

# Computer software for identification of honey bee subspecies and evolutionary lineages

Anna NAWROCKA<sup>1</sup>, İrfan KANDEMİR<sup>2</sup>, Stefan FUCHS<sup>3</sup>, Adam TOFILSKI<sup>1</sup>

<sup>1</sup>Department of Pomology and Apiculture, University of Agriculture in Krakow, Al. 29 Listopada 54, 31-425, Krakow, Poland

<sup>2</sup>Department of Biology, Faculty of Science, Ankara University, Tandoğan, 06100, Ankara, Turkey

<sup>3</sup>Institut für Bienenkunde, Polytechnische Gesellschaft, Faculty of Life Sciences, Goethe - Universität Frankfurt am Main, Karl-von-Frisch-Weg 2, 61440, Oberursel, Germany

Received 13 March 2017 – Revised 17 July 2017 – Accepted 2 August 2017

**Abstract** – Within the western honey bee (*Apis mellifera*), there are more than 20 recognised subspecies. It is well known that these subspecies differ in their wing venation patterns. However, there is a demand for efficient tools to identify honey bee subspecies, ecotypes, populations or hybrids. The aim of this study was to develop a fast and easy identification method based on analysing forewing vein patterns of honey bees by geometric morphometrics. Reference samples for the subspecies were obtained from the Morphometric Bee Data Bank in Oberursel, Germany. These contained 187 honey bee colonies allocated into 25 subspecies from four evolutionary lineages. The identification of evolutionary lineages of honey bees based on forewing venations proved to be highly reliable, which confirms earlier studies. The accuracy of honey bee subspecies identification was less consistent and ranged from 100 to 50% and was particularly low in African honey bees. The obtained identification data were exported to the IdentiFly computer software, which is freely available.

## honey bee / subspecies discrimination / geometric morphometry / wing morphometry

### 1. INTRODUCTION

The western honey bee (*Apis mellifera*) has a wide geographical range, which covers almost the entire area of Europe, Africa, the Near East and central Asia (Ruttner 1988; Sheppard and Meixner 2003; Chen et al. 2016). Within this range, there are markedly different climates and environments. Such diverse conditions and the history of spread and isolation of subpopulations resulted in a notable variation of morphological traits. These traits, however, are rarely distinct in single subspecies,

and multivariate statistical approaches were indispensable for reliable identification. By incorporating behavioural traits, numerous subspecies were defined (last reviewed by Engel 1999). These subspecies are grouped into four evolutionary lineages based on analysis of morphological characteristics (Ruttner 1988). Lineage A occurs in central and southern Africa, lineage C in southwest Europe and lineages M and O in Northern and Western Europe and lineage O in the Middle East (Kandemir et al. 2011; Meixner et al. 2013). Later molecular analyses mostly confirmed the morphological groupings, with few minor discrepancies (Franck et al. 1998; Alburaki et al. 2011).

Discrimination of the subspecies is important for the conservation of the honey bee biodiversity. Because of honey bee queen trade and migratory beekeeping, the natural ranges of subspecies are increasingly disturbed (De la Rúa et al. 2009). Furthermore, feral honey bee populations shrank

---

**Electronic supplementary material** The online version of this article (doi:[10.1007/s13592-017-0538-y](https://doi.org/10.1007/s13592-017-0538-y)) contains supplementary material, which is available to authorized users.

---

Corresponding author: A. Tofilski, [rotofilski@cyf-kr.edu.pl](mailto:rotofilski@cyf-kr.edu.pl)

Manuscript editor: Marina Meixner

or vanished in certain parts of Europe (Moritz et al. 2007). Monitoring local subspecies and limiting the introduction of non-native subspecies would help to conserve honey bee genetic biodiversity (De la Rúa et al. 2009). Moreover, discrimination of honey bee subspecies can be important in zoogeographic studies. In some parts of Asia and Africa, the distribution ranges of honey bee subspecies have not been fully explored (Meixner et al. 2013). Precise zoogeographical maps based on standardised identification methods would make an important contribution to future conservation programs.

Proper identification of the subspecies and evolutionary lineages is also important in honey bee breeding. Breeders often declare that their breeding lines belong to a particular subspecies. In order to keep those breeding lines as pure stocks, breeders need to eliminate colonies representing other subspecies (Kauhausen-Keller and Keller 1994). In some cases, beekeepers aim to avoid some subspecies and their hybrids, for example, Africanized honey bees in the American continents (Francoy et al. 2008; Sheppard et al. 1991, 1999).

There is a range of different methods of honey bee subspecies identification (Bouga et al. 2011), which differ in precision. Usually, methods that are more precise require more effort; therefore, for the purpose of biodiversity conservation and breeding, often rapid and less precise methods were used. Measurement of the cubital index and other traditional methods (Prabucki et al. 2002; Samborski et al. 2002; Rostecki et al. 2007) based on a small number of measurements are suitable for the discrimination of a limited number of subspecies. The precision and power of the identification were markedly improved by increasing the number of measured characters (DuPraw 1965a, b). Classical morphometry is based on measurements of 36 characters, which include the size of various body parts (forewings, abdomen and legs), colour and pilosity (Ruttner 1988). However, this method is highly labour intensive because it requires the preparation of several body parts and numerous measurements. Recently, molecular methods based on allozymes (Bouga et al. 2005), mitochondrial DNA (Smith and Brown 1990; Garnery et al. 1993) or microsatellites (Jensen et al. 2005) were used to study the intra-

specific variation of honey bees. Among these methods, only microsatellites proved to be suitable for subspecies identification. In general, molecular methods are relatively expensive, require sophisticated equipment and often are more time consuming than morphological methods. Moreover, validation of molecular methods again relies on traditional morphological subspecies identification, and it was demonstrated that morphometric and molecular methods provide similar results (Oleksa and Tofilski 2015).

