



Comparing classical and geometric morphometric methods to discriminate between the South African honey bee subspecies *Apis mellifera scutellata* and *Apis mellifera capensis* (Hymenoptera: Apidae)

Tomas BUSTAMANTE¹, Benjamin BAISER², James D. ELLIS¹

¹Entomology and Nematology Department, University of Florida, Gainesville, FL 32611, USA

²Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL 32611, USA

Received 28 June 2018 – Revised 11 February 2019 – Accepted 26 April 2019

Abstract – There are two endemic subspecies of western honey bees (*Apis mellifera* L.) in the Republic of South Africa (RSA), *A.m. capensis* and *A.m. scutellata*. They have traditionally been identified using morphometric characteristics, but geometric morphometric data from honey bee wings are easier to collect, possibly making them a useful alternative for identifying these subspecies. We compared the accuracy of both morphometric and geometric morphometric methods using linear discriminant and classification and regression tree analyses. We found that using geometric wing shape data from both forewings and hindwings resulted in a lower classification accuracy (73.7%) than did using models derived from the full set of standard morphometric data (97% accurate) in cross-validation. Tergite color and average ovariole number were the most important features for discriminating between the two subspecies. Finally, we used Kreiger interpolation to construct maps illustrating probable distributions of *A.m. capensis* and *A.m. scutellata* in the RSA.

Apis mellifera capensis / *Apis mellifera scutellata* / Morphometrics / Wing geometry / Discrimination

1. INTRODUCTION

Apis mellifera scutellata (Lepeletier 1836), the African or Savannah honey bee, and *Apis mellifera capensis* (Eschscholtz 1822), the Cape honey bee, are two western honey bee subspecies native to the Republic of South Africa (RSA) (Hepburn 2002) where their natural distributions overlap to create a stable hybrid zone (Crewe et al. 1994; Hepburn et al.

1998). These two subspecies are considered pests in some regions where they are not endemic (Allsopp 2003; Beekman et al. 2012; Dietemann et al. 2006; Schneider et al. 2003).

Apis mellifera capensis is native to the Fynbos ecoregion of RSA (Du Preez 2014). Its natural distribution is bound by mountain ranges and the natural distribution of *A.m. scutellata* to the north (Hepburn and Crewe 1991) and the Atlantic and Indian Oceans to the south. *Apis mellifera capensis* possess a unique phenotype whereby workers can produce diploid female offspring without mating (Onions 1912; Lattorff et al. 2005). This trait, known as thelytoky, enables *A.m. capensis* workers to act as social parasites when introduced into apiaries housing colonies of other honey bee subspecies (Velthuis et al. 1990). In this scenario, *A.m. capensis*

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13592-019-00651-6>) contains supplementary material, which is available to authorized users.

Corresponding author: T. Bustamante, g4spar@ufl.edu;
J. Ellis, jdellis@ufl.edu
Handling editor: David Tarpyjdellis@ufl.edu

workers can drift to other colonies where they reproduce and drain colony resources without benefiting the host (Hepburn and Allsopp 1994; Greeff 1997). Host colonies eventually die, at which point the remaining *A.m. capensis* workers disperse to other colonies (Allsopp 1998; Neumann et al. 2001). The ability of *A.m. capensis* to parasitize colonies of other subspecies of honey bees makes it a potential threat to manage honey bee populations outside of its native range.

Apis mellifera scutellata is native to south and south-central Africa, where it is the primary subspecies used for honey production because it outcompetes other introduced subspecies in that region (Johannsmeier 2001). It was introduced into Brazil in the late 1950s in an effort to increase honey production there (Schneider et al. 2003). Since that time, the feral population (called “African,” “Africanized,” or “killer” honey bees) has spread throughout South and Central America and into the southwestern United States of America (USA) and parts of FL, USA. It has become the subspecies of choice for beekeepers in South and Central America (Schneider et al. 2003). In these areas, the bees out-compete European subspecies as producers of honey and providers of pollination services (Winston 1993). *Apis mellifera scutellata* is considered an invasive pest in North America, however, because of its pronounced defensive behavior and general lack of productivity in temperate climates (Schneider et al. 2003; Guzman-Novoa and Page 1994). In North America, extensive efforts have been made to identify, monitor, and control the use of this subspecies in managed colonies (Rinderer et al. 1993, 1983).

Monitoring for either of these two subspecies is problematic in regions where they co-occur because they are difficult to distinguish from one another qualitatively (Ruttner 1988). Both bee subspecies have been classified historically using morphometric techniques outlined by Ruttner (1988), hereafter referred “standard morphometrics.” Standard morphometrics involves quantifying various external features of each specimen, such as wing length or number of hamuli, and using those measurements in multivariate analyses to identify populations of the bees. Some internal features, namely ovariole number and spermathecal size, also have been used in comparative studies of *A.m. capensis* and *A.m. scutellata*. Ruttner used principal component

analysis and clustering techniques to analyze the standard morphometrics data and establish two separate populations in RSA (Ruttner 1988). Hepburn et al. (1998) went further by using jackknifed linear discriminant analysis to estimate probability scores for each colony in order to classify some colonies as hybrids of the two subspecies.

Developing models that can discriminate between these two honey bee subspecies using standard morphometric data is laborious and time consuming due to the challenges associated with dissecting and measuring the various anatomical features of the bees. The purpose of our study was to attempt to overcome these difficulties using geometric morphometric data collected from wing vein intersections as landmarks (wing geometry).

Wing geometry has been used before to discriminate between several other subspecies of honey bees, namely *A.m. intermissa*, *A.m. sahariensis*, *A.m. carnica*, *A.m. ligustica*, *A.m. mellifera*, *A.m. scutellata*, *A.m. carnica*, *A.m. caucasica*, and Africanized honey bees (Barour and Baylac 2016; Francoy et al. 2008; Tofilski 2008). This approach works by laying morphologically common characteristics, forewings in this case, on a Cartesian plane. The resulting x – y coordinates of common landmarks, such as the wing vein intersections, are used in lieu of traditional morphometric data after being rotated, resized and aligned (Slice 2007). Both types of morphometric data, standard and wing geometry, can be analyzed similarly and used for subspecies identification purposes (Meixner et al. 2013).

