Quantifying morphological variation: Geometric Morphometrics data as a tool for discriminating genera and species in Tabaninae (Diptera: Tabanidae)

Diego Fernando Acevedo Chaparro

Trabajo de grado para optar al título de

Biólogo

Director

Daniel Rafael Miranda Esquivel

Doctor en Ciencias Naturales

Co-directora

Juliette Cristina Gualdrón Díaz

Bióloga

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Bucaramanga

2019

AGRADECIMIENTOS

A mis padres, Ángel Iván y Martha Cecilia, por su amor, comprensión y apoyo incondicional durante cada día de mi vida, por ser el más noble ejemplo de paciencia, humildad y entrega. A mis hermanos, Iván Camilo y Sergio Andrés, por ser mis consejeros, defensores y mi permanente motivación e inspiración. A mi tía Alcira y mis abuelos, por creer en mí. A Óscar, Carolina, Camila, Dante y Lucía, por ser mí segunda familia.

A mis compañeros y amigos, por las enseñanzas, correcciones, apoyo y paciencia durante mi formación académica, profesional y personal; especialmente a Ale, Yoja, Angie y Ambrosio, quienes de forma amable y desinteresada, brindaron su conocimiento y tiempo en el diseño y ejecución de la investigación.

A los docentes, personal administrativo y logístico de la Universidad, por la colaboración, asesoría y orientación cada semestre.

A todos aquellos que, con su apoyo económico, emocional y académico, me permitieron continuar hasta lograrlo.

List of contents

RESUMEN

TÍTULO: CUANTIFICACIÓN DE LA VARIACIÓN MORFOLÓGICA: DATOS DE FORMA GEOMÉTRICA PARA LA DISCRIMINACIÓN DE GÉNEROS Y ESPECIES EN TABANINAE (DIPTERA: TABANIDAE).[*]

AUTOR: DIEGO FERNANDO ACEVEDO CHAPARRO[**]

PALABRAS CLAVES: TAXONOMÍA, NEOTRÓPICO, FORMA DEL ALA.

DESCRIPCIÓN: La determinación taxonómica en Tabaninae (Diptera: Tabanidae) ha sido históricamente difícil de abordar debido a la variación de los caracteres morfológicos usados para la clasificación a nivel de subfamilia. Varios enfoques alternativos en las últimas décadas han sido utilizados para la identificación de géneros y especies. Por medio de métodos de Morfometría Geométrica digitalizamos configuraciones de 28, 29 y 30 landmarks tipo I sobre las intersecciones y terminaciones de las venas en las alas derechas de 1158 hembras adultas, pertenecientes a 38 especies colectadas en 15 áreas protegidas del país, para cuantificar la forma geométrica del ala como un grupo de caracteres para la correcta determinación taxonómica y estimamos el efecto del número y configuración de landmarks sobre la determinación taxonómica en Tabaninae. Los resultados obtenidos en los análisis de agrupamiento por el método UPGMA y los análisis de Componentes Principales no permitieron la clara separación de grupos a nivel de tribus y géneros. Sin embargo, los análisis de discriminantes permitieron la correcta determinación en casi la totalidad de casos evaluados y revelaron aquellos landmarks que deberían priorizarse para maximizar la discriminación taxonómica. Por lo tanto, la forma geométrica del ala permite la correcta discriminación taxonómica a nivel de tribu, género y especie en Tabaninae.

(*) Bachelor thesis

(**) Facultad de Ciencias. Escuela de Biología. Supervisor: Daniel Rafael Miranda Esquivel, PhD

ABSTRACT

TITLE: QUANTIFYING MORPHOLOGICAL VARIATION: GEOMETRIC MORPHOMETRICS DATA AS A TOOL FOR DISCRIMINATING GENERA AND SPECIES IN TABANINAE (DIPTERA: TABANIDAE) (*)

AUTHOR: DIEGO FERNANDO ACEVEDO CHAPARRO (**)

KEYWORDS: TAXONOMY, NEOTROPIC, WING SHAPE.

DESCRIPTION: Taxonomic recognition in Tabaninae (Diptera: Tabanidae) has been historically difficult to assess due to the variation of the morphological characters used for classification at tribe, genus and species level in some particular groups. Non-morphological (*i.e* molecular, enzimatic and biochemical) data have been collected in the last few decades to approach taxonomy within and between groups and ease identification of genera and species. Using GM methods and wing shape data, we identified and measured configurations of 28, 29 and 30 type I landmarks on the veins of the right wings of 1158 female adults belonging to 38 species, collected in 15 protected areas, to quantify wing shape as a set of characters for taxonomic discrimination and estimated the effect of both number and configuration of type I landmarks on taxonomic recognition in Tabaninae. We found unclear taxa identification through UPGMA cluster analysis and PCA analysis at tribe and genus level. Nevertheless, Discriminant Analysis allowed the correct identification of taxa in most cases and revealed those landmarks that might be prioritized to maximize taxonomic discrimination scores at each level. Thus, wing shape data is a feasible and reliable tool for taxonomic discrimination in the subfamily Tabaninae at tribe, genus and species level.

