

The deconvolution of pyrolysis mass spectra using genetic programming: application to the identification of some *Eubacterium* species

Janet Taylor ^a, Royston Goodacre ^{a,*}, William G. Wade ^b, Jem J. Rowland ^c,
Douglas B. Kell ^a

^a Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3DD, UK

^b Oral Microbiology Unit, Department of Oral Medicine Pathology, UMDS, Microbiology and Immunology, Guy's Hospital, London SE1 9RT, UK

^c Department of Computer Science, University of Wales, Aberystwyth, Ceredigion SY23 3DB, UK

Received 23 October 1997; revised 16 January 1998; accepted 22 January 1998

Abstract

Pyrolysis mass spectrometry was used to produce complex biochemical fingerprints of *Eubacterium exiguum*, *E. infirmum*, *E. tardum* and *E. timidum*. To examine the relationship between these organisms the spectra were clustered by canonical variates analysis, and four clusters, one for each species, were observed. In an earlier study we trained artificial neural networks to identify these clinical isolates successfully; however, the information used by the neural network was not accessible from this so-called 'black box' technique. To allow the deconvolution of such complex spectra (in terms of which masses were important for discrimination) it was necessary to develop a system that itself produces 'rules' that are readily comprehensible. We here exploit the evolutionary computational technique of genetic programming; this rapidly and automatically produced simple mathematical functions that were also able to classify organisms to each of the four bacterial groups correctly and unambiguously. Since the rules used only a very limited set of masses, from a search space some 50 orders of magnitude greater than the dimensionality actually necessary, visual discrimination of the organisms on the basis of these spectral masses alone was also then possible. © 1998 Federation of European Microbiological Societies. Published by Elsevier Science B.V.

Keywords: Chemometrics; *Eubacterium*; Genetic programming; Pyrolysis mass spectrometry

1. Introduction

The oral asaccharolytic *Eubacterium* species are a diverse group of organisms that are implicated in periodontitis, endodontic infections and dentoalveo-

lar abscesses [1,2]. They are slow-growing and difficult to identify by conventional means. The number of documented species continues to increase (for examples see [3,4]) and in a recent study [5] we used pyrolysis mass spectroscopy (PyMS) to confirm the taxonomic position of the three newly described species *E. exiguum*, *E. infirmum* and *E. tardum*.

PyMS is a 'whole cell fingerprinting' method

* Corresponding author. Tel.: +44 (1970) 621947;
Fax: +44 (1970) 622354; E-mail: rrg@aber.ac.uk

which can be used to produce a biochemical fingerprint of the substance or organism under study. However, the clustering of these multivariate fingerprints is not possible by simple visual inspection and has usually been carried out by conventional 'unsupervised' chemometric tools such as discriminant analysis and hierarchical cluster analysis. More recently the development of artificial neural networks (ANNs) has provided an alternative, 'supervised' learning method [6] and although ANNs have been shown to be effective tools for microbial identification and discrimination from PyMS data (for examples see [5,7,8]) the information in terms of which masses in the mass spectrum are important is not readily available, and ANNs are often perceived as a 'black box' approach to modelling spectra.

Genetic programming (GP) is an alternative supervised learning technique devised by Koza [9–11]. This method attempts to provide a solution in the form of a procedural program without the need for explicit coding (see Fig. 1 for a flowchart for GP). GP effectively performs a directed search through the abstract space of possible computer programs from random starting points. The programs generated are functions or expressions that are optimised using the Darwinian principles of selection and reproduction [12]. An initial population of computer programs is generated from a set of functions, which may be simple arithmetic functions or complex algebraic formulae, and a terminal set, which contain variables or constant values that may be used as input to the functions. Each function must accept and return the same data type, so that a working program is obtained (that is to say, the functions must have closure so that the function tree can actually be translated into a mathematical equation). Each program is generated in the form of a tree structure. Each individual in the population is tested via a fitness function, and a score obtained. These individuals are then selected, on the basis of their fitness score, for reproduction. There are two common reproduction processes: asexual reproduction, where a direct copy of the individual is made; and sexual reproduction, where random parts of two individuals are swapped to form two different offspring (see Fig. 2 for the crossover operation for GP). The child population then undergoes fitness testing, and this cycle is continued until an acceptable error level is

reached, producing a suitable computer program that solves the problem to a satisfactory degree. A GP optimises an individual by executing it using the training data as input to the program; the error is then calculated as the difference between the program output and the measured result. The ability to take multiple input values into a program makes GP an ideal candidate for multivariate analysis, where many x variables bear a relationship to one property (y variable). Another benefit of GP is that the output of a run is a procedural program whose instructions elucidate the variables used, and the required manipulation of those values. This allows reference back to the original spectrum for further analysis.

The aim of this study was to exploit GPs to deconvolute the pyrolysis mass spectra of strains previously identified by phylogenetic analyses as one of four *Eubacterium* species. To be confident in the GP's solution an independent test set of each of the four species was used along with five hospital isolates which had been recently identified as *E. exiguum* [5].

