Discriminant Analysis of Volatile Fatty Acids Produced in Culture Medium: a Novel Approach to the Identification of *Pseudomonas* Species

By F. PELADAN, J. C. TURLOT, AND H. MONTEIL

1 Institut de Bactériologie de la Faculté de Médecine, Université Louis Pasteur, 3 rue Koeberlé, 67000 Strasbourg, France

2 Centre d’Etudes Statistiques, IRMA, Université Louis Pasteur, 7 rue Descartes, 67000 Strasbourg, France

(Received 21 May 1984; revised 4 July 1984)

The volatile fatty acids produced in culture medium by 357 *Pseudomonas* strains belonging to eight species were determined quantitatively by GLC. The resultant chromatograms were submitted to discriminant analysis. Stable discriminant functions were computed and included in a computerized identification system which also involved some distinctive volatile fatty acids regarded as two-state qualitative characters (presence or absence characters). Using a test group of 249 strains belonging to the studied species, more than 89% of the identifications made by this system agreed with those made by conventional biochemical methods despite the relatively poor differentiation between *P. putida* and *P. fluorescens*. When the individual species within the matrices were weighted with prior probabilities reflecting results given by two simple biochemical tests, 96% of the 249 strains were correctly identified.

**INTRODUCTION**

The determination by GLC of volatile fatty acids (VFA) produced in culture medium is a classic method used for the identification of anaerobic bacteria (Deacon et al., 1978; Holdeman et al., 1977); however, the great number of species (more than 190) and the semi-quantitative interpretation of GLC data have considerably limited the discriminant power of this methodology. Nevertheless, we have recently shown that an accurate quantitative determination of eight VFA, containing two to six carbon atoms, was sufficient to achieve a good characterization of most of the *Pseudomonas* species of medical interest (Peladan & Monteil, 1984). Use of these results for the identification of unknown isolates requires statistical methods to ensure an objective analysis of chromatograms. Several such methods have been proposed (Drucker, 1974; Friedmann & MacLowry, 1973; MacFie et al., 1978; O’Donnell et al., 1980). These approaches lead, in general, to selection from among all the variables, of those which allow a good linear discrimination between species. However, beyond this descriptive aspect, one of these methods, called discriminant analysis, has a predictive aspect and discriminant functions achieved by the analysis of a representative sample of the studied species (base sample) could allow an accurate identification of unknown isolates (Darland, 1975; Gutteridge & Puckey, 1982).

This paper describes the application of discriminant analysis to GLC results from strains of eight *Pseudomonas* species. Furthermore an attempt has been made to establish an identification system and match unknown isolates of *Pseudomonas* against it.

**METHODS**

**Strains.** A total of 606 strains were used, belonging to eight species: *P. aeruginosa* (91 strains); *P. putida* (93 strains); *P. fluorescens* (80 strains); *P. cepacia* (90 strains); *P. acidovorans* (104 strains); *P. maltophilia* (88 strains); *P. stutzeri* (16 strains); *P. paucimobilis* (44 strains). Of these 606 strains, 597 were isolated from natural and hospital

**Abbreviation:** VFA, volatile fatty acids.
environments and were identified independently in two laboratories (Dr C. Richard, Institut Pasteur, Paris, France) by classical biochemical methods (Gilardi, 1978; Hugh, 1978; Lennette et al., 1980). Nine reference strains from the American Type Culture Collection (ATCC) were included: P. aeruginosa ATCC 10145; P. putida ATCC 12633; P. fluorescens ATCC 13525; P. cepacia ATCC 17759, 25609, 25414; P. acidovorans ATCC 15688; P. maltophilia ATCC 13637 and P. stutzeri ATCC 17588. The strains were randomly divided in two groups: group A (357 strains) was used to perform discriminant analysis (base group) and group B (249 strains) was used for the validation of the identification system (test group).

Culture and VFA extraction. Each organism was checked for purity, maintained on Mueller–Hinton plates and grown at 30 °C with shaking in 20 ml of a standard medium for 48 h (Peladan & Monteil, 1984a). The medium was that proposed by Veron (1975) but modified by the addition of (per litre): 100 g Casamino acids (Difco), 1.25 g sodium acetate and 1 g alcohols (0.2 g l−1 of each of the following: n-propyl, n-butyl, isobutyl, n-pentyl and iso- pentyl alcohol) and the omission of ammonium sulphate and cobalt nitrate. The VFA present in 10 ml of the supernatant of the centrifuged culture (15 min, 11000 g, 4 °C) were extracted with 10 ml diethyl ether after acidification of the medium with 0.5 ml H2SO4. A second extraction was then performed using 0.5 ml diethyl ether after dissolving the VFA in 1 ml 1 M-NaOH. The final volume was adjusted to 500 µl.

