Identifying Chagas disease vectors using Elliptic Fourier Descriptors of body contour: A case for the cryptic *dimidiata* complex

Daryl David Cruz Flores (daryldavidcf@gmail.com)
Centro de Investigacion en Biodiversidad y Conservacion https://orcid.org/0000-0002-7714-2459
Elizabeth Arellano Arenas
Centro de investigaciones en Biodiversidad y Conservación
Dennis Denis Ávila
Departamento de Biología Animal y Humana, Facultad de Biología, Universidad de la Habana
Carlos N. Ibarra-Cerdeña
Centro de Investigacion y de Estudios Avanzados Unidad Merida

Research

Keywords: triatomine, identification, morphometric analysis, contours, Fourier

Posted Date: December 19th, 2019

DOI: https://doi.org/10.21203/rs.2.19259/v1

Abstract

Background Triatoma dimidiata (Reduviidae: Triatominae) is an important vector of Chagas disease in various countries in the Americas. Phylogenetic studies have defined three lineages in Mexico and part of Central America, but methods for identifying them using morphometric analyses with landmarks have not yet been resolved. Elliptical Fourier Descriptors (EFDs), which mathematically describe the shape of any closed two-dimensional contour, could be a potentially useful alternative method. The objective of this work was to validate the use of EFDs for the identification of the three lineages of this species complex.

Method A total of 84 dorsal view photographs of individuals of the three lineages were used. The body contours were described with EFDs using between 5 and 30 harmonics. The number of obtained coefficients was reduced by a Principal Components Analysis and the first axis scores were used as shape variables. A Canonical Variances Analysis, a linear discriminant function analysis and a multilayer perceptron neural networks were then performed using the shape variables, identifying the minimum number of harmonics sufficient to produce efficiently classifications.

Results The first principal component explained 50% of the variability, regardless of the number of harmonics used, but the results of both, Principal Component Analysis and Canonical Variance Analysis get improved by increasing the number of harmonics and components considered. With 25 harmonics and 8 components, the identification of haplogroups was achieved with an overall efficiency greater than 97%. The 30 multilayer perceptron neural networks were also efficient in identification, reaching 91% efficiency with the validation data.

Conclusions The use of EFDs is a simple and useful method for the identification of major lineages of Triatoma dimidiata. This method outperform other novel approaches, therefore could serve as an automated identification method.

Background

Cryptic species are, at present, one of the great challenges for systematic biologists since, in many cases, speciation is not accompanied by distinctive morphological characters and allopatric distributions that facilitate the identification of different entities at the species level [1]. In the case of insects, the presence of cryptic species is a very frequent phenomenon in several orders such as Coleoptera and Hemiptera [2, 3]. In the field of epidemiology, the correct identification of species in insect groups with medical importance is a key component of mapping vector control and surveillance strategies [4]. This is mainly because different species may vary in terms of their competence as vectors and their epidemiological importance as well as in the right way to its control [5].

One of the most epidemiologically important groups of insects on the American continent is the triatomines (Triatominae: Reduviidae), the vectors of Chagas Disease (CD). In this group, the genus Triatoma is the most diverse genus [6]; approximately 70 species have been described and it is the genus with the largest geographic distribution within the subfamily [7, 8]. Multiple inter- and intraspecies

taxonomic questions have arisen in this group, with species repeatedly included and excluded from different complexes throughout the history of the study of their systematics and taxonomy [9, 10, 11, 12]. The combination of unresolved taxonomic relationships and the detection of cryptic species within this genus highlight the need to address the systematics of this group [13, 14, 15, 16]. The phenomenon of cryptic speciation is common in Triatominae [17, 18, 19, 20] and results in species that are nearly identical morphologically, which often makes identification based only on traditional morphological characters difficult or impossible.

The identification of triatomine species has usually been carried out using traditional morphometry [9, 12, 21]. However, the use of geometric morphometry has led to new techniques for evaluating morphological characters in a taxonomic context; it complements the use of other methods of discrimination [22] and has been used for the recognition of apparently cryptic species, including some of the Triatoma genus [5, 23].

The Triatoma dimidiata complex represents one of the major vectors of Chagas disease in all of the countries where it is distributed [16, 24, 25]. It is present in Mexico, all of the countries of Central America, Colombia, Ecuador and Perú [16, 26]. Throughout its range, it can be found in jungle, peridomestic and domestic habitats, where non-domiciled populations act as sources of reinfestation and participate in the transmission of the parasite to humans [26, 27, 28, 29].

Phylogenetic studies defined three major haplogroups for Mexico and part of Central America [13], which were recently reaffirmed by Pech-May et al. [16]. Although the three haplogroups are completely differentiable by molecular analysis, they have not yet been successfully delimited using morphometric approaches.

Through the use of geometric morphometry techniques by anatomical reference points, Gurgel-Goncalves et al. [30] implemented a method for the automated identification of 51 triatomine species, including the three haplogroups of T. dimidiata. They reached correct identification rates of 70.5%, 76.7% and 82.5% for haplogroups 1, 2 and 3 of this species, respectively. More recently, using TensorFlow, a deep learning algorithm, it was possible to increase the correct classification of specimens of the three haplogroups (84.1% H1, 86.7% H2 and 87.5% H3) [31]. Althought these methodologies are the cutting edge approach to the automatized species identification within the Triatomine group, this rate of identification is still insufficient and methods that guarantee higher power of correct discrimination are still necessary.

As an alternative, in this study we propose the use of Elliptical Fourier Descriptors (EFDs), which can delineate any shape with a two-dimensional closed contour, as suggested by Kuhl and Giardina [32]. Contour analysis is based on the digitalization of the silhouette of an object, which is expressed as a sequence of coordinates (x, y) that can be manipulated mathematically and adjusted to an equation derived from Fourier functions. This method has been widely applied to the analysis of various biological shapes [33, 34, 35] and more recently, as a tool for pattern detection and automated species identification [36, 37, 38, 39, 40]. Here, we apply EFDs in order to evaluate their ability to identify the three described Triatoma dimidiata haplogroups for Mexico and part of Central America. The results of this evaluation

contribute to the implementation of tools for accurate discrimination between triatomine species and potentially to the control and prevention of CD.

