COMPUTER-ASSISTED DISCRIMINATION OF HONEYBEE SUBSPECIES USED FOR BREEDING IN POLAND

Dariusz Gerula[1], Adam Tofilski[2], Paweł Węgrzynowicz[1], Wojciech Skowronek[1]

[1]Institute of Pomology and Floriculture, Apiculture Division,
Kazimierska 2, 24-100 Puławy, Poland, e-mail: dariusz.gerula@man.pulawy.pl
[2] Department of Pomology and Apiculture,
Agricultural University, 29 Listopada 54, 31-425 Krakow, Poland,
e-mail:rotofils@cyf-kr.edu.plor

Received 09 November; accepted 02 December 2009

Summary

The aim of this study was to develop a method for automated discrimination of three honeybee subspecies used for breeding in Poland: *A. mellifera carnica*, *A. m. caucasica* and *A. m. mellifera*. The method developed exploits characteristic venation of forewings. Samples of bees were collected from 1042 bee colonies of the three subspecies. From each of the colonies 20 right forewings were dissected and scanned with Nikon Coolscan 5000 ED. For every wing image the coordinates of 19 points were determined automatically using DrawWings software. Within each of the colonies the points were aligned using the Procrustes method, and the average configuration of points were calculated. Two methods of subspecie distinction are presented; one based on discriminant functions and the other based on canonical coefficients. Computer software was developed for automated discrimination of the three honeybee subspecies used for breeding in Poland.

Keywords: honeybee, *Apis mellifera*, subspecies, wing.

INTRODUCTION

Apis mellifera carnica, *A. m. caucasica* and *A. m. mellifera* are currently used in the national queen breeding program in Poland. Breeding lines have to be declared as belonging to one of these subspecies and therefore colonies are tested for identity. However, identification based on measurements of proboscis length, width of tergites and cubital index (Gromisz, 1967, 1981) are time consuming because of laborious preparations.

When measurements of many different body parts, e.g. forewing venation, proboscis length, wax plate size, number of hooks on hind wing and pigmentation of tergites were employed at the same time (Alpatov, 1929; DuPraw, 1964; Ruttner et al., 1978, 1988) accuracy of the discrimination increased but the time spent also increased. It is possible to identify honeybee subspecies on a forewing venation only. The cubital index (Goetze, 1940; Alpatov, 1948) and discoidal shift (Goetze, 1959) were used at first. Later, the venation was described with large number of angles between veins (DuPraw, 1964, 1965; Ruttner et al., 1978; Kauhausen-Keller, 1991). Originally the measurements of the wing venation were made using a microscope, but there were attempts to automate them (Daly et al., 1982; Batra, 1988). Indicating points on a computer screen can be used to calculate cubital index (Rostecki et al., 2007) or distances and

angles between wing veins (Bieńkowska and Szymula, 2003). There is also computer software for fully automated measurements of wings (Tofilski, 2004). The software detects characteristic points, which can be used to calculate distances and angles (Tofilski, 2008). The automated measurements are faster and more precise than manual measurements (Tofilski, 2007).

The aim of this study was to develop a software for automated discrimination of three honeybee subspecies used for queen breeding in Poland. The software should make the discrimination faster and more precise.

MATERIAL AND METHODS

A. m. carnica, *A. m. caucasica* and *A. m. mellifera* workers were sampled from 1042 colonies which represented 45 breeding lines kept in 30 Polish apiaries in August 2007. Samples were collected from the nest center of each of the colonies and were stored in alcohol.

First it was verified if the subspecies declared by breeders agree with the criteria of Gromisz (1981). Measurements were taken under microscope. Length of proboscis, width of fourth tergite and cubital index were made using ocular micrometer (Bornus et al.,1966). From every colony, at least 20 workers were measured. Only the colonies which agree with the criteria of Gromisz (1981) were selected for further analysis. There were 758 carniolan, 98 caucasian and 114 black bee colonies. The wings were mounted in glass photographic frames (Rowi 260). Then they were scanned with a Nikon Coolscan 5000 ED scanner equipped with an SF-210 slide feeder (image resolution 2400 dpi, grayscale). In every wing image the coordinates of 19 vein junctions (Fig. 1A) were determined automatically using DrawWing software (Tofilski, 2004). The coordinates of vein junctions were aligned using the Procrustes method according to generalized orthogonal least-squares procedures (Rohlf and Slice, 1990) using tspSuper software (Rohlf, 2004).

