

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/170511/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Palma, Leopoldo, Sauka, Diego Herman, Berry, Colin and Peralta, Cecilia 2024. Positive selection analysis of Cyt proteins from *Bacillus thuringiensis*: a conservative trend driven by negative (purifying) selection. *Toxicon* 247 , 107853. 10.1016/j.toxicon.2024.107853

Publishers page: <https://doi.org/10.1016/j.toxicon.2024.107853>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Positive selection analysis of Cyt proteins from *Bacillus thuringiensis*: a conservative trend driven by negative (purifying) selection.

Leopoldo Palma^{1,3,4*}, Diego Herman Sauka^{1,2}, Colin Berry⁵, Cecilia peralta^{1,3,4}.

¹Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ciudad Autónoma de Buenos Aires, C1425FQB, Argentina.

²Instituto Nacional de Tecnología Agropecuaria (INTA), Instituto de Microbiología y Zoología Agrícola (IMYZA), Hurlingham, Buenos Aires, 1686, Argentina.

³Instituto Multidisciplinario de Investigación y Transferencia Agroalimentaria y Biotecnológica (IMITAB), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Villa María (UNVM), Villa María, 1555, Argentina.

⁴Laboratorio de Control Biotecnológico de Plagas, Instituto BIOTECMED, Departamento de Genética, Universitat de València, Burjassot, València, 46100, Spain.

⁵Cardiff School of Biosciences, Cardiff University, Museum Avenue, Cardiff, CF10 3AX, UK.

*Corresponding author (L. Palma): leopoldo.palma@uv.es

Abstract

Bacillus thuringiensis is a Gram-positive entomopathogenic bacterium that produces different pesticidal proteins: vegetative insecticidal proteins (Vpb1/Vpa2, Vip3, and Vpb4) during vegetative growth and δ -endotoxins (Cry and Cyt) during sporulation, which accumulate into parasporal crystals. Cyt proteins are the smaller subset of δ -endotoxins targeting Diptera species. While Cry and Vip3 proteins undergo positive selection, our analysis suggests that Cyt proteins evolve following a conservative trend driven negative (purifying) selection.

Keywords: *Bacillus thuringiensis*, Cyt proteins, Diptera, molecular evolution, positive selection, negative (purifying) selection.

Bacillus thuringiensis (Bt) stands out as a Gram-positive entomopathogenic bacterium renowned for its ability to synthesize an array of insecticidal proteins throughout its life cycle. During the vegetative growth phase, Bt produces proteins such as Vpb1/Vpa2, Vip3, and Vpb4 (Chakraborty et al., 2016). Moreover, during sporulation, the bacterium synthesizes δ -endotoxins, which are categorized into Cry proteins, encompassing a diverse range of structurally distinct protein families, and a smaller group known as Cyt proteins (Palma et al., 2014). Within the Cyt protein family, subgroups include Cyt1, Cyt2, and Cyt3. The classification of Cyt proteins adheres to a four-rank nomenclature system, mirroring the rules employed for crystal and vegetative insecticidal proteins (Crickmore et al., 1998; Crickmore et al., 2018). This nomenclature system has been recently refined, transitioning to a robust structure-based approach where Cyt proteins maintain the previous Bt-toxin nomenclature (Crickmore et al., 2020). This intricate protein arsenal highlights the versatility and complexity of Bt's entomopathogenic properties, underscoring its significance in the realm of biological pest control.

