

HOST-PARASITE TRANSCRIPTOMICS IN MOSQUITO-BORNE FILARIASIS

By

Young-Jun Choi

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Comparative Biomedical Sciences)

at the
UNIVERSITY OF WISCONSIN-MADISON
2013

Date of final oral examination: 01/10/13

The dissertation is approved by the following members of the Final Oral Committee:

Bruce M. Christensen, Professor, Pathobiological Sciences
Susan M. Paskewitz, Professor, Entomology
Nicole T. Perna, Associate Professor, Genetics
Adel M. Talaat, Associate Professor, Pathobiological Sciences
Timothy P. Yoshino, Professor, Pathobiological Sciences

ACKNOWLEDGEMENTS

I am grateful to have had the opportunity to pursue my studies under the guidance of Bruce Christensen. It was truly an honor to be his student, his patience, encouragement, and support will always be remembered. Appreciation is expressed for the assistance of Jeremy Fuchs, Shelly Michalski, Sara Erickson, and George Mayhew, who have greatly contributed to my graduate training and preparation of this thesis. Thanks also are given to Tim Yoshino, Susan Paskewitz, Adel Talaat, and Nicole Perna for kindly serving on my committee. Finally, I owe my deepest gratitude to my family for their continuing love and support.

TABLE OF CONTENTS

Acknowledgements	i
Table of Contents	ii
Dissertation Abstract	iii
Chapter 1. Introduction	1
Chapter 2. Tissue-enriched expression profiles in <i>Aedes aegypti</i> identify hemocyte-specific transcriptome responses to infection	15
Chapter 3. A deep sequencing approach to comparatively analyze the transcriptome of lifecycle stages of the filarial worm, <i>Brugia malayi</i>	54
Chapter 4. Dual RNA-seq of parasite and host reveals gene expression dynamics during filarial worm-mosquito interactions	93
Chapter 5. Summary	133

DISSERTATION ABSTRACT

HOST-PARASITE TRANSCRIPTOMICS IN MOSQUITO-BORNE FILARIASIS

Young-Jun Choi

Under the supervision of Bruce M. Christensen

at the University of Wisconsin-Madison

Lymphatic filariasis (LF), caused by mosquito-borne filarial nematodes, is one of the most debilitating of the neglected tropical diseases. Great progress has been made in control programs based on mass drug administration, but sustaining the success requires a better understanding of parasite biology and host-parasite interactions. The purpose of the present study was to enhance our understanding of mosquito cellular immunity, filarial parasite development and the interaction of developing parasites with the mosquito host, using high-throughput transcriptomic approaches as a primary investigative tool. To this end, studies were designed to characterize transcripts enriched in *Aedes aegypti* hemocytes, the blood cells that mediate innate immune responses, including phagocytosis, melanization, and production of antimicrobial peptides. Genome-wide microarray analysis of genes that exhibit tissue specificity following immune challenge provided a detailed overview of the hemocyte molecular repertoire, likely involved in cell type specific functions that coordinate the actions of different infection responsive tissues. To better understand the parasite developmental processes required for the infectious cycle, research efforts were directed to comparatively analyze the different life stages of the filarial worm *Brugia malayi*. RNA-seq analysis revealed transitions in gene expression from eggs through larval stages to adults, including patterns of sex-biased or germline-enriched expression, potentially implicated in reproductive processes. Discrete transcriptional changes were observed

during maturation of microfilariae and the third- to fourth-stage larval transition that are vital for infection in mosquito vectors and vertebrate hosts, respectively. Finally, experiments were conducted to investigate the *in vivo* transcriptome dynamics of *B. malayi* during its obligatory intracellular developmental phase in the mosquito. Integrative analysis of the host transcriptome in a dual RNA-seq approach revealed the temporal organization of transcriptional events in both the nematode and the host tissue. Parasite gene transcription dynamics exhibited a highly ordered developmental program consisting of a series of cyclical and state-transitioning temporal patterns, which was contextualized in relation to the concurrent dynamics of the host tissue transcriptome. These data provided new perspective and insight into filarial worm-mosquito interactions, which will lead to future systems-level studies to further our understanding of this symbiotic relationship responsible for the perpetuation of LF.

CHAPTER 1

INTRODUCTION

Lymphatic filariasis (LF) is a neglected tropical disease (NTD) caused by parasitic filarial nematodes *Wuchereria bancrofti*, *Brugia malayi*, and *B. timori*, all of which are transmitted by mosquitoes. *W. bancrofti* is responsible for 90% of infections, with *B. malayi* accounting for most of the remainder. An estimated 120 million people in 81 countries are currently infected, and more than 1.3 billion are at risk of infection. Approximately 65% of those at risk live in Southeast Asia, 30% in Africa, and the rest in other tropical areas (WHO, 2009). LF is the second leading cause of chronic disability worldwide due to its stigmatizing and disabling clinical manifestations, including 15 million people with lymphoedema (elephantiasis) and 25 million men with urogenital swelling, principally scrotal hydrocele (Zeldenryk et al., 2011). As with other NTDs, LF disproportionately affects resource-constrained regions of the world living in inadequate sanitation, which contributes to a vicious circle of infection, decreased productivity, and poverty (Lustigman et al., 2012b; Streit and Lafontant, 2008).

Filarial worms are heteroxenous parasites alternating between mosquito vectors and vertebrate hosts. The adult parasites residing in the lymphatics are dioecious and reproduce sexually via copulation. Inseminated adult female worms are ovoviviparous and release live larvae known as microfilariae (mf) into the lymph, where they eventually circulate in the bloodstream to be taken up by mosquitoes during blood feeding. These mf exhibit a fascinating circadian rhythm in their circulating concentration in the peripheral blood (Hawking, 1967). In areas where the parasite is transmitted by nocturnally feeding mosquitoes, nocturnally periodic strains, with peak microfilaremias from 10 pm to 2 am, are endemic, and where vectors are diurnally active, non-periodic parasite strain persists (Hawking and Denham, 1976). Once ingested mf penetrate successfully the midgut of a susceptible vector, they migrate to the thoracic muscles, and develop intracellularly through two molts to achieve the developmentally

arrested third-stage larvae (L3) that exit the mosquito proboscis during bloodfeeding and subsequently penetrate the mammalian host. After L3s enter into the lymphatics and lymph nodes of the definitive host, they undergo two additional molts and mature to adults within 6-12 months.

LF can manifest itself in a variety of clinical and subclinical conditions. Although two thirds of infected individuals have subclinical infections, approximately 40 million have pathologic manifestations, including hydroceles and other forms of urogenital disease, episodic adenolymphangitis, tropical pulmonary eosinophilia, lymphedema, and (in its most severe form) elephantiasis (Taylor et al., 2010). Filarial infections are a classical example of host–parasite interactions resulting in an immune system-parasite homeostatic balance, the failure of which can result in disease (Babu and Nutman, 2012). Asymptomatic infections typically accompany several immune regulatory processes that are thought to promote the long-term survival of the parasite (Hoerauf et al., 2005). It is not well understood what triggers the expression of clinical disease, but the host inflammatory response to dead or dying adult worms is thought to play an important role (Pfarr et al., 2009). In the lymphatics and lymph nodes, adult worms induce changes that result in dilatation of lymphatics and thickening of the lymphatic vessel walls. The effects of these tissue alterations, and the host inflammatory response to living and nonliving parasites and their endosymbiont *Wolbachia*, contribute to progressive lymphatic damage and pathology, which is often further aggravated by secondary bacterial or fungal infections (Dreyer et al., 2002; Figueredo-Silva et al., 2002). The most effective treatment for elephantiasis and lymphedema is good hygiene. Regular cleansing of the affected extremity and treatment of acute skin infections with antibiotics can significantly reduce the progression of lymphedema. For men with advanced hydroceles, hydrocelectomy is the recommended treatment.

The Global Program to Eliminate Lymphatic Filariasis (GPELF) was created in 1997 and is based on mass drug administration (MDA) involving yearly administration of ivermectin or diethylcarbamazine (DEC), in combination with albendazole, for 4-6 years. The strategy relies on the use of these anthelmintics to reduce microfilaremias in the human population, which in turn will suppress transmission by interrupting the passage of mf to the mosquito host. Although partial killing of adult parasites occur in some patients following DEC treatment, none of these drugs are effective macrofilaricides, and therefore MDA programs must treat the population at risk until the adult parasites expire to prevent re-occurrence of circulating mf. Great progress has been made in this program, but sustaining the success will require better understanding of fundamental parasite biology and host-parasite interactions. Exclusive reliance on small number of drugs, with little knowledge of how chemotherapeutic pressure affects parasite population genetic structure, makes the control programs potentially vulnerable to the development of drug resistance. To accelerate progress in parasite population biology, and support research to develop new treatments and make current ones more sustainable, advances in genomics and functional genomics of LF parasites are important. Although a draft genome of *B. malayi* has been published (Ghedini et al., 2007), without improved annotation and post-genomic functional studies, it will not be truly useful to support the search for novel interventions. As control programs progress towards elimination goals, our understanding of vector-parasite interactions and their ecological, epidemiological, and evolutionary implications becomes even more crucial (Lustigman et al., 2012a). For example, to quantify epidemiologically valid endpoints for meeting program objectives, one needs a more thorough understanding of filarial parasite-vector-host infection processes in order to define and understand both the vector to human and human to vector transmission thresholds (Grady et al., 2007; Michael and Gambhir, 2010; Michael et al.,

2006). Stopping a program too soon results in re-establishment of infections and running a program too long is economically taxing to resource poor countries where LF is endemic.

Persistence of filarial infections within human populations is dependent on the presence of competent mosquito vectors representing several genera, including *Culex*, *Aedes*, *Mansonia* and *Anopheles*. As with all biologically transmitted, heteroxenous parasites, filarial worm development in the mosquito host is dependent upon the reciprocal morphological, physiological and biochemical compatibility of the mosquito and parasite. Accordingly, mosquito-filarial worm associations can be incompatible for several reasons. Microfilariae can be physically damaged in those mosquitoes possessing a well-developed cibarial armature (McGreevy et al., 1978), rapidly clotting blood within the midgut lumen can prevent the migration of mf (Christensen, 1977; Sutherland et al., 1986), and mf can be killed by factors in the midgut of certain mosquito species (Michalski et al., 2010; Obiamiwe, 1977). In some instances, mf migrate to the tissue site for development, but after becoming intracellular, they simply fail to develop. This refractory condition is often called a physiological incompatibility (Christensen and Tracy, 1989) and represents the mechanism of refractoriness commonly seen in *Aedes aegypti*. There is a genetic basis for this trait in *A. aegypti* that is controlled by at least two loci (Severson et al., 1994), with a sex-linked recessive gene playing a major role (designated *fm*) (MacDonald, 1962). Finally, mosquitoes can mount an active immune response against mf after they penetrate the midgut and enter the hemolymph environment, as manifested by the deposition of melanin on the surface of the parasite (Christensen et al., 2005a).

The mosquito innate immune response is a major factor governing the interaction between vector and pathogen; consequently, it is a primary determinant of vector competence (Beerntsen et al., 2000). Innate immunity in mosquitoes and other insects employs both cellular

and humoral components in response to invading pathogens, and can lead to pathogen death via three broadly defined mechanisms: lysis, melanization and hemocyte-mediated phagocytosis (Charroux and Royet, 2010; Cirimotich et al., 2010; Ganesan et al., 2011; Schmidt et al., 2010; Yassine and Osta, 2010). First-line defenses include cells or cell products associated with the external cuticle, gut, and tracheal lining. Organisms that breach these barriers and invade the host hemocoel encounter constitutive and inducible defenses including phagocytosis and encapsulation by hemocytes, and humoral immune factors produced by hemocytes, pericardial cells and fat body (Hillyer, 2010). Hemocytes can be circulating with the hemolymph or attached to visceral tissues, and are thought to act as immunosurveillance cells that initiate innate immune responses. Studies show that a broad range of immune factors are produced by hemocytes, which include pattern recognition receptors, melanization modulators and enzymes, signal transduction proteins, and antimicrobial peptides (Bartholomay et al., 2007; Baton et al., 2009; Pinto et al., 2009).

Invading pathogens are recognized by the molecular interaction between host-derived pattern recognition receptors (PRRs) and pathogen associated molecular patterns. Mosquito PRRs include thioester containing proteins involved in the killing of bacteria and *Plasmodium* ookinetes (Blandin et al., 2004; Fraiture et al., 2009), C-type lectins, gram-negative binding proteins that bind β -1,3-glucan and lipopolysaccharide, and fibrinogen-related proteins that have undergone massive lineage-specific expansion in mosquitoes (Wang et al., 2005). Recognition by PRRs can lead to pathogen destruction through constitutive effector mechanisms or the activation of intracellular signaling pathways such as Toll, Imd and JAK/STAT, which activate the transcription of effector genes, among which are the antimicrobial peptides (AMPs) (Hoffmann, 2003). Modulatory proteins such as serine proteases and serine protease inhibitors

further control the activity levels of these pathways. The main AMPs in mosquitoes are defensins, cecropins and gambicins. These are low molecular weight secreted proteins that were initially identified by *in vitro* antimicrobial assays, which showed that cecropins and gambicin are cytotoxic primarily against Gram(-) bacteria and defensins are cytotoxic primarily against Gram(+) bacteria (Lowenberger et al., 1999a; Lowenberger et al., 1999b; Vizioli et al., 2000). Melanization response, on the other hand, involves a series of reactions that include the conversion of tyrosine to melanin precursors and the cross-linking of proteins to form a layer of melanin that surrounds and sequesters invading pathogens. Although most often associated with defense against metazoan parasites, including filarial worms, melanization also can be the primary response of mosquitoes against certain species of bacteria and protozoans (Christensen et al., 2005b).

THE PRESENT STUDY

Mosquito-borne parasites continue to inflict a major global health burden. To support ongoing control efforts and aid in the development of more effective intervention strategies, the research community strives to better understand the biology of mosquito hosts, parasites, and their interactions. In this context, the mosquito immune system has received considerable attention, and is described as an integrated network of tissues, cells and effector molecules that protect the host from infection (and thereby influence vector competence) (Hillyer, 2010). Hemocytes are at the heart of this network, as they mediate immune mechanisms, including phagocytosis, melanization, and production of antimicrobial peptides (Blandin and Levashina, 2007; Christensen et al., 2005a; Strand, 2008). However, our understanding of hemocyte-specific molecular processes and their contribution to shaping the host immune response remains limited.

In order to gain insight into the immunophysiological features distinctive of hemocytes, studies were designed in chapter 2, to analyze hemocyte-enriched transcripts in the mosquito *Aedes aegypti* using genome-wide microarrays. In particular, the experiments addressed how tissue-enriched expression patterns change with the immune status of the host. Genes exhibiting tissue specificity during an immune response are particularly interesting in the context of hemocyte biology due to their possible involvement in cell type specific functions, such as intercellular signaling and communication that help coordinate the actions of different infection responsive tissues. In addition, conserved hemocyte-enriched molecular repertoires which might be implicated in core hemocyte function were identified by cross-species meta-analysis of microarray expression data from *A. gambiae* and *Drosophila melanogaster*.

In chapter 3, research efforts were directed to *B. malayi*, one of the three mosquito borne filarial nematodes that cause LF in humans. The purpose of the study in this chapter was to understand the developmental processes required for the infectious cycle by characterizing the transcriptional program of the parasite's lifecycle. To this end, high-throughput cDNA sequencing (RNA-seq) was used to study the transcriptome of seven *B. malayi* lifecycle stages: eggs & embryos, immature (≤ 3 days of age) and mature microfilariae (MF), third- and fourth-stage larvae (L3 and L4), and adult male and female worms. Comparative analysis across these stages provided a detailed overview of the molecular repertoires that define and differentiate distinct lifecycle stages of the parasite. Of particular interest was the identification of genes displaying sex-biased or germline-enriched profiles due to their potential involvement in reproductive processes. The study also revealed discrete transcriptional changes during larval development, namely those accompanying the maturation of MF and the L3 to L4 transition that

are vital in establishing successful infection in mosquito vectors and vertebrate hosts, respectively.

It is well accepted that the study of inter-species interaction is fundamental to understanding the biology of vector-borne parasites (Castillo et al., 2011; Lefevre and Thomas, 2008; Matthews, 2011), and filarial worms that cause LF are no exception. In humans, the interactions these parasites engage in with the host tissue and immune system dictate the pathobiology of LF (Bennuru and Nutman, 2009; Pfarr et al., 2009). Likewise, the interactions between filarial worms and their mosquito intermediate hosts determine the course of parasite development to the infective stage (Bartholomay and Christensen, 2002). The purpose of the studies presented in chapter 4, and that of this thesis overall was to develop a more complete understanding of this interplay of interactions that govern the filarial worm-mosquito symbiosis responsible for the perpetuation of LF as a major cause of long-term and permanent disability of millions of people. Past studies on filarial worm-mosquito interactions have been essentially one-sided investigations focusing on either of the partners, mostly on host factors without taking the parasite activity into explicit consideration. As a consequence, filarial worm molecular processes underlying development in, and interaction with, the mosquito host tissue remain poorly described. In order to generate molecular insights into the dynamic symbiotic processes of filarial worm development in the mosquito, a dual RNA-seq approach was employed using *B. malayi*-*A. aegypti* system to investigate the temporal organization of transcriptional events in both the nematode and the host tissue from the establishment of infection to the emergence of the L3. It was found that parasite gene transcription dynamics exhibit a highly ordered developmental program consisting of a series of cyclical and state-transitioning temporal patterns,

which was contextualized in relation to the concurrent dynamics of the host tissue transcriptome.

The overall conclusions and future directions of this work as a whole are discussed in chapter 5.

REFERENCES

- Babu, S., Nutman, T.B., 2012. Immunopathogenesis of lymphatic filarial disease. *Semin Immunopathol* 34, 847-861.
- Bartholomay, L.C., Christensen, B.M., 2002. Vector-parasite interactions in mosquito-borne filariasis., in: Klei, T., Rajan, T. (Eds.), *The Filaria*, 1 ed. Kluwer Academic Publishers, Boston, pp. 9-19.
- Bartholomay, L.C., Mayhew, G.F., Fuchs, J.F., Rocheleau, T.A., Erickson, S.M., Aliota, M.T., Christensen, B.M., 2007. Profiling infection responses in the haemocytes of the mosquito, *Aedes aegypti*. *Insect molecular biology* 16, 761-776.
- Baton, L.A., Robertson, A., Warr, E., Strand, M.R., Dimopoulos, G., 2009. Genome-wide transcriptomic profiling of *Anopheles gambiae* hemocytes reveals pathogen-specific signatures upon bacterial challenge and *Plasmodium berghei* infection. *BMC genomics* 10, 257.
- Beerntsen, B.T., James, A.A., Christensen, B.M., 2000. Genetics of mosquito vector competence. *Microbiology & Molecular Biology Reviews*. 64, 115-137.
- Bennuru, S., Nutman, T.B., 2009. Lymphatics in human lymphatic filariasis: in vitro models of parasite-induced lymphatic remodeling. *Lymphat Res Biol* 7, 215-219.
- Blandin, S., Shiao, S.H., Moita, L.F., Janse, C.J., Waters, A.P., Kafatos, F.C., Levashina, E.A., 2004. Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell* 116, 661-670.
- Blandin, S.A., Levashina, E.A., 2007. Phagocytosis in mosquito immune responses. *Immunological reviews* 219, 8-16.
- Castillo, J.C., Reynolds, S.E., Eleftherianos, I., 2011. Insect immune responses to nematode parasites. *Trends Parasitol* 27, 537-547.
- Charroux, B., Royet, J., 2010. *Drosophila* immune response: From systemic antimicrobial peptide production in fat body cells to local defense in the intestinal tract. *Fly* 4, 40-47.
- Christensen, B.M., 1977. Laboratory studies on the development and transmission of *Dirofilaria immitis* by *Aedes trivittatus*. *Mosquito News* 37, 367-372.
- Christensen, B.M., Li, J., Chen, C.C., Nappi, A.J., 2005a. Melanization immune responses in mosquito vectors. *Trends in parasitology* 21, 192-199.

Christensen, B.M., Li, J., Chen, C.C., Nappi, A.J., 2005b. Melanization immune responses in mosquito vectors. *Trends in Parasitology* 21, 192-199.

Christensen, B.M., Tracy, J.W., 1989. Arthropod-transmitted parasites: mechanisms of immune interaction, in: Yoshino, T., Damian, R. (Eds.), *Phylogeny of Immune Defense Mechanisms in Parasitic Infections*, pp. 387-398.

Cirimotich, C.M., Dong, Y., Garver, L.S., Sim, S., Dimopoulos, G., 2010. Mosquito immune defenses against *Plasmodium* infection. *Developmental and comparative immunology* 34, 387-395.

Dreyer, G., Addiss, D., Roberts, J., Noroes, J., 2002. Progression of lymphatic vessel dilatation in the presence of living adult *Wuchereria bancrofti*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 96, 157-161.

Figueredo-Silva, J., Noroes, J., Cedenho, A., Dreyer, G., 2002. The histopathology of bancroftian filariasis revisited: the role of the adult worm in the lymphatic-vessel disease. *Annals of tropical medicine and parasitology* 96, 531-541.

Fraiture, M., Baxter, R.H., Steinert, S., Chelliah, Y., Frolet, C., Quispe-Tintaya, W., Hoffmann, J.A., Blandin, S.A., Levashina, E.A., 2009. Two mosquito LRR proteins function as complement control factors in the TEP1-mediated killing of *Plasmodium*. *Cell Host Microbe* 5, 273-284.

Ganesan, S., Aggarwal, K., Paquette, N., Silverman, N., 2011. NF-kappaB/Rel proteins and the humoral immune responses of *Drosophila melanogaster*. *Current topics in microbiology and immunology* 349, 25-60.

Ghedini, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J.E., Delcher, A.L., Guiliano, D.B., Miranda-Saavedra, D., Angiuoli, S.V., Creasy, T., Amedeo, P., Haas, B., El-Sayed, N.M., Wortman, J.R., Feldblyum, T., Tallon, L., Schatz, M., Shumway, M., Koo, H., Salzberg, S.L., Schobel, S., Perte, M., Pop, M., White, O., Barton, G.J., Carlow, C.K., Crawford, M.J., Daub, J., Dimmic, M.W., Estes, C.F., Foster, J.M., Ganatra, M., Gregory, W.F., Johnson, N.M., Jin, J., Komuniecki, R., Korf, I., Kumar, S., Laney, S., Li, B.W., Li, W., Lindblom, T.H., Lustigman, S., Ma, D., Maina, C.V., Martin, D.M., McCarter, J.P., McReynolds, L., Mitreva, M., Nutman, T.B., Parkinson, J., Peregrin-Alvarez, J.M., Poole, C., Ren, Q., Saunders, L., Sluder, A.E., Smith, K., Stanke, M., Unnasch, T.R., Ware, J., Wei, A.D., Weil, G., Williams, D.J., Zhang, Y., Williams, S.A., Fraser-Liggett, C., Slatko, B., Blaxter, M.L., Scott, A.L., 2007. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 317, 1756-1760.

Grady, C.A., de Rochars, M.B., Direny, A.N., Orelus, J.N., Wendt, J., Radday, J., Mathieu, E., Roberts, J.M., Streit, T.G., Addiss, D.G., Lammie, P.J., 2007. Endpoints for lymphatic filariasis programs. *Emerging infectious diseases* 13, 608-610.

Hawking, F., 1967. The 24-hour periodicity of microfilariae: Biological mechanisms responsible for its production and control. *Proc R Soc London Ser B* 169, 56-76.

Hawking, F., Denham, D.A., 1976. The distribution of human filariasis throughout the world. Part I. the Pacific Region, including New Guinea. *Tropical diseases bulletin* 73, 347-373.

Hillyer, J.F., 2010. Mosquito immunity. *Adv Exp Med Biol* 708, 218-238.

Hoerauf, A., Satoguina, J., Saeftel, M., Specht, S., 2005. Immunomodulation by filarial nematodes. *Parasite immunology* 27, 417-429.

Hoffmann, J.A., 2003. The immune response of *Drosophila*. *Nature* 426, 33-38.

Lefevre, T., Thomas, F., 2008. Behind the scene, something else is pulling the strings: emphasizing parasitic manipulation in vector-borne diseases. *Infect Genet Evol* 8, 504-519.

Lowenberger, C., Charlet, M., Vizioli, J., Kamal, S., Richman, A., Christensen, B.M., Bulet, P., 1999a. Antimicrobial activity spectrum, cDNA cloning, and mRNA expression of a newly isolated member of the cecropin family from the mosquito vector *Aedes aegypti*. *The Journal of biological chemistry* 274, 20092-20097.

Lowenberger, C.A., Smartt, C.T., Bulet, P., Ferdig, M.T., Severson, D.W., Hoffmann, J.A., Christensen, B.M., 1999b. Insect immunity: molecular cloning, expression, and characterization of cDNAs and genomic DNA encoding three isoforms of insect defensin in *Aedes aegypti*. *Insect molecular biology* 8, 107-118.

Lustigman, S., Geldhof, P., Grant, W.N., Osei-Atweneboana, M.Y., Sripa, B., Basanez, M.G., 2012a. A research agenda for helminth diseases of humans: basic research and enabling technologies to support control and elimination of helminthiasis. *PLoS neglected tropical diseases* 6, e1445.

Lustigman, S., Prichard, R.K., Gazzinelli, A., Grant, W.N., Boatman, B.A., McCarthy, J.S., Basanez, M.G., 2012b. A research agenda for helminth diseases of humans: the problem of helminthiasis. *PLoS neglected tropical diseases* 6, e1582.

MacDonald, W.W., 1962. The genetic basis of susceptibility of infection with semi-periodic *Brugia malayi* in *Aedes aegypti*. *Annals of Tropical Medicine and Parasitology* 56, 373-382.

Matthews, K.R., 2011. Controlling and coordinating development in vector-transmitted parasites. *Science* 331, 1149-1153.

McGreevy, P., Bryan, J., Oothuman, P., Kolstrup, N., 1978. The lethal effects of the cibarial and pharyngeal armatures of mosquitoes on microfilariae. *Transactions of the Royal Society of Tropical Medicine & Hygiene*. 72, 361-368.

Michael, E., Gambhir, M., 2010. Transmission models and management of lymphatic filariasis elimination. *Adv Exp Med Biol* 673, 157-171.

Michael, E., Malecela-Lazaro, M.N., Kabali, C., Snow, L.C., Kazura, J.W., 2006. Mathematical models and lymphatic filariasis control: endpoints and optimal interventions. *Trends in parasitology* 22, 226-233.

Michalski, M.L., Erickson, S.M., Bartholomay, L.C., Christensen, B.M., 2010. Midgut barrier imparts selective resistance to filarial worm infection in *Culex pipiens pipiens*. *PLoS neglected tropical diseases* 4, e875.

Obiamiwe, B.A., 1977. The effect of anticoagulant on the early migration of *Brugia pahangi* microfilariae in *Culex pipiens* susceptible or refractory to *B. pahangi*. *Annals of Tropical Medicine and Parasitology* 71, 371-374.

Pfarr, K.M., Debrah, A.Y., Specht, S., Hoerauf, A., 2009. Filariasis and lymphoedema. *Parasite immunology* 31, 664-672.

Pinto, S.B., Lombardo, F., Koutsos, A.C., Waterhouse, R.M., McKay, K., An, C., Ramakrishnan, C., Kafatos, F.C., Michel, K., 2009. Discovery of Plasmodium modulators by genome-wide analysis of circulating hemocytes in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences of the United States of America* 106, 21270-21275.

Schmidt, O., Soderhall, K., Theopold, U., Faye, I., 2010. Role of adhesion in arthropod immune recognition. *Annual review of entomology* 55, 485-504.

Severson, D., Mori, A., Zhang, Y., Christensen, B., 1994. Chromosomal mapping of two loci affecting filarial worm susceptibility in *Aedes aegypti*. *Insect Molecular Biology* 3, 67-72.

