Acta Tropica 152 (2015) 165–169

Contents lists available at ScienceDirect

Acta Tropica

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

Artificial Neural Network applied as a methodology of mosquito species identification

Camila Lorenz[a,b,*], Antonio Sergio Ferraudo[c], Lincoln Suesdek[a,d]

[a] Instituto Butantan—Avenida Vital Brasil, 1500, Sao Paulo CEP 05509-300, Brazil
[b] Biologia da Relação Patógeno-Hospedeiro, Instituto de Ciências Biomédicas, University of Sao Paulo, Av. Prof. Lineu Prestes, 2415, Sao Paulo CEP 05508-000, Brazil
[c] Universidade Estadual Paulista Julio de Mesquita Filho, Av. Prof Paulo Donato Castellane s/n, Jaboticabal, Sao Paulo CEP 14884-900, Brazil
[d] Programa de Pós-Graduação em Medicina Tropical, Instituto de Medicina Tropical, Universidade de São Paulo, São Paulo, SP, Brazil

ARTICLE INFO

Article history:
Received 29 January 2015
Received in revised form 29 August 2015
Accepted 17 September 2015
Available online 21 September 2015

Keywords:
Wing
Geometric morphometrics
Parasite vector
Principal components
Artificial neural network

ABSTRACT

There are about 200 species of mosquitoes (Culicidae) known to be vectors of pathogens that cause diseases in humans. Correct identification of mosquito species is an essential step in the development of effective control strategies for these diseases; recognizing the vectors of pathogens is integral to understanding transmission. Unfortunately, taxonomic identification of mosquitoes is a laborious task, which requires trained experts, and it is jeopardized by the high variability of morphological and molecular characters found within the Culicidae family. In this context, the development of an automatized species identification method would be a valuable and more accessible resource to non-taxonomist and health professionals. In this work, an artificial neural network (ANN) technique was proposed for the identification and classification of 17 species of the genera *Anopheles*, *Aedes*, and *Culex*, based on wing shape characters. We tested the hypothesis that classification using ANN is better than traditional classification by discriminant analysis (DA). Thirty-two wing shape principal components were used as input to a Multilayer Perceptron Classification ANN. The obtained ANN correctly identified species with accuracy rates ranging from 85.70% to 100%, and classified species more efficiently than did the traditional method of multivariate discriminant analysis. The results highlight the power of ANNs to diagnose mosquito species and to partly automatize taxonomic identification. These findings also support the hypothesis that wing venation patterns are species-specific, and thus should be included in taxonomic keys.

1. Introduction

Currently, 3,543 mosquito species of the Culicidae family have been identified (http://mosquito-taxonomic-inventory.info/, accessed 01/Dec/2014), but only about 200 of them are recognized vectors of pathogens that cause disease in humans, including viruses (arboviruses), filarial worms (helminths), and protozoa (malaria). These identified species are primarily members of the genera *Anopheles*, *Aedes*, and *Culex*. The relationship between mosquitoes and public health has driven much research, which focuses primarily on species of medical interest (Reidenbach et al., 2009). Correct identification of these species is essential, to both recognize the vectors involved in pathogen transmission, and to develop efficient control strategies (World Health Organization, 1995). Morphological identification of specimens is currently restricted to a small number of highly specialized professionals, owing to difficulty in character interpretation by the laity. Molecular identification of mosquito DNA samples remains a slow and expensive process for most laboratories.

