Identification of Oxalotrophic Bacteria by Neural Network Analysis of Numerical Phenetic Data

N. SAHIN[a], S. AYDIN[b]

[a]*OFMA Biyoloji Egitimi, Egitim Fakultesi, Mugla University, 48170 Kotekli, Mugla, Turkey*
fax +90 252 223 8491
e-mail nsahin@mu.edu.tr

[b]*Department of Electrical Engineering, Faculty of Engineering, Celal Bayar University, 45140 Muradiye, Manisa, Turkey*

Received 1 March 2005

ABSTRACT. A new approach with artificial neural network (ANN) was applied to numerical taxonomy of bacteria using the oxalate as carbon and energy source. For this aim the characters effective in differentiating separate groups were selected from morphological, physiological and biochemical test results. Fourteen aerobic, Gram-negative, oxalate-utilizing isolates and four oxalate-utilizing reference strains (*Ralstonia eutropha* DSM 428, *Methylobacterium extorquens* DSM 1337^T, *Ralstonia oxalatica* DSM 1105^T, *Oxalicibacterium flavum* DSM 15506^T) were included in the study. ANN program used here was developed in Borland C^{++} language. Iterations were performed on an IBM compatible PC computer. ANN architecture having feed-forward backpropagation algorithm was used for teaching generalized δ rule. The results show that ANN can have a large potential in solving the taxonomic problems of oxalate-utilizing bacteria.

A well-established means for the assessment and evaluation of phenotypic data is numerical taxonomy (Sneath and Sokal 1973). The application of the concepts of numerical taxonomy was made possible only through the use of computers because of the heavy load of routine calculations (Sneath 1984). Rapid, reproducible and cost-effective methods are at premium for the classification, identification and typing of microorganisms. Artificial neural networks (ANNs) are well-known means of uncovering complex, non-linear relationships in multivariate data (Simpson 1990). ANN has an architecture analogous to the human brain. Some biological data such as the results of flow cytometry (Morris *et al.* 1992) or electrophoretic separation of cellular proteins (Kesters 1985), have been successfully analyzed using neural networks. Neural networks and ANNs, algorithms which mimic neural network function, are a form of problem solving that possesses a functional architecture of interconnected neurons in layers. Each neuron receives an input signal (information) from other connected neurons and makes a computation applied to an activation function. If the inputs exceed a set threshold, the neuron is activated, and the active neuron then passes an output signal to other neurons within the network (Widrow 1990).

Bacteria that are capable of using oxalate as a sole carbon and energy source are described as being “oxalotrophic”. Oxalotrophic bacteria do not constitute a homogeneous taxonomic group but they form a well-defined physiological group. The taxonomy of oxalotrophic bacteria has undergone many changes (Sahin 2003, 2004). Most of them are facultative methylotrophs and/or facultative hydrogen-oxidizing chemolithoautotrophs. Among the aerobic oxalotrophic bacteria *Ralstonia oxalatica* DSM 1105^T (Khambata and Bhat 1953; Sahin *et al.* 2000), *Ralstonia eutropha* (Jenni *et al.* 1988), *Methylobacterium extorquens* DSM 1337^T (Bassalik 1913; Bousfield and Green 1985) have been described most completely from the taxonomic and physiological points of view. Despite this progress, many questions concerning the taxonomy of this group remain unresolved, among them the identification of new isolates on the species level.

Determining the appropriate assignment of oxalotrophic strains to recognized species is still based on the biochemical characters and requires further assessment. Therefore, the present study reports on the ANN analysis and the results of numerical phenetic data of oxalate utilizing aerobic bacteria to propose an alternative solution in addition to other numerical taxonomic methods.

MATERIALS AND METHODS

Test strains. Fourteen test strains represented three phenotypically distinct groups of oxalotrophic bacteria (Table I; *cf.* Fig. 4). Culture conditions were described by Sahin *et al.* (2002).