Strand, M.R., 2008. Insect hemocytes and their role in immunity, in: Beckage, N.E. (Ed.), *Insect Immunology*. Academic Press, pp. 25-47.

Streit, T., Lafontant, J.G., 2008. Eliminating lymphatic filariasis: a view from the field. *Annals of the New York Academy of Sciences* 1136, 53-63.

Sutherland, D.R., Christensen, B.M., Lasee, B.A., 1986. Midgut barrier as a possible factor in filarial worm vector competency in *Aedes trivittatus*. *Journal of Invertebrate Pathology* 47, 1-7.

Taylor, M.J., Hoerauf, A., Bockarie, M., 2010. Lymphatic filariasis and onchocerciasis. *Lancet* 376, 1175-1185.

Vizioli, J., Bulet, P., Charlet, M., Lowenberger, C., Blass, C., Muller, H.M., Dimopoulos, G., Hoffmann, J., Kafatos, F.C., Richman, A., 2000. Cloning and analysis of a cecropin gene from the malaria vector mosquito, *Anopheles gambiae*. *Insect molecular biology* 9, 75-84.

Wang, X., Zhao, Q., Christensen, B.M., 2005. Identification and characterization of the fibrinogen-like domain of fibrinogen-related proteins in the mosquito, *Anopheles gambiae*, and the fruitfly, *Drosophila melanogaster*, genomes. *BMC genomics* 6, 114.

WHO, 2009. Global program to eliminate lymphatic filariasis. *Wkly Epidemiol Rec* 84, 437-444.

Yassine, H., Osta, M.A., 2010. *Anopheles gambiae* innate immunity. *Cell Microbiol* 12, 1-9.

Zeldenryk, L.M., Gray, M., Speare, R., Gordon, S., Melrose, W., 2011. The emerging story of disability associated with lymphatic filariasis: a critical review. *PLoS neglected tropical diseases* 5, e1366.

CHAPTER 2**TISSUE-ENRICHED EXPRESSION PROFILES IN *Aedes aegypti* IDENTIFY
HEMOCYTE-SPECIFIC TRANSCRIPTOME RESPONSES TO INFECTION**

Choi, Y-J, Fuchs, JF, Mayhew, GF, Yu, H and Christensen, BM (2012) *Insect Biochem Mol Biol* 42: 729-738. [PMID: 22796331]

ABSTRACT

Hemocytes are integral components of mosquito immune mechanisms such as phagocytosis, melanization, and production of antimicrobial peptides. However, our understanding of hemocyte-specific molecular processes and their contribution to shaping the host immune response remains limited. To better understand the immunophysiological features distinctive of hemocytes, we conducted genome-wide analysis of hemocyte-enriched transcripts, and examined how tissue-enriched expression patterns change with the immune status of the host. Our microarray data indicate that the hemocyte-enriched transcriptome is dynamic and context-dependent. Analysis of transcripts enriched after bacterial challenge in circulating hemocytes with respect to carcass added a dimension to evaluating infection-responsive genes and immune-related gene families. We resolved patterns of transcriptional change unique to hemocytes from those that are likely shared by other immune responsive tissues, and identified clusters of genes preferentially induced in hemocytes, likely reflecting their involvement in cell type specific functions. In addition, the study revealed conserved hemocyte-enriched molecular repertoires which might be implicated in core hemocyte function by cross-species meta-analysis of microarray expression data from *Anopheles gambiae* and *Drosophila melanogaster*.

INTRODUCTION

Hemocytes are fundamental elements of the mosquito host defense. They mediate immune mechanisms such as phagocytosis, melanization, and production of antimicrobial peptides (AMPs) (Blandin and Levashina, 2007; Christensen et al., 2005; Strand, 2008). Phagocytosis and melanization responses can be initiated within minutes of non-self recognition, albeit to a varying degree depending on the pathogen class (Hillyer et al., 2003b, 2004). In comparison, AMP responses are characterized by transcriptional induction through Rel/NF- κ B signaling pathways (Antonova et al., 2009; Shin et al., 2005). The emerging complexity of hemocyte biology and its multifaceted role in host defense lie not only in the temporally distinct and pathogen-specific manifestation of the effector mechanisms these cells mediate, but also because hemocyte responses occur in the context of an elaborate network of processes involving multiple physiological systems and immune-responsive tissues (Glenn et al., 2010; Hillyer, 2010; Schneider, 2009). The interactions between these components give rise to the function and behavior of the immune system as a whole. To begin to investigate the role of hemocytes in such interplay, a better understanding of the unique immunophysiological features of hemocytes becomes increasingly important. In this study, we utilized tissue-enriched expression profiles to critically evaluate hemocyte transcriptome responses to bacterial challenge by resolving patterns of transcriptional changes specific to circulating hemocytes from those that are likely shared by other immune responsive tissues. In addition, we identified conserved hemocyte-enriched molecular repertoires which might be implicated in core hemocyte function by cross-species meta-analysis of microarray expression data from *Anopheles gambiae* (Pinto et al., 2009) and *Drosophila melanogaster* (Irving et al., 2005).

Previous transcriptomic studies in dipteran species investigating the molecular physiology of circulating hemocytes (Baton et al., 2009; Irving et al., 2005; Pinto et al., 2009) have demonstrated that comparing hemocyte transcriptome profiles to carcass profiles can provide a useful metric for the screening of transcripts enriched in hemocytes. Tissue comparisons of this nature require careful analysis because (1) the numeric values of the resulting enrichment ratios do not have a readily interpretable biological meaning due to the undefined cellular composition of the carcass, and (2) high enrichment ratios themselves do not necessarily indicate tissue-specific gene expression. Nevertheless, relative rankings of the enrichment ratios in the context of a genome-wide screening can be highly informative because transcripts primarily or exclusively expressed in hemocytes will likely have higher enrichment ratios than most other transcripts. An added advantage of this approach is that it provides a potential means to guard against false-positive findings in studies comparing hemocyte samples, where results may be confounded by cell type heterogeneity. Provided that the proportion of contaminating cell type(s) in the hemocyte sample is lower than that in the carcass sample, enrichment ratios of the transcripts not expressed in hemocytes remain less than 1, and thus may be used as an additional criterion for critical evaluation of transcriptional profiles.

At any given time, the expression of many genes varies between different cell types and between different developmental and physiological states. Tissue-enriched genes, which are highly expressed in one particular tissue type and are either not expressed or are expressed at much lower levels in other tissues, have been hypothesized to be important in the specialized functions of the particular cell types in which they are expressed. Genes of the common host response may be induced in multiple tissue types during infection, whereas some clusters of genes are preferentially induced in specific cell types, likely reflecting their unique function in

response to infection. Genes exhibiting tissue specificity during an immune response are particularly interesting in the context of hemocyte biology due to their possible involvement in cell type specific functions, such as intercellular signaling and communication that help coordinate the actions of different infection responsive tissues. Within this conceptual framework, the present study provides a detailed molecular perspective into the characteristic features of the hemocyte transcriptome in the mosquito *Aedes aegypti* by actively harnessing tissue-enriched expression profiles following bacterial challenge.

MATERIALS AND METHODS

Mosquito rearing and colony maintenance. The *A. aegypti* Liverpool strain was originally obtained from a colony from the University of London in 1977 and was reared as previously described(Christensen and Sutherland, 1984). Adult female mosquitoes were used for experimentation within 3 days of eclosion.

Experimental design and replication. Material for each experimental condition was generated from three (bacterial challenge) or four (naïve) separate generations of mosquitoes, and four separate microarray hybridizations using carcass and hemocyte material were performed. The first hybridization compared naïve and *E.coli*-challenged conditions, the second compared naïve and *M. luteus*-challenged conditions, and the third and fourth compared all three conditions.

Bacterial cultures and injection into mosquitoes. Bacteria used for intrathoracic injections were *E. coli* DH5 α and *M. luteus* (University of Wisconsin-Madison). Cultures were grown to stationary phase in LB (Luria-Bertani) broth at 37 °C, with shaking at 300 rpm. A pulled glass capillary needle containing inoculum (0.5 μ l undiluted bacterial culture) was inserted through the cervical membrane between the head and the thorax and the fluid injected as previously

described (Hillyer et al., 2004). For each biological replicate, 50 individuals were injected per condition or remained naïve, and survivorship at 24 hours following injection was greater than 90%.

Mosquito tissue collection. At 24 hours post bacterial challenge, hemolymph was collected from 40 individuals by volume displacement (perfusion) as previously described (Beerntsen and Christensen, 1990). A tear was made above the penultimate abdominal segment of the mosquito, which was then placed on a vacuum saddle. A pulled glass capillary needle, attached to a syringe containing 1X HBSS (Invitrogen, Carlsbad, CA), was inserted through the cervical membrane between the head and the thorax. HBSS was slowly injected, and only the first drop of perfusate from each mosquito was collected into a microfuge tube containing cell lysis buffer (10% SDS, 1 M Tris pH 7.5, 5 mM EDTA), mixed well, and kept on ice. Each remaining carcass after perfusion was collected in parallel to the hemolymph sample by immediately freezing it in a tube on dry ice. Carcasses were stored at -80 °C.

RNA isolation and purification. Hemocyte total RNA was isolated from hemolymph immediately following collection. The isolation utilized a modification of a method originally developed for isolation of RNA from tissue culture cells (Peppel and Baglioni, 1990). Hemolymph was directly perfused into a 1.5 ml microfuge tube containing 150 µl 5X cell lysis buffer (10% SDS, 1 M Tris pH 7.5, 5 mM EDTA), mixed, and stored on ice. Once all the perfusate was collected, the solution was adjusted to 1X with HBSS and vortexed at medium speed for 5 seconds. Next, 0.3 volumes of ice-cold Solution 2 (42.9 g potassium acetate, 11.2 ml acetic acid, water to 100 ml) was added and the solution was vortexed at medium speed both upright and inverted for 10 seconds, centrifuged (14 k, 4 °C) for 15 min. and the supernatant transferred to a new tube. The solution was then centrifuged (14 k, 4 °C) for 15 min. and the

supernatant transferred to a new tube. One half volume of buffered phenol and one half volume of chloroform/isoamyl 24:1 were added, mixed, and centrifuged (14 k, 4 °C) for 10 min. and the supernatant transferred to a new tube. One volume of chloroform/isoamyl 24:1 was added, mixed, centrifuged (14 k, 4 °C) for 10 min. and the supernatant transferred to a new tube. One volume of ice-cold isopropanol was added, mixed, and the RNA was precipitated overnight at -20 °C. RNA was pelleted (14 k, 4 °C) for 30 min. and the supernatant discarded. RNA was washed with 750 µl ice-cold 80% ethanol, centrifuged (14 k, 4 °C) for 10 min. and the supernatant discarded. The RNA was then air dried and resuspended in 10 µl nuclease-free water. Carcass total RNA was isolated using the single-step acid guanidinium thiocyanate-phenol-chloroform extraction method (Chomczynski and Sacchi, 1987) and resuspended in nuclease-free water. The quantity and purity of all RNA samples were measured on a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA). The measured RNA concentrations and RNA integrity were verified by denaturing gel electrophoresis using GelRed™ (Phenix Research Products, Candler, NC) staining.

Microarray hybridization. Agilent (Santa Clara, CA) 4x44k whole-genome *Aedes aegypti* microarrays (Nene et al., 2007) were used for hybridizations. To generate amino allyl-modified cDNA (aRNA) coupled with CyDye (GE Healthcare, Piscataway, NJ), Amino Allyl MessageAmp™ II aRNA Kit (Ambion AM1753, Austin, TX) was used according to the manufacturers' instructions. The *in vitro* transcription was performed for 14 hours for all samples starting with 500 ng of total RNA. The quantity and purity of all resulting modified aRNA were measured on a NanoDrop spectrophotometer. The measured aRNA concentrations and aRNA size distributions were evaluated by denaturing gel electrophoresis using GelRed™ staining. In order to utilize dye-swap pairs for hybridization, each sample was coupled to both Cy3 and Cy5.

The same amount of input aRNA (5-7 μg) was used for all samples within a biological replicate. The dye labeled aRNA concentration, pmol/ μl dye and frequency of incorporation (FOI) were measured on a NanoDrop spectrophotometer. If the FOI of the labeled aRNA was outside of the desirable range the dye coupling reaction was repeated. Microarrays were hybridized using the Agilent Gene Expression Hybridization Kit according to manufacturers' instructions. Two color hybridizations were performed using 125 pmol (700-1200 ng) of each labelled aRNA. Once assembled, arrays were secured in Agilent hybridization chambers and hybridized for 17 hours at 65 °C in an Agilent hybridization oven rotating at 10 rpm. Microarrays were washed according to the manufacturers' instructions and scanned immediately.

Microarray data processing and Gene Ontology enrichment analysis. Fluorescence images were obtained from hybridized microarrays at 5 μm resolution using a GenePix 4000B array scanner (Molecular Devices, Foster City, CA) with the PMT gain settings automatically adjusted to a saturation tolerance level of 0.001%. Array features were quantified in GenePix Pro 6.1 (Axon Instruments) and low quality fluorescence spots were flagged and excluded from subsequent analyses. Extracted results were imported into R environment for differential transcript abundance analysis using the Bioconductor LIMMA package (Smyth, 2004). "Normexp" method (Ritchie et al., 2007) was used for background correction with an offset of 50 for both channels to damp down the variability of the log ratios for low intensity spots while avoiding negative corrected intensities. The A (average intensities) and M (log-ratios) values were averaged among within-array replicate spots, followed by between-array normalization using the "Aquantile" method (Smyth and Speed, 2003) such that the A-values have the same empirical distribution across arrays but the M-values remain unchanged. Following the approach developed in LIMMA for separate channel analysis of two-color microarrays, a linear model was

fit to log-intensity data rather than log-ratio data. Because the study utilized dye-swap technical replicates, each biological replicate was included as a coefficient in the linear model. After model fitting, relevant pairwise comparisons were made by specifying appropriate contrasts. Statistical significance of differential transcript abundance was assessed using moderated t-statistics, and p-values were adjusted for multiple testing to control false discovery rate using the Benjamini and Hochberg's method. The microarray data from this study were deposited to GEO under accession number GSE38744. Gene Ontology (GO) analysis was performed in ErmineJ (Gillis et al., 2010; Lee et al., 2005) using GO annotations retrieved from the UniProtKB-GOA database (Camon et al., 2004) and VectorBase (Lawson et al., 2009). The gene score resampling (GSR) method was used to identify statistically overrepresented GO categories among genes showing high hemocyte-enriched expression pattern. P-values were adjusted for multiple testing to control false discovery rate using the Benjamini and Hochberg's method. Hierarchical clustering analysis was performed in GeneSpring GX (Agilent Technologies) with average linkage using Pearson's uncentered correlation coefficient as distance metric.

Microarray validation. qPCR (quantitative polymerase chain reaction) was used to validate the microarray expression ratios for a selected group of transcripts. PrimeTime™ qPCR Assays (Integrated DNA Technologies, Coralville, IA) were designed and resuspended in TE (10mM Tris, 0.1mM EDTA pH 8.0) according to the manufacturer's recommendations. qPCR reactions (20 µl, 10 ng total RNA input) were executed using the TaqMan® RNA-to-C_t™ 1-Step Kit (Applied Biosystems, Carlsbad, CA) and an Applied Biosystems 7300 Real-Time PCR system according to the manufacturer's recommendations. The Comparative C_t Method was utilized using one endogenous control (AAEL000987; 60S ribosomal protein L8) and four target genes (AAEL007967, AAEL000636, AAEL006355, AAEL006361). Validation experiments

confirmed the endogenous control could be used for all four targets. All reactions were performed in triplicate. Primer sequences and results are summarized in Supplemental Dataset S1.

Orthologous group comparisons. *A. gambiae* hemocyte and carcass microarray data were obtained from (Pinto et al., 2009), and hemocyte enrichment ratios were computed using normalized intensity values reported in their S (stringent) list. *D. melanogaster* hemocyte and whole larvae data were obtained from (Irving et al., 2005), and processed similarly to generate tissue enrichment ratios, except that enrichment ratios were imputed for genes whose expression data exist only for hemocytes using the median (normalized intensity) value for the whole larvae microarray as the common denominator. Transcripts and their corresponding tissue enrichment ratios were mapped to genes and orthologous groups (OGs) to identify OGs that retained hemocyte-enriched expression patterns across *A. aegypti*, *A. gambiae* and *D. melanogaster*. An OG was considered as such, if there was at least one ortholog from each species that showed hemocyte enrichment. Insect non-supervised orthologous groups (inNOGs) from the eggNOG database (version 2.0) were used as a basis for these comparisons (Muller et al., 2010). For each species, an equal number of genes most highly enriched in hemocytes (as measured by enrichment ratios relative to the carcass or the whole animal within each dataset) were selected for Venn diagram analysis. To determine genes showing the most consistent pattern of high tissue enrichment across species, we compared orthologous groups using the geometric mean of the relative (i.e., within-study) enrichment ratio rankings of the constituent members.

RESULTS AND DISCUSSION

Hemocyte collection for transcriptome profiling. The methodologies previously employed to collect hemocytes for transcriptomic analyses undoubtedly contributed their own biases to both the transcripts that appear as significantly changed and the magnitude of their changes. In previous studies, hemocytes have been collected via displacement perfusion with cell lysis buffer (Bartholomay et al., 2007), collection from clipped probosces (Abraham et al., 2005; Chen and Laurence, 1987; Chun et al., 2000; Pinto et al., 2009), or collection in a capillary tube following the injection of buffer into the abdomen (Castillo et al., 2006). Not only do different collections contain varying proportions of contaminating cell types, but the hemocyte transcriptome itself can remain dynamic for differing amounts of time and be exposed to varying stimuli prior to either cell lysis or freezing in anticipation of RNA isolation. The displacement perfusion technique utilized in this study was designed to result in minimal contaminating cell types while rapidly initiating hemocyte lysis and capturing transcriptome RNA in the absence of influence from the collection conditions or factors outside of the mosquito hemocoel.

Microarray experimental design. Using *A. aegypti* whole-genome microarrays (Nene et al., 2007), we profiled the transcriptome of circulating hemocytes and the remaining carcass in both naïve and bacteria-challenged adult female mosquitoes (Fig. 1A). For each biological replication, hemolymph perfusate and the remaining carcass were collected as paired samples at 24 hours post intra-hemocoelic injection of either *Escherichia coli* or *Micrococcus luteus* overnight culture (0.5 µl) in parallel with an age-matched naïve group from the same cohort. A paired design was employed so that, within each treatment group, variation in physiological or immunological state of the experimental animals would not be confounded with tissue comparison. Concomitantly-collected hemocyte and carcass samples were co-hybridized on microarrays in accordance with our primary emphasis on the accurate delineation of tissue

enrichment ratios. However, ranking transcripts based on tissue enrichment was not the sole aim of the experimental design. We wanted to investigate how transcripts enriched in hemocytes change as a function of bacterial challenge, and conversely which infection-responsive transcripts show marked enrichment in hemocytes. To examine transcriptional changes associated with hemocoelic infection in distinct tissue samples, we adopted a linear model-based framework for analyzing dual-channel hybridizations (i.e., two-color microarrays) as separate single-channels (Smyth, 2004). This analytic approach benefited from consistent high quality of *in situ* synthesized Agilent microarrays and effective “within” and “between” array normalization methods in LIMMA (Bioconductor) (Ritchie et al., 2007; Smyth and Speed, 2003). The biological and technical dye-swap replicates of the 2×3 experimental conditions consisted of forty separate single-channel intensity profiles (Fig. S1). A hierarchical clustering of the complete dataset showed consistency among the biological replicates and, without exception, technical replicates clustered together suggesting that dye and array effects are smaller than the biological effects. Sample relations further indicated that tissue source is likely a more prominent factor influencing the overall transcriptome profile than infection status.

Hemocyte-enriched transcriptome in relation to systemic bacterial challenge. We assessed differential transcript abundance between circulating hemocytes and the remaining carcass in naïve, *E. coli*-challenged and *M. luteus*-challenged mosquitoes. In addition, we compared *E. coli*- and *M. luteus*-challenged hemocyte and carcass transcriptomes to their corresponding naïves to derive infection response profiles for each challenge and tissue type (Fig. 1A). The expression ratios from these comparisons were tested for statistical significance using LIMMA’s moderated t-statistics (Smyth, 2004). In total, 3,324 transcripts displayed significant differences in their abundance ($P < 0.01$ and fold change > 2) in at least one of the seven pair-wise comparisons. To

obtain a global view of the relationship among expression ratio profiles and partitioning of the transcripts in accordance with their tissue enrichment and infection-responsiveness, a two-way hierarchical clustering was performed on the differentially abundant transcripts (Fig. 1B). The resulting heat map indicates that the hemocyte-enriched transcriptome is dynamic and context-dependent. Responses to bacterial challenge in hemocytes and the carcass share common transcriptional patterns, but each also involves unique transcriptional changes, altering the tissue-enriched transcriptome in infected animals relative to the naïve. One implication of this finding is that transcripts specifically or preferentially induced in hemocytes during immune activation may not necessarily be tissue-enriched in naïve animals. Furthermore, the topology of the column dendrogram indicates that the hemocyte-enriched transcriptome profiles (i.e., expression ratios from tissue comparisons) in *E. coli*- and *M. luteus*-challenged mosquitoes are overall more similar to each other than either is to the naïve profile, suggesting common infection-associated transcriptional changes independent of the bacterial species used for immune challenge. However, because our hemocoelic infection model involved septic injury via injection, it is conceivable that some, if not most, of the shared transcriptional responses represent wound-healing processes.

To examine the distinctive features of the hemocyte-enriched transcriptome, we evaluated the statistical association between specific gene ontology (GO) terms and high tissue enrichment ratios using the gene score resampling (GSR) method in ErmineJ (Lee et al., 2005). One advantage of this permutation-based approach, making it well suited for the study, is that it takes into account all measured enrichment ratios without arbitrary thresholding. In addition to complementing observations made at the level of individual transcripts by highlighting overrepresented gene groups and functional categories, this analysis captures the compositional

changes in a hemocyte-enriched transcriptome following microbial challenge. Overall, our analyses identified 54 GO terms overrepresented among the hemocyte-enriched transcripts ($P < 0.01$; Fig. 2), which can be categorized into three broad groups: (1) GO terms that are significant in both naïve and the bacteria-challenged mosquitoes, (2) GO terms that are significant in the naïve mosquitoes but not in either of the bacteria-challenged mosquitoes, and (3) GO terms that are significant only in *E. coli*- or *M. luteus*-challenged mosquitoes.

Considering the rapid initiation of phagocytosis and melanization reactions upon microbial recognition, it is likely that the molecular components involved in these processes are already in place in naïve mosquitoes. Constitutively-expressed, but highly hemocyte-enriched factors and structural components are particularly interesting in this regard. Members of the prophenoloxidase (PPO), fibrinogen and fibronectin related protein (FREP) and class C scavenger receptor (SCRC) families are among the most highly hemocyte-enriched transcripts in both naïve and bacteria-challenged mosquitoes (Dataset S2). Similar to PPOs and FREPs, carboxylesterase (AAEL004022-RA) and apolipoprotein D (AAEL002520-RA) also rank high when genes are sorted by enrichment ratios. Examination of GO categories reveals oxygen transporter, extracellular matrix, scavenger receptor and lipase activities displaying particularly significant association with tissue enrichment in hemocytes across infection status. These observations recapitulate the significance of hemocytes as the primary source of pattern recognition molecules and components of the PPO cascade. Genes associated with lipid metabolism are noteworthy in our ranked lists considering their potential involvement in phagocytosis, hemolymph clotting and coagulation (Theopold et al., 2004; Yeung and Grinstein, 2007). Cysteine-type endopeptidase and endopeptidase regulator activities, on the contrary, are significant in naïve mosquitoes but not in either of the bacteria-challenged mosquitoes. GO

categories significant only in infected animals include integrin complex, integrin-mediated signaling pathway, vesicle-mediated transport, and proteasome complex. These are likely implicated in bacterial phagocytosis, cytoskeletal remodeling and degradation of ingested bacteria following phagocytosis (Hillyer et al., 2003b; Moita et al., 2006; Stuart et al., 2007).

An important limitation of ratio-based analyses using microarray data is the inherent inability to explicitly discern tissue-specific expression from tissue-enriched expression, which is further complicated by residual hemocytes in carcass samples. As a consequence, our data cannot be used to confirmatively establish tissue-specific expression of individual transcripts. Nonetheless, it remains reasonable to expect tissue-specific transcripts to display higher enrichment ratios than other transcripts in relative terms. Consistent with this idea is the observation that previously characterized hemocyte-specific transcripts, such as those encoding PPOs, rank particularly high in tissue enrichment as measured by our microarray approach. In addition, there is an apparent trend that, among the top-ranked transcripts, signal intensities in the carcass samples approach background levels, possibly reflecting their low abundance in most tissue types other than hemocytes (Fig. S2). Taken together, it is likely that our ranked lists capture biologically meaningful information concerning the characteristic features of the hemocyte transcriptome, and the top ranked transcripts represent viable candidates for future exploration of hemocyte-specific biomarkers.

Utilizing tissue-enriched expression profiles to elucidate hemocyte infection responses. To date, transcriptomic analyses on hemocyte infection responses have mostly taken a one-dimensional view by focusing on the differences between hemocyte samples collected under various infection conditions. This approach, while effective in characterizing transcriptome changes within hemocytes, often failed to contextualize the findings in relation to changes

occurring in other immune-responsive tissues. The lack of information regarding tissue specificity during infection response hinders assessment of whether an observed transcriptional change is likely to be distinctive of hemocytes or shared by multiple tissue types. In order to develop deeper insight into the transcriptome changes underlying hemocyte-specific processes, we took an integrative approach combining infection response profiles with tissue enrichment ratios. A two-dimensional representation of expression ratios, in which the infection response profiles are resolved along a second axis of post-infection tissue enrichment ratios, provides an insightful overview of hemocyte transcriptome dynamics (Fig. S3; Fig. 3). The distribution of genes on these Cartesian planes exposes the extent of heterogeneity in tissue enrichment among different infection responsive genes. For example, C-type lectins, CTLMA13 (AAEL011621-RA) and CTLMA14 (AAEL014382-RA) display similar levels of increase in hemocytes upon *E. coli* infection. However, their post-infection tissue enrichment ratios differ substantially. CTLMA13 is significantly enriched in hemocytes relative to the carcass, while CTLMA14 shows no enrichment, suggesting that transcriptional induction of CTLMA13 is likely more restricted to hemocytes.

Using our integrative approach that combines infection response profiles with tissue enrichment ratios, we investigated further the transcriptional characteristics of immune-related gene families in *A. aegypti* (Waterhouse et al., 2007). The pattern of spatial segregation between distinct gene families and the level of variation among family members, bring to light the immunological characteristics of the hemocyte molecular repertoire within the currently understood framework of insect immunity (Fig. 3).

Members of the gene families implicated in pattern recognition form a distinct group displaying strong tissue enrichment in hemocytes. Both cell surface receptors and secreted

proteins are represented, and many are constitutively expressed. Of particular interest is the FREP family, a phylogenetically conserved pattern recognition receptor that has undergone mosquito lineage-specific gene expansion with respect to *D. melanogaster* (Waterhouse et al., 2007). The present study indicates that at least 10 out of 37 FREPs in *A. aegypti* display a hemocyte-enriched expression to a varying degree (FREPs 3, 5, 10, 12, 16, 23, 26, 28, 35 and 38). Microbial binding of FREPs has been demonstrated in *Armigeres subalbatus* (Wang et al., 2005) and *A. gambiae* (Dong and Dimopoulos, 2009). FREP10, in particular, is a one-to-one ortholog of the *A. gambiae* FREP13 (FBN9) whose transcript knockdown alters *Plasmodium* infectivity as well as the host survival rate post bacterial challenge (Dong and Dimopoulos, 2009). Class C scavenger receptors, originally characterized in *D. melanogaster* hemocytes, recognize both gram-positive and -negative bacteria and a broad range of polyanionic ligands (Pearson et al., 1995; Ramet et al., 2001). SCRCs in *A. aegypti* contain two complement-control protein domains and one transmembrane domain like their homologs in *D. melanogaster*. These likely function as pattern recognition receptors for phagocytosis given the strong hemocyte-enriched expression. AAEL000636 and AAEL007967, homologs of the phagocytic receptors with EGF domains (eater and Nimrod C1) also display highly hemocyte-enriched expression patterns similar to that of the SCRCs (Kocks et al., 2005; Kurucz et al., 2007).

