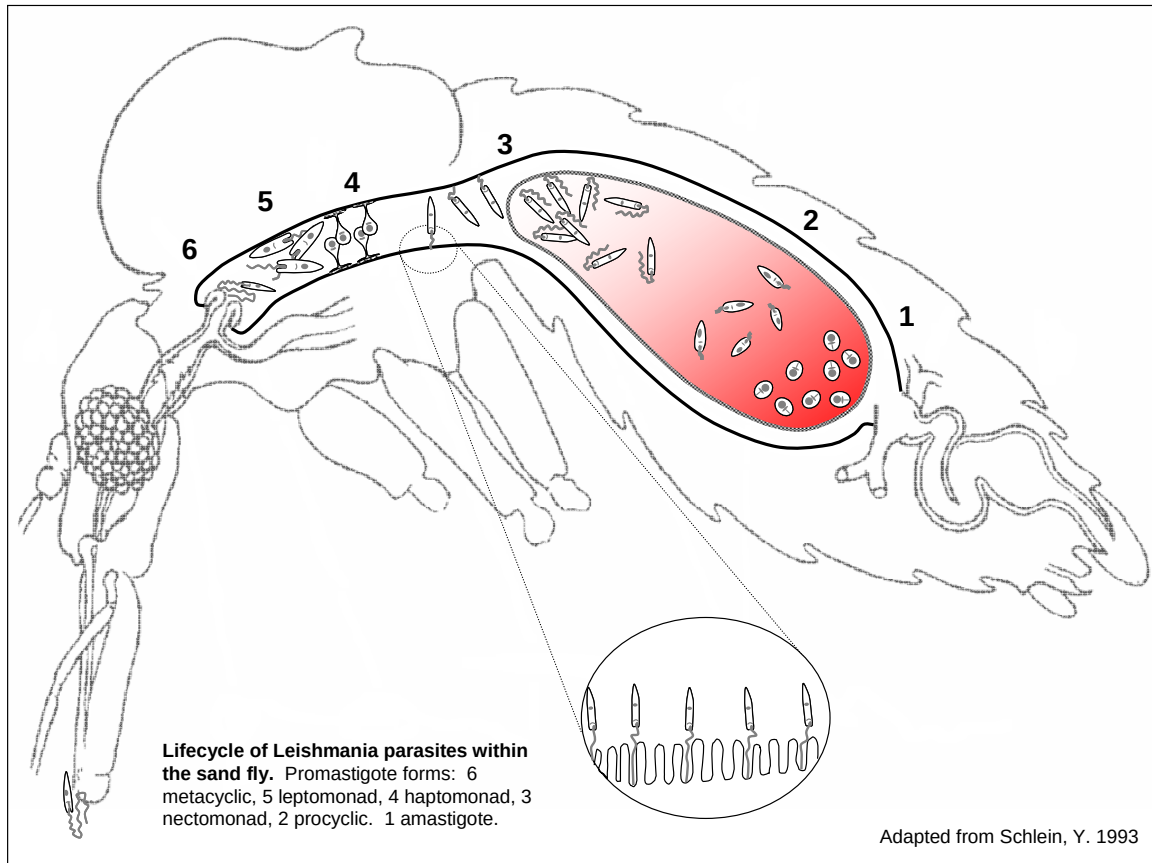


Figure 3. Life cycle of *Leishmania* parasites within the sand fly.

referred to as restrictive vectors, such as *P. sergenti* and *P. papatasi* [33, 34]. Sand flies that are considered permissive (they are able to harbor several different species of *Leishmania*) have been shown to possess a LPG-independent mechanism of *Leishmania* retention [35]. The next developmental form is the nectomonad promastigote (Figure 3.3), a non-dividing and highly motile stage that is most abundant about three days after the sand fly ingests the infected blood meal. In *Lu. longipalpis* infected with *L. mexicana*, the nectomonad promastigotes migrate to the anterior area of the PM; whereas, the procyclic promastigotes migrate to the posterior area of the PM [36]. At this point in development the nectomonad promastigotes must escape the PM and attach to the midgut epithelium (Figure 3.3). Escape from the PM is facilitated by a *Leishmania*-derived secretory chitinase and likely by the sand fly midgut chitinase [37, 38]. Electron microscopy has shown that the flagellum of nectomonad promastigotes appear to be inserted between the microvilli of the midgut epithelium and that there are even instances of flagella entering degenerating epithelial cells [30]. The exact role or outcome of this interaction is not clear; however, it is known that the parasite must bind to the midgut to prevent elimination during the defecation of the digested blood meal. The next parasite form in development is the leptomonad promastigote (Figure 3.5); the role of which is to expand the parasite population within the thoracic midgut and produce promastigote secretory gel (PSG), an important component in transmission [36]. Haptomonads, a promastigote that is likely derived from the leptomonad, bind to the cuticular surface of the stomodeal valve by a hemidesmosome-like interaction of the flagellum (Figure 3.4) [30]. Leptomonads also give rise to the infective form of parasite, the metacyclic promastigote (Figure 3.6). The metacyclic promastigote is a smaller, highly motile form

of the parasite that is covered in an altered configuration of LPG side chains, negating attachment to the midgut epithelium [39]. It is generally regarded as necessary for the parasite to develop to the metacyclic promastigote for transmission to occur to a mammalian host, as this form of the promastigote is not only the most infective but also greatly resists killing by normal human serum [40, 41].

The two main theories of transmission are direct inoculation via the proboscis with promastigotes originating from the anterior foregut, or regurgitation of promastigotes, which brings promastigotes from behind the pharynx from the posterior foregut and anterior midgut [42]. While the exact mechanics of transmission are still under debate, recent work has shown that about 10 times the number of promastigotes are egested during feeding than are present in the foregut, implicating regurgitation as the principal mechanism of transmission [43]. Two theories explaining the manner in which parasites are regurgitated are the formation of a plug and *Leishmania*-induced damage to the stomodeal valve. One proposed cause of regurgitation of parasites during blood-feeding was based on observations of damage to the cuticle of the stomodeal valve in *L. major*-infected *P. papatasi*, linked to a chitinase enzyme produced by the parasite [44]. The stomodeal valve is a sphincter muscle-controlled ring of cuticular epithelia that separates the cardia from the pharynx. The proposed theory is that damage to the stomodeal valve allows a flushing and regurgitation of the contents of the cardia into the pharynx, through the proboscis, and then into the bite site [44]. Additionally, this has been repeatedly observed in a number of sand fly species infected with different species of *Leishmania* [44, 45]. Prior to this observation, it was believed that the *Leishmania* parasites caused a blockage in the foregut, which caused regurgitation in a manner

synonymous with the transmission of *Yersinia pestis* by fleas. This older theory was reinforced by the discovery of a gel-like plug in the anterior midgut consisting of metacyclic promastigotes and promastigote secretory gel (PSG), a substance composed primarily of filamentous proteophosphoglycan [43]. Additionally, PSG was shown to greatly enhance cutaneous leishmaniasis lesion size and persistence; an attribute shared with molecules of sand fly saliva [25].

Sand flies that become infected with *Leishmania* parasites that then replicate and are successfully transmitted are considered competent vectors. Within the sand fly midgut, there are abundant molecular interactions that are the determinants of species-specific vector competence and include resisting digestive enzymes, escaping from the PM and binding to the midgut epithelium. The earliest obstacle faced by the recently ingested parasite is the attack by proteolytic enzymes. There are numerous proteases, including trypsins, chymotrypsins, amino- and carboxypeptidases, within the midgut lumen that are induced by blood-feeding, facilitate blood meal digestion and likely confer some immunity to ingested organisms. The presence of *Leishmania* parasites in the midgut lumen of sand flies infected using promastigotes changes the overall protease activity, inhibiting or delaying proteolytic enzymes [46, 47]. Infections initiated using amastigotes, a more natural representation of the infection route, caused a delay in trypsin and aminopeptidase activity [48]. A delay in protease activity would allow the ingested amastigotes to transform to the LPG-covered, and thus more protease resistant, promastigote form of the parasite. Until recently, it has been unclear which specific proteolytic molecules are up- or downregulated by the presence of the parasite within the sand fly midgut.

Further protection is offered by the peritrophic matrix during the first few hours after the blood meal is ingested. The addition of an exogenous chitinase, which prevented PM formation, resulted in high parasite death despite the low abundance of proteases immediately following the blood meal [28]. While the PM offers protection early in the establishment of the parasite within the midgut, failure of the *Leishmania* to escape the PM results in the parasite being defecated along with the digested blood meal. In uninfected *P. papatasi*, the posterior area of the PM begins to break down, likely attributed to intrinsic sand fly midgut chitinase, about four days post-blood meal and is excreted with the digested products. *Phlebotomus papatasi* infected with *L. major* shows break down of the PM at 48 hours post-blood meal in the posterior area and additionally in the anterior area, where there is a congregation of parasites [37]. Further evidence that escape from the PM requires a *Leishmania*-specific competence factor was demonstrated by the non-natural infection of *P. papatasi* with *L. panamensis*, which led to the riddance of the promastigotes enclosed in the PM when the sand fly excreted the digested blood meal — demonstrating a species-specific developmental barrier [49]. With chitinase genes identified and the molecules characterized in *Leishmania* and the sand fly midgut, determining the role of chitinolytic enzymes in species-specific vector competence would be best described using knockout mutant strains of parasites and RNA knockdown in sand flies. Additionally, it is prudent to theorize that the *Leishmania* parasite may influence the sand fly chitinase molecule and peritrophins, a protein component of the PM, in an effort to bypass this physical barrier prior to binding to the midgut epithelium.

Once free of the PM, the nectomonad promastigotes must adhere to the midgut epithelium to complete development within the sand fly host. The binding must occur

prior to defecation of the digested blood or the majority of parasites will be lost in the excrement and the sand fly may not become infective. It has been well described that lipophosphoglycan (LPG) on the surface of the *Leishmania* is responsible for the binding of parasites to the midgut microvilli in certain species of *Leishmania*-sand fly pairings. It was first documented that specific oligosaccharides on the phosphoglycan of *L. major* procyclic promastigotes could bind the midgut of *P. papatasi* [50]. A similar study demonstrated stage-specific binding of *L. donovani* in *P. argentipes* and used purified LPG from procyclics to show inhibition of procyclic binding; whereas, purified LPG from metacyclic parasites did not inhibit procyclic binding to the midgut [39]. Subsequently, the receptor for *L. major* in *P. papatasi* was identified as a galectin, PpGalec, binding specifically to the galactose residues of LPG [34]. Generation of an LPG deficient mutant of *L. major* showed the importance of parasite binding, mediated by LPG, in the development of a transmissible parasitemia in *P. papatasi* [51]. Recent work using LPG deficient *L. major* infections in the permissive vectors *Lu. longipalpis* and *P. arabicus* demonstrated that LPG is not required for the development of heavy promastigote infections in these sand flies [35]. Additional work generated the theory that GalNac-containing glycoproteins on the midgut epithelia are the ligands to which a parasite lectin receptor binds, retains and allows for full development of several different *Leishmania* species in permissive sand flies [35].

Sand fly saliva

When the sand fly bites a host, a cocktail of pharmacologically active molecules within the salivary glands are injected into the skin of the host, which facilitates

successful blood meal acquisition. The saliva cocktail contains molecules exhibiting anti-platelet, vasodilator, anticoagulant, anti-inflammatory and immunomodulatory activities [52-57]. Researching the saliva of hematophagous insects has improved the understanding of the evolution of blood-feeding, and more importantly, the impact of insect saliva on pathogen transmission and disease progression. During the transmission of *Leishmania* parasites by the bite of a sand fly, there is the ubiquitous co-inoculation of infective stage metacyclic promastigotes and saliva. The experiments of Titus and Ribeiro (1988), which demonstrated the enhancement of cutaneous leishmaniasis by the co-injection of *L. major* with *Lu. longipalpis* salivary gland homogenate, served as a stepping stone for further research on *Leishmania* susceptibility and resistance conferred by sand fly saliva [58]. The exacerbation of leishmaniasis by saliva may be a co-evolutionary mechanism that is essential for the propagation and maintenance of parasite populations in the wild.

Inhabitants of endemic areas show resilience to infection in comparison to individuals from non-endemic areas who acquire leishmaniasis when entering endemic areas. Converse to enhancing disease, pre-exposure to sand fly saliva and inoculation with specific salivary proteins confers protection against disease. This demonstrates the potential use of saliva-based vaccines against leishmaniasis [59-61]. The outcome of saliva-mediated disease exacerbation or protection is believed to be a product of immune modulation by specific salivary molecules. Saliva-based protection is characterized by the development of delayed-type hypersensitivity at the bite site with increased interferon- γ and interleukin-12, T-helper type 1 cellular immune response indicators [61]. A high-throughput approach to identifying potential salivary-based vaccine candidates

was developed, implementing the sequencing of salivary gland transcriptomes, proteomics and reverse antigen screening [62].

Premise

The use of functional transcriptomics (the exploitation of the wealth of knowledge produced by the massive amounts of cDNA sequencing data in an effort to formulate and test biologically pertinent hypotheses) can be instrumental in the identification of vaccine candidates for leishmaniasis as well as novel insect salivary molecules with potential pharmaceutical applications. In Chapter 2, I demonstrate the application of functional transcriptomics in the biochemical characterization of an adenosine deaminase (ADA) molecule in *Phlebotomus duboscqi* saliva, identified by the surprising presence of ADA sequences in a salivary gland transcriptome. Furthermore, the complex interactions occurring between the *Leishmania* parasite and the sand fly midgut go beyond simple ligand binding. Just as amastigotes invade and sabotage the microbicidal activity of the host macrophage, there are, presumably, interactions between vector and parasite surface and secreted molecules that influence *Leishmania* colonization and maturation within the sand fly. I propose that *Leishmania* colonization of the midgut affects the transcript abundance within the midgut tissue of the sand fly vector, whether due to subversion of normal physiological conditions of the midgut tissue in the sand fly for the benefit of the parasite or a response by the sand fly to the presence of *Leishmania* colonizing the alimentary canal. Functional transcriptomic methodologies were employed in the construction and comparative analysis of *P. papatasi* and *Lu. longipalpis* midgut cDNA libraries, as described in Chapters 3 and 4. Chapter 5 further addresses the influence of

L. infantum chagasi on midgut transcript abundance in *Lu. longipalpis* and provides insights into the molecular interactions between the parasite and the vector.

References

1. Desjeux P: **Leishmaniasis: current situation and new perspectives.** *Comp Immunol Microbiol Infect Dis* 2004, **27**(5):305-318.
2. Desjeux P: **Leishmaniasis. Public health aspects and control.** *Clin Dermatol* 1996, **14**(5):417-423.
3. Desjeux P: **Leishmaniasis.** *Nat Rev Microbiol* 2004, **2**(9):692.
4. Reithinger R, Dujardin JC, Louzir H, Pirmez C, Alexander B, Brooker S: **Cutaneous leishmaniasis.** *Lancet Infect Dis* 2007, **7**(9):581-596.
5. Mauricio IL, Stothard JR, Miles MA: **The strange case of *Leishmania chagasi*.** *Parasitol Today* 2000, **16**(5):188-189.
6. Piscopo TV, Mallia AC: **Leishmaniasis.** *Postgrad Med J* 2006, **82**(972):649-657.
7. Caceres AG, Villaseca P, Dujardin JC, Banuls AL, Inga R, Lopez M, Arana M, Le Ray D, Arevalo J: **Epidemiology of Andean cutaneous leishmaniasis: incrimination of *Lutzomyia ayacuchensis* (Diptera: Psychodidae) as a vector of *Leishmania* in geographically isolated, upland valleys of Peru.** *Am J Trop Med Hyg* 2004, **70**(6):607-612.
8. Lawyer PG, Perkins PV: **Leishmaniasis and Trypanosomiasis.** Norwell: Kluwer Academic Publishers; 2000.

9. Sadlova J, Hajmova M, Volf P: ***Phlebotomus (Adlerius) halepensis* vector competence for *Leishmania major* and *Le. tropica***. *Med Vet Entomol* 2003, **17**(3):244-250.
10. Sharma NL, Mahajan VK, Kanga A, Sood A, Katoch VM, Mauricio I, Singh CD, Parwan UC, Sharma VK, Sharma RC: **Localized cutaneous leishmaniasis due to *Leishmania donovani* and *Leishmania tropica*: preliminary findings of the study of 161 new cases from a new endemic focus in Himachal Pradesh, India**. *Am J Trop Med Hyg* 2005, **72**(6):819-824.
11. Silveira FT, Ishikawa EA, De Souza AA, Lainson R: **An outbreak of cutaneous leishmaniasis among soldiers in Belem, Para State, Brazil, caused by *Leishmania (Viannia) lindenbergi* n. sp. A new leishmanial parasite of man in the Amazon region**. *Parasite* 2002, **9**(1):43-50.
12. Lainson R, Rangel EF: ***Lutzomyia longipalpis* and the eco-epidemiology of American visceral leishmaniasis, with particular reference to Brazil: a review**. *Mem Inst Oswaldo Cruz* 2005, **100**(8):811-827.
13. Adler S, Ber M: **The transmission of *Leishmania tropica* by the bite of *Phlebotomus papatasi***. *Ind J Med Res* 1941, **29**:803-809.
14. Borovsky PE: **Sart Sore**. *Voenne-medicinskj zurnal (Military Medical Journal)* 1989, **195**(925).
15. Cunningham DD: **The presence of peculiar parasitic organism in the tissue of a speciment of Delhi Boil**. *Sci Mem Med Off Army* 1885, **1**(21).

16. Deane LM, Deane MP: **Encontro de leishmanias nas vísceras e na pele de uma raposa, em zona endêmica de calazar, nos arredores de Sobral, Ceará.**
Hospital 1954, **45**:419-421.
17. Deane MP, Deane LM: **Infecção natural do *Phlebotomus longipalpis* por leptomonas, provavelmente de *Leishmania*, em foco de calazar, no Ceará.**
Hospital 1954, **45**:697-702.
18. Deane MP, Deane LM: **Infecção experimental do *Phlebotomus longipalpis* em raposa (*Lycalopex vetulus*) naturalmente parasitada pela *Leishmania donovani*.** *Hospital* 1954, **46**:651-653.
19. Lainson R, Ward RD, Shaw JJ: **Experimental transmission of *Leishmania chagasi* causative agent of neotropical visceral leishmaniasis, by the sandfly *Lutzomyia longipalpis*.** *Nature* 1977, **266**:628-630.
20. Lutz A, Neiva A: **Contribuição para o conhecimento das espécies do genero *Phlebotomus* existentes no Brasil.** *Mem Inst Oswaldo Cruz* 1912, **4**:84-95.
21. Pressat A: **La paludism et les moustiques (prophylaxie).** 1905.
22. Rogers L: **On the development of flagellated organisms (trypanosomes) from the spleen protozoic parasites of cachexial fevers and kala-azar.** 1904.
23. Sergent E, Sergent E: **Sur un culicide nouveau, tres commun a Biskra.** *C R Soc Biol* 1905, **57**:673-674.
24. Sergent E, Sergent E, Parrott L, Donatien L, Beguet M: **Transmission du clou de Biskra Par le Phlebotome (*Phlebotomus papatasi*, Scop.).** *C R Acad Sci* 1921, **173**:1030-1032.
25. Wenyon CM: *J Lond Sch Trop Med* 1912, **1**:98.

26. Chagas E: **Primeira verificação em individuo vivo, da leishmaniose visceral no Brasil.** *Bras-Méd* 1936, **50**:221-222.
27. Adler S, Theodor O: **The experimental transmission of cutaneous leishmaniasis to man from *Phlebotomus papatasi*.** *Ann Trop Med Parasitol* 1925, **19**:365-371.
28. Pimenta PF, Modi GB, Pereira ST, Shahabuddin M, Sacks DL: **A novel role for the peritrophic matrix in protecting *Leishmania* from the hydrolytic activities of the sand fly midgut.** *Parasitology* 1997, **115 (Pt 4)**:359-369.
29. Gossage SM, Rogers ME, Bates PA: **Two separate growth phases during the development of *Leishmania* in sand flies: implications for understanding the life cycle.** *Int J Parasitol* 2003, **33**(10):1027-1034.
30. Killick-Kendrick R, Molyneux DH, Ashford RW: ***Leishmania* in phlebotomid sandflies. I. Modifications of the flagellum associated with attachment to the mid-gut and oesophageal valve of the sandfly.** *Proc R Soc Lond B Biol Sci* 1974, **187**(1089):409-419.
31. Killick-Kendrick R, Molyneux DH, Hommel M, Leaney AJ, Robertson ES: ***Leishmania* in phlebotomid sandflies. V. The nature and significance of infections of the pylorus and ileum of the sandfly by leishmaniae of the braziliensis complex.** *Proc R Soc Lond B Biol Sci* 1977, **198**(1131):191-199.
32. Descoteaux A, Turco SJ: **Glycoconjugates in *Leishmania* infectivity.** *Biochim Biophys Acta* 1999, **1455**(2-3):341-352.
33. Kamhawi S, Modi GB, Pimenta PF, Rowton E, Sacks DL: **The vectorial competence of *Phlebotomus sergenti* is specific for *Leishmania tropica* and is**

- controlled by species-specific, lipophosphoglycan-mediated midgut attachment.** *Parasitology* 2000, **121 (Pt 1)**:25-33.
34. Kamhawi S, Ramalho-Ortigao M, Pham VM, Kumar S, Lawyer PG, Turco SJ, Barillas-Mury C, Sacks DL, Valenzuela JG: **A role for insect galectins in parasite survival.** *Cell* 2004, **119**(3):329-341.
 35. Myskova J, Svobodova M, Beverley SM, Volf P: **A lipophosphoglycan-independent development of Leishmania in permissive sand flies.** *Microbes Infect* 2007, **9**(3):317-324.
 36. Rogers ME, Chance ML, Bates PA: **The role of promastigote secretory gel in the origin and transmission of the infective stage of *Leishmania mexicana* by the sandfly *Lutzomyia longipalpis*.** *Parasitology* 2002, **124**(Pt 5):495-507.
 37. Schlein Y, Jacobson RL, Shlomai J: **Chitinase secreted by *Leishmania* functions in the sandfly vector.** *Proc Biol Sci* 1991, **245**(1313):121-126.
 38. Ramalho-Ortigao JM, Kamhawi S, Joshi MB, Reynoso D, Lawyer PG, Dwyer DM, Sacks DL, Valenzuela JG: **Characterization of a blood activated chitinolytic system in the midgut of the sand fly vectors *Lutzomyia longipalpis* and *Phlebotomus papatasi*.** *Insect Mol Biol* 2005, **14**(6):703-712.
 39. Sacks DL, Pimenta PF, McConville MJ, Schneider P, Turco SJ: **Stage-specific binding of *Leishmania donovani* to the sand fly vector midgut is regulated by conformational changes in the abundant surface lipophosphoglycan.** *J Exp Med* 1995, **181**(2):685-697.
 40. da Silva R, Sacks DL: **Metacyclogenesis is a major determinant of *Leishmania* promastigote virulence and attenuation.** *Infect Immun* 1987, **55**(11):2802-2806.

41. Howard MK, Sayers G, Miles MA: ***Leishmania donovani* metacyclic promastigotes: transformation in vitro, lectin agglutination, complement resistance, and infectivity.** *Exp Parasitol* 1987, **64**(2):147-156.
42. Bates PA, Rogers ME: **New insights into the developmental biology and transmission mechanisms of *Leishmania*.** *Curr Mol Med* 2004, **4**(6):601-609.
43. Rogers ME, Ilg T, Nikolaev AV, Ferguson MA, Bates PA: **Transmission of cutaneous leishmaniasis by sand flies is enhanced by regurgitation of fPPG.** *Nature* 2004, **430**(6998):463-467.
44. Schlein Y, Jacobson RL, Messer G: ***Leishmania* infections damage the feeding mechanism of the sandfly vector and impede parasite transmission by bite.** *Proc Natl Acad Sci U S A* 1992, **89**(20):9944-9948.
45. Volf P, Hajmova M, Sadlova J, Votypka J: **Blocked stomodeal valve of the insect vector: similar mechanism of transmission in two trypanosomatid models.** *Int J Parasitol* 2004, **34**(11):1221-1227.
46. Borovsky D, Schlein Y: **Trypsin and chymotrypsin-like enzymes of the sandfly *Phlebotomus papatasi* infected with *Leishmania* and their possible role in vector competence.** *Med Vet Entomol* 1987, **1**(3):235-242.
47. Schlein Y, Romano H: ***Leishmania major* and *L. donovani*: effects on proteolytic enzymes of *Phlebotomus papatasi* (Diptera, Psychodidae).** *Exp Parasitol* 1986, **62**(3):376-380.
48. Dillon RJ, Lane RP: **Influence of *Leishmania* infection on blood-meal digestion in the sandflies *Phlebotomus papatasi* and *P. langeroni*.** *Parasitol Res* 1993, **79**(6):492-496.

49. Walters LL, Irons KP, Modi GB, Tesh RB: **Refractory barriers in the sand fly *Phlebotomus papatasi* (Diptera: Psychodidae) to infection with *Leishmania panamensis*.** *Am J Trop Med Hyg* 1992, **46**(2):211-228.
50. Pimenta PF, Turco SJ, McConville MJ, Lawyer PG, Perkins PV, Sacks DL: **Stage-specific adhesion of *Leishmania* promastigotes to the sandfly midgut.** *Science* 1992, **256**(5065):1812-1815.
51. Butcher BA, Turco SJ, Hilty BA, Pimenta PF, Panunzio M, Sacks DL: **Deficiency in beta1,3-galactosyltransferase of a *Leishmania major* lipophosphoglycan mutant adversely influences the *Leishmania*-sand fly interaction.** *J Biol Chem* 1996, **271**(34):20573-20579.
52. Charlab R, Valenzuela JG, Rowton ED, Ribeiro JM: **Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*.** *Proc Natl Acad Sci U S A* 1999, **96**(26):15155-15160.
53. Lerner EA, Shoemaker CB: **Maxadilan. Cloning and functional expression of the gene encoding this potent vasodilator peptide.** *J Biol Chem* 1992, **267**(2):1062-1066.
54. Ribeiro JM, Katz O, Pannell LK, Waitumbi J, Warburg A: **Salivary glands of the sand fly *Phlebotomus papatasi* contain pharmacologically active amounts of adenosine and 5'-AMP.** *J Exp Biol* 1999, **202**(Pt 11):1551-1559.
55. Lerner EA, Ribeiro JM, Nelson RJ, Lerner MR: **Isolation of maxadilan, a potent vasodilatory peptide from the salivary glands of the sand fly *Lutzomyia longipalpis*.** *J Biol Chem* 1991, **266**(17):11234-11236.

56. Kamhawi S: **The biological and immunomodulatory properties of sand fly saliva and its role in the establishment of *Leishmania* infections.** *Microbes Infect* 2000, **2**(14):1765-1773.
57. Belkaid Y, Valenzuela JG, Kamhawi S, Rowton E, Sacks DL, Ribeiro JM: **Delayed-type hypersensitivity to *Phlebotomus papatasi* sand fly bite: An adaptive response induced by the fly?** *Proc Natl Acad Sci U S A* 2000, **97**(12):6704-6709.
58. Titus RG, Ribeiro JM: **Salivary gland lysates from the sand fly *Lutzomyia longipalpis* enhance *Leishmania* infectivity.** *Science* 1988, **239**(4845):1306-1308.
59. Belkaid Y, Kamhawi S, Modi G, Valenzuela J, Noben-Trauth N, Rowton E, Ribeiro J, Sacks DL: **Development of a natural model of cutaneous leishmaniasis: powerful effects of vector saliva and saliva preexposure on the long-term outcome of *Leishmania major* infection in the mouse ear dermis.** *J Exp Med* 1998, **188**(10):1941-1953.
60. Valenzuela JG, Belkaid Y, Garfield MK, Mendez S, Kamhawi S, Rowton ED, Sacks DL, Ribeiro JM: **Toward a defined anti-*Leishmania* vaccine targeting vector antigens: characterization of a protective salivary protein.** *J Exp Med* 2001, **194**(3):331-342.
61. Kamhawi S, Belkaid Y, Modi G, Rowton E, Sacks D: **Protection against cutaneous leishmaniasis resulting from bites of uninfected sand flies.** *Science* 2000, **290**(5495):1351-1354.

62. Oliveira F, Kamhawi S, Seitz AE, Pham VM, Guigal PM, Fischer L, Ward J, Valenzuela JG: **From transcriptome to immunome: identification of DTH inducing proteins from a *Phlebotomus ariasi* salivary gland cDNA library.** *Vaccine* 2006, **24**(3):374-390.

Chapter 2

Identification and characterization of a salivary adenosine deaminase from the sand fly *Phlebotomus duboscqi*, the vector of *Leishmania major* in sub-Saharan Africa

Published as: Jochim RC*, Kato H*, Lawyer PG, Valenzuela JG. Identification and characterization of a salivary adenosine deaminase from the sand fly *Phlebotomus duboscqi*, the vector of *Leishmania major* in sub-Saharan Africa. *J Exp Biol* 2007 210(5):733-40. (* equally contributing authors)

Abstract

Two transcripts coding for an adenosine deaminase (ADA) were identified by sequencing a *Phlebotomus duboscqi* salivary gland cDNA library. Adenosine deaminase was previously reported in the saliva of the sand fly *Lutzomyia longipalpis*, but it was not present in the saliva of the sand flies *Phlebotomus papatasi*, *P. argentipes*, *P. perniciosus* and *P. ariasi*, suggesting that this enzyme is only present in the saliva of sand flies from the genus *Lutzomyia*. In the present work, we tested the hypothesis that the salivary gland transcript coding for ADA in *Phlebotomus duboscqi*, a sister species of *Phlebotomus papatasi*, produces an active salivary ADA. Salivary gland homogenates of *P. duboscqi* converted adenosine to inosine, suggesting the presence of ADA activity in the saliva of this species of sand fly; furthermore, this enzymatic activity was significantly reduced when using either salivary glands of recently blood-fed sand flies or punctured salivary glands, suggesting that this enzyme is secreted in the saliva of this insect. This enzymatic activity was absent from the saliva of *P. papatasi*. In contrast to other *Phlebotomus* sand flies, we did not find AMP or adenosine in *P. duboscqi* salivary glands as measured by HPLC-photodiode array. To confirm that the transcript coding for ADA was responsible for the activity observed in the saliva of this sand fly, we cloned this transcript into a prokaryotic expression vector and produced a soluble and active recombinant protein of approximately 60 kDa that was able to convert adenosine to inosine. Extracts of bacteria transformed with control plasmids did not show this activity. These results suggest that *P. duboscqi* transcripts coding for ADA are responsible for the activity detected in the salivary glands of this sand fly and that *P. duboscqi* acquired this activity independently from other *Phlebotomus* sand flies. This is

another example of a gene recruitment event in salivary genes of blood-feeding arthropods that may be relevant for blood-feeding and, because of the role of ADA in immunity, it may also play a role in parasite transmission.

Background

In their saliva, blood-feeding arthropods have potent pharmacologically active components that help them counteract the hemostatic and inflammatory system of the vertebrate host each time they attempt to take a blood meal [1]. Vasodilators, anticoagulants and inhibitors of platelet aggregation are part of this salivary mixture [2]. Recently, with the technological advances in DNA and protein sequencing, novel and unexpected molecules with potential biological activities have been isolated from the saliva of blood-feeding arthropods. Such molecules include hyaluronidase, nucleotidases, novel apyrases, amine-binding proteins, tissue-factor pathway inhibitors and others [2]. Another such protein is adenosine deaminase (ADA), which was identified from transcripts of a salivary gland cDNA library of the New World sand fly *Lutzomyia longipalpis* and the mosquitoes *Culex quinquefasciatus* and *Aedes aegypti* [3, 4]. This protein or the transcript coding for this protein also have been identified in other organisms including bacteria, fruit flies, mice and humans [5]. Adenosine deaminase (E.C. 3.5.4.4) catalyses the conversion of adenosine and 2'-deoxyadenosine to inosine and 2'-deoxyinosine, respectively [6]. This enzyme is evolutionarily conserved and has a beta alpha, 8 barrel structure and zinc ion in the catalytic site [7]. Adenosine deaminase deficiency in mice results in the impairment of T and B cell function due to the accumulation of adenosine resulting in severe combined immunodeficiency (SCID) [8].

The role of ADAs in insects, particularly in the saliva of blood-feeding insects, was proposed to be in the hydrolysis of adenosine, a molecule involved in pain perception [5]. The activity of this enzyme in blood-feeding insects was demonstrated in the saliva of *Lu. longipalpis*, *C. quinquefasciatus* and *Ae. aegypti* from the activity of the recombinant salivary ADA from *Lu. longipalpis* [3-5]. Of interest, ADA enzymatic activity or the transcripts coding for this enzyme were not present in the salivary gland of the sand flies *P. argentipes*, *P. papatasi*, *P. ariasi* and *P. perniciosus*, which belong to the genus *Phlebotomus* [3, 9]. Instead, the saliva of *P. papatasi* and *P. argentipes* contains large amounts of adenosine and adenosine monophosphate (AMP) [10]. Therefore, it appeared that ADA activity was only present in the saliva of *Lutzomyia* sand flies and not in *Phlebotomus* sand flies. Recently, transcriptome analysis of the salivary glands of the sand fly *Phlebotomus duboscqi*, a sibling species of *P. papatasi*, resulted in the identification of a transcript with homologies to ADA [11]. In the present work, we tested whether there is ADA activity in *P. duboscqi* and whether the identified transcript codes for this activity. Because *P. papatasi* does not have ADA activity, but has large amounts of adenosine and AMP in the saliva, we also tested for the presence of adenosine and AMP in the saliva of *P. duboscqi* sand flies.

Results

By sequencing a *P. duboscqi* salivary gland cDNA library, we have identified two transcripts coding for a protein homologous to ADA, an enzyme that metabolizes adenosine to inosine [11]. The first transcript (PduM73; NCBI accession number DQ835357) of 1846 bp codes for a secreted protein of 57.6 kDa with an isoelectric point

of 5.5; the second transcript (PduM74) of 1810 bp codes for a protein of 57.2 kDa with an isoelectric point of 5.8 (Figure 4). Multiple sequence comparison of *P. duboscqi* ADA with homologues from Dipterans such as the sand fly *Lu. longipalpis* and the mosquitoes *Ae. aegypti*, *Ae. albopictus* and *Culex pipiens* and from mammals such as mice, rats and humans shows an overall low level of identity. However, the amino acids forming part of the active site (His₁₁₆, His₁₁₈, Ala₁₂₁, Gly₃₂₈, His₃₅₅, Glu₃₅₈, Gly₃₈₁, Asp₄₄₀, Asp₄₄₁) are highly conserved (Figure 5). The ADA from *P. duboscqi* and from other insects is larger than the ADA from mice, rats or humans; a large string of approximately 80 amino acids at the N-terminal region is not present in the mammalian ADA. Additionally, the signal peptide sequence is not present in the mammalian ADA (Figure 5). Phylogenetic analysis of ADA from different organisms produced a tree with two distinct clades, one containing ADA from Dipteran blood-feeders and the other clade containing other organisms including *Leishmania*, *Plasmodium*, *Entamoeba*, mice, rats and humans (Figure 6). Within the Dipteran blood-feeders clade, sand flies form a distinct group separate from mosquitoes.

Salivary gland ADA activity

Because of the discovery of the ADA transcripts in the salivary gland cDNA library of *P. duboscqi*, we wanted to test whether the saliva of this sand fly had ADA activity. For this, SGH of *P. duboscqi* was incubated in the presence of adenosine and the reaction was followed spectrophotometrically by scanning from 220 nm to 300 nm every 3 min. The substrate adenosine absorbs at 265 nm, and the product of ADA activity, inosine, absorbs at 241 nm. The equivalent of 0.2 salivary gland pairs (0.2 g) of *P. duboscqi* converted

```

PduM73  MFPRLIVVLLAASAVHAVLDISNLIKPKRDYENFLQRYAEYADDEVDRSVGSDITLSLKEKFFVNQYLMDLKTEELKAGLKN
PduM74  MFPRLVFCLLAASAVHASL---KIKPKRSYDNFLEQYDNYATDEMDRSVGSDITLSHKEEIVNQYLMDLKTEELKAGLKN

PduM73  PSQFIPSNHFFSVLEDRINSSEIFKIIRRMPKGAILHAHDTALCSTDYVVSITTYRDELWQCADPKTGALQFRFSKESPKNT
PduM74  PSQFIPSNHFFSVLEDRINSSEIFKIIRRMPKGAILHAHDTALCSTDYVVSITTYRDELWQCADPKTGALQFRFSKESPKNT

PduM73  DTCQWTFVSEERKNQGEEQYNSKLRSQSLYNTDPIINRSRDVDSIWNDFMGLFGVNFGLLTYPVWKDYKQFLKEMMED
PduM74  DTCQWTFVSEERKNQGEEQYNSKLRSQSLYNTDPIINRSRDVDSIWNDFMGLFGVNFGLLTYPVWKDYKQFLKEMMED

PduM73  GVQYLELRGTLPPPLYDLGKIYNEEQQVVEIYYNVTEEFKKENSTFIGAKFIYAPVRFVNATGIKTLTTTVKQLHERFPDF
PduM74  GVQYLELRGTLPPPLYDLGKIYNEEQQVVEIYYNVTEEFKKENSTFIGAKFIYAPVRFVNATGIKTLTTTVKQLHERFPNF

PduM73  LAGFDLVGQEDKGGPLIGFSRELLELPESINFFHSGETNWNGMTDDNLIAAVTLGTKRIGHGYALFKHPRVLKQVKKDK
PduM74  LAGFDLVGQEDKGGPLIGFSRELLELPESINFFHSGETNWNGMTDDNLIAAVTLGTKRIGHGYALFKHPRVLKQVKKDK

PduM73  IAIEVCPISNQVLRRLVADMRNHPGSILLANKKYPMVISSDDPSFWEATPLSHDFYMAFMGLASYHQDLRMLKQLAINSLE
PduM74  IAIEVCPISNQVLRRLVADMRNHPGSILLANKKYPMVISSDDPSFWEATPLSHDFYMAFMGLASYHQDLRMLKQLAINSLE

PduM73  YSSMTLEEKTNAMKLWEAEWEKFIKELETEVFSLLE
PduM74  YSSMASEEKTNAMKLWEAEWEKFIKELETEV-SLLE

```

Figure 4. Amino acid alignment of the two adenosine deaminase (ADA) molecules derived from transcripts found in *P. duboscqi* salivary glands.

Amino acids shaded black are identical and those shaded gray are similar. The secretory signal peptide is italicized.

Figure 5. Clustal alignments of invertebrate putative salivary adenosine deaminase (ADA) of *P. duboscqi* (PduM73 and PduM74), *Lutzomyia longipalpis*, *Aedes aegypti*, *A. albopictus*, *Culex pipiens* and mammalian ADA (mouse, rat and human).

Arrowheads indicate conserved amino acids located in the active site of the mammalian enzyme. Black shading indicates amino acid sequence identity, and gray regions indicate conserved amino acid substitutions.

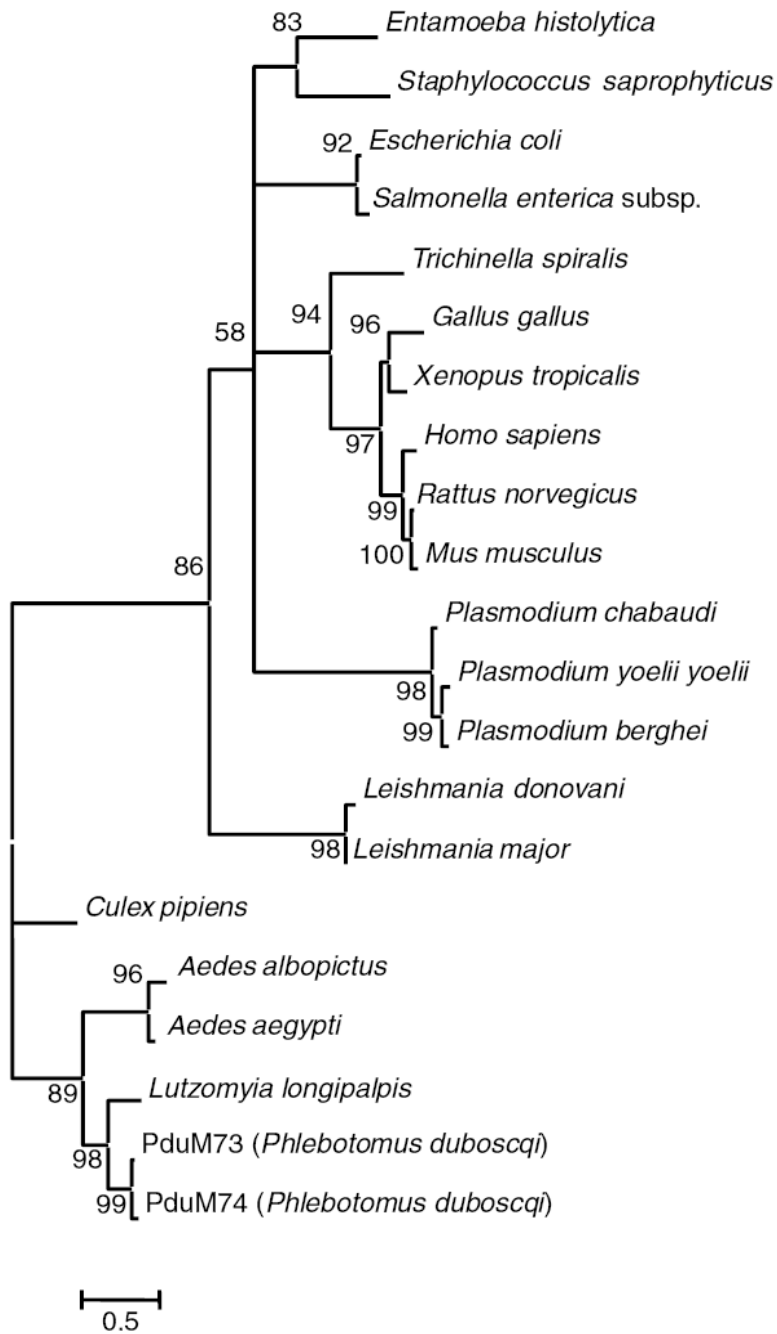


Figure 6. Phylogenetic tree analysis of putative adenosine deaminase (ADA).

Branch lengths are proportional to genetic distance calculated by the ClustalW program. The scale bar represents 0.5% divergence.

adenosine to inosine in 30 min (Figure 7A); by contrast, the same amount of SGH of *P. papatasi* had no effect on adenosine (Figure 7B). Differential spectrum shows, in better detail, the decrease of adenosine (265 nm) and, over time, the increase of inosine (241 nm) in the presence of *P. duboscqi* SGH (Figure 7C), indicating the presence of ADA activity in the salivary gland of this sand fly. To test whether this activity is secreted in the saliva of *P. duboscqi*, we compared ADA activity from SGH of unfed sand flies (intact saliva), from SGH of recently blood-fed sand flies (loss of secreted protein by salivation during feeding) and from punctured salivary glands (loss of all or the majority of the salivary contents and therefore any enzymatic activity). Salivary glands from unfed sand flies had the highest ADA activity while preparations from the salivary glands of recently blood-fed sand flies had approximately 70% less activity (Figure 8). Finally, ADA activity was not detected in the preparations of punctured salivary glands (Figure 8). Additionally, the amino-terminal sequence of the native protein was detected in the secreted fraction of the SGH of this sand fly [11]. These data suggest that the molecule responsible for this activity is secreted in the saliva of this sand fly.

Lack of adenosine and AMP in the saliva of *P. duboscqi*

It was previously shown that *P. papatasi* and *P. argentipes* do not have transcripts coding for the enzyme ADA in their salivary glands or the activity was not detected within their salivary glands; however, it was shown that these sand flies have large amounts of adenosine and AMP within their salivary glands [3, 9, 12]. Although counterintuitive, due

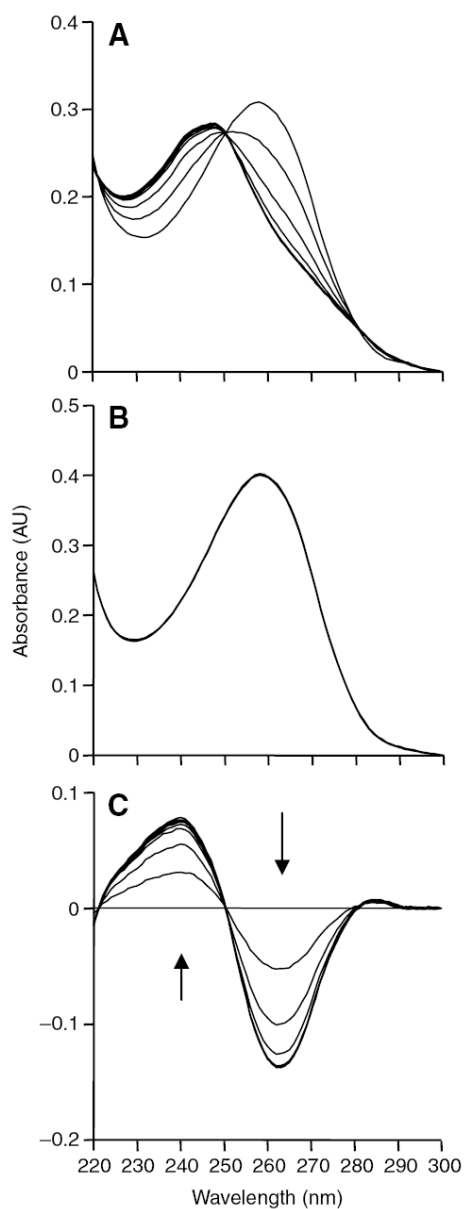


Figure 7. ADA activity of salivary homogenates of *P. duboscqi*.

A cuvette containing 20 μ M adenosine in PBS was scanned at 3 min intervals for 30 min following addition of salivary homogenate equivalent to 0.2 pairs of salivary gland from *P. duboscqi* (A) and *P. papatasi* (B). (C) Differential spectra of the data in (A) were obtained by subtracting each scan from the scan at time zero. The arrows indicate the direction of change of the spectrum over time.

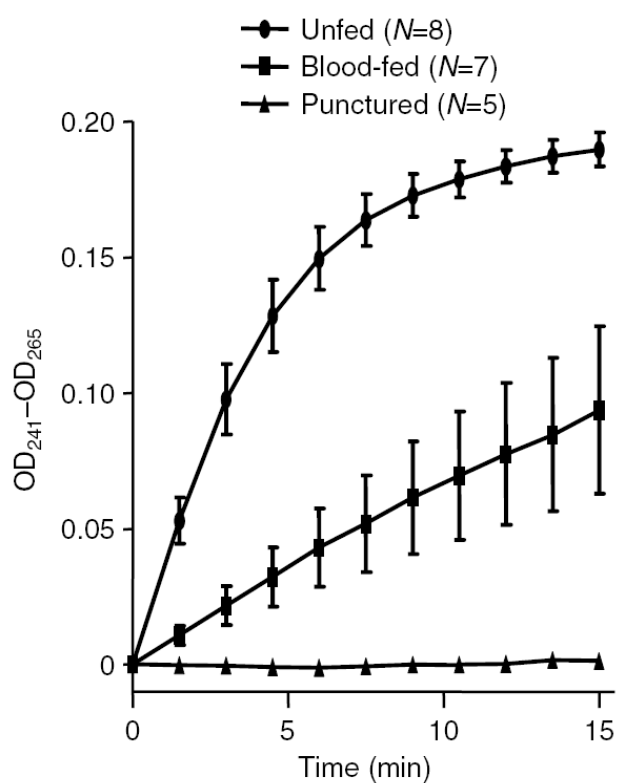


Figure 8. Salivary adenosine deaminase (ADA) activity from salivary gland homogenate (SGH) of unfed and blood-fed sand flies and from punctured salivary glands.

A cuvette containing 20 μ M adenosine in PBS was scanned at 1.5 min intervals for 15 min following addition of salivary homogenate equivalent to 0.2 pairs of salivary gland.

to the presence of ADA activity, *P. duboscqi* salivary glands were tested for the presence of adenosine and AMP by subjecting *P. duboscqi* SGH to molecular sieving-HPLC (MS-HPLC), and the eluted products were detected by photodiode array detection. As expected, and as previously shown, analysis of *P. papatasi* SGH resulted in the presence of two major peaks with the same retention times as adenosine (18.5 min) and AMP (14.5 min), respectively (Figure 9). By contrast, *P. duboscqi* SGH showed no peaks at the retention times of adenosine and AMP (Figure 9). Only a peak at 5 min was observed, which is the secreted proteins from the salivary glands, as determined by the retention time and the absorption spectra at 280 nm (Figure 9). These data suggest that, in contrast to *P. papatasi* and *P. argentipes*, *P. duboscqi* does not have AMP or adenosine in its salivary glands and that it contains the active salivary ADA.

