EXPERIMENTAL DESIGN AND BAYESIAN NETWORKS FOR ENHANCEMENT OF DELTA-ENDOTOXIN PRODUCTION BY *Bacillus thuringiensis*

KARIM ENNOURI1,2*, RAYDA BEN AYED1,2, HANEN BEN HASSEN3, MAURA MAZZARELLO4 and ENNIO OTTAVIANI4

1Laboratory of Probability and Statistics, Faculty of Sciences of Sfax, Sfax, Tunisia
2Centre of Biotechnology of Sfax, Sfax, Tunisia
3Physics-Mathematics and Applications, Faculty of Sciences of Sfax, Sfax, Tunisia
4On AIR s.r.l., Genova, Italy

(Received: 13 December 2014; accepted: 6 June 2015)

*Bacillus thuringiensis* (*Bt*) is a Gram-positive bacterium. The entomopathogenic activity of *Bt* is related to the existence of the crystal consisting of protoxins, also called delta-endotoxins. In order to optimize and explain the production of delta-endotoxins of *Bacillus thuringiensis kurstaki*, we studied seven medium components: soybean meal, starch, KH$_2$PO$_4$, K$_2$HPO$_4$, FeSO$_4$, MnSO$_4$, and MgSO$_4$ and their relationships with the concentration of delta-endotoxins using an experimental design (Plackett–Burman design) and Bayesian networks modelling. The effects of the ingredients of the culture medium on delta-endotoxins production were estimated. The developed model showed that different medium components are important for the *Bacillus thuringiensis* fermentation. The most important factors influenced the production of delta-endotoxins are FeSO$_4$, K$_2$HPO$_4$, starch and soybean meal. Indeed, it was found that soybean meal, K$_2$HPO$_4$, KH$_2$PO$_4$ and starch also showed positive effect on the delta-endotoxins production. However, FeSO$_4$ and MnSO$_4$ expressed opposite effect. The developed model, based on Bayesian techniques, can automatically learn emerging models in data to serve in the prediction of delta-endotoxins concentrations. The constructed model in the present study implies that experimental design (Plackett–Burman design) joined with Bayesian networks method could be used for identification of effect variables on delta-endotoxins variation.

**Keywords:** *Bacillus thuringiensis*, delta-endotoxins, Plackett–Burman, Bayesian networks

*Corresponding author; E-mail: karimennouri1@gmail.com
Introduction

Bacillus thuringiensis (Bt) is a Gram-positive bacterium, producing parasporal crystals including delta-endotoxins responsible for its insecticidal potency [1] through sporulation. Upon ingestion by vulnerable insects, the crystals are dissolved in the midgut lumen. With the action of proteases, the endotoxins are changed into toxins that attach to receptors on the microvilli, causing damages to the epithelial midgut [2]. These proteins are highly toxic to a huge number of insect pests, safe to the environment, innocuous to non-target insect species. In general, conditions for the culture of B. thuringiensis are adjusted to have both high microbial biomass and delta-endotoxins concentration. A fermentation improvement plan can begin by estimating product yield as a solution to factors as activity of medium ingredients. Nutritional needs may be conducted either by the usual or mathematical approach. Mathematical methods present many advantages over usual methods being quick and decreases total number of trials [3]. In fact, Plackett–Burman design is considered as a section of a two-level factorial design and permits the probe of $n-1$ variables in at least $n$ experiments [4]. The Plackett–Burman design is suggested when more than five factors have to be studied. These designs are helpful for detecting great main effects, supposing the fact that all interactions are insignificant when compared with the few significant main effects.

Recently, Bayesian networks (BN) have become an effective device for biological network reconstruction [5, 6, 7]. BNs constitute one of the most used formalisms for calculation and forecast under uncertainty. BNs offer a useful approach to illustrate the general dependency structure of a big number of variables, therefore removing the restriction of examining the relations between variables. The purpose of the present study is to develop a model that can automatically learn emerging models in data to serve in the prediction of delta-endotoxin concentrations.

