Data in brief 24 (2019) 103911

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Proteome data of female *Anopheles stephensi* antennae

Ajeet Kumar Mohanty [a], Gourav Dey [b], Manish Kumar [c], Sreelakshmi K. Sreenivasamurthy [b], Sandeep Garg [d], T.S. Keshava Prasad [e, **], Ashwani Kumar [a, *]

[a] *ICMR – National Institute of Malaria Research, Field Unit, Campal, Panaji, Goa 403001, India*
[b] *Institute of Bioinformatics, Discoverer Building, International Tech Park, Bangalore 560 066, India*
[c] *Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA*
[d] *Department of Microbiology, Goa University, Taleigao Plateau, Goa Pin – 403206, India*
[e] *Center for Systems Biology and Molecular Medicine, Yenepoya Research Center, Yenepoya University, Mangalore 575018, India*

ARTICLE INFO

Article history:
Received 2 February 2019
Received in revised form 26 March 2019
Accepted 2 April 2019
Available online 8 April 2019

ABSTRACT

Antennae of female *Anopheles stephensi* mosquitoes were dissected and lysed with 1% SDS. Proteins were extracted using ultra sonication and analyzed on high resolution mass spectrometer. Proteomic data was analyzed using two search algorithms SEQUEST and Mascot, resulting in the identification of 22,729 peptides corresponding to 3262 proteins. These proteins were characterized using different bioinformatics tools. VectorBase resource was used to assign Gene Ontology (GO) terms. Using Biomart tool ortholog information was fetched from the VectorBase database. Raw mass spectrometric data was deposited in ProteomeXchange Consortium via PRIDE partner repository in the public dataset PXD001128. Proteins involved in insecticide resistance and odorant binding were the most abundant in the antennae. The proteins identified in this study could be targeted for developing novel vector control strategy.

© 2019 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* Corresponding author.
** Corresponding author.
E-mail addresses: keshav@yenepoya.edu.in (T.S.K. Prasad), ashwani07@gmail.com (A. Kumar).

https://doi.org/10.1016/j.dib.2019.103911
2352-3409/© 2019 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications table

Subject area	Biology
More specific subject area	Mosquito proteomics
Type of data	Tables, Graphs and figures
How data was acquired	LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific) Proteome Discoverer 2.4 and MASCOT search engine (Matrix Science, London, UK; version 2.2) Protein database *Anopheles stephensi* Indian strain (www.VectorBase.org/, release date February 25, 2014)
Data format	Analyzed output data
Experimental factors	Antennae obtained from the laboratory reared female *Anopheles stephensi* mosquitoes
Experimental features	In-gel trypsin digestion of proteins followed by LC-MS/MS analysis using LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific).
Data source location	Goa and Bengaluru, India
Data accessibility	Data are available here and via a web application (ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD001128.
Related research article	Title: Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes. Author list: Prasad TS, Mohanty AK, Kumar M, Sreenivasamurthy S, Dey G, Nirujogi RS, Pinto SM, Madugundu AK, Patil AH, Advani J, Manda SS, Gupta MK, Dwivedi SB, Kelkar DS, Hall B, Jiang X, Peery A, Rajagopalan P, Yelamanchi SD, Solanki HS, Raja R, Sathe GJ, Chavan S, Verma R, Patel KM, Jain AP, Syed N, Datta KK, Khan AA, Dammalli M, Jayaram S, Radhakrishnan A, Mitchell CJ, Na CH, Kumar N, Sinnis P, Sharakhov IV, Wang C, Gowda H, Tu Z, Kumar A, Pandey A. Status: Published

Value of the data

- This data set is the largest catalogue of proteins identified from the antennae of female *Anopheles stephensi*.
- Data provides information about antennae proteins involved in different biological and molecular functions. Overall it enables better understanding of olfactory behaviour of this organ.
- This data could be utilized in future for the development of novel targets for control of disease transmission.

