Bulletin of Entomological Research (2008) **98**, 437–447
doi:10.1017/S0007485308005750
© 2008 Cambridge University Press
First published online 21 April 2008
Printed in the United Kingdom

Thrips (Thysanoptera) identification using artificial neural networks

P. Fedor[1], I. Malenovský[2], J. Vaňhara[3]*, W. Sierka[4] and J. Havel[5]

[1]Comenius University, Faculty of Natural Sciences, Department of Ecosozology, Mlynská dolina, 842 15 Bratislava, Slovak Republic: [2]Moravian Museum, Department of Entomology, Hviezdoslavova 29a, 627 00 Brno, Czech Republic: [3]Masaryk University, Faculty of Science, Department of Botany and Zoology, Kotlářská 2, 611 37 Brno, Czech Republic: [4]University of Silesia, Faculty of Biology and Environmental Protection, Department of Zoology, Bankowa 9, 400 07 Katowice, Poland: [5]Masaryk University, Faculty of Science, Department of Chemistry, Kotlářská 2, 611 37 Brno, Czech Republic

Abstract

We studied the use of a supervised artificial neural network (ANN) model for semi-automated identification of 18 common European species of Thysanoptera from four genera: *Aeolothrips* Haliday (Aeolothripidae), *Chirothrips* Haliday, *Dendrothrips* Uzel, and *Limothrips* Haliday (all Thripidae). As input data, we entered 17 continuous morphometric and two qualitative two-state characters measured or determined on different parts of the thrips body (head, pronotum, forewing and ovipositor) and the sex. Our experimental data set included 498 thrips specimens. A relatively simple ANN architecture (multilayer perceptrons with a single hidden layer) enabled a 97% correct simultaneous identification of both males and females of all the 18 species in an independent test. This high reliability of classification is promising for a wider application of ANN in the practice of Thysanoptera identification.

Keywords: artificial neural networks, routine identification, Thysanoptera, *Aeolothrips*, *Chirothrips*, *Dendrothrips*, *Limothrips*

(Accepted 20 November 2007)

Introduction

Biology, as well as its application in many fields, such as agriculture, forestry, human and veterinary medicine, relies heavily on the accurate identification of species, the basic units in the hierarchy of nature. For more than 250 years of modern taxonomic activity, the number of known species has been continuously increasing, and our knowledge of biodiversity is far from being complete. There are vast arrays of morphologically similar species, which are often so small in dimensions as to render establishing the species identity within many groups a difficult task. Insects are undoubtedly one of the most prolific of these groups. Their identification usually requires significant experience, encyclopaedic knowledge, a good reference collection and relevant literature. The process of becoming expert in a group is mostly pain-staking and time-consuming and the numbers of available taxonomists are limited for many groups and zoogeographical regions (Gaston & May, 1992).

Partly to compensate for the lack of expertise, several approaches have been proposed to simplify, particularly, routine identification in applied branches of entomology. Besides molecular diagnostic techniques, which are currently becoming more and more widespread, the

*Author for correspondence
E-mail: vanhara@sci.muni.cz

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

Table 1. List of examined Thysanoptera specimens.

		Males	Females	Total
Aeolothrips	*albicinctus* Haliday, 1836	6	11	17
	astutus Priesner, 1926	6	12	18
	ericae Bagnall, 1920	–	9	9
	fasciatus (Linnaeus, 1758)	4	13	17
	intermedius Bagnall, 1934	6	17	23
	versicolor Uzel, 1895	–	11	11
	vittatus Haliday, 1836	–	15	15
Chirothrips	*aculeatus* Bagnall, 1927	15	15	30
	ambulans Bagnall, 1932	–	15	15
	hamatus Trybom, 1895	15	15	30
	manicatus Haliday, 1836	15	15	30
	pallidicornis Priesner, 1925	15	15	30
Dendrothrips	*degeeri* Uzel, 1895	4	70	74
	ornatus (Jablonowski, 1894)	–	75	75
	saltatrix Uzel, 1895	2	12	14
Limothrips	*cerealium* Haliday, 1836	15	15	30
	consimilis Priesner, 1926	15	15	30
	denticornis Haliday, 1836	15	15	30

rapid progress in information technology has also opened opportunities for the development of computer-assisted taxonomy (Edwards & Morse, 1995; Chesmore, 1999). Computer-aided taxonomic applications can range from interactive multi-access taxonomic keys, distributed, for example, on CD-ROM (Cranston, 2005), to nearly fully automated identification systems (e.g. Weeks *et al*., 1997, 1999; Platnick *et al*., 2005). During the past ten years, artificial neural networks (ANN) seem to have been one of the most promising computing tools for the basis of such systems (Weeks & Gaston, 1997). Analogous to the structure of the human brain, the advantages of ANN include an ability to learn from examples and to generalize observed patterns. Compared to many traditional statistical methods, ANN are non-linear and make no general assumptions on the type or statistical distribution of data; and, thus, can be used for pattern recognition on practically any kind of multivariate data sets – transformed digital images (Do *et al*., 1999), optically sensed wing beat frequency spectra (Moore & Miller, 2002), near-infrared reflectance spectra (Aldrich *et al*., 2007), bioacoustic recordings (e.g. Chesmore, 2004), chemotaxonomy (Hernández-Borges *et al*., 2004) or morphometry (Marcondes & Borges, 2000; Vaňhara *et al*., 2007). Although the results from most case studies are very convincing, the practical use of ANN for species identification in current biology is still rather infrequent (Gaston & O'Neill, 2004).

Thrips (Thysanoptera) represent an order of tiny (usually 1–3 mm) insects with about 5500 species worldwide (Mound, 2001). Some phytophagous species are regarded as pests in agriculture, horticulture and forestry, as their feeding on plants which are economically important can cause damage, accompanied by symptoms such as silvering, leaf deformation, colour changes and scarring. Moreover, a limited number of thrips species are vectors of destructive tospoviruses, which can infect a wide range of host plants (Lewis, 1997; Jones, 2005). The correct identification down to species level is often needed to effectively control populations of such pests. This identification is, however, at times difficult because of their minute size and close morphological similarity.

Thrips can be identified in several ways. Printed dichotomous taxonomic keys (e.g. Schliephake & Klimt, 1979; zur Strassen, 2003 for European species) are the traditional methods, which often allow the identification of entire regional thrips faunas. However, their use is often not straightforward for inexperienced biologists. The work, of applied entomologists, has been facilitated by the introduction of more user-friendly pictorial keys (Mound & Kibby, 1998) and, for Thysanoptera of economic importance, computerized multi-access keys (Moritz *et al*., 2001). Genetic markers have proved to be a powerful tool in the identification of thrips pest species, including their immature stages (Moritz *et al*., 2000; Brunner *et al*., 2002; Toda & Komazaki, 2002; Rugman-Jones *et al*., 2006). The interactive electronic key by Moritz *et al*. (2004) combines both morphological and molecular information.