In recent years, the geometric morphometrics technique (Bookstein, 1991; Rohlf and Marcus, 1993) has proven to be a useful tool for identifying species based on polymorphic morphological characters (Calle et al., 2002; Lorenz et al., 2012; Villemant et al., 2007). In the case of mosquitoes, the wing is more commonly analysed by morphometrics, owing to its two-dimensional feature and the ease by which researchers can obtain anatomical landmarks (Gómez et al., 2013; Jaramillo et al., 2014; Demari-Silva et al., 2014). Typically, morphometrical data are statistically assessed using multivariate discriminant analyses (DA) (Dujardin, 2008; Lorenz et al., 2014). Artificial Neural Networks (ANN) is another methodology recently utilized to identify organisms based on morphological traits (Hopfield, 1988; Wang, 2003; Yegnanarayana, 2009). ANN

* Corresponding author at: Instituto Butantan—Avenida Vital Brasil, 1500, Sao Paulo CEP 05509-300, Brazil.
E-mail address: camilalorenz@usp.br (C. Lorenz).

http://dx.doi.org/10.1016/j.actatropica.2015.09.011
0001-706X/

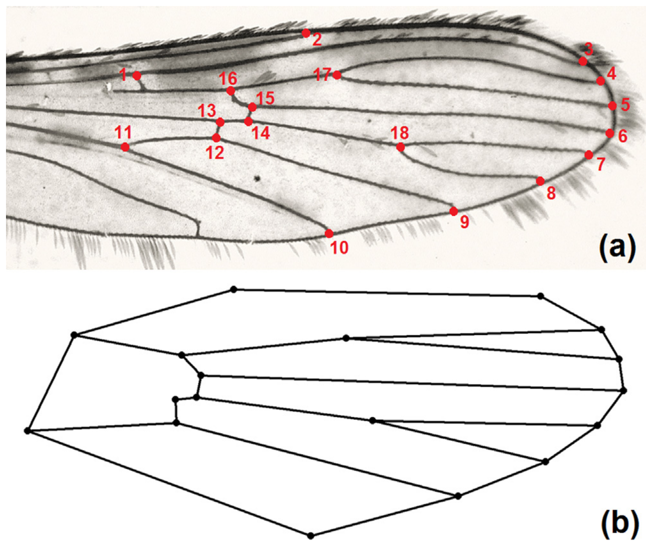

Fig. 1. Geometric Morphometrics. (a) Wing of *Anopheles triannulatus* showing the 18 landmarks used in analyses; (b) Hypothetical geometric diagram representing the wing portion considered in this study.

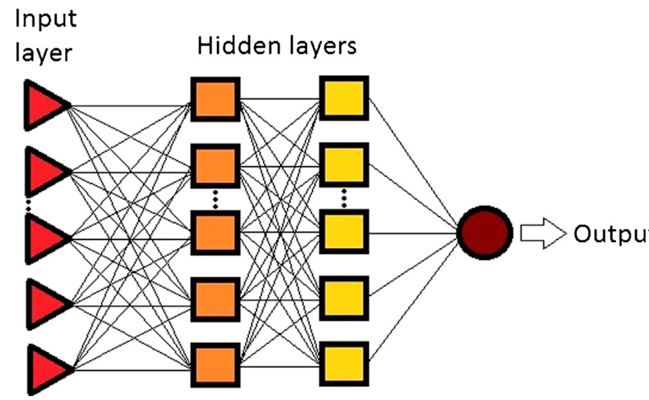

Fig. 2. Schematic representation of an artificial neural network with two hidden layers and one output.

has been adopted from the field of computational intelligence and is being used with increasing frequency to classify biological organisms (Dobigny et al., 2002; Marcondes and Borges, 2000), due to its efficiency in handling large datasets and its generalizability. Some types of ANN employ ratios between dimensions of morphological structures to classify and identify species (Marcondes and Borges, 2000). Recent studies have also used genetic characters in ANNs to identify species of mosquito (Banerjee et al., 2008; Venkateswarlu et al., 2012).

An efficient classification system should be able to extract relevant features from biological data in order to distinguish species with minimal to no error. In this work, we combined geometric morphometrics and ANNs to explore the efficiency of this identification methodology in distinguishing among 17 mosquito species of the genera *Anopheles*, *Aedes,* and *Culex* based on wing shape. We hypothesized that ANN is a superior classification method in comparison to DA, the traditional method. DA has been widely used to make statistical inferences from morphometric data, but it is inaccurate and inefficient in some cases (Calle et al., 2002; Lorenz et al., 2012). We wish to further the development of an automatized species identification method, which would provide non-taxonomists with the ability to accurately and feasibly identify mosquitoes.