The Cyt1 and Cyt2 proteins exhibit notable hemolytic activity in vitro and demonstrate specific toxicity against certain dipteran insects in vivo, particularly mosquitoes and black flies, while also displaying a broader spectrum of activity against hymenopteran, lepidopteran, coleopteran, and other dipteran insects (van Frankenhuyzen, 2009). While the toxic activity of the Cyt3 protein remains largely unexplored (Bravo et al., 2017), it is known that Cyt proteins may form cation-selective channels in lipid bilayers (Knowles et al., 1989). Despite these insights, the exact mechanisms underlying their action remain incompletely understood. Structurally, these proteins bear resemblance to the fungal toxin Volvatoxin A2, a membrane pore-forming toxin with hemolytic, cytotoxic, cardiotoxic, and neurotoxic properties (Bravo et al., 2017; Lin et al., 2004). Notably, the hemolytic activity pattern exhibited by the Cyt1A protein aligns with that of other pore-forming proteins (Cohen et al., 2011), in contrast to the proposed detergent-like mechanism (Butko, 2003). Cohen et al. (2011) proposed, based on the structure of Cyt1Aa, that its toxicity, as well as that of other unrelated proteins with a similar cytolysin fold, is linked to their capacity for essential conformational changes before membrane insertion and perforation. This particular fold facilitates the movement of α -helical layers, allowing them to move aside and

57 expose the β -sheet, which then inserts into the membrane. Additionally, the finding of a potential
58 'lipid binding pocket' situated between the β -sheet and the helical layer of Cyt1Aa (residues 171
59 to 221) supports this proposed mechanism, suggesting a specific interaction with membrane lipids
60 that may be crucial for toxicity (Cohen et al., 2011).

61 Beyond their intrinsic toxicity, Cyt proteins possess the capability to synergize with other
62 Bt and *Lysinibacillus sphaericus* insecticidal proteins (Bravo et al., 2017; Monnerat et al., 2014;
63 Wirth et al., 2000; Wu et al., 1994). In some instances, they may function as receptors for Cry
64 proteins in this synergistic mechanism (Pérez et al., 2005). Significantly, Cyt proteins play a
65 crucial role in thwarting the development of resistance in mosquitoes exposed to Bt strains,
66 serving as valuable assets in overcoming resistance developed against other insecticidal proteins
67 (Wirth et al., 2000). These multifaceted characteristics underscore the pivotal role of Cyt proteins
68 in insect control strategies.

69 All insecticidal proteins derived from Bt exhibit distinct toxic activity against their
70 respective target insects, subjecting these insects to selective pressures induced by the action of
71 the proteins. Similarly, entomopathogenic Bt strains, along with their insecticidal genes, undergo
72 selective pressures exerted by their target insects as these insects develop tolerance to the
73 insecticidal proteins. This observed phenomenon may shed light on, from an evolutionary
74 standpoint, the predominant and diverse array of Cry and vegetative insecticidal protein families
75 in comparison to Cyt proteins (Palma et al., 2014). Positive (Darwinian) or diversifying selection
76 stands out as an adaptive evolutionary force that rapidly accumulates non-synonymous (amino
77 acid changing) substitutions in protein-coding genes, potentially inducing alterations in protein
78 activity and facilitating adaptation. Positive selection can be inferred by assessing the non-
79 synonymous (d_N) to synonymous (d_S) substitution rate ratios ($\omega = d_N/d_S$). The ω value serves as a
80 sensitive metric for discerning selection pressures acting on codon sequences, signifying negative
81 or purifying selection ($\omega < 1$), neutral evolution ($\omega = 1$), and positive or diversifying selection (ω
82 > 1). Within this framework, a subgroup comprising 3-domain Cry toxins and the Vip3 proteins
83 has previously been identified as evolving under positive or adaptive selection (Wu et al., 2007a,
84 b). However, no positive selection analyses have been performed up to this date for Cyt proteins.

Therefore, the aim of this work was to analyze whether Cyt proteins are subjected to positive or diversifying selection by testing twenty complete coding sequences of *cyt* genes encoding for proteins of proven activity (Table 1), obtained from the Bacterial Pesticidal Protein resource Centre database (<https://bpprc-db.org>) (Crickmore et al., 2020) and GenBank (Sayers et al., 2021).

