MINISTRY OF EDUCATION AND TRAINING

VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY SCIENCE AND TECHNOLOGY

Le Thu Ngoc

ISOLATION AND EXPRESSION OF GENES ENCODING INSECTICIDAL PROTEINS AGAINST SOYBEAN POD BORER (*Etiella zinckenella* Treitschke) FROM NATIVE *Bacillus thuringiensis* BACTERIA

SUMMARY OF DISSERTATION ON APPLIED BIOLOGY

Major: Biotechnology

Code: 9.42.02.01

Ha Noi - 2024

The dissertation is completed at: Graduate University of Science and Technology, Vietnam Academy Science and Technology

Supervisors:
1. Supervisor 1: Dr. Le Thi Minh Thanh
Institute of Biotechnology
2. Supervisor 2: Assoc. Prof. Dr. Pham Bich Ngoc
Institute of Biotechnology

Referee 1: Assoc. Prof. Dr. Khuat Huu Trung
Agricultural Genetics Institute
Referee 2: Prof. Dr. Nghiem Ngoc Minh
Institute of Genome Research
Referee 3: Assoc. Prof. Dr. Nguyen Duc Bach
Vietnam National University of Agriculture

The dissertation is examined by Examination Board of Graduate University of Science and Technology, Vietnam Academy of Science and Technology at…….. (time, date……)

The dissertation can be found at:
1. Graduate University of Science and Technology Library
2. National Library of Vietnam

LIST OF THE PUBLICATIONS RELATED TO THE DISSERTATION

1. **Le Thu Ngoc**, Le Thi Minh Thanh, Pham Bich Ngoc, Trinh Thi Thu Ha, Dong Van Quyen, Ngo Dinh Binh, Chu Hoang Ha, Hoang Ha, Nguyen Van Dong (2022) Detection of a novel Cry2Ab toxin against *Etiella zinckenella* Treitschke from the *Bacillus thuringiensis serovar canadensis* SP142 strain. *Plant Prot Sci* 58:158–169. https://doi.org/10.17221/59/2021-PPS

2. **Le Thu Ngoc**, Tran Thi Huyen, Trinh Thai Vy, Nguyen Thi Tra, Chu Hoang Ha, Do Tien Phat, Le Thi Minh Thanh, Pham Bich Ngoc (2022) Transient expression of plant-codon-optimized *cry2Ab39* gene by agroinfiltration in *N. benthamiana*. *J Mod Agric Biotechnol* 1(2): 11. DOI: 10.53964/jmab.2022011.

3. Le Thi Minh Thanh, Trinh Thi Thu Ha, Ngo Dinh Binh, Chu Hoang Ha, Pham Bich Ngoc, **Le Thu Ngoc** (2020). Biologically pure *Bacillus thuringiensis* serovar canadensis strain SP14.2 carrying the gene encoding insecticidal protein Cry2Ab39 lethal for soybean pod borer. Patent granted by the Vietnam Intellectual Property Office (patent number 37733)

INTRODUCTION

1. The urgency of the thesis

Soybean (*Glycine max* L.) is one of the important crops with high economic value not only in Vietnam but also in many countries worldwide. However, with small-scale traditional farming methods, low-yielding varieties, domestically-produced soybeans cannot compete with imported ones in terms of cost. Currently, domestic soybean production only meets about 8-10% of the domestic consumption demand, with the remaining 90% relying on imported soybeans, mainly for animal feed processing. Pests and diseases, especially insect pests, are among the major factors affecting soybean yield. Among them, the damage caused by the pod borer, *Etiella zinckenella* Treitschke, to soybeans is considered severe and difficult to control.

Over the past decades, biopesticides have been widely used to effectively control pests and diseases in crops. Among them, the most common is the microbial insecticide *Bt* (*Bacillus thuringiensis*), accounting for up to 90% of the global market for biopesticides due to its broad spectrum and high efficiency in insect control, while still being environmentally friendly and safe for humans and other non-target organisms. Furthermore, *Bt* genes encoding insecticidal proteins have been isolated and integrated into the plant genome to produce transgenic crops conferred with insect resistance. The use of *Bt* crops in agriculture brings many benefits, including more effective pest control, reduced use of chemical insecticides, creation and maintenance of natural enemy populations in farming areas, and enabling more sustainable agricultural practices. Currently, there are six *Bt* soybean varieties approved for cultivation and commercialization worldwide, capable of resisting some pests of the Lepidoptera order such as the velvet bean caterpillar (*Anticarsia gemmatalis* Hübner), the soybean looper (*Chrysodeixis includens* Walker), and the cotton bollworm (*Helicoverpa armigera*). However, there have been almost no reports assessing the resistance of these six commercialized *Bt* soybean varieties to the pod borer *E. zinckenella* Treitschke, as well as other *Bt* soybean varieties under research.