Linear discriminant analysis (LDA) and classification and regression tree (CART) analysis are two ways of analyzing morphometric data that were tested in the present study. LDA produces a function, or equation, that can be used to give each colony a score based on the morphometric data measured. The benefit of LDA is that the score given by a discriminant function can be converted into a probability score showing the likelihood that a sample belongs to a given group, allowing the identification of specimens that fall between groups. LDA was used to develop the Universal System for Detecting Africanization Through Identification Based on Morphology (USDA-ID) and the Fast Africanized Bee Identification System (FABIS), two methods used to discriminate

Table I. Sampling information for each apiary in the Republic of South Africa from which *A.m. scutellata* and *A.m. capensis* samples were collected

City number	City	Sample size of apiaries	Sample size of colonies	Collection date
1	Beaufort West	3	9	5 May 2014
2	Bloemfontein	3	9	2 May 2014
3	Bredasdorp	3	9	2 May 2014
4	Cape Town	3	9	10 Jan 12, 8 Jan 12, 3 May 12
5	Citrusdaal	3	9	5 Apr 2014
6	East London	3	9	23 Apr 13, 24 Apr 13
7	George	3	9	30 Apr 2014
8	Graaff-Reinet	3	9	4 May 2014
9	Grahamstown	2	6	27 Apr 2013
10	Klawer	3	9	8 May 2014
11	Knysna	2	6	1 May 2013
12	Kroonstad	3	9	1 May 2014
13	Laingsburg	3	9	19 Apr 2014
14	Langebaan	3	9	27 Apr 2013
15	Modderfontein	1	3	28 Apr 2013
16	Moorreesburg	3	9	22 Mar 2014
17	Oudtshoorn	3	9	29 Apr 2014
18	Plettenburg Bay	2	6	30 Apr 2013
19	Port Elizabeth	1	3	29 Apr 2013
20	Pretoria	2	6	30 Apr 2014
21	Riversdale	3	9	1 May 2014
22	Springbok	3	9	8 May 2014
23	St. Francis	1	3	30 Apr 2013
24	Stellenbosch	3	9	28 Apr 2013
25	Swellendam	3	9	2 May 2014
26	Touwsriver	3	9	18 Apr 2014
27	Uppington	3	9	9 May 14, 10 May 14
28	Vryburg	1	3	11 May 2014
29	Worcester	3	9	5 May 2014
	Total	75	225	

between *A.m. scutellata*-derived and European-derived subspecies of *A. mellifera* in the Americas (Rinderer et al. 1986, 1993). Characteristics can be selected by repeating the analysis while adding and removing them in a stepwise fashion. This method is known as stepwise LDA.

CART analysis offers a nonparametric alternative to LDA. It works by following an algorithm that creates a binary tree where each node represents a measurement threshold for one morphometric feature. Decision trees produced by CART analysis

identify the most important variables for classifying specimens and can be used to classify new samples.

Here, we compared the ability of wing geometry and standard morphometric data to predict the identity of honey bees collected from RSA using LDA and CART analyses. For the sake of clarity, we refer to each combination of data and analysis as a different “method” herein. We will use the term “identification model” when generally referring to either a discriminant function produced by LDA or a classification tree produced by CART.

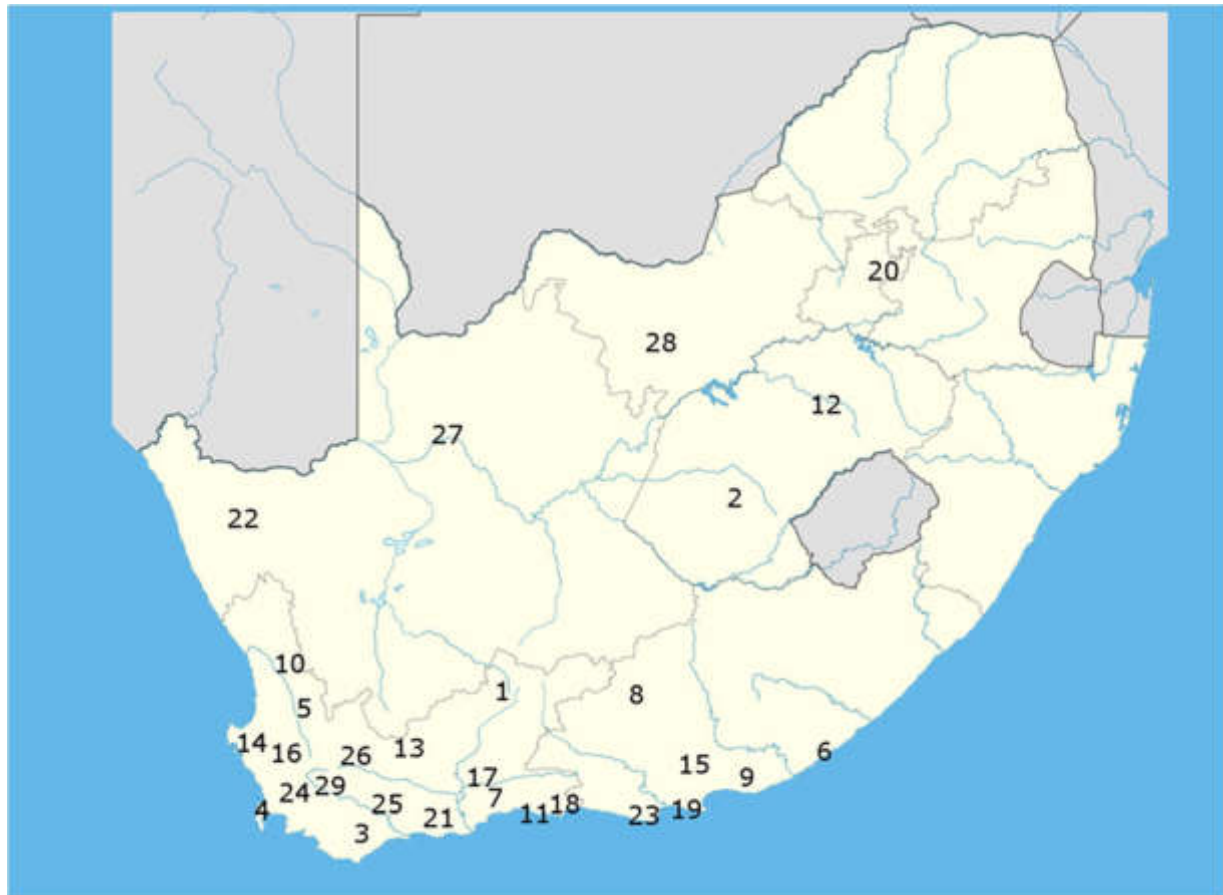


Fig. 1 Location of the cities in the Republic of South Africa where the honey bee samples were collected. Numbers correspond to city numbers shown in Table I. Base map downloaded from the Global Administrative Areas database (gadm.org).

Ultimately, we hoped to select a method that could be used to predict the identity of honey bee workers from RSA accurately and efficiently.

2. MATERIALS AND METHODS

For this study, it was necessary to (1) collect samples of honey bees from managed colonies in RSA, (2) analyze standard morphometric data from the colony samples to classify each colony as *A.m. capensis* or *A.m. scutellata*, (3) randomly select colonies of both subspecies as training samples to use to create identification models for the different methods, (4) test the models by using them to predict the subspecies identity of the remaining colonies (those identified in step 2 but not used for the identification models in step 3), and (5) compare the average accuracy of predictions across each method. Additionally, geospatial analysis software was used to create a map showing the average probability of a honey bee colony

being identified as *A.m. capensis* when sampled randomly from within RSA.

