

## MORPHOMETRIC IDENTIFICATION OF AN UNKNOWN HONEY BEE COLONY: AN EXAMPLE FROM NORTH INDIA

Hardeep Kaur<sup>1</sup> ORCID: 0000-0002-7291-9150

Sajad Ahmad Ganie<sup>2</sup> ORCID: 0000-0002-9290-8766

Adam Tofilski<sup>1\*</sup> ORCID: 0000-0002-3898-7029

<sup>1</sup>Department of Zoology and Animal Welfare, University of Agriculture in Krakow, Poland

<sup>2</sup>Division of Entomology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, India

\*corresponding author: [rotofilski@cyf-kr.edu.pl](mailto:rotofilski@cyf-kr.edu.pl)

Received: 26 September 2024; accepted: 29 October 2024

### Abstract

Identification of the honey bee (*Apis mellifera*) subspecies is an important aspect of bee breeding and biodiversity conservation. The identification can be based on molecular or morphological markers. For some markers, including the cytochrome c oxidase subunit, there is a well-established methodology allowing consistent subspecies identification in different laboratories. In the case of morphological markers, identification is hindered by a lack of reference data and a standardized methodology to reuse it. There is the Morphometric Bee Data Bank in Oberursel, Germany, which provides reliable reference data for many honey bee subspecies; however, it must be expanded and updated. Moreover, there is a need for standardized protocols that allow reliable identification of unknown samples. We show here that reference data for the identification of honey bees based on geometric morphometrics can be saved in an XML file. The information in this file can be easily extracted by other users for the identification of unknown samples. We illustrate this procedure using ten samples from north India. The samples were identified as *A. mellifera*; next, they were identified as lineage C; and finally, most of the samples had high similarity to honey bees from Croatia and Slovenia. We explained what data is required for such identification and how it can be reused. The method described here works correctly only if the identified species or subspecies is well represented in the reference samples. Without reference data from America, we could not confirm the importation of honey bees to India from this region.

Keywords: classification, geometric morphometrics, honey bee, wings, XML

### INTRODUCTION

Identifying honey bee (*Apis mellifera*) subspecies is crucial for the conservation of their diversity and breeding (De la Rúa et al., 2009; Meixner et al., 2010; Bouga et al., 2011). It can be achieved through the use of various methods based on molecular markers, including the COI-COII intergenic region of mtDNA (Cornuet et al., 1991; Meixner et al., 1993; Magnus et al., 2011), microsatellites (Estoup et al., 1995; Kandemir et al., 2006) and single nuclear polymorphisms (SNP) (Parejo et al., 2018; Momeni et al., 2021); for review, see Meixner et al. (2013). Identification of subspecies based on mtDNA is well documented, a detailed protocol is available for the users, and a small sample size can provide efficient results (Evans et al., 2013), but this

method is not effective in the detection of hybrids because of the maternal inheritance of mtDNA. Microsatellites can be used for the detection of hybrids (Jensen et al., 2005; Oleksa et al., 2011; Ellis et al., 2018), but in this case, the availability of reference material can be a limiting factor. The SNP markers are very efficient in the detection of hybrids (Parejo et al., 2018), and reference sequences are available for many subspecies. However, this method is more expensive, time-consuming, requires sophisticated equipment and involves more cumbersome storage of samples (Meixner et al., 2013). Due to the drawbacks of molecular markers, beekeepers and honey bee breeders often prefer identification based on morphological markers (Ruttner, 1988b). Insect wings are often used for morphometric

identification because the venation pattern has been shown to be species and subspecies specific. Moreover, they are flat and easy to collect and measure with the use of geometric morphometrics. In earlier studies, the identification of honey bees was based on the cubital index, which is a decimal number representing the proportion of two distances along the wing veins (Ruttner et al., 1978), making it easy to report and compare differences between subspecies. For example, the cubital index is higher in *A. m. carnica* (C lineage) than in *A. m. mellifera* (M lineage) (Maul & Hähnle, 1994; Kekecoglu et al., 2007), but its use to identify subspecies of the same lineage is less accurate (Ruttner, 1991).

Nowadays, identification of honey bees is usually based on a larger number of variables. The first multivariate identification of honey bees, based on thirteen angles and two distances, was introduced by DuPraw (1965b). A similar methodology was used later in standard morphometry (Ruttner et al., 1978; Meixner et al., 2007; Kekecoglu, 2018). Currently, the identification is often based on the coordinates of landmarks within a methodological approach called “geometric morphometrics” (Bookstein, 1991), which proved to be an effective, fast and accessible method to differentiate honey bee species and subspecies (Francoy et al., 2008; Tofilski, 2008; Bustamante et al., 2021; Rodrigues et al., 2022).

This method was recently used in many studies related to both honey bee breeding and biodiversity conservation (Galataud et al., 2020; Porrini et al., 2020, 2022; Litvinoff et al., 2023; Masaquiza et al., 2023; Oleksa et al., 2023; Kaur et al., 2024). The use of multiple variables improved the accuracy of identification but made reporting and reusing the results more difficult. For all identified subspecies and all variables, it is important to provide averages along with measures of variation. Moreover, classification functions or other identification data should also be provided to aid in identifying unknown samples. Unfortunately, such data were provided only in some of the earlier studies (DuPraw, 1965b, 1965a), but they were incomplete in the contribution by Ruttner (1988a), which covered

the global diversity of honey bees.

Later, the classification data were provided only in some studies (Rinderer et al., 1993; Francoy et al., 2008; Tofilski, 2008; Nawrocka et al., 2018; Bustamante et al., 2021). In most of those studies, the measurements were analyzed with the use of linear discriminant analysis (LDA), but other statistical methods are available, including the naive Bayes classifier and an artificial neural network (Silva et al., 2015). In order to make morphological identification of honey bees more reliable, there is a need for improvements in freely available reference samples and identification methods (Oleksa et al., 2023).