Therefore, identifying *Bt* strains harbouring insecticidal genes capable of effectively controlling *E. zinckenella* Treitschke is important and urgent, laying the groundwork for research on the application of genetic technology to develop pod borer-resistant soybean varieties. Currently, the Institute of Biotechnology maintains a *Bt* collection consisting of more than 3000 strains isolated from 52 provinces in Vietnam. With the advantage of being one of the most biologically diverse countries in the world, researching and exploiting native *Bt* strains of Vietnam to find specific insecticidal genes holds great promise. The results of these studies will allow Vietnamese scientists to proactively develop genetically modified soybean varieties resistant to pod borers that are well-adapted to the specific climatic conditions of Vietnam, thereby contributing to increased productivity and competitiveness of domestic soybeans. Based on the research situation and practical needs mentioned above, we have conducted the topic: "Isolation and expression of genes encoding insecticidal proteins against soybean pod borer (*Etiella zinckenella* Treitschke) from native *Bacillus thuringiensis* bacteria".

2. Research objectives of the thesis

Finding and isolating the genes encoding novel insecticidal proteins with high toxicity for *Etiella zinckenella* Treitschke from native *Bt* strains of Vietnam in order to provide genetic resources for generation of transgenic soybean resistant to this pod borer species.

3. Main research contents of the thesis

i) Selecting *Bt* strains effectively kill soybean pod borer *E. zinckenella* Treitschke from Vietnamese native *Bt* strains collection.

ii) Sequencing and mining of genome data of some *Bt* strains highly toxic to *Etiella zinckenella* Treitschke with efficacy more than 85% to predict gene candidates encoding novel insecticidal proteins.

iii) Isolating the genes encoding novel insecticidal proteins with potential to kill pod borer *E. zinckenella* Treitschke from selected native *Bt* strains.

iv)Expressing the genes encoding novel insecticidal proteins in *Escherichia coli* and evaluating insecticidal efficacy of recombined proteins against soybean pod borer *E. zinckenella* Treitschke.

v) Genetic modification and optimization of expression conditions for new *Bt* toxin gene expression in the plant expression system.

CHAPTER 1. OVERVIEW

1.1. Soybeans and harmful insects

The importance of soybeans

Soybeans (*Glycine max* (L.) Merrill), belonging to the legume family (Fabaceae), are the fourth most important crop among staple food crops after wheat, rice, and corn. They contribute 50% of the global edible oil supply and about two-thirds of the world's plant protein for human and animal consumption.

Soybean pod borer *Etiella zinckenella* Treitschke and control masures

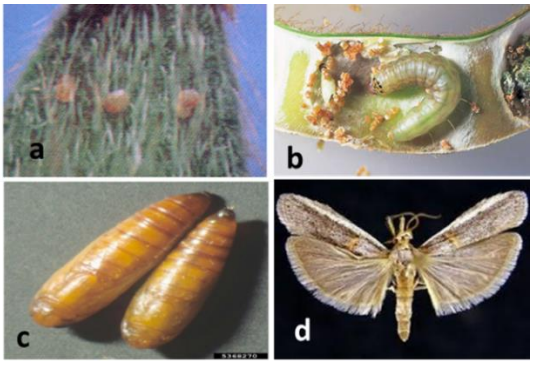


Figure 1.1. Lifecycle of the soybean pod borer *E. zinckenella* Treitschke

E. zinckenella Treischke is an insect from the order Lepidoptera, family Phycitidae. It is widely distributed and damages various host plants such as legumes and Panax, with soybeans being its preferred host. The lifecycle of the soybean pod borer ranges from 32-51 days, developing through four stages: egg, larva, pupa, and adult (Figure 1.1). Due to the larvae's habit of boring into the pods, conventional insect control methods, such as spraying insecticides, are generally ineffective against *Etiella zinckenella* Treitschke, as the insecticide cannot reach the larvae inside the pod. Currently, the most effective way to reduce damage caused by this insect is the use of pheromone traps or selecting disease-resistant soybean varieties.