Materials And Methods

In order to test the ability of EFDs to discriminate among the haplogroups of T. dimidiata, we used the images obtained by Gurgel-Goncalves et al. [30], which are available in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.br14k. Originally, 44, 30 and 40 images of haplogroups 1, 2 and 3 were obtained, respectively, with which the automated identification process tested by Gurgel-Goncalves et al. [30] was performed. For this study, only images that had the necessary characteristics to perform the contour analysis were selected, that is, only images with an unmodified contour and wings that were not broken or overlapped. This filtering process resulted in a total sample of 37, 23 and 36 images for haplogroups 1, 2 and 3, respectively. The conditions under which the photographs were taken are detailed in Gurgel-Goncalves et al. [30].

The images were preprocessed in Adobe Photoshop CS5. The legs and antennae were removed from each image to leave only the body contour. The brightness and contrast values were adjusted to their minimum and maximum values, respectively, to leave only a binary image (Fig. 1). All images were saved as bitmaps (BMP) in 24-bit RGB format.

SHAPE 1.3 software [41], designed to evaluate the contour shape based on Elliptical Fourier Transform, was then used to quantify the body contours. The mathematical description of contour extraction based on EFDs can be found in Iwata et al. [42].

SHAPE has four subprograms: ChainCoder, Chc2Nef, PrinComp, and PrinPrint, which together facilitate the processing of digital images, acquisition of the chain code and Fourier coefficients, and Principal Component Analysis. It also includes routines for the visualization of the shape from previously digitized data (ChcViewer and NefViewer).

The chain code is a coding system to describe the spatial information of the contours with numbers from 0 to 7 [43]: the digits indicate the direction of the next step around an outline: 0 = one step to the right, 2 = one step up, 4 = one to the left, 6 = one down, and the other digits are intermediate addresses. To obtain this code for each image, the ChainCoder subprogram was implemented for the images of the haplogroups. This subprogram reads the BMP images, converts them to grayscale, binarizes them from a threshold value selected in the image histogram, eliminates possible noise existing in the images using erosion-dilution filters and obtains the chain code, which is saved in an ASCII file with an extension chc.

Once the chain code file was generated for each image using the Chc2Nef program [32], the Fourier transform coefficients for 5, 10, 15, 20, 25 and 30 harmonics were calculated consecutively, using the ellipse as normalization of the first harmonic. Fourier coefficients were stored in an ASCII file of extension .nef that were used for subsequent multivariate analysis.

Given that a large number of variables is produced (four coefficients for each harmonic), a Principal Component Analysis (PCA) was performed using the variance-covariance matrices to reduce the dimensionality and create new derived variables that can be analyzed statistically. This was done using the PrinComp module, as proposed by Rohlf and Archie [34], and component scores were used as shape variables. These derived variables contain all of the information for each haplogroup body shape, as demonstrated by the fact that the contours can be graphically reconstructed from these values using an inverse Fourier transform in the PrinPrint module, according to the procedure of Furuta et al. [35].

To evaluate the ability of FED to discriminate among the three haplogroups of T. dimidiata, a Discriminant Function Analysis was performed. The minimum number of harmonics needed to produce satisfactory classifications was determined. For this, the scores of the principal components recovered from the PrinComp module were used. For the first five harmonics, the number of principal components was 16, while for 10, 15, 20, 25 and 30 harmonics, 30 principal components were recovered. A Canonical Variance Analysis (CVA) was performed with the minimun number of harmonics that allowed the best discrimination of the haplogroups and the confusion matrix was obtained to estimate the classification errors.

Finally, we compared each of the first five principal components among the three haplogroups to determine whether there were statistically significant differences among them. Because the data were not normally distributed, we used a Kruskal-Wallis test to compare among the three groups.

As an alternative method of discrimination and identification, a multilayer perceptron neural networks were trained. Artificial neural networks are mathematical models constructed by simulating the functioning of biological neural networks (the nervous system). They present a set of processing units called neurons, cells or nodes (formed by several mathematical equations), interconnected by connections that include a weight that modifies the values that pass through them between neurons [44].

The variables used to make the network were the scores of the principal components that contributed most to the total variance, obtained from the Fourier coefficients from 25 harmonics. For the basic topology, the automated search procedure of Statistica version 8.0 software was used, with an input layer of 30 neurons, corresponding to each shape variable, and the output layer with three neurons, one for each haplogroup to identify.

In the exploratory step, the most efficient network was evaluated by testing with hidden layers of between 10 and 40 neurons. Two error functions (sum of squares and cross-entropy) and four activation functions (identity, logistics, tangent and exponential) were used. The learning rate was 0.1, the inertia 0.66, and the stoping rule was set when the training error was below 0.001. Network learning was represented using the behavior of the maximum, average and minimum errors. Sixty percent of the data were randomly selected for network training and the remaining 40% was used for validation. Of the 30 networks, the one with the lowest classification error of the validation data was selected as best. The classification power for the species was analyzed using the confusion matrix and the calculation of the percentages of omission and commission errors.

Results

With the result of the first component, and when using the inversion of the Fourier transforms, the contour of the haplogroups of T. dimidiata was reconstructed and the variability between groups and within them was graphically characterized (Fig. 2). The greatest variability between haplogroups was observed in the posterior lobe of the pronotum and the terminal region of the head and neck. This pattern of variation was also observed internally for haplogroups 1 and 2. The greatest variation within haplogroup 3 specimens was in the anterior lobe and distal tubers (Fig. 2).

Regardless of the number of harmonics used to describe the contour, the first component explained about half of the contour variability (between 44 and 55%) (Fig. 3). As the number of harmonics used increased, more components were required to explain 90% of the variation in shape, but in general, this value was reached with 8 principal components (Fig. 3).

When performing the discriminant analysis using different numbers of harmonics, correct discrimination generally increased with the number of harmonics used, except for 30 harmonics when discrimination began to descend (Table 1). Haplogroup 1 was successfully differentiated from the analysis with 10 harmonics onwards with 100% correct discrimination. Haplogroup 2 was only 100% correctly discriminated when the contour was described with 20 and 25 harmonics. Haplogroup 3 could not be discriminated completely with any of the number of harmonics tested, but had correct discrimination vales of 88.24 to 94.12%. Overall, the best results were obtained by describing the contours of haplogroups using 25 harmonics.

Table 1
Percentage of correct discrimination of three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae) by discriminant analysis using 5, 10, 15, 20, 25 and 30 harmonics.