Efforts have been made to improve and facilitate the identification process based on morphometry. Computerised methods are considered as the most promising (Tofilski 2005; Francoy et al. 2008; Tofilski 2008). In order to reduce the number of dissected body parts, the new methods were only based on forewings and did not require the laborious preparation of sternites, mouthparts and legs. Among other methods of wing measurements, geometric morphometrics proved to be particularly useful. This method was highly effective in the discrimination of three honey bee subspecies used for breeding in Poland (*A. m. caucasica*, *A. m. carnica* and *A. m. mellifera*) (Tofilski 2008; Gerula et al. 2009). Later, a similar method was used for the discrimination of 24 known honey bee subspecies (Kandemir et al. 2011). Although it was often demonstrated that computerised methods can be useful for the identification of honey bee subspecies, for example, ID-BEES (Batra 1988), FABIS (Rinderer et al. 1990) and ABIS (Francoy et al. 2008), the software is usually not available to the general public. The two notable exceptions are the DrawWing software (Tofilski 2004, 2008) and the ApiClass website (Baylac et al. 2008); however, they cover only a small fraction of honey bee subspecies.

The aim of this study was to develop a fast and easily available method for the identification of a larger number of honey bee subspecies and lineages. The method is based on wing measurements. It requires the identification of 19 points on the forewing image. The method described here was implemented in the IdentiFly software (Tofilski 2017), which is able to identify 19 subspecies and four evolutionary lineages of the honey bee. The software and the data required for the identification are freely available to the

general public, including scientists, honey bee breeders and beekeepers. This should make the identification of honey bee subspecies easier and help in the efforts of their conservation.

## 2. MATERIALS AND METHODS

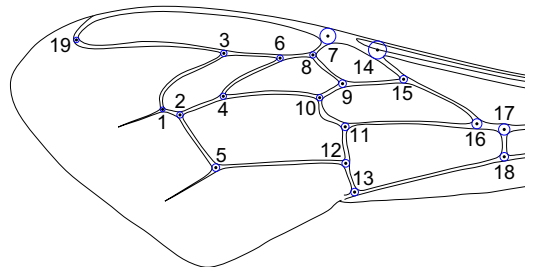
We have used images of honey bee worker forewings obtained from the Morphometric Bee Data Bank in Oberursel, Germany. Our dataset was similar to that used in the study of Kandemir et al. (2011) and included 1849 workers from 187 colonies, 25 subspecies and 4 evolutionary lineages (Supplementary Table 1). With very few later additions, this study used the samples, which the analyses in Ruttner's (1988) monography were based on. However, only the part of these samples where wing mounts had been preserved could be included, as the point coordinates used in geometric morphometry cannot be reconstructed from the wing venation angles stored in the original datasets. In order to base our study on a set of well-defined reference samples, we accepted restricted sample numbers in some subspecies, and the wings were re-measured to comply with our current method. We have updated some of the subspecies names used in Ruttner's monography according to the more recent Engel (1999) review (*A. m. iberiensis* for *A. m. iberica*, *A. m. jemenitica* for *A. m. yemenitica*) and added *A. m. ruttneri* (Sheppard et al. 1997). Affiliation of the subspecies to the lineages varies between studies (Ruttner 1988; Garnery et al. 1993; Franck et al. 2001; Kandemir et al. 2011; Meixner et al. 2013). We have used a more recent version of the honey bee taxonomy based in part on genetic studies (Meixner et al. 2013).

The available reference data set varied in the number of colonies per subspecies, given in brackets. In lineage A, there were the following subspecies: *A. m. adansonii* (12), *A. m. capensis* (3), *A. m. intermissa* (6), *A. m. lamarckii* (7), *A. m. litorea* (6), *A. m. major* (1), *A. m. monticola* (8), *A. m. ruttneri* (5), *A. m. sahariensis* (2), *A. m. scutellata* (14), *A. m. unicolor* (7) and *A. m. jemenitica* (14). In lineage C, there were the following subspecies: *A. m. carnica* (15), *A. m. cecropia* (9), *A. m. ligustica* (11) and *A. m. macedonica* (2). In lineage M, there were the

following subspecies: *A. m. iberiensis* (2) and *A. m. mellifera* (14). In lineage O, there were the following subspecies: *A. m. adami* (5), *A. m. anatoliaca* (5), *A. m. armeniaca* (6), *A. m. caucasica* (12), *A. m. cyprica* (4), *A. m. meda* (8) and *A. m. syriaca* (9). Each colony was represented by approximately ten workers. Subspecies with a sample size smaller than five colonies (*A. m. capensis*, *A. m. cyprica*, *A. m. iberiensis*, *A. m. macedonica*, *A. m. major* and *A. m. sahariensis*) were used only for the discrimination of lineages and they were not included in the discrimination of subspecies.

Wing pictures were obtained from the Oberursel collection, where they had been taken from mounts on microscopic slides using a camera attached to a stereomicroscope with a resolution of 650 pixels per centimetre (1651 pixels per inch). Each wing was measured by manual identification of 19 characteristic points (later referred to as landmarks, Figure 1). The position of the landmarks was similar to a previous study (Gerula et al. 2009) but was changed slightly for landmarks 7 and 14 in order to agree with that of standard morphometry (Ruttner 1988).

The coordinates of the landmarks were superimpositioned using full Procrustes fit (Dryden and Mardia 1998) in MorphoJ (Klingenberg 2011). Only the aligned coordinates, and not the wing size, were used for statistical analysis in Statistica software v.12 (StatSoft Inc 2011). The analysis was based on average values calculated for each colony. Canonical variate analysis was used for the discrimination of the lineages and subspecies. The correctness of the



**Figure 1.** Position of 19 landmarks on the forewing of honey bee workers. The landmarks should be placed in such a way that circles around them are tangent to the venation outline in at least three points.

classification was verified using leave-one-out cross-validation in the PAST 3.11 software (Hammer et al. 2001). All reported classification rates were cross-validated.