1.2. Insecticidal proteins derived from *Bacillus thuringiensis*

Classification and nomenclature

Bacillus thuringiensis (*Bt*) is a gram-positive, rod-shaped, aerobic or facultatively anaerobic soil bacterium. During their growth phases, they produce proteins toxic to insects. In 1998, Crickmore and colleagues established a Nomenclature Committee for *Bt* toxins and introduced a more comprehensive classification system based on amino acid sequence similarity of the toxin proteins for easy inclusion of newly discovered *Bt* toxins. Toxins with less than 45% amino acid sequence similarity are classified at level 1 (e.g., Cry1, Cry2, Cry3), while those with 45-78% similarity differ at level 2 (e.g., Cry1A, Cry1B). Level 3 (e.g., Cry1Ab, Cry1Ac) distinguishes toxins with 78%-95% similarity, and level 4 (e.g., Cry1Ac1, Cry1Ac3) for those with over 95% similarity (Figure 1.3). By 2020, the *Bt* toxin nomenclature committee had established a classification system comprising 16 classes based on the protein structures of the toxins.

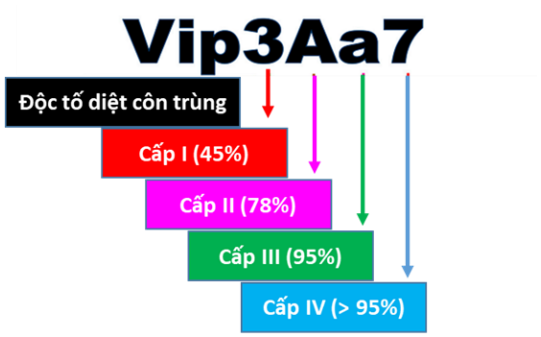


Figure 1.3. Schematic diagram of the four-level nomenclature system for identifying endotoxins (Cry and Cyt) and secretory toxins (Vip and Sip)

Activity and insecticidal spectrum

Cry toxins can kill insects from orders Lepidoptera, Coleoptera, Hemiptera, Diptera, non-arthropod invertebrates from the phylum Nematoda (including parasitic and free-living nematodes), and some mollusks. Cyt toxins form a small group of distinct crystalline proteins that kill certain dipteran larvae, especially mosquitoes and black flies. Vip proteins are secreted into the culture medium during the vegetative growth phase of the bacterium. Vip1 and Vip2 proteins form a binary toxin highly active against some coleopteran insects and the sap-sucking insect *Aphis gossypii* from Hemiptera. In contrast, Vip3 proteins are single-chain toxins with a broad spectrum of activity against many lepidopteran insects. The host range for the Vip4 subfamily has not yet been clearly identified.

Cry Protein Structure

A typical Cry toxin structure comprises three distinct domains. The domains are ordered from the N-terminus to the C-terminus of the polypeptide chain. Domain I, or the pore-forming domain, is highly conserved and located at the N-terminus. Domain II (also called the central or middle domain) is highly variable (hypervariable) and plays a crucial role in toxin-receptor interactions. Finally, Domain III (also called the galactose-binding domain) is located at the C-terminus and is involved in receptor binding and pore formation. Beyond the core toxic domain, the Cry1Ac protoxin structure includes an extension region (pro-region) forming four additional functional domains (Domain IV-VII) (Figure 1.5).