Materials and Methods

Microorganism

The used strain BUPM5 of B. thuringiensis subsp. kurstaki is known by a high delta-endotoxin production [8]. The strain was maintained by streak inoculating Luria Broth (LB) nutrient plates (g l$^{-1}$): yeast extract 5, peptone 10, NaCl 5 and agar 15, incubated at 30 °C for 24 h and stored at 48 °C for future use.
**Inocula preparation method**

Inocula preparation was previously optimised based on delta-endotoxin production yields [9]. One isolated colony was dispensed in 3 ml of LB medium and incubated overnight at 30 °C. 0.5 ml aliquots were used to inoculate 250 ml shake flasks including 50 ml of LB medium. After 6 h of incubation at 30 °C at 200 rev min⁻¹ in a rotary shaker (New Brunswick Scientific™, Edison, NJ, USA), the culture broth was used to inoculate the media. The O.D.600 was estimated using a SmartSpec™ 3000 UV-spectrophotometer (Bio-Rad Laboratories). The culture broth was used to inoculate the culture media to begin with a primary optical density at 600 nm (O.D.600) of 0.15. Considering that an O.D.600 of 1 in *B. thuringiensis* bacterium is previously estimated to represent approximately $2 \times 10^8$ CFU ml⁻¹ [10], the initial cell counts in the performed cultures are considered equal to approximately $3 \times 10^{-7}$ CFU ml⁻¹.

**Cultural conditions**

For delta-endotoxin production, *B. thuringiensis* strains were grown according to the medium compositions stated for each experiment (Table I). The 250 ml shake flasks, containing 50 ml of culture medium [11], were incubated for 72 h at 30 °C in a rotary shaker at 200 rev min⁻¹. The obtained values are the means (±SD) of three determinations of two different experiments.

<table>
<thead>
<tr>
<th>Nutrient code</th>
<th>Nutrient</th>
<th>Minimum value (–1)</th>
<th>Maximum value (+1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>KH$_2$PO$_4$</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>X2</td>
<td>K$_2$HPO$_4$</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>X3</td>
<td>MgSO$_4$</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>X4</td>
<td>MnSO$_4$</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>X5</td>
<td>FeSO$_4$</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>X6</td>
<td>Starch</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>X7</td>
<td>Soybean meal</td>
<td>20</td>
<td>30</td>
</tr>
</tbody>
</table>

Delta-endotoxin determination

Crystal proteins were dissolved before protein concentration assay as illustrated by Zouari et al. [10]. Crystal-spore pellets were washed twice with 1 M of sodium chloride solution (NaCl) and twice with distilled water. Subsequently, samples were incubated in 0.05 M NaOH (pH 12.5) for 2 h at 30 °C in a rotary shaker (200 rev min⁻¹). The soluble fractions were collected by centrifugation at 13000 rpm, for 10 min. The supernatant including the alkali-soluble insecticidal proteins was used to estimate delta-endotoxins concentration by the Bradford method [12] using bovine serum albumin (BSA) as a standard. Samples were estimated at 595 nm after 10 min.

Screening of important nutrient components

The significance of the different media components towards delta-endotoxin production was tested using Plackett–Burman experimental designs [4]. This technique is based upon the existence of Hadamard matrices, which are square matrices of order $N$ with entries at two levels, +1 and −1. These matrices are orthogonal such that for each column the number of +1 is equivalent to the number of −1. This statistical design is appropriate for screening the effect of a large number of factors in a trial and sufficient for the determination of main effects. With such experimental design, $N$ factors can be screened with only $N+1$ trials and screening up to 100 variables [13] is possible with the support of this technique. For screening aim, seven medium components have been tested using the Plackett–Burman design. The Plackett–Burman design was established on the first order model. The main effect was also estimated. Seven independent factors were evaluated in twelve trials and each factor was characterized by two levels: high and low concentrations. Statistical design analysis was carried out using the Minitab program package. The levels of independent and dependent variables evaluated in this study are listed in Table II [14]. The experimental designs, according to the Plackett–Burman method, are given in Table II.

Bayesian networks modelling

Bayesian networks (BNs) are a powerful framework for decision support under uncertain knowledge. They come out from artificial intelligence studies and constitute one of the most coherent techniques for the acquisition and the modelling of complex systems. They have been applied to a large range of prob-
problems and eventually in biology. BNs are directed acyclic graphs composed by
nodes (variables of the problem) and arcs that encode conditional probabilistic
independencies between the nodes. These graphical models are very attrac-
tive for their aptitude to explain probabilistic interactions connecting variables.
In fact, they have proven to capture causal relationships between variables and
they can show excellent forecast accuracy even with relatively small sample data
sizes [15, 16].

To achieve the mentioned objectives, Bayesian networks modelling was
used. We considered 8 nodes (7 variables representing medium components and
1 node representing delta-endotoxins concentration). Since the data are discon-
tinuous and experimental data produced using Plackett–Burman are limited, and
it is well known that the application of BN requires a lot of data for the learning
and testing procedures, our proposed methodology includes the following three
different stages for building model: 1) Data normalization; 2) Construction of an
undirected Gaussian graph; 3) Construction of Bayesian Network.