In our laboratories, we have recently demonstrated the successful use of ANN for the identification of parasitic flies (Diptera: Tachinidae) (Vaňhara *et al*., 2007). The aim of this paper is to examine further the potential use of ANN for the rather complicated task of Thysanoptera identification.

Material and methods

Selection of taxa

For our analysis, we selected 18 Thysanoptera species from four genera, i.e. *Aeolothrips* Haliday (Aeolothripidae), *Chirothrips* Haliday, *Dendrothrips* Uzel and *Limothrips* Haliday (all Thripidae). They represent common European species which are widely distributed (zur Strassen, 2005) and occur in a wide range of environmental conditions. Some species of *Aeolothrips* (*A. fasciatus, A. intermedius*) are predators feeding on mites or other thrips. *Dendrothrips* spp. live on leaves of trees or shrubs, whereas *Chirothrips* and *Limothrips* spp. are grass-feeders (Schliephake & Klimt, 1979). *L. cerealium* (the grain thrips) and *L. denticornis* (the barley thrips) attack cereals and can cause serious grain losses (e.g. Franssen & Mantel, 1965; Larsson, 2005).

In total, we included 498 specimens in our study. Individual species were represented in the data set by 9–75 specimens of both sexes, except for *Aeolothrips ericae*, *A. versicolor*, *A. vittatus*, *Chirothrips ambulans* and *Dendrothrips ornatus* for which only females were available to us (table 1). In many thrips species, it is the females that are more prevalent; males being more rarely found or absent. This is because the females are haplo-diploid, usually live longer than the males (e.g. they are often the only sex to overwinter) and some species or regional populations reproduce parthenogenetically (Lewis, 1973). The material was collected from various localities in Slovakia, Poland and the Czech Republic during several projects (e.g. Fedor *et al*., 2001; Pelikán *et al*., 2002).

Standard preparatory techniques were used for mounting; specimens were collected into AGA (a mixture of ethylalcohol, glycerine and acetic acid), macerated in warm 10% KOH, dehydrated in alcohol and clove oil and mounted on slides in Canada balsam.

The material was identified by P. Fedor and W. Sierka and is deposited in their collections (Comenius University Bratislava, University of Silesia Katowice).

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

Table 2. List of characters used for thrips identification.

No.	Character
1	head width
2	head length (dorsal side)
3	head length (ventral side, including mouthcone)
4	clavus length
5	clavus width
6	forewing length
7	forewing basal width
8	total body length (excluding antenna and penis)
9	pronotum width
10	pronotum length
11	eye length
12	ovipositor length
13	ovipositor width
14	antennal segment V length
15	antennal segment VI length
16	distance between the posterior pair of ocelli
17	distance between an anterior and posterior pair of ocelli
18	forewing with three light bands (yes/no)
19	clavus distinctly widest basally (yes/no)

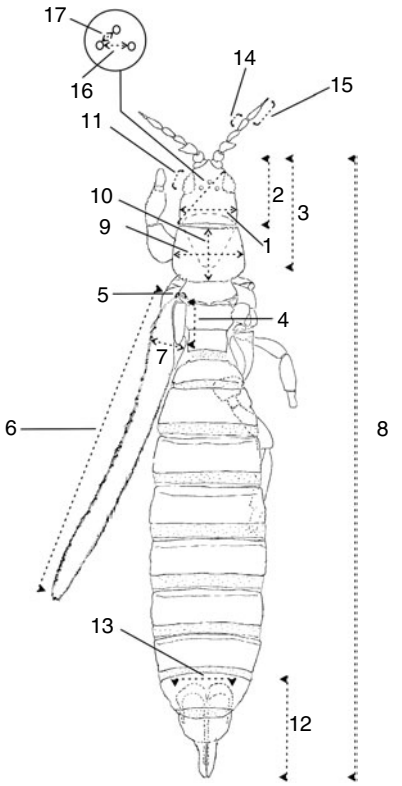


Fig. 1. Morphometric characters (1–17) used for thrips identification (explanation in table 2).

Selection of characters

We defined and recorded a total of 25 characters in at least one thrips genus. For further analysis, we avoided missing data and retained 19 characters, which were measured or determined for each specimen of the taxa (table 2). These included 17 quantitative morphometric characters on different parts of the thrips body (fig. 1) and two qualitative two-state characters (presence/absence). Most of the selected characters are commonly used for thrips identification (e.g. zur Strassen, 2003).

Quantitative characters were measured as linear distances on digital images taken from slide mounted specimens by P. Fedor and W. Sierka using the microscope NIKON ECLIPSE E600 and the image analyser software LUCIA net (Laboratory Imaging Ltd, Czech Republic). Experimental uncertainty of these measurements for the digital images was maximally 0.03 µm.

Many species of Thysanoptera exhibit a pronounced sexual dimorphism. Males of all the studied *Chirothrips* and *Limothrips* species are extremely brachypterous or apterous (this pertains also to the females of *C. ambulans* and both sexes of *Aeolothrips albicinctus*) and lack ocelli. The data corresponding to missing forewings and ocelli were included in the analysis and entered into the data matrix as zero values. The same approach was chosen for the characters of the ovipositor (length and width), which are applicable only to females. To distinguish between males and females, the sex was included as another two-state input character into the ANN model.

The data set used for ANN computations, thus, consisted of altogether 20 input variables (17 continuous, three binary) and a single nominal output variable, identification (the name of the species).

Software

ANN computation was performed using a Trajan Neural Network Simulator, version 3.0 D (Trajan Software, Ltd, 1996–1998). STATISTICA 7.0 (StatSoft, Inc., 2004) was used for other statistics. All calculations were performed on a standard PC computer with a Microsoft Windows 2000 Professional operating system.

Computational strategy and basics of ANN theory

The ANN computational strategy applied in this study is similar to that introduced by Vaňhara *et al*. (2007). Data was randomly divided into a learning (training) set, a verification set and a test set. Each set consisted of a number of samples (thrips specimens) characterized by input variables (characters) and identified to species (output).

