

Published in final edited form as:

*J Proteomics*. 2014 January 16; 96: 117–132. doi:10.1016/j.jprot.2013.10.028.

## Expression of the mevalonate pathway enzymes in the *Lutzomyia longipalpis* (Diptera: Psychodidae) sex pheromone gland demonstrated by an integrated proteomic approach

Natalia González-Caballero<sup>a</sup>, Andrés Rodríguez-Vega<sup>b</sup>, Geovane Dias-Lopes<sup>c</sup>, Jesus G. Valenzuela<sup>d</sup>, Jose M.C. Ribeiro<sup>e</sup>, Paulo Costa Carvalho<sup>f</sup>, Richard H. Valente<sup>g</sup>, Reginaldo P. Brazil<sup>a</sup>, and Patricia Cuervo<sup>b,\*</sup>

Patricia Cuervo: pcuervo@fiocruz.br

<sup>a</sup>Laboratório de Bioquímica e Fisiologia de Insetos, IOC, FIOCRUZ, Rio de Janeiro, RJ, Brazil

<sup>b</sup>Laboratório de Pesquisa em Leishmaniose, IOC, FIOCRUZ, Rio de Janeiro, RJ, Brazil

<sup>c</sup>Pós-graduação Biologia Parasitaria, IOC, FIOCRUZ, Rio de Janeiro, RJ, Brazil

<sup>d</sup>Vector Molecular Biology Section, Laboratory of Malaria and Vector Research, National Institutes of Health Rockville, MD, USA

<sup>e</sup>Vector Biology Section, Laboratory of Malaria and Vector Research, National Institutes of Health Rockville, MD, USA

<sup>f</sup>Laboratório de Proteômica e Engenharia de Proteínas, Instituto Carlos Chagas, FIOCRUZ, Curitiba, Brazil

<sup>g</sup>Laboratório de Toxinologia, IOC, FIOCRUZ, Rio de Janeiro, RJ, Brazil

### Abstract

In Latin America, *Lutzomyia longipalpis* is the main vector of the protozoan parasite *Leishmania infantum*, which is the causal agent of American Visceral Leishmaniasis. This insect uses male-produced pheromones for mate recognition. Elucidation of pheromone biogenesis or its regulation may enable molecular strategies for mating disruption and, consequently, the vector's population management. Motivated by our recent results of the transcriptomic characterization of the *L. longipalpis* pheromone gland, we performed a proteomic analysis of this tissue combining SDS-PAGE, and mass spectrometry followed by an integrative data analysis. Considering that annotated genome sequences of this sand fly are not available, we designed an alternative workflow searching MS/MS data against two customized databases using three search engines: Mascot, OMSSA and ProLuCID. A total of 542 proteins were confidently characterized, 445 of them using a Uniref100-insect protein database, and 97 using a transcript translated database. In addition, use of PEAKS for de novo peptide sequencing of MS/MS data confirmed ~90% identifications made with the combination of the three search engines. Our results include the identification of six of the seven enzymes of the mevalonate-pathway, plus the enzymes involved in sesquiterpenoid biosynthesis, all of which are proposed to be involved in pheromone production in *L. longipalpis*.

© 2013 Elsevier B.V. All rights reserved.

\*Corresponding author at: Laboratório de Pesquisa em Leishmaniose, Instituto Oswaldo Cruz, FIOCRUZ, Av. Brasil 4365, Manguinhos, Pav. Leônidas Deane, Sala 509 Rio de Janeiro, RJ 21040-360, Brasil. Tel.: +55 21 3865 8224.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jprot.2013.10.028>.

**Competing interests:** The authors declare that they have no competing interests.

**Biological significance**—*L. longipalpis* is the main vector of the protozoan parasite *L. infantum*, which is the causal agent of American Visceral Leishmaniasis. One of the control measures of such disease is focused on vector population control. As this insect uses male-produced pheromones for mate recognition, the elucidation of pheromone biogenesis or its regulating process may enable molecular strategies for mating disruption and, consequently, this vector's population management. On this regard, in this manuscript we report expression evidence, at the protein level, of several molecules potentially involved in the pheromone production of *L. longipalpis*. Our results include the identification of the mevalonate-pathway enzymes, plus the enzymes involved in sesquiterpenoid biosynthesis, all of which are proposed to be involved in pheromone production in *L. longipalpis*. In addition, considering that the annotated genome sequences of this sand fly are not yet available, we designed an alternative workflow searching MS/MS data against proteomic and transcript translated customized databases, using three search engines: Mascot, OMSSA, and ProLuCID. In addition, a de novo peptide sequencing software (PEAKS) was used to further analyze the MS/MS data. This approach made it possible to identify and annotate 542 proteins for the pheromone gland of *L. longipalpis*. Importantly, all annotated protein sequences and raw data are available for the research community in protein repositories that provide free access to the data.

## Keywords

*Lutzomyia longipalpis*; Male pheromone gland; Proteomics; Proteome; Mevalonate pathway

## 1. Introduction

Visceral leishmaniasis (VL) is the most severe form of leishmaniasis, with a global annual incidence of 0.2 to 0.4 million cases, of which 90% occur in India, Bangladesh, Sudan, South Sudan, Ethiopia and Brazil [1]. In the Neotropics, this disease is produced by the protozoan parasite *Leishmania infantum* [2], which is mainly disseminated by the sand fly *Lutzomyia longipalpis* (Lutz & Neiva, 1912) Psychodidae: Phlebotominae (reviewed in [3]). This sand fly species can be found in different habitats ranging from Mexico to northern Argentina [4–6]. As adult sand flies, females and males can find nutrients from natural sources such as honeydew or plant juices [7]. Alternatively, females can feed on animal blood, which is necessary for successful oviposition [8]. During the blood-feeding process, infected sand flies can transmit *Leishmania* to animals present in a variety of epidemiological cycles [9]. Although parasite transmission occurs primarily in rural areas, recent reports have shown the increase of VL incidence around large cities [10–12]. Such adaptation to human-modified environments reflects this vector's ability to occupy a wide range of ecologic niches (reviewed in [3]).

Male sand fly aggregation occurs near, or on, vertebrate hosts during female blood-feeding. Males arrive to the vertebrate hosts, before females, triggering lure dynamics that attract females. The first *L. longipalpis* male arriving to a vertebrate host primes the aggregation formation by producing pheromones that attract other males, which in turn produce more pheromones attracting even more flies [13]. Consequently, virgin females are attracted to male pheromones which are synergized by the host's odor [14].

The main components of sex pheromone blends in different populations of *L. longipalpis* have been characterized as terpenoid compounds. To date, two different homosesquiterpenes (C16) have been structurally identified as (S)-9-methylgermacrene-B (9MGB) and 3-methyl- $\alpha$ -himachalene (3MaH) [15–18]. Two other terpenoid monocyclic (C20) pheromones have been partially identified as cembrene isomers referred to as cembrene 1 and cembrene 2 [19,20]. Terpenoids (or isoprenoids) originate through the condensation of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are the assembling block

molecules (C<sub>5</sub>) of nearly all isoprenoid [21]. In higher eukaryotes and some bacteria, IPP and DMAPP are produced by a cellular metabolic pathway described as the mevalonate pathway or mevalonate dependent pathway MDP [22]. Three acetyl-CoA molecules need to be condensed to generate IPP and its isomer DMAPP. IPP then undergoes successive condensations with DMAPP to produce farnesyl pyrophosphate synthase FPP [23]. In most animals, the most studied final product of the mevalonate pathway is cholesterol because of its implication with cardiovascular disease [22]. Insects, however, cannot synthesize cholesterol *de novo* [24] because they are deficient in enzymes such as the squalene synthase and the lanosterol synthase [25–27]. Consequently, and contrary to those reactions needed to obtain FPP from acetyl-CoA, the enzymes associated to the mevalonate pathway are not homologous to those of cholesterol-producing organisms (reviewed in [28]). In insects, these reactions have been mostly described as result of the elucidation of the juvenile hormone (JH) biosynthetic steps, which occurs in a glandular tissue, the corpora allata (CA) [29–32].

In *L. longipalpis*, terpenoid sex pheromones are produced in glands located under the cuticle of pale patches on male abdominal segments. In these pale patch structures a number of gland cells are connected to the exterior via small cuticular ducts [33]. Recent studies correlating pheromone production to glandular cell morphogenesis have reported that pheromone biosynthesis begins approximately 12 h after adult emergence and increases constantly during the first three days, stabilizing afterward [34]. Different *L. longipalpis* populations produce and respond to different sex pheromones, a characteristic that has been exploited for phylogenetic purposes [20]. It has been proposed that pheromone biosynthesis in *L. longipalpis* involves the enzymes of the mevalonate pathway as well as prenyltransferases [35]; however, evidence for the expression of these enzymes, at the protein level, in the pheromone gland of the sand fly is lacking.

The elucidation of pheromone biogenesis or the mechanisms that may be regulating this process may enable molecular strategies for mating disruption and, consequently, sand fly population management. In this regard, describing the profile of proteins in the pheromone gland represents a first step toward comprehending the biosynthetic process occurring in this specialized tissue. Motivated by our recent results on the transcriptomic characterization of *L. longipalpis*' pheromone gland that reports mRNA evidence for several enzymes of the mevalonate pathway [35], we performed a robust and in-depth bioinformatics analysis of high-resolution proteomic data of this tissue to pinpoint protein evidence, correlated to transcriptomic data, and ultimately provide insights on the pheromone glands protein functions. As genome sequences of *L. longipalpis* are not yet available, to maximize the proteomic analysis of the pheromone gland, the MS/MS datasets obtained in this study were analyzed using three different search engines combined with a customized insect protein database and also our own transcript translated database. This workflow enabled a highly confident proteomic characterization of the pheromone gland and allowed the identification of mevalonate pathway enzymes potentially involved in the *L. longipalpis* pheromone production.

## 2. Methods

### 2.1. Chemicals

Laboratory chemicals were purchased from Sigma Chemicals St. Louis, MO, USA or Merck São Paulo, SP, Brazil. MilliQ-purified water Millipore Corp., Bedford, MA, USA was used to prepare all solutions.