2. Materials and methods

2.1. Organisms and cultivation

Details and origins of the organisms are given in Table 1. Strains were cultured on Fastidious Anaerobe agar (Lab M, Bury, UK) plus 5% sheep blood and incubated anaerobically in an atmosphere of N₂ 80%, CO₂ 10%, H₂ 10% for 72 h. The bacteria were harvested with a nichrome wire loop and suspended in phosphate buffered saline to 20 mg ml⁻¹.

2.2. Pyrolysis mass spectrometry

Bacterial samples (5 µl) were evenly applied on to iron-nickel foils to give a thin uniform surface coating. Prior to pyrolysis the samples were oven-dried at 50°C for 30 min. Each sample was analysed in triplicate. For full operational procedures see [7,13,14]. The sample tube carrying the foil was heated, prior to pyrolysis, at 100°C for 5 s. Curie-point pyrolysis was at 530°C for 3 s, with a temperature rise time of 0.5 s. These conditions were used for all experiments.

The data from PyMS were collected over the m/z range 51–200 and may be displayed as quantitative pyrolysis mass spectra (e.g. as in Fig. 3; here normalised to total ion count). The abscissa represents the m/z ratio whilst the ordinate contains information on the ion count for any particular m/z value ranging from 51 to 200.

2.3. Data pre-processing

The 150 normalised masses in each spectrum were reduced in number using their characteristicity values. Characteristicity is closely related to the Fisher (F) ratio [15], where:

$$F = \frac{\text{Between-group variance}}{\text{Within-group variance}}$$

and has been used to select relevant masses in PyMS spectra for multivariate analysis [16].

Characteristicity was calculated as described by Eshuis et al. [17]. The mass intensities are then ranked in order of their characteristicities; large values are more important, smaller ones less so. The 20 most characteristic masses were selected to form the set on which the GP was performed.

2.4. Discriminant function analysis

It is important that the 20 most characteristic masses contain as much information to separate the four *Eubacterium* species as do all 150 masses; therefore discriminant function analysis (DFA) was performed as described by Manly [15]. Ordination plots based on the first few DFs can be viewed and used to ascertain if the 24 *Eubacterium* strains, representative of the four bacterial groups, were clustered in a similar manner.

2.5. Genetic programming

A commercial GP package, *e* version 1.0 (System Dynamics International Inc., 512 Rudder Road, Fenton, MO 63026), was used in these experiments. The software performs genetic programming on one or more sets of randomly generated computer program functions. The components that comprise each program are selected from a set of operators provided by the package (Table 2). Each set of individ-

uals is known as an ecosystem. Each ecosystem is generated, and evolved in isolation, enabling different conditions for evolution to be set if desired. In a multi-ecosystem experiment individuals may transfer between isolated ecosystems using a process termed migration (the probability value for migration was set to 0.1; the probability values for all the genetic operations were on a scale of 0.0–1.0). If this process is enabled, an individual from one ecosystem is selected according to fitness and is copied into another ecosystem, replacing the individual with the worst fitness score. For the purposes of this experiment four ecosystems were created randomly, each with a different training set, but with other parameters set to be identical. Each ecosystem was set to contain a maximum of 15 different computer program functions (individuals) as its population. Each computer program function was constructed as a series of instructions, in order for relatively simple rules to be developed complex trees were penalised by setting the maximum number of instructions (tree length) in an individual to 15. The probability values for mutation and sexual reproduction (crossover) of instructions were set to 0.03 and 0.8 respectively.

The generated ecosystems were each evolved to recognise a specific *Eubacterium* species, using the parameters and operators detailed in Table 2. Ecosystem 0 was trained to identify *E. timidum*, ecosystem 1 to identify *E. infirmum*, ecosystem 2 to identify *E. exiguum*, and ecosystem 3 trained to identify *E. tardum*. The training set for the GPs contained only 45 spectra (15 samples in triplicate) described by 150 m/z intensities, and it is well known that if the number of parameters in calibration models such as ANNs and GPs are significantly higher than the number of exemplars in the training set then these methods have a tendency to overfit [18,19]. Therefore to obey the parsimony principle as described by Seasholtz and Kowalski [18] the number of inputs to the GP was reduced by ranking the masses according to their characteristicity values (see above for details). Twenty masses were chosen because (as detailed above) they adequately described the discrimination of these organisms in discriminant analysis space. The input data for each of the ecosystems were the first 20 most characteristic masses (61, 94, 91, 99, 136, 103, 62, 92, 69, 85, 150, 122, 68,