(*) Bachelor thesis

(**) Facultad de Ciencias. Escuela de Biología. Supervisor: Daniel Rafael Miranda Esquivel, PhD

Introduction

Horse-flies, deer-flies and clegs (Diptera: Brachycera: Tabanidae) are a monophyletic family (Morita *et al.* 2016; Wiegmann *et al.* 2000) listing about 4400 species with world-wide distribution (Lambkin *et al.* 2013; Morita *et al.* 2016; Pape *et al.* 2011). Tabanids are generally large flies with black, brown or grey bodies in which pale areas are common, bearing large calypteres and lacking body bristles (Chainey, 1993). Since nearly 25% of its known diversity occurs in the Neotropical region (Coscarón & Papavero, 2009; Henriques *et al.* 2012) a growing interest on the local tabanofauna has been aroused (Rodrigues Guimarães *et al.* 2017) and to date in Colombia 256 species have been recorded (Wolff & Miranda-Esquivel, 2016). Nowadays, the current classification scheme splits Tabanidae into four subfamilies (Tabaninae, Chrysopsinae, Pangoniinae, and Scepsidinae) with further division into 10 tribes (Coscarón *et al.* 1979; Mackerras, 1954). However, in some cases their high morphological variation makes taxonomy within the family difficult to assess (Chainey, 1983; Oldroyd, 1957; Usher, 1972) and today taxonomic research on Neotropical Tabanidae is a yet developing field (Torres & Miranda-Esquivel, 2016).

Taxonomic identification keys are based on external and internal morphological features of adults and the known immature stages (Coscarón & Papavero, 2014). Abdomen coloration patterns and wing venation characters (Maity *et al.* 2015) are diagnostic for classification at subfamily level (Chainey, 1993). However, taxonomic discrimination of genera or species is not always feasible since some characters do not exhibit enough variation or might be altered because of age and preservation methods (Morita *et al.* 2016). Therefore, problems related to taxonomic recognition based on morphological traits are common in Tabanidae (Coscarón & Papavero, 1993) and are especially evident for some of its subfamilies, such as Tabaninae.

For helping solve this problems related to taxonomic recognition non-morphological evidence, such as enzymatic (Hudson & Teskey, 1976), biochemical (Sutton & Carlson, 1997) and molecular data (Banerjee *et al.* 2015) have been gathered. These researches include the *Tabanus nigrovittatus* Macquart complex (Sofield *et al.* 1984), some species of *Stypommisa* Enderlein (Cárdenas *et al.* 2009) and the taxonomic discrimination of *Diachlorus* Osten-Sacken using wing shape data (Torres & Miranda-Esquivel, 2016).

In this taxonomic context, Geometric Morphometrics uses homologous morphological loci or *landmarks* (Thompson, 1917) to estimate and characterize biological shape (Arbour & Brown, 2014) through collecting and analyzing information otherwise impossible to be examined (Adams, Rohlf, & Slice, 2004; Bookstein, 1991, 1997). Thus, Geometric Morphometrics represents a suitable tool for discriminating taxa (Arbour & Brown, 2014) and approach taxonomic surveys (Baylac *et al.* 2003) in a wide range of life forms (Beolchini & Corti, 2004; Costa-Schmidt & de Araújo, 2010; Perrard *et al.* 2014). Tabanids wings, as in most taxa within the Insecta, present appropriate attributes for Geometric Morphometrics data acquisition, since they are mainly hyaline and its features are easy to recognize (Dujardin *et al.* 2012; Klingenberg *et al.* 2001; Torres & Miranda-Esquivel, 2016), as well they hold detectable and quantifiable interspecific variation (Houle *et al.* 2003).

1 Objectives

1.1 General objective

Quantify wing shape as a set of characters for taxonomic discrimination in Tabaninae.

1.2 Specific objective

Estimate the effect of number and configuration of type I landmarks on taxonomic discrimination in Tabaninae.

2 Materials and methods

2.1 Samples and data collection

We used 1158 adult females collected in 15 protected areas (See supplementary material Table 1) belonging to 38 species (12 genera of Diachlorini and 2 genera of Tabanini). We used females as blood-feeding females are more abundant than nectar-feeding males (Coscarón, & Papavero, 2009; Lessard *et al.* 2013).

We identified all samples using taxonomic keys to each genus and species (See supplementary material Table 2) and included into the analysis those specimens with complete and unfolded right wings. The specimens under study belong to the Entomology collections of Instituto de Investigación de Recursos Biológicos Alexander von Humboldt (IAvH-E), Instituto de Ciencias Naturales de La Universidad Nacional de Colombia (ICN) and Museo de Historia Natural de la Universidad Industrial de Santander (MHN-UIS).

Localities. We used samples from 15 localities belonging to five biogeographic provinces (See supplementary material figure 1).

Wings. We mounted each wing in dorsal view, not removing any from the axillary sclerites, over microscope slides and photographed them using a Canon Rebel Xs 10.1 megapixel camera attached to a Nikon SMZ 1000 stereoscopic microscope.

For selecting the sample size, we followed two criteria: availability of specimens in the museum collection (Cardini *et al.* 2015) and exploratory analyses (Bartlett *et al.* 2001; Cochran, 1977; Kar & Ramalingam, 2013).

Landmarks digitization. Based on venation patterns, we defined 28 homologous traits or Type-I landmarks (Bookstein, 1991) to represent wing shape at tribe, genus and species level (See supplementary material figure 2). A fork of variable length at R4 in some specimens provided 1 or 2 additional landmarks (29 or 30). In these late cases, the analysis were performed only at species level. We included all recognizable venation features and did not considered spots or coloration patterns (See supplementary material figure 2). We followed wing venation nomenclature as proposed by Brown *et al.* (2010).