There are several different types of supervised ANN which can be used for classification problems (e.g. Bishop, 1995). Preliminary experiments on the data set with some of these types (radial basis function, linear, probabilistic and multilayer perceptrons networks) suggested that the multilayer perceptrons (MLP) would be the most efficient for the purpose. MLP is generally one of the most commonly used types of ANN and can model functions of almost arbitrary complexity. The MLP has already been used in most taxonomic applications (Clark, 2003; Hernández-Borges *et al*., 2004; Chesmore, 2004; Vaňhara *et al*., 2007). Similar to other ANN, MLP is a computational model formed from a certain number of single units, artificial neurones (nodes), interconnected with each other while coefficients (weights, w_{ij}) are assigned to each connection. The MLP architecture is conventionally constructed with three or more feed forward layers, i.e. input, output and one or several hidden layers (e.g. fig. 2). Each layer might have a different number of

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

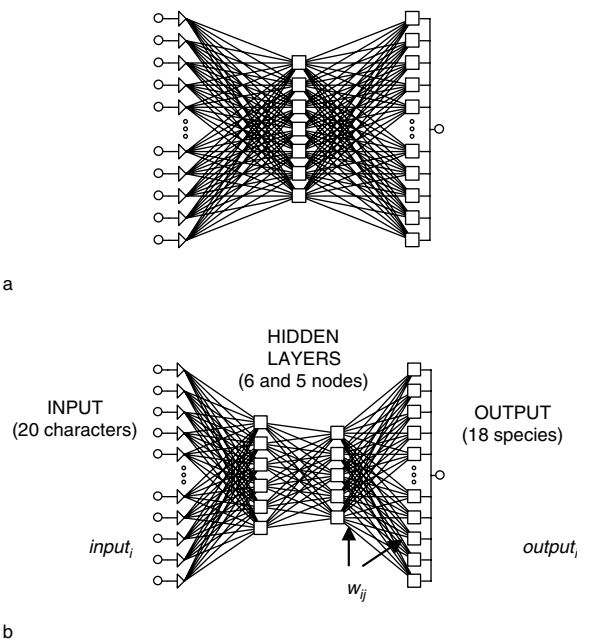


Fig. 2. Possible ANN architectures for thrips identification. (a) Three-layered MLP (20, 7, 18) used in the paper; (b) four-layered MLP (20, 6, 5, 18).

nodes. The input layer receives the information about the system (the nodes of this layer are simple distributive nodes, which do not alter the input value at all). The hidden layer processes the information initiated at the input, while the output layer is the observable response or behaviour. The input values, $input_i$, multiplied by connection weights, w_{ij}, are first summed and then passed through a transfer function to produce the output, out_i.

The learning or training process of MLP consists of searching for such values of w_{ij} weights to minimize the root mean square (RMS) value:

$$\mathrm{RMS} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j-1}^{M}(y_{ij} - out_{ij})^2}{N \times M}}$$

where y_{ij} is the element of the matrix ($N \times M$) for the training set, and out_{ij} is the element of the output matrix ($N \times M$) of the neural network, N is the number of variables in the pattern and M is the number of samples. By running the data on specimens from the training set, including the output variable (the identification), through the network and comparing the actual output generated with the desired or target outputs, the network automatically adjusts the weights and thresholds in order to minimize the overall error. This process is equivalent to fitting the model represented by the network to the training data available. Reaching the minimal RMS value indicates the best moment to stop the training procedure and is helpful in the search for the optimal network architecture. The latter largely consists of the estimation of an appropriate number of nodes in the hidden layer(s), which is one of the most critical tasks in ANN design. Unlike the input and output layers, the number and size of hidden layers are not predictable and will vary according to the complexity of the data.

The training of a MLP network can be executed by different algorithms. We used the back propagation, which is the best-known one and has relatively low memory requirements (Fausset, 1994; Patterson, 1996). We ran the training algorithm several times for 5000 to 10,000 iterations (epochs) with each configuration to ensure a proper convergation to RMS minimum and to avoid being stuck in a local minimum.

After obtaining the optimal architecture and minimal RMS, a number of randomly selected specimens from the learning set were excluded to form the verification set. They were used as a check for cross-validation of the training procedure to prevent over-training – the situation when the model is too complex and training achieves a low error but has a poor generalization when new samples are processed. The verification is a test of prediction power of the model (its efficiency is in identifying unknown specimens). However, unlike the test set, the verification set does actually play a role in selecting the final ANN model.

ANN computation also comprises pre- and post-processing stages. Pre-processing techniques used in our study included data standardization (scaling) and conversion of nominal input variables to numeric values (done automatically by Trajan software). Similarly, the output activation scores were transformed by post-processing into the name of thrips species. The classification by ANN is performed by checking output unit activation levels against two thresholds, the accept threshold and the reject threshold. To simply assign the classification to the species corresponding to the winning unit, irrespective of the settings of other units, we set the accept threshold to zero and ignored the reject threshold.

Results

Basic statistical analysis and structure of the data

Descriptive statistics of the data were examined first in order to understand what the inter- and intraspecific variabilities are in the individual characters.

The minimum and maximum values for each character are given separately in the Appendix for the males and females of each species. The range of variability highly exceeded the experimental uncertainty of measurements in all the morphometric characters. On the other hand, the morphometric characters largely overlapped between different species, especially those belonging to the same genus. A reliable distinction of included species, thus, would be practically impossible if only a single or a few morphometric characters were considered.

Correlation analysis of morphometric data

A correlation matrix of the morphometric characters is presented in table 3. High correlations were observed, particularly between characters 2 (head dorsal length) and 3 (head ventral length), and 3 and 8 (total body length). Also, other characters, more or less, correlated with the total body size, e.g. the length and width of ovipositor, whereas characters corresponding to antenna, ocelli or pronotum were statistically more independent. In spite of the high correlations between some of the variables, in order not to lose information, we did not reduce the data and kept all the available characters for the ANN analysis.

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

Table 3. Correlation matrix (Spearman rank correlation coefficient) of morphometric variables over the entire data set (measures missing due to the absence of wings and ocelli pairwise excluded).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1.00																
2	0.50	1.00															
3	0.54	0.95	1.00														
4	0.64	0.65	0.60	1.00													
5	0.41	0.82	0.81	0.57	1.00												
6	0.44	0.77	0.79	0.67	0.76	1.00											
7	0.69	0.68	0.70	0.79	0.65	0.68	1.00										
8	0.50	0.87	0.91	0.63	0.79	0.84	0.70	1.00									
9	0.31	0.62	0.66	0.43	0.74	0.73	0.53	0.76	1.00								
10	0.04	0.69	0.70	0.44	0.77	0.76	0.50	0.77	0.81	1.00							
11	0.76	0.66	0.66	0.75	0.59	0.65	0.73	0.64	0.45	0.32	1.00						
12	0.62	0.82	0.82	0.71	0.78	0.79	0.76	0.87	0.61	0.63	0.73	1.00					
13	0.59	0.86	0.89	0.63	0.77	0.75	0.72	0.89	0.70	0.67	0.66	0.89	1.00				
14	0.88	0.49	0.54	0.62	0.42	0.44	0.70	0.51	0.25	0.06	0.72	0.67	0.60	1.00			
15	0.44	0.63	0.66	0.48	0.59	0.72	0.48	0.65	0.42	0.45	0.59	0.63	0.57	0.53	1.00		
16	0.22	−0.30	−0.34	−0.08	−0.28	−0.30	−0.21	−0.35	−0.45	−0.50	0.05	−0.23	−0.36	0.01	−0.12	1.00	
17	0.82	0.41	0.41	0.62	0.22	0.30	0.58	0.39	0.09	0.02	0.63	0.51	0.42	0.73	0.29	0.28	1.00

Table 4. Factor analysis: eigenvalues (EV) and related statistics.