	Correct classification (%)			
No harmonics	Haplogroup 1	Haplogroup 2	Haplogroup 3	Total
5	97.14	60	88.24	85.39
10	100	85	94.12	94.38
15	100	90	94.12	95.51
20	100	100	91.18	96.63
25	100	100	94.12	97.75
30	100	95	94.12	96.63

The ordering diagram of the discriminating axes for the shape of the specimens, for the description of the contour with 25 harmonics and using 30 components is shown in Fig. 4. The separation of the minimum

convex polygons demonstrated the possibility of discriminating the haplogroups using the principal components scores as shape variables. Haplogroup 1 is separated perfectly from the rest, showing the greatest differentiation in canonical axis 1 from haplogroups 2 and 3. Haplogroups 2 and 3 presented greater variation along canonical axis 2. One individual from haplogroup 3 was located within the polygon of haplogroup 2, which was corroborated as an error of discrimination of the analysis in the confusion matrix (Table 3).

Figure 4. Canonical Variance Analysis using shape variables (30 principal components resulting from the Elliptical Fourier Coefficients of 25 harmonics) of three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae).

Table 3
Confusion matrix of the discrimination process of the three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae) by the Canonical Variance Analysis for 25 harmonics.

Haplogroups	H1	H2	H3	Total
H1	37	0	0	37
H2	0	23	0	23
H3	0	1	35	36

When comparing the PC scores among the three haplogroups, significant differences were found between at least two haplogroups for all principal components except for PC3 (Fig. 5). The greatest differences were always between haplogroups 1 and 3.

Figure 5. Differences in the scores of the first five principal components between three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae). H1: Haplogroup 1, H2: Haplogroup 2, H3: Haplogroup 3.

All trained networks achieved 100% correct classification with training data, but the most efficient with validation data was a perceptron of 13 neurons in the hidden layer, which reached 91% correct classification. This network used a BFGS18 training algorithm and an SOS error function. The activation function of the hidden layer was Logistics and for the output layer, Tangent. This network confused only one individual of haplogroup 1 (out of a total of 20) which was classified as haplogroup 2 (droping to 94% correct classification), and two haplogroup 2 individuals (of 16) were classified as haplogroup 3 (for 85% correct classification). The eight individuals belonging to haplogroup 2 used to validate the network were correctly classified (100%). Components 1, 5, 15 and 2 were the higest weight in the network.

Discussion

In entomological studies, much attention has been given to the use of different parts of the body to identify, name and classify insects [45]. To date, the wings have been the most commonly used structures to assess species discrimination through geometric morphometry methods, mainly through the use of anatomical reference points [5, 23]. Here, we use for the first time the whole body contour of an insect to discriminate between groups. The use of EFDs has been little explored, at least in triatomines, though on several occasions it has demonstrated its ability to discriminate among even closely-related species [46, 47]. However, recent works such as Santillán-Guayasamín et al. [48] have used semi-markings and EFDs for the delimitation of triatomine species using the contour of their eggs, demonstrating the usefulness of these methods for species recognition in this genus.

When using the inversion of the Fourier transforms, it was possible to visualize that the greatest differences in contour shape between the haplogroups were found in the pronotum and the head. Both of these structures have been used in morphometric studies, both traditional and geometric, because important variations in their shape have been detected [5, 12, 30]. In the case of the head, Bustamante et al. [12] consider that an important factor in the variability observed in this region is due to the geographic isolation of the populations of T. dimidiata, which has led to divergent evolution. The haplogroups used in this study mainly have allopatric populations, which could explain the morphometric differences found, although there are areas of sympatry [16]. In turn, differences in the head may have an evolutionary cause related to feeding strategies and growth patterns of this area of the body. Authors such as Patterson [49] have suggested that the shape of the head may reflect evolutionary mechanisms related to the ability to ingest blood. If the allopatry of the haplogroups populations of T. dimidiata is taken into account and that these must have diverged approximately 0.97 to 0.85 mya, according to results obtained from sequences of the ND4 gene [16], dissimilar feeding strategies may have been established among the haplogroups, which then generated morphological differences.

In the case of the pronotum, significant variability in the shape of the contour between the haplogroups was also observed. This structure has been used in the traditional morphological description of triatomine species [9] and has been used to try to discriminate between species [50]. Very few studies have used geometric morphometry techniques on this structure in triatomines, and in the future a comparative study of the pronotum could evaluating its utility for discriminating among triatomines.

When comparing the results obtained by Gurgel-Goncalves et al. [30], who reached correct discrimination values between 70.5 and 82.5% of the three haplogroups and the results obtained by Khalighifar et al. [31] (with correct discrimination values of 84.1% H1, 86.7% H2 and 87.5% H3) our results reached 100% correct discrimination values for haplogroup 1 and 2 and 94.12% for haplogroup 3, with total discrimination results of 97.75%, through discriminant function analysis. As with the discriminant function analysis, with the analysis of artificial neural networks, discrimination values greater than those reported by the previous authors were obtained. This is probably because, in comparison to the methods used in the aforementioned studies, EFDs can recover a greater variability of the shape through the contour analysis. Perhaps the integration of both, the methods of the previous works [30, 31] (which have demonstrated their ability with good values of correct discrimination in the recognition of these

haplogroups) and EFDs can help to establish an identification system of the haplogroups of T. dimidiata with higher values of correct identification.

This method of describing shapes and reconstructing images is advantageous when the analysis based on anatomical reference points fails to fully discriminate the objects of study. McLellan and Endler [51] suggested that the use of EFDs provides a precise reconstruction of the contour of the complex object and can explain the overall complexity of the shape with greater resolution than the methods of anatomical reference points and semi-marks. This has been demonstrated in other insect groups, where the use of EFDs has allowed the correct discrimination between species [52]. Francoy et al. [22] used both methods (anatomical reference points and Elliptical Fourier Descriptors) for the identification of euglossine bees. These authors found better results in the differentiation of species using EFD. However, they suggest the combined use of data matrices obtained by anatomical reference points and EFDs.

Species concepts and delimitation have always been highly controversial and complicated, especially when the focal organisms are considered cryptic or hypercryptic [45]. In the Triatoma genus, the presence of cryptic species has been widely addressed. Several authors have assembled the Triatoma species into different groups and complexes based on their external characters and the genitalia of both sexes [53, 54, 55]. Currently, the most accepted group was proposed by Schofield and Galvão [56], with the subdivision of Triatoma species into groups, complexes, and subcomplexes [57].

