

# Individuality of wing patterning in Giant honey bees (*Apis laboriosa*)

Gerald KASTBERGER<sup>a\*</sup>, Sarah RADLOFF<sup>b</sup>, Gerhard KRANNER<sup>c</sup>

<sup>a</sup> Department of Zoology, Karl-Franzens University of Graz, Austria

<sup>b</sup> Department of Statistics, Rhodes University, Grahamstown, South Africa

<sup>c</sup> eudaptics software gmbh, Kupelwiesergasse 27, 1130 Vienna, Austria

(Received 7 June 2002; revised 7 November 2002; accepted 29 November 2002)

**Abstract** – We investigated whether individual worker bees of a single *Apis laboriosa* colony can be re-identified by their wing patterns alone. In a sample of 183 bees we pre-selected 100 workers belonging to 12 intra-colonial patriline and re-identified them by 25 size-free wing characters. Re-identification was carried out by Self Organizing Map (SOM) reclassification and conventional discriminant analysis (DA) using the protocols of recognition (data for training and testing the model are equal or slightly modified by white noise), and prediction (test data are unknown to the model). SOM recognition of wing shaping was found to be more robust than that resulting from DA. If the test data were altered by white noise, SOM recognition success was 100% within a range of 3% modification which corresponded to the overall measurement error; under these conditions DA success was less than 40%. The SOM prediction capacity was tested using four test-training data ratios and reached 90% under a two-step reclassification protocol.

individuality / wing pattern / SOM re-classification / Giant honeybee / *Apis laboriosa*

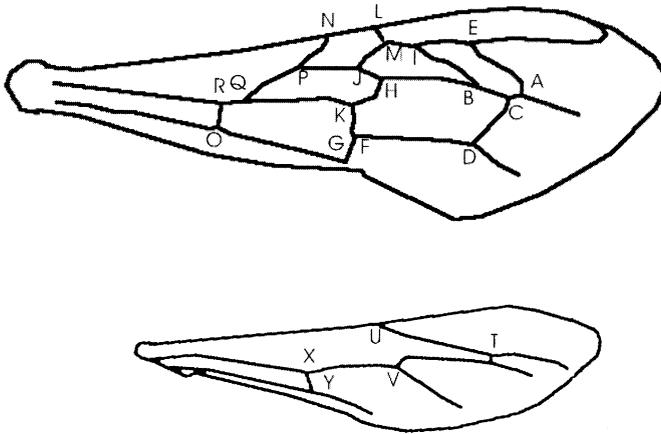
## 1. INTRODUCTION

Morphometrics has been widely used to classify honeybee ecotypes (Alpatov, 1929; Brückner, 1976; Louis and Lefebvre, 1968a, b; Daly, 1985; Ruttner, 1988; Damus and Otis, 1997; Kauhausen-Keller et al., 1997; Hepburn and Radloff, 1998; Radloff et al., 2002), in particular the polymorphy of wing patterns provides here one of the most informative set of characters (Ruttner, 1988). Although phenotypic characteristics are only an indirect and biased measure of genetic diversity (Cornuet and Garnery, 1991), they have been shown to be suitable for differentiating taxa at the macrolevel (Ruttner, 1988; Hepburn and Radloff, 1998; Damus and Otis, 1997). In addition, environmental factors may

select honeybee ecotypes (e.g. concerning geographical latitude, Alpatov, 1929; altitude, Verma et al., 1994).

Within the genetically determined frame of a taxon, individual variability of the phenotypic wing characters underlies further variance. While the main genetic reason for this variation is sexual recombination, the phenotypic expression of characters is varied throughout embryogenesis. Environmental factors may affect the development of imagines directly (e.g. by temperature, Soose, 1954; season, Nazzi, 1992). The stochasticity of this process has genetic and environmental roots and leads to the individual-bound uniqueness (Camazine et al., 2001) of phenotypic expressions, such as human finger prints (Cummins and Midlo, 1961; Cocchi and

\* Correspondence and reprints  
E-mail: gerald.kastberger@uni-graz.at



**Figure 1.** Fore and hind wing venation in *Apis laboriosa* (see text for further details).

Di Bacco, 1982; Prabhakar, 2001), zebra stripes (Klingel, 1974; Sander, 1986) or wing venation in honeybees.