Expression and activity of recombinant *P. duboscqi* salivary ADA

In order to determine if the ADA activity detected in *P. duboscqi* SGH was related to the transcript coding for this enzyme, we cloned the two transcripts coding for this protein into the PCRT7NT-TOPO bacterial expression vector. The soluble expressed proteins were purified from the supernatant of bacterial lysate by nickel magnetic beads, and an aliquot was subjected to western blot analysis and detected using anti-histidine antibody. This revealed a protein of approximately 60 kDa, which is the estimated molecular mass of the predicted ADA including the 4-kDa N-terminal addition that includes His₆G and Xpress™ peptide epitopes (Figure 10A, lanes 1 and 2). No protein of this molecular mass was detected in the supernatant of bacteria expressing the empty vector (Figure 10A, lane 3). Furthermore, a soluble expressed protein with the same migration pattern was

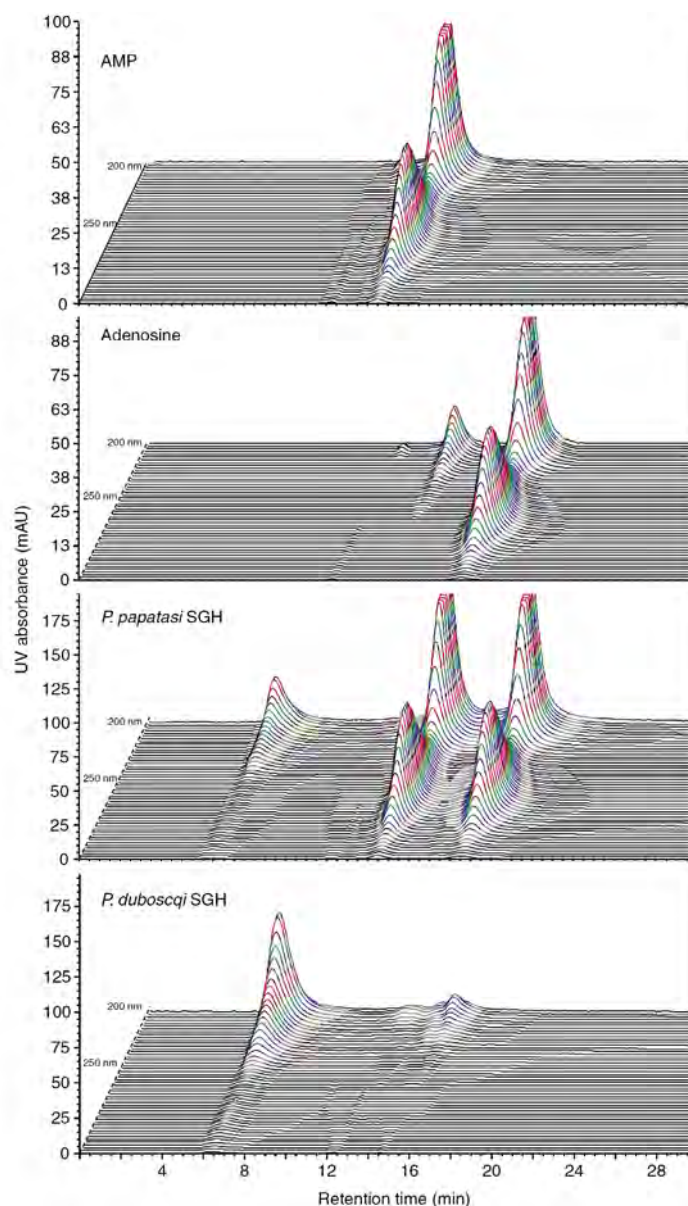


Figure 9. Three-dimensional chromatographic display of photodiode array data obtained from MS-HPLC of *P. papatasi* and *P. duboscqi* salivary gland homogenate (SGH).

The three dimensional data show retention time on the x-axis, UV absorbance on the y-axis (in milli-absorbance units) and the UV absorbance spectra on the z-axis (from 200 nm to 320 nm). Adenosine and AMP standards show retention times of 18.5 and 14.5 min, respectively.

detected by Coomassie blue staining (Figure 10B, lanes 1 and 2), and this protein was not detected in samples from bacteria transformed with control plasmid (Figure 10B, lane 3). The purified soluble recombinant proteins were then tested for the presence of ADA activity. Both sand fly recombinant proteins had a high level of ADA activity as detected spectrophotometrically by the conversion of adenosine to inosine (Figure 11A and 11B). This activity was not detected in the supernatant of bacteria transformed with the control plasmid (Figure 11C). These data suggest that the *P. duboscqi* transcript coding for an ADA is responsible for the ADA activity detected in the saliva of this sand fly.

Discussion

Phlebotomus duboscqi is a proven vector of *Leishmania major* in sub-Saharan Africa. It belongs to the subgenus *Phlebotomus*, together with the sand fly vector *P. papatasi*. Knowledge of the repertoire of salivary activities or molecules from the saliva of *P. duboscqi* is very limited. We recently sequenced a large number of transcripts from the salivary gland of *P. duboscqi* and identified a transcript coding for the enzyme ADA [11]. This is the first report of ADA in a sand fly from the genus *Phlebotomus*, including data from transcriptome analysis from the salivary glands of *P. papatasi*, *P. ariasi*, *P. argentipes* and *P. perniciosus* sand flies [9]. In the present work, we have demonstrated the presence of ADA activity in the saliva of *P. duboscqi*, and we also demonstrated that the soluble recombinant protein produced from the transcript coding for this enzyme exhibited ADA activity. Phylogenetic analysis placed *P. duboscqi* ADA in the same clade with ADA from other blood-feeding arthropods. This group belongs to the ADGF/CECR1 family of proteins identified previously in *Sarcophaga peregrina*,

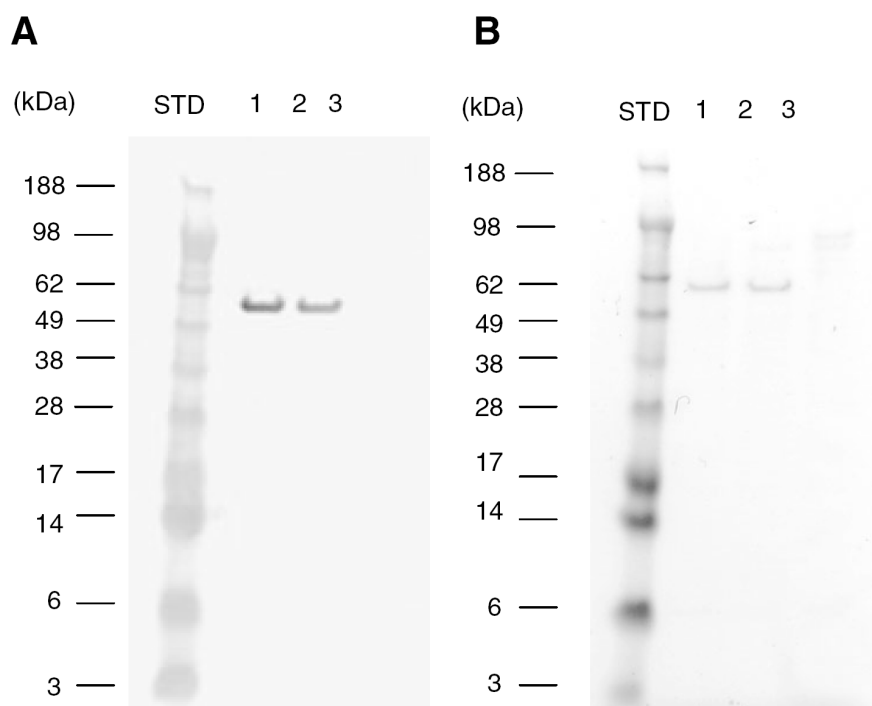


Figure 10. Expression of recombinant *P. duboscqi* salivary adenosine deaminase (ADA).

cDNAs encoding *P. duboscqi* salivary ADA (PduM73 and PduM74) were cloned into PCRT7NT-TOPO vector, and recombinant proteins were expressed in *E. coli*. The affinity column-purified proteins (lane 1, PduM73; lane2, PduM74; lane 3, empty plasmid vector) were analyzed by SDS-PAGE and then subjected to (A) western blotting and (B) Coomassie blue staining.

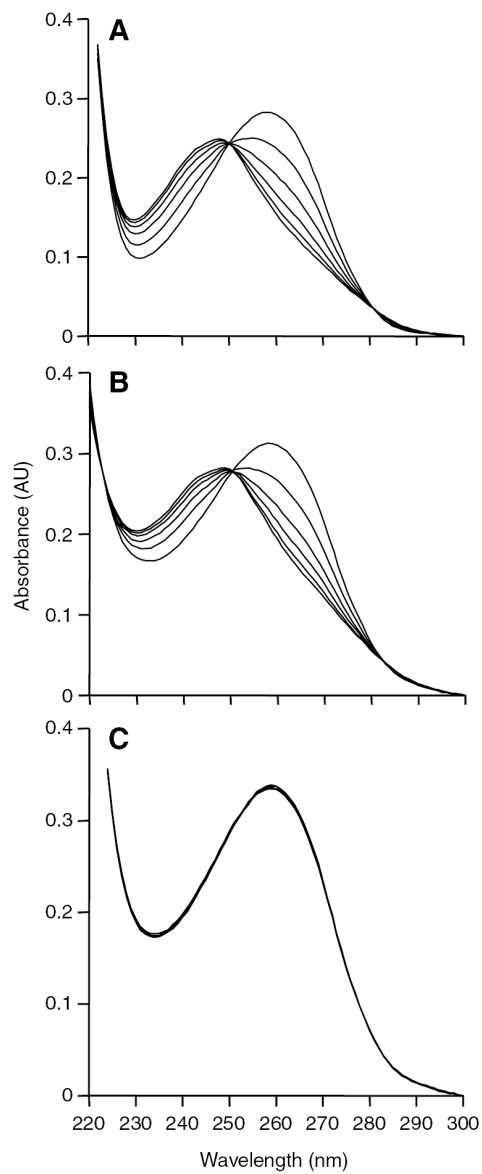


Figure 11. Enzymatic activity of recombinant *P. duboscqi* salivary adenosine deaminase (ADA).

A cuvette containing 20 μ M adenosine in PBS was scanned at 1.5 min intervals for 15 min following addition of recombinant ADA (A, PduM73; B, PduM74) or supernatant of bacteria transformed with control plasmid (C).

Lutzomyia longipalpis, *Drosophila*, *Aplysia* and humans [13]. This sub-family of ADAs has an extended N-terminus region and is targeted for secretion. These data suggest that *P. duboscqi* acquired this activity independent of other *Phlebotomus* sand flies. The question remains as to what the role of this protein is in blood-feeding. It was previously speculated that the activity may be related to the hydrolysis of adenosine, an important component in pain perception and in immunity [3]. What is puzzling is that other *Phlebotomus* sand flies do not have ADA in their salivary glands, but they have large amounts of adenosine and AMP, very active vasodilators and platelet inhibitors. Neither adenosine nor AMP was present in the saliva of *P. duboscqi*, as demonstrated in this chapter. *Lutzomyia longipalpis* also lacks adenosine and AMP in its saliva; however, it has maxadilan, a very potent vasodilator [14]. Maxadilan was not identified in the *P. duboscqi* cDNA library [11]. Therefore, it appears that *P. duboscqi* may contain a novel vasodilator that will replace the lack of vasodilatory activities exerted by AMP and adenosine in other *Phlebotomus* sand flies. The fact that ADA is present only in *P. duboscqi* and not in other *Phlebotomus* sand flies examined to date emphasizes the ability of blood-feeding arthropods to acquire independent strategies to overcome or modulate the host hemostatic, inflammatory and immune system. ADA has an important role in immunity as a result of the effects of adenosine, 2'-deoxyadenosine and the hydrolytic product of these compounds [6]. Further work will be necessary to determine the effect of this enzyme in parasite transmission. It may be possible that this enzyme changes the environment in the skin where the *Leishmania* parasite is deposited by the sand fly. Inosine is the primary metabolite of adenosine by ADA. Inosine has been shown to inhibit the production of proinflammatory cytokines including TNF- α , IL-1, IL-12,

MIP1- α and INF γ in stimulated macrophages and spleen cells [15]. Additionally, adenosine and inosine can alter cutaneous vasopermeability by activating A₃ receptors on mast cells [16]. Then, it may be possible that inosine may favor a Th2 environment that will benefit parasite establishment in the skin of the mammalian host.

Materials and methods

Sand flies and preparation of salivary gland homogenate (SGH)

A Phlebotomus duboscqi Theodor, Mali strain, were reared using a mixture of fermented rabbit food and rabbit feces as larval food. Adult sand flies were offered a cotton swab containing 20% sucrose and were used for dissection of salivary glands at 5–7 days after emergence. Salivary glands were stored in groups of 10 pairs in 10 μ l phosphate-buffered saline (PBS). Salivary glands were disrupted by ultrasonication in 1.5 ml conical tubes. Tubes were centrifuged at 10,000 *g* for 2 min and the resultant supernatant used for the studies.

Salivary gland cDNA library

The *P. duboscqi* salivary gland cDNA library was made as previously described [11]. Briefly, mRNA was isolated from 55 salivary gland pairs using the Micro-FastTrack mRNA isolation kit (Invitrogen, San Diego, CA, USA). The PCR-based cDNA library was made following the instructions for the SMART cDNA library construction kit (BD-Clontech, Palo Alto, CA, USA) with some modifications [11]. The *P. duboscqi* cDNA library was sequenced as previously described using an Applied

Biosystems 3730xl DNA Analyzer (Foster City, CA, USA) and a CEQ 2000XL DNA sequencing instrument (Beckman Coulter, Fullerton, CA, USA) [11].

Phylogenetic analysis

Consensus protein sequences were compared to related sequences from sand flies as well as non-sand fly species obtained from GenBank. Sequences were aligned using ClustalX and manually refined using BioEdit sequence editing software (<http://www.mbio.ncsu.edu/BioEdit/page2.html>) [17]. Phylogenetic analysis was conducted on protein alignments using Tree Puzzle version 5.2 [18]. Tree Puzzle constructs phylogenetic trees by maximum likelihood, using quartet puzzling, automatically estimating internal branch node support (1000 replications). Derived trees were visualized using MEGA (Molecular Evolutionary Genetics Analysis) version 3.1 (<http://www.megasoftware.net/>) [19].

Enzymatic assays

Measurement of activity was performed in quartz microcuvettes using 60 μ l samples (Starna Cells, Atascadero, CA, USA). 20 μ M adenosine in PBS was added to the cuvette, followed by addition of the enzyme source. After mixing the solution by pipetting, the absorbance between 220 and 300 nm was monitored at 1.5 or 3.0 min intervals using a Lambda 18 spectrophotometer from Perkin Elmer (Norwalk, CT, USA).

Molecular sieving-high-performance liquid chromatography

Molecular sieving-high-performance liquid chromatography (MS-HPLC) was carried out using a Dionex Summit system and Chromeleon software (Dionex, Sunnyvale, CA, USA). For analysis, 20 µl of sample was applied to a Superdex Peptide PC 3.2/30 column (Amersham Biosciences, Piscataway, NJ, USA) using 10 mM NaPO₄, 150 mM NaCl, pH 6.5 as the mobile phase at a flow rate of 150 µl min for 30 min of separation. Detection was performed using a photodiode array detector and a 3-D chromatogram generated using Chromeleon software. Adenosine and AMP standards (500 pmol) were applied separately. Single pairs of salivary glands from *P. papatasi* and *P. duboscqi* were sonicated in PBS, clarified by centrifugation and applied to HPLC.

Expression of *P. duboscqi* ADA

DNA fragments encoding mature *P. duboscqi* ADA protein were amplified and inserted into the cloning site of the pCRT7/NT-TOPO vector (Invitrogen). The primers used for PCR amplification of the mature ADA encoding fragments were PDBL_P02_G04_VF (5'-GTTTTGGACATTTCTGAACATTA-3') and PDBL_P02_G09_VR (5'-TGGCTCCAAATGATTCAGACA-3') for 2G4 (PduM73) and PDBL_P02_G09_VF (5'-CTTTGAAAATTAAACCGAAACGA-3' and PDBL_P02_G09_VR for 2G9 (PduM74). *Escherichia coli* strain BL21(DE3)pLysS cells (Invitrogen) were transformed with the recombinant plasmid and grown in LB broth containing ampicillin (50 µg/ml). Production of recombinant protein was induced by addition of IPTG to a final concentration of 1 mM and at 27°C for 3 h. The recombinant protein was purified from the supernatant of bacterial sonic lysate using a MagneHis

Protein Purification System (Promega, Madison, WI, USA) and dialysed with Centricon Plus-20 (Millipore, Bedford, MA, USA) to remove imidazole from the elution buffer before further enzymatic analysis.

Sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis (PAGE) and western blotting

The samples were treated with NuPAGE LDS sample buffer (Invitrogen) and analyzed on NuPAGE 10% Bis-Tris gels (Invitrogen) with NuPAGE MES SDS running buffer (Invitrogen). To estimate the molecular mass of the samples, SeeBlue markers from Invitrogen (myosin, bovine serum albumin, glutamic dehydrogenase, alcohol dehydrogenase, carbonic anhydrase, myoglobin, lysozyme, aprotinin, and insulin, chain B) were used. After electrophoresis, the gels were stained with SimplyBlue™ SafeStain Coomassie® (Invitrogen) or SilverQuest™ Silver Staining (Invitrogen). For the western blotting, the proteins in the gel were transferred to nitrocellulose membrane (Invitrogen) using NuPAGE transfer buffer (Invitrogen). After blocking with 5% milk in Tris-buffered saline containing 0.1% Tween-20, pH 8.0 (TBST), the membrane was incubated with alkaline phosphatase (AP)-conjugated anti-His₆/G antibody (Invitrogen) for 1 h at room temperature. After three washes with TBST, the blots were developed by addition of 5-bromo-4-chloro-3-indolyl-1-phosphate and nitro blue tetrazolium for visualization.

References

1. Ribeiro JM: **Role of saliva in blood-feeding by arthropods.** *Annu Rev Entomol* 1987, **32**:463-478.
2. Ribeiro JM, Francischetti IM: **Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives.** *Annu Rev Entomol* 2003, **48**:73-88.
3. Charlab R, Rowton ED, Ribeiro JM: **The salivary adenosine deaminase from the sand fly *Lutzomyia longipalpis*.** *Exp Parasitol* 2000, **95**(1):45-53.
4. Ribeiro JM, Charlab R, Valenzuela JG: **The salivary adenosine deaminase activity of the mosquitoes *Culex quinquefasciatus* and *Aedes aegypti*.** *J Exp Biol* 2001, **204**(Pt 11):2001-2010.
5. Charlab R, Valenzuela JG, Andersen J, Ribeiro JM: **The invertebrate growth factor/CECR1 subfamily of adenosine deaminase proteins.** *Gene* 2001, **267**(1):13-22.
6. Cristalli G, Costanzi S, Lambertucci C, Lupidi G, Vittori S, Volpini R, Camaioni E: **Adenosine deaminase: functional implications and different classes of inhibitors.** *Med Res Rev* 2001, **21**(2):105-128.
7. Wilson DK, Rudolph FB, Quijcho FA: **Atomic structure of adenosine deaminase complexed with a transition-state analog: understanding catalysis and immunodeficiency mutations.** *Science* 1991, **252**(5010):1278-1284.
8. Resta R, Hooker SW, Laurent AB, Jamshedur Rahman SM, Franklin M, Knudsen TB, Nadon NL, Thompson LF: **Insights into thymic purine metabolism and**

- adenosine deaminase deficiency revealed by transgenic mice overexpressing ecto-5'-nucleotidase (CD73).** *J Clin Invest* 1997, **99**(4):676-683.
9. Anderson JM, Oliveira F, Kamhawi S, Mans BJ, Reynoso D, Seitz AE, Lawyer P, Garfield M, Pham M, Valenzuela JG: **Comparative salivary gland transcriptomics of sandfly vectors of visceral leishmaniasis.** *BMC Genomics* 2006, **7**:52.
 10. Ribeiro JM, Modi G: **The salivary adenosine/AMP content of *Phlebotomus argentipes* Annandale and Brunetti, the main vector of human kala-azar.** *J Parasitol* 2001, **87**(4):915-917.
 11. Kato H, Anderson JM, Kamhawi S, Oliveira F, Lawyer PG, Pham VM, Sangare CS, Samake S, Sissoko I, Garfield M *et al*: **High degree of conservancy among secreted salivary gland proteins from two geographically distant *Phlebotomus duboscqi* sandflies populations (Mali and Kenya).** *BMC Genomics* 2006, **7**:226.
 12. Ribeiro JM, Katz O, Pannell LK, Waitumbi J, Warburg A: **Salivary glands of the sand fly *Phlebotomus papatasi* contain pharmacologically active amounts of adenosine and 5'-AMP.** *J Exp Biol* 1999, **202**(Pt 11):1551-1559.
 13. Dolezelova E, Zurovec M, Dolezal T, Simek P, Bryant PJ: **The emerging role of adenosine deaminases in insects.** *Insect Biochem Mol Biol* 2005, **35**(5):381-389.
 14. Ribeiro JM, Vachereau A, Modi GB, Tesh RB: **A novel vasodilatory peptide from the salivary glands of the sand fly *Lutzomyia longipalpis*.** *Science* 1989, **243**(4888):212-214.

15. Hasko G, Kuhel DG, Nemeth ZH, Mabley JG, Stachlewitz RF, Virag L, Lohinai Z, Southan GJ, Salzman AL, Szabo C: **Inosine inhibits inflammatory cytokine production by a posttranscriptional mechanism and protects against endotoxin-induced shock.** *J Immunol* 2000, **164**(2):1013-1019.
16. Tilley SL, Wagoner VA, Salvatore CA, Jacobson MA, Koller BH: **Adenosine and inosine increase cutaneous vasopermeability by activating A(3) receptors on mast cells.** *J Clin Invest* 2000, **105**(3):361-367.
17. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ: **Multiple sequence alignment with Clustal X.** *Trends Biochem Sci* 1998, **23**(10):403-405.
18. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**(3):502-504.
19. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5**(2):150-163.

Chapter 3

Exploring the midgut transcriptome of *Phlebotomus papatasi*: comparative analysis of expression profiles of sugar-fed, blood-fed and *Leishmania major*-infected sand flies

Published as: Jochim RC*, Ramalho-Ortigao M*, Anderson JM, Lawyer PG, Pham VM, Kamhawi S, Valenzuela JG: Exploring the midgut transcriptome of *Phlebotomus papatasi*: comparative analysis of expression profiles of sugar-fed, blood-fed and *Leishmania major*-infected sandflies. *BMC Genomics* 2007, 8(1):300. (* equally contributing authors)

Abstract

Background: In sand flies, the blood meal is responsible for the induction of several physiologic processes that culminate in egg development and maturation. During blood-feeding, infected sand flies also are able to transmit the parasite *Leishmania* to a suitable host. Many blood-induced molecules play significant roles during *Leishmania* development in the sand fly midgut, including parasite killing within the endoperitrophic space. In this work, we randomly sequenced transcripts from three distinct high quality full-length female *Phlebotomus papatasi* midgut-specific cDNA libraries from sugar-fed, blood-fed and *Leishmania major*-infected sand flies. Furthermore, we compared the transcript expression profiles from the three different cDNA libraries by customized bioinformatics analysis and validated these findings by semi-quantitative PCR and real-time PCR.

Results: Transcriptome analysis of 4010 cDNA clones resulted in the identification of the most abundant *P. papatasi* midgut-specific transcripts. The identified molecules included those with putative roles in digestion and peritrophic matrix formation, among others. Moreover, we identified sand fly midgut transcripts that are expressed only after a blood meal, such as microvilli associated-like protein (*PpMVP1*, *PpMVP2* and *PpMVP3*), a peritrophin (*PpPer1*), trypsin 4 (*PpTryp4*), chymotrypsin *PpChym2*, and two unknown proteins. Of interest, many of these overabundant transcripts such as *PpChym2*, *PpMVP1*, *PpMVP2*, *PpPer1* and *PpPer2* were of lower abundance when the sand fly was given a blood meal in the presence of *L. major*.

Conclusion: This tissue-specific transcriptome analysis provides a comprehensive look at the repertoire of transcripts present in the midgut of the sand fly *P. papatasi*. Furthermore, the customized bioinformatic analysis allowed us to compare and identify the overall transcript abundance from sugar-fed, blood-fed and *Leishmania*-infected sand flies. The suggested upregulation of specific transcripts in a blood-fed cDNA library were validated by real-time PCR, suggesting that this customized bioinformatic analysis is a powerful and accurate tool useful in analyzing expression profiles from different cDNA libraries. Additionally, the findings presented in this work suggest that the *Leishmania* parasite is modulating key enzymes or proteins in the gut of the sand fly that may be beneficial for its establishment and survival.

Background

Cutaneous leishmaniasis due to *L. major* is found throughout the Old World, including the Middle East and West Africa. *Phlebotomus papatasi* is the principal vector for this parasite and is refractory to the development of other species of *Leishmania*.

Upon taking a blood meal, hematophagous arthropods express a large number of molecules that participate in various physiologic processes ranging from blood digestion to egg development. Furthermore, many insects can either obtain or transmit pathogens during the acquisition of a blood meal. In blood-feeding arthropods, the midgut plays a crucial role as the primary organ involved in processing the blood meal and, in some instances, molecules expressed in the midgut of an insect vector have been shown to directly influence pathogen establishment [1, 2]. Certain pathogens, such as *Leishmania*,

appear able to modulate the activity of sand fly midgut proteases for their own benefit or survival [3, 4].

Sequenced data sets containing information regarding expression profiles of anopheline and culicine mosquitoes, such as *Anopheles gambiae* and *Aedes aegypti*, following a blood meal have become available [5, 6]. Other datasets now encompass insects such as *Pediculus humanus* [7] and *Culicoides sonorensis* [8]. In comparison, transcriptome information regarding sand flies is limited. Previous work has focused mainly on the sand fly salivary gland [9-11]; whereas, only a small number of sand fly-specific midgut cDNAs have been identified [12-16]. Recently, a large set of cDNA transcripts from the whole sand fly *Lutzomyia longipalpis* has been sequenced, providing greater information regarding molecules present in sand flies [17]. However, the information regarding sand fly midgut-specific transcripts remains poor.

In this work, we embarked on a comprehensive study of *P. papatasi* midgut-specific transcripts and compared the expression profile of these transcripts by directly comparing those obtained from midguts of females fed on sugar only, on blood or on blood containing *L. major*. With this approach, we have identified several *P. papatasi* midgut-specific transcripts that are differentially expressed after a blood meal and in the presence of *L. major*.

Results and discussion

The midgut is the tissue where *Leishmania* development takes place while within its sand fly vector. Within the midgut environment, *Leishmania* possibly interacts with various secreted molecules and cell types lining the midgut epithelia. In order to gain

greater insight into the repertoire of the proteins present in the midgut of *P. papatasi*, we constructed and sequenced three high quality full-length cDNA libraries from the midgut of sand flies fed either on sugar only (unfed), blood or blood containing *L. major*. A total of 4010 high quality sequenced clones obtained from the three cDNA libraries were combined and analyzed, resulting in the formation of 1382 clusters. Each cluster may contain a large number of transcripts, which creates a contig (high quality consensus sequence) or may have a single transcript that can be defined as a singleton. Therefore, we will utilize the nomenclature of “cluster” in the remainder of the manuscript to define either a consensus sequence from various transcripts or a singleton.

Consensus sequences were compared with various databases, and putative functions were assigned. The categories for the transcripts’ potential biologic functions included protein synthesis machinery, protein modification machinery, transcription machinery, transporters, extracellular matrix, signal transduction, immunity, adhesion, and conserved proteins of unknown function. Table 3 summarizes this analysis listing transcripts from female *P. papatasi* midguts fed on sugar, on blood, and on blood containing *L. major*. The first column shows the putative biological function, the first section of columns shows the number of clusters found in each of the three cDNA libraries in relation to this function. The second section of columns indicates the total number of sequences for these clusters, and the third section of columns shows the average of the number of sequences per cluster. The category of “conserved unknown function” had the largest number of clusters in all three of the cDNA libraries. These were followed by metabolism, energy in the sugar-fed library (95 clusters); metabolism, amino acid, which includes digestive enzymes, in the blood meal library (40 clusters);

Table 3: List of *Phlebotomus papatasi* midgut-specific sequences, clusters, and sequences per cluster of cDNA libraries made from flies sugar-fed, blood-fed, and blood-fed with *Leishmania major* parasites

Biological function	Number of Clusters			Number of Sequences			Sequences/cluster		
	Sugar	Blood-fed	L. major	Sugar	Blood-fed	L. major	Sugar	Blood-fed	L. major
protein synthesis machinery	88	39	51	281	73	159	3.19	1.87	3.12
protein modification machinery	19	20	8	39	22	9	2.05	1.10	1.13
protein export machinery	14	13	10	14	13	10	1.00	1.00	1.00
transcription machinery	4	2	1	4	2	1	1.00	1.00	1.00
transcription factors	13	6	3	22	9	4	1.69	1.50	1.33
proteasome machinery	13	7	7	21	7	9	1.62	1.00	1.29
transporters	24	16	11	37	19	33	1.54	1.19	3.00
extracellular matrix	6	12	5	164	122	59	27.33	10.17	11.80
cytoskeletal	16	15	14	61	291	210	3.81	19.40	15.00
signal transduction	28	20	13	38	20	15	1.36	1.00	1.15
protease inhibitor	3	1	3	8	5	9	2.67	5.00	3.00
immunity	3	4	3	8	5	4	2.67	1.25	1.33
adhesion	1	2	1	1	2	1	1.00	1.00	1.00
nuclear metabolism and regulation	17	7	6	18	9	6	1.06	1.29	1.00
metabolism, energy	95	38	42	194	54	62	2.04	1.42	1.48
metabolism, lipid	9	11	9	11	18	10	1.22	1.64	1.11
metabolism, carbohydrate	16	9	10	27	12	14	1.69	1.33	1.40
metabolism, amino acid	28	40	30	131	182	197	4.68	4.55	6.57
metabolism, nucleic acid and nucleotides	7	9	4	14	9	4	2.00	1.00	1.00
metabolism, heme	3	3	2	8	28	8	2.67	9.33	4.00
conserved of unknown function	262	167	322	405	282	470	1.55	1.69	1.46
Total	669	441	555	1506	1184	1294	Avg 3.23	Avg 3.27	Avg 3.01

and protein synthesis machinery in the *L. major* blood-meal library (51 clusters). The categories with the highest number of sequences per cluster differed between the three cDNA libraries and was highest among transcripts identified as extracellular matrix (27.33 seq/cluster) in the sugar-fed cDNA library and cytoskeletal transcripts for both the blood meal (19.40 seq/cluster) and *L. major* blood meal cDNA libraries (15.00 seq/cluster). The sugar-fed cDNA library has 669 clusters with an average of 3.23 sequences per cluster. The cDNA library constructed from blood-fed midguts consisted of 441 clusters, with an average of 3.27 sequences per cluster. Of *P. papatasi* midgut fed on blood containing *L. major*, this library produced 555 clusters, with an average of 3.01 sequences per cluster.

The number of sequences in each category for the three cDNA libraries is graphically represented in Figure 12. After blood-feeding, there is a decrease in the number of sequences in all categories other than cytoskeletal, amino acid metabolism, and heme metabolism. Noticeable differences in the number of sequences between the blood-fed and blood-fed containing *L. major* libraries occurs in the protein synthesis machinery, extracellular matrix, cytoskeletal, heme metabolism, and conserved of unknown function categories.

Table 4 gives a more detailed description of the different types of transcripts identified in the combined analysis of the three cDNA libraries. Only high quality sequences and, for the most part, full-length coding sequences submitted to GenBank are shown. This table shows the different clusters arranged in the order of cluster number in the combined analysis of the three cDNA libraries. The first column of Table 4 describes the cluster number, the second column shows the clone that produced the full-length

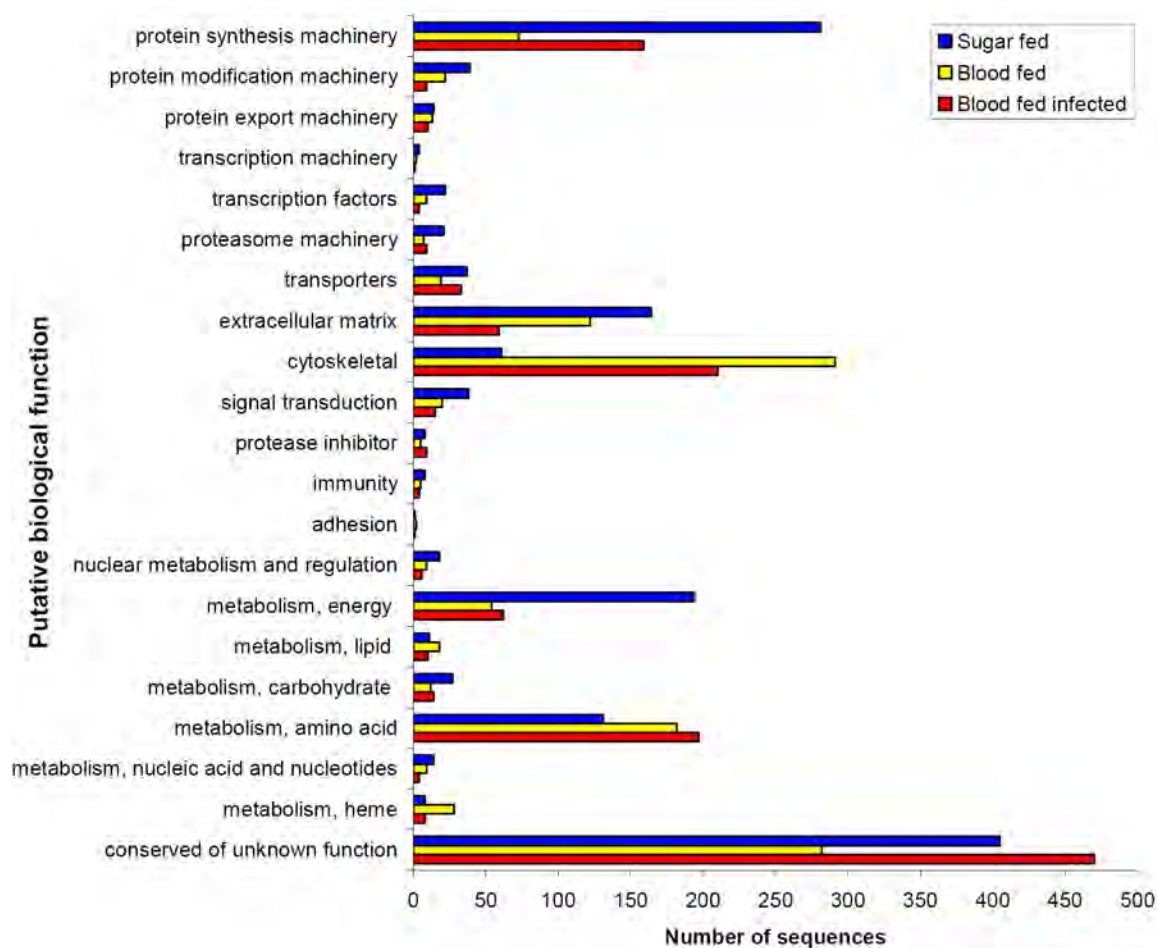


Figure 12. Distribution of sequences analyzed from each cDNA library separated by putative biologic function.

Table 4: Clusters of combined *P. papatasi* midgut cDNA libraries (sugar-fed, blood-fed and *Leishmania major* -infected) of transcripts with high quality sequences

Cluster	Clone	NCBI best match to NR database	E value	Putative function	GenBank
Cluster 1	B05_ppmglbl_p23	microvilli membrane protein [A. aegypti]	1.0E-47	Microvilli protein	EU031911
Cluster 2	F05_ppmglbl_p23	microvilli membrane protein [A. aegypti]	1.0E-47	Microvilli protein	EU031911
Cluster 3	PPMGBM189	microvilli membrane protein [A. aegypti]	7.0E-48	Microvilli protein	EU031911
Cluster 9	PPMGBL17	LP07759p [D. melanogaster]	8.0E-25	Peritrophin	EU031912
Cluster 10	A05_PPMGBS_P28	hypothetical protein 17 [L. obliqua]	3.0E-44	40S ribosomal S30 protein	EU040041
Cluster 11	A07_PPMGBS_P28	ENSANGP00000028746 [A. gambiae]	1.0E-33	Unknown	EU040042
Cluster 12	PPMGM173	similar to CG4778-PA [T. castaneum]	8.0E-11	Peritrophin	EU047543
Cluster 13	A02_PPINFIL_P30	similar to CG4778-PA [T. castaneum]	7.0E-11	Peritrophin	EU047543
Cluster 15	H08_PPMGBL_P31	chymotrypsin [P. papatasi]	1.0E-154	Chymotrypsin	AAM96939
Cluster 16	E10_ppmgbm_p21	carboxypeptidase B [A. aegypti]	4.0E-67	Carboxypeptidase	EU047544
Cluster 17	C07_PPMGS_P24	ribosomal protein S20 [B. mori]	9.0E-57	40S ribosomal protein S20	EU047545
Cluster 18	H10_ppmgbm_p22	trypsin 1 [P. papatasi]	1.0E-151	Trypsin	AAM96940
Cluster 20	H05_ppmgbm_p22	CG32276-PB, isoform B [D. melanogaster]	2.0E-23	Ribosome associated membrane protein	EU047546
Cluster 21	PPMGBS47	Ribosomal protein L19 [D. melanogaster]	1.0E-100	60S ribosomal protein L19	EU047547
Cluster 23	D05_PPMGL_P28	trypsin 2 [P. papatasi]	1.0E-157	Trypsin	AAM96941
Cluster 24	G04_PPMGM_P25	RE59709p [D. melanogaster]	7.0E-68	60S ribosomal protein L32	EU047548
Cluster 25	PPMGS_P31_B06	similar to D. melanogaster CG3203 [D. yakuba]	1.0E-91	60S ribosomal protein L17	EU045355
Cluster 26	PPMGL_P29_D06	peritrophin-like protein 1 [C. felis]	2.0E-36	Peritrophin	EU045354
Cluster 29	PPINFM-P7-G10	Ribosomal protein L29 [D. melanogaster]	4.0E-24	60S ribosomal protein L29	EU045353
Cluster 31	B03_PPMGM_P25	CG13551 [D. melanogaster]	3.0E-37	Unknown	EU049582
Cluster 32	H04_PPMGL_P23	LD17235p [D. melanogaster]	1.0E-93	60S ribosomal protein L11	EU045352
Cluster 34	PPMGM152	similar to D. melanogaster Rpl14 [D. yakuba]	5.0E-50	60S ribosomal protein L14	EU045351
Cluster 35	E07_ppmgs_p21	60S acidic ribosomal protein P1 [S. frugiperda]	1.0E-45	60S Acidic ribosomal protein P1	EU045350
Cluster 37	PPMGL197	ENSANGP00000019623 [A. gambiae]	4.0E-63	Astacin	EU045349
Cluster 40	F01_PPINFIL_P30	S7 ribosomal protein [C. pipiens quinquefasciatus]	3.0E-89	40S ribosomal protein S7	EU045348
Cluster 73	A03_PPMGBM_P25	unknown [C. sonorensis]	4.0E-25	Unknown	EU045347
Cluster 75	C04_ppmgs_p21	similar to D. melanogaster qm [D. yakuba]	1.0E-117	60S ribosomal protein L10	EU045346
Cluster 89	B11_PPINFIL_P32	trypsin 4 [P. papatasi]	1.0E-129	Trypsin	AAM96943
Cluster 94	F08_PPINFIL_P22	microvilli membrane protein [A. aegypti]	2.0E-35	Microvilli protein	EU047549
Cluster 96	C05_ppmgbm_p23	Cr-Pil [P. americana]	3.0E-20	Microvilli protein	EU047550
Cluster 98	B12_ppmglbl_p23	ENSANGP00000017713 [A. gambiae]	1.0E-17	Microvilli protein	EU047551
Cluster 99	A03_ppmglbl_p20	hypothetical protein [T. castaneum]	2.0E-23	Unknown	EU045345
Cluster 103	H03_PPMGL_P23	GA13179-PA [D. pseudoobscura]	2.0E-43	Ferritin	EU045344
Cluster 106	H07_ppmgbm_p23	TPA_inf. HDC07203 [D. melanogaster]	8.0E-14	Unknown	EU045343
Cluster 111	PPMGL_P34_H09	GA16408-PA [D. pseudoobscura]	2.0E-10	Kazal type serine protease inhibitor	EU045342
Cluster 113	B10_PPINFIL_P21	carboxypeptidase A [A. aegypti]	2.0E-87	Carboxypeptidase	EU045341
Cluster 119	PPMGL_P29_B04	midgut specific galectin [P. papatasi]	1.0E-145	Galectin	AAT11557
Cluster 122	C08_PPINFIL_P31	GA15307-PA [D. pseudoobscura]	3.0E-62	Ferritin	EU045340
Cluster 125	D06_PPMGL_P23	Glutathione S-transferase [M. domestica]	3.0E-86	Glutathione S-transferase	EU045339
Cluster 126	PPMGBL175	10 kDa salivary protein [P. ariasi]	6.0E-07	Unknown	EU045338
Cluster 127	H07_ppmgs_p21	ribosomal protein S8 [A. albopictus]	3.0E-94	40S ribosomal protein S8	EU045337
Cluster 128	C02_ppmgs_p21	60S acidic ribosomal protein P2 [A. aegypti]	8.0E-39	60S acidic ribosomal protein P2	EU045336
Cluster 129	PPMGL_P29_C01	ENSANGP00000016569 [A. gambiae]	5.0E-40	membrane LPS inducible TNF protein	EU035828
Cluster 134	C01_PPINFIL_P22	similar to D. melanogaster RpS18 [D. yakuba]	9.0E-70	40S ribosomal protein S18	EU035827
Cluster 135	D04_PPINFIL_P31	trypsin 3 [P. papatasi]	1.0E-144	Trypsin	AAM96942
Cluster 139	PPMGS_P31_A11	CG30415-PB, isoform B [D. melanogaster]	1.0E-27	Unknown	EU035823
Cluster 146	PPMGM132	similar to CG2998-PA [T. castaneum]	2.0E-27	40S ribosomal protein S28	EU035822
Cluster 147	A06_PPMGM_P25	Ribosomal protein L23 [D. melanogaster]	5.0E-73	60S ribosomal protein L23	EU035821
Cluster 149	B06_PPMGM_P25	similar to D. melanogaster RpS12 [D. yakuba]	1.0E-64	40S ribosomal protein S12	EU035820
Cluster 150	PPMGL_P29_D08	ENSANGP00000021011 [A. gambiae]	1.0E-111	Unknown	EU035819
Cluster 153	B07_PPMGS_P24	similar to D. melanogaster CG2033 [D. yakuba]	3.0E-67	40S ribosomal protein S15	EU035818
Cluster 158	D06_PPINFIL_P21	ENSANGP00000013724 [A. gambiae]	5.0E-70	Ryanodine receptor	EU035817
Cluster 163	PPMGM133	cyclophilin isoform [A. aegypti]	8.0E-82	Cyclophilin	EU032351
Cluster 165	A11_PPMGM_P25	60S ribosomal protein L40 [A. albopictus]	4.0E-68	Ubiquitin / ribosomal L40 fusion	EU032350
Cluster 167	PPMGL276	similar to ENSANGP00000002356 [A. mellifera]	2.0E-66	Na+/K+ ATPase	EU032348
Cluster 168	F05_ppmgbm_p21	chymotrypsin [P. papatasi]	1.0E-147	Chymotrypsin	AAM96938
Cluster 171	G12_ppmgs_p21	GA16582-PA [D. pseudoobscura]	8.0E-78	60S ribosomal protein L12	EU032349
Cluster 174	A06_ppmgbm_p22	similar to D. melanogaster CG2099 [D. yakuba]	1.0E-54	60S ribosomal protein L35A	EU032347
Cluster 176	A03_ppmgbm_p24	translation initiation factor SU11-like protein [A. aegypti]	1.0E-54	Translation initiation factor 1	EU032346
Cluster 177	A08_PPMGBS_P28	GA10714-PA [D. pseudoobscura]	5.0E-92	ADP ribosylation factor	EU032345
Cluster 182	PPMGS_P31_E04	similar to D. melanogaster CG10423 [D. yakuba]	2.0E-38	40S ribosomal protein S27	EU032344
Cluster 183	D12_ppmgbm_p22	ENSANGP000000026718 [A. gambiae]	1.0E-61	Cytochrome C oxidase subunit IV	EU032343
Cluster 184	C02_PPINFIL_P30	unknown [C. sonorensis]	4.0E-21	Unknown	EU032342
Cluster 185	D04_PPINFIL_P27	cytochrome b [P. papatasi]	1.0E-110	Cytochrome B	AF161214
Cluster 186	MGL69	similar to CG9916-PA isoform 1 [T. castaneum]	4.0E-81	Cyclophilin	EU032341
Cluster 187	PPINFIL-P7-F11	ribosomal protein S17e [Eucinetus sp.]	9.0E-31	40S ribosomal protein S17	EU032340
Cluster 188	H01_PPMGL_P23	10 kDa salivary protein [P. ariasi]	2.0E-06	Unknown	EU032339
Cluster 201	G11_PPMGBM_P26	larval chymotrypsin-like protein [A. aegypti]	5.0E-82	Chymotrypsin	EU035826
Cluster 228	PPMGM508	peroxiredoxin-like protein [A. aegypti]	1.0E-69	Peroxisome oxidoreductase	EU035825
Cluster 232	PPMGBL92	glutathione S-transferase [A. aegypti]	4.0E-59	Glutathione S-transferase	EU035824
Cluster 243	D11_ppmglbl_p24	midgut chitinase [P. papatasi]	1.0E-140	chitinase	AAV49322

sequence, the third column shows the best match in the non-redundant protein database (GenBank, NCBI), the fourth column shows the E-value for the best matching BLAST result in column 3, the fifth column shows the assigned putative function of that cluster, and the sixth column shows the accession number of the transcript submitted to GenBank. The four most abundant transcripts were microvilli-associated like protein, followed by peritrophin-like protein, 40 S ribosomal protein S30 and a transcript coding for a protein of unknown function. Still, other abundant transcripts include those coding for various ribosomal proteins, chymotrypsins, carboxypeptidases, trypsins, a zinc metalloprotease astacin, a Kazal-type serine protease inhibitor, Glutathione S-transferase (GST) and various proteins of unknown function (Table 4). All the sequences generated from these three cDNA libraries have been deposited as an EST database at the National Center of Biological Information (NCBI), accession numbers ES346912 - ES351350 and ES351429). The following is a more detailed description of relevant transcripts identified in the cDNA libraries:

Microvilli-associated like proteins

Of the most abundant transcripts found in the combined analysis of all three libraries were transcripts coding for proteins with similarities to microvilli membrane proteins from *A. aegypti* and *A. gambiae*. These transcripts are also homologous to major allergens identified in the cockroaches *Blattella germanica* and *Periplaneta americana* [18] and to a nitrile-specifier protein (PrNSP) from the midgut of *Pieris rapae*. PrNSP has a role of converting toxic compounds, such as isothiocyanate, into less toxic

compounds, such as nitriles, that are excreted in the feces of larval stages of this Lepidopteran [19].

Four different putative microvilli-associated proteins were identified in the three *P. papatasi* midgut cDNA libraries (Figure 13). Clusters 1, 2, and 3 represent likely polymorphisms of the same transcript named here “microvilli protein 1” (*PpMVP1*), which has a predicted molecular weight of 23.7 kDa. Another three transcripts coding for microvilli proteins and derived from clusters 94, 96 and 98 were named *PpMVP2*, *PpMVP3*, and *PpMVP4*, respectively. The predicted molecular weight for these microvilli-associated like proteins is 24.0, 25.6, and 25.6 kDa, respectively. Additionally, each of these microvilli proteins has a potential signal peptide as predicted by SignalP 3.0 and no evidence of transmembrane helices as predicted using the TMHMM 2.0 server. Identity between the amino acid sequences of these microvilli proteins ranges from 21 to 36 percent (Figure 13, black-shaded amino acids) and similarity from 45 to 57 percent (Figure 13, grey-shaded amino acids). The degree of conservation may indicate that these are biochemically distinct from one another and only commonly named based on the previous annotation of other organisms with similar sequences. Searching the translated assembled sequences from an EST database of *Lu. longipalpis* identified NSFM-139c08, NSFM-18h11, NSFM-68e08, and NSFM-47h07 as having high sequence homology to the microvilli-associated like proteins *PpMVP1*, *PpMVP2*, *PpMVP3*, and *PpMVP4*, respectively [17].

