Data in Brief 32 (2020) 106243

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Proteomics dataset of adult *Anopheles Stephensi* female brain

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

[a] *Institute of Bioinformatics, International Tech Park, Bangalore 560 066, India*
[b] *Manipal Academy of Higher Education (MAHE), Manipal 576104, India*
[c] *ICMR-National Institute of Malaria Research, Field Station, Campal, Panaji, Goa 403001, India*
[d] *ICMR-Vector Control Research Centre, Indira Nagar, Puducherry-605 006, UT, India*
[e] *Center for Systems Biology and Molecular Medicine, Yenepoya Research Center, Yenepoya (Deemed to be University), Deralakatte, Mangalore-575018, India*

ARTICLE INFO

Article history:
Received 10 July 2020
Revised 11 August 2020
Accepted 25 August 2020
Available online 3 September 2020

Keywords:
Vector Borne-Disease
Network Analysis
Malaria
Host-Seeking Behavior
Vector-Pathogen Interaction
Infectious disease

ABSTRACT

Mosquitoes with their ability to transmit several pathogens of human disease pose a serious threat to healthcare worldwide. Although much has been done to prevent the disease transmission by mosqitos. The rising rate of resistance in mosquitos towards conventionally used control strategies necessitates developing of novel strategies to counter disease transmission. The mosquito brain plays a key role in host-seeking, finding mates and selection of oviposition sites. However, not much is know about the underlying physiological processes in mosquito brain. The data presented in this study describes the proteins that have been identified in the brain tissue of adult female *Anopheles stephensi* and their associated processes. Interpretation of the data can be related to the previously published article "Integrating transcriptomics and proteomics data for accurate assembly and annotation of genomes" [1].

* Corresponding author(s).
E-mail addresses: directorvcrc@icmr.gov.in (A. Kumar), keshav@yenepoya.edu.in (T. S.K. Prasad).

https://doi.org/10.1016/j.dib.2020.106243

Specifications Table

Subject	Biotechnology
Specific subject area	Proteomics and Vector Biology
Type of data	Table Chart Graph Figure
How data were acquired	LTQ-Orbitrap Velos and LTQ-Orbitrap-Elite mass spectrometers (Thermo Scientific, Bremen, Germany) Proteome Discoverer 2.1, with SEQUEST and MASCOT search engine (Matrix Science, London, UK; version 2.2) *Anopheles stephensi* (Indian strain) protein database (www.VectorBase.org, release February 25, 2014)
Data format	Raw, Analysed
Parameters for data collection	*An. stephensi* mosquitoes were maintained under laboratory conditions at 27 ± 2°C, 75% relative humidity and photoperiod of 12h of day - night cycle. The adult female mosquitoes were fed on 10% glucose solution soaked in cotton pad. Brain tissues were dissected under dissecting microscope using 0.65% normal saline and further stored at -80°C.
Description of data collection	Brains were homogenized in using sonication and extracted proteins were fractionated at peptide and protein level. Digested peptides were analysed using tandem mass spectrometry. Data obtained from tandem mass spectrometry was searched against the *An. stephensi* reference protein database of identification of proteins. Identified proteins were then analysed for their functions and interacting partners using bioinformatic tools.
Data source location	Panaji, Goa, India Bangalore, Karnataka, India
Data accessibility	Repository name: PRIDE (PRotein IDEntification Database) Data identification number: PXD001128 Direct URL to data: https://www.ebi.ac.uk/pride/archive?keyword=PXD001128 Analyzed data is provided along with this article as excel sheets.
Related research article	Prasad et al. Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes. *Genome Res.* 2017;27(1):133–144. doi:10.1101/gr.201368.115

Value of the Data

- The data provides information on proteins expressed and provides insights into the underlying processes in the adult female *An. stephensi* brain.
- The data provides a platform for researchers working towards identification of key proteins associated with mosquito vector-pathogen interactions, host-seeking behavior and transmission blocking vaccines.
- The data provides a comprehensive list of genes that can be studied further in a variety of experiments to understand mosquito physiology, immunity, complexities of host attraction processes (host seeking behavior), avoidance of insecticide sprayed surfaces, resting, feeding and reproductive behavior and also oviposition site preferences of female *An. stephensi* mosquitoes in nature.

