Journal of Chromatography, 349 (1985) 39–48
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 641

USE OF COMPUTERIZED PATTERN RECOGNITION IN THE STUDY OF THE CUTICULAR HYDROCARBONS OF IMPORTED FIRE ANTS

II. COMPARISON OF THE CUTICULAR HYDROCARBON PATTERNS BETWEEN DIFFERENT COLONIES OF *SOLENOPSIS RICHTERI*

JEFFREY H. BRILL, TOM MAR, HOWARD T. MAYFIELD and WOLFGANG BERTSCH
Department of Chemistry, University of Alabama, University, AL 35486 (U.S.A.)

SUMMARY

Gas chromatography (GC) data obtained from the cuticular hydrocarbons of the black imported fire ants are treated by methods of pattern recognition. Based on a recently described sample preparation procedure, GC data are normalized to eliminate slight variations in chromatographic conditions, and converted to the proper format for discriminant analysis by computer. The results of several methods of data treatment and display are discussed, based on the chemometrics system package, ARTHUR.

INTRODUCTION

The cuticular hydrocarbons of insects have been shown to be important compounds in the chemical communication of insects, besides their major function of preventing desiccation[1–3]. The cuticular hydrocarbons of the imported fire ants have been characterized. They have been shown to be complex mixtures containing homologous series of monomethyl and dimethyl alkanes[4,5]. It has been suggested that the cuticular hydrocarbons are involved in species and caste recognition[3,6]. Another function of the cuticular hydrocarbons might be in colony recognition. Worker and soldier ants are extremely sensitive to alien colony odors. Alien intruders in a colony appear to be detected by contact chemoreception[3,7,8], which would indicate that the cuticle plays a role in recognition. If the cuticular hydrocarbons do play a role in intraspecific colony recognition, then it can be expected that there might be differences in the cuticular hydrocarbon profiles from different colonies.

The cuticular hydrocarbon profiles from a number of different colonies of the black imported fire ant, *Solenopsis richteri*, were investigated using a number of chemometric data treatment methods, in order to determine if there are significant differences between the profiles from different colonies. These methods have been discussed in an earlier publication[9].

Initially, a relatively small data set incorporating the hydrocarbon profiles from three different colonies was used. After examining the results, it was then de-

0021-9673/85/$03.30

cided to expand the data set to contain the profiles of ten different colonies in order to determine whether similar results could be obtained with a relatively large data set.

EXPERIMENTAL

Samples

The black imported fire ants used were collected from nests in Pickens county, west central Alabama, U.S.A. In the first part of these studies, samples from three different nests were collected. Two of the nests were nearest neighbors (located *ca.* 5 m apart) and the third nest was located *ca.* 24 km away. For convenience, the two nearest neighbor colonies will be designated A and B, while the third colony will be designated Y.

In the second part, 10 colonies were collected in Pickens and Lamar counties, Alabama, as well as in east-central Mississippi (Moxube and Lowndes counties). These included four pairs of nearest neighbor colonies.

Samples were prepared using a recently described solventless procedure. Details of this method have been described elsewhere[9,10].

Gas chromatography

The samples were examined by gas chromatography (GC), using a Hewlett-Packard 5830A gas chromatograph, fitted with an injector port suitable to accept the Pyroprobe insert. The column used was a 16 m × 0.25 mm I.D. glass capaillary, coated with a 0.25-μm film of immobilized OV-1 (prepared according to the method described by Grob and Grob[11]). The column was temperature-programmed from 80° to 300°C at 8°C/min. The carrier gas was helium, at a flow-rate of *ca.* 1.0 ml/min. The split ratio was set to *ca.* 100:1. Each sample was treated with an internal standard, containing 250 ng each of tricosane ($C_{23}H_{48}$) and dotriacontane ($C_{32}H_{66}$).

GC–mass spectrometry (GC–MS) was also performed on a number of the samples. The system used was a Hewlett-Packard 5985A instrument, operated in the electron impact (EI) mode. The column was a 15 m × 0.32 mm I.D. fused-silica capillary, coated with a 0.10-μm film of DB-5 phase (J&W Scientific, Rancho Cordova, CA, U.S.A.). Other chromatographic conditions were as above.