The focus of this paper lies in the phenomenon of individuality, which is expressed by a dissimilarity of wing pattern even in closely related individuals. Using the giant honeybee *Apis laboriosa* Smith 1871 as experimental animal, we questioned to what extent the offspring of one queen differ from each other. Therefore, we checked the re-identification success of the sampled individuals by comparing two different routines, the Self Organizing Map (SOM) reclassification and the conventional discriminant analysis (DA) using the protocols of *recognition* (data for training and testing the model are equal or slightly modified by white noise) and *prediction* (data for training are unknown to the model). A consecutive paper (Kastberger and Paar, in preparation) investigates the aspect of dissimilarity regarding the same bee sample but on the next higher discrimination level, namely on that of subfamilies.

## 2. MATERIALS AND METHODS

### 2.1. Characterization of wing pattern

By means of image analysis we digitized the size of the fore and hind wings, the points A to R on the fore wing venation (according to Ruttner, 1988), and 9 points of the hind wings (Fig. 1). In total, 31 characters were calculated; from the right fore wing: (1) F\_wing length, (2) F\_wing width, (3)

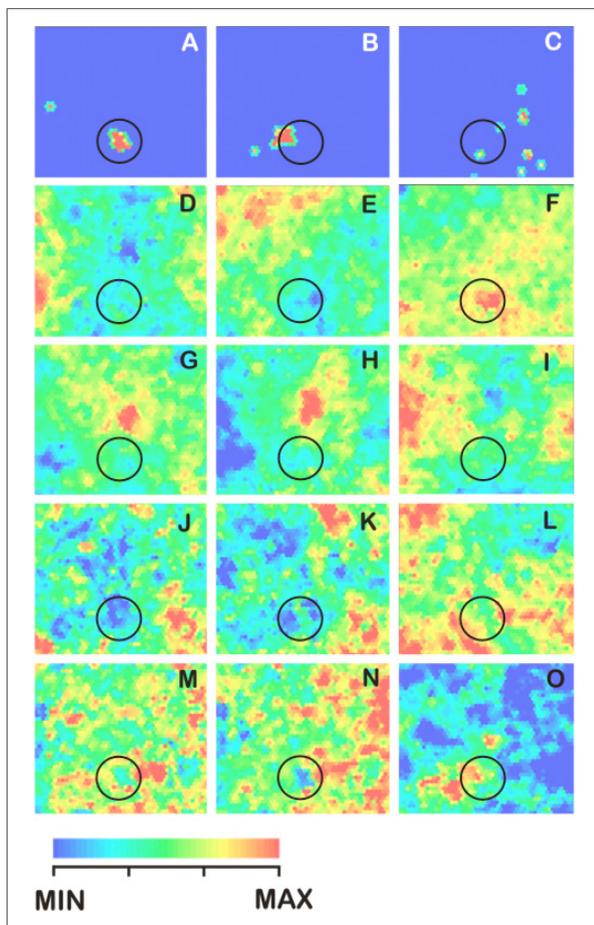
F\_ratio\_LW, (4) F\_wing area; as distance ratios: (5) AC\_BC, (6) AB\_BE, (7) BD\_DF, (8) BE\_BH, (9) JK\_JP, (10) GK\_KQ; as angles (rear numbers according to Ruttner 1988): (11) EAB\_1, (12) ABE\_4, (13) BDG\_7, (14) IEH\_9, (15) HJI\_10, (16) ELI\_13, (17) MJN\_16, (18) FGO\_18, (19) FKO\_19, (20) JNP\_23, (21) ROQ\_26; from the right hind wing, (22) H\_wing length, (23) H\_wing width, (24) H\_ratio\_LW, (25) H\_wing area; as distance ratios: (26) UT\_TV, (27) XY\_VX, (28) UX\_UV; as angles: (29) UTV, (30) UXV, (31) VXY. The characters 1, 2, 4 and 22, 23, 25 represent “size” measures and are strongly coupled; the other 25 characters are size-free and correlate with less than 0.6.

### 2.2. Input data

Each operational record collected by a single run of measurements included 25 size-free wing characters. For each of the 100 individual workers, 10 (for some individuals 11; in total: 1010) runs were conducted.