Geometric morphometrics analysis. First, we removed the effect of size, rotation and orientation from our landmark data set through Generalized Procrustes Analysis (GPA) under the least squares criterion (Gower, 1975; Rohlf & Slice, 1990) and, using the GPA values, we calculated the consensus shapes (CS). Then, we evaluated whether the taxonomic groups are recovered through Cluster Analysis under the UPGMA method (Sokal & Michener, 1958) based on the Mahalanobis distances. For selecting the optimal number of clusters, we tested different Gaussian distribution models and selected the model with the highest Bayesian Information Criterion (BIC) value (Fraley & Raftery, 2002). Next, we reduced data dimensions and observed trends in shape variation through Principal Component Analysis (Pearson, 1901) based on the GPA values. Finally, we estimated taxonomic differences between groups in terms of wing shape through Discriminant Analysis (DA) (Fisher, 1936).

We split our data set so that GPA, CS and DA were calculated at each taxonomic level, regardless of the number of landmarks. For the analyses at tribe and genus level, we only included all taxa in which 28 landmarks were measured. At genus level we took a sample of some species of *Diachlorus* Osten-Sacken, *Phaeotabanus* Lutz, *Dichelacera* Macquart, *Poeciloderas* Lutz and *Tabanus* Linnaeus (See supplementary material Table 3). At species level, we also included those in which 29 and 30 landmarks were digitized, performing the comparisons between groups of four, three or two species. The four groups of species with 28 landmarks were **group 1:** *D. curvipes* Fabricius, *D. leucotibialis* Wilkerson & Fairchild, *T. secundus* Walker and *T. simplex* Walker*;* **group 2:** *T. occidentalis* Linnaeus, *T. secundus* and *T. simplex;* **group 3:** *D. curvipes, P. magnificus* Kröber, *T. secundus* and *T. simplex;* **group 4:** *P. serenus* Kröber and *T. simplex.* The three groups of species with 29 landmarks were **group 5:** *C. inanis* Fabricius and *T. occidentalis;* **group 6:** *D. leucotibialis, D. marginata* Macquart and *T. simplex;* **group 7:** *D. curvipes, D. fasciata* Walker and *D. nuneztovari* Fairchild & Ortiz. The three groups of species with 30 landmarks were **group 8:** *T. thiemeanus* Enderlein and *T. occidentalis;* **group 9:** *T. thiemeanus* and *T. surifer* Fairchild*;* **group 10:** *D. nuneztovari*, *T. occidentalis var. modestus* Kröber, *T. secundus* and *T. surifer.* (See supplementary material Table 4).

Optimal configuration. We performed 905 pseudo replicates extracting landmarks from the data set to measure their contribution to taxonomic recognition based on both, number and configuration. As we aimed to find configurations that optimize the analyses through estimating their effect on taxonomic discrimination we included those close landmarks since they could be potentially functional.

First, we removed one individual landmark at a time, starting from L1 to L28, L29 or L30. Then, we removed pairs of landmarks as follows: L1 and L2, L1 and L3 and so on. We did not remove landmark sequences starting form L2, L3 or any other landmark. Next, we removed sequences of greater numbers of landmarks representing particular wing regions: basal region: L1-L6; Upper margin region: L7-L12; Lower margin region: L13-18; Radial-Cubital region: L19-L24; Radial-Medial region: L25-L30. Finally, we subtracted random landmark arrangements (10 to 24) until finding a minimum viable number of landmarks that allow discrimination scores greater than 50 % at each taxonomic level. Landmarks digitized on cross-veins (L1-L16) are referred to as inner landmarks while landmarks digitized on the end of veins (L17-L30) are referred to as margin landmarks (See supplementary material figure 2). All the analyses and comparisons were performed as mentioned in the *Geometrics Morphometrics Analysis* section.

Software. For landmark digitization or data collection we used the package *geomorph* (Adams & Otárola-Castillo, 2013). For the geometric morphometrics analyses we used the libraries *geomorph* (Adams et al., 2004) *cluster* (Maechler *et al.* 2018), *DiscriMiner* (Sánchez, 2013), *ellipse* (Murdoch & Chow, 2018), *colorspace* (Ihaka *et al*. 2016), *ggplot2* (Wickham, 2016), *mclust* (Fraley *et al.* 2018), *shapes* (Dryden, 2018) and *vegan* (Oksanen *et al.* 2018). Landmark digitization and all geometric morphometrics analyses were performed using The R software environment for statistical computing (R Core Team, 2017).

2.2 Results

Landmarks digitization. Specimens with configurations of 28 landmarks are present in 14 genera occurring in all localities. In 9 of these localities and 12 of these genera we also found specimens with configurations of 29 and 30 landmarks. Thus, we found no evidence to suggest that the number of landmarks of a given taxon may be defined by geography but, instead, could be related to taxonomy, since only in two genera (*Stypommisa* and *Poeciloderas*) the number of landmarks was 28 in all cases.

Geometric morphometric analyses.

Cluster analysis. Classification at tribe level (Diachlorini and Tabanini) based on the BIC value showed six groups represented in our dataset (See supplementary material figure 3).

Principal component analysis. At tribe level, the first two components accounted for 55% of the variance. We found two overlapping groups that did not coincide with the two tribes. We could not observe Diachlorini nor Tabanini occupying independent regions in the morphospace, as all genera here included are distributed mainly on the PC2 (See supplementary material figure 4).

We did not observe independent clusters, but instead, *Dichelacera* and *Tabanus* appeared overlapped, such as *Phaeotabanus* and *Poeciloderas*. *Diachlorus* was scattered over a wider range than any other genus, accounting for more shape variation on the first two PC axis, compared to the other genera (See supplementary material figure 4).

Conversely, most of the groups compared at species level were clustered independently in the morphospace. All groups of species with 28 landmarks could be recognized except for *T. simplex* and *T. occidentalis* (See supplementary material figure 5) and all groups of species with 29 landmarks formed independent clusters (See supplementary material figure 6).