One of the methods, which can be used for the classification of an unknown individual, is linear discriminant analysis (LDA). In this method, a set of features of the unknown individual is measured, and the measurements are used together with the discriminant functions to calculate the discriminant scores for each of the groups (Fisher 1936). In honey bees, colonies, and not individuals, are usually classified and the mean values of all workers from the colony are used. This allows to achieve higher precision because measurement errors are reduced. The information that is required for classification using this method is the reference configuration and classification functions. Those data are provided as supplementary material (Supplementary Tables 2 and 3) and can be used for classification in both IdentiFly and other software, which might be created in the future. The classification is based on superimposed landmarks. The first step of the classification is Procrustes superposition of the landmarks with the reference configuration. The next step is obtaining discriminant scores for each of the subspecies or lineages. The score is calculated by multiplying the superimposed coordinates with the corresponding discriminant functions and summing them together, including the constant of the discriminant functions. The colony is classified as the subspecies or lineage with the highest classification score. More informative than the classification score is the probability of belonging to a particular class. The probability provided by the IdentiFly software was based on the Mahalanobis distance between an identified colony and the mean of particular classes in the canonical space. For example, if a colony is identified as *A. m. mellifera* with a probability of 0.01, it means that 1% of colonies in the reference sample of *A. m. mellifera* is expected to be at a greater distance from the average of this subspecies than the identified colony. In short, a high probability of classification as a particular subspecies means higher similarity of the identified colony to the reference sample of this subspecies. The IdentiFly (Tofilski 2017) software can be

downloaded from <http://drawwing.org/identify>. It is free for non-commercial use, including research, education and personal use by beekeepers.

### 3. RESULTS

#### 3.1. Evaluation of the reference database

In the reference data set, the shape of the forewing venation differed markedly between evolutionary lineages (MANOVA  $F = 34$ ;  $P < 0.0001$ ). In pairwise comparisons, all lineages differed significantly from each other (Table I). The largest difference was between lineage C and lineage M (Table I). The smallest difference was between lineages A and lineage O (Table I).

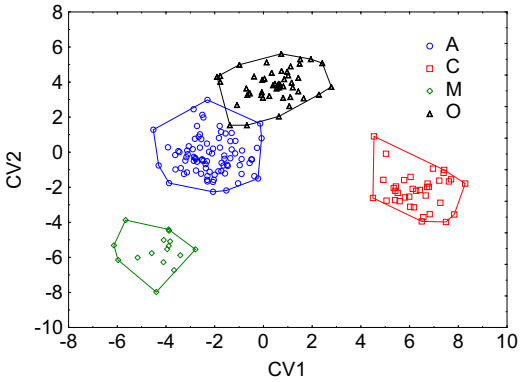
Three of the honey bee lineages (A, C and M) were classified correctly without error (100%) (Figure 2). In lineage O, one *A. m. anatoliaca* colony was incorrectly classified as lineage A. Overall, the cross-validated correct identification rate of lineages was 99.5%.

The shape of the forewing venation also differed significantly among subspecies (MANOVA  $F = 7$ ;  $P < 0.0001$ ). In pairwise comparisons, all subspecies differed significantly from each other (Table II). The largest difference was between *A. m. mellifera* and *A. m. cecropia* (Table II). The smallest difference was between *A. m. jemenitica* and *A. m. litorea* (Table II). Overall, the cross-validated correct identification rate of subspecies was 88.4%. Nine subspecies were classified without error: *A. m. adami*, *A. m. anatoliaca*, *A. m. caucasica*, *A. m. lamarckii*, *A. m. meda*, *A. m. mellifera*, *A. m. ruttneri*,

**Table I.** Squared Mahalanobis distances between honey bee evolutionary lineages (lower triangle) and statistical significance of pairwise comparisons between the lineages (upper triangle)

Evolutionary lineage	A	C	M	O
A	–	***	***	***
C	77.9	–	***	***
M	75.9	149.1	–	***
O	37.3	75.5	119.1	–

\*\*\* $P < 0.0001$



**Figure 2.** Discrimination of honey bee evolutionary lineages based on canonical variate analysis of honey bee forewing shape.

*A. m. syriaca* and *A. m. unicolor*. In the other subspecies, the correct identification rate ranged from 93 to 50% (Table III).

Identification rates were high in lineages C and O (91.4 and 97.7%, respectively), where subspecies formed well-separated clusters (Figures 3 and 4). However, they were low (79.7%) in lineage A where most subspecies from central and southern Africa are similar to each other and could not be resolved reliably (Figure 5).

### 3.2. Program description

To use the *IdentiFly* software for the identification of lineages and subspecies, forewing images of at least ten workers from one colony need to be obtained as described in the “[Materials and methods](#)” section. Each of the wings has to be measured by indicating 19 landmarks in the appropriate part of the wing. In order to do this, the software should be in the mode “Add landmarks”. In this mode, clicking the image will result in adding a landmark. Careful positioning of the landmarks (Figure 1) is essential for correct classification. The positions of the landmarks are then saved within the file and can later be verified, corrected and exported to other software. When all wing images from one colony are measured, the colony can be classified. Though it is also possible to classify a single wing from one worker instead of colony samples, this is not recommended because the results are unreliable (Tofilski 2008). In order to properly assign a sample to a

subspecies, the classification probabilities, as displayed in a results window, should be taken into account, where a high probability of assignment indicates a high similarity of the classified colony to a particular group of reference samples.

