Available online at www.sciencedirect.com
ScienceDirect
Genomics 88 (2006) 831–840
GENOMICS
www.elsevier.com/locate/ygeno

# Analysis of ESTs from *Lutzomyia longipalpis* sand flies and their contribution toward understanding the insect–parasite relationship ☆

Rod J. Dillon [a,*], Al C. Ivens [b], Carol Churcher [b], Nancy Holroyd [b], Michael A. Quail [b], Matthew E. Rogers [a,1], M. Bento Soares [c], Maria F. Bonaldo [c], Thomas L. Casavant [d], Mike J. Lehane [a,2], Paul A. Bates [a,2]

[a] *Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK*
[b] *The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*
[c] *Children's Memorial Research Center and Northwestern University, Chicago, IL 60611, USA*
[d] *Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA*

Received 3 May 2006; accepted 20 June 2006
Available online 1 August 2006

**Abstract**

An expressed sequence tag library has been generated from a sand fly vector of visceral leishmaniasis, *Lutzomyia longipalpis.* A normalized cDNA library was constructed from whole adults and 16,608 clones were sequenced from both ends and assembled into 10,203 contigs and singlets. Of these 58% showed significant similarity to known genes from other organisms, <4% were identical to described sand fly genes, and 42% had no match to any database sequence. Our analyses revealed putative proteins involved in the barrier function of the gut (peritrophins, microvillar proteins, glutamine synthase), digestive physiology (secreted and membrane-anchored hydrolytic enzymes), and the immune response (gram-negative binding proteins, thioester proteins, scavenger receptors, galectins, signaling pathway factors, caspases, serpins, and peroxidases). Sequence analysis of this transcriptome dataset has provided new insights into genes that might be associated with the response of the vector to the development of *Leishmania.*


*Keywords:* *Lutzomyia*; Expressed sequence tag; *Leishmania*; Genomics; Immunity; Parasite; Midgut

The leishmaniases are a group of important neglected diseases with approximately 2,000,000 new cases every year and one-tenth of the world's population at risk of infection (www.who.int/tdr/diseases/leish). Symptoms of leishmaniasis range from relatively benign cutaneous disease through to potentially fatal visceral disease. The various parasites are all transmitted by certain species of female phlebotomine sand flies, and of these, *Lutzomyia longipalpis* is particularly significant, being the main vector of visceral leishmaniasis in South America [1]. The global risk of leishmaniasis is increasing, and the colonization of urban areas by *Lu. longipalpis* appears to be a significant factor in the recent increase in visceral leishmaniasis in South America [2]. Unfortunately, there are no vaccines or prophylactic drugs for leishmaniasis currently available, and chemotherapy is reliant on a small number of drugs. These factors indicate that control of the sand fly vector will remain an important component of leishmaniasis control for the foreseeable future [3].

Rearing the diminutive sand fly under laboratory conditions is a challenging process, and the limited amount of biological material that can be obtained from sand flies, for example, in

*Abbreviations:* GALE, galectin; AMP, antimicrobial peptide; GNBP, gram-negative binding protein; PAMP, pathogen-associated molecular pattern; PGRP, peptidoglycan recognition proteins; PRR, pattern recognition receptor; TEP, thioester-containing protein family; SCR, scavenger receptor.

☆ Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AM088777–AM0109845.

* Corresponding author. Fax: +44 151 705 3371.
*E-mail address:* r.j.dillon@liv.ac.uk (R.J. Dillon).

[1] Present address: Department of Immunology, Imperial College of Science and Technology, London, UK.

[2] These authors contributed equally to this work.

0888-7543 
doi:10.1016/j.ygeno.2006.06.011

comparison to mosquitoes and *Drosophila,* has been an obstacle to the study of responses to the *Leishmania* parasite. A major step forward will be to develop transcriptome information for the sand fly vector to accompany that now available for the parasites. For example, following the publication of the *Drosophila* and *Anopheles* mosquito genomes there has been rapid progress in the use of transcriptome information from these insects and the development of microarrays to study insect gut–microbe interactions (e.g., [4,5]). In contrast, there have been remarkably few molecular studies of any kind examining sand fly genes that might influence *Leishmania* development. Those performed to date include a study of secreted salivary gland proteins [6], characterization of certain midgut digestive enzymes [7,8], a differential expression study [9], characterization of a sand fly defensin [10], and the identification of a midgut epithelial galectin implicated in binding of the *Leishmania* parasite [11].