Data normalization

To normalize data from different experiences, a random sample with 116
data was created from experimental data (each new observation is the average of
30 observations).

Table II. Study of variables on delta-endotoxin productions by Plackett–Burman design

<table>
<thead>
<tr>
<th>Trial</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>X7</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+1</td>
<td>+1</td>
<td>–1</td>
<td>+1</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>2130±90</td>
</tr>
<tr>
<td>2</td>
<td>–1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>–1</td>
<td>+1</td>
<td>+1</td>
<td>3810±75</td>
</tr>
<tr>
<td>3</td>
<td>–1</td>
<td>–1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>–1</td>
<td>+1</td>
<td>2350±100</td>
</tr>
<tr>
<td>4</td>
<td>+1</td>
<td>+1</td>
<td>–1</td>
<td>+1</td>
<td>+1</td>
<td>–1</td>
<td>+1</td>
<td>3060±110</td>
</tr>
<tr>
<td>5</td>
<td>+1</td>
<td>–1</td>
<td>+1</td>
<td>+1</td>
<td>–1</td>
<td>+1</td>
<td>–1</td>
<td>2300±85</td>
</tr>
<tr>
<td>6</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>1725±75</td>
</tr>
<tr>
<td>7</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>–1</td>
<td>+1</td>
<td>+1</td>
<td>–1</td>
<td>1975±80</td>
</tr>
<tr>
<td>8</td>
<td>+1</td>
<td>–1</td>
<td>+1</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>+1</td>
<td>3490±95</td>
</tr>
<tr>
<td>9</td>
<td>+1</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>3500±105</td>
</tr>
<tr>
<td>10</td>
<td>–1</td>
<td>+1</td>
<td>+1</td>
<td>–1</td>
<td>+1</td>
<td>–1</td>
<td>–1</td>
<td>1700±80</td>
</tr>
<tr>
<td>11</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>–1</td>
<td>1240±60</td>
</tr>
<tr>
<td>12</td>
<td>–1</td>
<td>+1</td>
<td>–1</td>
<td>–1</td>
<td>–1</td>
<td>+1</td>
<td>+1</td>
<td>3890±100</td>
</tr>
</tbody>
</table>
Construction of an undirected Gaussian graph

In order to define the different conditional interactions between variables (medium components) and delta-endotoxin concentration variable, a Gaussian graphical model [17, 18] was developed. The principle of the model is based on conditional independence. Indeed, from a full connected graph, an edge is removed between two nodes if the conditional independence between two variables is accepted. The graph obtained is called conditional independence graph. The used distribution of the variables is considered as Gaussian (implying normality of each variable). Therefore, two variables are independent only if the partial correlation coefficient is null [see (I)]. We consider a graph consisting of \( n \) nodes, each node \( i \) is associated with a random variable \( X_i \). It is assumed that the vector \( X = (X_1, ..., X_n) \) is Gaussian with mean “\( m \)” and matrix empirical variance-covariance \( \Sigma = (\sigma_{ij})_{1 \leq i \leq j \leq n} \). We denote \( \Sigma^{-1} = ((w_{ij})) \) the precision matrix. The partial correlations can be defined as following:

\[
p^i_j = \text{Corr}(X_i, X_j | X_1, ..., X_k) \text{ for } 1 \leq i \leq j \leq n \quad (I).
\]

Under the normality hypothesis, \( X_i \) and \( X_j \) are conditionally dependent only if \( p^i_j \neq 0 \), it is known that \( p^i_j = - \frac{w_{ij}}{\sqrt{w_{ii} w_{jj}}} \). This formula estimates the partial correlations [19] and constructs a conditional independence graph.

Construction of Bayesian Network

A Bayesian network [20] is generated as following: two nodes \( i \) and \( j \) having a partial correlation are connected by a non-oriented edge. The orientation is determined by a heuristic method based on the following test: If \( B^i_j = \frac{w_{ij} \sigma_{ii}}{w_{ii} \sigma_{jj}} > 1 \), the arc is then oriented from \( i \) to \( j \) and if \( B^i_j = \frac{w_{ij} \sigma_{ii}}{w_{ii} \sigma_{jj}} < 1 \), the arc is then oriented from \( j \) to \( i \). The other edges with \( B^i_j = \frac{w_{ii} \sigma_{jj}}{w_{ij} \sigma_{jj}} = 1 \) remained undirected. The graph with all directed arcs constituted the Bayesian network. It is imperative to note that it does not necessarily include all nodes contained in the network [21]. The advantage of Bayesian network is to deduct all parent nodes (nutritional components) which are directly dependent on child nodes (delta-endotoxins concentration). Matlab program was used to analyze obtained data.
Results and Discussion