The ratios of non-synonymous (d_N) and synonymous (d_S) substitutions (ω) were estimated by using a combination of three approaches: two maximum likelihood methods such as CODEML program included in PAML 4 (version 4.8) (Yang, 2007) and FEL (Fixed Effects Likelihood) approach included on Data Monkey server (Pond and Frost, 2005), along with Fast Unconstrained Bayesian AppRoximation (FUBAR), a Bayesian approach also included in Data Monkey server for Inferring Selection (Murrell et al., 2013). The CODEML program in PAML 4 enables the use of a maximum likelihood approach supported with codon-based models, such as M0, M1, M2, M3, M7 and M8, for estimating ω (Weaver et al., 2018). Models that are nested together (e.g. M0-M1, M2-M3 and M7-M8) are statistically compared using a likelihood ratio test (LRT) between log-likelihood values ($\ln L$) of two models and a chi-square distribution (χ^2), providing a p value that either supports or rejects the null hypothesis (no positive selection was found). Positive selection is inferred when models M2, M3 or M8 identify a group of codons with $\omega > 1$ and the LRT for positive selection is significant ($p < 0.05$). Five duplicated sequences were removed from the original set before analyses (Table 1). Subsequently, the fifteen *cyt* sequences were aligned using PAL2NAL (Suyama et al., 2006) and analyzed with PAML 4 and Data Monkey using FEL and FUBAR approaches for detecting positive selection. PAL2NAL facilitates the construction of a multiple codon sequence alignment from previously aligned proteins and their corresponding unaligned nucleotide coding sequences (Supplementary Material). The multiple sequence alignment is suitable for conducting nucleotide substitution analyses in various evolutionary studies. A corresponding Neighbor-Joining phylogenetic tree (Supplementary Material) was constructed using *cyt* sequences with the Geneious tree builder tool included in Geneious R11 (www.geneious.com). Subsequently, the codon alignment and the phylogenetic tree were used to estimate selection using PAML 4. Additionally, FEL

(<https://www.datamonkey.org/fel>) and FUBAR (<https://www.datamonkey.org/fubar>) analyses were performed on the server, using the PAL2NAL codon alignment file as input.

For the dataset used in this study, PAML 4 ran smoothly in a Mac (Apple) laptop computer (Apple M2 pro, 16 GB RAM LPDDR5 and 500 GB SSD), while FEL and FUBAR analyses were conducted on-line at Data Monkey server (<https://www.datamonkey.org>). Installation and execution of the PAML 4 software require Unix command line knowledge, which is performed through the Mac command line terminal. However, for users unfamiliar with Unix command line, a graphical interface called PamlX is available (Xu and Yang, 2013). The CODEML program included in PAML 4 was executed using the F3×4 model (the codeml.ctl control file is deposited at Supplementary Material), which estimates codon frequencies from average nucleotide frequencies at the three codon positions with a significance threshold set at $p < 0.05$ (Yang, 2007). FEL and FUBAR programs were run under default parameters, with a significance threshold determined by the likelihood ratio test at $p < 0.05$ for FEL and the posterior probability set at 0.9 for FUBAR. Only those sites identified and selected as significant by at least two of the three approaches, were considered as evolving under positive or negative selection. However, the three programs could not detect significant positive selection acting on Cyt proteins, whereas a high proportion of residues (24%) were found to be evolving under negative or purifying selection. This was observed at a significance level of $p < 0.05$ for PAML 4 and FEL, while FUBAR used a posterior probability of 0.9 (refer to Supplementary Material and Figure 1).

The Cyt1a protein sequence was also analyzed using ConSurf (Ashkenazy et al., 2016), a program that predicts functional regions in macromolecules (protein/DNA/RNA) by inferring the evolutionary dynamics of non-synonymous substitutions among homologous sequences. The results showed that some negatively selected sites were additionally predicted as conserved by ConSurf, suggesting a potential role of these residues as functional or structural amino acids in the protein.

As mentioned earlier, positive selection drives the fixation of mutations or non-synonymous substitutions that confer fitness advantages to a given organism. Within the 3-domain Cry proteins family, a diverse array of proteins exhibit activity against several insect

orders, as well as nematodes and human-cancer cells (van Frankenhuyzen, 2009). These pesticidal proteins exhibit toxicity upon proteolytic activation and binding to specific receptors of the epithelial cells of the insect gut. These processes, essential for toxicity, may play a significant role in inducing adaptive evolution, as insect resistance can arise from alterations in both the toxin-binding sites and the proteolytic processing sites (Ferré and Van Rie, 2002).