Data analysis

From each of the three colonies, 10–15 workers of roughly the same size were selected. These workers were analyzed individually by Pyroprobe dynamic headspace analysis[10]. The retention time and absolute area for each peak in the diagnostic (hydrocarbon) region of each chromatogram were encoded onto computer cards and the data, together with the marker peaks for the SETUP program[12], were transferred to the mainframe computer. The SETUP parameters for transducing the data set into multivariate form are listed in Table I. The resulting data set consisted of 33 data vectors, comprising 40 features. Colony A was represented by thirteen data vectors and was designated as category 1. Colony B was represented by ten data vectors and was assigned category 2, while colony Y consisted of ten data vectors and was designated category 3.

The raw data set (designated as data set 1a) was submitted to the ARTHUR

TABLE I

ADJUSTABLE PARAMETERS USED FOR DATA SET TRANSDUCTION BY SETUP

Parameter	
Maximum allowed retention time error for matching peaks	0.10 min
Number of marker peaks per chromatogram	7
Minimum retention time distance between non-redundant features*	0.05 min
Minimum frequency of occurrence of acceptable features*	0%

* Features are based on normalized peak areas.

program for chemometric analysis. Various preprocessing and feature selection procedures were examined, in an attempt to enhance the discriminating ability of the data in the raw data set, 1a. A new data set, designated 2a, was obtained by first autoscaling the data in 1a, then weighting all the data vectors, using the Fisher weighting method. Finally, a third data set, designated 3a, was obtained from data set 2a by determining the number of weighed features which gave the greatest discrimination between the categories. After visual inspection of the non-linear maps, four features were found to be the most significant.

Various classification methods were tested, to determine the discriminating ability of the data in set 3a. Table II lists the training set–test set combinations used in the classification tests. Classification testing was performed using the *K*-nearest neighbor method (KNN), the linear learning machine method (LLM), the SIMCA method and the Bayes method.

The data from the GC profiles of the ten colonies were treated in the same manner as that for the three colonies, with the following changes: the classification tests were omitted, since the non-linear maps showed discernible clustering of the different categories. The data set resulting from the SETUP treatment consisted of

TABLE II

TRAINING SET–TEST SET COMBINATIONS FOR THE CLASSIFICATION METHODS TESTED

R = (No. of training set patterns)/(No. of features).

Method	*Category 1 (colony A)*		*Category 2 (colony B)*		*Category 3 (colony Y)*	
	Training set	*Test set*	*Training set*	*Test set*	*Training set*	*Test set*
KNN ($K = 1$)	52 ($R = 13$)	13	40 ($R = 10$)	10	40 ($R = 10$)	10
LLM	104 ($R = 26$)	26	80 ($R = 20$)	20	80 ($R = 20$)	20
SIMCA	52 ($R = 13$)	13	40 ($R = 10$)	10	40 ($R = 10$)	10
Bayes	52 ($R = 13$)	13	40 ($R = 10$)	10	40 ($R = 10$)	10

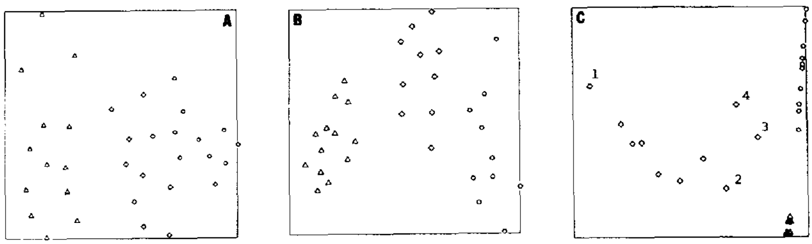

Fig. 1. Non-linear maps of hydrocarbon profiles from three colonies of the black imported fire ant. A, raw data set 1a; B, Fisher weighted data set 2a; C, best four features by Fisher weight data set 3a. 1–4 in C are outliers (see text). △ = Colony A; ○ = colony B; ◇ = colony Y.

160 data vectors, comprising 52 features. Each category (colony) was represented by an average of 16 data vectors.