Nevertheless, in some of the groups of species with 30 landmarks we observed partial or total overlapping (*i.e. Tabanus occidentalis var. modestus, T. secundus* and *T. simplex;* and *T. surifer* and *T. thiemeanus*) (See supplementary material figure 7).

Discriminant analysis. Wing shape data allowed the successful taxonomic recognition of both tribes, since discrimination score for Diachlorini was 98% and for Tabanini was 99%. These scores were similar at genus level, in which *Diachlorus*, *Dichelacera*, *Phaeotabanus*, *Poeciloderas* and *Tabanus* had 100% discrimination scores (See supplementary material Table 6).

For all the groups of species discrimination scores were, in almost all cases, over 95%. The only species under this score was *T. occidentalis var. modestus* (group 10) which discrimination score was 50%.

Optimal configuration. For most of the configurations tested at tribe, genus and species level, removing landmarks individually or by the pairs used showed little or no effect on taxonomic recognition. For Diachlorini and Tabanini, we observed negative deltas in discrimination scores when removing some particular landmarks or pairs of landmarks. A similar behavior was observed when subtracting arrangements of up to ten or fifteen landmarks, in which for Diachlorini discrimination scores remained above 78% in all cases, and for Tabanini remained above 89% (See supplementary material Table 5).

For the groups of species, in only one case a positive delta in discrimination scores was caused after subtracting landmark R_1 and C for *T. occidentalis var. modestus*. In contrast, when the landmark on appendix vein R_4 was removed the discrimination score dropped to 25%. (See supplementary material Table 7).

These arrangements of landmarks do not belong to a particular region of the wing, but instead, are found at different vein junctions and ending of veins. We found that using only four landmarks: h and Sc, R_4 and C, Petiole A1 and CuA_2 and br and R_{2+3} discrimination scores remained above 65% for both tribes. In addition, some of the configurations allowing the highest discrimination scores for both tribes included the Alular incision, M and Cu, h and Sc, Sc and C, R_4 and C, CuA, A1 and CuA_2, bm-cu and CuA1, bm-cu and M_3, M_1 and M_3, br and R_{2+3}, r-m and R_{4+5} and R_5 and R_4.

We observed no compromise in taxonomic identification when removing individual or pairs of landmarks at genus level, since in all cases, discrimination scores remained at 100%. In contrast, removing arrangements of four and up to twenty landmarks caused negative deltas on discrimination scores for each genus, depending on which landmarks were subtracted (See supplementary material Table 4). Even in these cases, the five genera were correctly discriminated when using a configuration of ten landmarks, including h and Sc, h and C, R_1 and R_{2+3}, R_1 and C, R_4 and C, M_1, Petiole A1 and CuA_2, M_1 and M_3, r-m and M_{1+2} and r-m and R_{4+5}. We found that, some particular configurations of only four landmarks allow distinguishing *Diachlorus*: M and Cu, Sc and C, R_4 and C, bm-Cu and CuA1 and *Poeciloderas*: Alular incision, bm-cu and CuA1, L14 h and Sc, L17 A1 and CuA_2 (See supplementary material Table 6)

Depending on the species, removing sequences or arrangements of up to 20 landmarks caused different levels of impact on taxonomic recognition. However, in several cases, discrimination scores remained above 75% (See supplementary material Table 7). As our most remarkable result, we found that using five different configurations of four landmarks, discrimination scores for *T. occidentalis var. modestus* raised from 50% to 100%.

We obtained perfect discrimination using only R_1 and C, R_{2+3} and C, bm-cu and CuA_1 and br and R_{2+3} for *Diachlorus curvipes* Fabricius, *D. nuneztovari* Fairchild & Ortiz and *Dichelacera fasciata* Walker. Similarly, *T. occidentalis* and *P. thiemeanus* had %100 discrimination scores when including only h and C, R_5, CuA_1 and CuA_2 and r-m and R_{4+5}.

2.3 Discussion

Landmarks digitization. The variation in the general venation pattern of 28 landmarks in Tabaninae, represented by the fork of veins R_4 and R_5 (L29) and the appendix veins R_4 (L30), could be, as found in other taxa within the Insecta, related to environmental factors (Oguz *et al.* 2017) as localities are composed of different vegetation species and differ in their climatic conditions, or seasonality (Prudhomme *et al.* 2016) since samples were collected in different periods, instead of geographical variation (Bai *et al.* 2015). Nevertheless, as these two additional landmarks are not common to all taxa under study they should be used cautiously for taxonomic research through Geometric Morphometrics methods.

Cluster Analysis. We did not observe Diachlorini and Tabanini as independent tribes. However, the groups formed are clear in terms of separation, but they do not coincide with the known taxa within Tabaninae, as most of the genera of both tribes are clustered together. Genera of Diachlorini appeared to be more similar to those of Tabanini, since they are recovered within *Tabanus.* Thus, wing shape, as seen for morphological characters, might be highly variable within a given group (Lorenz *et al.* 2017).

Principal Component Analysis. In Diachlorini, the most diverse and varied tribe, in morphological terms, in the family (Bayless *et al.* 2009) shape behavior at the morphospace might be caused by higher levels of variation at genus or species level, which might be related to the absence of redundant information (Kucharczyk *et al.* 2012) or to sample size (Kocovsky, 2009). At this taxonomic level, inner and margin landmarks (L1-L16 & L17-L28) account for the overall within tribe variation, since both reveal deformations (Fig. 4).