Of the 8 peptidoglycan recognition proteins (PGRPs) annotated in *A. aegypti*, our data indicate that PGRPLB and PGRPS1 (PGRP-SA in *D. melanogaster*) display substantial increases in transcript abundance in hemocytes, both after *E. coli*- and *M. luteus*-challenge. In *Drosophila*, the former degrades gram-negative bacteria peptidoglycan with its amidase activity in a negative feedback regulation of the Imd pathway (Zaidman-Remy et al., 2006), whereas the latter is involved in the recognition of Gram-positive bacteria and activation of the Toll pathway (Michel

et al., 2001). Interestingly, our tissue enrichment data suggest that the induction of PGRPLB is more specific to hemocytes.

Members of intracellular immune signaling pathways, such as Toll and IMD, do not show particular enrichment in hemocytes, and are likely expressed in diverse (immune-responsive) tissues. Because signal transduction is mostly mediated by protein interactions, transcriptome data allow only indirect monitoring of the activity of these immune pathways via transcriptional changes of downstream target genes. In *Drosophila* larvae, hemocytes express the Toll ligand spätzle (SPZ) in a tissue-specific manner after bacterial challenge (Irving et al., 2005) and its knockdown in hemocytes blocks the expression of drosomycin gene in the fat body (Shia et al., 2009). Our data indicate a moderate induction of SPZ3A in both hemocytes and the carcass after bacteria challenge, and a carcass-enriched constitutive expression of SPZ1A. The latter has been shown previously to be induced in the fat body following fungal challenge (Shin et al., 2006). A large number of CLIP-domain serine proteases and serine protease inhibitors (serpins) are encoded in the genome of *A. aegypti*, many of which are induced upon infection. The present study indicates that only a few specific members of these families (e.g., CLIPB27 and CLIPD1) display distinct hemocyte-enriched expression patterns, suggesting their possible involvement in hemocyte-mediated immune mechanisms.

Tissue enrichment ratios of the AMP transcripts after bacterial challenge indicate that transcriptional induction of AMPs is not restricted to hemocytes. Moreover, the overall direction and magnitude of transcriptional change are highly similar between hemocytes and carcass samples. One notable exception to this pattern is a transcript encoding attacin that shows strong preferential increase in the carcass relative to hemocytes. A gambicin transcript also displays a similar pattern, but at a more moderate level. In our previous study, we reported dramatic

increase in AMP transcripts in hemocytes upon bacterial challenge (Bartholomay et al., 2007). The current data suggest that, despite the comparatively large magnitude of induction, none of the AMP responses are unique to hemocytes. Although both studies highlight the contribution of hemocytes to AMP production during systemic infection, lack of tissue specificity suggests that AMP responses may not be central in directly mediating cellular function unique to hemocytes. This expression pattern is in sharp contrast to that of PPOs, the majority of which display strong hemocyte-enriched expression in both naïve and infected animals. Overall, these observations conform to earlier findings in mosquitoes and other insect species (Haine et al., 2008; Jiang et al., 2010; Lemaitre and Hoffmann, 2007; Yassine and Osta, 2010).

In addition to transcripts showing tissue enrichment following bacterial challenge (i.e., restricted expression in hemocytes), it is of interest to examine transcripts showing a disproportionate response in hemocytes relative to that observed in the carcass. Direct comparison of infection responses in hemocytes versus those in the carcass (Fig. S3) can identify transcriptional changes that are more pronounced in hemocytes. We compiled a list of genes showing infection-responsive transcriptional induction that is more restricted to or more pronounced in hemocytes; out of 625 transcripts that showed a greater than 2-fold increase ($P < 0.01$) in hemocytes at 24 hours post *E. coli*-challenge, 70 transcripts displayed a greater than 3-fold post-infection tissue enrichment ($P < 0.01$) or a magnitude of transcriptional response at least 3 times greater than that in the carcass ($P < 0.01$). *M. luteus*-challenge increased abundance levels of 534 hemocyte transcripts to more than 2-fold ($P < 0.01$) at 24 hours post immune challenge, and among these, 109 transcripts satisfied the above mentioned criteria of tissue enrichment or differential induction in hemocytes, and 47 transcripts were common between the two gene lists (*E. coli* and *M. luteus* challenge).

Our compiled list contains a number of cytoskeletal components, adhesion molecules, cell-cycle associated proteins, calmodulins and a nitric oxide synthase, in addition to the members of the canonical immune gene families (Table 1; Dataset S3). The most noticeable group, however, is the small heat shock proteins (HSPs) with alpha-crystallin domain. These likely function as molecular chaperones alleviating malformed protein aggregation during cellular stress. In vertebrate systems, increasing evidence suggests that extracellular and membrane-bound HSPs display immunomodulatory properties (Henderson and Pockley, 2010), and it remains to be investigated whether any of the hemocyte-derived HSPs contribute to systemic immune responses beyond maintaining the cellular homeostasis of phagocytes. It may be possible that these HSPs constitute “danger” signals for the host to modulate the level of immune activation based on the balance between nonself elicitors and endogenous danger signals (Lazzaro and Rolff, 2011).

Hemocyte transcriptome responses to *E. coli* vs. *M. luteus*. We examined transcripts showing differential response to *E. coli* as compared to *M. luteus* challenge. Because we based our analysis of differential transcript abundance on linear models (Smyth, 2004), it was possible to directly contrast a pair of expression ratio profiles (i.e., “ratio of ratios”), rather than making a comparison at the level of gene lists. Interestingly, the number of transcripts with a differential response ($P < 0.01$ and fold change > 3) was greater in hemocytes than in carcass (51 and 16, respectively; Fig. S4; Dataset S2). A closer look at these transcripts suggests that hemocyte response differs between the two immune elicitors in three distinct ways: (1) elevation of HSP transcript abundance is more pronounced following *E. coli* challenge, and (2) PPO and (3) cell-cycle related transcripts are present at higher levels after *M. luteus* challenge. These observations are consistent with our GO analysis indicating significant overrepresentation of the GO terms,

response to temperature stimulus (GO:0009266) and DNA replication (GO:0006260), among hemocyte-enriched transcripts after *E. coli* and *M. luteus* challenge, respectively (Fig. 2). The transcriptomic differences are intriguing, especially the differences in PPO levels, because *E. coli* predominantly elicits phagocytosis, whereas *M. luteus* induces a melanization response (but not in a mutually exclusive manner) (Hillyer et al., 2003a, b). In addition, the variation in HSP levels could be a reflection of the different phagocytic loads in hemocytes. However, the inferences we can make on the molecular basis of the differential hemocyte response remain limited, not only because of the shortcomings of our experimental design (e.g., temporal resolution, dosage coverage, etc.) but also because the initiation of phagocytosis and melanization responses likely depend primarily on protein level interactions that cannot be examined using a transcriptomic approach.

Hemocyte-enriched transcripts conserved among *A. aegypti*, *A. gambiae* and *D. melanogaster*. A meta-analysis was performed on published data to investigate the hemocyte-enriched transcripts conserved among *A. aegypti*, *A. gambiae* and *D. melanogaster*. Transcripts enriched in hemocytes in adult female *A. gambiae* (relative to the carcass) and in larval *D. melanogaster* (relative to the whole animal) were obtained from previous microarray studies (Irving et al., 2005; Pinto et al., 2009). A three-way comparison was made at the level of orthologous groups delineated by eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) database (Muller et al., 2010). An orthologous group was considered to retain hemocyte-enriched expression pattern if there was at least one ortholog from each species that showed tissue enrichment. Transcripts with the highest enrichment ratio were mapped to genes and orthologous groups when alternatively spliced transcripts or in-paralogs were present. Relative ranking of tissue enrichment ratios within each study then was used as a common basis

for comparison to alleviate the effects of study specific differences in transcriptome quantification. The most highly hemocyte enriched transcripts (with an orthologous group assignment) from each species were compared to identify orthologous groups of genes that retained hemocyte enriched expression patterns among these dipterans. A Venn diagram analysis including the top 500 transcripts from each species indicated that 117 and 87 *A.aegypti* transcripts had orthologs with conserved hemocyte-enriched expression in *A. gambiae* and *D. melanogaster*, respectively, suggesting a rather low level of overall conservation (Fig. 4). This observation is in agreement with previous comparisons in which the *A. aegypti* hemocyte transcriptome was represented by EST datasets (Baton et al., 2009; Pinto et al., 2009). However, considering biological and methodological factors that confounded our analysis (i.e., variation in developmental stages, hemocyte collection methods, mRNA amplification methods, microarray platforms, normalization strategies, etc.) the observed level of differences in hemocyte molecular repertoires among these dipterans may not be due solely to their phylogenetic divergence, making it difficult to analyze the distinguishing transcriptional characteristics of each species with high confidence. In particular, stringent data filtering performed in each study could have resulted in an increase in false-negatives that might have inflated the number of hemocyte-enriched transcripts that appear unique to each species. On the other hand, the current meta-analysis revealed 38 orthologous groups with hemocyte-enriched expression conserved in all three species. We further determined genes showing the most consistent pattern of high tissue enrichment across species using the geometric mean of the relative ranks (Table 2; Dataset S3). Given the nature of our meta-analytic approach, these likely represent true positives, albeit a partial list. Prophenoloxidase (AAEL011763; AGAP006258; CG5779), extracellular ligand-gated ion channel (AAEL004958; AGAP010580; CG6698), Wnt-protein binding (AAEL007967;

AGAP009762; CG6124: eater/Nimrod homologs) and C-type scavenger receptor (AAEL006361; AGAP011974; CG4099) were at the top of this ranked list. Several serine proteases, structural proteins such as laminin (implicated in cell migration and adhesion), and calcium ion binding proteins also were represented. Phylogenetically conserved tissue-enriched expression could imply functional constraints (Piro et al., 2011), possibly suggesting that these genes are indispensable for hemocyte function. Because our approach effectively captured genes previously described in association with hemocyte-specific processes (e.g., PPOs and phagocytic receptors), top ranked orthologous groups with uncharacterized function strongly merit further experimental investigation in relation to hemocyte biology.

SUPPORTING INFORMATION

See the following online documents at: <http://dx.doi.org/10.1016/j.ibmb.2012.06.005>

Figure S1. Hierarchical clustering of single-channel intensity profiles resulting from biological and technical dye-swap replicates of the 2×3 experimental conditions.

Figure S2. Microarray signal intensity distribution for the top 100 hemocyte-enriched transcripts from *Aedes aegypti* (naïve condition). Signal intensities from hemocyte and carcass samples are colored blue and grey, respectively. The trend lines were fitted by moving average with a period of 5.

Figure S3. Hemocyte response profiles were compared to tissue enrichment and carcass response profiles for *Aedes aegypti*. The axes indicate expression ratios. Pairing of expression

ratios further resolved the hemocyte response profiles, highlighting transcriptional changes distinctive of hemocytes. Transcripts enriched in hemocytes after bacterial challenge or transcripts displaying a differential response in hemocytes relative to the carcass are colored blue ($P < 0.01$ and fold change > 3).

Figure S4. *Aedes aegypti* hemocyte and carcass response to *Escherichia coli*- and *Micrococcus luteus*-challenge. The axes indicate expression ratios. Transcripts displaying a differential response to *E. coli*- relative to *M. luteus*-challenge are colored blue ($P < 0.01$ and fold difference > 3).

Dataset S1. qPCR validation results and primer sequences.

Dataset S2. Log expression ratios (logFC) and p-values from differential transcript abundance analysis using LIMMA.

Dataset S3. Complete Table 1 and 2.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge support from NIH grants AI 067698 and AI 019769.

REFERENCES

Abraham, E.G., Pinto, S.B., Ghosh, A., Vanlandingham, D.L., Budd, A., Higgs, S., Kafatos, F.C., Jacobs-Lorena, M., Michel, K., 2005. An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. *Proc Natl Acad Sci U S A* 102, 16327-16332.

- Antonova, Y., Alvarez, K.S., Kim, Y.J., Kokoza, V., Raikhel, A.S., 2009. The role of NF-kappaB factor REL2 in the *Aedes aegypti* immune response. *Insect Biochem Mol Biol* 39, 303-314.
- Bartholomay, L.C., Mayhew, G.F., Fuchs, J.F., Rocheleau, T.A., Erickson, S.M., Aliota, M.T., Christensen, B.M., 2007. Profiling infection responses in the haemocytes of the mosquito, *Aedes aegypti*. *Insect Mol Biol* 16, 761-776.
- Baton, L.A., Robertson, A., Warr, E., Strand, M.R., Dimopoulos, G., 2009. Genome-wide transcriptomic profiling of *Anopheles gambiae* hemocytes reveals pathogen-specific signatures upon bacterial challenge and *Plasmodium berghei* infection. *BMC Genomics* 10, 257.
- Berntsens, B.T., Christensen, B.M., 1990. *Dirofilaria immitis*: effect on hemolymph polypeptide synthesis in *Aedes aegypti* during melanotic encapsulation reactions against microfilariae. *Exp Parasitol* 71, 406-414.
- Blandin, S.A., Levashina, E.A., 2007. Phagocytosis in mosquito immune responses. *Immunol Rev* 219, 8-16.
- Camon, E., Magrane, M., Barrell, D., Lee, V., Dimmer, E., Maslen, J., Binns, D., Harte, N., Lopez, R., Apweiler, R., 2004. The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res* 32, D262-266.
- Castillo, J.C., Robertson, A.E., Strand, M.R., 2006. Characterization of hemocytes from the mosquitoes *Anopheles gambiae* and *Aedes aegypti*. *Insect Biochem Mol Biol* 36, 891-903.
- Chen, C.C., Laurence, B.R., 1987. In vitro study on humoral encapsulation of microfilariae: establishment of technique and description of reactions. *Int J Parasitol* 17, 781-787.
- Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162, 156 - 159.
- Christensen, B., Li, J., Chen, C., Nappi, A., 2005. Melanization immune responses in mosquito vectors. *Trends Parasitol* 21, 192 - 199.
- Christensen, B.M., Sutherland, D.R., 1984. *Brugia pahangi*: exsheathment and midgut penetration in *Aedes aegypti*. *Trans Am Microsc Soc* 103, 423-433.
- Chun, J., McMaster, J., Han, Y., Schwartz, A., Paskewitz, S.M., 2000. Two-dimensional gel analysis of haemolymph proteins from *Plasmodium*-melanizing and -non-melanizing strains of *Anopheles gambiae*. *Insect Mol Biol* 9, 39-45.
- Dong, Y., Dimopoulos, G., 2009. *Anopheles* fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. *J Biol Chem* 284, 9835-9844.

Gillis, J., Mistry, M., Pavlidis, P., 2010. Gene function analysis in complex data sets using ErmineJ. *Nat Protoc* 5, 1148-1159.

Glenn, J.D., King, J.G., Hillyer, J.F., 2010. Structural mechanics of the mosquito heart and its function in bidirectional hemolymph transport. *J Exp Biol* 213, 541-550.

Haine, E.R., Moret, Y., Siva-Jothy, M.T., Rolff, J., 2008. Antimicrobial defense and persistent infection in insects. *Science* 322, 1257-1259.

Henderson, B., Pockley, A.G., 2010. Molecular chaperones and protein-folding catalysts as intercellular signaling regulators in immunity and inflammation. *J Leukoc Biol* 88, 445-462.

Hillyer, J.F., 2010. Mosquito immunity. *Adv Exp Med Biol* 708, 218-238.

Hillyer, J.F., Schmidt, S.L., Christensen, B.M., 2003a. Hemocyte-mediated phagocytosis and melanization in the mosquito *Armigeres subalbatus* following immune challenge by bacteria. *Cell Tissue Res* 313, 117-127.

Hillyer, J.F., Schmidt, S.L., Christensen, B.M., 2003b. Rapid phagocytosis and melanization of bacteria and *Plasmodium* sporozoites by hemocytes of the mosquito *Aedes aegypti*. *J Parasitol* 89, 62-69.

Hillyer, J.F., Schmidt, S.L., Christensen, B.M., 2004. The antibacterial innate immune response by the mosquito *Aedes aegypti* is mediated by hemocytes and independent of Gram type and pathogenicity. *Microbes Infec* 6, 448-459.

Irving, P., Ubeda, J.M., Doucet, D., Troxler, L., Lagueux, M., Zachary, D., Hoffmann, J.A., Hetru, C., Meister, M., 2005. New insights into *Drosophila* larval haemocyte functions through genome-wide analysis. *Cell Microbiol* 7, 335-350.

Jiang, H., Vilcinskas, A., Kanost, M.R., 2010. Immunity in lepidopteran insects. *Adv Exp Med Biol* 708, 181-204.

Kocks, C., Cho, J.H., Nehme, N., Ulvila, J., Pearson, A.M., Meister, M., Strom, C., Conto, S.L., Hetru, C., Stuart, L.M., Stehle, T., Hoffmann, J.A., Reichhart, J.M., Ferrandon, D., Ramet, M., Ezekowitz, R.A., 2005. Eater, a transmembrane protein mediating phagocytosis of bacterial pathogens in *Drosophila*. *Cell* 123, 335-346.

Kurucz, E., Markus, R., Zsamboki, J., Folkl-Medzihradzky, K., Darula, Z., Vilmos, P., Udvardy, A., Krausz, I., Lukacsovich, T., Gateff, E., Zettervall, C.J., Hultmark, D., Ando, I., 2007. Nimrod, a putative phagocytosis receptor with EGF repeats in *Drosophila* plasmatocytes. *Curr Biol* 17, 649-654.

Lawson, D., Arensburger, P., Atkinson, P., Besansky, N.J., Bruggner, R.V., Butler, R., Campbell, K.S., Christophides, G.K., Christley, S., Dialynas, E., Hammond, M., Hill, C.A., Konopinski, N.,

Lobo, N.F., MacCallum, R.M., Madey, G., Megy, K., Meyer, J., Redmond, S., Severson, D.W., Stinson, E.O., Topalis, P., Birney, E., Gelbart, W.M., Kafatos, F.C., Louis, C., Collins, F.H., 2009. VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res* 37, D583-587.

Lazzaro, B.P., Rolff, J., 2011. Danger, Microbes, and Homeostasis. *Science* 332, 43-44.

Lee, H.K., Braynen, W., Keshav, K., Pavlidis, P., 2005. ErmineJ: tool for functional analysis of gene expression data sets. *BMC Bioinformatics* 6, 269.

Lemaitre, B., Hoffmann, J., 2007. The host defense of *Drosophila melanogaster*. *Annu Rev Immunol* 25, 697-743.

Michel, T., Reichhart, J.-M., Hoffmann, J.A., Royet, J., 2001. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. *Nature* 414, 756-759.

Moita, L.F., Vriend, G., Mahairaki, V., Louis, C., Kafatos, F.C., 2006. Integrins of *Anopheles gambiae* and a putative role of a new beta integrin, BINT2, in phagocytosis of *E. coli*. *Insect Biochem Mol Biol* 36, 282-290.

Muller, J., Szklarczyk, D., Julien, P., Letunic, I., Roth, A., Kuhn, M., Powell, S., von Mering, C., Doerks, T., Jensen, L.J., Bork, P., 2010. eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res* 38, D190-195.

Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., Tu, Z.J., Loftus, B., Xi, Z., Megy, K., Grabherr, M., Ren, Q., Zdobnov, E.M., Lobo, N.F., Campbell, K.S., Brown, S.E., Bonaldo, M.F., Zhu, J., Sinkins, S.P., Hogenkamp, D.G., Amedeo, P., Arensburger, P., Atkinson, P.W., Bidwell, S., Biedler, J., Birney, E., Bruggner, R.V., Costas, J., Coy, M.R., Crabtree, J., Crawford, M., Debruyn, B., Decaprio, D., Eiglmeier, K., Eisenstadt, E., El-Dorry, H., Gelbart, W.M., Gomes, S.L., Hammond, M., Hannick, L.I., Hogan, J.R., Holmes, M.H., Jaffe, D., Johnston, J.S., Kennedy, R.C., Koo, H., Kravitz, S., Kriventseva, E.V., Kulp, D., Labutti, K., Lee, E., Li, S., Lovin, D.D., Mao, C., Mauceli, E., Menck, C.F., Miller, J.R., Montgomery, P., Mori, A., Nascimento, A.L., Naveira, H.F., Nusbaum, C., O'Leary, S., Orvis, J., Pertea, M., Quesneville, H., Reidenbach, K.R., Rogers, Y.H., Roth, C.W., Schneider, J.R., Schatz, M., Shumway, M., Stanke, M., Stinson, E.O., Tubio, J.M., Vanzee, J.P., Verjovski-Almeida, S., Werner, D., White, O., Wyder, S., Zeng, Q., Zhao, Q., Zhao, Y., Hill, C.A., Raikhel, A.S., Soares, M.B., Knudson, D.L., Lee, N.H., Galagan, J., Salzberg, S.L., Paulsen, I.T., Dimopoulos, G., Collins, F.H., Birren, B., Fraser-Liggett, C.M., Severson, D.W., 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316, 1718-1723.

Pearson, A., Lux, A., Krieger, M., 1995. Expression cloning of dSR-CI, a class C macrophage-specific scavenger receptor from *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 92, 4056-4060.

Peppel, K., Baglioni, C., 1990. A simple and fast method to extract RNA from tissue culture cells. *Biotechniques* 9, 711-713.

Pinto, S.B., Lombardo, F., Koutsos, A.C., Waterhouse, R.M., McKay, K., An, C., Ramakrishnan, C., Kafatos, F.C., Michel, K., 2009. Discovery of *Plasmodium* modulators by genome-wide analysis of circulating hemocytes in *Anopheles gambiae*. *Proc Natl Acad Sci U S A* 106, 21270-21275.

Piro, R.M., Ala, U., Molineris, I., Grassi, E., Bracco, C., Perego, G.P., Provero, P., Di Cunto, F., 2011. An atlas of tissue-specific conserved coexpression for functional annotation and disease gene prediction. *Eur J Hum Genet* 19, 1173-1180.

Ramet, M., Pearson, A., Manfrulli, P., Li, X., Koziel, H., Gobel, V., Chung, E., Krieger, M., Ezekowitz, R.A., 2001. *Drosophila* scavenger receptor CI is a pattern recognition receptor for bacteria. *Immunity* 15, 1027-1038.

Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A., Smyth, G.K., 2007. A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 23, 2700-2707.

Schneider, D., 2009. Physiological integration of innate immunity., in: Rolff, J., Reynolds, S.E. (Eds.), *Insect Infection and Immunity: Evolution, Ecology, and Mechanisms*. Oxford University Press, pp. 106-116.

Shia, A.K., Glittenberg, M., Thompson, G., Weber, A.N., Reichhart, J.M., Ligoxygakis, P., 2009. Toll-dependent antimicrobial responses in *Drosophila* larval fat body require Spatzle secreted by haemocytes. *J Cell Sci* 122, 4505-4515.

Shin, S.W., Bian, G., Raikhel, A.S., 2006. A toll receptor and a cytokine, Toll5A and Spz1C, are involved in toll antifungal immune signaling in the mosquito *Aedes aegypti*. *J Biol Chem* 281, 39388-39395.

Shin, S.W., Kokoza, V., Bian, G., Cheon, H.M., Kim, Y.J., Raikhel, A.S., 2005. REL1, a homologue of *Drosophila* dorsal, regulates toll antifungal immune pathway in the female mosquito *Aedes aegypti*. *J Biol Chem* 280, 16499-16507.

Smyth, G.K., 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3, Article3.

Smyth, G.K., Speed, T., 2003. Normalization of cDNA microarray data. *Methods* 31, 265-273.

Strand, M.R., 2008. Insect hemocytes and their role in immunity, in: Beckage, N.E. (Ed.), *Insect Immunology*. Academic Press, pp. 25-47.

Stuart, L.M., Boulais, J., Charriere, G.M., Hennessy, E.J., Brunet, S., Jutras, I., Goyette, G., Rondeau, C., Letarte, S., Huang, H., Ye, P., Morales, F., Kocks, C., Bader, J.S., Desjardins, M.,

Ezekowitz, R.A.B., 2007. A systems biology analysis of the *Drosophila* phagosome. *Nature* 445, 95-101.

Theopold, U., Schmidt, O., Soderhall, K., Dushay, M.S., 2004. Coagulation in arthropods: defence, wound closure and healing. *Trends Immunol* 25, 289-294.

Wang, X., Zhao, Q., Christensen, B.M., 2005. Identification and characterization of the fibrinogen-like domain of fibrinogen-related proteins in the mosquito, *Anopheles gambiae*, and the fruitfly, *Drosophila melanogaster*, genomes. *BMC Genomics* 6, 114.

Waterhouse, R.M., Kriventseva, E.V., Meister, S., Xi, Z., Alvarez, K.S., Bartholomay, L.C., Barillas-Mury, C., Bian, G., Blandin, S., Christensen, B.M., Dong, Y., Jiang, H., Kanost, M.R., Koutsos, A.C., Levashina, E.A., Li, J., Ligoxygakis, P., Maccallum, R.M., Mayhew, G.F., Mendes, A., Michel, K., Osta, M.A., Paskewitz, S., Shin, S.W., Vlachou, D., Wang, L., Wei, W., Zheng, L., Zou, Z., Severson, D.W., Raikhel, A.S., Kafatos, F.C., Dimopoulos, G., Zdobnov, E.M., Christophides, G.K., 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* 316, 1738-1743.

Yassine, H., Osta, M.A., 2010. *Anopheles gambiae* innate immunity. *Cell Microbiol* 12, 1-9.

Yeung, T., Grinstein, S., 2007. Lipid signaling and the modulation of surface charge during phagocytosis. *Immunological Reviews* 219, 17-36.

Zaidman-Remy, A., Herve, M., Poidevin, M., Pili-Floury, S., Kim, M.S., Blanot, D., Oh, B.H., Ueda, R., Mengin-Lecreulx, D., Lemaitre, B., 2006. The *Drosophila* amidase PGRP-LB modulates the immune response to bacterial infection. *Immunity* 24, 463-473.

FIGURE 1. (A) Transcriptome profile comparisons among hemocyte and carcass samples from naïve and bacteria-challenged *Aedes aegypti*. (B) Hierarchical clustering of 3,324 transcripts displaying differential abundance in at least one of the seven pair-wise comparisons ($P < 0.01$ and fold change > 2).

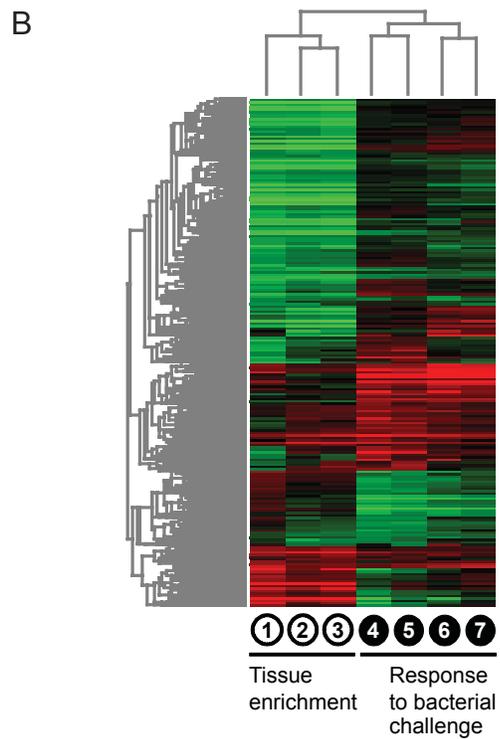
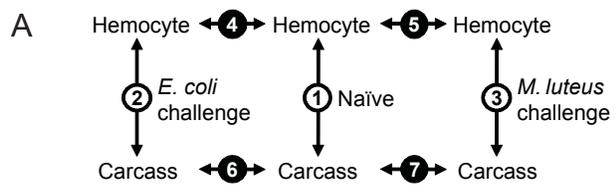
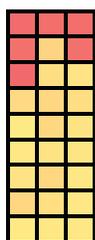
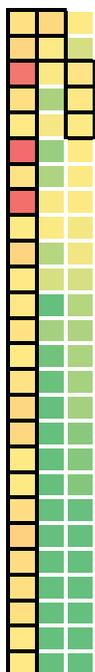


FIGURE 2. GO categories overrepresented among hemocyte-enriched transcripts from *Aedes aegypti*. N: naïve, E: *Escherichia coli*-challenged, and M: *Micrococcus luteus*-challenged conditions. Heatmap represents the degree of statistical significance as determined by gene score resampling. P-value < 0.01 is boxed, and the color gradient of red to green indicates increasing P-values.