In India, the western honey bee, *A. mellifera*, is a non-native species. It was introduced in the 1960s in Himachal Pradesh, which is a neighboring state of Jammu and Kashmir (Atwal & Goyal, 1973; Goyal, 1990; Rao et al., 1993; Rahman, 2017). Since then, it has been used in commercial apiculture. Both Italy and the USA are considered to be the places of origin of bees imported to India (Ibrahim et al., 2017). In 1984-1985, the western honey bee was introduced to Jammu and Kashmir (Abrol, 2004), which is currently important beekeeping region in India (Goyal, 1974). In this region, native eastern honey bees (*A. cerana*) are found in the wild or managed by beekeepers.

In this study, we demonstrate how previously published data about wing measurements can be used to identify the origin of an unknown sample. For identification, we have used samples of western honey bees from northern India. First, the samples were classified as *Apis mellifera* with the use of data provided by Bustamante et al. (2021). Next, they were classified as evolutionary lineage C with the use of data provided by Nawrocka et al. (2018). Finally, the samples were classified as originating from Croatia and Slovenia with the use of data provided by Oleksa et al. (2022). We explain what data are required for such identification and how they can be analyzed with the use of the R package described here.

## MATERIAL AND METHODS

### Sampling

Honey bee samples were collected from the union

territory of Jammu and Kashmir in northern India (Fig. 1). The sample consisted of 175 workers, representing ten colonies from ten locations (Fig 1). Both the left and right wings were used (Janczyk & Tofilski, 2021), and in total, 350 wings were measured. The data were averaged within colonies, and the averages were used for identification.

The worker bees were preserved in ethanol. The forewings were carefully detached at their base and mounted between two clean microscope slides; their images were acquired with the use of a USB camera equipped with a 25-mm lens (FL-CC2514-2M, Ricoh). The right-wing images were flipped horizontally before measurements and had a resolution of 94488 pixels per meter. Nineteen

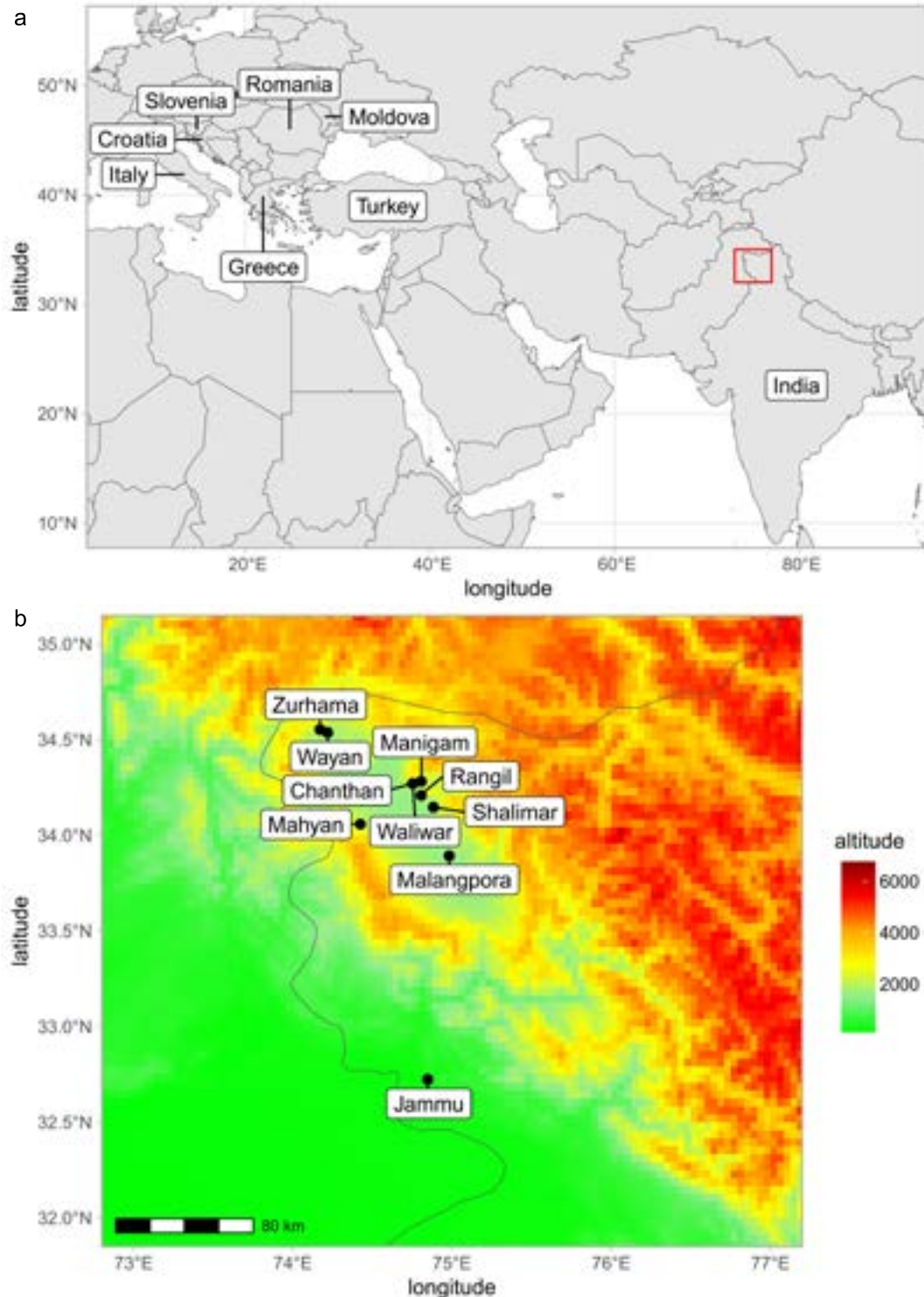


Fig. 1. Map of India and countries used as a reference for lineage C (a). A red rectangle indicating the study area was enlarged (b). Sampling sites are indicated with labeled black circles.

landmark coordinates were digitized on each wing with the use of IdentiFly software (Nawrocka et al., 2018). The position of the landmarks is the same as in the study of Nawrocka et al. (2018) and Oleksa et al. (2023). The dataset used in this study, including forewing images, landmark coordinates and geographic coordinates of sampling locations, is available on the Zenodo website (Kaur et al., 2023) under Open Data Commons Open Database License (ODbL). According to this license, the data can be adapted as long as the adapted version is also available under ODbL.