No.	Eigenvalue	% Total variance	Cumulative EV	Cumulative %
1	10.23	51.16	10.23	51.16
2	4.46	22.29	14.69	73.45
3	1.73	8.63	16.42	82.08
4	0.99	4.97	17.41	87.05
5	0.74	3.72	18.15	90.77
6	0.53	2.66	18.69	93.43
7	0.40	2.01	19.09	95.43
8	0.28	1.42	19.37	96.86
9	0.17	0.86	19.54	97.72
10	0.15	0.74	19.69	98.46
11	0.07	0.37	19.77	98.83
12	0.06	0.28	19.82	99.12
13	0.04	0.20	19.86	99.31
14	0.04	0.18	19.90	99.49
15	0.03	0.16	19.93	99.65
16	0.02	0.12	19.95	99.76
17	0.02	0.08	19.97	99.85
18	0.01	0.07	19.98	99.91
19	0.01	0.05	19.99	99.96
20	0.01	0.04	20.00	100.00

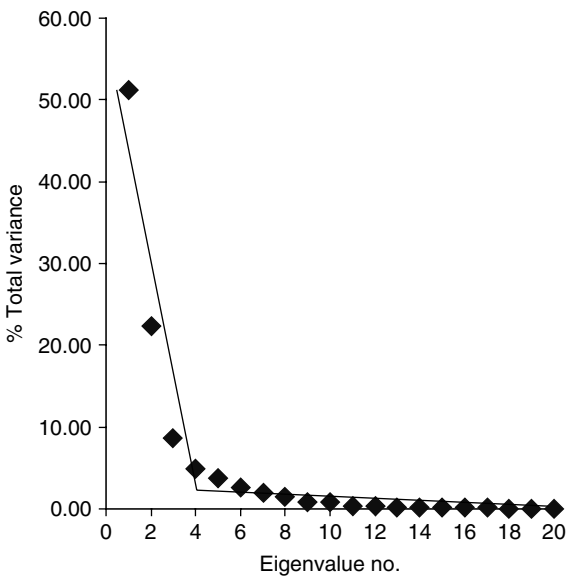


Fig. 3. Variability in the thrips data matrix as a function of the number of eigenvalues.

Basic factor analysis

The basic factor analysis (FA, sometimes also called Eigenvalues analysis) was used here as an exploratory technique to examine the data structure and also to find the number of factors responsible for the variability in the whole data matrix. It is known that the number of non-zero Eigenvalues (EV), or the rank of the matrix, gives a sound estimate for the number of factors influencing the data. The rank has a strong and quite real meaning. FA for the complete data set here revealed that 87% of the total variance in the whole data (characters 1–19 and sex) can be explained by the first four factors (table 4). In other words, just the first four EV correspond to nearly 90% of the total variability. This is illustrated in fig. 3. The biological meaning is that four thrips genera in the data are responsible for most of the variability and vice versa. To explain 99.9% variability, a much higher number of EV is needed; and this number approaches the total number of different species present in the whole data set, i.e. 18. Thus, to explain the complete variability in the data set, we really need a much higher number of EV. Thus, the rank of the data matrix used is near to the total number of the species, and this indicates that the data we have selected really contains information about just 18 species. There is, of course, the question whether ANN would be able to make the distinction between each individual species present when the number of morphological characters and the total number of inputs (20) is only slightly higher than the total number of species (18).

ANN analysis of the thrips database

As already mentioned above, the use of ANN consists of several steps; search for a suitable ANN architecture,

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

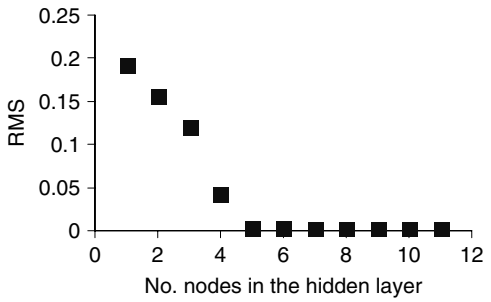


Fig. 4. Dependence of RMS error on the number of nodes in the hidden layer for a three-layered MLP architecture.

training (modelling), verification and identification (prediction) of unknown specimens. The individual steps were carefully examined.

Search for the optimal ANN architecture

First, 100 samples were randomly removed from the original data to compose the test set. These samples were reserved for a single unrepeated testing of the final ANN model, which had been selected and optimized using the learning and the verification sets. The test set did not play any role in the search of the model and ensured that the results of the training and verification sets were real and not artefacts of the training process.

The remaining 398 samples were used to search for a suitable and/or optimal ANN architecture and for corresponding network training.

For simplicity, we started the search for a suitable ANN architecture with a three-layered MLP structure. For this purpose, we studied the dependence of RMS error on the number of nodes in the hidden layer in the architecture of the general form (20, *n*, 18), where 20 is the number of nodes in the input layer, *n* is the number of nodes in the hidden layer, and 18 is the number of nodes in the output layer. The effect of *n* on the RMS error value is shown in fig. 4. The minimal value of RMS (*ca*. 1.10^{-3}) was achieved for $n \geq 5$–7. We chose the network (20, 7, 18) with seven nodes in a single hidden layer (fig. 2a) for further computations because only for $n \geq 7$, a 100% correct classification of samples in the training process was achieved in most of the runs.

Another suitable architecture found in training experiments was with two hidden layers, (20, 6, 5, 18), cf. fig. 2b, which also gave good results and a 100% prediction. The total number of unknown weights, w_{ij}, is lower by 16 in the latter than in the (20, 7, 18) structure (266 unknowns). However, the simpler three-layered (20, 7, 18) architecture was preferred, as it was found to be experimentally more resistant and robust in performance.

Further experiments suggested that the optimal ANN architecture was independent of the number of samples in the training set. Even when 50% of the available data was used in the training set, the effect of *n* on the RMS error value remained very similar.

Training and verification

The training and verification was done with (20, 7, 18) architecture. An example of the results, using a training set with 378 samples and a verification set with 20 samples, is summarized in table 5. A 100% correct classification was reached over both the training and verification sets.

Identification of the test samples

The trained (20, 7, 18) model was then tested with the test set data. Only three specimens of Thysanoptera (one specimen of each *Chirothrips hamatus*, *C. manicatus* and *Dendrothrips degeeri*) were misidentified (table 5). The achieved 97% success rate indicates the potential effectiveness of the trained network in an independent test.

Number of specimens in the training set and the accuracy of identification

Further, we examined the effect of the number of samples in the training set on a successful identification using repeated verification experiments. For this purpose, we used the recombined original data set and the same (20, 7, 18) architecture as in the preceding test, which was, however, repeatedly trained on different numbers of samples.

Successively, we randomly excluded 10, 50, 100 and/or up to 200 samples for the verification set. For each of these numbers, we simulated ten random training runs. If these verification experiments do not provide entirely independent tests, they can, at least, give an estimate of the prediction power of the network under different sizes of the training set.