2. Material and methods

Specimen identification was based on larvae, pupae, and adult mosquito characteristics compiled from the identification key proposed by Forattini (2002) and features suggested by Sallum et al. (2009). The geometric morphometrics analyzes were based on 18 points, called landmarks, marked on the right wing of each mosquito (Fig. 1). We used only right wings to follow the standard of other morphometric studies on culicids (Lorenz et al., 2012; Gómez et al., 2013). We believe that the use of solely right wing is enough representative, given that all species sampled showed no directional asymmetry (data not shown). We used 388 female specimens, because these are the ones that are hematophages, representing 17 species of the three major genera of disease vectors: *Aedes*, *Anopheles*, and *Culex*. The species sampled and the sampling sites are described in Table 1.

Wings were photographed using a Leica DFC320 digital camera coupled to a Leica S6 microscope with 40× magnification. The landmarks were digitized using TpsDig v 1.40 software (Rohlf, 2006) and coordinate images were plotted onto a Cartesian plane for morphometric analysis. All digital images were scored by the principal author (CL) and the repeatability test considered the marking system acceptable. To reduce the original space of variables while preserving maximum information, we generated new latent variables called principal components. These are eigenvectors, built

Table 1
Information about mosquito species used in this study.

Genus	Species	N	Sampling site	Date	Geographic coordinates
Aedes	*Ae. aegypti*	35	São Paulo/SP	Jan-13	23°29′52.6″S/46°31′33.1″W
	Ae. albopictus	35	São Paulo/SP	Jan-13	23°29′52.6″S/46°31′33.1″W
Anopheles	*An. homunculus*	24	Pariquera-Açu/SP	Jul-14	24°53′25.6″S/47°50′16.0″W
	An. triannulatus s.l.	23	Minas Gerais/MG	Jan-14	20°03′56.3″S/48°58′10.3″W
	An. albitarsis s.l.	18	Minas Gerais/MG	Jan-14	20°03′56.3″S/48°58′10.3″W
	An. aquasalis	36	São Paulo/SP	Mar-14	23°29′52.6″S/46°31′33.1″W
	An. bellator	36	Pariquera-Açu/SP	Jul-14	24°53′25.6″S/47°50′16.0″W
	An. cruzii	24	Pariquera-Açu/SP	Jul-14	24°53′25.6″S/47°50′16.0″W
	An. darlingi	28	Manaus/AM	Nov-13	3°00′13.6″S/59°55′07.5″W
Culex	*Cx. sacchetae*	9	Pariquera-Açu/SP	Jul-14	24°53′25.6″S/47°50′16.0″W
	Cx. usquatus	17	Pariquera-Açu/SP	Jul-14	24°53′25.6″S/47°50′16.0″W
	Cx. atratus	10	Rio de Janeiro/RJ	Nov-14	22°56′46.8″S/43°17′09.6″W
	Cx. coronator	16	Manaus/AM	Nov-13	3°00′13.6″S/59°55′07.5″W
	Cx. ribeirensis	15	Rio de Janeiro/RJ	Nov-14	22°56′46.8″S/43°17′09.6″W
	Cx. cornirger	27	São Paulo/SP	Mar-14	23°29′52.6″S/46°31′33.1″W
	Cx. nigripalpus	19	Manaus/AM	Nov-13	3°00′13.6″S/59°55′07.5″W
	Cx. quinquefasciatus	16	São Paulo/SP	Mar-14	23°29′52.6″S/46°31′33.1″W