1. Data Description

The dataset provided in this article describes the proteins identified in female *Anopheles stephensi* brain and their associated processes and function Fig. 1. Tandem mass spectrometer-based acquisition of peptides derived from adult female *An. stephensi* brain samples led to the

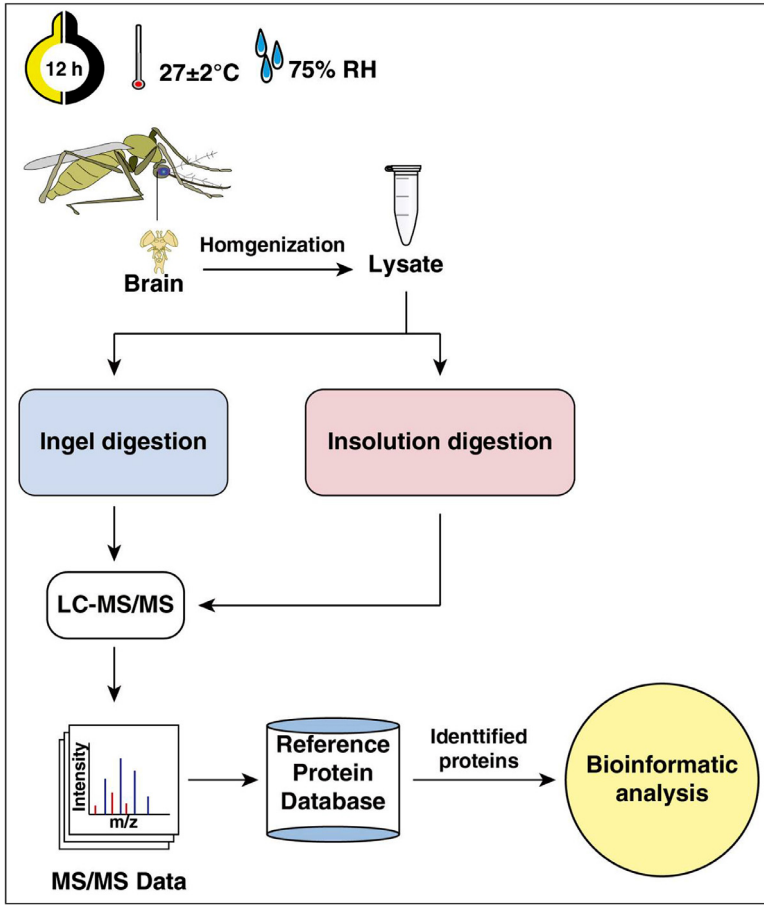

Fig. 1. Representation of the overall study with illustration of mosquito insectary conditions and experimental workflow.

identification of 4,662 proteins Supplementary Table S1, S2. The data provided in this article can be related to our previously published article "Integrating transcriptomics and proteomics data for accurate assembly and annotation of genomes" [1]. The identified proteins were further categorized based on their Gene Ontology term associations. Details of the Gene Ontology categorization for the identified proteins are provided in Fig. 2 and Supplementary Table S3. Fig. 3 illustrates the predicted interacting protein clusters along with their associated processes for the proteins identified in female *Anopheles stephensi* brain. The complete details of predicted protein-protein interaction have been provided in Supplementary Table S4. List of genes expressed in the female *An. stephensi* brain and with potential role in sporogony [2] have been provided in Supplementary Table S5. Processes and pathways for these 28 genes have been illustrated in Fig. 4 and the complete details are provided in Supplementary Table S6 and S7. The complete list of proteins identified in female *Anopheles stephensi* brain and mapping to Immunodb data is provided in Supplementary Table S8. Further analysis of these 54 genes for their associated pathways have been illustrated Fig. 5 and the complete details of the analysis result have been provided in Supplementary Table S9 and S10.