### 2.3. Data representation with Self-Organizing Maps (SOMs)

SOMs are non-linear data representations which can be trained and tested, and which provide a parameter-free reclassification methodology for the paradigms recognition and prediction (see below). A SOM establishes a topological concept of order, that is based on similarity between the records containing the selected components (for summarization of SOMs, see Kohonen, 1999); it locates more similar records closer together, and more dissimilar records further apart. Since this order reflects the neighbourhood within the data, attributes and features of the data distribution can be



**Figure 2.** Example of a Self-Organizing Map (SOM) produced for wing pattern analysis to demonstrate the specific topology concept for visualization of data. Map characteristics: size, principal plane ratio, 100/85 with 46 columns/45 rows with a total of 1010 test data displayed in 2047 nodes, normalized distortion = 0.185, quantization error = 0.0699. For this example, only size-free wing characters ( $n = 25$ ), had an impact on the ordering process of the topology of the map (priority = 1), all other parameters were set at priority = 0. This priority setup spots the individuals ( $n = 100$ ) as single or multiple patches of up to 11 measurements per individual, clustering them according to the similarity principles based on the impacting wing characters. An individual is represented by a cluster of  $2047/100 = 20.47$  nodes in the average. The grey scale bar below displays the values of the components from their minimum (white) to maximum (black) values, each component map refers here to its own scale (not shown for simplification, but given in the following details). A-C, individual bees, whereas the individual bees A and B distribute in a main and a second small patch, whereas bee C is distributed in two main and four smaller spots. In total, out of 100, 74 individuals showed a main patch with  $1.31 \pm 0.069$  (mean and mean error; for comparison, a similar number of individuals – 71 – was re-identified by the first step of reclassification (Fig. 4) at 0.25 test-training level) and 26 individuals were distributed without a main patch in  $6.34 \pm 0.29$  spots in the average. The D-F, ratio of two lengths between venation

points (see methods): D (BE\_BH, scaling from 0.43...0.58), E (UX\_UV\_3, 1.80...2.25), F (UT\_TV\_1, 1.07...1.25); G-L angles: G (BDG\_7, 90-108 degrees), H (ABE, 80-105), I (EAB\_1, 30...42), I (HJI\_10, 32...43), J (IEH\_9, 17...22), K (MJN\_16, 55...82); M-N, size-dependent parameters: M (wing\_area\_FW, 39...45mm<sup>2</sup>), N (wing\_width\_FW, 4.7...5.1 mm); O (ratio-length-width, 2.9...3.1). Note, the main patch of individual bee A is marked by a circle, it is represented in all other components maps at the same topological spot. The size-free characters contribute to an intrinsic data shape which leads to sinks (white areas) and peaks (black areas). A good survey on the differentiating power is only achieved by colour maps with a rainbow palette (see the same figure in colour under [www.edpsciences.org/apido](http://www.edpsciences.org/apido)). As expected, some wing characters reveal topologies similar to others signalling a tighter positive correlation, such as G – H with  $r = +0.456$ ; D – I with  $r = +0.635$ , J – K with  $r = +0.782$ ; others correlate negatively, such as H – I with  $r = -0.715$ , N – O with  $r = -0.6755$ ; others are strictly uncorrelated such as F – O with  $r = 0.084$ , or D – E with  $r = -0.019$ .

read from the “landscape” of weight vectors, as a non-linear representation of the underlying data distribution. A SOM is commonly made up of units placed on a two-dimensional, hexagonal grid. At the beginning of the training process, the weight vectors of the units are regularly placed on the principal plane spanned by the two longest eigenvectors of the autocorrelation matrix of the data distribution. The graphical units of the grid are

called nodes and represent microclusters of the source data set. The SOM training process works with variable scalings for the internal data distribution, and can be multiplied with user-defined priority factors that change the impact of the respective components on the ordering process, in particular those with a selected priority  $> 0$ . Figure 2 serves for further explanation, e.g. in which way SOMs typically visualize a

data distribution (data base: 100 individuals, 12 patrines; for further details, see Kastberger and Paar, in preparation).

## 2.4. SOMs in wing pattern analysis

In the present study, SOMs were generated by the Viscovery<sup>®</sup> scientific software (for details see www.eudaptics.com; Kastberger and Kranner, 2000). We generated the SOMs as a rectangular matrix of hexagonal nodes (e.g. 2047, the number of nodes is automatically chosen by the program) and parameterized them accordingly to optimize the reclassification procedure. In our model, a single data record acting as the operational unit, is typically represented by roughly 2 (e.g. = 2047/1010) nodes, and the data of the wing of one individual (the main object of the investigation) by 20 (= 2047/100) nodes according to ten runs of data acquisition. Once the SOM had been established by the training data set, test data was recalled on it, by use of 25 size-free wing characters. The results are expressed as probability ( $0 \leq P \leq 1$ ) of membership to the respective category of individual identity. In general, a procedure that proves the membership of subjects of known origin to such a category can be termed as re-identification.