The gut of the hematophagous insect is a potentially nutrient-rich but highly specialized environmental niche, and the successful development of ingested potential pathogens or parasites such as *Leishmania* depends on their ability to avoid or adapt to the dramatic changes in the physicochemical environment accompanying blood-meal and sugar-meal digestion. The strategy of the mosquito-borne malaria parasite is to exit rapidly through the gut epithelium and continue development in the hemocoel. In contrast African trypanosomes in tsetse flies [12] and *Leishmania* [13] have adapted to remaining and developing in the insect gut. The *Leishmania* parasite is supremely adapted to the gut environment of the sand fly, secreting a unique gel-like material composed mainly of a high-molecular-weight filamentous proteophosphoglycan (fPPG [14]). *Leishmania* fPPG serves a dual function, first blocking the fly gut and improving chances for transmission and subsequently aiding survival of the parasite in the mammalian host [15,16]. Although the parasites are confined to the gut lumen, *Leishmania* is expected to have a wider impact on gene regulation in other tissues such as the fat body and ovaries. Therefore, a whole-body-derived cDNA library was generated in the current study.

Interpretation of the resulting data is helped by the order Diptera containing the two best studied insect genomes, *Anopheles gambiae* and *Drosophila melanogaster,* and information on two other hematophagous Diptera has also recently become available: the tsetse fly *Glossina morsitans* (http://www.sanger.ac.uk/Projects/G_morsitans/) [17], and the mosquito *Aedes aegypti* (http://www.tigr.org/msc/aedes/aedes.shtml) [18]. For comparative purposes, it should be noted that phlebotomine sand flies are more closely related to *Anopheles* and *Aedes,* belonging to the dipteran suborder Nematocera along with many bloodsucking insects (mosquitoes, blackflies, and biting "midges"), whereas *Glossina* and *Drosophila* are found in the other suborder (Brachycera). The availability of these dipteran genome resources has facilitated the sequence identification and annotation of the *Lutzomyia* data described here and online (http://www.sanger.ac.uk/Projects/L_longipalpis/). This transcriptome study will provide the platform for the development of microarrays and will provide further impetus to identifying the insect genes involved in regulating *Leishmania* development in the vector.

## Results and discussion

A total of 16,608 cDNA clones from a normalized library of female *Lu. longipalpis* were sequenced from both ends. A total of 33,216 reads were attempted and of these 26,495 were successful (80% pass rate). An additional 1728 reads were attempted from the unnormalized cDNA library, with 1433 obtained. The assembly with Phrap [19] generated 5210 contigs (mean length, 1225 bp) and 4993 singlets (mean length, 605 bp), giving a total of 10,203 ESTs. The average number of reads per contig was 4.4. Most of the sequences were novel sand fly sequences; a comparison of assembled contigs with the 1309 *Lutzomyia* spp. (highly redundant) DNA sequences available in the public databases (August 2005) revealed only 222 similar sequences occurring among the contigs with an E value of less than $10^{-25}$. Comparison with 379 *Phlebotomus* spp. DNA sequences revealed 160 hits with the same E value.

The sequences were compared using Blastx to the UniProt database to identify the number of transcripts without a significant match and thereby obtain an estimate of putative novel genes (Supplementary Table S1). A total of 5962 (58.4%) sequences had matches at the $E=10^{-5}$ cutoff, i.e., were similar to known genes; 1624 had no hits and 2617 failed to meet the E value threshold; thus up to 4241 (41.6%) of the sequences may be novel. Comparison with *Drosophila* and *Anopheles* databases gave similar results, with estimates of sequences possessing no similarity of 44.7 and 45.9%, respectively.

The sand fly expressed sequence tags (ESTs) were categorized by selected GO terms (Supplementary Fig. S1). The proportions of GO terms are similar to those found in the *Drosophila* proteome analysis (http://www.ebi.ac.uk/integr8/). A total of 6460 (63.3%) sequences were assigned a putative molecular function term by transitive annotation of GO terms. It was not possible to give an accurate estimate of the proportion of the *Lutzomyia* transcriptome sequenced in the EST study. Some transcripts, for example, may represent nonoverlapping parts of the same gene and the annotation was largely automated. However, the current release of Ensembl (version 36—December 2005) lists over 14,300 genes in the *An. gambiae* genome and 5517 of our sequences had matches with *Anopheles* proteins; therefore, considering the library was derived solely from adult female sand flies it was apparent that a large proportion of the predicted *Lutzomyia* genes were represented in this study.

The cDNA was synthesized from a pool of RNA extracted from whole bodies of sand flies, some of which were infected with *Leishmania infantum, Le. mexicana,* or bacteria. The rationale was to produce a wide range of cDNAs that could be used to construct a cDNA microarray to explore gene expression throughout the whole insect in response to *Leishmania* or microbial infections. *Lu. longipalpis* is a permissive vector allowing the development of *Le. mexicana* as well as the naturally occurring species *Le. infantum;* thus including *Le. mexicana*-infected insects will allow comparisons