RESULTS AND DISCUSSION

Comparison of the profiles of three colonies

The clustering of the categories is improved with the various prepocessing procedures. The non-linear map (NLM) of the raw data for the three colonies, 1a, shows incomplete separation of the three categories (see Fig. 1A). The data for the two nearest neighbor nests appear to be separated, but the pattern vectors for category 3 are interspersed between them. Thus, there is a poor separation among the three categories.

After the first preprocessing step, in which the data vectors were weighed using the Fisher method, there was a marked improvement in the separation in data set 2a. Fig. 1B shows a separation beginning to appear between all three categories, as the clustering becomes tighter.

The final preprocessing step was feature selection. It was determined that four features (by Fisher weight) gave the greatest discrimination among the categories. Fig. 1C shows the NLM for data set 3a, in which the three colonies were compared, using the four best features as the basis for discrimination. The two nearest neighbor nests are well separated, colony A showing a very tight clustering. Colony Y, which was located some 24 km from the other two colonies, shows a rather scattered clustering.

Fig. 2 shows the cuticular hydrocarbon profiles for the three colonies. The four most discriminating features are indicated by the arrows. Feature 1 appeared to be a branched hydrocarbon eluted between heptacosane and octacosane. Feature 2 was identified as *n*-hexacosane, while feature 3 proved to be *n*-heptacosane. Feature 4 appeared to be a branched hydrocarbon, eluted between pentacosane and hexacosane.

The clustering of the data for colony Y is rather scattered. This is due to a number of factors. There are differences in the cuticular hydrocarbon profiles of individual ants from the same colony[13]. This would preclude the data vectors in a colony profile from overlapping completely, even if the experimental errors were negligible. The GC profiles for colony Y were examined visually. Sometimes, small and apparently insignificant peaks (features) have a great diagnostic value in discrim-

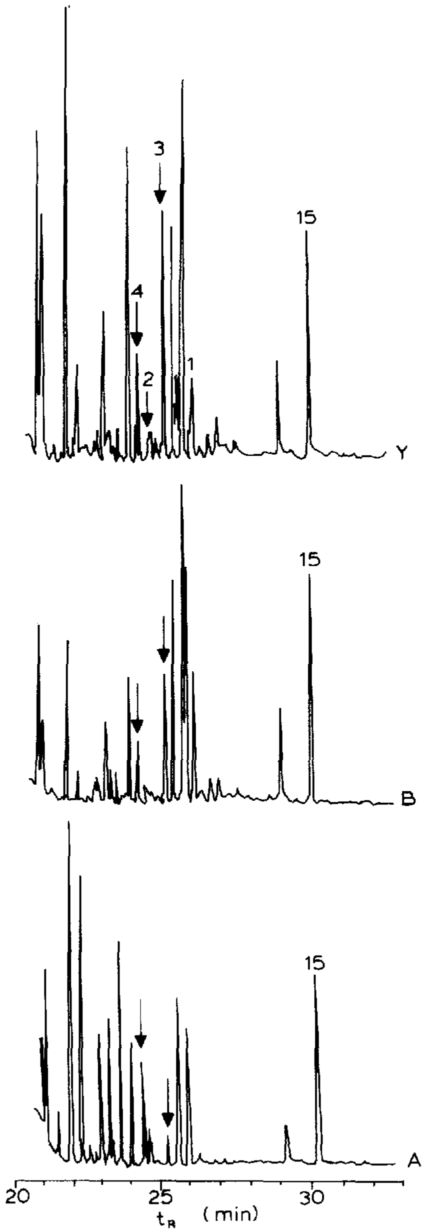

Fig. 2. Comparison of the cuticular hydrocarbon profiles of three colonies of the black imported fire ant. Colonies A and B are nearest neighbors (5 m apart). Colony Y was located *ca.* 24 km away. The four best features (by Fisher weight) are indicated by 1–4. IS, internal standard (dotriacontane).

inating between two or more patterns. This is the case in these studies. Feature 1 in the profile of colony Y is a small peak on the shoulder of a large peak (see Fig. 2Y). This presents an integration problem. In a few of the chromatograms, the integrator

TABLE III

RESULTS OF CLASSIFICATION TEST ON THREE COLONIES OF THE BLACK IMPORTED FIRE ANT (DATA SET No. 3a)

Number or runs: category 1 = 13; category 2 = 10; category 3 = 10. Category 1 = colony A, category 2 = colony B, category 3 = colony Y.