Triatominae species show high morphological variation, which suggests that ecological factors may be the main force driving speciation in Triatominae [21]. Very closely related species can develop rapid morphological changes in adaptation to new environments. Conversely, similar morphs adapted to the same ecotope could be derived from different ancestors [21]. Thus the existence of morphologically similar species could be reflecting their evolution from a common ancestor or convergent adaptation to the same ecological niche. This phenotypic flexibility leads to the misidentification of distinct genetic units by morphological convergence, resulting in taxonomic uncertainties in the description of new subspecies, species or even genera. Considering that the Triatominae species groupings into complexes and subcomplexes are mainly based on morphological similarities [58], the morphological plasticity complicates both species identification and the establishment of evolutionarily related groups. In this sense, traditional morphological analysis has failed to clarify the differences that other sources of evidence, such as genetic, chromosomal, karyotype analyses, etc., have contributed to the clarification of the cryptic species complexes.

Specifically, in T. dimidiata, wide distribution and variation in morphology (historically explained by wide clinal variation along its distribution range) [9], has resulted in a long history of reconsiderations of its taxonomic status, from a single species to a species complex of distinct taxonomic groups [59]. Studies focused on the analysis of morphological variation using classical morphometry techniques have led to the inclusion of T. dimidiata populations within other species complexes such as phyllosoma [12]. However, these considerations have been rejected due to genetic evidence that has demonstrated the

presence of different haplogroups within the dimidiata complex; this demonstrates that it is impossible for classical morphological techniques to correctly discriminate among these haplogroups.

Conclusions

The use of Elliptic Fourier Descriptors allows the identification of three haplogroups of Triatoma dimidiate with higher precision than previous works. The main advantage is its easy application from easily obtainable digital images with minimal and uncomplicated processing, which guarantees its replicability. Despite its relative mathematical complexity, it can be fully automated, which minimizes the researcher manipulation errors when processing the samples. Its ability to reconstruct the shape automatically, after statistical processing, is also attractive and does not require any drawing skills from the researcher, allowing the visual identification of the location of the differences detected. Also, in principle, it uses a holistic approach to the true shape, regardless of the absolute physical dimensions. The assessment of the identification ability of this method in other triatomine species is a necessary aspect to advance procedures that allow the automation of the identification of these important vectors of Chagas disease.

Declarations

Acknowledgments

Authors want to thank MSc. Daily Martínez Borrego who provided comments on the manuscript.

Authors' contributions

DDCF, EAA conceived the study. DDCF, DDA conducted all statistical analyses. DDCF, EAA, DDA, and CNIB wrote the manuscript. All authors contributed to the final draft of the manuscript. All authors read and approved the final manuscript.

Funding

The first author was supported by a CONACYT scholarship program 2018-000012-01NACF-11846.

Availability of data and materials

Data supporting the conclusions of this article are included in the article. Raw data are available upon request to the first author. Also, all data derived from this investigation are deposited in the Figshare repository (https://doi.org/10.6084/m9.figshare.11344073.v1)

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Rivera PC, González-Ittig R, Robainas A, Trimarchi LI, Levis S, Calderón G, Gardenal C. Molecular phylogenetics and environmental niche modeling reveal a cryptic species in the *Oligoryzomys flavescens* complex (Rodentia, Cricetidae). Journal of Mammalogy. 2018;99(2):363-376.
2. Jackson JK, Resh VH. Morphologically cryptic species confound ecological studies of the caddifishy genus *Gumaga* (Trichoptera: Sericostomatidae) in Northern California. Aquat. Insect. 1998;20:69-84.
3. Schonrogge K, Barr B, Wardlaw JC, Napper E, Gardner MG, Breen J, Elmes GW, Thonas JA. When rare species become endangered: cryptic speciation in myrmecophilous hoverfishes. Biol. J. Linn. Soc. 2002;75:291-300.
4. Abad-Franch F, Monteiro FA. Molecular research and the control of Chagas disease vectors. Anais da Academia Brasileira de Ciências. 2005;77(3),437-454.
5. Gurgel-Goncalves R, Ferreira JBC, Rosa AF, Bar ME, Galvao C. Geometric morphometrics and ecological niche modelling for delimitation of near-sibling triatomine species. Medical and Veterinary Entomology. 2011;25(1):84-93.
6. Martínez FH, Villalobos GC, Cevallos AM, De la Torre P, Laclette JP, Alejandre-Aguilar R, Espinoza B. Taxonomic study of the *Phyllosoma* complex and other triatomines (Insecta: Hemiptera: Reduviidae) species of epidemiological importance in the transmission of Chagas disease using ITS-2 and mtCytB sequences. Molecular Phylogenetics and Evolution. 2006;41:279-287.
7. Panzera F, Hornos S, Pereira J, Cestau R, Canale D, Diotaiutu L, Dujardin JP, Pérez R. Genetic variability and geographic differentiation among three species of triatomine bugs (Hemiptera: Reduviidae). *J. Trop. Med. Hyg*. 1997;57:732-739.
8. Ramsey JM, Peterson AT, Carmona-Castro O, Moo-Llanes DA, Nakazawa Y, Butrick M, Tun-Ku E, de la Cruz-Félix K, Ibarra-Cerdeña CN. Atlas of Mexican Triatominae (Reduviidae: Hemiptera) and vector transmission of Chagas disease. Mem Inst Oswaldo Cruz. 2015;Vol.110(3):339-352.
9. Lent H, Wygodzinsky P. Revision of Triatominae (Hemiptera: Reduviidae) and their significance as vector of Chagas' disease. Bull. Am. Museum. Nat. His. 1979;163: 123–520.
10. Schofield CJ. Triatominae: Biology and Control. Eurommunica Publications, W. Sussex, UK. 1994;80 pp.
11. Flores A, Magallón-Gastélum E, Bosseno MF, Ordoñez R, Kasten FL, Espinoza B, Ramsey J, Breniere F. Isoenzyme variability of five principal triatomine vector species of Chagas disease in Mexico. Genet. Evol. 2001;1:21-28.