1. Data

The data presented here is the raw as well as processed data of the proteomic analysis of antennae from female *Anopheles stephensi* [1]. In total, thirty one LC-MS/MS runs were performed using high-resolution mass spectrometry (Fig. 1). A total of 207,034 MS/MS spectra were generated from all the runs. The resulting mass spectral data was searched against the predicted protein database of *An. stephensi* Indian strain, using the two different search algorithms Mascot and Sequest. Peptide spectrum matches were filtered with 1% false discovery rate (FDR). The processed data set contains 207,034 MS/MS spectra, which led to identification of 22,729 non-redundant peptides and 3262 proteins. Of the total proteins, 242 proteins (7.4%) were identified with single PSM evidence. On the other hand, 92.6% of the proteins identified had two or more PSMs (Fig. 2). A large fraction of the identified proteins had more than 10 PSMs and this increased the confidence level in the peptides identified through mass spectrometer. All the peptides and proteins identified in this study are listed in Supplementary Tables S1 and S2. In addition, 12 odorant binding proteins (OBPs) have also identified Supplementary Table S3. These proteins are known for their involvement in chemosensory function and have been identified in the antennae and legs of *An. gambiae*, *Aedes aegypti* and several other mosquito species [2,3]. The odorant binding proteins are candidates for developing novel vector control strategy for reducing malaria transmission by blocking sensory functions of antennae. Seventy nine proteins involved in insecticide resistance mechanisms have also been identified in the antennae Supplementary Table S4. These proteins are reported in *An. gambiae* associated with insecticide resistance mechanism which

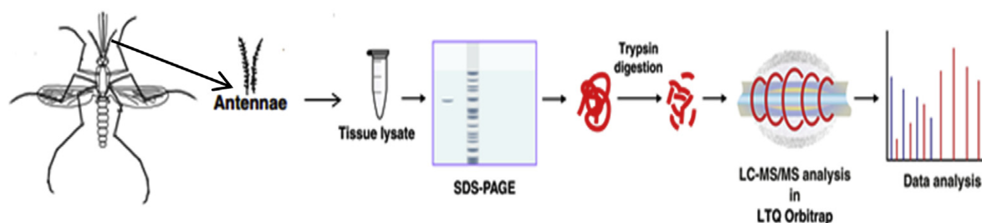

Fig. 1. The workflow illustrating the steps involved in proteomic analysis of antennae of female *Anopheles stephensi.* Proteins were extracted from the antennae and subjected to in-gel trypsin digestion followed by analyses on LTQ-Orbitrap Velos mass spectrometer. Mascot and SEQUEST algorithms were used for database searches.

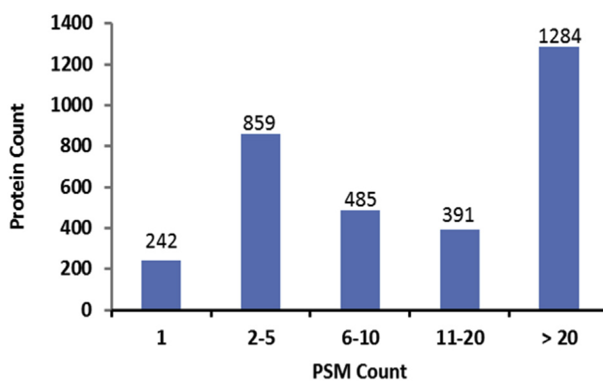

Fig. 2. Distribution of number of proteins identified with the extent of Peptide Spectra Matches (PSMs) obtained from the antennae of *An. stephensi.*

helps insects to metabolize insecticides at a higher rate [4]. The information for the identified proteins was fetched from the VectorBase database. To assign molecular functions, biological process and cellular localization to these proteins, VectorBase resource to assign Gene Ontology (GO) terms was utilized [5–7] Supplementary Table S5. Protein-protein interaction networks were mapped using STRING (version 1.1.0) [8] Supplementary Table S6.

2. Experimental design

2.1. Sample preparation

Antennae were dissected from the 600 female *An. stephensi.* The antennae collected were homogenized in 200 μl of 1% SDS using ultrasonication. The extracted proteins were then quantified according to Bicinchoninic acid assay (Pierce®.Cat#: 23225).

2.2. In-gel digestion

Two hundred micrograms of proteins from antennae were resolved on 10% SDS-PAGE gel. The gel was stained using Colloidal Coomassie 33 stain (Invitrogen, Carlsbad, CA).

With 10% methanol solution, excess stain was removed by giving multiple washings. The entire protein lanes were cut into 31 gel bands and then subjected to in-gel trypsin digestion as described earlier [5–7]. In brief, gel bands were destained with 40mM Ammonium Bicarbonate (ABC), pH- 8, 40%