The Cyt group of insecticidal proteins constitute a smaller sub-division of the δ -endotoxins exhibiting a narrow host range mainly specific to some dipteran species (de Maagd et al., 2003). In contrast to Cry proteins, Cyt proteins may not bind to specific cell receptors but instead directly interact with saturated membrane lipids (Xu et al., 2014). Such functional features might impose less restrictive selective pressures and could also explain the lack of literature reports describing insect resistance against Cyt proteins. Tetreau et al. (2013) confirmed previous studies demonstrating the absence of *B. thuringiensis* var *israelensis* (Bti) resistance in mosquito populations treated for decades with Bti-based bioinsecticides (Tetreau et al., 2013). More recently, Becker et al. (2018), reported the continued absence of *Aedes vexans* resistance following 36 years of Bti treatment in the Upper Rhine Valley in Germany.

Cyt proteins share structural similarities with volvatoxin and also have homologs active against aphids in the Gram-negative bacterium *Dickeya dadantii* (Grenier et al., 2006; Loth et al., 2015). Modeling studies have suggested further structural homology with the P20 accessory protein from Bti (Berry and Board, 2016). However, the Cyt homologs from *D. dadantii* and P20 exhibited low pairwise identity percentages compared to the Cyt proteins from Bt and were not included in this study.

In conclusion, the results of this study strongly suggest a conservative evolutionary trend of Cyt proteins from Bt, which may explain the small number and low variability of Cyt proteins described to date in Bt. In the complex evolutionary struggle between hosts and pathogens, hosts often play an active role rather than remaining passive in the battle for survival. Over time, they have developed various mechanisms to counteract the lethal effects of virulence factors such as enzymes and toxins produced by the pathogens. This co-evolutionary dynamic has played a significant role in driving the diversification of the pesticidal proteins from *B. thuringiensis*.

However, in the case of Cyt proteins, where no resistance events have been documented even after several decades of use, the immediate necessity for improvement or refinement of these mosquitocidal proteins may not be evident. This suggests the need for further investigations including site-directed mutagenesis modifications and bioassays. Gathering such information will enrich our understanding about Cyt proteins, their evolutionary patterns and ecological implications arising from a small group of insecticidal proteins that remain effective against mosquitoes.

Acknowledgements

This research was funded by FONCYT, Plan Argentina Innovadora 2020, grant number PICT 2017-0087, by the CONICET, grant number PIP 2017 GRUPO INVES, and by the Generalitat Valenciana grant (PROMETEO/2020/010). Leopoldo Palma would like to express gratitude to the Spanish Government-Universities Ministry, the Next Generation EU and Recovery, Transformation, and Resilience plans for funding his awarded María Zambrano contract (ref. ZA21-003).

