AMINO ACID SEQUENCE ALIGNMENT AND ESTIMATION OF A PHYLOGENETIC TREE FOR *Bacillus thuringiensis* δ-ENDOTOXINS RETRIEVED FROM DATABANKS.

Marina Ferreira da Silva[1]; Flavia França Teixeira[1]; Leandro Lopes Loguercio[1]; Luiz Pacheco Motta[1]; Fernando Hercos Valicente[1]; Edilson Paiva[1] & José Domingos Fabris[2]. [1] – Núcleo de Biologia Aplicada – Embrapa Milho e Sorgo; [2] – Depto. de.Química – UFMG.

Key-words: *Zea mays*, *Bacillus thuringiensis*, *Spodoptera frugiperda*, computational biology, biological control

The necessity to cope with agricultural pests, with a minimum impact to the environment and human health leads us to aim at alternative methods, such as biological control and use of transgenic plants carrying appropriate resistance genes. In this sense, the commercial use of transgenic crops expressing insecticidal δ-endotoxins from *Bacillus thuringiensis* (B.t.) is a current reality in several countries, for the control of various economically important insect pests (e.g., Koziel *et al*., 1993). It has been shown that *Spodoptera frugiperda*, the most important insect pest of the maize crop in Brazil, can be efficiently controlled by B.t. δ-endotoxins (Höefte and Whiteley, 1989; Li et al.,1991; Gill et al., 1992). In this work, we attempted to organize a local database of δ-endotoxin B.t. genes (*Cry*) to serve as a faster and more efficient reference for our local prospection and characterization of B.t. genes. The basic information was searched and collected from general sequence databases such as EMBL, Genbank, Swiss-Prot, Pir and PDB, and analyzed further in terms of protein homology relationships by using the 'ClustalW' (Thompson *et al*., 1994) and 'Phylip' software packages. We noticed that the vast majority of the existing data in those databases refer mostly to primary sequences, with very little information available concerning the crystallographic structure of the proteins. Moreover, the database files commonly do not display a standard nomenclature and format that would significantly facilitate computer searches. For example, the sub-species *kurstaki* displayed the highest amount of information registered in databanks, which led us to focus our analysis on its data only. By using the software packages described above, protein alignments (Table 1) and an estimated phylogenetic tree (Figure 1) were constructed for the δ-endotoxins analyzed. The amino acid similarity shared by these B.t. toxins from *kurstaki* demonstrated that the ones with GenBank accession number. D17518 and U43605 displayed the highest scores relative to the endotoxin described by the 1ciy entry in the PDB database (Table 1). This was reflected clearly by their topological proximity in the resulting phylogram, in which a more divergent cluster with those three peptides could be identified (Figure 1). Interestingly, the entry A09398 appeared to have diverged from the 1ciy cluster earlier in the evolution, despite having shown similar levels of amino acid identity evidenced by a high alignment score (Table 1). Nevertheless, this result, the other clusters and the overall topology observed in our analysis can probably be explained by the fact that we have estimated the phylogenetic tree using full amino acid sequences in our alignments (Figure 1); therefore, a similar level of identity between proteins can be produced by a different set of amino acids within this same protein. More complete and detailed phylogenetic relationships among B.t. δ-endotoxin family proteins have been estimated recently, in which the alignments and tree reconstructions were based upon the distinct functional domains of these toxins (Bravo, 1997).

Table 1: Percentage of amino acid identity between the B.t. δ-endotoxins of the sub-specie *kurstaki* obtained from GenBank, relative to the PDB database entry corresponding to the protein 1ciy.

