Journal of Applied Microbiology

Journal of Applied Microbiology ISSN 1364-5072

ORIGINAL ARTICLE

Interpreting the SDS-PAGE protein patterns with self-organizing maps: application for the characterization of mosquito-pathogenic *Bacillus* strains

M.A. Sreshty, S. Misra and U.S.N. Murty

Biology Division, Indian Institute of Chemical Technology (IICT), Hyderabad, India

Keywords
Bacillus sphaericus, *Bacillus thuringiensis* ssp. *israelensis*, characterization, SDS-PAGE, self-organizing map.

Correspondence
Upadyayula S.N. Murty, Biology Division, Indian Institute of chemical Technology, Hyderabad 500 607, Andhra Pradesh, India.
E-mail: usnmurty@iict.res.in

2010/1323: received 30 July 2010, revised 10 September 2010 and accepted 21 September 2010

doi:10.1111/j.1365-2672.2010.04880.x

Abstract

Aims: To present the pairwise comparison of potential mosquito-pathogenic *Bacillus* strains based on their SDS-PAGE protein patterns and to evaluate their characteristic toxicity patterns.

Methods and Results: In this work, 20 *Bacillus* strains were subjected to qualitative toxicity tests against *Aedes aegypti* and *Culex quinquefasciatus* larvae. The selected strains were then characterized by SDS-PAGE protein profiles. The highly heterogeneous multiple protein components of protein patterns were analysed using self-organizing map (SOM), a 'visualization and clustering' tool. Members of mosquitocidal *Bacillus* species were classified in four distinct clusters, and then toxicity patterns were examined. Cluster (1, 1) comprised of three highly toxic strains of *Bacillus sphaericus*: SPH88, 1593 and KSD-4; cluster (1, 2) consisted of two *B. sphaericus* strains: SSII-1 and Bsp-R that showed weak larvicidal activity; cluster (2, 1) constituted two *B. sphaericus* strains: WHO2297 and ISPC-5 that possessed moderate toxicity; and cluster (2, 2) contained four *B. thuringiensis* ssp. *israelensis* strains: ONR-60A, HD500, IPS70 and IPS82 belonging to serotype H14 but exhibited moderate to high mosquito larvicidal toxicity.

Conclusions: SOM served as a colour-coded alternate for easy visualization of similarities or dissimilarities between the strains even at the infra subspecies level. Furthermore, characteristic toxicity patterns of *Bacillus* strains of different clusters were determined.

Significance and Impact of the Study: Analysis of electrophoretic protein patterns using SOM provides a better insight into the inter-relationships of bacterial strains through similarity-based clustering and pairwise comparison of two strains.

Introduction

Mosquitoes transmit a myriad of human diseases by acting as vectors of important pathogens, including the aetiological agents of malaria, filariasis, dengue, yellow fever, encephalitis and chikungunya. For efficient management of mosquito populations, World Health Organization (WHO) has recommended the use of bioinsecticides based on *Bacillus thuringiensis* ssp. *israelensis* and *Bacillus sphaericus* in control programmes (World Health Organization 1985). For this, several indigenous strains were isolated from different geographical regions and screened for potential larvicidal activity (Manonmani *et al.* 1990; Khetan 2001). Because of high genetic variability existing between the *Bacillus* strains of same or different serotypes, disparity in toxicity is also noted. This necessitates a thorough identification and classification of the strains.

As cellular protein profiles provide second-level information for a cell, characterization based on electrophoretic protein patterns of SDS-PAGE correlated closely with the genotyping results that suggest that it could be an effective method for rapid bacterial classification (Vauterin *et al.*

1990; Niemi *et al.* 1993; Berber 2004). Previous studies based on SDS-PAGE combined with computerized analysis of protein profiles provided an effective approach to investigate the taxonomic relationships among many bacterial species (Kersters 1985; Costas 1992). Conventional methods involved the numerical analysis of electrophoretic protein patterns and visualized in the form of dendrograms to describe the taxonomical relationships (Costas 1992; Höfling *et al.* 2001; Berber 2004), but pairwise comparison based on similarities was not performed so far.

Computer technology has simplified the complexity in analysing scientific data providing new dimensions and easy understanding, which was otherwise impossible with numerical data. In this arena, a new approach is used to visualize the scientific data as colour-coded images that undergo qualitative changes to convey information for better pattern recognition (DeFanti and Brown 1999). Many times, visualization provides a better understanding of the results and especially in studies for inferring the inter-relationships. 'self-organizing map' (SOM) is an invaluable data mining tool. SOM (or Kohonen map) is an unsupervised learning process based on similarity comparisons in a continuous space, which results in a system that associates similar inputs close to each other in the two-dimensional grid called the map (Kohonen 2001). SOM processes complex data sets to simple, readily explainable and in an easily visualized format, where similarity matching plays an important role. So far, SOM was successfully implemented in different biomedical fields for disease diagnosis (Papadimitriou *et al.* 2001), study of genome signatures (Gatherer 2007), interpreting patterns of gene expression (Tamayo *et al.* 1999), blood plasma lipoprotein lipid profiling (Kaartinen *et al.* 1998) and epidemiological studies (Murty *et al.* 2008), and study of phylogeny is no exception (Andrade *et al.* 1997; Banerjee *et al.* 2009).