Results of the simulations are summarized in table 6. For ten verified specimens, a 100% correct classification was achieved in all ten runs; 50–100 verified specimens were identified on average with 98% to almost 100% accuracy. Even when the number of simultaneously verified specimens increased to 200, the verification was excellent and failed only in *ca*. 4% of the cases which were misidentified. We, therefore, consider the predictive reliability of the ANN model as highly satisfactory.

An extrapolation of the data obtained from the verification tests indicated that for a 95% correct identification ($\alpha = 0.05$) of Thysanoptera, a training set with *ca*. 300–350 specimens in total and at least six specimens of each species is sufficient. Using the complete information from the 498 specimens in the data set, the thrips identification by ANN is unproblematic and reliable. If only a few of these 498 specimens were eliminated from the training and used for the prediction, a fully successful identification was practically always reached.

Discussion and conclusions

The high percentage of correctly identified specimens from our data set is promising for a wider use of ANN for thrips identification in practice. Our example enables the identification of 18 thrips species including two pests, *Limothrips cerealium* and *L. denticornis*. When using the same or similar methodology and including other taxa into the training set, the system could be extended or modified for the identification of further Thysanoptera taxa, e.g. a model could specifically be designed for the identification of tospoviruses vectors.

The use of ANN for identification in taxonomy requires first the existence of a training database in which many specimens, correctly identified by experts, are included.

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

Table 5. Results of the classification using the final ANN model in an independent test.

	Training set			Verification set			Test set		
	Total	Correct	Wrong	Total	Correct	Wrong	Total	Correct	Wrong
A. albicinctus	16	16	0	0	0	0	1	1	0
A. astutus	11	11	0	1	1	0	6	6	0
A. ericae	7	7	0	0	0	0	2	2	0
A. fasciatus	12	12	0	2	2	0	3	3	0
A. intermedius	20	20	0	0	0	0	3	3	0
A. versicolor	9	9	0	0	0	0	2	2	0
A. vittatus	10	10	0	0	0	0	5	5	0
C. aculeatus	26	26	0	0	0	0	4	4	0
C. ambulans	14	14	0	0	0	0	1	1	0
C. hamatus	23	23	0	1	1	0	6	5	1
C. manicatus	21	21	0	1	1	0	8	7	1
C. pallidicornis	25	25	0	1	1	0	4	4	0
D. degeeri	56	56	0	5	5	0	13	12	1
D. ornatus	50	50	0	3	3	0	22	22	0
D. saltatrix	9	9	0	0	0	0	5	5	0
L. cerealium	23	23	0	0	0	0	7	7	0
L. consimilis	26	26	0	3	3	0	1	1	0
L. denticornis	20	20	0	3	3	0	7	7	0
Total	378	378	0	20	20	0	100	97	3
%	100	100	0	100	100	0	100	97	3

Table 6. Summary of the results concerning ANN verification experiments with different sizes of the training set ($N = 10$).

No. samples excluded for verification	No. samples in the training set	Mean % of correct identification	Min–Max % of correct identification
10	488	100.00	100
50	448	99.98	96–100
100	398	98.20	97–100
200	298	95.80	92–99

Each specimen has to be characterized by diagnostic variables which are adequate to classify the specimen to a species and discriminate the same from other taxa in the database. Second, an ANN model is designed to find a relationship between the characters (= *input*) and species (= *output*). This model can then be applied for the identification of unknown specimens.

It might be laborious to create the training data set which should consist of a relatively large number of specimens to assure a correct identification. However, the more specimens that are included into the training, the better the real variation in morphology that is captured and the higher the prediction success, i.e. the proportion of correctly identified specimens (Gaston & O'Neill, 2004; Vaňhara *et al*., 2007). However, our experiment on thrips suggests that identification is possible if at least several specimens of a species are included into the training data set (the lowest number of specimens in this study was nine for *Aeolothrips ericae*). The effort needed would, thus, be manageable in a practical application. The same is true for the ANN modelling, which requires knowledge of the basic principles, e.g. how to select and prepare data, how to select an appropriate network and how to interpret the results. The level of user knowledge needed for its successful application is, however, lower than in many other statistical methods.

Our case study on thrips uses, for the main part, morphometric data. The advantage is that homologous characters can be objectively defined as distances on the thrips body in different taxa. Measurement of such distances requires only limited experience in slide-mounting techniques and a basic knowledge of thrips morphology. With the help of a pictorial schema, which specifies the distances, it would be without difficulty for a non-expert. Technical requirements are limited to a microscope with an eye-piece graticule or a digital camera and image analysis software, which would allow more comfort.

The measurements on thrips would probably have to be made by a human operator. It would probably be difficult to fully automate the process unlike, for example, in the wings of Hymenoptera and Diptera, which have a contrasting two-dimensional pattern and where different vein lengths can be extracted from a digital picture automatically (Houle *et al*., 2003; Tofilski, 2004). Also, an analysis of information extracted directly from a digital image as proposed for the wings of Ichneumonid parasitic wasps by Weeks *et al*. (1997) or for genital structures of spiders by Do *et al*. (1999), probably would turn out to be very complicated in thrips as they are much smaller. Consequently, it would be more difficult to mount exactly the same position before image acquisition. Acquiring the values for the 20 input characters used in our study took approximately ten minutes for each thrips specimen. This time is, however, to be compared to the duration of the identification process using a traditional determination key. Since the classification of a specimen using a trained ANN is then almost immediate, the identification by ANN can be achieved in less time.

The practical use of automated systems and supervised ANN for species identification has several limitations, which have been thoroughly discussed in general by Weeks & Gaston (1997) and Gaston & O'Neill (2004). For Thysanoptera, errors in identification could be caused by intraspecific variation due to geographical differences in morphology between populations, as is observed for the females of *Limothrips cerealium*, which are always macropterous in central Europe, while pterygopolymorphism has

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

been recorded in southern Europe (zur Strassen, 2003). For apterous *L. cerealium*, in order to have females correctly identified by ANN, several such specimens would probably have to be included in the training set. The quality of the training set is an essential prerequisite to obtaining reliable identifications, and any known variation in morphology should be taken into account and represented in the training set.

For the same reason, only the taxa included in the training set can be identified by an ANN model. Specimens of other species will, thus, tend to be identified as belonging to one of the training set, and this would lead to false identification. This problem can partly be eliminated by adjusting the post-processing stage of ANN modelling, setting the accept and reject thresholds for the output activation scores as confidence limits. Only if the output score for a corresponding network unit exceeded the accept level and all other units failed to exceed the reject threshold would the identification be made. Otherwise, the specimen's identity would stay undecided (unknown). Alternatively, samples of other similar species occurring, e.g. in a certain geographical region, could be included in the training set and the identifier extended to also include those taxa. The model used by us works as a non-hierarchical system which identifies simultaneously 18 species belonging to four genera. Possible larger identification systems could take advantage of a hierarchical approach, in which specimens would first be identified, for example, to genus and then to species, which would also be computationally more efficient (Do *et al*., 1999; Gaston & O'Neill, 2004).