Forward stepwise discriminant function analysis (tolerance = 0.01; F to enter = 1.0) (StatSoft 2001) was used to determine classification functions. The discriminant analysis was followed by canonical analysis. The contribution of the different variables to canonical variables was assessed using standardized coefficients. Differences between subspecies were tested using MANOVA of partial warp scores produced by the tpsRelw software package (Rohlf, 2005).

RESULTS

Classification based on discriminant functions

Shape of forewing venation differed significantly between the three subspecies (MANOVA: Wilks' Lambda = 0.182; $F_{68,1868} = 36.931$; $p < 0.001$) Fig. 1B, C, D. Discrimination function analysis based on 29 of 38 variables allowed for correct classification of 94.4% of the colonies. The percent of correctly classified colonies ranged from 97.5% in *A. m. carnica* to 71.4% in *A. m. caucasica* (Tab. 1).

In order to classify an unknown colony to one of the three subspecies, coordinates of the 19 points should be determined for each of the bees from the sample, and the mean configuration should be calculated. The mean configuration should be aligned with the reference configuration (Tab. 2). After alignment, 29 variables selected for the discrimination should be multiplied by respective classification functions (Tab. 3). The classification score should be obtained within subspecies by summing up the products of the multiplication and adding

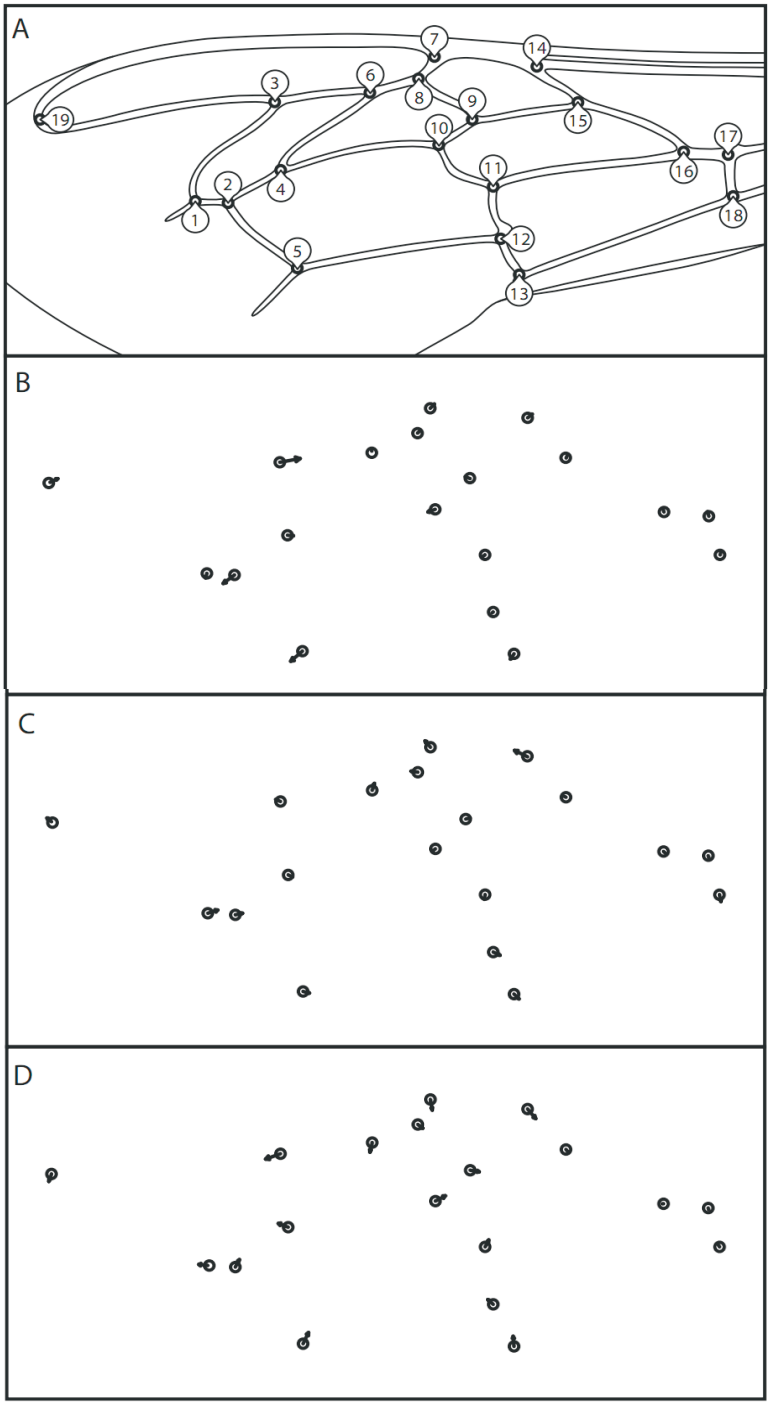

Fig. 1. Honeybee forewing diagram with the vein junctions numbered (A). Vectors indicate the magnitude and direction of the differences between the average wing and the wing of *A. m. carnica* (B), *A. m. caucasic*a (C) and *A. m. mellifera* (D). Lengths of the vectors were exaggerated 5 times to make them more visible.

Table 1

Identifications of honeybee subspecies based on discriminant functions

Subspecies	Correctly classified %	Number of colonies classified as		
		A. m. carnica	*A. m. caucasica*	*A. m. mellifera*
A. m. carnica	97.5	739	11	8