## 2.2. Insects

The experiments were performed using 3 to 7 day old *L. longipalpis* Diptera: Psychodidae one-spot adult males. The insects were captured from Gruta da Lapinha, Municipality of Lagoa Santa, Minas Gerais, Brazil, 19° 38S; 43° 53 W and colonized at the Laboratório de Bioquímica e Fisiologia de Insetos at the Instituto Oswaldo Cruz, Rio de Janeiro, Brazil. Colony maintenance procedures followed those previously described [8] including a temperature of 26 °C, 80% relative humidity, and 12:12 LD photoperiod. We used insects from the 1st to 7th generations after checking for their (S)-9-methylgermacrene-B production according to [15]. Permission for sand fly collection was obtained from the Brazilian Ministry of Environment (SISBIO#26066-1).

## 2.3. Pheromone gland dissection and protein extraction

Cold-anesthetized male insects were carefully dissected using a stereoscopic microscope 6× to 40× in sterilized conditions in order to minimize any risk of contamination. Following decapitation and midgut removal, insect abdomens were washed in a sterilized 0.9% saline solution and ultrapure water. Pheromone glands from one hundred males were dissected, pooled and lysed in 100 µL of 2 M thiourea, 7 M urea, 4% CHAPS, 65 mM DTT and a protease inhibitor cocktail (complete protease inhibitor cocktail, Roche) that is intended to inhibit serine, cysteine and metalloproteases. Lysis proceeded with ten cycles of freeze (liquid nitrogen)/thaw/sonication. Sonication was carried out with a Fisher Sonic Dismembrator with an output frequency of 20 kHz, 15% amplitude for 20 s. To remove insoluble material, the samples were centrifuged at 10,000 ×g for 10 min at 4 °C, the resulting supernatant was precipitated with methanol: chloroform, 3:1. The proteins were resuspended in SDS sample buffer and the total protein concentration was measured using the Thermo Scientific Pierce 660 nm Protein Assay.

## 2.4. SDS-PAGE, protein digestion and peptide extraction

Aliquots of pheromone gland extracts containing 40 µg of proteins were separated by 12% SDS-PAGE 30% acrylamide, 0.8% bis-acrylamide. The proteins were stained using Coomassie Brilliant Blue and photo-documented using a GS-800™ calibrated imaging densitometer Bio-Rad. The proteins were enzymatically digested following procedures previously described [36] with some modifications. Briefly, lanes containing proteins were finely sliced and each slice was destained three times in 400 µL of 50% acetonitrile, 25 mM NH<sub>4</sub>HCO<sub>3</sub> pH 8.0 for 15 min. The proteins were reduced in 65 mM of DTT for 30 min at 56 °C and then alkylated with 200 mM iodoacetamide at 25 °C in darkness for 30 min. The gel slices were washed with 100 mM NH<sub>4</sub>HCO<sub>3</sub> for 10 min followed by dehydration with acetonitrile for 5 min. The slices were rehydrated with a solution of 20 ng/µL of sequencing grade modified porcine trypsin (Promega, USA) in 50 mM NH<sub>4</sub>HCO<sub>3</sub> and incubated overnight at 37 °C. Peptides were extracted using 0.1% formic acid in 50% v/v acetonitrile, desalted and concentrated with Poros oligo R3 resin (Applied Biosystems, USA).

## 2.5. Mass spectrometry analysis

Tryptic digests, obtained as described in the previous step, were submitted to reversed-phase nanochromatography coupled to nanoelectrospray high resolution mass spectrometry for identification. For each sample, 4 µL of desalted tryptic peptide digest were initially applied to a 2-cm long (100 µm internal diameter) trap column packed with 5 µm, 200A Magic C18 AQ matrix (Michrom Bioresources, USA) followed by separation on a 10.5-cm long (75 µm internal diameter) separation column that was packed with the same matrix, directly on a self-pack 15 µm PicoFrit empty column (New Objective, USA). Chromatography was carried out on an EASY-nLC II instrument (Thermo Scientific, USA). The samples were loaded onto the trap column at 2000 nL/min while chromatographic separation occurred at

200 nL/min. Mobile phase A consisted of 0.1% (v/v) formic acid in water while mobile phase B consisted of 0.1% (v/v) formic acid in acetonitrile and gradient conditions were as follows: 2 to 40% B in 32 min; up to 80% B in 4 min, maintaining at this concentration for 2 min more. Eluted peptides were directly introduced to an LTQ/Orbitrap XL mass spectrometer (Thermo, USA) for analysis. Source voltage was set to 1.9 kV, capillary temperature to 200 °C and tube lens voltage to 100 V. Ion trap full and MSn AGC target values were 30,000 and 10,000, respectively, while the FTMS full AGC target was set to 500,000. MS1 spectra were acquired on the Orbitrap analyzer (300 to 1,700  $m/z$ ) at a 60,000 resolution (for  $m/z = 445.1200$ ). For each spectra, the 10 most intense ions were submitted to CID fragmentation (minimum signal required of 10,000; isolation width of 2.5; normalized collision energy of 35.0; activation Q of 0.25 and activation time of 30 ms) followed by MS2 acquisition on the linear trap analyzer. Dynamic exclusion option was enabled and set with the following values for each parameter: repeat count = 1; repeat duration = 30 s; exclusion list size = 500; exclusion duration = 45 s and exclusion mass width = 10 ppm. Charge state rejection was enabled for unassigned charges and for those equal to one. The experiments were performed using three biological replicates.

## 2.6. Peaklist settings and databases searching

MS raw files for each lane were converted to mzXML or MGF format using Mass Matrix MS Data File converter V. 3.9 <http://www.massmatrix.net/mm-cgi/downloads.py> and to .MS2 format using RawReader in the PatternLab platform tools [37] <http://proteomics.fiocruz.br/Softwares.aspx>. In order to maximize search sensitivity and protein identification, data from each replicate were merged into a single file and searched using ProLuCID 1.3 in the PatternLab platform [38], OMSSA [39] in the Proteomic platform 1.2.1 [40] <http://www.proteomatic.org/download.html>, and Mascot Daemon 2.3.02 [http://www.matrixscience.com/daemon\\_support\\_v2\\_3.html](http://www.matrixscience.com/daemon_support_v2_3.html). As annotated genome sequences of *L. longipalpis* are not available, data were searched against a customized database including non-redundant sequences for the Insecta Class downloaded October 2012 from UniRef100 [41] <http://www.uniprot.org/> with 592,178 entries, and against the pheromone gland transcriptome protein database previously generated for *L. longipalpis* (1387 translated sequences [35]). Searches estimated peptide identification false discovery rates (FDRs) using identification from the reversed decoy database. Searches were performed with one missed cleavage, fixed modification of cysteine (carbamidomethylation), variable modification of methionine (oxidation) and mass tolerances of 2.0 and 0.8 Da in OMSSA, 10 ppm and 0.4 Da in Mascot, 50 ppm and 500 ppm in ProLuCID for precursor and fragmentions, respectively.

The validity of the peptide sequence matches (PSMs) searched with ProLuCID were assessed using the Search Engine Processor (SEPro)[37]. Briefly, identifications were grouped by charge state (+2 and +3) and then by tryptic status (fully tryptic, semi-tryptic), resulting in four distinct subgroups. For each result, the ProLuCID XCorr, DeltaCN and ZScore values were used to generate a Bayesian discriminator. The identifications were sorted in a non-decreasing order according to the discriminator score. A cutoff score was established to accept a false-discovery rate (FDR) of 1% based on the number of labeled decoys. This procedure was independently performed on each data subset, resulting in a false-positive rate that was independent of tryptic status or charge state. Additionally, a minimum sequence length of six amino acid residues was required. The results were post-processed to only accept PSMs with less than 10 ppm and proteins supported by two or more independent evidences (e.g., identification of a peptide with different charge states, a modified and a non-modified version of the same peptide, or two different peptides). This last filter led to a 0%FDR in all search results.

For the other search engines, peptide candidates were filtered with 2% peptide FDR and 2 peptides per protein minimum. In addition, data resulting from OMSSA searches were manually analyzed and cured. Peptides identified with each search engine were manually re-checked and clustered together if peptide identifications were shared among different insect species, as this indicates considerable sequence similarity. The identification in the closest species to *Lutzomyia* was considered as the leader protein. This leader protein was considered a unique protein identifier. Furthermore, proteins identified and cured from the multiple search engines were clustered and organized in a single table. In addition, MS data were searched against our own transcript translated FASTA database using ProLuCID tool [35]. Furthermore, de novo peptide sequencing and peptide-spectrum matching (PSM) of MS/MS spectra data were performed with PEAKS 6 (build 20120620) [90], using PEAKS De Novo and PEAKS DB options, respectively. PEAKS DB tool combines PSM results and de novo peptide sequencing data to match proteins in the target database to be searched. Such a tool was used in combination with our customized Insecta database. Proteins identified by PEAKS DB were filtered for a minimum score ( $-10\lg P$ ) of 40 and at least 2 peptides per protein. The acceptable false discovery rate for peptide identification was set to 1% at most.

The Gene Ontology Explorer tool at the PatternLab platform [42] was used to assign molecular and biological function to identified proteins. Moreover, for assignment of predicted functional domain, proteins of unknown function were analyzed with the InterProScan engine <http://www.ebi.ac.uk/Tools/pfa/iprscan/> [43,44]. In addition, functional associations among proteins potentially involved in pheromone production were confirmed by an in silico analysis using the STRING 9.0 server, a database that provides information about known and predicted protein interactions <http://www.string-db.org> [57,58]. The data was analyzed following the standards proposed under the Minimum Information About a Proteomic Experiment MIAPE consensus [45]. MS/MS data and identified proteins (Mascot and OMSSA) were deposited in the ProteomeXchange via the PRIDE repository [46]. The raw data together with all identifications via ProLuCID are available at <http://max.ioc.fiocruz.br/natalia130430/>.