References

- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 44, W344-350.
- Berry, C., Board, J., 2016. The use of structural modelling to infer structure and function in biocontrol agents. *J. Invertebr. Pathol* In Press.
- Bravo, A., Pacheco, S., Gómez, I., Garcia-Gómez, B., Onofre, J., Soberón, M., 2017. Insecticidal Proteins from *Bacillus thuringiensis* and Their Mechanism of Action. *In: Fiuza L., Polanczyk R., Crickmore N. (eds) Bacillus thuringiensis and Lysinibacillus sphaericus*. Springer, Cham.
- Butko, P., 2003. Cytolytic toxin Cyt1A and its mechanism of membrane damage: data and hypotheses. *Appl Environ Microbiol* 69, 2415-2422.
- Chakroun, M., Banyuls, N., Bel, Y., Escriche, B., Ferré, J., 2016. Bacterial Vegetative Insecticidal Proteins (Vip) from Entomopathogenic Bacteria. *Microbiol. Mol. Biol. Rev.* 80, 329-350.
- Cohen, S., Albeck, S., Ben-Dov, E., Cahan, R., Firer, M., Zaritsky, A., Dym, O., 2011. Cyt1Aa toxin: crystal structure reveals implications for its membrane-perforating function. *J Mol Biol* 413, 804-814.
- Crickmore, N., Berry, C., Panneerselvam, S., Mishra, R., Connor, T.R., Bonning, B.C., 2020. A structure-based nomenclature for *Bacillus thuringiensis* and other bacteria-derived pesticidal proteins. *J Invertebr Pathol*, 107438.
- Crickmore, N., Zeigler, D.R., Feitelson, J., Schnepf, E., Van Rie, J., Lereclus, D., Baum, J., Dean, D.H., 1998. Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* 62, 807-813.
- Crickmore, N., Zeigler, D.R., Schnepf, E., Van Rie, J., Lereclus, D., Baum, J., Bravo, A., Dean, D.H., 2018. *Bacillus thuringiensis* toxin nomenclature. www.lifesci.sussex.ac.uk/Home/Neil_Crickmore/Bt/. Acc. 2017.
- Ferré, J., Van Rie, J., 2002. Biochemistry and genetics of insect resistance to *Bacillus thuringiensis*. *Annu Rev Entomol* 47, 501-533.
- Grenier, A.M., Duport, G., Pages, S., Condemine, G., Rahbe, Y., 2006. The phytopathogen *Dickeya dadantii* (Erwinia chrysanthemi 3937) is a pathogen of the pea aphid. *Appl Environ Microbiol* 72, 1956-1965.
- Knowles, B.H., Blatt, M.R., Tester, M., Horsnell, J.M., Carroll, J., Menestrina, G., Ellar, D.J., 1989. A cytolytic delta-endotoxin from *Bacillus thuringiensis* var. *israelensis* forms cation-selective channels in planar lipid bilayers. *FEBS Lett* 244, 259-262.
- Lin, S.C., Lo, Y.C., Lin, J.Y., Liaw, Y.C., 2004. Crystal structures and electron micrographs of fungal volvatoxin A2. *J. Mol. Biol.* 343, 477-491.

218 Loth, K., Costechareyre, D., Effantin, G., Rahbe, Y., Condemine, G., Landon, C., da Silva, P.,
 219 2015. New Cyt-like delta-endotoxins from *Dickeya dadantii*: structure and aphicidal activity. Sci
 220 Rep 5, 8791.

221 Monnerat, R., Pereira, E., Teles, B., Martins, E., Praca, L., Queiroz, P., Soberon, M., Bravo, A.,
 222 Ramos, F., Soares, C.M., 2014. Synergistic activity of *Bacillus thuringiensis* toxins against
 223 *Simulium* spp. larvae. J Invertebr Pathol 121, 70-73.

224 Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L., Scheffler,
 225 K., 2013. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol
 226 Evol 30, 1196-1205.

227 Palma, L., Muñoz, D., Berry, C., Murillo, J., Caballero, P., 2014. *Bacillus thuringiensis* toxins:
 228 an overview of their biocidal activity. Toxins (Basel) 6, 3296-3325.

229 Pérez, C., Fernandez, L.E., Sun, J., Folch, J.L., Gill, S.S., Soberón, M., Bravo, A., 2005. *Bacillus*
 230 *thuringiensis* subsp. *israelensis* Cyt1Aa synergizes Cry11Aa toxin by functioning as a membrane-
 231 bound receptor. Proc Natl Acad Sci U S A 102, 18303-18308.

232 Pond, S.L.K., Frost, S.D.W., 2005. Datamonkey: rapid detection of selective pressure on
 233 individual sites of codon alignments. Bioinformatics 21, 2531-2533.

234 Sayers, E.W., Cavanaugh, M., Clark, K., Pruitt, K.D., Schoch, C.L., Sherry, S.T., Karsch-
 235 Mizrahi, I., 2021. GenBank. Nucleic Acids Res 49, D92-D96.