Thus, the aim of this work is to study the relative similarities or dissimilarities by relative comparison of the protein patterns of *B. thuringiensis* ssp. *israelensis* and *B. sphaericus* strains using a 'visualization and clustering' tool (SOM) for discrimination and deciphering their inter-relationships. For this quest, mosquito-pathogenic *Bacillus* strains were screened to study their phylogeny from SDS-PAGE protein patterns by visualizing their degree of resemblances. Furthermore, the activity profiles of these *Bacillus* strains were also investigated.

Materials and methods

Bacillus strains

Twenty *Bacillus* strains were obtained from various culture collections for this study. Of which, one *B. thuringiensis* ssp. *israelensis* (Acc no: 869) and five *B. sphaericus* (Acc nos: 511, 1486, 1914, 3047 and 3672) strains were obtained from Microbial Type Culture Collection (MTCC) of IMTECH, Chandigarh, India; three *B. thuringiensis* ssp. *israelensis* (Acc nos: 2513, 2978 and 5132) strains were obtained from National Collection of Industrial Microorganisms (NCIM), Pune, India; four *B. thuringiensis* ssp. *israelensis* (Acc nos: 4Q1, 4Q2, 4Q3 and 4Q4) and five *B. sphaericus* (Acc nos: 13A1, 13A2, 13A3, 13A4 and 13A69) strains were obtained from *Bacillus* Genetic Stock Center (BGSC), US. The standard strains of *B. thuringiensis* ssp. *israelensis* (Acc no: 1884) and *B. sphaericus* (Acc no: 2362) were received from Institut Pasteur, Paris, France.

Culture conditions

All strains were maintained and cultured in nutrient yeast salt mineral medium (NYSM) (containing 5 g glucose (bacteriological), 5 g peptone, 5 g NaCl, 3 g beef extract, 5 g yeast extract, 203 mg $MgCl_2$, 10 mg $MnCl_2$ and 103 mg $CaCl_2$) (Geetha *et al.* 2007). Seed cultures were prepared by inoculating 20 ml of broth with a loopful of culture and incubated in a rotary shaker at 35 ± 2°C, 150 rev min^{-1} for 6 h. The seed culture was transferred to 1000 ml medium in a 2-l Erlenmeyer flask (2% v/v). Culture conditions were maintained in a shaking incubator (Lab Tech) for 48–72 h until complete sporulation was achieved. The spores were harvested by centrifugation in a Microfuge 22R centrifuge (Beckman Coulter Inc., Fullerton, CA) at 3200 ***g*** for 15 min. The supernatant was discarded and the pellet was washed with 0·1 mol l^{-1} NaCl. Subsequently, the wet biomass was lyophilized for dry powder preparation. Stock solutions were prepared using sterile distilled water at 5 mg ml^{-1} concentration.

Preliminary bioassays

Qualitative bioassays were performed to determine the pathogenicity of *Bacillus* strains to *A. aegypti* larvae and *C. quinquefasciatus* larvae. Bioassays were carried out as previously described by Manonmani *et al.* (1990). One millilitre of stock suspension from each test strain was added to 20 III instars placed in 100 ml distilled water in a 250-ml plastic cup. All test treatments were carried out in triplicates along with appropriate controls. Mortality was scored 48 h post-treatment. The strains resulting in 100% mortality were considered for further studies.

Preparation of alkali soluble proteins

Only 11 strains have been chosen for electrophoresis studies. For the analysis of protein profile, the soluble toxin proteins of *Bacillus* strains were prepared by alkali

Journal of Applied Microbiology **110**, 239–247

solubilization procedure (De Leon and Ibarra 1995). Samples (1·5 ml) of spore crystal complex were suspended in 0·1 mol l^{-1} Na_2CO_3 buffer (pH 10). The suspensions were incubated for 3 h at 37°C under continuous shaking at 200 rev min^{-1}, followed by centrifugation at 13 000 ***g*** for 30 min. The proteins were then precipitated by adding cold 100% acetone to the sample to obtain a final concentration of 80% (v/v) (Olson and Markwell 2007).

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

Protein profiling was performed using SDS-PAGE on 8% resolving gel, following the procedure of Laemmli (1970). Fifteen microlitres of the prepared sample was separated by electrophoresis at current of 50 mA and 120 V. The broad range protein molecular weight marker (29–205 kDa) (PMWH from Bangalore Genei, India) was used to determine the protein weights. Proteins were fixed by immersing the gels in an aqueous of 10% (v/v) trichloracetic acid solution for 1 h and stained overnight in Coomassie Blue stain [0·25% (w/v) Coomassie Blue R-250, 50% (v/v) methanol and 10% (v/v) acetic acid]. The excess stain was washed out by destaining the gel with a solution of 25% (v/v) methanol and 10% (v/v) acetic acid. The gel was documented using a BioVis geldoc equipped with a Kodak Camera. Protein profiles were analysed using BioVis Gel 1D (Expert Vision Pvt, Ltd, Mumbai, India). The protein molecular weights and Rf (Relative front) values of different protein bands were noted.

Analysis of protein patterns

The molecular weight of proteins in electrophoregrams and coefficients of similarity (Jaccard coefficient) were calculated as stated by Ghazi *et al.* (2009). The coefficient of similarity is the match value between the two strains, and it ranges in value from zero to unity.

$$Sab = Ns/Ns + Nd$$

Sab: coefficients of similarity between two strains (a) and (b).

Ns: Number of proteins having similar molecular weight between two strains (a) and (b) in electrophoregram.

Nd: Number of proteins having different molecular weights between two strains (a) and (b) in electrophoregram.