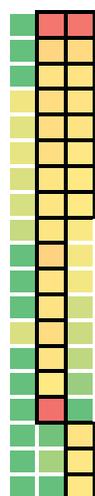
N E M GO TERMS



oxygen transporter activity
 extracellular matrix
 lipase activity
 scavenger receptor activity
 transferase activity, transferring amino-acyl groups
 protein amino acid glycosylation
 ligase activity, forming carbon-oxygen bonds
 triglyceride lipase activity
 glycoprotein metabolic process



tRNA aminoacylation for protein translation
 acyl carrier activity
 cysteine-type endopeptidase activity
 inositol or phosphatidylinositol phosphatase activity
 proteinaceous extracellular matrix
 extracellular region part
 immune system process
 endopeptidase regulator activity
 gamma-glutamyltransferase activity
 serine-type endopeptidase inhibitor activity
 carboxylesterase activity
 carboxylic acid binding
 calcium-dependent cysteine-type endopeptidase activity
 acyl-CoA dehydrogenase activity
 cysteine-type peptidase activity
 serine family amino acid metabolic process
 monosaccharide metabolic process
 carbon-nitrogen ligase activity, with glutamine as amido-N-donor
 oxidoreductase activity, acting on the CH-CH group of donors
 metalloendopeptidase activity
 hexose metabolic process
 peptidase inhibitor activity
 transferase activity, transferring hexosyl groups
 FAD or FADH2 binding
 vitamin binding
 hydrolase activity, acting on carbon-nitrogen bonds



proteasome complex
 proteasome core complex
 vesicle-mediated transport
 oligosaccharyl transferase activity
 integrin-mediated signaling pathway
 integrin complex
 protein amino acid N-linked glycosylation
 peptidyl-asparagine modification
 receptor complex
 calcium-dependent phospholipid binding
 proteasome core complex, alpha-subunit complex
 protein folding
 tRNA metabolic process
 phospholipid binding
 unfolded protein binding
 response to temperature stimulus
 endomembrane system
 endoplasmic reticulum
 DNA replication

FIGURE 3. Transcriptional response of immune-related gene families in *Aedes aegypti* hemocytes after bacterial challenge. ImmunoDB gene families were grouped into recognition, signaling and effector molecules (upper, middle and lower panels, respectively). X-axis: tissue enrichment relative to carcass after bacterial challenge. Y-axis: hemocyte response to bacterial challenge relative to naïve.

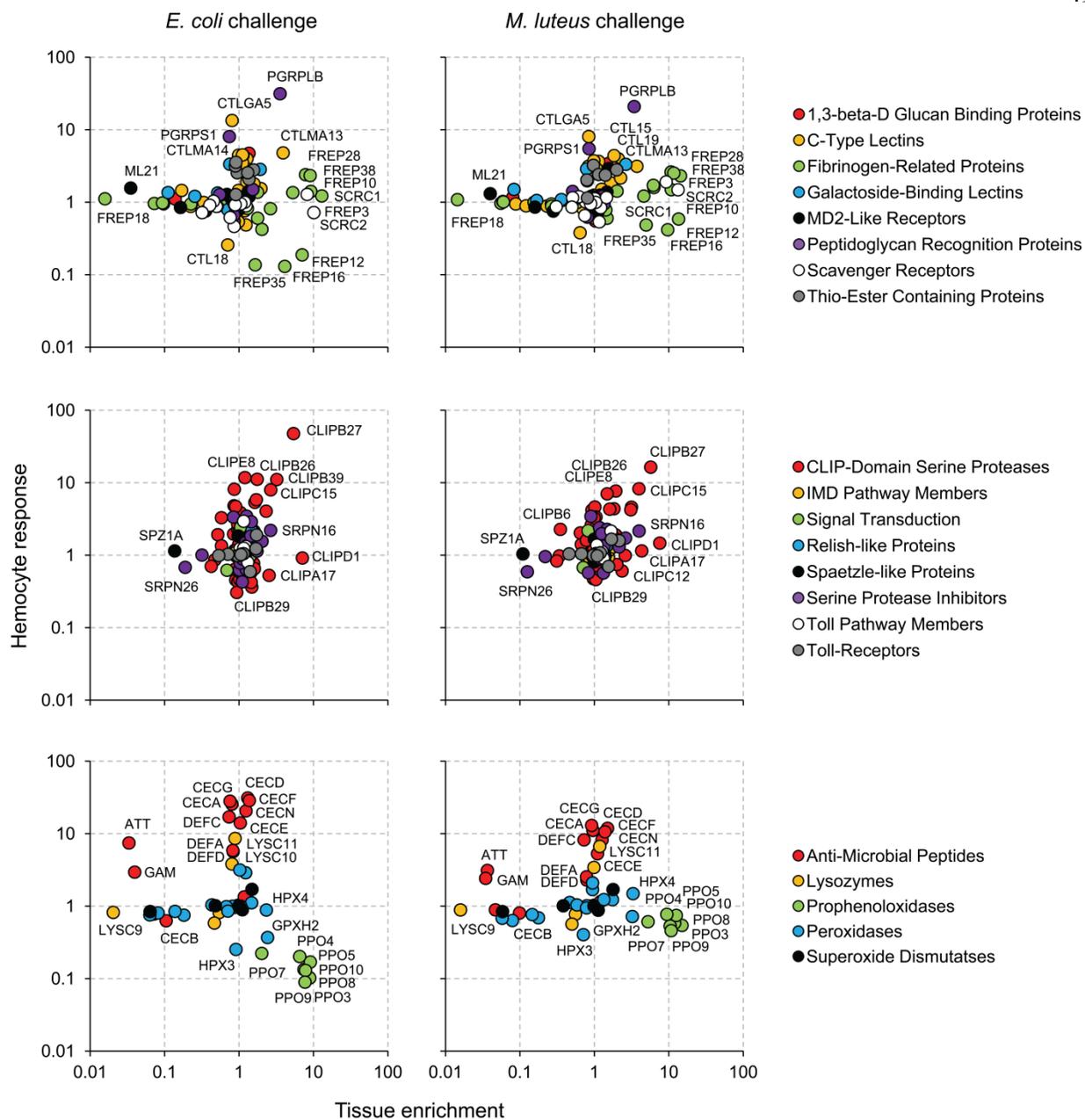


FIGURE 4. Orthologous groups displaying hemocyte-enriched expression identified through meta-analysis of microarray studies.

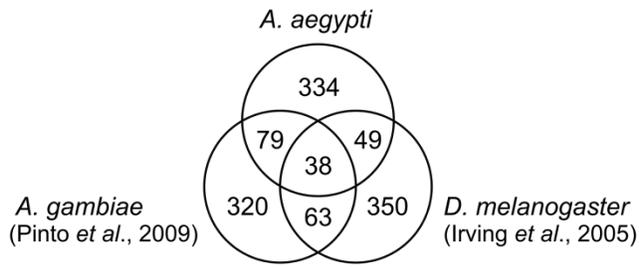


TABLE 1. Genes displaying transcriptional induction (logFC) more restricted to or more pronounced in *Aedes aegypti* hemocytes, with respect to carcass following *Escherichia coli* or *Micrococcus luteus* challenge^a.

VectorBase ID	Description	Hemocyte response ^b		Tissue enrichment ^c		Differential response ^d	
		<i>E. coli</i>	<i>M. luteus</i>	<i>E. coli</i>	<i>M. luteus</i>	<i>E. coli</i>	<i>M. luteus</i>
AAEL013350-RA	heat shock protein 26kD	6.49	3.10	3.04	1.98	2.95	1.91
AAEL013281-RA	serine-type endopeptidase	3.16	1.71	2.73	1.82	2.62	1.71
AAEL010171-RA	peptidoglycan recognition protein sb2	4.97	4.39	1.83	1.77	2.56	2.46
AAEL006479-RA	conserved hypothetical protein	3.59	2.65	2.28	1.93	2.18	1.87
AAEL009537-RA	hypothetical protein	5.79	4.35	3.98	3.81	1.65	
AAEL001241-RA	conserved hypothetical protein	3.65	2.50	3.67	2.93	2.31	
AAEL006921-RA	Calmodulin	4.51	3.47	2.56	2.52	1.63	
AAEL006533-RA	Ets domain-containing protein	3.61	2.07	2.45	1.90	1.98	
AAEL010802-RB	conserved hypothetical protein	3.02	2.01	2.41	1.82	2.03	
AAEL013349-RA	lethal(2)essential for life protein	5.46	2.12	3.11		3.17	1.61
AAEL013344-RA	lethal(2)essential for life protein	5.18	1.65	2.16		3.93	1.68
AAEL008489-RA	calcyphosine/tpp	4.20	3.04	1.86		2.76	2.20

^a see Dataset S3 for a complete list.

^b Hemocyte response to bacterial challenge relative to naive.

^c Tissue enrichment relative to carcass after bacterial challenge.

^d Differential response in hemocytes relative to carcass.

TABLE 2. Orthologous groups among *Aedes aegypti*, *Anopheles gambiae* (Pinto et al., 2009) and *Drosophila melanogaster* (Irving et al., 2005) with hemocyte-enriched expression^a.

Orthologous group ^b	Description	<i>A. aegypti</i>			<i>A. gambiae</i>			<i>D. melanogaster</i>			Geometric mean rank
		ID ^c	logFC	Rank	ID ^c	logFC	Rank	ID ^c	logFC	Rank	
inNOG07770	Prophenoloxidase	11763	3.96	1	006258	6.81	1	5779	3.75	32	3.2
inNOG10178	Extracellular ligand-gated ion channel	04958	3.25	7	010580	5.37	2	6698	5.44	3	3.5
inNOG06548	Wnt-Protein binding (Eater/Nimrod)	07967	2.65	13	009762	4.73	6	6124	5.45	2	5.4
inNOG09691	Scavenger receptor class C, type	06361	3.70	4	011974	3.85	18	4099	4.67	13	9.8
inNOG07282	Serine protease	07796	2.73	12	002422	4.26	9	9372	4.42	17	12.2
inNOG10070	Unknown function	02520	3.38	6	000790	5.17	4	4139	2.92	87	12.8
inNOG10383	CG10824-PA protein	08658	2.63	15	007045	2.87	41	4950	5.05	5	14.5
inNOG09987	Unknown function	09086	2.59	17	005820	4.24	11	6310	3.23	61	22.5
inNOG08715	Beta chain protein	02848	2.19	28	010510	2.88	40	3401	4.22	20	28.2
inNOG07005	Protein involved in regulation of embryonic development	08773	2.08	32	004993	3.15	31	10236	3.60	40	34.1
inNOG07544	Calcium ion binding protein	13656	2.16	30	000305	2.73	46	6378	3.78	29	34.2
inNOG06226	Serine-Type peptidase	05521	1.04	145	000994	4.26	10	2493	3.70	34	36.7
inNOG08907	Beta-1,3-Glucuronyltransferase	06254	1.97	39	001367	2.92	39	6207	3.13	67	46.7
inNOG05564	Laminin beta protein	03658	1.03	152	001381	2.98	37	7123	4.13	22	49.8
inNOG08874	Procollagen-Lysine protein	08099	1.41	77	001507	2.67	51	6199	3.51	43	55.3

^a see Dataset S3 for a complete list.

^b eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) database

^c AAEL00, AGAP0, and CG were omitted from *A. aegypti*, *A. gambiae*, and *D. melanogaster* gene IDs, respectively.

CHAPTER 3**A DEEP SEQUENCING APPROACH TO COMPARATIVELY ANALYZE THE
TRANSCRIPTOME OF LIFECYCLE STAGES OF THE FILARIAL WORM,
*BRUGIA MALAYI***

Choi, Y-J, Ghedin, E, Berriman, M, McQuillan, J, Holroyd, N, Mayhew, GF, Christensen, BM and Michalski, ML (2011) PLoS Negl Trop Dis 5: 13. [PMID: 22180794]

ABSTRACT

Background. Developing intervention strategies for the control of parasitic nematodes continues to be a significant challenge. Genomic and post-genomic approaches play an increasingly important role for providing fundamental molecular information about these parasites, thus enhancing basic as well as translational research. Here we report a comprehensive genome-wide survey of the developmental transcriptome of the human filarial parasite *Brugia malayi*.

Methodology/Principal Findings. Using deep sequencing, we profiled the transcriptome of eggs & embryos, immature (≤ 3 days of age) and mature microfilariae (MF), third- and fourth-stage larvae (L3 and L4), and adult male and female worms. Comparative analysis across these stages provided a detailed overview of the molecular repertoires that define and differentiate distinct lifecycle stages of the parasite. Genome-wide assessment of the overall transcriptional variability indicated that the cuticle collagen family and those implicated in molting exhibit noticeably dynamic stage-dependent patterns. Of particular interest was the identification of genes displaying sex-biased or germline-enriched profiles due to their potential involvement in reproductive processes. The study also revealed discrete transcriptional changes during larval development, namely those accompanying the maturation of MF and the L3 to L4 transition that are vital in establishing successful infection in mosquito vectors and vertebrate hosts, respectively.

Conclusions/Significance. Characterization of the transcriptional program of the parasite's lifecycle is an important step towards understanding the developmental processes required for the infectious cycle. We find that the transcriptional program has a number of stage-specific pathways activated during worm development. In addition to advancing our understanding of

transcriptome dynamics, these data will aid in the study of genome structure and organization by facilitating the identification of novel transcribed elements and splice variants.

AUTHOR SUMMARY

Lymphatic filariasis, also known as elephantiasis, is a tropical disease affecting over 120 million people worldwide. More than 40 million people live with painful, disfiguring symptoms that can cause severe debilitation and social stigma. The disease is caused by infection with thread-like filarial nematodes (roundworms) that have a complex parasitic lifecycle involving both human and mosquito hosts. In the study, the authors profiled the transcriptome (the set of genes transcribed into messenger RNA rather than all of those in the genome) of the human filarial worm *Brugia malayi* in different lifecycle stages using deep sequencing technology. The analysis revealed major transitions in RNA expression from eggs through larval stages to adults. Using statistical approaches, the authors identified groups of genes with distinct life stage dependent transcriptional patterns, with particular emphasis on genes displaying sex-biased or germline-enriched patterns and those displaying significant changes during larval development. This study presents a first comprehensive analysis of the lifecycle transcriptome of *B. malayi*, providing fundamental molecular information that should help researchers better understand parasite biology and could provide clues for the development of more effective interventions.

INTRODUCTION

Wuchereria bancrofti, *Brugia malayi* and *Brugia timori* are mosquito-borne filarial nematode parasites that cause the tropical disease lymphatic filariasis (LF). The manifestation of the disease ranges from swelling of the lymph nodes to elephantiasis and hydrocele. LF is a major

cause of clinical morbidity and disability, leading to significant psychosocial and psychosexual burden in endemic countries. *B. malayi* is the primary organism for the study of LF because it has a tractable lifecycle that can be replicated in a laboratory setting. Like other filarial nematodes it is a heteroxenous parasite alternating between arthropod vectors and vertebrate hosts. Filarial nematodes are dioecious and reproduce sexually via copulation. Inseminated adult female worms are ovoviviparous and release live larvae (microfilariae) into the lymph, where they eventually circulate in the bloodstream to be taken up by mosquitoes during blood feeding. After a microfilaria (MF) successfully penetrates the midgut of a susceptible vector, it migrates to the thoracic muscles, and develops intracellularly through two molts to achieve the developmentally arrested third-stage larva (L3) that exits the mosquito proboscis during bloodfeeding and subsequently penetrates the mammalian host. Once L3s enter the definitive host, they undergo two additional molts and mature to adults in the lymphatics.

Characterization of the transcriptional program over the complete lifecycle is necessary to clearly understand the development of the parasite and could help devise better target strategies for control. From the standpoint of possibly designing drug-based or vaccine interventions that prevent infection or curtail parasite transmission, there is particular interest in understanding the biology of the L3 to L4 transition in the mammalian host, and the reproductive biology of filarial worms. The completion of the draft genome of *B. malayi* (Ghedini et al., 2007) has ushered in the possibility to use whole-genome gene expression profiling. With that goal in mind, we used next-generation sequencing to comparatively analyze the transcriptome of seven *B. malayi* lifecycle stages: eggs & embryos, immature MF (of less than 3 days of age), mature MF, L3, L4, adult male and adult female. We find that the transcriptional program has a number

of stage-specific pathways activated during worm development and that a number of these are potential targets for drugs or vaccines.

METHODS

Ethics statement. All animal work was conducted according to relevant national and international guidelines outlined by the National Institutes of Health Office of Laboratory Animal Welfare, and was approved under UWO Institutional Animal Care and Use Protocol 0-03-0026-000246-4-6-11; and UWM Research Animal Resource Center Protocol V00846-0-10-09.

Parasites. *Brugia malayi* adults and MF were obtained from the peritoneal cavities of patently infected dark-clawed Mongolian gerbils (*Meriones uguiculatus*) by peritoneal flush with prewarmed RPMI media (Fisher Scientific, Piscataway, NJ). MF were purified by centrifugation through Ficoll-Paque® lymphocyte isolation media (Amersham Pharmacia Biotech, Piscataway, NJ), and washed in PBS three times prior to flash freezing at -80°C. Adult worms were separated by gender, washed three times in RPMI, and flash frozen. Egg and embryo preparations were made by repeated cutting of 10 female worms with a scalpel to release eggs and embryos into a small volume of cold RPMI. The sample was examined microscopically and pieces of uterine tissue were removed using watchmaker's forceps. The sample was washed three times in cold RPMI prior to flash freezing. Immature MF (≤ 3 days old) were generated and purified as previously described (Griffiths et al., 2009). L4s were isolated from gerbils 12-13 days post peritoneal infection and were processed as described for adult worms. L3s were obtained from the NIAID-NIH Filariasis Research Reagent Resource Center at University of Georgia, Athens, GA.

RNA isolation. Total RNA was isolated from the majority of samples using a previously described protocol (Griffiths et al., 2009) that combines organic extraction with Trizol LS (Invitrogen, Carlsbad, CA) and column purification (RNAqueous-Micro®, Applied Biosystems, Foster City, CA). Samples were treated with DNase I (Ambion, Austin, TX, USA) according to the manufacturer's instructions, and the absence of background DNA confirmed by using a portion of each sample in a PCR designed to amplify the *B. malayi* GPX gene [GenBank:X69128] (data not shown). Isolation of RNA from L3s often produces low yields therefore we used a modified protocol employing homogenization of tissue combined with organic extraction in RNazol (Chomczynski and Sacchi, 1987) followed by cleaning, concentration and DNase treatment using a Zymo Research RNA column (Zymo Research Corp, Orange, CA). For all samples RNA integrity was confirmed visually by agarose gel electrophoresis (data not shown) and purity and concentration determined spectrophotometrically (NanoDrop ND-1000, ThermoFisher Scientific); samples were stored at -80°C. Total RNA was lyophilized under vacuum for transport on dry ice to the Wellcome Trust Sanger Institute Genome Facility.

RNA library creation. Polyadenylated mRNA was purified from total RNA using oligo-dT dynabead selection followed by metal ion hydrolysis fragmentation with the Ambion RNA fragmentation kit. First strand synthesis, primed using random oligonucleotides, was followed by 2nd strand synthesis with RNaseH and DNAPolI to produce double-stranded cDNA using the Illumina mRNA Seq kit.. Template DNA fragments were end-repaired with T4 and Klenow DNA polymerases and blunt-ended with T4 polynucleotide kinase. A single 3' adenosine was added to the repaired ends using Klenow exo- and dATP to reduce template concatemerization and adapter dimer formation, and to increase the efficiency of adapter ligation. Adapters (containing primer sites for sequencing) were then ligated and fragments size-selected (200-

275bp) by agarose gel electrophoresis. DNA was extracted using a Qiagen gel extraction kit protocol but with dissolution of gel slices at room temperature (rather than 50°C) to avoid heat induced bias. Libraries were then amplified by PCR to enrich for properly ligated template strands, to generate enough DNA, and to add primers for flowcell surface annealing. AMPure SPRI beads were used to purify amplified templates before quantification using an Agilent Bioanalyser chip and Kapa Illumina SYBR Fast qPCR kit.

Sequencing. Libraries were denatured with 0.1M sodium hydroxide and diluted to 6pM in a hybridization buffer to allow the template strands to hybridize to adapters attached to the flowcell surface. Cluster amplification was performed on the Illumina cluster station or the Illumina cBOT using the V4 cluster generation kit following the manufacturer's protocol. A SYBRGreen QC was performed to measure cluster density and to determine whether to pass or fail the flowcell for sequencing. This was followed by linearization, blocking and hybridization of the R1 sequencing primer. The hybridized flowcells were loaded onto the Illumina Genome Analyser IIx for 54 cycles of sequencing-by-synthesis using Illumina's v4 or v5 SBS sequencing kit then, *in situ*, the linearization, blocking and hybridization step was repeated to regenerate clusters, release the 2nd strand for sequencing and to hybridize the R2 sequencing primer followed by another 54 cycles of sequencing to produce paired end reads. These steps were performed using proprietary reagents according to manufacturer's recommended protocol (<https://icom.illumina.com/>). Data were analyzed using the RTA1.6 or RTA1.8 Illumina pipeline and submitted to Array Express (<http://www.ebi.ac.uk/arrayexpress/>) under the accession number E-MTAB-811.

QC Analysis. Each lane of Illumina sequence was assessed for quality based on %GC content, average base quality and Illumina adapter contamination. To assess the quality of the lane, the

mean base quality at each base position in the read was computed over all reads from the lane. To assess %GC content of the reads a frequency distribution of values was plotted. For a single sample in a lane, a GC plot with a normal distribution around the expected GC for the organism would be expected. Any lanes containing a contamination could therefore be identified by the presence of multiple peaks in the %GC plot. To screen for adapter contamination, the sequence reads were aligned to the set of Illumina adapter sequences using BLAT v.34 with default parameters (Kent, 2002). Any reads matching these sequences were reported as being contaminated with adapter sequence.

Sequence alignment and transcript quantification. Sequence reads from each lifecycle stage were aligned to the genome assembly [GenBank:DS236884-DS264093] using TopHat v1.0.14, a splice junction mapper built upon the short read aligner Bowtie (Langmead et al., 2009; Trapnell et al., 2009). The pipeline utilized exon records in the genome annotation (Ghedini et al., 2007) to build a set of known splice junctions for each gene model, complementing its *de novo* junction mapping algorithm. Default parameters were used except for the following: minimum intron length was set to 50; minimum isoform fraction filter was disabled; closure-search, coverage-search, microexon-search and butterfly-search were enabled for maximum sensitivity. The resulting alignment files were converted to BAM format and low quality alignments with mapping quality scores less than 5 were removed before downstream analyses (Li et al., 2009b; Li et al., 2008). No replicate samples were sequenced and all data were combined per lifecycle stage. Reads aligned to exonic regions were enumerated for each gene model using the HTSeq package (v0.4.7) in Python (www-huber.embl.de/users/anders/HTSeq). Reads overlapping more than one gene model were counted as ambiguous with the mode parameter set as “union”. Following Mortazavi et al. (2008), transcript abundance estimates were computed as RPKMs

(Reads Per Kilobase of exon model per Million mapped reads) with the following modifications:

(i) a set of paired-end reads were counted as one in compiling sequence counts to represent a single sampling event and (ii) TMM (trimmed mean of M)-normalized values were used in place of the nominal library size to account for compositional biases (Robinson and Oshlack, 2010). The correction factors for TMM-normalization (i.e., the weighted trimmed mean of M values to the reference) were calculated using the Bioconductor edgeR package (Robinson et al., 2010). The weights were from the delta method on binomial data, and the library whose upper quartile is closest to the mean upper quartile was used as the reference.

Differential expression analysis. Differential expression analysis was performed in edgeR by fitting a negative binomial model to the sequence count data. Using the quantile-adjusted conditional maximum likelihood method, dispersion parameters were estimated for each gene as a measure of the overall stage-to-stage variability to facilitate between-gene comparisons. All hypothesis testing was carried out using exact test for the negative binomial distribution with a common dispersion term for all genes. P-values less than 0.01 were considered significant. Dispersion parameters were estimated directly from the count data for comparisons contrasting a single stage or two related stages relative to all other stages. For comparisons between pairs of lifecycle stages, a common dispersion value of 0.2 was used, which is equivalent to allowing within-stage variations in expression levels of up to 45%. This value was chosen based on the level of variability observed between the immature and mature MF samples. Because longer transcripts give more statistical power for detecting differential expression between samples (Oshlack and Wakefield, 2009), Gene Ontology (GO) analysis was performed using the goseq package that adjusts transcript length bias in deep sequencing data (Young et al., 2010). GO annotation was retrieved from the UniProtKB-GOA database (Camon et al., 2004), and

statistically over-represented GO terms in a given gene list were identified using the Wallenius non-central hypergeometric distribution. Hierarchical clustering analysis was performed using GeneSpring GX (Agilent Technologies). RPKM values for each gene were baseline transformed to the median of all samples, and hierarchically clustered with centroid linkage using Pearson's uncentered correlation coefficient as distance metric.

RESULTS

Transcriptome quantification by deep sequencing. In total, 104 million paired-end reads (2×54 bp) were generated from polyA-tailed mRNA using the Illumina Genome Analyser IIX (Table S1). Sequence reads were aligned to the genome assembly using TopHat (Trapnell et al., 2009), and the number of reads aligned to each gene model was summed yielding relative transcript levels for individual genes. Approximately 50% of the sequenced reads were mapped to the reference genome after low quality alignments were removed; 10% of which were aligned to genomic regions outside of the current gene models. Sequencing depth varied between the lifecycle stage libraries, affecting gene model coverage and the distribution of the read counts per gene model for each library (Figure 1 and Figure S1). Overall, in each library, 8,000-10,000 genes (equivalent to 70 to 90% of the currently annotated gene models) had 5 or more mapped reads. Sequence counts were RPKM (Reads Per Kilobase of exon model per Million mapped reads)-transformed and TMM (trimmed mean of M)-normalized to assist in the interpretation of transcript abundance comparisons between stages and genes (Mortazavi et al., 2008; Robinson and Oshlack, 2010). For statistical inferences, however, raw read counts were directly used. Further analysis of our sequence data from a genomics perspective, covering issues related to

missing, incomplete or incorrect gene models of the 2007 assembly (Ghedin et al., 2007) will be published elsewhere (in preparation).

Our sequencing libraries contained reads that map to the *Wolbachia* genome [GenBank:AE017321]. However, the study was not adequately designed such that one could quantitatively analyze these reads in a biologically meaningful way. Abundance estimates (inferred from read counts) of these transcripts most likely deviate substantially from their true *in vivo* levels. Poly-A selection directly affects the relative abundance of non-poly-A *Wolbachia* transcripts with respect to *B. malayi* transcripts. Moreover, the nature and extent of the biases introduced by oligo-dT method to the relative abundance levels among the non-poly-A species (with respect to each other) is not well understood, and one cannot assume that these biases would remain uniform among different sample preparations. Another layer of uncertainty stems from the possibility that these “*Wolbachia*” sequences were transcribed from the *B. malayi* nuclear genome rather than the endosymbiont as a consequence of the past horizontal gene transfer events, leading to a differential capture of (presumably) poly-A tailed “*Wolbachia*” transcripts of the *B. malayi* nuclear origin. However, given the incomplete draft nature of the *B. malayi* genome assembly and the inherent difficulty in mapping short reads originating from multiple loci that are similar in sequences, it remains challenging to rigorously test this hypothesis *in silico*.