## 2.5. Testing reclassification capacities of two procedures (SOM and DA) by two protocols (recognition and prediction)

For testing the repeatability of a model, the discrimination quality should be independent of the composition of the training data set (cf. Reby et al., 1997). This demand has been realized in the present paper by using two specific protocols for testing the *recognition* and the *prediction* capacity of a reclassification model. *Recognition* means here reclassifying data records by a model which had been defined (= trained) by a data set which is essentially the same as the test set. In our approach the test data differed from the training set by white noise which was superimposed on it (see Eqs. (1, 2)). For proving the prediction capacity the reclassification model was first trained by a certain data set and subsequently tested by other – fresh – data records, which are not known to the model.

Investigation of the *recognition capacity* uses the whole set of wing data for training ( $T_w$ ) and testing ( $t_w$ ) comprising 1010 records. The test data have been calculated from the training data according to equations (1, 2); the resulting values  $t_w$  deviate symmetrically and equally distributed around the original data  $T_w$  within a fraction of  $T_w$  (e.g. DEV = 0.00, 0.05, 0.10, 0.15, ...).

$$f_{\text{NOISE}} = \text{sign}_{r_1} \times r_2 \times \text{DEV} \quad (1)$$

$$t_w = T_w + T_w \times f_{\text{NOISE}} \quad (2)$$

where the factor  $f_{\text{NOISE}}$  represents a double randomization regarding direction and gain:  $r_1, r_2$  are the random variables with ( $0 < r < 1$ ) with  $\text{sign}_{r_1} = +1$  if  $r_1 > 0.5$  and  $\text{sign}_{r_1} = -1$  if  $r_1 \leq 0.5$ ; in this model, the values of  $f_{\text{NOISE}}$  turned out to alter the values of 25 size-free wing characters and 1010 records by  $f_{\text{NOISE}} = -0.00126 \pm 0.5758$  ( $0 > r_2 > 1$ ). Therefore,  $f_{\text{NOISE}}$  biased the variances but not essentially the means of the training data  $T_w$ . The quality of *white-noise* modification here refers, firstly, to the fact that randomization around the original data  $T_w$  bases on equal (and not Gauss) distribution, and secondly, that all wing parameters are of equal significance.

In the second approach we tested the *prediction capacity*. Here, the training set comprised a certain percentage of the original data records, while the test data set consisted of the complimentary part. It is important to know that each multiple data record of one individual worker bee was here divided into that part which was used for the training set and the complimentary part, which was used for the test data. For example, if a training data set was selected as 200 out of 1010 data records, it was subsequently tested with the complementary set of  $1010 - 200 = 810$  data records, representing a *test-Training Data Ratio (tTDR)* of  $810/200 = 4.05$  (cf. Fig. 4). Expectedly, this example at the level  $tTDR = 4.05$  can only deliver good prediction results if the data are friendly (cf. Fig. 2A,B) and differentiate individuals. For noisy data (cf. Fig. 2C) maps have to be much smaller than the number of training samples.

## 2.6. Stepwise discriminant analysis

The conventional stepwise discriminant analysis (DA) procedure was carried out using the same training and test data arrangements as described above for the SOM *recognition* test. For each of the discriminant functions associated with the 100 bees, the estimated coefficients were determined and used to calculate the classification scores of the test data set of unclassified data records.

## 2.7. Measurement error

The overall normalized measurement error (mean\_ME  $\pm$  mean\_error\_ME) of all wing character values ( $n_{\text{ch}} = 31$ ) in all individuals tested ( $\text{ind} = 100$ ) was assessed on the basis of the normalization

$$\text{ME} [\text{ind}] = s_w [\text{ind}] / m_w [\text{ind}] \quad (3)$$

where  $m_w$ ,  $s_w$  are the arithmetical means and the standard deviations of the wing data records per individual.

## 2.8. Second-step reclassification procedure

Both reclassification protocols, testing the *recognition* and the *prediction* (see definitions above) performance of SOM and DA, aim for each data record to (re-)find the membership to the respective category of *individual identity*. One way to improve the matching success of both protocols, independent of the method (SOM or DA) is, to run a *second step* of data reclassification. This uses the number of independent measurements (up to 11 runs) per individual thus enlarging the data base accordingly. The *second step of reclassification* was carried out on the identity level of individuals using the additional “package” information of the data records, namely the membership to one and the same individual. It simply defines that *individual* as re-identified which gained most of the scores in the first step of reclassification per “package”.