Data mining – self-organizing maps

SOM v1.0, a data mining tool based on artificial neural networks, was used for constructing the SOM. In SOM, the neurons are organized in a lattice, typically as one- or two-dimensional array, which is placed in the input space and is spanned over the input distribution. It is feasible to achieve a map of input space where imminence between units or clusters in the map represents closeness of the input data using a two-dimensional SOM network. Processing units in the SOM lattice are associated with weights of the same dimension of the input data. Using the weights of each processing unit as a set of coordinates, the lattice can be positioned in the input space. Throughout the learning stage, the weights of the units change their position and 'move' towards the input points. Progress of the movement acquires a gradually slower pace, and network is almost 'frozen' in the input space at the end of the learning stage. On the completion of the learning stage, the inputs can be associated with the nearest network unit. On visualization, the inputs can be associated with each cell on the map. Cells that evidently contain analogous entities can be considered as a cluster on the map. These clusters are generated during the learning phase without any prior information. The main application of the SOM is the visualization of high-dimensional data in a two-dimensional way and the construction of abstractions akin to other clustering techniques.

Steps involved in the algorithm

1 Initialization: Randomly initialize a weight vector (W_i) for each neuron I $W_i = [w_{i1}; w_{i2};...; w_i\ n]$; n denotes the dimension of input data.
2 Sampling: Select an input vector X=[x1, x2,..., xn]
3 Similarity matching: Find the winning neuron whose weight vector best matches with the input vector $j(t)$= arg min $\{||X - W_i||\}$
4 Updating: Update weight vector of winning neuron, such that it becomes still closer to the input vector. Also, update weight vectors of neighbouring neurons-the further the neighbour, the lesser the degree of change.

$$W_i(t+1) = W_i(t) + \alpha(t)Xhij(t)X[X(t) - W_i(t)]$$

$\alpha(t)$: learning rate that decreases with time t, $0 < \alpha(t) = 1$

$$hij(t) = \exp(-||rj - ri||2/2, X\,\sigma(t)2)$$

$||rj - ri||2$ = distance between winning neuron and other neurons
$\sigma(t)$ = neighbourhood radius that decreases with time t.
5 Continuation: Repeat steps 2–4 until there is no change in weight vectors or up to certain number of iterations. For each input vector, find the best matching weight vector and allot the input vector to the corresponding neuron/cluster.

The coefficient of similarities obtained from the electrophoretic protein pattern served as input parameters for constructing a grid map. The dimensions of the SOM are set to x columns by y rows, weight vectors initializing for each of the xy cells were selected at random from the entire set of match values. Clusters were generated on 2×2 grid using a learning constant of 0·01 for 10 000 iterations for studying the magnitude of similarity between the strains.

Quantitative determination of larvicidal toxicity

Insecticidal activity of *Bacillus* strains was determined on III instar larvae of *A. aegypti* and *C. quinquefasciatus* mosquito species following the guidelines of World Health Organization (2005). Different concentrations resulting in mortality between 10 and 99% were recorded. Each concentration was tested in triplicates along with appropriate controls on three days. Larval mortality was recorded after 24 and 48 h post-treatment for *B. thuringiensis* ssp. *israelensis* and *B. sphaericus*, respectively. Control mortality was corrected using Abbott's formula (Abbott 1925). The mean LC_{50} and LC_{90} for all the *Bacillus* strains were estimated using Probit analysis software (IICT ver. 1.0).

Results

Initially, selective bioassays were performed to determine the pathogenicity of *Bacillus* strains to *A. aegypti* and *C. quinquefasciatus* mosquito larvae. Four *B. thuringiensis* ssp. *israelensis* strains (ONR60A, HD500, IPS70 and IPS82) and six *B. sphaericus* strains (1593, SSII-1, WHO2297, ISPC-5, KSD-4 and SPH88) showed toxicity against *A. aegypti* larvae, whereas three *B. thuringiensis* ssp. *israelensis* strains (ONR60A, HD500 and IPS82) and seven *B. sphaericus* strains (1593, SSII-1, *Bsp*-R, ISPC-5, WHO2297, KSD-4 and SPH88) were toxic to *C. quinquefasciatus* species. The details of the selected *Bacillus* strains used in this study are listed in Table 1.

SDS-PAGE of eleven *Bacillus* strains under study produced Coomassie blue staining protein patterns containing up to 20 distinct bands (Fig. 1). The components ranged in molecular weight from about 18 to 214 kDa in 8% homogeneous separating gels. Each species produced a characteristic band pattern easily identified by the position of the major protein bands. The patterns were reproducible for the replicates of the strains in different gels, with electrophoretic mobilities do not showing variations greater than ± 0·02, for Rf values. Similarities and differences between the patterns with respect to the number and electrophoretic mobility of protein fractions are shown. The isolates of each species generated different patterns and, in most instances, the differences between the isolates were clearly evident. The electrophoretic gel patterns of *B. thuringiensis* ssp. *israelensis* strains showed 9 or 10 protein bands ranging between 18 and 168 kDa, with six or seven major bands (18, 28, 43, 54, 63, 78 and 137 kDa). On other hand, 8–10 proteins were observed in each *B. sphaericus* strain with the presence or absence of some of the proteins belonging to the molecular weights of approximately, 18, 28, 43, 54, 78, 89, 98, 107, 113, 123, 137, 144, 158, 177, 189, 191 and 214 kDa bands.

