Talanta 83 (2010) 216–224

5085

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Analysis of chemical signals in red fire ants by gas chromatography and pattern recognition techniques

Barry K. Lavine[a,*], Nikhil Mirjankar[a], Robert K. Vander Meer[b]

[a] Department of Chemistry, Oklahoma State University, 107 Physical Science, Stillwater, OK 74078, United States
[b] Center for Agricultural, Medical, and Veterinary Entomology, Agricultural Research Service, U.S. Department of Agriculture, Gainesville, FL 32611, United States

ARTICLE INFO

Article history:
Received 17 June 2010
Received in revised form 31 August 2010
Accepted 8 September 2010
Available online 16 September 2010

Keywords:
Chemometrics
Pattern recognition
Chemical communication
Classification
Social insects

ABSTRACT

Gas chromatographic (GC) profiles of cuticular hydrocarbon extracts obtained from individual and pooled ant samples were analyzed using pattern recognition techniques. Clustering according to the biological variables of social caste and colony were observed. Pooling individual extracts enhanced the recognition of patterns in the GC profile data characteristic of colony. Evidently, the contribution of the ant's individual pattern to the overall hydrocarbon profile pattern can obscure information about colony in the GC traces of cuticular hydrocarbon extracts obtained from red fire ants. Re-analysis of temporal caste and time period data on the cuticular hydrocarbon patterns demonstrates that sampling time and social caste must be taken into account to avoid unnecessary variability and possible confounding. This and the fact that foragers could not be separated from reserves and brood-tenders in all five laboratory colonies studied suggests that cuticular hydrocarbons as a class of sociochemicals cannot model every facet of nestmate recognition in *Solenopsis invicta* which in turn suggests a potential role for other compounds in the discrimination of alien conspecifics from nestmates.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Cuticular hydrocarbons serve several purposes in insects, such as prevention of desiccation and regulation of cuticular permeability [1]. In addition, this class of chemically inert compounds can have semiochemical functions, such as alarm, recruitment, defense, sex attractants, and host attractants [2]. There is also evidence to suggest a supporting role for cuticular hydrocarbons in social insect nestmate recognition [3,4], which is defined as the ability of a worker to discriminate nestmates from alien conspecifics. Nestmate recognition is a well documented phenomenon in the majority of social insect species [5]. In social Hymenoptera (ants, bees, and wasps) nestmate recognition cues are thought to be chemical signals [6].

It is generally accepted that nestmate recognition in ants and other social insects involves the detection of specific cues on the cuticle [7]. Both environmental and/or heritable compounds can contribute to nestmate recognition. Heritable nestmate recognition cues are biosynthesized under genetic control, whereas environmental cues are derived from food, nest material, and anything else with which individual social insects come into contact. In addition, recognition cues may be derived from genetic–environment interactions.

Although there is no direct evident for the role of cuticular hydrocarbons in the nestmate recognition of *Solenopsis invicta* (red imported fire ants), cuticular hydrocarbons have been used as a model to study the quantitative variation in heritable components of colony odor [8,9]. Indirect, correlative evidence regarding a potential role for cuticular hydrocarbons in nestmate recognition of *S. invicta* came from a study of a myrmecophilous beetle [10]. The beetle had been reported to co-habit nests of four *Solenopsis* species, without the ants recognizing them as intruders. It was determined experimentally that the beetle acquired the species-specific cuticular hydrocarbons of its original host, *S. richteri*. After the beetles were removed from the *S. richteri* colony, they lost the hydrocarbon pattern of the host species. On transfer to *S. invicta* colonies, the surviving beetles had acquired the hydrocarbon patterns of the new host species, *S. invicta*. Although the mechanism of the hydrocarbon transfer is unknown, it is obvious that acquisition of host hydrocarbons is correlated with the acceptance of the beetles into fire ant colonies.

For *S. invicta*, cuticular hydrocarbons represent only one genetically controlled chemical class of compounds that contributes to colony odor, and thus far have only been shown to be correlated to nestmate recognition. If cuticular hydrocarbons of *S. invicta* play an important role in nestmate recognition, there must be significant variation in the hydrocarbon pattern from one colony to another. The combination of gas chromatography and pattern recognition analysis has been used to study the relationship between cuticular hydrocarbon patterns obtained from pooled ant samples in *S.*

* Corresponding author. Tel.: +1 405 744 5945; fax: +1 405 744 6007.
E-mail address: bklab@chem.okstate.edu (B.K. Lavine).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.09.010

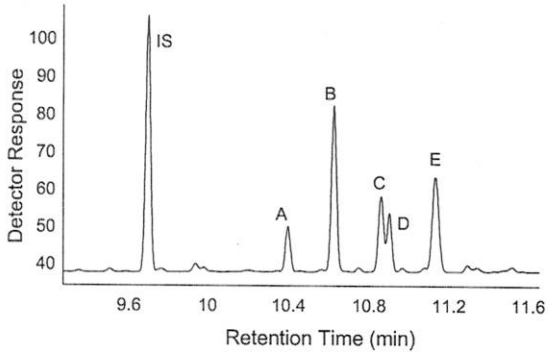

Fig. 1. Gas chromatographic trace of cuticular hydrocarbons from *S. invicta*. The compounds eluting off the capillary column were identified and quantified by GC/MS: (a) heptacosane, (b) 13-methylheptacosane, (c) 13,15-dimethylheptacosane, (d) 3-methylheptacosane, and (e) 3,9-dimethylheptacosane. Hexacosane was added for quantitation as an internal standard (IS).

invicta workers as a function of developmental changes, e.g., temporal social castes, colony of origin, and time Here we explore the use of a variety of multivariate analysis methods to separate three temporal (age dependent) social castes (foragers, reserves, and broods) and to visualize colony cuticular hydrocarbon changes that occur over time. The dynamic nature of these heritable characters in relationship to nestmate recognition confirms previous studies (3), and is briefly discussed. The focus of this report is on the data analysis techniques used to identify the fingerprint patterns in the GC data characteristic of temporal social caste and colony of origin for *S. invicta*.

2. Experimental

For this study, 125 individuals and 235 pooled ant samples were obtained from laboratory colonies maintained at the USDA-ARS Fire Ant Project Laboratory in Gainesville, FL. Ants from each colony were fed with sugar–water (1:1) and crickets. Three temporal worker categories were represented in the data: foragers, reserves, and brood tenders. Brood tenders were identified by disturbing a colony and observing the workers that were carrying broods. When disturbing a colony, workers that remained within the colony (mound) structure and were not near the broods were reserves. The very first workers that came to food baits setup near the mound were the foragers. Brood tenders, reserves, and foragers are part of a temporal caste system. As the workers age, they transition from brood tenders to reserves and finally to foragers. Further details about age polymorphism in *S. invicta* can be found elsewhere [11].