Method	*Category*	*Training set*		*Test set*		*Overall percent correct*
		No. of misses	*Percent correct*	*No. of misses*	*Percent correct*	
KNN	1	0	100	0	100	100
($K = 1$)	2	0	100	0	100	100
	3	6	85	1	90	86
LLM	1	0	100	0	100	100
	2	0	100	0	100	100
	3	0	100	2	90	98
SIMCA	1	0	100	0	100	100
	2	0	100	0	100	100
	3	9	77.5	2	80	78
Bayes	1	0	100	0	100	100
	2	0	100	0	100	100
	3	0	100	4	60	87.9

did not detect the peak as a separate peak. As a result, it was considered to be missing from these chromatograms by the pattern recognition procedure. Feature 2 in the Colony Y profile is a small peak located in the middle of a group of five small peaks which are very close together (see Fig. 2Y). Once again, this situation presents integration problems. In addition, in some of the chromatograms the peaks were fused, and it was difficult to determine whether there were three, four or five peaks in the group.

Four outliers were identified in the colony Y data cluster (see Fig. 1C). The chromatograms of these outliers were examined carefully in order to determine why they were possibly outliers. For example, the chromatogram for outlier 1 showed a problem with feature 1, which was too small a shoulder peak to be integrated and was thus considered to be missing by the pattern recognition routine. In addition, feature 2 was larger than average. In the other outliers, similar problems were observed, in which one or more of the features differed significantly from the average for the category profile.

Table III shows the results of the various classification tests performed on data set 3a. The results show excellent classification for colonies A and B (nearest neighbors). Classification was less than 100% for colony Y. This is perhaps due to the rather scattered clustering of the data, the reasons for which have been discussed above. The overall results are quite acceptable.

Comparison of the profiles of ten colonies

Fig. 3 shows a comparison of NLM for the n best features, by Fisher weight (where $4 \leqslant n \leqslant 8$). Fig. 3A shows the NLM of the raw data set, before weighing.