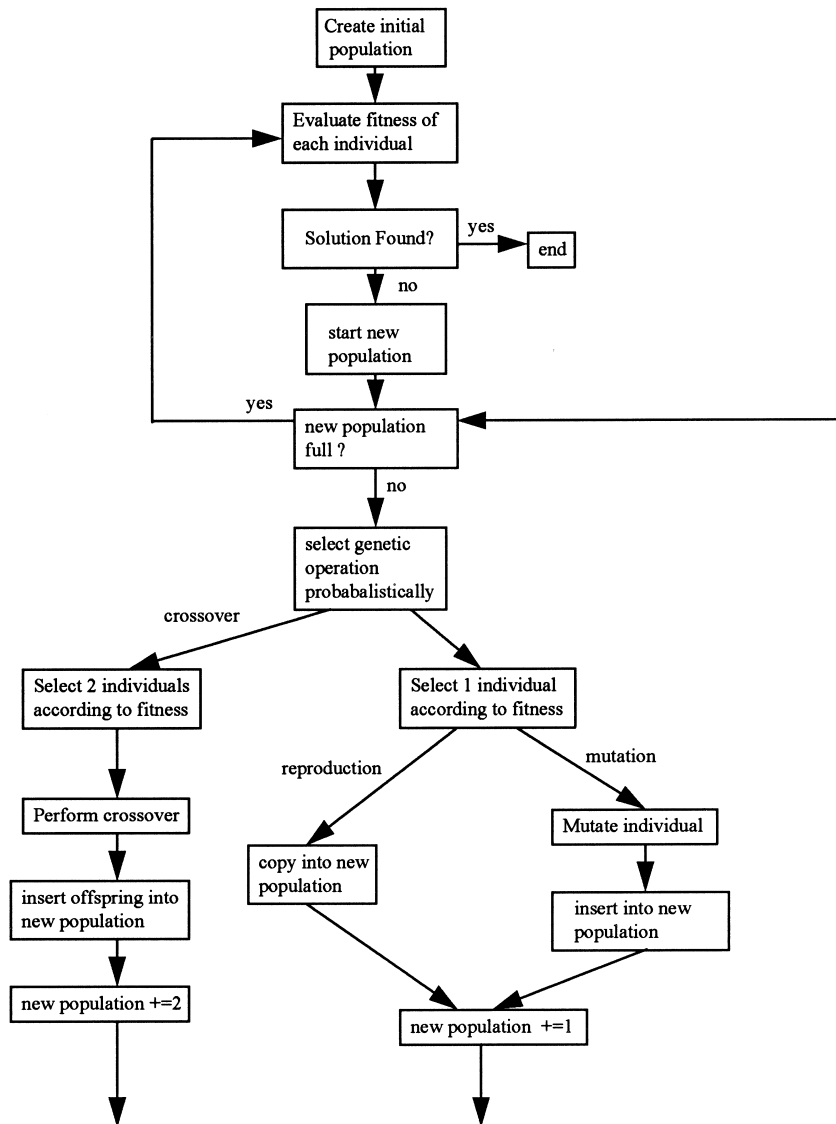


Fig. 1. Flowchart showing the overall scheme for genetic programming (adapted from The Genetic Programming Tutorial Notebook, URL: <http://www.geneticprogramming.com/Tutorial/index.html>).

84, 162, 65, 88, 58, 107 and 102) and these were scaled so that the sum of all 20 masses was unity.

From each of the four 'known' species, one of the samples (in triplicate) was reserved to form a test set (details are given in Table 1). The remaining samples were used as training data for the GPs. The single output of each of the ecosystems was encoded so that 1 was taken as the bacterium under study belonging to that *Eubacterium* species and 0 was taken

as a negative recognition for that *Eubacterium* class. Training was over a number of 'generations' where a single generation was one full reproduction cycle of the ecosystem, and training was stopped when the raw error of prediction on the training set reached 0.1. This error was calculated by the proportion of incorrect 'hits' by the program in recognising a species. After training the GPs both the training and test sets were used to challenge the GPs.

3. Results and discussion

Typical pyrolysis mass spectra are shown in Fig. 3. These and the pyrograms from all 24 bacteria show very little qualitative difference between the spectra, although some complex quantitative differences between them were observed; for example mass 58 in the *E. exiguum* spectrum has a much higher ion intensity than that in any of the other *Eubacterium* species, and a large mass 91 appears to be characteristic for *E. timidum*.

After collection of the data the first stage was to observe the relatedness between all these strains using DFA. It was evident in the DFA plots (data not shown) that four clusters, one for each *Eubacterium* species, were seen. The next stage was to perform DFA on the first 20 most characteristic masses, and it was also evident that these ordination plots (data not shown) allowed sufficient separation between the four different *Eubacterium* species.