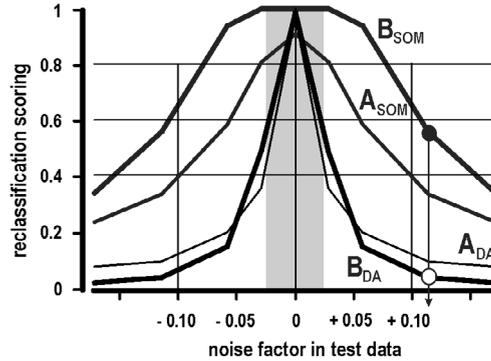
## 3. RESULTS

### 3.1. Recognition capacity of SOM

The data records ( $n = 1010$ ) of 100 individuals were used as training data set and modified under defined symmetrical white noise conditions ( $f_{\text{NOISE}} = 0.00$  to  $\pm 0.15$ ; see formula 1, 2 in methods). SOM first-step recognition success of unchanged ( $f_{\text{NOISE}} = 0.00$ ) data was 91%, under white noise it decreased accordingly, at  $f_{\text{NOISE}} = 0.03$ , recognition was 80%, at  $f_{\text{NOISE}} = 0.06$  it was 59% (Fig. 3,  $A_{\text{SOM}}$ ). The success of SOM second-step recognition was 100% within a white noise range of  $\pm 0.03$ ; this range corresponded well with the measurement error (formula 3, Fig. 3, shaded zone) of the wing character values ( $\text{ME} = 0.0263 \pm 0.0007$ ; min: 0.0068; max: 0.2532;  $n = 3100$  for 31 wing characters and 100 individuals).

### 3.2. Recognition capacity of DA

In comparison to SOM reclassification, DA was better for the first-step recognition (DA: 97%, versus SOM: 91%) but only for unmodified data ( $f_{\text{NOISE}} = 0.0$ , Fig. 3  $A_{\text{DA}}$ ). In addition, the second-step reclassification improved

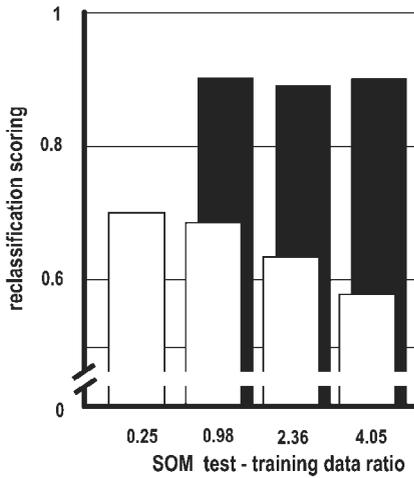


**Figure 3.** Performance of SOM reclassification and discriminant analysis (DA) using the recognition protocol. For training, the whole data set (1010 data records of 25 size-free wing characters, multiply assessed from 100 individuals) was used; for testing, the data was modified by white noise ( $f_{\text{NOISE}}$ , see methods, formula 1, 2). A, first-step reclassification with the data records as operational units, SOM and DA procedures only refer to this step; B, second-step reclassification with the individual identity as operational unit; a given individual identity is defined as reclassified if it gains most of the scores of the multiple runs per individual. The gray area corresponds to the range of measurement error (see methods, formula 3) of data acquisition. Abscissa, the noise factor by which the original data was modified,  $f_{\text{NOISE}} = 0$  gives unchanged condition meaning training and test data are equal; ordinate, the scoring of correct re-classification. For explanation, circles refer to second-step reclassification abilities at the modification level  $f_{\text{NOISE}} = +0.12$ ; for that, the first-class procedure was SOM (full circle) and DA (open circle).

DA recognition capacity, but only for test data equal to the training data. With test data modified by white noise, DA is much less successful than SOM (Fig. 3  $B_{\text{DA}}, B_{\text{SOM}}$ ). Even low modification, in the range of measurement error ( $f_{\text{NOISE}} = \pm 0.03$  and shaded zone in Fig. 3), lowered the success of DA *recognition* markedly down to 46% (Fig. 3  $B_{\text{DA}}$ ), where SOM reclassification still had 100% matching success (Fig. 3  $B_{\text{SOM}}$ ).

