Quantitative Imagery Analysis of Spot Patterns for Haplogroup Classification of Triatoma Dimidiata (Latreille, 1811) (Hemiptera: Reduviidae) an Important Vector of Chagas Disease

Daryl David Cruz Flores (daryldavidcf@gmail.com)
Centro de Investigación en Biodiversidad y Conservación https://orcid.org/0000-0002-7714-2459
Dennis Denis Ávila
Universidad de la Habana Facultad de Biologia
Elizabeth Arellano
Centro de Investigacion en Biodiversidad y Conservacion
Carlos N. Ibarra-Cerdeña
CINVESTAV IPN: Centro de Investigacion y de Estudios Avanzados del Instituto Politecnico Nacional

Research

Keywords: cryptic species, coloring pattern, species complex, taxonomy, triatomine

Posted Date: September 18th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-78108/v1

Version of Record: A version of this preprint was published on January 29th, 2021. See the published version at https://doi.org/10.1186/s13071-021-04598-5.

Abstract

Background: Spots and coloring patterns evaluated quantitatively can be used to discrimination and identify possible cryptic species. *Triatoma dimidiata* (Reduviidae: Triatominae) is one of the main complexes of Chagas disease vector species. Phylogenetic studies have defined three haplogroups for Mexico and part of Central America. In this work, we evaluate the possibility of correctly discriminating among these *T. dimidiata* haplogroups using the pattern of dorsal spots.

Methods: Digital images of the dorsal region of individuals from the three haplogroups were used. We used image processing to extract primary and secondary variables characterizing the dorsal spot pattern. The statistical analysis of the variables included descriptive statistics, non-parametric Kruskal-Wallis tests, Discriminant Function Analysis (DFA), and a neural classification network.

Results: A distinctive spot pattern was found for each haplogroup. The most differentiated pattern was presented by haplogroup 2, mainly apparent in the notably larger central spots. Haplogroups 1 and 3 were more similar to each other, but there were consistent differences in the shape and orientation of the spots. Significant differences were found among haplogroups in almost all the variables analyzed, these differences being greater in relative spot area, mean relative area of central spots, central spots Feret diameters, lateral spots Feret diameters and aspect ratio. Both the DFA and the neural network had correct discrimination values above 90%.

Conclusions: We conclude that the spot pattern can be reliably used to discriminate among haplogroups, and possibly among triatomine species.

Background

Genetic and morphological divergences associated with speciation processes may not appear at the same time or progress at the same rate [1]. The emergence of new species usually results from the isolation of populations due to geographic, ecological, or behavioral barriers, which act individually or synergistically [2]. This can lead to populations that have substantial genetic differentiation that has not been expressed phenotypically (at least not obviously), giving rise to cryptic species [3]. Identifying cryptic species complexes is one of the most important challenges facing taxonomy in recent years [4].

The correct delimitation of cryptic species has important implications for research in many fields of biology, such as studies on biodiversity, conservation, and behavioral ecology [4]. Frequently, these cryptic species delimitations are achieved using different types of data such as molecular, ecological, behavioral, and geometric morphometric data [5]. This combination of methods, known as integrative taxonomy [6], is the surest and most precise way of determining species limits [7, 8].

Cryptic species in the genus *Triatoma* (main vectors of Chagas disease) have mainly been recognized using molecular tools [9, 10, 11, 12], although both ecological and morphometric analyses have also been used [13]. Within Triatoma, the *dimidiata* complex has received considerable attention, in part because it

is one of the most widely distributed triatomine species complexes. It is the only triatomine bug that naturally occurs throughout the northern neotropical realm of North, Central, and South America [14]. Within the *dimidiata* complex, at least five new species have been proposed based on genetic data [15]. Furthermore, the species in this complex have different morphological patterns [16, 17]. In the field of epidemiological entomology, the delimitation of species of medical importance is vital for the establishment of efficient control strategies [13, 18]. Geometric morphometric techniques using landmarks [19, 20, 21, 22] or body contour descriptors [23, 24, 25, 26] have been used for this purpose, mainly because of its superiority over traditional morphometric methods [27] and because it is a cheaper method than, for example, molecular ones.

Spot patterns are widely used to describe species in traditional taxonomy [28]. However, because spot pattern is highly variable due to its ecological functions, it is usually described in subjective, qualitative terms. Alternatively, using digital tools to quantify spot patterns can minimize bias, increase precision, and allow automated identification processes [29]. However, since few studies use quantitative measurements of pattern properties for taxonomic purposes, evidence on the usefulness of color patterns to separate (or discriminate) species is still lacking [30].

The general body color of triatomines is black or spruce, with pattern elements ranging from light yellow to light brown, orange, or red shades [31]. The lighter pattern elements can be present on any area of the body or appendages, and the color, intensity, and distribution of these elements are of considerable importance for systematic purposes. The pattern of the connexivum region is particularly marked [32]. However, despite their taxonomic importance, there have been very few quantitative studies of color and pattern variation in *Triatoma*. The first study to quantify color patterns in a species of *Triatoma* was by [33], who analyzed the melanic and non-melanic forms of domestic and peridomestic populations of *Triatoma infestans*, and [34], in addition to other aspects, explored pattern variation as a function of elevation in triatomines from El Salvador, including populations of *T. dimidiata*. To date, the utility of the spot pattern to discriminate among haplogroups within the *dimidiata* complex or any other triatomine complex has not been explored.

Although there are no obvious external morphological differences based on our observations of *dimidiata* complex specimens, we hypothesize that the evaluation of more detailed quantitative differences in the spot pattern between haplogroups could be used to distinguish them morphologically. This method could potentially be useful to improve the separation criteria for cryptic species in this group without requiring genetic data. In this work, we evaluate the reliability of discriminating among three *T. dimidiata* haplogroups using the dorsal spot pattern.

If successful, this technique could be extended to other species in this (and other) group and lay the foundations for an automated identification system to facilitate correct species recognition within the genus *Triatoma*.

Methods

Sample information

Images of individuals from each of the haplogroups of *Triatoma dimidiata* were obtained from Gurgel-Gonçalves et al. [20]. These images are part of a collection of images of 51 triatomine species from Mexico and Brazil available for public use in the Dryad repository (http://dx.doi.org/10.5061/dryad.br14k). The original series of images that represent the species distributed in Mexico was taken from the following entomological collections in Mexico: Regional Center for Health Research, National Institute of Public Health of Mexico, Guanajuato State Public Health Laboratory, Benito Juárez Autonomous University of Oaxaca, and the Autonomous University of Nuevo León, Monterrey, and details of how these images were taken are described in the referenced publication. We obtained a total of 44, 30, and 40 images of individuals belonging to haplogroups 1, 2, and 3 respectively; the haplogroup assignment of these individuals was corroborated genetically, and this corroboration constitute a major factor to use this images in quantitative analysis like our work. From the 114 images, we selected only high-quality images that clearly captured the spot pattern, eliminating the cases where the spots were fused or covered by hyperchromatic wings (Additional file 1). This resulted in a final sample of 101 images (39 from H1, 23 from H2, and 39 from H3).

Image processing

The images were processed to facilitate the extraction of standardized measurements of the spot pattern (Fig. 1). The abdomens were clipped manually, removing the legs and cutting off the head at the thorax level. Subsequently, the images were aligned and re-scaled, using the insertion angles of the abdomen and thorax and the back of the body as references for alignment and scaling all individuals to the width of the first individual that was taken as a reference (image H10355). These transformations may slightly alter the shape and absolute values of the spot measurements, but they are essential to standardize the spatial patterns of the spots and make them comparable, eliminating differences due to body shape or size, whose identifying value has been tested in previous works [20, 26]. For this reason, the quantitative estimates of areas were always expressed relative to the total area of the abdomen and linear measurements are relative to the square root of the total abdomen area.

Processing for spot pattern extraction included removing color information (desaturation) and reduction of levels to the central 50% of the image histogram. In some cases, noise produced by surface reflectance of the specimens or shadows that artificially connected adjacent spots during the binarization of the images were manually eliminated.

In the ImageJ program [35] a macro (Additional file 2) was programmed to automate image processing and measurements. This included 8-bit image conversion, binarization with a minimum automatic threshold, background removal, mask conversion, and gap filling. The outlier points, both black and white (using radius 6 and threshold 50) were then removed and the resulting particles (spots) were measured.