Since there was no degradation in clustering by

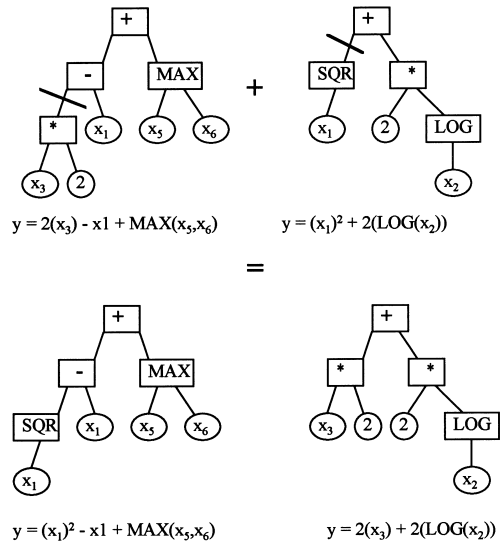
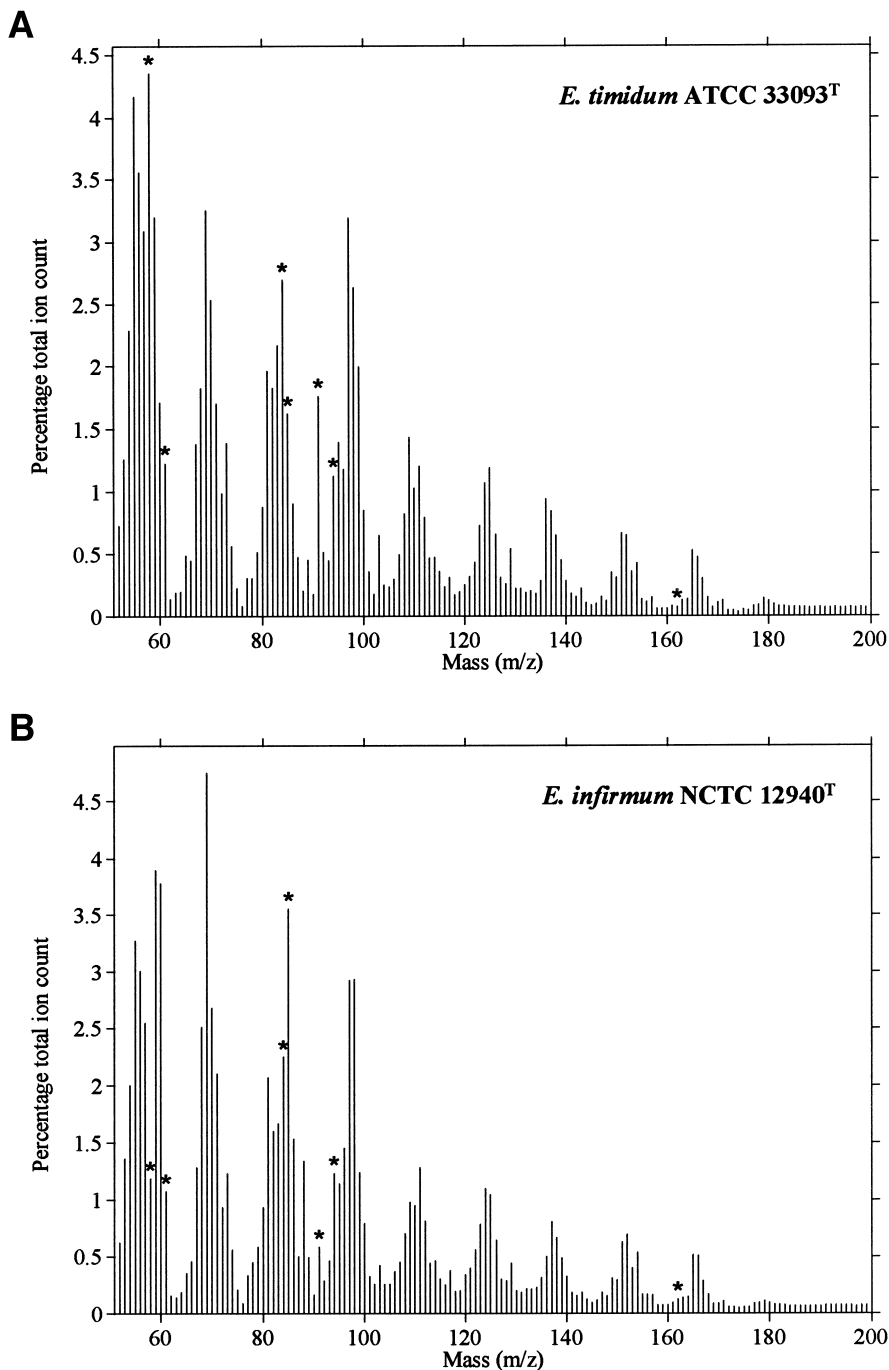


Fig. 2. An example of the crossover operation for genetic programming. The following tree represents the crossover of parents $y = 2(x_3) - x_1 + \text{MAX}(x_5, x_6)$ and $y = (x_1)^2 + 2(\text{LOG}(x_2))$ at the point indicated by a 'slash', leading to the production of progeny with equations $y = (x_1)^2 - x_1 + \text{MAX}(x_5, x_6)$ and $y = 2(x_3) + 2(\text{LOG}(x_2))$.