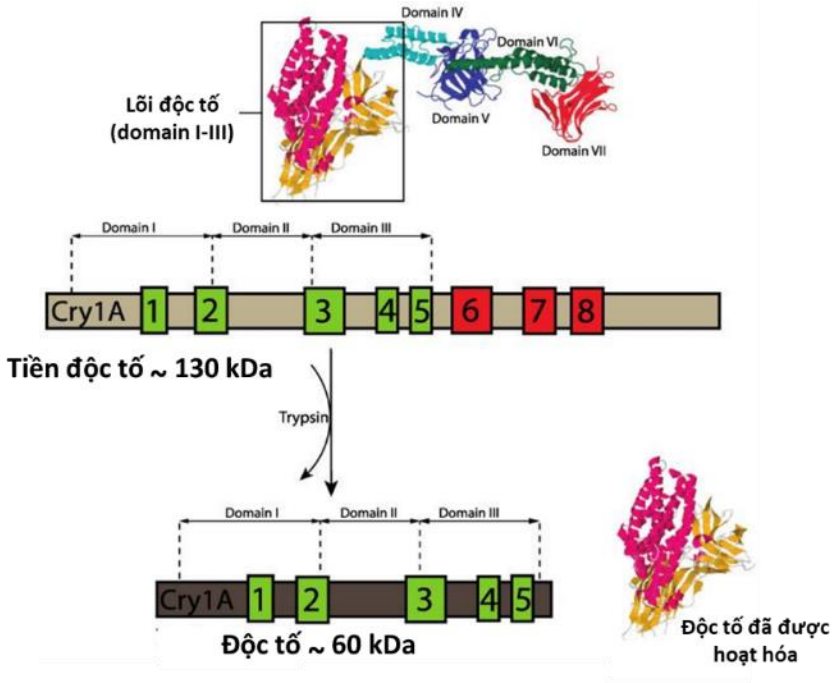


Figure 1.5. Three-dimensional structural changes during Cry1Ac protoxin activation

Insecticidal mechanism

The most widely accepted mechanism of Cry toxin insecticidal action is the pore-forming model. When larvae ingest the toxin crystals, they dissolve at a high pH and are processed by proteases in the larval midgut to form activated toxins of approximately 60-70 kDa. The activated toxin passes through the peritrophic membrane to contact the midgut microvilli, recognize, and bind to membrane receptors to form pores. These ion

channel pores lead to osmotic imbalance, cell lysis, and eventually insect death.

1.3. Achievements in developing genetically modified *Bt* soybeans

Most *Bt* genetically modified (GM) soybean varieties currently cultivated worldwide are developed by two GM crop companies: Monsanto and Dow AgroSciences. Monsanto developed the GM soybean variety MON87701 (expressing Cry1Ac from *B. thuringiensis* kurstaki HD73), which shows significant resistance to the caterpillar *H. armigera*. Through breeding, Monsanto further developed the GM soybean Intacta™ Roundup Ready™ 2 Pro, combining MON87701 and MON89788 (herbicide resistance gene). Recently, the GM soybean MON87751 carrying recombinant genes cry1A.105 and cry2Ab2 has shown resistance to certain Lepidoptera insects. Building on these GM soybeans, Monsanto created MON 87751 × MON 87701 × MON 87708 × MON 89788, a multi-gene Bt combination approved for cultivation in Taiwan and South Korea. Similarly, the GM soybeans DAS81419 and DAS81419 × DAS44406, developed by Dow AgroSciences, utilize *Agrobacterium*-mediated transformation to incorporate cry1Ac, cry1F, and the phosphinothricin acetyltransferase (PAT) gene.

In Vietnam, since 2000, genetic modification research in soybeans has been successfully conducted at many centers and research institutes, although most efforts have focused on genes related to drought tolerance and herbicide resistance.

CHAPTER 2. MATERIALS AND METHODS

2.1. Materials, chemicals and equipment

- 221 indigenous *Bt* bacteria strains derived from the Vietnam Bt Collection provided by the VAST- Culture Collection of Microorganisms, Institute of Biotechnology.
- 4-week-old *N. benthamiana* tobacco plants were grown hydroponically in a growth chamber
- Larvae of soybean pod borer *E. zinckenella* Treitschke provided by the Plant Protection Research Institute.

2.2. Methods

2.2.1. Screening Bt strains with insecticidal activity against soybean pod borer

- The activation method of *Bt* strains
- Microbial biomass cultivation
- Crystallography method
- Evaluating the insecticidal activity of *Bt* strains against *E. zinckenella* Treitschke

2.2.2. Genomic sequencing of Bt strains using PacBio SMRT Sequencing Technology and mining the Bt genome database for genes encoding insecticidal proteins using bioinformatics tools

- Extraction of total DNA from *Bt*
- DNA library preparation
- PacBio whole genome sequencing
- *De novo* genome assembly
- Rapid annotation method for *denovo* assembled genome
- Prediction of *Bt* toxin genes from the genome database using BtToxin_scanner

2.2.3. Isolation and gene cloning

- Amplifying target genes by PCR
- Gene cloning method.
- Transformation of plasmid into *E. coli* using the heat shock method
- Plasmid DNA extracion from *E. coli*
- Sanger sequencing method

2.2.4. Expression of the genes encoding Bt toxins in E. coli BL21 and evaluation of insecticidal activity of the recombinant proteins.