Cuticular hydrocarbons were obtained by soaking individual or pooled ant samples for at least 10 min at room temperature in enough hexane (with n-$C_{26}H_{56}$ added for quantitation as an internal standard) to just cover them. After the rinses were complete, the soaks were processed using an Agilent 6890N Network Gas Chromatograph System (Palo, Alto, CA). The Agilent System was equipped with a split–splitless injector, a flame ionization detector, and a DB-1 fused silica capillary column (30 m, 0.25 mm id, 0.25 μm film thickness, J&W Scientific Inc., Folsom, CA). The injector and detector were set at 300 °C, and the oven temperature was programmed from 150 °C to 285 °C at 10°/min and then held at 285 °C for 4 min. Hydrogen was used as the carrier gas and nitrogen was used as the makeup gas. The chromatographic data (see Fig. 1) were processed using Agilent Technologies GC Chemstation G2071AA A.10.01 (Agilent Technologies, Palo Alto). Peak retention times were compared to standard cuticular hydrocarbons from *S. invicta*. If there was ambiguity in a peak assignment, then mass spectra were obtained on an Agilent 5973 Network Mass Selective Detector US10480853 using Agilent 6890N Network Gas Chromatography System US10124023. For the GC/MS runs, the injector was set at 300 °C and the oven temperature was programmed from 100 °C to 285 °C at 10°/min, and then held at 285 °C for 10 min with the transfer line set at 285 °C. Helium was used as the carrier gas for the column. GC/MS data were processed using Agilent Enhanced GC/MS Chemstation software G1701DA version D.00.00.38.

2.1. Pattern recognition methodology

For pattern recognition analysis, each gas chromatogram was translated into a data vector $X=(x_1, x_2, x_3, x_4, x_5)$ using as descriptors the mass (expressed in nanograms) of each major hydrocarbon component as determined by the internal standard ($C_{26}H_{54}$). Each gas chromatogram was normalized to the weight of the corresponding ant sample. Expressing each peak in nanograms and normalizing the GC data to weight of the ant sample provided more information about colony of origin, time period, and social caste than using the area of each GC peak as a descriptor because inclusion of the internal standard facilitated removal of information about sample size from the data.

The GC data were analyzed using the Advanced Data Analysis and Pattern Recognition Toolkit (ADAPT) which was written in MATLAB 7.6.0.324(R2008a) using the graphical user interface development environment (GUIDE). The toolkit consists of a collection of MATLAB M-files and MATLAB Figure files that control the GUI's computational and graphical components. The M-file provides both a code to initialize the GUI and a framework for the GUI routines that execute in response to user-generated events. The main GUI module is a MATLAB M-file called ADAPTv5.M. Invoking this file opens up the main GUI which has menus to perform different tasks and graphical objects like fields and buttons to display information related to a particular dataset and the results of various pattern recognition analysis performed on the data set.

The GUI area is divided into two fields: one for training and the other for prediction. All the training set information is displayed in the training field whereas the prediction set information is displayed in the prediction field. The information displayed in each field includes the name of the dataset file loaded, number of features in the data set, type of data preprocessing done, and the pattern recognition analysis method used. Each field also displays the number of samples and the number of misclassified samples in each class. The sample id number and true class membership of each misclassified sample is also displayed as well as the fitted or predicted class membership value.

The four main types of pattern recognition methods are mapping and display, discriminant development, clustering, and modeling. ADAPT has routines in all four areas and most were used in this study. ADAPT has routines to perform principal component analysis, canonical varaiate analysis, hierarchical clustering, FCV clustering, variance and Fisher weights (for feature selection in classification), linear discriminant analysis, quadratic discriminant analysis, regularized discriminant analysis, K-NN, and back propagation neural networks with one or two hidden layers using the Levenberg Marquardt algorithm or adaptive learning with momentum. Bootstrapped and cross-validated error rates can be computed for each trained model. A description of the various pattern recognition routines that comprise ADAPT can be found elsewhere [12–14].

3. Results and discussion

Several questions have been addressed in this study using a variety of pattern recognition methods. First, is there an advantage to analyze the cuticular hydrocarbon extracts of pooled versus

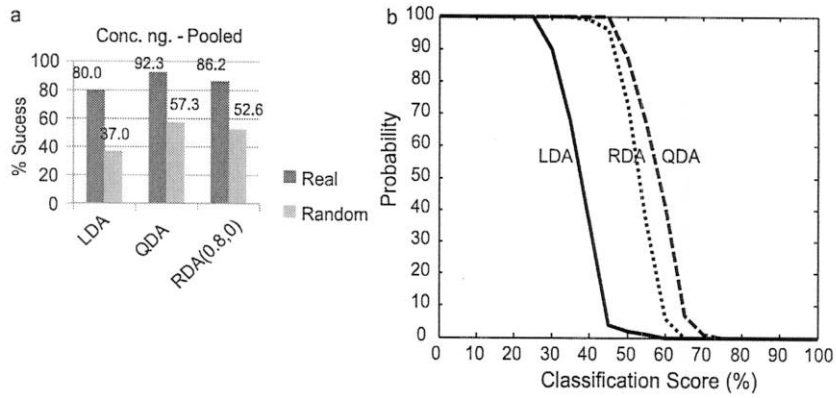

Fig. 2. (a) Comparison of the classification scores for the pooled ant samples versus the average degree of separation in the data due to chance. (b) Probability of achieving any degree of separability due to chance for the pooled ant samples with RDA (0.8, 0), LDA, and QDA.

individual ant samples? Second, is the cuticular hydrocarbon patterns of *S. invicta* workers correlated with their age-linked temporal caste? Third, do the hydrocarbon profiles of *S. invicta* significantly differ for each laboratory colony? And fourth, can the same methods used to distinguish colony of origin be used to track colony cuticular hydrocarbon changes over time? Previous studies [15–20,9] performed in our laboratories on differences in cuticular hydrocarbon profiles for carpenter ants and for *Cataglyphis niger* have shown separation by colony, and temporal caste. For *S. invicta*, it has been previously reported, albeit in a preliminary study, that cuticular hydrocarbon patterns are consistent within colonies for a given sampling time, but they vary sufficiently from colony to colony. The cuticular hydrocarbon profiles of *S. invicta* colonies also change over time [21–23]. In these studies, the results were reported on a subset of the data collected and/or the multivariate methods used were limited in their ability to extract information from the cuticular hydrocarbon profiles. Furthermore, there was no attempt to deconvolve the confounding effects of the biological variables investigated. For these reasons, a more exhaustive investigation of the biological variables that influence the cuticular hydrocarbon profiles of *S. invicta* was undertaken.

During the course of this study, the cuticular hydrocarbon data were investigated for curvilinear relationships among the five measurement variables through analysis of pair wise plots of these measurement variables which revealed only linear relationships. This would suggest that a linear pattern recognition methodology would be sufficient for analysis of this data. Nevertheless, self-organizing maps and back propagation neural networks were applied to the data. However, classification results obtained with so-called nonlinear methods were inferior to those obtained using linear methods and linear models, e.g., linear and quadratic discriminant analysis, canonical variates, and principal component analysis.