Usually, electrophoretic profiles were studied in the form of dendrograms, but here we have employed SOM to visualize the degree of resemblances between the strains through grid map. The SOM grid shows the pairwise comparison for interpreting the similarities and dissimilarities between different larvicidal *Bacillus* strains (Fig. 2a). In this figure, each cell at the intersection of two strains represented the degree of matching between them. The match values of protein patterns scaling

Table 1 List of selected mosquito larvicidal *Bacillus* strains used in the study

Acc. No.	Strains	*Bacillus* species	Susceptible mosquito species	Serotypes	Place of isolation	References
13A1	1593	Bs	*Aedes*, *Culex*	H5a5b	Indonesia	Priest *et al.* (1997)
13A2	SSII-1	Bs	*Aedes*, *Culex*	H2a2b	Singapore	Singer (1973)
13A3	Bsp-R	Bs	*Culex*	NA	NA	Kiss *et al.* (1977)
13A4	WHO2297	Bs	*Aedes*, *Culex*	H25	Sri Lanka	Wickremesinghe and Mendis (1980)
13A69	ISPC-5	Bs	*Aedes*, *Culex*	H26a26b	India	Menon *et al.* (1982)
3672	KSD-4	Bs	*Aedes*, *Culex*	NA	India	Dhindsa *et al.* (2002)
2362	SPH88	Bs	*Aedes*, *Culex*	H5a5b	Nigeria	Weiser (1984)
4Q1	ONR60A	Bti	*Aedes*, *Culex*	H14	Israel	Goldberg and Margalit (1977)
4Q2	HD500	Bti	*Aedes*, *Culex*	H14	Israel	Goldberg and Margalit (1977)
4Q3	IPS70	Bti	*Aedes*	H14	Israel	Goldberg and Margalit (1977)
1884	IPS82	Bti	*Aedes*, *Culex*	H14	France	de Barjac and Larget-Thiery (1984)

NA, not available; Bs, *Bacillus sphaericus*; Bti, *B. thuringiensis* ssp. *israelensis.*

Journal of Applied Microbiology **110**, 239–247

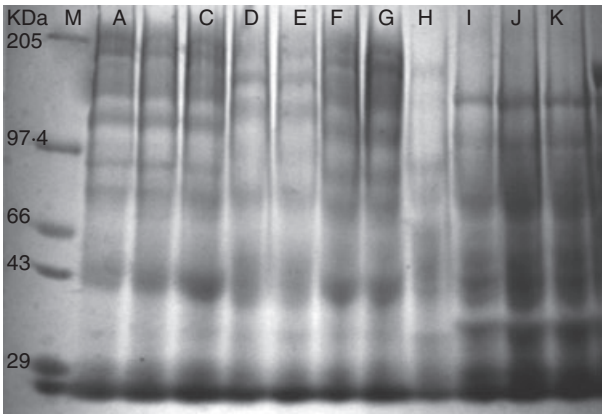

Figure 1 SDS-PAGE profiles of *Bacillus* spore proteins. Lanes M: Protein marker, A: 1593, B: SSII-1, C: Bsp-R, D: WHO2297, E: ISPC-5, F: KSD-4, G: SPH88, H: IPS82, I: ONR60A, J: HD500, K:IPS70.

between 0 and 1 represent low to high similarity between the strains. Protein patterns showing high levels of matching are indicated by red and low values by black. Likewise, all the data were visualized in a single pass that enables the study of inter-relationships. Depending on the degree of likeliness in the protein constitution, the intensity of colour in each cell of the grid varies from black to red. Thus, the magnitude of similarities between the strains was easily visualized by the colour gradation from black to red.

Classification was obtained by similarity-based clustering that resulted into four clusters. Cluster (1, 1) was formed with three *B. sphaericus* strains: 1593, KSD-4 and SPH88 (Fig. 2b); cluster (1, 2) contained two *B. sphaericus* strains: SSII-1 and Bsp-R (Fig. 2c); Fig. 2d represents the cluster (2, 1) that comprised of two *B. sphaericus* strains: WHO2297 and ISPC-5 and cluster (2, 2) constituted four *B. thuringiensis* ssp. *israelensis* strains (IPS82, ONR60A, HD500 and IPS70) (Fig. 2e).

The toxicity patterns of mosquitocidal *Bacillus* strains belonging to different clusters generated by SOM were determined through quantitative bioassays. Virulence was determined in terms of LC_{50} against *A. aegypti* and *C. quinquefasciatus* larvae. Table 2 shows the insecticidal activities of four *B. thuringiensis* ssp. *israelensis* strains (viz., IPS82, ONR60A, HD500 and IPS70) and six *B. sphaericus* strains (1593, SSII-1, WHO2297, ISPC-5, KSD-4 and SPH88) against *A. aegypti* larvae. Similarly, the larvicidal activities of three *B. thuringiensis* ssp. *israelensis* (ONR60A, HD500 and IPS82) and seven *B. sphaericus* strains (1593, SSII-1, Bsp-R, WHO2297, ISPC-5, KSD-4 and SPH88) against *C. quinquefasciatus* larvae are represented in Table 3.