- Construction of expression vector pET32a(+) carrying *cry1Na*, *cry1Be*, *cry2Ab*, and *cry2Ah*
- Expression of recombinant proteins in *E. Coli*
- SDS-PAGE and Western blot
- Purification of recombinant proteins using affinity chromatography
- Evaluation of insecticidal activities against *Etiella zinckenella* Treitschke of purified recombinant proteins

2.2.5. Genetic modification and expression of cry2Ab39 in N. benthamiana by agroinfiltration

- Codon optimization of *cry2Ab39* for suitable expression in *N. benthamiana*
- Design of the construct pFGC/35S-cry2Ab39opt
- Design of the construct pFGC/rbcS1A-pro_cry2Ab39opt
- Generation of Agrobacterium strains carrying expression vectors
- Inoculation of Agrobacterium into tobacco leaves using vacuum infiltration
- Extraction and investigation of Cry2Ab39 expression in infected N. benthamiana tobacco leaves.
- Purification of recombinant Cry2Ab39 expressed in tobacco

CHAPTER 3. RESULTS

3.1. Screening *Bt* strains with high insecticidal activity against soybean pod borer from native *Bt* Collection in Vietnam Activation and biomass cultivation of *Bt* strains

A total of 221 *Bt* strains derived from the native *Bt* collection in Vietnam (collected from 4 mountainous provinces in the North including Lao Cai, Tuyen Quang, Ha Giang, and Dien Bien) were thawed on ice and cultured on LB agar medium plates, then incubated at 28°C in an incubator for 2 days. From activated *Bt* plates, with each strain, one pure culture colony was selected for primary culture. Primary culture colonies were then transferred to medium for biomass cultivation. The results of the measurement showed that after 72 hours of cultivation, the *Bt* strains' cultures all had OD600 $\geq$ 2, ensuring suitability for subsequent soybean pod borer insecticidal activity studies.

Insecticidal activity test

Statistical results of dead larvae due to toxin ingestion after 7 days of testing showed that among the 221 *Bt* strains studied, the group with insecticidal activity ranging from 0-20% included 138 strains, accounting for the highest proportion (62.44%), followed by the group with insecticidal activity ranging from 21-70%, accounting for 31.22% with 69 strains. The number of *Bt* strains highly toxic to soybean pod borer larvae obtained through screening was relatively low with 10 strains having activity from 71 - 84% (4.52%). Only 4 strains (1.81%) including SP14.2,

Đ6.1, Đ6.2, and BD8.2 had larval mortality rates above 85% (Figure 3.2, table 3.2).

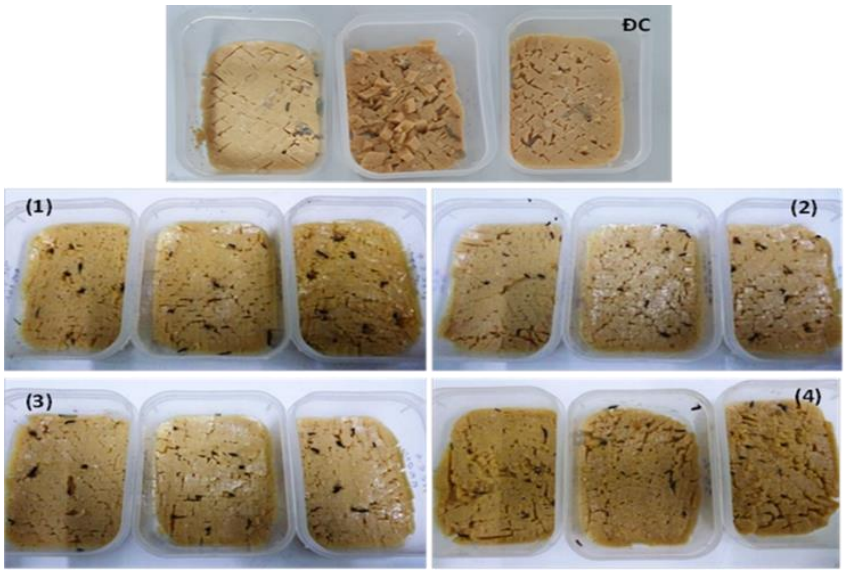


Figure 3.2. Results of insecticidal activity test on soybean pod borer larvae *Etiella zinckenella* Treitschke of some *Bt* strains studied

Table 3.2. Insecticidal activity against soybean pod borer of tested *Bt* strains