Weeks & Gaston (1997) envisaged a combination of ANN-based automated identification systems with multi-access keys. This combination seems particularly suitable for Thysanoptera, for which such keys already exist, at least for species of economic importance (Moritz *et al*., 2001, 2004). ANN could refine their performance, in that they can compare multivariate continuous morphometric data efficiently or simply provide an independent check, e.g. for critical taxa (Clark, 2003), specimens partly damaged by rough collecting methods (sticky traps, aeroplanctonic traps, Tullgren photoeclectors, etc.) or old slides. Thus, ANN probably have the potential to enhance the practice of routine Thysanoptera identification. In a similar way, they could also serve for other insect pests, e.g. aphids for which morphometrics is also extensively used for identification.

Acknowledgements

For financial support, the Ministry of Education and the Masaryk University (grant No. MSM 0021622416) are acknowledged. The paper was also partly supported by the grant (No. MK 00009486201) from the Ministry of Culture of the Czech Republic to the Moravian Museum and the VEGA grant no. 1/4339/07 to Peter Fedor. We thank Dr Laurence Mound (Canberra, Australia) for his kind comments.

References

Aldrich, B.T., Magirang, E.B., Dowell, F.E. & Kambhampati, S. (2007) Identification of termite species and subspecies of the genus *Zootermopsis* using near-infrared reflectance spectroscopy. *Journal of Insect Science* **7**, 18. Available online: insectscience.org/7.18 (accessed 25 October 2007).

Bishop, C. (1995) *Neural Networks for Pattern Recognition*. 504 pp. New York, Oxford University Press.

Brunner, P.C., Fleming, C. & Frey, J.E. (2002) A molecular identification key for economically important thrips species (Thysanoptera: Thripidae) using direct sequencing and a PCR-RFLP-based approach. *Agricultural and Forest Entomology* **4**, 127–136.

Clark, J.Y. (2003) Artificial neural networks for species identification by taxonomists. *BioSystems* **72**, 131–147.

Chesmore, D. (1999) Technology transfer: applications of electronic technology in ecology and entomology for species identification. *Natural History Research* **5**, 111–126.

Chesmore, D. (2004) Automated bioacoustic identification of species. *Anais da Academia Brasileira de Ciençias* **76**, 435–440.

Cranston, P.S. (2005) Ancient and modern – toolboxes for e-bugs. *Systematic Entomology* **30**, 183–185.

Do, M.T., Harp, J.M. & Norris, K.C. (1999) A test of a pattern recognition system for identification of spiders. *Bulletin of Entomological Research* **89**, 217–224.

Edwards, M. & Morse, D.R. (1995) The potential for computer-aided identification in biodiversity research. *Trends in Ecology and Evolution* **10**, 153–158.

Fausett, L. (1994) *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*. 461 pp. New York, Prentice Hall.

Fedor, P.J., Pelikán, J., Cyprich, D. & Krumpál, M. (2001) Thrips (Thysanoptera) in the nests of birds and mammals of the NPR Jurský Šúr. *Folia Faunistica Slovaca* **6**, 69–73.

Franssen, C.J.H. & Mantel, W.P. (1965) *Thripsen in graangewassen (levenswijze, economische betekenis en bestrijding) I. Levenswijze* (Thrips in cereal crops (biology, economic importance and control) I. Biology). Instituut voor Plantenziektenkundige Onderzoek, Wageningen, Mededeling No. 381.

Gaston, K.J. & May, R.M. (1992) Taxonomy of taxonomists. *Nature* **356**, 281.

Gaston, K.J. & O'Neill, M.A. (2004) Automated species identification: why not? *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **359**, 655–667.

Hernández-Borges, J., Corbella-Tena, R., Rodriguez-Delgado, M.A., Garcia-Montelongo, F.J. & Havel, J. (2004) Content of aliphatic hydrocarbons in limpets as a new way for classification of species using Artificial Neural Networks. *Chemosphere* **54**, 1059–1069.

Houle, D., Mezey, J., Galpern, P. & Carter, A. (2003) Automated measurement of *Drosophila* wings. *BMC Evolutionary Biology* **3**, 25. http://www.biomedcentral.com/1471-2148/3/25 (accessed 2 April 2007).

Jones, D.R. (2005) Plant viruses transmitted by thrips. *European Journal of Plant Pathology* **113**, 119–157.

Larsson, H. (2005) Economic damage by *Limothrips denticornis* in rye, triticale and winter barley. *Journal of Applied Entomology* **129**, 386–392.

Lewis, T. (1973) *Thrips: Their Biology, Ecology and Economic Importance*. 349 pp. London, New York, Academic Press.

Lewis, T. (Ed.) (1997) *Thrips as Crop Pests*. 740 pp. Oxford and New York, CAB International.

Marcondes, C.B. & Borges, P.S. (2000) Distinction of males of the *Lutzomyia intermedia* (Lutz & Neiva, 1912) species complex by ratios between dimensions and by an Artificial Neural Network (Diptera: Psychodidae, Phlebotominae). *Memorias del Instituto de Oswaldo Cruz* **95**, 685–688.

Moore, A. & Miller, R.H. (2002) Automated identification of optically sensed aphid (Homoptera: Aphidae) wingbeat

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

waveforms. *Annals of the Entomological Society of America* **95**, 1–8.

Moritz, G., Delker, C., Paulsen, M., Mound, L.A. & Burgermeister, W. (2000) Modern methods for identification of Thysanoptera. *EPPO Bulletin* **30**, 591–593.

Moritz, G., Morris, D.C. & Mound, L.A. (2001) *ThripsID – pest thrips of the world. CD ROM*. ACIAR, CSIRO Publishing.

Moritz, G., Mound, L.A., Morris, D.C. & Goldarazena, A. (2004) *Pest thrips of the world, visual and molecular identification of pest thrips. CD ROM*. Center for Biological Information Technology AUD, Lucid, University of Queensland.

Mound, L.A. (2001) So many thrips – so few tospoviruses? pp. 3–6 *in* Marullo, R. & Mound, L.A. (*Eds*) *Thrips and Tospoviruses: Proceedings of the 7th International Symposium on Thysanoptera*. http://www.ento.csiro.au/thysanoptera/symposium.html) (accessed 15 February 2007).

Mound, L.A. & Kibby, G. (1998) *Thysanoptera: An Identification Guide*. 2nd edn. 70 pp. Oxford and New York, CAB International.

Patterson, D. (1996) *Artificial Neural Networks: Theory and Applications*. 506 pp. Singapore, Prentice Hall.

Pelikán, J., Fedor, P., Krumpál, M. & Cyprich, D. (2002) Thrips (Thysanoptera) in nests of birds and mammals in Slovakia. *Ekológia (Bratislava)* **21**, 275–282.

Platnick, N.I., Russell, K.N. & Do, M.T. (2005) SPIDA-web. Species Identified Automatically. A neural network based automated identification system for biological species. https://research.amnh.org/invertzoo/spida/common/index.htm (accessed 15 February 2007).