As an example, we present the classification results of one colony collected in Cyprus. First, we have used the lineage classification file and the colony was classified as O lineage (Suppl. Fig. 1). The probability of classification to this group was 0.16, which is much higher than the values for lineages A, C and M:  $2.0 \times 10^{-10}$ ,  $1.1 \times 10^{-8}$  and  $1.5 \times 10^{-17}$ , respectively. In this case, it is safe to conclude that the identified colony was from lineage O. In the graph showing CV1 and CV2, the black point representing the colony is within the O ellipse.

Next, we have used the subspecies classification file, and the same colony was classified as *A. m. armeniaca* (Suppl. Fig. 2). The classification probability ( $1.5 \times 10^{-3}$ ) is much lower than previously. The second highest probability ( $1.3 \times 10^{-8}$ ) is for *A. m. syriaca*. In this case, we should conclude that the identified colony is similar to *A. m. armeniaca* but markedly different from the reference sample of this subspecies, and it either is a hybrid with *A. m. syriaca* or it belongs to a subspecies not covered by the identification model. We know that the latter explanation is true, because by using full standard morphometry, this colony had been identified as *A. m. cyprica*. The colony was correctly classified as belonging to lineage O. However, it could not be classified as *A. m. cyprica* because this subspecies was not included in the classification model due to a small sample size.

As a second example of identification, we have used a colony from Austria. The colony was classified as lineage M (Suppl. Fig. 3) with a probability of 0.42. This number is much higher than the probabilities for other lineages, which were smaller than  $10^{-11}$ ; therefore, it is safe to conclude that the classification was correct. When the subspecies classification file was used, the same colony was classified as *A. m. mellifera* (Suppl. Fig. 4) with a probability of 0.97. The probabilities for all other subspecies were smaller than  $10^{-12}$ . In this situation, it can be concluded that the subspecies classification is also correct. However, in

**Table II.** Squared Mahalanobis distances between honey bee subspecies (lower triangle) and statistical significance of pairwise comparisons between the subspecies (upper triangle)

Subspecies	<i>A. m. adami</i>	<i>A. m. adansonii</i>	<i>A. m. anatoliaca</i>	<i>A. m. armeniaca</i>	<i>A. m. carnica</i>	<i>A. m. caucasica</i>	<i>A. m. cecropia</i>	<i>A. m. intermissa</i>	<i>A. m. lamareckii</i>	<i>A. m. ligustica</i>
<i>A. m. adami</i>	—									
<i>A. m. adansonii</i>	154.8	—								
<i>A. m. anatoliaca</i>	81.8	45.2	—							
<i>A. m. armeniaca</i>	136.6	69.2	44.2	—						
<i>A. m. carnica</i>	193.4	184.2	139.5	129.8	—					
<i>A. m. caucasica</i>	148.8	92.5	47.9	53.9	175.3	—				
<i>A. m. cecropia</i>	203.7	164.7	140.1	136.0	30.5	185.3	—			
<i>A. m. intermissa</i>	71.9	154.4	97.2	142.8	174.9	158.1	216.3	—		
<i>A. m. lamareckii</i>	60.5	139.1	87.5	149.8	171.6	120.1	190.6	56.5	—	
<i>A. m. ligustica</i>	193.3	162.7	133.4	108.4	26.0	153.5	56.7	146.6	174.0	—
<i>A. m. litorea</i>	15.7	129.6	73.0	126.7	200.2	143.7	217.1	66.7	58.1	199.3
<i>A. m. meda</i>	134.1	59.2	68.4	47.6	164.4	114.6	146.4	133.1	148.9	147.8
<i>A. m. mellifera</i>	125.2	180.3	121.4	201.8	208.3	180.7	240.8	84.5	87.0	190.8
<i>A. m. monticola</i>	26.0	126.2	67.8	130.1	185.5	123.1	185.2	99.9	59.8	196.7
<i>A. m. rutheri</i>	96.0	164.9	93.9	132.3	122.3	147.6	176.1	72.0	90.0	135.5
<i>A. m. scutellata</i>	14.2	106.7	50.5	88.2	156.6	111.0	159.5	65.8	61.4	150.9
<i>A. m. syriaca</i>	63.7	85.9	53.6	86.2	190.3	106.4	196.8	104.6	75.0	201.3
<i>A. m. unicolor</i>	105.7	120.0	99.1	117.6	209.0	160.9	215.4	105.4	120.5	198.3
<i>A. m. jementica</i>	24.5	141.2	84.9	136.8	186.9	161.7	212.3	66.8	67.3	195.9



Table II (continued)

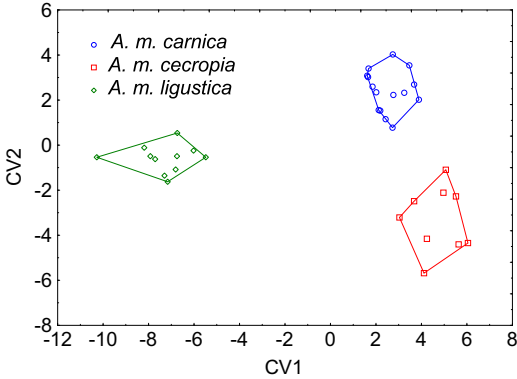
Subspecies	<i>A. m. litorea</i>	<i>A. m. meda</i>	<i>A. m. mellifera</i>	<i>A. m. monticola</i>	<i>A. m. rutneri</i>	<i>A. m. scutellata</i>	<i>A. m. syriaca</i>	<i>A. m. unicolor</i>	<i>A. m. jemenitica</i>
<i>A. m. adami</i>	***	***	***	***	***	***	***	***	***
<i>A. m. adansonii</i>	*	***	***	***	***	***	***	***	***
<i>A. m. anatoliaca</i>	***	***	***	***	***	***	***	***	***
<i>A. m. armeniaca</i>	***	***	***	***	***	***	***	***	***
<i>A. m. carnica</i>	***	***	***	***	***	***	***	***	***
<i>A. m. caucasica</i>	***	***	***	***	***	***	***	***	***
<i>A. m. cecropia</i>	***	***	***	***	***	***	***	***	***
<i>A. m. intermissa</i>	***	***	***	***	***	***	***	***	***
<i>A. m. lamareckii</i>	***	***	***	***	***	***	***	***	***
<i>A. m. ligustica</i>	***	***	***	***	***	***	***	***	***
<i>A. m. litorea</i>	—	***	***	***	***	***	***	***	*
<i>A. m. meda</i>	125.6	—	***	***	***	***	***	***	***
<i>A. m. mellifera</i>	120.4	220.3	—	***	***	***	***	***	***
<i>A. m. monticola</i>	26.4	118.7	129.2	—	***	***	***	***	***
<i>A. m. rutneri</i>	80.2	150.8	122.3	96.5	—	***	***	***	***
<i>A. m. scutellata</i>	21.4	85.0	133.2	23.0	87.0	—	***	***	***
<i>A. m. syriaca</i>	52.4	63.1	172.5	61.7	109.1	48.7	—	***	***
<i>A. m. unicolor</i>	95.1	97.3	171.0	97.1	152.5	77.9	114.4	—	***
<i>A. m. jemenitica</i>	15.4	127.7	133.4	43.5	88.9	28.6	55.9	97.6	—