Chromatography. The VFA extracts were analysed with a Perkin-Elmer Sigma 3B gas chromatograph fitted with flame ionization detectors. One µl of extract was injected onto a 2 m × 1.75 mm internal diameter glass column packed with Chromosorb W, HP (AW, DMSC, 80–100 mesh) coated with 8% (w/w) SP-1000 (Supelco, Bellefonte, USA) and 1% (w/w) H3PO4. The instrument was run isothermally for 5 min at 150 °C. Injector and detector temperatures were 210 °C and 235 °C respectively. Nitrogen was used as carrier gas at a flow rate of 50 ml min−1 and hydrogen and air were supplied at 135 kPa and 175 kPa respectively. The resultant peaks were recorded and integrated using a Perkin-Elmer Sigma 10B integrator and were identified by comparing retention times of test samples with those obtained from a series of chromatographically pure standards.

Data collection. Samples of uninoculated sterile culture medium were included as controls in every set of analyses. Peak areas from such controls were subtracted from the respective peak areas for the test samples. The resulting values were normalized using a standardized method whereby each peak area was divided by the sum of the areas of the selected peaks and the resultant quotient multiplied by 100, so as to express each acid produced as a percentage of the total acid output of the strain under examination. All strains in group A were cultivated twice and each culture analysed twice. The average concentration of each acid determined from these four repeated until either all the variables are included or introduction of any of the remaining variables does not increase the Wilks Lambda criterion (Jennrich & Sampson, 1981). Discriminant functions, which are linear combinations of the original variables, could be calculated for a strain x with respect to each k species:

\[
 f_k(x) = S_k(a_k' \cdot X + c_k)
\]

where \( f_k(x) \) is the discriminant score of the strain x relative to species k, \( S_k \) is a prior probability assigned to the species k, \( a_k' \) is the vector transpose of the coefficients from the discriminant function relative to species k, \( c_k \) is a constant and \( X \) represents the observation vector of the strain x. Thus a strain x would be allocated to the species for which it achieves the highest discriminant score. Moreover, an evaluation of the similarity between a strain x and each of the k species is given by the a posteriori probability \( P_k(x) \):

\[
 P_k(x) = e^{f_k(x)} / \sum_{i=1}^{n} e^{f_i(x)}
\]

where \( P_k(x) \) is the a posteriori probability that x belongs to species k, \( f_k(x) \) is the discriminant score of strain x for species k and n represents the total number of species. This probability was used to point out strains which were very distant from all the species present in the study. These strains were considered as outliers or belonging to unknown species.

'Jacknifing' procedure. This technique is used at each step of the discriminant analysis to test the stability of the discriminant functions obtained. Each strain is allocated to a species according to the discriminant functions computed from all the data except the case being classified. Comparison of the percentage of the correct allocations before and after 'jacknifing' is an estimate of the stability of the discriminant functions (Lachenbruch & Mickey, 1968).
Identification of Pseudomonas strains by GLC

Fig. 1. Typical chromatographic profiles given by eight Pseudomonas species and the control (uninoculated medium). VFA: 1, acetic; 2, propionic; 3, isobutyric; 4, n-butyric; 5, isovaleric; 6, n-valeric; 7, isocaproic; 8, n-caproic.

Table 1. Concentrations of VFA produced by 357 strains belonging to eight Pseudomonas species, determined by GLC.