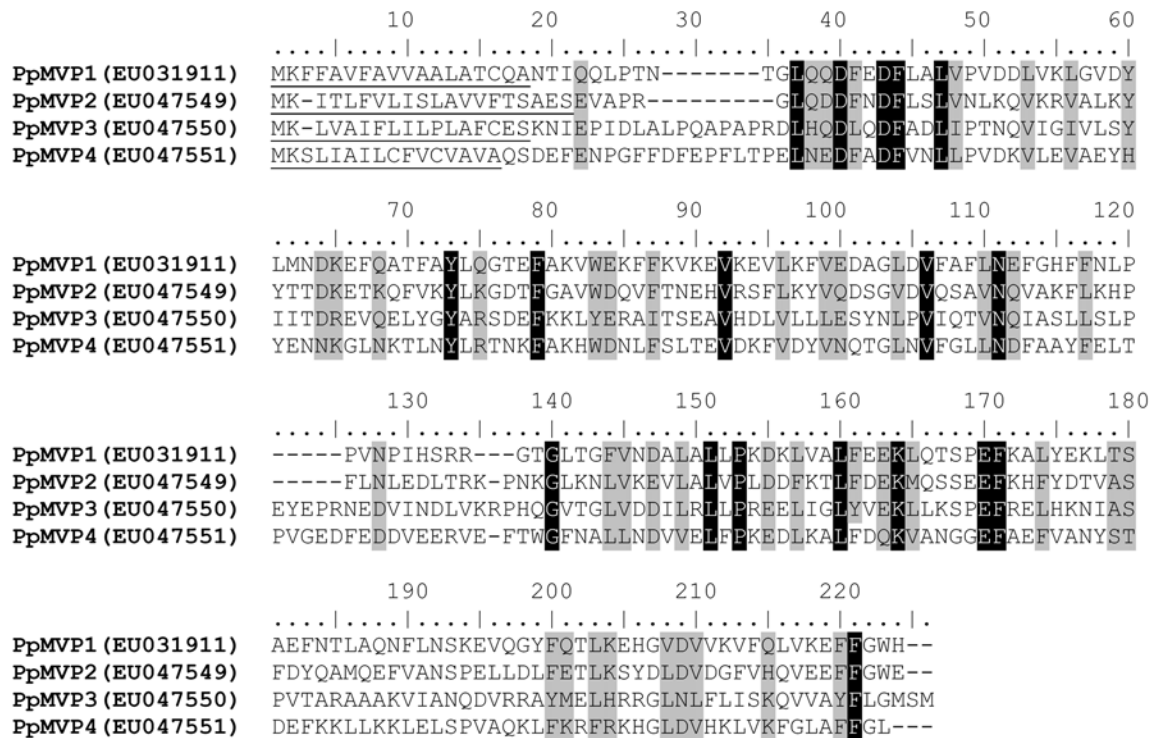


Figure 13. Multiple sequence alignment of the four putative microvilli associated-like proteins found in the midgut of *Phlebotomus papatasi*.

Predicted signal peptide sequence is underlined and the accession numbers given in parentheses.

Peritrophin-like proteins

Transcripts coding for three different putative peritrophin-like molecules were identified in the midgut of *P. papatasi*. *PpPer1* (cluster 9) and *PpPer2* (clusters 12 and 13) transcripts code for secreted proteins with predicted molecular masses of 29.8 and 9.6 kDa, respectively. *PpPer1* is comprised of four potential chitin-binding peritrophin-A domains (Figure 14A). *PpPer2* is a much smaller predicted protein and has only one potential chitin-binding domain (Figure 14A). A third putative peritrophin, *PpPer3*, was identified from cluster 26 with an apparent molecular mass of approximately 32 kDa (Figure 14A) and contains two distant putative chitin-binding domains. Phylogenetic analysis using the chitin binding domains of *PpPer1*, *PpPer2*, *PpPer3* and those of peritrophins from several insects (Figure 14B) suggests a low level of conservation between the domains. Insect peritrophins have been reported to bind to chitin fibers via multiple chitin-binding domains, forming the scaffold that maintains the molecular structure of the peritrophic matrix (PM) in the mosquito gut [20]. In addition to their role in the formation of the PM, peritrophins may also play a role in preventing the toxic effects of heme, a bi-product of blood meal digestion. In *Ae. aegypti*, *AeIMUC1*, a mucin that encodes putative chitin-binding domains was recently shown to bind heme [21]. Although peritrophins have been characterized from several insects, including *Ae. aegypti* and *An. gambiae* [20, 22], no information exists related to sand fly midgut-specific peritrophins. *PpPer1* and *PpPer3* have high sequence similarity, at the protein level, to the translated sequences SFM-03c06 and SFM-02h07 of the *Lu. longipalpis* EST database. However, *PpPer2* has lower sequence similarity to any assembled and

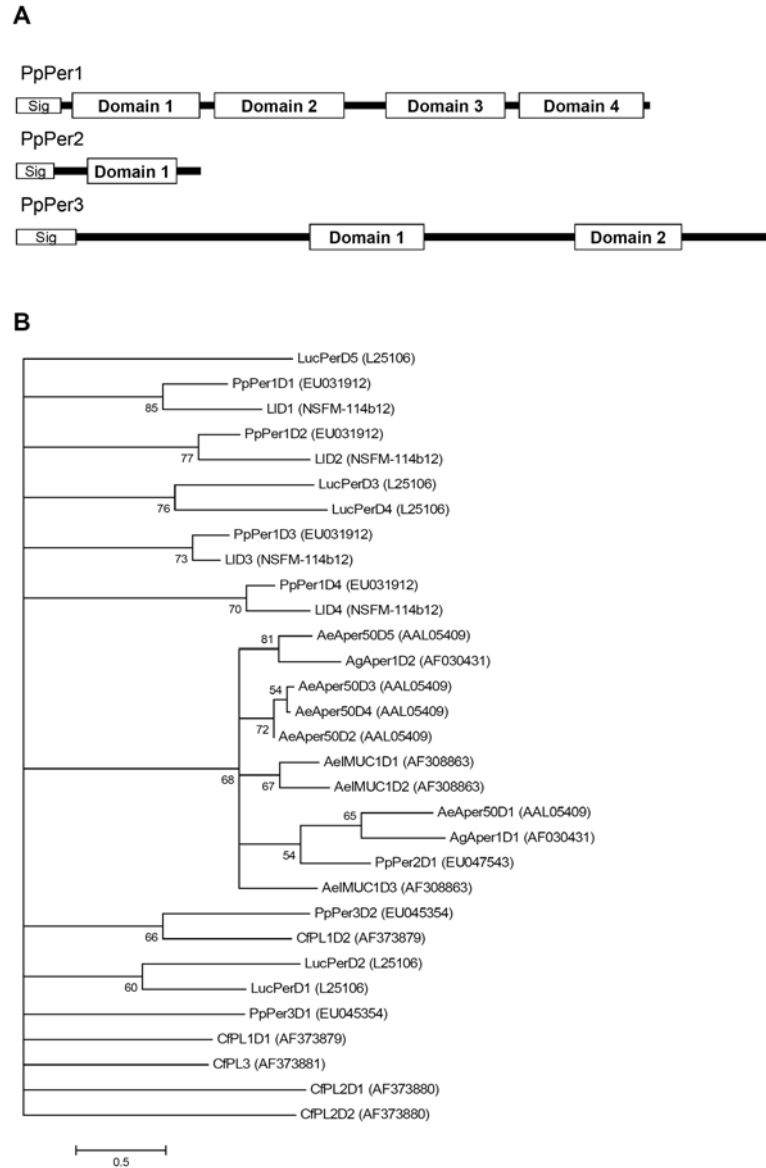


Figure 14. Characterization of peritrophin sequences.

(A) Diagrammatic representation of *Phlebotomus papatasi* peritrophin-like molecules showing the predicted signal peptide and chitin-binding domains. (B) Phylogenetic analysis of chitin-binding domains of peritrophin molecules from *Aedes aegypti* (Ae), *Anopheles gambiae* (Ag) *Ctenocephalides felis* (Cf), *Lucilia cuprina* (Luc), *Phlebotomus papatasi* (Pp), *Lutzomyia longipalpis* (Ll). Accession numbers are indicated in parentheses and bootstrap values at the nodes.

translated sequences of the *Lu. longipalpis* database, suggesting a more divergent or novel molecule.

Trypsin

Among the most abundant transcripts in the cDNA libraries were the previously characterized *P. papatasi* trypsin-like, *PpTryp1*, (Cluster 18 with 158 sequences), and *PpTryp4* (Cluster 89 with 114 sequences) [13]. *PpTryp2* (Cluster 23) and *PpTryp3* (Cluster 135) were less abundant with 12 and 8 sequences, respectively. Phylogenetic analysis of trypsins from *P. papatasi* and from other organisms resulted in the formation of two major clades, each supported by maximum likelihood analysis (Figure 15). *P. papatasi* trypsins co-localized in the clade I containing other insect trypsins, while their mammalian counterparts were found in clade II (Figure 15). As detected previously, [13] *PpTryp1* and *PpTryp2* form a different clade apart from a clade formed by *PpTryp3* and *PpTryp4* (Figure 15). The *P. papatasi* trypsins *PpTryp1*, *PpTryp2*, *PpTryp3*, *PpTryp4*, show high protein sequence similarity to *Lu. longipalpis* ESTs NSFM-02a01, NSFM-113h06, NSFM-94b08, and NSFM-165c07, respectively.

Chymotrypsin

Two previously characterized *P. papatasi* chymotrypsin-like cDNA, *PpChym1* and *PpChym2* [13], as well as a novel chymotrypsin-like, *PpChym3* (Cluster 113) were found also in the transcriptome database. This newly identified novel chymotrypsin-like molecule was found in low abundance in the blood-fed midgut library. The predicted *PpChym3* has 36% amino acid identity to *PpChym1* and 30% amino acid identity to

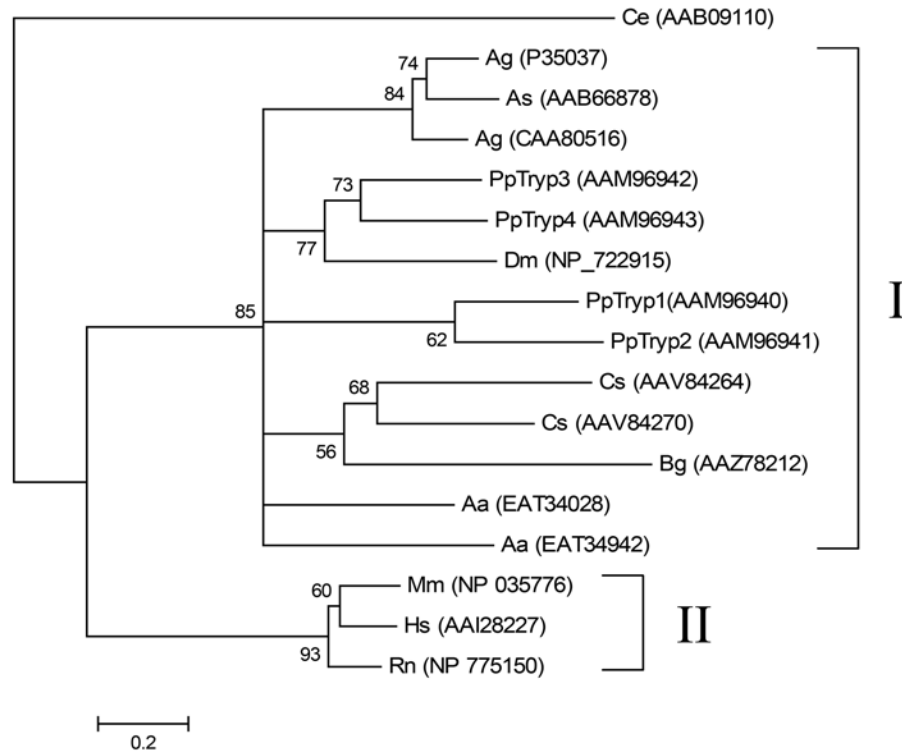


Figure 15. Phylogenetic analysis of trypsins.

Caenorhabditis elegans (Ce), *Rattus norvegicus* (Rn), *Mus musculus* (Mm), *Homo sapiens* (Hs), *Blattella germanica* (Bg), *Anopheles gambiae* (Ag), *Anopheles stephensi* (As), *Aedes aegypti* (Aa), *Drosophila melanogaster* (Dm), *Culicoides sonorensis* (Cs), and *Phlebotomus papatasi* (Pp). The accession number of the sequence used is in parenthesis and node support indicated by the bootstrap values.

Ppchym2. Furthermore, Ppchym3 has a signal secretory peptide (Figure 16A) and has the required His/Asp/Ser amino acid triad necessary for catalytic activity (Figure 16B). Ppchym1 and Ppchym2 both share sequence homology from the assembled sequence NSF01d03 from the *Lu. longipalpis* EST database, while Ppchym3 is most similar to sequence SFM-01b03.

Carboxypeptidase

A number of sequences were identified with homology to carboxypeptidases. The full-length transcript of a putative carboxypeptidase B, *PpCpepB*, was found from 37 sequences in cluster 16 and has high homology to a carboxypeptidase B identified in *Ae. aegypti* (GenBank accession# AAT36733). The predicted amino acid sequence of PpCpepB contains a signal peptide, a propeptide domain, and a carboxypeptidase domain. A putative carboxypeptidase A, *PpCpepA*, was also identified from cluster 113 based on amino acid sequence homology. Phylogenetic analysis shows that the identified *P. papatasi* putative carboxypeptidases are separated into distinct clades (Figure 17A). Comparison of sequence homology indicates the potential for these molecules to have substrate specificities of either carboxypeptidases A or B (Figure 17B). Sequence alignment of the two carboxypeptidases depicts the difference in amino acid composition; however, both sequences contain the zinc ion binding motifs of metallo-carboxypeptidases (Figure 17B). Additionally, the presence of a putative signal peptide alludes that these molecules are midgut digestive enzymes. Similarity between these carboxypeptidases and those present in *Lu. longipalpis* EST database are evident by the high homology between PpCpepA and SFM-05c11 and between PpCpepB and NSF01d09.



Figure 16. Chymotrypsin sequence analysis.

(A) Diagrammatic representation of PpChym3 sequence showing the predicted signal peptide (underlined) and the residues of the catalytic triad (H/D/S) marked with a triangle. (B) Sequence alignment of the three *Phlebotomus papatasi* chymotrypsin-like sequences. Identical residues are highlighted in black and similar residues highlighted in grey. The predicted signal peptides are underlined, the catalytic residues marked with (*) and the accession numbers are in parentheses.

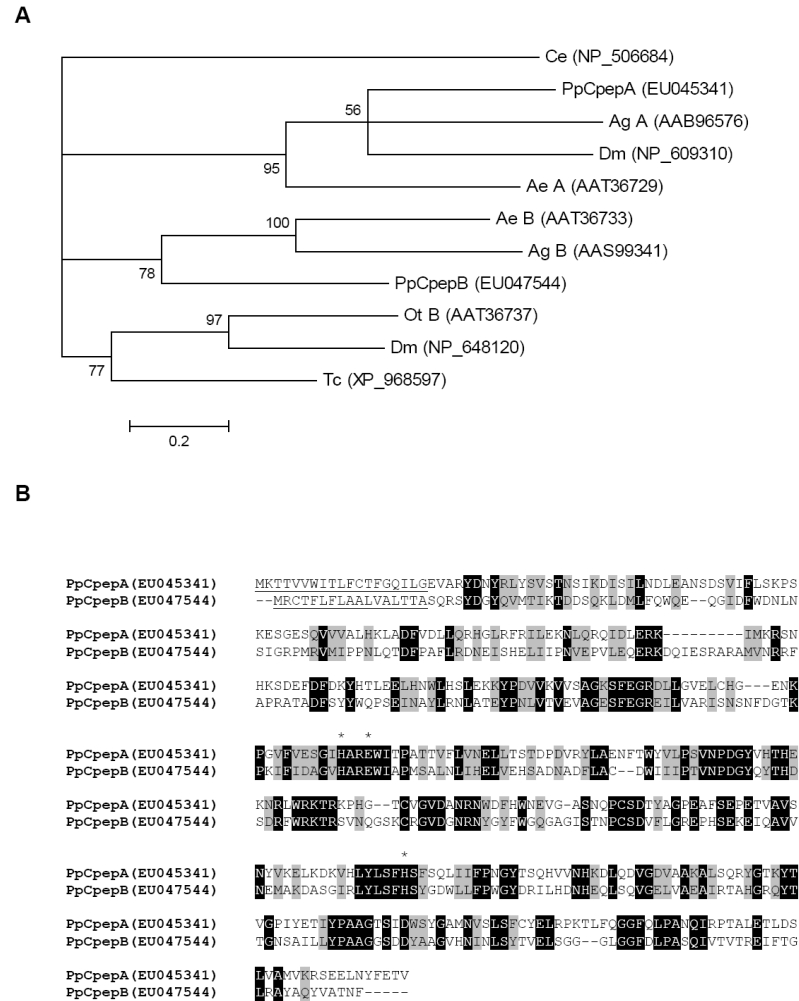


Figure 17. *Phlebotomus papatasi* midgut carboxypeptidase-like proteins.

(A) Phylogenetic analysis of carboxypeptidases from *Caenorhabditis elegans* (Ce), *Aedes aegypti* (Ae), *Anopheles gambiae* (Ag), *Drosophila melanogaster* (Dm), *Ochlerotatus triseriatus* (Ot), *Tribolium castaneum* (Tc), and *Phlebotomus papatasi* (Pp). Accession numbers are indicated in parentheses and node support indicated by the bootstrap values. (B) Sequence comparison of midgut *Phlebotomus papatasi* carboxypeptidase A (PpCpepA) and carboxypeptidase B (PpCpepB). The predicted signal peptide is underlined, and the residues necessary for zinc binding (H and E) are indicated by (*).

Astacin-like zinc metalloprotease

A putative astacin-like zinc metalloprotease (*PpAstacin*) was identified from cluster 37, a product of five sequences. This putative astacin-like protein displays a predicted signal peptide and a slightly modified form of the signature zinc binding catalytic domain for proteins in the astacin family (HEXXHXXGFXHEXXRXDR). In *PpAstacin*, changes in two residues (E to M and R to A) resulted in the motif HEFLHALGFFHMQSASDR (Figure 18). Although the altered residues may be involved in target specificity the zinc-binding catalytic domain remains conserved. The likely role of this putative protein is blood meal digestion, as astacins molecules have not been implicated in immune functions and a considerable number of transcripts constituting this cluster were derived from the blood-fed midgut cDNA library. This is the first report of this type of protease from the gut of a sand fly, though NSF-127b08 of the *Lu. longipalpis* EST database was identified based on sequence homology.

Kazal-type serine protease inhibitor

Two Kazal-type serine protease inhibitors were identified from cluster 111 (*PpKZL1*) and 859 (*PpKZL2*) in the cDNA midgut libraries. *PpKZL1* codes for a small peptide of 78 amino acids while *PpKZL2* codes for a peptide of 89 amino acids. Both proteins are predicted to be secreted based on the presence of signal peptides (Figure 19). *PpKZL1* is similar to various small Kazal-type inhibitors found in *Drosophila pseudoobscura* (gi: 125986397), *C. sonorensis* (gi: 56199538) and the mosquitoes *Ae. aegypti* and *An. gambiae*, and to larger Kazal-type molecules such as infestin [23] from *Triatoma infestans* (Figure 19A). There is only 28% identity and 42 % similarity

Figure 18. Multiple sequence analysis of astacin-like proteins.

Sequence alignment of zinc proteases astacin-like sequences from *Phlebotomus papatasi* (Pp), *Aedes aegypti* (Ae), *Anopheles gambiae* (Ag), *Culicoides sonorensis* (Cs), *Drosophila melanogaster* (Dm), *Glossina morsitans morsitans* (Gm), *Astacus astacus* (As), *Caenorhabditis elegans* (Ce), *Mus musculus* (Mm), and *Homo sapiens* (Hs).

Arrows indicate the residues likely necessary for catalytic activity. Accession numbers are shown.

```

      70      80      90      100     110     120
Pp EU045349 -----GNFEGDMILSPRQMDLR-----FRTGLINLKYRWPNK-
Ae EAT36347 -----GNFEGDMILSKEQRQALAG-----MRNGLFDDQYRWPNN-
Ag XP_318543 -----GQFEGDIVLSEEQERSLSN-----RRNGLIATYRWPGN-
Cs AAV84221 -----GKFEVDIELTPEQIRAMK-----QRNGLLLVTKRWNNN-
Dm AAY55427 -----GFVEGDMMLTEEQQRNLEQGA-----PKARNGLINTEKRWPGN-
Gm AAK07478 -----GYVEGDMVLNSRQ-----RNLGRDEVWRWPNN-
As CAA64981 -----DGMFEGDIKLRAGQPARGV-----AAILGDEYLVSGG-
Ce AAB53827 IKVKDDPTIGNYSEGDILLESPPKKFVEENNK-----LGRNAIKQIYRRWPNN-
Mm CAM13879 -----VFQDRDIPAINQGLISEETPESSFLVEGDIIRPSPFRLLSVTNKKWPKGV
Hs AAI07128 -----ASGDKDIPAINQGLILEETPESSFLIEGDIIRPSPFRLLSATSNKKWPMGG

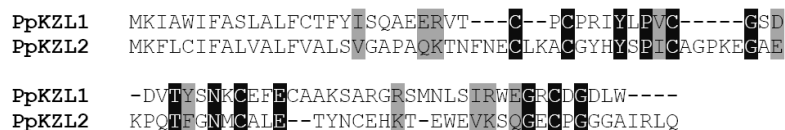
      130     140     150     160     170     180
Pp EU045349 ---LVPYQLSSE-FTREESEFIREALDSIECVSCLRFVEKNSSH--SDFVKVSREVDSGC
Ae EAT36347 ---TVYYRIISDNFTTEQVNYIRRLDTISDVSCIRFVEAAENS--TAYIRVLGN-EGGC
Ag XP_318543 ---TVPVMIVEEDFTPEQIEHIIKRLRQIESVTCLFVTRTEEP---DYVRVIGT-GSGC
Cs AAV84221 ---TVDYIITGN-YSQEQKNYIRKGLDTLQLVTCLFIGHDNATGLTDYVEVSS-GGGC
Dm AAY55427 ---VVYRISDD-FDTAHKKAITGTIDTLELHTCLRFREATDED--KAYLTVTAK-SGGC
Gm AAK07478 ---TVYKFFTV-FDEDHNYILRGMKIIIEISCIRFEEADATT--PNYVNTIGF-VGGC
As CAA64981 ---VIPYTFAG-VSGADQSAILSGMQELEEKTCIRFVPRTTES---DYVEIFTS-GSGC
Ce AAB53827 ---EIPYTLSSQYGSYARSVIANAMNEYHTKTCVKEFVARDPK-HHDYLVWHPD--EGC
Mm CAM13879 GGFVEIPFLLSRKYDELSRRVIMDAFAEFERFTCIRFVAYHGQR---DFVSLIPM--AGC
Hs AAI07128 SGVVEVPFLLSSKYDEPSQVILEALAEFERSTCIRFVTYQDQR---DFISILIPM--YGC

      190     200     210     220     230     240
Pp EU045349 FSSVGYQAG-EQQNLNAPNELGTGCFRKGTTIHEFLHALGFYHMQSASIRDDYVTVVWEN
Ae EAT36347 FSEVGYTGT-VQDLNAPNELENGCFRLGTIMHEFLHALGFYHMQSASIRDDFVTVVWEK
Ag XP_318543 YSSVGHRRG-AQTLNLEPYDVTGCFRLATIVHEFLHAGFYHMQSASIRDDFVQTVVDN
Cs AAV84221 SSTVGRKGG-RQTLNLQSYFVEEGCFRLATIMHEFLHALGFYHMQSTYNRDEYVDVKYEN
Dm AAY55427 YTAAGYQGA-PQEMNLEIYPLGEGCFRPGTILHEFMHALGFYHMQSSIRDDFINVIYEN
Gm AAK07478 YSEVGLWLEGAQAYNLEMYALDTGCFRLGTIVHEFLHTLGFYHMQSATNRDDYVHVVEGN
As CAA64981 WSYVGRISG-----AQQVSLQANGCVYHGTTIHELMHALGFYHEHTRMDRDNVVTINYQN
Ce AAB53827 YSLVGKTGG-----KQPVSLSGCIQVGTIVHELMHALGFYHMQSRQERDSYIDVYVWQN
Mm CAM13879 FSGVGRSGG-MQVVS LAPTCLRK---GRGIVLHELMHVLGFWHEHSRADRDRIYQVNWNE
Hs AAI07128 FSSVGRSGG-MQVVS LAPTCLQK---GRGIVLHELMHVLGFWHEHTRADRDRIYQVNWNE

      250     260     270     280     290     300
Pp EU045349 INPQHVHNFKKYNESVITHFGVKYDYSSVMHYHKTAFSMND--EDTIIVP---KDPNAEIG
Ae EAT36347 IEQQHQHNFKEYNSSFVSAFNVEYDYGSVLHYPRVSFSIDG--SATIIP---KVAGVTIG
Ag XP_318543 IEDGKEHNFNIYDSDTVDFSVQYDYSSVMHYSSTAFSKNG--EKTIVP---KDPNATIG
Cs AAV84221 IEPGKENNFNKYTETDVTVDYGIEYDLYNSVMHYGRTGFSING--EFTLVPI--KDPEAKIG
Dm AAY55427 IVPGKEFNFKYADTVVTDFEVGYDYSSCLHYRPGAFSING--EDTIIVPL---DSSAVIG
Gm AAK07478 IDPRNLHNFNKYNETQVNDFDQBYDYSSVMHYGPKAFSING--EDTIIPLYENEAAGNMG
As CAA64981 VDPSMTSNFDIDTYS--RYVGEDYCYYSIMHYGKYSFSIQWGVLETIVPLQNGIDLTDPY
Ce AAB53827 VMNGADDQFEKYNLNVISHLDEPYDYASIMHYGPYAFSGSG--KKTIVPK---KSGSERMG
Mm CAM13879 ILPGFEINFIKRSR---TNMLVPYDYSSVMHYGRFAFSWRG--QFTIIPL--WTSSVHIG
Hs AAI07128 ILPGFEINFIKRSR---SNMLTPYDYSSVMHYGRFAFSRRG--LFTIIPL--WAPSVHIG

      310     320     330     340     350     360
Pp EU045349 QRIGLSDGDIKRLNKMYQCDEM-----
Ae EAT36347 QRKEMSTSDITKLNRMHYHCE-----
Ag XP_318543 QRVGMSERDISKLNHMYKC-----
Cs AAV84221 QRVGLSRDIEKLNKMYDCPL-----
Dm AAY55427 QRVGLSSKIDIDKINIMYKCPILL-----
Gm AAK07478 QRLGMSEKDINKLNLMYRCPIEV-----
As CAA64981 DKAHMLQTDANQINNLYTNECSLRH-----
Ce AAB53827 QRVKFSDDIDVRKINKLYNCPGVSGNNNNNNNNQINSNSIVNHPQV-----
Mm CAM13879 QRWNLSSTDITRVCRLYNCSRSVPDSHGRGFEAQSDGSSLTASISRLQRLLLEALSEESG
Hs AAI07128 QRWNLSASDITRVLQLYGCSPSGPRPRGRGSHAHSTGRSPAPASLS-LQRLLLEALS AESR

```

A**B****Figure 19.** Sequence analysis of Kazal-type proteins.

(A) Sequence alignment of Kazal-type proteins from *Phlebotomus papatasi* (Pp), *Aedes aegypti* (Ae), *Culicoides sonorensis* (Cs), *Drosophila melanogaster* (Dm) and *Triatoma infestans* (Ti). The predicted signal peptide sequences are underlined and the conserved cysteine residues denoted by #. Identical residues are highlighted in black and similar residues highlighted in grey. PpKZL1 accession number is EU045342 (B) Sequence comparison of the two Kazal-type proteins (PpKZL1 and PpKZL2) from *Phlebotomus papatasi* found in the midgut cDNA libraries. Identical residues are highlighted in black and similar residues highlighted in grey.

between PpKZL1 and PpKZL2 (Figure 19B) suggesting these may have different functions. Additionally, these two Kazal-type cDNAs are similar to two previously characterized thrombin inhibitors, rhodniin and infestin, from the triatomines *Rhodnius prolixus* [24] and *T. infestans* [23], respectively. Due to their anti-hemostatic effect, rhodniin and infestin are believed to play a role in the fluidity of the blood within the midgut of these vectors. It is conceivable that one or both transcripts coding for Kazal-type thrombin inhibitors identified in *P. papatasi* may play a role in blood fluidity within the sand fly midgut, allowing it to be fully digested by the various proteases secreted within the midgut following the blood meal. These represent the first Kazal-type serine protease inhibitors identified from sand flies. PpKZL2 shares low sequence similarity with SFM-0406 from the *Lu. longipalpis* EST database and no significant similarities were identified for PpKZL1.

Ferritin

Two transcripts encoding putative ferritin light (*PpFLC*) and heavy (*PpFHC*) chain subunits were identified in clusters 103 and 122, respectively (Figure 20). After the ingestion of a blood meal, the fly encounters a tremendous dose of iron and heme, which would be fatal to most organisms. Ferritin is one of the important factors in controlling the high iron load in hematophagous insects. The midgut of blood-feeding insects envelopes the blood meal and consequently makes the midgut tissue the most likely site of iron regulatory molecules. However, ferritin also may be important for oxidative stress not related to the presence of iron or heme, as it is induced by the presence of H₂O₂

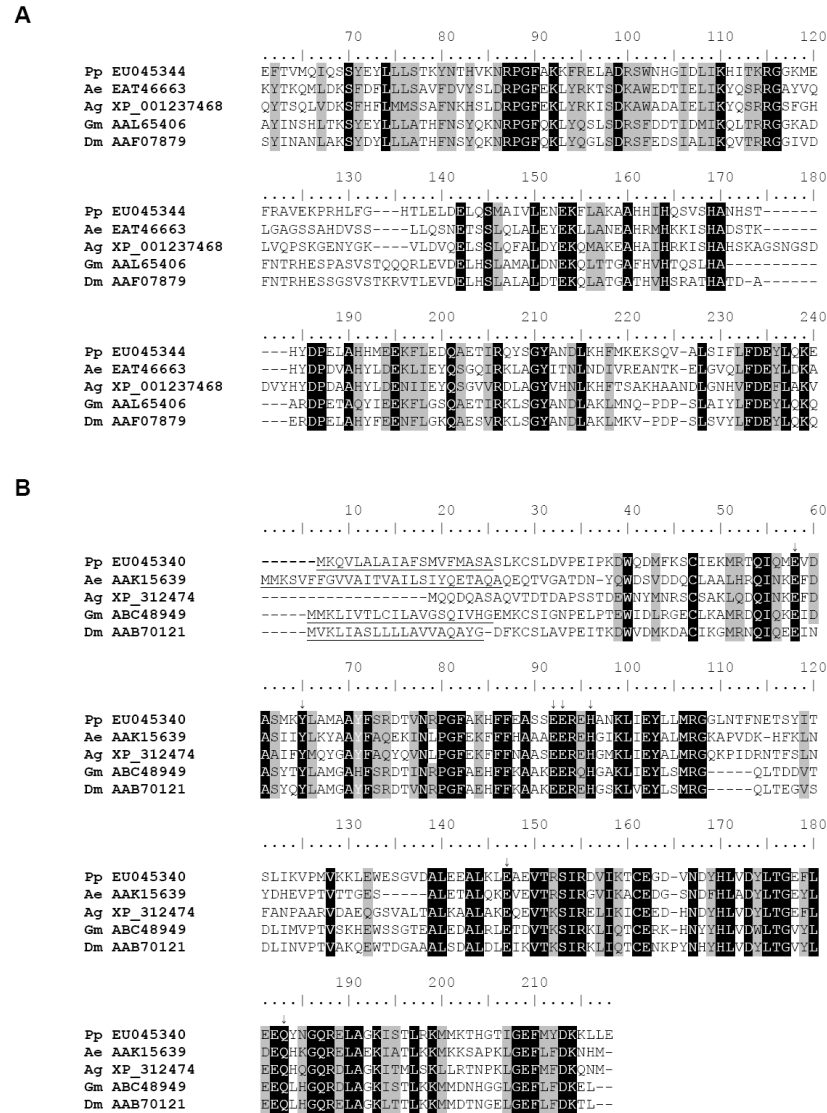


Figure 20. Sequence analysis of ferritin heavy and light chain molecules.

Sequence alignment of sequences from *Aedes aegypti* (Ae), *Anopheles gambiae* (Ag), *Glossina morsitans morsitans* (Gm), *Drosophila melanogaster* (Dm), and *Phlebotomus papatasi* (Pp). (A) Light-chain ferritin subunits. (B) Heavy-chain ferritin subunit. Arrows indicate residues associated with the ferroxidase center, the predicted signal peptide sequence is underlined and the accession numbers are given.

in *Ae. aegypti* [25]. PpFLC and PpFHC are similar to NSF-144g07 and NSF-146d09, respectively; molecules identified by searching the *Lu. longipalpis* EST database.

Glutathione S-transferase (GST)

From clusters 125 and 232, two transcripts were identified to encode putative GSTs with homology to other Dipteran GSTs in the Sigma and Delta/Epsilon classes, respectively. The predicted molecular weights of the two putative proteins are similar at 23.2 kDa for cluster 125 and 24.5 kDa for cluster 232. Within the midgut, these proteins may play an important role in the regulation of reactive oxygen species which occur as a by-product of hemoglobin digestion. Cluster 125 and 232 share high protein sequence similarity with *Lu. longipalpis* ESTs NSF-105e10 and NSF-74c11, respectively.

Unknown proteins

A large number of clusters produced by the three cDNA libraries have no sequence similarity to other known proteins. This has also been observed in the analysis of the *Chironomus tentans* midgut, with good evidence that the unknown transcripts contained coding sequences [26]. It is also possible that the abundance of unidentifiable sequences may be caused by the sequence quality of the transcripts or that the captured sequences are 3' untranslated regions, non-coding small nuclear RNA, or sequences of uncharacterized organisms such as bacteria and yeast present in the sand fly midgut. A number of clusters with unknown functions were identified as coding sequences that exhibited signal peptides, such as clusters 11 and 126.

Functionally characterized proteins

From the three cDNA libraries, we identified chitinase transcripts that were then expressed as recombinant proteins for the demonstration of activity in the midgut of *P. papatasi* sand flies [15]. Another product of the cDNA libraries was the identification and characterization of a galectin protein as the first arthropod receptor for a parasite; specifically, *L. major* within the *P. papatasi* sand fly midgut [1].

Comparative analysis of transcripts that significantly differ from the sugar-fed and blood-fed midgut cDNA libraries

To investigate the effects of blood-feeding on the midgut expression profile in *P. papatasi*, we compared the abundance of transcripts in sugar and blood-fed cDNA libraries. We hypothesized that a blood meal will have an effect on the expression of sand fly midgut transcripts that will be reflected in the relative abundance of sequences forming a cluster in the two libraries. Chi-square statistical analysis was used to evaluate the significance of the differences in the abundance of midgut transcripts from unfed and blood-fed cDNA libraries; thereby, identifying different expression profiles of selected midgut transcripts in each cDNA library.

We observed a significant difference (P value ≤ 0.05) in the abundance of a number of midgut transcripts when we compared the sugar-fed and blood-fed sand fly midgut cDNA library. Table 5 shows a list of selected transcripts that were either more abundantly or less abundantly expressed in these two cDNA libraries.

Table 5: Clusters overrepresented in the sugar-fed and blood-fed midgut cDNA libraries as determined by χ^2 statistical analysis

Putative function	Sugar fed	Blood fed	P value	Genbank
Microvilli protein (PpMVP1)	0	195	4.3E-58	EU031911
Microvilli protein (PpMVP2)	1	60	1.9E-17	EU047549
Microvilli protein (PpMVP3)	39	8	1.1E-04	EU047550
Microvilli protein (PpMVP4)	0	18	2.4E-06	EU047551
Peritrophin (PpPer1)	0	54	1.9E-16	EU031912
Peritrophin (PpPer2)	152	45	1.1E-10	EU047543
Ferritin light chain (PpFLC)	6	18	2.9E-03	EU045344
Chymotrypsin (Ppchym2)	0	36	2.1E-11	AY128107
Trypsin (PpTryp1)	86	10	4.4E-14	AY128108
Trypsin (PpTryp4)	0	52	6.9E-16	AY128111
Unknown (Cluster 73)	13	21	4.6E-02	EU045347
Unknown (Cluster 99)	0	29	1.9E-09	EU045345

As expected, transcripts coding for proteolytic enzymes such as trypsin (*PpTryp4*), and chymotrypsin (*PpChym2*) were more abundantly represented in the blood-fed cDNA library than in the sugar-fed cDNA library (Table 5). Other transcripts coding for peritrophin and microvilli-like proteins and ferritin also were more abundantly represented in the blood-fed cDNA library. Also, we observed a number of transcripts that were less abundantly represented in the blood-fed cDNA library, such as trypsin 1 (*PpTryp1*), and peritrophin (*PpPer2*).

Validation of transcript abundance of selected sequences by real-time PCR

In order to validate the results observed by the chi-square analysis, we further characterized several transcripts by semi-quantitative end-point reverse-transcriptase PCR as well as by real-time PCR. These were utilized to assess the relative abundance of transcripts in the midgut tissue under sugar-fed and blood-fed conditions. The investigated transcripts included peritrophins *PpPer1* and *PpPer2*, as well as microvilli proteins *PpMVP1*, *PpMVP2*, and *PpMVP4*.

The results of semi-quantitative PCR can be seen in Figures 21B and 21D and the induction of *PpPer1* is clearly seen. *PpPer2* abundance between the two midguts conditions is less clear. Figure 21A shows the transcript abundance of *PpPer1* as fold abundance over the control gene in non blood-fed and post-blood meal ingestion as measured by real-time PCR. Figure 21C shows the same real-time PCR analysis of the *PpPer2* transcript. The profile of the peritrophin transcripts by real-time PCR strongly correlates with the profile found in the libraries based on the number of sequences.

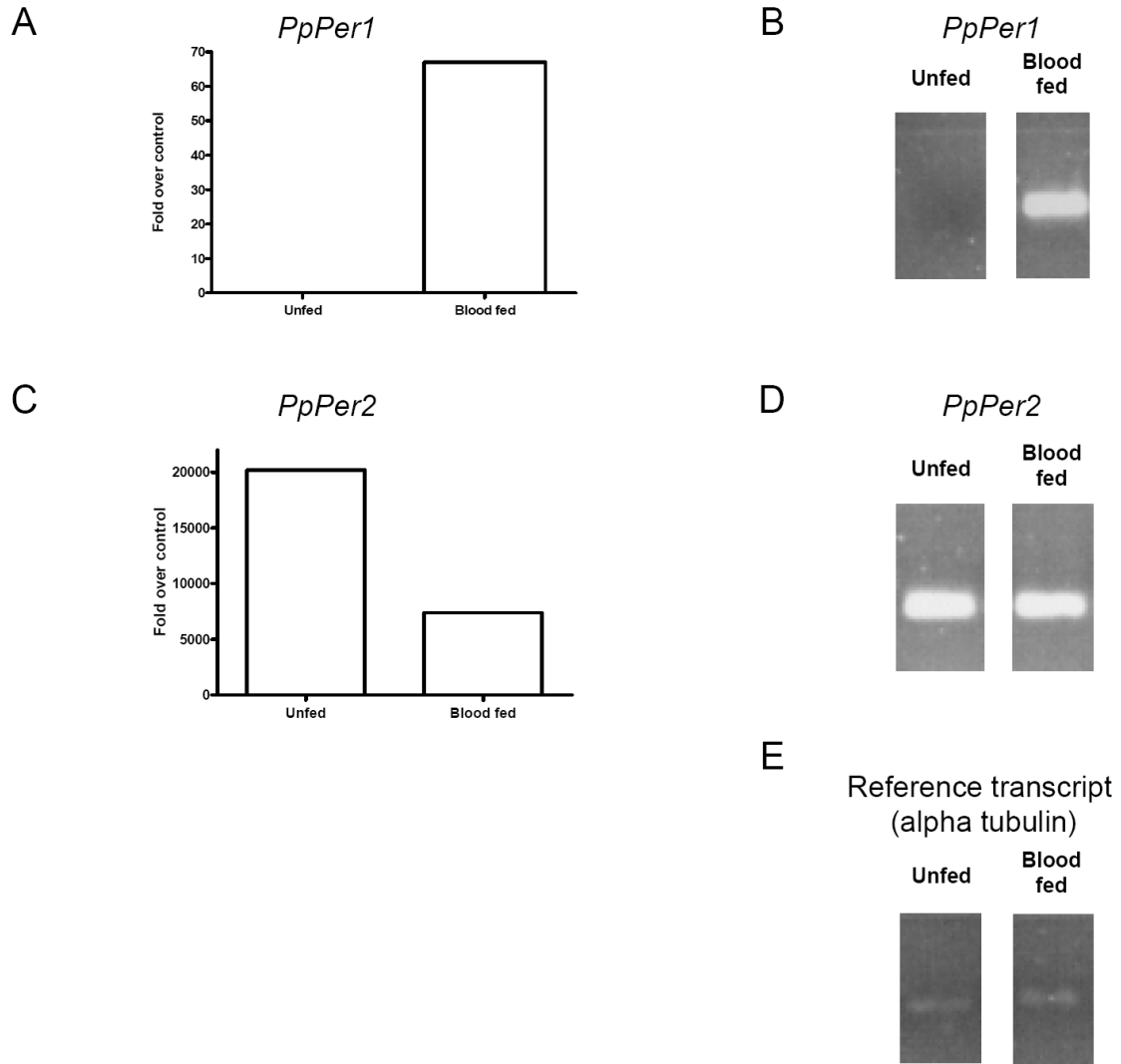


Figure 21. Comparative abundance of peritrophin transcripts in sugar-fed or blood-fed sand flies.

(A, C) *PpPer1* and *PpPer2* transcripts fold over control (reference transcript = alpha tubulin) in unfed and blood-fed *P. papatasi* midgut. (B, D) Semi-quantitative PCR amplified *PpPer1* and *PpPer2* transcripts separated by agarose electrophoresis.

Based on real-time PCR, *PpPer1* expression is induced by blood digestion, and it is not detected in sugar-fed midguts, corresponding with the lack of any sequences produced in the sugar-fed midgut cDNA library, compared to 54 sequences found in the blood meal library. As predicted by the high sequence abundance of *PpPer2* in the sugar-fed cDNA library the expression of this transcript is highest in unfed sand flies and seems to be downregulated by the ingestion of a blood meal.

Transcription levels of mRNAs coding for microvilli-like proteins (*PpMVP1*, *PpMVP2*, and *PpMVP4*) tested by semi-quantitative PCR and real-time PCR are shown in Figure 22 and illustrate the induction of transcription by the ingestion of a blood meal. This mirrors what is seen by the sequence abundance of the cDNA library, in which only one sequence of *PpMVP2* was observed in the sugar-fed cDNA library. The remaining sequences were contributed by the cDNA library produced from blood-fed sand flies.

Pptryp1 and *Pptryp4* low and high transcript abundance, respectively, were in accordance with the results of previously published endpoint reverse-transcriptase PCR [13]. Additionally, the previously characterized chitinase molecule, *PpChit1*, was identified in cluster 243 produced by three sequences contributed by the blood-fed cDNA library with none present in the sugar-fed cDNA library. The mRNA expression levels of *PpChit1* peak at 72 hours post-blood meal ingestion [15].

Comparative analysis of transcripts significantly differs from the blood-fed and *L. major*-infected midgut cDNA libraries

During its development within the midgut of the sand fly, *Leishmania* is faced with various potential barriers that may prevent the establishment of the infection.

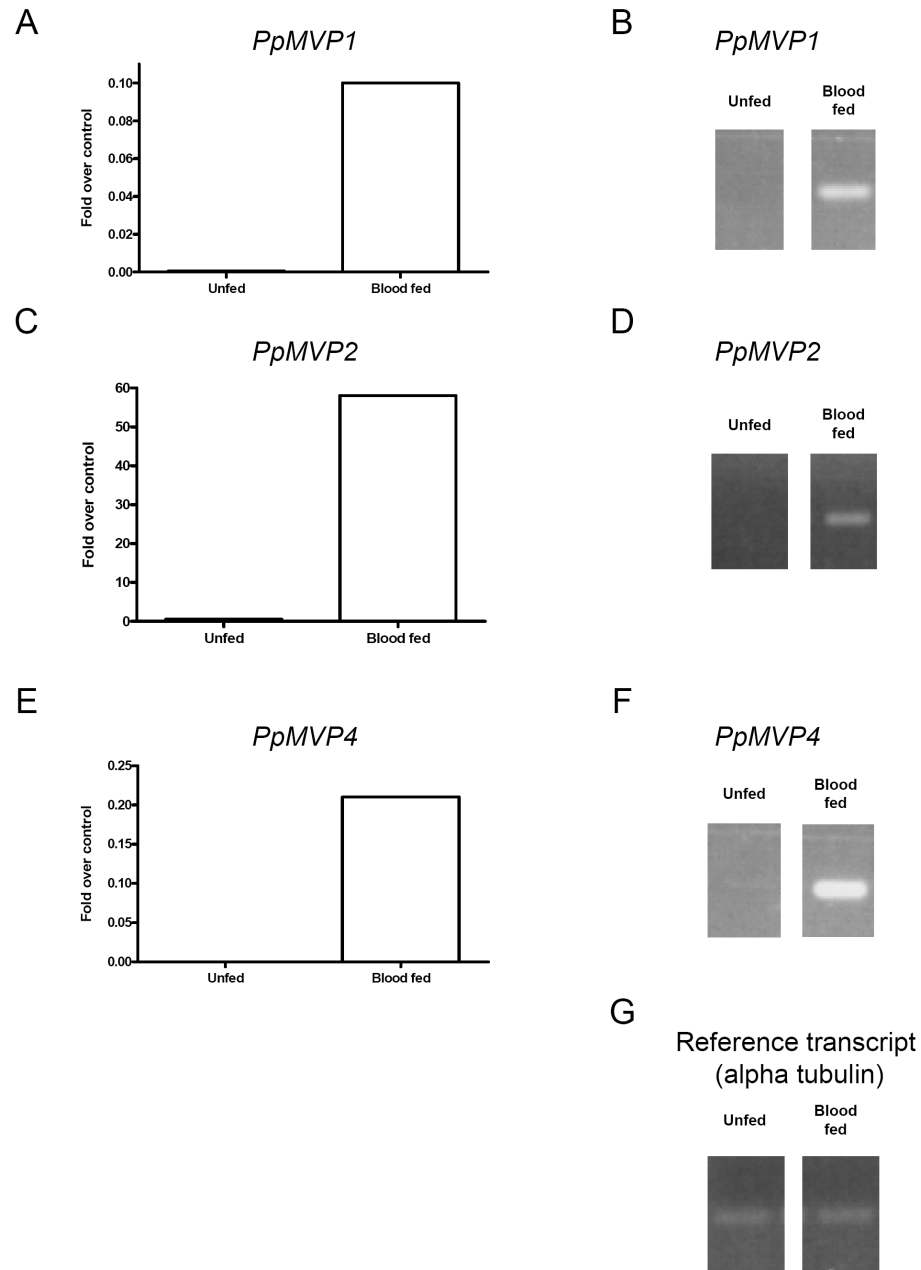


Figure 22. Transcript abundance of microvilli associated-like proteins compared between unfed and blood-fed sand flies.

A, C, E: *PpMVP1*, *PpMVP2*, and *PpMVP4* transcript fold over control (reference transcript = alpha tubulin) in unfed and blood-fed *P. papatasi* midgut. B, C, F: *PpMVP1*, *PpMVP2*, and *PpMVP4* semi-quantitative PCR amplified transcripts separated by agarose electrophoresis.