### 3.3. Prediction capacity of SOM

Four test-training data ratios ( $t\text{TDR} = 0.25, 0.98, 2.36, 4.05$ , see methods) were produced from the original data set ( $n = 1010$  recordings, 100 individuals). The



**Figure 4.** Performance of SOM reclassification using the prediction protocol assessing the correct individual identity. For training, the data set (1010 data records of 25 size-free wing characters, multiply assessed from 100 individuals) was divided into two parts for training and testing according to four test-training ratios (abscissa); note, the partition referred to each of the data packages, of the multiply assessed data records per individual. White columns, first-step reclassification of the data records; black columns, second-step reclassification of individual identities as operational units.

first-step reclassification scores (Fig. 4) were maximal (70%) for  $tTDR = 0.25$  and decreased with increasing  $tTDR$  down to 58% (for  $tTDR = 4.05$ ). The second reclassification step improved SOM prediction rate up to a final success of 90%, independent of the  $tTDR$  level. This is at least true for all three  $tTDR$  categories (0.98, 2.36, 4.05) tested here; for  $tTDR = 0.25$  the second step of reclassification could not be applied, as the number of test data records per individual identity package were smaller than 3.

#### 4. DISCUSSION

It can be hypothesized that each honeybee has its own self organized (Camazine et al., 2001; Bains, 2001; Bray, 2001; Shinbrot and Muzzio, 2001) wing pattern. However, the underlying basis of individuality has not been formally tested. This is even true for human finger prints (Prabhakar, 2001) although

various classification paradigms of false acceptance or false rejection, examined the distinguishing nature of the physical characteristics. The prediction accuracies achieved for human finger prints typically lie between 85% (Fitz and Green, 1996) and 96.5% (Chong et al., 1997), using different data bases (such as singular points, orientation field, ridge lines) and paradigms (such as rules, supervised learning, discrete class assignments based on knowledge-based features; see Prabhakar, 2001).

To investigate *individuality* in honeybee wings we used a two-step procedure of classification. For the first step, two protocols, recognition and prediction, were conducted by two classification routines, SOM and DA. Furthermore, the principle of multiple measurements from one wing was introduced to allow a more uneven data assessment, which may be present if the wing data are collected under different perspectives, by different persons, from preserved wings or from living animals (however, this paper does not focus on that kind of data biasing). On the first step of reclassification, each record is treated separately as an operational unit. In the second step, the focus of interest lies on the next higher data level and utilizes the “package” information based on the multiple measurements. It notifies which data records originate from one individual (the identity of the individual itself is here still unclassified and unknown to the operation).

We compared the reclassification success of SOM and DA by testing their recognition capacities. For that, we artificially modified the training data by white noise to re-use them as test data. SOM reclassification proved here to be much more robust than DA reclassification for the following reason: under white noise modification of the test data SOM recognition showed much higher matching success; even a low level of noise, in the range of measurement error, led to a massive loss in DA, but to 100% success in SOM analysis.

However, for re-identification, the prediction protocol was the more important technique, in particular because training and testing data derive from different sources. For wing samples tested, SOM prediction provided accuracy rates of at least 90%, this is as high as found for dermatoglyphics (Mardia

et al., 1997; Prabhakar, 2001) or for other applications (Reby et al., 1997) in similarly sized data bases.

Two further aspects should be mentioned to value the SOM reclassification abilities appropriately: first, this paper investigated only the capacity of reclassification on two levels of operation, the level of individual identity and that of data records; it did not consider which of the wing parameters are more or less important for individuality. Nevertheless, we found the 25 size-free parameters to have low interdependencies and to inter-correlate by less than 0.50, which proves all, and not only some of them, good for distinguishing individuals. Second, to some extent, the study also covers the aspect how SOM reclassification works on data bases greater than 100. Although the number of individuals had been restricted to 100, the number of data records was actually tenfold higher. The respective SOM capacity has been demonstrated in particular, by the first step of reclassification, with both recognition and prediction protocols. The interesting feature of SOM recognition (Fig. 3) is demonstrated using “unfriendly” test data revealing a reclassification success of more than 0.8; this included all 1010 data records, independently measured and independently modified by white noise in the range of measurement error ( $f_{\text{NOISE}} = 0.03$ ). In addition, the prediction protocol (Fig. 4) delivered a scoring of 0.70 if the test-training data ratio was 0.25. This means that the model has been trained by 810 data records, and 200 data records had been tested by it. For lower tTDRs, e.g. using less data records for testing and more for training, even a higher reclassification success can be expected (see Fig. 4).

## ACKNOWLEDGEMENTS

Grant from the Austrian Science Fund (FWF), project P 13210-BIO; thanks to Karl Crailsheim (Graz), John Cram (Newcastle), Randall Hepburn (Grahamstown), Ilse Kranner (Graz), Ben Oldroyd (Sydney) for improving earlier versions of the manuscript; Ernst Hüttinger (Lunz, Austria) for valuable discussions and sampling, Beatrice Schöller (Graz) for digitizing the wing samples.