### 3. Results and discussion

#### 3.1. Analysis of the pheromone gland proteome using Insecta database searching

Pheromone gland tissues from *L. longipalpis* males were dissected, pooled for protein extraction, and fractionated by SDS-PAGE (Fig. 1). These preparations represent a reliable tool for studying the pheromone gland of this sand fly. In general, when annotated genome sequences are available, the identification of MS-based proteomics data relies on the match of spectra to these sequences. However, there are no annotated genome sequences of *L. longipalpis* yet and data related to gene expression in pheromone glands from insects are also scarce. To deal with this obstacle, we designed an alternative data analysis method for protein identification searching MS/MS data against a database that included sequences available for Insecta Class in UniProt (UniRef100), totalizing 592,178 entries (October 2012). The UniRef100 database contains non-redundant reference clusters for proteins from all insect species found in UniProt. Our rationale was that this search space would be broad enough to allow identification of *L. longipalpis* proteins by similarity, but also narrow enough to diminish both the false positives and the time consumed by the search. In addition, in order to maximize our results, we used three different search engines that allowed a significant increase in peptide identification (Fig. 2). In total, 445 proteins were identified against the Insecta database (Supplementary Table 1). Among these proteins, 140 (31%) were commonly identified by the three engines and 316 (71%) were identified by at least two search engines (Fig. 3, Supplementary Table 2). The results obtained with each engine are presented in Supplementary Tables 3, 4 and 5. The use of the Insecta database

obtained from UniRef100 guaranteed us avoiding redundancies in protein identification as different insect species can exhibit many isoforms of the same protein. Therefore, the 445 proteins identified in the pheromone gland preparations correspond to unique protein entries. Based on these non-redundant clusters we proceeded to analyze the insect species or insect families that contributed more for the similarity-based identification of *L. longipalpis* pheromone gland proteins. We observed that *L. longipalpis* proteins were identified by similarity mostly to proteins of insects from the Diptera order (68%). Among Diptera, the Culicidae family, followed by the Drosophilidae family contributed 55% and 27% of the identifications, respectively (Fig. 4, Supplementary Table 6). The Psychodidae, which is the family that the *Lutzomyia* genus belongs to, contributed only 6% of the identifications (Fig. 4A). In fact, only 15 proteins were identified by match with *L. longipalpis* deposited sequences (Supplementary Table 1). Among the Culicidae species, *Aedes aegypti*, *Culex quinquefasciatus* and *Anopheles gambiae* contributed ~30% of the identifications (Fig. 4B). Our multi-engine search and identification approach was further verified by the PEAKS 6 de novo sequencing and database search (Fig. 2), which corroborated 90% of the identifications obtained with at least two engines (Supplementary Table 7).

### 3.2. Proteins involved in organic cyclic and heterocyclic compound binding enrich the pheromone gland proteome

Very little is known about *L. longipalpis* global or tissue specific protein expression patterns. Much less is known about protein function and functional associations. The use of the Gene Ontology Explorer tool showed that among the identified proteins ~46% have a molecular function assigned, ~35% participate in a known biological process, and only 19% could be assigned to a cellular component (Supplementary Fig. 1; Supplementary Table 8). Gene ontology classification based on molecular function showed that proteins involved in the binding of organic cyclic and heterocyclic compounds are enriched in our analysis, accounting for ~37% proteins. In addition, transferases and oxidoreductases accounted for ~12% of proteins with molecular functions assigned. The enzymes of the mevalonate pathway related to pheromone metabolism are classified in these categories. Additionally, the molecular function of hydrolases could be assigned to ~8% identified proteins. Hydrolases, specifically endopeptidases, are involved in the proteolysis process for farnesal production [47,48], which is an intermediate in the sesquiterpenoid backbone synthesis [49,50]. Our results also show that half of the identified proteins (49%, 218/445) are annotated as uncharacterized proteins or do not have any associated putative function (Supplementary Table 1). In an effort to obtain more information about these uncharacterized proteins we performed a search using the InterProScan tool to scan the PFAM Protein families database and the SMART domain database. Our results show that it was possible to assign PFAM signatures and functional domains to 94% (205/218) of the uncharacterized proteins (Supplementary Table 9). Enrichment of GO terms for the predicted functional domains of such uncharacterized proteins revealed that 43% are involved in binding, primarily in the binding of proteins and heterocyclic compounds (Supplementary Fig. 2). In addition, 13% of predicted domains are involved in oxidoreductase activity, 12% in catalytic activity, 10% in hydrolase activity, and 6% in transferase activity, among the more represented categories.

### 3.3. Pheromone gland protein identification is increased by the use of a customized transcript translated database

In order to enrich our identifications, we additionally used a ProLuCID search engine to analyze the MS/MS data against our own pheromone gland transcript translated database (1387 translated sequences [35]) (Fig. 2). This search yielded 97 protein identifications (Supplementary Table 10) among which ~12% (12/97) were common to proteins identified by ProLuCID via the Insecta protein database. We observed that the two searching

approaches obtained different portions of the proteome with little overlap between them. Searching against the Insecta Protein database gives a larger number of identifications than the transcript translated database. As transcriptome data may contain numerous short sequences and we set our identification to a minimum of two peptides, many protein identifications were discarded, explaining the smaller number found from the second database. However, examination using the transcript translated database allowed the identification of *L. longipalpis* proteins that could not be present in the Insecta protein database (Supplementary Table 10). Together, our proteomic analysis of the *L. longipalpis* pheromone gland generated 530 new non-redundant proteins. All data were deposited in the ProteomeXchange database.

The knowledge about *L. longipalpis* protein expression is very limited. Only 665 unreviewed protein entries for *Lutzomyia* spp. were available at the UniProt database when we conducted the present study. The workflow used here for proteomic characterization of pheromone gland tissue in *L. longipalpis* allowed us to report 515 new proteins for this species after discounting the 15 proteins identified as being of this species when the Insecta protein database was used. In addition, our analysis contributed to the annotation of a large portion of uncharacterized proteins. Searching against the two databases (Insecta protein database and transcript translated database) made possible the identification of proteins by increasing the chances of matches between the experimental and theoretical peptide data. Searching against the protein database derived from translation of transcript sequences of the *L. longipalpis* pheromone gland previously generated by our group [35] allowed a 16% (85/530) increase in protein identification using the mass spectra of this sand fly. In addition, the use of three search engines resulted in a 2-fold increase in the number of proteins identified with a 2 peptide minimum relative to a single search engine.

#### **3.4. The pheromone gland proteome includes enzymes of the mevalonate pathway, enzymes involved in sesquiterpenoid biosynthesis, and proteins associated to the pheromone production**

The proteins identified in the pheromone gland of *L. longipalpis* provide interesting molecule candidates for pheromone production in this insect (Table 1). We identified six of the seven enzymes of the mevalonate pathway as well as two enzymes involved in the sesquiterpenoid backbone biosynthesis (Fig. 5) [51,52]. In addition, we identified twelve proteins that are functionally related to the mevalonate pathway or have been reported as involved in pheromone or odor metabolism such as glutaminyl-tRNA synthetase, phosphoglycerate kinase, succinyl Co-A synthetase alpha subunit, transketolase, glutathione S transferases and cytochrome family proteins [53–56] (Table 1). Functional associations among proteins potentially involved in pheromone production could be confirmed by in silico analysis using the STRING database 9.0 [57,58]. Using this tool, it was possible to obtain protein association networks for 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) of *A. aegypti* and *C. quinquefasciatus*, which is the rate controlling enzyme of the mevalonate pathway [59] (Supplementary Fig. 3). All these potentially isoprenoid-pathway related sequences found on the pheromone gland tissue of *L. longipalpis* have previously been observed in other insects in the site where terpenoid production actively occurs [55,60,61]. Proteins with putative thiolase-2/SCP2 sterol-binding domains were identified by similarity with 4 peptides in the ant *Solenopsis invicta* and the bee *Apis mellifera*. This enzyme has been characterized in the pheromone gland of *Ostrinia scapularis* [62]. The 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) was identified with 3 peptides by similarity to the enzyme of *Blattella germanica* in the Insecta protein database. For this insect, it has been observed that the HMGS expression is developmentally regulated and the mevalonate synthesis parallels vitellogenin production [63,64]. The HMGR, identified with 17 and 2 distinct peptides in *C. quinquefasciatus* and *Drosophila persimilis*, respectively,

using the transcript translated database, has been reported in various Coleoptera to be involved in terpenoid pheromone production [65–69]. It has been shown that expression of HMGR in an established *Drosophila* cell line is regulated by mevalonate [70]. Despite having identified mevalonate kinase by similarity to *Bombix mori* with only 1 peptide in the Insecta protein database, we believe that this enzyme is present in the pheromone gland of *L. longipalpis* based on the reports of the presence of this enzyme in several sex pheromone-producing tissues in other insects [71]. The phosphomevalonate kinase (PMK) and the isopentenyl-diphosphate delta-isomerase (IDI) identified by similarity to *A. aegypti* proteins with 3 and 11 peptides, respectively, using the transcript database, have been described in the isoprenoid synthesis pathway of *B. mori* [72] and in the juvenile hormone pathway in *A. aegypti* [73]. Furthermore we also identified an additional type of isopentenyl-diphosphate delta-isomerase having 8 peptides (Table 1, Supplementary Table 10). These results reinforce our presumption of two IDI isoforms occurring in the pheromone gland of *L. longipalpis* [35]. In general, the enzymes of the mevalonate pathway have also been proposed to be involved in monoterpenoid pheromone production in the beetle *Ips pini* and other bark beetle species [61,71]. In these insects, mevalonate pathway enzymes are regulated by feeding and by the levels of the juvenile hormone JH III [74].

The *L. longipalpis* pheromones (S)-9-methylgermacrene-B as well as 3-methylhimachalene are homosesquiterpenes (C16) that have a branching pattern reminiscent to those observed in the faranal, the trail pheromone of *Monomorium pharaoni* [75], and the lepidopteran juvenile hormones JH I and JH II [76,77]. Based on the knowledge about the biosynthesis route of these natural homoterpenoid compounds, the construction of *L. longipalpis* homosesquiterpenes backbone possibly requires molecules derived from homomevalonate, resulting in sesquiterpene backbone molecules with an extra methylene group (CH<sub>2</sub>), such as the 2E,6E,8S-8-methylfarnesol proposed by Tashiro et al. in 2000 [78]. As demonstrated by Schooley et al. [79], the homomevalonate can be obtained by the condensation of propionic acids instead of acetic acid. Then, as in the classic mevalonate pathway, the homomevalonate derivatives advance through this pathway producing homofarnesyl pyrophosphate to finally generate homoterpene compounds [80]. Therefore, the occurrence of substrate specificity in *L. longipalpis* prenyltransferase and the cyclization of the products in the final reactions for pheromone biosynthesis represent interesting subjects for future studies.