Table 6: Clusters overrepresented in the blood-fed and *Leishmania major*-infected sand fly midgut cDNA libraries as determined by χ^2 statistical analysis

Putative function	Blood fed	<i>L. major</i>	P value	Genbank
Microvilli protein (PpMVP1)	134	70	5.8E-07	EU031911
Microvilli protein (PpMVP2)	60	42	4.1E-02	EU047549
Peritrophin (PpPer1)	54	16	1.7E-06	EU031912
Peritrophin (PpPer2)	45	35	1.8E-02	EU047543
Ferritin light chain (PpFLC)	18	3	7.1E-04	EU045344
Chymotrypsin (Ppchym2)	36	8	1.1E-05	AY128107
Trypsin (PpTryp1)	10	82	1.0E-13	AY128108
Unknown (Cluster 73)	21	6	2.6E-03	EU045347
Unknown (Cluster 99)	29	5	1.9E-05	EU045345

Among such potential barriers are digestive proteases (trypsins and chymotrypsins), the peritrophic matrix and the requirement for parasite attachment to the midgut epithelia to prevent excretion of parasites with remnants of the digested blood. Previous data suggested that *Leishmania* is able to downregulate proteolytic activity in the sand fly midgut [4]. Also, chitinases produced either by the sand fly [15] or by the *Leishmania* [27] facilitates parasites in the escape from the peritrophic matrix. Attachment to the midgut epithelia occurs via the presence of *L. major* lipophosphoglycan receptors, such as PpGalec [1] or, in the case of permissive sand flies, via the presence of midgut glycoproteins bearing terminal N-acetyl-galactosamine [28].

In sand flies, only a handful of midgut proteins have been clearly implicated in *Leishmania* development. Previous data indicated that *Leishmania* is able to manipulate the activity of certain digestive proteases, inhibiting or delaying their peak activity, possibly in order to survive the proteolytic attack it faces in the midgut of the vector [3, 27]. We hypothesized that a blood meal containing *L. major* will affect the expression profile of midgut transcripts altering the abundance of the different transcripts in each of these cDNA libraries. Table 6 shows the results of the chi-square analysis when transcripts from the blood-fed and *L. major*-infected blood-fed cDNA libraries were compared. Of interest, the abundance of transcripts coding for proteolytic enzymes were dramatically decreased in the midgut cDNA library of sand flies fed on *L. major*-infected blood. Additionally, other transcripts that also appear to have their number reduced included those coding for microvilli-associated like proteins and peritrophins. Transcripts such as the one corresponding to *PpTryp1* (trypsin 1) and one corresponding to *PpPer2* (peritrophin 2) were more abundant. Other transcripts coding for unknown

proteins also were less abundant in the *L. major*-infected cDNA library than in the blood-fed cDNA library. These data suggest that the parasite may be affecting the expression profile of these transcripts, and this inhibition, particularly of proteolytic enzymes, may be advantageous for the survival and establishment of the parasite in the midgut of the sand fly.

Conclusion

Development of *Leishmania* within its sand fly host is largely restricted to the vector midgut. Within the midgut *Leishmania* begins its development confined within a peritrophic matrix and subjected to the onslaught of digestive enzymes. Later, the parasites attach to the epithelia to prevent excretion with remnants of the blood meal and detach as they develop into the infective metacyclic form before being transmitted to a suitable host during a subsequent blood meal. The sand fly midgut presents to the parasite a number of biological barriers the *Leishmania* parasite must circumnavigate or defeat to proliferate and develop inside the insect vector. Acquiring a better understanding of the molecules present in this organ will illuminate the potential molecular interactions occurring between the *Leishmania* parasite and the sand fly vector. Comparative transcriptome analysis provides a powerful global approach as demonstrated by the repertoire of molecules identified from a whole organism or from a specific tissue and the generation of new hypotheses from these data. Large scale genome analyses benefit from data generated from transcriptome analyses, for example, by aiding in the annotation of exons and introns.

The results of the present work provide insights into the repertoire of the molecules present in the midgut of the sand fly *P. papatasi*, the natural vector of *L. major*. We identified a variety of molecules and obtained high quality, full-length sequences from many of them. The high quality sequences were deposited at NCBI, significantly augmenting the available midgut-specific coding sequences. A large number of non-annotated sequences were deposited in the EST database for the scientific communities to access these transcripts.

The global changes in sand fly midgut expression profile were assessed by comparing data generated from randomly sequenced midgut cDNA clones obtained from cDNA libraries of adult females fed on sugar only, blood or blood with the addition of *L. major*. Our approach allowed for the identification of transcripts that are induced by blood-feeding and likely participate in the digestion of the blood meal and events leading to egg production. Digestion of blood as a nutritional source is complicated by the cellular and molecular response and components of the blood itself, once ingested by the insect vector. Transcripts identified in the *P. papatasi* midgut, such as ferritin, Kazal-type serine protease inhibitors, and GST, are examples of the molecules identified on the gut of this insect. Additionally, the inclusion of a *L. major*-infected midgut cDNA library provides insight into genes potentially regulated by this parasite during its development within the sand fly midgut. The random sequencing approach followed by the *in silico* analysis of the transcript abundance was supported by experimental analyses obtained via real-time PCR.

Overall, this analysis will contribute to the understanding of the molecular interactions between *Leishmania* and the sand fly vector and may open new avenues for basic research towards the control of this neglected vector-borne disease.

Methods

Sand flies

Phlebotomus papatasi sand flies (Saudi Arabia strain) were obtained from colonies maintained at Walter Reed Army Institute for Research (WRAIR) and at NIAID-NIH. Three- to 5-day old female sand flies were fed either on 20% sucrose solution (sugar-fed) or on BALB/c mouse whole blood, via artificial meals [1], with or without the addition of 2×10^6 *L. major* (V1 strain) amastigotes per ml.

Messenger RNA extraction and cDNA library construction

Phlebotomus papatasi female midguts (10 midguts) were dissected from sugar-fed only, from blood-fed at 6h (6 midguts), 24h, 48h and 72h post-blood meal PBM (5 midguts each) and from *L. major*-infected at 16h (3 midguts), 22h and 96h (5 midguts each) post infection (p.i.). For blood-fed and for *L. major*-infected, groups of midguts were pooled for RNA extraction. Pooling was done for the sugar-fed group as well. Messenger RNA was purified with the Micro-FastTrack mRNA isolation kit (Invitrogen-Life Technologies, Carlsbad, CA), and 100 ng of mRNA was used to produce a first strand cDNA. A cDNA library, enriched for full-length cDNA, was synthesized using the SMART cDNA library construction kit (Clontech Laboratories, Mountain View, CA). One microgram of double stranded DNA for each original library (sugar-fed, blood-fed,

L. major-infected) was fractionated using a Chromaspin 1000 column (Clontech Laboratories, Mountain View, CA) into small (S), medium (M) and large (L) transcripts based upon their electrophoresis profile on a 1.1% agarose gel. Pooled fractions were ligated into Lambda TriplEx2 vector (Clontech, Mountain View, CA) and packaged into lambda phage (Stratagene, La Jolla, CA). Individual libraries were plated on LB agar plates in order to achieve roughly 200–300 plaques per 182mm plates.

Random sequencing

Unidirectional sequencing of randomly selected clones was completed as previously described [10]. Single, isolated plaques were picked from the plate using sterile wooden sticks and placed into 70µl of water. Amplification of the cDNA was performed using Platinum PCR SuperMix (Invitrogen), 4µl template, and primers PT2F1 (AAG TAC TCT AGC AAT TGT GAG C) and PT2R1 (CTC TTC GCT ATT ACG CCA GCT G). PCR amplification products were cleaned using either Multiscreen PCR cleaning plates (Millipore) or Edge Biosystems PCR cleaning plates and three washes with ultra pure water. The cleaned PCR product was resuspended in 25µl of water, of which 4µl were used for cycle sequencing with PT2F3 primer (TCT CGG GAA GCG CGC CAT TGT) and either DTCS reaction kit (Beckman) or Big Dye 3.1 (Applied Biosystems). Sequencing reaction products were cleaned using Sephadex G-50 (GE Healthcare) in a multiscreen cleaning plate (Millipore) and analyzed using either CEQ8000 (Beckman Coulter) or ABI3700 (Applied Biosystems) DNA sequencing instrument.

Bioinformatic analysis

Detailed description of the bioinformatic analysis of the data appear in [10, 29]. Briefly, prior to analysis the vector sequence was removed from the cDNA nucleotide sequences. Sequence data from the three libraries were grouped together and aligned to generate clusters of contiguous sequences or contigs based on 90% homology over 90 nucleotides, after sequences with more than 5% Ns were discarded. Three frame translations of the consensus sequence of each contig were subjected to comparison using the appropriate BLAST algorithm to the NCBI non-redundant protein database, conserved domain database [30] that contains the eukaryotic clusters of orthologous groups (COG), Simple Modular Architecture Tool (SMART) and Protein Family Database (Pfam), and the Gene Ontology database [31]. Nucleotide sequences were directly compared with two customized databases, mitochondrial and ribosomal RNA (rRNA) nucleotide databases using BlastN. Determination of the presence of a signal secretion peptide or transmembrane helices was accomplished by the submission of sequence peptides to the SignalP server [32] or TMHMM server [33], respectively. The *Lu. longipalpis* BLAST server was utilized to determine homology between the *P. papatasi* clusters and *Lu. longipalpis* ESTs [34]. The number of transcripts each library contributed to a particular contig was derived using a custom program, Count Libraries (JMC Ribeiro, personal communication). Comparisons between the sugar-fed and blood-fed midgut cDNA sequences and comparisons between blood-fed and *L. major*-infected midgut cDNA sequences were based on separate Chi-square analysis [35]. The grouped and assembled sequences, BLAST results and signal peptide results were combined in an Excel spreadsheet and the putative function, if any, was manually verified and annotated.

Sequences were aligned using Clustal X, version 1.83, and converted to graphical aligned sequences using BioEdit, version 7.0.5.3 [36]. Phylogenetic analysis was conducted on amino acid alignments using TREE-PUZZLE, version 5.2, generating trees by maximum likelihood using quartet puzzling to calculate node support [37].

Quantitative PCR

Quantitative PCR (qPCR) was performed in selected clones using the first-strand cDNA, obtained from 100ng total RNA isolated from midguts dissected from *P. papatasi* females fed on sugar (unfed) or dissected after a blood meal (24–72h post-blood meal or PBM). cDNAs were synthesized using the 1st Strand cDNA Synthesis kit (Invitrogen, San Diego CA). Transcript levels were measured with SYBR green dye using a LightCycler 2.0 (Roche Diagnostics, Mannheim, Germany). For qPCR reactions, samples were subjected to an initial holding step at 95°C for 15 minutes, followed by an amplification step consisting of 35 cycles of 95°C for 10 seconds, 54°C for 20 seconds and 72°C for 20 seconds with a single acquisition. The reaction continued with a single-cycle melting step of 95°C for 10 seconds, 67°C for 30 seconds and 95°C for 10 seconds, prior to cooling for 1 minute. Equal amounts of cDNA were amplified using gene-specific primer sets targeting individual transcripts as well as a *P. papatasi* alpha tubulin, as control or reference transcript. Reactions were routinely done in duplicate. The relative expression ratio of the target transcript and control or reference transcript (fold over control) was calculated using the LightCycler relative quantification software (Roche).

Semi-quantitative PCR

Semi quantitative RT-PCR reactions were performed with selected transcripts to further demonstrate the differential expression of these genes in *P. papatasi* midgut. In this case, 100ng of total RNA isolated from midguts dissected from *P. papatasi* females fed on sugar (unfed) or dissected after a blood meal (48h PBM) were used to synthesize a cDNA using the 1st Strand cDNA Synthesis kit (Invitrogen). PCR reactions were carried out by an initial hot start at 95°C for 5 minutes followed by 25 cycles of 95°C for 30 seconds, 54°C for 1 minute and 72°C for 1.5 minutes and a final extension cycle of 72°C for 5 minutes. PCR products were separated on 1.5% agarose.

References

1. Kamhawi S, Ramalho-Ortigao M, Pham VM, Kumar S, Lawyer PG, Turco SJ, Barillas-Mury C, Sacks DL, Valenzuela JG: **A role for insect galectins in parasite survival.** *Cell* 2004, **119**(3):329-341.
2. Pal U, Li X, Wang T, Montgomery RR, Ramamoorthi N, Desilva AM, Bao F, Yang X, Pypaert M, Pradhan D *et al*: **TROSPA, an Ixodes scapularis receptor for Borrelia burgdorferi.** *Cell* 2004, **119**(4):457-468.
3. Borovsky D, Schlein Y: **Trypsin and chymotrypsin-like enzymes of the sandfly Phlebotomus papatasi infected with Leishmania and their possible role in vector competence.** *Med Vet Entomol* 1987, **1**(3):235-242.
4. Dillon RJ, Lane RP: **Influence of Leishmania infection on blood-meal digestion in the sandflies Phlebotomus papatasi and P. langeroni.** *Parasitol Res* 1993, **79**(6):492-496.

5. Ribeiro JM: **A catalogue of *Anopheles gambiae* transcripts significantly more or less expressed following a blood meal.** *Insect Biochem Mol Biol* 2003, **33**(9):865-882.
6. Dana AN, Hong YS, Kern MK, Hillenmeyer ME, Harker BW, Lobo NF, Hogan JR, Romans P, Collins FH: **Gene expression patterns associated with blood-feeding in the malaria mosquito *Anopheles gambiae*.** *BMC Genomics* 2005, **6**(1):5.
7. Pedra JH, Brandt A, Li HM, Westerman R, Romero-Severson J, Pollack RJ, Murdock LL, Pittendrigh BR: **Transcriptome identification of putative genes involved in protein catabolism and innate immune response in human body louse (Pediculicidae: *Pediculus humanus*).** *Insect Biochem Mol Biol* 2003, **33**(11):1135-1143.
8. Campbell CL, Wilson WC: **Differentially expressed midgut transcripts in *Culicoides sonorensis* (Diptera: ceratopogonidae) following Orbivirus (reoviridae) oral feeding.** *Insect Mol Biol* 2002, **11**(6):595-604.
9. Valenzuela JG, Garfield M, Rowton ED, Pham VM: **Identification of the most abundant secreted proteins from the salivary glands of the sand fly *Lutzomyia longipalpis*, vector of *Leishmania chagasi*.** *J Exp Biol* 2004, **207**(Pt 21):3717-3729.
10. Anderson JM, Oliveira F, Kamhawi S, Mans BJ, Reynoso D, Seitz AE, Lawyer P, Garfield M, Pham M, Valenzuela JG: **Comparative salivary gland transcriptomics of sandfly vectors of visceral leishmaniasis.** *BMC Genomics* 2006, **7**:52.

11. Kato H, Anderson JM, Kamhawi S, Oliveira F, Lawyer PG, Pham VM, Sangare CS, Samake S, Sissoko I, Garfield M *et al*: **High degree of conservancy among secreted salivary gland proteins from two geographically distant *Phlebotomus duboscqi* sandflies populations (Mali and Kenya).** *BMC Genomics* 2006, **7**:226.
12. Ramalho-Ortigao JM, Temporal P, de Oliveira SM, Barbosa AF, Vilela ML, Rangel EF, Brazil RP, Traub-Cseko YM: **Characterization of constitutive and putative differentially expressed mRNAs by means of expressed sequence tags, differential display reverse transcriptase-PCR and randomly amplified polymorphic DNA-PCR from the sand fly vector *Lutzomyia longipalpis*.** *Mem Inst Oswaldo Cruz* 2001, **96**(1):105-111.
13. Ramalho-Ortigao JM, Traub-Cseko YM: **Molecular characterization of Llchit1, a midgut chitinase cDNA from the leishmaniasis vector *Lutzomyia longipalpis*.** *Insect Biochem Mol Biol* 2003, **33**(3):279-287.
14. Ramalho-Ortigao JM, Kamhawi S, Rowton ED, Ribeiro JM, Valenzuela JG: **Cloning and characterization of trypsin- and chymotrypsin-like proteases from the midgut of the sand fly vector *Phlebotomus papatasi*.** *Insect Biochem Mol Biol* 2003, **33**(2):163-171.
15. Ramalho-Ortigao JM, Kamhawi S, Joshi MB, Reynoso D, Lawyer PG, Dwyer DM, Sacks DL, Valenzuela JG: **Characterization of a blood activated chitinolytic system in the midgut of the sand fly vectors *Lutzomyia longipalpis* and *Phlebotomus papatasi*.** *Insect Mol Biol* 2005, **14**(6):703-712.

16. Boulanger N, Lowenberger C, Volf P, Ursic R, Sigutova L, Sabatier L, Svobodova M, Beverley SM, Spath G, Brun R *et al*: **Characterization of a defensin from the sand fly *Phlebotomus duboscqi* induced by challenge with bacteria or the protozoan parasite *Leishmania major*.** *Infect Immun* 2004, **72**(12):7140-7146.
17. Dillon RJ, Ivens AC, Churcher C, Holroyd N, Quail MA, Rogers ME, Soares MB, Bonaldo MF, Casavant TL, Lehane MJ *et al*: **Analysis of ESTs from *Lutzomyia longipalpis* sand flies and their contribution toward understanding the insect-parasite relationship.** *Genomics* 2006, **88**(6):831-840.
18. Pomes A, Melen E, Vailes LD, Retief JD, Arruda LK, Chapman MD: **Novel allergen structures with tandem amino acid repeats derived from German and American cockroach.** *J Biol Chem* 1998, **273**(46):30801-30807.
19. Wittstock U, Agerbirk N, Stauber EJ, Olsen CE, Hippler M, Mitchell-Olds T, Gershenzon J, Vogel H: **Successful herbivore attack due to metabolic diversion of a plant chemical defense.** *Proc Natl Acad Sci U S A* 2004, **101**(14):4859-4864.
20. Shao L, Devenport M, Fujioka H, Ghosh A, Jacobs-Lorena M: **Identification and characterization of a novel peritrophic matrix protein, Ae-Aper50, and the microvillar membrane protein, AEG12, from the mosquito, *Aedes aegypti*.** *Insect Biochem Mol Biol* 2005, **35**(9):947-959.
21. Devenport M, Alvarenga PH, Shao L, Fujioka H, Bianconi ML, Oliveira PL, Jacobs-Lorena M: **Identification of the *Aedes aegypti* peritrophic matrix**

- protein AeIMUCI as a heme-binding protein.** *Biochemistry* 2006, **45**(31):9540-9549.
22. Devenport M, Fujioka H, Jacobs-Lorena M: **Storage and secretion of the peritrophic matrix protein Ag-Aper1 and trypsin in the midgut of *Anopheles gambiae*.** *Insect Mol Biol* 2004, **13**(4):349-358.
 23. Campos IT, Amino R, Sampaio CA, Auerswald EA, Friedrich T, Lemaire HG, Schenkman S, Tanaka AS: **Infestin, a thrombin inhibitor presents in *Triatoma infestans* midgut, a Chagas' disease vector: gene cloning, expression and characterization of the inhibitor.** *Insect Biochem Mol Biol* 2002, **32**(9):991-997.
 24. Friedrich T, Kroger B, Bialojan S, Lemaire HG, Hoffken HW, Reuschenbach P, Otte M, Dodt J: **A Kazal-type inhibitor with thrombin specificity from *Rhodnius prolixus*.** *J Biol Chem* 1993, **268**(22):16216-16222.
 25. Geiser DL, Chavez CA, Flores-Munguia R, Winzerling JJ, Pham DQ: ***Aedes aegypti* ferritin.** *Eur J Biochem* 2003, **270**(18):3667-3674.
 26. Arvestad L, Visa N, Lundeborg J, Wieslander L, Savolainen P: **Expressed sequence tags from the midgut and an epithelial cell line of *Chironomus tentans*: annotation, bioinformatic classification of unknown transcripts and analysis of expression levels.** *Insect Mol Biol* 2005, **14**(6):689-695.
 27. Schlein Y, Romano H: ***Leishmania major* and *L. donovani*: effects on proteolytic enzymes of *Phlebotomus papatasi* (Diptera, Psychodidae).** *Exp Parasitol* 1986, **62**(3):376-380.

28. Myskova J, Svobodova M, Beverley SM, Volf P: **A lipophosphoglycan-independent development of *Leishmania* in permissive sand flies.** *Microbes Infect* 2007, **9**(3):317-324.
29. Ribeiro JM, Arca B, Lombardo F, Calvo E, Phan VM, Chandra PK, Wikel SK: **An annotated catalogue of salivary gland transcripts in the adult female mosquito, *Aedes aegypti*.** *BMC Genomics* 2007, **8**:6.
30. NCBI Conserved Domain Database (CDD)
[<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>].
31. Mappings of External Classification Systems to GO
[<http://www.geneontology.org/GO.indices.shtml>].
32. SignalP 3.0 Server [<http://www.cbs.dtu.dk/services/SignalP/>].
33. TMHMM Server, v. 2.0 [<http://www.cbs.dtu.dk/services/TMHMM/>].
34. Lu. longipalpis Blast Server [http://www.sanger.ac.uk/cgi-bin/blast/submitblast/l_longipalpis].
35. Ribeiro JM, Alarcon-Chaidez F, Francischetti IM, Mans BJ, Mather TN, Valenzuela JG, Wikel SK: **An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks.** *Insect Biochem Mol Biol* 2006, **36**(2):111-129.
36. Tippmann HF: **Analysis for free: comparing programs for sequence analysis.** *Brief Bioinform* 2004, **5**(1):82-87.
37. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**(3):502-504

Chapter 4

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

Published as: Jochim RC, Teixeira CR, Laughinghouse A, Mu J, Oliveira F, Gomes RB, Elnaïem DE, Valenzuela JG. 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. 2008 Jan 14;9(1):15.

Abstract

Background: In the life cycle of *Leishmania* within the alimentary canal of sand flies, the parasites have to survive the hostile environment of blood meal digestion, escape the blood bolus and attach to the midgut epithelium before differentiating into the infective metacyclic stages. The molecular interactions between the *Leishmania* parasites and the gut of the sand fly are poorly understood. In the present work, we sequenced five cDNA libraries constructed from midgut tissue from the sand fly *Lutzomyia longipalpis* and analyzed the transcripts present following sugar-feeding, blood-feeding and after the blood meal had been processed and excreted, both in the presence and absence of *Leishmania infantum chagasi*.

Results: Comparative analysis of the transcripts from sugar-fed and blood-fed cDNA libraries resulted in the identification of transcripts differentially expressed during blood-feeding. These included upregulated transcripts such as four distinct microvillar-like proteins (LuloMVP1, 2, 4 and 5), two peritrophin like proteins, a trypsin like protein (Lltryp1), two chymotrypsin like proteins (LuloChym1A and 2) and an unknown protein. Downregulated transcripts by blood-feeding were a microvillar-like protein (LuloMVP3), a trypsin like protein (Lltryp2) and an astacin-like metalloprotease (LuloAstacin). Furthermore, a comparative analysis between blood-fed and *Leishmania*-infected midgut cDNA libraries resulted in the identification of transcripts that were differentially expressed due to the presence of *Leishmania* in the gut of the sand fly. This included downregulated transcripts such as four microvillar-like proteins (LuloMVP1,2, 4 and 5), a chymotrypsin (LuloChym1A) and a carboxypeptidase (LuloCpepA1), among others.

Upregulated midgut transcripts in the presence of *Leishmania* were a peritrophin-like protein (LuloPer1), a trypsin-like protein (Lltryp2) and an unknown protein.

Conclusion: This transcriptome analysis represents the largest set of sequence data reported from a specific sand fly tissue and provides further information of the transcripts present in the sand fly *Lutzomyia longipalpis*. This analysis provides the detailed information of molecules present in the midgut of this sand fly and the transcripts potentially modulated by blood-feeding and by the presence of the *Leishmania* parasite. More importantly, this analysis suggests that *Leishmania infantum chagasi* alters the expression profile of certain midgut transcripts in the sand fly during blood meal digestion and that this modulation may be relevant for the survival and establishment of the parasite in the gut of the fly. Moreover, this analysis suggests that these changes may be occurring during the digestion of the blood meal and not afterwards.

Background

Leishmaniasis is a spectrum of diseases caused by numerous species of the kinetoplastid parasite *Leishmania*, which are transmitted by Phlebotomine sand flies. Different forms of disease presentation can be linked with the various species of *Leishmania* parasites, with the visceral form of the disease being caused mainly by the Old World *Leishmania infantum* or the New World variant *Leishmania infantum (chagasi)*. Visceral leishmaniasis is a disease that is commonly fatal if left untreated. Currently, there is no licensed vaccine for the prevention of visceral disease in humans,

and current drug treatment with antimonials and other components is a lengthy and arduous procedure with undesirable secondary effects [1].

The sand fly *Lutzomyia longipalpis*, the principal vector of the parasite *Leishmania infantum chagasi*, is the most significant source of American visceral leishmaniasis. As with many other arthropod-borne diseases, transmission of the *Leishmania* parasite occurs during the act of vector blood-feeding upon a vertebrate host. Upon blood meal ingestion, a large number of events are induced, including digestion, metabolism, diuresis, and ultimately oogenesis. Unlike the arboviruses, *Plasmodium* or *Borrelia*, *Leishmania* can complete the necessary developmental changes and propagate to numbers sufficient for transmission and infection solely within the confines of the midgut tissue of the sand fly [2]. Several sand fly proteases involved in blood meal digestion and implicated in the species specificity between *Leishmania* and the respective vectors have been characterized and include trypsins, chymotrypsins and chitinases from both *Lu. longipalpis* and *Phlebotomus papatasi* [3, 4].

A more global approach to identifying and characterizing sand fly molecules has been accomplished through the sequencing of whole sand fly-derived expressed sequence tags [5]. While that study contributes to the knowledge of the molecular components of the sand fly, it does not provide the specific molecules of the midgut tissue that would interact with the developing parasites. The construction and sequencing of midgut tissue-specific cDNA libraries aims therefore, to identify those molecules involved in blood meal digestion and metabolism, peritrophic matrix formation and possible parasite associations. Here, we have generated and sequenced five cDNA libraries from the midgut tissue of *Lu. longipalpis*, investigated the molecules present

during sugar and blood-feeding as well as after the blood meal has been processed and excreted, both in the presence and absence of *L. infantum chagasi*. In addition to the identification of midgut-associated molecules, sequence analysis and phylogenetic comparison of the sequences of *Lu. longipalpis* allows a better understanding of blood meal processing in sand flies and the differences between visceral (*Lutzomyia longipalpis*) and cutaneous leishmaniasis (*Phlebotomus papatasi*) sand fly vectors.

Results and discussion

As the midgut is the primary organ of the sand fly in which the *Leishmania* parasite develops, cDNA libraries of the midgut tissue were constructed, sequenced and analyzed to investigate the molecules present that may provide for important interactions between these two organisms. In total, five cDNA libraries were constructed from the midgut tissue of female *Lu. longipalpis* during different conditions of feeding and digestion. These conditions included one library combining the midguts from sand flies allowed to feed on a sucrose solution (SF), a pool of midgut tissue from sand flies fully engorged from an artificial blood meal 1, 2 and 3 days post-blood meal ingestion (BF), and a pool of midguts from gravid sand flies 5, 6 and 7 days post-blood meal digestion (PBMD). The conditions chosen and the pooling of those times after blood meal ingestion allows better coverage of the most abundant molecules transcribed in the midgut as well as a comparison of the molecules present prior to blood-feeding, while the blood bolus is present, during digestion of the blood meal and after the blood byproducts have been excreted. Two cDNA libraries were constructed from the equivalent pools of time points after blood-feeding in *Lu. longipalpis* midgut tissue from sand flies that had

ingested amastigote-infected macrophages in an artificial blood meal (BFi and PBMDi), a more natural presentation of parasites to the blood-feeding sand fly.

Once constructed, approximately 2300 phage plaques were picked and ultimately sequenced for each of the five cDNA libraries; generating a total of 9601 high quality sequences from the midgut tissue of *Lu. longipalpis*. These sequences have been submitted to the NCBI EST database under the accession numbers EW987149 - EW996682. Table 7 summarizes the results of sequence quality and bioinformatics analysis of each library and the combination of all libraries by the number of sequences analyzed the number of high quality sequences used in the bioinformatics analysis, the number of contigs, the number of singletons and the average number of sequences per contig. Each library generated a similar number of sequences, and sequence recovery from the phage plaques ranged from 79-85%. After discarding low quality sequences, each library retained 71-80% sequences with an average of 73% of the total 11,520 phage producing high quality sequence data. Clustering similar sequences into contigs, based on sequence homology, produced a comparable number of contigs for each library as well as a similar number of singletons. The comparable number of high quality sequences, contigs and singletons produced from each library allows for a better comparison between the sequence abundance of specific molecules of interest and the respective biological condition of the midgut under which they were recovered. The average number of sequences in the clusters of contigs varied slightly between libraries. The BF, PBMD and PBMDi cDNA libraries contained an average sequence per cluster

Table 7: Overall examination of the 5 individual cDNA libraries and the combined analysis

	SF	BF	BFi	PBMD	PBMDi	Combined
Sequences analyzed	1822	1970	1928	1953	1928	9601
High quality sequences	1646	1845	1683	1650	1647	8471
Contigs	148	137	156	125	117	655
Singletons	631	694	694	638	622	2279
Sequences/contig	6.86	8.40	6.34	8.10	8.76	9.45

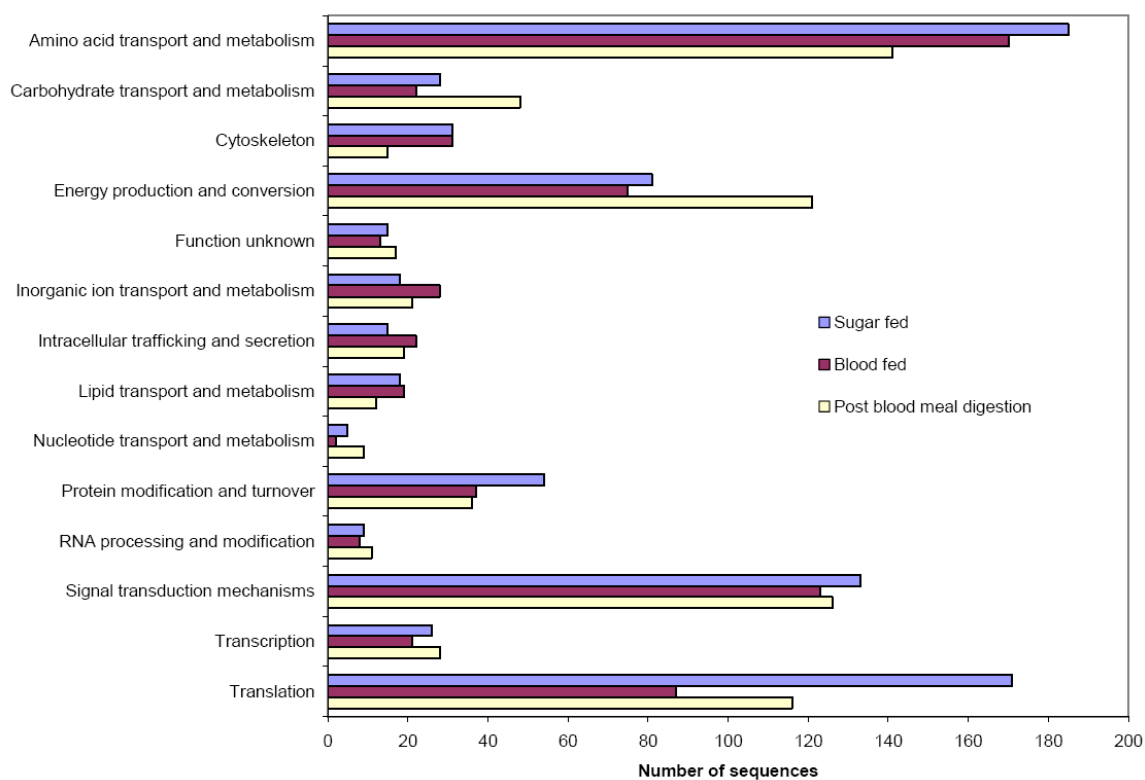


Figure 23. Histogram of the number of sequences grouped into functional classes from the sugar-fed, blood-fed and post-blood meal digestion cDNA libraries.

Sequences from clusters of those three cDNA libraries, with an E-value less than $10E-5$ result of the COG BLAST grouped into the general functional class as assigned by COG.

ratio of 8.4, 8.10 and 8.76, respectively. The SF cDNA library had a sequence per cluster ratio of 6.86, and the BFi cDNA library produced an average of 6.34 sequences per cluster. The combining of all cDNA library sequences produced 655 contigs, 2279 singletons and an average of 9.45 sequences per contig. Each cluster was assigned a putative function and placed in a functional class based on the sequence homology to molecules identified by the BLAST results from the NCBI non-redundant protein, the Gene Ontology, the conserved domain, rRNA and mitochondrial databases. Figure 23 shows an overall view of sequence abundance of functional classes that occur during the processes of sugar-feeding, blood-feeding and after the digestion of the blood meal. The clusters of those three cDNA libraries, with an E-value less than $10E-5$ as determined by KOG BLAST, were grouped according to the general functional class. Although this is a summation of a large number of different clusters, the total number of sequences in each functional class can highlight overall trends that are potentially important in the processes of blood-feeding and digestion.

Following is a more detailed description of the most abundant transcripts identified in this analysis:

Proteases

Proteases were among the most abundant transcripts captured in the random sequencing of the midgut cDNA libraries and included trypsin-like serine proteases, chymotrypsins, carboxypeptidases, and an astacin-like metalloprotease. Table 8 shows the putative proteases identified in the midgut transcriptome. The Sanger Institute's *Lutzomyia longipalpis* EST database was searched using BLAST to find the best matches

Table 8: Putative midgut-associated proteases; best matched results and corresponding E-values from BLAST inquiries of a GenBank-derived non-redundant protein database and *Lutzomyia longipalpis* EST database

Cluster	Best match to non-redundant protein database	NR E value	Best match to <i>Lutzomyia</i> EST database	<i>Lutzomyia</i> E value	GenBank
35	trypsin 4 [<i>P. papatasi</i>]	1.E-101	NSFM-61a01	3.E-140	ABM26904
18	trypsin 1 [<i>P. papatasi</i>]	8.E-79	SFM-03g02	3.E-132	ABM26905
83	trypsin 3 [<i>P. papatasi</i>]	6.E-94	NSFM-113g08	1.E-125	EUI24590
60	trypsin 2 [<i>P. papatasi</i>]	4.E-67	NSFM-48a06	2.E-137	EUI24582
291	trypsin-eta, putative [<i>A. aegypti</i>]	7.E-55	NSFM-15d03	6.E-157	EUI24595
33	chymotrypsin [<i>P. papatasi</i>]	1.E-96	NSFM-95b07	5.E-143	EUI24576
32	chymotrypsin [<i>P. papatasi</i>]	8.E-97	NSFM-61f07	1.E-137	EUI24575
64	larval chymotrypsin-like protein precursor [<i>A. aegypti</i>]	1.E-79	SFM-01b03	1.E-130	EUI24583
87	chymotrypsin [<i>P. papatasi</i>]	3.E-79	NSFM-96h06	2.E-148	EUI24591
30	chymotrypsin [<i>P. papatasi</i>]	2.E-94	NSFM-29b07	5.E-133	EUI24573
31	chymotrypsin [<i>P. papatasi</i>]	5.E-96	NSFM-129f09	1.E-141	EUI24574
58/59	ENSANGP00000019623 [<i>A. gambiae</i>]	3.E-57	NSFM-121h10	6.E-131	EUI24581
104	carboxypeptidase [<i>A. aegypti</i>]	1.E-126	SFM-05c11	1.E-221	EUI24592
107	carboxypeptidase [<i>A. aegypti</i>]	1.E-114	NSFM-146a05	1.E-226	EUI24593
91	similar to CG8560-PA [<i>T. castaneum</i>]	2.E-82	NSFM-32d09	8.E-199	EUI24594

Table 9: Putative midgut-associated proteases; putative function and sequence distribution contributed from each cDNA library

Cluster	Clone	Putative function	SF	BF	Number of sequences			Total
					BFi	PBMD	PBMDi	
35	LJGFIM23_B07	Trypsin	3	55	34	0	0	92
18	LJGUL-P03_G08	Trypsin	136	6	15	109	168	434
83	LJGFM-P03_E11	Trypsin	8	7	3	7	1	26
60	LJGU-I-5_D05	Trypsin	8	4	8	10	8	38
291	LJGFIM26_A01	Serine protease	2	0	1	0	2	5
33	LJGFM-P04_B01	Chymotrypsin	3	51	22	1	0	77
32	LJGFL10_C10	Chymotrypsin	0	2	5	0	0	7
64	LJGFM-P03_C01	Chymotrypsin	0	17	17	0	1	35
87	LJGF-I-8_E03	Chymotrypsin	1	14	8	0	2	25
30	LJGDIL5_B09	Chymotrypsin	12	1	3	1	4	21
31	LJGFM-P01_C04	Chymotrypsin	3	4	2	0	0	9
58/59	LJGUL-P01_B07	Astacin-like metalloprotease	28	7	1	1	4	41
104	LJGFL_P01_F01	Carboxypeptidase	0	14	3	0	0	17
107	LJGFL_P03_G11	Carboxypeptidase	6	5	7	0	0	18
91	LJGFIM22_C05	Carboxypeptidase	1	8	8	1	1	19

and results are shown with the corresponding E-value. The proteases described here are most similar to those described in the sand fly *Phlebotomus papatasi* and the mosquitoes *Aedes aegypti* or *Anopheles gambiae*, with the exception that cluster 91 encodes a putative carboxypeptidase that shares homology with a molecule from the beetle *Tribolium castaneum*. Table 9 shows the transcript producing a full length, high quality sequence for each cluster and the putative function of the identified transcripts. The number of sequences that each cluster contributed to each of the cDNA libraries also is shown and from this, it can be seen that proteases are more abundant, as expected, in the blood-fed (BF) and blood-fed *Leishmania*-infected (BFi) libraries. An interesting observation is that cluster 18, which encodes a putative trypsin, is more abundant in the SF, PBMD and PMBDi cDNA libraries, indicating that this putative trypsin may have a role other than blood meal digestion or is produced and stored prior to the ingestion of a blood meal. Table 10 describes the predicted localization, molecular weight and isoelectric point of these proteases. All of the identified proteases possess a potential signal peptide, and the molecular weight and isoelectric point given is that of the predicted mature protein.

Trypsin

Four trypsin-like transcripts were identified in the transcriptome with high homology to the described *P. papatasi* midgut trypsins [3, 6]. Clusters 18, 35, 60 and 83 are similar to *P. papatasi* *Pptryp1*, *Pptryp4*, *Pptryp2* and *Pptryp3*, respectively. Recently, two transcripts from *Lu. longipalpis* midgut EST sequencing were partially characterized and named *Lltryp1*, which corresponds with Cluster 35 identified in our cDNA libraries, and *Lltryp2*, which corresponds with Cluster 18 [3]. *Lltryp2* was found

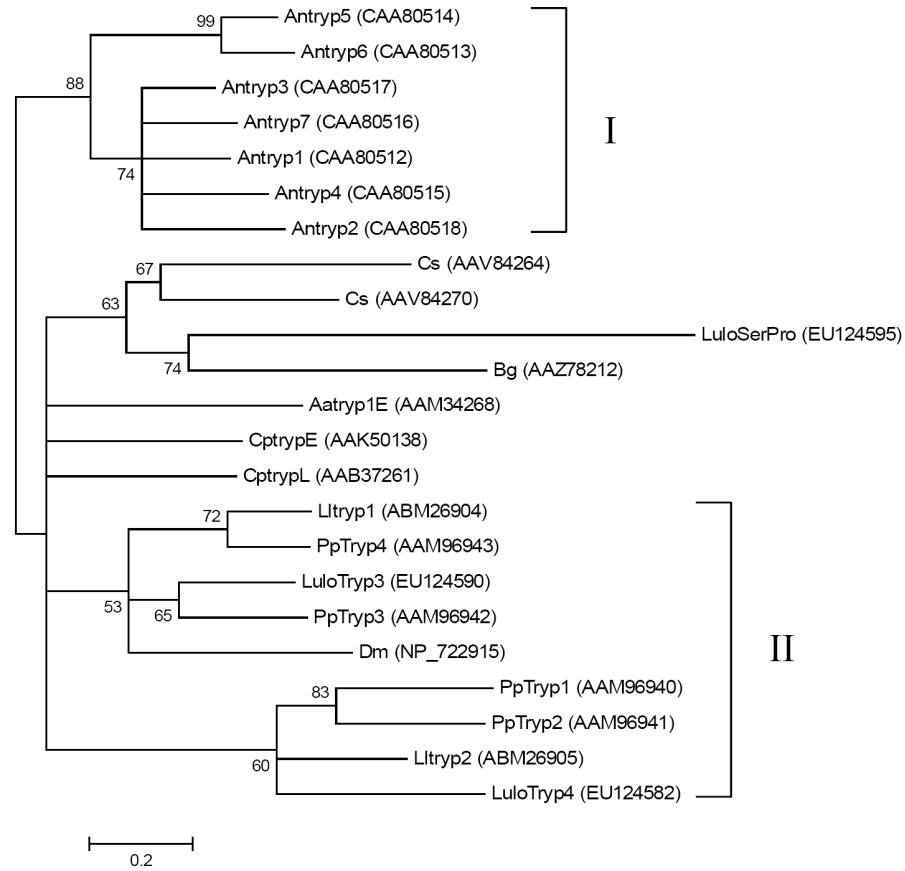
Table 10: Putative midgut-associated proteases; localization, molecular weight and isoelectric point of putative midgut proteins

Cluster	Putative function	Gene name	Localization	Molecular weight (kDa)	Isoelectric point
35	Trypsin	<i>Lltryp1</i>	Secreted	26.2	6.32
18	Trypsin	<i>Lltryp2</i>	Secreted	26.0	4.95
83	Trypsin	<i>LuloTryp3</i>	Secreted	26.0	5.67
60	Trypsin	<i>LuloTryp4</i>	Secreted	26.1	5.52
291	Serine protease	<i>LuloSerPro</i>	Secreted	29.0	8.26
33	Chymotrypsin	<i>LuloChym1A</i>	Secreted	26.8	6.55
32	Chymotrypsin	<i>LuloChym1B</i>	Secreted	26.6	6.40
64	Chymotrypsin	<i>LuloChym2</i>	Secreted	25.8	6.74
87	Chymotrypsin	<i>LuloChym3</i>	Secreted	27.6	4.77
30	Chymotrypsin	<i>LuloChym4</i>	Secreted	26.9	5.86
31	Chymotrypsin	<i>LuloChym5</i>	Secreted	26.9	6.19
58/59	Astacin-like metalloprotease	<i>LuloAstacin</i>	Secreted	28.0	5.04
104	Carboxypeptidase	<i>LuloCpepA1</i>	Secreted	45.8	5.36
107	Carboxypeptidase	<i>LuloCpepA2</i>	Secreted	46.0	5.41
91	Carboxypeptidase	<i>LuloCpepB</i>	Secreted	45.9	4.73

in highest abundance, 434 sequences, with the unique sequence distribution among the five cDNA libraries in that most sequences were contributed by the sugar-fed and post-blood meal digestion groups. Sequence abundance of trypsins varied; listed in order of decreasing abundance are *Lltryp1*, *LuloTryp4*, and *LuloTryp3*. *LuloTryp3* and *LuloTryp4* had relatively homogenous sequence distribution among the cDNA libraries, although *LuloTryp3* was underrepresented in the PMBDi cDNA library with only one sequence identified. The distribution of *Lltryp1* sequences between the cDNA libraries correlates with reverse transcriptase-PCR results published showing the expression of *Lltryp1* during the presence of a blood meal in the female sand fly midgut [3]. Further information about the putative trypsin molecules can be found in Table 10, showing the range of molecular weight from 26.0 to 26.2 kDa. The isoelectric points (pI) of these putative trypsins vary with *Lltryp1* having a higher pI of 6.32, *Lltryp* with a lower pI of 4.95, and *LuloTryp3* and *LuloTryp4* having similar pIs of 5.67 and 5.52, respectively. Phylogenetic analysis of amino acid sequences from Dipteran trypsin molecules and a trypsin from *Blattella germanica* resulted in two major clades, one containing the *An. gambiae* trypsin molecules (group I) and another containing the remaining sequences. Within the other major clade the sand fly trypsins from *Lu. longipalpis* and *P. papatasi* form two subclades (Group II) (Figure 2A). As previously published [3], *Pptryp1* and *Pptryp2* form a clade apart from the clade containing *Pptryp3* and *Pptryp4*. The putative trypsin molecules identified in *Lu. longipalpis* midgut share a high homology with the *P. papatasi* molecules, being grouped into the same clades. Multiple sequence alignment of the trypsin molecules of *Lu. longipalpis* depicts the potential secretory signal peptide, the H/D/S catalytic site residues and substrate specifying residues (Figure 24B).

Figure 24. Sequence analysis of trypsin-like serine proteases.

(A) Phylogenetic analysis of amino acid sequences from *Anopheles gambiae* (Antryp), *Culicoides sonorensis* (Cs), *Blattella germanica* (Bg), *Lutzomyia longipalpis* (Lulo and Ll), *Phlebotomus papatasi* (Pp), *Aedes aegypti* (Aa), *Drosophila melanogaster* (Dm) and *Culex pipiens quinquefasciatus* (Cp). Node support is indicated by bootstrap values and accession numbers given in parenthesis. (B) Multiple sequence alignment of *Lutzomyia longipalpis* putative trypsin molecules. Predicted secretion signal peptides are underlined, catalytic residues marked by (*) and residues determining substrate specificity marked by (#).

A**B**

Lltryp2	MLNQAAAILAILCISASAA---VFRCP--IQRVGGKPVNIEDIPYQVSNYFG-----
LuloTryp4	MWHHALALAILFLSVNAA---SLREP--IHRVGGIAVNIEDFPHQVSLIFNN-----
Lltryp1	-MHSVAILCLLPLAVLAG-PAFLPRRLDGRIVGGFEVDIRHVPYQVSLQTS-----
LuloTryp3	-MIRFVVLSTLLVGVWGANVKFFPKRLDGRIVGGYVINIEDTPYQLSLQRSN-----
LuloSerPro	-MVSSTYFIVLFLAITFTSNVWSQQIPPETRIVGGIPVQDALRQQAIRYKSSDSTFGN
*	
Lltryp2	QHLCGGSILSERFILTAAHCTIG-ATPDFTVRTG-----SNYSSTDGEVHKVRQIIS
LuloTryp4	SHICGGSIIIDKFLLSAAHCYIG-RTAFFQVRVG-----STNSTAGGTLTYVKDIHA
Lltryp1	GHECGGSIIISHNFVFTAAHCTDGQDASRLKVRVG-----SNEHGTGGDFPKVKVHQ
LuloTryp3	WHICGASLIISEFVLTAHCTFGSSANSFTVRTQ-----TSFHGRGGVVVGKRIIQ
LuloSerPro	GHVCGNLISSNNTVLSAAHCFVDDLGRIRISVSDIRVVGGNLLLTEQTQNTFTSDVARLII
*	
Lltryp2	HELYDEETTDYDFSIFELEDPTTFS-DTRRAVOLPEADEDIAVGTVLKVSGWGHTKNHQE
LuloTryp4	HITSYNDTTYDYDFSIFEINDAIALDNVTSRVVRLSEINEYLPNGSNLTISGWGHTRNPTTE
Lltryp1	HPSFNYQTYDYDFSILLEESITFN-SVRYFVRLPEKDDVDYDGAILLVSGWGNTQSSQE
LuloTryp3	HPKFDYSTTDYDFSILELAAPVEFN-EKLQPIRLPEQDEVEDGTHLLVTGWGNTQNAQE
LuloSerPro	HPGYNRNTYADDIAILLIKTIIPASHSSFRPITLRSVP--VDSNTTCELSGWGHTSENGQ
## *	
Lltryp2	SNYHVRVAVSVKVNQLEQANYIFRAITISEQMCAGYQEGGKDSQCGDSGGPVVDDNNVQ
LuloTryp4	SEEHTRAVSVKLEQLBCTIASYLFSGFITDRMCAGYRCGGKDACQGDSSGGPVLDNDNNVQ
Lltryp1	SNKHTRATVVEKYNDQCNKAYTQYGGITKTMICAGFEEGEKDACQGDSSGGPLTHGD-VL
LuloTryp3	SREQTRAAIVPKSNDEVCNKAYGQFGGITARMICAGLPEGGEKDACQGDSSGGPLASDG-VL
LuloSerPro	GSNSLLTANVSIIDFNVCNNAYGNS--LRSGMLCAGRMCGGVDTCQGDSSGGPLVCNG-LL
Lltryp2	HGVVSWGKGCALPSYPGVYAKVSAVRNWIREISNV-----
LuloTryp4	VGVVSWGLECALARYPGVYGRITSSVRQWRDITNV-----
Lltryp1	VGVVSWGFGCAEPKYPGVYSRVSSVREWHGIVGF-----
LuloTryp3	VGVVSWGYGCAVRGYPGVYSRVASVRDWINASTNI-----
LuloSerPro	TGITSEGNCAKPGFPGVYTAVHSYHRWIIANAQQQLRINYFLSICLLLGARNIARVL

A novel midgut-associated serine protease, *LuloSerPro*, was identified in the sequencing and annotation of these midgut cDNA libraries. *LuloSerPro* is predicted to be secreted and have a mature molecular weight of 29.0 kDa, slightly larger than the other trypsin-like serine proteases in the midgut, and has an unusually high predicted pI of 8.26 (Table 10). This molecule, while found in low abundance, was present in the sugar-fed, blood-fed *Leishmania*-infected, and post-blood meal digestion-*Leishmania*-infected cDNA libraries (Table 9). Phylogenetic analysis and multiple sequence alignments of the midgut trypsin molecules and *LuloSerPro* show that while this molecule is very similar to other trypsin molecules and retains the catalytic residues, this is a distinctly different serine protease (Figure 24). Additionally, there is a difference in the residues that determine the substrate specificity (Lys to Val) between the other midgut tryptins and *LuloSerPro* (Figure 24B).