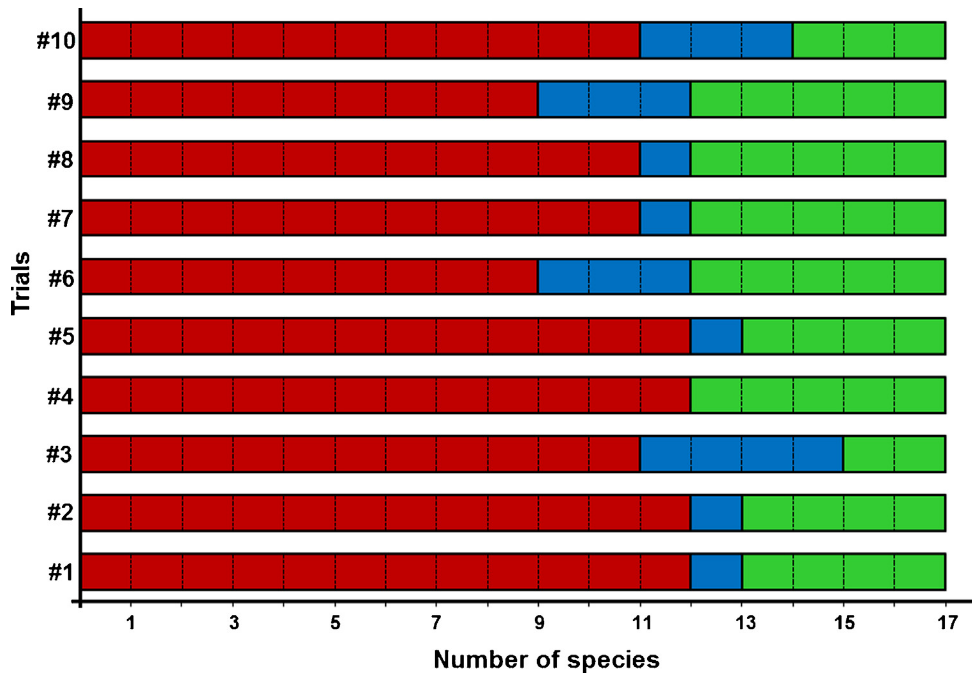

Fig. 3. Number of species that scored highest in correct classification using trials of ANNs. Each single block represents a species. The colours represent a better classification using ANN (red), DA (blue), or both approaches (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