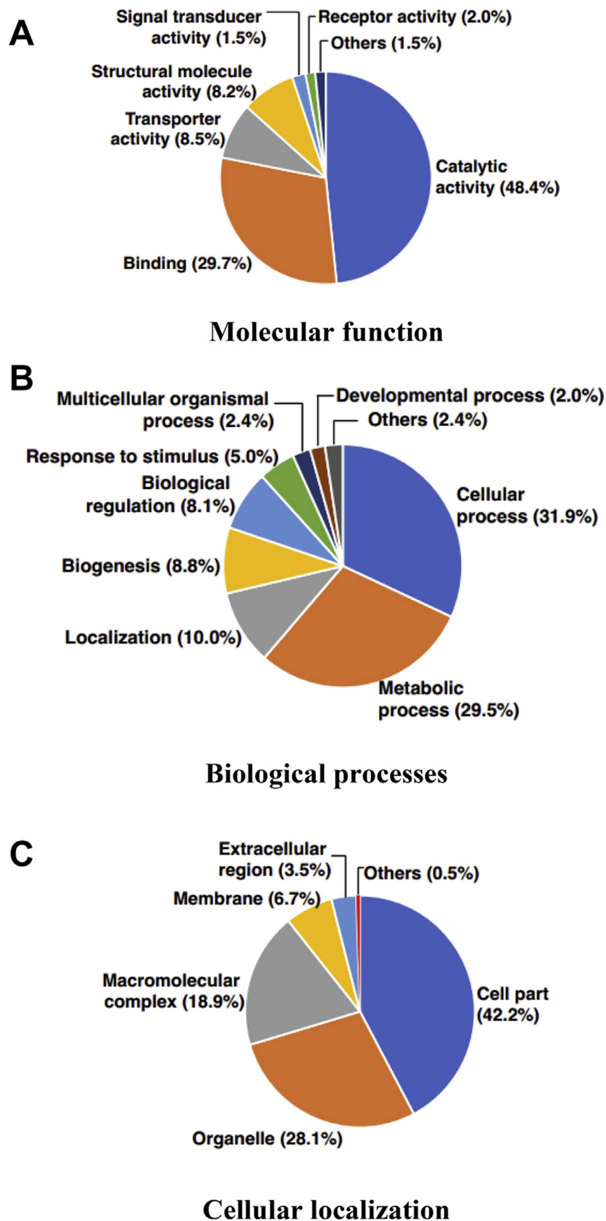

Fig. 3. Classification of proteins based on Molecular functions (A) Biological processes (B) and Subcellular localization (C).

Acetonitrile (ACN). In-gel proteins were reduced using 10mM Dithiothreitol (DTT) for 60 °C for 10 min and alkylated by adding 10mM Iodoacetamide (IAA) in 40mM ABC for 10 min in dark. In-gel digestion was carried out using sequencing grade trypsin (Promega) with enzyme: substrate ration of 1:20 and incubate at 37 °C for 12–16 hrs. Post trypsin digestion, peptides were extracted from the gel, dried in speedvac and cleaned using C_{18} Sep-Pak column (Waters) prior to analysis on mass spectrometer.

2.3. Mass spectrometry analysis

In this study, a total of 31 LC-MS/MS analyses were performed on LTQ-Orbitrap Velos mass spectrometer interfaced with Proxeon Easy nLC system (Thermo Scientific, Bremen, Germany) as described previously [1,5–7]. The peptides from each fraction were reconstituted in 0.1% formic-acid and loaded on to a pre-column (2cm, 5u –100 Å), followed by separation on an analytical column (11cm, 3u –100 Å) packed with magic C_{18} AQ (Michrom Bioresources, USA). The solvent used contained 0.1% aqueous formic acid as solvent A and 95% acetonitrile, 0.1% formic acid as solvent B. The peptides were loaded on the trap column using solvent A followed by resolving at 250 nl/min flow rate using a linear gradient of 10–35% solvent B (0.1% formic acid in 95% Acetonitrile) over 75 minutes on an analytical column. Mass spectrometry analysis was carried out in a data dependent manner with a full scans in the range of m/z 350 to 2000. Both MS and MS/MS were acquired and measured using Orbitrap mass analyzer. Full MS scans were measured at a resolution of 30,000 at m/z 400 and fifteen most intense precursors were selected for MS/MS and were fragmented using higher-energy collisional dissociation (HCD) method and detected at a mass resolution of 15,000 at m/z 400. Automatic gain control for full MS was set to 1 million ions and for MS/MS was set to 0.1 million ions with a maximum ion injection time of 100 and 200 ms respectively. Internal calibration was carried out using lock mass option (m/z 445.1200025) from ambient air.

2.4. Data analysis

The data obtained was processed using Proteome Discoverer software (version 2.4, Thermo Fisher Scientific, Bremen, Germany) workflow and searched using Sequest and Mascot search algorithm against VectorBase protein database Astel2.2 and consists of 11,789 protein sequences as previously described [1]. The search parameters included trypsin as the proteolytic enzyme with single missed cleavage, oxidation of methionine was set as a variable modification while carbamidomethylation of cysteine was set as static modification. Precursor and fragment mass tolerance were set to 10 ppm and 0.05 Da respectively. For the entire data set, false discovery rate (FDR) was calculated by enabling the peptide sequence analysis using a decoy database and a cut-off of 1% was used for identifications.