Chymotrypsin

Chymotrypsin is another serine protease found in abundance in the midgut of this hematophagous. This study identified five clusters with homology to chymotrypsin molecules described in *P. papatasi* and one cluster with homology to a putative larval chymotrypsin found in *Ae. aegypti* (Tables 8-10). Clusters 33, 32, 64, 87, 30 and 31 were named *LuloChym1A*, *LuloChym1B*, *LuloChym2*, *LuloChym3*, *LuloChym4* and *LuloChym5*, respectively. *LuloChym4* was found in higher abundance in the sugar-fed cDNA library and *LuloChym5* sequences were found in relatively equal numbers between blood-fed and sugar-fed cDNA libraries. In contrast the other chymotrypsin molecules appear in highest abundance in the blood-fed and blood-fed *Leishmania*-infected cDNA

libraries (Table 9). According to sequence numbers between the cDNA libraries it appears that chymotrypsin transcription is quiescent after the blood meal has been digested and excreted. The *Lu. longipalpis* chymotrypsin sequences have a predicted molecular weight of secreted protein ranging from 25.8 to 27.6 kDa (Table 10).

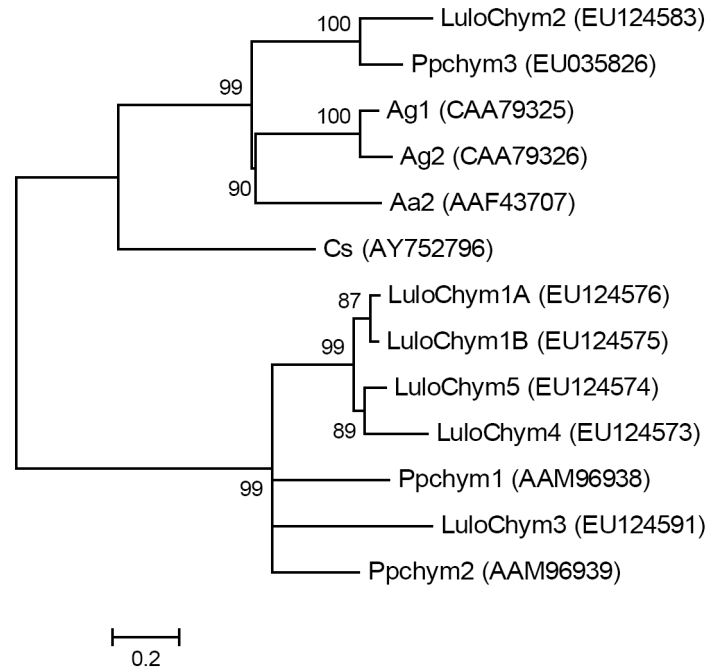
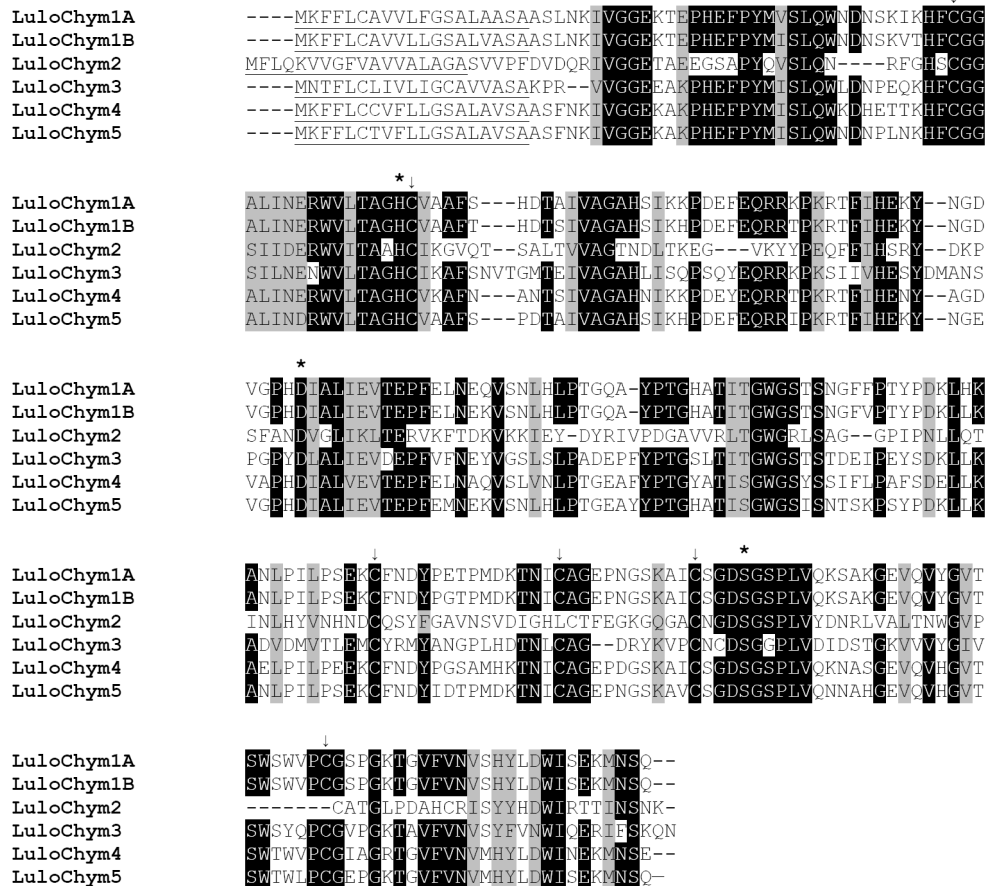
Phylogenetic analysis of chymotrypsin amino acid sequences show that there is conservation in sequence homology between *Lu. longipalpis* chymotrypsin and *P. papatasi* chymotrypsin molecules (Figure 25A). LuloChym1A, LuloChym1B, LuloChym4 and LuloChym5 form a subclade within a clade containing only sand fly chymotrypsin molecules. The short phylogenetic distance between LuloChym1A and LuloChym1B and the 95% amino acid identity they share suggests that these transcript sequences may represent polymorphisms. Further comparisons between the amino acid sequences of the midgut-associated chymotrypsin molecules show that the cysteine and catalytic residues H/D/S are conserved (Figure 25B).

Carboxypeptidases

The three longest transcripts encoding putative proteases identified in the analysis are similar to zinc metallocarboxypeptidases found in other insects as well as having significant similarity to ESTs from the Sanger Institute database (Table 8). These transcripts from clusters 104, 107 and 91, were named *LuloCpepA1*, *LuloCpepA2* and *LuloCpepB*, they have molecular weights of 45.8, 46.0 and 45.9 kDa and a pI of 5.36, 5.41 and 4.73, respectively (Table 10). Although *LuloCpepA2* appears to be an incomplete transcript with a 5' truncation, based on homology and predicted signal peptide sequences, a putative mature protein can be used in further characterization and

Figure 25. Chymotrypsin sequence analysis.

(A) Phylogenetic analysis of chymotrypsin sequences from *Phlebotomus papatasi* (Pp), *Lutzomyia longipalpis* (Lulo), *Anopheles gambiae* (Ag), *Aedes aegypti* (Aa), and *Culicoides sonorensis* (Cs). Accession numbers are shown in parentheses and node support indicated by the bootstrap values. (B) Sequence comparison of midgut putative chymotrypsin molecules. The probably signal peptide is underlined, the catalytic residues indicated by (*) and conserved cysteine residues marked with (↓).

A**B**

comparison. Most of the sequences grouped to produce the carboxypeptidase clusters were captured from the blood-fed library, suggesting that these molecules are likely induced by the ingestion or presence of blood in the midgut of the sand fly (Table 9). The classification of these molecules as members of the A or B class of metallo-carboxypeptidases was determined by the output from phylogenetic analysis of the amino acid sequences (Figure 26A). The phylogenetic tree produced by this analysis shows distinct clades containing insect sequences nearly all annotated as either carboxypeptidase A or carboxypeptidase B molecules. The high node support values of the sand fly carboxypeptidases in the phylogenetic tree imply conservation of these molecules when comparing the Old World sand fly *P. papatasi* and the New World sand fly *Lu. longipalpis*. Similarity between the two sand flies, with regards to the carboxypeptidase molecules, can be seen in amino acid sequence alignments depicting the high level of identity and retention of the catalytic residues necessary for metallo-carboxypeptidase activity (Figure 26B, 26C). Furthermore, the amino acid sequence alignment depicts the incongruousness that separates LuloCpepA1 from LuloCpepA2 (Figure 26B).

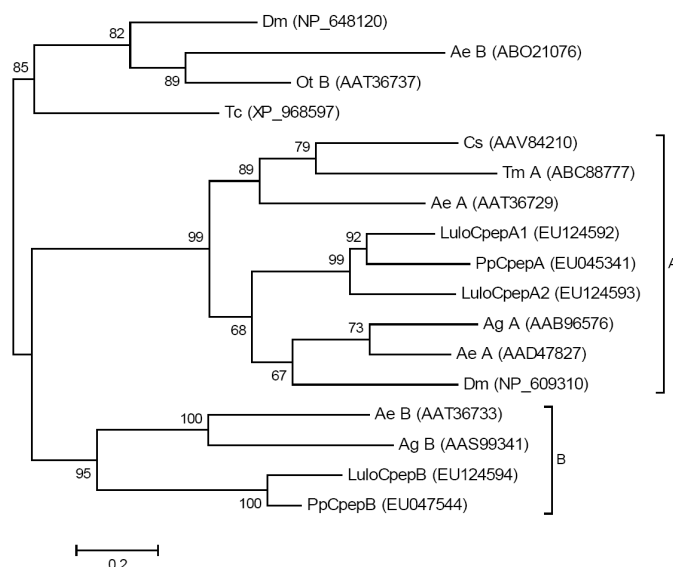
Astacin

A putative zinc metalloprotease was identified as a likely astacin-like molecule based on results from a search of the conserved domains database. This astacin molecule was derived from clusters 58 and 59, both encoding the same putative protein, but separated due to differing lengths of 5'- and 3'- UTRs by the bioinformatics software. The astacin-like metalloprotease was named LuloAstacin and is predicted to have a

Figure 26. Analysis of putative carboxypeptidase molecules.

(A) Phylogenetic analysis of carboxypeptidases from *Lutzomyia longipalpis* (Lulo), *Phlebotomus papatasi* (Pp), *Ochlerotatus triseriatus* (Ot), *Aedes aegypti* (Ae), *Anopheles gambiae* (Ag), *Drosophila melanogaster* (Dm), *Tribolium castaneum* (Tc), *Tenebrio molitor* (Tm) and *Culicoides sonorensis* (Cs). GenBank accession numbers are in parentheses and node support is indicated by bootstrap values. (B) Sequence alignment of putative carboxypeptidase A molecules identified from the midgut of *Lutzomyia longipalpis* (Lulo) and *Phlebotomus papatasi* (Pp). Predicted catalytic residues are marked with (*). (C) Sequence alignment of putative carboxypeptidase B molecules identified from the midgut of *Lutzomyia longipalpis* (Lulo) and *Phlebotomus papatasi* (Pp). Predicted catalytic residues are marked with (*).

A



B

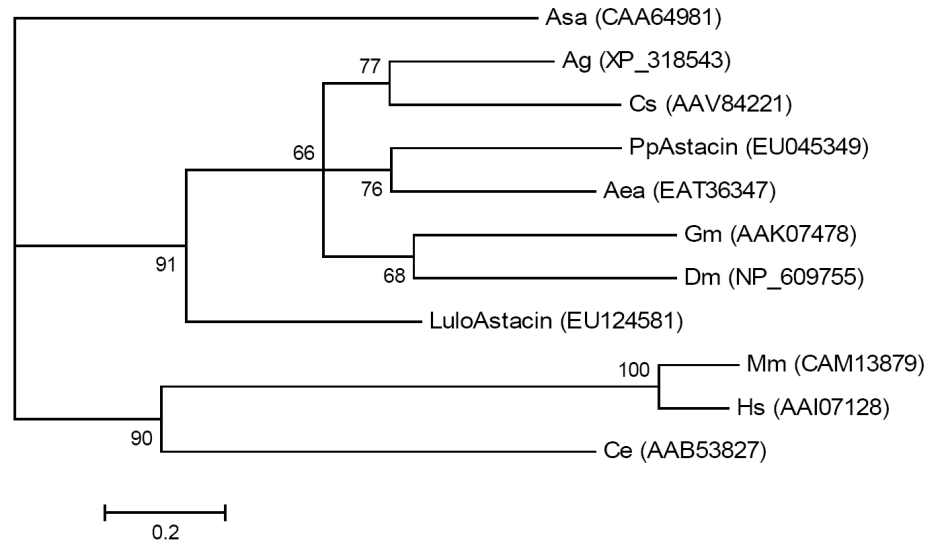
LuloCpepA1	MTISSALLVVLSCVLAQSLGEVARYDNYRIYGVTVNTHEDIKALQSLGTSGLIFIIYMTSKLAEDSQIVVAPHKFADFE
LuloCpepA2	--SALLVLIASCVFGHTLGEVARYDNYRIYGVTVNSVEDFKALESLGTSDDGYIFLSMPSKLSKFAAVVAPHKFDIFE
PpCpepA	-MKTTVVWITLCTFGQILGEVARYDNYRIYGVTVNSIKIDTSILNLEANSQSVIFLSKPSKESGESQVVVAHKLADFV
LuloCpepA1	DYLRKNCMTCKVLAENLQQLIDMERRMMNLAKSRGCGDFDNYHTLLEIHAWLKSLEQDHPDVVSVISAGNSYEGRDLLG
LuloCpepA2	SFLDSKSIKHRVTCKNLQRMIDMERRSMDFRKRSNGDFDNYHTLEEINAWLKSLEEAHPDVVSVITAGKSYDERDILG
PpCpepA	DLLQRHGLRFRILEKNLQRCIDLERKIMKRSNHKSDGFDFRYHTLEEHNWLSLEKKYDPDVVKVVSAGKSEGRDLLG
LuloCpepA1	VKLSHGAGRPATFVESGIHAREWITPATVFLVNNELLTSEDAVKDLAENYDWYVFPSPNPDGYVYTHEKLRMRKTRCP
LuloCpepA2	VKLSEGADKPGIFVEAGIHAREWITPATVFLVNNELLTSADQGVKDLAENYDWYFFPNINPDGYVYTHEKNRMWRKTLKP
PpCpepA	VELCHGENKPGVFVESGIHAREWITPATVFLVNNELLTSDPDVRYLAENFTWYVLPSPNPDGYVYTHEKNRMWRKTRKP
LuloCpepA1	YG---SCVGADANRNWDFHWNEVGASNPQCSDTYAGPFAFSEPEALAVSSYAEKLDKDVRLYSFHSFSQLIFPNGYTA
LuloCpepA2	DPDIAAGCGVDANRNWDFHWTSSETTNTPCSDIKYAGSPFASEPEVAVVANYMTSKDKLHLFLSFHSFSQLIFPNGYTE
PpCpepA	HG---TCVGVDANRNWDFHWNEVGASNPQCSDTYAGPFAFSEPETVAVSNYVKELDKDVHLYLSFHSFSQLIFPNGYTS
LuloCpepA1	QHVVDNHKDLKIGDVAAKALAKRYGTYKTVGDIYSTIYPAGTSDIYTHGVLGIDLSYCYELRPKNVFOGGFTLPKQIR
LuloCpepA2	ELVEHYNDLKDIGDSAAKALAQRYGTEHTVGDYIYSTIYPAGTSDIDWAYGALGIKLSFYELRPTSNLEGGHTLPADQIR
PpCpepA	QHVVDNHKDLKIGDVAAKALAKRYGTYKTVGDIYSTIYPAGTSDIDWSYGAMNVSLSFCYELRPKTLFQGGFTLPANQIR
LuloCpepA1	PTSLETLDLSLVALVNRKAKELNYFGEN
LuloCpepA2	PVALETIDSLVALVNRKAKELNYFDRQ
PpCpepA	PTALETLDLSLVALVNRSEELNYFETV

C

LuloCpepB	MRCSVLFLAAAFALATAAHRASYDGYQVMTIKVNDKQKLDLFEWQEKGIDFWDNLNSIGRPMRVMPPTLTDFPAFLRD
PpCpepB	MRCTFLFLAALVALTTASQRSYDGYQVMTIKTLDKQKLDLFEWQEKGIDFWDNLNSIGRPMRVMPPTLTDFPAFLRD
LuloCpepB	NEMSYELTIIPNVEITVFOERMVQMRARRAMLRARSISPRATADFSYYWQPAEINEYLRNLGTEYPNLITVEVAGNSYEGR
PpCpepB	NEISHELIIIPNVEITVLEQERKDLQESRARAMNRRFAPRATADFSYYWQPSSEINAYLRNLATEYPNLITVEVAGESYEGR
LuloCpepB	EILVVRISNIGFDGTPKIFIDAGIHAREWIAPMSALNLIHELVEHSAENTDLFACDWIIPSPVNPBGYQYTHSDRMWR
PpCpepB	EILVARISNENFDGTPKIFIDAGIHAREWIAPMSALNLIHELVEHSAENADFLACDWIIPSPVNPBGYQYTHSDREWR
LuloCpepB	KTRSVNQGSSCRGVDGNRRNYGYRWGCGGISTNPCSDIFLGRPFSELEVOQAVVNELARDAAGIRLYLSFHSYGDWLLYP
PpCpepB	KTRSVNQGSSCRGVDGNRRNYGYRWGCGAGISTNPCSDIFLGRPFSEKELQAVVNEMAKDASGIRLYLSFHSYGDWLLYP
LuloCpepB	WGYDRVLHDNNEEDLVATTVANATFOAHGROYTNGNSAILLYPAAGSDDFAAAEHDIQLSYTVELTGGGPRGFDLPAD
PpCpepB	WGYDRVLHDNNEEDLSQWSELVAEATRTAHGROYTNGNSAILLYPAAGSDDYAAGVHNINLSYTVELSGGGLGFDLPAS

Figure 27. Astacin-like metalloprotease sequence comparison and analysis.

(A) Phylogenetic analysis of amino acid sequences from *Lutzomyia longipalpis* (Lulo), *Phlebotomus papatasi* (Pp), *Mus musculus* (Mm), *Homo sapiens* (Hs), *Glossina morsitans morsitans* (Gm), *Drosophila melanogaster* (Dm), *Aedes aegypti* (Aea), *Caenorhabditis elegans* (Ce), *Anopheles gambiae* (Ag), *Astacus astacus* (Asa) and *Culicoides sonorensis* (Cs). Node support is indicated by the bootstrap values. (B) Multiple sequence alignment of Dipteran astacin-like molecules. Predicted signal peptide sequence is underlined and the residues likely necessary for catalytic activity are marked with (*).

A**B**

LuloAstacin -MKCLIVLSALAVATLASPVAKSNILM-----QPIEEMGDYFQGDMMVLTESQRR
 PpAstacin MIGVQVVILCIFSVLAKPFLPY SINWNFP-----YDTDNAEEKSGNEEGDMILSPROMI
 Aea EAT36347 -MMNKFAICAVFVALCAFRIFALPVVP-----NASDAEQSGNEEGDMILSKEQRR
 Gm AAK07478 -MRLYGVLVLLFIDFGRSLPAQNYVE-----KDPELSAGYVEGDMVLNSRQ--
 Ag XP_318543 MVVFRGLLLLVLAVAFVQAGVIKNTPENAAARLRLRPDELAELSGQFEGDIVLSEEQER
 Dm NP_609755 -MSRSGIIYVLLQVVLNSGKPLPAGV-----YDPEEAGGFVEGDMMLTEEQQR
 Cs AAV84221 -MIKFGCLKLVIFALIGSSFAFPSPKIA-----YDAEELAGKEEGDIELTPEQIR

LuloAstacin AFFNGRTDSRTGLLNERERWPNMNLVYDFAN-DVNQEOKDYIELALRNISASTCLTFSKR
 PpAstacin DLR-----FRTGLINLKYRWPKNLVPYQLSS-EFTREESEFIREALDSIECVSCLRFVE-
 Aea EAT36347 ALAG-----MRNGLFDDQYRWPNNIVYYRIISDNFTTEQVNYIRGLDTISDVSCIRFVE-
 Gm AAK07478 -----RNLGRDEVWRWPNNTVYYKFFT-VFDEDHNNYILRGMKIIIEISCI RFEE-
 Ag XP_318543 SLLSN---RRNGLIATTYRWPGNTVPVMIVEEDFTPEQIEHIKRGLRQIESVCLKFFVT-
 Dm NP_609755 NLEQGAPKARNGLINTEKRWPNVNVVYRISD-DFDTAHKKAIQTGIDTLELHTCLRFREA
 Cs AAV84221 AMKQ-----RNLGLLVTKRWNNNTVDYIITG-NYSQEOKNYIRKGLDTLQLVCLKFTIGH

LuloAstacin TNEK---DYVKVTTSS- GCSNVGRVGG-MQM RLANNEVGSGCFRFGTIVHEFIHAIG
 PpAstacin -KNSSHDFVKVSREVDSGCFSSVGYQAG-EQQINLAPNELGTGCFRKGTTIHEFLHAIG
 Aea EAT36347 -AAENSTAYIRVLGNEG-GCFSEVGYTGT-VQDINLAPNELENGCFRLGTIMHEFLHAIG
 Gm AAK07478 -ADATTNPNVNTIGFVG-GCYSEVGWLNEGAQAYNLEMYALDTGCFRLGTIVHEFLHTIG
 Ag XP_318543 -RTE-EPDYVRVIGTGS-GCYSSVGHRRG-AQTINLEPYDVTGCFRLATIVHEFIHAVG
 Dm NP_609755 TDDE--KAYLTVTAKSG-GCYTAVGYQGA-PEEMNLEIYPLGEGCFRPGTIIHEFMHAIG
 Cs AAV84221 DNATGLTDYVEVSSGG-GCSSTVGRKGG-RQTINLQSYFVEEGCFRLATIVHEFIHAIG

LuloAstacin FVHQAQSAITRDDYVLIKWENIQKGTEFNFEKEDSSKTTMFNLEDYDYGSMVHYSNKAFSIN
 PpAstacin FEHMQSASDRDDYVTIVWENINPQHVENFEKKYNESVITHFGVKYDYESVMVHYHKTAFSMN
 Aea EAT36347 FEHMQSASDRDDFVTIVWEKIEQQHQHNEKYNSSFSAFNVBYDYGSVLHYPRVSFSID
 Gm AAK07478 FEHMQSATNRDDYVHIVEGNIDPRNLHNFNKYNETQVNIQFDQBYDYGSVMVHYGPKAFSIN
 Ag XP_318543 FEHMQSASDRDFVQIVWDNIEDGKEHNFNIYDSDTVTTFVQYDYDYGSMVHYSSSTAFSIN
 Dm NP_609755 FEHMQSSSIRDDFINVIYENIVPGKEFNFKYADTVVITFEVGYDYDSCLHYRPGAFSIN
 Cs AAV84221 FEHMQSTYNRDEYVDVYENIEPGKENFNKYTEDTVTDYGIBYDYNVSMVHYGRTGFSIN

LuloAstacin DEDTIVPLQ--DG-VTIGQREMSSELDIKRLNQMYNCPDQ-
 PpAstacin DEDTIVP---KDPNAEIGQRIGLSDGIKRLNRMVYQCDEM-
 Aea EAT36347 GSATIIP---KVAGVTIGQKEMSTSDITKLNRMVYHCE---
 Gm AAK07478 GEDTIIIPLYENEAAGNMGQRLGMSKDIKLNLMYRCPIEV
 Ag XP_318543 GEKTIIVP---KDPNATIGQRVGMSERDISKLNRMVYKC---
 Dm NP_609755 GEDTIVPL---DSSAVIGQRVGLSSKDIDKINIMYKCPILL
 Cs AAV84221 GEFTLVPIK--DPEAKIGQRVGLSRRDIEKLNRMVYDCPL-

molecular weight or 28 kDa once secreted and pI of 5.36 (Table 10). LuloAstacin was most abundant in the sugar-fed cDNA library in contrast to PpAstacin, an astacin-like molecule identified in *P. papatasi* midgut, which was most abundant in the blood-fed cDNA library (Table 9). Phylogenetic analysis of other putative astacin amino acid sequences illustrate that one clade is an assemblage of the Dipteran sequences.

LuloAstacin branches out of the subclade containing PpAstacin and away from the other Dipteran sequences (Figure 27A). Further differences in amino acid sequence can be visualized in the multiple sequence alignment of Dipteran astacins and while LuloAstacin diverges from the other astacin molecules, the residues responsible for zinc-binding and activity are conserved (Figure 27B).

Peritrophin-like proteins

A number of molecules were identified as containing chitin-binding domains based on results from the conserved domains database (Tables 11-13). Three of the transcripts resembled previously identified peritrophin molecules based on sequence homology with peritrophin-A domains. The most abundant of these putative peritrophin transcripts was named *LuloPer1* (Cluster 77/78) and was overrepresented in the blood-fed *Leishmania*-infected cDNA library and likely encodes a secreted protein of 27.8 kDa (Tables 12 and 13). *LuloPer1* consists of four chitin-binding domains (Figure 28A); contrasting the other two peritrophin molecules, *LuloPer2* and *LuloPer3*, which are molecules of a single chitin-binding domain (Figure 28). *LuloPer2* and *LuloPer3* sequences originated in higher numbers from blood-fed midgut cDNA libraries and were

Table 11: Putative midgut-associated peritrophin proteins; best matched results and corresponding E-values from BLAST inquiries of a GenBank-derived non-redundant protein database and *Lutzomyia longipalpis* EST database

Cluster	Best match to non-redundant protein database	NR E value	Best match to <i>Lutzomyia</i> EST database	<i>Lutzomyia</i> E value	GenBank
77/78	similar to CG7248-PA [<i>T. castaneum</i>]	1.E-17	NSFM-114b12	5.E-155	EU124588
114	similar to CG4778-PA [<i>T. castaneum</i>]	1.E-12	NSFM-67f02	4.E-06	EU124602
171	ENSANGP00000013237 [<i>A. gambiae</i>]	4.E-10	NSFM-67f02	6.E-07	EU124607
274	conserved hypothetical protein [<i>A. aegypti</i>]	6.E-67	NSFM-35c11	3.E-28	EU124616

Table 12: Putative midgut-associated peritrophin proteins; putative function and sequence distribution contributed from each cDNA library

Cluster	Clone	Putative function	SF	BF	Number of sequences			Total
					BFi	PBMD	PBMDi	
77/78	LJGFIM27_H09	Peritrophin	0	6	22	0	0	28
114	LJGUM-P04_G10	Peritrophin	1	7	9	0	0	17
171	LJGFL_P03_H05	Peritrophin	1	4	4	1	0	10
274	LJGFIM24_D03	Chitin binding	1	0	4	1	0	6

Table 13: Putative midgut-associated peritrophin proteins; localization, molecular weight and isoelectric point of putative midgut proteins

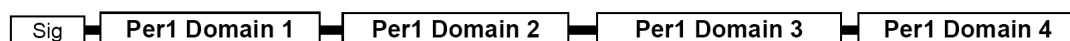
Cluster	Putative function	Gene name	Localization	Molecular weight (kDa)	Isoelectric point
77/78	Peritrophin	<i>LuloPer1</i>	Secreted	27.8	5.00
114	Peritrophin	<i>LuloPer2</i>	Secreted	9.2	4.38
171	Peritrophin	<i>LuloPer3</i>	Secreted	7.5	3.80
274	Chitin binding	<i>LuloChiBi</i>	Secreted	20.9	6.65

Figure 28. Characterization of peritrophin sequences.

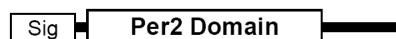
(A) Diagrammatic representation of *Lutzomyia longipalpis* peritrophin-like molecules showing the predicted signal peptide and chitin binding domains. (B) Phylogenetic analysis of predicted chitin binding domains of peritrophin molecules from *Aedes aegypti* (Ae), *Anopheles gambiae* (Ag), *Ctenocephalides felis* (Cf), *Lucilia cuprina* (Luc), *Phlebotomus papatasi* (Pp), *Lutzomyia longipalpis* (Lulo). Accession numbers are given in parentheses and bootstrap values indicate node support.

A

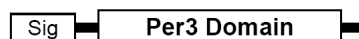
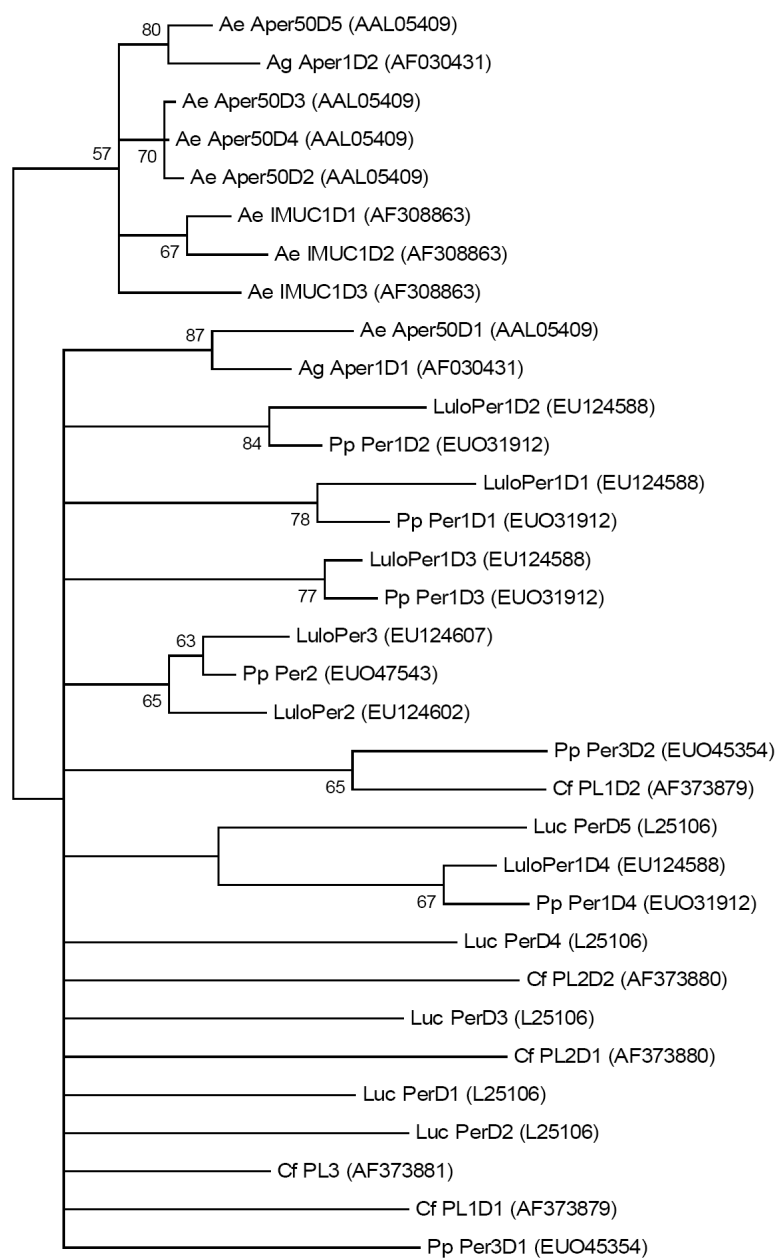
LuloPer1



LuloPer2



LuloPer3

**B**

in relatively equal numbers between the infected and uninfected sand flies. These small putative peritrophins are predicted to have a mature molecular weight of 9.2 and 7.5 kDa and isoelectric points of 4.38 and 3.8 for LuloPer2 and LuloPer3, respectively (Table 13). LuloPer1 is likely to have a role in cross linking chitin fibrils that will form the peritrophic matrix around the ingested blood bolus. LuloPer2 and LuloPer3 may have roles in capping the ends of chitin fibrils or sequestering free chitinous molecules within the midgut lumen. However, the two sequences share only 39% identity and 44% similarity, conserving primarily the cysteine residues, suggesting they may have very different ligand specificities or roles in peritrophic matrix formation and/or chitin management within the midgut. Phylogenetic analysis of the individual chitin-binding domains from several other insect peritrophin and mucin molecules demonstrate conservation of the LuloPer1 domain arrangement when compared with *P. papatasi* PpPer1, suggesting that if the domains are gene duplication events then those events occurred prior to speciation (Figure 28). Additionally, the small putative peritrophin molecules domain from LuloPer2 and LuloPer3 form a clade containing another chitin-binding domain from a small peritrophin of *P. papatasi* (Figure 28).

In addition to the putative peritrophin molecules a transcript with homology to a predicted chitin-binding domain was identified from the clustering of 6 sequences collected primarily from the blood-fed *Leishmania*-infected cDNA library. This domain has homology to a much larger chitin-binding domain than those found in the putative peritrophin molecules. The identified transcript, LuloChiBi, has one of these chitin-binding domains and is predicted to have a mature molecular weight of 20.9 kDa (Table 13).

Microvillar proteins

Among the most abundant sequences identified in the cDNA libraries were transcripts encoding putative microvillar-associated proteins with homology to insect allergens identified in *Periplaneta americana* and *Blattella germanica* (Table 14). By BLAST analysis high homology also was found to molecules in the mosquito *Aedes aegypti*. In order of decreasing overall sequence abundance, clusters 27, 29, 48, 66 and 36 were named *LuloMVP1*, *LuloMVP2*, *LuloMVP3*, *LuloMVP4* and *LuloMVP5*, respectively (Table 15). In general, the microvillar proteins were most abundant in the blood-fed cDNA libraries; although, *LuloMVP3* (cluster 48) sequences were underrepresented in the blood-fed cDNA libraries and were relatively equally identified in the sugar-fed and post-blood meal ingestion cDNA. *LuloMVP1*, *LuloMVP2* and *LuloMVP5* have nearly equal mature molecular weights of 21 kDa based on the cleavage of the predicted signal peptide present in all of the microvillar proteins while *LuloMVP3* and *LuloMVP4* are slightly larger, around 23 kDa. A notable difference in the isoelectric point among the microvillar proteins was observed. There was a predicted value of 8.84 for *LuloMVP3*; whereas, the other microvillar molecules isoelectric point ranges from 4.46 to 5.12 (Table 16).

The *Lu. longipalpis* microvillar proteins share respective homology with similar molecules identified in the midgut of *P. papatasi*, as demonstrated by amino acid phylogenetic analysis (Figure 29). The sand fly microvillar proteins are separated from the clade containing cockroaches. Additionally, *LuloMVP2* and *LuloMVP5* are in a subclade with the microvillar proteins of *Ae. aegypti* and *An. gambiae*, while the other molecules pair with the *P. papatasi* microvillar proteins (Figure 29A). Sequence

Table 14: Putative midgut-associated microvillar proteins; best matched results and corresponding E-values from BLAST inquiries of a GenBank-derived non-redundant protein database and *Lutzomyia longipalpis* EST database

Cluster	Best match to non-redundant protein database	NR E value	Best match to <i>Lutzomyia</i> EST database	<i>Lutzomyia</i> E value	GenBank
27	conserved hypothetical protein [A. aegypti]	2.E-46	NSFM-126e12	2.E-100	EU124571
29	conserved hypothetical protein [A. aegypti]	2.E-36	NSFM-19a10	6.E-99	EU124572
48	Cr-PII allergen [P. americana]	9.E-22	NSFM-68e08	1.E-101	EU124579
66	conserved hypothetical protein [A. aegypti]	4.E-27	NSFM-47h07	1.E-117	EU124584
36	putative protein G12 [A. aegypti]	6.E-41	NSFM-154e02	3.E-106	EU124577

Table 15: Putative midgut-associated microvillar proteins; putative function and sequence distribution contributed from each cDNA library

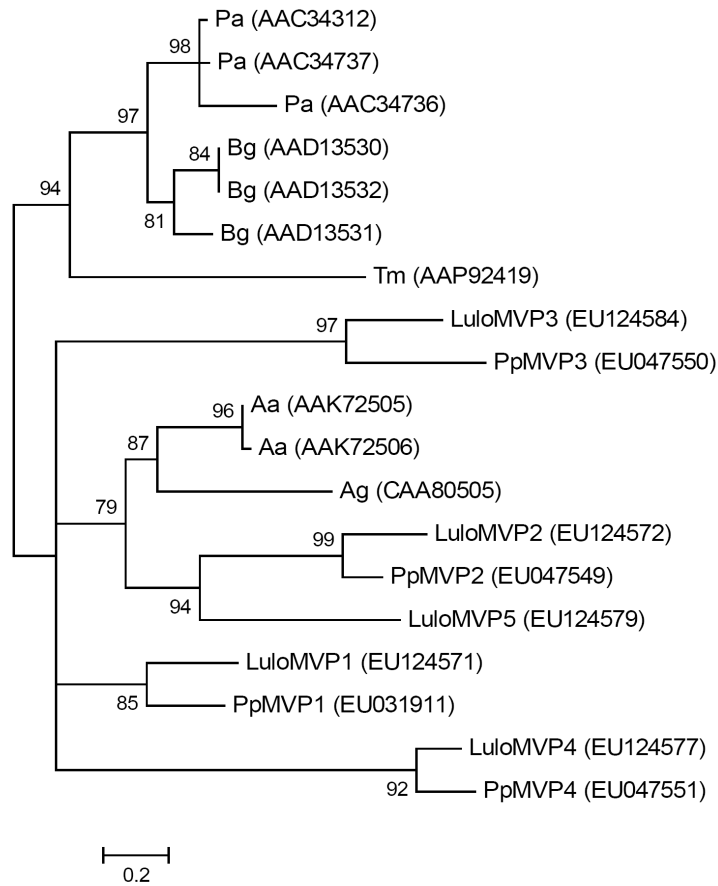
Cluster	Clone	Putative function	SF	BF	Number of sequences			Total
					BFi	PBMD	PBMDi	
27	LJGFIL9_B01	Microvillar protein	5	109	55	0	0	169
29	LJGFM9_F05	Microvillar protein	3	87	40	0	0	130
48	LJGFS_P01_C07	Microvillar protein	15	6	5	18	18	62
66	LJGFIM27_D08	Microvillar protein	1	24	7	0	0	32
36	LJGFS_P04_B01	Microvillar protein	1	60	27	0	0	88

Table 16: Putative midgut-associated microvillar proteins; localization, molecular weight and isoelectric point of putative midgut proteins

Cluster	Putative function	Gene name	Localization	Molecular weight (kDa)	Isoelectric point
27	Microvillar protein	<i>LuloMVP1</i>	Secreted	21.6	5.09
29	Microvillar protein	<i>LuloMVP2</i>	Secreted	21.5	5.12
48	Microvillar protein	<i>LuloMVP3</i>	Secreted	23.1	8.84
66	Microvillar protein	<i>LuloMVP4</i>	Secreted	23.6	4.46
36	Microvillar protein	<i>LuloMVP5</i>	Secreted	21.7	4.67

Figure 29. Sequence analysis of microvillar proteins.

(A) Phylogenetic analysis of amino acid sequences from *Blattella germanica* (Bg), *Periplaneta americana* (Pa), *Tenebrio molitor* (Tm), *Aedes aegypti* (Aa), *Anopheles gambiae* (Ag), *Phlebotomus papatasi* (Pp) and *Lutzomyia longipalpis* (Lulo). Bootstrap values indicated node support and accession numbers are given in parentheses. (B) Multiple sequence alignment of the microvillar proteins of *Lutzomyia longipalpis*. The predicted signal secretion peptide is underlined.

A**B**

LuloMVP2	--MKTSVVVLFALVAFVA-----AVSAKPAVRGLKEDADDFLNLVNLAQVKRVALKY
LuloMVP5	---MKIIVACILLSCAVS-----ALAAQP--RSLQDDFKFEALIPADATQSVVTKY
LuloMVP1	MKFFALFAIIATVAASQA-----FTVEKISTRGLDQDFQDFLDDLVPVEATIKLAVDY
LuloMVP4	--MKTIFAIFCLVCSVVAQSGDSESYFTVDPFNTPELNQDFADFVNLLPVDTVLDIVDDH
LuloMVP3	-MKLIFLVFLPLALCGEINSETDDMNRSTGVVQRGLQDDLKEFVDLIPMNKIMSVALNY
LuloMVP2	YAKDKETKQFVKYLLKGETFSAVWEQVETNEHVKHFLKYVQDSGVDVLSVLNQVAKFLKKP
LuloMVP5	YIIDGETRNFVKYIKGAQFRKVDQVETHPITKDVLEYILVSKDVPDPTSLINQLADLLGLP
LuloMVP1	LGSDKEFQAAYAYLASPEFAKLWEGEEKLNDVKECAKFLQAAGLDVYDLLNQFGGLFNLP
LuloMVP4	YENNAGINKTLNLYLKTNKEAKHWDNLESLREVHNFFVYLNESGLNIFGVLNEFAEYFELT
LuloMVP3	IVADKEVQALYNYIRSEEFRELYTAAIKTSAVRELIEILESGLPVVKYINQITPLLMLP
LuloMVP2	FVDFGDVTRKPS-----GAGLNELVKELLAEVFPDEFKALVVEKLESSPEFKQFFDTVSG
LuloMVP5	HVQP--TIANIH-----SRSLGLLYNEIVRLLPLDKFEALLNDKMENSEDFQELYQKITS
LuloMVP1	PVKP---SIVRR-----GSGLEGFLNDVLALLPKDKLVALFKKELQTSPEFKAFVDKVT
LuloMVP4	PVGFFVVEDAPEEKIEYTWGFNALVNDVMDVLPKDDLKALFDQKVANGEDFANFVECFST
LuloMVP3	TYPQRMSAEKAP-----TKGVNGLVDEILHVLPRDKLIGLFIRKTITSPDFRQLLKNLGS
LuloMVP2	FDYAAIKEFVANSNELLDLVNTVKSYSVDLDGYLHEVEQFFGWN-
LuloMVP5	IDYKVVEFVTNSAEIQDFVQRLRNKIPVDVLEGEVGTQFFGWTY
LuloMVP1	PEFEKLTQNVLNNKEFDYLLKELKAHGEDLKKFFEIVKNFFGLN-
LuloMVP4	SEFKEVLKKLELSPVAQKLFKRFRKHGLDVHKLVLQALAVFGLN-
LuloMVP3	EQTLRAVIKFAFNKEVRRAYAE LNKRGVDIIVTISKKAITYFLNLC

alignment of the *Lu. longipalpis* microvillar proteins show little sequence homology, suggesting that the classification of microvillar proteins is rather broad and perhaps that these molecules have different functions altogether.

Oxidative stress molecules

The sand fly, being an obligate blood-feeding insect, must cope with the physiological challenges posed by the digestion of blood, which includes the generation of reactive oxygen species (ROS) released by free heme and metabolic radicals produced in abundance during the digestion of the blood meal [7]. Five molecules were identified in the midgut cDNA libraries that have putative roles as antioxidants such as glutathione s-transferase (GST), catalase, copper-zinc superoxide dismutase (SOD) and peroxiredoxin (PRX) (Table 17). In addition to the protection these molecules may impart on the regulation of ROS due to blood meal digestion, there is evidence that antioxidants interact with and can impact the outcomes of infection by bacterial and parasitic agents [8]. Two transcripts were identified with homology to GST molecules of the Class Sigma and Class Delta/Epsilon subfamilies and were named *LuloGST1* and *LuloGST2*, respectively. Phylogenetic analysis of the putative GST molecules supports the separation and classification into the subfamily classes of Sigma and Delta/Epsilon. Additionally, *LuloGST1* is grouped in a subclade with other Dipteran GST molecules while *LuloGST2* diverges from the Dipteran Delta/Epsilon GST molecules (Figure 30). The *LuloGST1* cluster was generated from sequences from each of the cDNA libraries made and analyzed while *LuloGST2* consists of one sequence from the sugar-fed library

Table 17: Putative midgut-associated oxidative stress molecules; best matched results and corresponding E-values from BLAST inquiries of a GenBank-derived non-redundant protein database and *Lutzomyia longipalpis* EST database

Cluster	Best match to non-redundant protein database	NR E value	Best match to <i>Lutzomyia</i> EST database	<i>Lutzomyia</i> E value	GenBank
221	glutathione s-transferase [<i>A. aegypti</i>]	8.E-89	NSFM-105e10	9.E-105	EU124611
419	glutathione s-transferase [<i>A. aegypti</i>]	2.E-50	NSFM-95g05	2.E-116	EU124621
76	GA13179-PA [<i>D. pseudoobscura</i>]	1.E-43	NSFM-144g07	6.E-103	EU124587
79	ferritin heavy chain-like [<i>G. morsitans</i>]	2.E-66	NSFM-146d09	7.E-111	EU124589
781	catalase [<i>A. aegypti</i>]	1.E-119	NSFM-142e04	3.E-286	EU124624
1709	ENSANGP00000015824 [<i>A. gambiae</i>]	1.E-50	NSFM-39d09	1.E-91	EU124625
2557	peroxiredoxins, prx-1, prx-2, prx-3 [<i>A. aegypti</i>]	1.E-105	NSFM-34h03	3.E-126	EU124629

Table 18: Putative midgut-associated oxidative stress molecules; putative function and sequence distribution contributed from each cDNA library

Cluster	Clone	Putative function	Number of sequences					Total
			SF	BF	BFi	PBMD	PBMDi	
221	LJGFM9_A02	Glutathione s-transferase	1	2	1	2	1	7
419	LJGFIL9_G06	Glutathione s-transferase	1	0	2	0	0	3
76	LJGFL_P04_E07	Ferritin light-chain	5	7	8	3	5	28
79	LJGFIM24_C11	Ferritin heavy-chain	3	6	7	4	4	24
781	LJGFIM21_C02	Catalase	0	0	1	0	1	2
1709	LJGD-L10_H10	Cu/Zn superoxide dismutase	0	0	0	1	0	1
2557	LJGDIM21_G01	Peroxiredoxin	0	0	0	0	1	1

Table 19: Putative midgut-associated oxidative stress molecules; localization, molecular weight and isoelectric point of putative midgut proteins

Cluster	Putative function	Gene name	Localization	Molecular weight (kDa)	Isoelectric point
221	Glutathione s-transferase	<i>LuloGST1</i>	Intracellular	23.3	5.00
419	Glutathione s-transferase	<i>LuloGST2</i>	Intracellular	24.8	6.41
76	Ferritin light-chain	<i>LuloFLC</i>	Secreted	24.4	6.68
79	Ferritin heavy-chain	<i>LuloFHC</i>	Secreted	21.9	4.92
781	Catalase	<i>LuloCat</i>	Intracellular	57.7	8.11
1709	Cu/Zn superoxide dismutase	<i>LuloSOD</i>	Secreted	19.8	5.63
2557	Peroxiredoxin	<i>LuloPRX</i>	Secreted	25.0	6.66

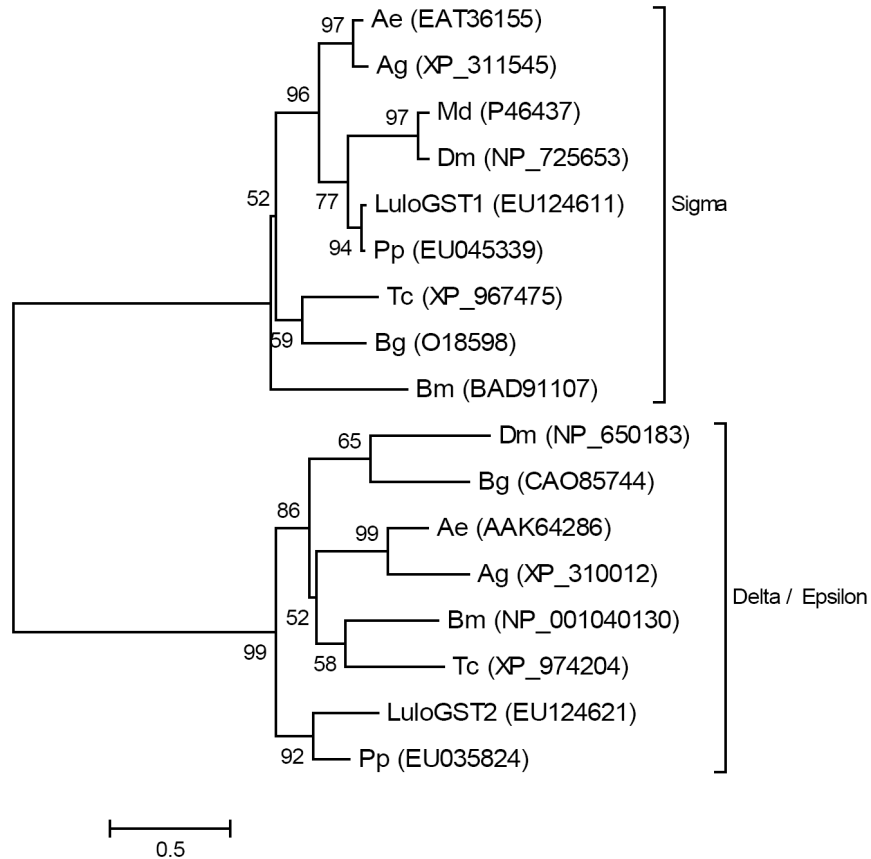


Figure 30. Phylogenetic analysis of glutathione s-transferase molecules.