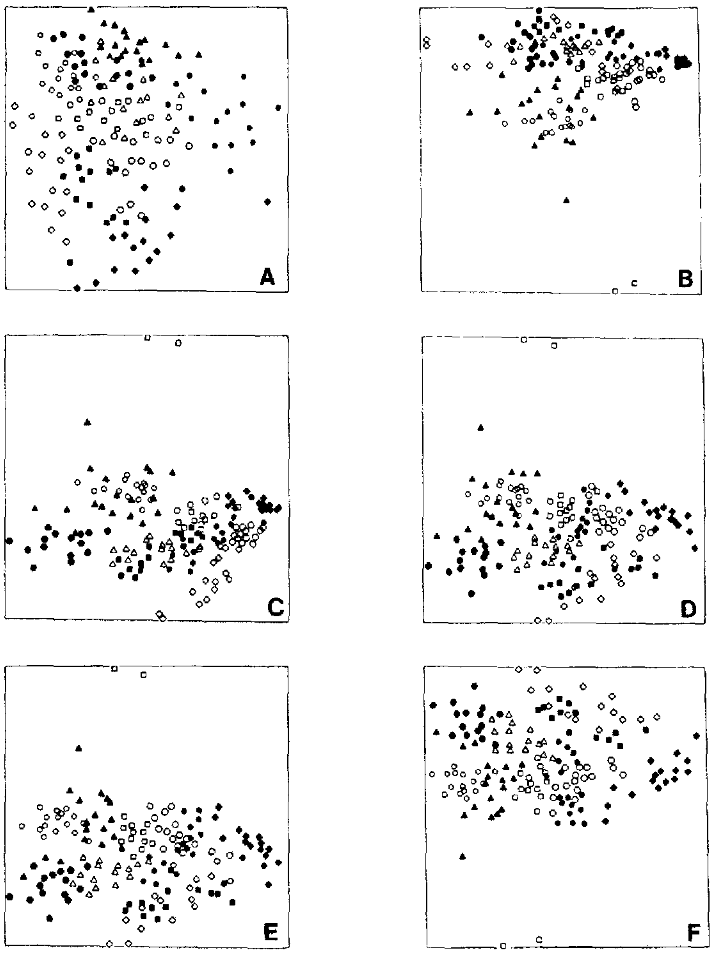

Fig. 3. Non-linear maps of the hydrocarbon profiles of 10 colonies of the imported fire ant. A, raw data; B, best four features; C, best five features; D, best six features; E, best seven features; F, best eight features (by Fisher weight). Nearest neighbor nests are ● and △; ◆ and ⬡; ■ and □; ○ and ◇. The category assignments were as follows: 1, ◆; 2, △; 3, ◆; 4, ⬡; 5, ■; 6, □; 7, ▲; 8, ○; 9, ◇; 10, ●.

The separation between the ten categories is poor in this plot. After Fisher weighting, feature selection shows an improvement as the number of features selected is increased. An optimum number of features is reached, where the separation among the different categories is maximized. Thereafter, increasing the number of features results in a deterioration in the category separations, due to an increase in noise in the system. This can be seen by comparing Fig. 3B through F. The optimum number of features was *ca.* 5 (see Fig. 3C). However, the categories are not completely separated from one another. This is possibly due to there being a relatively large number (10) of categories. Moreover, the ten profiles are from the same species and are not greatly different from one another. Nonetheless, as was observed in the comparison of three colonies, the nearest neighbor colony profiles showed the greatest separation (see Fig. 3C).

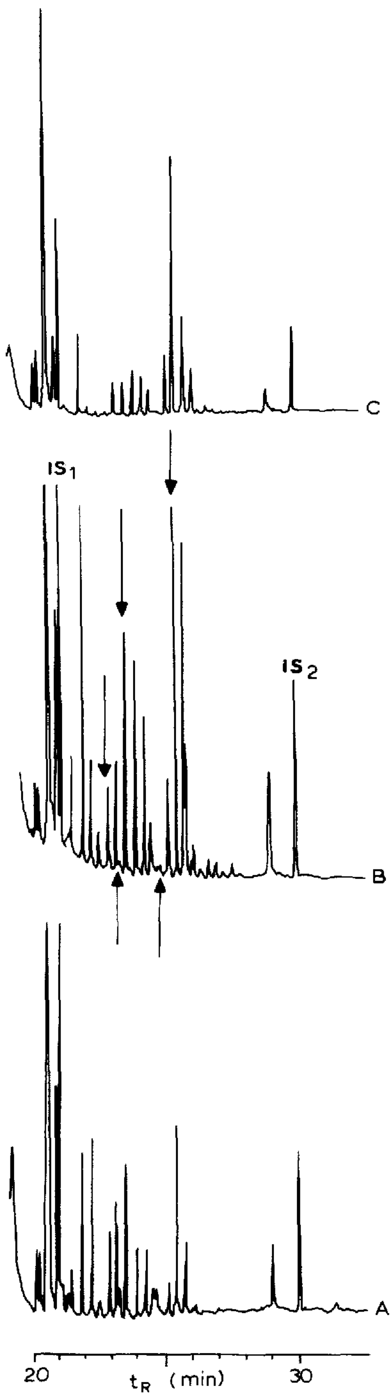

Fig. 4. Hydrocarbon profiles from three of the ten colonies of the black imported fire ant. Arrows indicate the best five features, by Fisher weight. B and C were nearest neighbor colonies.

In the NLM's (see Fig. 3, category 6) it appeared that some outliers were present. The profiles corresponding to these outliers were examined more closely. The same problems as discussed in the comparison of the three colonies were observed here, *i.e.*, in the chromatogram of each of the outliers there were some peaks which were larger and/or smaller than average. In addition, there were also small peaks located on the shoulders of much larger peaks, which created integration problems (see Fig. 4).