12. Bustamante DM, Monroy C, Menes M, Rodas A, Salazar-Schettino PM, Rojas G, Pinto N, Guhl F, Dujardin JP. Metric variation among geographic populations of the Chagas vector *Triatoma dimidiata* (Hemiptera: Reduviidae: Triatominae) and related species. J Med Entomol. 2004;41:296-301.
13. Bargues MD, Klisiowicz DR, Gonzalez-Candelas F, Ramsey JM, Monroy C, Ponce C, Salazar-Schettino PM, Panzera F, Abad-Franch, Sousa OE, Schofield CJ, Dujardin JP, Guhl F, Mas-Coma S. Phylogeography and genetic variation of *Triatoma dimidiata*, the main Chagas disease vector in Central America, and its position within the genus *Triatoma*. PLoS Negl. Trop. Dis. 2008;2(5):e233. DOI:10.1371/journal.pntd.0000233.
14. Ibarra-Cerdeña CN, Zaldívar-Riverón A, Peterson AT, Sánchez-Cordero V, Ramsey JM. Phylogeny and niche conservatism in North and Central American triatomine bugs (Hemiptera: Reduviidae: Triatominae), vectors of Chagas' disease. PLoS neglected tropical diseases. 2014;8(10),e3266.
15. Justi SA, Russo CA, Mallet JR, Obara MT, Galvao C. Molecular phylogeny of Triatomini (Hemiptera: Reduviidae: Triatominae). Parasit Vectors. 2014;7:149.
16. Pech-May A, Mazariegos-Hidalgo CJ, Izeta-Alberdi A, López-Cancino SA, Tun-Ku E, De la Cruz-Félix K, et al. Genetic variation and phylogeography of the *Triatoma dimidiata* complex evidence a potential center of origin and recent divergence of haplogroups having differential *Trypanosoma cruzi* and DTU infections. PLoS neglected tropical diseases. 2019,13(1)e0007044.
17. Monteiro FA, Barrett TV, Fitzpatrick S, Cordon-Rosales C, Feliciangeli D, Beard CB. Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *robustus*. Molecular ecology. 2003;12(4),997-1006.
18. Gardim S, Almeida CE, Takiya DM, Oliveira J, Araújo RF, Cicarelli RM, da Rosa JA. Multiple mitochondrial genes of some sylvatic Brazilian *Triatoma*: non-monophyly of the *brasiliensis* subcomplex and the need for a generic revision in the Triatomini. Infection, Genetics and Evolution. 2014;23:74-79.
19. Gonçalves TCM, Teves-Neves SC, Santos-Mallet JRD, Carbajal-de-la-Fuente AL, Lopes CM. *Triatoma jatai sp. nov*. in the state of Tocantins, Brazil (Hemiptera: Reduviidae: Triatominae). Memórias do Instituto Oswaldo Cruz. 2013;108(4),429-437.
20. Jurberg J, Cunha V, Cailleaux S, Raigorodschi R, Lima MS, Rocha DDS, Moreira F. *Triatoma pintodiasi sp. nov.* do subcomplexo *rubrovaria* (Hemiptera, Reduviidae, Triatominae). Rev Pan-Amaz Saude. 2013;4(1):43-56
21. Dujardin JP, Panzera P, Schofield CJ. Triatominae as a Model of Morphological Plasticity under Ecological Pressure. Mem Inst Oswaldo Cruz. 1999;Vol. 94,Suppl.I: 223-228.
22. Francoy TM, Silva RAO, Nunes-Silva P, Menezes C, Imperatriz-Fonseca VL. Gender identification of five genera of stingless bees (Apidae, Meliponini) based on wing morphology. Genet Mol Res. 2009;8:207-214.
23. Nouvellet P, Ramirez-Sierra MJ, Dumonteil E, Gourbiere S. Effects of genetic factor and infection status on wing morphology of *Triatoma dimidiata* species complex in the Yucatan peninsula,

Mexico. Infection, Genetics and Evolution. 2011;11:1243-1249.

24. Dujardin JP, Schofield J, Panzera F, Matias A, De La Riva J. Los vectores de la enfermedad de Chagas. Bruxelles : Académie Royale des Sciences d'Outre-Mer. 2002; 25(3),189p.(Mémoire in-8°.Nouvelle Série ; 3).ISBN 90-75652-27-5.
25. Galvão C, Carcavallo R, Da Silva Rocha D, Jurberg J. A check-list of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution with nomenclatural and taxonomic note. Zootaxa. 2003;202:1-36.
26. Zeledón, R. El *Triatoma dimidiata* (La treille, 1811): y su relación con la enfermedad de chagas (No. 595.754 Z49t). San José, CR: EUNED. 1981.
27. Acevedo F, Godoy E, Schofield CJ. Comparison of intervention strategies for control of *Triatoma dimidiata* in Nicaragua. Memorias do Instituto Oswaldo Cruz. 2000;95(6),867-871.
28. Monroy MC, Bustamante DM, Rodas AG, Enriquez ME, Rosales RG. Habitats, dispersion and invasion of sylvatic *Triatoma dimidiata* (Hemiptera: Reduviidae: Triatominae) in Peten, Guatemala. J Med Entomol. 2003;40:800–806.
29. Nakagawa J, Juárez J, Nakatsuji K et al. Geographical characterization of the triatomine infestations in north-central Guatemala. Annals of Tropical Medicine and Parasitology. 2005;99,307–315.
30. Gurgel-Gonçalves R, Komp E, Campbell LP, Khalighifar A, Mellenbruch J, Mendonça VJ, Owens HL, de la Cruz Felix K, Peterson AT, Ramsey JM. Automated identification of insect vectors of Chagas disease in Brazil and Mexico: the Virtual Vector Lab. 2017;5:e3040.
31. Khalighifar A, Komp E, Ramsey JM, Gurgel-Gonçalves R, Peterson AT. Deep Learning Algorithms Improve Automated Identification of Chagas Disease Vectors. Journal of medical entomology. 2019;1-7.
32. Kuhl FP, Giardina CR. Elliptic Fourier Features of a Closed Contour. Computer Graphics and Image Processing. 1982;18:236-258.
33. Kincaid DT, Schneider RB. Quantification of leaf shape with a microcomputer and Fourier transformation. Canadian Journal of Botany. 1983;61,2333-2342.
34. Rohlf FJ, Archie JW. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). Systematic Zoology. 1984; 33:302-317.
35. Furuta N, Ninomiya S, Takahashi N, Ohmori H, Yasuo U. Quantitative evaluation of soybean (*Glycine max* Merr.) leaflet shape by principal component scores based on elliptic Fourier descriptor. Japanese Journal of Breeding. 1995;45(3):315-320.
36. Neto J C, Meyer GE, Jones DD, Samal AK. Plant species identification using Elliptic Fourier leaf shape analysis. Computers and Electronics in Agriculture.2006;50,121-134.
37. Andrade IM, Mayo SJ, Kirkup D, Van den Berg C. Elliptic Fourier analysis of leaf outline shape in forest fragment populations of *Anthurium sinuatum* and *pentaphyllum* (Araceae) from northeast Brazil. Kew Bulletin. 2010;65,1-18.