Table 1
Eubacterium species studied

Identifier	Species/group	Strain number ^a	Origin/reference	Training or test set
Ta	<i>E. timidum</i>	ATCC 33093 ^T	[21]	Training
Tb		W557		Training
Tc		W690		Training
Td		W693		Training
Te		W2847		Test
1a	<i>E. infirmum</i>	NCTC 12940 ^T	[3]	Training
1b		W687		Training
1c		W1475		Training
1d		W1470		Test
2a	<i>E. exiguum</i>	SC142	[22]	Training
2b		SC108		Training
2c		W1365		Training
2d		W733		Training
2e		W2848		Test
Na	<i>E. tardum</i>	SC68	[3]	Training
Nb		SC88P		Training
Nc		SC41B		Training
Nd		SC37		Training
Ne		NCTC 12941 ^T		Test
Ha	Hospital isolates	SBH463	All bacteria from oral abscess at St. Bartholomew's Hospital, London	Test
Hb		SBH481		Test
Hc		SBH462		Test
Hd		SBH403		Test
He		SBH477		Test

^aStrains have been deposited in the ATCC (American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD 20852-1776, USA) or NCTC (National Collection of Type Cultures, Central Public Health Laboratory, 61, Colindale Avenue, London NW9 5HT, UK).



using only 20 masses the four GP ecosystems were trained using GPs with these data to a raw error of 0.1. This process was fast, and took only 1 min and 12 s on a 486 DX66 IBM compatible running under

Microsoft NT 4.01. After calibration the following mathematical expressions were produced by each ecosystem:

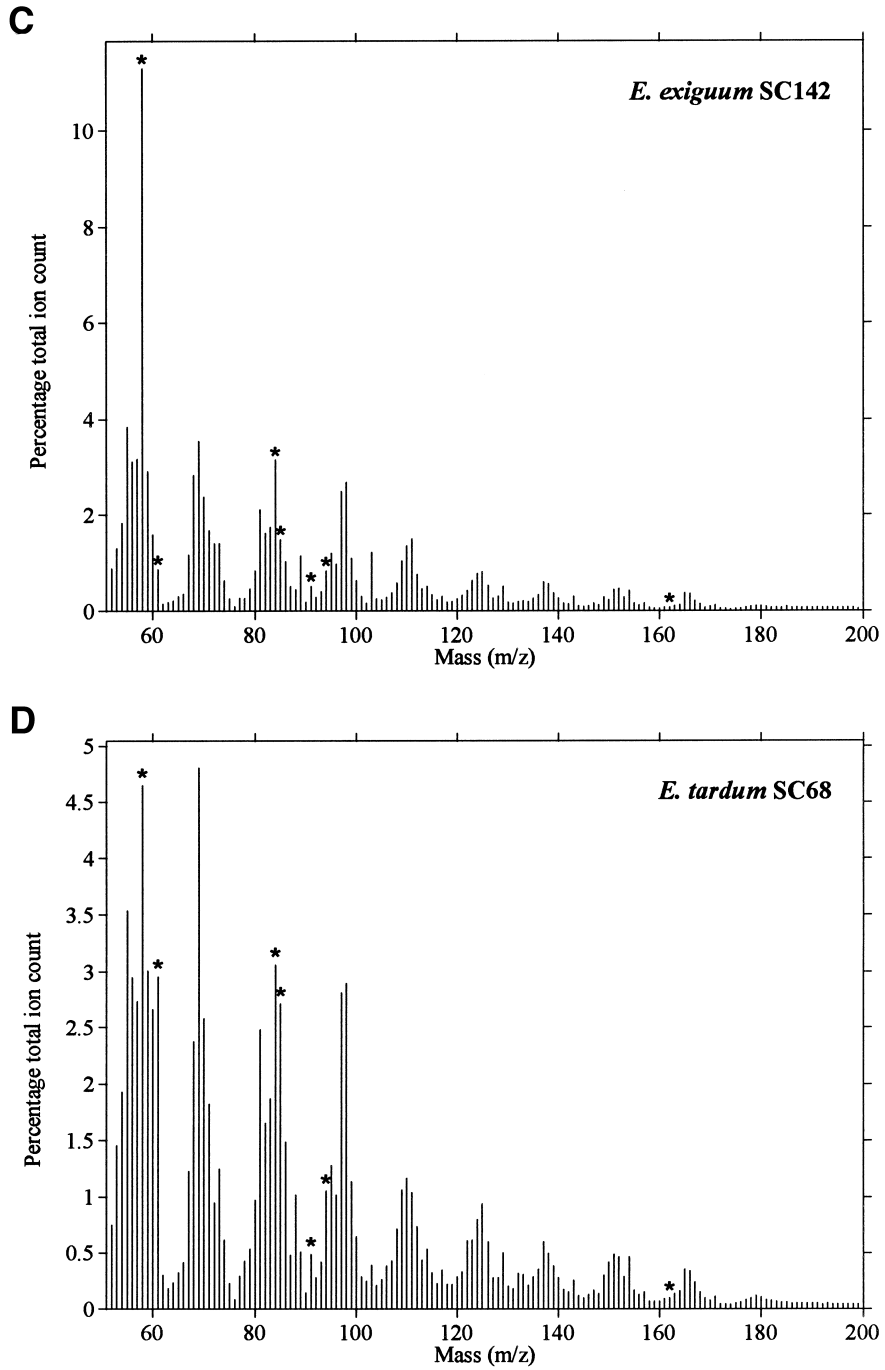


Fig. 3. Deconvolution and interpretation of the pyrolysis mass spectra of *E. timidum* ATCC 33093^T, *E. infirmum* NCTC 12940^T, *E. exiguum* SC142 and *E. tardum* SC68. * indicates the peaks chosen by the GPs.

- Ecosystem 0 was trained to identify *E. timidum*, calibration took 1032 generations and the formula that evolved used only mass 91 and was:
 - **IF** (m91–AVG(x,20)) <= 0.0, **THEN** identity = 0, **ELSE** identity = 1.
 - The AVG function calculates the average scaled mass intensity of a single spectrum (x).
- Ecosystem 1 was trained to identify *E. infirmum*, calibration took 66 generations and the formula that evolved used masses 84, 85 and 162 and was:
 - **IF** (m85–m162–m84) <= 0.0, **THEN** identity = 0, **ELSE** identity = 1.
- Ecosystem 2 was trained to identify *E. exiguum*, calibration took 165 generations and the formula that evolved used only mass 58 and was:
 - **IF** (m58–0.23835) <= 0.0, **THEN** identity = 0, **ELSE** identity = 1.
- Ecosystem 3 was trained to identify *E. tardum*, calibration took 171 generations and the formula that evolved used masses 61 and 94 and was:
 - **IF** (m61–2*(m94)) <= 0.0, **THEN** identity = 0, **ELSE** identity = 1.

After training, each calibrated ecosystem was challenged with the training and test sets. As expected the GP identified all 15 bacteria in the training set; more importantly however is that all nine isolates in the *unknown* (unseen) test set were also correctly identified; there was 100% correct recognition of the nine species by each ecosystem (scored as a 1), and 0% recognition of species which were not in that ecosystem (scored as a 0). Ecosystem 2 also placed each of the five unseen hospital isolates into the category of *E. exiguum*, a result which has previously been achieved with ANNs [5].

The above result was very encouraging because although both ANNs and GP can be used successfully to identify these *Eubacterium* isolates, the GP has also provided an insight into the analysis of the original spectrum, which by contrast is not accessible to the user from ANN analyses. Seven of the 20 most characteristic masses were used and these were 58, 61, 84, 85, 91, 94 and 162 and were ranked according to their characteristicity index as 18th, 1st, 14th, 10th, 3rd, 2nd and 15th most important respectively. These masses are highlighted by asterisks on the PyMS spectra in Fig. 3.

That the GPs all produced different expressions, and that they were rather simplistic mathematical formula, has permitted the deconvolution of the PyMS in terms of which masses were important. The next stage was therefore to interpret these rules from the mass spectra. Two of the ecosystems used only a single mass intensity and so are consequently very easy to interpret; ecosystem 0 was trained to identify *E. timidum* and the formula produced was **IF** (m91–AVG(x,20)) <= 0.0, **THEN** identity = 0, **ELSE** identity = 1. When the spectra are inspected one can see that mass 91 is highest in spectra from *E. timidum* (as noted above); indeed the average magnitude of this scaled mass for all the *E. timidum* strains was 0.099, whilst the same mass in all the other eubacteria was only < 0.024. The other single mass expression was from ecosystem 2 which was trained to identify *E. exiguum* and the formula produced was **IF** (m58–0.23835) <= 0.0, **THEN** identity = 0, **ELSE** identity = 1. When the spectra are inspected (Fig. 3) one can see that mass 58 is highest in spectra from *E. exiguum*, the average for this mass for all the *E. exiguum* strains was 0.39, whilst the same mass in all the other eubacteria was < 0.16; moreover the intensity of mass 58 for the hospital isolates was also high and was typically 0.36. For the identification of *E. tardum* the expression produced

Table 2
Parameters and operators used in the GP

Parameter	Setting
Number of ecosystems	4
Population size	15
Migration	Enabled
Operators	Meaning
+	add
–	subtract
*	multiply
/	divide
MAX(x,y)	larger of x, y
MIN(x,y)	smaller of x, y
NOT(x)	1–x
AVG(x[,num])	average value of an array x (all mass intensities for a given organism) where num is the number of elements (or masses) in that array

Also included is an IF...THEN...ELSE statement.