All analyses were conducted in R (v. 4.3.2) (R Core Team, 2018) using RStudio (v. 2024.09.0). The whole statistical analysis is available as an R script (Supplementary online document 1). Classification into species, lineages and regions was based on LDA which is also referred to as canonical variate analysis (CVA). The process of classifying an unknown sample involves two major steps: (1) preparation of classification data and (2) application of this data to a sample of unknown wings.

### Classification data

In this study, the classification was based on the LDA of landmark coordinates, and the classification data consisted of: reference configuration (mean of all landmark configurations), mean configurations for all classes (groups), variance-covariance matrices for all classes and classification coefficients.

In order to classify species from the genus *Apis*, we have used data provided by Bustamante et al. (2021). The classification data were obtained from supplementary files: ESM1, ESM2, and ESM3 (Bustamante et al., 2021). The data were converted to XML format and saved as the file “apis-species.dw.xml” (available at: <https://zenodo.org/records/14053812>). For the classification of evolutionary lineages and geographic regions, the data were calculated from raw landmark coordinates provided by Nawrocka et al. (2018) and Oleksa et al. (2023), respectively. They consist of x and y coordinates of nineteen landmarks, which gives thirty-eight columns (or variables). Each row was labeled with a name or class identifier; for lineages, it was: A, C, M or O.

The identification data was prepared in the “gmLdaData2xml” function. In the first step,

the landmark configurations were aligned with the use of a generalized Procrustes analysis in the geomorph package (Baken et al., 2021). The aligned coordinates were averaged to obtain the reference configuration and mean configuration for each class (often called a group). At this stage, the LDA could be applied; however, the aligned coordinates contain redundant information because, after Procrustes alignment, the degrees of freedom are reduced by four. The covariance matrices obtained from all aligned coordinates are singular, which complicates the process of matrix inversion. In order to remove this problem, we used principal component analysis (PCA) and removed the last four principal components from the analysis. The obtained principal components were uncorrelated with one another, which improved the predictive power of LDA, calculated using package MASS (Venables & Ripley, 2002). Because we used both PCA and LDA, the classification coefficients were calculated through matrix multiplication of the principal coefficients by the discriminant coefficients. Finally, the linear discriminant scores obtained from LDA were used to obtain variance-covariance matrices for all classes.

At this stage we obtained the identification data, which could be described as a collection of tables: reference configuration, mean configurations for all classes, variance-covariance matrices for all classes and classification coefficients. Those tables can be considered as a short summary of the reference data and can be used for classification of an unknown sample without direct access to the reference data. The identification data can be saved in various formats but we used XML, which is widely used by various software. In this study the identification data were encoded in two XML files: “apis-mellifera-lineage.dw.xml” (available at: <https://zenodo.org/records/14054009>) and “apis-mellifera-lineage\_c\_region.dw.xml” (available at: <https://zenodo.org/records/14054162>), which correspond to the identification of four honey bee lineages and five geographic regions in lineage C, respectively.

### Classification of unknown data

The classification data and the unknown data should have the same number of landmarks that

are homologues to each other. In the case of the data provided by Bustamante et al. (2021), the order of the landmarks was different than in our data, and therefore were reordered. The unknown data can consist of the landmarks from a single wing, but the results are more precise when it is an average of the landmarks from a large number of wings. If both the left and right wings were measured, their landmarks could be averaged to represent one individual. Classification is often based on a sample of 10-20 workers collected from one colony or one location. Before averaging, the landmark configurations have to be aligned with one another.

The data can be easily classified with the use of the “xml2id” function. The name of the XML file with classification data, the coordinates of the landmarks, and an optional “average” argument indicates if the input data should be averaged before analysis. Inside the function, the input data was aligned with the reference configuration with use of ordinary Procrustes analysis in the shapes package (Dryden, 2021). The aligned input data was multiplied by classification coefficients in order to obtain linear discriminant scores, which were used to compute the squared Mahalanobis distance (MD) between the unknown sample and each of the classes through the means and variance-covariance matrices of all classes.

MD follows the chi-squared distribution and can be used to calculate the probability that the unknown sample belongs to a particular class. The unknown sample is usually assumed to belong to the class for which MD was the smallest and the classification probability was the highest; however, it is important to review classification results for all classes. The interpretation of the classification results is discussed below. The functions for saving and reading classification data are part of the IdentiFlyR package, which can be downloaded from GitHub (<https://github.com/DrawWing/IdentiFlyR>) and easily reused in R code.

## RESULTS

Principal component analysis revealed that both workers (individuals) and samples (colonies) from India formed a single cluster of points, but samples 3 and 10 differed to some degree from others (Supplementary document 1). The divergence of sample 10 can be explained by geographic distance from other samples, but the divergence of sample 3 is difficult to explain and can be related to a more recent introduction from another location. Through the use of data provided by Bustamante et al. (2021), all workers and samples were classified as *Apis mellifera* (Fig. 2). The classification probabilities were relatively high (Supplementary document 1), which confirms that the collected material represented *A. mellifera* and not *A. cerana*, which is present in the study area.