Medium component effects for delta-endotoxin production using Plackett–Burman design

A total of seven components were screened through twelve experimental runs. The main effects of the components in the medium for delta-endotoxins production are presented in Figure 1. The soybean meal showed the maximum positive effect on delta-endotoxin production, followed by starch, K2HPO4 and KH2PO4. The effect of FeSO4 and yeast MnSO4 were negative indicating that these components are required in the medium for delta-endotoxins production but in lower concentration than the low level. MgSO4 had neutral effect on delta-endotoxin production.

Table III shows the results of regression (coefficient, standard error, t and p-values) generated by applying experimental design technique. The significance of coefficients was estimated by Student’s t-test and p-values. The higher the level of the t-value and the lower the p-value, the more significant is the coefficient [22]. All but MgSO4 were considered significant or marginally significant having an influence on the delta-endotoxin production on the 0.1 significance level (Table III). In our study, the soybean meal, starch, FeSO4 and KH2PO4 were the most significant nutrients for improvement of delta-endotoxin production by Bt. However, KH2PO4, MnSO4 and MgSO4 were considered as insignificant. The use of high concentrations of KH2PO4, K2HPO4, starch and soybean meal and exclusion of FeSO4 and MnSO4 increased the level of delta-endotoxin production. MgSO4 has no effect on production of delta-endotoxins. Besides, the coefficient of determination (R² = 98.42%) explains the high degree of collinearity between simulated and measured data. Likewise, R² illustrates the proportion of the variance in measured data explained by the model. R² ranges from 0 to 1, with higher values indicating less error variance, and usually values greater than 0.5 are judged acceptable [23].

The statistical significance of the ratio, between the mean square variation (MS), due to regression, and the mean square residual error, was tested using analysis of variance (ANOVA) (Table IV). ANOVA is a statistical technique that subdivides the total variation of a set of data into component associated to specific sources of variation for the purpose of testing hypotheses for the modelled parameters. According to the ANOVA, the F-value was high, which indicates that variation on the response variable can be explained by the regression model. The associated p-value is used to estimate whether F is large enough to indicate statistical significance. A p-value (0.002) is lower than 0.01, which indicates that the model is considered to be statistically significant [24].
Figure 1. Effect of media components on delta-endotoxins production in submerged culture

Table III. Estimated regression coefficients for delta-endotoxin production

<table>
<thead>
<tr>
<th>Term</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2597.5</td>
<td>54.75</td>
<td>47.45</td>
<td>0.000</td>
</tr>
<tr>
<td>KH$_2$PO$_4$</td>
<td>145.0</td>
<td>54.75</td>
<td>2.65</td>
<td>0.057</td>
</tr>
<tr>
<td>K$_2$HPO$_4$</td>
<td>163.3</td>
<td>54.75</td>
<td>2.98</td>
<td>0.041</td>
</tr>
<tr>
<td>MgSO$_4$</td>
<td>6.7</td>
<td>54.75</td>
<td>0.12</td>
<td>0.909</td>
</tr>
<tr>
<td>MnSO$_4$</td>
<td>-115.8</td>
<td>54.75</td>
<td>-2.12</td>
<td>0.102</td>
</tr>
<tr>
<td>FeSO$_4$</td>
<td>-293.3</td>
<td>54.75</td>
<td>-5.36</td>
<td>0.006</td>
</tr>
<tr>
<td>Starch</td>
<td>188.3</td>
<td>54.75</td>
<td>3.44</td>
<td>0.026</td>
</tr>
<tr>
<td>Soybean meal</td>
<td>752.5</td>
<td>54.75</td>
<td>13.75</td>
<td>0.000</td>
</tr>
</tbody>
</table>

R$^2$ = 98.42%; R$^2$(pred) = 85.82%; R$^2$(adj) = 95.67%

The matching quality of the data obtained by the proposed model was evaluated by considering the correlation coefficient (R$^2$) between the experimental and modelled data. The statistical adjustment of those values generated an R$^2$ = 0.9836, revealing that the model could not explain only 1.64% of the overall effect and showing that it is a robust statistical model.
Table IV. Analysis of variance for delta-endotoxins