An additional consideration concerns the NLM procedure itself. NLM is a method whereby a multi-dimensional data space is reduced to a two-dimensional space for display purposes. This reduction can be accomplished only approximately. Every data point in the multi-dimensional data space (where $n \gg 3$) has a distance from every other data point. In NLM, these distances are calculated and then considered to be constants. Upon reduction to two dimensions, NLM attempts to preserve the interpoint distances by minimizing an error function, using a non-linear minimization method[14,15]. It must be remembered that NLM is a method for displaying multi-dimensional data in a form which is easily perceived by the human observer. If classes overlap in the display, they may still be separated in multi-dimensional space by anyone of a number of classification methods.

Fig. 4 shows the hydrocarbon profiles from three of the ten colonies. The five most discriminating features are indicated by the arrows. The profiles from the nearest neighbor colonies show marked differences from each other (compare Fig. 4B and C). This is also the case in the comparison of the three colonies (compare Fig. 2A and B).

These studies show that there are distinct and significant differences between the hydrocarbon profiles of different colonies of the black imported fire ant. These differences appear to be most pronounced between colonies which are nearest neighbors. Workers from neighboring colonies are much more likely to encounter one another while out foraging than workers from more distant nests. Therefore, if the cuticular hydrocarbons are involved in colony recognition, it is expected that differences in the hydrocarbon profiles would be observed between neighboring colonies. However, the degree of territoriality depends on the number of queens in the colony[7,16]. Nests which contain only a single egg-laying queen display a well-defined territoriality and react aggressively towards towards other colonies of the same species. In contrast, polygynous colonies show little aggressiveness and are quite amicable toward neighboring colonies of the same species. It was not determined whether the colonies sampled in these studies were monogynous or polygynous.

Workers and brood from two neighboring colonies were collected and maintained in the laboratory for a while. The hydrocarbon profiles of the two colonies were very different from each other. It was observed that workers (especially the foragers) from these two colonies quickly recognized and reacted aggressively toward each other. The queens from these two colonies wsere not taken; thus it was not possible to determine whether these colonies were monogynous or polygynous.

ACKNOWLEDGEMENTS

This work was supported by a BRS grant from the University of Alabama. We also wish to acknowledge the loan of a pyroprobe from Chemical Data Systems.

REFERENCES

1 N. F. Hadley, G. L. Blomquist and U. N. Lanham, *Insect Biochem.*, 11 (1981) 173.
2 G. L. Blomquist and L. L. Jackson, *Prog. Lipids Res.*, 17 (1979) 319.
3 R. W. Howard and G. L. Blomquist, *Ann. Rev. Entomol.*, 27 (1982) 149.
4 J. B. Lok, E. W. Cupp and G. L. Blomquist, *Insect Biochem.*, 5 (1975) 821.
5 D. R. Nelson, C. L. Fatland, R. W. Howard, C. A. McDaniel and G. L. Blomquist, *Insect Biochem.*, 10 (1980) 409.
6 R. K. Vander Meer and D. P. Wojcik, *Science* (Washington, D.C.), 218 (1982) 806.
7 J. W. S. Bradshaw and P. E. Howse, in J. W. Bell and R. T. Carde (Editors), *Chemical Ecology of Insects*, Sinauer Assoc., Sunderland, MA, 1984, p. 443.
8 E. O. Wilson, *The Insect Societies*, Harvard University Press, Cambridge, MA, 1971, p. 272–278.
9 J. H. Brill, T. Mar, H. T. Mayfield and W. Bertsch, *J. Chromatogr.*, in press.
10 J. H. Brill and W. Bertsch, *Insect Biochem.*, 15 (1985) 49.
11 K. Grob and G. Grob, *J. Chromatogr.*, 213 (1981) 211.
12 H. T. Mayfield and W. Bertsch, *Comput. Appl. Lab.*, 2 (1983) 130.
13 J. H. Brill, *Ph.D. Dissertation*, 1985, in preparation.
14 B. R. Kowalski and C. F. Bender, *J. Am. Chem. Soc.*, 94 (1972) 5632.
15 B. R. Kowalski and C. F. Bender, *J. Am. Chem. Soc.*, 95 (1973) 686.
16 D. J. C. Fletcher, personal communication.