With the use of data from Nawrocka et al. (2018), the average configuration from all samples from India led to a classification as evolutionary lineage C with a probability of  $1.381 \times 10^{-3}$  (Tab. 1) and the smallest MD 13.17. When the same classification was based on individual samples, nine of them were classified as lineage C, and one

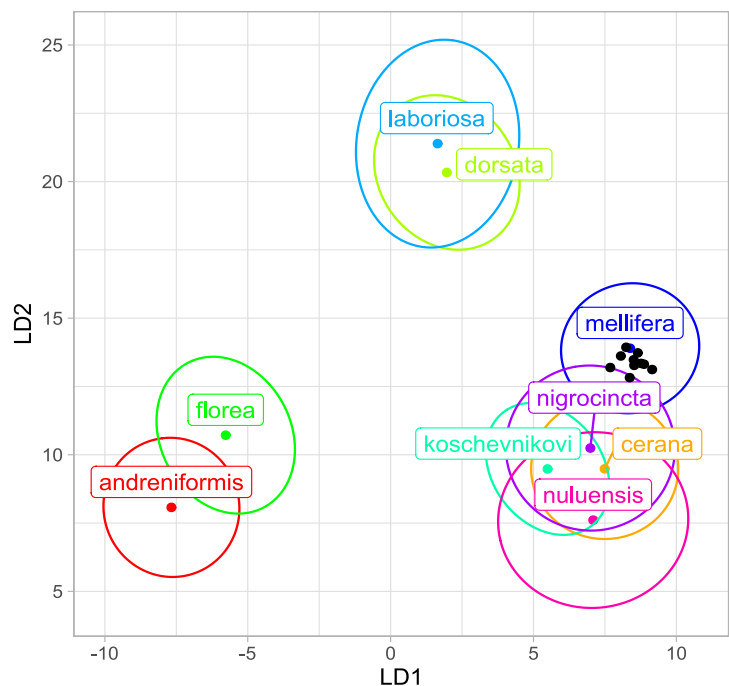


Fig. 2. Classification of honey bee samples from India as species from genus *Apis* based on the first two linear discriminants (LD). Each black dot indicates one sample. Ellipses indicate 95% confidence regions.

was classified as lineage A (Fig. 3). In the case of some samples, the classification probabilities were relatively low (Tab. 2).

With the use of data, which was a collection of wing images, especially from different European countries (Oleksa et al., 2023; Nawrocka et al., 2018), the average configuration of all samples from India was classified as Croatia and Slovenia with a probability of  $3.825 \times 10^{-2}$  (Tab. 3) and smallest MD 8.410. When the same classification was based on individual samples, seven of them were classified as Croatia and Slovenia, two samples were classified as Romania and Moldova cluster and one sample was classified as Greece (Tab. 4, Fig. 4).

## DISCUSSION

The results presented here indicate that the western honey bees (*A. mellifera*) samples from the Jammu and Kashmir region of northern India belong to lineage C (Fig. 3), which naturally occurs in south-eastern Europe. Comparison of the obtained data with those from the countries of this region showed the highest similarity with honey bees from Croatia and Slovenia (Fig. 4). It had earlier been reported that *A. mellifera* was introduced to India from Italy and the USA (Rahman, 2017). Unexpectedly, the analyzed samples here differed markedly from *A. m. ligustica* from Italy. In comparison to honey bees originating from other countries,

Table 1.

Dissimilarity of the honey bees from India to evolutionary lineage.  
MD: Squared Mahalanobis distance

| Evolutionary Lineage | MD     | Probability             |
|----------------------|--------|-------------------------|
| A                    | 37.86  | $5.995 \times 10^{-9}$  |
| C                    | 13.17  | $1.381 \times 10^{-3}$  |
| M                    | 115.07 | $1.032 \times 10^{-25}$ |
| O                    | 71.54  | $2.926 \times 10^{-16}$ |

Table 2.

Classification of honey bee samples from India as evolutionary lineages. The probability corresponds to the lineage, which is the result of the classification. The probabilities for all other lineages were lower.

| Sample | Evolutionary Lineage | Probability            |
|--------|----------------------|------------------------|
| 1      | C                    | $1.055 \times 10^{-3}$ |
| 2      | C                    | $6.053 \times 10^{-3}$ |
| 3      | A                    | $2.139 \times 10^{-5}$ |
| 4      | C                    | $1.866 \times 10^{-1}$ |
| 5      | C                    | $4.528 \times 10^{-4}$ |
| 6      | C                    | $3.292 \times 10^{-2}$ |
| 7      | C                    | $9.116 \times 10^{-5}$ |
| 8      | C                    | $3.557 \times 10^{-3}$ |
| 9      | C                    | $6.220 \times 10^{-6}$ |
| 10     | C                    | $1.201 \times 10^{-5}$ |

Table 3.

Dissimilarity of honey bee samples from India to regions in lineage C.  
MD: Squared Mahalanobis distance

| Region            | MD      | Probability             |
|-------------------|---------|-------------------------|
| Greece            | 22.591  | $4.914 \times 10^{-5}$  |
| Croatia- Slovenia | 8.410   | $3.825 \times 10^{-2}$  |
| Italy             | 110.800 | $7.380 \times 10^{-24}$ |
| Romania- Moldova  | 9.727   | $2.103 \times 10^{-2}$  |
| Turkey            | 33.193  | $2.933 \times 10^{-7}$  |

Table 4.

Classification of honey bee samples from India as regions in lineage C. The probability corresponds to the region, which is the result of the classification. The probabilities for all other regions were lower.

| Sample | Region           | Probability |
|--------|------------------|-------------|
| 1      | Romania- Moldova | 0.0172      |
| 2      | Croatia-Slovenia | 0.0671      |
| 3      | Greece           | 0.0092      |
| 4      | Croatia-Slovenia | 0.4042      |
| 5      | Croatia-Slovenia | 0.0499      |
| 6      | Croatia-Slovenia | 0.0035      |
| 7      | Croatia-Slovenia | 0.0198      |
| 8      | Croatia-Slovenia | 0.0100      |
| 9      | Croatia-Slovenia | 0.0125      |
| 10     | Romania- Moldova | 0.0322      |