In Tabanini, several characters used in taxonomic recognition tend to overlap, causing identification problems (Wilkerson, 1978) because of morphological similarity. This might be reflected in wing shape data since some body variables could be correlated to shape variables (Dujardin, 2008).

Considering the overlapping of both tribes on the first PC2 axes, it could be suggested that there are wing landmarks that could share variation patterns (Kucharczyk *et al.* 2012). In geometric morphometrics terms, those landmarks allowing a clear group separation at tribe level are not represented by veins junctions, but instead, by endings of veins on the wing margin. Therefore, visualizing data through the morphospace is a practical approach for distinguishing particular features that might be functional for taxonomic recognition (Chartier *et al.* 2014; Egan, 2015; Osman *et al.* 2011) even at higher-level taxa.

Discriminant analysis. Wing venation patterns are used for description at order or family level (Comstock & Needham, 1899) but through Geometric Morphometrics data we found differences that enable correct taxonomic discrimination at tribe, genus and species level, suggesting that Geometric Morphometrics methods represent a reliable tool for taxonomic recognition in Tabaninae.

Despite the high variation presented at subfamily level, wing shape data allowed the successful recognition of each tribe, representing a suitable means for suprageneric identification, as it is for the Scarabaeoidea (Ren *et al.* 2017). However, these results opposite to what have been tested in other orders of Insecta, such as the Hymenoptera, since most of the variation is found at the interspecific level (Perrard *et al.* 2014) or at the genus level, as in the Hemiptera (Li *et al.* 2017), but coincide traditional morphology within Tabanidae (Torres & Miranda-Esquivel, 2016).

As found at tribe level, despite not having a clear group separation through PCA, we correctly identified all genera under study (See supplementary material Table 6). Therefore, at this taxonomic level, differences in wing shape are descriptive and functional to achieve a clear group discrimination when all the initial landmarks are included, making possible to detect between-group differences without losing accuracy on taxonomic recognition caused by lower-level variation (Pretorius & Scholtz, 2001). Additional landmarks do not improve discrimination scores at species level (See supplementary material Table **7**).

Optimal configuration. Contrary to what we found through PCA analyses, in which margin landmarks seem to be critical for taxonomic identification, there is no rule about whether inner or margin landmarks should be used for taxonomic recognition, since particular arrangements of both are necessary to optimize between-group discrimination at tribe, genus and species level. This might suggests that partial datasets are useful for taxa identification (Gupta *et al.* 2018) since shape variation in Tabaninae could be summarized by particular landmarks taken from the initial configurations.

Using only 14% of the original landmark dataset (*i.e.* four landmarks) discrimination scores in both tribes are similar, or even higher (See supplementary material Table 5) than those found for other taxa in which a larger number of landmarks was used (Mooney & Kingsford, 2016; Ruane, 2015; Sontigun *et al.* 2017). At genus level, we observed a similar behavior to that of tribes regarding taxonomic discrimination based on wing shape data, in which inner and margin landmarks improve recognition when used together (See supplementary material Table 6). Compared to tribes, genera need more landmarks to be discriminated. However, such as at tribe level, functional configurations might be selected from the dataset. For optimizing species discrimination, particular configurations of different number of landmarks are required; although there are particular landmarks present in several configurations (See supplementary material Table

7), we could not detect a common configuration allowing the correct taxonomic discrimination for all the groups of species under study.

3 Conclusions

Our results present a practical way for an efficient taxon recognition within Tabaninae, since discrimination scores after removing landmarks suggest the viability for the correct identification of tribes, genera and species using particular landmarks instead of the entire dataset. Therefore, correct taxonomic discrimination based on Geometric Morphometrics data is feasible even when only partial information is available or wings have been damaged, since particular landmarks might summarize shape variation within the subfamily. These landmarks differ at each taxonomic level; thus, there is at least one useful configuration for recognizing tribes, genera or species, suggesting that wings shape varies independently at each taxonomic level.

Nevertheless, as in several of these configurations some landmarks are relatively close, the effect of distance should be measured in further analyses. In addition, the correct taxonomic discrimination in Tabaninae would be determined by how landmarks are arranged, rather than how many of them are measured. Then, such exploratory analyses become relevant for detecting useful information for taxonomic surveys in Tabaninae.

Bibliographic references

Adams, D.C. & Otárola-Castillo, E. (2013). Geomorph: An R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4(4), 393–399.

Adams, D.C. Rohlf, F.J. & Slice, D.E. (2004). Geometric morphometrics: Ten years of progress following the 'revolution.' *Italian Journal of Zoology*, 71(1), 5–16.

Arbour, J.H. & Brown, C.M. (2014). Incomplete specimens in geometric morphometric analyses. *Methods in Ecology and Evolution*, *5*(1), 16–26.

Bai, Y., Ma, L. B., Xu, S.Q. & Wang, G.H. (2015). A Geometric Morphometric Study of the Wing Shapes of *Pieris rapae* (Lepidoptera: Pieridae) from the Qinling Mountains and Adjacent Regions: An Environmental and Distance-Based Consideration. *Florida Entomologist*, 98(1), 162–169.

Banerjee, D., Kumar, V., Maity, A., Ghosh, B., Tyagi, K., Singha, D., Kundu, S., Laskar, B., Naskar, A. & Rath, S. (2015). Identification through DNA barcoding of Tabanidae (Diptera) vectors of surra disease in India. *Acta Tropica*, 150, 52–58.

Bartlett, J.E., Kotrlik, J.W. & Higgins, C.C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, 19(1), 43–50.