In a previous study of the pheromone gland isolated of *L. longipalpis* and *Lutzomyia cruzi*, which is a putative *L. longipalpis* sibling species [81], numerous lipid droplets were described through transmission electron microscopy. Such findings were suggestive of a potential role for lipids in the pheromone synthesis, as precursors of acetyl-CoA [34,82–84]. Accordingly, our proteomic data revealed enrichment in several enzymes and proteins involved in lipid metabolism such, 3-ketoacyl-coenzyme A thiolase, acyl-coenzyme A oxidase (peroxisome), acyl-coenzyme A dehydrogenase (mitochondria), hydroxyacylcoenzyme A dehydrogenase, 2-hydroxyacyl-CoA lyase 1, acetyl-coenzyme A carboxylase, fatty acid synthase, fatty acid synthase S-acetyltransferase, short-chain dehydrogenase among others. We also identified acetyl-coenzyme A synthetase, which synthesizes acetyl CoA from acetate, adenosine triphosphate, and coenzyme A through an acetyl-adenosine monophosphate (AMP) intermediate [85,86]. Given the proximity of the fat body to the glandular cells [34], it is plausible suggest that this tissue may supply potential precursors, in the form of triacylglycerols stored in lipid droplets, for pheromone biosynthesis, as observed in moths [87,88]. Further studies may help to explore other functions of those lipid droplets in the *L. longipalpis* pheromone-producing cells, such as transient storage depots for proteins as observed in *Drosophila* [89].

## 4. Conclusion

The combination of SDS-PAGE, LC-MS/MS and a confident workflow for data mining, enabled us to perform a proteomic characterization of the 4th abdominal segment, containing the pheromone (9-methy-germacrene B)-producing cells of *L. longipalpis*. Considering that the annotated genome sequences of this sand fly are not yet available, we designed an alternative workflow searching MS/MS data against proteomic and transcript translated customized databases, using three search engines: Mascot, OMSSA, and ProLuCID, and further verification by PEAKS. This approach made it possible to obtain expression evidence, at the protein level, of several molecules potentially involved in the pheromone production of *L. longipalpis*. Having identified six of the seven enzymes of the classical mevalonate pathway and associated enzymes involved in the sesquiterpene biosynthetic pathway, we can now use a variety of tools to dissect their molecular correlation with pheromonogenesis by means of biochemical, genetic, and behavioral measurements. Additionally, our analysis also revealed a list of interesting candidate proteins that could be related to lipid metabolism, whose activity may result in providing precursors for the pheromone biosynthetic process. Most importantly, our results unraveled a range of molecules that might be directly associated with the pheromone production in *L. longipalpis*; such molecules can now be explored as potential targets for vector control.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Programa de Desenvolvimento Tecnológico em Insumos para Saúde (PDTIS-Fiocruz). J.M.C.R. and J.G.V. were supported by the Intramural Research Program of the Division of Intramural Research, NIAID, National Institutes of Health. Because J.M.C.R. and J.G.V. are government employees and this is a government work, the work is in the public domain in the United States. Notwithstanding any other agreements, the NIH reserves the right to provide the work to PubMedCentral for display and use by the public, and PubMedCentral may tag or modify the work consistent with its customary practices. You can establish rights outside of the U.S. subject to a government use license. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank Francisco de Assis Marra and Andressa Fuzari Rodrigues for their assistance with the insect maintenance and dissection, respectively. We are grateful also to Dr. Aniel Sanchez (CIGB, Cuba) for the help with the Mascot searches, and to the colleagues of the Laboratório de Pesquisa em Leishmaniose for their help with the protein purification and discussions.

## References

1. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, et al. Leishmaniasis worldwide and global estimates of its incidence. PLoS One. 2012; 7:e35671. [PubMed: 22693548]
2. Kuhls K, Alam MZ, Cupolillo E, Ferreira GEM, Mauricio IL, Oddone R, et al. Comparative microsatellite typing of new world *Leishmania infantum* reveals low heterogeneity among populations and its recent old world origin. PLoS Negl Trop Dis. 2011; 5:e1155. [PubMed: 21666787]
3. Lainson R, Rangel EF. *Lutzomyia longipalpis* and the eco-epidemiology of American visceral leishmaniasis, with particular reference to Brazil: a review. Mem Inst Oswaldo Cruz. 2005; 100:811–27. [PubMed: 16444411]
4. Galati, E.; Morfologia, E. taxonomia: Classificação de flebotominae. In: Rangel, E.; Lainson, R., editors. Flebotomíneos do Brasil. Rio de Janeiro: Editora Fiocruz; 2003. p. 23-52.
5. Young, DG.; Duncan, MA. Memoirs of the American Entomological Institute. Gainesville: Associated Publishers; 1994. Guide to the identification and geographic distribution of *Lutzomyia* sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae).

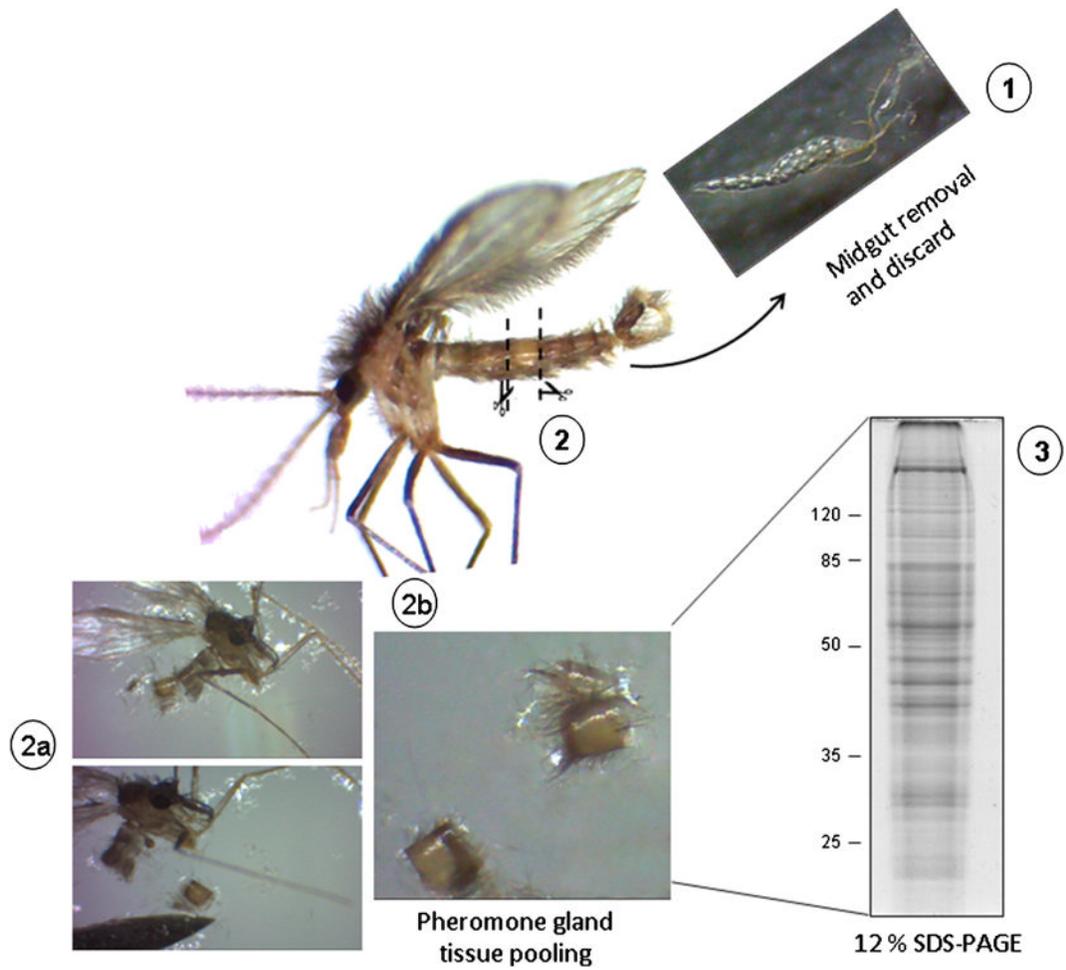
6. Grimaldi G, Tesh RB, McMahon-Pratt D. A review of the geographic distribution and epidemiology of leishmaniasis in the New World. *Am J Trop Med Hyg.* 1989; 41:687–725. [PubMed: 2701633]
7. Killick-Kendrick R. The biology and control of phlebotomine sand flies. *Clin Dermatol.* 1999; 17:279–89. [PubMed: 10384867]
8. Brazil, RP.; Brazil, BG. *Biologia de Flebotomíneos Tropicais*. In: Rangel, E.; Lainson, R., editors. *Flebotomíneos do Brasil*. Rio de Janeiro: Editora Fiocruz; 2003. p. 257-74.
9. Oliveira EF, Silva EA, Casaril AE, Fernandes CES, Paranhos Filho AC, Gamarra RM, et al. Behavioral aspects of *Lutzomyia longipalpis* (Diptera: Psychodidae) in urban area endemic for visceral leishmaniasis. *J Med Entomol.* 2013; 50:277–84. [PubMed: 23540114]
10. Jeronimo SMB, Duggal P, Braz RFS, Cheng C, Monteiro GRG, Nascimento ET, et al. An emerging peri-urban pattern of infection with *Leishmania chagasi*, the protozoan causing visceral leishmaniasis in Northeast Brazil. *Scand J Infect Dis.* 2004; 36:443–9. [PubMed: 15307565]
11. Werneck GL. Expansão geográfica da leishmaniose visceral no Brasil. *Cad Saude Publica.* 2010; 26:644–5. [PubMed: 20512199]
12. Humberg RM, Oshiro ET, Cruz Mdo S, Ribolla PE, Alonso DP, Ferreira AM, et al. *Leishmania chagasi* in opossums (*Didelphis albiventris*) in an urban area endemic for visceral leishmaniasis, Campo Grande, Mato Grosso do Sul, Brazil. *Am J Trop Med Hyg.* 2012; 87:470–2. [PubMed: 22802435]
13. Kelly DW, Dye C. Pheromones, kairomones and the aggregation dynamics of the sandfly *Lutzomyia longipalpis*. *Anim Behav.* 1997; 53:721–31.
14. Bray DP, Hamilton JGC. Host odor synergizes attraction of virgin female *Lutzomyia longipalpis* (Diptera: Psychodidae). *J Med Entomol.* 2007; 44:779–87. [PubMed: 17915508]
15. Hamilton JGC, Dawson GW, Pickett JA. 3-Methyl- $\alpha$ -himachalene: proposed structure for novel homosesquiterpene sex pheromone of *Lutzomyia longipalpis* (Diptera: Psychodidae) from Jacobina Brazil. *J Chem Ecol.* 1996; 22:2331–40. [PubMed: 24227307]
16. Hamilton JGC, Dawson GW, Pickett JA. 9-Methylgermacrene-B; proposed structure for novel homosesquiterpene from the sex pheromone glands of *Lutzomyia longipalpis* (Diptera: Psychodidae) from Lapinha, Brazil. *J Chem Ecol.* 1996; 22:1477–91. [PubMed: 24226250]
17. Hamilton JGC, Ibbotson HC, Hooper AM, Pickett JA. 9-Methylgermacrene-B is confirmed as the sex pheromone of the sandfly *Lutzomyia longipalpis* from Lapinha, Brazil, and the absolute stereochemistry defined as S. *Chem Commun.* 1999:2335–6.
18. Hamilton JGC, Hooper AM, Pickett JA, Mori K, Sano S. 3-Methyl- $\alpha$ -himachalene is confirmed, and the relative stereochemistry defined, by synthesis as the sex pheromone of the sandfly *Lutzomyia longipalpis* from Jacobina, Brazil. *Chem Commun.* 1999:355–6.
19. Hamilton JG, Brazil RP, Maingon R. A fourth chemotype of *Lutzomyia longipalpis* (Diptera: Psychodidae) from Jaibas, Minas Gerais State, Brazil. *J Med Entomol.* 2004; 41:1021–6. [PubMed: 15605640]
20. Hamilton JG, Maingon RD, Alexander B, Ward RD, Brazil RP. Analysis of the sex pheromone extract of individual male *Lutzomyia longipalpis* sandflies from six regions in Brazil. *Med Vet Entomol.* 2005; 19:480–8. [PubMed: 16336313]
21. Cane, DE. Isoprenoid biosynthesis: overview. In: Otto Meth-Cohn, SDB.; Koji, N., editors. *Comprehensive natural products chemistry*. Oxford: Pergamon; 1999. p. 1-13.
22. Goldstein JL, Brown MS. Regulation of the mevalonate pathway. *Nature.* 1990; 343:425–30. [PubMed: 1967820]
23. Bochar, DA.; Freisen, J.; Stauffacher, CV.; Rodwell, VW. Biosynthesis of mevalonic acid from acetyl-CoA. In: Otto Meth-Cohn, SDB.; Koji, N., editors. *Comprehensive natural products chemistry*. Oxford: Pergamon; 1999. p. 15-44.
24. Kircher, H. Cholesterol systems in insects and animals. Dupont, J., editor. Boca Raton: CRC Press Inc.; 1982. p. 1-50.
25. Beenackers AM, Van der Horst DJ, Van Marrewijk WJ. Insect lipids and lipoproteins, and their role in physiological processes. *Prog Lipid Res.* 1985; 24:19–67. [PubMed: 3916237]
26. Clark AJ, Bloch K. The absence of sterol synthesis in insects. *J Biol Chem.* 1959; 234:2578–82. [PubMed: 13810427]