Khalighifar A, Mellenbruch J, Mendonça VJ, et al. Automated identification of insect vectors of Chagas disease in Brazil and Mexico: the virtual vector lab. PeerJ. 2017;5:e3040 [20]

Heat maps were obtained by superimposing the images of the spot patterns of all individuals per haplogroup, using the PAT-GEOM v1.0.0 package, developed by Chan et al. [36]. This package allows the analysis of different measures of the coloration pattern quantitatively, and it was designed to work with macros on ImageJ. These maps allowed us to visually explore and qualitatively describe the general patterns that characterized each haplogroup.

Quantitative characterization of the spot pattern

The spots were numbered consecutively for identification; spots 1 and 2 were the central spots, and the spots on the edge of the abdomen were numbered with consecutive odd numbers on the left and even numbers on the right. To quantitatively describe the pattern of spots, a series of primary variables were taken at the spot level, as well as derived variables that included both the spot and individual levels.

The variables measured are shown in Fig. 2. The total body area (Ta) was used for standardization purposes only. The relative area (Ra) was the area of each spot relativized as a percentage of Ta (%). The sum of the Euclidean distances (SED) was calculated by taking the centroid coordinate of each spot and calculating, at the individual level, the distance between the central and lateral spots after making a Procrustes record of the complete configurations. The maximum and minimum Feret diameters (MaxFd and MinFd, respectively), as well as the Feret angle (Fa), were calculated for each spot. These variables refer to the maximum and minimum distances between any pair of contour points of a shape, and although they are identified as diameter, they are not strictly analogous to a diameter, since they do not pass through the center of the figure or divide it into symmetrical sections. The Fa refers to the angle of the vector of the MaxFd and indicates the general directionality of the spot (its inclination). The aspect ratio (Ar) of each spot (ratio of the minor to the major diameter) was used as an indicator of its shape.

For each individual, the averages of the variables per spot, the sum of the total Ra of the spots, and the ratio of the mean Ra of the central spots to the lateral spots were calculated as derived variables. For the calculation of the average inclination angle, both for the central and lateral spots, the angles of the spots from the left to the right quadrant (0–90°) were reflected.

Data analysis

Non-parametric descriptive statistics (median, quartiles, and range) were used because the distribution of the data was not normal, and traditional descriptors gave a false impression of precision and marked differences. Statistical comparisons among haplogroups were done using Kruskal-Wallis tests in Statistica v8 software. Also, a Linear Discriminant Function Analysis (forward stepwise) (LDFA) was performed to estimate the ability to discriminate haplogroups based on the variables used. Since this method has a series of restrictive premises and can only linearly differentiate the groups, a multilayer perceptron type neural classification network was used as an alternative method. Neural networks are

supervised machine learning procedures and do not have statistical premises on the nature of the data, making them more powerful and capable of exploring nonlinear relationships in complex sets of variables. The most efficient topology for the network was found by the automated search procedure of the Statistica 8.0 software, considering all of the variables analyzed. The network was trained with 60% of the individuals by haplogroup and validated with the remaining 40%. Assignment to each group was random, except for individuals wrongly classified by the LDFA, who were forced into the validation sample for a more robust check of network performance. The weight assigned by the neural network to each variable was estimated to identify those of greatest importance in the discrimination process.

Results

The heat maps generated by superimposing all the individuals within each haplogroup show the spot patterns that characterize each haplogroup and evidence a well-differentiated pattern between them (Fig. 3). The most differentiated pattern was presented by haplogroup 2, mainly apparent in the notably larger central spots. Haplogroups 1 and 3 were more similar to each other, but there were consistent differences in the shape and orientation of the spots.

The ratio of spotted area to Ta differed among haplogroups. The highest Ra was presented by haplogroup 2 with 15.6%, while haplogroup 3 had only 8.7% (Fig. 4A). When comparing the ratio of the area of the central spots to the lateral spots, haplogroups 1 and 3 had higher relative lateral spot areas. In haplogroup 2, the lateral and central spots contributed almost equally to the total spot area, while the percentage of the central spots area was slightly higher. When statistically comparing the mean Ra of the central and lateral spots, only the central spot area differed significantly among haplogroups (Fig. 4B and C).

The average spot size, characterized by Feret diameters, was significantly different among haplogroups, both for the central and lateral spots (Fig. 5). In the case of the central spots, haplogroup 2 was the most strongly differentiated (Fig. 5A), while for the lateral spots, haplogroup 1 presented the most notable differences (Fig. 5B).

The orientation of the abdominal spots, expressed by the Fa, were markedly different between haplogroup 1 and the other groups. The largest differences were observed in the orientations of the first three pairs of spots (3/4, 5/6, and 7/8), which tended to be more forward oriented. For the remaining spots, although differences in orientation were observed, these were less noticeable, both in the Fa value and in its variation among individuals (Fig. 6).

When comparing the mean orientations of the lateral spots of the abdomen (Fa), significant differences were found between the three haplogroups. Haplogroup 1 was the most distinct and had less variation in Fa than the remaining haplogroups (Fig. 7).

The shapes of the central and lateral spots (Ar), differed among haplogroups, both for the lateral and central spots (Fig. 8). The shape of the central spots in haplogroup 2 showed the greatest differences

among the haplogroups, while the most differentiated lateral spots were from haplogroup 3.

The LDFA correctly assigned 93.8% of the individuals into the correct haplogroups based on spot pattern (Fig. 9). One individual from H2 (H2 0504) was erroneously assigned to H3, and one H3 individual (H3 0847) was misassigned as H2. Haplogroups 1 and 3 had more overlap in the ranking space, with three H1 individuals (H1 0367; H1 0372; H1 0374) assigned as H3 and two H3 individuals (H3 0388 and H3 0395) assigned as H1.

The most efficient neural classification network had a topology with 20 neurons in the hidden layer. This achieved an overall performance of 94.7% with a BFGS-12 training algorithm and an entropy error function. 100% of the training data were correctly identified, and considering only the validation data, the correct identification was 87.2% of the individuals. The hidden layer had sine activation functions and the output layer logistic functions, with Sum of Squares as an error function. This network achieved 100% correct classification of H2 specimens and misclassified three H1 individuals (H1 0367, H1 0372 and H1 0374; from a total of 16 in the validation sample) as H3, and two H3 individuals (H3 0847 and H3 0862) as H2. The remaining three individuals that had been incorrectly classified by LDFA were correctly assigned to their haplogroups by the neural network (H2 0504; H3 0388 and H3 0395).

Classification methods made similar use of variables. The LDFA used only five variables in the final model: the size of the central spots, the shape, angle, and diameter of the lateral spots and the total relative spot area. The neural network assigned greater importance to these same variables, and additionally included the relative area of the lateral spots.

When analyzing the weights assigned to each variable used in the neural network procedure (Fig. 10), the most important in the classification process was the Feret diameter of the lateral spots and the aspect ratio of the lateral spots, respectively. The variable that contributed the least to the classification was the ratio of the central spot area to lateral spot area.

Discussion

The cryptic *dimidiata* species complex has been largely supported using molecular tools, which has led to the identification of three phylogenetically well-differentiated haplogroups in Mexico and part of Central America [9, 11, 12], and two taxa have been formally described as new species [37, 38]. For the first time, we used the spot pattern presented by this complex to discriminate among haplogroups (possible cryptic species) by extracting and analyzing quantifiable variables from digital images.

Our results demonstrate the ability to use these measures to correctly recognize the haplogroups analyzed. Of the variables used for discrimination, only one (mean relative area of the lateral spots) did not differ significantly among haplogroups, indicating that overall, pattern variables were useful for delimitation. This was verified both by the discriminant analysis ordination plot and the results obtained by the most efficient neural network.

The study of coloration in triatomines and its application in taxonomy has mainly been used in traditional qualitative approaches [32]. This has led to the assumption of a lack of clear morphological diagnostic characters to facilitate recognition and formal descriptions at the species level [15]. However, using heat maps, three well-differentiated spot patterns were evident, corresponding with the three haplogroups. This, once again, highlights the importance of using quantitative tools to study complex patterns such as coloration, where subtle aspects such as the orientation of groups of spots or other patterns may not be apparent or easily distinguishable to a human observer.