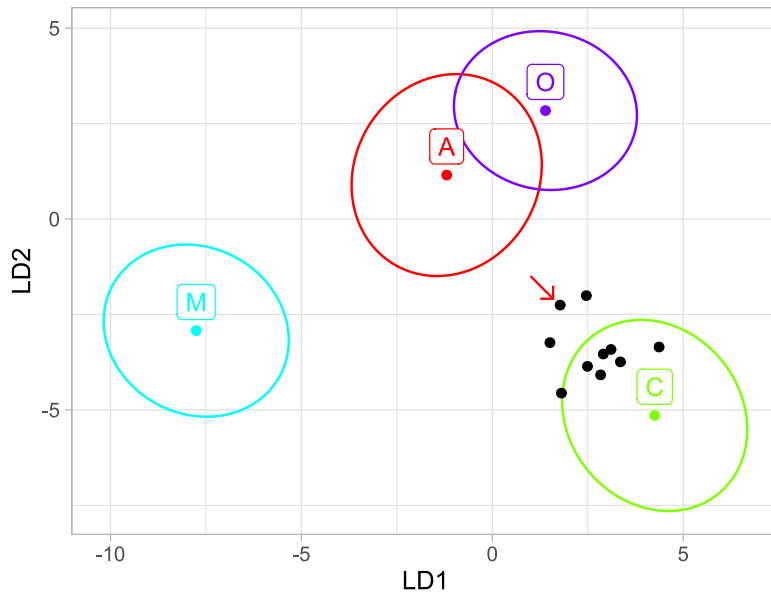


Fig. 3. Classification of honey bee samples from India as honey bee lineages based on the first two linear discriminants (LD). Each black dot indicates one sample. Ellipses indicate 95% confidence regions. Nine samples were classified as lineage C, and one sample indicated with a red arrow was classified as lineage A.

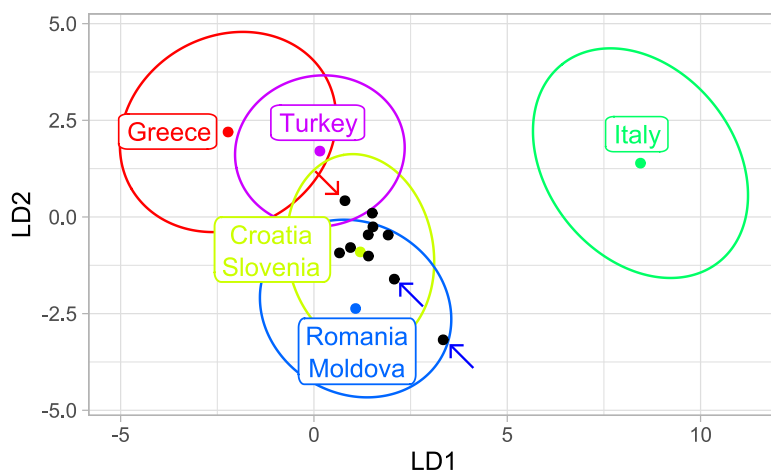


Fig. 4. Classification of honey bee samples from India as regions in lineage C based on the first two linear discriminants (LD). Each black dot indicates one sample. Ellipses indicate 95% confidence regions. Seven samples were classified as Croatia-Slovenia, two samples indicated with blue arrows were classified as Romania-Moldova and one sample indicated with a red arrow was classified as Greece.

the similarity to those from Italy was the lowest (Tab. 3, Fig. 4). However, the reference sample of *A. m. ligustica* is relatively small and unexpectedly differed from all other regions from lineage C. In order to conclude that the honey bees from India do not originate from Italy, there is a need

for larger and more representative reference samples from this country. It cannot be excluded that the analyzed sample had ancestors from the USA where numerous honey bee subspecies had been introduced at various times (Carpenter & Harpur, 2021). *A. m. ligustica*, and *A. m. carnica* (both from lineage C) were introduced to the USA in the 1870s and 1880s from north Italy and Croatia, respectively (Carpenter & Harpur, 2021). The current honey bee population in the USA primarily exhibits the genetic characteristics of those two subspecies (Carpenter & Harpur, 2021). Therefore, possibly the investigated honey bees originated from North America, but we do not have reference samples to verify this assumption. It is also possible that the bees introduced to India originated from more than one geographic region. Because they could hybridize with each other, it is difficult to reconstruct their origin using morphological data. Therefore, analysis based on molecular markers could be more conclusive.

#### Probability of classification

It is often assumed that the classification model includes all possible groups. Such an approach has been used, among others, by Bustamante et al. (2021), in which the success of classification is presented as a percent, calculated by adding the classification probabilities for all groups and finding what percentage of this total probability corresponds to the group with the highest probability. In this approach, the classification probability is often large and close to 100%, even if the similarity between the unknown sample and the focal group is relatively low.

This problem was recognized to some degree in the web application for the identification of *Apis* species (Bustamante et al., 2021), in which results can be accompanied by a warning that “shape is abnormal”. This happens when the measured wing

was obtained from a different genus, for example, *Bombus*.

The use of percentages or proportions in identification results was called the restricted approach to classification (Albrecht, 1992). It was argued that it is better to use an unrestricted method in which there is a provided probability of classification for all identified groups. (Albrecht, 1992). In most realistic scenarios, the reference data used for classification represents only part of the variation. For example, in the case of honey bee subspecies classification, some geographic regions are often unrepresented, but bees from this region can be used as an unknown sample. We should attempt to detect situations when the unknown sample is not similar to any of the classes. The unrestricted method of classification of honey bee subspecies was used in this study and the study by Nawrocka et al. (2018).

The interpretation of the results should not be limited to obtaining the most probable group, but the probabilities of classification should be verified for all groups. In particular, the two largest probabilities should be compared; in the case of classification into lineages, the largest probability value ( $1.381 \times 10^{-3}$ ) was relatively low, but the next largest value ( $5.995 \times 10^{-9}$ ) was much smaller. The largest probability was relatively low because the reference sample of lineage C consisted of only thirty-seven colonies (Nawrocka et al., 2018) and did not adequately cover the natural variation. In this situation, many samples within the range of lineage C will be relatively dissimilar to the reference sample. In the case of classification into species, the sample size of *A. mellifera* was 615 colonies (Bustamante et al., 2021) and in the case of classification into regions in lineage C, the sample size of regions Croatia and Slovenia was 181 colonies (Oleksa et al., 2023). These reference samples are more representative and the maximum classification probabilities corresponding to them are greater.