Table I. Bacterial strains studied

Strain		Collection no.[a]	Reference
Methylobacterium extorquens		DSM 1337[T]	Bassalik 1913
Methylobacterium sp.	NS06, NS07 NS08, NS09		Sahin *et al.* 2002
Oxalicibacterium flavum	TA17	DSM 15506[T]	Tamer *et al.* 2002
	NS13	DSM 15507	*ditto*
Pseudomonas sp.	KOx	CCM 2766	Chandra and Shethna 1975
	YOx	CCM 2767	*ditto*
	OD1	IFO 13594	Jayasuriya 1955
Ralstonia eutropha	H16	DSM 428	Davis *et al.* 1969
	TA6	DSM 4182	Tamer *et al.* 1980
oxalatica	Ox1	DSM 1105[T]	Sahin *et al.* 2000
Xanthobacter sp.	NS14		Sahin 2002

[a]DSM – *Deutsche Sammlung von Mikroorganismen und Zellkulturen* (DSMZ), Braunschweig, Germany; CCM – *Czech Collection of Microorganisms, Masaryk University Brno*, Czechia; IFO – *Institute for Fermentation*, Osaka, Japan; [T] indicates type strain of the species.

Conventional data analysis. The results of tests that were positive or negative for all of the strains were eliminated from subsequent calculations. Cluster analysis was computed into the phenogram by using average linkage (UPGMA) clustering method (Sneath and Sokal 1973). The phenotypic properties of all studied strains were described by Sahin (2002). The minimum number of characters required to differentiate between the phena was selected from the 48 unit characters used to define clusters. From the several separation indices provided by the CHARSEP program (Sneath 1979), the VSP (variance of separation potential) index was chosen to find the best differentiating characters. Sneath (1979) recommends selection of characters with a VSP > 25 % where possible. A part of the data derived from the first set of cultures was used for a training set of ANN.

Supervised artificial neural network. Since supervised learning relies on a particularly sure and reliable training set, it has been decided to use for the learning set numerical phenetic data coming from type strains present in the collection. The numerical phenetic data of the other strains presented in the collection were used for the test set.

A common structure of ANN neuron is shown in Fig. 1. An ANN neuron is a simple work element and has a local memory. A neuron takes a multidimensional input and then delivers it to the other neurons according to their weights. This gives a scalar result at the output of a neuron. The neuron has only one output but the number of the outputs can be increased. The program used in this paper was written in the C^{++} language and run on an IBM compatible PC having Pentium CPU.

Multilayered feed-forward network with backpropagation algorithm (also known as multilayer perceptrons) was used (Chauvin and Rumelhart 1995; Karlik *et al.* 1998). The ANN topology was designed by using the trial and error method (Rumelhart *et al.* 1986). The ANN structure is in the form of 60 : 70 : 70 : 70 : 1. Several hierarchical ANNs were tested and the best results were obtained using 3 hidden layers with 70 units in each hidden layer. Sigmoid function was used as a threshold function, since its derivative could be taken and its magnitude increased monotonically.

Numerical phenetic data having high separation potential of 14 test strains were used as inputs for the ANN. All of 15 separating unit characters were used: pigment production, urease, nitrate reduction, D-fructose, D-mannose, D-ribose, formate, acetate, DL-lactate, malonate, (+)-tartrate, citrate, phenol, glycine and L-serine utilization as a source of carbon and energy. Distance values between two strains were obtained from the output layer.

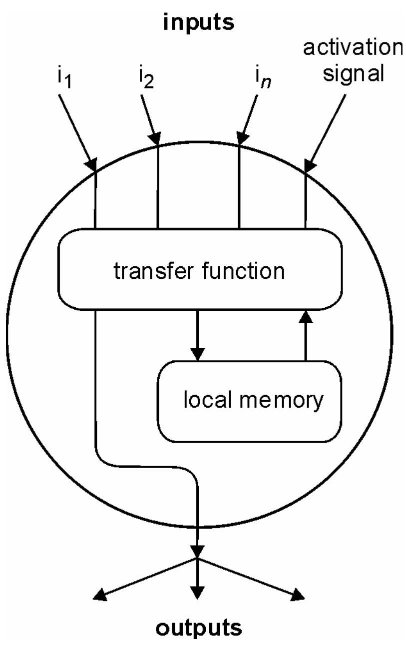

Fig. 1. The basic building block of artificial neural networks.

Taxonomical similarity values between two strains were obtained from the output layer. In this ANN architecture, learning speed and rate was given to program as 0.9 and 0.7, respectively. In the training phase, 4000 iterations were performed.