The variation found among the haplogroups in spot pattern may respond to different processes. In other groups of insects such as butterflies, coloration patterns have been shown to vary depending on environmental conditions such as temperature [39, 40], associated with processes of genetic assimilation of phenotypic changes [41]. Although there are populations where the three *T. dimidiata* haplogroups analyzed in this study are found sympatrically [see 11], their distributions are mostly allopatric, therefore the pattern of variation among these haplogroups may reflect adaptation to environments with different characteristics in response to environmental stress. Recently, genetic assimilation in the evolution of phenotypic plasticity has been demonstrated not only for butterflies but also for various groups of organisms [42, 43, 44, 45]. However, corroborating this phenomenon in *T. dimidiata* will require specifically designed studies.

Another important aspect that this research demonstrates is the value of the combination of digital image analysis and machine learning for taxonomy [46]. The potential of this combination of approaches in species delimitation has been broadly demonstrated [47]. However, even though its utility is clear and, in many cases, superior to the traditional taxonomy, it is still relatively rarely used.

Classical taxonomy is a science that is practically in danger of extinction, especially due to the lack of expert taxonomists and specialists in species identification, which require many years of training and experience [48]. In the era of big data, image pattern recognition is a new technology that provides many potential advantages for taxonomists, including speeding up and automating the classification process, reducing error, and assimilating quantitative information that would be impossible for a human observer [49].

Specifically, with triatomines, there have been recent efforts to employ these methods to establish automated identification systems. These include the studies of Gurgel-Goncalves et al. [20] and Khalighifar et al. [50], who, using geometric morphometry techniques and deep learning algorithms, respectively, have taken the first steps in this regard. In another example, Cruz et al. [26] were able to discriminate *T. dimidiata* haplogroups with high correct discrimination values by characterizing the entire body contour using Fourier Elliptical Descriptors, and this same method has been successfully used to generate automated identification systems in other groups of insects [51]. The integration of this method with the analysis of the spot pattern is potentially a novel and powerful tool to generate a computer-based approach for species identification in cryptic groups. Although classification processes still need improvement, these novel works bring new challenges and perspectives in the field of epidemiological

entomology, and the integration of methods should be a central aspect in the future of automated identification in this group, given its epidemiological importance, as well as in other groups of insects.

Although this research was focused on evaluating the possibility of correctly discriminating three haplogroups of *T. dimidiata* using the dorsal spot pattern, the value of coloration patterns in species biology cannot be forgotten. Color in insects has important biological functions including mate choice, intra-sexual competition, dominance relationships, and other social interactions [52]. Therefore, its study is relevant in many contexts beyond taxonomy, and research should be increased to explore the role that coloration patterns play in nature. In relation to other groups of insects such as Coleoptera, Lepidoptera and Hymenoptera, in Hemiptera and especially the Triatominae subfamily, there are very few studies associated with the coloring patterns. [52].

Conclusions

The importance of the correct recognition of insect species of epidemiological importance is vital for the establishment of good control measures [13, 18]. The results obtained in this investigation allow us to conclude that the spot pattern in triatomines constitutes a significant source of information that can be used directly in the taxonomic analysis of this group of insects. If we consider that the haplogroups used here may constitute phylogenetically close cryptic species [9, 11, 53], similar pattern analysis in a larger number of less closely related species will likely find larger, more easily distinguishable differences in the spotting pattern than those found here. A fruitful avenue for future research would be comparing spot patterns among multiple species in order to discriminate among them. If similarly, successful to this study, pattern recognition could allow, in the not too distant future, the development of a reliable automated identification system to use as a tool for the recognition of vectors of Chagas disease, one of the most important on the American continent.

Abbreviations

Ta: Total area; Csp: Central spots; Lsp: Lateral spots; A: area; c: centroid; Ra: Relative area; Ar: Aspect ratio; MaxFd: Maximun Feret diameter; MinFd: Minimun ferret diameter; Fa: ferret angle; SED: Sum of the Euclidean distances; H1: Haplogroup 1; H2: Haplogroup 2; H3: Haplogroup 3; LDFA: Linear Discriminant Function Analysis

Declarations

Acknowledgments

We would like to extend our thanks to Félix N. Estrada, Alejandro López Michelena, and Daily Martínez who offered useful comments to improve the manuscript.

Authors’ contributions

DDC, DDA conceived the study. DDC, 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

Financial support for this study was provided by a CONACYT scholarship program 2018-000012-01NACF-11846 to the first author. We acknowledge financial support by the Secretaría de Educación Pública-Cinvestav (Project: FIDSC2018/160) to CNIC.

Availability of data and materials

Data supporting the conclusions of this article are included in the article. Also, all data derived from this investigation are deposited in the Figshare repository (https://doi.org/10.6084/m9.figshare.12909968.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. de Jesús-Bonilla VS, García-París M, Ibarra-Cerdeña CN, Zaldívar-Riverón A. Geographic patterns of phenotypic diversity in incipient species of North American blister beetles (Coleoptera: Meloidae) are not determined by species niches, but driven by demography along the speciation process. Invertebr Syst. 2018;32(3):672–88.
2. Yoder AD, Yang Z. Estimation of primate speciation dates using local molecular clocks. Mol Biol Evol. 2000;17(7):1081–90.
3. Struck TH, Feder JL, Bendiksby M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson K, Liow LH, Nowak MD, Stedje B, Bachmann L, Dimitrov D. Finding evolutionary processes hidden in cryptic species. Trends Ecol Evol. 2018;33(3):153–63.

4. Bickford D, Lohman DJ, Sodhi NS, Ng PK, Meier R, Winker K, Ingram K, Das I. Cryptic species as a window on diversity and conservation. Trends Ecol Evol. 2007;22(3):148–55. https://doi.org/10.1016/j.tree.2006.11.004.
5. Dayrat B. Towards integrative taxonomy. Biol J Linn Soc. 2005;85:407–15. https://doi.org/10.1111/j.1095-8312.2005.00503.x.
6. Jörger KM, Schrödl M. How to describe a cryptic species? Practical challenges of molecular taxonomy. Front Zoo. 2013;10(1):59.
7. Brower AV. Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). Syst Biodivers. 2006;4(2):127–32. https://doi.org/10.1017/S147720000500191X.
8. Smith MA, Rodriguez JJ, Whitfield JB, Deans AR, Janzen DH, Hallwachs W, Hebert PD. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc Natl Acad Sci. 2008;105(34):12359–64. https://doi.org/10.1073/pnas.0805319105.
9. Bargues MD, Klisiowicz DR, Gonzalez-Candelas F, Ramsey JM, Monroy C, Ponce C, et al. 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.
10. Justi SA, Russo CA, Mallet JR, Obara MT, Galvao C. Molecular phylogeny of Triatomini (Hemiptera: Reduviidae: Triatominae). Parasit Vectors. 2014;7:149.
11. Pech-May A, Mazariegos-Hidalgo CJ, Izeta-Alberdi A, López-Cancino SA, Tun-Ku E, De la Cruz 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 Negl Trop Dis. 2019;13(1):e0007044.
12. Aguilera-Uribe M, Meza-Lázaro RN, Kieran TJ, Ibarra-Cerdeña CN, Zaldívar-Riverón A. Phylogeny of the North-Central American clade of blood-sucking reduviid bugs of the tribe Triatomini (Hemiptera: Triatominae) based on the mitochondrial genome. Infect Genet Evol. 2020;202084:104373.
13. Gurgel-Gonçalves R, Ferreira JBC, Rosa AF, Bar ME, Galvao C. Geometric morphometrics and ecological niche modelling for delimitation of nearsibling triatomine species. Med Vet Entomol. 2011;25:84–93.
14. Monteiro FA, Peretolchina T, Lazoski C, Harris K, Dotson EM, Abad-Franch F, et al. Phylogeographic pattern and extensive mitochondrial DNA divergence disclose a species complex within the Chagas disease vector *Triatoma dimidiata*. PloS one. 2013;8(8):e70974.
15. Bargues MD, Schofield C, Dujardin JP. The phylogeny and classification of the triatominae. In: Telleria J, Tibayrenc M, editors. American trypanosomiasis: Chagas disease, one hundred years of research. Amsterdam: Elsevier; 2017.
16. Bustamante DM, Monroy C, Menes M, Rodas A, Salazar-Schettino PM, Rojas G, et al. Metric variation among geographic populations of the Chagas vector *Triatoma dimidiata* (Hemiptera: Reduviidae: Triatominae) and related species. J Med Entomol. 2004;41:296–301.