Baylac, M., Villemant, C. & Simbolotti, G. (2003). Combining geometric morphometrics with pattern recognition for the investigation of species complexes. *Biological Journal of the Linnean Society*, 80(1), 89–98.

Bayless, K., Morita, S. & Wiegmann, B. (2009). The evolution of Diachlorini (Diptera: Tabanidae) based on phylogenetic analyses using morphological and molecular data: Conference:

Entomological Society of America Annual Meeting. Available from: https://esa.confex.com/esa/2009/webprogram/Paper44425.html (Accessed 18 August 2018).

Beolchini, F. & Corti, M. (2004). The taxonomy of the genus *Tachyoryctes*: A geometric morphometric approach. *Italian Journal of Zoology*, 71(1), 35–43.

Bookstein, F.L. (1991). Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press, 455 pp.

Bookstein, F.L. (1997). Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(3), 225–243.

Brown, B.V., Borkent, A., Cumming, J.M., Wood, D.M., Woodley, N.E. & Zumbado, M.A. (2010). *Manual of Central American Diptera*. NRC Research Press, Ottawa, Ontario, Canada, 714 pp.

Cárdenas, R.E., Buestán, J. & Dangles, O. (2009). Diversity and distribution models of horse flies (Diptera : Tabanidae) from Ecuador. *Annales de La Société Entomologique de France (N.S.): International Journal of Entomology*, 45(4), 511–528.

Cardini, A., Seetah, K. & Barker, G. (2015). How many specimens do I need? Sampling error in geometric morphometrics: testing the sensitivity of means and variances in simple randomized selection experiments. *Zoomorphology*, 134(2), 149–163.

Chainey, J.E. (1983). Afrotropical Tabanidae (Diptera): the genus *Philoliche* Wiedemann, subgenus *Ommatiosteres* Enderlein. *Annals of the Natal Museum*, 25(2), 453–474.

Chainey, J.E. (1993). Horse-flies, deer-flies and clegs (Tabanidae). *In:* Lane R.P. & Crosskey R.W. (eds) *Medical Insects and Arachnids.* Springer, Dordrecht, pp. 310–332.

Chartier, M., Jabbour, F., Gerber, S., Mitteroecker, P., Sauquet, H., von Balthazar, M., Staedler, Y., Crane, P.R. & Schönenberger, J. (2014). The floral morphospace – a modern comparative approach to study angiosperm evolution. *New Phytologist,* 204(4), 841-853.

Cochran, W.G. (1977). The estimation of sample size. *In:* John Wiley & Sons, Inc (eds) *Sampling techniques 3rd edition*. John Wiley & Sons, New York, pp. 72-85.

Comstock, J.H. & Needham, J.G. (1899). The wings of insects. *The American Naturalist*, 374(32), 81–89.

Coscarón, S. & Papavero, N. (1993). An Illustrated Manual for the Identification of the Neotropical Genera and Subgenera of Tabanidae (Diptera), Museu Paraense Emilio Goeldi, Coleção Emilie Snethlage, Belém, 150 pp.

Coscarón, S. & Papavero, N. (2009). Manual of Neotropical Diptera. Tabanidae. *Neotropical Diptera*, 6, 1–137.

Coscarón, S. & Papavero, N. (2014). Key to the known immature stages of Neotropical Tabanidae. *Neotropical Diptera*, *24*, 1–22.

Coscarón, S., Philip C. & Fairchild, G.B. (1979). Further notes on the Pangoniini of the austral region of South America (Diptera: Tabanidae). *Florida Entomologist*. 62: 301-304.

Costa-Schmidt, L.E. & de Araújo, A.M. (2010). Genitalic variation and taxonomic discrimination in the semi-aquatic spider genus *Paratrechalea* (Araneae: Trechaleidae). *The Journal of Arachnology*, 38(2), 242–249.

Dryden, I. (2018). Shapes package. R Foundation for Statistical Computing, Vienna, Austria. Contributed package. Version 1.2.4. Available from: http://www.R-project.org (Accessed 01 November 2018).

Dujardin, J.P. (2008). Morphometrics applied to medical entomology. *Infection, Genetics & Evolution*, 8, 875–890.

Dujardin, J.P., Le Pont, F. & Baylac, M. (2012). Geographical versus interspecific differentiation of sand flies (Diptera: Psychodidae): a landmark data analysis. *Bulletin of Entomological Research*, 93(01), 87-90.

Egan, A.N. (2015). Species delimitation and recognition in the *Pediomelum megalanthum* complex (Fabaceae) via multivariate morphometrics. *PhytoKeys*, 44, 65–87.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(7), 179–188.

Fraley, C. & Raftery, A.E. (2002). Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association*, 97(458), 611–631.

Fraley, C., Raftery, A.E. & Scrucca, L. (2018). Gaussian Mixture Modelling for Model-Based Clustering, Classification and Density Estimation. The R Journal. Available from: https://CRAN.R-project.org/package=mclust (Accessed 01 November 2018).

Gower, J. C. (1975). Generalized Procrustes Analysis. *Psychometrika*, 40(1), 33–51.

Gupta, O., Das, A.J., Hellerstein, J. & Raskar, R. (2018). Machine learning approaches for large scale classification of produce. *Scientific Reports*, 8(5226), 1-8.

Henriques, A.L., Krolow, T.K. & Rafael, J.A. (2012). Corrections and additions to Catalogue of Neotropical Diptera (Tabanidae) of Coscarón & Papavero (2009). *Revista Brasileira de Entomologia, 56*(3), 277–280.