| Accession no.	Alignment score	Amino acid no.	Number of identical amino acids	Percentage of identity
A09398	88.00	1155	512	88.73
D00117	88.00	1155	512	88.73
D17518	100.00	1176	577	100.00
L36338	40.00	719	247	42.81
M11068	73.00	1178	427	74.00
M12661	80.00	1156	476	82.50
M23723	15.00	633	124	21.49
M23724	15.00	633	119	20.62
M37263	87.00	1155	508	88.04
M73248	73.00	1177	427	74.00
M73249	73.00	1178	428	74.18
M98544	40.00	719	248	42.98
S43931	6.00	223	37	16.59
S49203	7.00	210	20	9.52
U03552	5.00	372	65	17.47
U03553	7.00	122	14	11.48
U43605	100.00	620	577	100.00
U43606	73.00	607	428	74.18
U87793	73.00	1178	427	74.00
X59797	29.00	649	182	31.54
X62821	40.00	719	247	42.81
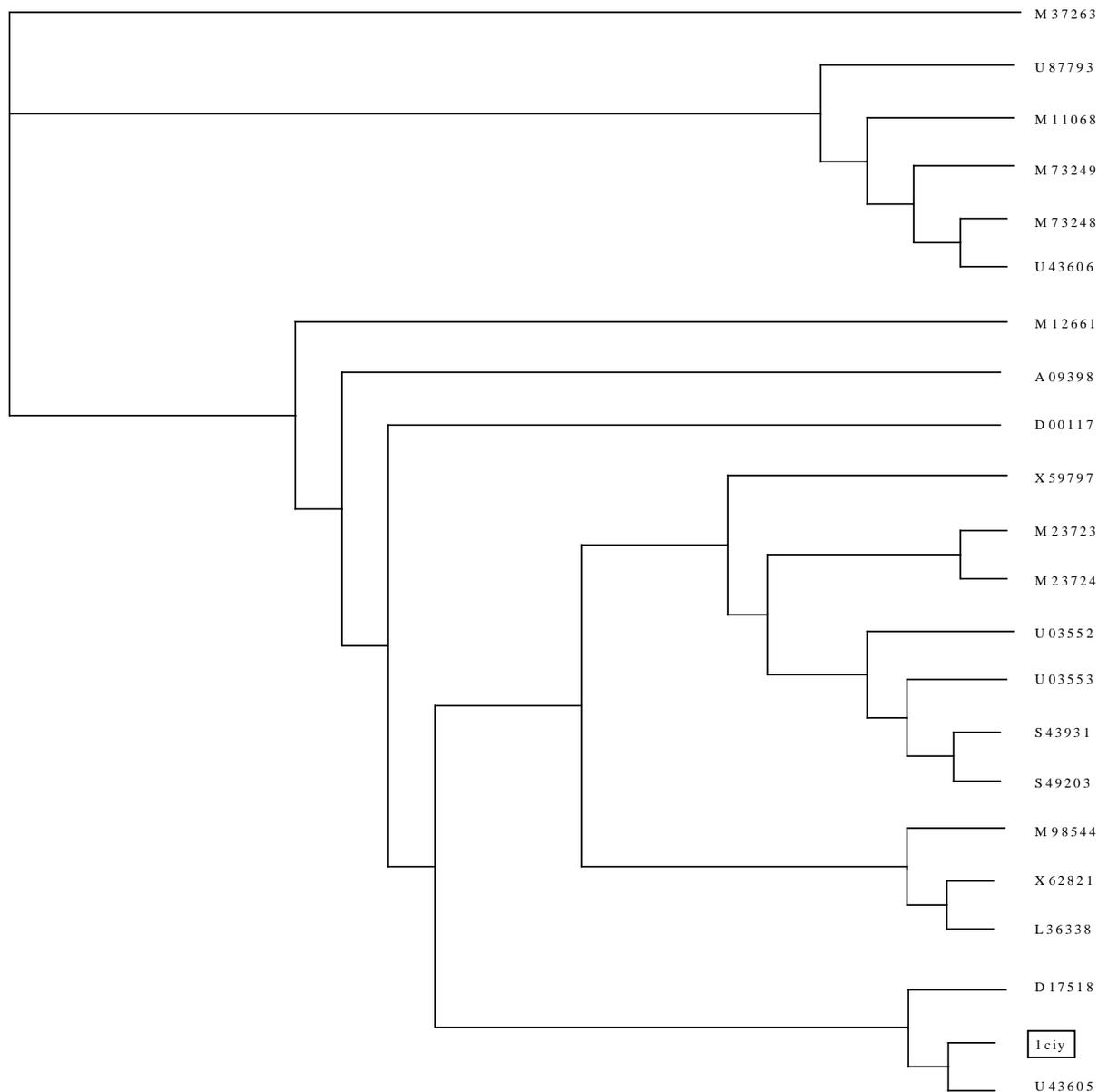
Figure 1: Estimated phylogenetic tree of δ-endotoxins from *Bacillus thuringiensis*, sub-species *kurstaki*, using multiple alignment (ClustalW program) of the whole protein (total number. of amino acids).

References:

Bravo, A. Phylogenetic relationships of *Bacillus thuringiensis* δ-endotoxin family proteins and their functional domains. *J Bacteriol*, 179: 2793-2801. 1997.

Gill, S.S.; Cowles, E.A. & Pietrantonio, P.V. The mode of action of *Bacillus thuringiensis* endotoxins. *Annu Rev Entomol*, 37: 615-636. 1992.

Höfte, H. & Whiteley, H.R. Insecticidal crystal proteins of *Bacillus thuringiensis. Microbiol Rev*, 53(2): 242-255. 1989.

Koziel, M.G.; Beland, G.L.; Bowman, C.; Carozzi, N.B.; Crenshaw, R.; Crossland, L.; Dawson, J.; Desai, N.; Hill, M.; Kadwell, S.; Launis, K.; Lewis, D.; Maddox, D.; McPherson, K.; Meghji, M.R.; Merlin, E.; Rhodes, R.; Warren, G.W.; Wright, M. and Evola, S.V. Field performance of elite transgenic maize plants expressing insecticidal protein derived from *Bacillus thuringiensis. Bio/Technology*, 11:194-200. 1993.

Li, J.; Carrol, J. and Ellar D.J. Crystal structure of insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 A° resolution. *Nature*, 353: 815-821. 1991.

Thompson, J.D.; Higgins, D.G. and Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22: 4673-4680. 1994.