27. Santos AC, Lehmann R. Isoprenoids control germ cell migration downstream of HMGCoA reductase. *Dev Cell*. 2004; 6:283–93. [PubMed: 14960281]
28. Bellés X, Martin D, Piulachs MD. The mevalonate pathway and the synthesis of juvenile hormone in insects. *Annu Rev Entomol*. 2005; 50:181–99. [PubMed: 15355237]
29. Feyereisen, R. Regulation of juvenile hormone titer: synthesis. In: Kerkut, G.; Gilbert, L., editors. *Comprehensive insect physiology, biochemistry and pharmacology*. Oxford: Pergamon Press; 1985. p. 391-429.
30. Schooley, D.; Baker, F. Juvenile hormone biosynthesis. In: Kerkut, G.; Gilbert, L., editors. *Comprehensive insect physiology, biochemistry and pharmacology*. Oxford: Pergamon Press; 1985. p. 363-89.
31. Ueda H, Shinoda T, Hiruma K. Spatial expression of the mevalonate enzymes involved in juvenile hormone biosynthesis in the corpora allata in *Bombyx mori*. *J Insect Physiol*. 2009; 55:798–804. [PubMed: 19414014]
32. Nouzova M, Edwards MJ, Mayoral JG, Noriega FG. A coordinated expression of biosynthetic enzymes controls the flux of juvenile hormone precursors in the corpora allata of mosquitoes. *Insect Biochem Mol Biol*. 2011; 41:660–9. [PubMed: 21554954]
33. Lane RP, Bernardes DS. Histology and ultrastructure of pheromone secreting glands in males of the phlebotomine sandfly *Lutzomyia longipalpis*. *Ann Trop Med Parasitol*. 1990; 84:53–61. [PubMed: 2331176]
34. Spiegel CN, Batista-Pereira LG, Bretas JAC, Eiras ÁE, Hooper AM, Pelxoto AA, et al. Pheromone gland development and pheromone production in *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae). *J Med Entomol*. 2011; 48:489–95. [PubMed: 21661306]
35. Gonzalez-Caballero N, Valenzuela JG, Ribeiro JM, Cuervo P, Brazil RP. Transcriptome exploration of the sex pheromone gland of *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae). *Parasit Vectors*. 2013; 6:56. [PubMed: 23497448]
36. Cuervo P, De Jesus JB, Saboia-Vahia L, Mendonça-Lima L, Domont GB, Cupolillo E. Proteomic characterization of the released/secreted proteins of *Leishmania (Viannia) braziliensis* promastigotes. *J Proteomics*. 2009; 73:79–92. [PubMed: 19703603]
37. Carvalho PC, Fischer JSG, Xu T, Cociorva D, Balbuena TS, Valente RH, et al. Search engine processor: filtering and organizing peptide spectrum matches. *Proteomics*. 2012; 12:944–9. [PubMed: 22311825]
38. Carvalho, PC.; Yates, JR., III; Barbosa, VC. Analyzing shotgun proteomic data with PatternLab for proteomics *Current protocols in bioinformatics*. John Wiley & Sons, Inc.; 2010.
39. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, et al. Open mass spectrometry search algorithm. *J Proteome Res*. 2004; 3:958–64. [PubMed: 15473683]
40. Specht M, Kuhlert S, Fufezan C, Hippler M. Proteomics to go: proteomatic enables the user-friendly creation of versatile MS/MS data evaluation workflows. *Bioinformatics*. 2011; 27:1183–4. [PubMed: 21325302]
41. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*. 2007; 23:1282–8. [PubMed: 17379688]
42. Carvalho PC, Fischer JS, Chen EI, Domont GB, Carvalho MG, Degraeve WM, et al. GO Explorer: a gene-ontology tool to aid in the interpretation of shotgun proteomics data. *Proteome Sci*. 2009; 7:6. [PubMed: 19239707]
43. Zdobnov EM, Apweiler R. InterProScan — an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 2001; 17:847–8. [PubMed: 11590104]
44. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res*. 2010; 38:W695–9. [PubMed: 20439314]
45. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK, Jones AR, et al. The minimum information about a proteomics experiment (MIAPE). *Nat Biotech*. 2007; 25:887–93.
46. Jones P, Côté RG, Cho SY, Klie S, Martens L, Quinn AF, et al. PRIDE: new developments and new datasets. *Nucleic Acids Res*. 2008; 36:D878–83. [PubMed: 18033805]

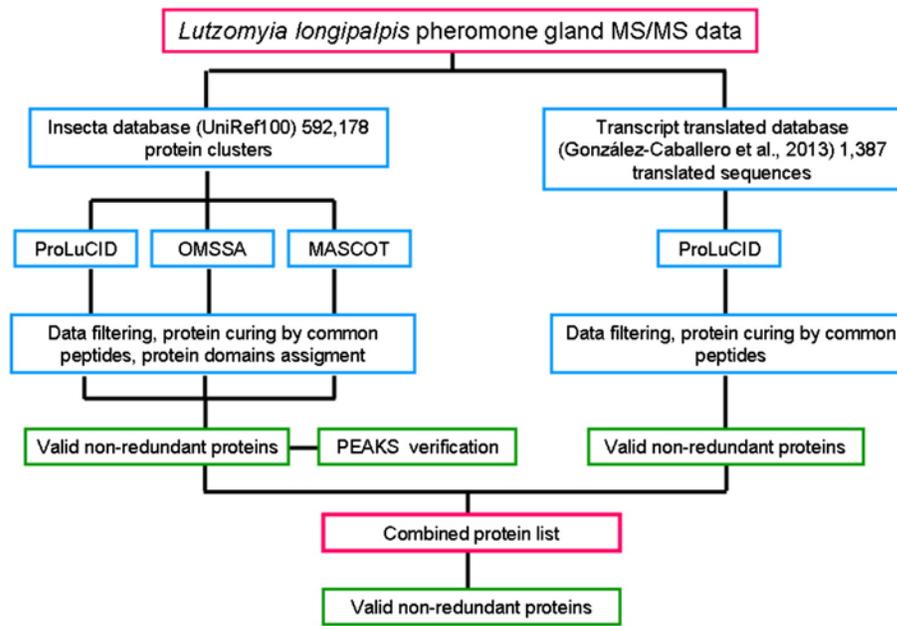
47. Huizinga DH, Omosogbon O, Omery B, Crowell DN. Isoprenylcysteine methylation and demethylation regulate abscisic acid signaling in *Arabidopsis*. *Plant Cell*. 2008; 20:2714–28. [PubMed: 18957507]
48. Freije JM, Blay P, Pendas AM, Cadinanos J, Crespo P, Lopez-Otin C. Identification and chromosomal location of two human genes encoding enzymes potentially involved in proteolytic maturation of farnesylated proteins. *Genomics*. 1999; 58:270–80. [PubMed: 10373325]
49. Fraga BM. Natural sesquiterpenoids. *Nat Prod Rep*. 2008; 25:1180–209. [PubMed: 19030608]
50. Mayoral JG, Nouzova M, Navare A, Noriega FG. NADP<sup>+</sup>-dependent farnesol dehydrogenase, a corpora allata enzyme involved in juvenile hormone synthesis. *Proc Natl Acad Sci U S A*. 2009; 106:21091–6. [PubMed: 19940247]
51. Vandermoten S, Haubruge E, Cusson M. New insights into short-chain prenyltransferases: structural features, evolutionary history and potential for selective inhibition. *Cell Mol Life Sci*. 2009; 66:3685–95. [PubMed: 19633972]
52. Bhandari J, Fitzpatrick AH, Crowell DN. Identification of a novel abscisic acid-regulated farnesol dehydrogenase from *Arabidopsis*. *Plant Physiol*. 2010; 154:1116–27. [PubMed: 20807998]
53. Sutherland TD, Unnithan GC, Andersen JF, Evans PH, Murataliev MB, Szabo LZ, et al. A cytochrome P450 terpenoid hydroxylase linked to the suppression of insect juvenile hormone synthesis. *Proc Natl Acad Sci U S A*. 1998; 95:12884–9. [PubMed: 9789009]
54. López MF, Cano-Ramírez C, Cesar-Ayala AK, Ruiz EA, Zúñiga G. Diversity and expression of P450 genes from *Dendroctonus valens* LeConte (Curculionidae: Scolytinae) in response to different kairomones. *Insect Biochem Mol Biol*. 2013; 43:417–32. [PubMed: 23454142]
55. Keeling CI, Henderson H, Li M, Yuen M, Clark EL, Fraser JD, et al. Transcriptome and full-length cDNA resources for the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major insect pest of pine forests. *Insect Biochem Mol Biol*. 2012; 42:525–36. [PubMed: 22516182]
56. Robertson H, Martos R, Sears C, Todres E, Walden K, Nardi J. Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. *Insect Mol Biol*. 1999; 8:501–18. [PubMed: 10620045]
57. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8 — a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*. 2009; 37:D412–6. [PubMed: 18940858]
58. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 2011; 39:D561–8. [PubMed: 21045058]
59. Wang Z, Dong Y, Desneux N, Niu C. RNAi silencing of the HaHMG-CoA reductase gene inhibits oviposition in the *Helicoverpa armigera* cotton bollworm. *PLoS One*. 2013; 8:e67732. [PubMed: 23844078]
60. Hojo M, Maekawa K, Saitoh S, Shigenobu S, Miura T, Hayashi Y, et al. Exploration and characterization of genes involved in the synthesis of diterpene defence secretion in nasute termite soldiers. *Insect Mol Biol*. 2012; 21:545–57. [PubMed: 22984844]
61. Keeling C, Blomquist G, Tittiger C. Coordinated gene expression for pheromone biosynthesis in the pine engraver beetle, *Ips pini* (Coleoptera: Scolytidae). *Naturwissenschaften*. 2004; 91:324–8. [PubMed: 15257386]
62. Fujii T, Ito K, Katsuma S, Nakano R, Shimada T, Ishikawa Y. Molecular and functional characterization of an acetyl-CoA acetyltransferase from the adzuki bean borer moth *Ostrinia scapulalis* (Lepidoptera: Crambidae). *Insect Biochem Mol Biol*. 2010; 40:74–8. [PubMed: 20043999]
63. Buesa C, Martínez-González J, Casals N, Haro D, Piulachs MD, Bellés X, et al. *Blattella germanica* has two HMG-CoA synthase genes Both are regulated in the ovary during the gonadotrophic cycle. *J Biol Chem*. 1994; 269:11707–13. [PubMed: 7909314]
64. Casals, Nr; Buesa, C.; Piulachs, MD.; Cabanó, J.; Marrero, PF.; Belles, X., et al. Coordinated expression and activity of 3-hydroxy-3-methylglutaryl coenzyme a synthase and reductase in the fat body of *Blattella germanica* (L during vitellogenesis). *Insect Biochem Mol Biol*. 1996; 26:837–43. [PubMed: 9014330]