Lifecycle stage dependent changes in the transcriptome. To investigate the global transcriptional differences between stages and between genes during development, a negative binomial (NB) based model (Robinson et al., 2010) was fit to sequence count data. First, the degree of between-stage differences was assessed globally using a multidimensional scaling (MDS) of all-against-all comparisons in the NB model (Figure 2). The resulting sample relations

appear consistent with the expected biological differences between the samples. The MDS plot indicates that, in relative terms, the transcriptome profiles of the immature and mature MF are more similar to each other than either is to other stages. Likewise, the eggs & embryos sample is closely related to the adult female sample, part of which consists of the germ-line cells. Interestingly, this plot also shows how different the transcriptome profiles of adult male and female worms are to each other.

Next, we made between-gene comparisons in terms of overall transcriptional variability across stages. It is generally hypothesized that while some genes are expressed constitutively, genes with specific developmental functions are expressed at specific stages. To quantify the level of transcriptional variation for each gene across the seven lifecycle stages, the NB dispersion parameters were estimated for each gene, and used as a measure of the extra-Poisson, stage-to-stage variability. Genome-wide distribution of the dispersion parameter estimates suggests that the level of transcriptional variation is not uniform across all genes (Figure S2). Although the majority of genes show low to moderate levels of variation, certain groups of genes exhibit a significantly greater level of variation. Approximately 25% of genes have NB dispersion parameter values larger than 1. After ranking by dispersion, genes were partitioned into quarters and designated as Q1 through Q4 in the order of decreasing variability.

To examine genes displaying life stage dependent transcriptional patterns in greater detail, the top 25% most variable genes according to the NB dispersion (i.e., Q1) were subjected to an unsupervised hierarchical clustering (Figure 3A). The resulting heatmap and dendrogram suggest that there are four major transcriptional patterns, each of which corresponds to an increased transcript abundance in (i) female and/or eggs & embryos, (ii) male, (iii) microfilariae, or (iv) late larval stages. The transcriptional patterns identified through the clustering analysis largely

recapitulate the sample relations revealed in the MDS plot (Figure 2). To classify genes into these broad but distinct co-expression groups in a statistically robust manner, we performed a series of exact tests for the NB distribution using raw read counts for all genes (Figure 3B). Relying solely on the “shape” of expression patterns derived from RPKM values, without considering how many reads contributed to each pattern, may lead to false-positive findings. We first identified genes preferentially transcribed during single stages by performing exact tests contrasting each individual stage relative to the mean of all other stages. The resulting gene lists were augmented by additional exact tests to include genes displaying increased transcript abundance in two (related) stages with respect to all other stages. At the level of p-value < 0.01 , mutually-exclusive, non-redundant gene lists were compiled for each group. In total, we cataloged 2,430 genes into groups with distinct life stage dependent transcriptional patterns. Comparing the gene lists to the highly variable genes in the Q1 group suggests that members of the four main expression groups account for ~80% of the top 25% most variable genes (Figure 3C). Genes that are highly variable in transcript abundance, yet are not assigned to any of the four main groups (n=563) likely display complex transcriptional patterns falling outside of the four categories. In addition, five direct pairwise comparisons were made between relevant stages to gain further insights into the transcriptomic features associated with (1) sex differences, (2) intrauterine reproductive processes, (3) MF maturation, and (4) late larval development (Figure 3D). Cross-referencing with the previously defined coexpression groups (Figure 3B) indicates that stage specificity is not homogeneous within each group of differentially transcribed genes, highlighting the complexity of the relative transcriptome differences among the lifecycle stages examined in the study. The results outlined above are described in further detail in the following sections.

Genes displaying high levels of transcriptional variation over lifecycle. We identified and compared statistically overrepresented GO terms in groups of genes that differ in their level of transcriptional variation over the lifecycle (i.e., Q1 to Q4) to investigate specific gene sets and functional categories distinctly associated with high levels of transcriptional variation (Table S2 and Figure S2). This analysis identified ‘structural constituent of cuticle’ (GO:0042302) as the most significantly overrepresented GO category among Q1 genes that exhibit high levels of between-stage transcriptional variation. Forty-six cuticle collagen genes are annotated with this GO term, and thirty-three of these have distinct lifecycle stage dependent transcriptional patterns (18 late larval, 12 female/eggs, 2 male and 1 microfilarial; Dataset S1). Additional GO terms overrepresented among Q1 genes include those related to serine type endopeptidase inhibitor (serpin), structural molecule, and kinase/phosphatase activity. By contrast, GO categories associated with protein metabolism, such as translation, protein transport and proteasome complex are significantly overrepresented among genes displaying relatively little transcriptional variation over lifecycle stages (i.e., Q2-4).

Sex-biased and germline-enriched transcriptome. Although transcript levels of 990 genes are significantly higher during larval stages, 886 and 554 genes display elevated transcript abundance in adult male, and adult female and/or eggs & embryos, respectively (Figure 3B). A direct pairwise comparison of male versus female transcriptome further identified 1,279 genes with male-biased expression and 651 genes with female-biased expression (Figure 3D). At the level of GO categories, structural molecular activity and those associated with protein phosphorylation and dephosphorylation are prominent among genes preferentially transcribed in adult male. A closer look at individual genes with male-biased expression reveals that major sperm proteins are largely responsible for driving the statistical significance of structural

molecular activity (GO:0005198) in these comparisons. By contrast, structural constituents of cuticle (collagens), transcription factor/regulator activity, nuclear receptor activity and serpin activity constitute a main theme of the overrepresented functional categories among genes preferentially transcribed in adult female and/or eggs & embryos.

In an effort to elucidate female germline-enriched transcripts and gain insight into intrauterine reproductive processes, the transcriptome profile of a library enriched for eggs and embryos was compared with that of whole adult female (Figure 3D). However, because the eggs & embryos transcriptome is inherently a subset of the adult female transcriptome, this pairwise comparison is almost subtractive in nature and is likely biased against identifying transcripts enriched in germline tissues. On the contrary, detection of female transcripts either not expressed or expressed at lower levels in eggs and embryos likely remains unaffected by this asymmetric sample relation. For this reason, we used the adult male transcriptome profile as an additional reference point to better identify genes showing a germline-enriched expression pattern. We performed a Venn diagram analysis with three datasets: (1) genes with enriched expression in adult female relative to eggs & embryos, (2) genes with enriched expression in eggs & embryos relative to adult male, and (3) genes with enriched expression in adult female and/or eggs & embryos relative to all other stages (Figure S3). We considered genes belonging to the first set to exhibit somatic tissue-enriched expression pattern, and those belonging to either of the last two sets, but excluded from the first set, to exhibit germline-enriched expression pattern. Based on these criteria, 788 and 239 genes show enriched expression in female germline and somatic tissues, respectively. GO term overrepresentation analysis indicates that functional categories, such as transcription factor activity, DNA binding, regulation of transcription and nuclear receptor activity are more frequently found among genes displaying germline-enriched

expression. On the contrary, genes implicated in chloride transport, lipid binding, and proteolysis are overrepresented among those with somatic tissue-enriched expression pattern (Table S3). Interestingly, structural constituents of cuticle (GO:0042302) is overrepresented among both genes with germline-enriched and somatic tissue-enriched expression patterns. A closer look at individual genes reveals that mutually exclusive subsets of collagens are overrepresented in each gene set.

MF, L3 and L4 transcriptome. When compared across all stages, transcript levels of 148 genes are distinctly elevated during the MF stage. Overrepresented GO terms in this group include zinc ion binding, nucleic acid binding, chitinase activity, and proteolysis (Figure 3B and Table 1). Most notably, among these are 44 genes that encode proteins with C₂H₂-type zinc finger domains. There are 195 zinc finger protein genes annotated in the *B. malayi* draft genome, some of which have high transcript levels in stages other than MF (i.e., 3 late larval, 17 male and 6 female/eggs). In a similarly biased manner, 3 out of 4 endochitinase genes identified in the current *B. malayi* genome show transcriptional increase during MF stages. Diverse classes of proteases are also represented in this gene set (e.g., cathepsin L-like proteases including *Bm-cpl-6*, papain cysteine protease family, metalloprotease I, aspartyl protease and trypsin-like protease). Direct comparison of immature and mature MF (IM and MM) indicates that 126 genes show differential transcript abundance between the two samples (Figure 3D). Many different metabolic genes are found in the IM overexpressed gene set, while the endochitinases are overrepresented in the MM.

We identified 842 genes displaying increased transcript abundance during L3 and/or L4 stages relative to other lifecycle stages (Figure 3B). Functional categories overrepresented among these genes include structural components of the cuticle, oxidoreductase activity, serpin activity, chloride transport, hedgehog receptor activity, glycogen biosynthetic process, and

proteolysis. As suggested by the last GO category, various proteases (e.g., metalloprotease, papain family peptidase, zinc carboxypeptidase family and cathepsin-like cysteine proteases, including *Bm-cpl-1,4* and 5) are prominently represented in this gene set, a pattern similarly found in the MF transcriptome. A pairwise comparison of the transcriptomes of late larval stages indicates that 342 genes have elevated transcript levels in L3s, and 155 in L4s. At the level of functional categories, cysteine-type peptidase activity (e.g., cathepsin-z and -L like proteases) and serpin activity are overrepresented among L3-enriched transcripts, whereas structural constituents of the cuticle and cellular component organization are overrepresented among L4-enriched transcripts (Table S3). In addition, our data indicate that abundant larval transcripts (*Alt1.2* and *Alt2*) show increased abundance in L3s relative to L4s.

DISCUSSION

Using high-throughput sequencing, we have undertaken a comprehensive genome-wide survey of the developmental transcriptome of the human filarial parasite *B. malayi*. Although deep sequencing data are highly informative in identifying novel transcribed elements and splice variants that help improve genome annotation (Wang et al., 2009), the present study aims to characterize transcriptome changes along the progression of the parasite's lifecycle. Transcriptome changes mediating cuticular molting likely represent one of the most notable developmental transitions in RNA expression. Like all nematodes, *Brugia* spp. have five lifecycle stages that are punctuated by molting of the collagenous cuticle. The tightly regulated process of molting involves cell signaling within the hypodermis to cue secretion of the new collagenous cuticle, shedding of the old cuticle and proteolytic remodeling of the new cuticle (Craig et al., 2007; Frand et al., 2005). Analysis of overrepresented GO terms highlights

structural cuticle components, extracellular matrix components and cysteine-peptidase inhibitors, among others, in genes with high levels of transcriptional variation over the lifecycle (Table S2). In particular, the cuticle collagen gene family displays distinct dynamic transcriptional patterns over the course of the lifecycle, likely reflecting compositional variation in cuticular structure among different life stages. Besides these structural components, genes displaying the most dramatic transcriptional variation in our data set are likely associated with developmental processes that differ between the larval and the adult stages and/or between the genders (e.g., gametogenesis). By contrast, genes constitutively expressed over the developmental period studied frequently have predicted cellular functions related to protein expression, modification and transport, possibly representing core cellular processes that are essential to the survival of cells independent of the lifecycle stage.

The present study indicates that genes exhibiting adult male enriched transcriptional pattern (relative to adult female and/or other stages) show strong statistical bias towards GO categories related to cytoskeleton, structural molecule activity, protein phosphorylation and dephosphorylation (Table 1). Many of these gene sets and functional categories are highly represented among classes of male-enriched transcripts in parasitic nematodes (Bennuru et al., 2011; Boag et al., 2003; Li et al., 2011; Nisbet et al., 2008) and have been identified in the *Caenorhabditis elegans* male and hermaphrodite germline as being involved in spermatogenesis (Reinke et al., 2004). Nematode sperm are unique in that they utilize a nematode-specific cytoskeletal element, major sperm protein, for ameboid motility. It is hypothesized that because mature nematode sperm lack ribosomal elements, the phosphorylation and dephosphorylation of molecules by a host of enzymes within the differentiated cells could promote maturation and pseudopod extension (Reinke et al., 2004). Seven of the genes found to be differentially

expressed in male worms in our study were also found in a microarray comparison of adult male and female worms (Li et al., 2005), and were shown by *in situ* localization to be expressed either in sperm or vas deferens tissue of adult male worms and not in gravid adult female worms (Jiang et al., 2008). If we compare our RNA-seq data with recent microarray work comparing gene expression in adult male and female *B. malayi* (Li et al., 2011), 515 of our 1,276 (40%) genes with male-biased expression match with male up-regulated genes found in the microarray comparison, and 150 out of the 651 (23%) genes with female-biased expression match the microarray findings.

In filarial nematodes, fertilization is internal and gravid females hold oocytes, sperm, zygotes, developing embryos, and MF in their uteri. Structural constituents of cuticle, transcription factor activity, DNA binding, and regulation of transcription emerged as notable themes in our analysis of overrepresented functional categories among genes with increased transcript levels in adult female and/or eggs & embryos (Table 1). These are likely relevant in the context of embryogenesis. Pairwise comparison of adult female with adult male presents us with a similar but more expanded view on features of genes displaying female-enriched expression (Table S3). Further comparisons with genes displaying germline-enriched expression patterns suggest that many of the female-biased transcripts, and more importantly, the majority of the above mentioned functional categories are attributable to the characteristics of the germline transcriptome. For instance, 33 out of 34 genes annotated with transcription factor activity (e.g., nuclear hormone receptors and homeobox domain containing proteins) that are enriched in female and/or eggs & embryos, have a distinctly germline-enriched expression pattern. *Bm-fab-1* (Bm1_33050), an embryonic fatty acid binding protein transcript previously found to be female-associated by differential display PCR and whose protein localizes to

embryos (Michalski et al., 2002; Michalski and Weil, 1999) also exhibits a germline-enriched expression pattern.

Much of our current information on molecular aspects of filarial reproduction comes from microarray and PCR-based transcriptome comparisons between whole adult male and female worms. These studies were based on the assumption that gender-associated transcripts arise from the reproductive organs and their contents. Our data suggest that such an assumption is not wholly unreasonable but may not always hold true. Out of 651 female-enriched transcripts we identified (in comparison to male), 82 display somatic tissue-enriched expression patterns, and it is likely that some of these transcripts are truly not derived from the germline tissues. Spatial expression patterns have not been confirmed for the majority of gender-associated *B. malayi* genes, and a growing body of research on nematode neurobiology and extracellular signaling lends support to the idea that some gender-associated genes can be expressed in non-reproductive tissues. For example, free-living and parasitic nematodes use gender-specific receptors to sense environmental signals, as demonstrated by the presence of anterior chemosensors in male worms that specifically bind female pheromones (Boag et al., 2001; Portman, 2007). Nematodes also store fat in intestinal cells, which may act as endocrine organs involved in germline signaling and are triggered by activation of intestinal cell nuclear receptors by lipophilic hormones (Kohler et al., 2007; Motola et al., 2006; Rottiers et al., 2006).

On the other hand, these observations are not inconsistent with the possibility that some somatic tissue derived transcripts play an essential role in embryonic development or intrauterine reproductive processes. The current study suggests that components incorporated into the embryonic cuticle and the eggshell membrane may be in some part maternal in origin. This interpretation is supported in at least one case where MF sheath protein transcripts in *Brugia* are

detectable by *in situ* hybridization only in adult female tissues and not in eggs or embryos (Jiang et al., 2008), while the encoded protein is found on the surface of *in utero* sheathed MF but not in maternal tissues (Selkirk et al., 1991). Other notable transcripts showing enrichment in female somatic tissues in our study include Juv-p120 excretory/secretory proteins and astacin proteases (Bm1_30065; Bm1_13915). Homologs of the latter in *C. elegans*, *nas-4* and *nas-9* are found in pharyngeal marginal cells, and in the hypodermis and reproductive tract, respectively (Park et al., 2010). Their functions are unknown but the localizations suggest roles in cuticle and eggshell remodeling.

After expulsion from females, developmentally arrested *Brugia* MF must undergo a maturation process within the mammalian host to become infective to the mosquito vector (de Hollanda et al., 1982; Fuhrman et al., 1987; Griffiths et al., 2009). *Brugia* MF are sheathed in a remnant of the eggshell membrane that is acellular and insoluble, and is composed of chitin and a variety of cross-linking proteins, lipids and polysaccharides (Araujo et al., 1994; Araujo et al., 1993; Laurence and Simpson, 1974; Simpson and Laurence, 1972). Our data indicate that a large number of transcripts representing DNA-binding proteins with zinc finger motifs as well as several endochitinase transcripts are significantly elevated in MF. Proteomic analysis also revealed a significant enrichment of zinc finger proteins in this stage of the lifecycle (Bennuru et al., 2011). Although the precise role of these DNA binding proteins is unknown, it is tempting to speculate on their involvement in maintaining the developmentally arrested state of circulating MF. Transcriptional increase in proteases and chitin-associated enzymes in MF could be important in the process of casting off the chitinous sheath during or after mosquito midgut penetration (Christensen and Sutherland, 1984; Fuhrman et al., 1987; Yamamoto et al., 1983). Immunolocalization studies have shown that in sheathed MF, chitinase is stored in the inner

body of the MF and secreted to the surface to degrade the sheath upon mosquito infection (Wu et al., 2008). Microfilarial maturation is accompanied by transcriptional transitions and changes in the composition of the microfilarial surface (Fuhrman et al., 1987; Griffiths et al., 2009). Despite the remarkable change in infectivity, our data suggest that transcriptional differences between IM and MM are relatively small; it is the least pronounced of all pairwise comparisons made in this study (Figure 2 and 3D). Genes involved in ATP synthase activity, tRNA production and cytoskeleton are overrepresented among those that show transcriptional change between IM and MM (Table S3). Although it is difficult to further characterize the exact nature of these changes due to a high proportion of genes with no functional annotation, we hypothesize that a metabolic shift is likely part of the maturation process in anticipation of the transition from the blood of a homeothermic host to the inhospitable midgut and hemocoel of the poikilothermic mosquito vector. It is important to consider that both populations of MF used in this experiment were derived from the peritoneal cavities of infected gerbils. Although we have previously shown a dramatic difference in mosquito infectivity between peritoneally-derived immature and mature MF (de Hollanda et al., 1982; Fuhrman et al., 1987; Griffiths et al., 2009), it is clear that intraperitoneally-derived MF, regardless of age, are considerably less infective than those found in circulating blood (Schrater et al., 1982). It is possible that the transcriptional profile of mature circulating MF differs from those that are derived from the peritoneal cavity.

Following the introduction of L3s into the peritoneal cavity of gerbils, the L3 to L4 transition requires no migration and occurs approximately 8 days post infection (unpublished). This particular lifecycle transition is of great interest to researchers trying to identify parasite molecules that mediate interactions with the host immune system, and that could be exploited with vaccines to confer protective immunity, or with drugs to prevent infection. Antigens that

historically have been of interest in this regard are the ALT (abundant larval transcript) family of potentially secreted larval acidic proteins found predominately in L2 and L3 stages (Gregory et al., 2000; Gregory et al., 1997; Wu et al., 2004); the L3 cystatin cysteine protease inhibitor family, Bm-SPN-2, TGF- β homologues, macrophage inhibition factor and Bm-VAL-1 (Gregory et al., 1997); troponin, tropomyosin and cuticular collagens (Hunter et al., 1999); *Onchocerca volvulus* activation associated secreted protein (Ov-ASP-1) (MacDonald et al., 2004), onchocystatin (Ov-CPI-2) (Cho-Ngwa et al., 2010) and Ov-SPI-1 (Ford et al., 2005), and *B. malayi* glutathione-s-transferase (Veerapathran et al., 2009). One hypothetical protein found to be L3 specific in our experiment, Bm1_38105, was also highly ranked as a potential drug target (Kumar et al., 2007).

In the present study, the transcriptome of developmentally arrested, vector-derived L3s was compared to that of peritoneally-derived L4s at 12-13 days post infection. Comparing our RNA-seq data to a recent microarray experiment (Li et al., 2009a) that assessed transcriptomes of vector-derived L3s to cultured and irradiated L3s, shows that 29 genes are shared, and likely constitute genes required for L3 survival in mosquitoes. These include Alt-2 and Alt1.2 proteins, cathepsin L precursors, Bm-*col-2*, cystatin, microfilarial surface associated protein, metabolic proteins and BmSERPIN. The differential expression of cathepsins *Bm-cpl-1*, 4, and 5 in vector stage L3s is supported by EST sequences and these genes are grouped phylogenetically into a distinct clade (Ia) separate from other nematode cathepsin-like proteases (Guiliano et al., 2004). There is strong evidence that these proteins play important roles in the L3 to L4 molt, because targeting the *cpl-1* gene in *O. volvulus* by RNAi decreased the rate of molting (Lustigman et al., 2004), and suppression of the cathepsin L-like cysteine protease transcript by injection of siRNA or dsRNA into infected mosquitoes carrying L2 and L3 stages of *B. malayi* retarded worm

growth, disrupted development and resulted in cuticular sloughing (Song et al., 2010). It is important to point out that the L4s we used were from the peritoneal cavity of gerbils, and did not follow the normal behavioral pathway of intradermal passage and migration to the lymphatics. It is possible that the transcriptional profile of intraperitoneally-derived L4s is different than that of worms found in lymphatics; indeed Chirgwin et al. (2008) showed different transcriptional profiles for three L3 genes at 3 days post infection in groups that had been injected intradermally and allowed to migrate naturally to the popliteal lymph node in the gerbil model, and those that were confined to the peritoneum.

In this study we provide a detailed overview of the molecular repertoires that define and differentiate distinct lifecycle stages of the parasite, extending and complementing previously published work on stage-specific gene expression (Bennuru et al., 2011; Bennuru et al., 2009; Griffiths et al., 2009; Jiang et al., 2008; Li et al., 2009a; Li et al., 2011; Moreno and Geary, 2008). Inclusion of seven different developmental stage samples uniquely allows us to place specific between-stage transcriptional differences into the broader context of the transcriptomic landscape during the lifecycle of *B. malayi*. It is important to emphasize, however, that this is just an overview of observations and that these data will be mined by the community to provide specific information on particular gene sets to bring these deep sequencing data into more complete biological context.

Because expression dynamics is an important consideration in the genome-wide assessment of candidate targets for control (Crowther et al., 2010; Kumar et al., 2007; Taylor et al., 2011), our comprehensive analysis of transcript abundance over developmental time is a valuable addition to a growing body of genomic and post-genomic resources that guide and support the concerted efforts to develop better intervention strategies.

SUPPORTING INFORMATION

See the following online documents at: <http://dx.doi.org/10.1371/journal.pntd.0001409>

Dataset S1. Gene-level RNA-seq read counts and RPKM values for *Brugia malayi* lifecycle transcriptome.

Figure S1. Histogram of raw read counts per gene model and scaling-normalized RPKM values. To facilitate transcript abundance comparisons between genes and stages, read counts were RPKM-transformed and TMM-normalized (Mortazavi et al., 2008; Robinson and Oshlack, 2010). The distribution of transcript level estimates indicated 4 to 5 logs of dynamic range.

Figure S2. Genome-wide distribution of the dispersion parameter estimating stage-to-stage variability in transcript abundance. Based on the NB dispersion parameter, genes were ordered from the most variable to the least variable, and partitioned into four equal-sized groups as indicated by the horizontal dotted lines. GO term enrichment tests were performed on each of the four groups (Table S2).

Figure S3. Venn diagram analysis to identify genes with female somatic tissue- or germline-enriched expression pattern. (1) genes with enriched expression in adult female relative to eggs & embryos; (2) genes with enriched expression in eggs & embryos relative to adult male; (3) genes with enriched expression in adult female and/or eggs & embryos relative to all other stages.

Table S1. Total number of reads sequenced and mapped to the genome.

Table S2. List of over-represented GO terms in groups of genes with different levels of overall variability in their transcript abundance across lifecycle stages. Genes were partitioned into four equal-sized groups designated as Q1 to 4 in the order of decreasing variability (Figure S2). Over-represented GO terms with p-value < 0.01 were included.

Table S3. List of over-represented GO terms and their corresponding p-values among genes that show relative transcriptional enrichment in pairwise comparisons (Figure 3D). Female somatic tissue- and germline-enriched patterns were defined using Venn diagram analysis (Figure S3).

REFERENCES

Araujo, A., Souto-Padron, T., De Souza, W., 1994. An ultrastructural, cytochemical and freeze-fracture study of the surface structures of *Brugia malayi* microfilariae. *International journal for parasitology* 24, 899-907.

Araujo, A.C., Souto-Padron, T., de Souza, W., 1993. Cytochemical localization of carbohydrate residues in microfilariae of *Wuchereria bancrofti* and *Brugia malayi*. *J Histochem Cytochem* 41, 571-578.

Bennuru, S., Meng, Z., Ribeiro, J.M., Semnani, R.T., Ghedin, E., Chan, K., Lucas, D.A., Veenstra, T.D., Nutman, T.B., 2011. Stage-specific proteomic expression patterns of the human filarial parasite *Brugia malayi* and its endosymbiont *Wolbachia*. *Proceedings of the National Academy of Sciences of the United States of America* 108, 9649-9654.

Bennuru, S., Semnani, R., Meng, Z., Ribeiro, J.M., Veenstra, T.D., Nutman, T.B., 2009. *Brugia malayi* excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling. *PLoS Negl Trop Dis* 3, e410.

Boag, P.R., Newton, S.E., Gasser, R.B., 2001. Molecular aspects of sexual development and reproduction in nematodes and schistosomes. *Advances in parasitology* 50, 153-198.

Boag, P.R., Ren, P., Newton, S.E., Gasser, R.B., 2003. Molecular characterisation of a male-specific serine/threonine phosphatase from *Oesophagostomum dentatum* (Nematoda: Strongylida), and functional analysis of homologues in *Caenorhabditis elegans*. *Int J Parasitol* 33, 313-325.

Camon, E., Magrane, M., Barrell, D., Lee, V., Dimmer, E., Maslen, J., Binns, D., Harte, N., Lopez, R., Apweiler, R., 2004. The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res* 32, D262-266.

Chirgwin, S.R., Coleman, S.U., Klei, T.R., 2008. *Brugia pahangi*: in vivo tissue migration of early L3 alters gene expression. *Exp Parasitol* 118, 89-95.

Cho-Ngwa, F., Liu, J., Lustigman, S., 2010. The *Onchocerca volvulus* cysteine proteinase inhibitor, Ov-CPI-2, is a target of protective antibody response that increases with age. *PLoS Negl Trop Dis* 4, e800.

Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162, 156-159.

Christensen, B.M., Sutherland, D.R., 1984. *Brugia pahangi*: Exsheathment and Midgut Penetration in *Aedes aegypti*. *Transactions of the American Microscopical Society* 103, 423-433.

Craig, H., Isaac, R.E., Brooks, D.R., 2007. Unravelling the moulting degradome: new opportunities for chemotherapy? *Trends Parasitol* 23, 248-253.

Crowther, G.J., Shanmugam, D., Carmona, S.J., Doyle, M.A., Hertz-Fowler, C., Berriman, M., Nwaka, S., Ralph, S.A., Roos, D.S., Van Voorhis, W.C., Agüero, F., 2010. Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. *PLoS Negl Trop Dis* 4, e804.

de Hollanda, J.C., Denham, D.A., Suswillo, R.R., 1982. The infectivity of microfilariae of *Brugia pahangi* of different ages to *Aedes aegypti*. *Journal of helminthology* 56, 155-157.

Ford, L., Guiliano, D.B., Oksov, Y., Debnath, A.K., Liu, J., Williams, S.A., Blaxter, M.L., Lustigman, S., 2005. Characterization of a novel filarial serine protease inhibitor, Ov-SPI-1, from *Onchocerca volvulus*, with potential multifunctional roles during development of the parasite. *The Journal of biological chemistry* 280, 40845-40856.