Insecticidal activity	**No. strains**				**Total**	**Ratio (%)**
	Lao Cai	**Tuyen Quang**	**Ha Giang**	**Đien Bien**		
0-20%	79	11	36	12	138	62,44
21-40%	25	3	10	2	40	18,10
41-60%	12	1	7	3	23	10,41
61-70%	2	1	1	2	6	2,71
71-84%	4	0	4	2	10	4,52
85-100%	1	2	0	1	4	1,81
Total	123	18	58	22	**221**	100

Characterization of toxin crystals and insecticidal action dynamic of *Bt* strains with insecticidal activity against soybean pod borer ≥ 85%

SEM images showed that three strains Đ6.1, Đ6.2, and BD8.2 all produced spherical crystal shapes. In contrast, the crystal of strain SP14.2 formed a biphasic shape (Figure 3.3). To determine at which growth phase the SP14.2, Đ6.1, Đ6.2, and BD8.2 strains could cause experimental larvae death, thus predicting whether larvae death was due to the action of Bt crystal toxins or other toxins, insecticidal activity against soybean pod

borer larvae was evaluated at 24 hours, 48 hours, 72 hours, and 96 hours of cultivation. The results after 3 repetitions of the experiment showed that the bacterial solution after 24 hours of cultivation of all 4 Bt strains did not have the ability to kill non-damaged soybean pod borer larvae. However, after 48 hours of cultivation, all four strains showed high insecticidal activity (above 80% -90%), notably strain BD8.2 after 48 hours of cultivation had a maximum mortality rate (100%). The remaining three strains Đ6.1, Đ6.2, and SP14.2, the mortality rate increased gradually and reached maximum (93.33% - 100%) after 72 hours and 96 hours. Meanwhile, at the 96-hour time point, the insecticidal activity of strain BD8.2 slightly decreased (83.33%).

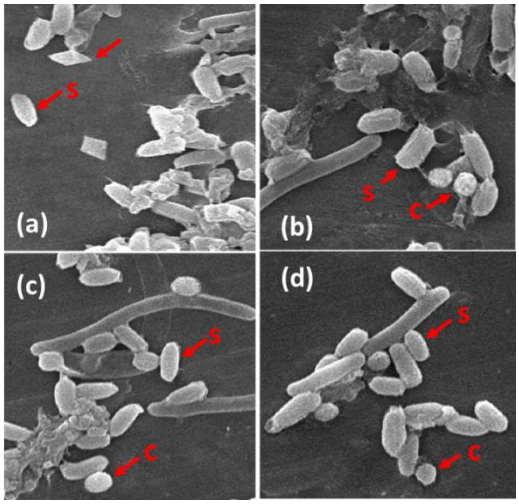


Figure 3.3. Spore and crystal morphology of *Bt* strains on scanning electron microscope at 5000x magnification.

3.2. Analysis, identification of potential new *Bt* toxin genes from genomic database of 4 *Bt* strains with high toxicity to soybean pod borer

Sequencing of 4 *Bt* Strains using PacBio SMRT Sequencing Technology

The results of genome sequencing and de novo assembly of SP14.2, Đ6.1, Đ6.2, and BD8.2 *Bt* strains obtained the number of contigs for each strain in the respective order as 57, 38, 61, 48; the length of the largest contigs was from 1,320,742 to 1,802,229 bp, the total length of contigs assembled in each strain ranged from 6,868,840 bp to 7,025,211 bp, the G+C content was nearly 35%. Based on the proportion of shorter reads less than 100 bp accounting for less than 2% and the average overlap degree of internal control samples over 0.8, it can be confirmed that the

genomic sequencing of 4 *Bt* strains achieved good results (table 3.5). When evaluating the assembly quality using BUSCO, we obtained an estimated completion percentage of the gene set for each Bt strain ranging from 90.1 - 92.9%.