3.1. Pooled versus individual ant samples

To answer the first question, gas chromatograms of cuticular hydrocarbon extracts obtained from 65 pooled, reserve ant samples from five laboratory colonies were collected and analyzed using linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and regularized discriminant analysis (RDA). The goal was to separate one colony from another. The results of this study are summarized in Fig. 2a.

Because there was no validation set, Monte Carlo simulation studies were performed to assess the statistical significance of the classification scores. The goal was to estimate the separation in the data due to chance using LDA, QDA, and RDA. For these studies, data sets comprised of random numbers were generated. Both Gaussian and uniform distributions were employed. A method described in previous publications [24,25] was used to compute the expected level of chance classification for both the pooled and individual ant samples. For each chance classification study, 100 data sets consisting of random numbers were generated. The statistical properties of the simulated data (i.e., dimensionality, number of samples, class membership distribution, and covariance structure) were identical to the actual data for which we wish to determine

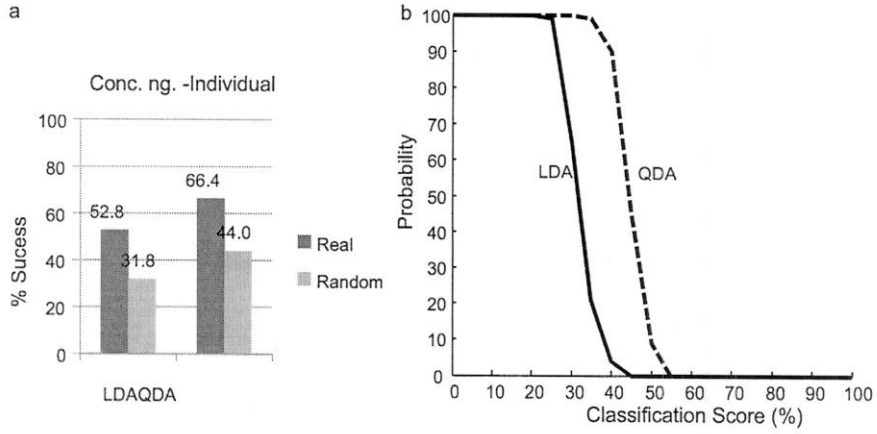

Fig. 3. (a) Comparison of the classification scores for the individual ant samples versus the average degree of separation in the data due to chance. (b) Probability of achieving any degree of separability due to chance for the pooled ant samples with LDA, and QDA.

its degree of classification due to chance. For each random data set, its degree of separability was assessed. The number of occurrences of several degrees of separation (e.g., at least 70% of the patterns were correctly classified or at least 80% of the patterns were correctly classified) was noted and the fraction of the total number of occurrences (cumulative probability) for each degree of separation was plotted against the percentage of patterns correctly classified. These cumulative distribution curves provide information about the likelihood that a particular classification result is due to chance. For example, if the classification score obtained for real data is 80% but the mean classification success rate for the simulated data is only 37% and the probability of achieving 65% correct classification due to chance is zero (see Fig. 2b), the score obtained using the real data (80%, see Fig. 2a) would be considered statistically significant.

Results from these Monte Carlo simulation studies are summarized in Fig. 2a and b. For the QDA classification study involving the pooled ant samples (see Fig. 2b), 100 data sets consisting of random numbers were generated. The statistical properties of the simulated data (i.e., dimensionality, number of samples, class membership distribution, and covariance structure) were identical to those of the 65 pooled ant samples. The separabililty of each random data set was assessed using QDA and a cumulative probability plot was generated for the random data. The mean classification score of the 100 random data sets was also computed and compared to the classification score obtained in the QDA study for the GC data. Since the mean classification success rate of the simulated data was only 57.3%, the classification score obtained for QDA using GC data expressed in nanograms was judged to be statistically significant (see Fig. 2b).

Fig. 3a summarizes the results obtained for 125 individual ant samples collected from the same laboratory colonies as the pooled ant samples. Each colony is represented by 25 reserve workers. Fig. 3a and b summarizes the results of the chance classification studies for this data. Results for RDA were not reported because the values of γ and λ that gave the best classification for colony were 0, 0 which corresponds to QDA.

For the pooled ant samples, there were 65 independent samples equally distributed among five classes. For the individual ants, there were 125 independent samples distributed equally among five classes. As the number of objects in a data set increases, the degree of separation due to chance will decrease. For this reason, chance classifications are lower for the individual ants than for the pooled ant samples.

An examination of Figs. 2 and 3 reveals that differences between the classification-success rates obtained for real data versus random data are smaller for the individual ant samples. This suggests that pooling the samples enhances the recognition of patterns indicative of colony in the cuticular hydrocarbon profiles of *S. invicta* when pattern recognition techniques are used to analyze the data. Evidently, the contribution of the ant's individual pattern to the overall hydrocarbon profile pattern obscures information about colony of origin in GC traces obtained from cuticular hydrocarbon extracts. For these reasons, we strongly suggest that cuticular hydrocarbon profiles from pooled ant samples, not individual ant samples be studied to seek meaningful relationships between cuticular hydrocarbon profiles and biological variables such as colony of origin and temporal (social) caste.

3.2. Social caste

To address the question about patterns in the hydrocarbon profiles indicative of temporal caste, it was necessary to collect additional data. A set of 170 gas chromatograms of cuticular hydrocarbon extracts were obtained from 170 *S. invicta* samples. Each ant sample contains hydrocarbons extracted with hexane from the cuticle of 100 individual ants. The ant samples were obtained

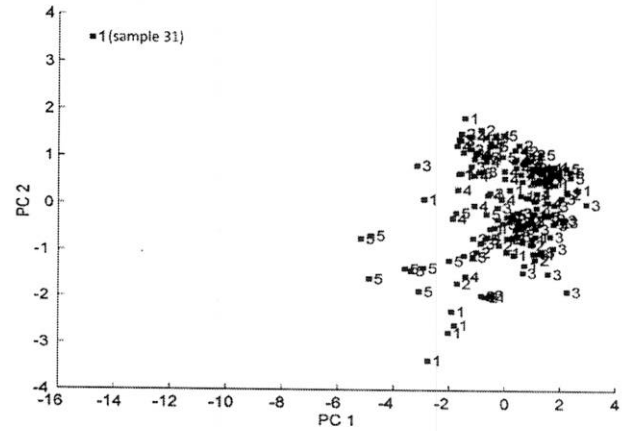

Fig. 4. A plot of the two largest principal components of the 170 pooled red fire ant samples and the five high molecular weight hydrocarbon compounds that characterize the cuticle of *S. invicta*. Each ant sample is represented as a point in the principal component map of the data. 1 is a pooled ant sample from colony 1; 2 is a pooled ant sampled from colony 2; 3 is a pooled ant sample from colony 3; 4 is a pooled ant sample from colony 4; 5 is a pooled ant sample from colony 5.

from five laboratory colonies (which were not the same laboratory colonies used in the pooled versus individual ant sample study), three temporal castes (foragers, reserves, and brood tenders), and the colonies were sampled at four different time periods (three in the spring and summer, and one in the winter).