17. Panzera F, Ferrandis I, Ramsey J, Ordonez R, Salazar-Schettino PM, Cabrera M, et al. Chromosomal variation and genome size support existence of cryptic species of *Triatoma dimidiata* with different epidemiological importance as Chagas disease vectors. Trop Med Int Health. 2006;11:1092–103.
18. Abad-Franch F, Monteiro FA. Molecular research and the control of Chagas disease vectors. Anais Acad Bras Ciências. 2005;77:437–54. Dujardin JP, Beard CB, Ryckman R. The relevance of wing geometry in entomological surveillance of Triatominae, vectors of Chagas disease. Infect Genet Evol. 2007;7:161–7.
19. Gurgel-Gonçalves R, Komp E, Campbell LP, Khalighifar A, Mellenbruch J, Mendonça VJ, et al. Automated identification of insect vectors of Chagas disease in Brazil and Mexico: the virtual vector lab. PeerJ. 2017;5:e3040.
20. Oliveira J, Marcet PL, Takiya DM, Mendonça VJ, Belintani T, Bargues MD, et al. Combined phylogenetic and morphometric information to delimit and unify the *Triatoma brasiliensis* species complex and the *Brasiliensis* subcomplex. Acta Trop. 2017;170:140–8.
21. Nattero J, Piccinali RV, Lopes CM, Hernández ML, Abrahan L, Lobbia PA, Rodríguez CS, de la Fuente ALC. Morphometric variability among the species of the Sordida subcomplex (Hemiptera: Reduviidae: Triatominae): evidence for differentiation across the distribution range of Triatoma sordida. Parasit Vectors. 2017;10:412.
22. Dujardin JP, Kaba D, Solano P, Dupraz M, McCoy KD. Jaramillo-O N. Outline-based morphometrics, an overlooked method in arthropod studies? Infect Genet Evol. 2014;28:704–14.
23. Santillán-Guayasamín S, Villacís AG, Grijalva MJ, Dujardin JP. The modern morphometric approach to identify eggs of Triatominae. Parasit Vectors. 2017;10:55.
24. Santillán-Guayasamín S, Villacís AG, Grijalva MJ, Dujardin JP. Triatominae: does the shape change of non-viable eggs compromise species recognition. Parasit Vectors. 2018;11:543.
25. Cruz DD, Arellano E, Ávila DD, Ibarra-Cerdeña CN. Identifying Chagas disease vectors using elliptic Fourier descriptors of body contour: a case for the cryptic *dimidiata* complex. Parasit Vectors. 2020;13(1):1–12.
26. Tatsuta H, Takahashi KH, Sakamaki Y. Geometric morphometrics in entomology: basics and applications. Entomol Sci. 2018;21:164–84.
27. Padula V, Bahia J, Stöger I, Camacho-García Y, Malaquias MAE, Cervera JL, Schrödl M. A test of color-based taxonomy in nudibranchs: Molecular phylogeny and species delimitation of the *Felimida clenchi* (Mollusca: Chromodorididae) species complex. Mol Phyl Evol. 2016;103:215–29.
28. Yu X, Wang J, Kays R, Jansen PA, Wang T, Huang T. Automated identification of animal species in camera trap images. EURASIP J Image Video Process. 2013;(1):52.
29. Chan IZ, Chang JJM, Huang D, Todd PA. Colour pattern measurements successfully differentiate two cryptic Onchidiidae Rafinesque, 1815 species. Mar Biodivers. 2019;49(4):1743–50. https://doi.org/10.1007/s12526-019-00940-4.
30. Kieran TJ, Gordon ERL, Zaldívar-Riverón A, Ibarra-Cerdeña CN, Glenn TC, Weirauch C. Ultraconserved elements reconstruct the evolution of the Chagas disease-vectoring kissing bugs (Hemiptera:

Reduviidae: Triatominae). Syst. Entomol. 2020.Under Review.

31. 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.
32. Nattero J, de la Fuente ALC, Piccinali RV, Cardozo M, Rodríguez CS, Crocco LB. Characterization of melanic and non-melanic forms in domestic and peridomestic populations of *Triatoma infestans* (Hemiptera: Reduviidae). Parasit Vectors. 2020;13(1):47.
33. Carmona-Galindo VD, Recinos MFM, Hidalgo SAG, Paredes GR, Vaquerano EEP, Magaña ALR, Ayala AKC. Morphological variability and ecological characterization of the Chagas disease vector *Triatoma dimidiata* (Hemiptera: Reduviidae) in El Salvador. Acta Trop. 2020;205:105392. https://doi.org/10.1016/j.actatropica.2020.105392.
34. Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–5. https://doi.org/10.1038/nmeth.2089.
35. Chan IZW, Stevens M, Todd PA. PAT-GEOM: a software package for the analysis of animal patterns. Methods Ecol Evol. 2018;0:1–10. https://doi.org/10.1111/2041-210X.13131.
36. Dorn PL, et al. Description of *Triatoma mopan* sp. n. from a cave in Belize (Hemiptera, Reduviidae, Triatominae). ZooKeys. 2018;775:69–95.
37. Justi SA, Cahan S, Stevens L, Monroy C, Lima-Cordón R, Dorn PL. Vectors of diversity: Genome wide diversity across the geographic range of the Chagas disease vector *Triatoma dimidiata* sensu lato (Hemiptera: Reduviidae). Mol Phyl Evol. 2018;120:144–50.
38. Otaki JM, Yamamoto H. Color-pattern modifications and speciation in lycaenid butterflies. Trans Lepidopterol Soc Jpn. 2003;54:197–205.
39. Otaki JM, Yamamoto H. Species-specific color-pattern modifications on butterfly wings. Dev Growth Differ. 2004;46:1–14.
40. Otaki JM, Hiyama A, Iwata M, Kudo S. Phenotypic plasticity in the range-margin population of the lycaenid butterfly *Zizeeria maha*. BMC EvolBiol. 2010;10:252. doi:10.1186/1471-2148-10-252.
41. Aubret F, Shine R. Genetic assimilation and the post colonization erosion of phenotypic plasticity in island tiger snakes. CurrBiol. 2009;19:1932–6.
42. Buckley J, Bridle JR, Pomiankowski A. Novel variation associated with species range expansion. BMC EvolBiol. 2010;10:382. doi:10.1186/1471-2148-10-382.
43. Scoville AG, Pfrender ME. Phenotypic plasticity facilitates recurrent rapid adaptation to introduced predators. Proc. Natl. Acad. Sci. 2010;107:4260–4263.
44. Muschick M, Barluenga M, Salzburger W, Meyer A. Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC EvolBiol. 2011;11:116. doi:10.1186/1471-2148-11-116.
45. Lukhtanov VA. Species delimitation and analysis of cryptic species diversity in the XXI century. Entomological Review. 2019;99(4):463–72. DOI:10.1134/S0013873819040055.

46. Wäldchen J, Mäder P. Machine Learning for Image Based Species Identification. Methods Ecol Evol. 2018;9:2216–25. doi:10.1111/2041-210X.13075.
47. Martineau M, Conte D, Raveaux R, Arnault I, Munier D, Venturini G. A survey on image-based insect classification. Pattern Recognit. 2017;65:273–84.
48. Zerdoumi S, Sabri AQM, Kamsin A, Hashem IAT, Gani A, Hakak S, Al-garadi M, Chang V. Image pattern recognition in big data: taxonomy and open challenges: survey. Multimed. 2018;77(8):10091–121. DOI 10.1007/s11042-017-5045-7.
49. Khalighifar A, Komp E, Ramsey JM, Gurgel-Gonçalves R, Peterson AT. Deep learning algorithms improve automated identification of Chagas disease vectors. J Med Entomol. 2019;56:1404–10.
50. Yang HP, Ma CS, Wen H, Zhan QB, Wang XL. A tool for developing an automatic insect identification system based on wing outlines. Sci Rep. 2015;5:12786.
51. Mora R, Hanson PE. Widespread occurrence of Black-Orange-Black color pattern in Hymenoptera. J Insect Sci. 2019;19(2):13. doi:10.1093/jisesa/iez021.
52. Gómez-Palacio A, Arboleda S, Dumonteil E, Peterson AT. Ecological niche and geographic distribution of the Chagas disease vector, *Triatoma dimidiata* (Reduviidae: Triatominae): Evidence for niche differentiation among cryptic species. Infect Genet Evol. 2015;36:15–22.