2.1. Collecting honey bee samples

We collected 50–75 worker honey bees from beekeeper-managed colonies in or near 29 cities in RSA over two collection trips (collection 1: April/May 2013; collection 2: May, 2014; Table I, Figure 1). The bees were collected from hive frames into 50-ml vials containing 100% ethanol. Three colony samples were analyzed from each apiary. Ten bees were used from each colony sample. Thus, $10 \text{ bees/colony} \times 3 \text{ colonies/apiary} \times 75 \text{ apiaries}$ were used in the study for a total of 225 colonies or 2250 bees.

2.2. Classifying samples

Ruttner (1988) noted which morphometric features are used to distinguish between the

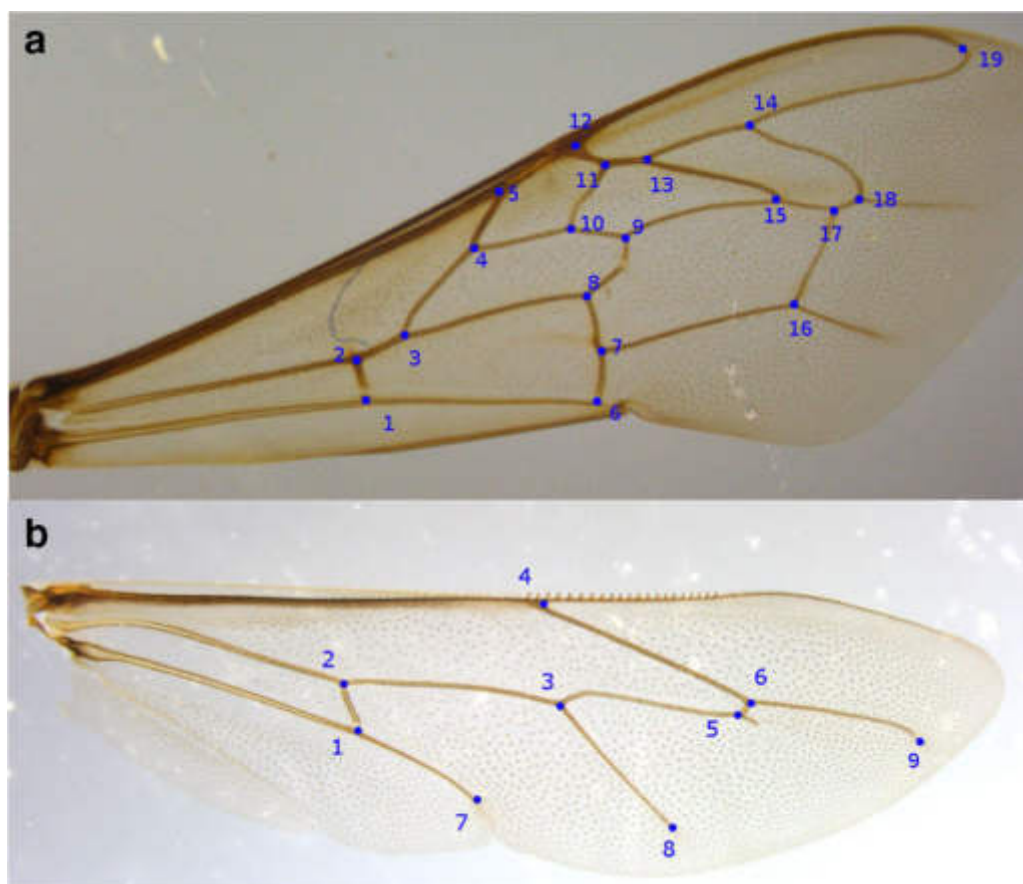


Fig. 2 Geometric morphometric landmarks on the right forewing (a) and hindwing (b).

A. mellifera subspecies. Hepburn et al. (1998) showed that only eight of the Ruttner features are necessary for distinguishing between *A.m. capensis* and *A.m. scutellata* colonies. These features include (1) the length of cover hair on tergite A6, (2) the width of the wax plate on sternite A4, (3) the transverse length of the wax plate on sternite A4, (4) the pigmentation of the scutellum, (5) the pigmentation of the scutellar plate, (6) the pigmentation of tergite A3, (7) wing angle N23, and (8) wing angle O26. Forewing length and number of ovarioles per ovary were used as ninth and tenth features here due to their reported discriminatory power or use in current honey bee pest subspecies monitoring programs respectively (Phiancharoen et al. 2010; Rinderer et al. 1986). Coordinates of forewing vein intersections (19 points, Smith et al. 1997; Francoy et al. 2008, Figure 2a) and hindwing vein intersection and end-points (i.e., hindwing landmarks, nine points, Figure 2b) were also used. The end-points were used on the hind wings because they were easily identifiable and added more shape information. Furthermore, we collected data on

the diameter of spermathecae in all dissected workers (Carreck et al. 2013). While we measured 119 spermathecae in the sampled *A.m. capensis* workers, too few were measured from *A.m. scutellata* workers ($N = 3$); so, they were not included in any analysis.

Photographs of all features for each bee were taken using a Leica M205 microscope with a Leica MC170 camera and Leica Application Suite software. Images were 1600×1200 pixels. The magnification being used varied by feature. All of the bees were dissected to facilitate image collection (Carreck et al. 2013). Forewings were removed proximal to the sagittal hook. Hindwings also were removed as close to the base as possible. Scutellum, scutellar plate, and abdominal tergite hair images were taken after removing wings from the thorax. Abdominal hair images were taken such that the target hair was orthogonal to the camera view and its base was visible. Abdominal tergite A3 was removed from each bee, cleaned of tissue using a paintbrush, and cut in half to facilitate flattening it under a slide cover. One ovary was removed from each bee and placed on a well



Fig. 3 The results from a principle component analysis and k-means clustering based on standard morphometrics data, including ovariole number and wing length, collected from dissected bees. K-means clustering defined two groups, labeled *Apis mellifera capensis* (cap) and *A.m. scutellata* (scut).

slide. Neither the left nor the right ovaries were favored. The ovarioles were separated using minuten insect pins and counted. A picture of each ovary was taken for validation purposes. Spermathecae also were dissected out of the bee when present, placed on a well slide, and imaged. Abdominal sternite A4 was removed, cleaned with a paintbrush and soaked in KOH to remove any remaining tissue.

Forewings, hindwings, sternites, and tergites were mounted on 25 × 75 mm Fisher glass slides (S17466A) for imaging. They air-dried on a covered Kim-tech wipe prior to mounting. Sternites were stained with Bioquip double stain (6379B), dabbed against Kim-tech wipes and washed in 90% ethanol to remove the excess stain. All of these parts were mounted in Euparal Mounting Media from Bioquip (6372B). Concave tergite pieces were inverted to keep them from folding

in on themselves under the cover glass. No. 1 cover glasses from Fisher (12-518-105H) were used. Finished slides were warmed at 60 °C for 3 days on a slide warmer (models: Premiere xh-2004 or C.S. & E. 26020). Each anatomical feature was imaged individually.