Houle, D., Mezey, J., Galpern, P. & Carter, A. (2003). Automated measurement of *Drosophila* wings. *BMC Evolutionary Biology, 3*(1), 1-13.

Hudson, A. & Teskey, H.J. (1976). Morphological and biochemical characteristics of two forms of *Hybomitra typhus* (Diptera:Tabanidae). *The Canadian Entomologist*, 108, 737–740.

Ihaka, R., Murrell, P., Hornik, K., Fisher, J. C. & Zeileis, A. (2016). Colorspace: Color Space Manipulation. R package version 1.3-2. Avaliable from: https://CRAN.R-project.org/package=colorspace (Accessed 01 November 2018).

Kar, S.S. & Ramalingam, A. (2013). Is 30 the Magic Number ? Issues in Sample Size. *National Journal of Community Medicine*, *4*(1), 175–179.

Klingenberg, C.P., Badyaev, A.V., Sowry, S.M. & Beckwith, N.J. (2001). Inferring Developmental Modularity from Morphological Integration: Analysis of Individual Variation and Asymmetry in Bumblebee Wings. *The American Naturalist*, *157*(1), 11–23.

Kocovsky, P.M. (2009). The Effect of Sample Size on the Stability of Principal Components Analysis of Truss-Based Fish Morphometrics. *Transactions of the American Fisheries Society*, 138, 487–496.

Kucharczyk, H., Kucharczyk, M., Stanisławek, K. & Fedor, P. (2012). *Application of PCA in Taxonomy Research- Thrips (Insecta: Thysanoptera) as a Model Group*. *In*: Sanguansat, P. (ed) *Principal Component Analysis*, IntechOpen, Rijeka, pp. 111-126.

Lambkin, C.L., Sinclair, B.J., Pape, T., Courtney, G.W., Skevington, J.H., Meier, R., Yeates, D., Blagoderov, V. & Wiegmann, B.M. (2013). The phylogenetic relationships among infraorders and superfamilies of Diptera based on morphological evidence. *Systematic Entomology*, 38, 164–179.

Lessard, B.D., Cameron, S.L., Bayless, K.M., Wiegmann, B.M. & Yeates, D.K. (2013). The evolution and biogeography of the austral horse fly tribe Scionini (Diptera: Tabanidae: Pangoniinae) inferred from multiple mitochondrial and nuclear genes. *Molecular Phylogenetics and Evolution*, 68, 516–540.

Li, R., Li, M., Li, S. & Zhang, H. (2017). Further geometric morphometric analysis on the genus *Eysarcoris* (Hemiptera: Pentatomidae) from China. *Zoological Systematics*, *42*(4), 446–462.

Lorenz, C., Almeida, F., Almeida-Lopes, F., Louise, C., Pereira, S., Petersen, V., Vidal, P.O, Virginio, F. & Suesdek, L. (2017). Geometric morphometrics in mosquitoes : What has been measured ? *Infection, Genetics and Evolution*, *54*, 205–215.

Mackerras, I. M. (1954). The Classification and Distribution of Tabanidae (Diptera). I. General review. *Australian Journal of Zoology*, 39, 439- 511.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2018). Cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1. Retrieved from: http://cran.r-project.org/package=cluster. (Accessed 01 November 2018).

Maity, A., Naskar, A., Mukhopadhyay, E., Hazra, S., Sengupta, J., Ghosh, S. & Banerjee, D. (2015). Taxonomic studies on Tabanidae (Insecta : Diptera) from Himachal Pradesh, India. *International Journal of Fauna and Biological Studies*, *2(4)*, 43–52.

Mooney, C.J. & Kingsford, M.J. (2016). Statolith Morphometrics Can Discriminate among Taxa of Cubozoan Jellyfishes. *PLoS ONE*, *11(5)*, 1–16.

Morita, S.I., Bayless, K.M., Yeates, D.K. & Wiegmann, B.M. (2016). Molecular phylogeny of the horse flies: A framework for renewing tabanid taxonomy. *Systematic Entomology*, *41*, 56–72.

Morrone, J. (2014). Biogeographical regionalisation of the Neotropical region. *Zootaxa*, 3782, 1-110.

Murdoch, D. & Chow E.D. (2018). Functions for Drawing Ellipses and Ellipse-Like Confidence Regions. R package version 0.4.1. Retrieved from: http://cran.r-project.org/package=ellipse. (Accessed 01 November 2018).

Oguz, G., Kasap, O.E. & Alten, B. (2017). Wing morphology variations in a natural population of *Phlebotomus tobbi* Adler and Theodor 1930. *Journal of Vector Ecology*, 42(2), 223–232.

Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlin, D., Minchin, P.R., O´Hara, R.B; Simpson, G., Solymos, Peter; Stevens, H., Szoecs, E. & Wagner, H. (2018). Vegan: Community Ecology Package. R package version 2.5-3. Retrieved from: http://cran.r-project.org/package=vegan. (Accessed 01 November 2018).

Oldroyd, H. (1957). The Horse-flies (Diptera: Tabanidae) of the Ethiopian region. III. Subfamiles Chrysopsinae, Scepsidinae and Pangoniinae, and a revised classification. British Museum (*Natural History*), London, UK. 499 pp.

Osman, A., Zaki, M., Hamed, S. & Hussein, N. (2011). Numerical Taxonomic Study of Some Tribes of Gramineae from Egypt. *American Journal of Plant Sciences,* 2, 1–14.

Pape, T., Blagoderov, V. & Mostovski, M. B. (2011). Order Diptera Linnaeus, 1758. *In:* Zhang, Z.-Q. (Ed.) *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness*, Magnolia Press, Auckland, New Zealand, 222–229.