236 Suyama, M., Torrents, D., Bork, P., 2006. PAL2NAL: robust conversion of protein sequence
 237 alignments into the corresponding codon alignments. Nucleic Acids Res 34, W609-612.

238 Tetreau, G., Stalinski, R., David, J.P., Després, L., 2013. Monitoring resistance to *Bacillus*
 239 *thuringiensis* subsp. *israelensis* in the field by performing bioassays with each Cry toxin
 240 separately. Mem Inst Oswaldo Cruz 108, 894-900.

241 van Frankenhuyzen, K., 2009. Insecticidal activity of *Bacillus thuringiensis* crystal proteins. J
 242 Invertebr Pathol 101, 1-16.

243 Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V., Kosakovsky Pond, S.L., 2018.
 244 Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other
 245 Evolutionary Processes. Mol Biol Evol 35, 773-777.

246 Wirth, M.C., Federici, B.A., Walton, W.E., 2000. Cyt1A from *Bacillus thuringiensis* synergizes
 247 activity of *Bacillus sphaericus* against *Aedes aegypti* (Diptera: Culicidae). Appl. Environ.
 248 Microbiol. 66, 1093-1097.

249 Wu, D., Johnson, J.J., Federici, B.A., 1994. Synergism of mosquitocidal toxicity between CytA
 250 and CryIVD proteins using inclusions produced from cloned genes of *Bacillus thuringiensis*. Mol
 251 Microbiol 13, 965-972.

252 Xu, B., Yang, Z., 2013. pamlX: A Graphical User Interface for PAML. Molecular Biology and
 253 Evolution 30, 2723-2724.

254 Xu, C., Wang, B.C., Yu, Z., Sun, M., 2014. Structural insights into *Bacillus thuringiensis* Cry,
255 Cyt and parasporin toxins. *Toxins (Basel)* 6, 2732-2770.
256 Yang, Z.H., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology*
257 *and Evolution* 24, 1586-1591.
258
259

Table legends

Table 1. The 20 nucleotide *cyt* gene sequences intended to be used in this study including GenBank accession numbers. Five sequences were not unique (duplicates) and removed.

Gene name	GenBank Accession No.
<i>cyt1Aa1</i>	X03182
<i>cyt1Aa2</i>	X04338 (Removed)
<i>cyt1Aa3</i>	Y00135 (Removed)
<i>cyt1Aa4</i>	M35968
<i>cyt1Aa5</i>	AL731825 (Removed)
<i>cyt1Aa7</i>	KF152888
<i>cyt1Ab1</i>	X98793
<i>cyt1Ba1</i>	U37196
<i>cyt1Ca1</i>	AL731825
<i>cyt1Da1</i>	ADV33305
<i>cyt1Da2</i>	JN226105 (Removed)
<i>cyt2Aa1</i>	Z14147
<i>cyt2Aa2</i>	AF472606 (Removed)
<i>cyt2Aa3</i>	ACF35049
<i>cyt2Ba1</i>	U52043
<i>cyt2Ba8</i>	AF215646
<i>cyt2Ba14</i>	FJ205866
<i>cyt2Ba15</i>	JF283552
<i>cyt2Bb1</i>	U82519
<i>cyt3Aa1</i>	HM596591

Figure Legends

Figure 1. Negatively selected and conserved sites found on Cyt proteins. A) Negatively selected sites are marked with red rectangles under Cyt1Aa1 protein sequence (only residues identified both with FEL and FUBAR approaches are shown). Colored rectangles indicate residues predicted by ConSurf as conserved and potentially involved in structural or functional roles. Buried residue: an amino acid located within the protein structure (not accessible to solvent, e.g., water). Exposed residue: an amino acid located on the surface of the protein, accessible to the solvent (e.g., water). B) Three-dimensional structure of Cyt1Aa (PDB accession number 3RON) protein exhibiting negatively selected zones (red colored) spread throughout the cytolsin fold toxin structure (Cohen et al., 2011).

316 **Figure 1**