Custom-built, assistive measuring software was used to measure wax plate length, wax plate width, wing length, spermatheca diameter, and hair length of the abdominal tergite from images. Pixel coordinates of wing venation intersection points in images also were recorded using the same software. A known length of 1 mm was digitized across several images to estimate error that may affect landmark digitization precision. A rate of .04% was measured. Wing angles (N23 and O26) were calculated based on the associated intersection points. The color of the abdominal tergite, scutellum and scutellar plate was ranked

Table II. Mean values for each morphometric characteristic used in the multivariate analysis of honey bees collected from the Republic of South Africa. Data are mean $\pm$ s.e. ($\pm$ 95% CI). The means were compared using Bonferroni corrected t tests. An asterisk (*) in the final column indicates the row means for both subspecies are significantly different from one another at $\alpha \leq 0.05$. ns = not significant. Spermatheca diameter data were not analyzed because only three *A.m. scutellata* spermatheca were collected (Na)

Statistic	<i>A. m. capensis</i> ($N = 1670$)	<i>A. m. scutellata</i> ($N = 580$)	Significance ($p < 0.005$)
Ovariole number	12.4 $\pm$ 0.3 (0.5)	5.6 $\pm$ 0.3 (0.5)	*
Abdomen hair (mm)	0.2 $\pm$ 0.002 (0.004)	0.2 $\pm$ 0.003 (0.007)	*
Scutellar plate	2.7 $\pm$ 0.1 (0.2)	7.3 $\pm$ 0.1 (0.2)	*
Scutellum	0.9 $\pm$ 0.03 (0.06)	1.1 $\pm$ 0.05 (0.09)	*
Sternite width (mm)	2.1 $\pm$ 0.006 (0.01)	2.2 $\pm$ 0.008 (0.02)	ns
Sternite length (mm)	1.2 $\pm$ 0.005 (0.009)	1.2 $\pm$ 0.006 (0.01)	ns
Tergite	5.5 $\pm$ 0.07 (0.1)	8.2 $\pm$ 0.07 (0.1)	*
Angle N23	23.6 $\pm$ 0.04 (0.07)	23.7 $\pm$ 0.07 (0.1)	ns
Angle O26	57.4 $\pm$ 0.2 (0.4)	56.5 $\pm$ 0.3 (0.5)	ns
Forewing length (mm)	8.5 $\pm$ 0.01 (0.02)	8.7 $\pm$ 0.02 (0.04)	*
Spermatheca diameter (mm)	0.4 $\pm$ 0.02 (0.05), $N = 119$	0.2 $\pm$ 0.02 (0.1), $N = 3$	Na

per the qualitative index produced by Ruttner (1988). Each tergite, scutellar plate, and scutellum color was ranked by three different people. The average rank of each was used for analysis to eliminate any bias one person may have due to the qualitative nature of the index.

Data were analyzed using R (version: 3.2.1). Generalized Procrustes alignment (from package “geomorph,” Adams and Otárola-Castillo 2013) was used to align all forewing points with one another and all hindwing points with one another. Non-geometric numeric data for bees were averaged for each colony ($N = 225$) before analysis (Hepburn et al. 1998).

A reference population was established, giving a cohort of bees to which the results of each method could be compared. This was done in a similar fashion to the population structure analyses of Hepburn et al. (1998). Principal component analysis, through the “princomp” function in R, combined with K-means clustering, via the “cascadeKM” function (Vegan package, Oksanen et al. 2016), grouped samples into morphoclusters based on standard morphometric data (the ten features mentioned above). A Calinski criterion was used to choose the optimal number of groups (Calinski and Harabasz 1974).

A multiple response permutation procedure (MRPP, function mrpp from Vegan package, Oksanen et al. 2016) was used to test for significance between groups. Groups were assigned to a subspecies based on geographic location. LDA with jackknife leave-one-out cross validation was used to generate posterior probability scores (i.e., percent chance *A.m. capensis* or percent chance *A.m. scutellata*) for each colony based on principal components (MASS package, Venables and Ripley 2002). Differences in the averages between *A.m. capensis* and *A.m. scutellata* were tested for each measured characteristic using Bonferroni corrected t tests.

2.3. Testing methods for identification

Combinations of five different subsets of data and two types of analysis were evaluated as methods for their ability to discriminate between the two subspecies. The data sets included (1) standard morphometric data with ovariole number, (2) standard morphometric data alone (no ovarioles), (3) ovariole number alone (ovarioles only), (4) forewing length alone (forewing only), and (5) geometric morphometric (wing shape) forewing and hindwing data. LDA and CART

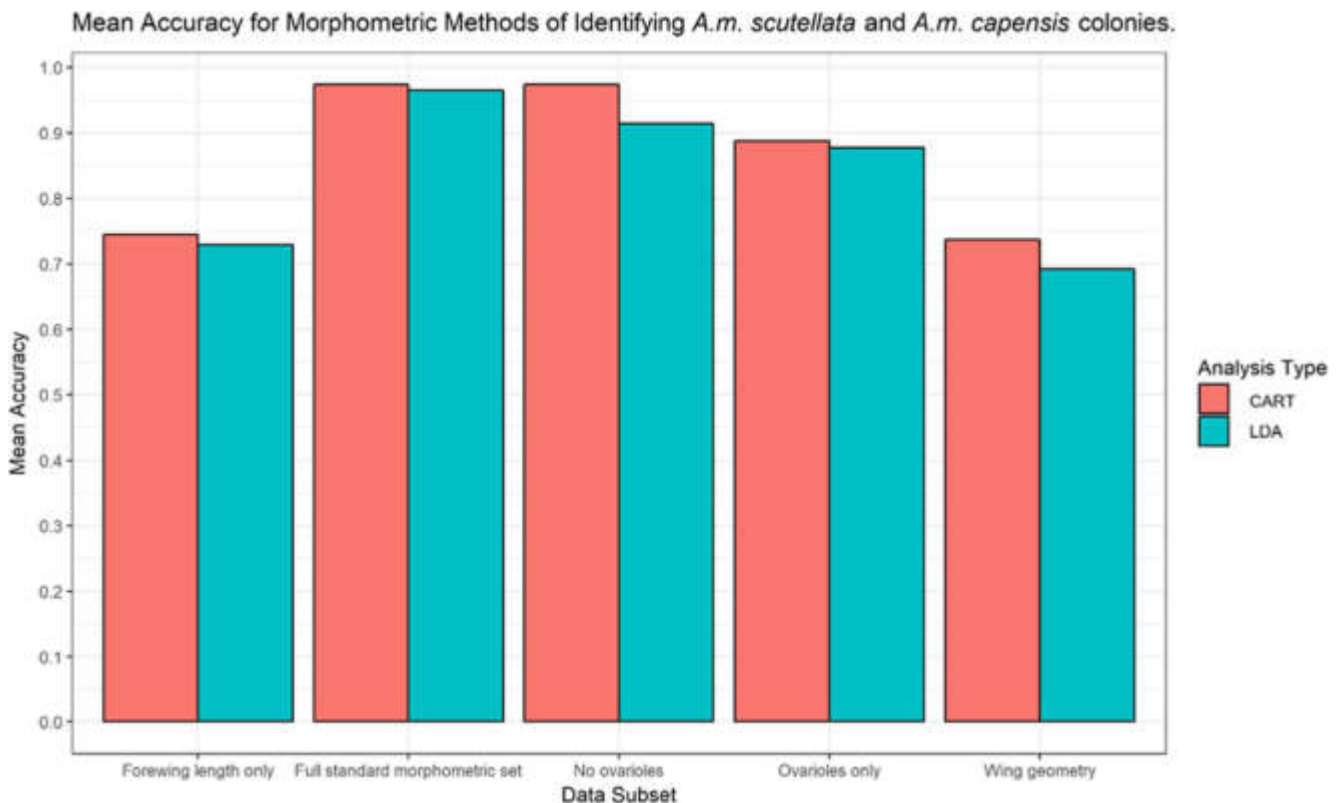


Fig. 4 The mean accuracy (proportion) for each method tested.

analysis were both used to analyze each data set. Thus, ten different methods for identification, each consisting of one data subset and one analysis, were compared in total.