The first step was to analyze the data using principal component analysis (PCA). This technique can be summarized as a method for transforming the original measurement variables into new, uncorrelated variables called principal components. Each principal component is a linear combination of the original measurement variables. Using this procedure is analogous to finding a set of orthogonal axes that represent the directions of largest variance in the data. PCA can furnish information about trends present in a data set.

Fig. 4 is a plot of the two largest principal components of the 170 pooled *S. invicta* samples and the five GC peaks that characterize each sample. Each pooled ant sample is represented as a point in the principal component map of the data. It is evident from the plot that sample 31 (colony 1) is an outlier, and this sample was subsequently deleted from the analysis because of the adverse effect that outliers can have on the performance of pattern recognition methods.

For each laboratory colony, the data were divided into three categories according to temporal caste. Previous analyses of the cuticular hydrocarbons [22,23] using PCA to analyze the GC profiles of the hydrocarbon soaks revealed patterns indicative of the temporal caste of the *S. invicta* samples in only one of the five laboratory colonies investigated. Therefore, canonical variate analysis (CVA) was performed to separate the pooled ant samples in each colony by temporal caste. The results of this study are summarized in Figs. 5–9. Each pooled ant sample is represented as a point in the CVA map of the data. Foragers, which are represented by the symbol 1, could be readily differentiated from brood tenders (represented by the symbol 2) and reserves (represented by the symbol 3) in four of the five laboratory colonies (colonies 1, 2, 4, and 5) investigated. Because reserves can assume the role of brood tenders, it is plausible that both reserves and the brood tenders could have similar hydrocarbon profiles.

Fig. 10 shows a CVA plot of the GC data from colonies 1, 2, 4, and 5. The data were divided into three classes according to social caste. Again, separation of the foragers from the reserves and brood tenders is evident. When social caste is investigated on a per colony basis, separation of the foragers from the reserves and the brood

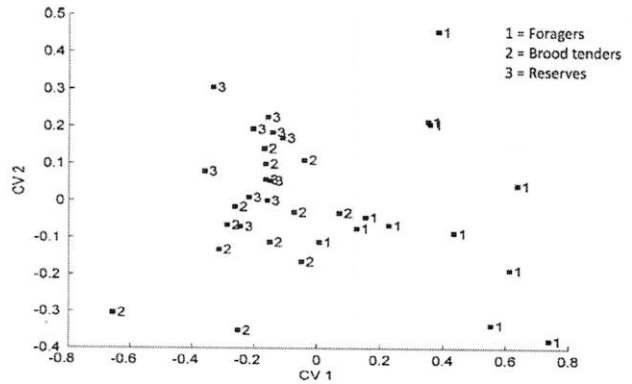

Fig. 5. A plot of the two largest canonical variates of the pooled ant samples obtained from colony 1. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled forager ant sample; 2 is a pooled reserve ant sample; and 3 is a pooled brood tender ant sample. Separation of the foragers from brood tenders and reserves in the plot is evident.

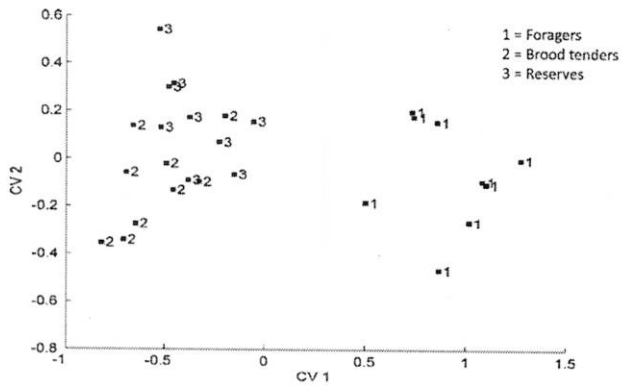

Fig. 6. A plot of the two largest canonical variates of the pooled ant samples obtained from colony 2. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled forager ant sample; 2 is a pooled brood tender ant sample; and 3 is a pooled reserve ant sample. Separation of the foragers from brood tenders and reserves in the plot is evident.

tenders occurred on the first canonical variate. Upon investigating social caste as the class variable using GC data from colonies 1, 2, 4, and 5, separation of the foragers from the reserves and brood tenders occurred on the second canonical variate. These results (see Figs. 5–9 versus Fig. 10) confirm that patterns correlated to temporal caste are present in the cuticular hydrocarbon profiles of *S. invicta*, but are not the major source of variation in the hydrocarbon profiles obtained from pooled ant samples.

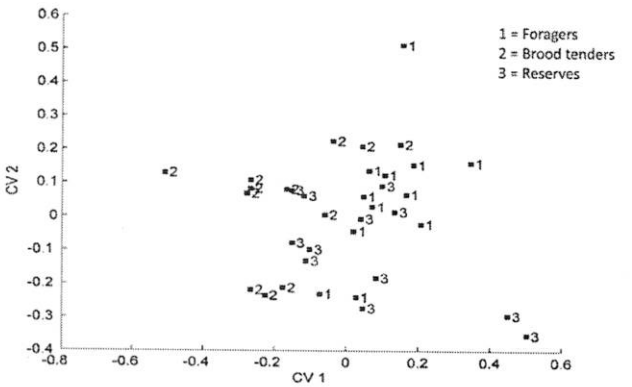

Fig. 7. A plot of the two largest canonical variates of the pooled ant samples obtained from colony 3. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled forager ant sample; 2 is a pooled brood tender ant sample; and 3 is a pooled reserve ant sample. Clustering of the pooled ant samples on the basis of social caste is not observed in this plot.

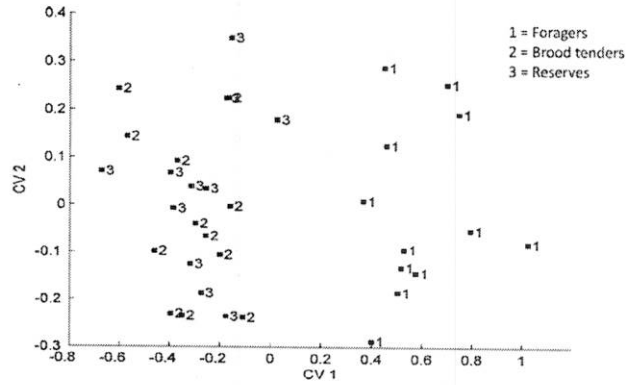

Fig. 8. A plot of the two largest canonical variates of the pooled ant samples obtained from colony 4. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled forager ant sample; 2 is a pooled brood tender ant sample; and 3 is a pooled reserve ant sample. Separation of the foragers from brood tenders and reserves in the plot is evident.