Sample 3 was classified as lineage A but classified as originating from Greece. This can be considered inconsistent because Greece is within the range of lineage C. The classification into regions can be interpreted as more reliable because it is based on a large reference sample. This inconsistency

is suspected to be related to an inadequate sample size in lineage classification. In figure 4, sample 3 was closer to Croatia and Slovenia than to Greece but still classified as Greece. This seeming inconsistency is related to the fact that the graph depicts only the first two of four linear discriminants. In the two invisible dimensions, sample 3 was closer to Greece than to other regions (Supplementary document 1). In this context, it is better to pay more attention to probabilities than to graphs.

### Providing and using identification data

In this study, we show how to produce classification data and provide it to others in a format that is easy to reuse. The classification data were saved in an XML files, which are readable by various software. XML is widely used in various operating systems and is supported by many software libraries, among others by R (R Core Team, 2018) and IdentiFly software (Nawrocka et al., 2018). We provide an IdentiFlyR package with functions that can be used to create new classification data, read it, and reuse it for the identification of unknown samples.

In order to create new classification data, a user should prepare a table with reference data. The columns of the table should contain the x and y coordinates of landmarks. Moreover, there is a required column with class identifiers or group names. If the table is in csv format, it can be imported into R using the “read.csv” function. The “gmLdaData2xml” function should be used to create the XML file.

In order to classify an unknown sample in R, an array of input data is required, which can be imported in the same way as the reference data mentioned above. The input data should consist of landmark coordinates formatted in the same way as the reference data. In order to classify the input data, the “xml2id” function should be used. Although R software is freely available, some users find it difficult to use because it is based on scripts in text format. Those who prefer a graphical user interface can open the XML file in the IdentiFly software and use it for identification. In the latter case, the user needs to determine landmarks on wing images that are saved in one directory.

Next, the file with the extension “.dw.xml”, which contains classification data, should be chosen using the option “Set classification” in the menu “Classification”, and finally, the option “Classify directory” should be used from the same menu. The results of the classification will be provided in the new window. More information about using IndentiFly software can be found in Nawrocka et al. (2018).

### Providing raw data

An unknown sample can be identified through the use of summary statistics, including reference configuration, class means, variance-covariance matrices and classification coefficients. However, providing all raw data, including wing images and landmark coordinates allows the use of different statistical identification models, which can be more adequate (Silva et al., 2015). Moreover, different studies used different configurations of landmarks (Nawrocka et al., 2018; Bustamante et al., 2021). When wing images are provided, they can be reanalyzed, and additional landmarks can be determined. The reuse of datasets from earlier studies allowed us to compare our sample with a much larger reference dataset. We provided both wing images and landmark coordinates, which were newly obtained in this study (Kaur et al., 2023).

When classifying unknown samples, it is better to use an unrestricted rather than a restricted approach to calculate the probability of classification. In order to obtain those probabilities based on LDA of landmark coordinates, the following data is required: reference configuration, mean configurations for all classes, classification coefficients, and variance-covariance matrices for all classes. We provide an R package that can be used to produce the classification data, save it in an XML file and reuse it in another analysis. The method described here can be applied not only to honey bee wings but also to all data based on landmark coordinates.

### ACKNOWLEDGMENTS

The authors gratefully acknowledge the Principal investigator IDP NAHEP- SKUAST-Kashmir for his

valuable guidance, support, encouragement and sanctioning of overseas training. We are also grateful to the beekeepers for providing honey bee samples for the study. This research was funded by the National Science Centre, Poland, grant number 2021/41/B/NZ9/03153.

### REFERENCES

- Abrol, D. P. (2004). Introduction of *Apis mellifera* in Jammu and Kashmir: Present status and future prospects. *Bee World*, 85(2), 37-40. <https://doi.org/10.1080/0005772X.2004.11099616>
- Albrecht, G. H. (1992). Assessing the affinities of fossils using canonical variates and generalized distances. *Human Evolution*, 7(4), 49-69. <https://doi.org/10.1007/BF02436412>
- Atwal, A. S., & Goyal, N. P. (1973). Introduction of *Apis mellifera* in Punjab plains. *Indian Bee Journal*, 35(1/4), 1-9.
- Baken, E. K., Collyer, M. L., Kaliontzopoulou, A., Adams, D. C. (2021). geomorph v4.0 and gmShiny: Enhanced analytics and a new graphical interface for a comprehensive morphometric experience. *Methods in Ecology and Evolution*, 12(12), 2355-2363. <https://doi.org/10.1111/2041-210X.13723>
- Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press.
- Bouga, M., Alaux, C., Bienkowska, M., Büchler, R., Carreck, N. L., Cauia, E., ... Wilde, J. (2011). A review of methods for discrimination of honey bee populations as applied to European beekeeping. *Journal of Apicultural Research*, 50(1), 51-84. <https://doi.org/10.3896/IBRA.1.50.1.06>
- Bustamante, T., Fuchs, S., Grünwald, B., Ellis, J. D. (2021). A geometric morphometric method and web application for identifying honey bee species (*Apis* spp.) using only forewings. *Apidologie*, 52(3), 697-706. <https://doi.org/10.1007/s13592-021-00857-7>