\**P* < 0.05; \*\*\**P* < 0.001

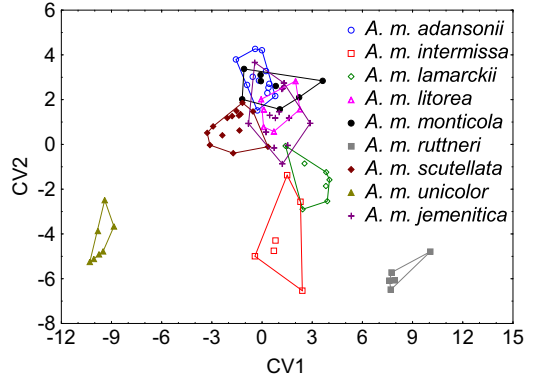
**Table III.** Cross-validated classification rate of honey bee subspecies. Only subspecies with classification rate lower than 100% were included

Subspecies	Classified as											Correctly classified %
	<i>A. m. adansonii</i>	<i>A. m. anatoliaca</i>	<i>A. m. armeniaca</i>	<i>A. m. carnica</i>	<i>A. m. cecropia</i>	<i>A. m. intermissa</i>	<i>A. m. ligustica</i>	<i>A. m. litorea</i>	<i>A. m. monticola</i>	<i>A. m. scutellata</i>	<i>A. m. jemenitica</i>	
<i>A. m. adansonii</i>	11	0	0	0	0	0	0	1	0	0	0	91.6
<i>A. m. armeniaca</i>	0	1	5	0	0	0	0	0	0	0	0	83.3
<i>A. m. carnica</i>	0	0	0	14	1	0	0	0	0	0	0	93.3
<i>A. m. cecropia</i>	0	0	0	1	8	0	0	0	0	0	0	88.8
<i>A. m. intermissa</i>	0	0	0	0	0	5	0	1	0	0	0	83.3
<i>A. m. ligustica</i>	0	0	0	1	0	0	10	0	0	0	0	90.9
<i>A. m. litorea</i>	1	0	0	0	0	0	0	3	0	2	0	50.0
<i>A. m. monticola</i>	1	0	0	0	0	0	0	1	6	0	0	75.0
<i>A. m. scutellata</i>	2	0	0	0	0	0	0	1	0	10	1	71.4
<i>A. m. jemenitica</i>	1	0	0	0	0	0	0	4	0	0	9	64.2





**Figure 3.** Discrimination of honey bee subspecies from lineage C.



**Figure 5.** Discrimination of honey bee subspecies from lineages A.

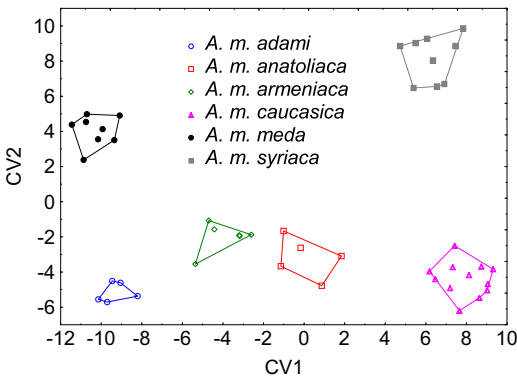
M, there are two subspecies *A. m. mellifera* and *A. m. iberiensis*, with the latter one not included in the classification model due to a small sample size. However, Austria is far outside the natural distribution range of *A. m. iberiensis*; thus, the colony is most probably *A. m. mellifera*, unless it is *A. m. iberiensis* imported by a beekeeper.

#### 4. DISCUSSION

The results presented here show that measurements of the forewings of honey bee workers can be successfully used for the identification of evolutionary lineages and subspecies based on the newly developed Identifly software. This can be useful for the conservation of endangered

subspecies. In a practical context, this may help beekeepers eliminate colonies that differ the most from their preferred native subspecies. The average identification rate of subspecies was close to 90%. This relatively high error margin is mainly due to the A lineage subspecies, where only 79% of colonies were correctly identified. If lineage A subspecies were excluded, the average identification rate would be satisfactorily high (95%). In particular, the identification rates for the commercially used bee subspecies (*A. m. mellifera*, *A. m. carnica* and *A. m. caucasica*) were satisfactory (Table III). Only the identification of *A. m. ligustica* is below average (Table III). In comparison to the subspecies, the identification of evolutionary lineages is much more reliable. The discrimination of evolutionary lineages can be particularly useful for the identification of Africanized bees. We did not find a general improvement of subspecies classification results from a two-step procedure, determining the lineage first, and then determining the subspecies within that particular lineage.