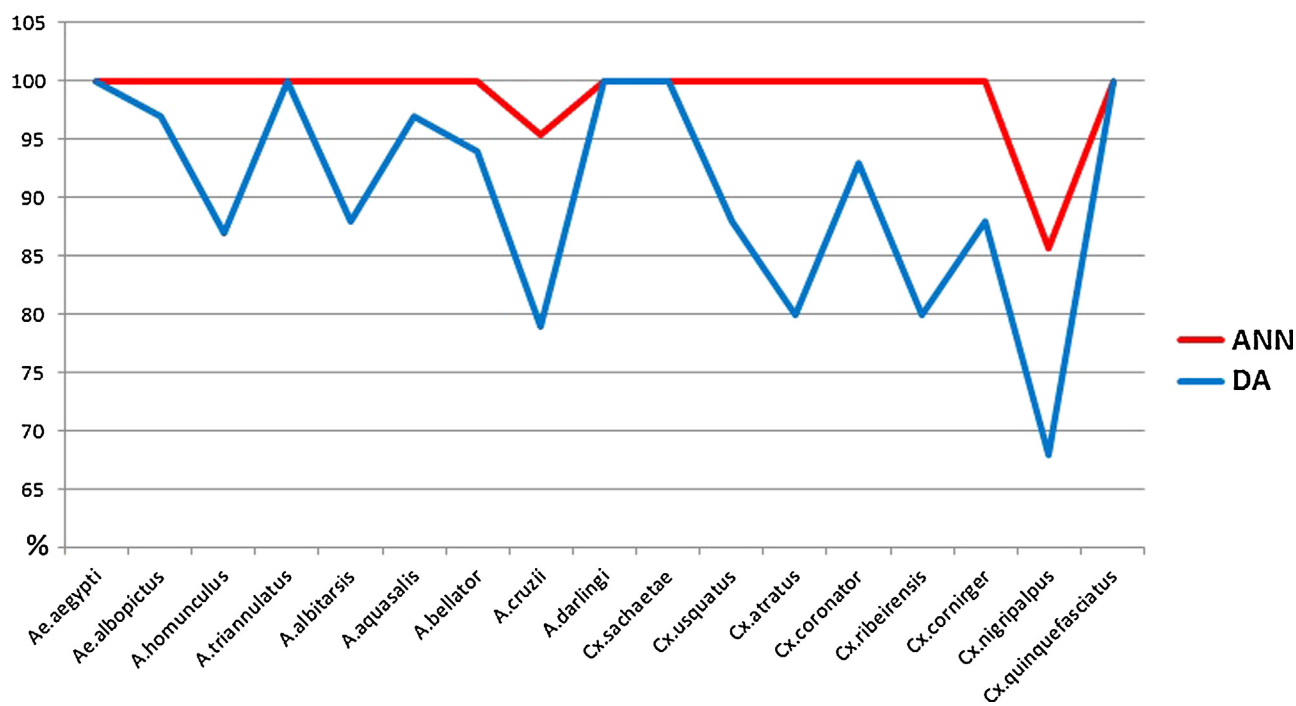

Fig. 4. A comparison between classification by Artificial Neural Network (ANN) and Discriminant Analysis (DA). We used 17 species; their scores of correct classification are represented on the y-axis (%).

with eigenvalues of the correlation matrix of the original variables. This Principal Component Analysis (PCA) was processed in the software MorphoJ 1.02 (Klingenberg, 2011). The 32 obtained principal components were used as input to an ANN using the software StatSoft, Inc. (2004). We calculate the allometric effect of all samples and it was not significant (2.54%; $p = 0.466$). The network model used was multi-layered for classification ("Multilayer Perceptron", Fig. 2) during the modelling process to determine the network topology. Phase 1 of the network had 100 epochs through a back propagation model, and phase 2 had 500 epochs through a conjugate gradient descendent model. Inner layers ranged from 1 to 4, containing up to 20 neurons each. Learning rates ranged from 0.1 to 0.5 and the momentum rate was 0.4. We used the logistic function as the activation function. A group of 300 individuals, randomly selected, were used as model training and 88 were reserved for validation. Discriminant analyses were performed to compare with

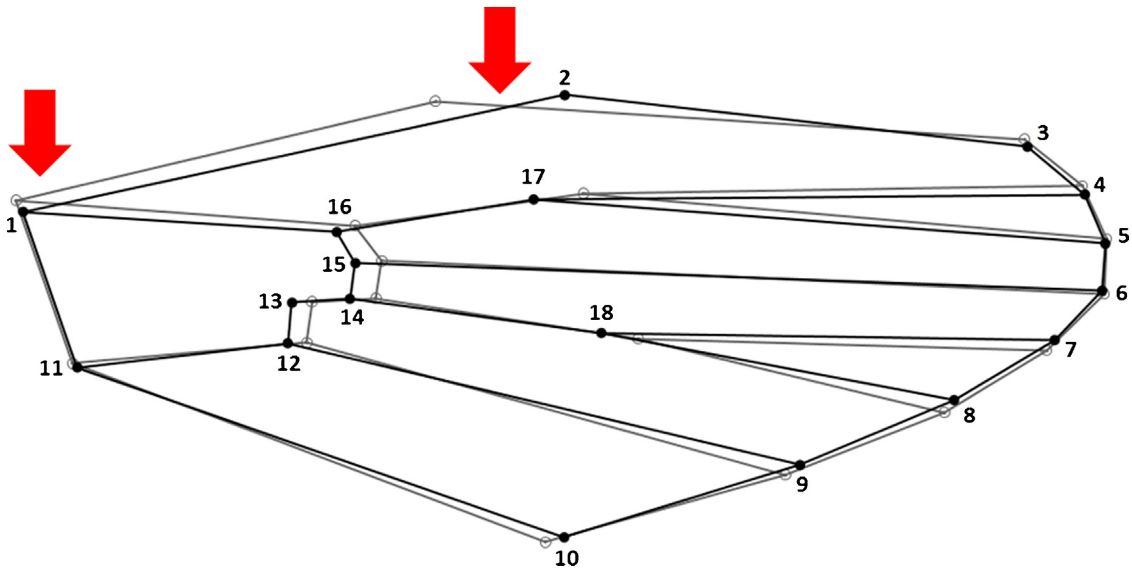

Fig. 5. Variation of the wing. The contours represent the most distinct extremes within the 17 sampled species. The arrows indicate landmarks of greatest variation in shape analysis after Procrustes superimposition.