Figures

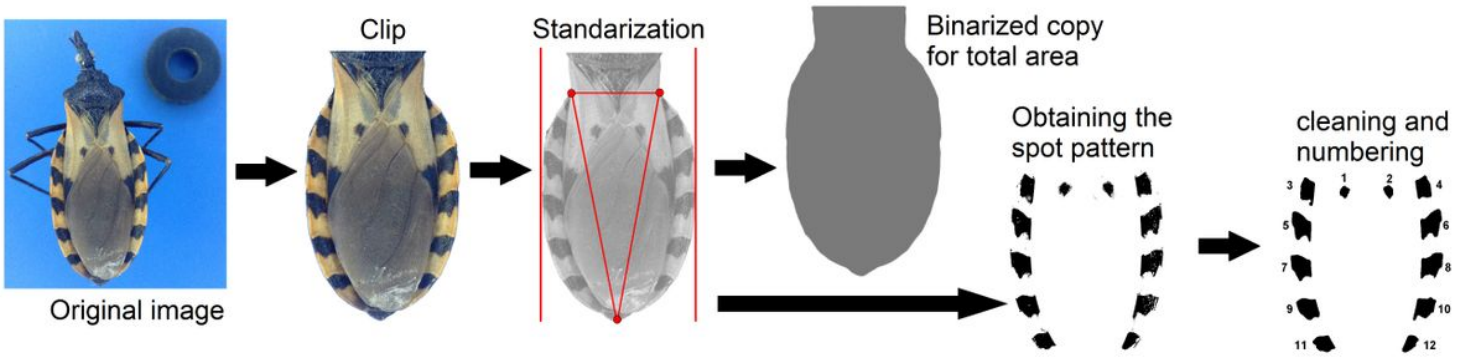

Figure 1

Steps in the image processing of three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae) for the analysis of spot patterns. Standardization includes isometric rescaling, translation, and rotation following common reference guidelines. Spot removal and cleaning were done with a macro code in ImageJ. Copyright: Creative Commons Attribution 1.0 Universal (CC0 1.0) Public Domain Dedication license (https ://creat iveco mmons .org/licen ses/by/1.0). Images modified from Gurgel-Gonçalves R, Komp E, Campbell LP, Khalighifar A, Mellenbruch J, Mendonça VJ, et al. Automated identification of insect vectors of Chagas disease in Brazil and Mexico: the virtual vector lab. PeerJ. 2017;5:e3040 [20]

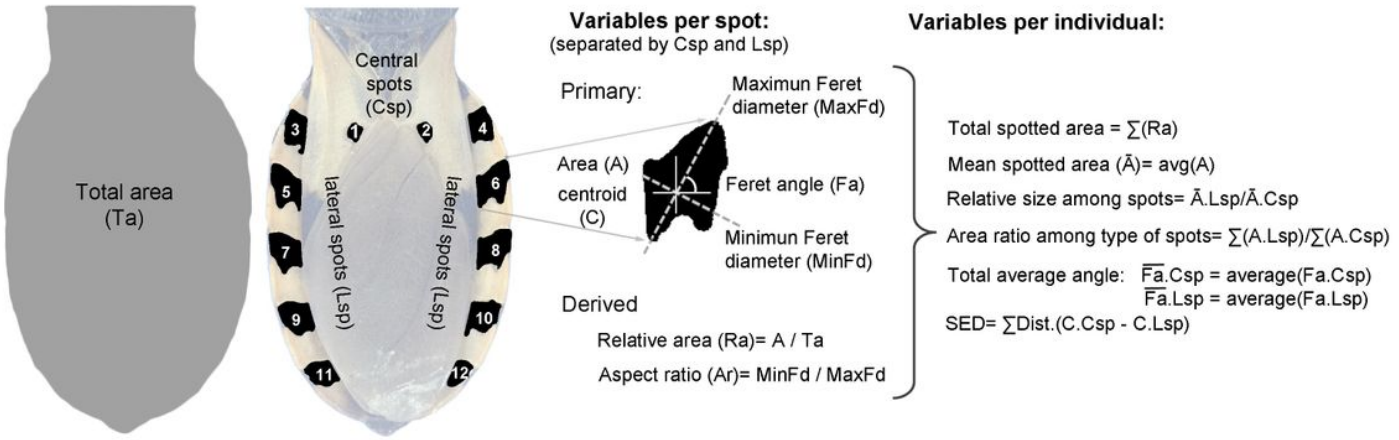

Figure 2

Variables used for the quantitative description of the spot pattern in three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae).

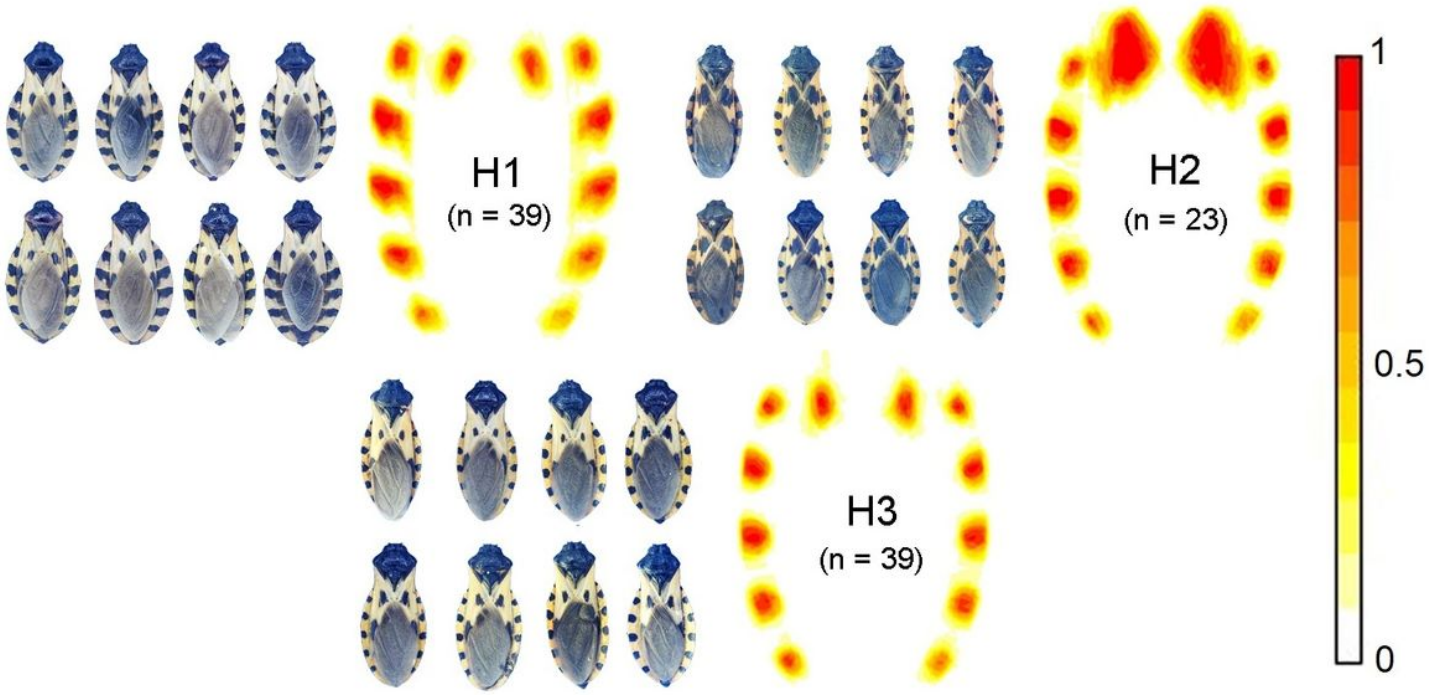

Figure 3

Heat maps obtained by superimposing the spot patterns of all the individuals from the samples of three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae), and a sample of the images used, to visualize the differences between the haplogroups. H1: Haplogroup 1, H2: Haplogroup 2, H3: Haplogroup 3. The color scale represents the areas of overlap between spots. High values represent the most shared pattern among individuals, while low values represent extreme patterns.