Table 3.5. Denovo assembly results of *Bt* strains

Bt strain	**SP14.2**	**Đ6.1**	**Đ6.2**	**BD8.2**
No. contig	57	38	61	48
# contigs (>= 0 bp)	57	38	61	48
# contigs (>= 1000 bp)	57	38	61	48
# contigs (>= 5000 bp)	52	36	55	43
# contigs (>= 10000 bp)	38	31	46	39
# contigs (>= 25000 bp)	28	19	31	25
# contigs (>= 50000 bp)	27	15	25	23
Largest contig (bp)	1.709.205	1.802.229	1.330.436	1.320.742
Genome size (bp)	7.018.726	7.025.211	6.868.840	6.932.040
N50	330824	771356	512572	658180
GC (%)	34,98	34,94	34,91	34,94
# N's	0	0	0	0

Rapid annotation results of *de novo* genome assemblies for strains SP14.2, Đ6.1, Đ6.2, and BD8.2. We obtained the following number of coding sequences (CDSs) for the strains SP14.2, Đ6.1, Đ6.2, and BD8.2: 8,329, 7,641, 7,980, and 7,781, respectively. Additionally, the numbers of tRNA and rRNA sequences were 108 and 26, 127 and 38, 112 and 28, and 108 and 34, respectively. Among the coding sequences, over 2000 genes for each strain were allocated to subsystems. The sequences associated with the metabolism system accounted for the highest proportion, followed by systems related to protein processing and energy.

Prediction of *Bt* toxin gene sequences from the genome database of four *Bt* strains using BtToxin_scanner and identification of potential new genes for soybean pod borer control

The prediction of *Bt* toxin protein-encoding genes were conducted using the BtToxin_scanner tool. The number of genes found in the strains ranged from 11 to 22 sequences. In terms of toxin composition, most sequences found in these four strains encoded toxin proteins belonging to the Cry crystal group (47 sequences, accounting for 78.34%). Additionally, there

were 9 sequences in the Vip3A vegetative insecticidal protein group (accounting for 15%) and 4 sequences in the Cyt toxin group (accounting for 6.66%). Within the Cry crystal group, Cry1 predominated with 34 sequences, while the remaining 13 sequences belonged to the Cry2 subfamily (Figure 3.9).

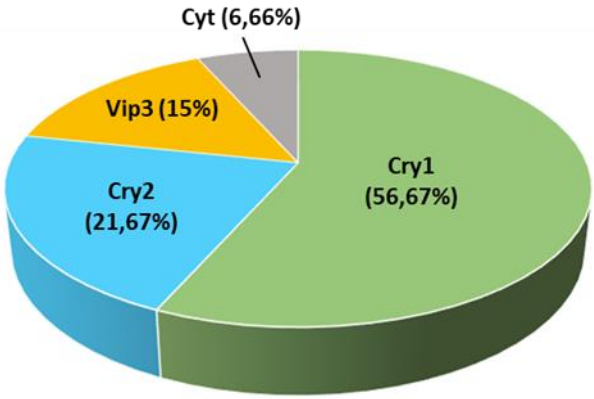


Figure 3.9. Composition and ratio of Bt toxin protein forms extracted from the genome sequence database of 4 *Bt* strains using the BtToxin_scanner tool

Theoretically, to maximize the probability of successfully isolating novel toxin-encoding genes from the genome database of the four studied *Bt* strains, we prioritized sequences with inferred amino acid compositions having less than 100% similarity to published *Bt* toxins. Moreover, for applications in controlling soybean pod borers, candidate sequences were also compared with data related to genes encoding proteins toxic to Lepidoptera (primarily stem, branch, ear, and root borers) pests of crops. Based on these criteria, six predicted novel toxin genes were selected, including four genes from the *cry1* group (*cry1Aa*, *cry1Ac*, *cry1Be*, *cry1Na*) and two genes from the *cry2A* subgroup (*cry2Ab* and *cry2Ah*). The inferred amino acid sequences from these six genes had similarities ranging from 94.01% to 99.26% with published *Bt* toxins. Notably, for *cry1Na*, although the inferred protein exhibited 100% similarity to Cry1Na3, it had 39 additional amino acids at the N-terminal compared to the reference protein, indicating that the novelty of the predicted protein requires further consideration.

3.3. Isolation, cloning, and sequencing of *cry* genes encoding novel insecticidal proteins potential for soybean pod borer control from selected *Bt* strains

Isolation and Cloning of *cry genes*

Twelve specific primer pairs were designed to fully amplify the CDS regions of the six genes (*cry1Aa*, *cry1Ac*, *cry1Be*, *cry1Na*, *cry2Ab*, and *cry2Ah*) from the genomic DNA of the selected *Bt* strains SP14.2, Đ6.1, Đ6.2, and BD8.2. The results of PCR product electrophoresis (Figure 3.13) showed successful amplification of six DNA fragments corresponding to the predicted sizes of the target *Bt* toxin genes (more than 3.5 kb for *cry1Aa*, *cry1Ac*, *cry1Be* genes, and nearly 2 kb for *cry1Na*, *cry2Ab*, *cry2Ah* genes).