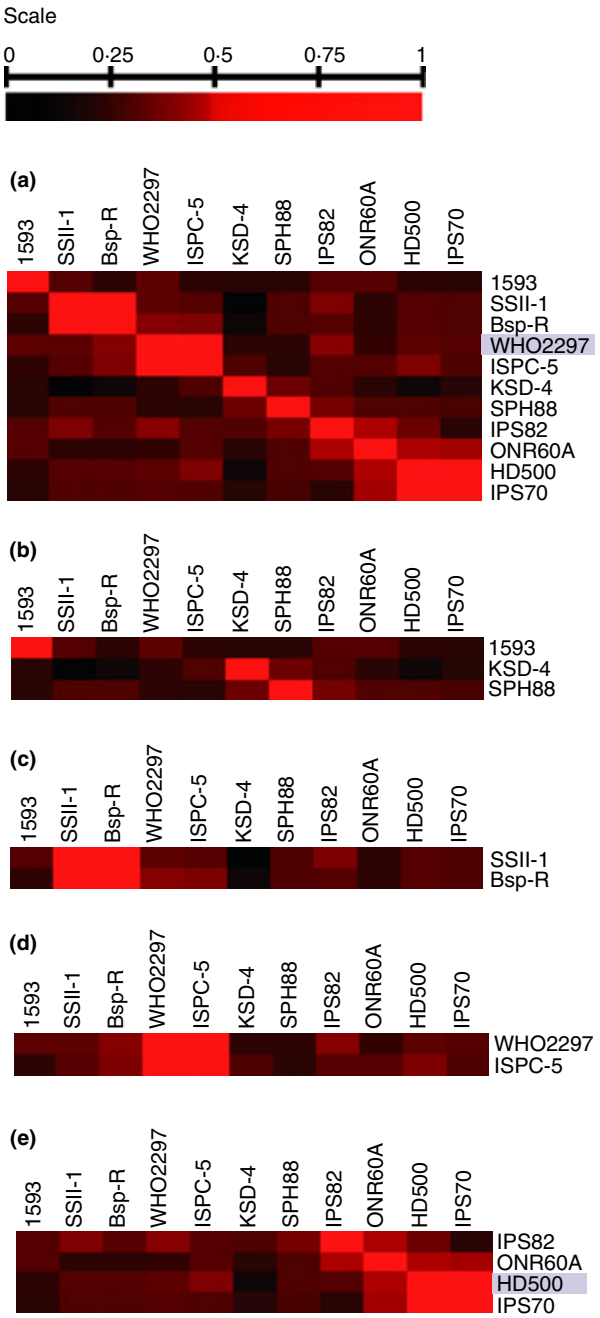

Figure 2 Self-organizing map (SOM) analysis of *Bacillus* spore proteins (a) SOM main grid, (b) Cluster (1, 1), (c) Cluster (1, 2), (d) Cluster (2, 1), (e) Cluster (2, 2).

Discussion

The growing significance for *Bacillus* biolarvicides in mosquito management programmes has encouraged the screening of a large number of isolates from different sources. Heterogeneity in toxicity is obvious within the strains of *B. thuringiensis* ssp. *israelensis* (strains belonging

Table 2 Larvicidal activities of *Bacillus* strains against *Aedes aegypti* mosquito species

Strain	*Bacillus* species	LC_{50} ± SE (fiducial limits) ($\mu g\ ml^{-1}$)	LC_{90} ± SE (fiducial limits) ($\mu g\ ml^{-1}$)	Chi-square	Slope
1593	Bs	2·53 ± 0·42 (1·69–3·37)	14·17 ± 3·35 (7·59–20·75)	5·29	1·19
SSII-1	Bs	4·20 ± 0·74 (2·73–5·66)	29·17 ± 8·56 (12·38–5·97)	5·13	0·81
WHO2297	Bs	0·68 ± 0·09 (0·49–0·86)	2·61 ± 0·56 (1·49–3·72)	1·19	2·35
ISPC-5	Bs	0·49 ± 0·07 (0·35–0·64)	2·32 ± 0·616 (1·11–3·53)	2·03	2·40
KSD-4	Bs	11·40 ± 1·89 (7·70–15·11)	70·87 ± 23·25 (25·30–116·44)	6·50	1·56
SPH88	Bs	0·075 ± 0·016 (0·042–0·107)	0·578 ± 0·158 (0·267–0·888)	2·09	4·46
ONR60A	Bti	2·48 ± 0·38 (1·86–3·56)	8·18 ± 0·42 (7·08–9·47)	3·12	9·07
HD500	Bti	0·61 ± 0·06 (0·48–0·74)	1·59 ± 0·28 (1·02–2·16)	7·92	4·83
IPS70	Bti	5·12 ± 0·70 (3·74–6·49)	19·30 ± 3·97 (11·50–27·10)	3·43	3·01
IPS82	Bti	0·013 ± 0·003 (0·006–0·020)	0·142 ± 0·049 (0·045–0·239)	3·11	3·09

Bs, *Bacillus sphaericus*; Bti, *B. thuringiensis* ssp. *israelensis*.

Table 3 Larvicidal activities of *Bacillus* strains against *Culex quinquefasciatus* mosquito species