ANN results and effectiveness. The software PADWIN (Dujardin, 2002) was used with cross-validation among groups.

3. Results and discussion

After heuristically testing several alternatives and comparing their respective performances, the most efficient topology was found to be that of 32 input neurons (corresponding to 32 principal components of wing shape), two hidden layers with 17 neurons each, and one output neuron. It is important to emphasize that ANN is a heuristic classification method, not an exhaustive one; therefore, different results from the same data set may be obtained. After the best network topology was chosen, we reran the same data for a total of 10 trials to verify the efficiency of ANNs (Fig. 3). All of the trials displayed better results than did the traditional discriminant analysis, and in all networks at least nine species obtained a better reclassification score.

We choose the best ANN (Fig. 3—Trial #4) to demonstrate its efficiency compared to discriminant analysis. Classification results ranged from 85.70% to 100%, a better score than the one obtained using discriminant analysis for most of the species (Fig. 4). Only *Anopheles cruzii* and *Culex nigripalpus* showed ranking scores lower than other species. These ANN results agreed with the DA scores, which were also lower in these two species.

The largest variation between species was observed in landmarks #1 and #2, located in the border of the wing (Fig. 5).

The combination of many principal components through ANNs allowed for highly accurate and reliable species identification, more efficient for most species than DA methodology. To statistically evaluate the geometric morphometrics data, the ANN method seems to be a good substitute for traditional DA; all ANN tests here resulted in greater than or equal identification accuracy. The great advantage of using ANN instead of traditional DA is the ability to handle large amounts of data quickly and the possibility of automating the process with constructed models.

This is the first study that has associated geometric morphometrics analysis with ANNs for identifying mosquito species. Using wing shape data as a network input speeds and reduces the process in comparison to the use of genetic data. For mosquitoes analysed in this study there are unique features in the wing venation that allow the correct identification of species. Our results also open a new question to be investigated: which other species present a species-specific pattern?

For a more complete and accurate identification, technique combination would be ideal, as would employing more specimens for method validation. Here, we show that it is possible to distinguish among mosquito species using the wing; however, further study is required that focuses on the combination of ANN with other methodologies to increase the accuracy of results. The classification method proposed here can aid in initial insect screening from the field, and is useful as a complementary tool of traditional identification keys. This technique should be tested in other species groups, such as complex or cryptic species, and in different populations of the same species to verify its effectiveness and generalizability.

Acknowledgements

Thanks to a number of people of the Butantan Institute and School of Public Health of USP who provided field and laboratory assistance or mosquito samples used in this study. C.L. is a fellow of Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Grant #2013/05521-9. This investigation was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Grants #23038.005274/2011-24 and 1275/2011.

References

Banerjee, A.K., Kiran, K., Murty, U.S.N., Venkateswarlu, C., 2008. Classification and identification of mosquito species using artificial neural networks. Comput. Biol. Chem. 32, 442–447.
Bookstein, F.L., 1991. Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press.
Calle, L., David, A., Quiñones, M.L., Erazo, H.F., Jaramillo, O., 2002. Morphometric discrimination of females of five species of Anopheles of the subgenus *Nyssorhynchus* from Southern and Northwest Colombia. Mem. Inst. Oswaldo Cruz 97, 1191–1195.
Demari-Silva, B., Suesdek, L., Sallum, M.A.M., Marrelli, M.T., 2014. Wing geometry of Culex coronator (Diptera: Culicidae) from South and Southeast Brazil. Parasit Vectors 7, 174.
Dobigny, G., Baylac, M., Denys, C., 2002. Geometric morphometrics, neural networks and diagnosis of sibling *Taterillus* species (*Rodentia Gerbillinae*). Biol. J. Linn. Soc. 77, 319–327.
Dujardin, J.P., (2002). Permutations and discriminant analises (PAD). Available at: <mpl.ird.fr/morphometrics>.
Dujardin, J.P., 2008. Morphometrics applied to medical entomology. Infect. Gen. Evol. 8, 875–890.