<table>
<thead>
<tr>
<th>Species*</th>
<th>Acetic</th>
<th>Propionic</th>
<th>Isobutyric</th>
<th>Butyric</th>
<th>Isovaleric</th>
<th>Isocaproic</th>
<th>Caproic</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. acidovorans</td>
<td>28.0</td>
<td>5.2</td>
<td>10.0</td>
<td>3.0</td>
<td>19.2</td>
<td>5.0</td>
<td>11.6</td>
</tr>
<tr>
<td>(59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>25.0</td>
<td>5.8</td>
<td>5.6</td>
<td>1.0</td>
<td>16.9</td>
<td>2.0</td>
<td>2.5</td>
</tr>
<tr>
<td>(54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
<td>5.0</td>
</tr>
<tr>
<td>P. cepacia</td>
<td>0</td>
<td>4.4</td>
<td>1.0</td>
<td>14.1</td>
<td>3.0</td>
<td>50.3</td>
<td>8.5</td>
</tr>
<tr>
<td>(57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. fluorescens</td>
<td>0</td>
<td>10.2</td>
<td>5.0</td>
<td>17.2</td>
<td>4.5</td>
<td>14.6</td>
<td>6.0</td>
</tr>
<tr>
<td>(41)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. putida</td>
<td>0</td>
<td>7.7</td>
<td>2.0</td>
<td>28.6</td>
<td>3.5</td>
<td>8.1</td>
<td>2.0</td>
</tr>
<tr>
<td>(57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. maltophilia</td>
<td>0</td>
<td>4.0</td>
<td>1.0</td>
<td>8.0</td>
<td>1.0</td>
<td>10.0</td>
<td>3.0</td>
</tr>
<tr>
<td>(53)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. paucimobilis</td>
<td>0</td>
<td>5.5</td>
<td>1.0</td>
<td>15.0</td>
<td>3.0</td>
<td>22.8</td>
<td>4.5</td>
</tr>
<tr>
<td>(26)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. stutzeri</td>
<td>0</td>
<td>0.7</td>
<td>0.0</td>
<td>18.8</td>
<td>5.6</td>
<td>13.0</td>
<td>5.1</td>
</tr>
<tr>
<td>(10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The numbers in parentheses represent the numbers of individual isolates.
† The results are normalized concentrations (%). M, mean; S², within-group variance.

Canonical variates analysis. The aim of this descriptive method is to obtain the best separation among all the studied species. Thus, for each strain, a set of new co-ordinates expressed as linear combinations of the original variables (canonical variates axes) is determined. These canonical axes are given by the principal components of the species means calculated with the Mahalanobis D² generalized distance (Anderson, 1958; Jennrich & Sampson, 1981).

The canonical variates analysis, stepwise discriminant analyses and 'jacknifing' programs used in this study are available in the BMDP statistical computing package (Biomedical Computer Programs - University of California - Los Angeles) and were run on an UNIVAC 1110 computer at the Centre de Calcul du CNRS, Strasbourg, France.

RESULTS

Chromatographic results

The average concentration of each acid determined from the four chromatograms was considered for each strain of group A. In these analyses, the observed variation in each of the detected peak areas was less than 10%. Although the typical traces obtained for each species showed many similarities (Fig. 1), there were a number of obvious qualitative and quantitative differences between the species (Table 1), particularly for propionic, n-caproic and isocaproic acids. Presence or absence of these acids allowed an immediate identification of P. paucimobilis, P. maltophilia and P. stutzeri strains.
A Canonical axis 2
-3 -5 -7 -5 -3 1 3 5 Canonical axis 1
P. aeruginosa
P. acidovorans
P. putida
P. fluorescens

Fig. 2. Plot of the 357 isolates of group A relative to the first two canonical axes. (The ellipses do not represent theoretical confidence regions but they encompass all the strains belonging to the same species.)

**Canonical variates analysis**

The three species identified by two-state qualitative characters (*P. paucimobilis*, *P. maltophilia*, *P. stutzeri*) were excluded from the canonical variates analysis and the stepwise discriminant analysis, which were performed only with the strains belonging to *P. acidovorans*, *P. aeruginosa*, *P. cepacia*, *P. fluorescens* or *P. putida*. Canonical variates analysis showed that these five species were distributed between two groups: *P. aeruginosa* and *P. acidovorans* on the one hand and *P. cepacia*, *P. fluorescens* and *P. putida* on the other (Fig. 2). The cluster corresponding to the *P. aeruginosa* strains was included in the *P. acidovorans* cluster although both were completely separated on the third canonical axis. Clusters representing *P. putida* and *P. cepacia* were also completely separated but each of these overlapped with *P. fluorescens*.

**Stepwise discriminant analysis**

At each step of the discriminant analysis an estimation of the quality of discrimination between the five remaining species was given by the percentage of strains correctly allocated to their original species by 'jacknifing' (Table 2). Thus the four most reliable discriminant variables among those used (i.e. propionic, n-butyric, isobutyric, isovaleric acid) were chosen. The classification achieved by this analysis was based on a set of five discriminant functions. A global evaluation of the quality of this classification was given by 'jacknifing' of the strains belonging to group A (Table 3), in which more than 83% of those strains were correctly allocated to their original species. Furthermore these results corroborated those achieved by the canonical variates analysis, i.e. *P. putida* and *P. cepacia* were perfectly separated and the overlap between *P. putida* and *P. fluorescens* was more evident (23%) than between *P. fluorescens* and *P. cepacia* (9%).