Strains	*Bacillus* species	LC_{50} ± SE (fiducial limits) ($\mu g\ ml^{-1}$)	LC_{90} ± SE (fiducial limits) ($\mu g\ ml^{-1}$)	Chi-square	Slope
1593	Bs	0·0010 ± 0·0004 (0·0002–0·0019)	0·0798 ± 0·0485 (0·0152–0·1749)	5·33	15·67
SSII-1	Bs	1·12 ± 0·19 (0·74–1·50)	7·27 ± 2·13 (3·08–11·46)	2·36	1·37
Bsp-R	Bs	3·55 ± 0·39 (2·78–4·33)	10·34 ± 1·73 (6·95–13·74)	3·23	4·05
WHO2297	Bs	0·074 ± 0·019 (0·036–0·112)	0·850 ± 0·362 (0·141–1·560)	2·80	4·83
ISPC-5	Bs	0·54 ± 0·08 (0·37–0·71)	2·88 ± 0·74 (1·42–4·35)	1·74	2·19
KSD-4	Bs	0·0035 ± 0·0009 (0·0017–0·0053)	0·037 ± 0·014 (0·008–0·065)	1·06	13·85
SPH88	Bs	0·0005 ± 0·0001 (0·0001–0·0008)	0·0235 ± 0·0146 (0·0005–0·0522)	3·22	19·46
ONR60A	Bti	0·57 ± 0·11 (0·35–0·79)	4·23 ± 1·35 (1·58–6·89)	7·24	1·83
HD500	Bti	0·016 ± 0·009 (0·002–0·036)	11·52 ± 1·48 (9·99–14·03)	5·34	6·37
IPS82	Bti	0·0012 ± 0·0003 (0·0005–0·0020)	0·0305 ± 0·0001 (0·0038–0·0572)	1·66	16·60

Bs, *Bacillus sphaericus*; Bti, *B. thuringiensis* ssp. *israelensis*.

to the serotype H14) and *B. sphaericus* (only 6 of 48 serotypes were known to be pathogenic) because of high genetic diversity (Berry *et al.* 1993). As members of highly toxic groups have certain positive characteristics like tolerance to harsh environments, differentiating them from other strains is very essential (Yap 1990; Baumann *et al.* 1991). Thus, discrimination of the mosquito-pathogenic *Bacillus* strains and understanding their disparity in toxicity patterns is of paramount importance.

In the present study, we have observed that few proteins of molecular weight *c.* 18, 28, 43 and 78 kDa were common in all *Bacillus* strains, whereas proteins of 63 and 214 kDa were found specific to *B. thuringiensis* ssp. *israelensis* and *B. sphaericus* strains, respectively. Usually, SDS-PAGE proteins profiles were subjected to numerical analysis and visualized in the form of dendrograms to study the phylogenetic inter-relationships, but here we have postulated the use of SOM for definitive recognition of electrophoretic types, which has already proved its proficiency in classification studies in a myriad of scientific areas (Andrade *et al.* 1997; Arab *et al.* 2004; Murty *et al.* 2009).

With SOM, the level of protein pattern heterogeneity between the strains was clearly evident, and the reason for their distinct grouping was easily deciphered from the pairwise comparison of similarities or dissimilarities observed between the strains of same cluster or different clusters. Using Jaccard coefficients of similarities, SOM grid was generated representing the similarity matrix of the eleven *Bacillus* strains (Fig. 2a). As all typing methods are substantiated by fundamental traits, the characteristic toxicity patterns of the *Bacillus* strains of different clusters were studied based on their virulence to *A. aegypti* (Table 2) and *C. quinquefasciatus* (Table 3) larvae.

Eleven *Bacillus* strains were classified into four distinct clusters using SOM tool, which are discussed as follows:

Cluster (*1*, *1*): This cluster contained three *B. sphaericus* strains: 1593, KSD-4 and SPH88 that possessed five protein bands (18, 28, 43, 78 and 214 kDa) in common (Fig. 2b). The pairwise comparison of SPH88 and KSD-4 showed that they share more proteins in common than with 1593. SPH88 and KDS-4 shared nine protein bands; whereas 1593 shared six proteins with SPH88 and five

proteins with KSD-4. Cluster (1, 1) was represented by highly toxic strains. Their toxicity patterns revealed that the two strains 1593 and SPH88 (belonging to the serotype H5a5b) demonstrated high activities against *A. aegypti* and *C. quinquefasciatus* species, whereas KSD-4 (unknown serotype) exhibited low toxicity to the former larval species and higher toxicity to the latter.

Cluster (*1*, *2*): This cluster is characterized by weakly toxic strains. The two strains of *B. sphaericus* species: Bsp-R (serotype not known) and SSII-1 (H2a2b) belong to this cluster (Fig. 2c). These strains possessed similar protein constitution showing eight protein bands in common (18, 28, 43, 78, 98, 123, 146 and 192 kDa). Activity profile of SSII-1 showed comparatively low toxicity against *A. aegypti* and *C. quinquefasciatus* species, which substantiated the previous findings that the strains of serotype H2 are weakly toxic (Louis *et al.* 1984; de Barjac *et al.* 1985). Whereas Bsp-R of unknown serotype was found to be toxic to *Culex* only and nontoxic to *Aedes* sp.

Cluster (*2*, *1*): This cluster constituted WHO2297 (H25) and ISPC-5 (H26a26b) belonging to *B. sphaericus* (Fig. 2d). Both the strains shared eight protein bands in common with only one protein band of 107 kDa exceptionally present in ISPC-5. Also, the proteins of 158 and 189 kDa molecular weight were exclusively seen in this group. Both ISPC-5 and WHO2297 exhibited toxicity to both *A. aegypti* and *C. quinquefasciatus* larvae. This cluster was characterized by moderately toxic strains (Tables 2 and 3).