65. Taban AH, Fu J, Blake J, Awano A, Tittiger C, Blomquist GJ. Site of pheromone biosynthesis and isolation of HMG-CoA reductase cDNA in the cotton boll weevil, *Anthonomus grandis*. Arch Insect Biochem Physiol. 2006; 62:153–63. [PubMed: 16933277]
66. Hall GM, Tittiger C, Andrews GL, Mastick GS, Kuenzli M, Luo X, et al. Midgut tissue of male pine engraver, *Ips pini*, synthesizes monoterpenoid pheromone component ipsdienol de novo. Naturwissenschaften. 2002; 89:79–83. [PubMed: 12046626]
67. Hall GM, Tittiger C, Blomquist GJ, Andrews GL, Mastick GS, Barkawi LS, et al. Male Jeffrey pine beetle, *Dendroctonus jeffreyi*, synthesizes the pheromone component frontalin in anterior midgut tissue. Insect Biochem Mol Biol. 2002; 32:1525–32. [PubMed: 12530220]
68. Nardi J, Young AG, Ujhelyi E, Tittiger C, Lehane M, Blomquist G. Specialization of midgut cells for synthesis of male isoprenoid pheromone components in two scolytid beetles, *Dendroctonus jeffreyi* and *Ips pini*. Tissue Cell. 2002; 34:221–31. [PubMed: 12176306]
69. Ivarsson P, Tittiger C, Blomquist C, Borgeson CE, Seybold SJ, Blomquist GJ, et al. Pheromone precursor synthesis is localized in the metathorax of *Ips paraconfusus* Lanier (Coleoptera: Scolytidae). Naturwissenschaften. 1998; 85:507–11.
70. Silberkang M, Havel CM, Friend DS, McCarthy BJ, Watson JA. Isoprene synthesis in isolated embryonic *Drosophila* cells. I. Sterol-deficient eukaryotic cells. J Biol Chem. 1983; 258:8503–11. [PubMed: 6863298]
71. Blomquist GJ, Figueroa-Teran R, Aw M, Song M, Gorzalski A, Abbott NL, et al. Pheromone production in bark beetles. Insect Biochem Mol Biol. 2010; 40:699–712. [PubMed: 20727970]
72. Kinjoh T, Kaneko Y, Itoyama K, Mita K, Hiruma K, Shinoda T. Control of juvenile hormone biosynthesis in *Bombyx mori*: cloning of the enzymes in the mevalonate pathway and assessment of their developmental expression in the corpora allata. Insect Biochem Mol Biol. 2007; 37:808–18. [PubMed: 17628279]
73. Diaz ME, Mayoral JG, Priestap H, Nouzova M, Rivera-Perez C, Noriega FG. Characterization of an isopentenyl diphosphate isomerase involved in the juvenile hormone pathway in *Aedes aegypti*. Insect Biochem Mol Biol. 2012; 42:751–7. [PubMed: 22782071]
74. Bearfield J, Henry A, Tittiger C, Blomquist G, Ginzl M. Two regulatory mechanisms of monoterpenoid pheromone production in *Ips* spp of bark beetles. J Chem Ecol. 2009; 35:689–97. [PubMed: 19554371]
75. Ritter FJ, Brüggemann-Rotgans IEM, Verwiel PEJ, Persoons CJ, Talman E. Trail pheromone of the Pharaoh's ant, *Monomorium pharaonis*: isolation and identification of faranal, a terpenoid related to juvenile hormone II. Tetrahedron Lett. 1977; 18:2617–8.
76. Meyer AS, Schneiderman HA, Hanzmann E, Ko JH. The two juvenile hormones from the *Cecropia* silk moth. Proc Natl Acad Sci U S A. 1968; 60:853–60. [PubMed: 16591663]
77. Meyer AS, Hanzmann E, Schneiderman HA, Gilbert LI, Boyette M. The isolation and identification of the two juvenile hormones from the *Cecropia* silk moth. Arch Biochem Biophys. 1970; 137:190–213. [PubMed: 5435056]
78. Tashiro T, Bando M, Mori K. Pheromone synthesis, CCVIII: synthesis of (1S, 3S, 7R)-3-methyl- $\alpha$ -himachalene, the sex pheromone of the sandfly *Lutzomyia longipalpis* from Jacobina, Brazil. Synthesis. 2000; 13:1852–62.
79. Schooley DA, Judy KJ, Bergot BJ, Hall MS, Siddall JB. Biosynthesis of the juvenile hormones of *Manduca sexta*: labeling pattern from mevalonate, propionate, and acetate. Proc Natl Acad Sci U S A. 1973; 70:2921–5. [PubMed: 16592112]
80. Francke W, Schulz S. Pheromones. Comprehensive natural products chemistry. 1999; 8:197–261.
81. Vigoder FM, Araki AS, Bauzer LGSR, Souza NA, Brazil RP, Peixoto AA. Lovesongs and period gene polymorphisms indicate *Lutzomyia cruzi* (Mangabeira, 1938) as a sibling species of the *Lutzomyia longipalpis* (Lutz and Neiva, 1912) complex. Infect Genet Evol. 2010; 10:734–9. [PubMed: 20478408]
82. Spiegel CN, Brazil RP, Soares MJ. Ultrastructure of male sex pheromone glands in abdominal tergites of five *Lutzomyia* sandfly species (Diptera: Psychodidae). Arthropod Struct Dev. 2002; 30:219–27. [PubMed: 18088957]

83. Spiegel CN, Brazil RP, Soares MJ. Ultrastructural cytochemistry of the sex pheromone glands of *Lutzomyia cruzi* male sand flies (Diptera: Psychodidae: Phlebotominae). *Arthropod Struct Dev.* 2004; 33:399–404. [PubMed: 18089046]
84. Boufana B, Ward R, Phillips A. Development of the tergal pheromone gland in male *Lutzomyia longipalpis* (Diptera: Psychodidae). *Trans R Soc Trop Med Hyg.* 1986; 80:333–44.
85. Millerd A, Bonner J. Acetate activation and acetoacetate formation in plant systems. *Arch Biochem Biophys.* 1954; 49:343–55. [PubMed: 13159282]
86. Starai VJ, Celic I, Cole RN, Boeke JD, Escalante-Semerena JC. Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine. *Science.* 2002; 298:2390–2. [PubMed: 12493915]
87. Matsumoto S, Fónagy A, Yamamoto M, Wang F, Yokoyama N, Esumi Y, et al. Chemical characterization of cytoplasmic lipid droplets in the pheromone-producing cells of the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol.* 2002; 32:1447–55. [PubMed: 12530212]
88. Fónagy A, Yokoyama N, Matsumoto S. Physiological status and change of cytoplasmic lipid droplets in the pheromone-producing cells of the silkworm, *Bombyx mori* (Lepidoptera, Bombycidae). *Arthropod Struct Dev.* 2001; 30:113–23. [PubMed: 18088949]
89. Cermelli S, Guo Y, Gross SP, Welte MA. The lipid-droplet proteome reveals that droplets are a protein-storage depot. *Curr Biol.* 2006; 16:1783–95. [PubMed: 16979555]
90. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics.* 2012; 11 M111.010587.