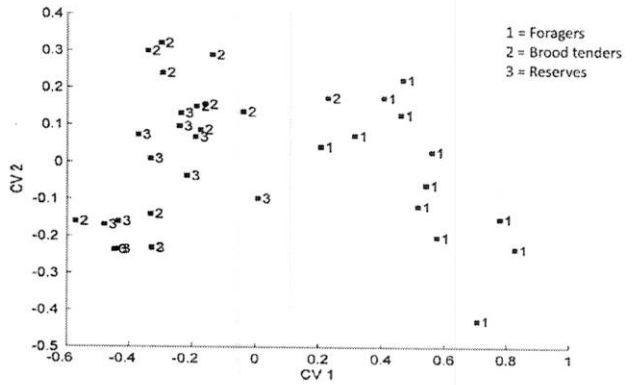

Fig. 9. A plot of the two largest canonical variates of the pooled ant samples obtained from colony 5. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled forager ant sample; 2 is a pooled brood tender ant sample; and 3 is a pooled reserve ant sample. Separation of the foragers from brood tenders and reserves in the plot is evident.

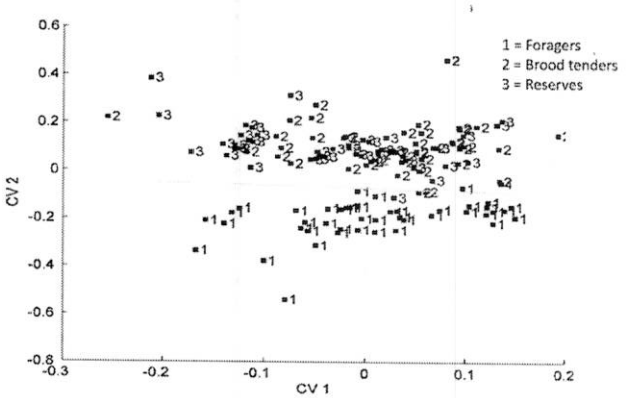

Fig. 10. A plot of the two largest canonical variates of the pooled ant samples obtained from all five colonies. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled forager ant sample; 2 is a pooled brood tender ant sample; and 3 is a pooled reserve ant sample. Separation of the foragers from the brood tenders and reserves in the plot is evident.

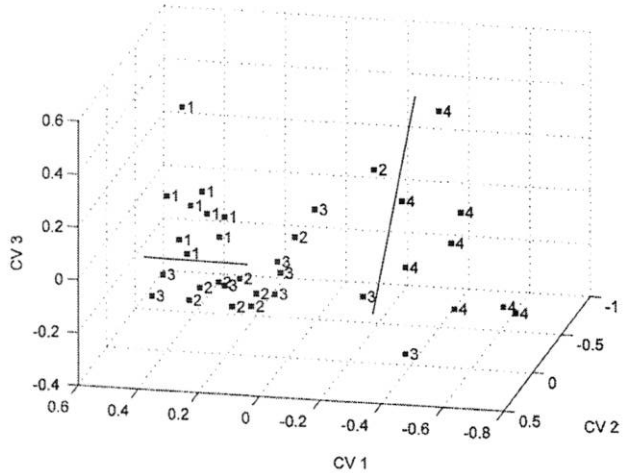

Fig. 11. A plot of the three largest canonical variates of the pooled ant samples obtained from colony 1. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled ant sample from time period 1; 2 is a pooled ant sample from time period 2; 3 is a pooled ant sample from time period 3; and 4 is a pooled ant sample from time period 4. Clustering of the pooled ant samples by time period is evident in this plot.

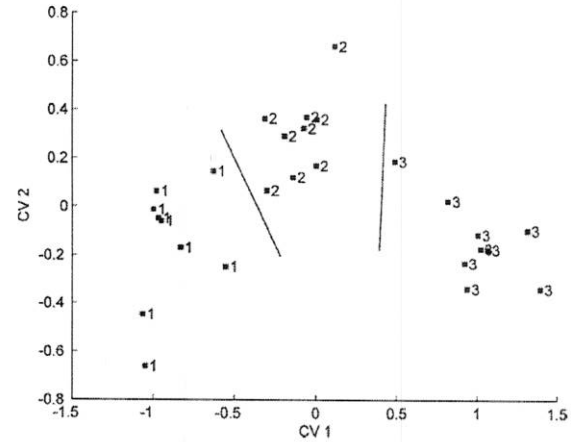

Fig. 13. A plot of the three largest canonical variates of the pooled ant samples obtained from colony 2. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled ant sample from time period 1; 2 is a pooled ant sample from time period 2; 3 is a pooled ant sample from time period 3; and 4 is a pooled ant sample from time period 4. Clustering of the pooled ant samples by time period is evident in this plot.

3.3. Time period

For each colony, the data were divided into four categories according to the time period of sampling. CVA was performed to separate pooled ant samples in each colony by time period. Monte Carlo simulation experiments were also performed in tandem to assess the degree of separation in the data due to chance. One hundred data sets comprised of random numbers with Gaussian distributions that had statistical properties (i.e. dimensionality, number of samples, class membership distribution, and covariance structure) identical to those of the real data were generated. CVA was performed on a data set that was an average of the 100 random data sets generated. The results are summarized in Figs. 11–20. It is evident from the Monte Carlo simulation experiments that separation of the pooled ant samples by time period in the CVA plots cannot be attributed to chance.

Each laboratory colony exhibited a different pattern of change with time. In our previous studies [22,23], we were able to determine that *S. inivcta* cuticular hydrocarbon profiles from time period four were different from the cuticular hydrocarbon profiles of the other time periods and that only one laboratory colony exhibited a systematic change in its their cuticular hydrocarbon profile over time. The results obtained in the present study indicate that cuticular hydrocarbon profiles of each *S. invicta* colony change with time. However, the pattern of change as shown in each CVA plot is different for each colony. In some instances, all of the time periods are well separated whereas in other instances only two of the four time periods are well separated. This should not come as a surprise for the cuticular hydrocarbon profiles of *S. invicta* may be a dynamic system that undergoes changes with time and the nature of this change will be different for each colony.

3.4. Laboratory colony

The cuticular hydrocarbon profiles of *S. invicta* were also found to be characteristic of the individual colony. For each time period, the data were divided into five categories according to the colony of origin of the pooled ant samples. Again, decision surfaces were developed from the five major hydrocarbon components. QDA was used to classify the data by colony for each time period. Monte

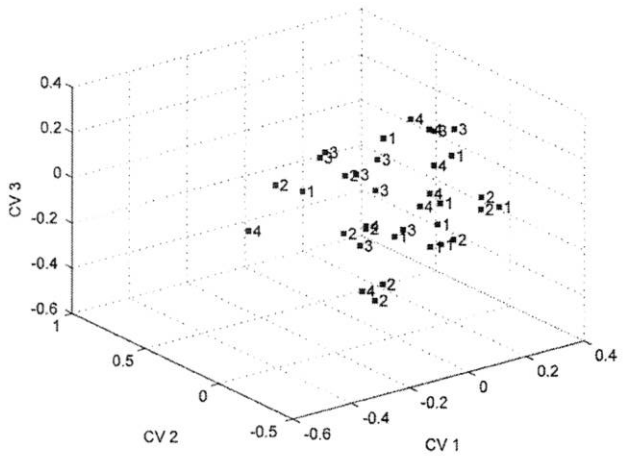

Fig. 12. A plot of the three largest canonical variates of the simulated data sets for colony 1. Clustering of the pooled ant samples by time period is not evident in this plot.