A. m. caucasica	71.4	28	70	0
A. m. mellifera	93.8	5	2	107
Total	94.4	772	83	115

the respective constant value. The colony should be considered as belonging to the subspecies for which the calculation score was the highest.

Classification based on canonical analysis

Canonical variate analysis showed that *A. m. carnica* and *A. m. caucasica* overlap to a larger degree than *A. m. carnica* and *A. m. mellifera* (Fig. 2A). The first canonical variable discriminated mainly between *A. m. carnica* and *A. m. mellifera*. The variable that contributed most to this discrimination was coordinate Y of point 2. The second canonical variable discriminated mainly between

Table 2

Mean coordinates of 19 vein junctions (shown in Fig. 1A) for three honeybee subspecies

Vein junction	*A. m. carnica*		*A. m. caucasica*		*A. m. mellifera*		All 3 subspecies (reference configuration)	
	x	y	x	y	x	y	x	y
1	-0.2819	-0.0612	-0.2806	-0.0520	-0.2839	-0.0599	-0.2819	-0.0609
2	-0.2495	-0.0636	-0.2464	-0.0548	-0.2456	-0.0601	-0.2486	-0.0630
3	-0.1841	0.0610	-0.1887	0.0658	-0.1930	0.0590	-0.1858	0.0607
4	-0.1779	-0.0189	-0.1788	-0.0144	-0.1819	-0.0179	-0.1784	-0.0188
5	-0.1638	-0.1464	-0.1630	-0.1403	-0.1595	-0.1419	-0.1629	-0.1457
6	-0.0733	0.0728	-0.0707	0.0757	-0.0739	0.0702	-0.0733	0.0726
7	0.0017	0.1193	0.0025	0.1198	0.0012	0.1161	0.0014	0.1190
8	-0.0147	0.0918	-0.0146	0.0921	-0.0140	0.0906	-0.0149	0.0917
9	0.0494	0.0427	0.0506	0.0406	0.0530	0.0418	0.0499	0.0425
10	0.0051	0.0087	0.0066	0.0088	0.0095	0.0105	0.0058	0.0090
11	0.0694	-0.0404	0.0686	-0.0429	0.0709	-0.0385	0.0696	-0.0402
12	0.0797	-0.1019	0.0790	-0.1045	0.0786	-0.1007	0.0798	-0.1018
13	0.1056	-0.1482	0.1038	-0.1509	0.1063	-0.1448	0.1059	-0.1478
14	0.1246	0.1090	0.1230	0.1064	0.1256	0.1062	0.1243	0.1088
15	0.1726	0.0652	0.1734	0.0604	0.1728	0.0643	0.1725	0.0651
16	0.2955	0.0070	0.2964	-0.0020	0.2953	0.0065	0.2956	0.0068
17	0.3518	0.0028	0.3523	-0.0082	0.3524	0.0014	0.3520	0.0025
18	0.3670	-0.0390	0.3658	-0.0513	0.3663	-0.0394	0.3670	-0.0393
19	-0.4774	0.0391	-0.4793	0.0518	-0.4801	0.0367	-0.4781	0.0388