All data used for identification were based on colony means. A holdout method was used to test accuracy. Half of the colony samples for each subspecies ($N = 112$) were selected from the total sample set ($N = 225$) to be used as the “training” sample set. The remaining samples were used as a “test” set. Either LDA or CART analysis was used to create an identification model using the training samples. The *corrplot* package was used to find correlation in morphometric data (Wei and Simko 2016). Scutellum color data were excluded from all LDA methods because it had a high Pearson’s correlation coefficient (> 0.6) with tergite color, making it inappropriate for LDA. Forewing coordinate data X2, Y2, Y4, Y5, X6, Y6, X7, Y7, X11, Y11, X13, Y13, X14, Y17, and hindwing coordinate data X4, X5, X6, X9, and Y9 were excluded from LDA methods for the same reasons. Five outlier colonies were removed when checking for data normality. The data met the assumptions for LDA otherwise. Final

characteristics for LDA were chosen in a stepwise manner using the *stepclass* method from the *klaR* library.

The “*rpart*” package (Therneau et al. 2015) was used to build classification trees. Each tree resulting from the CART analysis was pruned to the minimum number of nodes that was within one standard error of the minimum estimated error based on a V-fold cross validation analysis.

The percent classification accuracy for each model was calculated by dividing the number of correctly identified honey bees by the total number of honey bees in the “test” set. The holdout method was repeated 1000 times per method to estimate an average percent accuracy for each method (Kohavi 1995).

2.4. Creating an interpolation map

ArcMap (10.3.1) was used to create a simple Kriger interpolation map of subspecies distribution based on the average colony posterior probability scores of each apiary from the jackknifed LDA analysis. Parameters (nugget, etc.) for the analysis were chosen using the iterative cross-

Table III. Raw and standard coefficients for the most accurate discriminant functions for the three most accurate methods of discriminating between *Apis mellifera capensis* and *A.m. scutellata* tested (numbered 1–3)

Raw coefficients			
(1) Full morphometric data set		(2) Excluding ovarioles	
Tergite	− 0.9776053	Tergite	1.128432
Ovariole number	0.1339441		
		(3) Ovarioles only	
		Ovariole number	0.320683
Standardized coefficients			
(1) Full morphometric data set		(2) Excluding ovarioles	
Tergite	− 0.8647	Tergite	− 0.3652
Ovariole number	0.4670		
		(3) Ovarioles only	
		Ovariole number	0.3038

validation method provided by ArcMap. Posterior probability scores were arc-sin transformed prior to analysis.

3. RESULTS

The reference population consisted of two morphoclusters (Figure 3). MRPP confirmed a difference between these two groups ($A = 0.64$, $P = 0.001$). The northernmost group was classified as *A.m. scutellata* and the southernmost as *A.m. capensis*. Of the 225 colonies analyzed, 166 were classified as *A.m. capensis* and 59 were classified as *A.m. scutellata*.

Sternite height, sternite width, wing angle O26, and wing angle N23 were not significantly different between both bee subspecies while the remaining traits were (Table II). According to our results, *A.m. capensis* workers had more ovaries, shorter abdomen hair, darker coloration, and shorter forewings than did *A.m. scutellata* workers (Table II).

Tergite color and ovariole number had the highest absolute value for standardized coefficients in a linear discriminant analysis containing all morphometric traits. These were − 1.2 and 0.68, respectively.

Mean accuracy of each method are illustrated in Figure 4. Here, we define “accuracy” as the percentage of bees classified into the correct subspecies. Models using the full set of standard

morphometric data were the most accurate on average, with the LDA being 96.6% accurate and CART being 97.4% accurate, on average. In comparison, models using forewing and hindwing geometry were less accurate, with 69.7% for LDA and 73.3% for CART on average. The average accuracy of the LDA with the number of ovarioles only and forewing length only and CART with the number of ovarioles only and forewing length were, 87.77%, 72.92%, 88.74%, and 74.48%, respectively. The most accurate classification tree using ovariole data alone consisted of a single threshold of 8.7 ovarioles per ovary per bee and was 94.7% accurate in discriminating subspecies. Raw and standardized coefficients for the most accurate discriminant functions of the three best methods using LDA are noted in Table III. The most accurate classification trees produced by CART consisted of a single node using only tergite color. The most accurate trees produced by the three best methods using CART analysis are shown in Figure 5.

The geographic distribution of *A.m. capensis* predicted by Kriger interpolation of the average posterior probabilities for colonies in each apiary is illustrated in Figure 6. A likely hybrid zone similar to the one illustrated by Hepburn et al. (1998) is shown. The eastern part of the hybrid zone shown in Figure 6, however, is located further north than is the one of Hepburn et al. (1998).