Sequences analyzed from *Lutzomyia longipalpis* (Lulo), *Phlebotomus papatasi* (Pp), *Drosophila melanogaster* (Dm), *Aedes aegypti* (Ae), *Anopheles gambiae* (Ag), *Musca domestica* (Md), *Bombyx mori* (Bm), *Tribolium castaneum* (Tc) and *Blattella germanica* (Bg). Accession numbers are given in parentheses and the clades labeled with the respective glutathione s-transferase class.

and two sequences from the blood-fed *Leishmania*-infected cDNA library. Additional antioxidant molecules include a catalase (*LuloCAT*), copper-zinc superoxide dismutase (*LuloSOD*) and peroxiredoxin (*LuloPRX*) of which *LuloSOD* and *LuloPRX* are both predicted to be secreted based on the presence of a likely signal peptide sequence. ROS and reactive nitrogen oxide species (RNOS) are important in host defenses against microorganisms and *LuloCAT*, *LuloSOD* and *LuloPRX* are molecules which may serve to regulate and prevent damage of the sand fly midgut by the ROS and RNOS defenses similar to the protective effect of peroxiredoxin in *An. stephensi* [9].

Upon the ingestion of a blood meal by a hematophagous insect, a large amount of iron and heme are released during digestion. To combat the toxic effects of free iron and the generation of damaging reactive oxygen species, ferritin is produced to sequester the iron and hemoglobin that is liberated by the digestion of red blood cells. Ferritin molecules are commonly associated with iron metabolism, and it is likely that the molecules identified in this transcriptome engage in metabolic function. However, given the relative size of the blood meal in comparison with the sand fly, ferritin molecules within the midgut likely serve a large role in preventing the generation of oxygen radicals by the Fenton reaction. Two transcripts from clusters 76 and 79 were identified with homology to ferritin light-chain and ferritin heavy-chain molecules and were named *LuloFLC* and *LuloFHC*, respectively (Tables 17 and 19). The expression of *LuloFLC* and *LuloFHC* appears to be constitutive based on the number of sequences generated in each cDNA library spanning the condition of sugar-fed, blood-fed, and post-blood meal digestion (Table 18).

Serine protease inhibitors

Two types of serine protease inhibitors were identified in the cDNA libraries; a single sequence with homology to SERPIN and a cluster of 17 sequences with homology to a Kazal-type serine protease inhibitor (Tables 20-22). SERPIN molecules within the midgut of the sand fly may serve to counteract damaging proteases produced by microorganisms; however, LuloSRPN lacks a predicted signal peptide sequence and thus may serve an intracellular housekeeping function. *LuloKZL*, identified from cluster 112, is a small molecule of 6.3 kDa and is predicted to be secreted. Comparison of *LuloKZL* with Kazal-type serine protease inhibitors found in a transcriptome analysis of the midgut of *P. papatasi* identified PpKZL1 as a highly conserved homolog. Kazal-type protease inhibitors, such as rhodniin and infestin identified in *Rhodnius prolixus* and *Triatoma infestans*, respectively, have been characterized as thrombin inhibitors; thereby, these molecules would prevent coagulation of ingested blood to facilitate successful digestion of the blood meal [10, 11]. *LuloKZL* sequences are more abundant prior to and during blood meal digestion based on the number of sequences in the sugar-fed, blood-fed and post-blood meal digestion cDNA libraries. Additionally, *LuloKZL* was not identified in an EST analysis of whole sand fly *Lu. longipalpis* and is therefore more likely a midgut-specific molecule found in abundance only in the alimentary tissue [5]. Thus, a prudent hypothesis would be that *LuloKZL* serves a similar function, allowing the blood bolus to remain in a colloidal suspension within the gut to facilitate peristalsis and digestion.

Table 20: House keeping and low abundant transcripts from the midgut of *Lutzomyia longipalpis*; best matched results and corresponding E-values from BLAST inquiries of a GenBank-derived non-redundant protein database and *Lu. longipalpis* EST database

Cluster	Best match to non-redundant protein database	NR E value	Best match to <i>Lutzomyia</i> EST database	E value	GenBank
128	GAPDH II [<i>D. pseudoobscura</i>]	1.E-168	SFM-03d02	2.E-161	EUI24605
195	fructose-bisphosphate aldolase [<i>A. aegypti</i>]	1.E-138	NSFM-99b02	5.E-153	EUI24609
189	sugar transporter [<i>A. aegypti</i>]	0.E+00	NSFM-46e10	1.E-255	EUI24608
200	ENSANGP00000018531 [<i>A. gambiae</i>]	0.E+00	SFM-05b09	4.E-225	EUI24610
292	cytochrome c oxidase subunit iv [<i>A. aegypti</i>]	3.E-58	NSFM-43a12	6.E-97	EUI24618
97	ADP/ATP translocase [<i>L. cuprina</i>]	1.E-154	NSFM-64b06	2.E-161	EUI24598
69	Vacuolar ATP synthase 16 kDa proteolipid subu [<i>A. aegypti</i>]	2.E-77	NSFM-95b05	6.E-70	EUI24586
67/192	Actin 87E CG18290-PA, isoform A [<i>D. melanogaster</i>]	0.E+00	NSFM-41f08	5.E-202	EUI24585
112	GA16408-PA [<i>D. pseudoobscura</i>]	4.E-10			EUI24601
2287	serine protease inhibitor 4 [<i>A. aegypti</i>]	3.E-27	NSFM-73e11	3.E-185	EUI24627
358	RAS, putative [<i>A. aegypti</i>]	1.E-90	NSFM-155h05	2.E-93	EUI24620
2556	ENSANGP00000016718 [<i>A. gambiae</i>]	1.E-103	NSFM-83c08	4.E-99	EUI24628
500	conserved hypothetical protein [<i>A. aegypti</i>]	7.E-08	NSFM-154d08	3.E-09	EUI24623
235	peptidoglycan recognition protein LB [<i>G. morsitans</i>]	8.E-69	NSFM-81b08	4.E-109	EUI24614
1960	defensin isoform B1 [<i>A. aegypti</i>]	2.E-12			EUI24626
269	40S ribosomal protein S7 ribosomal protein [<i>C. pipiens</i>]	3.E-89	NSFM-15f05	1.E-97	EUI24615
423	ribosomal protein S20 [<i>B. mori</i>]	7.E-56	NSFM-41g09	2.E-59	EUI24622
226	ribosomal protein S8 [<i>A. albopictus</i>]	6.E-93	NSFM-01c05	9.E-97	EUI24612
125	LD16326p [<i>D. melanogaster</i>]	1.E-100	NSFM-52a05	2.E-84	EUI24604
304	Ribosomal protein L32 CG7939-PC, isoform C [<i>D. melanogaster</i>]	1.E-67			EUI24619
101	GA20389-PA [<i>D. pseudoobscura</i>]	1.E-153	SFM-03h12	2.E-133	EUI24599
108	60S acidic ribosomal protein P1 [<i>S. frugiperda</i>]	3.E-48	NSFM-163b12	5.E-27	EUI24600
119	similar to Drosophila melanogaster CG2099 [<i>D. yakuba</i>]	2.E-54	NSFM-100a07	6.E-62	EUI24603
40	similar to Neurospecific receptor kinase CG4007-PA [<i>A. mellifera</i>]	8.E-01	NSFM-57e04	6.E-140	EUI24578
54/55	14.5 kDa salivary protein [<i>P. duboscqi</i>]	8.E-41			EUI24580
88	bS11M [<i>A. aegypti</i>]	2.E-03	NSFM-149f10	1.E-58	EUI24596
151	CG14401-PA [<i>D. melanogaster</i>]	3.E-06	NSFM-114e07	1.E-11	EUI24606
230	conserved hypothetical protein [<i>A. aegypti</i>]	8.E-22			EUI24613
90	Hypothetical protein C30H6.11 [<i>C. elegans</i>]	2.E-10	NSFM-55h01	7.E-64	EUI24597
276	CG32644-PB [<i>D. melanogaster</i>]	2.E-11	NSFM-23d08	8.E-27	EUI24617

Table 21: House keeping and low abundant transcripts from the midgut of *Lutzomyia longipalpis*; putative function and sequence distribution contributed from each cDNA library

Cluster	Clone	Putative function	SF	BF	Number of sequences			Total
					BFI	PBMD	PBMDi	
128	LJGDiM22_F06	Glyceraldehyde-3-phosphate dehydrogenase	1	1	0	4	7	13
195	LJGDIL8_D01	Fructose-bisphosphate aldolase	0	0	0	3	3	6
189	LJGFL_P04_A09	Sugar transporter	2	3	0	2	1	8
200	LJGD-L2_G11	Enolase	0	1	0	3	2	6
292	LJGFIM27_F04	Cytochrome c oxidase IV	1	2	1	0	1	5
97	LJGDiM25_C06	ADP/ATP translocase	4	0	2	3	6	15
69	LJGDIL9_F01	V-ATPase C-subunit	2	6	4	7	7	26
67/192	LJGFL_P01_G02	Actin	10	16	4	6	2	38
112	LJGUM-P03_G07	Kazal-type serine protease inhibitor	6	4	5	2	0	17
2287	LJGFIM21_F02	Serine protease inhibitor 4	0	0	1	0	0	1
358	LJGFMS_B11	Ras	1	1	2	0	0	4
2556	LJGDiM21_F11	Aquaporin	0	0	0	0	1	1
500	LJGFIL8_B01	Galectin	0	0	1	2	0	3
235	LJGFS_P02_C04	Peptidoglycan recognition protein	3	2	0	0	1	6
1960	LJGDM27_A10	Defensin	0	0	0	1	0	1
269	LJGUM-P03_F07	40S ribosomal protein S7	1	1	1	2	1	6
423	LJGDM25_A04	40S ribosomal protein S20	1	1	0	1	0	3
226	LJGDiM26_A12	40S ribosomal protein S8	1	2	1	2	1	7
125	LJGFIM25_D10	60S ribosomal protein L19	5	2	2	2	2	13
304	LJGF-L8_E06	60S ribosomal protein L32	1	2	0	0	2	5
101	LJGFIL3_D04	60S acidic ribosomal protein P0	3	0	4	5	5	17
108	LJGU-m-5_A09	60S acidic ribosomal protein P1	6	0	2	6	4	18
119	LJGUS_P03_A12	60S Ribosomal protein L35Ae	10	1	1	3	0	15
40	LJGFIL7_D12	Unknown	6	4	13	25	22	70
54/55	LJGF-I-10_A05	Unknown	7	13	13	11	4	48
88	LJGU-I-10_D11	Unknown	5	2	2	5	2	16
151	LJGFIL1_H05	Unknown	0	2	6	1	2	11
230	LJGFIM22_H04	Unknown	0	1	3	1	2	7
90	LJGU-I-7_C03	Unknown	5	5	5	4	5	24
276	LJGFIL2_F05	Unknown	0	1	2	3	0	6

Table 22: House keeping and low abundant transcripts from the midgut of *Lutzomyia longipalpis*; localization, molecular weight and isoelectric point of putative midgut proteins

Cluster	Putative function	Gene name	Localization	Molecular weight (kDa)	Isoelectric point
128	Glyceraldehyde-3-phosphate dehydrogenase		Intracellular	35.2	7.84
195	Fructose-bisphosphate aldolase		Intracellular	30.8	6.73
189	Sugar transporter		Transmembrane	53.7	7.03
200	Enolase		Intracellular	46.7	6.5
292	Cytochrome c oxidase IV		Intracellular	20.7	9.26
97	ADP/ATP translocase		Transmembrane	33.3	9.87
69	V-ATPase C-subunit		Transmembrane	16.0	8.41
67/192	Actin		Intracellular	41.8	5.29
112	Kazal-type serine protease inhibitor	<i>LuloKZL</i>	Secreted	6.3	4.83
2287	Serine protease inhibitor 4	<i>LuloSRPN</i>	Intracellular	42.1	4.95
358	Ras		Intracellular	20.5	5.20
2556	Aquaporin		Transmembrane	27.7	8.71
500	Galectin	<i>LuloGalec</i>	Intracellular	17.2	7.33
235	Peptidoglycan recognition protein	<i>LuloPGRP</i>	Intracellular	21.9	6.75
1960	Defensin	<i>LuloDEF</i>	Secreted	7.2	6.89
269	40S ribosomal protein S7		Intracellular	21.9	9.82
423	40S ribosomal protein S20		Intracellular	13.4	10.44
226	40S ribosomal protein S8		Intracellular	23.6	10.72
125	60s ribosomal protein L19		Intracellular	24.0	11.13
304	60S ribosomal protein L32		Intracellular	16.0	11.77
101	60S acidic ribosomal protein P0		Intracellular	34.2	6.23
108	60S acidic ribosomal protein P1		Intracellular	11.5	4.08
119	60S Ribosomal protein L35Ae		Intracellular	16.8	11.24
40	Unknown		Secreted	29.2	9.59
54/55	Unknown		Secreted	14.3	9.13
88	Unknown		Secreted	14.5	7.78
151	Unknown		Secreted	11.9	4.72
230	Unknown		Secreted	11.6	9.95
90	Unknown		Secreted	19.0	3.8
276	Unknown		Secreted	16.6	3.41

Anti-bacterial molecules

Two molecules, originating from clusters 235 and 1960, encode a putative peptidoglycan recognition protein (*LuloPGRP*) and defensin (*LuloDEF*), respectively. *LuloPGRP* is similar to other predicted peptidoglycan recognition proteins found in *Glossina morsitans morsitans* and mosquitoes, while it is phylogenetically distinct from Lepidopteran molecules (Figure 31). This is the first report of a putative PGRP identified in sand flies, and in searching a midgut transcriptome database of *P. papatasi*, a molecule was identified with 87% identity. *LuloPGRP* may serve as a pattern recognition protein, specifically for the conserved structure of peptidoglycan indicated by the conservation of the amino acid sequence among insects, as a component of the sand fly immune system defense against bacterial pathogens (Figure 31). PGRP molecules characterized in *Bombyx mori* and *Trichoplusia ni* have been shown to be expressed primarily in the fat body and hemocytes, and it is conceivable that the identification of *LuloPGRP* transcripts arose due to a contamination of the tissue sample [12, 13]. It is possible that the midgut tissue of sand flies express a PGRP for protection against microorganisms ingested during sugar and blood-feeding as a PGRP was identified as preferentially expressed in the midgut of *Samia cynthia ricini* [14].

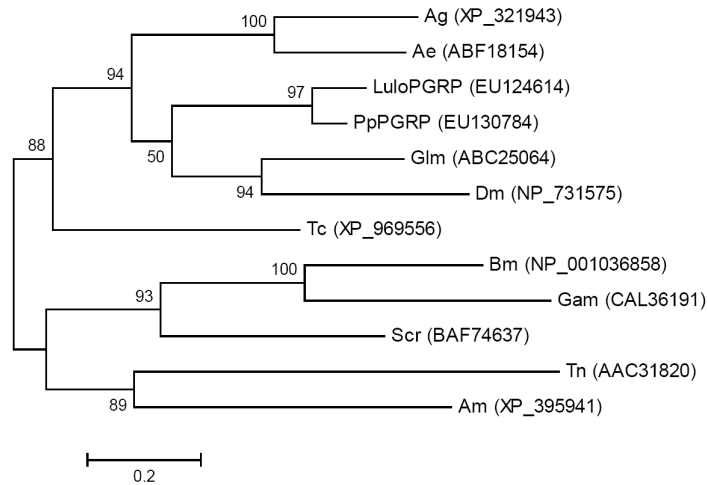
Defensins are another type of innate immune defense that insects possess to ward off pathogenic bacteria. A single sequence, named *LuloDEF*, was identified in the post-blood meal digestion midgut cDNA library with homology to a defensin molecule characterized in *Ae. aegypti*. Like other insect defensin molecules, *LuloDEF* has a predicted secretion signal peptide, and most sequence homology is given by the carboxyl half of the sequence and conservation of cysteine residues (Figure 32). *LuloDEF* shares

Figure 31. Sequence analysis of peptidoglycan recognition proteins.

(A) Phylogenetic analysis of amino acid sequences of peptidoglycan recognition proteins from *Lutzomyia longipalpis* (Lulo), *Phlebotomus papatasi* (Pp), *Anopheles gambiae* (Ag), *Aedes aegypti* (Ae), *Glossina moristans moristans* (Glm), *Drosophila melanogaster* (Dm), *Tribolium castaneum* (Tc), *Apis mellifera* (Am), *Bombyx mori* (Bm), *Galleria mellonella* (Gam), *Trichoplusia ni* (Tn) and *Samia cynthia ricini* (Scr).

Accession numbers are in parentheses and bootstrap values indicate node support. (B)

Multiple sequence alignment of peptidoglycan recognition proteins. Identical amino acid residues are highlighted black and similar residues are highlighted grey.

A**B**

Lulo	-----MPGTDFAARVLYFS-----ALS-----
Pp	----MPGVDFAAALLYFS-----AMT-----
Glm	---MTAFGFVLLSMMGWMN-----ETTPHLS-----
Dm	---MTALGLVLLSMMGYSQ-----HMQQANLGDGVATA
Ag	LRAYCQLDSLIMIVYVIVIIASVIQLHAAVIRDAVMELFFPSDDSDTTTAPTMTYGANPV
Ae	--MYGKGGVLIVTVLQIVS-----HFEAGSSCL
Tc	---MYTFLFCFAAFLATGQMS-----EL
Bm	--MFLSFCIFIVFCAYTSSHP-----RLIEKDHLVSDF
Gam	-----IDGKDYPF
Scr	MFNLSIGLFVTIIMNVKAYPSIF-----SGESVENEVPSYDF
Am	MTKLI AVLFLLVNCQILFCS-----VHETPVRP
Tn	---MEILFVLFFVFTVSG-----DC

Lulo	DIVSRCDWGAPPKLIVETENG--PSQEVIIHHSYIIPGECHDSAEICIKAMQSMQDFHQNDR
Pp	DVVSRCBWGAAPPKLIVETENG--PAPFVLIHHSYIIPGECHSSSECIKAMQNMORFHQNDR
Glm	RITARSBWGARDPLIVEKFIG--PSAFVLIHHSYIIPACYTDDCKKAMRSMQDFHQLER
Dm	RLLSRSBWGARLPKSVETFOG--PAPYVLIHHSYIIPAVCYSTPDCKMSMRQDFHQLER
Ag	PYVTRDWSALPPKRIEHPAG--PIPYVLIHHSYIIPACYNGLQCIAMQSMQKHQDER
Ae	PYTBREAWDAKPPKSVETHPAG--PIPYVLIHHSYIIPACLTGSECCBAMLSMQKHQDER
Tc	VVVEREGWHARPPPTATEPMAN--PVPEVLIHHSYIIPACHTPEACVQSMQTMQDMHQLQN
Bm	PVCSRDGWGAVPSKDTPLNK--PVPIYVLIHHTAIPVTCNTTTQCMRDMRSMQKHNS-L
Gam	PFYSRADWNTDPTDVRPLIT--LTPYVLIHHSYIIPACNDTAQCEAMQSMQDYHKS-L
Scr	PFVSRQWSARQPNQTLPLKT--PVPIYVLIHHSYIIPACHTRETCCAMRSMQNFHMDGH
Am	RITSRSEWGARKPTTITRALAQNPPEVLIHHSAT--DSCITQAI CNARVRSFQNYHIDEK
Tn	GVVTKDEWDGLTPIHVEYLAR--PVELVLIHHTVT--STCNTDAACQIVRNISYHMDNL

Lulo	GWNDIGYSFVGGDGNIIYGRGFSVIGAHAPRYNDKSVGICLIGDWREKLPPKMLRAVR
Pp	GWNDIGYSFVGGDGNIIYGRGFSVIGAHAPRYNDKSVGICLIGDWREKLPPKMLQAVQ
Glm	GWNDIGYSFVGGDGNIIYGRGFNVIGAHAPRYNDKSVGICLIGDWRLDFTDKMLQATR
Dm	GWNDIGYSFVGGDGMIIYGRGFNVIGAHAPRYNDKSVGIVLIGDWRTLEPPKMLDAAK
Ag	QWNDIGYSFVGGDGHVYIGRGFNVIAGAHAPRYNNRSVGICLIGDWVADLPPKNMLTAAQ
Ae	GWNDIGYSFVGGDGRVYEGRGFNVIAGAHAPRYNDKSVGICLIGDWTVDLPPENMLTAAQ
Tc	GWNDIGYSFVGGDGNAYEGRGWSKVIAGAHAPRYNNISIGICVIGDWTKELPENQNTVH
Bm	GWNDIGYHFCVGGDGVAYEGRWNVIGTHAGPANKLSIGICLIGDWRVETPPAEQLATTK
Gam	DWGDIGYHFCVGSBGGAYEGRWNVIGTHAGKANSISIGICLIGDWREQLPPVQQLETTK
Scr	QWNDIGYHFCVSSDGTIVYEGRGWSTLGAHALHNSVSIIGICLIGDWRVSLPPADQIKATK
Am	GWNDIGYQFVGGDGNIIYEGRWDKHGAHSISYNSKSIIGICLIGFVGHTEPNAATPATK
Tn	NYWDIGSSFIIGGKGVYEGAGWLHVGAHTYGYNRKSIGITFTGNYNNDKETQKSLDALR

Lulo	KLIQYGVEEGHIERNYTLGLHRQVRDTECPGQRLFEIKTWPHFSRPPSPGPNNDYKAFR-
Pp	KLIQSGVEEGYIAKNYTLGLHRQVRDTECPGQRLFEIKTWPHFSRPPSPGPNNDYRPPS-
Glm	DLIAFGVLSQGIYHRQYQLGLHRQVRATECPGQRLYKEIKTWPHFSRPPVNNNNNNNNNN--
Dm	NLIAFGVFKGIYDPAYKLLGLHRQVRDTECPGQRLFAEISSWPHFTINDTEGVSTTAPV
Ag	NLIEYGVNRNLIQAQNYTLGLHRQVRDTECPGQRLFEIKTWPHFETMTDIVDQNSV----
Ae	SLIEYGVNRNIIAANYTLGLHRQVRDTECPGQRLFEIKTWPHFSMLDVVDPNISIRT--
Tc	KLIAFGVEKGWIRETYKLLGLHRQVRDTECPGQRLFEIKTWPHFEGVKGERRQENNGVEDN
Bm	KLLESTGVEMGAISSDYKLLGHNOAMTTECPGQALFEISTWDNYHFGHVNPRELNKQTKF
Gam	ALIEEGVKGREISSDYKLLGHQALFTECPGTALFFPRGGWGFVPPQMYPPPLKKTHTK-
Scr	SLIAAGVELGISPCYKLVGHROVRATECPGDALYENIKTWTHWSAFPSSVKDLIHVKEL
Am	NLISYGVRAIGRIQSNYTLGLHQRTTRTECPGDSLYELIKTWPHSSSI-----
Tn	ALLRCGVERGHLTANYHIVGHRLISTEPGRKLYNEIRRWDFHLDN-----

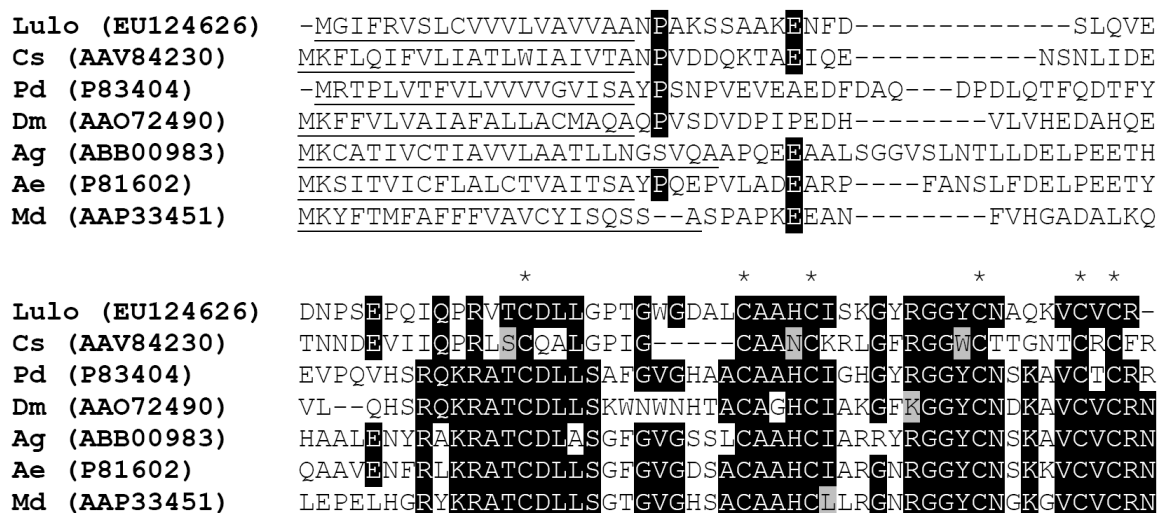


Figure 32. Multiple sequence alignment of putative defensin sequences.

Aligned sequences from *Lutzomyia longipalpis* (Lulo), *Culicoides sonorensis* (Cs), *Phlebotomus duboscqi* (Pd), *Drosophila melanogaster* (Dm), *Anopheles gambiae* (Ag), *Aedes aegypti* (Ae) and *Mus mus domestica* (Md). Conserved cysteine residues are marked with (*).

47% identity and 61% similarity with a defensin characterized in *Phlebotomus duboscqi*, which is induced by the presence of wild type *Leishmania major* [15]. Both immunity-associated genes, *LuloPGRP* and *LuloDEF*, may have an impact on the progression and result of a midgut infection by *Leishmania* parasites, either directly or by indirect effects if co-colonization of the midgut with bacteria is an intermediary confounding factor.

Transcripts differentially expressed by blood-feeding and digestion

A comparison between the sugar-fed and blood and between the blood-fed and post-blood meal digestion libraries was conducted using Pearson's chi-square equation to identify overrepresented transcripts within each cluster. As was previously seen in *P. papatasi* a number of digestion-associated transcripts were overabundant in the blood-fed cDNA library [6]. We envisioned similar results in the analysis of the *Lu. longipalpis* midgut cDNA libraries with the enhanced advantage of a cDNA library produced from midguts that had fully processed and excreted the blood meal byproducts. It was our hypothesis that the post-blood meal midgut transcript abundance would be most similar to the sugar-fed midgut transcript abundance prior to a blood meal. Overall, the number of sequences per cluster was similar in the sugar-fed cDNA library to those in the post-blood meal digestion cDNA library and more transcripts are overrepresented in the blood-fed library (Table 23). Several exceptions to both overall observations do occur, however. Most of the microvillar protein transcripts are abundant in the blood-fed cDNA library except for *LuloMVP3*, which is highly represented in the sugar-fed and post-blood meal digestion cDNA libraries. This reinforces the suggestion that the microvillar

Table 23: Sequence distribution altered during sugar-feeding and blood meal digestion; clusters overrepresented in the sugar-fed, blood-fed and post-blood meal digestion midgut cDNA libraries as determined by χ^2 statistical analysis

Putative function	Cluster #	SG	BF	PBMD	P value	GenBank
Microvillar protein (<i>LuloMVP1</i>)	27	5	109	0	6.8E-03	EUI24571
Microvillar protein (<i>LuloMVP2</i>)	29	3	87	0	3.2E-06	EUI24572
Microvillar protein (<i>LuloMVP3</i>)	48	15	6	18	4.2E-02	EUI24579
Microvillar protein (<i>LuloMVP4</i>)	66	1	24	0	3.5E-02	EUI24584
Microvillar protein (<i>LuloMVP5</i>)	36	1	60	0	3.2E-04	EUI24577
Trypsin (<i>Utryp1</i>)	35	3	55	0	4.9E-08	ABM26904
Trypsin (<i>Utryp2</i>)	18	136	6	109	4.2E-02	ABM26905
Chymotrypsin (<i>LuloChym1A</i>)	33	3	51	1	1.1E-17	EUI24576
Chymotrypsin (<i>LuloChym2</i>)	64	0	17	0	2.4E-02	EUI24583
Astacin-like metalloprotease (<i>LuloAstacin</i>)	58	23	5	0	3.6E-16	EUI24581
Unknown	40	6	4	25	3.4E-02	EUI24578

Table 24: Sequence distribution altered during sugar-feeding and blood meal digestion; clusters that appear overabundant, but are not statically significant by χ^2 analysis, in the sugar-fed, blood-fed and post-blood meal digestion midgut cDNA libraries

Putative function	Cluster #	SG	BF	PBMD	P value	GenBank
Peritrophin (<i>LuloPer1</i>)	77	0	6	0	4.3E-02	EUI24588
Peritrophin (<i>LuloPer2</i>)	114	1	7	0	2.0E-05	EUI24602
Chymotrypsin (<i>LuloChym3</i>)	87	1	14	0	4.2E-02	EUI24591
Chymotrypsin (<i>LuloChym4</i>)	30	12	1	1	1.7E-03	EUI24573
Carboxypeptidase (<i>LuloCpepA1</i>)	104	0	14	0	1.4E-02	EUI24592
Carboxypeptidase (<i>LuloCpepA2</i>)	107	6	5	0	3.40E-02	EUI24593
Carboxypeptidase (<i>LuloCpepB</i>)	91	1	8	1	2.4E-07	EUI24594
60S acidic ribosomal protein P0	101	3	0	5	1.80E-02	EUI24599
60S acidic ribosomal protein P1	108	6	0	6	9.2E-05	EUI24600
60S Ribosomal protein L35Ae	119	10	1	3	4.2E-02	EUI24603

proteins are likely functionally different molecules grouped solely on homology to previously annotated sequences. In general, proteases appear to be induced by the act of blood-feeding or the presence of a blood meal within the midgut; with the exception of *Lltryp2*, which is significantly more abundant in the sugar-fed and post-blood meal digestion cDNA libraries and also *LuloAstacin* that is more abundant in the sugar-fed cDNA library (Tables 23 and 24). These molecules may be produced and stored prior to blood-feeding for immediate use in digestion or perhaps have a role other than digestion altogether, such as immunity. Other proteases such as *LuloChym4* and *LuloCpepA2* are present in higher or near equal numbers in the sugar-fed library when compared with that of the blood-fed library. Molecules, such as the peritrophins *LuloPer1* and *LuloPer2*, are also more plentiful in the blood-fed cDNA library, suggesting that these molecules may be transcribed only in response to blood-feeding. A transcript encoding a predicted protein of unknown function derived from cluster 40 was identified as being most abundant in the post-blood meal digestion cDNA library, signifying it may play a role outside of blood meal digestion, such as oogenesis.

Transcripts differentially expressed by the presence of *Leishmania infantum chagasi*

To evaluate the effects of the presence of *L. infantum chagasi* parasites on the transcript abundance in the midgut tissue of the sand fly we compared the number of sequences in each cluster between the blood-fed and blood-fed *Leishmania*-infected cDNA library as well as between the post-blood meal digestion and post-blood meal digestion *Leishmania*-infected cDNA library using Chi-square analysis (Tables 25-27).

Table 25: Sequence distribution altered by *Leishmania infantum chagasi*; clusters overrepresented in the blood-fed and blood-fed *L. infantum chagasi*-infected midgut cDNA libraries as determined by χ^2 statistical analysis

Putative function	Cluster #	BF	BFi	P value	GenBank
Microvillar protein (<i>LuloMVP1</i>)	27	109	55	2.0E-04	EU124571
Microvillar protein (<i>LuloMVP2</i>)	29	87	40	2.0E-04	EU124572
Microvillar protein (<i>LuloMVP4</i>)	66	24	7	4.9E-03	EU124584
Microvillar protein (<i>LuloMVP5</i>)	36	60	27	1.6E-03	EU124577
Peritrophin (<i>LuloPer1</i>)	77/78	6	22	1.0E-03	EU124588
Trypsin (<i>Lltryp2</i>)	18	6	15	2.9E-02	ABM26905
Chymotrypsin (<i>LuloChym1A</i>)	33	51	22	2.4E-03	EU124576
Carboxypeptidase (<i>LuloCpepA1</i>)	104	14	3	1.3E-02	EU124592
Actin	67/192	16	4	1.3E-02	EU124585
Unknown	40	4	13	1.7E-02	EU124578

Table 26: Sequence distribution altered by *Leishmania infantum chagasi*; clusters overrepresented in the post-blood meal digestion and post-blood meal digestion *L. infantum chagasi*-infected midgut cDNA libraries as determined by χ^2 statistical analysis

Putative function	Cluster #	PBMD	PBMDi	P value	GenBank
Trypsin (<i>Lltryp2</i>)	18	109	168	2.0E-04	ABM26905

Table 27: Sequence distribution altered by *Leishmania infantum chagasi*; LuloTryp3 appears underrepresented in the post-blood meal digestion *L. infantum chagasi*-infected midgut cDNA library

Putative function	Cluster #	PBMD	PBMDi	GenBank
Trypsin (<i>LuloTryp3</i>)	83	7	1	EU124590

We hypothesized that the effects of the parasite's presence in the blood engorged sand fly would likely mirror what we had observed in a similar comparison of *P. papatasi* infected with *L. major*. Additionally, we hypothesized that the analysis of the post-blood meal digestion midgut tissue would reveal a large number of differentially abundant transcripts, as during this time period *Leishmania* parasites are interacting with the midgut epithelium, replicating and differentiating to the metacyclic form. In accordance with what we observed previously in blood engorged *P. papatasi* infected with *L. major*, there was an under representation of the microvillar protein transcripts [6]. Similar trends in abundance between infected *P. papatasi* and infected *Lu. longipalpis* also occur for transcripts encoding the putative digestion enzymes trypsin (*Lltryp2*) and chymotrypsin (*LuloChym1A*). Two other digestive proteases, *LuloAstacin* and *LuloCpepA1*, were identified as differentially abundant in the presence of *L. infantum chagasi* with a reduction in the number of transcripts captured in the blood-fed *Leishmania*-infected library; however, only the *LuloCpepA1* difference was statistically significant. There is a striking contradiction in the modulated abundance of peritrophin transcripts. In the midgut of infected *P. papatasi*, peritrophin transcripts decrease; whereas, *Lu. longipalpis* infected with *L. infantum chagasi* has a significant overrepresentation of peritrophin (*LuloPer1*) and overrepresentation of the putative chitin-binding molecule (*LuloChiBi*). There appears to be a downregulation of actin transcripts by the presence of the *L. infantum chagasi* parasites in the midgut. We speculate that this could be a tactic of the parasite to decrease the cytoskeletal rearrangement that occurs after blood-feeding as a means of decreasing peristalsis, which may aid in the retention of the parasite within the gut of the sand fly.

In the context of abundant transcripts, the post-blood meal digestion midgut infected with *L. infantum chagasi* is relatively quiescent. Only one transcript, encoding a putative trypsin molecule, was identified as significantly different in abundance. *Lltryp2* sequences were 1.54 times more abundant in the *L. infantum chagasi*-infected post-blood meal digestion cDNA library, which corroborates the observed overrepresentation of *Lltryp2* sequences in the blood-fed infected cDNA library. It is possible that the increase in sand fly *Lltryp2* occurs due to the presence of a perceived pathogen or as a consequence of a non-specific perception of contents within the midgut. Conversely, *LuloTryp3* transcripts were captured at a lower frequency in the *L. infantum chagasi*-infected midgut after blood meal digestion.

Conclusion

Leishmania parasites develop to a transmissible and infective form entirely within the confines of the alimentary tract of the sand fly, in contrast to numerous other arthropod-borne pathogens. We wished to further investigate the response of the sand fly midgut tissues that are occurring in reaction to blood meal ingestion and interactions with *Leishmania* parasites. The previously reported extensive sequencing of whole sand fly *Lutzomyia longipalpis* ESTs provided a large overview of the transcripts present in this vector. However, it did not provide information regarding tissue specific transcripts, particularly from the sand fly midgut or information regarding the midgut molecules that may be transcribed in response to blood-feeding and digestion or interact with the *Leishmania* parasite. In the present work, the production of five different cDNA libraries generated a large number of redundant tissue specific transcripts for analysis as well as

provided the capability of a comparative analysis between these cDNA libraries. Several molecules were identified in this midgut-specific transcriptome that were not identified in the EST database of whole sand fly sequences, including *LuloKZL* and *LuloDEF*.

The present analysis of midgut tissue from *Lu. longipalpis* further increases our knowledge of the molecular events that occur throughout the adult life cycle of the sand fly. In general, it appears that the midgut reverts, after complete digestion and excretion of the blood meal, to a state nearly mimicking the midgut of a sand fly that has only taken a sugar meal. Comparing data generated from the sugar-fed and blood-fed sand fly midguts resulted in comparable global changes found in the similar analysis of the midgut of *P. papatasi* [6]. Microvillar proteins, digestive proteases and peritrophin molecules are some of the transcripts identified as differentially represented between cDNA libraries when comparing unfed and blood-fed sand flies. Interestingly, many molecules, such as microvillar proteins and digestive proteases, were found to be over- or underrepresented when comparing the blood-fed with the blood-fed *Leishmania*-infected cDNA libraries. Similar results were observed in the midgut of *P. papatasi* when infected with *L. major*. This not only demonstrates the reproducibility of this technique of analyzing transcript abundance across cDNA libraries, but the redundancy present in the biology of blood-feeding and digestion in sand flies as well as the *Leishmania*-vector interactions occurring between Old World and New World sand fly species. When comparing the uninfected and *L. infantum chagasi*-infected post-blood meal digestion library we were astounded by the scarcity of differentially abundant transcripts when considering the number and volume of *Leishmania* parasites present in the midgut at the time points encompassed by the cDNA library. These data suggest that the *Leishmania*

parasite affects the midgut expression profile during the blood digestion process and not afterwards. It is likely that *Leishmania* parasites modulate the expression profile of other molecules, but our approach was not able to detect these proteins, probably as a result of their low abundance. Further testing employing more direct techniques such as real-time PCR or other expression profile approaches are still required to test the hypothesis that *L. infantum chagasi* is altering the expression of specific gut transcripts from the sand fly *Lutzomyia longipalpis*. However, the information presented in the current work and previous work on *P. papatasi* and *L. major* strongly suggest that *Leishmania* parasites can alter the expression of midgut transcripts that may be relevant for the survival and establishment of the parasite in the gut of the fly. Also, these changes may be occurring during the digestion of the blood meal and not afterwards.

Methods

Sand flies

Lutzomyia longipalpis sand flies (Jacobina strain) were maintained at the Laboratory of Malaria and Vector Research at the National Institute of Allergy and Infectious Diseases. Three to four-day post eclosion sand flies were allowed a 20% sucrose solution (sugar-fed/unfed) or fed blood on anesthetized BALB/c mice (blood-fed).

***Leishmania* and sand fly infection**

The infection of *Lu. longipalpis* using an artificial blood meal containing *Leishmania infantum chagasi*-infected macrophages (blood-fed, infected) was based on

the work of Tesh and Modi [16]. Briefly, *L. infantum chagasi* MHOM/BROO/MER/Strain 2 promastigote cultures were maintained in M199 (Sigma-Aldrich, St. Louis, MO) containing 20% (V\V) fetal bovine serum(FBS) (Invitrogen, Carlsbad, CA) and 100 units/ml Penicillin, 100 µg/ml Streptomycin and 0.292 mg/ml Glutamine (PSG) (Invitrogen, Carlsbad, CA) at 25°C. Macrophage cell line J774A.1 (American Type Culture Collection, Manassas, VA) was cultured in RPMI (Invitrogen, Carlsbad, CA) containing 10% FBS and PSG at 37.0°C, 95% air, 5% CO₂. At confluency, the macrophages were scraped from the culture flask and washed twice by centrifugation in phosphate buffered saline (PBS) at 380 x g for 10 minutes before resuspension in culture media. The washed macrophages were then placed in 5 wells of a 24-well culture plate at a concentration of 2×10^6 cells/ml and allowed to adhere for 90 minutes at 37°C, 5% CO₂. Stationary-phase *L. infantum chagasi* culture was washed by centrifugation in PBS at 1200 x g for 15 minutes and resuspended in macrophage culture media. Nonadherent macrophages were removed by the replacement of the culture media and *Leishmania* parasites added at a 5:1 ratio of parasite to macrophage. The parasites were co-cultured with the macrophages for 5 hours at 26°C. The culture was then washed to remove extracellular parasites and the macrophages scraped from the wells. Macrophages were confirmed to contain intracellular amastigotes by staining with QUICK III (Astral Diagnostics, Inc., West Deptford, NJ) according to the manufacture's protocol and visualized by light microscopy. The infected macrophage culture was centrifuged at 380 x g for 10 minutes and resuspended in 500µl fresh whole mouse blood collected in heparin. The blood containing amastigote-infected macrophages was used for artificial blood-feeding of sand flies as described [17].

cDNA library construction

The conditions of the *Lutzomyia longipalpis* midguts harvested for the construction of the five cDNA libraries included: unfed/sugar-fed four-day post-eclosion female midguts; midguts containing blood 1, 2, and 3 days post-blood meal from female flies allowed to feed on BALB/c mice; midguts containing blood 1, 2, and 3 days post-blood meal from female flies allowed an artificial blood meal containing *L. infantum chagasi* infected macrophages; midguts devoid of blood 5, 6, and 7 days post-blood meal from gravid flies allowed to feed on BALB /c mice; midguts devoid of blood 5, 6, and 7 days post-blood meal from gravid flies allowed an artificial blood meal containing *L. infantum chagasi* infected macrophages. All midguts used for the construction of cDNA libraries containing infected sand fly midguts were verified by microscopy as carrying *Leishmania* parasite infections comparable to mature infections seen in sand flies that are used routinely in transmission experiments. *Lu. longipalpis* midguts were dissected in phosphate buffered saline (PBS), placed in RNAlater (Sigma-Aldrich, St. Louis, MO) and stored at 4°C prior to cDNA library construction. Libraries constructed using midguts at different time points consisted of two midguts at each day the midguts were dissected. *Lu. longipalpis* midgut mRNA was isolated from six midguts using the MicroFastTrack mRNA isolation kit (Invitrogen, San Diego, CA). The cDNA libraries were constructed using the SMART cDNA Library Construction Kit (Clontech, Mountain View, CA) as described previously [18].

DNA Sequencing

Phage plaques lacking β -galactosidase activity were picked from the soft top agar using a sterilized wooden stick and placed into 75 μ l of ultrapure water in a 96-well v-bottom plate. PCR was used to amplify the cDNA insert from 3 μ l of the phage in water using FastStart PCR Master premixed PCR reagent (Roche Applied Science, Indianapolis, IN) and primers PT2F1 (AAGTACTCTAGCAATTGTGAGC) and PT2R1 (CTCTTCGCTATTACGCCAGCTG). Reaction conditions were 75°C, 3 min; 94°C, 4 min; 33 cycles of 94°C, 1 min; 49°C, 1 min; 72°C 2 min; a final extension of 72°C for 7 minutes. The PCR products were cleaned of buffering salts, dNTPs, and primers using ExcelaPure 96-well UF PCR purification plates (Edge Biosystems, Gaithersburg, MD) using three washes of 100 μ l of ultrapure water and recovery in 30 μ l of ultrapure water. Cycle sequencing was accomplished using BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA), primer PT2F3 (TCTCGGGAAGCGCGCCATTGT), and 5 μ l of the cleaned PCR product. The cycle sequencing products were prepared for sequencing by centrifugation through hydrated Sephadex G-50 (Amersham, Piscataway, NJ), desiccation, and rehydration with 10 μ l sequencing buffer. Sequencing was performed using a 3730xl DNA analyzer (Applied Biosystems, Foster City, CA).

Bioinformatics

Detailed reports of the bioinformatic analysis of the data were previously reported [19, 20]. Succinctly, high N (unidentified nucleotide) content was removed at the 5' and 3' ends of each sequence, as well as any primer or vector nucleotides removed. Sequences from all five libraries were combined and contigs constructed from the

clustering of homologous sequences based on 100% identity over 64 nucleotides, while sequences with greater than 5% N's were discarded. Three frame translated sequences were supplied to the appropriate BLAST algorithm for comparison to the contents of the NCBI non-redundant protein database, the Gene Ontology database [21] and the conserved domain database [22], which contains the eukaryotic clusters of orthologous groups (COG), Simple Modular Architecture Tool (SMART) and Protein Family Database (Pfam) [23, 24]. Customized databases of mitochondrial and ribosomal RNA nucleotide sequences also were used for the comparison of cDNA sequences. The predicted presence of a signal secretion peptide or transmembrane helices was determined using the SignalP [25] or TMHMM server [26], respectively. A custom program, Count Libraries, was used to identify the number of transcripts that each library contributed to the formation of a contig (JMC Ribeiro). The contigs, information for each contig, the BLAST and SignalP results were combined in a hyperlinked Excel spreadsheet and each contig annotated by manually assigning the most likely predicted function based on BLAST results. Sequences were aligned using Clustal X, version 1.83, and converted to graphically aligned sequences using BioEdit, version 7.0.5.3 [27]. Phylogenetic analysis was conducted on amino acid alignments using TREE-PUZZLE, version 5.2, generating trees by maximum likelihood using quartet puzzling with 10,000 puzzling steps to calculate node support [28]. Statistical significance in the number of transcripts per cluster within that same cluster, between cDNA libraries, was analyzed using Pearson's Chi-square test.

References

1. den Boer M, Davidson RN: **Treatment options for visceral leishmaniasis.** *Expert Rev Anti Infect Ther* 2006, **4**(2):187-197.
2. Kamhawi S: **Phlebotomine sand flies and *Leishmania* parasites: friends or foes?** *Trends Parasitol* 2006, **22**(9):439-445.
3. Ramalho-Ortigao JM, Kamhawi S, Rowton ED, Ribeiro JM, Valenzuela JG: **Cloning and characterization of trypsin- and chymotrypsin-like proteases from the midgut of the sand fly vector *Phlebotomus papatasi*.** *Insect Biochem Mol Biol* 2003, **33**(2):163-171.
4. Schlein Y, Jacobson RL: **Resistance of *Phlebotomus papatasi* to infection with *Leishmania donovani* is modulated by components of the infective bloodmeal.** *Parasitology* 1998, **117** (Pt 5):467-473.
5. Dillon RJ, Ivens AC, Churcher C, Holroyd N, Quail MA, Rogers ME, Soares MB, Bonaldo MF, Casavant TL, Lehane MJ *et al*: **Analysis of ESTs from *Lutzomyia longipalpis* sand flies and their contribution toward understanding the insect-parasite relationship.** *Genomics* 2006, **88**(6):831-840.
6. Jochim RC, Ramalho-Ortigao M, Anderson JM, Lawyer PG, Pham VM, Kamhawi S, Valenzuela JG: **Exploring the midgut transcriptome of *Phlebotomus papatasi*: comparative analysis of expression profiles of sugar-fed, blood-fed and *Leishmania major*-infected sandflies.** *BMC Genomics* 2007, **8**(1):300.
7. Vincent SH: **Oxidative effects of heme and porphyrins on proteins and lipids.** *Semin Hematol* 1989, **26**(2):105-113.