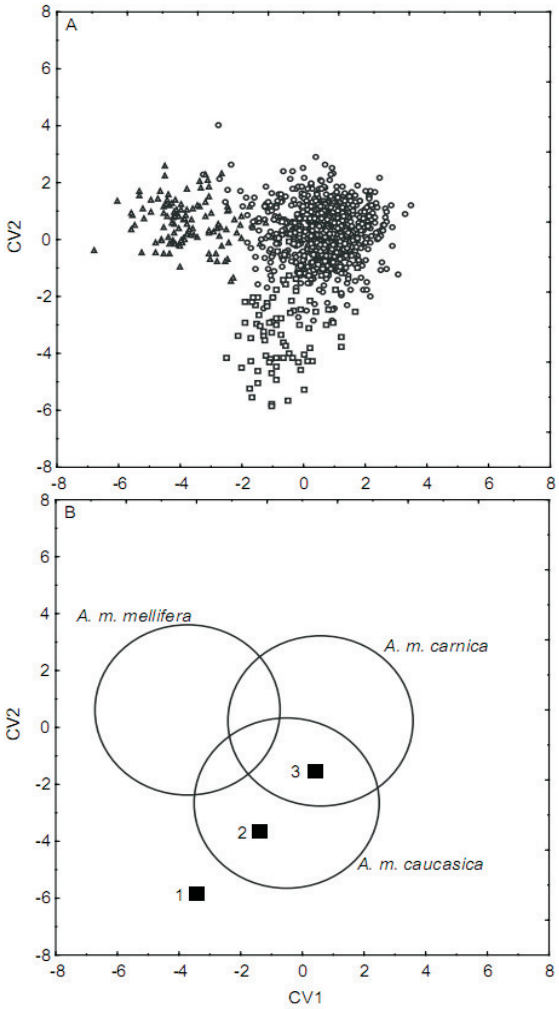

Fig. 2. Classification of honeybee subspecies based on canonical analysis. Canonical scores of *A. m. carnica*, *A. caucasica* and *A. m. mellifera* are marked with circles, squares and triangles, respectively (A). Circles of radius three were drawn around the mean canonical scores (B). If the colony is within a circle it is assumed not to differ from a particular subspecies. Point 1 represents the colony which cannot be classified as any of the three subspecies. Point 2 represents the colony which can only be classified as *A. m. caucasica*. Point 3 represents the colony which can be classified both as *A. m. caucasica* and *A. m. carnica.* The horizontal and vertical axes correspond to the first and second canonical variables, respectively.

A. m. caucasica and the other two subspecies. The variable that contributed most to the second canonical variable was coordinate Y of point 5.