**Classification of unknown isolates**

The identification system used here involved a set of five discriminant functions and three VFA (propionic, caproic and isocaproic acids) regarded as two-state qualitative characters. The ability of this system to identify unknown isolates from group B, which had previously been characterized to species level on conventional biochemical criteria, was tested. All of the strains belonging to the species *P. paucimobilis*, *P. maltophilia* and *P. stutzeri* were correctly identified from their qualitative characters only. For the other species, 82% of the strains were correctly identified (Table 4) but if *P. putida* and *P. fluorescens* were regarded as a single species, 92% of the identifications agreed with those performed using conventional biochemical methods. The
Table 2. Stepwise selection of acids and estimation of discriminant power at each step by means of 'jacknife' classification

<table>
<thead>
<tr>
<th>Step</th>
<th>Selected acid</th>
<th>Percentage of strains correctly reallocated to their original species</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P. acidovorans</td>
<td>P. aeruginosa</td>
</tr>
<tr>
<td>1</td>
<td>Butyric</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Isovaleric</td>
<td>+ Butyric</td>
</tr>
<tr>
<td>3</td>
<td>Isobutyric</td>
<td>+ Isovaleric</td>
</tr>
<tr>
<td>4</td>
<td>Propionic</td>
<td>+ Isobutyric</td>
</tr>
</tbody>
</table>

Table 3. ‘Jacknife’ classification of strains belonging to group A by means of discriminant functions

<table>
<thead>
<tr>
<th>Original group*</th>
<th>Discriminant group:</th>
<th>P. acidovorans</th>
<th>P. aeruginosa</th>
<th>P. cepacia</th>
<th>P. fluorescens</th>
<th>P. putida</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. acidovorans</td>
<td>85</td>
<td>15</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>5</td>
<td>95</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>P. cepacia</td>
<td>–</td>
<td>–</td>
<td>91</td>
<td>9</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>P. fluorescens</td>
<td>–</td>
<td>–</td>
<td>12</td>
<td>68</td>
<td>20</td>
<td>–</td>
</tr>
<tr>
<td>P. putida</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>23</td>
<td>77</td>
<td>–</td>
</tr>
</tbody>
</table>

Total percentage of strains correctly identified = 83

* Original groups were established on the basis of conventional biochemical tests.

Table 4. Identification of 249 unknown isolates in group B by means of discriminant functions and qualitative characters

All strains of P. paucimobilis, P. stutzeri and P. maltophilia were identified using three qualitative characters: the presence or absence of propionic, isocaproic and caproic acids.

<table>
<thead>
<tr>
<th>Discriminant group:</th>
<th>P. acidovorans</th>
<th>P. aeruginosa</th>
<th>P. cepacia</th>
<th>P. fluorescens</th>
<th>P. putida</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. acidovorans</td>
<td>82</td>
<td>18</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>4</td>
<td>96</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>P. cepacia</td>
<td>–</td>
<td>–</td>
<td>98</td>
<td>12</td>
<td>–</td>
</tr>
<tr>
<td>P. fluorescens</td>
<td>–</td>
<td>–</td>
<td>12</td>
<td>62</td>
<td>25</td>
</tr>
<tr>
<td>P. putida</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>19</td>
<td>81</td>
</tr>
</tbody>
</table>

Total percentage of strains correctly identified = 82

Table 5. Sets of a priori probabilities assigned to the different species on the basis of the results given by two biochemical tests

Each number represents the a priori probability that an unknown isolate would be assigned to the species in column 1, on the basis of results for production of pyoverdin (P) and gelatinase (G): +, production; –, non-production.

<table>
<thead>
<tr>
<th>Species</th>
<th>P+ /G+</th>
<th>P+ /G-</th>
<th>P- /G+</th>
<th>P- /G-</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. acidovorans</td>
<td>0.05</td>
<td>0.05</td>
<td>0.20</td>
<td>0.60</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>0.35</td>
<td>0.25</td>
<td>0.10</td>
<td>0.25</td>
</tr>
<tr>
<td>P. cepacia</td>
<td>0.05</td>
<td>0.05</td>
<td>0.50</td>
<td>0.05</td>
</tr>
<tr>
<td>P. fluorescens</td>
<td>0.50</td>
<td>0.05</td>
<td>0.15</td>
<td>0.05</td>
</tr>
<tr>
<td>P. putida</td>
<td>0.05</td>
<td>0.60</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Isocaproic acid positive
negative
P. maltophilia (100%)