Frand, A.R., Russel, S., Ruvkun, G., 2005. Functional genomic analysis of *C. elegans* molting. *PLoS Biol* 3, e312.

Fuhrman, J.A., Urioste, S.S., Hamill, B., Spielman, A., Piessens, W.F., 1987. Functional and antigenic maturation of *Brugia malayi* microfilariae. *The American journal of tropical medicine and hygiene* 36, 70-74.

Ghedini, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J.E., Delcher, A.L., Guiliano, D.B., Miranda-Saavedra, D., Angiuoli, S.V., Creasy, T., Amedeo, P., Haas, B., El-Sayed, N.M., Wortman, J.R., Feldblyum, T., Tallon, L., Schatz, M., Shumway, M., Koo, H., Salzberg, S.L., Schobel, S., Pertea, M., Pop, M., White, O., Barton, G.J., Carlow, C.K.S., Crawford, M.J., Daub, J., Dimmic, M.W., Estes, C.F., Foster, J.M., Ganatra, M., Gregory, W.F.,

Johnson, N.M., Jin, J., Komuniecki, R., Korf, I., Kumar, S., Laney, S., Li, B.-W., Li, W., Lindblom, T.H., Lustigman, S., Ma, D., Maina, C.V., Martin, D.M.A., McCarter, J.P., McReynolds, L., Mitreva, M., Nutman, T.B., Parkinson, J., Peregrin-Alvarez, J.M., Poole, C., Ren, Q., Saunders, L., Sluder, A.E., Smith, K., Stanke, M., Unnasch, T.R., Ware, J., Wei, A.D., Weil, G., Williams, D.J., Zhang, Y., Williams, S.A., Fraser-Liggett, C., Slatko, B., Blaxter, M.L., Scott, A.L., 2007. Draft Genome of the Filarial Nematode Parasite *Brugia malayi*. *Science* 317, 1756-1760.

Gregory, W.F., Atmadja, A.K., Allen, J.E., Maizels, R.M., 2000. The abundant larval transcript-1 and -2 genes of *Brugia malayi* encode stage-specific candidate vaccine antigens for filariasis. *Infection and immunity* 68, 4174-4179.

Gregory, W.F., Blaxter, M.L., Maizels, R.M., 1997. Differentially expressed, abundant transcribed cDNAs from larval *Brugia malayi*. *Molecular and biochemical parasitology* 87, 85-95.

Griffiths, K.G., Mayhew, G.F., Zink, R.L., Erickson, S.M., Fuchs, J.F., McDermott, C.M., Christensen, B.M., Michalski, M.L., 2009. Use of microarray hybridization to identify *Brugia* genes involved in mosquito infectivity. *Parasitol Res* 106, 227-235.

Guiliano, D.B., Hong, X., McKerrow, J.H., Blaxter, M.L., Oksov, Y., Liu, J., Ghedin, E., Lustigman, S., 2004. A gene family of cathepsin L-like proteases of filarial nematodes are associated with larval molting and cuticle and eggshell remodeling. *Molecular and biochemical parasitology* 136, 227-242.

Hunter, S.J., Martin, S.A., Thompson, F.J., Tetley, L., Devaney, E., 1999. The isolation of differentially expressed cDNA clones from the filarial nematode *Brugia pahangi*. *Parasitology* 119 (Pt 2), 189-198.

Jiang, D., Li, B.W., Fischer, P.U., Weil, G.J., 2008. Localization of gender-regulated gene expression in the filarial nematode *Brugia malayi*. *Int J Parasitol* 38, 503-512.

Kent, W.J., 2002. BLAT--the BLAST-like alignment tool. *Genome research* 12, 656-664.

Kohler, H.R., Kloas, W., Schirling, M., Lutz, I., Reye, A.L., Langen, J.S., Triebkorn, R., Nagel, R., Schonfelder, G., 2007. Sex steroid receptor evolution and signalling in aquatic invertebrates. *Ecotoxicology (London, England)* 16, 131-143.

Kumar, S., Chaudhary, K., Foster, J.M., Novelli, J.F., Zhang, Y., Wang, S., Spiro, D., Ghedin, E., Carlow, C.K., 2007. Mining predicted essential genes of *Brugia malayi* for nematode drug targets. *PLoS ONE* 2, e1189.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.

Laurence, B.R., Simpson, M.G., 1974. The ultrastructure of the microfilaria of *Brugia*, Nematoda: Filarioidea. *International journal for parasitology* 4, 523-536.

- Li, B.-W., Rush, A., Mitreva, M., Yin, Y., Spiro, D., Ghedin, E., Weil, G., 2009a. Transcriptomes and pathways associated with infectivity, survival and immunogenicity in *Brugia malayi* L3. *BMC Genomics* 10, 267.
- Li, B.W., Rush, A.C., Crosby, S.D., Warren, W.C., Williams, S.A., Mitreva, M., Weil, G.J., 2005. Profiling of gender-regulated gene transcripts in the filarial nematode *Brugia malayi* by cDNA oligonucleotide array analysis. *Molecular and biochemical parasitology* 143, 49-57.
- Li, B.W., Rush, A.C., Jiang, D.J., Mitreva, M., Abubucker, S., Weil, G.J., 2011. Gender-associated genes in filarial nematodes are important for reproduction and potential intervention targets. *PLoS Negl Trop Dis* 5, e947.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009b. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Li, H., Ruan, J., Durbin, R., 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research* 18, 1851-1858.
- Lustigman, S., Zhang, J., Liu, J., Oksov, Y., Hashmi, S., 2004. RNA interference targeting cathepsin L and Z-like cysteine proteases of *Onchocerca volvulus* confirmed their essential function during L3 molting. *Molecular and biochemical parasitology* 138, 165-170.
- MacDonald, A.J., Tawe, W., Leon, O., Cao, L., Liu, J., Oksov, Y., Abraham, D., Lustigman, S., 2004. Ov-ASP-1, the *Onchocerca volvulus* homologue of the activation associated secreted protein family is immunostimulatory and can induce protective anti-larval immunity. *Parasite Immunol* 26, 53-62.
- Michalski, M.L., Monsey, J.D., Cistola, D.P., Weil, G.J., 2002. An embryo-associated fatty acid-binding protein in the filarial nematode *Brugia malayi*. *Molecular and biochemical parasitology* 124, 1-10.
- Michalski, M.L., Weil, G.J., 1999. Gender-specific gene expression in *Brugia malayi*. *Molecular and biochemical parasitology* 104, 247-257.
- Moreno, Y., Geary, T.G., 2008. Stage- and gender-specific proteomic analysis of *Brugia malayi* excretory-secretory products. *PLoS Negl Trop Dis* 2, e326.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* 5, 621-628.
- Motola, D.L., Cummins, C.L., Rottiers, V., Sharma, K.K., Li, T., Li, Y., Suino-Powell, K., Xu, H.E., Auchus, R.J., Antebi, A., Mangelsdorf, D.J., 2006. Identification of ligands for DAF-12 that govern dauer formation and reproduction in *C. elegans*. *Cell* 124, 1209-1223.

- Nisbet, A.J., Cottee, P.A., Gasser, R.B., 2008. Genomics of reproduction in nematodes: prospects for parasite intervention? *Trends Parasitol* 24, 89-95.
- Oshlack, A., Wakefield, M.J., 2009. Transcript length bias in RNA-seq data confounds systems biology. *Biol Direct* 4, 14.
- Park, J.O., Pan, J., Mohrlen, F., Schupp, M.O., Johnsen, R., Baillie, D.L., Zapf, R., Moerman, D.G., Hutter, H., 2010. Characterization of the astacin family of metalloproteases in *C. elegans*. *BMC Dev Biol* 10, 14.
- Portman, D.S., 2007. Genetic control of sex differences in *C. elegans* neurobiology and behavior. *Adv Genet* 59, 1-37.
- Reinke, V., Gil, I.S., Ward, S., Kazmer, K., 2004. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development* 131, 311-323.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.
- Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11, R25.
- Rottiers, V., Motola, D.L., Gerisch, B., Cummins, C.L., Nishiwaki, K., Mangelsdorf, D.J., Antebi, A., 2006. Hormonal control of *C. elegans* dauer formation and life span by a Rieske-like oxygenase. *Dev Cell* 10, 473-482.
- Schrater, A.F., Rossignol, P.A., Hamill, B., Piessens, W.F., Spielman, A., 1982. *Brugia malayi* microfilariae from the peritoneal cavity of jirds vary in their ability to penetrate the mosquito midgut. *The American journal of tropical medicine and hygiene* 31, 292-296.
- Selkirk, M.E., Yazdanbakhsh, M., Freedman, D., Blaxter, M.L., Cookson, E., Jenkins, R.E., Williams, S.A., 1991. A proline-rich structural protein of the surface sheath of larval *Brugia* filarial nematode parasites. *The Journal of biological chemistry* 266, 11002-11008.
- Simpson, M.G., Laurence, B.R., 1972. Histochemical studies on microfilariae. *Parasitology* 64, 61-88.
- Song, C., Gallup, J.M., Day, T.A., Bartholomay, L.C., Kimber, M.J., 2010. Development of an in vivo RNAi protocol to investigate gene function in the filarial nematode, *Brugia malayi*. *PLoS Pathog* 6, e1001239.
- Taylor, C.M., Fischer, K., Abubucker, S., Wang, Z., Martin, J., Jiang, D., Magliano, M., Rosso, M.N., Li, B.W., Fischer, P.U., Mitreva, M., 2011. Targeting protein-protein interactions for parasite control. *PLoS One* 6, e18381.

- Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*.
- Veerapathran, A., Dakshinamoorthy, G., Gnanasekar, M., Reddy, M.V., Kalyanasundaram, R., 2009. Evaluation of *Wuchereria bancrofti* GST as a vaccine candidate for lymphatic filariasis. *PLoS Negl Trop Dis* 3, e457.
- Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews* 10, 57-63.
- Wu, Y., Egerton, G., Pappin, D.J., Harrison, R.A., Wilkinson, M.C., Underwood, A., Bianco, A.E., 2004. The Secreted Larval Acidic Proteins (SLAPs) of *Onchocerca* spp. are encoded by orthologues of the alt gene family of *Brugia malayi* and have host protective potential. *Molecular and biochemical parasitology* 134, 213-224.
- Wu, Y., Preston, G., Bianco, A.E., 2008. Chitinase is stored and secreted from the inner body of microfilariae and has a role in exsheathment in the parasitic nematode *Brugia malayi*. *Molecular and biochemical parasitology* 161, 55-62.
- Yamamoto, H., Ogura, N., Kobayashi, M., Chigusa, Y., 1983. Studies on filariasis II: exsheathment of the microfilariae of *B. pahangi* in *Armigeres subalbatus*. *Jap J Parasit* 32, 287–292.
- Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A., 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11, R14.

FIGURE 1. Library size and gene model coverage. Library size refers to the total number of reads unambiguously mapped to gene models. At the sequencing depth of the current study, $\geq 8,000$ gene models had 5 or more mapped reads in each lifecycle stage library.

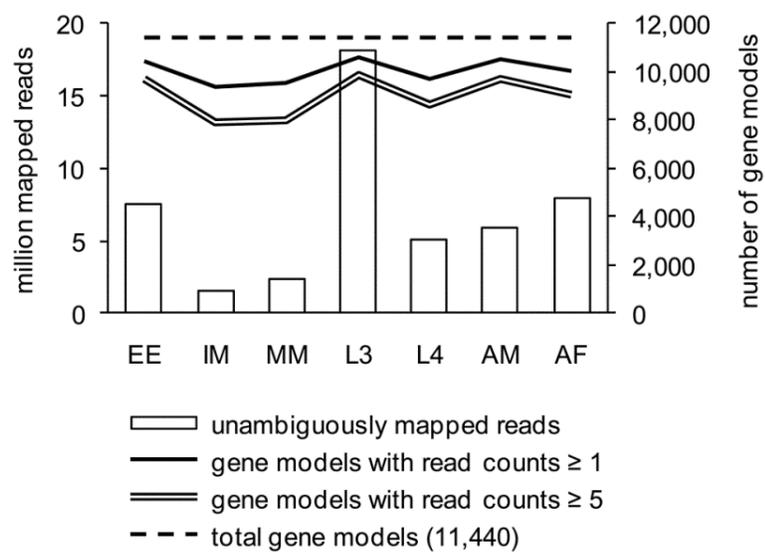


FIGURE 2. Multidimensional scaling (MDS) plot showing sample relations. The distance between each pair of samples was the square root of the common dispersion for the top 500 genes that best distinguished that pair of samples. These top 500 genes were selected according to the tagwise dispersion of all the samples.

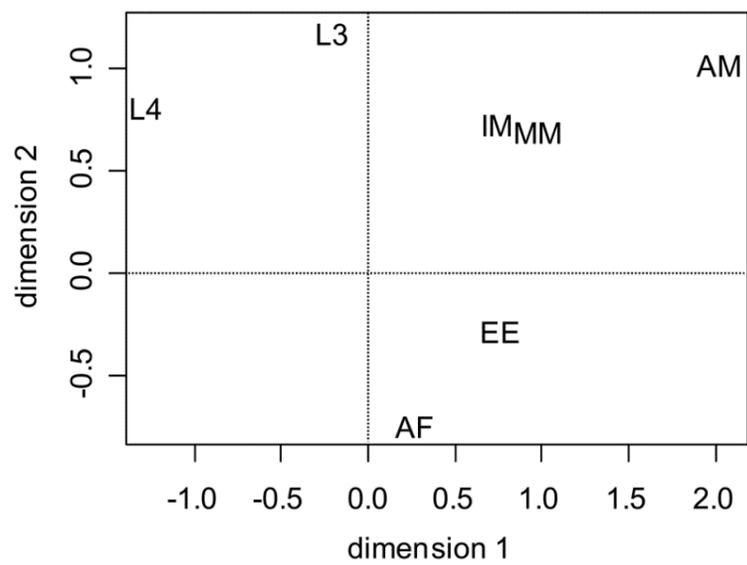


FIGURE 3. Comparison of transcriptome profiles between lifecycle stages. (A) Unsupervised hierarchical clustering of the RPKM values of the top 25% most variable genes according to the NB dispersion parameter (see Figure S2). (B) The number of genes classified into groups with distinct life stage dependent transcriptional patterns using a series of exact tests for the NB distribution (p -value < 0.01). (C) Venn diagram showing that members of the expression groups identified through the exact tests account for $\sim 80\%$ of the top 25% genes most variable in transcript abundance. (D) The number of differentially transcribed genes. Five direct pairwise comparisons were made between relevant stages to gain insights into the transcriptomic features associated with (1) sex differences, (2) intrauterine reproductive processes, (3) MF maturation, and (4) late larval development. Differentially transcribed genes were cross-referenced with the previously defined coexpression groups (Figure 3B).

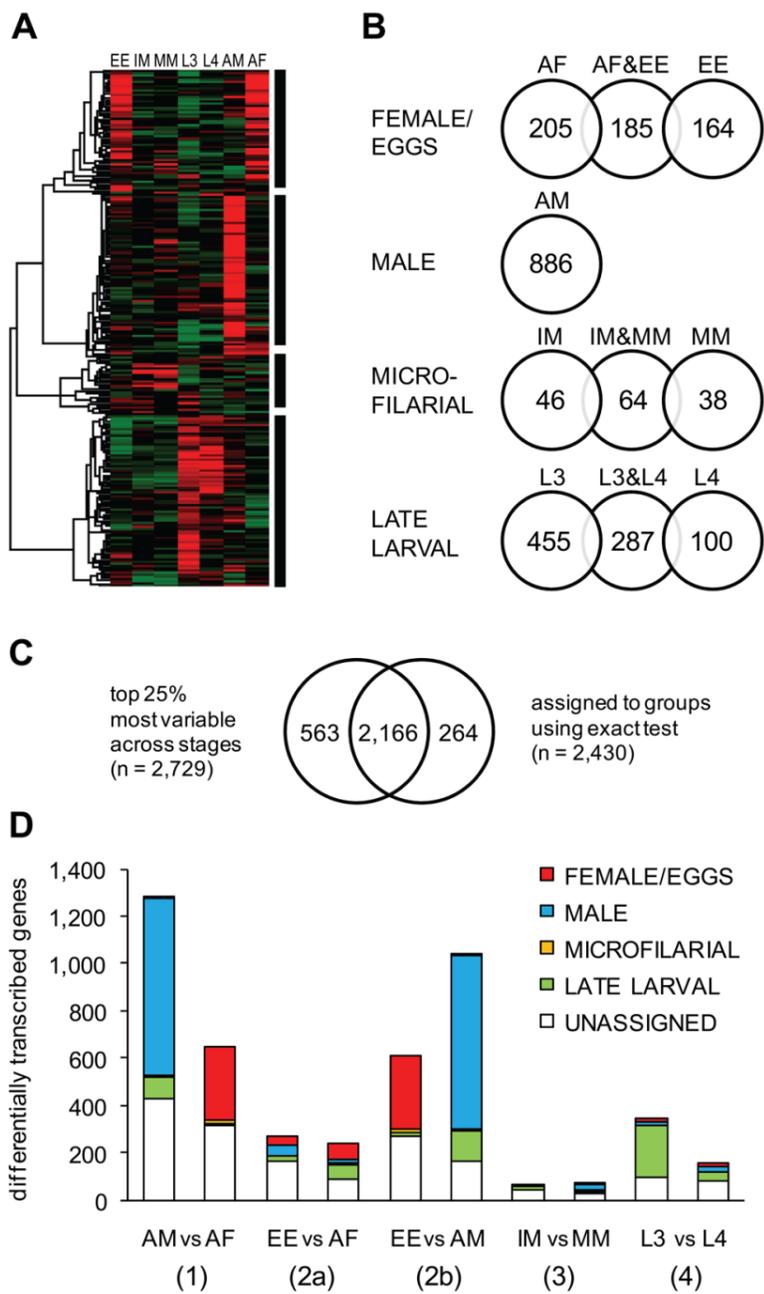


TABLE 1. List of over-represented GO terms in gene groups with distinct lifecycle stage dependent transcriptional patterns.

MICRO-FILARIAL	LATE LARVAL	MALE	FEMALE/EGGS	GO terms
				cellular component
1.8E-22				intracellular GO:0005622
	1.2E-05			integral to membrane GO:0016021
	2.1E-04			membrane GO:0016020
	2.9E-04			intermediate filament GO:0005882
		7.8E-05		cytoskeleton GO:0005856
			6.2E-07	nucleus GO:0005634
				molecular function
1.2E-23				zinc ion binding GO:0008270
1.4E-10				nucleic acid binding GO:0003676
9.7E-06				chitinase activity GO:0004568
1.6E-05				cation binding GO:0043169
2.5E-05				hydrolase activity, acting on glycosyl bonds GO:0016798
2.5E-05	1.8E-06			cysteine-type peptidase activity GO:0008234
7.6E-04				hydrolase activity, hydrolyzing O-glycosyl compounds GO:0004553
	9.2E-10		1.8E-06	structural constituent of cuticle GO:0042302
	2.9E-06			cysteine-type endopeptidase activity GO:0004197
	1.6E-05			transporter activity GO:0005215
	3.3E-05			oxidoreductase activity GO:0016491
	4.5E-05			extracellular matrix structural constituent GO:0005201
	9.5E-05		3.8E-04	serine-type endopeptidase inhibitor activity GO:0004867
	1.2E-04			voltage-gated chloride channel activity GO:0005247
	4.7E-04			calcium ion binding GO:0005509
	8.8E-04			hedgehog receptor activity GO:0008158
		9.5E-18		phosphoprotein phosphatase activity GO:0004721
		1.1E-13		protein kinase activity GO:0004672
		6.7E-12		structural molecule activity GO:0005198
		6.3E-10		kinase activity GO:0016301
		2.8E-08		protein tyrosine phosphatase activity GO:0004725
		1.0E-06		protein tyrosine kinase activity GO:0004713
		5.7E-06		phosphatase activity GO:0016791

TABLE 1. (continued)

MICRO-FILARIAL	LATE LARVAL	MALE	FEMALE/EGGS	GO terms	
		5.8E-06		ATP binding	GO:0005524
			1.6E-11	transcription factor activity	GO:0003700
			5.5E-10	sequence-specific DNA binding	GO:0043565
			4.4E-07	DNA binding	GO:0003677
			6.1E-05	ligand-dependent nuclear receptor activity	GO:0004879
			1.0E-04	transcription regulator activity	GO:0030528
				biological process	
1.0E-04	1.9E-04			proteolysis	GO:0006508
6.2E-04				chitin catabolic process	GO:0006032
	1.4E-06			metabolic process	GO:0008152
	6.0E-06			oxidation reduction	GO:0055114
	4.0E-05			transport	GO:0006810
	1.2E-04			chloride transport	GO:0006821
	3.2E-04			cell adhesion	GO:0007155
	5.8E-04			glycogen biosynthetic process	GO:0005978
		3.7E-13		protein amino acid phosphorylation	GO:0006468
		2.0E-07		protein amino acid dephosphorylation	GO:0006470
		1.7E-06		dephosphorylation	GO:0016311
			2.4E-10	regulation of transcription	GO:0045449
			5.3E-10	regulation of transcription, DNA-dependent	GO:0006355
			9.1E-05	transcription	GO:0006350

The gene groups used for the analysis were defined through a series of exact tests for NB distribution (Figure 3B). Over-represented GO terms with p-value < 0.01 were included.

CHAPTER 4**DUAL RNA-SEQ OF PARASITE AND HOST REVEALS GENE EXPRESSION
DYNAMICS DURING FILARIAL WORM-MOSQUITO INTERACTIONS**

INTRODUCTION

Lymphatic filariasis (LF) results from infection with mosquito-borne nematodes, *Wuchereria bancrofti*, *Brugia malayi*, or *Brugia timori*. With stigmatizing and disabling clinical manifestations including elephantiasis and hydrocele, LF is one of the most debilitating of the neglected tropical diseases (WHO, 2009). It is well accepted that the study of inter-species interaction is fundamental to understanding the biology of vector-borne parasites (Castillo et al., 2011; Lefevre and Thomas, 2008; Matthews, 2011), and filarial worms that cause LF are no exception. In humans, the interactions these parasites engage in with the host tissue and immune system dictate the pathobiology of LF (Bennuru and Nutman, 2009; Pfarr et al., 2009). Likewise, the interactions between filarial worms and their mosquito intermediate hosts (in the genera *Culex*, *Aedes*, *Mansonia* and *Anopheles*) determine the course of parasite development to the infective stage. (Bartholomay and Christensen, 2002)

In susceptible mosquitoes, ingested microfilariae (mf) traverse the midgut epithelium, migrate to the thorax, and become intracellular in the indirect flight muscles by day 1 post blood meal. In the next several days, mf differentiate into first-stage larvae (L1) and, at approximately day 5, L1s complete their first molt to second-stage larvae (L2), at which time the stoma opens and parasites actively feed on mosquito muscle tissue. At days 8-9, parasite development within the mosquito is complete with a second molt to the infective third-stage larva (L3). L3s then migrate to the head and proboscis from which they exit the mosquito during blood-feeding to infect the mammalian host (Erickson et al., 2009). Larval development within individual flight muscle fibers implies highly intimate interaction between host cell and parasite. For mf to successfully reach the L3 stage, host tissue must provide all of the nutritional needs of the rapidly growing nematodes, while surviving the mechanical damage caused by parasite

migration and development. Infection processes as such accompany histological changes in the host muscle fibers, ranging from depletion of glycogen granules to swelling and disintegration of nuclei or mitochondria (likely altering the metabolism of these organelles) to a complete cellular degeneration following the exit of L3s (Beckett, 1971, 1973).

Although efforts to understand factors that influence the developmental success of filarial worms in their mosquito vectors have led to a deeper appreciation of the complexity of the parasite-host interplay, the difficulty in making quantitative molecular-level observations on the parasite and the host simultaneously during infection continues to limit our ability to develop a coherent understanding of the filarial worm-mosquito interactions in their fine-grained details. Past studies have been essentially one-sided investigations focusing on either of the partners, mostly on host factors without taking the parasite activity into explicit consideration. As a consequence, filarial worm molecular processes underlying development in, and interaction with, the mosquito host tissue remain poorly described. By its very nature, parasite biology cannot be understood without integrating it with that of the host, nor can the host response be adequately explained without considering the parasite activity. In order to generate molecular insights into the dynamic symbiotic processes of filarial worm development in the mosquito, we conducted gene expression analysis of both the parasite and the host simultaneously over the course of infection.

Recent advances in sequencing-based approaches to quantitative transcriptome profiling (RNA-seq) facilitate direct, integrated analysis of mixed-species samples obtained from *in vivo* infection studies, in which host tissue samples often contain minute quantities of parasite material (Westermann et al., 2012). With the increase in sequencing depth, RNA-seq offers improved levels of sensitivity and dynamic range of detection, without the need of predefined

species-specific probes (or selective amplification of parasite RNA), unlike methods that rely on hybridization of targeted oligonucleotides (e.g., northern blotting, RT-PCR and microarrays). Here, we present a time-resolved dual RNA-seq analysis of *Brugia malayi* and *Aedes aegypti*, which investigates the temporal organization of transcriptional events in both the nematode and the mosquito tissue from the establishment of infection to the emergence of L3. We report the parasite gene transcription dynamics, exhibiting a highly ordered developmental program consisting of a series of cyclical and state-transitioning temporal patterns, which we contextualize in relation to the concurrent dynamics of the host tissue transcriptome.

MATERIALS AND METHODS

Mosquito strains and colony maintenance. Liverpool (LVP) and RED strains of *A. aegypti* were reared as previously described (Christensen and Sutherland, 1984). Briefly, mosquitoes were maintained on 0.3 M sucrose in an environmental chamber at $26.5\pm 1^\circ\text{C}$, $75\pm 10\%$ RH, and with a 16 hour light and 8 hour dark photoperiod with a 90 minute crepuscular period at the beginning and end of each light period. Three- to four-day-old mosquitoes were sucrose starved for 14 to 16 hours prior to blood feeding.

***B. malayi* infection and mosquito dissection.** Mosquitoes were exposed to *B. malayi* (originally obtained from NIAID/NIH Filariasis Research Reagent Resource Center) by feeding on ketamine/xylazine anesthetized, dark-clawed Mongolian gerbils, *Meriones unguiculatus*. Microfilaremiias ranged from 90-204 mf per 20 μl of blood as determined in samples obtained by orbital punctures immediately before feeding. Control mosquitoes were exposed to anesthetized, uninfected gerbils. Mosquitoes that fed to repletion were separated and maintained on 0.3 M sucrose in 30 \times 30 \times 30 cm colony cages. Thoracic tissues were collected every 24 hours after

exposure from infected and uninfected blood-fed mosquitoes for 4 (RED) and 8 (LVP) consecutive days, and also from non-blood-fed mosquitoes at the time of exposure (i.e., time-zero sample). For each time point, 30 thoraces were collected per strain per condition. Dissected tissues were immediately flash-frozen in microfuge tubes on dry ice and stored at -80 °C. At 8 days post infection, thoraces, abdomen, and heads were dissected for each strain as previously described (Erickson et al., 2009) to assess infection intensity and prevalence (Table 1). A complete set of replicate samples was collected using a separate generation of mosquitoes blood-fed on the same infected animals in an independent exposure.

RNA isolation. Total RNA was isolated from thoraces with MasterPure RNA Purification Kit (Epicenter, Madison, WI) including built-in Proteinase K and DNase treatments following the manufacturer's recommended protocols. Labgen7b electric tissue grinder (Cole-Parmer, Vernon Hills, IL) was used for tissue homogenization, and RNA purity and concentration were determined spectrophotometrically (NanoDrop ND-1000; Thermo Scientific, Waltham, MA).

For all samples, RNA integrity was visually assessed by denaturing gel electrophoresis using GelRed staining (Phenix Research Products, Candler, NC) and further confirmed by Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA).