- Carpenter, M. H., & Harpur, B. A. (2021). Genetic past, present, and future of the honey bee (*Apis mellifera*) in the United States of America. *Apidologie*, 52(1), 63-79. <https://doi.org/10.1007/s13592-020-00836-4>
- Cornuet, J. M., Garnery, L., Solignac, M. (1991). Putative origin and function of the intergenic region between COI and COII of *Apis mellifera* L. mitochondrial DNA. *Genetics*, 128(2), 393-403. <https://doi.org/10.1093/genetics/128.2.393>
- De la Rúa, P., Jaffé, R., Dall'Olio, R., Muñoz, I., Serrano, J. (2009). Biodiversity, conservation and current threats to European honeybees. *Apidologie*, 40(3), 263-284. <https://doi.org/10.1051/apido/2009027>
- Dryden, I. L. (2021). *Shapes package*. R Foundation for Statistical Computing.
- DuPraw, E. J. (1965a). Non-Linnean taxonomy and the systematics of honeybees. *Systematic Zoology*, 14(1), 1-24.
- DuPraw, E. J. (1965b). The recognition and handling of honeybee specimens in non-Linnean taxonomy. *Journal of Apicultural Research*, 4(2), 71-84.
- Ellis, J. S., Soland-Reckeweg, G., Buswell, V. G., Huml, J. V., Brown, A., Knight, M. E. (2018). Introgression in native populations of *Apis mellifera mellifera* L: implications for conservation. *Journal of Insect Conservation*, 22(3), 377-390. <https://doi.org/10.1007/s10841-018-0067-7>
- Estoup, A., Garnery, L., Solignac, M., Cornuet, J. M. (1995). Microsatellite variation in honey bee (*Apis mellifera* L.) populations: Hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics*, 140(2), 679-695. <https://doi.org/10.1093/genetics/140.2.679>
- Evans, J. D., Schwarz, R. S., Chen, Y. P., Budge, G., Cornman, R. S., Rua, P. D. la, ... Pinto, M. A. (2013). Standard methods for molecular research in *Apis mellifera*. *Journal of Apicultural Research*, 52(4), 1-54. <https://doi.org/10.3896/IBRA.1.52.4.11>
- Francoy, T. M., Wittmann, D., Drauschke, M., Müller, S., Steinhage, V., Bezerra-Laure, M. A., De Jong, D., Gonçalves, L. S. (2008). Identification of Africanized honey bees through wing morphometrics: Two fast and efficient procedures. *Apidologie*, 39(5), 488-494. <https://doi.org/10.1051/apido:2008028>
- Galataud, J., Delatte, H., Techer, M. A., Simiand, C., Sookar, P., Reynaud, B., Clémencet, J. (2020). When European meets African honeybees (*Apis mellifera* L.) in the tropics: Morphological changes related to genetics in Mauritius Island (South-West Indian Ocean). *PLoS ONE*, 15(11), e0242053. <https://doi.org/10.1371/journal.pone.0242053>
- Goyal, N. P. (1974). *Apis cerana indica* and *Apis mellifera* as complementary to each other for the development of apiculture. *Bee World*, 55(3), 98-101. <https://doi.org/10.1080/0005772X.1974.11097506>
- Goyal, N. P. (1990). Introduction of *Apis mellifera* in India: Background and present status. *Indian Bee Journal*, 52(1-4), 57-59.
- Ibrahim, M. M., Chandel, Y. S., Anil. (2017). Morphometrics of *Apis mellifera* after five decades of its introduction in North-Western Himalayan region of India. *Pakistan Journal of Zoology*, 49(4), 1397-1403. <https://doi.org/10.17582/journal.pjz/2017.49.4.1397.1403>
- Janczyk, A., & Tofilski, A. (2021). Monthly changes in honey bee forewings estimated using geometric morphometrics. *Journal of Apicultural Science*, 65(1), 139-146. <https://doi.org/10.2478/jas-2021-0002>
- Jensen, A. B., Palmer, K. A., Boomsma, J. J., Pedersen, B. V. (2005). Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe. *Molecular Ecology*, 14(1), 93-106. <https://doi.org/10.1111/j.1365-294X.2004.02399.x>

- Kandemir, İ., Meixner, M. D., Ozkan, A., Sheppard, W. S. (2006). Genetic characterization of honey bee (*Apis mellifera cypria*) populations in northern Cyprus. *Apidologie*, 37(5), 547-555. <https://doi.org/10.1051/apido:2006029>
- Kaur, H., Ganie, S. A., Tofilski, A. (2023). Fore wings of honey bee (*Apis mellifera*) from Jammu and Kashmir, India. *Zenodo*. <https://doi.org/10.5281/zenodo.8071014>
- Kaur, H., Nedić, N., Tofilski, A. (2024). Influence of honey bee (*Apis mellifera*) breeding on wing venation in Serbia and neighbouring countries. *PeerJ*, 12, e17247. <https://doi.org/10.7717/peerj.17247>
- Kekecoglu, M. (2018). Morphometric divergence of anatolian honeybees through loss of original traits: A dangerous outcome of Turkish apiculture. *Sociobiology*, 65(2), 232-243. <https://doi.org/10.13102/sociobiology.v65i2.1895>
- Kekecoglu, M., Bouga, M., Soysal, M. İ., Harizanis, P. (2007). Morphometrics as a tool for the study of genetic variability of honey bees. *Journal of Tekirdag Agricultural Faculty*, 4(1), 7-15.
- Litvinoff, L., Menescardi, F., Porrini, L., Russo, R., Liendo, M. C., Nucci, A., ... Galindo-Cardona, A. (2023). Morphometric and genetic characterization as tools for selection of *Apis mellifera* (Hymenoptera: Apidae) stocks in an area of natural hybridization in Argentina. *Frontiers in Insect Science*, 2. <https://doi.org/10.3389/finsc.2022.1073999>
- Magnus, R. M., Tripodi, A. D., Szalanski, A. L. (2011). Mitochondrial DNA diversity of honey bees, *Apis mellifera* L. (Hymenoptera: Apidae) from queen breeders in the United States. *Journal of Apicultural Science*, 55(1), 37-46.
- Masaquiza, D., Ferrán, M. O., Guamán, S., Naranjo, E., Vaca, M., Curbelo, L. M., Arenal, A. (2023). Geometric morphometric analysis of wing shape to identify populations of *Apis mellifera* in Camagüey, Cuba. *Insects*, 14(3), 306. <https://doi.org/10.3390/insects14030306>
- Maul, V., & Hähnle, A. (1994). Morphometric studies with pure bred stock of *Apis mellifera carnica* Pollmann from Hessen. *Apidologie*, 25(2), 119-132. <https://doi.org/10.1051/apido:19940201>
- Meixner, M. D., Costa, C., Kryger, P., Hatjina, F., Bouga, M., Ivanova, E., Büchler, R. (2010). Conserving diversity and vitality for honey bee breeding. *Journal of Apicultural Research*, 49(1), 85-92. <https://doi.org/10.3896/IBRA.1.49.1.12>
- Meixner, M. D., Pinto, M. A., Bouga, M., Kryger, P., Ivanova, E., Fuchs, S. (2013). Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. *Journal of Apicultural Research*, 52(4), 1-28. <https://doi.org/10.3896/IBRA.1.52.4.05>
- Meixner, M. D., Sheppard, W. S., Poklukur, J. (1993). Asymmetrical distribution of a mitochondrial DNA polymorphism between 2 introgressing honey bee subspecies. *Apidologie*, 24(2), 147-153. <https://doi.org/10.1051/apido:19930207>
- Meixner, M. D., Worobik, M., Wilde, J., Fuchs, S., Koeniger, N. (2007). *Apis mellifera mellifera* in eastern Europe - morphometric variation and determination of its range limits. *Apidologie*, 38(2), 191-197. <https://doi.org/10.1051/apido:2006068>
- Momeni, J., Parejo, M., Nielsen, R. O., Langa, J., Montes, I., Papoutsis, L., ... Estonba, A. (2021). Authoritative subspecies diagnosis tool for European honey bees based on ancestry informative SNPs. *BMC Genomics*, 22(1), 101. <https://doi.org/10.1186/s12864-021-07379-7>
- Nawrocka, A., Kandemir, İ., Fuchs, S., Tofilski, A. (2018). Computer software for identification of honey bee subspecies and evolutionary lineages. *Apidologie*, 49(2), 172-184. <https://doi.org/10.1007/s13592-017-0538-y>
- Oleksa, A., Căuia, E., Siceanu, A., Puškadija, Z., Kovačić, M., Pinto, M. A., ... Tofilski, A. (2022). Collection of wing images for conservation of honey bees (*Apis mellifera*) biodiversity in Europe.