Pearson, K. (1901). On Lines and planes of Closest Fit to Systems of Points in Space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1, 559–572.

Perrard, A., Baylac, M., Carpenter, J. M. & Villemant, C. (2014). Evolution of wing shape in hornets: Why is the wing venation efficient for species identification? *Journal of Evolutionary Biology*, 27(12), 2665–2675.

Pretorius, E. & Scholtz, C. H. (2001). Geometric morphometrics and the analysis of higher taxa : a case study based on the metendosternite of the Scarabaeoidea (Coleoptera). *Biological Journal of the Linnean Society*, 74, 35–50.

Prudhomme, J., Cassan, C., Hide, M., Toty, C., Rahola, N., Vergnes, B., Dujardin, J.P., Alten, B., Sereno, D. & Bañuls, A.L. (2016). Ecology and morphological variations in wings of *Phlebotomus ariasi* (Diptera: Psychodidae) in the region of Roquedur (Gard, France): a geometric morphometrics approach. *Parasites and Vectors*, *9*, 1–13.

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.r-project.org/. (Accessed 01 November 2018).

Ren, J., Bai, M., Yang, X., Zhang, R., & Ge, S. (2017). Geometric morphometrics analysis of the hind wing of leaf beetles : proximal and distal parts are separate modules. *ZooKeys* (658), 131–149.

Rodrigues Guimarães, R., Storti Rodrigues, H.R. & Rodrigues Guimarães Júnior, R. (2017). Tabanids In South America. *In:* Shields, V (ed) *Insect Physiology and Ecology,* IntechOpen, Rijeka, pp. 55-111.

Rohlf, J. & Slice, D. (1990). Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Systematic Zoology*, 39(391), 40–59.

Ruane, S. (2015). Using geometric morphometrics for integrative taxonomy : an examination of head shapes of milksnakes (genus *Lampropeltis*). *Zoological Journal of the Linnean Society*, 174, 394–413.

Sánchez, G. (2013). DiscriMiner: Tools of the Trade for Discriminant Analysis. Available from: http://cran.r-project.org/package= DiscriMiner. (Accessed 01 November 2018).

Sofield, R.K., Douglas, M.E., Hansens, E.J. & Vrijenhoek, R.C. (1984). Diagnosis and detection of cryptic species: the *Tabanus nigrovittatus* complex (Diptera: Tabanidae) in Coastal New Jersey. *Annals of the Entomological Society of America*, 77(5), 587–591.

Sokal, R. & Michener, C. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 28, 1409–1438.

Sontigun, N., Sukontason, K.L., Zajac, B.K., Zehner, R., Sukontason, K., Wannasan, A. & Amendt, J. (2017). Wing morphometrics as a tool in species identification of forensically important blow flies of Thailand. *Parasites and Vectors*, 10(1), 229.

Sutton, B. D. & Carlson, D. A. (1997). Cuticular hydrocarbon variation in the Tabanidae (Diptera): *Tabanus nigrovittatus* complex of the North American Atlantic coast. *Annals of the Entomological Society of America*, 90(5), 542–549.

Thompson, D.W. (1917). *On Growth and Form*. Cambridge University Press, Cambridge 793 pp.

Torres, A. & Miranda-Esquivel, D.R. (2016). Wing Shape Variation in the Taxonomic Recognition of Species of Diachlorus Osten-Sacken (Diptera: Tabanidae) from Colombia. *Neotropical Entomology*, *45*(2), 180–191.

Usher, P.J. (1972). A review of the South African horsefly fauna (Diptera: Tabanidae). *Annals of the Natal Museum*, 21(2), 459–507.

Wickham, H. (2016). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. New York, USA: Springer-Verlag. Available from: http://cran.r-project.org/package=ggplot2. (Accessed 01 November 2018).

Wiegmann, B.M., Tsaur, S.C., Webb, D.W., Yeates, D.K. & Cassel, B.K. (2000). Monophyly and Relationships of the Tabanomorpha (Diptera: Brachycera) Based on 28S Ribosomal Gene Sequences. *Annals of the Entomological Society of America*, 93(5), 1031–1038.

Wilkerson, R.C. (1978). Horse Flies (Diptera: Tabanidae) of the Colombian Departments of Chocó, Valle and Cauca. University of Florida pp. 974.

Wolff, M. & Miranda-Esquivel, D. R. (2016). Family Tabanidae. *Zootaxa*, *362*(7), 249–301.

List of appendices

(See attached appendices on the CD. You can find them at the database of UIS library)

Apendix A. Sample localities of Tabaninae species. Map modified from Morrone (2014).

Apendix B. Nomenclature of landmarks and position on wing veins.

Apendix C. Dendrogram of Tabaninae through hierarchical clustering under UPGMA method.

Apendix D. Principal Component Analysis of tribes Diachlorini and Tabanini.

Apendix E. Principal Component Analysis of groups of species with 28 landmarks.

Apendix F. Principal Component Analysis of groups of species with 29 landmarks.

Apendix G. Principal Component Analysis of groups of species with 30 landmarks.

Apendix H. Sampling localities for Tabaninae species used in Geomtric Morphometric analysis.

Apendix I. List of keys used for taxonomic determination.

Apendix J. Number of specimens used for Geometric Morphometrics analyses at genus level.

Apendix K. Groups of species based on the number of landmarks.

Apendix L. Discrimination scores for each tribe.

Apendix M. Discrimination scores for genera.

Apendix N. Discrimination scores for groups of species based on the number of landmarks.