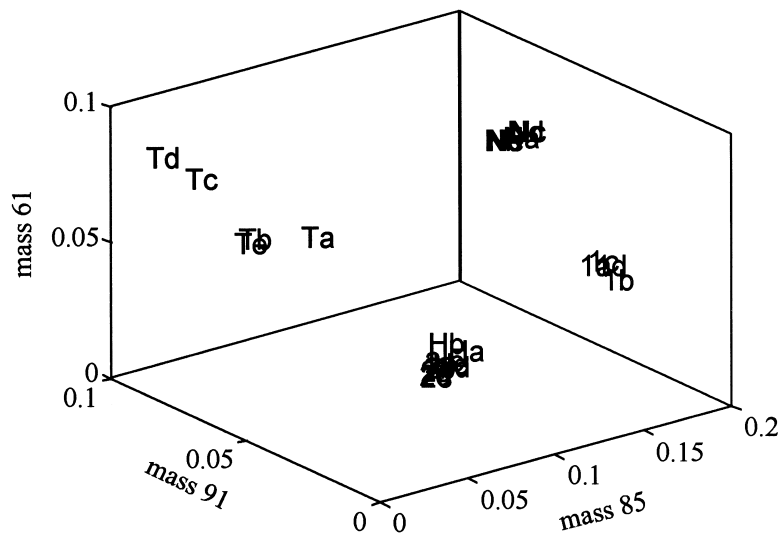


Fig. 4. Pseudo-3-D plot of the scaled mass intensities 85 vs 91 vs 61. Data points are the averages for the replicate samples and their identifiers are given in Table 1.

by ecosystem 3 was **IF** $(m61 - 2 \cdot (m94)) \leq 0.0$, **THEN** identity = 0, **ELSE** identity = 1. Mass 61 was highest in *E. tardum* and the average for this species was 0.11, compared with < 0.05 for the other *Eubacterium* spp. There was no obvious trend in mass 94 and the scaled intensities for these masses ranged between 0.03 and 0.06 and the intensity for *E. tardum* was 0.04. However, as was evident from the expression, the weight ratio of these two masses was important. Finally, the most complex expression was **IF** $(m85 - m162 - m84) \leq 0.0$, **THEN** identity = 0, **ELSE** identity = 1, and was produced from ecosystem 1 which was trained to identify *E. infirmum*. As one might expect from this formula mass 85 was highest in *E. infirmum* and was 0.16 whilst for the other eubacteria this mass intensity was < 0.097 , although there was no obvious relationship between masses 84 and 162 other than the ratio between these masses being important.

Finally the significance of the GPs choosing masses 58, 61, 85 and 91 as the most important inputs can be seen when three of these (61, 85 and 91) are plotted against one another (Fig. 4); each of the four *Eubacterium* spp. are recovered in separate clusters, and the hospital isolates unequivocally group with the five *E. exiguum* strains. An exhaustive search of all 2^{150} (10^{45}) possible permutations (where a mass is either used or not) would be computation-

ally prohibitive, and to effect visible inspection of all possible three dimensional plots would necessitate graphing 5.5×10^5 plots since the number of combinations needed to pick 3 masses out of 150 = $150! / [(150-3)!3!]$.

It is very significant that the GP has highlighted which masses are predominantly important in separating these four *Eubacterium* species because it will now allow the deconvolution in terms of what (bio)-chemical information is different in each of the bacteria and thereby could allow specific characteristic biochemical markers to be developed for each of the eubacteria. Due to the complex nature of the pyrolysis process this will be difficult, since on pyrolysis the weakest bonds between all molecules in and on the cell are broken. In addition, it is likely that each mass is not of unique origin and so a more detailed analysis of the peaks highlighted by the GP might have been effected using pyrolysis GC-MS or pyrolysis tandem MS-MS, but this facility was not available to us. However, the following information is available (Prof. Jaap J. Boon, personal communication); mass 58 is likely to be a radical recombination product of methyl radical and CH_3CO and so a specific pyrolysis fragment polysaccharide, mass 61 is the most basepeak of glycerol from lipids, mass 85 is often the largest peak in pentose anhydrosugar products, and mass 91 is the base peak of toluene,

a pyrolysis products of phenylalanine found in proteins [20].

In conclusion, this is the first study that has shown that the genetic programming approach can be applied successfully to the accurate identification of bacteria by analysis of their PyMS spectra. This approach provided identities which were as good as those from other supervised learning methods such as ANNs, but with the enormous additional benefit of enabling the deconvolution of the pyrograms in terms of which masses were characteristic for each of the bacterial species studied.

Acknowledgments

J.T., J.J.R. and D.B.K. thank the UK BBSRC, EPSRC and Science in Finance Ltd for financial support. R.G. is indebted to the Wellcome Trust for financial support (Grant 042615/Z/94/Z).