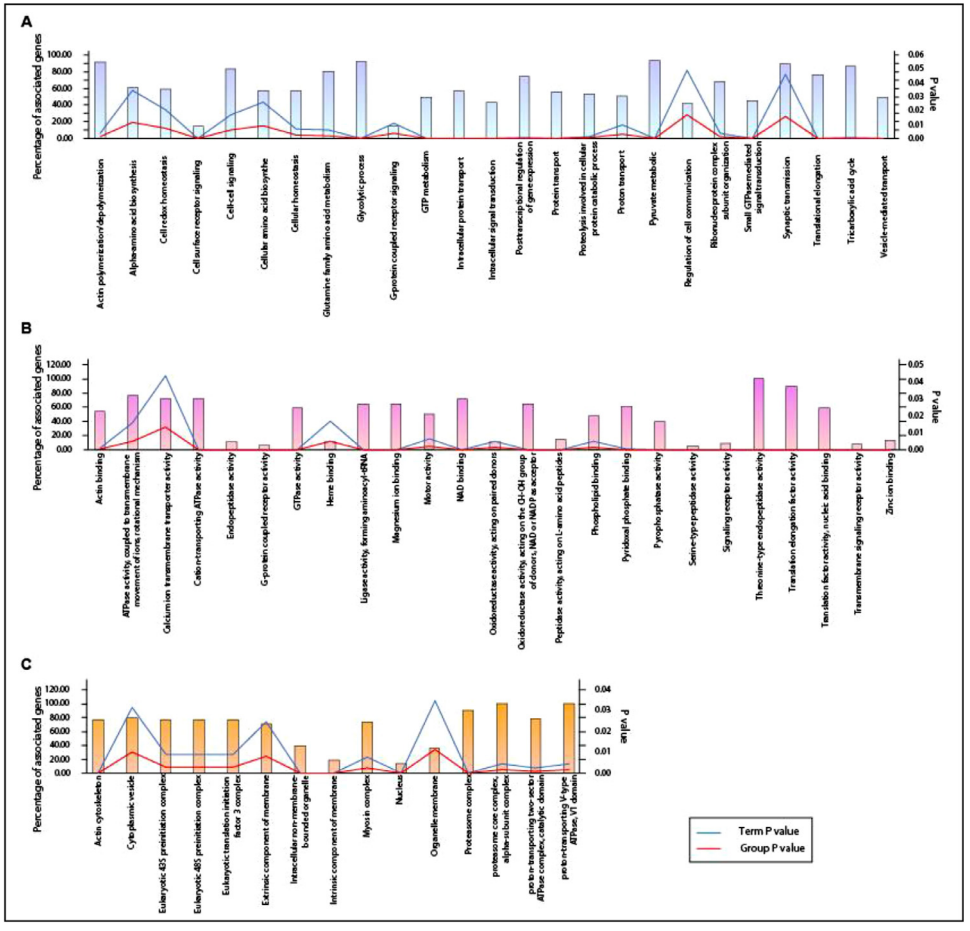

Fig. 2. Gene Ontology analysis of the identified protein in *An. stephensi* brain. A) Biological processes; B) Molecular function; C) Cellular components of the identified proteins. Only p value significant terms and groups were considered. Blue line represents term p value while red line represents group p value.

2. Experimental Design, Materials, and Methods

2.1. Maintenance of mosquito colony

An. stephensi mosquitoes required for the study were obtained from the insectary of ICMR- National Institute of Malaria Research, Field Station Goa. Mosquitoes were maintained under laboratory conditions and dissection of mosquito tissues were carried out as mentioned in our previous study [1].

2.2. Protein extraction

As previously discussed, proteins were extracted from adult female brains by homogenization using a probe sonicator in 200 µl of 4% SDS buffer. The extracted protein amount was quantified using Lowry's method (Bio-Rad DC Protein assay) [3].