The obtained results are consistent with those obtained in earlier studies (Kauhausen-Keller et al. 1997; Kandemir et al. 2011) using similar research material. The earlier studies demonstrated that the identification of the subspecies was possible, but it did not provide enough information for the identification of unknown samples. A lack of this information hindered the research on honey bee biogeography and the conservation of endangered subspecies (Meixner et al. 2013). The



**Figure 4.** Discrimination of honey bee subspecies from lineage O.

main novelty of this study is to provide data and methods for the identification of honey bee evolutionary lineages and subspecies. The new method is fast and easily available to beekeepers, which should play a key role in the conservation of honey bee subspecies. The alternative method based on standard morphometry (Ruttner 1988) is more time consuming and reference samples of honey bee subspecies are not widely available.

Although the identification of honey bee subspecies based on molecular methods became more available recently, identification based on morphological measurements has an indispensable place. For one, subspecies definitions are historically linked to morphology, and molecular identification needs to be validated against these definitions. Further on, wing morphometry has some advantages over molecular methods. The identification based on wing measurements is relatively easy and fast. In order to obtain the wing images, either a scanner or a digital camera can be used. It is possible to obtain wing images without detaching them from the valuable specimen (Houle et al. 2003), but in most cases, it is easier and faster to dissect them. The equipment required is relatively inexpensive, and nowadays, it is accessible to both scientists and beekeepers.

The number and the position of landmarks used for the wing measurements differed between studies (Meixner et al. 2013). In some studies, the landmark 19 at the distal part of the marginal cell was not included (point 19, Figure 1) (Tofilski 2004). The position of this landmark also differed between studies and it was located either at the front margin of the wing (Kandemir et al. 2011) or at the apex of the marginal cell (Miguel et al. 2011). The position of landmark 14 also varied between studies. Its position was either in the middle of the radius vein (Tofilski 2008; Gerula et al. 2009; Miguel et al. 2011; Kandemir et al. 2011) or at the midpoint between the posterior edge of the radius vein and the anterior edge of the wing (Ruttner 1988). The numbering of the landmarks also varies between studies (Tofilski 2008; Gerula et al. 2009; Miguel et al. 2011; Kandemir et al. 2011). It was suggested that the position and numbering of the landmarks should be standardised in order to facilitate a comparison between studies (Meixner et al. 2013). The

landmark numbering scheme presented here may become such a standard because the data can be easily exported to other formats.

The software provides information about the similarity of a colony to different groups. However, it is difficult to make the decision if an investigated colony belongs to a particular subspecies or not. Bees belonging to different subspecies can interbreed and produce hybrids, and there is possibly a range of intermediate phenotypes between the typical phenotypes of “pure” subspecies. Further, a particular sample may not be a member of any of the subspecies in the data set. The interpretation of the allocation probabilities may give some indications. In an optimal outcome, the probability for one subspecies will be high, and in all remaining subspecies—very low. In this clear case, the focal colony is without doubt classified as that particular subspecies or lineage. However, it is difficult to recommend a fixed threshold value of the classification probability above which an allocation can be considered unequivocal. In another scenario, classification probabilities might be relatively high in two or several subspecies. In this case, the focal colony should be considered as a hybrid between the subspecies with high classification probabilities. Finally, the classification probability can be low in all subspecies. In this case, the classification of the focal colony should be concluded as being doubtful. This can happen in two cases: when the focal colony is very different from all reference samples or when there is an error in the measurements. The first case can occur when the colony belongs to species other than *A. mellifera*, for example, *A. cerana*, or the colony belongs to subspecies not included in the classification model, for example, *A. m. pomonella*. This should not happen very often because the classification presented here covers most of the honey bee subspecies. Low classification probabilities in all subspecies most often occur because there is an error in the measurements, most commonly by swapping the position of two landmarks. It is important to notice that the classification probabilities indicate the similarity of the investigated colony to a reference sample. Unfortunately, the reference sample set of some subspecies is small and may not cover their entire variation range. A colony within the true

variation range of a subspecies may thus be assigned only a low similarity to a reference sample.

The current analysis clearly pointed out difficulties with the discrimination of some subspecies. This might partly come from the restriction of the analysis of wing venation (as for example the confusion of *A. m. ligustica* with *A. m. carnica*, where colour is a distinctive discriminative body character), but partly also points to deficiencies in the current state of knowledge, as in the African bees, where morphological variation and distribution patterns are notoriously defying distinctive subspecies classifications (Hepburn and Radloff 1998).

A crucial component of the identification tool is the selection of reference samples on which the classification is based. We have used a part of the samples, which were used by Ruttner (1988) in his monography on the taxonomy of honey bees. This first comprehensive description of honey bee subspecies is still widely accepted and still provides a valid framework, which stood the test of time. It thus can be regarded as a lucky circumstance that the original wing mounts of that crucial first morphometric description of the subspecies were still available. Most samples dated back to the times of less intense colony transportation, and quite a number originated from Brother Adam collections (Brother Adam 1983). To this set, we added *A. m. ruttneri*, but not the more recently described *A. m. pomonella* (Sheppard and Meixner 2003) and *A. m. simensis* (Meixner et al. 2011).

Though well documented and preserved in the Oberursel collection and data bank, and covering almost all known subspecies, this reference data set also faces obvious limitations. The most apparent limitation is a low sample size, which compelled us to exclude six subspecies from the set. The estimation of the shape based on geometric morphometrics requires relatively large sample sizes (Cardini and Elton 2007). It is recommended that, in multivariate analysis, the sample size of each group should markedly exceed the number of variables (Arnold 1983). In the case of honey bee wings described by 19 landmarks, the

sample size should be markedly larger than 38 colonies, because each landmark consists of two coordinates. This criterion is not met in the case of subspecies. In order to obtain more precise results, the sample size for most of the subspecies should be increased markedly.