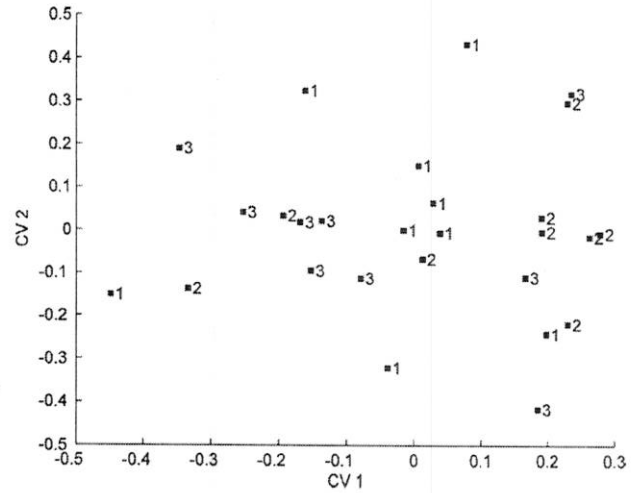

Fig. 14. A plot of the three largest canonical variates of the simulated data sets for colony 2. Clustering of the pooled ant samples by time period is not evident in this plot.

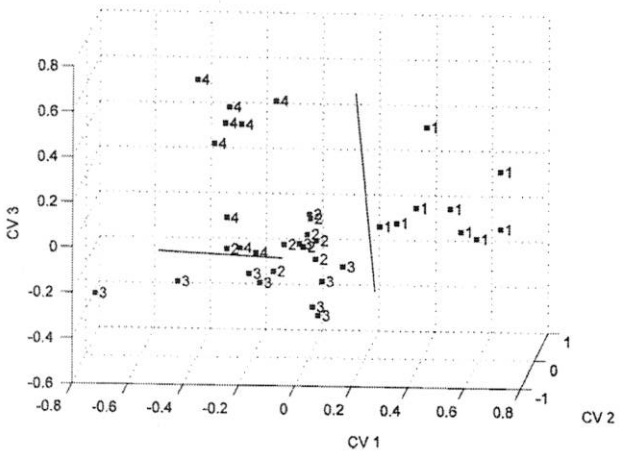

Fig. 15. A plot of the three largest canonical variates of the pooled ant samples obtained from colony 3. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled ant sample from time period 1; 2 is a pooled ant sample from time period 2; 3 is a pooled ant sample from time period 3; and 4 is a pooled ant sample from time period 4. Clustering of the pooled ant samples by time period is evident in this plot.

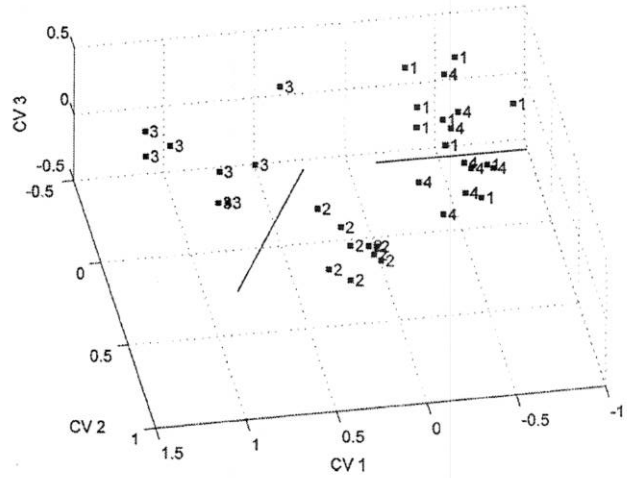

Fig. 17. A plot of the three largest canonical variates of the pooled ant samples obtained from colony 4. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled ant sample from time period 1; 2 is a pooled ant sample from time period 2; 3 is a pooled ant sample from time period 3; and 4 is a pooled ant sample from time period 4. Clustering of the pooled ant samples by time period is evident in this plot.

Carlo simulation studies were also performed to assess the degree of separation in the data due to chance. The results of these studies are summarized in Fig. 21. Clearly, the cuticular hydrocarbon profiles of the red fire ants are characteristic of the colony of origin for a given time period. However, it was surprising that our Monte Carlo simulations revealed high chance classification success rates for the case of 45 samples distributed equally among four classes with each sample characterized by five measurements using QDA. Chance classification may be a more serious problem with QDA than was previously thought.

QDA was also used to classify the data by colony across all time periods. The results of this study are summarized in Fig. 22. To assess the significance of these classifications, Monte Carlo simulation studies were performed. The results of these studies are summarized in Figs. 22 and 23. Differences in chance classification across all time periods versus individual time periods were due to the larger number of samples involved in colony classification across all time periods. Using the Monte Carlo simulation studies as a benchmark, it is evident that classifications obtained in the quadratic discriminant analysis study across all time periods are significant for four of the five laboratory colonies. When the classifications for colony from each time period (see Fig. 21) are compared to the classifications for colony across all time periods (see Fig. 22), it is evident that cuticular hydrocarbon profiles of *S. invicta* change with time, which can confound the classification of GC profile data by colony using pattern recognition techniques. This is most evident in the cuticular hydrocarbon profiles of pooled samples of *S. invicta* from colonies 4 and 5. The changes in the cuticular hydrocarbon profiles that occurred in laboratory colonies 4 and 5 over time caused their cuticular hydrocarbon profiles to overlap. For example, cuticular hydrocarbon profiles from colony 5 at time period 1 were similar to those of colony 4 at time period 3.

We have previously reported [22,23] that four of five laboratory colonies could be differentiated on the basis of their cuticular hydrocarbon profiles. These studies were carried out by formulating the problem as a series of binary classifications using the linear learning machine and related linear nonparametric methods of classification. In the current study, better multivariate analysis methods have been used and the analysis of the cuticular hydrocarbon data was more detailed in its scope. From the current study, we have learned that all five colonies could be separated on

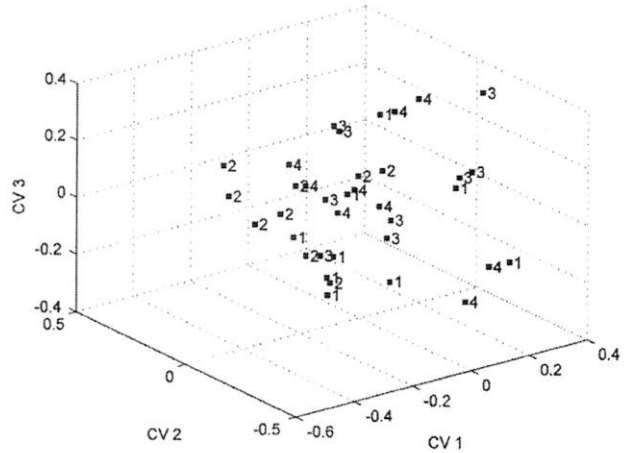

Fig. 16. A plot of the three largest canonical variates of the simulated data sets for colony 3. Clustering of the pooled ant samples by time period is not evident in this plot.