RESULTS AND DISCUSSION

In this ANN approach study, 496 normalized data obtained from phenotypic tests were used, 320 data of the phenotypic matrix being used in the training set. Then, trained ANN architecture was tested with 176 data but these data were not used in training. Mean square error (MSE) *vs.* iterations-numbers-graphic is shown in Fig. 2 for the training set phase (from Fig. 2, above the 2000 iterations, the MSE approximates zero). In addition, mean error from the test phase is approximately 1 %. The change of error according to test data is shown in Fig. 3.

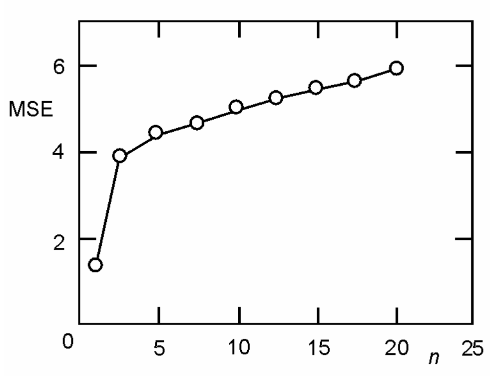

Fig. 2. Learning phase mean square error (MSE) graphic; n – number of iterations (×100).

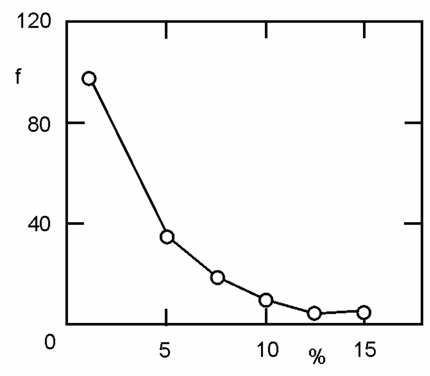

Fig. 3. Percentage of correct results on test data; f – frequency of test data.

The neural network correctly identified the oxalotrophic bacteria. The ANN approach showed a similar correlation to DNA homology based phenogram of oxalate-utilizing bacteria reported previously by Tamer *et al.* (1993). Results from this study were examined in Fig. 4 from the phenograms; three main clusters were obtained in both genomic and ANN approach.

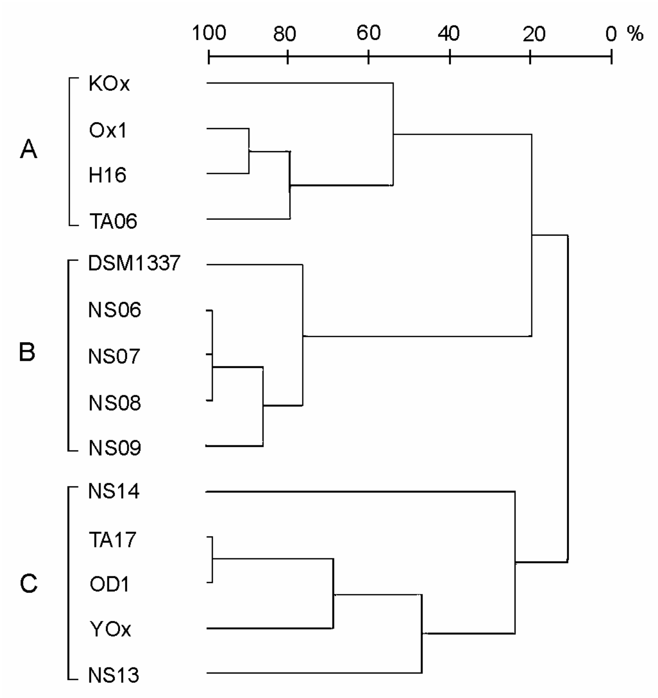

Fig. 4. Phenogram based on the numerical phenetic data from the second data set (data were analyzed using ANN); scale shows the similarity values (in %); A–C – clusters.

Cluster A contains nonpigmented strains and is rather heterogeneous. The taxonomic status of this cluster was discussed by Jenni *et al.* (1988). Cluster B contains pink-pigmented strains belonging to the genus *Methylobacterium*. Cluster C contains yellow-pigmented strains belonging to the genera *Oxalicibacterium* and *Xanthobacter*.

Several taxonomic ANN studies were performrd on different physiological groups of bacteria (Duerden *et al.* 1989; Rataj *et al.* 1991; Kennedy *et al.* 1993; Goodacre *et al.* 1996, 1998; Giacomini *et al.* 1997, 2000; Mariey *et al.* 2001). It is very encouraging that the ANN was able to detect characteristic patterns of the test strains.