Most of the colonies of a given subspecies lay within a circle of radius 3 around the means of this subspecies (Fig. 2B). We assumed that if a colony is within the circle it belongs to the subspecies. In order to determine the position of a colony on the plot of canonical scores, the 19 points should be aligned with the reference configuration (Tab. 2). The aligned configuration should

Journal of Apicultural Science

Table 3

Classification functions for three honeybee subspecies

Variable	*A. m. carnica*	*A. m. caucasica*	*A. m. mellifera*
2y	-1132340	-1132058	-1129958
1x	-1012821	-1012725	-1013836
1y	53889	54421	52294
14x	-57446	-57877	-57978
4x	-716049	-716417	-717062
10x	-247663	-247607	-247628
6y	2401	2997	1522
3x	-564007	-564269	-564363
11y	-261195	-261510	-260934
12y	-294392	-294211	-295391
5y	-730960	-730448	-731705
12x	-35706	-35474	-36609
11x	215619	214945	215301
14y	-174645	-174371	-175860
8y	93095	93262	94321
13x	138613	138519	138801
15y	204787	205572	205786
19x	-1176634	-1176773	-1177037
8x	-474616	-475218	-475801
16x	425122	424999	424610
13y	-514469	-514290	-514007
9y	-190566	-191085	-191730
15x	155645	155828	155863
6x	-372892	-372597	-372884
5x	-620016	-619906	-620263
17x	1130549	1130175	1130317
19y	-399150	-398949	-399427
7y	426041	425840	425532
18y	327055	326944	326701
Constant	-1046435	-1046284	-1046844

be multiplied by canonical variables coefficients (Tab. 4). Canonical scores should be obtained by summing up the products of the multiplication and adding the respective constant values. Coefficients of difference are calculated as the distance between a point and the mean canonical scores of the subspecies (Tab. 5). The colony can be classified as belonging to the subspecies if the coefficient of difference for that subspecies is less than three. If the colony is within a circle it is assumed not to differ from the particular subspecies (Fig. 2B).

Software for automated identification of honeybee subspecies

Data obtained during the research were used to build Skrzydlak software (KCHZ, 2008). The software allows a colony to be classified into one of the subspecies used

Table 4

Canonical coefficients for three honeybee subspecies

Variable	CV1	CV2
2y	-538.557	107.525
1x	223.929	-116.956
1y	342.554	-309.310
14x	129.487	97.181
4x	235.350	35.922
10x	-9.211	-15.596
6y	181.275	-270.135
3x	85.850	56.123
11y	-50.144	125.866
12y	218.197	-143.748
5y	153.347	-231.271
12x	195.585	-152.507
11x	87.819	195.225
14y	264.141	-192.262
8y	-277.796	48.169
13x	-39.467	46.836
15y	-242.422	-174.596
19x	93.424	11.946
8x	279.630	98.528
16x	117.506	-2.510
13y	-107.386	-20.157
9y	272.658	73.020
15x	-53.303	-41.754
6x	-9.169	-96.362
5x	52.232	-57.235
17x	61.110	103.607
19y	56.743	-89.743
7y	118.695	23.486
18y	81.746	6.799
Constant	85.356	-85.212

Table 5

Mean canonical scores of three honeybee subspecies

Subspecies	CV 1	CV 2
A. m. carnica	0.62453	0.25920
A. m. caucasica	-0.48842	-2.69262
A. m. mellifera	-3.73271	0.59123

for breeding in Poland. Before the analysis a folder with images of fore wings from one colony needs to be prepared. The images are analyzed by the software in order to detect wings. The wings are cropped out and saved as separate files. Position of 19 landmarks is determined automatically in each of the wings. The user can verify the position of the points and correct it. All the calculations are performed automatically and the user is provided with the coefficients of difference. The coefficients show the difference between the tested bee colony and the three subspecies.

DISCUSSION

We presented here a new method of identification of three honeybee subspecies used for breeding in Poland. Despite the large number of measurements (38 variables) not all the colonies were classified correctly. The percent of correctly classified colonies was particularly low in the case of *A. m. caucasica* (Tab. 1). This suggests that not all of them are racially pure-breeds. This, however, should not happen because in Polish breeding apiaries artificial insemination is used on a large scale. It is also possible that *A. m. carnica* differs from *A. m. caucasica* mainly in proboscis length, and wing venation is similar in both subspecies. The observed overlap between subspecies can result from the method of their discrimination in the past. The criteria of Gromisz (1981) were not mutually exclusive. A colony with a cubital index of 50% (in Alpatov notation) could be classified as either *A. m. Carnica* or *A. m. caucasica*. In this situation some colonies from the *A. m. caucasica* breeding lines could be more similar to the average colony of *A. m. carnica* than to the average colony of *A. m. caucasica*. In order to avoid classification of a large proportion of colonies, believed to be *A. m. caucasica*, as *A. m. carnica* the classification was based on canonical analysis. This classification is similar to that of Gromisz (1981). It is less restrictive; however, it cannot prevent hybridization between the subspecies.