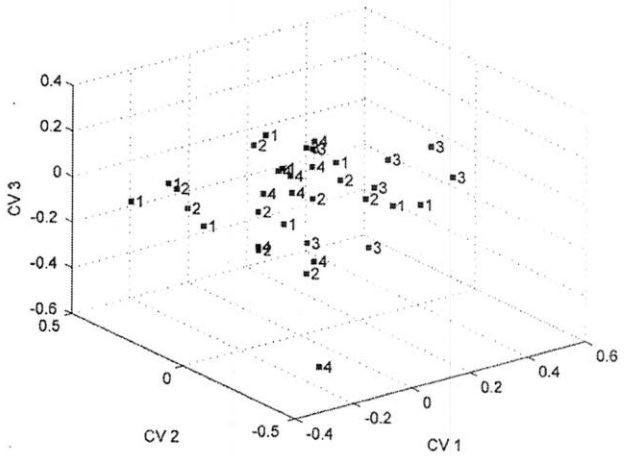

Fig. 18. A plot of the three largest canonical variates of the simulated data sets for colony 4. Clustering of the pooled ant samples by time period is not evident in this plot.

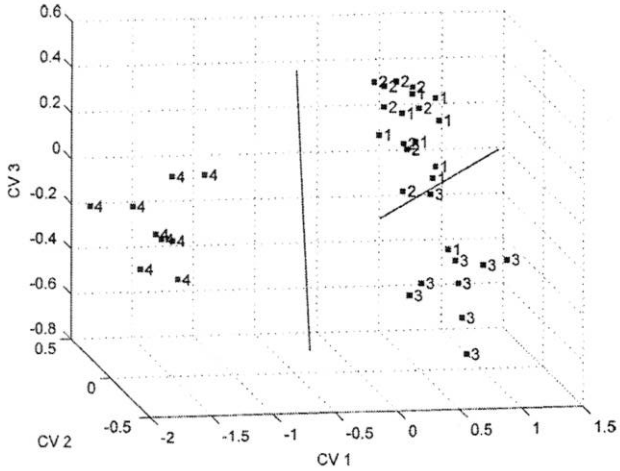

Fig. 19. A plot of the three largest canonical variates of the pooled ant samples obtained from colony 5. Each pooled ant sample is represented as a point in the CVA map of the data. 1 is a pooled ant sample from time period 1; 2 is a pooled ant sample from time period 2; 3 is a pooled ant sample from time period 3; and 4 is a pooled ant sample from time period 4. Clustering of the pooled ant samples by time period is evident in this plot.

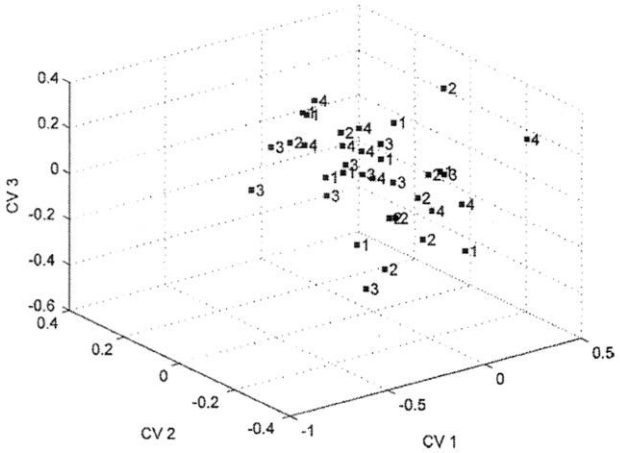

Fig. 20. A plot of the three largest canonical variates of the simulated data sets for colony 5. Clustering of the pooled ant samples by time period is not evident in this plot.

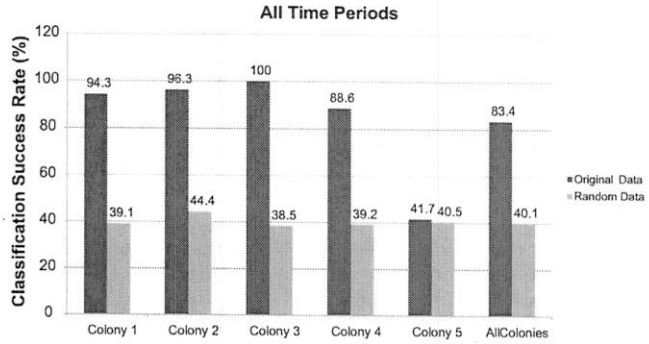

Fig. 22. A comparison of the classification scores for colony across all time periods versus the degree of separation in the data due to chance.

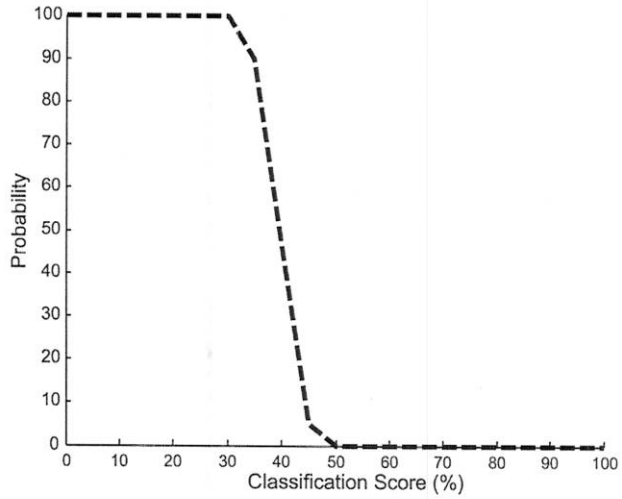

Fig. 23. Probability of achieving any degree of separation in the data due to chance for all five laboratory colonies using QDA. There is a 50% probability of achieving a classification score of 40.1%.

the basis of their cuticular hydrocarbon profiles when data from each time period is analyzed separately. When colonies are analyzed using data from all of the time periods, the classifications become confounded which considerably strengthens the previ-

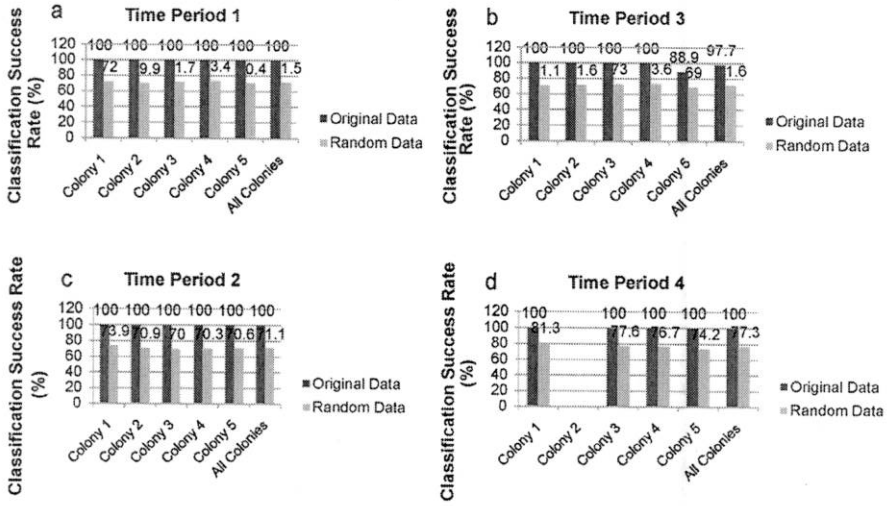

Fig. 21. A comparison of the classification scores for colony versus the degree of separation in the data due to chance at (a) time period 1, (b) time period 2, (c) time period 3, and (d) time period 4.

ously stated conclusion that cuticular hydrocarbon profiles of red fire ants change over time.