38. Kim SJ, Kim BW, Kim DP. Tree recognition for landscape using by combination of features of its leaf, flower and bark. Proceedings of SICE Annual Conference (SICE). 2011;1147-1151.
39. Singh K, Gupta I, Gupta S. Classification of Bamboo Species by Fourier and Legendre Moment. International Journal of Advanced Science and Technology. 2013;50,61-70.
40. Zhan QB, Wang XL. Elliptic Fourier analysis of the wing outline shape of five species of antlion (Neuroptera: Myrmeleontidae: Myrmeleontini). Zoological Studies. 2012;51(3):399-405.
41. Iwata H, Ukai Y. SHAPE: A computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors. Journal of Heredity. 2002;93:384-385.
42. Iwata H, Ebana K, Uga Y, Hayashi T. Genomic prediction of biological shape: elliptic Fourier analysis and kernel partial least squares (PLS) regression applied to grain shape prediction in rice (*Oryza sativa*). PloS one. 2015;10(3):e0120610.
43. Freeman H. Computer processing of line-drawing images. Computing. Surveys. 1974;6:57-97.
44. Samarasinghe S. Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. Auerbach publications. 2016.
45. Tatsuta H, Takahashi KH, Sakamaki Y. Geometric morphometrics in entomology: Basics and applications. Entomological Science. 2018;21(2):164-184.
46. Arribas P, Andújar C, Sánchez-Fernández D, Abellán P, Millán A. Integrative taxonomy and conservation of cryptic beetles in the Mediterranean region (Hydrophilidae). Zoologica Scripta. 2013;42(2):182-200.
47. Polášek M, Godunko RJ, Rutschmann S, Svitok M, Novikmec M, Zahrádková S. Integrative taxonomy of genus *Electrogena* (Ephemeroptera: Heptageniidae): the role of innovative morphological analyses for species delimitation. Arthropod Systematics & Phylogeny. 2018;76:449-462.
48. Santillán-Guayasamín S, Villacís AG, Grijalva MJ, Dujardin JP. The modern morphometric approach to identify eggs of Triatominae. Parasites & vectors. 2017;10(1):55.
49. Patterson JS. Comparative morphometrics and molecular genetic analyses of Triatominae (Hemiptera: Reduviidae). Ph.D. thesis, London School of Hygiene and Tropical Medicine, University of London, London.
50. Dos Santos CMD, Jurberg J, Galvão C, Martínez M. Morfometria comparada de *Triatoma infestans*, *rubrovaria* e *T. platensis* (Hemiptera, Reduviidae, Triatominae) do Uruguai. Iheringia, Sér. Zool., Porto Alegre. 2009;99(1):56-60.
51. McLellan T, Endler JA. The relative success of some methods for measuring and describing the shape of complex objects. Systematic Biology. 1998;47:264-281.
52. Chávez B, Cruz DD. Valor taxonómico de la forma del ala en seis especies de esfíngidos (Lepidoptera: Sphingidae). Revista Cubana de Ciencias Biológicas. 2015;4(3).
53. Cortez MBDS, Sforça DA, Alves FDM, Vidal JDD, Alves-Pereira A, Mori GM, et al. Elucidating the *Clusia criuva* species 'complex': cryptic taxa can exhibit great genetic and geographical variation. Botanical Journal of the Linnean Society. 2019;190(1): 67-82.

54. Ryckman RE. Biosystematics and hosts of the *Triatoma protracta* complex in North America. Calif. Publ. Entomol. 1962;27:93–239.
55. Usinger RL, Wygodzinsky P, Ryckman RE. The biosystematics of Triatominae. Annu. Entomol. 1966;11:309–330.
56. Carcavallo RU, Jurberg J, Lent H, Noireau F, Galvão C. Phylogeny of the Triatominae (Hemiptera Reduviidae). Proposals for taxonomic arrangements. Entomol. Vectores. 2000;7:1–99.
57. Schofield CJ, Galvão C. Classification, evolution and species groups within the Triatominae. Acta Trop. 2009;110:88–100.
58. Pita S, Lorite P, Nattero J, Galvão C, Alevi KC, Teves SC, et al. New arrangements on several species subcomplexes of *Triatoma* genus based on the chromosomal position of ribosomal genes (Hemiptera-Triatominae). Infection, Genetics and Evolution. 2016;43:225-231.
59. Dorn PL, Monroy C, Curtis A. *Triatoma dimidiata* (Latreille, 1811): A review of its diversity across its geographic range and the relationship among populations. Infect Genet Evol. 2007;7:343–352.

Figures

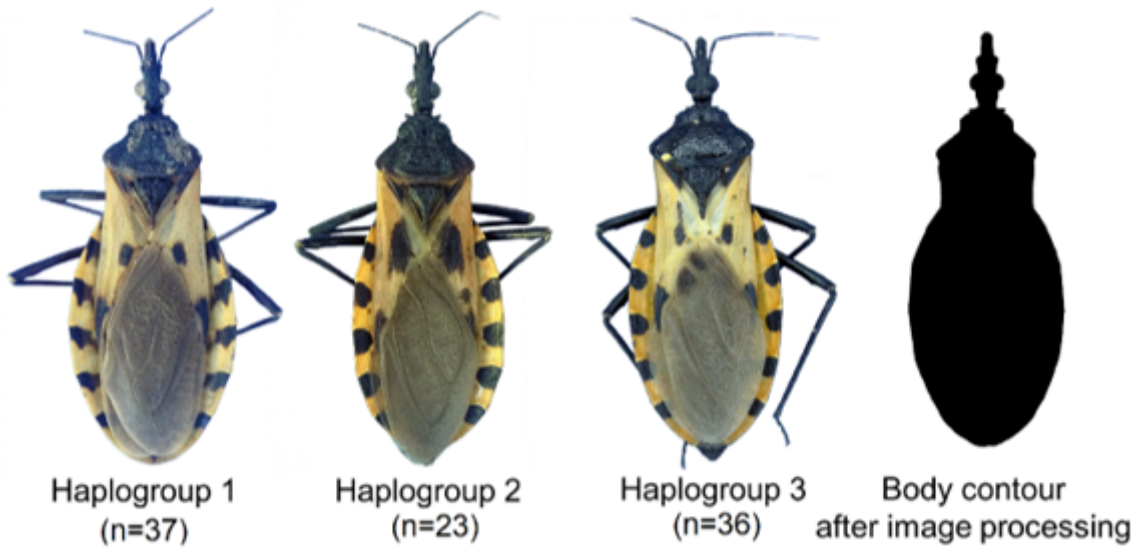

Figure 1

Examples of the three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae) and an example of a body contour after image processing for the analysis of the Elliptical Fourier Descriptors. Sample sizes per haplogroup are shown in parentheses under each image. Images from Gurgel-Goncalves et al. [30].