8. Kumar S, Christophides GK, Cantera R, Charles B, Han YS, Meister S, Dimopoulos G, Kafatos FC, Barillas-Mury C: **The role of reactive oxygen species on *Plasmodium* melanotic encapsulation in *Anopheles gambiae*.** *Proc Natl Acad Sci U S A* 2003, **100**(24):14139-14144.
9. Peterson TM, Luckhart S: **A mosquito 2-Cys peroxiredoxin protects against nitrosative and oxidative stresses associated with malaria parasite infection.** *Free Radic Biol Med* 2006, **40**(6):1067-1082.
10. Friedrich T, Kroger B, Bialojan S, Lemaire HG, Hoffken HW, Reuschenbach P, Otte M, Dodt J: **A Kazal-type inhibitor with thrombin specificity from *Rhodnius prolixus*.** *J Biol Chem* 1993, **268**(22):16216-16222.
11. Campos IT, Amino R, Sampaio CA, Auerswald EA, Friedrich T, Lemaire HG, Schenkman S, Tanaka AS: **Infestin, a thrombin inhibitor presents in *Triatoma infestans* midgut, a Chagas' disease vector: gene cloning, expression and characterization of the inhibitor.** *Insect Biochem Mol Biol* 2002, **32**(9):991-997.
12. Ochiai M, Ashida M: **A pattern recognition protein for peptidoglycan. Cloning the cDNA and the gene of the silkworm, *Bombyx mori*.** *J Biol Chem* 1999, **274**(17):11854-11858.
13. Kang D, Liu G, Lundstrom A, Gelius E, Steiner H: **A peptidoglycan recognition protein in innate immunity conserved from insects to humans.** *Proc Natl Acad Sci U S A* 1998, **95**(17):10078-10082.
14. Hashimoto K, Mega K, Matsumoto Y, Bao Y, Yamano Y, Morishima I: **Three peptidoglycan recognition protein (PGRP) genes encoding potential amidase**

- from eri-silkworm, *Samia cynthia ricini*. *Comp Biochem Physiol B Biochem Mol Biol* 2007.
15. Boulanger N, Lowenberger C, Volf P, Ursic R, Sigutova L, Sabatier L, Svobodova M, Beverley SM, Spath G, Brun R *et al*: **Characterization of a defensin from the sand fly *Phlebotomus duboscqi* induced by challenge with bacteria or the protozoan parasite *Leishmania major***. *Infect Immun* 2004, 72(12):7140-7146.
 16. Tesh RB, Modi GB: **A simple method for experimental infection of phlebotomine sand flies with *Leishmania***. *Am J Trop Med Hyg* 1984, 33(1):41-46.
 17. Harre JG, Dorsey KM, Armstrong KL, Burge JR, Kinnamon KE: **Comparative fecundity and survival rates of *Phlebotomus papatasi* sandflies membrane fed on blood from eight mammal species**. *Med Vet Entomol* 2001, 15(2):189-196.
 18. Valenzuela JG, Belkaid Y, Rowton E, Ribeiro JM: **The salivary apyrase of the blood-sucking sand fly *Phlebotomus papatasi* belongs to the novel *Cimex* family of apyrases**. *J Exp Biol* 2001, 204(Pt 2):229-237.
 19. Valenzuela JG, Pham VM, Garfield MK, Francischetti IM, Ribeiro JM: **Toward a description of the sialome of the adult female mosquito *Aedes aegypti***. *Insect Biochem Mol Biol* 2002, 32(9):1101-1122.
 20. Valenzuela JG, Francischetti IM, Pham VM, Garfield MK, Mather TN, Ribeiro JM: **Exploring the sialome of the tick *Ixodes scapularis***. *J Exp Biol* 2002, 205(Pt 18):2843-2864.

21. **Mappings of External Classification Systems to GO**
[<http://www.geneontology.org/GO.indices.shtml>].
22. **NCBI Conserved Domain Database (CDD)**
[<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>].
23. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z *et al*: **CDD: a Conserved Domain Database for protein classification**. *Nucleic Acids Res* 2005, **33**(Database issue):D192-196.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. *Nat Genet* 2000, **25**(1):25-29.
25. **SignalP 3.0 Server** [<http://www.cbs.dtu.dk/services/SignalP/>].
26. **TMHMM Server, v. 2.0** [<http://www.cbs.dtu.dk/services/TMHMM/>].
27. Tippmann HF: **Analysis for free: comparing programs for sequence analysis**. *Brief Bioinform* 2004, **5**(1):82-87.
28. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing**. *Bioinformatics* 2002, **18**(3):502-504.

Chapter 5

Temporal profiling of transcripts in the midgut tissue of the sand fly *Lutzomyia longipalpis*: effects of blood-feeding and *Leishmania infantum chagasi* infection

Abstract

The midgut of a sand fly is a key organ for *Leishmania* parasites successful proliferation and differentiation into an infective and transmissible form, but little is known about the molecular interactions that occur between *Leishmania infantum chagasi* and the principal vector, *Lutzomyia longipalpis*. A comparative analysis of five cDNA libraries identified transcripts that may be differentially modulated at certain time points after blood meal ingestion in the presence or absence of *L. infantum chagasi*. Several transcripts, which were found in higher numbers in midgut cDNA libraries from sugar-fed, blood-fed, or *L. infantum chagasi*-infected sand flies, were chosen for multiplexed quantitative RT-PCR analysis. Temporal expression profiles of transcripts, such as those encoding putative chymotrypsin, carboxypeptidase, astacin-like serine protease, peritrophin-like and microvillar proteins, showed strong modulation by the ingestion of a blood meal. Actin and a trypsin molecule showed very subtle changes in transcript abundance in the presence of *L. infantum chagasi*. Colonization of *Lu. longipalpis* by *L. infantum chagasi* induced changes both during blood digestion and at later time points associated with metacyclogenesis and transmission. Furthermore, transcript profile analysis allowed the comparison between blood-fed and *L. infantum chagasi* infected sand flies as the blood meal is digested and the parasite proliferates within the midgut. Temporal profiling of specific transcripts provides insights into the processes of blood meal digestion as well as demonstrating the effect *L. infantum chagasi* has on the transcripts of *Lu. longipalpis* midgut tissue.

Background

The leishmaniasis are world-wide tropical diseases infecting an estimated 12 million people with over 350 million people at risk. Visceral leishmaniasis is one of several manifestations of disease caused by the protozoan parasite *Leishmania*. Visceral leishmaniasis is a fatal disease if untreated, and the traditional treatment, administration of the toxic drug class of pentavalent antimonials, commonly causes adverse effects and sometimes life-threatening conditions such as cardiac arrhythmia and acute pancreatitis. There is one licensed vaccine for canine leishmaniasis, which may function as an immunotherapeutic and transmission blocking vaccine as well; however, no vaccine is available for use in humans [1, 2]. Current pursuits for a vaccine against visceral leishmaniasis are focusing on *Leishmania*-host interactions, sand fly-host interactions and sand fly-*Leishmania* interactions. Vaccines based on *Leishmania*-host interactions, vaccination using parasite proteins, are the largest of the three aforementioned efforts and have produced a vast amount of information. Exploration of the sand fly-host interaction as a means of vaccination, namely, the use of sand fly saliva to elicit a protective immune response, has been very productive. Comparably, the workforce researching sand fly-*Leishmania* interactions as potential vaccine targets attempting to block transmission are smaller and have focused on *Leishmania*-derived molecules. There is a fundamental need to fill this knowledge gap for both scientific and medical importance.

The sand fly *Lutzomyia longipalpis* is a vector of the protozoan agent known to cause visceral leishmaniasis, *Leishmania infantum chagasi*. Infection of the sand fly occurs in females following the ingestion of an infected blood meal. The *Leishmania* parasite then develops solely within the midgut through several stages and defeats or

avoids life cycle barriers to ultimately propagate and mature into the infective metacyclic promastigote. A transmission blocking vaccine exploits the necessity for specific molecular interactions that occur during the development of a transmissible infection. Unlike arboviruses or *Plasmodium* that migrate and infect the salivary glands of their respective vectors, the *Leishmania* parasite develops within the confines of the alimentary canal of the sand fly. Through the sequencing of midgut tissue-specific cDNA libraries of the sand fly *Phlebotomus papatasi*, the receptor necessary for *Leishmania major* parasite binding, PpGalec, was identified and shown to be required for complete development of a transmissible infection [3, 4].

Sequencing cDNA libraries of the midgut of *Lutzomyia longipalpis* identified numerous molecules that likely interact with and impact the life cycle of the pathogen *Leishmania infantum chagasi* [5]. Differential comparison of the transcript abundance in cDNA libraries constructed from midgut tissues under conditions of sugar-fed, blood-fed, post-blood meal digestion and in the presence or absence of *L. infantum chagasi* identified molecules potentially altered by each of these conditions. Elucidation of the molecular events occurring during blood-feeding and digestion, as well as the interactions between the sand fly midgut tissue and *Leishmania* parasites, warrants further research. In this chapter, we provide an in depth temporal quantitation of ten transcripts identified in the midgut transcriptome and reveal the impact on transcription by colonization of *L. infantum chagasi* within the midgut.

Results and discussion

Previously, differential comparison and sequencing of midgut cDNA libraries of the sand fly *Lutzomyia longipalpis* found a number of transcripts were over- or under-expressed when *Leishmania infantum chagasi* was present within the midgut [5]. Ten transcripts were chosen, based on their potential importance in *L. infantum chagasi* infection of the sand fly and sequence abundance in the cDNA libraries, for further investigation by quantitative multiplex reverse transcriptase-PCR (*qm* RT-PCR) (Table 28). The transcripts analyzed consist of proteins categorized as proteases, peritrophins, and microvillar proteins as well as actin. The proteases are likely integral to blood meal digestion and have potential importance in the establishment of successful *Leishmania* colonization. Peritrophic matrix formation begins immediately following blood ingestion so as to envelope the blood bolus. While the peritrophic matrix is important in blood meal digestion, it serves as a barrier to *Leishmania* parasites that must cross the peritrophic matrix and bind the midgut epithelium to complete the life cycle within the sand fly. Microvillar proteins, identified as some of the most abundant midgut transcripts, are functionally uncharacterized proteins that are over- or under-expressed in the different conditions evaluated by cDNA library sequencing. Table 29 provides the gene name and putative translational product, NCBI accession number, PCR product size without the additional nucleotides provided by the multiplex universal primers, primer sequences and the concentration of primer used in the multiplex reaction. The cDNA libraries generating the previously reported transcript abundances are derived from the pooling of midgut tissue from the general conditions of blood-fed (visible blood within

Table 28: Transcript abundance in the cDNA libraries of *Lutzomyia longipalpis* sugar-fed (SF), blood-fed (BF), blood-fed *L. infantum chagasi*-infected (BFi), post-blood meal digestion (PBMD), post-blood meal digestion *L. infantum chagasi*-infected

Gene	SF	BF	BFi	PBMD	PBMDi
<i>MVP3</i>	15	6	5	18	18
<i>MVP4</i>	1	24	7	0	0
<i>Per1</i>	0	6	22	0	0
<i>Per2</i>	1	7	9	0	0
<i>CpepA1</i>	0	14	3	0	0
<i>CpepB</i>	1	8	8	1	1
<i>Chym3</i>	1	14	8	0	2
<i>Astacin</i>	28	7	1	1	4
<i>Tryp3</i>	8	7	3	7	1
<i>Actin</i>	10	16	4	6	2

Table 29: Quantitative multiplex RT-PCR primers

Gene (product)	Accession	Product Size	Forward Sequence	Reverse Sequence	Concentration
<i>MVP3</i> (Microvillar protein)	EU124579	149	CGGCTGCAATTAAGACTTCC	GGTAGGTGCCTTCTCAGCAC	1 μ M
<i>MVP4</i> (Microvillar protein)	EU124584	100	GTTGCCAATGGTGAGGATTT	TCTGGGCAACTGGAGATAGC	15.6 nM
<i>Per1</i> (Peritrophin)	EU124588	209	GCTCTGAGGGACACTCGTTC	GTAGCCTTCTGGGCATTTGA	15.6 nM
<i>Per2</i> (Peritrophin)	EU124602	177	CTTGTTATGTTCCCGCATGA	AGCCTCAGTATTGCCAGGTG	15.6 nM
<i>CpepA1</i> (Carboxypeptidase A)	EU124592	184	ACTGGAATGAAGTTGGTGCC	GCGGTGTAACCATTAGGGAA	15.6 nM
<i>CpepB</i> (Carboxypeptidase B)	EU124594	128	TTCTCCAGCTCTTACCGAA	AGCACGTCTCTGCATCTGAA	15.6 nM
<i>Chym3</i> (Chymotrypsin)	EU124591	191	GCCAGTATGAGCAGAGGAGG	TGTGATTGTGAGGCTTCCAG	15.6 nM
<i>Astacin</i> (Astacin-like metalloprotease)	EU124581	223	CCTCACATTTCAGCAAGCGTA	ATAATCATCGCGGGGTAGG	1 μ M
<i>Tryp3</i> (Trypsin)	EU124590	230	TACACAGACGAGCTTCCACG	CATTCTGCGTATTTCCCCAT	1 μ M
(40S ribosomal protein S7)	EU124615	202	TTGATTGCGACAGCAAGAAAG	GAACGTGGTCGCTTTTGT	1 μ M
(Actin)	EU124585	268	AATCCTGCGGTATTCACGAG	TTGGAGATCCACATCTGCTG	1 μ M
(<i>L. infantum</i> alpha-tubulin)	XM_001464016	296	AGTGCATCGGTGTTGAGGAT	GAACACCATAAAGCCCTGGA	1 μ M

the midgut) and post-blood meal digestion (no visible traces of blood in gravid sand fly midgut). The following is a comparative and temporal profiling of transcripts in the midgut tissue of sugar-fed and 6 hours, daily from 1-7 days and 13 days post-blood meal ingestion in artificially blood-fed and *L. infantum chagasi* infected blood-fed female *Lu. longipalpis* sand flies.

***L. infantum chagasi* is most abundant prior to the defecation of digested blood.**

Leishmania infection was verified and quantified using *qm* RT-PCR to correlate the presence of parasite with alterations in midgut transcript abundance. *L. infantum chagasi* replicates to its highest number two days after the infectious blood meal is taken, as measured by the presence of *Leishmania* alpha-tubulin (TUA) transcripts (Figure 33). Parasite numbers, or TUA transcripts, were below the detectable limit six hours post-blood meal (PBM) and are first detected ~24 hours PBM. TUA abundance peaks at day 2 PBM, corresponding with rapid parasite replication within the blood meal. The TUA transcripts then decrease at day 3, likely due to parasite death and a decrease in TUA expression. Excretion of the parasites in the feces of the sand fly likely contributes to the considerable decrease in TUA transcripts; however, most of the blood meal byproducts were excreted between days 3 and 4. There is then a slight and steady increase in TUA expression as the parasites continue to replicate in the midgut.

Microvillar protein transcripts are affected by the presence of *L. infantum chagasi*.

The two microvillar proteins selected for analysis demonstrate contrasting expression patterns in the cDNA libraries with *MVP3* being most abundant in the absence

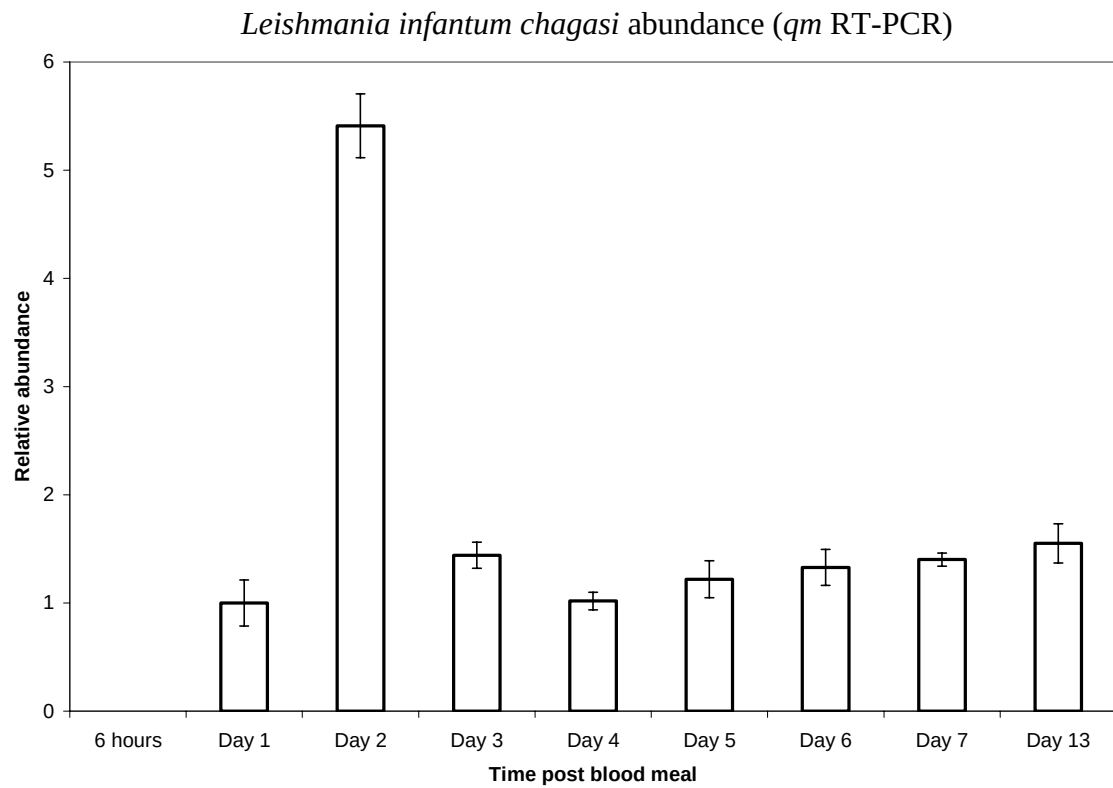


Figure 33. Temporal profile of *L. infantum chagasi* infection in the midgut of *Lu. longipalpis*.

L. infantum chagasi alpha-tubulin relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, after artificial blood-feeding of *L. infantum chagasi*-infected macrophages. Error bars represent standard deviation.

of a blood meal and *MVP4* found almost solely in the presence of blood within the midgut (Table 28). *MVP3* is more abundant in sugar-fed sand flies than during the first three days succeeding blood meal ingestion and *MVP3* increases with peak abundance reached at five days after blood-feeding and returns to a nearly equivalent level to that found in the sugar-fed fly by 13 days PBM (Figure 34A). The impact of *Leishmania* on *MVP3* transcript abundance in the cDNA libraries was immeasurable, but there appears to be a significant increase during blood digestion, followed by a nearly two-fold decrease five days PBM. It appears that the presence of *L. infantum chagasi* parasites within the midgut advances the expression of *MVP3*. We also investigated *MVP4* abundance. As early as six hours after blood-feeding, *MVP4* is dramatically induced and reaches peak levels two days PBM and is then maintained at very low amounts 13 days after feeding (Figure 35A). There was a slight impact on *MVP4* expression in the infected sand fly, with a detectable increase at one and five days PBM and a decrease at three days. To reaffirm the impact of *Leishmania* on transcript abundance as measured by quantitative multiplex RT-PCR the blood-fed and *L. infantum chagasi*-infected samples were analyzed by real-time RT-PCR (*rt* RT-PCR). Overall, the results of the *rt* RT-PCR show the same relative abundance of transcripts at all time points assayed by *qm* RT-PCR (Figures 34B and 35B). The inherent sensitivity of *rt* RT-PCR revealed stronger impacts on *MVP3* and *MVP4* expression. Expanding on the dichotomy of microvillar protein expression, it is more evident that *MVP3* expression is induced by *L. infantum chagasi* during blood meal digestion, while *MVP4* is induced after the digestion and excretion of blood meal products. Annotated as microvillar proteins due to the

Figure 34. Comparative abundance of microvillar protein *MVP3* in uninfected and *L. infantum chagasi*-infected *Lu. longipalpis* midguts.

Transcript abundance measured by (A) quantitative multiplex RT-PCR and (B) real-time RT-PCR. Bars indicate *MVP3* relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, in sugar-fed, blood-fed (white bars) or blood-fed infected (striped bars) sand flies. Diamonds indicate the *MVP3* expression ratio of infected to uninfected sand flies. Error bars represent 95% confidence intervals.

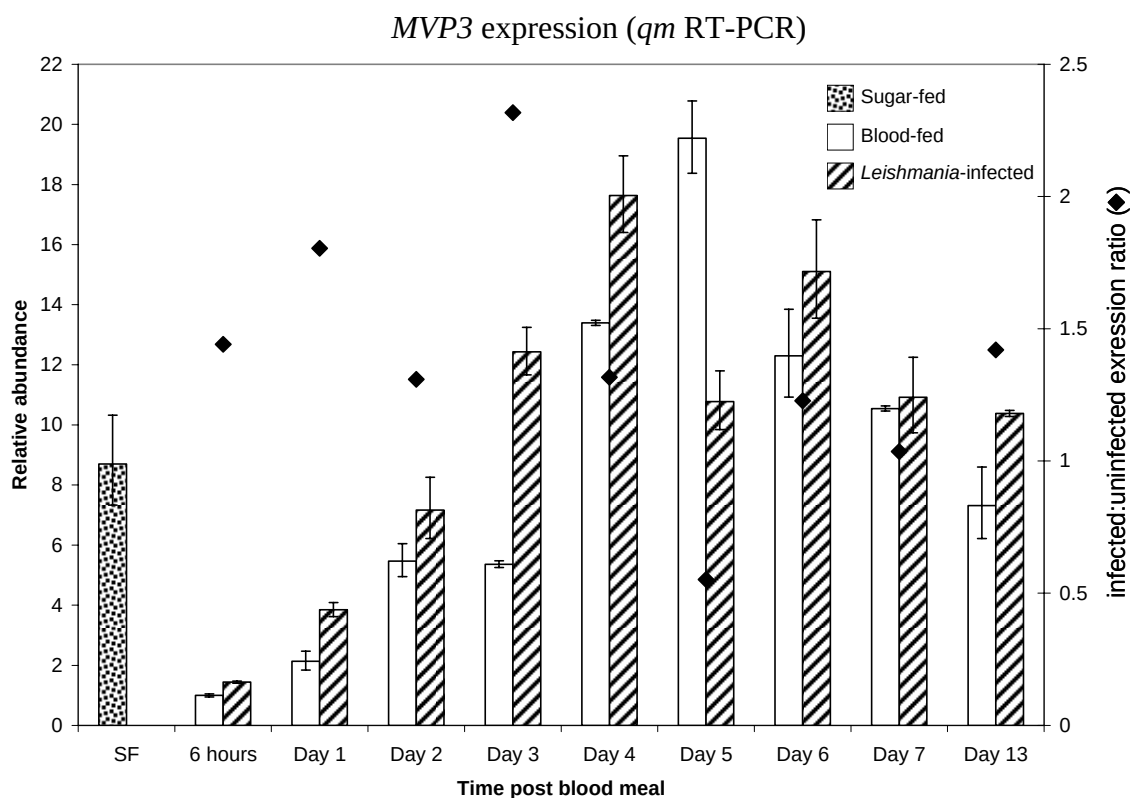
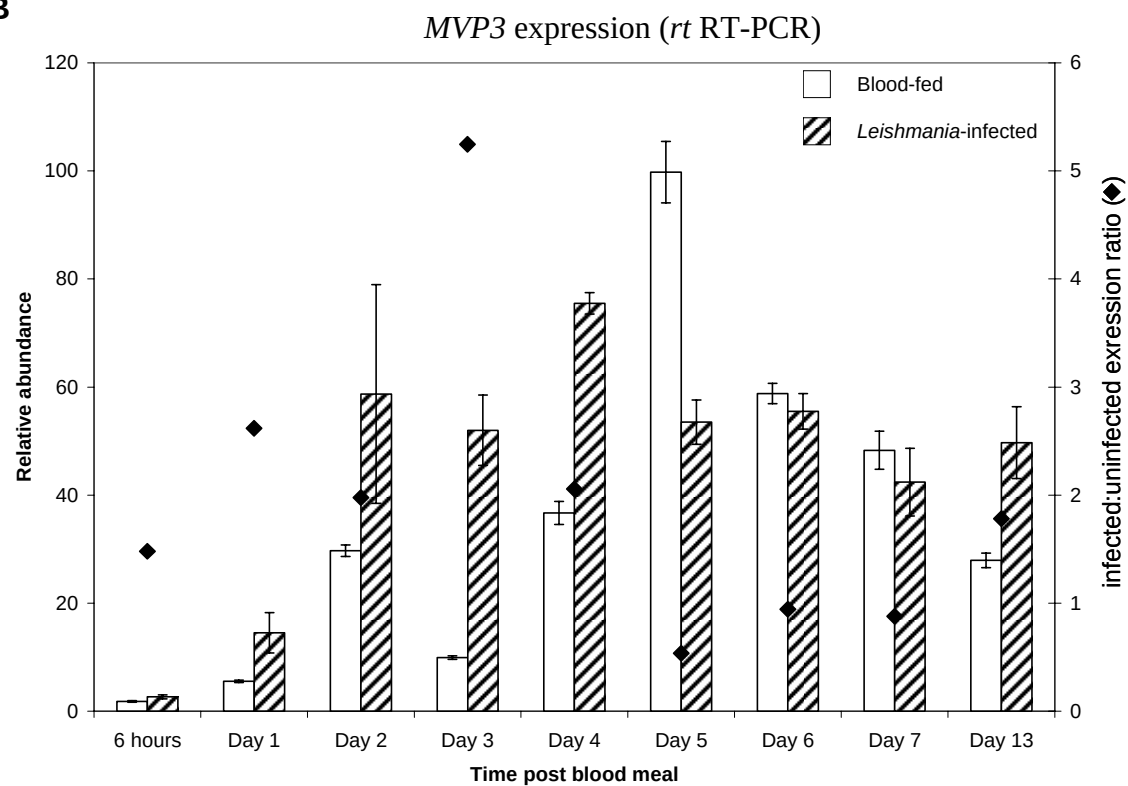
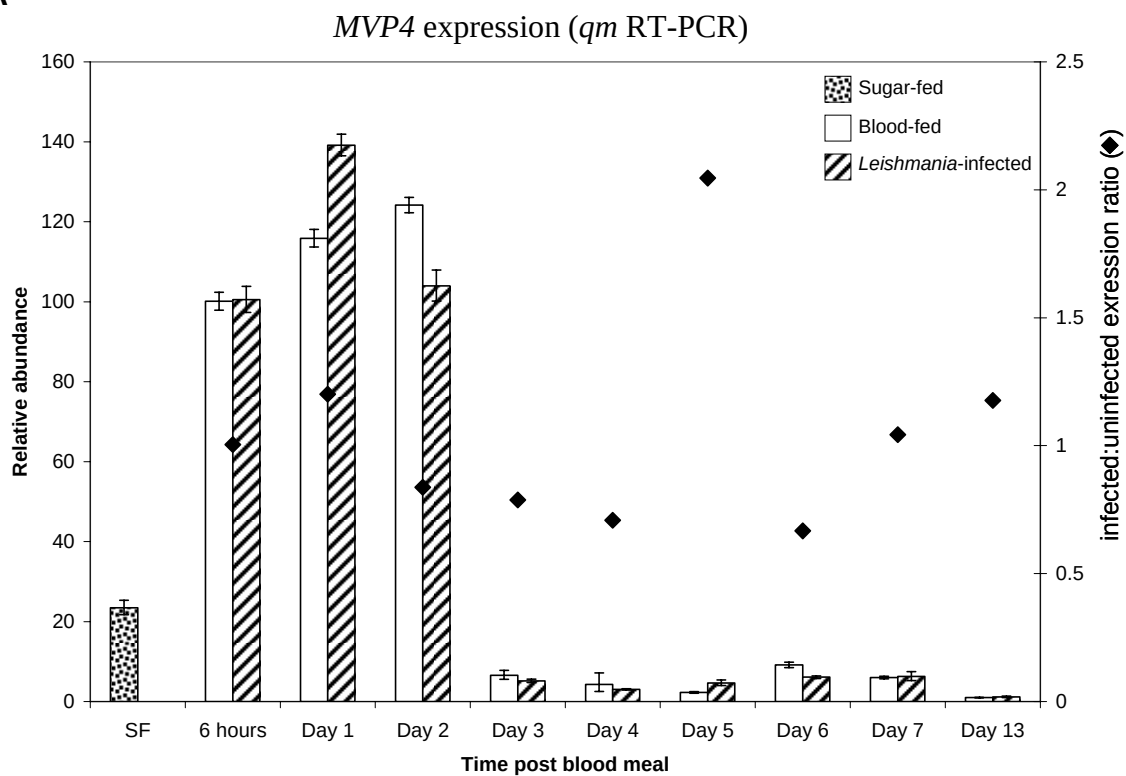
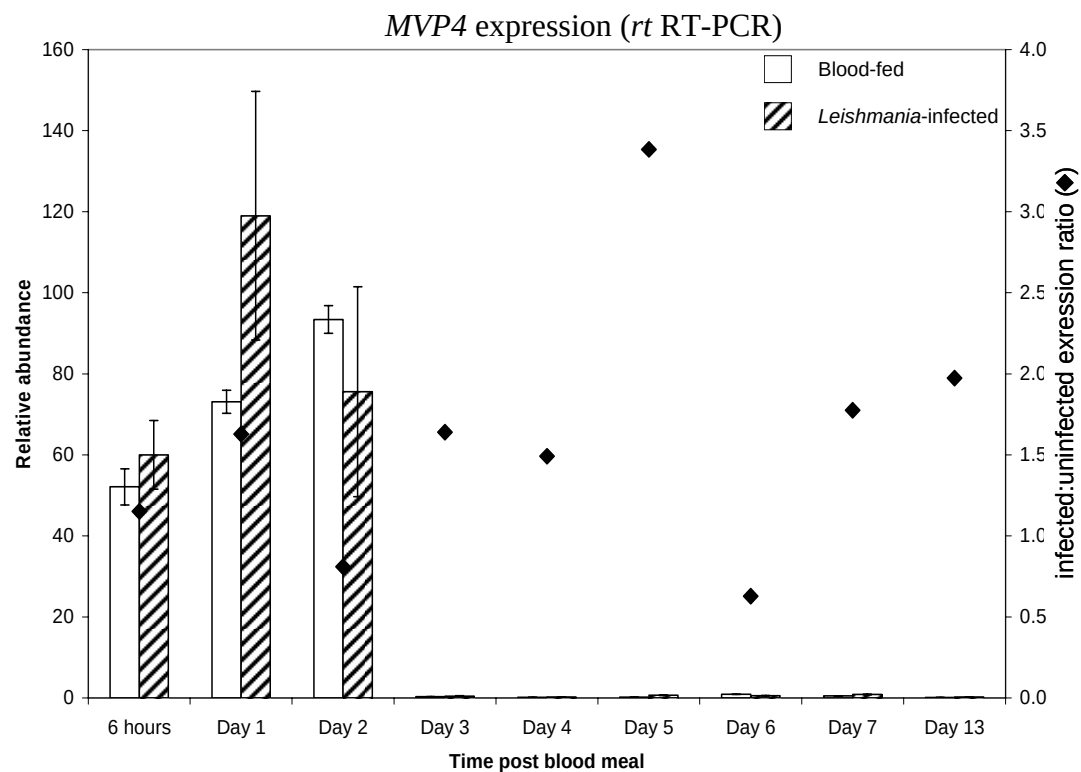
A**B**

Figure 35. Comparative abundance of microvillar protein *MVP4* in uninfected and *L. infantum chagasi*-infected *Lu. longipalpis* midguts.

Transcript abundance measured by (A) quantitative multiplex RT-PCR and (B) real-time RT-PCR. Bars indicate *MVP4* relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, in sugar-fed, blood-fed (white bars) or blood-fed infected (striped bars) sand flies. Diamonds indicate the *MVP4* expression ratio of infected to uninfected sand flies. Error bars represent 95% confidence intervals.

A**B**

homology with *Blattella germanica* and *Periplaneta americana* microvilli proteins, sand fly microvillar proteins contain a number of conserved insect allergen repeats [6, 7]. The function of microvillar proteins is currently unknown; however, the impact of blood-feeding and the presence of *L. infantum chagasi* within the midgut on MVP expression levels warrant further research on the importance of these molecules.

***L. infantum chagasi* accelerates and upregulates peritrophin transcription.**

Two putative peritrophin molecules were assayed for temporal abundance, *Per1* and *Per2*. One and two days after blood ingestion *Per1* was abundant, and at all other time points analyzed, the transcript was below the detectable limit of the quantitative multiplex RT-PCR (Figure 36A). In response to the presence of *L. infantum chagasi*, *Per1* transcript increases 1.7 or 3.6 fold one day PBM as measured by multiplex or *rt* RT-PCR, respectively, and is below levels normally found in the uninfected *Lu. longipalpis* two days after blood-feeding. Due to the sensitivity of *rt* RT-PCR, the minute amounts of *Per1* detected from day 4 onward after blood-feeding showed an average of 1.8 fold increase in transcription when infected with *L. infantum chagasi*; whereas, the multiplex was unable to detect the transcript (Figure 36). An additional molecule identified as a putative peritrophin, *Per2*, also demonstrates an induction by blood ingestion and decreases after blood digestion (Figure 37). *Per2* is found earlier in the midgut tissue than *Per1*, reaching peak levels at both six hours and two days after feeding with low basal levels maintained at times when the midgut contains no blood. There is a subtle decrease in *Per2* transcript abundance in blood-fed infected sand flies at both two and

Figure 36. Comparative abundance of peritrophin *Per1* in uninfected and *L. infantum* *chagasi*-infected *Lu. longipalpis* midguts.

Transcript abundance measured by (A) quantitative multiplex RT-PCR and (B) real-time RT-PCR. Bars indicate *Per1* relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, in sugar-fed, blood-fed (white bars) or blood-fed infected (striped bars) sand flies. Diamonds indicate the *Per1* expression ratio of infected to uninfected sand flies. Error bars represent 95% confidence intervals.

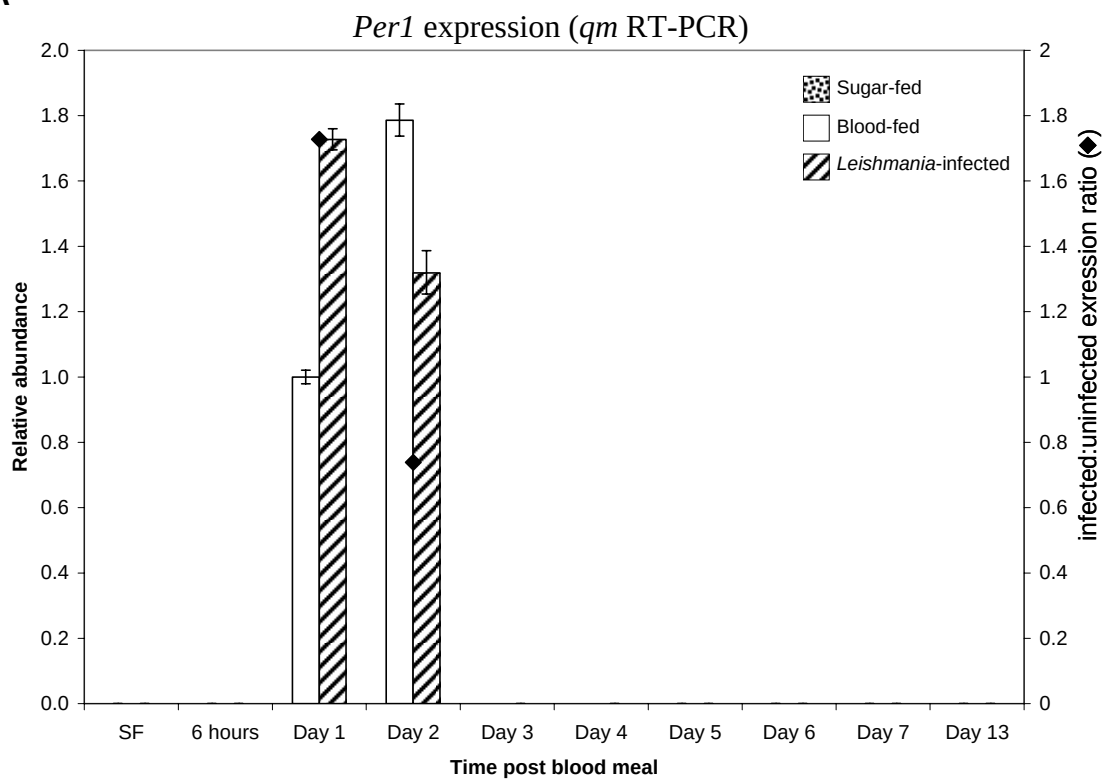
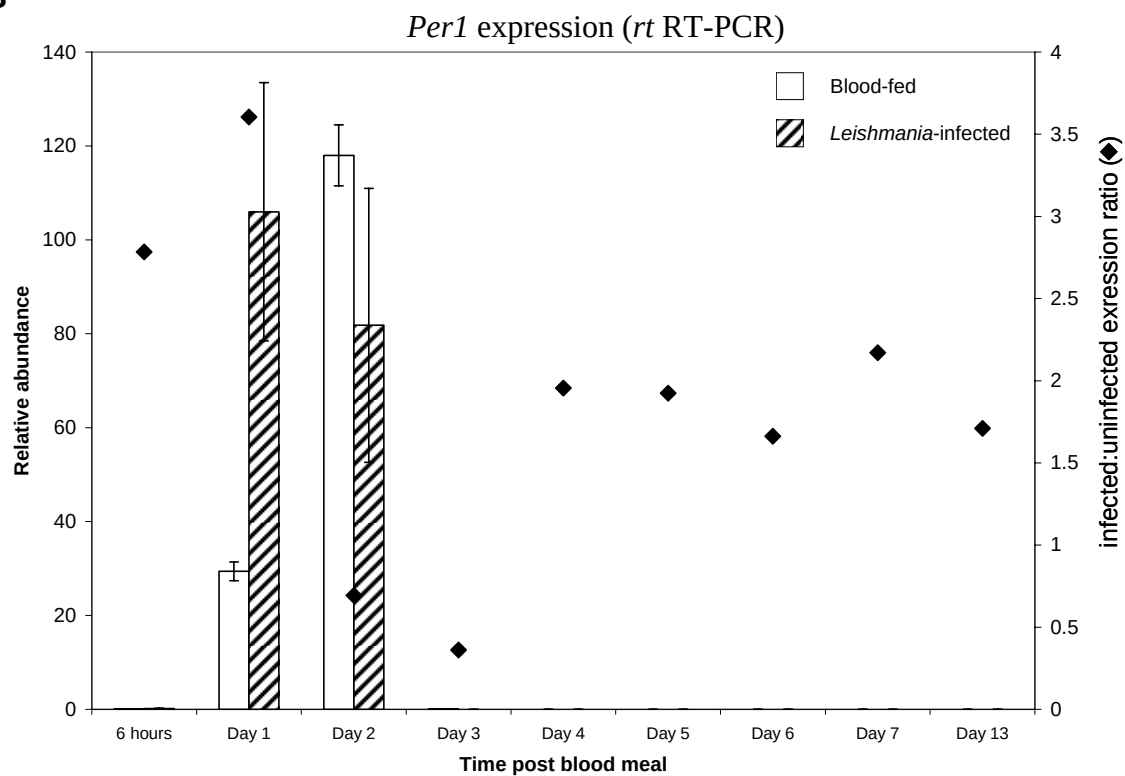
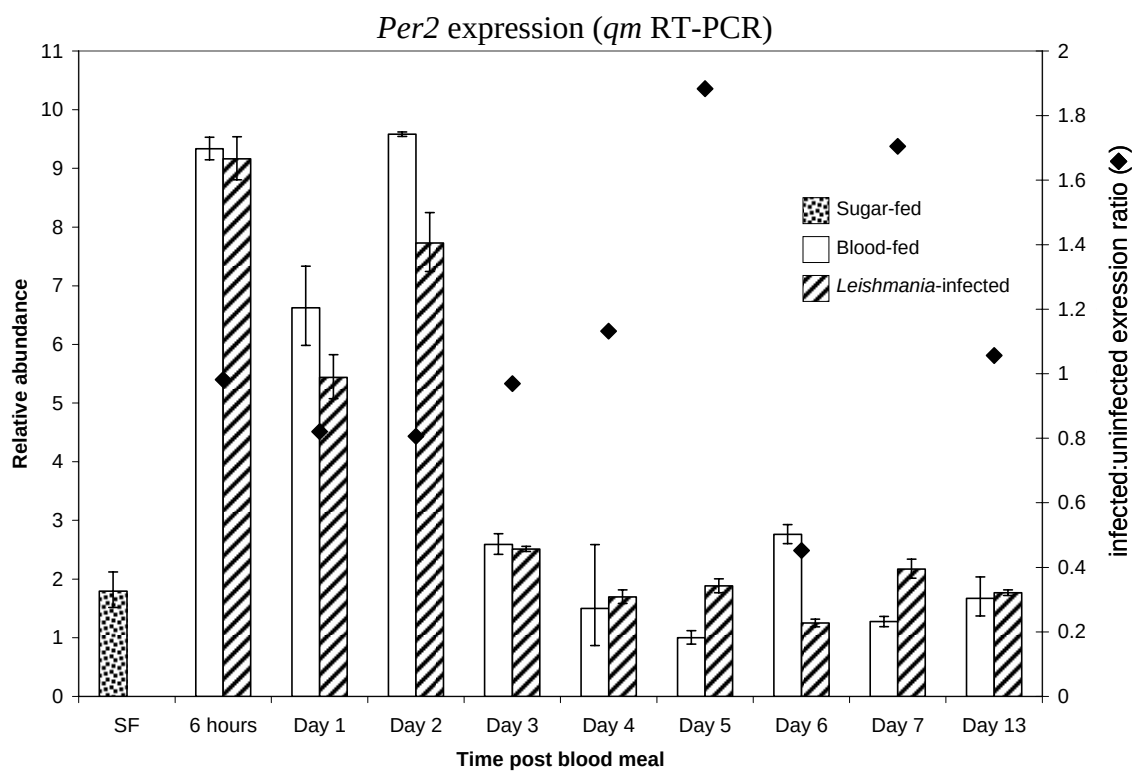
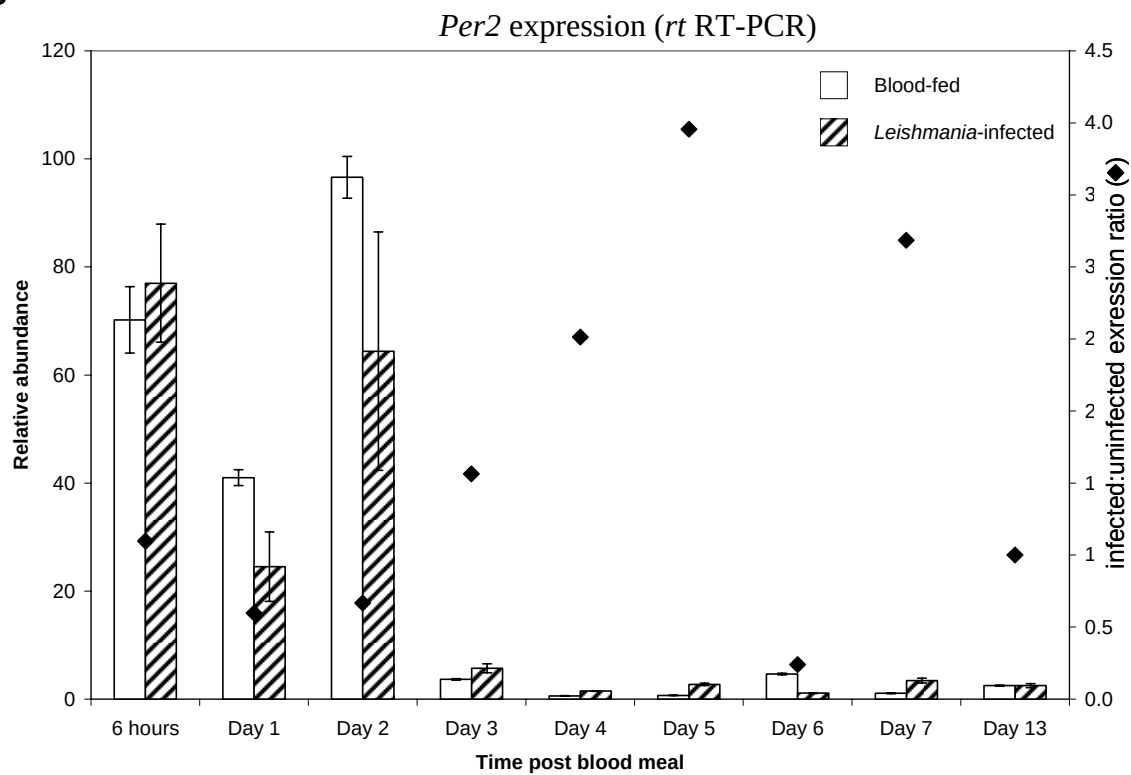
A**B**

Figure 37. Comparative abundance of peritrophin *Per2* in uninfected and *L. infantum* *chagasi*-infected *Lu. longipalpis* midguts.

Transcript abundance measured by (A) quantitative multiplex RT-PCR and (B) real-time RT-PCR. Bars indicate *Per2* relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, in sugar-fed, blood-fed (white bars) or blood-fed infected (striped bars) sand flies. Diamonds indicate the *Per2* expression ratio of infected to uninfected sand flies. Error bars represent 95% confidence intervals.

A**B**

three days PBM, and then a significant increase is notable after blood digestion, particularly at Days 5 and 7 (Figure 37). *Per1* contains four possible chitin-binding domains, presumably integral to the cross-linking of the chitin fibrils of the peritrophic matrix. It is clear that *Per1* is highly regulated with response to both induction and repression in response to blood-feeding and digestion, which is to be expected of a molecule involved in the temporary existence of a specialized foundation, the peritrophic matrix. *Per2* is unique in that it contains only one potential chitin-binding domain and thus cannot act in the cross linking of chitin fibrils. It is possible that *Per2* and other single domain chitin-binding molecules may act to cap the end of chitin fibrils, sequester free chitinous molecules, or regulate fungal pathogens as innate immune molecules within the midgut. It appears that the presence of *L. infantum chagasi* promotes the earlier formation of the peritrophic matrix in an effort to restrict the locality of the parasite within the blood bolus. Prior research indicates that the peritrophic matrix plays a vital role in protecting *Leishmania* parasites from midgut proteases as they differentiate from amastigotes to promastigotes [8]. Coincidentally, the parasites must also escape the peritrophic membrane prior to the passing of the blood meal for the continuation of infection [8]. A hypothesis explaining the early formation and subsequent early dissociation of the peritrophic matrix is that *Leishmania* requires quick shelter of transitional-stage parasites from proteases and then escape from the peritrophic matrix to facilitate attachment to the midgut epithelium.

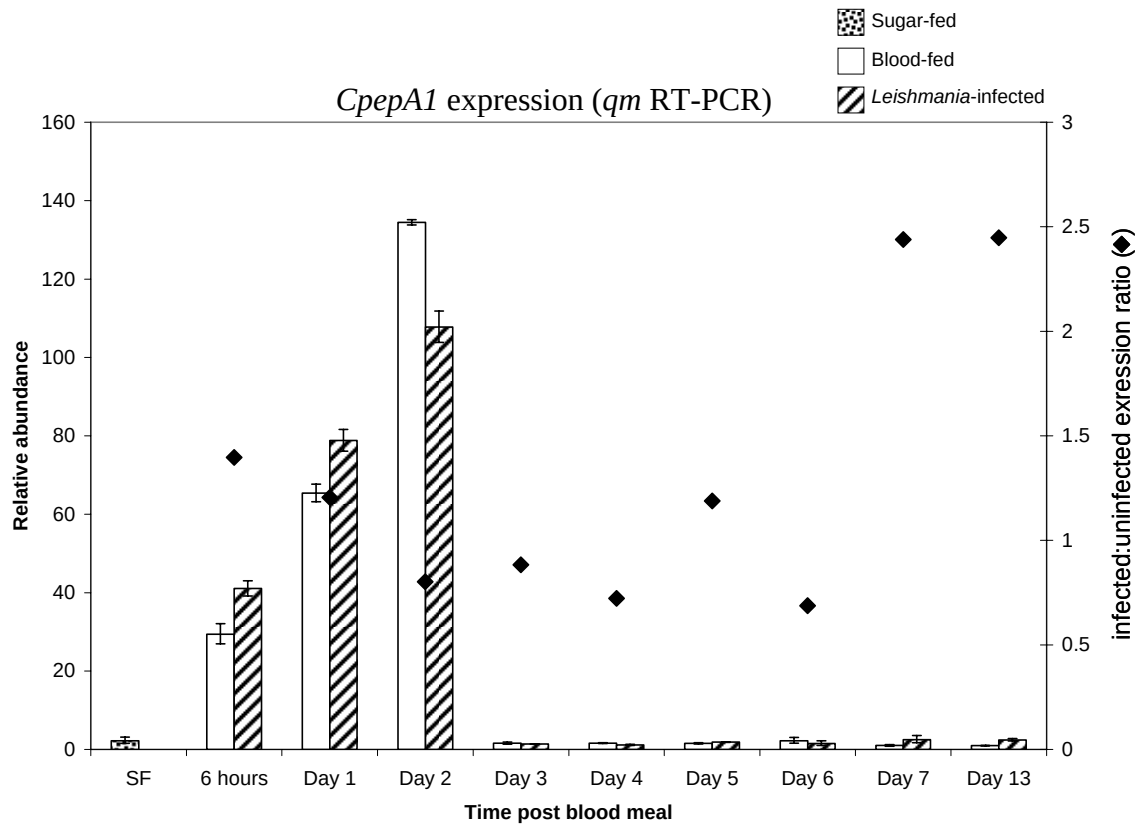


Figure 38. Comparative abundance of carboxypeptidase *CpepA1* in uninfected and *L. infantum chagasi*-infected *Lu. longipalpis* midguts.