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Seq SS</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main effects</td>
<td>7</td>
<td>8.9478</td>
<td>8987217</td>
<td>1283888</td>
<td>35.70</td>
<td>0.002</td>
</tr>
<tr>
<td>Residual error</td>
<td>4</td>
<td>143858</td>
<td>143858</td>
<td>35965</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>9131075</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R² = 98.36%; R²(pred) = 85.23%; R²(adj) = 95.49%

Bayesian networks

BN is a graph in which nodes represent variables and arcs represent dependencies among these variables. Usually, assigning a value to a variable determines the state of the variable. The obtained graphs for delta-endotoxins production are shown in Figures 2 and 3. The network contains 8 nodes: experimental variables, namely FeSO₄, MnSO₄, MgSO₄, KH₂PO₄, K₂HPO₄, soybean meal (S.meal) and starch, and the measured variable known as delta-endotoxins concentrations (delta endo).

Figure 2 illustrates an interconnection between different nodes in the network (with a minimum partial correlation of 0.45). The network also demonstrates that the output variable (delta endo) is strongly correlated with starch,
soybean sulphate. These variables were the most influential on delta-endotoxin production.

The given result by the obtained Bayesian network (Figure 3) showed that the output node of delta-endotoxins was mainly determined by the “parents” identified as soybean meal, starch and potassium dihydrogen phosphorus. Other elements such as FeSO₄, MgSO₄, MnSO₄ affected the production of the toxin, but indirectly. We also note that iron sulfate attracts the majority of nodes suggesting that Bt consumes iron in a complexed form. These observations are in line with several studies indicating high external validity. Indeed, soybean meal, the main source of nitrogen for Bt, is considered as the most important nutrient during bacterial cell multiplication due to its participation in the construction of cellular proteins and nucleic acid synthesis. Nitrogen sources have direct control on crystal formation by stimulating the production of delta-endotoxin [10, 25]. The degradation of starch, considered as a carbon source, produced acetic acid which is directly consumed by the bacterial cell and also the precursor of PHB production. Acetates are partly converted into intracellular PHB. Subsequently, acetates and PHB are assimilated in tricarboxylic acid cycle (TCA), a fundamental energy source for cell growth [26, 27]. According to Braun [28], the absence of manganese (Mn) in the medium reduces the capacity of delta-endotoxin synthesis, suggesting the opposing direction of the arc. Similarly, Yang and Wang [29] deducted that the presence of PO₄³⁻ in the culture medium is required to ensure better turnover of metabolic pathways. In fact, Gupta et al. [30] showed that phosphate has a role of regulator on the metabolite synthesis of Bt and is furthermore implicated in stimulation of delta-endotoxin production. This is confirmed by the Bayesian network indicating a causal relationship between potassium dihydrogen phosphate and delta-endotoxin production. On the other hand, the addition of ions such as Fe³⁺, Mg²⁺, Cu²⁺, Co⁺ in the culture medium can improve the growth of Bt [31].

The effects of the parent nodes on output variable are estimated by multiple linear regression. ANOVA based on Bayesian method was carried out, the significance of each variable (parent) on delta-endotoxins production was determined using the Student’s t-test and the accepted confidence level was 95% (Table V). The regression results, t-values and p-values are given in Table V. Generally, factors having the larger t-value and smaller p-value were considered to be more significant in comparison to the factors whose t-value and p-value were different [32, 33]. These results showed that the values of KH₂PO₄, starch and soybean meal were highly significant (p < 0.05). The high value of adjusted R square (R² = 0.941) suggests that 94.1% of the delta-endotoxins production
is mainly explained by these three variables. Besides, the regression model based on the Bayesian network displayed a higher value of predicted R squared (R²(pred) = 93.94%), which indicates that the model fits well to the data and provides higher prediction power for future observations.

From Figure 4, we note an important influence of KH₂PO₄ on the delta-endotoxin production. Quantitatively, it is approximately estimated at 6 times greater than starch and 3 times higher than soybean meal.

**Conclusions**

The model developed in this study suggests that experimental design (Plackett–Burman design) coupled with Bayesian networks method could be employed for identification of effect variables on response variation. However, it has been demonstrated only for very simple systems. A wider range of parameters of nutritional conditions should be examined in order to establish the impact on response variables, such as delta-endotoxins in our case.
Acknowledgement

This study was supported by grants from the “Tunisian Ministry of Higher Education, Scientific Research and Technology”.

Conflict of Interest

No conflict of interest.

References