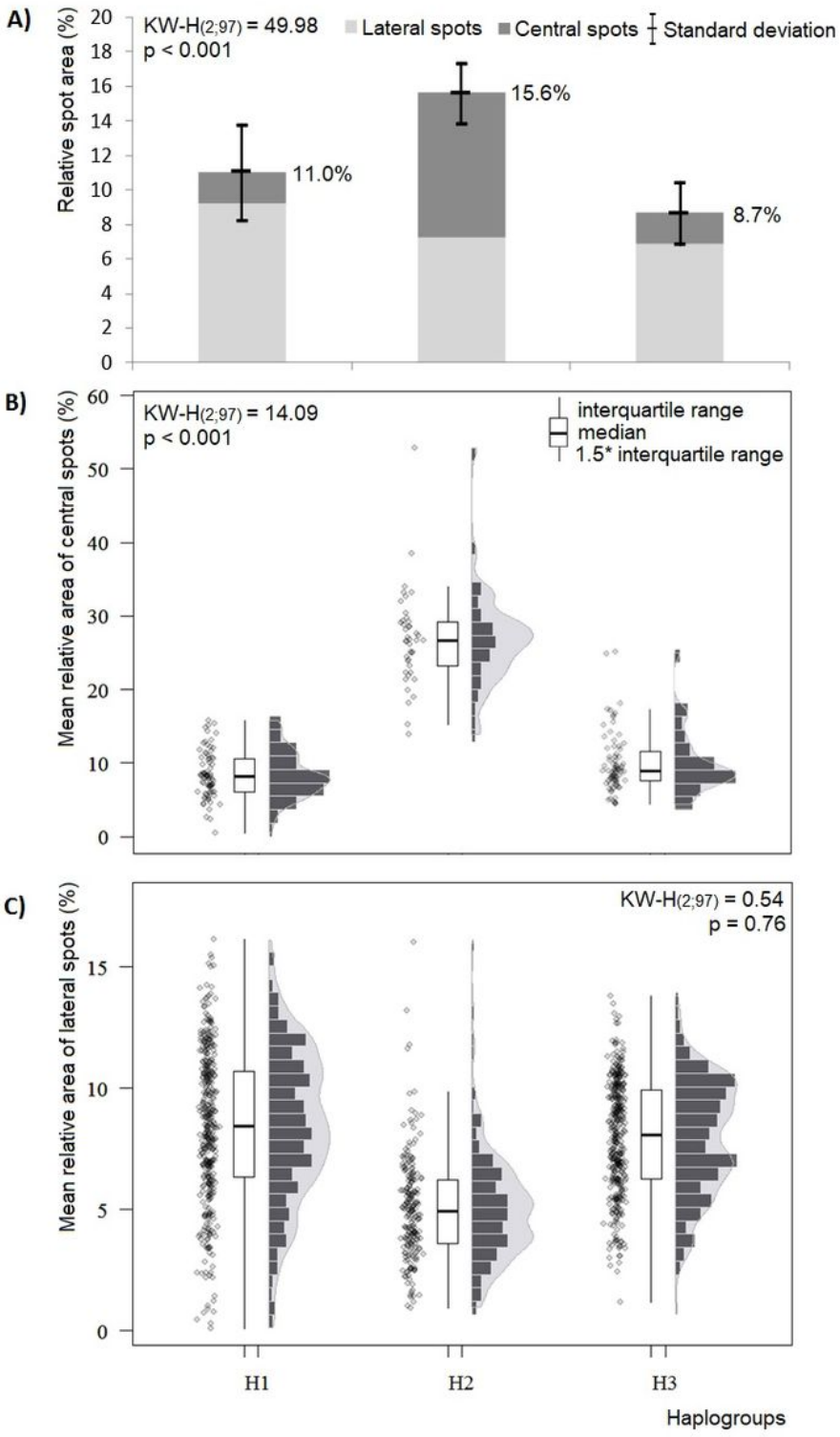

Figure 4

Differences in relative spot area of the abdomen (A) and the mean relative areas of the central (B) and lateral (C) spots in the three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae). H1: Haplogroup 1, H2: Haplogroup 2, H3: Haplogroup 3.

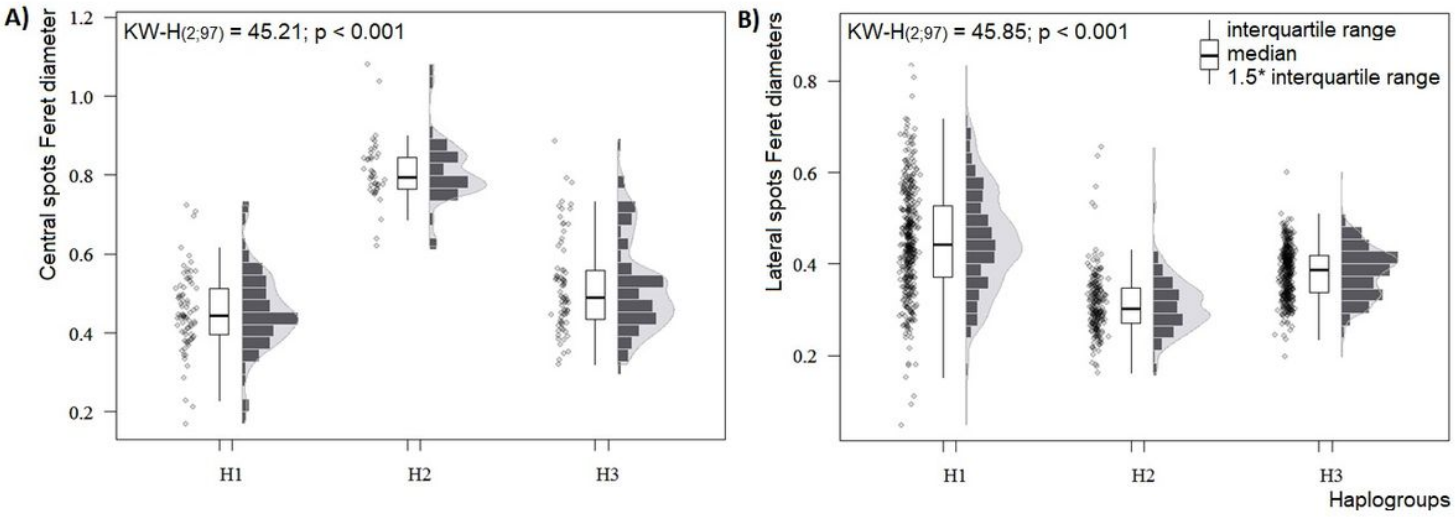

Figure 5

Differences in the average size (Feret diameters), of the central (A) and lateral (B) spots on the abdomen of three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae). H1: Haplogroup 1, H2: Haplogroup 2, H3: Haplogroup 3.