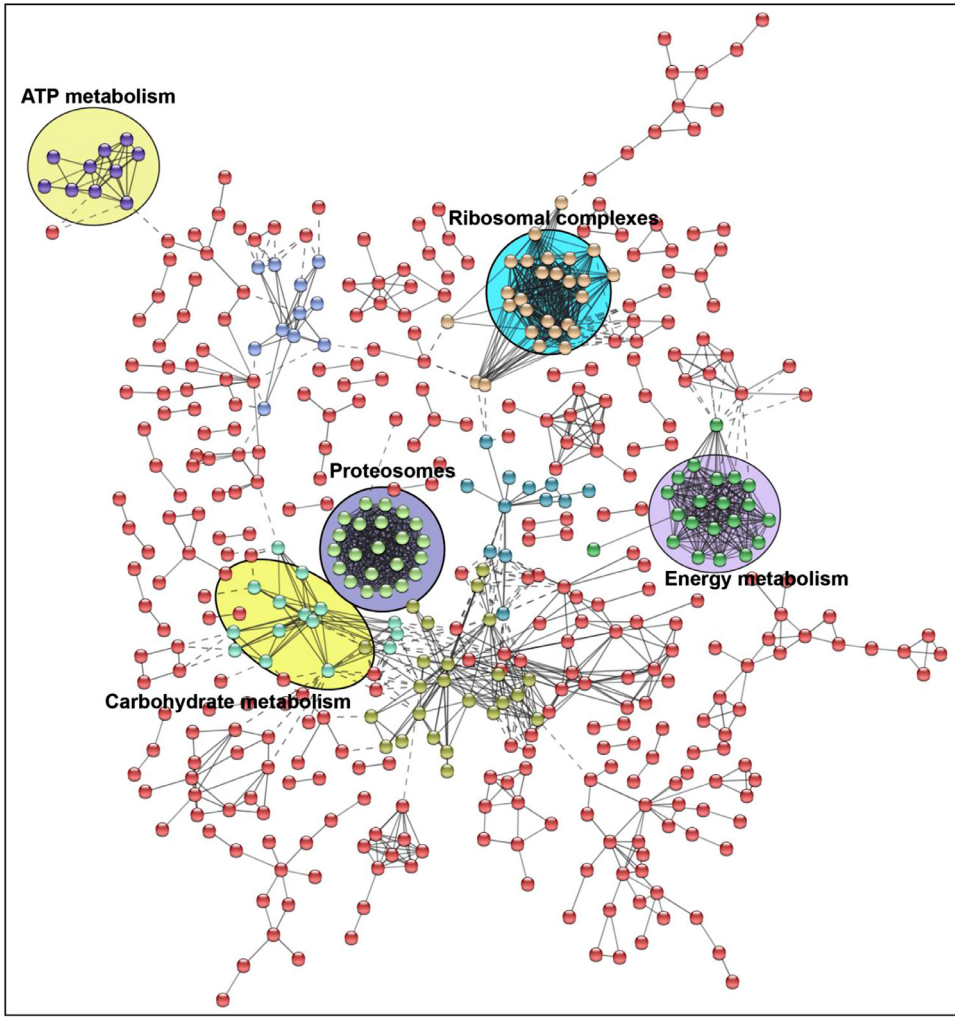

Fig. 3. Predicted protein-protein interaction map of sub-set of proteins identified in the *An. stephensi* brain along with the associated processes. The various processes involved is depicted in different color shades with the details provided in the box provided in the right-hand corner below.

2.3. Fractionation

Fractionation was carried out as discussed in our previously published study [1]. Briefly, for in-gel digestion, extracted proteins (300 µg) was fractionated on 10% SDS PAGE stained with Coomassie blue and bands were cut out based on intensity of protein bands. . Excised bands were processed for overnight trypsin digestion using sequencing grade trypsin (Promega). For in-solution digestion, 500 µg of protein were trypsin digested and lyophilized prior to bRPLC fractionation. Fractionated peptides were concatenated into 14 fractions.

2.4. Mass spectrometry analysis

As discussed in our previously published study, fractions were reconstituted in 0.1% formic acid prior to mass spectrometry analysis [1]. Fractions were analysed on LTQ-Orbitrap Velos and