Bars indicate *CpepA1* relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, in sugar-fed, blood-fed (white bars) or blood-fed infected (striped bars) sand flies. Diamonds indicate the *CpepA1* expression ratio of infected to uninfected sand flies. Error bars represent 95% confidence intervals.

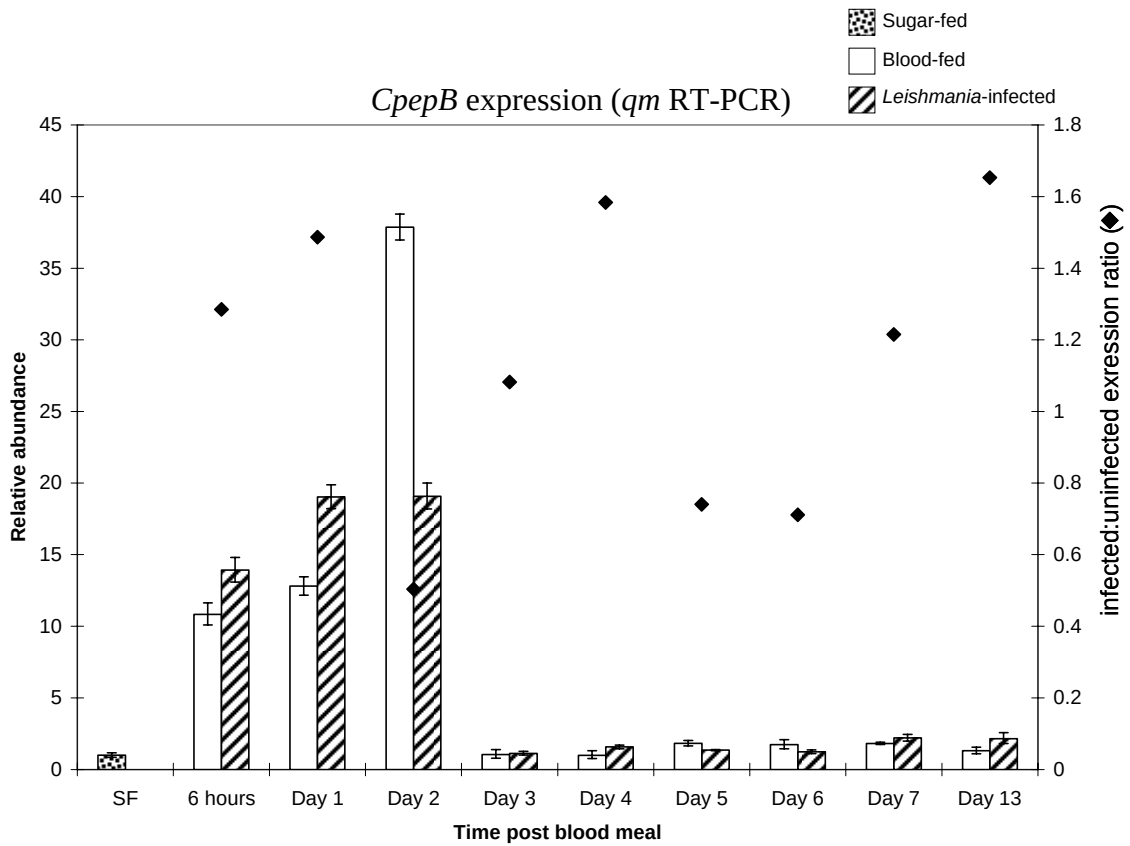


Figure 39. Comparative abundance of carboxypeptidase *CpepB* in uninfected and *L. infantum chagasi*-infected *Lu. longipalpis* midguts.

Bars indicate *CpepB* relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, in sugar-fed, blood-fed (white bars) or blood-fed infected (striped bars) sand flies. Diamonds indicate the *CpepB* expression ratio of infected to uninfected sand flies. Error bars represent 95% confidence intervals.

Carboxypeptidases are increased in response to *L. infantum chagasi* colonization.

We recently described several carboxypeptidases in the transcriptome of the *Lu. longipalpis* midgut in blood-fed sand flies [5]. This temporal analysis of *CpepA1* and *CpepB* carboxypeptidase transcripts confirmed the correlation of *Cpep* expression with blood-feeding as observed in the cDNA library comparative analysis. Both carboxypeptidases are scarce in sugar-fed sand flies but are induced as early as six hours PBM and peak in abundance two days after a blood meal is taken (Figures 38 and 39). There is a striking reduction in *Cpep* transcripts three days PBM and expression remains at a low basal level, likely until a subsequent blood meal is taken. The similarities in the expression profiles of these carboxypeptidases are observed in the *L. infantum chagasi* infected sand flies as well. There is an increase in transcript abundance between six hours and one day PBM with expression peaking at day two. The presence of *Leishmania* parasites within the midgut appears to initiate an increased amount of these digestive proteases early in blood digestion. The decrease in transcripts two days PBM could be an effort of the parasite to regulate the amount of proteases present within the midgut lumen as a protective measure. However, it is also possible that the parasite induces increased transcript production early in blood digestion and the increased transcript or protein product acts as a negative feedback on translation of these carboxypeptidases. Strikingly, late in *L. infantum chagasi* colonization of the sand fly, there is a significant increase in *CpepA1* transcripts – this can be seen at seven and 13 days post infection. While increases in digestive proteases are likely a result of the presence of parasite protein in the lumen of the midgut it is possible that protease

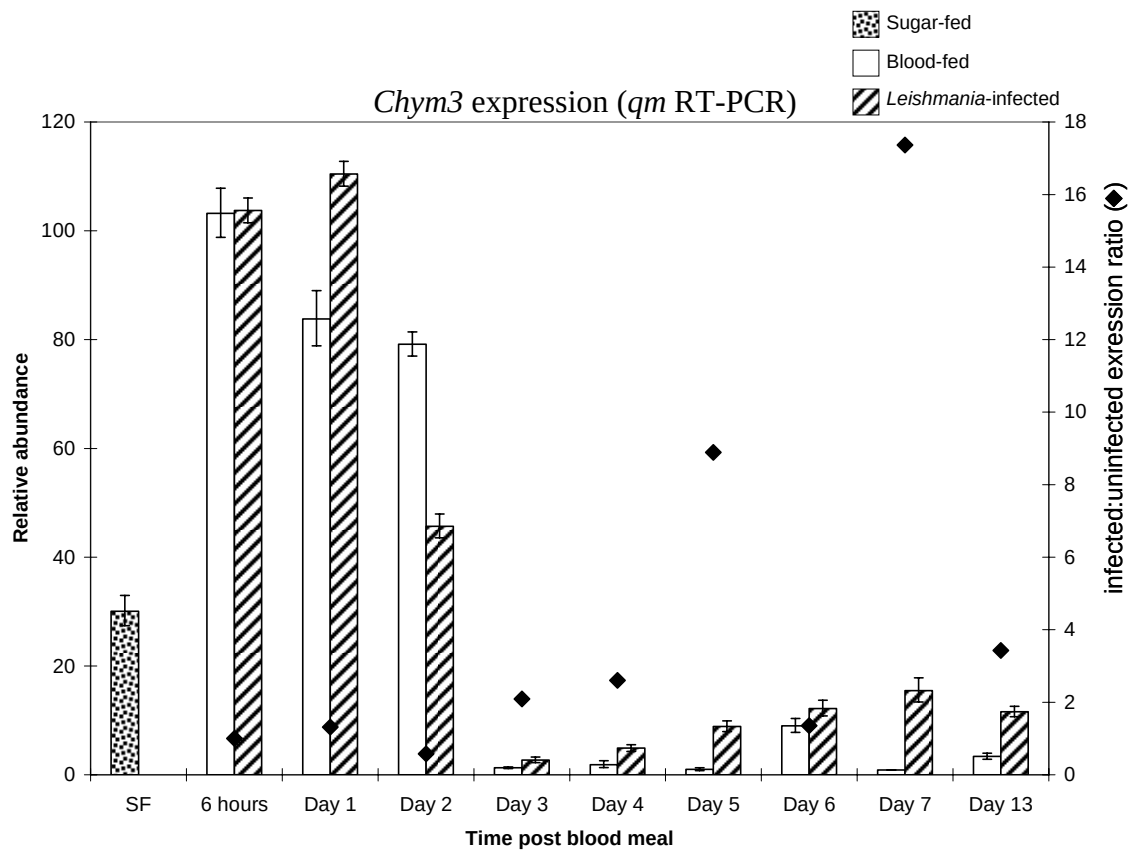


Figure 40. Comparative abundance of chymotrypsin, *Chym3*, in uninfected and *L. infantum chagasi*-infected *Lu. longipalpis* midguts.

Bars indicate *Chym3* relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, in sugar-fed, blood-fed (white bars) or blood-fed infected (striped bars) sand flies. Diamonds indicate the *Chym3* expression ratio of infected to uninfected sand flies. Error bars represent 95% confidence intervals.

induction is a host defense mechanism of the sand fly. Subtle increases in *CpepB* expression also are seen late in *L. infantum chagasi* colonization.

Late-stage *L. infantum chagasi* infection increases chymotrypsin abundance.

A chymotrypsin analyzed, *Chym3*, represents another blood-feeding-induced digestive protease. There is little impact by *L. infantum chagasi* on *Chym3* expression immediately following feeding but a two-fold decrease is observed two days PBM with a subsequent increase in *Chym3* expression reaching 8- and 17-fold increases in transcript abundance five and seven days after infection, respectively (Figure 40). Thirteen days after ingesting *L. infantum chagasi* the sand fly is producing 3.4 times the normal amount of *Chym3*. While *Chym3* is most abundant in the same time frame as the carboxypeptidases there is a variation in the profile of expression; most abundant six hours after feeding and decreasing until there is a drop to low basal level three days PBM ingestion. The variation in expression profiles may be due to separate transcriptional regulation such that carboxypeptidases are induced by the presence of blood within the midgut while *Chym3* is induced by the act of feeding or probing. If the increase in *Chym3* transcripts correlates with a large increase in chymotrypsin activity in the midgut of sand flies with mature transmissible *L. infantum chagasi*, then there is the possibility of midgut enzymes exerting some effect on disease and transmission once the parasite is regurgitated into the host tissue.

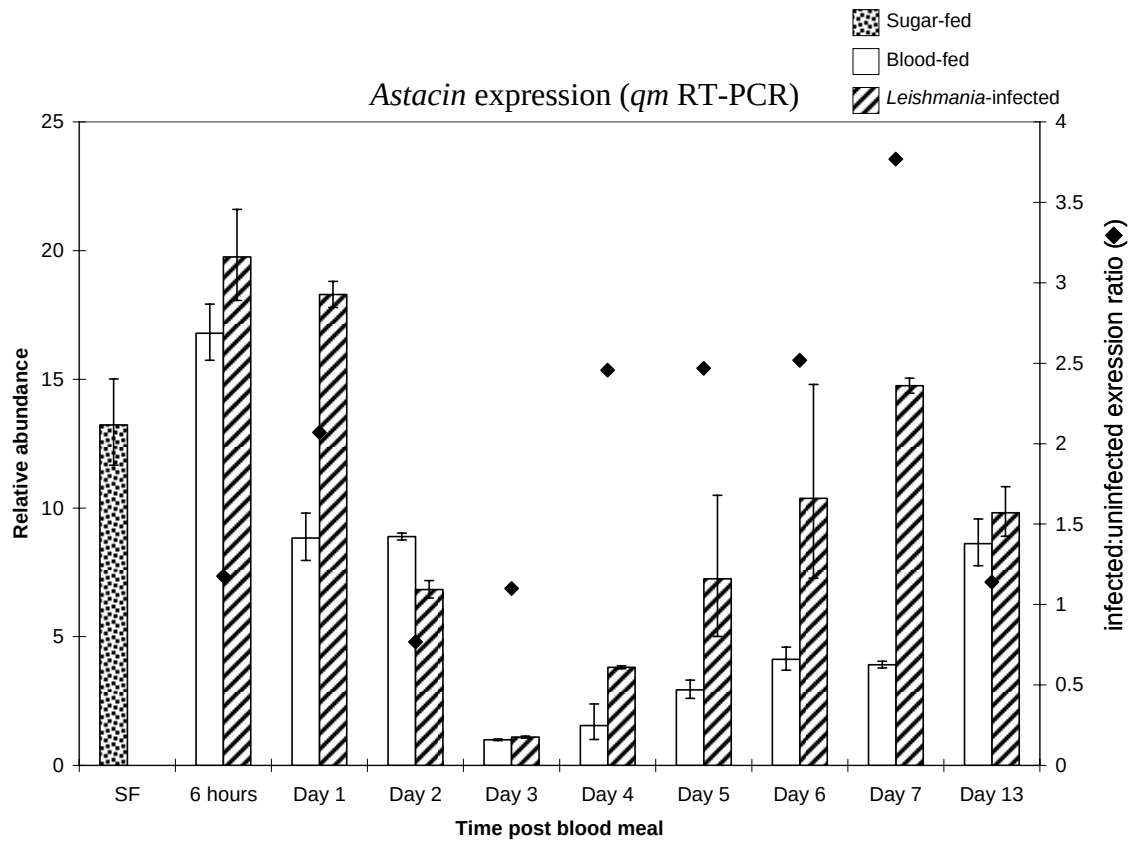


Figure 41. Comparative abundance of *astacin* in uninfected and *L. infantum chagasi*-infected *Lu. longipalpis* midguts.

Bars indicate *astacin* relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, in sugar-fed, blood-fed (white bars) or blood-fed infected (striped bars) sand flies. Diamonds indicate the *astacin* expression ratio of infected to uninfected sand flies. Error bars represent 95% confidence intervals.

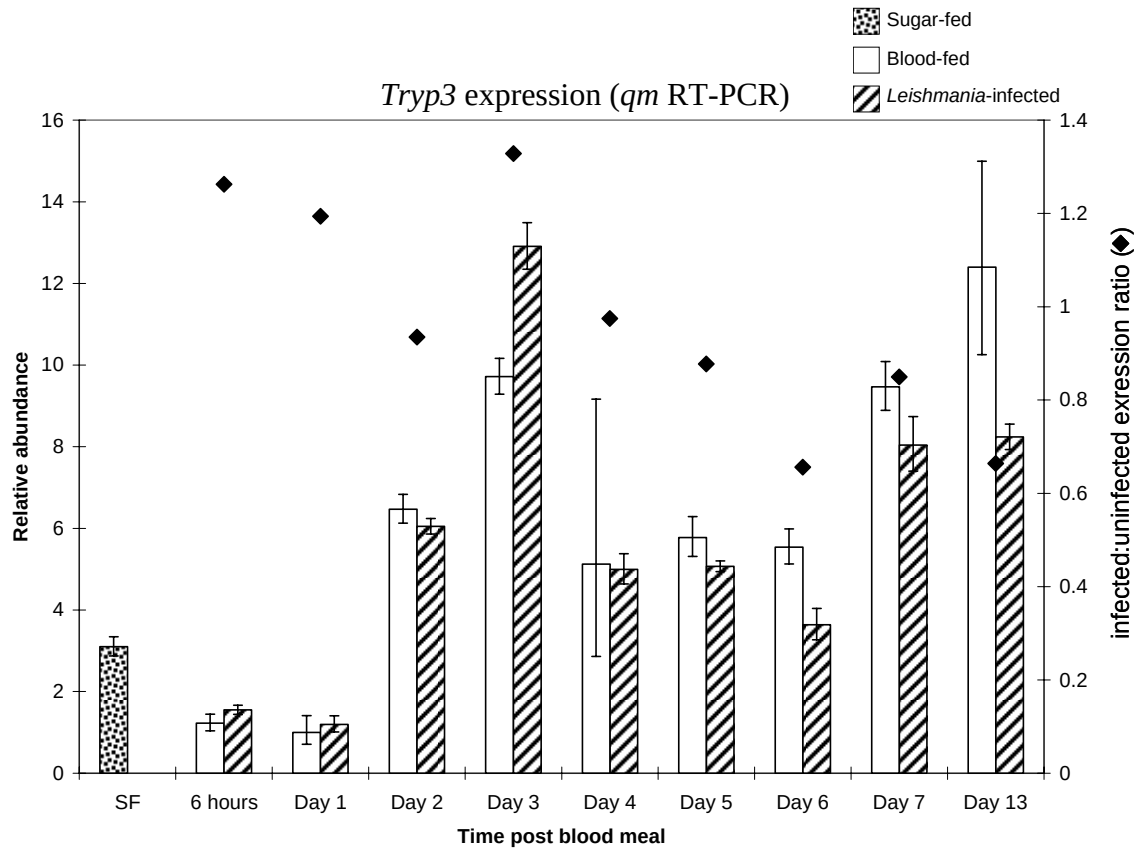


Figure 42. Comparative abundance of trypsin *Tryp3* in uninfected and *L. infantum chagasi*-infected *Lu. longipalpis* midguts.

Bars indicate *Tryp3* relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, in sugar-fed, blood-fed (white bars) or blood-fed infected (striped bars) sand flies. Diamonds indicate the *Tryp3* expression ratio of infected to uninfected sand flies. Error bars represent 95% confidence intervals.

Astacin transcripts increase but trypsin is relatively unaffected by *L. infantum chagasi*.

The expression profiles of two other digestive proteases, *Astacin* and *Tryp3*, were also investigated and differed in several ways to the previously described proteases. *Astacin* is abundant in sugar-fed sand flies, correlating with sequence abundance in the transcriptome, and shows a decrease as the blood meal is digested, reaching the lowest levels detected at day three (Table 28 and Figure 41). After the blood meal is passed *Astacin* begins increasing and by 13 days after the last blood meal the transcripts are almost as abundant as found in sugar-fed sand flies. *Leishmania infantum chagasi* appears to stimulate an overall induction of *Astacin*, with two-fold increases seen for several days in a row. Dipteran astacin molecules are most likely digestive enzymes, and there have been no previous reports of astacin proteases acting as antimicrobials; however, this expression profiling convincingly demonstrates an affect by *L. infantum chagasi* on *Astacin* abundance. *Tryp3* shares homology with molecules in a class of insect trypsins that has been referred to previously as constitutive or early trypsins; having high transcript abundance prior to blood-feeding and then reducing as the blood meal is digested [9]. Blood-feeding causes a reduction in transcript abundance during the early stages of digestion 6-24 hours after feeding (Figure 42). Surprisingly, in uninfected flies, transcript abundance increases at two days and has two peaks of high abundance at Days 3 and 7 PBM, with the highest *Tryp3* abundance 13 days after feeding. There is very little effect on *Tryp3* transcripts in *L. infantum chagasi* infected sand flies (Figure 42). While statistically significant differences occur at Days 3, 6 and 13, the fold change in expression may result in a negligible change in protein expression. The induction of

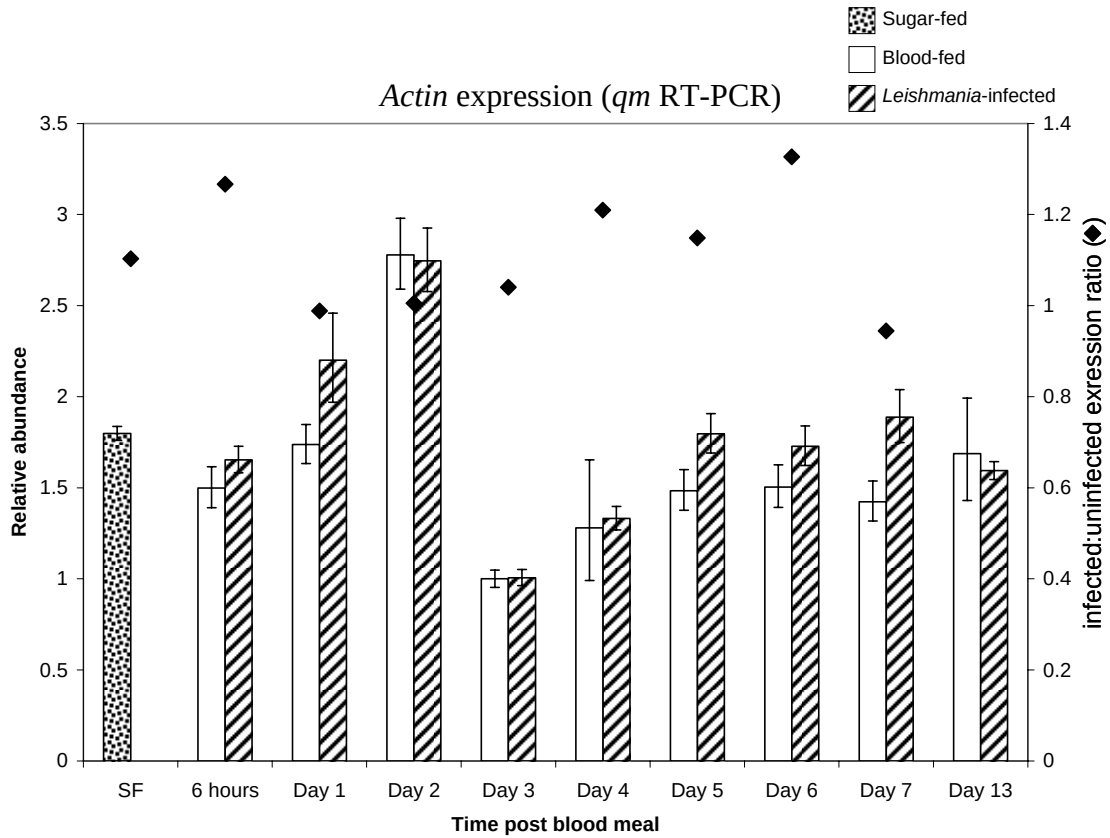


Figure 43. Comparative abundance of *actin* in uninfected and *L. infantum chagasi*-infected *Lu. longipalpis* midguts.

Bars indicate *actin* relative transcript abundance, normalized to *Lu. longipalpis* S7 rRNA, in sugar-fed, blood-fed (white bars) or blood-fed infected (striped bars) sand flies. Diamonds indicate the *actin* expression ratio of infected to uninfected sand flies. Error bars represent 95% confidence intervals.

Tryp3 as early as two days after blood-feeding is not seen in the transcriptional analysis of trypsin in the midgut of *Aedes aegypti* or *P. papatasi*, and additionally, this early induction is not observed in the two characterized *Lu. longipalpis* trypsins, *Lltryp1* and *Lltryp2* [9-11]. Previous studies investigating the temporal expression levels of Dipteran trypsins utilized semi-quantitative PCR; whereas, the quantitative multiplex used here detected oscillations in the expression of what might be considered a “constitutive” trypsin [9-11].

Blood meal digestion induces actin transcription.

Actin transcript abundance was assessed as a result of the significant reduction in sequences captured in the *L. infantum chagasi*-infected midgut cDNA libraries (Table 28). Surprisingly, a statistically significant increase in transcript abundance is found one, five, and seven days after blood-feeding when comparing uninfected and *L. infantum chagasi*-infected sand fly midguts when using *qm* RT-PCR (Figure 43). However, the ratio of actin transcript abundance in the infected compared with uninfected midguts is minute (Figure 43). The discrepancy between cDNA library sequence abundance and transcript quantity measured by RT-PCR demonstrates the importance in evaluating the expression profile of each transcript identified as differentially expressed by the sequencing of individual cDNA libraries. Actin transcripts increase after blood-feeding, peaking at two days after ingestion, then drop to the lowest measured level three days post-blood meal. Once the blood meal is excreted, actin transcription appears to be maintained at a steady basal level. Blood-induced transcription of actin has been noted in

the mosquito *Anopheles gambiae*, and this induction correlates with cytoskeletal and cell morphological changes in the midgut epithelium [12].

Conclusion

The establishment and propagation of a transmissible *Leishmania* infection within the sand fly vector occurs within the lumen of the midgut. In comparison, *Plasmodium* ookinetes transit through the midgut epithelium of mosquitoes one day after blood-feeding. Previous work identified gene induction during *Plasmodium* invasion of epithelial cells of mosquito midgut tissue [13]. *Leishmania* parasites can persist in the midgut of a competent sand fly host for the remaining life span of the vector. Although *Leishmania* do not actively invade midgut epithelial cells, they must adhere to the cell surface to prevent being excreted with the digested blood meal. During *Leishmania* adhesion to the midgut epithelium, there is likely responsive cellular signaling occurring. The abundance of parasites and extracellular *Leishmania* molecules within in the lumen also impacts the type and quantity of molecules transcribed by the midgut tissue. We wished to further understand the impact of *L. infantum chagasi* colonization on *Lu. longipalpis* sand flies by evaluating a number of transcripts important in peritrophic matrix formation and digestion.

Artificial infection of the sand flies was accomplished using a mouse macrophage cell line infected with *L. infantum chagasi* amastigotes in an attempt to replicate the natural presentation of the parasite to the midgut environment. This method resulted in all of the dissected flies colonized with an abundance of metacyclic promastigotes 13 days after blood-feeding; equating to a transmissible infection as observed in other

studies. This work vindicates the modulation of *Lu. longipalpis* midgut transcript abundance of specific molecules by *L. infantum chagasi* as was first reported in the differential comparison of cDNA library sequence abundance [5]. More specifically, the presence of *L. infantum chagasi* within the midgut of *Lu. longipalpis* caused notable induction or repression of transcripts necessary for peritrophic matrix formation and blood meal digestion. The manipulation of the sand fly midgut transcriptome by *Leishmania* parasites may serve to benefit the parasites survival and necessary differentiation to the final, infective metacyclic form. Additionally, the presence of the parasite within the midgut of the sand fly may trigger protective responses, such as increased protease expression, eluding that the *Leishmania* colonization of the sand fly is neither mutualistic nor commensal. There are significant increases in protease transcription at time points associated with transmission. We hypothesize that upregulated proteases could be present in the regurgitated parasite milieu and attribute to enhanced transmission. Included in this work is the first reported molecular-based quantitation of *Leishmania* parasites as they develop within the midgut of the sand fly.

Methods

Sand flies

Lutzomyia longipalpis sand flies (Jacobina strain) were maintained at the Laboratory of Malaria and Vector Research at the National Institute of Allergy and Infectious Diseases. Three- to four-day post eclosion sand flies were allowed a 20% sucrose solution (sugar-fed/unfed) or fed on the blood of anesthetized BALB/c mice (blood-fed).

***Leishmania* and sand fly infection**

Lu. longipalpis sand flies were infected by an artificial blood meal containing *Leishmania infantum chagasi*-infected macrophages (blood-fed, infected) as reported previously with minor modifications [5]. In brief, *L. infantum chagasi* MHOM/BROO/MER/Strain 2 promastigote cultures were maintained in M199 (Sigma-Aldrich, St. Louis, MO) containing 20% (V/V) fetal bovine serum(FBS) (Invitrogen, Carlsbad, CA) and 100 units/ml Penicillin, 100 µg/ml Streptomycin, and 0.292 mg/ml Glutamine (PSG) (Invitrogen, Carlsbad, CA) at 25°C. Macrophage cell line J774A.1 (American Type Culture Collection, Manassas, VA) was cultured in RPMI (Invitrogen, Carlsbad, CA) containing 10% FBS and PSG at 37.0°C, 95% air, 5% CO₂. At confluency, the macrophages were scraped from the culture flask and washed twice by centrifugation in phosphate buffered saline (PBS) at 380 x g for 10 minutes before resuspension in culture media. The washed macrophages were then placed in 2 wells of a 24-well culture plate, 2×10^6 cells per well, and allowed to adhere for 1 hour at 37°C, 5% CO₂. Stationary-phase *L. infantum chagasi* culture was washed by centrifugation in PBS at 1200 x g for 15 minutes and resuspended in macrophage culture media. Nonadherent macrophages were removed by the replacement of the culture media and *Leishmania* parasites in macrophage added at a 5:1 ratio of parasite to macrophage. The parasites were co-cultured with the macrophages for 5 hours at 26°C. The culture was then washed to remove extracellular parasites and the macrophages scraped from the wells. Macrophages were confirmed to contain intracellular amastigotes by staining with QUICK III (Astral Diagnostics, Inc., West Deptford, NJ) according to the manufacture's

protocol and visualized by light microscopy. The infected macrophage culture was centrifuged at 380 x g for 10 minutes and resuspended in 500µl fresh whole mouse blood collected in heparin. The blood containing amastigote-infected macrophages was used for artificial blood-feeding of sand flies as described [14].

RNA extraction

Midgut tissue from female *Lu. longipalpis* was dissected at 6 hours and 1, 2, 3, 4, 5, 6, 7, and 13 days after artificial blood-feeding with or without the presence of *L. infantum chagasi*-infected macrophages. For each time point, ten midguts were placed in 20 µl of RNAlater (Sigma-Aldrich, St. Louis, MO) and stored at 4°C prior to RNA extraction. Total RNA was extracted using Agencourt RNAdvance Tissue RNA extraction kit (Beckman, Beverly, MA) according to the provided instructions. To prepare the tissues for RNA extraction, 200 µl of tissue homogenization buffer was added and the tissues homogenized using a Kontes pestle (Fischer, Itasca, IL) and another 200 µl of homogenization buffer was added prior to incubation. The optional DNase procedure was performed according to instructions using RNase-free DNase (New England Biolabs, Ipswich, MA). RNA was washed from the magnetic beads using molecular grade water and samples stored at -70 °C prior to dilution and experimental use.

Multiplex RT-PCR

Primers and genes used in the multiplex reaction are described in Table 29. Each forward and reverse primer was designed using GeXP Express Profiler, Primer Design

software (Beckman, Fullerton, CA) to produce PCR products with lengths from 100 to 300 bp in length and 4 to 7 bp apart. Additionally, the primers used in the multiplex reaction are chimeric with universal primer sequences of 18 and 19 nucleotides incorporated onto the forward and reverse primer, respectively. Transcript expression profiles were assessed using the GenomeLab GeXP Analysis System Multiplex RT-PCR assay (Beckman). The GenomeLab GeXP Start Kit (Beckman) was used for the reverse transcription reaction, 3 μ l DNase/RNase free H₂O, 4 μ l RT buffer, 5 μ l Kan^r RNA with RI, 1 μ l reverse transcriptase, 5 μ l sample RNA (2ng/ μ l), and 2 μ l reverse primer plex (concentration noted in Table 29). For RT-minus and no-template control reactions DNase/RNase free H₂O was used as a substitute. The reverse transcription reactions were incubated at 48°C for 1 minute, 37°C for 5 minutes, 42°C for 60 minutes, 95°C for 5 minutes and stored at 4°C. The PCR amplification of each sample included 4 μ l PCR buffer (contains the universal primers), 4 μ l 25 mM MgCl₂ (Abgene, Rockford, IL), 2 μ l forward primer plex (200 nM of each primer), 0.7 μ l DNA polymerase (Abgene) and 9.3 μ l cDNA from the RT-PCR reaction. The PCR samples were incubated at 95°C for 10 minutes and 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 68°C for 1 minute and then stored at 4°C. Sample PCR products were pre-diluted 1:20 using 10 mM Tris-HCl, pH 8.0, while no-template and RT-minus reactions were not pre-diluted. All samples were then diluted by the addition of 1 μ l sample and 0.5 μ l DNA Size Standard-400 into 38.5 μ l Sample Loading Solution, mixed by pipetting, and covered by 1 drop of mineral oil. The samples were placed in the GenomeLab GeXP Genetic Analysis System for capillary electrophoresis and fragment size analysis. Fragment results were analyzed using the eXpress Analysis software of the GeXP Genetic Analysis System.

Real-time RT-PCR

Primer sequences used in real-time RT-PCR analysis are identical to those used in the multiplex quantitative RT-PCR and are shown in Table 29. cDNA synthesis was accomplished using Transcriptor First Strand cDNA Synthesis Kit (Roche, Indianapolis, IN) and the included anchored-oligo(dT)₁₈ primer according to the manufacturer's instructions using 12 µl sample RNA (2ng/µl). PCR was performed using Lightcycler[®] 480 SYBR Green I Master (Roche) according to the manufacturer's protocol and using 4 µl of cDNA. Real-time quantitation of transcript abundance was performed using the Lightcycler[®] 480 Instrument under the following conditions: 95°C for 5 minutes, 45 cycles of 95°C for 10 seconds, 55°C for 10 seconds, 72°C for 20 seconds. Fluorescence was acquired after amplification at the end of each cycle.

Data analysis

The Kan^r RNA served as an internal control for all multiplex reactions, and the *Lu. longipalpis* 40S ribosomal protein S7 housekeeping gene served as a control for normalization. Transcript expression levels analyzed by multiplex RT-PCR were normalized by dividing the peak area result of each gene by the peak area result of the housekeeping gene and then log₂-transformed. Relative abundance of transcripts quantitated using real-time PCR were analyzed by the log₂-back-transformation of the C_t ratio of the sample amplicon to the control amplicon and reported as the fold increase in sample transcript over that of the control transcript. Three multiplex assays or real-time reactions were completed for each experiment and significant differences were identified

between *L. infantum chagasi*-infected and non-infected *Lu. longipalpis* midgut tissue samples using Student's t-test and reporting p-values below 0.05 as significant.

References

1. Borja-Cabrera GP, Cruz Mendes A, Paraguai de Souza E, Hashimoto Okada LY, de ATFA, Kawasaki JK, Costa AC, Reis AB, Genaro O, Batista LM *et al*: **Effective immunotherapy against canine visceral leishmaniasis with the FML-vaccine.** *Vaccine* 2004, **22**(17-18):2234-2243.
2. Saraiva EM, de Figueiredo Barbosa A, Santos FN, Borja-Cabrera GP, Nico D, Souza LO, de Oliveira Mendes-Aguiar C, de Souza EP, Fampa P, Parra LE *et al*: **The FML-vaccine (Leishmune) against canine visceral leishmaniasis: a transmission blocking vaccine.** *Vaccine* 2006, **24**(13):2423-2431.
3. Ramalho-Ortigao M, Jochim RC, Anderson JM, Lawyer PG, Pham VM, Kamhawi S, Valenzuela JG: **Exploring the midgut transcriptome of *Phlebotomus papatasi*: comparative analysis of expression profiles of sugar-fed, blood-fed and *Leishmania major*-infected sandflies.** *BMC Genomics* 2007, **8**:300.
4. Kamhawi S, Ramalho-Ortigao M, Pham VM, Kumar S, Lawyer PG, Turco SJ, Barillas-Mury C, Sacks DL, Valenzuela JG: **A role for insect galectins in parasite survival.** *Cell* 2004, **119**(3):329-341.
5. Jochim RC, Teixeira CR, Laughinghouse A, Mu J, Oliveira F, Gomes RB, Elnaiem DE, Valenzuela JG: **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* 2008, **9**(1):15.
6. Wu CH, Hsieh MJ, Huang JH, Luo SF: **Identification of low molecular weight allergens of American cockroach and production of monoclonal antibodies.** *Ann Allergy Asthma Immunol* 1996, **76**(2):195-203.
 7. Pomes A, Melen E, Vailes LD, Retief JD, Arruda LK, Chapman MD: **Novel allergen structures with tandem amino acid repeats derived from German and American cockroach.** *J Biol Chem* 1998, **273**(46):30801-30807.
 8. Pimenta PF, Modi GB, Pereira ST, Shahabuddin M, Sacks DL: **A novel role for the peritrophic matrix in protecting *Leishmania* from the hydrolytic activities of the sand fly midgut.** *Parasitology* 1997, **115** (Pt 4):359-369.
 9. Noriega FG, Pennington JE, Barillas-Mury C, Wang XY, Wells MA: ***Aedes aegypti* midgut early trypsin is post-transcriptionally regulated by blood feeding.** *Insect Mol Biol* 1996, **5**(1):25-29.
 10. Ramalho-Ortigao JM, Kamhawi S, Rowton ED, Ribeiro JM, Valenzuela JG: **Cloning and characterization of trypsin- and chymotrypsin-like proteases from the midgut of the sand fly vector *Phlebotomus papatasi*.** *Insect Biochem Mol Biol* 2003, **33**(2):163-171.
 11. Telleria EL, Pitaluga AN, Ortigao-Farias JR, de Araujo AP, Ramalho-Ortigao JM, Traub-Cseko YM: **Constitutive and blood meal-induced trypsin genes in *Lutzomyia longipalpis*.** *Arch Insect Biochem Physiol* 2007, **66**(2):53-63.

12. Sodja A, Fujioka H, Lemos FJ, Donnelly-Doman M, Jacobs-Lorena M: **Induction of actin gene expression in the mosquito midgut by blood ingestion correlates with striking changes of cell shape.** *J Insect Physiol* 2007, **53**(8):833-839.
13. Vlachou D, Schlegelmilch T, Christophides GK, Kafatos FC: **Functional genomic analysis of midgut epithelial responses in *Anopheles* during *Plasmodium* invasion.** *Curr Biol* 2005, **15**(13):1185-1195.
14. Harre JG, Dorsey KM, Armstrong KL, Burge JR, Kinnamon KE: **Comparative fecundity and survival rates of *Phlebotomus papatasi* sandflies membrane fed on blood from eight mammal species.** *Med Vet Entomol* 2001, **15**(2):189-196.

Chapter 6

Discussion and future directions

Overview

The objective of this dissertation was to further understand the relationship between the sand fly vector and the host as well as the relationship between the sand fly and the *Leishmania* parasite. As much of research is moving well into the “post-genomic era,” the field of sand fly biology is anxiously awaiting the completion of genome sequencing. In order to formulate and test important biological hypotheses, in the absence of genome data, we utilized functional transcriptomic technologies. Making use of this cost effective, productive and efficient method of high-throughput cDNA library sequencing, functional transcriptomics provides nearly the equivalent information of genomics. Even after the release of an organism’s genome, there is little useful data without further annotation of introns and exons, a task that often relies on proteomics and transcriptomics. By embracing functional transcriptomics, we are able to formulate hypotheses and answer essential research questions regarding disease vectors in a field that currently remains in the “pre-genomic era.” Accomplishing this research has had a significant beneficial impact on the current knowledge of sand fly biology and vector-pathogen interactions while advancing the frontier of functional transcriptomics. This work effectively characterized a unique salivary enzyme in *Phlebotomus duboscqi*, catalogued and analyzed a large set of midgut-specific transcripts from *Phlebotomus papatasi* and *Lutzomyia longipalpis* and additionally demonstrated the impact of *Leishmania infantum chagasi* and *L. major* on the temporal profile of sand fly midguts transcripts.

Phlebotomus duboscqi adenosine deaminase

Prior to our research, it was commonly accepted that adenosine deaminase (ADA), described in *Lu. longipalpis*, was only a component of the New World *Lutzomyia* sand flies. This concept was nullified with the discovery of two transcripts in the Old World sand fly *P. duboscqi* that likely encode an enzyme previously only considered a New World salivary molecule. To rule out the possibility that the transcripts were evolutionary relics encoding an inactive enzyme, we demonstrated enzymatic activity of ADA in the saliva. Additionally, purified recombinant proteins produced from the transcript sequences demonstrated ADA activity. The saliva of *P. duboscqi* does not contain any adenosine or adenosine-monophosphate (AMP), which are abundant molecules in the saliva of other *Phlebotomus* sand flies. The copious amount of salivary adenosine and AMP in *P. papatasi* functions as vasodilators; whereas, the enzyme maxadilan acts as a vasodilator in *Lu. longipalpis*. For that reason, there is likely a novel vasodilator molecule in the saliva of *P. duboscqi* that may be of importance in disease transmission, saliva-based vaccine research and potential uses in pharmacology. It is unclear what role ADA plays in blood-feeding. ADA has an important role in immunity as a result of the effects of adenosine, 2'-deoxyadenosine and the hydrolytic product of these compounds [1]. Further work will be necessary to determine the effect of this enzyme in parasite transmission and disease progression. It may be possible that this enzyme changes the environment in the skin where the *Leishmania* parasite is deposited by the sand fly. Inosine is the primary metabolite of adenosine by ADA and has been shown to inhibit the production of proinflammatory cytokines including TNF- α , IL-1, IL-12, MIP1- α and INF γ in stimulated macrophages and spleen cells [2]. Additionally,

adenosine can stimulate mast-cell degranulation, causing the release of histamine and serotonin [3]. The presence of ADA in the saliva of *P. duboscqi* may reduce the potential degranulation of mast-cells and the generation of an unfavorable feeding site due to the potent and immediate effects of histamine and serotonin. Another hypothesis, regarding the function of ADA in sand fly saliva, is that the enzyme plays an intermediary role in the degradation of adenosine to uric acid. If ingested by the sand fly, uric acid would act as an extremely effective peroxynitrate scavenger, protecting the midgut tissue from the toxic effects of peroxynitrate [4]. Ultimately, it may be possible that inosine generation by ADA activity favors a Th2 environment that will benefit parasite establishment in the skin of the mammalian host. The identification of these ADA transcripts in *P. duboscqi* demonstrates the independent acquisition of distinctive blood-feeding strategies. It is now apparent that the careful analysis of vector salivary components needs to be considered on a species specific level with due respect to the evolutionary force of host hemostasis on hematophagous arthropods. Foremost, the identification and characterization of ADA in *P. duboscqi* demonstrates how functional transcriptomics can be used to gain knowledge about the biological functioning of an organism. Therefore, future work may focus on the effect of ADA on parasite transmission and/or disease progression.

Functional transcriptomics of the sand fly midgut

We constructed and sequenced several cDNA libraries generated from the midgut tissue of female *P. papatasi* and *Lu. longipalpis*. Analyzing cDNA libraries created from different conditions in the sand fly midgut allowed for the comparison between sugar-fed,

blood-fed, and blood-fed *L. major*-infected *P. papatasi*. Using a similar approach with five *Lu. Longipalpis* midgut cDNA libraries, we were able to compare sugar-fed, blood-fed, and post-blood meal digestion conditions in the absence or presence of *L. infantum chagasi*. We then utilized this information to reproducibly identify the up- or downregulation of novel midgut proteins in each sand fly vector. These include microvillar proteins, peritrophins, trypsin, chymotrypsin and several uncharacterized proteins. These studies are also the first reports of transcript modulation by the presence of *Leishmania* parasites within the midgut of sand flies. It is likely that *Leishmania* parasites modulate the expression profile of other molecules, but our approach was not able to detect these proteins, probably as a result of their low abundance. Use of a normalized or cDNA subtraction library would yield far more information about low abundance transcripts, but would not allow the analysis of transcriptional modulation occurring under the conditions assessed by our research. The microvillar proteins (MVPs) are among the more interesting molecules characterized. The function of MVPs is unknown; however, based on the divergent amino acid sequences of these proteins they may have very different roles. In both *P. papatasi* and *Lu. longipalpis*, the presence of *L. major* or *L. infantum chagasi*, respectively, resulted in significant changes in blood-induced MVP sequences captured by the cDNA library. Future work will include the expression of *Lu. longipalpis* MVPs and the generation of polyclonal antibodies. The recombinant proteins and antibodies could then be used to investigate the effect of MVPs on *L. infantum chagasi* as well as the effect of antibody on blood digestion and sand fly longevity. Additionally, we would like to study the cellular localization of MVPs within the midgut tissue.

Another interesting class of molecules identified by functional transcriptomics is the peritrophins. *Leishmania* colonization induced changes in transcript abundance of peritrophin proteins in both *P. papatasi* and *Lu. longipalpis*. However, *PpPer1* was downregulated while *LuloPer1* was induced when the sand flies were correspondingly infected with *L. major* or *L. infantum chagasi*. Future work will likely include the expression of peritrophin molecules of *Lu. longipalpis* to demonstrate chitin-binding activity. Additionally, it would be intriguing to see if peritrophin-specific antibody could abrogate peritrophic matrix formation and if so, what effects this could have on the development of *L. infantum chagasi*.

The submitted sequences in NCBI's GenBank in March of 2007 contained eight midgut-specific sequences from *P. papatasi* and one from *Lu. longipalpis*. Subsequently, we have submitted 65 and 59 high quality sequences, primarily full length nucleotide and protein sequences, to GenBank from the midgut transcriptomes of *P. papatasi* and *Lu. longipalpis*, respectively. The sequence data submitted to GenBank from this research comprises over half of the total current available annotated nucleotide and protein sequence data for *P. papatasi* and *Lu. longipalpis*. Additionally, 4,439 and 9,533 ESTs were submitted to NCBI's EST database for *P. papatasi* and *Lu. longipalpis*, respectively. The transcriptomic approach presents a broader view and thus provides a more complete picture of an organism's functioning. In fact, the functional transcriptomic approach taken here offers both a global vantage (systematics) of the midgut transcriptome and the very exhaustive detailing (reductionism) of individual sequence elements of the "-ome." The research presented in this dissertation addresses both systematic and reductionist viewpoints; (1) the processes of blood-feeding and

digestion and the impact of *Leishmania* colonization of the sand fly midgut and (2) the comparison of amino acid sequences from the sequenced transcripts to confer putative function and contrast evolutionary dissemination.

Temporal profiling of Lutzomyia longipalpis midgut transcripts

Functional transcriptomics can provide an overwhelming amount of information. Using a Chi-square analysis to compare sequence abundance between cDNA libraries, we added order to the database in the context of blood digestion and *Leishmania* impact on sand fly midgut transcripts. Identifying molecules that have an altered expression due to the presence of *Leishmania* as measured by this methodology was an excellent preliminary finding. We selected sequences identified in the *Lu. longipalpis* cDNA libraries that were modulated by blood-feeding and parasite infection and reaffirmed this transcript alteration with greater sensitivity. This task was performed using a recently developed multiplex reverse transcriptase PCR system, analyzing midgut RNA that was collected at daily intervals to generate an informative temporal transcript profile. Our findings confirm that blood feeding and *L. infantum chagasi* manipulates a number of transcripts including carboxypeptidases, microvillar proteins, and peritrophins. The manipulation of the sand fly midgut transcriptome by *Leishmania* parasites may serve to benefit the parasites survival and necessary differentiation to the final, infective metacyclic form. Additionally, the presence of the parasite within the midgut of the sand fly may trigger protective responses, such as increased protease expression, eluding that the *Leishmania* colonization of the sand fly is neither mutualistic nor commensal. With the data provided by the cDNA libraries and the temporal profiling techniques used here,

another avenue of sand fly-*Leishmania* interactions can be researched. Our future endeavors include the temporal profiling of *P. papatasi* transcripts using normal blood-fed sand flies and *L. major*-infected sand flies. Moreover, we will analyze the impact of non-natural sand fly-parasite pairings (*Lu. longipalpis* infected with *L. major* and *P. papatasi* infected with *L. infantum chagasi*) in order to better understand molecular determinants of vector competence.

In addition to the aforementioned benefits of this research and the future goals, the identification of molecules necessary for successful colonization of the sand fly by *Leishmania* has important public health implications. A molecule that is vital for the generation of transmissible parasites, by allowing the full development of the parasite within the sand fly, is a potential transmission-blocking vaccine. A transmission-blocking vaccine may be of limited use to control transmission of parasites acquired from zoonotic reservoirs, such as rodents. However, in endemic areas in which canines are the principal reservoir for *L. infantum chagasi*, it is very likely that a canine-targeted transmission-blocking vaccine would greatly reduce the potential spread of a fatal parasitic disease.

References

1. Cristalli G, Costanzi S, Lambertucci C, Lupidi G, Vittori S, Volpini R, Camaioni E: **Adenosine deaminase: functional implications and different classes of inhibitors.** *Med Res Rev* 2001, **21**(2):105-128.

2. Ribeiro JM, Francischetti IM: **Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives.** *Annu Rev Entomol* 2003, **48**:73-88.
3. Hasko G, Cronstein BN: **Adenosine: an endogenous regulator of innate immunity.** *Trends Immunol* 2004, **25**(1):33-39.
4. Radi R, Beckman JS, Bush KM, Freeman BA: **Peroxynitrite oxidation of sulfhydryls. The cytotoxic potential of superoxide and nitric oxide.** *J Biol Chem* 1991, **266**(7):4244-4250.