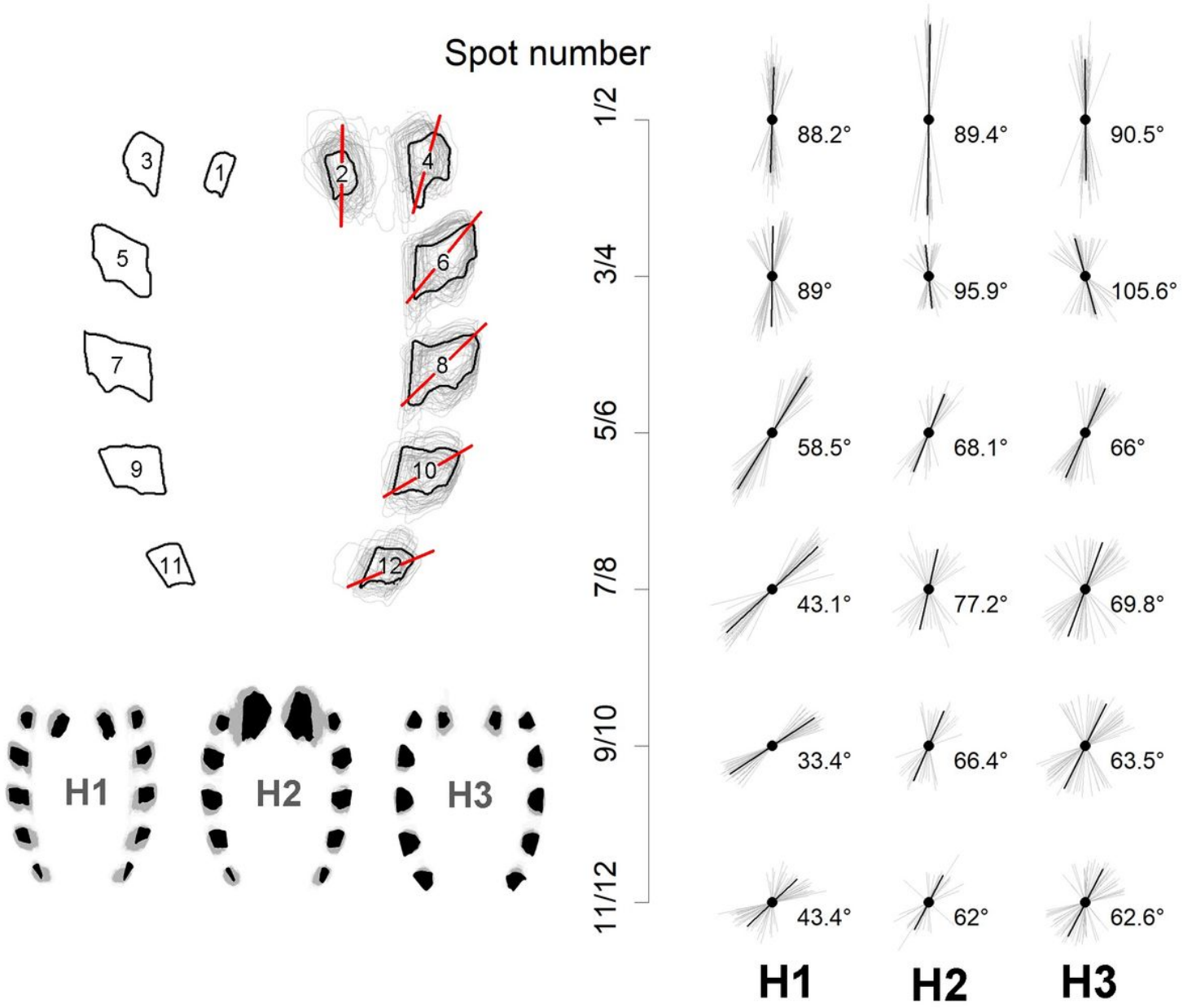

Figure 6

Orientation patterns of the spots (Feret diameter angle) H1: Haplogroup 1, H2: Haplogroup 2, H3: Haplogroup 3.

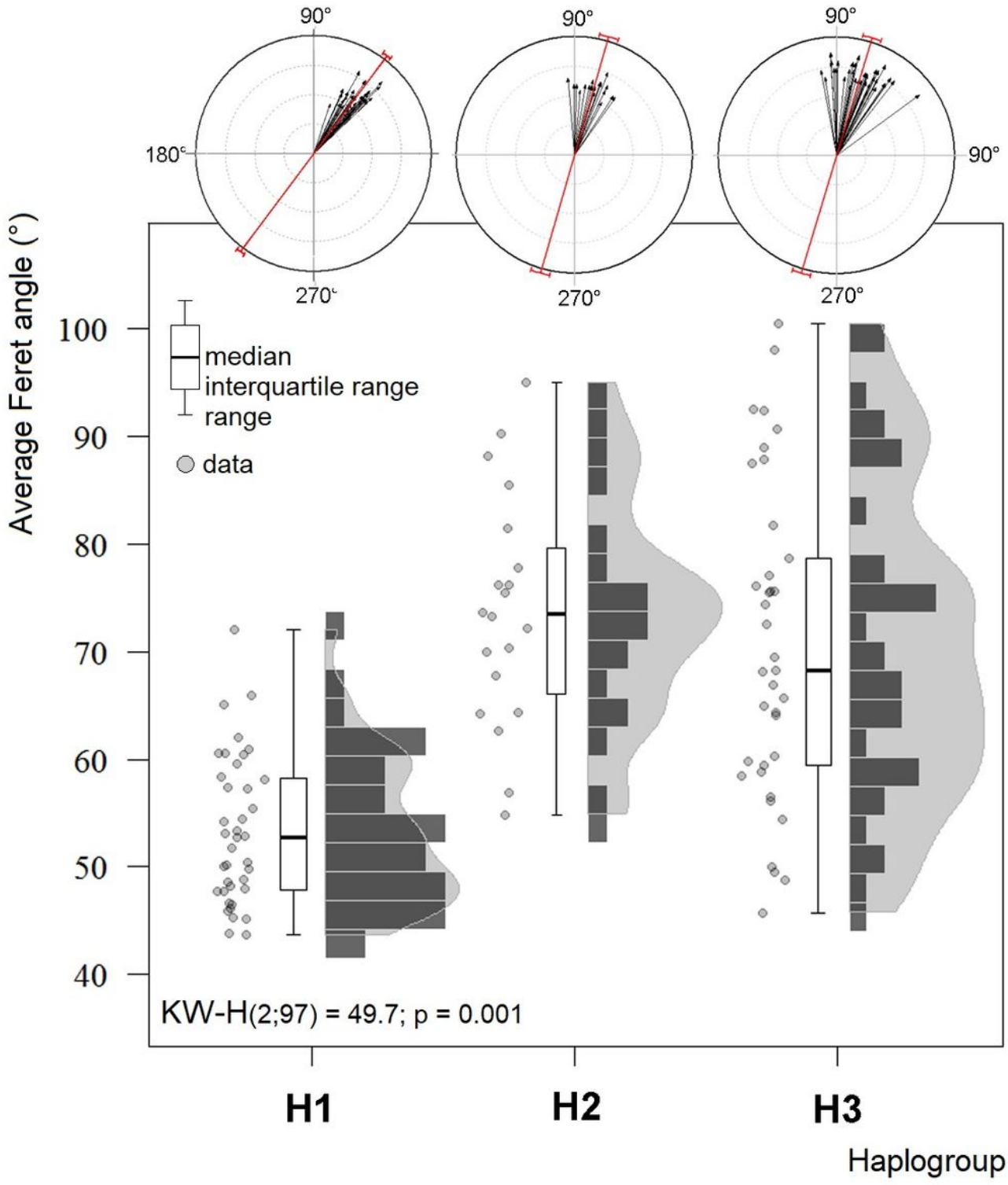

Figure 7

Comparison of the mean orientations of the lateral spots of the abdomen (Feret angle) between three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae). H1: Haplogroup 1, H2: Haplogroup 2, H3: Haplogroup 3.