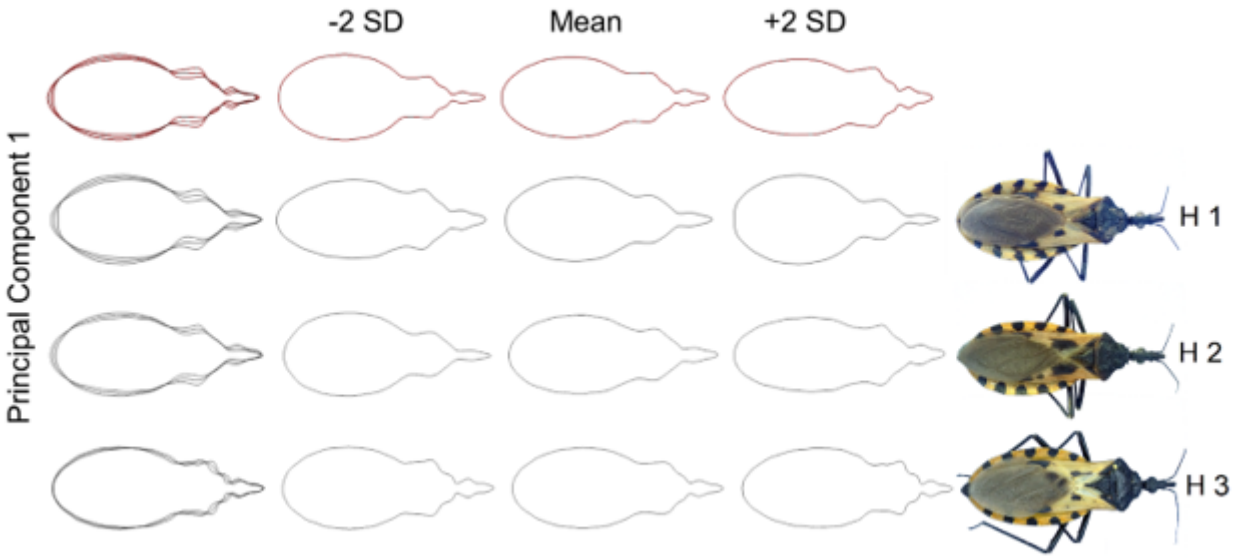

Figure 2

Digital reconstruction and variability of the contours in three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae). The contours were obtained from the first principal component obtained with the Elliptical Fourier Descriptors. Red contours represent the consensus of the three haplogroups. H 1: Haplogroup 1; H 2: Haplogroup 3; H 3: Haplogroup 3.

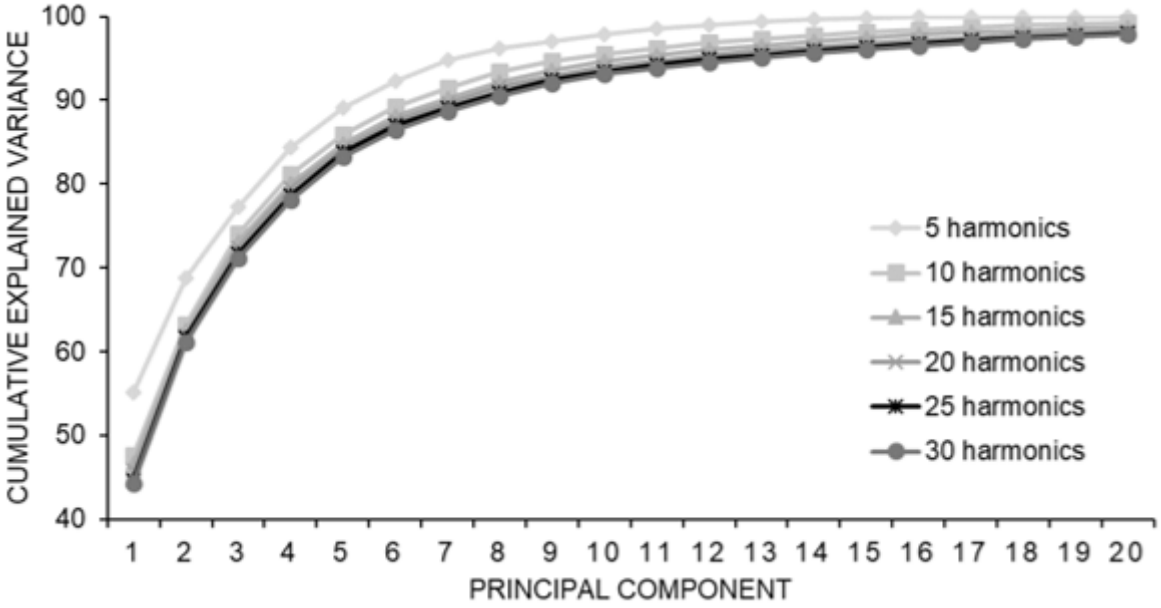

Figure 3

Behavior of the variance explained by the principal components as the number of harmonics used to characterize the shape of three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae) is increased.