- Zenodo. <https://doi.org/10.5281/zenodo.7244070>
- Oleksa, A., Căuia, E., Siceanu, A., Puškadija, Z., Kovačić, M., Pinto, M. A., ... Tofilski, A. (2023). Honey bee (*Apis mellifera*) wing images: A tool for identification and conservation. *GigaScience*, 12, giad019. <https://doi.org/10.1093/gigascience/giad019>
- Oleksa, A., Chybicki, I., Tofilski, A., Burczyk, J. (2011). Nuclear and mitochondrial patterns of introgression into native dark bees (*Apis mellifera mellifera*) in Poland. *Journal of Apicultural Research*, 50(2), 116-129. <https://doi.org/10.3896/IBRA.1.50.2.03>
- Parejo, M., Henriques, D., Pinto, M. A., Soland-Reckeweg, G., Neuditschko, M. (2018). Empirical comparison of microsatellite and SNP markers to estimate introgression in *Apis mellifera mellifera*. *Journal of Apicultural Research*, 57(4), 504-506. <https://doi.org/10.1080/00218839.2018.1494894>
- Porrini, L. P., Quintana, S., Brasesco, C., Maggi, M. D., Porrini, M. P., Garrido, M. P., Fernández Iriarte, P. J., Eguaras, M. J. (2022). Current genetic diversity of managed and commercially produced *Apis mellifera* colonies in Argentina inferred by wing geometric morphometrics and COI-COII mtDNA locus. *Apidologie*, 53(5), 61. <https://doi.org/10.1007/s13592-022-00970-1>
- Porrini, L. P., Quintana, S., Brasesco, C., Porrini, M. P., Garrido, P. M., Eguaras, M. J., Müller, F., Fernandez Iriarte, P. (2020). Southern limit of Africanized honey bees in Argentina inferred by mtDNA and wing geometric morphometric analysis. *Journal of Apicultural Research*, 59(4), 648-657. <https://doi.org/10.1080/00218839.2019.1681116>
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rahman, A. (2017). *Apiculture in India*. Directorate of Knowledge Management in Agriculture, Indian Council of Agricultural Research.
- Rao, G. M., Rao, K. S., Chaudhary, O. P. (1993). Introduction of *Apis mellifera* in India. *Khadi Gramodyog*, 34(11 & 12), 815-819.
- 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. <https://doi.org/10.1051/apido:19930605>
- Rodrigues, P. J., Gomes, W., Pinto, M. A. (2022). DeepWings©: Automatic wing geometric morphometrics classification of honey bee (*Apis mellifera*) subspecies using deep learning for detecting landmarks. *Big Data and Cognitive Computing*, 6(3), Article 3. <https://doi.org/10.3390/bdcc6030070>
- Ruttner, F. (1988a). *Biogeography and taxonomy of honeybees*. Springer.
- Ruttner, F. (1988b). *Breeding techniques and selection for breeding of the honeybee*. British Isles Bee Breeders' Association.
- Ruttner, F. (1991). Biometrical control of breeding. *Journal of Apicultural Research*, 30(2), 113-115.
- Ruttner, F., Tassencourt, L., Louveaux, J. (1978). Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie*, 9(4), 363-381. <https://doi.org/10.1051/apido:19780408>
- Silva, F. L. da, Grassi Sella, M. L., Franco, T. M., Costa, A. H. R. (2015). Evaluating classification and feature selection techniques for honeybee subspecies identification using wing images. *Computers and Electronics in Agriculture*, 114, 68-77. <https://doi.org/10.1016/j.compag.2015.03.012>
- Tofilski, A. (2008). Using geometric morphometrics and standard morphometry to discriminate three honeybee subspecies. *Apidologie*, 39(5), 558-563. <https://doi.org/10.1051/apido:2008037>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). Springer. <https://www.stats.ox.ac.uk/pub/MASS4/>