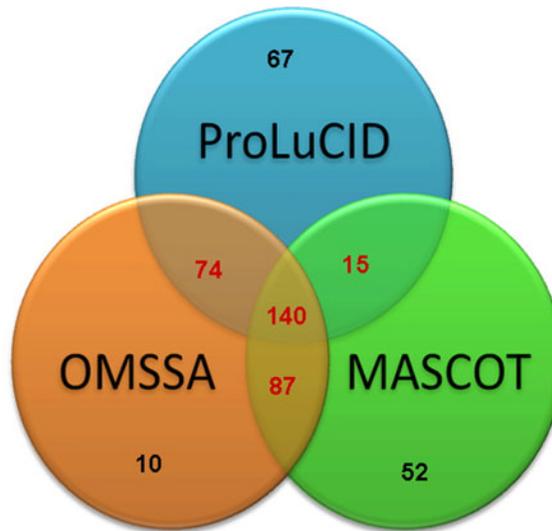


**Fig. 1.** Dissection strategy used to isolate the 4th abdominal segment of *L. longipalpis* males which contains the pheromone gland. First, the integral midgut was carefully removed and discarded (1). Next, the 4th abdominal segment was cut out from the remaining abdominal segments (2) with the help of a tip of a needle of 20 mm (2a). After thoroughly washing in 0.9% saline solution, a pool of approximately 100 segments (2b) was used for protein extraction and SDS-PAGE analysis (3).



**Fig. 2.**

Proteomics data analysis workflow. Mass spectra data from proteins of *L. longipalpis* pheromone gland was searched against the Insecta database (UniRef100; 592,178 protein clusters), downloaded from UniProt, using three search engines: Mascot, OMSSA and ProLuCID. Proteins identified with these search engines were verified by de novo sequencing and peptide spectrum matching using PEAKS (left). In addition, MS/MS data were analyzed against the pheromone gland transcript translated database previously obtained and reported (1387 translated sequences; [35]) using the ProLuCID search engine (right). Results were manually filtered and validated separately as described in the Methods section. The data obtained with the two databases were compared and combined in a list containing all non-redundant validated proteins.

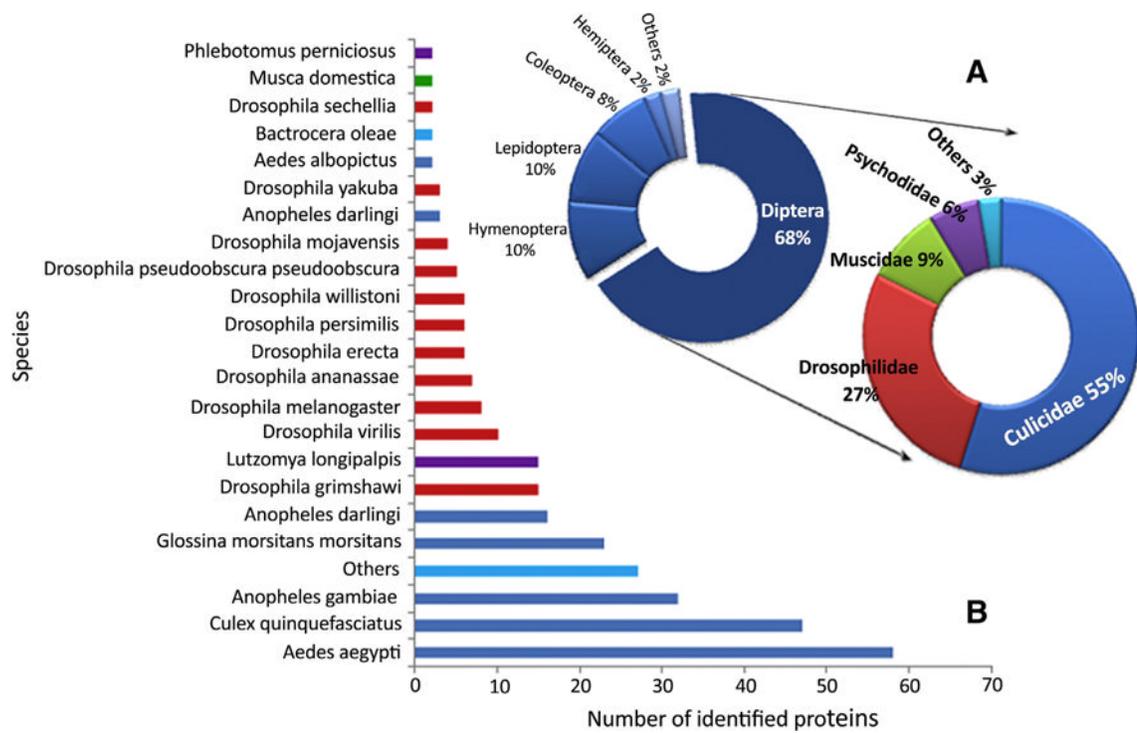


**Total identified proteins: 445**

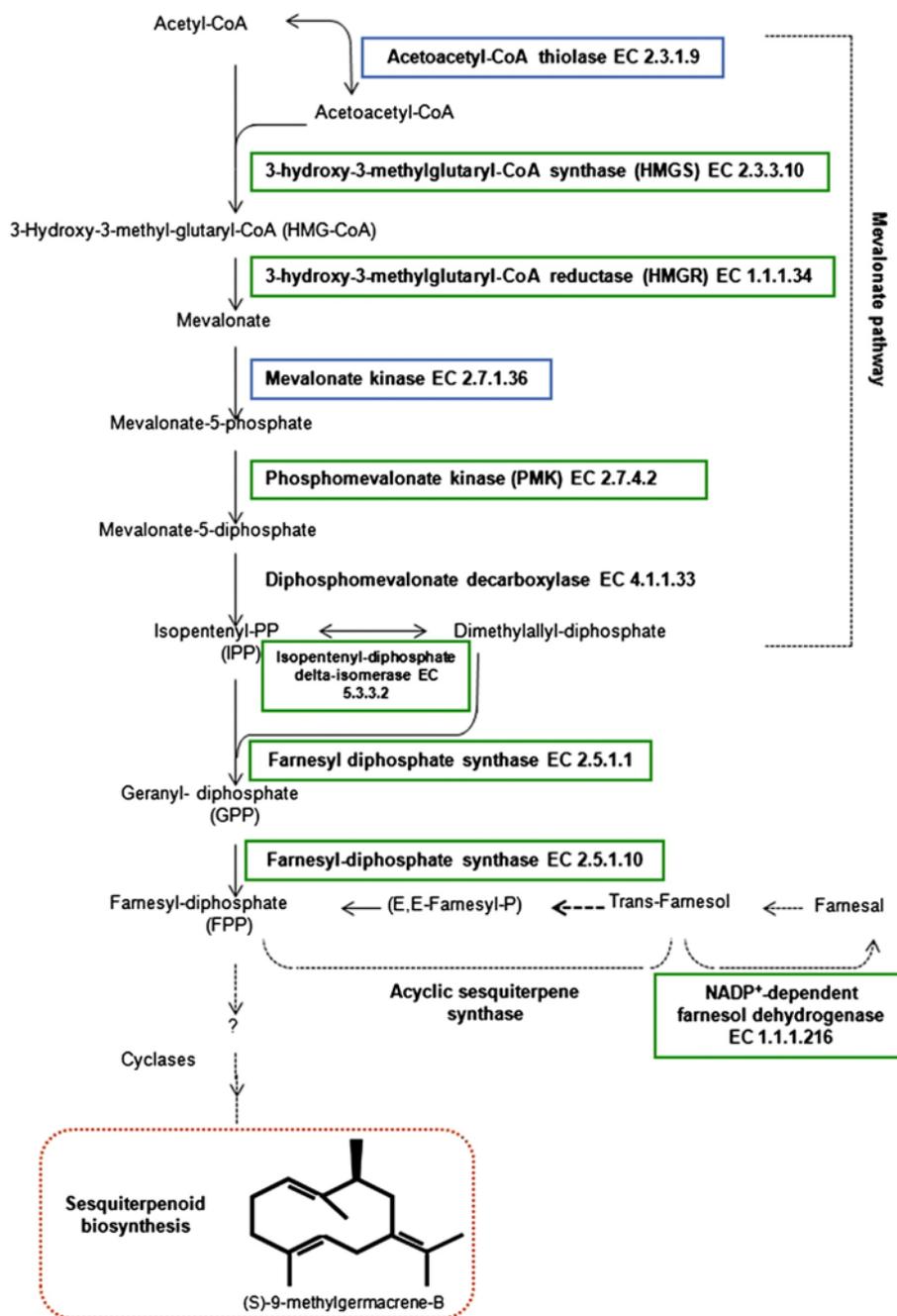
**Proteins commonly identified by at least two search engines: 316**

**Fig. 3.**

Venn diagram showing the proteins commonly identified against the Insecta (Uniref100) database using the three search engines: Mascot, OMSSA and ProLuCID. A total of 445 unique proteins were identified, 316 of which were commonly identified by at least two search engines.



**Fig. 4.** Insecta orders, families and species that contributed to the identification, by similarity, of the *L. longipalpis* pheromone gland proteins. (A) The Diptera order contributed 68% identification; among Diptera, the Culicidae family contributes most of the identifications (55%) of the *L. longipalpis* pheromone gland proteins. (B) Species of the Culicidae (blue), Drosophilidae (red), Muscidae (green), Psychodidae (violet) and others (light blue) families that contributed for the identification of proteins from the *L. longipalpis* pheromone gland.



**Fig. 5.** Diagram showing the mevalonate pathway and prenyltransferase enzymes possibly related to the *L. longipalpis* pheromone biosynthesis. The names enclosed in boxes indicate the proteins identified in this study. Green boxes: enzymes confidently identified by at least two peptides. Blue boxes: Protein with putative thiolase-2/SCP2 sterol-binding domains identified by similarity with 4 peptides in the ant *Solenopsis invicta* and the bee *Apis mellifera* (acetoacetyl-CoA thiolase) or identified by one peptide with confident score (mevalonate kinase). For homoterpene biosynthesis, the propionyl-CoA can substitute acetyl-CoA leading to the formation of homomevalonate and consequently homofarnesyl [79].

Table 1

Proteins potentially related to the *Lutzomyia longipalpis* pheromone biosynthesis or metabolism.