The classification of honeybee subspecies based on wing venation is faster because it only requires preparation of wings. Moreover, the measurements of wing venation can be automated. In consequence the measurements can be faster and more precise (Tofilski, 2007). However, the wing measurements are not able to completely replace proboscis measurement which provided important information for identification of *A. m. caucasica* (Gromisz, 1967, 1981).

When subspecies discrimination was based on relatively small sample size (30 colonies or less per subspecies), all colonies were classified correctly (Bieńkowska and Szymula, 2003; Tofilski, 2008). In this study, based on 970 colonies, 94.4% of colonies were classified correctly. In all these three studies the bees were collected from queen-breeding apiaries in Poland and the classification was based on discriminant function analysis of forewing measurements. The only difference was sample size. When the sample is large it is more difficult to find criteria to differentiate between all the colonies. Despite a lower percent of correctly classified colonies the classification functions presented here should be more reliable than those mentioned in the earlier studies. This is because data presented here represent a larger proportion of variation within the subspecies.

CONCLUSIONS

Identification of honeybee subspecies *A. m. carnica, A. m. caucasica* and *A. m. mellifera* based on discriminant function analysis of forewing venation allowed for correct classification of 94.4% of the colonies.

A large proportion of colonies believed to be *A. m. caucasica* were classified as *A. m. carnica.*

The discrimination of honeybee subspecies can be automated by using a scanner with a slide feeder and image analysis.

ACKNOWLEDGEMENTS

This research was supported by the European Commission, ARR agreement No. ZB pp/076/604/2007/08 and MNiSW grant No. N N311 292436.

REFERENCES

Alpatov V.V. (1929) – Biometrical studies on variation and races of the honey bee (*Apis mellifera* L.). *Q. Rev. Biol*., 4:1-58.

Alpatov V.V. (1948) – Porody miedonosnej pčely. Sredi prirody. 4. Moskva.

Batra S.W.T. (1988) – Automatic image analysis for rapid identification of Africanized honey bees. In: Needham GR, editors. Africanized honey bees and bee mites. Ellis Horwood Series in Entomology and Acarology. *Halsted Press,* New York. pp. 260-263.

Bornus L., Demianowicz A., Gromisz M. (1966) – Morfologiczne badania krajowej pszczoły miodnej *Apis mellifica* L. *Pszczeln. Zesz. nauk*., 10(1-4):1-46.

Bieńkowska M., Szymula J. (2003) – Morphological evaluation of Polish breeding lines of *Apis mellifera* bees using scanning techniques. Proceedings of the XXXVIIIth International Apicultural Congress. Lubljana: 574.

Daly H.V., Hoelmer K, Norman P, Allen T (1982) – Computer-assisted measurement and identification of honey bees (Hymenoptera: Apidae). *Ann. Entomol. Soc*., 75: 591-594.

DuPraw E. J. (1964) – Non-Linnean taxonomy. *Nature,* 202:849-852.

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

Goetze G. (1940) – Die beste Biene. Liedloft, Loth u. Michaelis, Leipzig.

Goetze G. (1959) – Die Bedeutung des Flügelgeäders für züchterische Beuerteilugder Honigbiene. Zeitschrift f. *Bienenforsch*., 4 (7):141-148.

Gromisz M. (1967) – Przydatność niektórych cech morfologicznych w systematyce wewnątrz gatunku *Apis mellifera* L. *Pszczeln. Zesz. nauk*., 11(1-3):37-50.