4. Conclusions

The GC traces representing ant cuticle extracts can be related to colony of origin and temporal caste. These results support a correlative role for cuticular hydrocarbons in nestmate recognition. However, it remains for specific behavioral bioassays with purified hydrocarbons to determine if cuticular hydrocarbons are in fact used by *S. invicta* in nestmate recognition. In addition, the re-analysis of temporal caste and time on cuticular hydrocarbon patterns demonstrates that sampling time and social caste must be taken into account to avoid unnecessary variability and possible confounding. This and the fact that foragers could not be separated from reserves and brood-tenders in all five laboratory colonies suggests that cuticular hydrocarbons as a class of compounds cannot model every facet of nestmate recognition in *S. invicta* which in turn suggests a potential role for other compounds in the discrimination of alien conspecifics from nestmates.

It is truly remarkable that all of this information (social caste, colony of origin, and time period) is contained in the concentration pattern of five high molecular weight hydrocarbons which comprise a dynamic system that changes with time with the nature of these changes being different for each colony. Neither colony of origin, social caste, nor time period is the major source of variation in the data although distinct patterns in the concentration profiles of the five hydrocarbons characteristic of these biological variables can be identified.

This study also demonstrates the importance of using pattern recognition methods to analyze complex chromatographic data sets and to seek meaningful relations between chemical constitution and biological variables. The classification of complex biological samples on the basis of their GC profiles can be complicated by two factors: (1) confounding of the desired group information by other systematic variations present in the data and (2) random or chance classification effects. The existence of these complicating relationships is an inherent part of fingerprint type data.

Many of the pattern recognition methods used in this study relied heavily on graphics for the presentation of results. It is the opinion of the authors that multivariate analysis techniques should be used to extend the ability of human pattern recognition. This allows the user to directly interpret the meaning of the underlying relationships present in data.

Acknowledgements

The authors acknowledge Douglas Henry and William T. Rayens for helpful discussions. This work was partially supported by the OSU Research Foundation through start up funds provided to BKL for purchase of a computer cluster. NM acknowledges the financial support of the OSU Chemistry Department via a teaching assistantship.

References

[1] K.H. Lockey, Comp. Biochem. Physiol. B 89 (1980) 595–645.
[2] R.W. Howard, G.J. Blomquist, Ann. Rev. Entomol. 50 (2005) 371–393.
[3] R.K. Vander Meer, L. Morel, Nestmate recognition in ants, in: R.K. Vander Meer, M. Breed, M. Winston, K.E. Espelie (Eds.), Pheromone Communication in Social Insects, Westview Press, Boulder, CO, 1998.
[4] M.D. Breed, Chemical cues in kin recognition; criteria for identification, experimental approaches, and the honey bee as an example, in: Pheromone Communication in Social Insects, Westview Press, Boulder, CO, 1998.
[5] M. Breed, B. Bennet, Kin recognition in highly exotic insects, in: D.J.C. Fletcher, C.D. Michener (Eds.), Kin Recognition in Animals, Wiley, New York, 1986, pp. 243–286.
[6] B. Holldobler, C.D. Michener, Mechanisms of identification and discrimination in social hymenoptera, in: H. Markl (Ed.), Evolution of Social Behavior: Hypothesis and Empirical Testing, Verlag Chemie, Weinheim, 1980, pp. 433–439.
[7] E.O. Wilson, The Insect Societies, Harvard University Press, Cambridge, MA, 1971.
[8] K.G. Ross, R.K. Vander Meer, D.J.C. Fletcher, E.L. Vargo, Evolution 41 (1987) 280–293.
[9] L. Morel, R.K. Vander Meer, B.K. Lavine, Behav. Ecol. Sociobiol. 22 (1988) 175–183.
[10] R.K. Vander Meer, D.P. Wojcik, Science 218 (1982) 806–808.
[11] J.T. Mirenda, S.B. Vinson, Anim. Behav. 29 (1981) 410–420.
[12] D.L. Massart, B.G.V.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics: A Textbook, Data Handing in Science and Technology, Elsevier, New York, 1988.
[13] G.L. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, Wiley Interscience, NY, 1992.
[14] K. Varmuza, P. Filzmoser, Introduction to Multivariate Statistical Analysis in Chemometrics, CRC Press, Boca Raton, FL, 2009.
[15] B.K. Lavine, L. Morel, R.K. Vander Meer, R.W. Gunderson, J.H. Han, A. Bonanno, A. Stine, Chemometr. Intell. Lab. Syst. 9 (1990) 107–114.
[16] B.K. Lavine, R.K. Vander Meer, L. Morel, R.W. Gunderson, J.H. Han, A. Stine, Microchem. J. 41 (1990) 288–295.
[17] B.K. Lavine, C. Davidson, R.K. Vander Meer, S. Lahav, V. Soroker, A. Hefetz, J. Chem. 66 (2003) 51–62.
[18] C.A. Johnson, H. Topoff, R.K. Vander Meer, B. Lavine, Anim. Behav. 64 (2002) 807–815.
[19] C.A. Johnson, H. Topoff, R.K. Vander Meer, B.K. Lavine, Behav. Ecol. Sociobiol. 57 (2005) 245–255.
[20] C.A. Johnson, R.K. Vander Meer, B. Lavine, J. Chem. Ecol. 27 (2001) 1787–1804.
[21] R.K. Vander Meer, D. Saliwanchik, B. Lavine, J. Chem. Ecol. 15 (1989) 2115–2125.
[22] B.K. Lavine, P.C. Jurs, D.R. Henry, R.K. Vander Meer, J.A. Pino, J.E. McMurry, Chemometr. Intell. Lab. Syst. 3 (1988) 79–89.
[23] P.C. Jurs, B.K. Lavine, T.R. Stouch, J. Res. Nat. Bur. Stand. 90 (6) (1985) 543–549.
[24] B.K. Lavine, D.R. Henry, P.C. Jurs, J. Chem. 2 (1988) 1–10.
[25] B.K. Lavine, D.R. Henry, J. Chem. 2 (1988) 85–89.