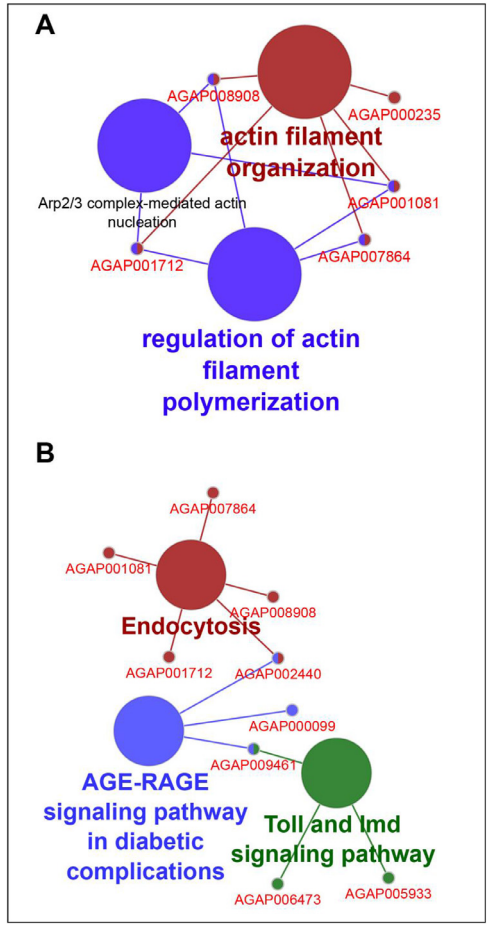

Fig. 4. Representation of various processes and pathways and their associated genes that are expressed in *An. stephensi* brain and have a role in vector-pathogen interactions. A) Biological processes associated with *An. stephensi* brain related genes that play key roles in vector-pathogen interactions. B) Associated pathways and their genes of the genes involved in vector-pathogen interactions and identified in *An. stephensi* brain.

LTQ-Orbitrap-Elite mass spectrometers (Thermo Scientific, Bremen, Germany) interfaced with Easy-nLCII (Thermo Scienific, Bremen, Germany). Data acquisition was carried out in a data dependent manner as described in our previous study [1].

2.5. Data analysis

Mass spectrometry data obtained from our previous study [1] was searched against the reference database containing 11,789 *An. stephensi* proteins along with common mass spectrometry contaminants. Data was analyzed using Proteome Discoverer, version 2.1 (Thermo Fischer Scientific, Bremen, Germany). Overall, 4,662 proteins were identified. Database search parameters included trypsin as digestion enzyme. Amino acid modification considered were Carbamidomethylation of cysteine as static and oxidation of methionine as variable. Results were retrieved using 1% false discovery rate (FDR).