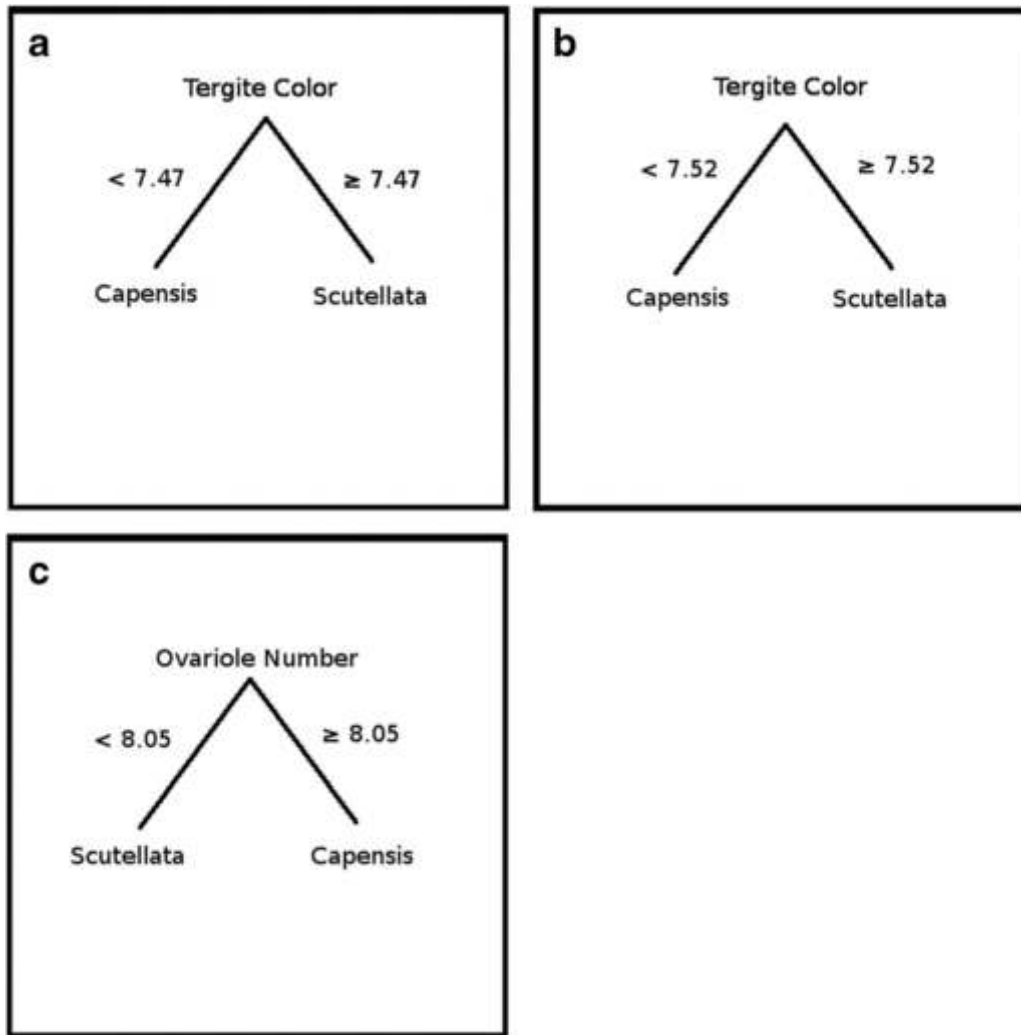


Fig. 5 The most accurate identification models created by the three most accurate CART methods. These include the full morphometric data set (**a**), full morphometric data set without ovariole count (**b**), and ovariole count only (**c**).

4. DISCUSSION

Identification models using the full set of standard morphometric characteristics were the most accurate on average and identified about 97% of the worker bees correctly for both LDA and CART on average. This was expected since these morphometric features were used to classify the original reference population. The most accurate classification trees produced by these methods consisted of a single node using only tergite color.

The most accurate classification tree using ovariole data alone consisted of a single threshold of about nine ovarioles, which is higher than that reportedly used by Ruttner (five ovarioles per ovary; Hepburn and Radloff 2002). Ovariole number is believed to serve as a proxy for thelytoky (Ruttner 1988; Hepburn et al. 1998; Hepburn and Radloff 2002), which could also

make it an important characteristic for monitoring for that trait specifically. There are practical drawbacks that should be considered when using ovariole number. First, collecting these data require a large amount of initial practice before one can extract ovaries from bees quickly and reliably. However, once the initial training period has passed, it is much simpler and quicker to dissect and count ovarioles alone than collecting the other morphometric features. Second, ovarioles are more sensitive to degradation than are the external features. Thus, samples must be processed and stored more carefully. The bees used in this study were stored in 100% ethanol, but some still showed signs of internal degradation and could not be used for the analysis. Alternatively, freshly collected bees could be used. One should collect more bees from a colony than are needed for the analysis so spare

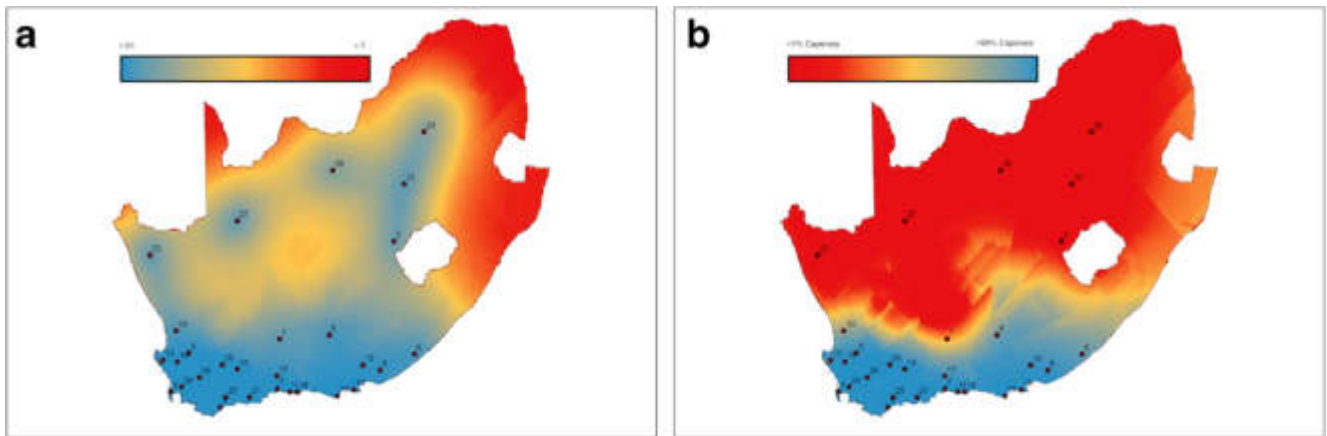


Fig. 6 Predicted percent chance colonies in an area are *Apis mellifera capensis* (**a**, with bluer colors indicating a higher percent chance the sampled colonies would be *A.m. capensis* and redder colors indicating a lower one) and predicted error (**b**, with bluer colors indicating low uncertainty and yellow to redder colors a higher one) of Kriger analysis based on multivariate analysis of Hepburn (1998) characteristics. For example, there is a high probability that sampled colonies in red areas (left image) would be *A.m. scutellata* rather than *A.m. capensis*. The numbers on the map correspond to the city numbers shown in Table 1 and Fig. 1.

samples are available should more data be necessary.

The identification models created using exclusively forewing and hindwing shape correctly identified the test samples 70% (with LDA) and 73% (with CART) of the time. This is significantly lower than the predictive power of the models based on the full set of morphometric characteristics or just ovarioles alone. The low accuracy of the classification techniques using these data suggests that *A.m. capensis* and *A.m. scutellata* wing geometry is too similar to be used in methods developed to discriminate between the two subspecies. The results may have also been affected by colonies that were collected in the hybrid zone. Wing geometry could still prove useful for monitoring *A.m. capensis* in areas where *A.m. scutellata* is invasive. The invasive *A.m. scutellata* populations in North and South America may not necessarily be morphometrically the same as the original *A.m. scutellata* population in RSA (Bucu et al. 1987). Comparisons between Africanized honey bees collected in 1968 and 2002 indicate a change in wing shape over time (Francoy et al. 2009).