**Résumé – Individualité de la veination alaire chez l’Abeille géante (*Apis laboriosa*).** La morphométrie a été largement utilisée pour classer les écotypes d’abeilles et la veination alaire a fourni l’un des ensembles de caractères les plus informatifs. Ce travail étudie le phénomène d’individualité qui s’exprime par une dissemblance de la veination alaire chez des individus étroitement apparentés. Nous avons vérifié si des ouvrières, prélevées dans une même colonie d’*Apis laboriosa* Smith 1871, pouvaient être re-identifiées individuellement à l’aide de la seule veination alaire. On a choisi dans un échantillon de 183 individus 100 ouvrières qui appartenaient à 23 lignées paternelles, issues donc de la même reine mais d’un père différent. Nous les avons ré-identifiées au moyen de 25 caractères alaires indépendants de la taille à l’aide de la reclassification par Cartes Auto-Organisatrices (SOM) et par l’analyse discriminante (DA) classique. Nous avons en outre choisi deux protocoles d’étude (i) de reconnaissance (les données pour apprendre et tester le modèle sont égales ou légèrement modifiées par le bruit blanc) et (ii) de prédiction (le modèle est testé par de nouvelles données inconnues de lui). Le processus de reconnaissance s’est montré plus robuste dans la reclassification par les SOM que par la DA. Si les données testées étaient modifiées par le bruit blanc, la reconnaissance par les SOM obtenait 100 % de succès avec une fourchette de modifications de 3 % correspondant à l’ordre de grandeur de l’erreur de mesure. Dans ces conditions la DA obtenait moins de 40 % de succès. La méthode de prédiction par les SOM a été appliquée dans un protocole de reclassification en deux étapes ; même avec un rapport donné de test-données d’apprentissage défavorable elle obtient un succès de 90 %.

## *Apis laboriosa* / reclassification par SOM / veination alaire / individualité

**Zusammenfassung – Individualität des Flügelgeäders von Riesenhonigbienen (*Apis laboriosa*).** Morphometrie ist ein gängiges Verfahren für die Klassifizierung von Honigbienen-Ökotypen. Trotz relativ schwach entwickelter Polymorphie ist das Flügelgeäder einer phenotypischen Variabilität unterworfen, die genetisch und umweltabhängig gesteuert ist. Sie führt zu individueller Einzigartigkeit, wie es auch für menschliche Fingerabdrücke oder Zebrastrifen bekannt ist. Die Arbeit untersucht das Prinzip dieser Individualität von Honigbienen; wir prüften, ob einzelne Arbeiterinnen einer ausgewählten Kolonie von *Apis laboriosa* Smith 1871 anhand des Flügelmusters re-identifiziert werden können. Aus einem Sample von 183 Individuen wurden 100 Arbeiterinnen ausgewählt, die zu 12 Subfamilien gehörten, von derselben Königin aber von unterschiedlichen Drohnen abstammten. Wir re-identifizierten sie mittels 25 größenunabhängiger Flügelparameter durch SOM

(Self Organizing Maps) – Reklassifizierung und konventioneller Diskriminantenanalyse (DA). Dazu wählten wir zwei Untersuchungsprotokolle, (a) die Rekognition, für die Trainings- und Testdaten entweder gleich oder durch weißes Rauschen modifiziert waren, und (b) die *Prediktion*, bei der ein Modell durch neue, ihm unbekannte Daten geprüft wird. Das Verfahren der *Rekognition* erwies sich in der SOM-Analyse als viel robuster als durch DA. Bei einer 3 % igen Verrauschung der Testdaten (was der Größenordnung des Messfehlers entsprach) betrug die nach einem zweistufigen Verfahren ermittelte SOM-Re-Identifizierungsrate 100 %, jene der DA-gestützten nur 40 %. Zusätzlich wurde die SOM-Prediktionsmethode in einem zweistufigen Reklassifizierungsverfahren angewandt; es erbrachte selbst bei ungünstiger Test-Trainingsdaten-Relation einen Re-Identifizierungserfolg von 90 %.