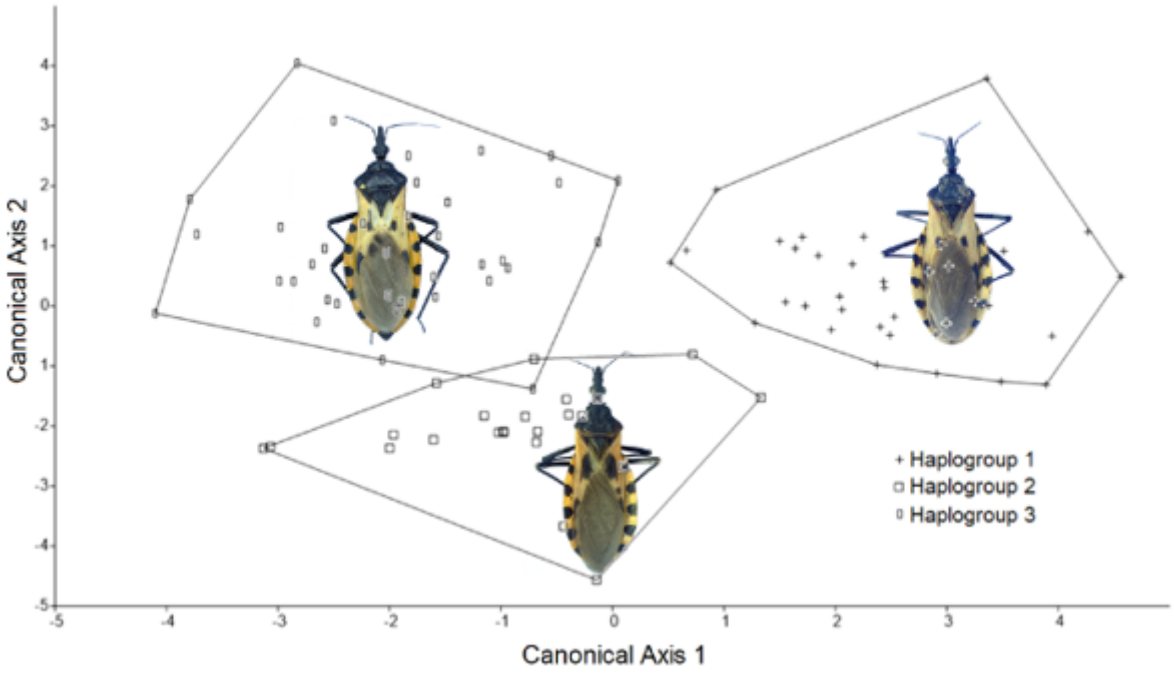

Figure 4

Canonical Variance Analysis using shape variables (30 principal components resulting from the Elliptical Fourier Coefficients of 25 harmonics) of three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae).

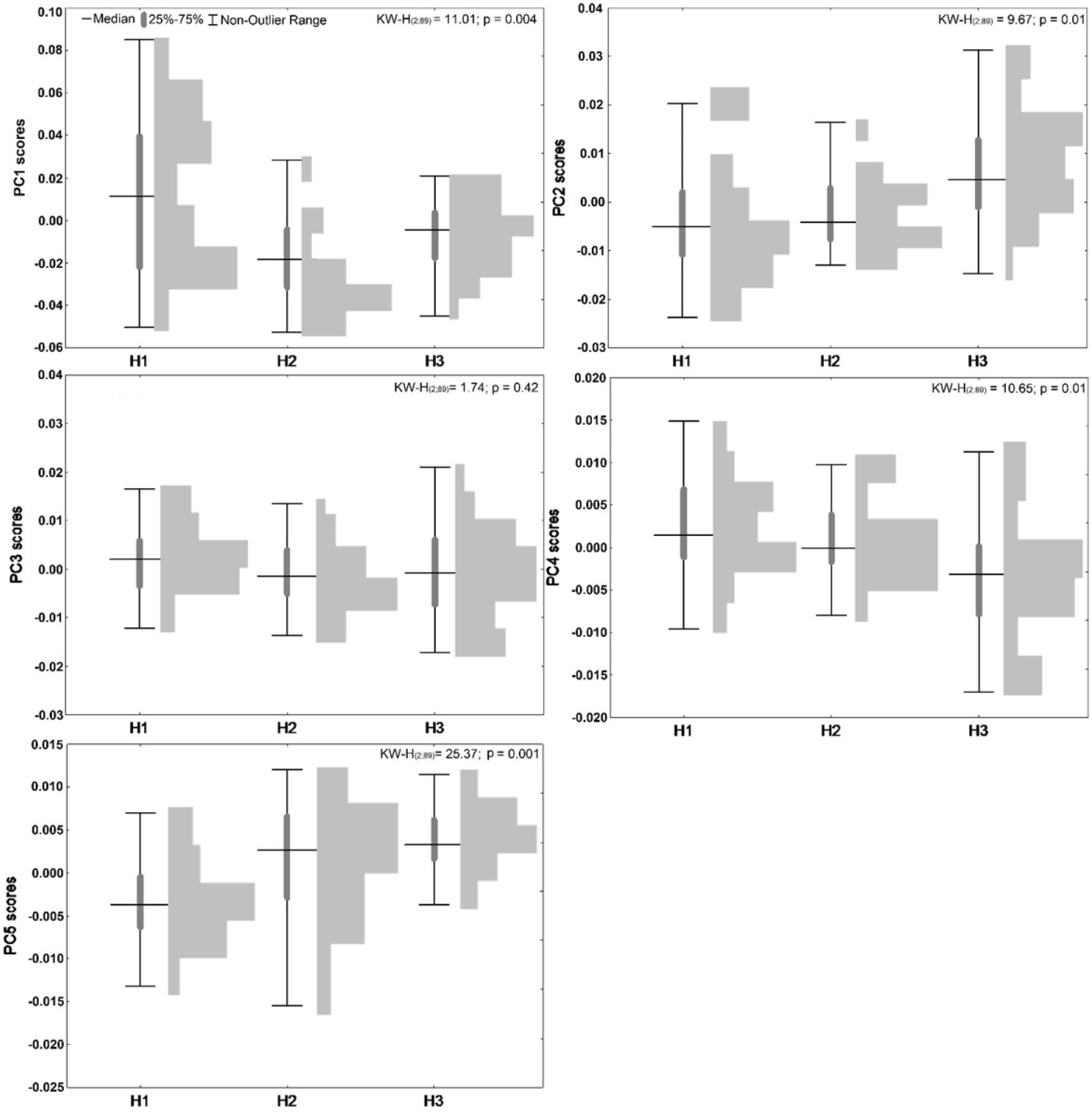

Figure 5

Differences in the scores of the first five principal components between three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae). H1: Haplogroup 1, H2: Haplogroup 2, H3: Haplogroup 3.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Graphicalabstract.png