Caproic acid positive
negative
P. pan genomicis (100%)

Propionic acid positive
P. stutzeri (100%)

Selection of a set of a priori probabilities on the basis of the results given by two tests:
production of pyoverdin
production of gelatinase

Original groups
Discriminant groups:
P. acidovorans 91% 9%
P. aeruginosa 9% 9%
P. cepacia 9% 94%
P. fluorescens 8% 8% 8%
P. putida 1% 99%

Total percentage of strains from group B correctly identified = 96%

Fig. 3. Schematic diagram of the identification system. The numbers indicate the percentages of strains correctly identified. Positive, production of the volatile fatty acid; negative, no production of the volatile fatty acid.

percentage of strains from group A belonging to P. acidovorans, P. aeruginosa, P. cepacia, P. fluorescens or P. putida classified by means of discriminant functions (83%), was very similar to that found for the strains in group B belonging to the same species (82%). This result confirmed that the error rate after 'jacknifing' was an unbiased estimate of the error rate in discriminating between the whole populations of the species (Lachenbuch & Mickey, 1968).

In order to reduce overlap between P. putida and P. fluorescens an a priori probability was introduced in the $S_0$ coefficient of each of the discriminant functions. This was computed on the basis of the results given by two tests: production of pyoverdin and production of gelatinase. Previously every unknown isolate had the same a priori probability of being assigned to any of the eight species before chromatographic analysis. Now, according to the results given by those two tests, the probabilities for an unknown strain being assigned to one of the species would not be identical. These a priori probabilities were computed for all species on the basis of our own results and those reported by Bergan (1981) and Gilardi (1978) (Table 5). As a consequence, the separation between P. fluorescens and P. putida was increased and 96% of the 249 strains from group B were correctly identified by this system (Fig. 3).

DISCUSSION

The results from this study illustrate a valuable approach to bacterial identification using discriminant analysis of the chromatographic determination of VFA produced in culture medium. Although the mathematical and chromatographic analyses may seem cumbersome at first, many laboratories already use GLC for the identification of bacteria. Moreover, several computer programs that perform the necessary calculations for discriminant analysis are now available (Jennrich & Sampson, 1981; Romeder, 1973). This methodology, as well as other methods using computerized algorithms (Dybrowsky & Franklin, 1968; Friedmann & MacLowry, 1973; Sielaff et al., 1979), requires three conditions. First, by performing some preliminary tests, it must be confirmed that the unknown isolate is a Pseudomonas and of a species susceptible to identification by the proposed procedure. Second, when selecting strains to establish baseline values, extreme care must be taken to ensure that the strains accurately represent the different
species populations. Finally, all the selected strains must be accurately identified by conventional biochemical methods.

Within the above limitations, discriminant analysis on the basis of their VFA production allowed the classification of strains of eight species of Pseudomonas and enabled the accurate identification of unknown strains belonging to those species to be made. In spite of the relatively poor differentiation between P. fluorescens and P. putida, 89% of all the 'unknown' strains were correctly identified by this technique. The similarities between P. fluorescens and P. putida have already been reported (Bergan, 1981; Palleroni & Douderoff, 1979) and Rhodes (1959) suggested that these two species should be merged into one. In the same way, many of the automated identification systems available do not make a distinction between these two species (Sieff et al., 1979) and it is often more important for the bacteriologist to distinguish between P. aeruginosa and other fluorescent pseudomonads than between P. putida and P. fluorescens (Bergan, 1981). Nevertheless, we have shown that in such cases where species are very similar, the accuracy of our system can be increased by the use of a priori probabilities based on a few simple biochemical tests.

Unknown isolates not belonging to the reference species (outliers) pose a significant problem. Such strains would be assigned to the nearest species and thus misidentified. In such cases, only a careful inspection of Mahalanobis distances and posteriori probabilities would reveal their true identity. Nevertheless from these preliminary results in the identification of 249 unknown isolates, we can assume that GLC analysis of VFA produced in culture medium, when used in conjunction with discriminant analysis, provides an accurate means for characterizing and identifying some of the most frequently encountered Pseudomonas species. Although only eight species were included in this study, the extension of this procedure to other species or genera should not create any obstacle. In fact, it would be possible to increase the number of detectable organic acids so that the results of biochemical tests could be combined with chromatographic data to obtain more accurate identification in doubtful instances.

The authors gratefully acknowledge Miss M. Zinsius for her excellent and skilful technical assistance. This work was supported in part by a grant from 'Piren-Eau Alsace'.

REFERENCES