One of the significant differences in this study compared with similar studies is the number of hidden neurons used in the ANN design. It would be expected that a design with fewer hidden neurons would result in networks capable of correctly classifying strains to a higher degree than those with a larger number of hidden neurons. However, the data presented in Figs 2 and 3 contradict this notion; the best results were obtained using 3 hidden layers with 70 units in each hidden layer. Additionally, other studies have attempted the classification (or identification) of microorganisms through a variety of methods and provided successful results but such networks had less than 20 input neurons (Simpson *et al.* 1992; Carson *et al.* 1995) which is considerably less than the 60 input neurons utilized in this study.

For reliable bacterial classification and identification a polyphasic approach should be recommended (Vandamme *et al.* 1996; Zahran *et al.* 2003). A polyphasic approach includes a set of different methods, from genotypic to phenotypic chemotaxonomic methods. The combination of numerical phenetic data and ANN provides practical means for identifying large numbers of isolates for ecological and industrial purposes. To obtain numerical phenetic data by ANN is more economical and time-saving than other methods.

REFERENCES

BASSALIK K.: Über die Verarbeitung der Oxalsaüre durch *Bacillus extorquens* n.sp. *Jahrb.Wiss.Bot.* **53**, 255–302 (1913).

BOUSFIELD I.J., GREEN P.N.: Reclassification of bacteria of the genus *Protomonas* URAKAMI and KOMAGATA 1984 in the genus *Methylobacterium* (PATT, COLE and HANSON) emend. GREEN and BOUSFIELD 1983. *Internat.J.Syst.Bacteriol.* **35**, 209 (1985).

CARSON C.A., KELLER J.M., MCADOO K.K., WANG D., HIGGINS B., BAILEY C.W., THORNE J.G., PAYNE B.J., SKALA M., HAHN A.W.: *Escherichia coli* O157:H7 restriction pattern recognition by artificial neural network. *J.Clin.Microbiol.* **33**, 2894–2898 (1995).

CHAUVIN Y., RUMELHART D.E.: *Backpropagation: Theory, Architectures, and Applications*. Lawrence Erlbaum Associates, Hillsdale (USA) 1995.

DUERDEN B.I., ELEY A., GOODWIN L., MAGEE J.T., HINDMARCH J.M., BENNET K.W.: A comparison of *Bacteroides ureolyticus* isolates from different clinical sources. *J.Med.Microbiol.* **29**, 63–73 (1989).

GIACOMINI M., RUGGIERO C., BERTONE S., CALEGARI L.: Artificial neural network identification of heterotrophic marine bacteria based on their fatty-acid composition. *IEEE T.Bio-Med.Eng.* **44**, 1185–1191 (1997).

GIACOMINI M., RUGGIERO C., CALEGARI L., BERTONE S.: Artificial neural network based identification of environmental bacteria by gas-chromatographic and electrophoretic data. *J.Microbiol.Meth.* **43**, 45–54 (2000).

GOODACRE R., TIMMINS E.M., ROONEY P.J., ROWLAND J.J., KELL D.B.: Rapid identification of *Streptococcus* and *Enterococcus* species using diffuse reflectance-absorbance Fourier transform infrared spectroscopy and artificial neural networks. *FEMS Microbiol.Lett.* **140**, 233–239 (1996).

GOODACRE R., TIMMINS E.M., BURTON R., KADERBHAI N., WOODWARD A.M., KELL D.B., ROONEY P.J.: Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiology* **144**, 1157–1170 (1998).

JAYASURIYA G.C.N.: The isolation and characteristics of an oxalate-decomposing organism. *J.Gen.Microbiol.* **12**, 419–428 (1955).

JENNI B., REALINI L., ARAGNO M., TAMER A.U.: Taxonomy of non H_2-lithotrophic, oxalate utilizing bacteria related to *Alcaligenes eutrophus*. *Syst.Appl.Microbiol.* **10**, 126–133 (1988).

KARLIK B., OZKAYA E., AYDIN S., PAKDEMIRLI M.: Vibrations of a beam-mass systems using artificial neural networks. *Computers & Structures* **69**, 339–347 (1998).

KENNEDY M.J., THAKUR M.S.: The use of neural networks to aid in microorganism identification: a case study of *Haemophilus* species identification. *Antonie van Leeuwenhoek* **63**, 35–38 (1993).