Rugman-Jones, P.F., Hoddle, M.S., Mound, L.A. & Stouthamer, R. (2006) Molecular identification key for pest species of *Scirtothrips* (Thysanoptera: Thripidae). *Journal of Economic Entomology* **99**, 1813–1819.

Schliephake, G. & Klimt, K. (1979) *Thysanoptera. Die Tierwelt Deutschlands 66*. 477 pp. Jena, G. Fisher Verlag.

StatSoft, Inc. (2004) STATISTICA 7.0. Computer program.

Toda, S. & Komazaki, S. (2002) Identification of thrips species (Thysanoptera: Thripidae) on Japanese fruit trees by polymerase chain reaction and restriction fragment length polymorphism of the ribosomal ITS2 region. *Bulletin of Entomological Research* **92**, 359–363.

Tofilski, A. (2004) DrawWing, a program for numerical description of insect wings. *Journal of Insect Science* **4**, 17. Available online: insectscience.org/4.17 (accessed 25 October 2007).

Trajan Software, Ltd (1996–1998) Trajan Neural Network Simulator, version 3.0D. Computer program.

Vaňhara, J., Muráriková, N., Malenovský, I. & Havel, J. (2007) Artificial Neural Networks for fly identification: a case study from the genera *Tachina* and *Ectophasia* (Diptera, Tachinidae). *Biologia* **62**, 462–469.

Weeks, P.J.D. & Gaston, K.J. (1997) Image analysis, neural networks, and the taxonomic impediment to biodiversity studies. *Biodiversity and Conservation* **6**, 263–274.

Weeks, P.J.D., Gauld, I.D., Gaston, K.J. & O'Neill, M.A. (1997) Automating the identification of insects: a new solution to an old problem. *Bulletin of Entomological Research* **87**, 203–211.

Weeks, P.J.D., O'Neill, M.A., Gaston, K.J. & Gauld, I.D. (1999) Automating insect identification: exploring the limitations of a prototype system. *Journal of Applied Entomology* **123**, 1–8.

zur Strassen, R. (2003) *Die terebranten Thysanopteren Europas und des Mittelmeer-Gebietes*. 277 pp. Keltern, Goecke and Evers.

zur Strassen, R. (2005) *Thysanoptera. Fauna Europaea, version 1.2*. http://www.faunaeur.org (accessed 9 March 2007).

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

Appendix: Table A1. Minimum and maximum values recorded for characters used for ANN Thysanoptera identification (measurements in μm) for males.

character		1	2	3	4	5	6	7	8	9	10	11	14	15	16	17	18	19
A. albicinctus	min	210.6	171.6	315.9	–	–	–	–	1627.5	220.4	163.8	74.1	134.6	60.5	35.1	25.4	yes	no
	max	218.4	179.4	323.7	–	–	–	–	1674.0	224.3	167.7	78.0	140.4	66.3	39.0	27.3		
A. astutus	min	175.5	144.3	253.5	159.9	37.1	829.3	116.3	1581.0	214.5	132.6	72.2	48.8	60.5	23.4	25.4	no	no
	max	187.2	159.9	273.0	175.5	42.9	852.5	131.8	1643.0	253.5	156.0	78.0	58.5	74.1	25.4	27.3		
A. fasciatus	min	177.8	158.9	272.4	151.3	37.8	796.9	90.2	1563.6	211.8	124.8	60.5	71.9	62.4	30.3	22.7	no	no
	max	190.1	169.9	291.2	161.8	40.4	851.9	96.4	1671.6	226.5	133.5	64.7	76.8	66.7	32.4	24.3		
A. intermedius	min	198.9	163.8	266.7	191.1	45.7	821.5	121.1	1550.0	205.8	135.3	76.6	68.6	53.3	31.2	27.3	no	no
	max	210.6	175.5	292.5	200.9	47.7	868.7	126.4	1782.5	214.6	141.1	79.9	74.1	62.4	39.0	29.3		
C. aculeatus	min	105.6	96.1	183.8	–	–	–	–	1489.1	210.0	168.0	49.6	24.5	58.4	–	–	no	no
	max	113.3	107.4	205.1	–	–	–	–	1581.3	223.0	178.4	52.7	26.0	62.1	–	–		
C. hamatus	min	101.1	91.3	173.0	–	–	–	–	1003.2	181.1	129.9	53.1	25.1	59.7	–	–	yes	yes
	max	126.3	143.5	222.3	–	–	–	–	1517.6	225.2	180.1	71.6	31.2	72.8	–	–		
C. manicatus	min	79.7	62.2	136.0	–	–	–	–	729.6	161.4	124.5	37.2	12.8	34.9	–	–	no	no
	max	110.8	95.2	196.3	–	–	–	–	1307.2	236.2	169.3	51.2	18.5	44.7	–	–		
C. pallidicornis	min	94.9	74.3	167.1	–	–	–	–	1069.3	192.3	134.6	38.6	16.1	40.7	–	–	yes	no
	max	101.1	85.5	188.5	–	–	–	–	1261.6	216.5	161.4	44.5	21.0	48.6	–	–		
D. degeeri	min	139.9	50.5	126.3	101.1	25.3	642.6	56.4	765.0	136.0	69.0	56.4	29.2	56.4	23.3	18.5	yes	no
	max	149.6	76.8	141.9	106.9	31.1	749.7	64.1	979.2	165.2	79.7	63.2	33.0	58.3	38.9	24.3		
D. saltatrix	min	140.9	68.0	145.8	97.2	27.2	642.6	67.1	979.2	155.5	77.7	52.5	29.2	55.4	24.3	15.5	yes	no
	max	143.8	68.0	146.7	98.1	31.1	657.9	68.0	994.5	160.3	77.7	58.3	35.0	57.3	25.3	16.5		
L. cerealium	min	149.9	155.6	256.1	–	–	–	–	1468.9	172.9	119.1	65.3	29.4	70.2	–	–	yes	no
	max	157.2	163.2	268.7	–	–	–	–	1541.0	181.4	125.0	68.5	30.8	73.6	–	–		
L. consimilis	min	121.5	140.7	229.6	–	–	–	–	1231.6	169.1	119.1	60.5	26.6	47.4	–	–	no	no
	max	149.2	177.0	268.5	–	–	–	–	1773.4	221.6	153.1	72.5	29.8	55.7	–	–		
L. denticornis	min	181.0	188.5	301.6	–	–	–	–	1562.9	229.1	141.3	76.4	32.0	73.5	–	–	no	no
	max	188.4	196.3	314.0	–	–	–	–	1627.2	238.5	147.1	79.5	33.4	76.5	–	–		

https://doi.org/10.1017/S0007485308005750 Published online by Cambridge University Press

Table A2. Minimum and maximum values recorded for characters used for ANN Thysanoptera identification (measurements in μm) for females.