Proteins potentially related to the <i>Lutzomyia longipalpis</i> pheromone biosynthesis or metabolism				Insecta protein database (Uniprot)				Search engines				Search on protein domain families (Pfam, SUPERFAMILY, PROSITE) databases				Transcript translated database			
Identified	Code	Protein names (Uniprot)	General annotation (Uniprot)	Species	OMSSA (PopA)	Mascot (PopB)	Protein ID (PopC)	Description	Identified	Code	Description	Identified	Code	Description	Protein ID (PopD)	Search engine			
<i>Enzymes from the mevalonate pathway</i>																			
+	E02D9	Putative uncharacterized protein		<i>Scaptomyza imicola</i>	1	4	3	Thiolase-2, PROSITE-SCP2 acetyl-binding domain/Thiolase, N-terminal (Pfam) — Thiolase-like (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		
+	H9K0B	Uncharacterized protein		<i>Aedes aegypti</i>	-	-	4	SCP2 acetyl-binding domain/Thiolase, N-terminal (Pfam) — Thiolase-like (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		
-	Q17HE1	AALI002687-PA		<i>Aedes aegypti</i>	-	-	-	SCP2 acetyl-binding domain/Thiolase-like (SUPERFAMILY) — Thiolase-2, PROSITE — Thiolase, N-terminal/Thiolase, C-terminal (Pfam)	+	LL-207		+	LL-207			6			
+	P54961	Hydroxymethylglutaryl-CoA synthase 1	Belongs to the HMG-CoA synthase family.	<i>Blattella germanica</i>	-	-	3	Hydroxymethylglutaryl-CoA synthase, N-terminal/Thiolase-2, PROSITE — Thiolase, N-terminal/Thiolase-like (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		
-	B4G517	GL23258	Belongs to the HMG-CoA reductase family.		-	-	-		+	LL-400	3-Hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase/Keg database	+	LL-400	3-Hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase/Keg database		2			
-	B0WAH3	3-hydroxy-3-methylglutaryl-CoA reductase	HMG-CoA reductase family.		-	-	-		+	LL-215	3-Hydroxy-3-methylglutaryl-CoA reductase/NR-LIGHT database	+	LL-215	3-Hydroxy-3-methylglutaryl-CoA reductase/NR-LIGHT database		17			
+	A5A7A0	Mevalonate kinase	Belongs to the GMP kinase family	<i>Bombix mori</i>	-	-	1	SCP2 acetyl-binding domain/GMP kinase, N-terminal domain/GMP kinase, C-terminal domain (Pfam)/Ribosomal protein S5 domain 2-like/GMP Kinase, C-terminal domain (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		
-	Q1G2B	AALI013044-PA			-	-	-		+	LL-417	Phosphomevalonate kinase/NR-LIGHT database	+	LL-417	Phosphomevalonate kinase/NR-LIGHT database		3			
-	Q1HQH4	Isopentenyl pyrophosphate dimethylallyl pyrophosphate isomerase			-	-	-		+	LL-53	Isopentenyl pyrophosphate dimethylallyl pyrophosphate isomerase/Keg database	+	LL-53	Isopentenyl pyrophosphate dimethylallyl pyrophosphate isomerase/Keg database		11			
-	Q1HQH4	Isopentenyl pyrophosphate dimethylallyl pyrophosphate isomerase			-	-	-		+	LL-118	Isopentenyl pyrophosphate dimethylallyl pyrophosphate isomerase/Keg database	+	LL-118	Isopentenyl pyrophosphate dimethylallyl pyrophosphate isomerase/Keg database		8			
<i>Enzymes involved in the sesquiterpene biosynthesis</i>																			
+	Q5KGE9	Farnesyl diphosphate synthase	Belongs to the FPP/GGPP synthase family.	<i>Ilysi puti</i>	1	2	-	Poly(prenyl) synthase (Pfam) — Terpenoid synthases (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		
-	B3NSB9	G022667	Belongs to the FPP/GGPP synthase family		-	-	-		+	LL-257	Farnesyl diphosphate synthase/Keg database	+	LL-257	Farnesyl diphosphate synthase/Keg database		13			
-	B3MP90	G112409	Belongs to the FPP/GGPP synthase family.		-	-	-		+	LL-94	Poly(prenyl) synthase/Keg database	+	LL-94	Poly(prenyl) synthase/Keg database		10			
+	G0YZR4	Mitochondrial isoprenyl diphosphate synthase	Belongs to the FPP/GGPP synthase family	<i>Rhynchosiphum pallidum</i>	-	2	-	Poly(prenyl) synthase (Pfam) — Terpenoid synthases (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		
-	D2WKED9	NADP-dependent farnesyl dehydrogenase 1	Belongs to the short-chain dehydrogenase/reductases (SDR) family.		-	-	-		+	LL-314	Predicted dehydrogenase/Keg database	+	LL-314	Predicted dehydrogenase/Keg database		8			
+	Q7Q577	AGAP006564-PA	Belongs to the short-chain dehydrogenase/reductases (SDR) family.	<i>Anopheles gambiae</i>	1	2	2	Short-chain dehydrogenase/reductase SDR (Pfam) — SCP2 acetyl-binding domain (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		
+	E3WXS6	Uncharacterized protein	Belongs to the short-chain dehydrogenase/reductases (SDR) family	<i>Anopheles darlingi</i>	1	4	2	Short-chain dehydrogenase/reductase SDR (Pfam)	-	-	-	-	-	-	-	-	-		
+	Q7QR65	AGAP006667-PA	Belongs to the short-chain dehydrogenase/reductases (SDR) family.	<i>A. gambiae</i>	1	-	3	Short-chain dehydrogenase/reductase SDR (Pfam) — SCP2 acetyl-binding domain (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		
+	E3X7U8	Uncharacterized protein	Belongs to the short-chain dehydrogenase/reductases (SDR) family.	<i>A. darlingi</i>	3	7	3	Short-chain dehydrogenase/reductase SDR (Pfam) — SCP2 acetyl-binding domain (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		
+	B4M521	G11044	Belongs to the short-chain dehydrogenase/reductases (SDR) family.	<i>Drosophila virilis</i>	-	3	2	Short-chain dehydrogenase/reductase SDR (Pfam) — SCP2 acetyl-binding domain (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		
+	B4J1Y0	G112555	Belongs to the short-chain dehydrogenase/reductases (SDR) family.	<i>Drosophila grimshawi</i>	1	3	3	Short-chain dehydrogenase/reductase SDR (Pfam)	-	-	-	-	-	-	-	-	-		
+	H9K6D	Uncharacterized protein	Belongs to the short-chain dehydrogenase/reductases (SDR) family.	<i>A. aegypti</i>	1	2	-	Short-chain dehydrogenase/reductase SDR (Pfam)	-	-	-	-	-	-	-	-	-		
+	H9G50	Uncharacterized protein	Belongs to the short-chain dehydrogenase/reductases (SDR) family.	<i>Aedes cephalis</i>	1	3	-	Short-chain dehydrogenase/reductase SDR (Pfam)	-	-	-	-	-	-	-	-	-		
<i>Proteins functionally associated to mevalonate pathway enzymes or pheromone metabolism</i>																			
+	Q1GQ55	AALI011395-PA	Belongs to the class I (aminoacyl)-tRNA synthetase family.	<i>A. aegypti</i>	1	5	-	Gluaminyl-tRNA synthetase, class Ib, non-specific RNA-binding domain (Pfam) — Ribosomal protein L25-like (SUPERFAMILY)	-	-	-	-	-	-	-	-	-		

Protein potentially related to the <i>Zn. longidigitus</i> pheromone biosynthesis or metabolism		Insecta protein database (Uniprot ID)				Search engines			Search on protein domain families (Pfam, SUPERFAMILY, PROSITE) databases			Transcript translated database		Search engine	
Identified	Code	Protein names (Uniprot)	Gene/annotation (Uniprot)	Species	OMSA #/seq#	Mascot #/seq#	ProteinCD #/seq#	Description	Identified	Code	Description	Code	Description	ProteinCD #/seq#	Search engine
+	E5XBN5	Phosphoglycerate kinase	Belongs to the phosphoglycerate kinase family.	<i>A. duranii</i>	2	-	-	Phosphoglycerate kinase (SUPERFAMILY)	-	-	-	-	-	-	-
-	P0AGE9	Succinyl-CoA synthetase alpha subunit	Belongs to the succinyl-CoA synthetase subunit alpha family.	-	-	-	-	Succinyl-CoA synthetase, alpha subunit/Kog database	+	LL-149	Succinyl-CoA synthetase, alpha subunit/Kog database	-	-	17	-
+	Q17CT0	AALD00434-PA	-	<i>A. oryzae</i>	2	-	4	Transketolase, N-terminal/Transketolase-like pyridoxal-binding oxidoreductase, domain II	+	LL-363	Transketolase/Kog database	-	-	7	-
+	A8KWEE	Putative glutathione S-transferase	Belongs to the GST superfamily.	<i>Lacworysta longidigitus</i>	3	7	4	Glutathione S-transferase, N and C-terminal (Pfam) — Glutathione S-transferase, C-terminal like (SUPERFAMILY)	-	-	-	-	-	-	-
+	A8KWBE	Putative glutathione S-transferase	Belongs to the GST superfamily.	<i>L. longidigitus</i>	3	-	3	Glutathione S-transferase, N and C-terminal (Pfam) — Glutathione S-transferase, C-terminal like (SUPERFAMILY)	-	-	-	-	-	-	-
+	B0WUG8	Glutathione S-transferase 1-5	Belongs to the GST superfamily.	<i>Culex quinquefasciatus</i>	-	2	-	Glutathione S-transferase, N and C-terminal (Pfam) — Glutathione S-transferase, C-terminal like (SUPERFAMILY)	-	-	-	-	-	-	-
+	B4DYJ3	GH14004	Contains 1 GST C-terminal domain/Contains 1 GST N-terminal domain.	<i>D. grimshawi</i>	1	3	-	β1-Glutathione S-transferaseS and C-terminal (Pfam) — Glutathione S-transferase, C-terminal like (SUPERFAMILY)	-	-	-	-	-	-	-
-	A8KWEE	Putative glutathione S-transferase	Belongs to the GST superfamily	-	-	-	-	Oxidoreductase (ADP/NADPH)-binding (Pfam) — Riboflavin synthase-like beta-barrel (SUPERFAMILY)	+	LL-158	Glutathione S-transferase/Diptera database	-	-	7	-
+	Q07994	NADPH-cytochrome P450 reductase	In the C-terminal section; belongs to the flavoprotein pyridine nucleotide cytochrome reductase family.	<i>Musca domestica</i>	-	-	2	-	-	-	-	-	-	-	-
+	B4DX14	GH16186	-	<i>D. grimshawi</i>	1	2	1	Cytochrome c1 (Pfam) — Cytochrome c domain (SUPERFAMILY)	-	-	-	-	-	-	-
-	Q174F5	AALD00824-PA	Belongs to the cytochrome P450 family	-	-	-	-	-	+	LL-42	Cytochrome P450 CYP4CY19/CYP26 subfamilies/Kog database	-	-	5	-