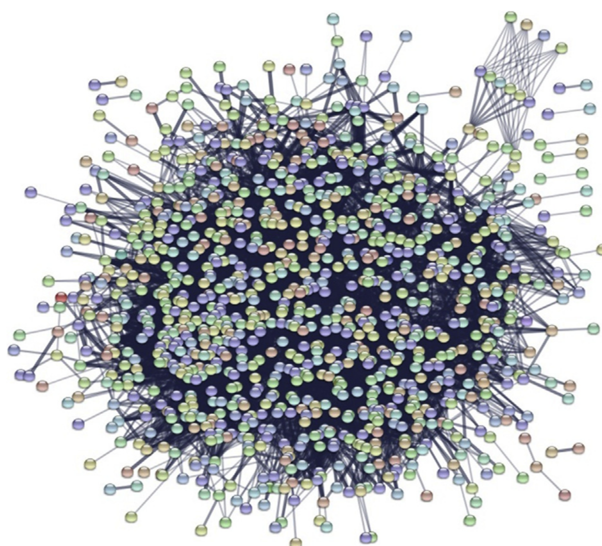

Fig. 4. Representation of predicted protein-protein interaction map of proteins identified in female *An. stephensi* antennae. The interaction map was generated using online STRING tool with default parameters. Proteins identified with multiple PSMs and peptides were used for generating the map.

The information for all the identified proteins was fetched from the VectorBase database [9]. The identified proteins were functionally categorized based on their subcellular localization; biological processes and molecular function using Gene Ontology (GO) based annotation available for *An. stephensi* in the VectorBase database. Proteins identified were found to be involved in different molecular functions such as catalytic activity (48.4%), binding activity (30%), transporter activity (8.5%), structural (8.2%), receptors (2%) and others (1.5%). Biological process-based categorization showed that a majority of proteins played a role in metabolism (30%), cellular processes (32%), localization (10%), biogenesis (9%), biological regulation (8%), response to stimulus (5%), development (2%), multicellular organismal process (2.4%) and others (2.4%). The proteins have been described based on their cellular localization as shown in (Fig. 3A–C, Supplementary Table S5). The antennae proteins identified were analyzed using online STRING tool to generate an interacting map for all the proteins [8] (Fig. 4, Supplementary Table S6).

Acknowledgements

We thank ICMR – National Institute of Malaria Research; Institute of Bioinformatics; and Yenepoya (Deemed to be University) for institutional support. We thank Science and Engineering Research Board (SERB), Government of India, for funding (EMR/2014/000444) to NIMR and IOB to characterize proteome and transcriptome of mosquito vectors. This manuscript has been approved by the NIMR publication committee 021/2018.

Transparency document

Transparency document associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2019.103911.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.103911.

References

[1] T.S. Prasad, et al., Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes, Genome Res. 27 (1) (2017) 133–143.
[2] H. Biessmann, et al., The *Anopheles gambiae* odorant binding protein 1 (AgamOBP1) mediates indole recognition in the antennae of female mosquitoes, PLoS One 5 (3) (2010) e9471.
[3] M. Manoharan, et al., Comparative genomics of odorant binding proteins in *Anopheles gambiae, Aedes aegypti*, and *Culex quinquefasciatus*, Genome Biol. Evol. 5 (1) (2013) 163–180.
[4] V.A. Ingham, et al., Transcriptomic meta-signature identified in *Anopheles gambiae* populations reveal previously undetected insecticide resistance mechanism, Nat. Commun. 11 (9) (2018) 5282.
[5] G. Dey, et al., Proteome data of *Anopheles stephensi* hemolymph using high resolution mass spectrometry, Data Brief 18 (2018) 1441–1447.
[6] R. Dhawan, et al., Data from salivary gland proteome analysis of female *Aedes aegypti* Linn, Data Brief 13 (2017) 274–277.
[7] M. Kumar, et al., Dataset on fat body proteome of *Anopheles stephensi* Liston, Data Brief 22 (2019) 1068–1073.
[8] D. Szklarczyk, et al., The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, Nucleic Acids Res. 45 (D1) (2017) D362–D368.
[9] G.I. Giraldo-Calderón, et al., VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases, Nucleic Acids Res. 43 (2015) D707–D713.