Illumina library preparation and sequencing. Multiplex sequencing libraries were generated from 10 µg total RNA (per library) using Illumina's mRNA-seq sample prep kit and multiplexing sample preparation oligonucleotide kit (Illumina Inc., San Diego, CA) following the manufacturer's instructions. Separate libraries were constructed for each replicate sample from infected mosquitoes. Equal amounts of total RNA were pooled from replicates prior to library construction for uninfected blood-fed and non-blood-fed samples. In brief, polyadenylated mRNA was purified from total RNA using oligo-dT dynabead selection followed

by metal ion hydrolysis fragmentation. First strand synthesis, primed using random oligonucleotides, was followed by 2nd strand synthesis with RNaseH and DNAPolII to produce double-stranded cDNA. Template DNA fragments were end-repaired with T4 and Klenow DNA polymerases and blunt-ended with T4 polynucleotide kinase. A single 3' adenosine was added to the repaired ends using Klenow exo- and dATP. Index paired-end adapters were ligated and fragments size-selected (300-350 bp) by agarose gel electrophoresis. DNA was extracted using a Qiagen gel extraction kit, and libraries were amplified by PCR to enrich for properly ligated template strands, to generate enough DNA, and to add indexed primers for flowcell surface annealing, according to the manufacturer's protocol. Zymo DNA columns (ZRC163472) were used for purification before quantification using Agilent 2100 bioanalyzer and Kapa library quantification kit. Indexed libraries were denatured, diluted and combined for multiplexing. Cluster amplification was performed on the Illumina cBOT following the manufacturer's protocol. Sequencing-by-synthesis on the Illumina Genome Analyser Iix generated 2×50 bp paired-end reads. CASAVA v1.7 was used for demultiplexing, and sequence quality was assessed based on %GC content, average base quality and sequence duplication levels.

Multiplexing. To minimize the confounding effects of lane-to-lane or run-to-run technical variations, instead of sequencing each library using a single lane, libraries were sequenced over a number of lanes and runs by the use of Illumina's multiplexing capability (Auer and Doerge, 2010). Illumina Genome Analyzer Iix platform allowed 12 different sequencing libraries to be multiplexed in a single lane, and 7-8 separate lanes were available during each sequencing run. Because the total number of unique libraries exceeded the maximum number of usable indices, each index was used to label multiple libraries according to the scheme outlined in Fig. 1. Index #1, for example, was used to tag the first library of the sample blocks A, E and I, where each

block consisted of 4 individual libraries of sequential time points. This scheme allowed multiplexing of 12 libraries from the sample blocks A, B and C in a single lane, and also a different combination of 12 libraries from the sample blocks A, D and G in a single lane. In fact, the following combinations of sample blocks were multiplexed without indexing conflicts: [A, B, C], [D, E, F], [G, H, I], [A, D, G], [B, E, H] and [C, F, I]. Generation of sequence reads using these combinations of libraries (occupying 6 of the lanes during each run) contributed to a more robust dataset for longitudinal as well as cross-sectional analyses.

Sequence alignment and transcript quantification. Sequence reads from technical replicates (i.e., reads from the same library but sequenced in different lanes and runs) were combined per library prior to mapping as they share the same insert-size distribution. Alignment to the combined genomes of *A. aegypti* [VectorBase: AaegL1] and *B. malayi* [GenBank:DS236884-DS264093] was performed using TopHat v1.0.14, a splice junction mapper built upon the short read aligner Bowtie (Langmead et al., 2009; Trapnell et al., 2009). The pipeline utilized exon records in the genome annotation to build a set of known splice junctions for each gene model, complementing its *de novo* junction mapping algorithm. Low quality alignments with mapping quality scores less than 5 were removed before downstream analyses (Li et al., 2009; Li et al., 2008). Reads aligned to exonic regions were enumerated for each gene model using the HTSeq package (v0.4.7) in Python (www-huber.embl.de/users/anders/HTSeq). A set of paired-end reads were counted as one to represent a single sampling event, and reads overlapping more than one gene model were counted as ambiguous with the mode parameter set as “union”.

Differential transcription analysis. Statistical analysis of count data was performed with edgeR in Bioconductor (McCarthy et al., 2012; Robinson et al., 2010; Robinson and Smyth, 2007). We used generalized linear models (GLMs) with the negative binomial (NB) distribution to conduct

inference on differential expression. Genes were retained for analysis if read counts were greater than 0.5 cpm (counts per million) in at least two libraries and the overall within-replicate NB dispersion was lower than 0.5. Count data were normalized to account for compositional biases using TMM (trimmed mean of M) method, in addition to adjusting for differing library sizes (Robinson and Oshlack, 2010). GLM with explanatory variables of host strain, infection status and time was fit to the counts for each mosquito gene, and the biological coefficient of variation (BCV) was inferred from how much the variance of the counts exceeds the variance that would arise from Poisson sampling using the Cox-Reid profile-adjusted likelihood method. For the analysis of filarial worm genes, only the host strain and time were included as factors in the GLM. Information sharing was used so that genes could take individual values for the BCV, but stabilized towards a common value. This approach greatly improves inference, and ensures rigorous results even for experiments with minimal biological replication (McCarthy et al., 2012). A likelihood ratio test was then used to compute p-values for differential expression within the GLM framework. To identify parasite and host genes that are differentially transcribed between any of the time points (within each time-series), we used nested interaction models, and conducted an ANOVA-like test for any differences by testing for multiple coefficients being equal to zero. Using full interaction models, we then assessed host genes that responded differently at any of the time points in the infected host relative to the uninfected host. Similarly, we tested parasite genes for differential transcription at any of the time points in the susceptible host relative to the refractory host. Furthermore, interaction between effects of host stain and infection status was tested in mosquito genes over all time points. Adjusted p-values were reported after correction for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995).

Clustering of temporal expression profiles. Because log fold-change (logFC) estimates for genes with small read counts can be highly variable, we moderated logFC estimates using preFC function in edgeR. Thus, undefined values were avoided and poorly defined fold-changes for low counts were shrunk towards zero. We then used logFC values between sequential time points for each gene as a basis for clustering. After median-centering, K-means clustering was performed using (uncentered) Pearson correlation as a distance metric to group and summarize expression profiles into common temporal patterns. Figure of merit (FOM) plots were utilized in determining the appropriate number of clusters to ensure informative partitioning. Using goseq in Bioconductor, Gene Ontology (GO) terms statistically over-represented in each cluster were compiled to help interpret the biological implications of the expression patterns (Young et al., 2010). To assist in interpretation and visualization, we systematically summarized significant GO terms into a representative subset using an algorithm implemented in REVIGO (revigo.irb.hr) that relies on semantic similarity measures (Supek et al., 2011).

RESULTS AND DISCUSSION

To study gene expression dynamics during filarial worm-mosquito interactions using RNA-seq, we collected thoracic tissue samples from *B. malayi*-infected and uninfected blood-fed *A. aegypti* (Liverpool strain; LVP) every 24 hours after exposure for 8 days, and also from non-blood-fed mosquitoes at the time of exposure. LVP is a genetically selected susceptible strain that supports the complete development of *B. malayi* to the infective stage. In addition, a matching set of tissue samples was concurrently collected for 4 days using refractory mosquitoes (RED) to comparatively evaluate the transcriptome dynamics in an incompatible parasite-host association (Table 1). The RED strain is a multiple marker strain previously used in QTL mapping of filarial

worm susceptibility to *B. malayi* (Hickey and Craig, 1966; Severson et al., 1994). In this refractory mosquito, *B. malayi* mf traverse the midgut and become intracellular within the thoracic musculature, but after reaching their site for development, parasites fail to develop and die within a few days (Christensen et al., 1984).

Poly(A)-selected mRNA samples isolated from these tissues were subjected to Illumina sequencing (38 libraries across 48 lanes), and the resulting paired-end reads (2X50 bp) were aligned to the combined genomes of *A. aegypti* and *B. malayi* using a splice-junction aware aligner. Out of 1.17 billion sequenced reads, 987 million reads (84%) mapped unambiguously to the reference genomes (Fig. 2A). In the susceptible mosquito (LVP), the proportion of parasite reads relative to host reads increased from 0.1 to 9.0% during the course of infection, consistent with the nematode growth from mf to L3 (Murthy and Sen, 1981) (Fig. 2B). Interestingly, these data suggest that, during larval development leading up to the first molt (days 1-4), the rate of increase in the parasite's total mRNA output substantially exceeds that of the body growth (Fig. 3). This burst of transcriptional activity coincides with the initiation of tissue and organ differentiation following recovery from a developmental arrest in the mf stage. In the refractory mosquito (RED), the proportion of parasite reads remained at 0.1% and declined to lower levels by day 4, in line with the observed failure of parasite development and survival (Fig. 2B).

Next, we examined, in each organism, genome-wide temporal changes in transcript expression patterns in an effort to better understand the dynamic progression of infection processes. Genes displaying statistically significant differences in mRNA abundance levels across time points were identified (negative binomial generalized linear model, likelihood ratio test, $p < 0.01$ and fold-difference > 2). These “non-flat” profiles were grouped and summarized into common temporal patterns using k-means clustering. Mean expression changes between

sequential time points were computed and visualized to better evaluate the temporal dynamics across clusters in reference to specific time periods. In addition, Gene Ontology (GO) terms statistically over-represented in each cluster were compiled to help interpret the biological implications of expression patterns.

An overall assessment of the transcriptional dynamics in *B. malayi* during successful development from mf to L3 indicates a progressive and complex turnover in transcriptional contents (Fig. 4). Abundant transcripts at day 1 diminish over time, and new sets of transcripts emerge as the parasite matures. Although some expression clusters display patterns of gradual incremental changes, a more prevalent type is a pattern marked with disproportionately larger change at specific time intervals followed by transition to a new steady level. Collectively, these state-transiting patterns form a series of transcriptional “waves”, characteristic of developmental programs imposing order on cellular biogenesis (Yosef and Regev, 2011). Cyclical patterns are also evident, consistent with gene transcription associated with recurrent processes underlying nematode molting (Frandsen et al., 2005). The thoracic tissue of the mosquito in which filarial worms develop also undergoes extensive transcriptional changes (Fig. 5). Comparison of infected and uninfected host profiles indicates that normal physiological activities of the mosquito, most notably those related to blood-feeding responses, are accountable for a large proportion of these time-dependent expression changes (Bonizzoni et al., 2011). Concordantly, overall transcriptome dynamics are considerably higher during the first two days after blood meal. Infection-induced gene expression changes, on the other hand, exhibit distinct temporal kinetics that appears to reflect closely the histological phenotypes previously described in filarial worm infected mosquito muscle tissues (Fig. 6). In the following sections, we describe and

discuss the temporal coordination of parasite and host gene expression along the developmental timeline of filarial worm-mosquito interactions.

After ingestion with the blood meal, mf traverse the midgut epithelium within 2 hours and begin their migration through the hemocoel to the thoracic flight muscles (Christensen and Sutherland, 1984). Myofibrils are parted to form tunnels as larva enters the host cell, within which it lies uncoiled and parallel to the muscle fibers (Beckett, 1971). Over the next 2 days, mf transform into shorter, sausage form larvae. This process involves rearrangement and growth of preexisting microfilarial structures, as well as extensive cuticular reorganization (Lehane and Laurence, 1977). RNA-seq expression profile of *B. malayi* indicates that, between day 1 and 2, transcriptional induction is prominent among gene clusters P1, P5 and P6. GO categories associated with glycolysis, mitochondrial hydrogen-transporting ATP synthase, signal peptidase complex, DNA replication, and phosphoric diester hydrolase activity are overrepresented in these clusters. These data suggest that filarial worms transcriptionally respond to their changing metabolic and energy needs in the intracellular environment of their new host during re-initiation of larval development. Induction of genes involved in DNA replication is likely important in supporting increased cell division required for tissue differentiation. During this period, a decreasing trend is observed in clusters P13 and P14, and in the latter group, the pattern extends to day 3. Genes with nucleic acid or zinc ion binding activity are highly represented in these clusters, many of which are C2H2-type zinc finger proteins that may serve as *trans* regulators of gene expression (Razin et al., 2012).

In the mosquito host, a proportion of both parasitized and non-parasitized muscle fibers show signs of structural abnormality. Within the first 2 days after infection, nuclear and mitochondrial enlargement occur, and the number of affected fibers, the proportion of organelles

involved, and the degree of damage progressively increase as larval development proceeds. Although the proportion is very low, severe damage and cellular degeneration, likely caused by migrating mf, also appear early in infection, but without progression in the number of affected fibers (Beckett, 1973). The mosquito host response profile, representing transcript levels of infected animals relative to the uninfected controls, indicates that gene clusters HR1 and HR2, containing groups of stress response genes, such as small heat shock proteins with alpha-crystallin domain, display impulse-like patterns with acute dynamics during early infection. The distinct temporal pattern of induction among these genes, peaking at days 1, 6 and 8, suggests a close link between the kinetics of transcriptional response and the degree of cellular damage and mechanical disruption the host experience as parasites penetrate into and out of the muscle fiber, or actively ingest the host cell contents.

Middle to late L1 development, from day 3 to the first molt that occurs between day 4 and 5, is characterized by numerous mitotic divisions, lengthening of the body, and differentiation of internal structures, including a well-defined intestine and a divided-type esophagus (the anterior region muscular, the posterior glandular) that is formed around the pharyngeal thread of the mf (Laurence and Simpson, 1971). *B. malayi* gene clusters P7 and P8, enriched with genes implicated in ion channel activity and transmembrane transport, exhibit transcriptional increase between day 2 and 4, and their abundance levels are maintained on subsequent days. In addition, transient transcriptional increase is observed among clusters P2, P3 and P9, in which GO categories involved in calcium ion binding, response to stress, serpin activity, cuticle, metalloproteinase and steroid hormone receptor activity are overrepresented. Interestingly, these clusters show a second peak during L2 development, forming a cyclical pattern with respect to the molting events. By filtering on the direction and magnitude of

transcript abundance changes between sequential time points, we identified genes showing strong periodic patterns where transcript levels oscillate between high levels during intermolt periods and low levels during ecdysis (Fig. 7). Given our sampling rate, we could discern three groups with distinct kinetics, reaching their maximum levels at different times.

Genes with pulsatile transcription dynamics identified in the present study include a number of structural and regulatory components (such as cuticular collagens, cuticle-digesting proteases and nuclear receptor transcription factors) that have been predicted to participate in various aspects of the molting process (Ghedini et al., 2007). *B. malayi*, like all nematodes, progresses through its life stages through molts, each of which involves synthesis and secretion of a new cuticle, followed by separation and shedding of the old cuticle. From intracellular signaling to extracellular execution, molting requires a series of complex molecular reactions under tight spatiotemporal regulation at the level of the hypodermis. Also, equally critical to successful molting is a precise coordination of tissue development throughout the animal (Monsalve and Frand, 2012). In this context, the specific ordering of gene clusters with cyclical patterns is highly intriguing and informative, shedding light on the temporal organization of molecular events underlying the periodic episodes of molting in relation to the progressive life stage transitions.

In addition to the initial wave of acute reactions to the invading parasites, the mosquito host transcriptional responses to developing larvae also contain relatively moderate, but sustained temporal patterns of induction, as illustrated in cluster HR4. These expression changes, extending mostly over the 6 to 7 day period, likely reflect long-term chronic effects of filarial worm infection, part of which may represent a compensatory host response to energy and nutrient imbalances caused by intracellular parasite development. Overrepresented GO terms in

this cluster include oxidoreductase activity, purine nucleotide binding, phosphoenolpyruvate carboxykinase (PEPCK) activity, cAMP biosynthesis process, and gluconeogenesis. Because parasite-host interactions inherently involve two interconnected biological systems with a net flow of energy and nutrients, a metabolic perturbation is likely inevitable in the host (Kafsack and Llinas, 2010). It has been reported that filarial worm-infected muscle fibers show a large decrease in the amount of glycogen granules (Lehane and Laurence, 1977), and our data point to a transcriptional induction in the pathway of gluconeogenesis, by which cells synthesize glucose from metabolic precursors, such as glycogen. Data further suggest that this shift in metabolic state likely involves regulation at the level of cAMP synthesis and PEPCK activity, the latter of which controls a rate-limiting step of the pathway. Another notable feature of the infected host tissue transcriptome is the expression changes in glutathione transferase and glutathione peroxidase, whose main function is to inactivate toxic products of oxygen metabolism. Induction of these antioxidative enzymes could enhance protection from oxidative damage, but this may also change the redox state of the host cell environment in favor of the parasite survival.

By day 5, the first molt is complete, which marks the transition between the L1 and L2. At this time, the L2 start to feed through their open stoma and newly developed pharynx, with a complete, cuticularized lumen. Flight muscle mitochondria of the mosquito appear within the larval esophagus and midgut after 6 days of development, indicating an active tissue ingestion (Beckett and Boothroyd, 1970). In the following days, the larvae elongate and the gut undergoes further development. The rectum, however, remains closed with an anal plug, which is a vesicular structure preventing the egress of larval gut contents into the host tissue. This could represent one of the mechanisms by which filarial worms restrict trauma to the host cell during their intracellular development (Beckett, 1990b). The lumen of the rectum is then formed within

the anal vesicle just before the molt to the third stage (Laurence and Simpson, 1971). Between days 5 and 6, the genital primordium is formed, the position of which differs in males and females; it is at or just behind the esophago-intestinal junction in males, and at the midesophageal level in females. In middle to the late L2, the body wall consists of cuticle, chords (differentiated small dorso-ventral chords and broad lateral chords) and muscle components. A slight loosening of the cuticle at the head and tail is typically observed at day 7. By day 8, most larvae have either completed the second molt or are in the process of molting. Within a day or two, L3 migrate from the thoracic muscles through the head to the labium of the proboscis, from which they exit the mosquito during blood-feeding to infect the mammalian host. During L2 development in *B. malayi*, a diverse set of genes in clusters P2, P3 and P9, with oscillatory expression patterns, exhibit their second peak of transcription. Clusters P4, P10 and P11, on the other hand, show state-transitioning patterns with transcriptional induction between day 5 and 7. These likely represent specific processes that are initiated during the L2. Overrepresented GO categories in these clusters include cysteine-type peptidase activity, structural constituent of cuticle and glycolysis. Subsequently, between day 7 and 8, transcript levels rise in cluster P12, in which gene sets associated with cuticle components, transmembrane transport and chloride transport are enriched. Together, these clusters encompass a large array of functional categories, including genes with unknown function. Although the magnitude of change is comparatively small, clusters P15 and P16 show a decreasing trend over the period of L2 development. Enrichment of genes implicated in ribosome, translation and nucleic acid binding in these clusters suggests that genes involved in basal cellular activity are expressed in lower levels in late L2 and L3 compared to those of earlier stages. Overall, these data suggest an increase in the organism's transcriptional diversity in the second stage during which high levels

of cellular differentiation and tissue development occur. Nevertheless, such interpretation requires caution because our temporal data are confounded with variations in transcript detection limits, due to the increase in sequence sampling depth over time.

The mosquito transcriptional response during L2 development leading to L3 emergence is largely characterized by further induction of the stress response genes of the heat shock protein family, many of which were responsive at the onset of infection (clusters HR1 and HR2). These, serving as molecular chaperones, may play a crucial role in the repair process of the host cell, and thereby possibly suppress or delay necrosis during intracellular parasite development (Fulda et al., 2010). One striking observation from the past histological studies is that, despite the large increase in parasite size and host tissue consumption, almost all infected cells harboring live larvae do not undergo cellular breakdown or degeneration until the larval development is complete and the L3 migrate out of the flight muscle fibers (Beckett, 1971, 1973). It is thought that some physical or chemical factor associated with the terminal phase of larval development causes significant damage above a tolerable threshold, initiating necrosis. It remains unknown if filarial worms actively modulate host cell survival in an effort to conserve their habitat. In *A. aegypti* about 5-15% of all fibers eventually degenerate completely, which represents a permanent loss of contractile capacity, likely contributing to decreased flight activity and longevity (Beckett, 1990a; Berry et al., 1986; Hockmeyer et al., 1975).

B. malayi gene expression changes discussed thus far underlie successful developmental progression to the infective stage, which is dependent on the mosquito host tissue environment. Changing the genetic background of the host to an incompatible strain (RED) results in a failure of the parasite to develop and survive. Using a negative binomial generalized linear model, we assessed time dependent transcriptional changes between day 1 and 4 in *B. malayi* that were

different between the two different host environments, i.e., compatible (LVP) versus incompatible (RED). The most significant transcriptional alterations, as judged by statistical evidence, are presented in Table 2 ($P < 0.05$). Our data indicate that transcripts encoding hAT family dimerization domain containing protein, transmembrane protein, Leucine Rich Repeat family protein, dehydrogenases, and spectrins, among others, show an increasing trend during successful parasite development. In *Caenorhabditis elegans*, a mutation in β spectrin results in retarded growth and paralysis, suggesting that it is required for normal nematode development (Hammarlund et al., 2000). Of particular interest is the presence of putative transcriptional regulators in this list, such as nuclear hormone receptor and C2H2 type Zinc finger protein (Wilson et al., 2008). Although their specific function in *B. malayi* has not been elucidated, these are possibly involved in the control of key developmental pathways that are initiated during the early larval development in the mosquito. Among the transcripts that show an increasing trend in the non-developing or dying worms, is a calpain gene that encodes a member of the Ca^{2+} -dependent cysteine protease family. Calpain activation is an integral component of necrosis, and has been implicated in apoptotic cell death (Golstein and Kroemer, 2007), which implies that the refractory host environment of the RED strain not only prevents a normal filarial worm development, but also is lethal to the parasite. However, because of the extensively divergent genetic backgrounds between RED and LVP, the host genetic factors that confer this lethal environment could not be reasonably inferred from the between-strain differences in the host tissue transcriptome.

Transcriptomic approaches have been instrumental in studying the developmental regulation of gene expression in the context of filarial worm life cycle progression (Choi et al., 2011; Griffiths et al., 2009; Li et al., 2011; Li et al., 2012). Advances in high-throughput

nucleotide sequencing now enable us to interrogate the *in vivo* transcriptome dynamics of this metazoan parasite during its obligatory intracellular developmental phase in the mosquito host. Integrative analysis of the host transcriptome in a dual RNA-seq approach provides high-resolution overview of the parasite-host system, in which each partner's transcriptional state is dependent on the other partner. Studies of this nature will grow increasingly important in investigating the unique and common biological mechanisms that enable parasites to maintain their symbiotic association with the host, as well as the host strategies to counter parasitic infections, within a unified framework.

REFERENCES

- Auer, P.L., Doerge, R.W., 2010. Statistical design and analysis of RNA sequencing data. *Genetics* 185, 405-416.
- Bartholomay, L.C., Christensen, B.M., 2002. Vector-parasite interactions in mosquito-borne filariasis., in: Klei, T., Rajan, T. (Eds.), *The Filaria*, 1 ed. Kluwer Academic Publishers, Boston, pp. 9-19.
- Beckett, E.B., 1971. Histological changes in mosquito flight muscle fibres associated with parasitization by filarial larvae. *Parasitology* 63, 365-372.
- Beckett, E.B., 1973. Some quantitative aspects of flight muscle damage in mosquitoes infected with filarial larvae. *Annals of tropical medicine and parasitology* 67, 455-466.
- Beckett, E.B., 1990a. The non-specific nature of the response of mosquito flight muscle to filarial parasitization. *Parasitology research* 76, 360-366.
- Beckett, E.B., 1990b. Species variation in mosquito flight-muscle damage resulting from a single filarial infection and its repercussions on a second infection. *Parasitology research* 76, 606-609.
- Beckett, E.B., Boothroyd, B., 1970. Mode of nutrition of the larvae of the filarial nematode *Brugia pahangi*. *Parasitology* 60, 21-26.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 289-300.

Bennuru, S., Nutman, T.B., 2009. Lymphatics in human lymphatic filariasis: in vitro models of parasite-induced lymphatic remodeling. *Lymphat Res Biol* 7, 215-219.

Berry, W.J., Rowley, W.A., Christensen, B.M., 1986. Influence of developing *Brugia pahangi* on spontaneous flight activity of *Aedes aegypti* (Diptera: Culicidae). *Journal of medical entomology* 23, 441-445.

Bonizzoni, M., Dunn, W.A., Campbell, C.L., Olson, K.E., Dimon, M.T., Marinotti, O., James, A.A., 2011. RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, *Aedes aegypti*. *BMC genomics* 12, 82.

Castillo, J.C., Reynolds, S.E., Eleftherianos, I., 2011. Insect immune responses to nematode parasites. *Trends Parasitol* 27, 537-547.

Choi, Y.J., Ghedin, E., Berriman, M., McQuillan, J., Holroyd, N., Mayhew, G.F., Christensen, B.M., Michalski, M.L., 2011. A deep sequencing approach to comparatively analyze the transcriptome of lifecycle stages of the filarial worm, *Brugia malayi*. *PLoS neglected tropical diseases* 5, e1409.

Christensen, B.M., Sutherland, D.R., 1984. *Brugia pahangi*: Exsheathment and Midgut Penetration in *Aedes aegypti*. *Transactions of the American Microscopical Society* 103, 423-433.

Christensen, B.M., Sutherland, D.R., Gleason, L.N., 1984. Defense reactions of mosquitoes to filarial worms: comparative studies on the response of three different mosquitoes to inoculated *Brugia pahangi* and *Dirofilaria immitis* microfilariae. *Journal of invertebrate pathology* 44, 267-274.

Erickson, S.M., Xi, Z., Mayhew, G.F., Ramirez, J.L., Aliota, M.T., Christensen, B.M., Dimopoulos, G., 2009. Mosquito infection responses to developing filarial worms. *PLoS neglected tropical diseases* 3, e529.

Frand, A.R., Russel, S., Ruvkun, G., 2005. Functional genomic analysis of *C. elegans* molting. *PLoS biology* 3, e312.

Fulda, S., Gorman, A.M., Hori, O., Samali, A., 2010. Cellular stress responses: cell survival and cell death. *International journal of cell biology* 2010, 214074.

Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J.E., Delcher, A.L., Giuliano, D.B., Miranda-Saavedra, D., Angiuoli, S.V., Creasy, T., Amedeo, P., Haas, B., El-Sayed, N.M., Wortman, J.R., Feldblyum, T., Tallon, L., Schatz, M., Shumway, M., Koo, H., Salzberg, S.L., Schobel, S., Pertea, M., Pop, M., White, O., Barton, G.J., Carlow, C.K., Crawford, M.J., Daub, J., Dimmic, M.W., Estes, C.F., Foster, J.M., Ganatra, M., Gregory, W.F., Johnson, N.M., Jin, J., Komuniecki, R., Korf, I., Kumar, S., Laney, S., Li, B.W., Li, W., Lindblom, T.H., Lustigman, S., Ma, D., Maina, C.V., Martin, D.M., McCarter, J.P., McReynolds, L., Mitreva, M., Nutman, T.B., Parkinson, J., Peregrin-Alvarez, J.M., Poole, C., Ren, Q., Saunders, L., Sluder, A.E., Smith, K., Stanke, M., Unnasch, T.R., Ware, J., Wei, A.D., Weil, G., Williams, D.J., Zhang,

- Y., Williams, S.A., Fraser-Liggett, C., Slatko, B., Blaxter, M.L., Scott, A.L., 2007. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 317, 1756-1760.
- Golstein, P., Kroemer, G., 2007. Cell death by necrosis: towards a molecular definition. *Trends in biochemical sciences* 32, 37-43.
- Griffiths, K.G., Mayhew, G.F., Zink, R.L., Erickson, S.M., Fuchs, J.F., McDermott, C.M., Christensen, B.M., Michalski, M.L., 2009. Use of microarray hybridization to identify *Brugia* genes involved in mosquito infectivity. *Parasitology research* 106, 227-235.
- Hickey, W.A., Craig, G.B., Jr., 1966. Genetic distortion of sex ratio in a mosquito, *Aedes aegypti*. *Genetics* 53, 1177-1196.
- Hockmeyer, W.T., Schiefer, B.A., Redington, B.C., Eldridge, B.F., 1975. *Brugia pahangi*: effects upon the flight capability of *Aedes aegypti*. *Experimental parasitology* 38, 1-5.
- Kafsack, B.F., Llinas, M., 2010. Eating at the table of another: metabolomics of host-parasite interactions. *Cell host & microbe* 7, 90-99.
- Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Laurence, B.R., Simpson, M.G., 1971. The microfilaria of *Brugia*: a first stage nematode larva. *Journal of helminthology* 45, 23-40.
- Lefevre, T., Thomas, F., 2008. Behind the scene, something else is pulling the strings: emphasizing parasitic manipulation in vector-borne diseases. *Infect Genet Evol* 8, 504-519.
- Lehane, M.J., Laurence, B.R., 1977. Flight muscle ultrastructure of susceptible and refractory mosquitoes parasitized by larval *Brugia pahangi*. *Parasitology* 74, 87-92.
- Li, B.W., Rush, A.C., Jiang, D.J., Mitreva, M., Abubucker, S., Weil, G.J., 2011. Gender-associated genes in filarial nematodes are important for reproduction and potential intervention targets. *PLoS neglected tropical diseases* 5, e947.
- Li, B.W., Wang, Z., Rush, A.C., Mitreva, M., Weil, G.J., 2012. Transcription profiling reveals stage- and function-dependent expression patterns in the filarial nematode *Brugia malayi*. *BMC genomics* 13, 184.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Li, H., Ruan, J., Durbin, R., 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research* 18, 1851-1858.