A second limitation comes from the unsystematic coverage of the subspecies' geographic range. Within populations believed to belong to one subspecies, some variation occurs, as in black bees (*A. m. mellifera*) from Western and Eastern Europe. This problem also calls for an enlargement of the reference data set, but such a crucial step should be considered with utmost care. Currently, there is no general consensus about subspecies reference samples, and the definition of such a set should be based on a more general experts' agreement. A future reference sample set should be large and cover the whole range of a subspecies. This may be facing some difficulties because the exact borders of the distribution may be unclear; the feral populations can be extinct and geographic variation could be affected strongly by beekeeping (De la Rúa et al. 2009). In addition, some of the subspecies may occur in remote and inaccessible geographic regions, or the present subspecies definitions are under debate. In this situation, the preliminary results about subspecies identification were provided here in hope that they will prove to be useful; however, they should be interpreted with some caution until a more expanded and validated reference sample set is available.

## 5. CONCLUSIONS

The Identifly computer software presented here was shown to be highly effective in the discrimination of the evolutionary lineages of honey bees. It was also satisfactorily effective in discriminating most honey bee subspecies, though it failed in some, particularly within the African A lineage. The reference sample size of many honey bee subspecies is relatively low and needs to be increased in order to achieve higher precision of identification. The software is freely available and can easily be used by both scientists and beekeepers.

## ACKNOWLEDGEMENTS

We thank the editor Marina Meixner and two anonymous reviewers for the helpful comments on an earlier version of the manuscript. This study was supported by the Polish National Science Centre (NCN) grant number UMO-2013/10/E/NZ9/00682 and the Ministry of Science and Higher Education grant number DS-3500.

## AUTHOR'S CONTRIBUTION

AN performed wing measurements. AN, İK, SF and AT equally contributed to analysis of data and writing of the manuscript. All authors read and approved the final manuscript.

## OPEN ACCESS

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

**Logiciel d'identification des sous-espèces et des lignées évolutives de l'abeille.**

***Apis mellifera* / détermination à la sous-espèce / morphométrie géométrique / morphométrie de l'aile**

**Ein Computerprogramm zur Identifizierung von Unterarten und Evolutionslinien der Honigbiene**

**Honigbiene / Unterscheidung von Unterarten / geometrische Morphometrie / Flügelmorphometrie**

## REFERENCES

Alburaki, M., Moulin, S., Legout, H., Alburaki, A., Garnery, L. (2011) Mitochondrial structure of Eastern honeybee populations from Syria, Lebanon and Iraq. *Apidologie* **42**, 628–641

Arnold, S.J. (1983) Morphology, performance and fitness. *Am. Zool.* **23**, 347–361

Batra, S.W.T. (1988) Automatic image analysis for rapid identification of Africanized honey bees. in: Needham G.R., Page R.E., Delfinado-Baker M., Bowman C.E.

(Eds.), Africanized honey bees and bee mites. Ellis Horwood Limited, Chichester, pp. 260–263

Baylac, M., Garnery, L., Tharavy, D., Pedraza-Acosta, J., Rortais, A., Arnold, G. (2008) ApiClass, an automatic online wing morphometric expert system for honey bee worker identification. Retrieved from <http://apiclass.mnhn.fr>

Bouga, M., Kiliyas, G., Harizanis, P.C., Papatotiropoulos, V., Alahiotis, S. (2005) Allozyme variability and phylogenetic relationships in honey bee (Hymenoptera: Apidae: *Apis mellifera*) populations from Greece and Cyprus. *Biochem. Genet.* **43**, 471–483

Bouga, M., Alaux, C., Bienkowska, M., Büchler, R., Carreck, N. L., et al. (2011) A review of methods for discrimination of honey bee populations as applied to European beekeeping. *J. Apic. Res.* **50**, 51–84.

Brother Adam (1983) In Search of the best Strains of Bees, and the Results of the Evaluations of the Crosses and Races. Northern Bee Books, UK; Dadant & Sons, USA

Cardini, A., Elton, S. (2007) Sample size and sampling error in geometric morphometric studies of size and shape. *Zoomorphology* **126**, 121–134

Chen, C., Liu, Z., Pan, Q., Chen, X., Wang, H., Guo, H., Liu, S., Lu, H., Tian, S., Li, R., Shi, W. (2016) Genomic analyses reveal cenographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinixinyuan* n. ssp. *Mol. Biol. Evol.* **33**, 1337–1348

De la Rúa, P., Jaffé, R., Dall'Olio, R., Muñoz, I., Serrano, J. (2009) Biodiversity, conservation and current threats to European honeybees. *Apidologie* **40**, 263–284

Dryden, I.L., Mardia, K.V. (1998) Statistical Shape Analysis. J. Wiley, Chichester

DuPraw, E.J. (1965a) The recognition and handling of honeybee specimens in non-Linnean taxonomy. *J. Apic. Res.* **4**, 71–84

DuPraw, E.J. (1965b) Non-Linnean taxonomy and the systematics of honeybees. *Syst. Zool.* **14**, 1–24

Engel, M. S. (1999) The Taxonomy of Recent and Fossil Honey Bees (Hymenoptera: Apidae: *Apis*). *J. Hymenoptera Res.* **8** (2), 165–169

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

Franck, P., Garnery, L., Solignac, M., Cornuet, J.-M. (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): New insights from microsatellite and mitochondrial data. *Evolution* **52**, 1119–1134

Franck, P., Garnery, L., Loiseau, A., Oldroyd, B.P., Hepburn, H.R., et al. (2001) Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. *Heredity* **86**, 420–430