It is important to note that our results with wing geometry are based on the use of wing vein intersection points and a few other landmarks to describe the wing shape. There are other methods of describing geometry that may have more

predictive capabilities. For example, relative warp analysis based on the outline of the cells created by veins has been useful in the geometric analysis of other bee species (Francoy et al. 2012; Mendez et al. 2007). Including the centroids sizes in analyses has increased the success of using geometric techniques to discriminate between some European and African subspecies of honey bees (Tofilski 2008; Barour et al. 2011; Barour and Baylac 2016; Oleksa and Tofilski 2015). Additionally, Barour and Baylac (2016) present a slightly different set of landmarks for the hindwing. Some methods have led to the development of web application (Baylac et al. 2008). When comparing these methods, it should be noted that while there is general consensus about where landmarks are placed, there are some differences in landmark numeration (Miguel et al. 2011; Barour and Baylac 2016; Francoy et al. 2008; Kandemir et al. 2011). More advanced methods of classifying images, such as using artificial intelligence (AI) neural networks, also may potentially be useful because of AI's ability to solve difficult pattern recognition problems (Rowley et al. 1998; Egmont-Peterson et al. 2002). For the time being, however, our results suggest that one should use either LDA or CART analyses with the full set of morphometric data or simply ovariole count to discriminate between *A.m. capensis* and *A.m. scutellata*.

ACKNOWLEDGMENTS

The authors would like to thank the current and former members of the University of Florida Honey Bee Research and Extension Laboratory who collected honey bee samples across the Republic of South Africa. This includes Daniel Schmehl, Ashley Mortenson, and Mark Dykes. We graciously acknowledge Mike Allsopp (ARC-Plant Protection Research Institute, South Africa), Christian Pirk (University of Pretoria, South Africa), Garth Cambray (beekeeper), and Eddie Hart (beekeeper) for the assistance they provided in coordinating field sample collections and/or providing samples and the South African beekeepers who allowed us to sample their colonies. We also thank Ekaterina Bustamante, Joe Cicero, Siria Gamez, Quiana Vidal, William Brandon Jones, James Snyder, Vince and Cynthia Alderman, Matt Johnson, Brandi Simmons, and Andrew Grimm for the technical assistance they provided.

AUTHORS' CONTRIBUTION

TB, BB, and JE conceived this research and designed experiments; TB performed experiments and analysis; TB, BB, and JE wrote the paper and participated in the revisions of it. All authors read and approved the final manuscript. Funding information

This project was financed through a cooperative agreement provided by the United States Department of Agriculture, Animal and Plant Health Inspection Service (USDA-APHIS, Projects 14-8130-0414-CA and 16-8130-0414-CA) and by the Florida Department of Agriculture and Consumer Services through the guidance of the Honey Bee Technical Council.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest The authors declare that they have no potential conflict of interest.

Comparaison des méthodes morphométriques classique et géométrique pour distinguer les sous-espèces d'abeilles sud-africaine *Apis mellifera scutellata* et *Apis mellifera capensis* (Hymenoptera: Apidae).

***Apis mellifera capensis* / *Apis mellifera scutellata* / morphométrie / géométrie des ailes**

Ein Vergleich klassischer und geometrischer morphometrischer Methoden zur Unterscheidung zwischen der südafrikanischen Unterart *Apis mellifera scutellata* und *Apis mellifera capensis* (Hymenoptera: Apidae)

***Apis mellifera capensis* / *Apis mellifera scutellata* / Morphometrie / Flügelgeometrie**

REFERENCES

- Adams, D.C., Otarola-Castillo, E. (2013). geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* 4, 393–399.
- Allsopp, M. (2003) Cape honeybee (*Apis mellifera capensis* Eshscholtz) and varroa mite (*Varroa destructor* Anderson & Trueman) threats to honeybees and beekeeping in Africa. *Int. J. Trop. Insect Sci.* 24, 87–94. <https://doi.org/10.1079/IJT20041>
- Barour, C., Baylac, M. (2016). Geometric morphometric discrimination of the three African honeybee subspecies *Apis mellifera intermissa*, *A. m. sahariensis* and *A. m. capensis* (Hymenoptera, Apidae): Fore wing and hind wing landmark configurations. *J. Hymenopt. Res.*, 52(52)
- Barour, C., Tahar, A., Baylac, M. (2011). Forewing shape variation in Algerian honey bee populations of *Apis mellifera intermissa* (Buttel-Reepen, 1906) (Hymenoptera: Apidae): a landmark-based geometric morphometrics analysis. *Afr. Entomol.*, 19(1), 11–22
- Baylac M, Garnery L, Tharavy D, Pedraza-Acosta J, Rortais A, Arnold G (2008) ApiClass, an automatic online wing morphometric expert system for honeybee worker identification. [online] <http://apiclass.mnhn.fr> (accessed on 10 Feb 2019)
- Beekman, M., Allsopp, M. H., Wossler, T. C., Oldroyd, B. P. (2008) Factors affecting the dynamics of honeybee (*Apis mellifera*) hybrid zone of South Africa. *Heredity* 100, 13–18
- Beekman, M., Allsopp, M. H., Holmes, M. J., Lim, J., Noach-Pienaar, L. A., Wossler, T. H., Oldroyd, B. P. (2012) Racial mixing in South African honeybees: the effects of genotype mixing on reproductive traits of workers. *Behav. Ecol. Sociobiol.* 66, 897–904. doi: <https://doi.org/10.1007/s00265-012-1338-z>
- Buco, S. M., Rinderer, T. E., Sylvester, H. A., Collins, A. M., Lancaster, V. A., Crewe, R. M. (1987) Morphometric differences between South American Africanized and South African (*Apis mellifera scutellata*)