References

- [1] Wade, W.G., Lewis, M.A.O., Cheeseman, S.L., Absi, E.G. and Bishop, P.A. (1993) An unclassified *Eubacterium* taxon in acute dentoalveolar abscess. *J. Med. Microbiol.* 40, 115–117.
- [2] Wade, W.G. (1996) The role of *Eubacterium* species in periodontal disease and other oral infections. *Microb. Ecol. Health Dis.* 9, 367–370.
- [3] Cheeseman, S.L., Hiom, S.J., Weightman, A.J. and Wade, W.G. (1996) 6S ribosomal DNA sequence comparison and proposal of *Eubacterium infirmum* sp. nov. and *Eubacterium tardum* sp. nov. *Int. J. Syst. Bacteriol.* 46, 957–959.
- [4] Wade, W.G., Lewis, M.A.O., Cheeseman, S.L., Absi, E.G. and Bishop, P.A. (1994) An unclassified *Eubacterium* taxon in acute dentoalveolar abscess. *J. Med. Microbiol.* 40, 115–117.
- [5] Goodacre, R., Hiom, S.J., Cheeseman, S.L., Murdoch, D., Weightman, A.J. and Wade, W.G. (1996) Identification and discrimination of oral asaccharolytic *Eubacterium* spp. by pyrolysis mass spectroscopy and artificial neural networks. *Curr. Microbiol.* 32, 77–84.
- [6] Goodacre, R. and Kell, D.B. (1996) Pyrolysis mass spectrometry and its applications in biotechnology. *Curr. Opin. Biotechnol.* 7, 20–28.
- [7] Goodacre, R., Neal, M.J., Kell, D.B., Greenham, L.W., Noble, W.C. and Harvey, R.G. (1994) Rapid identification using pyrolysis mass spectrometry and artificial neural networks of *Propionibacterium acnes* isolated from dogs. *J. Appl. Bacteriol.* 76, 124–134.
- [8] Freeman, R., Goodacre, R., Sisson, P.R., Magee, J.G., Ward, A.C. and Lightfoot, N.F. (1994) Rapid identification of species within the *Mycobacterium tuberculosis* complex by artificial neural network analysis of pyrolysis mass spectra. *J. Med. Microbiol.* 40, 170–173.
- [9] Koza, J.R. (1992) *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA.
- [10] Koza, J.R. (1994) Genetic programming as a means for programming computers by natural selection. *Statist. Comput.* 4, 87–112.
- [11] Koza, J.R., Goldberg, D.E., Fogel, D.B. and Riolo, R.L. (1996) *Genetic Programming 1996: Proceedings of the First Annual Conference*. MIT Press, Cambridge, MA.
- [12] Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection*. John Murray, London.
- [13] Goodacre, R., Trew, S., Wrigley-Jones, C., Saunders, G., Neal, M.J., Porter, N. and Kell, D.B. (1995) Rapid and quantitative analysis of metabolites in fermentor broths using pyrolysis mass spectrometry with supervised learning: application to the screening of *Penicillium chrysogenum* fermentations for the overproduction of penicillins. *Anal. Chim. Acta* 313, 25–43.
- [14] Goodacre, R., Neal, M.J. and Kell, D.B. (1994) Rapid and quantitative analysis of the pyrolysis mass spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks. *Anal. Chem.* 66, 1070–1085.
- [15] Manly, B.F.J. (1994) *Multivariate Statistical Methods: A Primer*. Chapman and Hall, London.
- [16] Gutteridge, C.S. (1987) Characterization of microorganisms by pyrolysis mass spectrometry. *Methods Microbiol.* 19, 227–272.
- [17] Eshuis, W., Kistemaker, P.G. and Meuzelaar, H.L.C. (1977) Some numerical aspects of reproducibility and specificity. In: *Analytical Pyrolysis* (Jones, C.E.R. and Cramers, C.A., Eds.), pp. 151–156. Elsevier, Amsterdam.
- [18] Seasholtz, M.B. and Kowalski, B. (1993) The parsimony principle applied to multivariate calibration. *Anal. Chim. Acta* 277, 165–177.
- [19] Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford.
- [20] Meuzelaar, H.L.C., Haverkamp, J. and Hileman, F.D. (1982) *Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials*. Elsevier, Amsterdam.
- [21] Holdeman, L.V., Cato, E.P., Burmeister, J.A. and Moore, W.E.C. (1980) Descriptions of *Eubacterium timidum* sp. nov., *Eubacterium brachy* sp. nov. and *Eubacterium nodatum* sp. nov. isolated from human periodontitis. *Int. J. Syst. Bacteriol.* 30, 163–169.
- [22] Poco, S.E.J., Nakazawa, F., Ikeda, T., Sato, M., Sato, T. and Hoshino, E. (1996) *Eubacterium exiguum* sp. nov., isolated from human oral lesions. *Int. J. Syst. Bacteriol.* 46, 1120–1124.