Forattini, O.P. (2002). Culicidologia Médica: Identificação. Biologia, Epidemiologia. São Paulo: Edusp, 2.
Gómez, G., Jaramillo, L., Correa, M.M., 2013. Wing geometric morphometrics and molecular assessment of members in the Albitarsis Complex from Colombia. Mol. Ecol. Resour. 13 (6), 1082–1092.
Hopfield, J.J., 1988. Artificial neural networks. Circuits Devices Mag. IEEE 4, 3–10.
James Rohlf, F., Marcus, L.F., 1993. A revolution morphometrics. Trends Ecol. Evol. 8, 129–132.
Jaramillo, O.N., Fonseca-González, I., Chaverra-Rodríguez, D., 2014. Geometric morphometrics of nine field isolates of *Aedes aegypti* with different resistance levels to lambda-cyhalothrin and relative fitness of one artificially selected for resistance. PLoS One 9 (5), 1–15.
Klingenberg, C.P., 2011. MorphoJ: an integrated software package for geometric morphometrics. Mol. Ecol. Res. 11, 353–357.
Lorenz, C., Marques, T.C., Sallum, M.A.M., Suesdek, L., 2012. Morphometrical diagnosis of the malaria vectors *Anopheles cruzii*, *An. homunculus* and *An. Bellator*. Parasit Vectors 5, 257.
Lorenz, C., Marques, T.C., Sallum, M.A.M., Suesdek, L., 2014. Altitudinal population structure and microevolution of the malaria vector *Anopheles cruzii*. Parasit Vectors 7, 581.
Marcondes, C.B., Borges, P.S., 2000. Distinction of males of the Lutzomyia intermedia (Lutz & Neiva, 1912) species complex by ratios between dimensions and by an artificial neural network (Diptera: Psychodidae, Phlebotominae). Mem. Inst. Oswaldo Cruz 95 (5), 685–688.
Reidenbach, K.R., Cook, S., Bertone, M.A., Harbach, R.E., Wiegmann, B.M., Besansky, N.J., 2009. Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. BMC Evol. Biol. 9, 298.
Rohlf, F.J., 2006. TpsDig. State University of New York at Stony Brook, Department of Ecology and Evolution, version 2.10.
Sallum, M. A., Dos Santos, C. L., Wilkerson, R.C., (2009). Studies on Anopheles (Kerteszia) homunculus Komp (Diptera: Culicidae). Walter Reed Army Inst of Research Silver Spring MD Dept of Entomology.
StatSoft, Inc., (2004). Statistica for Windows Computer Program Manual. Tulsa, OK: StatSoft, Inc.
Venkateswarlu, C., Kiran, K., Eswari, J.S., 2012. A hierarchical artificial neural system for genera classification and species identification in mosquitoes. Appl. Art. Intell. 26, 903–920.
Villemant, C., Simbolotti, G., Kenis, M., 2007. Discrimination of Eubazus (*Hymenoptera, Braconidae*) sibling species using geometric morphometrics analysis of wing venation. Syst. Entomol. 32 (4), 625–634.
Wang, Sun-Chong, 2003. Artificial neural network. In: Interdisciplinary Computing in Java Programming, Springer US, pp. 81–100.
World Health Organization. (1995). Vector Control for Malaria and Other Mosquito-Borne Diseases: report of a WHO study group.
Yegnanarayana, B. (2009). Artificial neural networks. PHI Learning Pvt. Ltd.