- honey bees. *Apidologie* 1987 18, 217–222. doi: <https://doi.org/10.1051/apido:19870301>
- Calinski, T., Harabasz, J. (1974) A dendrite method for cluster analysis. *Commun. Stat.* 3 (1), 1–27. doi: <https://doi.org/10.1080/03610927408827101>
- Carreck, N. L., Andree, M., Colin, B. S., Cox-Foster, D., Dade, H. A., Ellis, J. D., Hatjina, F., van Englesdorp, D. (2013) Standard methods for *Apis mellifera* anatomy and dissection. *J. Apic. Res.* 52 (4), 1–40 doi: <https://doi.org/10.3896/IBRA.1.52.4.03>
- Crewe, R. M., Hepburn, H. R., Moritz, R. F. A. (1994) Morphometric analysis of 2 southern African races of honeybee. *Apidologie* 25, 61–70. <https://doi.org/10.1051/apido:19940107>
- Dade, H. A. (1962) Anatomy and dissection of the honey bee. International Bee Research Association, Cardiff.
- Dietemann, V., Lubbe, A., Crewe, R. M. (2006) Human Factors Facilitating the Spread of a Parasitic Honey Bee in South Africa. *J. Econ. Entomol.* 99, 7–13. doi: [https://doi.org/10.1603/0022-0493\(2006\)099\[0007:HFFTSO\]2.0.CO;2](https://doi.org/10.1603/0022-0493(2006)099[0007:HFFTSO]2.0.CO;2)
- Egmont-Peterson, M., Ridder, D., Handels, H. (2002) Image processing with neural networks—a review. *Pattern Recogn.* 35, 2279–2301. doi: [https://doi.org/10.1016/S0031-3203\(01\)00178-9](https://doi.org/10.1016/S0031-3203(01)00178-9)
- Francoy, T. M., Wittmann, D., Drauschke, M., Müller, S., Steinhage, V., Bezerra-Laure, M. A. F., De Jong, D., Gonçalves, L. S. (2008) Identification of Africanized honey bees through wing morphometrics: two fast and efficient procedures. *Apidologie* 39, 488–494. doi: <https://doi.org/10.1051/apido:2008028>
- Francoy, T. M., Witmn, D., Steinhage, V., Drauschke, M., Müller, S., et al. (2009) Morphometric and genetic changes in a population of *Apis mellifera* after 34 years of Africanization. *Genet. Mol. Res.* 8, 709–717
- Francoy, T. M., De Franco F. F., Roubik, D. W. (2012) Integrated landmark and outline-based morphometric methods efficiently distinguish species of *Euglossa* (Hymenoptera, Apidae, Euglossini). *Apidologie* 43, 609–617. doi: <https://doi.org/10.1007/s13592-012-0132-2>
- Guzman-Novoa, R. E., Page, R. E. (1994) Genetic dominance and worker interactions affect honeybee colony defense. *Behav. Ecol.* 4, 91–97
- Hepburn, H. R., Radloff, S. E. (1998) *Honeybees of Africa*. Springer, New York, New York.
- Hepburn, R., Radloff, S. E. (2002) *Apis mellifera capensis*: an essay on the subspecific classification of honeybees. *Apidologie* 33, 105–127. doi: <https://doi.org/10.1051/apido:2002001>
- Hepburn, H. R., Radloff, S. E., Fuchs, S. (1998) Population structure and the interface between *Apis mellifera capensis* and *Apis mellifera scutellata*. *Apidologie* 29, 333–346. doi: <https://doi.org/10.1051/apido:19980404>
- Jarek, S. (2012). *mvnrmtest*: Normality test for multivariate variables. R package version 0.1-9. <https://CRAN.R-project.org/package=mvnrmtest>
- Johannsmeier, M. F. (2001). *Beekeeping in South Africa*. Agricultural Research Council of South Africa, Plant Protection Research Institute.
- Kandemir, İ., Özkan, A., Fuchs, S. (2011) Reevaluation of honeybee (*Apis mellifera*) microtaxonomy: a geometric morphometric approach *Apidologie* 42: 618–627. doi: <https://doi.org/10.1007/s13592-011-0063-3>
- Kohavi, Ron (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *Int. J. Conf. Artifici. Intel.* 4 (2), 1137–1145
- Meixner, M. D., Pinto, M. A., Bouta, M., Kryger, P., Ivanova, E., Fuchs, S. (2013) Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. *J. Apic. Res.* 52, 1–28. doi: <https://doi.org/10.3896/IBRA.1.52.4.05>
- Mendez, M. F. M., Francoy, T. M., Nunes-Silvia, P., Menezes, C., Imperatriz-Fonseca, V. L. (2007) Intrapopulation variability of *Nannotrigona testaceicornis* Lepeletier, 1836 (Hymenoptera, Meliponini) using relative warp analysis. *Biosci. J.* 23, 147–152
- Miguel, I., Baylac, M., Iriondo, M. et al. (2011) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie.* 42, 150–161. doi: <https://doi.org/10.1051/apido/2010048>
- Oksanen, O., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., et al. (2016). *vegan*: Community Ecology Package. R package version 2.4-1. <https://cran.r-project.org/package=vegan>
- Oleksa, A., Tofilski, A. (2015) Wing geometric morphometrics and microsatellite analysis provide similar discrimination of honey bee subspecies. *Apidologie* 46: 49–60. doi: <https://doi.org/10.1007/s13592-014-0300-7>
- Phiancharoen, M., Pirk, C. W. W., Radloff, S. E., Hepburn, R. (2010) Clinal nature of the frequencies of ovarioles and spermathecae in Cape worker honeybees, *Apis mellifera capensis*. *Apidologie* 41, 129–134. <https://doi.org/10.1051/apido/2009054>
- R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.r-project.org/>.
- Rinderer, T. E., Sylvester, H. A., Brown, M. A., Villa, J. D., Pesante, D., Collins, A. M., Spencer, R., Kleinpeter, S., Lancaster, V. (1986) Field and simplified techniques for identifying Africanized and European honey bees. *Apidologie.* 17 (1), 33–48.
- Rinderer, T. E., Buco, S. M., Rubink, W. L., Daly, H. V., Stelzer, J. A., Riggio, R. M., Baptista, F. C. (1993) Morphometric identification of Africanized and European honey bees using large reference populations. *Apidologie.* 24 (6), 569–585
- Rowley, H. A., Baluja, S., Kanade, T. (1998) Neural network-based face detection. *Pattern Anal. Machine Intell.* 20, 23–38. doi: <https://doi.org/10.1109/34.655647>

- Ruttner, F. (1988) Biogeography and taxonomy of honeybees. Springer, Berlin, Heidelberg.
- Schneider, S. S., DeGrandi-Hoffman, G., Smith, D. R. (2003) The African honey bee: Factors contributing to a successful biological invasion. *Annu. Rev. Entomol.* 49, 351–376 doi: <https://doi.org/10.1146/annurev.ento.49.061802.123359>.
- Slice DE (2007) Geometric morphometrics. *Annu. Rev. Anthropol.* 36, 261–281 <https://doi.org/10.1146/annurev.anthro.34.081804.120613>
- Smith D.R., Crespi B.J., Bookstein F.L. (1997) Fluctuating asymmetry in the honey bee, *Apis mellifera*: effects of ploidy and hybridization. *J. Evol. Biol.* 10, 551–574. doi: <https://doi.org/10.1007/s000360050041>
- Therneau, T., Atkinson, B., Ripley, B. (2015). *rpart*: Recursive Partitioning and Regression Trees. R package. version 4.1-10. <https://CRAN.R-project.org/package=rpart>
- Tofilski, A. (2008) Using geometric morphometrics and standard morphometry to discriminate three honeybee subspecies. *Apidologie* 39, 558–563. DOI: <https://doi.org/10.1051/apido:200803>
- Venables, W. N., Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Wei T, Simko V (2016) *corrplot*: Visualization of a Correlation Matrix. R package version 0.77. <https://CRAN.R-project.org/package=corrplot>
- Weihs, C., Ligges, U., Luebke, K., Raabe, N. (2005). *klaR* Analyzing German Business Cycles, in Baier, D., Decker, R. and Schmidt-Thieme, L. (eds.). *Data Analysis and Decision Support*. Springer-Verlag, Berlin, pp. 335–343
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.