- Matthews, K.R., 2011. Controlling and coordinating development in vector-transmitted parasites. *Science* 331, 1149-1153.
- McCarthy, D.J., Chen, Y., Smyth, G.K., 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40, 4288-4297.
- Monsalve, G.C., Frand, A.R., 2012. Toward a unified model of developmental timing: A “molting” approach. *Worm* 1, 25-34.
- Murthy, P.K., Sen, A.B., 1981. Sequential development changes in microfilariae of subperiodic *Brugia malayi* to infective larvae in susceptible strain of *Aedes aegypti* (Macdonald). *The Journal of communicable diseases* 13, 102-109.
- Pfarr, K.M., Debrah, A.Y., Specht, S., Hoerauf, A., 2009. Filariasis and lymphoedema. *Parasite Immunol* 31, 664-672.
- Razin, S.V., Borunova, V.V., Maksimenko, O.G., Kantidze, O.L., 2012. Cys2His2 zinc finger protein family: classification, functions, and major members. *Biochemistry. Biokhimiia* 77, 217-226.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.
- Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11, R25.
- Robinson, M.D., Smyth, G.K., 2007. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* 23, 2881-2887.
- Severson, D.W., Mori, A., Zhang, Y., Christensen, B.M., 1994. Chromosomal mapping of two loci affecting filarial worm susceptibility in *Aedes aegypti*. *Insect molecular biology* 3, 67-72.
- Supek, F., Bosnjak, M., Skunca, N., Smuc, T., 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one* 6, e21800.
- Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*.
- Westermann, A.J., Gorski, S.A., Vogel, J., 2012. Dual RNA-seq of pathogen and host. *Nature reviews. Microbiology* 10, 618-630.
- WHO, 2009. Global program to eliminate lymphatic filariasis. *Wkly Epidemiol Rec* 84, 437-444.
- Wilson, D., Charoensawan, V., Kummerfeld, S.K., Teichmann, S.A., 2008. DBD--taxonomically broad transcription factor predictions: new content and functionality. *Nucleic acids research* 36, D88-92.

Yosef, N., Regev, A., 2011. Impulse control: temporal dynamics in gene transcription. *Cell* 144, 886-896.

Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A., 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology* 11, R14.

FIGURE 1. (A) Description of the Illumina sequencing libraries. Except for J and K, each sample block (A through I) represents 4 distinct libraries. (B) Indices #1-12 were used to tag 36 different sequencing libraries.

A

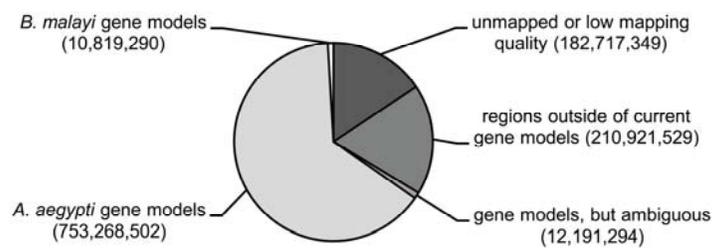
infection	biological replication	host strain	days post blood-feeding										
			0	1	2	3	4	5	6	7	8		
infected	1	susceptible					A		B				
		refractory					C						
	2	susceptible					D		E				
		refractory					F						
uninfected	pooled	susceptible	J			G			H				
		refractory	K			I							

B

index #	sample block
1 - 4	A,E,I
5 - 8	B,F,G
9 - 12	C,D,H

FIGURE 2. (A) Overall composition of Illumina reads based on sequence alignment to reference genomes. (B) Total number of reads that aligned unambiguously to the gene models of *Aedes aegypti* and *Brugia malayi*.

A



B

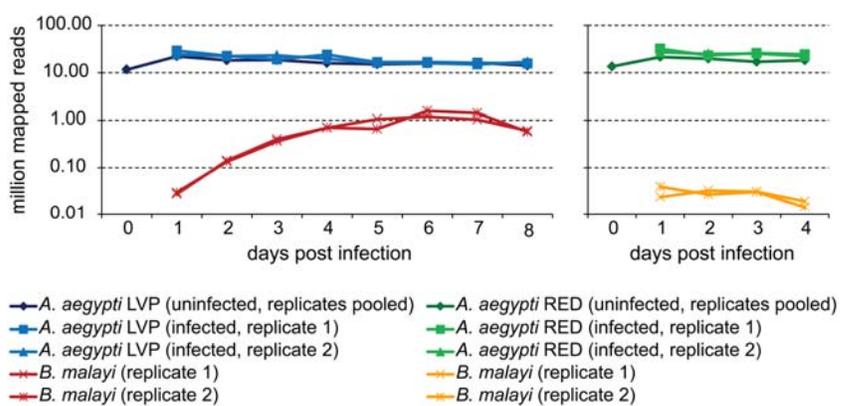


FIGURE 3. Relative change in *Brugia malayi* mRNA output and body volume during development in *Aedes aegypti* (LVP). Change in transcriptional output was estimated from the parasite to host ratio of total read counts, assuming that the host mRNA output is stable over time. Volumetric calculations were carried out based on measurements reported by Murthy and Sen (1981) after approximating the shape of the worm to a cylinder. Error bars denote range.

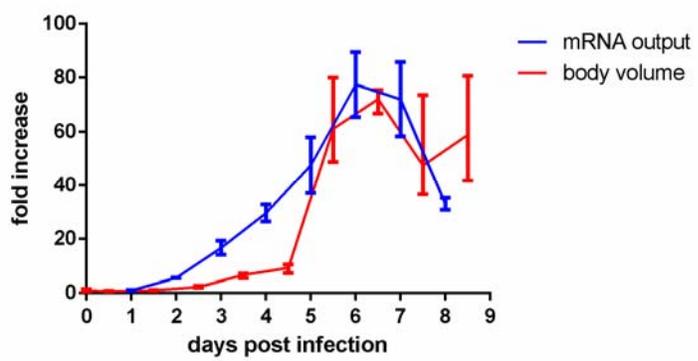


FIGURE 4. Time-course profiles of *Brugia malayi* transcript abundance during development in *Aedes aegypti* (LVP). “Non-flat” profiles (negative binomial generalized linear model, likelihood ratio test, $p < 0.01$ and fold-difference > 2) were grouped into common temporal patterns using k-means clustering. The vertical scale shows \log_2 fold-change in increments of 2, and the horizontal scale shows time in days. Heatmaps represent mean expression changes between sequential time points (red-blue), and over-represented Gene Ontology terms (green-white) in each cluster.

FIGURE 5. Time-course profiles of *Aedes aegypti* (LVP) transcript abundance during *Brugia malayi* infection. “Non-flat” profiles (negative binomial generalized linear model, likelihood ratio test, $p < 0.01$ and fold-difference > 2) were grouped into common temporal patterns using k-means clustering. The vertical scale shows \log_2 fold-change in increments of 2, and the horizontal scale shows time in days. Heatmaps represent mean expression changes between sequential time points (red-blue), and over-represented Gene Ontology terms (green-white) in each cluster.

FIGURE 6. Time-course profiles of *Aedes aegypti* (LVP) host response to *Brugia malayi* infection. These represent transcript levels of infected animals relative to uninfected controls. Differential profiles (negative binomial generalized linear model, likelihood ratio test, $p < 0.01$) were grouped into common temporal patterns using k-means clustering. The vertical scale shows \log_2 fold-change in increments of 2, and the horizontal scale shows time in days. Heatmaps represent mean expression changes between sequential time points (red-blue), and over-represented Gene Ontology terms (green-white) in each cluster.

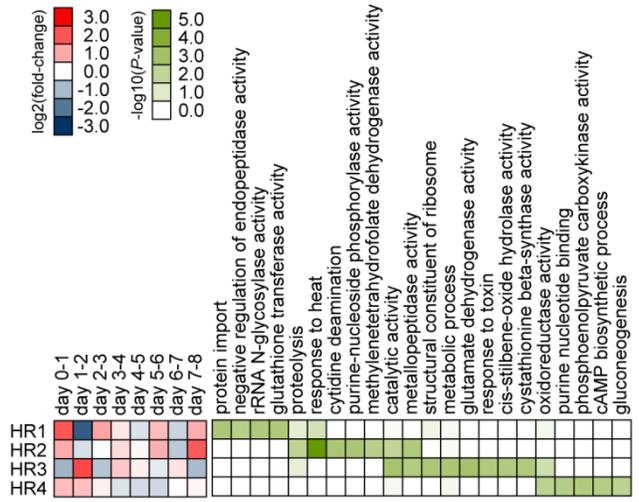
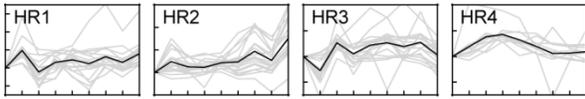
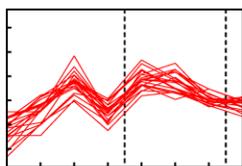
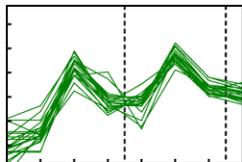


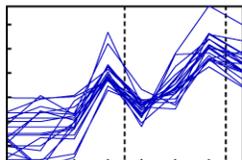
FIGURE 7. Time-course profiles of *Brugia malayi* transcript abundance, displaying periodic patterns during development in *Aedes aegypti* (LVP). For each temporal pattern with distinct kinetics (red:early, green:middle, and blue:late), top 20 genes showing high magnitude of change are plotted along with a list of its representative constituents. Mean expression profiles of these gene sets are displayed in the bottom panel. The vertical scale shows \log_2 fold-change in increments of 2, and the horizontal scale shows time in days.

**Early**

Calcium binding EGF domain containing protein
 Nuclear receptor RXR
 Steroid receptor seven-up type 2
 Fasciclin domain containing protein
 IQ calmodulin-binding motif family protein

**Middle**

Kunitz/Bovine pancreatic trypsin inhibitor domain containing protein
 PAN domain containing protein
 Cuticle collagen
 EF hand family protein
 Roller protein 8
 Lipase family protein
 Cell death specification protein 2

**Late**

Nematode astacin protease
 DB module family protein
 LBP / BPI / CETP family, C-terminal domain containing protein
 Ground-like domain containing protein
 Glycine-rich cell wall structural protein 1
 Sybindin-like family protein
 Collagen protein

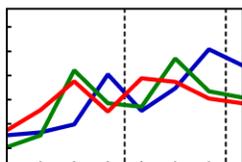


TABLE 1. Infection prevalence and intensity of *Brugia malayi* in different strains of *Aedes aegypti* during the generation of present RNA-seq dataset. Parasites were observed in mosquito dissections by microscopy at 8 days after ingestion of microfilaremic blood (90-204 mf per 20 μ l).

Biological replication	<i>Aedes aegypti</i> strain	Prevalence (n=30)	Mean intensity (SD)
1	LVP	93.3%	14.7 (7.3)
	RED	6.7%	1.0 (0.0)
2	LVP	97.0%	14.5 (7.1)
	RED	6.7%	1.0 (0.0)

TABLE 2. *B. malayi* genes showing time dependent transcriptional changes that are different in *A. aegypti* LVP with respect to *A. aegypti* RED between day 1 and 4 after infection.

Gene ID		<i>P</i> -value	Log ₂ (fold-change) ^a		
			day 2	day 3	day 4
Bm1_38815 ^b	Zinc finger, C2H2 type family protein	4.9E-02	3.8	4.9	6.6
Bm1_41565	hAT family dimerisation domain containing protein	1.9E-02	5.2	2.0	5.7
Bm1_41500	hypothetical protein	1.9E-02	4.4	3.5	3.9
Bm1_55295	Transmembrane protein	1.9E-02	3.9	4.7	3.1
Bm1_38105	hypothetical protein	4.0E-03	2.2	6.4	2.3
Bm1_54290	Dehydrogenases, short chain protein 2, isoform a	3.6E-02	0.7	4.3	5.7
Bm1_32800	hypothetical protein	2.4E-02	2.8	1.5	4.3
Bm1_55880 ^b	Nuclear hormone receptor E75	1.9E-02	4.0	-0.2	3.8
Bm1_40135	Leucine Rich Repeat family protein	2.3E-02	3.0	3.8	-0.5
Bm1_52790	beta-G spectrin	2.4E-02	1.8	2.3	2.2
Bm1_21675	Timeless protein	4.4E-02	-0.4	1.9	4.3
Bm1_07350	DNA translocase ftsK	3.6E-02	1.6	-0.3	4.1
Bm1_27855	Spectrin alpha chain	6.8E-03	1.2	1.4	1.7
Bm1_08800	Rab GDP dissociation inhibitor alpha	4.1E-02	1.9	0.6	1.5
Bm1_28665	hypothetical protein	2.4E-02	3.4	0.1	0.2
Bm1_40715	myosin heavy chain	2.9E-02	0.5	1.0	2.2
Bm1_47310	acetyl-Coenzyme A synthetase 2	2.2E-02	1.0	1.0	1.5
Bm1_02940	Glycine-rich cell wall structural protein 1 precursor	4.4E-02	-2.9	1.2	4.8
Bm1_35580	hypothetical protein	1.9E-02	1.7	1.1	-0.6
Bm1_41425	BM20	4.8E-05	-0.4	0.6	1.7
Bm1_30555	Proteasome A-type and B-type family protein	2.4E-02	-0.5	2.4	-0.8
Bm1_34965	Conserved hypothetical protein	1.7E-02	2.4	0.4	-3.4
Bm1_22365	DB module family protein	2.3E-02	0.8	0.6	-2.4
Bm1_14260	40S ribosomal protein S17	2.9E-02	-0.3	-0.9	-1.5
Bm1_16090	hypothetical protein	1.9E-02	-3.7	1.1	-1.1
Bm1_42225	CG7773-PA	4.9E-02	-3.2	-0.9	-2.5
Bm1_51055	Immunoglobulin I-set domain containing protein	2.4E-02	-4.2	-2.4	-1.7
Bm1_10280	transketolase	4.4E-02	-2.8	-2.2	-3.2
Bm1_10650	resolvase	1.9E-02	-3.1	-4.8	-1.3
Bm1_14050	calpain 7, putative	2.4E-02	-3.9	-4.6	-4.8
Bm1_19605 ^b	hypothetical protein	2.4E-02	-4.6	-4.4	-4.8
Bm1_55075 ^b	Zinc finger, C2H2 type family protein	1.9E-02	-4.0	-5.5	-6.0

^a abundance changes relative to day 1 in LVP with respect to (abundance changes relative to day 1 in) RED

^b putative transcription factor predicted by DBD database (Wilson et al., 2008)

CHAPTER 5

SUMMARY

Mosquitoes serve as vectors of some of the most debilitating parasitic diseases that exact a devastating toll on global health (WHO, 2002). The failure of traditional control methods and the difficulty of developing effective interventions emphasize our limited knowledge of the biology of parasites, hosts and parasite-host interrelationships. The present work aimed to fill in these knowledge gaps by conducting studies on mosquito cellular immunity, filarial parasite development and its interaction with the mosquito host, using high-throughput transcriptomic approaches as a primary investigative tool. In particular, hemocyte-enriched transcripts were analyzed in the mosquito *Aedes aegypti* using genome-wide microarrays, with emphasis on genes exhibiting tissue specificity following immune challenge (Chapter 2). An integrative approach, combining infection response profiles with tissue enrichment ratios, provided a detailed overview of the transcriptional characteristics of the hemocyte molecular repertoire and immune-related gene families within the currently understood framework of insect immunity (Hillyer, 2010; Waterhouse et al., 2007). In addition, hemocyte-enriched transcripts conserved among *A. aegypti*, *Anopheles gambiae* (Pinto et al., 2009) and *Drosophila melanogaster* (Irving et al., 2005) were identified. Phylogenetically conserved tissue-enriched expression could imply functional constraints (Piro et al., 2011), possibly suggesting that these genes are indispensable for hemocyte function. Because this approach effectively captured genes previously described in association with hemocyte-specific processes (e.g., prophenoloxidasases and phagocytic receptors), top ranked orthologous groups with uncharacterized function strongly merit further experimental investigation in relation to hemocyte biology.

The development of high-throughput sequencing technologies in recent years has rapidly transformed the field of transcriptomics, making it possible to study global RNA expression with extraordinary coverage and base-level resolution (Wang et al., 2009). Using a sequencing-based

approach to quantitative RNA profiling (RNA-seq), the lifecycle transcriptome of the human filarial worm *Brugia malayi* was characterized to uncover developmental transitions in gene expression from eggs through larval stages to adults (Chapter 3). The analysis revealed groups of genes with distinct life stage dependent transcriptional patterns, including those displaying sex-biased or germline-enriched expression, complementing previously published expression data on *B. malayi* (Bennuru et al., 2011; Bennuru et al., 2009; Griffiths et al., 2009; Moreno and Geary, 2008). From the standpoint of designing drug-based or vaccine interventions that prevent infection or curtail parasite transmission, there is high interest in understanding the biology of the L3 to L4 transition, and the reproductive biology of filarial worms (Li et al., 2011; Li et al., 2009). Because expression dynamics is an important consideration in the genome-wide assessment of candidate targets for control, this work represents a valuable resource for basic as well as translational research (Crowther et al., 2010; Kumar et al., 2007; Taylor et al., 2011).

In Chapter 4, the *in vivo* transcriptome dynamics of *B. malayi* during its obligatory intracellular developmental phase in the mosquito host was investigated (Bartholomay and Christensen, 2002; Erickson et al., 2009). Integrative analysis of the host transcriptome in a dual RNA-seq approach provided high-resolution overview of the parasite-host system, in which each partner's transcriptional state is dependent on the other partner. It was the advances in RNA-seq technology that allowed direct analysis of mixed-species samples, in which host tissues contain minute quantities of parasite material (Westermann et al., 2012). With the increase in sequencing depth, RNA-seq offers improved levels of sensitivity and dynamic range of detection, without the need of predefined species-specific probes, unlike methods that rely on hybridization of targeted oligonucleotides. A time-course analysis from the establishment of infection to the emergence of L3 revealed that *B. malayi* transcriptome exhibits a highly ordered developmental

program consisting of a series of cyclical and state-transitioning temporal patterns. In the host tissue during infection, transcriptional changes were observed in metabolic and stress response genes with distinct kinetics that seem to reflect the degree and nature of cellular insults that occur at different stages of parasite development. These data will be foundational for future studies on the molecular processes underlying filarial worm development in, and interaction with, the mosquito host tissue, which remain poorly understood despite their critical importance in transmission and completion of the parasite lifecycle.

The extensive RNA-seq data generated in the present work (Chapters 3 and 4) effectively cover the entire lifecycle stages of *B. malayi*. These present a unique opportunity to further evaluate the genome-wide transcriptional changes in relation to the host transitions in this heteroxenous parasite, where different life stages pass through radically different host environments. By comparing mRNA expression during the vector stages relative to the mammalian stages, one could investigate the transcriptional features distinctive of the mosquito-borne parasitic lifestyle. Particularly useful in this analysis would be the rapidly growing genomic and transcriptomic information from other parasitic and free-living nematodes (Blaxter et al., 2012). Joint analysis of expression data and phylogenetic history could provide deeper insight into the molecular basis of filarial worm-mosquito symbiosis and nematode parasitism.

An exciting recent development in research methodology is a novel approach to parasite transcript suppression that introduces an RNAi trigger into the site of nematode development, i.e., the mosquito thoracic tissue, as opposed to the conventional *in vitro* soaking approach (Song et al., 2010). This protocol has been demonstrated to effectively knock down *B. malayi* transcripts despite the fact that animal parasitic nematodes are notoriously intractable to RNAi-based gene silencing (Geldhof et al., 2007; Maule et al., 2011). This approach opens the possibility to disrupt

gene expression in either the parasite or the mosquito, or both at the same time in their *in vivo* settings. The method allows targeting of specific genes and their associated pathways at discrete stages of parasite development in a least intrusive way such that the progression of the mosquito-parasite interaction can be studied post-gene perturbation (Song et al., 2010). By critically assessing knockdown effects, both at the organismal level and at the molecular level using dual RNA-seq approaches, one can test the hypothesis that the effects of perturbing the gene network of one organism during infection will impact the gene network of its interacting partner, and begin to uncover clues to the nature and topology of the inter-species gene interactions that underlie filarial worm development and mosquito vector competence.

REFERENCES

- Bartholomay, L.C., Christensen, B.M., 2002. Vector-parasite interactions in mosquito-borne filariasis., in: Klei, T., Rajan, T. (Eds.), *The Filaria*, 1 ed. Kluwer Academic Publishers, Boston, pp. 9-19.
- Bennuru, S., Meng, Z., Ribeiro, J.M., Semnani, R.T., Ghedin, E., Chan, K., Lucas, D.A., Veenstra, T.D., Nutman, T.B., 2011. Stage-specific proteomic expression patterns of the human filarial parasite *Brugia malayi* and its endosymbiont *Wolbachia*. *Proceedings of the National Academy of Sciences of the United States of America* 108, 9649-9654.
- Bennuru, S., Semnani, R., Meng, Z., Ribeiro, J.M., Veenstra, T.D., Nutman, T.B., 2009. *Brugia malayi* excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling. *PLoS Negl Trop Dis* 3, e410.
- Blaxter, M., Kumar, S., Kaur, G., Koutsovoulos, G., Elsworth, B., 2012. Genomics and transcriptomics across the diversity of the Nematoda. *Parasite immunology* 34, 108-120.
- Crowther, G.J., Shanmugam, D., Carmona, S.J., Doyle, M.A., Hertz-Fowler, C., Berriman, M., Nwaka, S., Ralph, S.A., Roos, D.S., Van Voorhis, W.C., Agüero, F., 2010. Identification of attractive drug targets in neglected-disease pathogens using an *in silico* approach. *PLoS Negl Trop Dis* 4, e804.
- Erickson, S.M., Xi, Z., Mayhew, G.F., Ramirez, J.L., Aliota, M.T., Christensen, B.M., Dimopoulos, G., 2009. Mosquito infection responses to developing filarial worms. *PLoS neglected tropical diseases* 3, e529.

Geldhof, P., Visser, A., Clark, D., Saunders, G., Britton, C., Gilleard, J., Berriman, M., Knox, D., 2007. RNA interference in parasitic helminths: current situation, potential pitfalls and future prospects. *Parasitology* 134, 609-619.

Griffiths, K.G., Mayhew, G.F., Zink, R.L., Erickson, S.M., Fuchs, J.F., McDermott, C.M., Christensen, B.M., Michalski, M.L., 2009. Use of microarray hybridization to identify *Brugia* genes involved in mosquito infectivity. *Parasitol Res* 106, 227-235.

Hillyer, J.F., 2010. Mosquito immunity. *Adv Exp Med Biol* 708, 218-238.

Irving, P., Ubeda, J.M., Doucet, D., Troxler, L., Lagueux, M., Zachary, D., Hoffmann, J.A., Hetru, C., Meister, M., 2005. New insights into *Drosophila* larval haemocyte functions through genome-wide analysis. *Cellular microbiology* 7, 335-350.

Kumar, S., Chaudhary, K., Foster, J.M., Novelli, J.F., Zhang, Y., Wang, S., Spiro, D., Ghedin, E., Carlow, C.K., 2007. Mining predicted essential genes of *Brugia malayi* for nematode drug targets. *PLoS ONE* 2, e1189.

Li, B.W., Rush, A.C., Jiang, D.J., Mitreva, M., Abubucker, S., Weil, G.J., 2011. Gender-associated genes in filarial nematodes are important for reproduction and potential intervention targets. *PLoS neglected tropical diseases* 5, e947.

Li, B.W., Rush, A.C., Mitreva, M., Yin, Y., Spiro, D., Ghedin, E., Weil, G.J., 2009. Transcriptomes and pathways associated with infectivity, survival and immunogenicity in *Brugia malayi* L3. *BMC genomics* 10, 267.

Maule, A.G., McVeigh, P., Dalzell, J.J., Atkinson, L., Mousley, A., Marks, N.J., 2011. An eye on RNAi in nematode parasites. *Trends in parasitology* 27, 505-513.

Moreno, Y., Geary, T.G., 2008. Stage- and gender-specific proteomic analysis of *Brugia malayi* excretory-secretory products. *PLoS Negl Trop Dis* 2, e326.

Pinto, S.B., Lombardo, F., Koutsos, A.C., Waterhouse, R.M., McKay, K., An, C., Ramakrishnan, C., Kafatos, F.C., Michel, K., 2009. Discovery of Plasmodium modulators by genome-wide analysis of circulating hemocytes in *Anopheles gambiae*. *Proc Natl Acad Sci U S A* 106, 21270-21275.

Piro, R.M., Ala, U., Molineris, I., Grassi, E., Bracco, C., Perego, G.P., Provero, P., Di Cunto, F., 2011. An atlas of tissue-specific conserved coexpression for functional annotation and disease gene prediction. *Eur J Hum Genet* 19, 1173-1180.

Song, C., Gallup, J.M., Day, T.A., Bartholomay, L.C., Kimber, M.J., 2010. Development of an in vivo RNAi protocol to investigate gene function in the filarial nematode, *Brugia malayi*. *PLoS Pathog* 6, e1001239.

Taylor, C.M., Fischer, K., Abubucker, S., Wang, Z., Martin, J., Jiang, D., Magliano, M., Rosso, M.N., Li, B.W., Fischer, P.U., Mitreva, M., 2011. Targeting protein-protein interactions for parasite control. *PLoS One* 6, e18381.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews* 10, 57-63.

Waterhouse, R.M., Kriventseva, E.V., Meister, S., Xi, Z., Alvarez, K.S., Bartholomay, L.C., Barillas-Mury, C., Bian, G., Blandin, S., Christensen, B.M., Dong, Y., Jiang, H., Kanost, M.R., Koutsos, A.C., Levashina, E.A., Li, J., Ligoxygakis, P., Maccallum, R.M., Mayhew, G.F., Mendes, A., Michel, K., Osta, M.A., Paskewitz, S., Shin, S.W., Vlachou, D., Wang, L., Wei, W., Zheng, L., Zou, Z., Severson, D.W., Raikhel, A.S., Kafatos, F.C., Dimopoulos, G., Zdobnov, E.M., Christophides, G.K., 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* 316, 1738-1743.

Westermann, A.J., Gorski, S.A., Vogel, J., 2012. Dual RNA-seq of pathogen and host. *Nature reviews. Microbiology* 10, 618-630.

WHO, 2002. *The World Health Report 2002: reducing risks promoting healthy life*. The World Health Organization, Geneva, Switzerland, pp. 186-197.