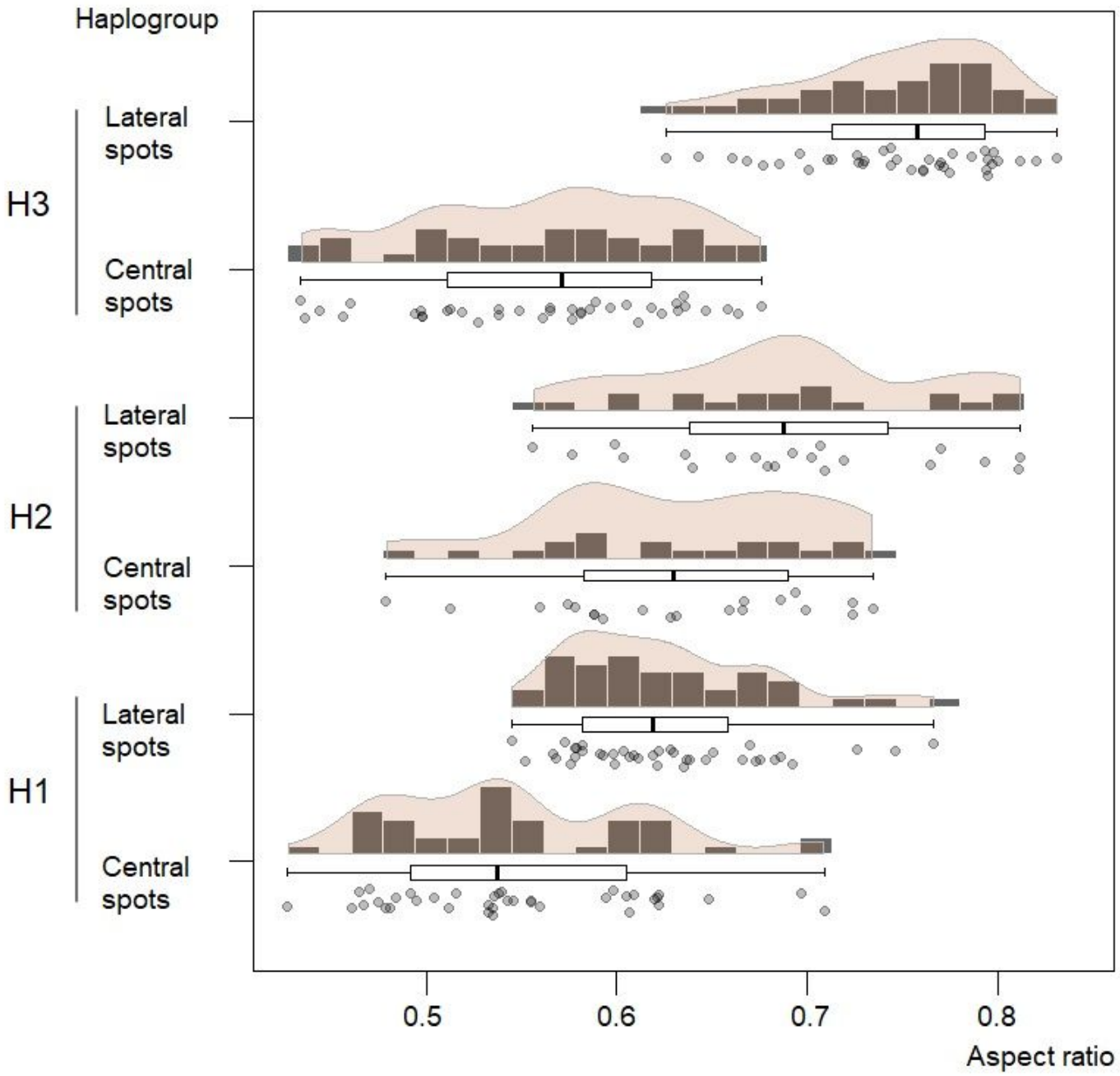

Figure 8

Comparison of the shape (aspect ratio) of the central and lateral spots of the abdomen in three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae). The aspect ratio differed among haplogroups, both for the central (KW-H (2; 96) = 15.45; p <0.001) and lateral spots (KW-H (2; 97) = 46.25; p = 0.001). H1: Haplogroup 1, H2: Haplogroup 2, H3: Haplogroup 3.

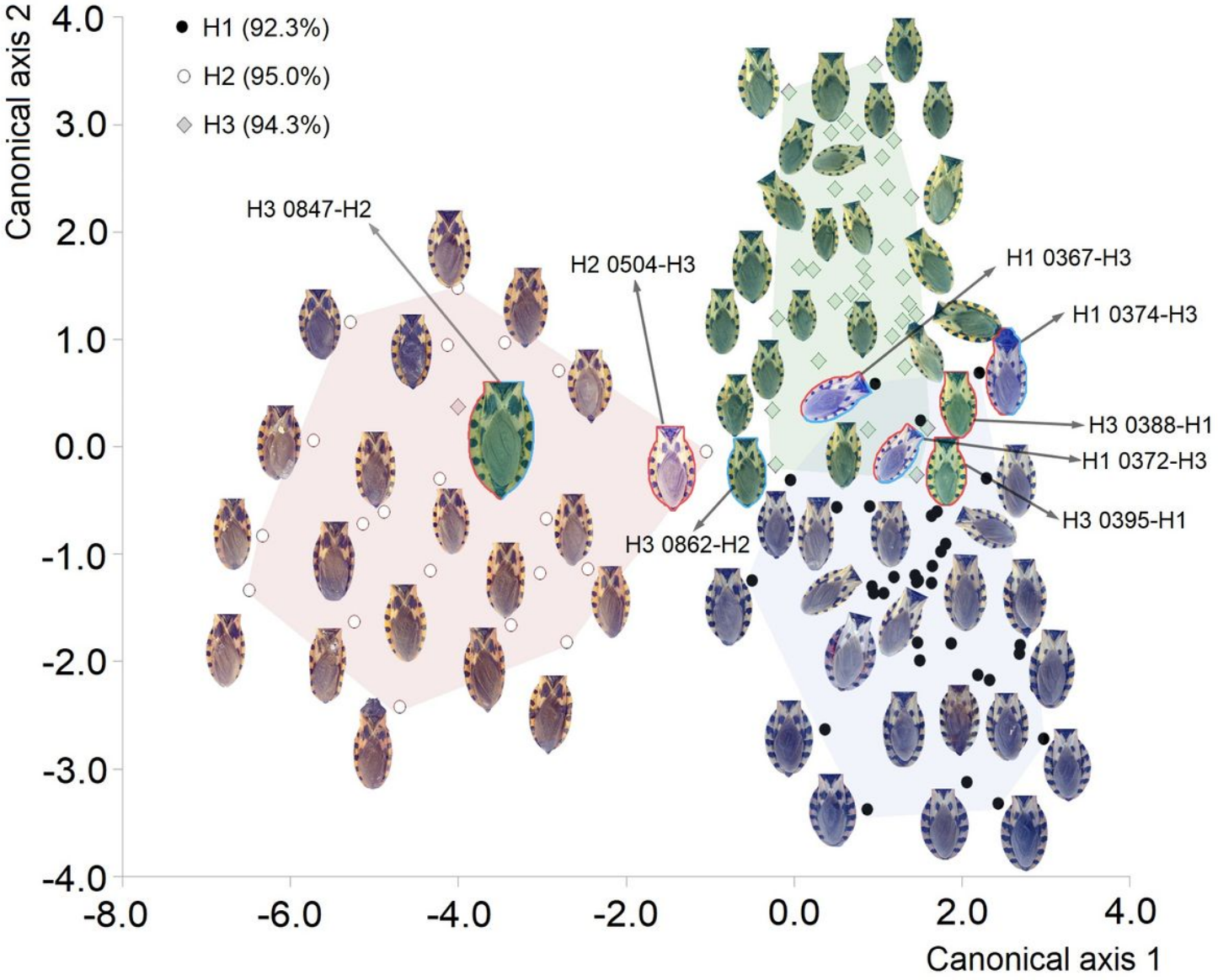

Figure 9

Ordination plot of the discriminant analysis of the individuals of Triatoma dimidiata (Hemiptera: Reduviidae), according to the analysis of discriminant function, from the properties of their spots. The correct classification percentages are shown, and the poorly classified individuals are outlined (red: according to the Linear Discriminant Function Analysis, blue: according to the neural network), as well as their numbers and the haplogroup with which they were confused. H1: Haplogroup 1, H2: Haplogroup 2, H3: Haplogroup 3.

Predictor variables	F to enter	p-level	No. Of Var	Lambda	F-value
Aspect ratio of central spot	0.405	0.668	-	-	-
Ratio of lateral spots area / central spots area	1.491	0.231	8.000	0.057	34.012
Mean relative area of central spots (%)	1.045	0.356	7.000	0.059	38.446
Mean relative area of lateral spots (%)	2.785	0.067	6.000	0.060	44.661
Spotted area (%)*	8.057	0.001	5.000	0.064	52.016
Feret diameter of lateral spots*	5.314	0.007	4.000	0.076	58.662
Feret angle of lateral spots*	17.513	0.000	3.000	0.085	73.111
Aspect ratio of lateral spot*	41.437	0.000	2.000	0.118	87.184
Feret diameter of central spots*	158.730	0.000	1.000	0.225	158.730

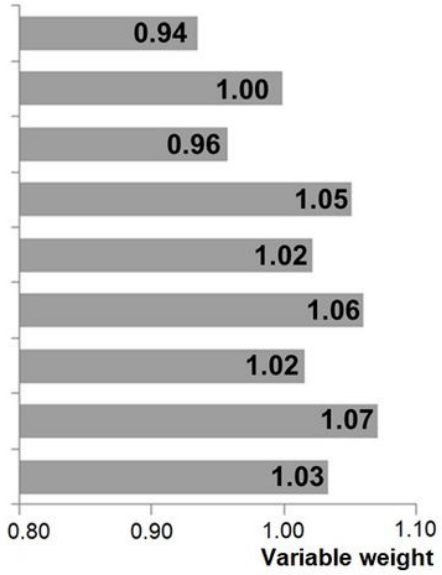

Figure 10

Relative importance of each variable in the classification procedures used to assign individuals to three haplogroups of Triatoma dimidiata (Hemiptera: Reduviidae) based on spot pattern. The statistical results of the Linear Discriminant Function Analysis are included, in which only the variables identified with a red asterisk and the weights assigned to the variables by the neural classification network obtained were included in the model. Wilks Lambda: 0.057; approx. F (16, 170) = 34.01; $p < 0.001$.

Supplementary Files

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

- SupplementaryInformationSI1Unusedindividuals.jpg
- SupplementaryInformationSI2MacroImageJ.ijm