character		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A. albicinctus	min	238.3	215.6	334.6	189.2	49.4	947.2	135.3	2180.1	245.9	181.6	83.2	410.7	227.0	94.6	64.3	34.0	24.7	yes	no
	max	253.1	229.0	353.5	200.9	53.2	1005.8	143.7	2394.8	279.7	200.9	92.4	492.4	249.1	104.4	68.3	36.2	27.6		
A. astutus	min	195.0	158.9	315.9	190.5	37.8	1052.5	150.4	2362.4	278.2	189.2	83.8	461.0	194.3	62.9	79.4	22.7	28.6	no	no
	max	212.9	169.5	333.1	208.9	40.6	1131.5	160.4	2602.3	317.3	201.8	100.4	502.1	220.9	68.3	85.8	24.2	32.6		
A. ericae	min	227.0	158.9	336.7	83.2	49.2	902.1	105.8	2025.1	249.1	166.5	68.1	454.0	219.4	75.7	64.3	34.0	23.4	no	no
	max	242.2	169.5	359.2	88.8	52.5	962.6	120.3	2181.8	266.4	177.6	72.7	484.4	234.1	80.7	68.6	36.3	24.2		
A. fasciatus	min	198.9	171.5	316.3	188.6	36.2	991.9	121.1	2271.5	214.5	159.9	72.4	468.0	207.7	62.4	74.1	34.3	23.4	no	no
	max	212.6	184.8	333.8	201.9	41.0	1023.0	134.3	2407.1	237.9	178.1	78.0	498.8	218.6	76.1	83.1	39.0	33.2		
A. intermedius	min	209.1	133.1	304.2	190.1	45.6	952.1	136.0	1858.8	201.5	152.1	76.1	467.7	232.0	62.7	66.7	38.0	26.6	no	no
	max	240.4	202.7	359.7	218.6	51.2	1093.0	192.9	2385.0	282.2	190.8	87.4	608.4	245.7	71.5	72.8	40.4	31.8		
A. versicolor	min	190.5	167.7	262.9	133.4	38.1	795.0	90.9	1689.5	209.6	133.4	72.4	401.7	198.1	55.2	50.7	27.3	24.8	no	no
	max	206.7	175.5	292.5	148.2	44.9	837.0	110.6	1840.1	218.6	143.1	76.1	437.2	241.8	62.4	54.6	31.2	27.3		
A. vittatus	min	195.0	136.5	253.5	156.0	36.2	837.0	100.8	1782.5	218.4	128.7	70.2	386.7	187.2	58.5	52.7	35.1	23.4	no	no
	max	220.5	168.4	273.8	172.9	46.8	961.0	143.4	2077.5	248.6	148.2	74.4	425.0	244.6	70.2	62.4	42.9	30.2		
C. aculeatus	min	123.8	117.5	217.9	125.1	43.6	904.0	113.7	1630.2	249.5	191.9	59.5	230.3	115.1	27.5	58.7	28.4	14.2	no	no
	max	142.6	136.6	245.9	148.5	49.5	1019.9	134.7	1872.6	281.5	228.6	88.2	281.5	140.7	31.0	67.3	32.7	19.8		
C. ambulans	min	120.5	107.9	203.5	–	–	–	–	1246.4	251.9	168.4	59.9	160.9	94.5	22.2	46.2	23.3	14.0	no	no
	max	137.7	116.0	218.3	–	–	–	–	1431.7	278.9	186.0	66.1	173.6	118.3	26.3	56.5	29.2	15.7		
C. hamatus	min	110.8	103.0	194.4	126.3	38.9	914.5	77.7	1444.0	220.4	189.0	63.0	222.4	118.1	25.3	67.1	23.3	11.7	yes	yes
	max	155.5	132.2	242.9	137.4	48.6	1059.1	88.9	1869.6	283.4	240.1	78.7	271.6	145.7	33.0	77.7	31.1	17.5		
C. manicatus	min	106.9	93.3	174.9	97.2	23.0	729.6	67.1	1276.8	196.8	157.5	53.1	104.3	84.6	20.4	46.8	24.9	14.4	no	no
	max	126.3	106.9	212.7	133.2	45.7	1041.8	87.5	1617.2	261.8	189.3	60.4	197.3	120.1	24.9	59.6	35.8	19.9		
C. pallidicornis	min	123.4	105.4	201.2	101.6	36.4	794.4	72.8	1678.6	232.9	166.9	60.2	211.5	120.3	20.1	54.6	27.8	15.3	yes	no
	max	128.3	117.1	224.4	116.6	42.9	854.6	78.1	1922.9	276.7	196.8	63.2	221.3	128.4	27.3	67.3	31.2	15.6		
D. degeeri	min	162.3	62.2	139.9	117.6	31.1	657.9	68.0	849.2	161.3	85.5	52.5	165.2	81.6	29.2	35.0	36.9	17.5	yes	no
	max	194.4	97.2	194.4	143.8	40.8	918.0	99.1	1269.9	219.6	104.9	71.9	206.0	128.3	35.0	62.2	46.6	27.2		
D. ornatus	min	153.5	64.1	145.8	104.9	27.2	673.2	58.3	1025.1	190.5	81.6	58.3	155.5	87.5	28.2	52.5	37.9	19.4	yes	no
	max	178.8	81.6	186.6	134.1	38.9	841.5	141.9	1208.7	217.7	104.9	103.0	194.4	114.7	35.0	64.1	48.6	28.2		
D. saltatrix	min	144.8	64.1	153.5	106.9	29.2	726.8	77.7	1071.0	174.9	77.7	55.4	149.6	87.5	34.0	58.3	25.3	16.5	yes	no
	max	169.1	81.6	174.9	112.7	33.0	818.6	85.5	1147.5	191.4	87.5	65.1	165.2	108.8	36.0	62.2	29.2	19.4		
L. cerealium	min	173.2	155.9	269.4	135.9	42.1	928.3	86.2	1760.7	210.5	140.3	62.0	280.6	131.8	40.4	84.7	30.6	19.2	yes	no
	max	191.3	200.2	315.6	153.9	51.3	1018.1	98.6	2128.9	239.7	179.8	87.9	387.5	167.8	43.4	91.2	41.4	25.6		
L. consimilis	min	164.4	167.4	298.2	141.2	45.4	892.3	92.8	1745.6	210.5	144.2	72.5	292.3	123.4	42.3	77.0	34.8	19.3	no	no
	max	176.9	188.5	320.7	147.7	48.6	988.0	98.1	2052.0	228.3	177.1	80.7	358.2	139.7	48.6	89.4	42.8	25.3		
L. denticornis	min	186.2	194.0	293.5	114.4	40.7	834.3	81.5	1714.2	196.8	149.3	70.9	269.7	149.6	42.7	81.6	40.7	22.4	no	no
	max	225.4	239.1	388.7	159.4	62.2	1124.8	112.7	2280.0	295.2	196.8	90.5	307.0	159.4	52.5	104.9	48.6	28.2		