**Individualität / Flügelmuster / SOM Reklassifikation / Riesenhonigbiene / *Apis laboriosa***

**REFERENCES**

- Alpatov W.W. (1929) Biometrical studies on variation and races of the honeybee *Apis mellifera* L., Q. Rev. Biol. 4, 1–57.
- Bains W. (2001) The parts list of life, Nature Biotechnol. 19, 401–402.
- Bray D. (2001) Reasoning for results, Nature 412, 883.
- Brückner D. (1976) The influence of genetic variability on wing symmetry in honeybees (*Apis mellifera*), Evolution 30, 100–108.
- Camazine S., Deneubourg J.L., Franks N.R., Sneyd J., Theraulaz G., Bonabeau E. (2001) Self-Organization in Biological Systems, Princeton University Press, New Jersey, ISBN 0 691 0211 3.
- Chong M.M.S., Ngee T.H., Jun L., Gay R.K.L. (1997) Geometric framework for fingerprint classification, Pattern Recognit. 30, 1475–1488.
- Cocchi D., Di Bacco M. (1982) Modelling fingerprint pattern inheritance, Acta Anthropogenet. 6, 115–124.
- Cornuet J.M., Garnery L. (1991) Mitochondrial DNA variability in honeybees and its phylogeographic implications, Apidologie 22, 627–642.
- Cummins H., Midlo C. (1961) Fingerprints, Palms and Soles: An Introduction to Dermatoglyphics, Dover Publication, Inc New York.
- Daly H.V. (1985) Insect morphometrics, Annu. Rev. Entomol. 30, 415–438.
- Damus M.S., Otis G.W. (1997) A morphometric analysis of *Apis cerana* F. and *Apis nigrocincta* Smith populations from Southeast Asia, Apidologie 28, 309–323.
- Fitz A.P., Green R.J. (1996) Fingerprint classification using hexagonal fast fourier transform, Pattern Recognit. 29, 1587–1597.
- Hepburn H.R., Radloff S.E. (1998) Honeybees of Africa, Springer Berlin.
- Kastberger G., Kranner G. (2000) Visualization of multiple influences on ocellar flight control in giant honey bees with the data-mining tool ViscoveySOMine, Behav. Res. Methods, Instrum. Comput. 32, 157–168.
- Kauhausen-Keller D., Ruttner F., Keller R. (1997) Morphometric studies on the microtaxonomy of the species *Apis mellifera* L., Apidologie 28, 295–307.
- Klingel H. (1974) Soziale Organisation und Verhalten des Grevy-Zebras (*Equus grevyi*), Z. Tierpsychol. 36, 37–70.
- Kohonen T. (1999) Self-Organizing Maps, 3rd ed., Springer-Verlag, Berlin Heidelberg.
- Louis J., Lefebvre J. (1968a) Étude quantitative de la divergence dans l'évolution morphologique de certaines entités infraspécifiques d'abeilles domestiques (*A. mellifera* L.), C.R. Acad. Sci., Paris 266, 1131–1133.
- Louis J., Lefebvre J., Moratille R., Fresnaye J. (1968b) Essai de discrimination de lignées consanguines d'abeilles domestiques (*Apis mellifera* L.) obtenues par insémination artificielle, C.R. Acad. Sci., Paris 267, 526–528.
- Mardia K.V., Baczkowski A.J., Feng X., Hainsworth T.J. (1997) Statistical methods for automatic interpretation of digitally scanned finger prints, Pattern Recognit. Lett. 18, 1197–1203.
- Nazzi F. (1992) Fluctuation of forewing characters in hybrid honey bees from north-eastern Italy, J. Apic. Res. 31, 27–31.
- Prabhakar S. (2001) Fingerprint classification and matching using a filterbank, Doctoral thesis Michigan State University, Computer Science Engineering.
- Radloff S.E., Hepburn H.R., Neumann P., Moritz R.F.A., Kryger P. (2002) A method for estimating variation in the phenotypic expression of morphological characters by thelytokous parthenogenesis *Apis mellifera capensis*, Apidologie 33, 129–137.
- Reby D., Lek S., Dimopoulos I., Joachim J., Lauga J., Aulagnier S. (1997) Artificial neural networks as a classification method in the behavioural sciences, Behav. Process. 40, 35–43.
- Ruttner F. (1988) Biogeography and taxonomy of honeybees, Springer-Verlag, Berlin.
- Sander L.M. (1986) Fractal growth processes, Nature 322, 789–793.
- Shinbrot T., Muzzio F.J. (2001) Noise to order, Nature 410, 251–258.
- Soose E. (1954) Effect of temperature on the wing index and chitin colour of the honeybee, Arch. Bienenkd. 31, 49–66.
- Verma L.R., Mattu V.K., Daly H.V. (1994) Morphometrics of Indian honeybee in northwest Himalayas, Apidologie 25, 220–223.