KESTERS K.: Numerical methods in the classification of bacteria by protein electrophoresis, pp. 337–368 in M. Goodfellow, D. Jones (Eds): *Computer-Assisted Bacterial Systematics*. Academic Press, London 1985.

KHAMBATA S.R., BHAT J.V.: Studies on a new oxalate decomposing bacterium, *Pseudomonas oxalaticus*. *J.Bacteriol.* **66**, 505–507 (1953).

MARIEY L., SIGNOLLE J.P., AMIEL C., TRAVERT J.: Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics. *Vib.Spectrosc.* **26**, 151–159 (2001).

MORRIS C.W., BODDY L., ALLMAN R.: Identification of basidiomycete spores by neural network analysis of flow cytometry data. *Mycol.Res.* **96**, 697–701 (1992).

RATAJ T., SCHINDLER J.: Identification of bacteria by a multilayer neural network. *Binary* **3**, 159–164 (1991).

RUMELHART D.E., HINTON G.E., WILLIAMS R.J.: Learning internal representations by error propagation, pp. 318–362 in D.E. Rumelhart, J.L. McClelland (Eds): *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Vol. 1*. MIT Press, Cambridge (USA) 1986.

SAHIN N.: Oxalotrophic bacteria. *Res.Microbiol.* **154**, 399–407 (2003).

SAHIN N.: Isolation and characterization of mesophilic, oxalate-degrading *Streptomyces* from plant rhizosphere and forest soils. *Naturwissenschaften* **91**, 498–502 (2004).

SAHIN N., ISIK K., TAMER A.U., GOODFELLOW M.: Taxonomic position of "*Pseudomonas oxalaticus*" strain Ox1^T (DSM 1105^T) (KHAMBATA and BHAT 1953) and its description in the genus *Ralstonia* as *Ralstonia oxalatica* comb.nov. *Syst.Appl.Microbiol.* **23**, 206–209 (2000).

SAHIN N., GOKLER I., TAMER A.U.: Isolation, characterization and numerical taxonomy of novel oxalate-oxidizing bacteria. *J.Microbiol.* **40**, 109–118 (2002).

SIMPSON P.K.: *Artificial Neural System*. Pergamon Press, Oxford (UK) 1990.

SIMPSON R., WILLIAMS R., ELLIS R., CULVERHOUSE P.: Biological pattern recognition by neural networks. *Mar.Ecol.Prog.Ser.* **79**, 303–308 (1992).

SNEATH P.H.A.: BASIC program for character separation indices from an identification matrix of percent positive characters. *Comput. Geosci.* **6**, 21–26 (1979).

SNEATH P.H.A., SOKAL R.R.: Numerical taxonomy: the principles and practice of numerical classification. W.H. Freeman & Co., San Francisco (USA) 1973.

SNEATH P.H.A.: Numerical taxonomy, pp. 5–7 in N.R. Krieg, J.G. Holt (Eds): *Bergey's Manual of Systematic Bacteriology, Vol. 1*. Williams & Wilkins, Baltimore–London 1984.

TAMER A.U., ARAGNO M.: Isolement, caractérisation et essai d'identification de bactéries capables d'utiliser l'oxalate comme seule source de carbon et d'énergie. *Bull.Soc.Neuchatel Sci.Nat.* **103**, 91–104 (1980).

TAMER A.U., ARAGNO M., JENNI B.: Genomic relation among nine strains of oxalate utilizing bacteria. *Turk.J.Biol.* **17**, 115–119 (1993).

TAMER A.U., ARAGNO M., SAHIN N.: Isolation and characterization of a new type of aerobic, oxalic acid utilizing bacteria, and proposal of *Oxalicibacterium flavum* gen.nov., sp.nov. *Syst.Appl.Microbiol.* **25**, 513–519 (2002).

VANDAMME P., POT B., GILLIS M., DE VOS P., KERSTERS K., SWINGS J.: Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microb.Rev.* **60**, 407–438 (1996).

WIDROW B.: 30 years of adaptive neural networks: perception, madaline, and backpropagation. *Proc.IEEE* **78**, 1415–1441 (1990).

ZAHRAN H.H., ABDEL-FATTAH M., AHMAD M.S., ZAKY A.Y.: Polyphasic taxonomy of symbiotic rhizobia from wild leguminous plants growing in Egypt. *Folia Microbiol.* **48**, 510–520 (2003).