Cluster (*2*, *2*): As the mosquitocidal *B. thuringiensis* strains only belong to the serotype H14 (Margalit *et al.* 1983), the four *B. thuringiensis* ssp. *israelensis* strains, IPS82, ONR60A, HD500 and IPS70, of the same serotype were exclusively assembled in this cluster (Fig. 2e). These strains showed seven proteins in common (18, 28, 43, 54, 63, 78 and 137 kDa), of which a protein of 63 kDa was exceptionally found in *B. thuringiensis* ssp. *israelensis* strains. Although they belong to the same serotype (H14), variation in the toxicity patterns was observed. The toxicity profiles revealed that ONR60A, HD500 and IPS82 were highly toxic to the larvae of both mosquito species; however, IPS82 was the most toxic strain among all. On the other hand, relatively low activity was observed with IPS70 against *A. aegypti* species and found to be nontoxic to *C. quinquefasciatus* species.

Conclusion

Our work is a new paradigm in the classification studies based on electrophoretic profiles studied through data mining tool. Classification derived from SOM tool resulted in discrete clustering of members of mosquito-pathogenic *Bacillus* species and visualization of their inter-relationships through pairwise comparison, which was otherwise not possible with dendrograms. SOM effectively yielded 4 clusters that were distinct from each other and marked by their characteristic toxicity patterns. Likewise, SOM can be applied for the characterization of a large number of strains using their electrophoresis patterns. Based on the data presented here, it could be concluded that SOM serves as a colour-coded alternate to study the relative similarities of the bacterial strains as well as to differentiate the bacteria both at the species and at infra subspecific levels.

Acknowledgements

Authors are grateful to UGC for awarding the fellowship, Dr J.S. Yadav, Director IICT for his encouragement and the facilities provided. Dr J.F. Charles, Director, Institut Pasteur, and Dr Daniel R. Zeigler, Director, BGSC, for gifting the *Bacillus* strains required for this study.

References

Abbott, W.S. (1925) A method of computing the effectiveness of an insecticide. *J Econ Entomol* **18**, 265–267.

Andrade, M.A., Casari, G., Sander, C. and Valencia, A. (1997) Classification of protein families and detection of the determinant residues with an improved self organizing map. *Biol Cybern* **76**, 441–450.

Arab, A., Lek, S., Lounaci, A. and Park, Y.S. (2004) Spatial and temporal patterns of benthic invertebrate communities in an intermittent river (North Africa). *Ann Limnol* **40**, 317–327.

Banerjee, A.K., Arora, N. and Murty, U.S. (2009) Clustering and classification of anopheline spacer sequences using self organizing maps. *Internet J Genom Proteom* **4**. http://www.ispub.com/ostia/index.php?xmlFilePath=journals/ijpg/vol4n1/spacer.xml.

de Barjac, H. and Larget-Thiery, I. (1984) Characterization of IPS82 as standard for biological assay of *Bacillus thuringiensis* H-14 preparation. *Document Mimiogr* WHO/VBC/84.892, Genéve.

de Barjac, H., Larget-Thiery, I., Cosmao-Dumanoir, V. and Ripouteau, H. (1985) Serological classification of *Bacillus sphaericus* strains in relation with toxicity to mosquito larvae. *Appl Microbiol Biotechnol* **21**, 85–90.

Baumann, P., Clark, M.A., Baumann, L. and Broadwell, A.H. (1991) *Bacillus sphaericus* as a mosquito pathogen: properties of the organism and its toxin. *Microbiol Rev* **55**, 425–436.

Berber, I. (2004) Characterization of *Bacillus* species by numerical analysis of their SDS-PAGE protein profiles. *J Cell Mol Biol* **3**, 33–37.

Berry, C., Hindley, J., Ehrhardt, A.F., Grounds, T., de Souza, I. and Davidson, E.W. (1993) Genetic determinants of the

host range of the *Bacillus sphaericus* mosquito larvicidal toxins. *J Bacteriol* **175**, 510–518.

Costas, M. (1992) Classification, identification and typing of bacteria by the analysis of their one-dimensional polyacrylamide gel electrophoretic protein patterns. In *Advances in Electrophoresis* ed. Chrambach, A., Dunn, N.J. and Radola, B.J. pp. 351–408. New York/Weinheim/Cambridge: VCH Publishers.

De Leon, T. and Ibarra, J.E. (1995) Alternative bioassay technique to measure activity of cry III proteins of *Bacillus thuringiensis*. *J Econ Entomol* **88**, 1596–1601.

DeFanti, T.A. and Brown, M.D. (1999) Visualization expanding scientific and engineering research opportunities. In *Readings in Information Visualization Using Vision to Think* ed. Card, S.K., Mackinlay, J.D. and Schneiderman, B. pp. 37–56. San Diego, CA, USA: Academic Press.

Dhindsa, K.S., Sangodkar, U.M.X. and Kumar, A. (2002) Novel cost effective method of screening soils for the presence of mosquito pathogenic bacilli. *Lett Appl Microb* **35**, 457–461.

Gatherer, D. (2007) Genome signatures, self-organizing maps and higher order phylogenies: a parametric analysis. *Evol Bioinform* **3**, 211–236.

Geetha, I., Prabakaran, G., Paily, K.P., Ma, A.M. and Balaraman, K. (2007) Characterisation of three mosquitocidal *Bacillus* strains isolated from mangrove forest. *Biol Control* **42**, 34–40.

Ghazi, F., Henni, D.E., Benmechernene, Z. and Kihal, M. (2009) Phenotypic and whole cell protein analysis by SDS-PAGE for identification of dominants lactic acid bacteria isolated from algerian raw milk. *World J Dairy Food Sci* **4**, 78–87.

Goldberg, L.Y. and Margalit, J. (1977) A bacterial spore demonstrating rapid larvicidal activity against *Anopheles sergentii*, *Uranotaenia unguiculata*, *Culex univittatus*, *Aedes aegypti* and *Culex pipiens*. *Mosq News* **37**, 355–358.