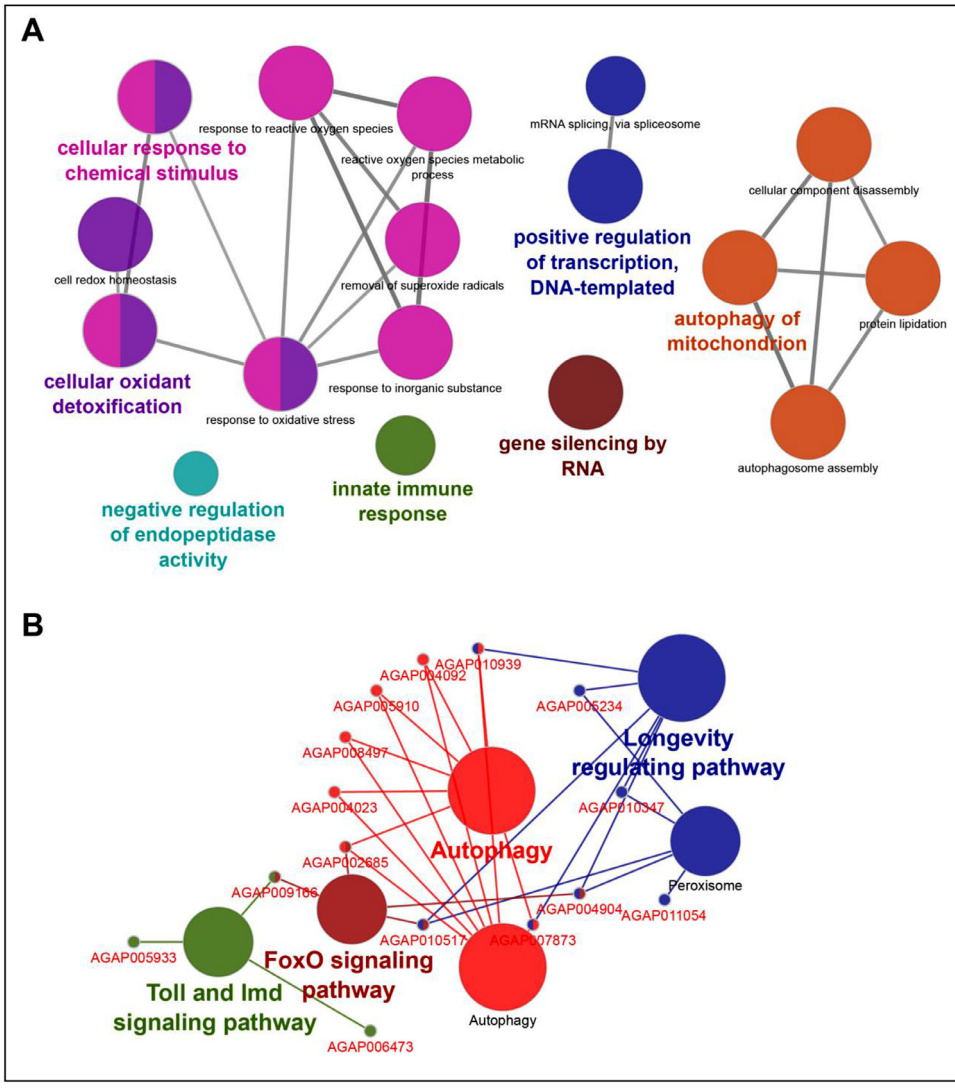

Fig. 5. Immunity associated proteins identified in *An. stephensi* brain as predicted by Immunodb and their associated processes and interactions. A) Biological processes associated with the identified proteins related to immune functions. B) Associated pathways for the proteins associated with immune-related processes.

2.6. Functional categorization and prediction of interaction map

The identified *An. stephensi* proteins were used to fetch corresponding ortholog genes of *An. gambiae* using Biomart online tool available through Vector Base [4-5]. The corresponding *An. gambiae* genes were then used to fetch Gene Ontology terms using Cluego plugging and the resulting figure was generated using Cluepedia plugging in Cytoscape [6–8]. A predicted protein-protein interaction map for the identified proteins were also generated using STRING (Search Tool for the Retrieval Interacting Genes/Proteins) online tool (version 10.5).plugin (version 1.1.0) [9]. Highest confidence parameter was used to retrieve interaction data from STRING database.

Possible immunological associations of the identified proteins were inferred from the information provided in Immunodb [10].

Ethics Statement

No human subjects or lab animals were used for this study. Mosquitoes used for the study were obtained after approval from National Institute of Malaria Research (NIMR) Scientific Research and Ethics Committee.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

We thank Institute of Bioinformatics and ICMR- National Institute of Malaria Research for institutional support and research facilities. 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. The NIMR publication committee has approved this manuscript. We also thank Karnataka Biotechnology and Information Technology Services (KBITS), Government of Karnataka for the support to the Center for Systems Biology and Molecular Medicine at Yenepoya (Deemed to be University) under the Biotechnology Skill Enhancement Programme in Multiomics Technology (BiSEP GO ITD 02 MDA 2017).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2020.106243.

References

[1] T S Keshava Prasad, et al., "Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes, Genome Res. 27 (1) (2017) 133–144, doi:10.1101/gr.201368.115.
[2] Sreelakshmi K Sreenivasamurthy, et al., "A compendium of molecules involved in vector-pathogen interactions pertaining to malaria, Malar. J. 12 (26 Jun. 2013) 216, doi:10.1186/1475-2875-12-216.
[3] Sutopa B Dwivedi, et al., "Brain proteomics of Anopheles gambiae, Omics: J. Integr. Biol. 18 (7) (2014) 421–437, doi:10.1089/omi.2014.0007.
[4] Damian Smedley, et al., The BioMart community portal: an innovative alternative to large, centralized data repositories, Nucleic Acids Res. 43 (W1) (2015) W589–W598, doi:10.1093/nar/gkv350.
[5] Gloria I Giraldo-Calderón, et al., "VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases, Nucleic Acids Res. 43 (Database issue) (2015) D707–D713, doi:10.1093/nar/gku1117.
[6] Paul Shannon, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504, doi:10.1101/gr.1239303.
[7] Gabriela Bindea, et al., ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, Bioinformatics 25 (8) (2009) 1091–1093, doi:10.1093/bioinformatics/btp101.
[8] Gabriela Bindea, et al., CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data, Bioinformatics 29 (5) (2013) 661–663, doi:10.1093/bioinformatics/btt019.
[9] Damian 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, doi:10.1093/nar/gkw937.
[10] Robert M Waterhouse, et al., "Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes, Science (New York, N.Y.) 316 (5832) (2007) 1738–1743, doi:10.1126/science.1139862.