Gromisz M. (1981) – Morfologiczna ocena populacji rojów w pasiekach zarodowych. *Pszczeln. Zesz. nauk*., 25:51-66.

Kauhausen-Keller D. (1991) - Biometrische Unterscheidung zwichen *Apis mellifera carnica* Pollm. Und allen anderen Rassen von *Apis mellifera* L. *Apidologie,* 22:97-103.

KCHZ (2008) – Program do identyfikacji rasowej pszczół. Licencja Krajowe Centrum Hodowli Zwierząt w Warszawie.

Rohlf F.J. (2004) – tpsSuper, superimposition and image averaging, version 1.13.

Rohlf F.J. (2005) – tpsRelw, relative warps, version 1.42.

Rohlf F. J. and Slice D. (1990) – Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zool*., 39:40-59.

Rostecki P., Samborski J., Prabucki J., Chuda-Mickiewicz B. (2007) – A comparison of various hardware for the measurements of the cubital index. *J. apic. Sci*., 51(1):49-53.

Ruttner F. (1988) – Biogeography and taxonomy of honey bees, Springer-Verlag, Berlin, Germany.

Ruttner F., Tassencourt L., Louveaux J. (1978) – Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. *Apidologie,* 9:363-381.

StatSoft (2001) – Statistica (data analysis software), version 6.

Tofilski A. (2004) – DrawWing, a program for numeral description of insect wings. *J. Insect Sci*., 4:1-5.

Tofilski A. (2007) – Automatic measurements of honeybee wings, in: MacLeod N. (Ed.), automated object identification in systematics: theory approaches and applications. *CRC Press*, Boca Ration, Florida, pp 289-298.

Tofilski A. (2008) – Using morphometrics and standard morphometry to discriminate three honeybee subspecies. *Apidologie*, 39:558-563.

WSPOMAGANE KOMPUTEROWO ROZRÓŻNIANIE PODGATUNKÓW PSZCZÓŁ MIODNYCH HODOWANYCH W POLSCE

Gerula D., Tofilski A., Węgrzynowicz P., Skowronek W.

S t r e s z c z e n i e

Celem pracy było opracowanie nowej, automatycznej metody rozróżniania podgatunków pszczół *A. m. carnica, A. m. caucasica* oraz *A. m. mellifera* hodowanych w Polsce. Metoda ta wykorzystuje charakterystyczne użyłkowanie przedniego skrzydła robotnic, poprzez automatyczne wyznaczanie 19 punktów przecięcia się żyłek za pomocą programu komputerowego. Zaletą tej metody jest mniejsza pracochłonność w porównaniu z metodą mikroskopową pomiarów długości języczka, szerokości IV tergitu oraz indeksu kubitalnego.

Pierwszym etapem pracy było opracowanie matematycznego modelu rozróżniania ras pszczół, a następnie wykorzystanie w budowie modułu rozróżniania ras w programie komputerowym „Skrzydlak".

Do badań pobrano 1042 próby pszczół robotnic z rodzin, które reprezentowały 45 linii hodowlanych pszczół kraińskich, kaukaskich oraz środkowoeuropejskich. Z każdej próby pszczół wypreparowano przednie prawe skrzydło i wykonano obrazy skanerem Nikon Coolscan 5000 ED. Współrzędne punktów przecięcia się żyłek wygenerowano automatycznie za pomocą programu DrawWings. W obrębie jednej rodziny punkty te nakładano na siebie metodą Prokrusta i obliczono ich średnie położenie. Średnie położenie punktów przecięcia się żyłek potraktowano jako wzór dla każdej rasy pszczół. Matematyczny model rozróżniania ras pszczół oparto na analizie kanonicznej. Obliczono dwa pierwiastki kanoniczne, które pozwoliły na odróżnienie badanych ras pszczół. Pierwszy pierwiastek kanoniczny pozwala na odróżnienie pszczół *A. m. mellifera* od pozostałych natomiast drugi pierwiastek pszczół *A. m. carnica* od *A. m. caucasica.*

Słowa kluczowe: pszczoła miodna, *Apis mellifera,* podgatunek, skrzydło.