Höfling, J.F., Rosa, E.A.R., Pereira, C.V., Boriollo, M.F.G. and Rodrigues, J.A.O. (2001) Differentiation and numerical analysis of oral yeasts based on SDS-PAGE profiles. *Braz J Biol* **61**, 507–516.

Kaartinen, J., Hiltunen, Y., Kovanen, P.T. and Ala-Korpela, M. (1998) Application of self organizing maps for the detection and classification of human blood plasma lipo-protein lipid profiles on the basis of 1H NMR spectros-copy data. *NMR Biomed* **11**, 168–176.

Kersters, K. (1985) Numerical methods in the classification of bacteria by protein electrophoresis. In *Computer Assisted Bacterial Systematic* ed. Goodfellow, M., Jones, D. and Priest, F.G. pp. 337–368. London: Academic Pres.

Khetan, S.K. (2001) *Microbial Pest Control*. New York: Marcel Dekker.

Kiss, A., Sain, B., Csordás-Toth, E. and Venetianer, P. (1977) A new sequence-specific endonuclease (Bsp) from *Bacillus sphaericus*. *Gene* **1**, 323–329.

Kohonen, T. (2001) *Self-Organizing Maps*. Berlin: Springer Press.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

Louis, J., Jayaraman, K. and Szulmajster, J. (1984) Biocide gene(s) and biocide activity in different strains of *Bacillus sphaericus*. *Mol Gen Genet* **195**, 23–28.

Manonmani, A.M., Hoti, S.L. and Balaraman, K. (1990) Characterization and larvicidal activity of indigenous isolates of *Bacillus sphaericus* from natural breeding habitats. *Indian J Med Res* **91**, 223–227.

Margalit, J., Zomer, E., Erel, Z. and Barak, Z. (1983) Development and application of *Bacillus thuringiensis* subsp. *israelensis* serotype H14 as an effective biological control agent against mosquitoes in Israel. *Biotechnology* **1**, 74–76.

Menon, K.K.R., Rao, A.S. and Amonkar, S.V. (1982) Isolation, identification & toxicity of a spore-bearing Bacillus (ISPC-5) from diseased mosquito larvae. *Indian J Exp Biol* **20**, 312–315.

Murty, U.S.N., Rao, M.S. and Misra, S. (2008) Prioritization of malaria endemic zones using self-organizing maps in the Manipur state of India. *Inform Health Soc Care* **33**, 170–178.

Murty, U.S.N., Banerjee, A.K. and Arora, N. (2009) An in silico approach to cluster CAM kinase protein sequences. *J Proteomics Bioinform* **2**, 97–107.

Niemi, R.M., Niemela, S.I., Bamford, D.H., Hantula, J., Hyvarinen, T., Forsten, T. and Raateland, A. (1993) Presumptive fecal *Streptococci* in environmental samples characterized by one-dimensional sodium dodecyl sulfate-polyacrilamide gel electrophoresis. *Appl Environ Microbiol* **59**, 2190–2196.

Olson, B.J.S.C. and Markwell, J. (2007) Assays for determination of protein concentration. *Curr Protoc Protein Sci* **3**, 3.4.1–3.4.29.

Papadimitriou, S., Mavroudi, S., Vladutu, L. and Bezenrianos, A. (2001) Ischemia detection with a self organizing map supplemented by supervised learning. *IEEE Trans Neural Netw* **12**, 503–515.

Priest, F.G., Eldrup, L., Zahner, V. and Carter, P. (1997) Distribution and characterization of mosquitocidal toxin genes in some strains of *Bacillus sphaericus*. *Appl Environ Microbiol* **63**, 1195–1198.

Singer, S. (1973) Insecticidal activity of recent bacterial isolates and their toxins against mosquito larvae. *Nature* **244**, 110–111.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S. and Golub, T.R. (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* **96**, 2907–2912.

Vauterin, L., Vantomme, R., Pot, B., Hoste, B., Swings, J. and Kersters, K. (1990) Taxonomic analysis of *Xhantomonas campestris pv. begonidae* and *X. campestris pv. pelargonii* by means of phythopathological, phenotypic, protein

electrophoretic and DNA hybridization methods. *Syst Appl Bacteriol* **13**, 166–167.

Weiser, J. (1984) A mosquito virulent *Bacillus sphaericus* in adult *Simulium damnosum* from Northern Nigeria. *Zbl Mikrobiol* **139**, 57–59.

Wickremesinghe, R.S.B. and Mendis, C.L. (1980) *Bacillus sphaericus* spore from Sri Lanka demonstrating rapid larvicidal activity on Culex *quinquefasciatus*. *Mosq News* **40**, 387–389.

World Health Organization. (1985) Informal consultation on the development of *Bacillus sphaericus* as a microbial larvicide. Special Programme for Research and Training in tropical Diseases. TDR/BCV/SPHAERICUS/85.3.

World Health Organization. (2005) Guidelines for laboratory and field testing of mosquito larvicides. WHO/CDS/WHOPES/GCDPP/2005.13. Geneva.

Yap, H.H. (1990) Field trials of Bacillus sphaericus for mosquito control. In *Bacterial Control of Mosquitoes and Blackflies: Biochemistry, Genetics, and Applications of Bacillus thuringiensis and Bacillus sphaericus* ed. De Barjac, H. and Sutherland, D. pp. 307–320. New Brunswick: Rutgers University Press.