Francoy, T.M., Wittmann, D., Drauschke, M., Müller, S., Steinhage, V., Bezerra-Laure, M.A., De Jong, D., Gonçalves, R.L.S. (2008) Identification of Africanized honey bees through wing morphometrics: two fast and efficient procedures. *Apidologie* **39**, 488–494

- Garnery, L., Solignac, M., Celebrano, G., Cornuet, J.-M. (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. *Experientia* **49**, 1016–1021
- Gerula, D., Tofilski, A., Węgrzynowicz, P., Skowronek, W. (2009) Computer-assisted discrimination of honey bee subspecies used for breeding in Poland. *J. Apic. Sci.* **53**, 105–114
- Hammer, Ø., Harper, D.A.T., Ryan, P.D. (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **4**(1), 9
- Hepburn, H. R., Radloff, S. E. (1998) *Honeybees of Africa*. Springer, Berlin
- Houle, D., Mezey, J., Galpern, P., Carter, A. (2003) Automated measurement of *Drosophila* wings. *BMC Evol. Biol.* **3**, 1
- 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. *Mol. Ecol.* **14**, 93–106
- Kandemir, İ., Özkan, A., Fuchs, S. (2011) Reevaluation of honeybee (*Apis mellifera*) microtaxonomy: a geometric morphometric approach. *Apidologie* **42**, 618–627
- Kauhausen-Keller, D., Keller, R. (1994) Morphometrical control of pure race breeding in the honeybee (*Apis mellifera* L.). *Apidologie* **25**, 133–133
- Kauhausen-Keller, D., Ruttner, F., Keller, R. (1997) Morphometric studies on the microtaxonomy of the species *Apis mellifera* L. *Apidologie* **28**, 295–308
- Klingenberg, C.P. (2011) MorphoJ: an integrated software package for geometric morphometrics. *Mol. Ecol. Res.* **11**, 353–357
- Meixner, M. D., Leta, M. A., Koeniger, N., Fuchs, S. (2011) The honey bees of Ethiopia represent a new subspecies of *Apis mellifera*—*Apis mellifera simensis* n. ssp. *Apidologie* **42**, 425–437
- 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*. *J. Apic. Res.* **52**, 1–28
- Miguel, I., Baylac, M., Iriondo, M., Manzano, C., Garnery, L., Estonba, A. (2011) Both geometric morphometric and microsatellite data consistently support the differentiation of the *Apis mellifera* M evolutionary branch. *Apidologie* **42**, 150–161
- Moritz, R.F., Kraus, F.B., Kryger, P., Crewe, R.M. (2007) The size of wild honeybee populations (*Apis mellifera*) and its implications for the conservation of honeybees. *J. Insect Conserv.* **11**, 391–397
- Oleksa, A., Tofilski, A. (2015) Wing geometric morphometrics and microsatellite analysis provide similar discrimination of honey bee subspecies. *Apidologie*, **46**, 49–60
- Prabucki, J., Samborski, J., Chuda-Mickiewicz, B. (2002) The use of three taxonomic characters for race identification of Middle European bee. *J. Apic. Sci.* **46**, 41–46
- Rinderer, T.E., Daly, H.V., Sylvester, H.A., Collins, A.M., Bucu, S.M., Hellmich, R.L., Danka, R.G. (1990) Morphometric differences among Africanized and European honey bees and their F1 hybrids (Hymenoptera: Apidae). *Ann. Entomol. Soc. Am.* **83**, 346–351
- Rostecki, P., Samborski, J., Prabucki, J., Chuda-Mickiewicz, B. (2007) A comparison of various hardware for the measurement of the cubital index. *J. Apic. Sci.* **51**, 49–53
- Ruttner, F. (1988) *Biogeography and Taxonomy of Honeybees*. Springer, Berlin
- Samborski, J., Prabucki, J., Chuda-Mickiewicz, B., Perużyński, G. (2002) Operation rate and sensitivity of devices used for determining cubital index value. *J. Apic. Sci.* **46**, 35
- Sheppard, W. S., Meixner, M.D. (2003) *Apis mellifera pomonella*, a new honeybee subspecies from central Asia. *Apidologie* **34**, 367–375
- Sheppard, W. S., Soares, A. E. E., DeJong, D., Shimanuki, H. (1991) Hybrid status of honey bee populations near the historic origin of Africanization in Brazil. *Apidologie* **22**, 643–652.
- Sheppard, W. S., Arias, M. C., Grech, A., Meixner, M. D. (1997) *Apis mellifera ruttneri*, a new honey bee subspecies from Malta. *Apidologie* **28**, 287–293
- Sheppard, W. S., Rinderer, T. E., Garnery, L., Shimanuki, H. (1999) Analysis of Africanized honey bee mitochondrial DNA reveals further diversity of origin. *Genet. Mol. Biol.* **22**, 73–75
- Smith, D.R., Brown, W.M. (1990) Restriction endonuclease cleavage site and length polymorphisms in mitochondrial DNA of *Apis mellifera mellifera* and *A. m. carnica* (Hymenoptera: Apidae). *Ann. Entomol. Soc. Am.* **83**, 81–88
- StatSoft Inc. (2011) STATISTICA (data analysis software system), version 10
- Tofilski, A. (2004) DrawWing, a program for numerical description of insect wings. *J. Insect Sci.* **4**, 17
- Tofilski, A. (2005) Automatic measurements of honeybee wings. in: MacLeod N. (Ed.), *Automated taxon identification in systematics: theory, approaches and applications*. CRC Press, Boca Raton, pp. 289–298
- Tofilski, A. (2008) Using geometric morphometrics and standard morphometry to discriminate three honeybee subspecies. *Apidologie* **39**, 558–563
- Tofilski, A. (2017) Identifly software, version 0.31. available at <http://drawwing.org/identifly>