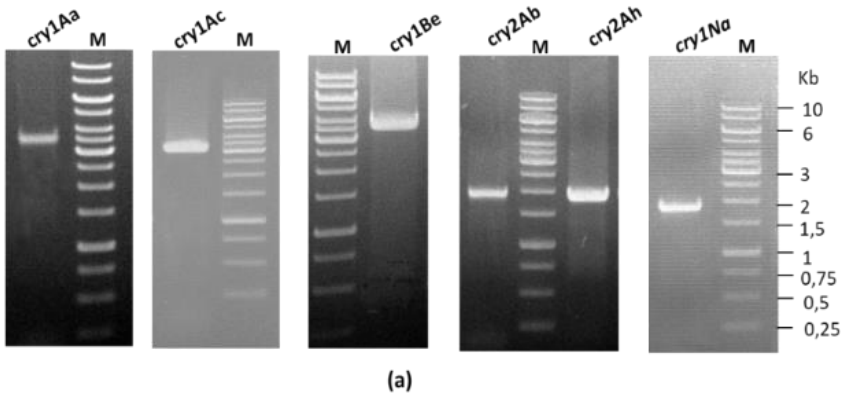


Figure 3.13. Isolation and cloning of genes encoding *Bt* insecticidal proteins

Sequencing and comparison of isolated *cry* genes with published *Bt* toxin genes

Sanger sequencing results and analysis using BioEdit software, as well as comparisons via the BLAST tool, confirm the successful isolation of six toxin protein-encoding genes: *cry1Aa*, *cry1Ac*, *cry1Be*, *cry1Na*, *cry2Ab*, and *cry2Ah* from four selected *Bt* strains from the Biotechnology Institute's Bt collection. Except for the *cry1Na* gene, which matched the predicted sequence exactly, the actual sequences of the other isolated genes showed significant differences from the predicted *Bt* genome database sequences. This discrepancy is likely due to the higher error rate in genome sequencing and assembly compared to short DNA segment sequencing using the Sanger method. Four genes (*cry1Na*, *cry1Be*, *cry2Ab*, and *cry2Ah*) predicted to be novel were selected for further studies (Table 3.9).

Table 3.9. Similarity of six isolated insecticidal proteins with published GenBank proteins

No.	Isolated Gene	Gene Size (bp)	Homologous reference proteins in Genbank	Similarity Level (%)	Novelty Rank
1	*cry1Aa*	3531	*Cry1Aa8*	99.21	Not novel
2	*cry1Ac*	3537	*Cry1Ac9*	95.93	Not novel
3	*cry1Na*	1890	*Cry1Na3*	100 (with 39 additional amino acids)	To be considered
4	*cry1Be*	3681	*Cry1Be3*	99.26	**Novel rank 4**
5	*cry2Ab*	1902	*Cry2Ab31*	94.01	**Novel rank 4**
6	*cry2Ah*	1899	*Cry2Ah1*	99.53	**Novel rank 4**

3.4. Vector construction and expression of *cry* genes in *E. coli* and evaluation of insecticidal activity against soybean pod borer of recombinant proteins

Construction of expression vectors pET32a(+) carrying *cry1Na, cry2Ab, cry2Ah,* and *cry1Be*

DNA materials were treated with *Bam*HI and *Xho*I, followed by ligation using T4-DNA ligase. Recombinant plasmids were verified by digestion with *Bam*HI and *Xho*I.

Expression of *cry1Be, cry1Na, cry2Ab, and cry2Ah* in *E. coli* BL21

Under IPTG induction, total protein extracted from *E. coli* BL21 strains carrying pET32/cry1Be, pET32/cry1Na, pET32/cry2Ab, and pET32/cry2Ah vectors showed distinct protein bands corresponding to Trx-His-Stag-Cry1Be (95 kDa), Trx-His-Stag-Cry1Na (88 kDa), Trx-His-Stag-Cry2Ab, and Trx-His-Stag-Cry2Ah (89 kDa). No such bands were observed in the control *E. coli* BL21. Successful expression of the desired recombinant proteins was further confirmed by Western blot using anti-hexahistidine (6xHis) antibodies (Figure 3.14).

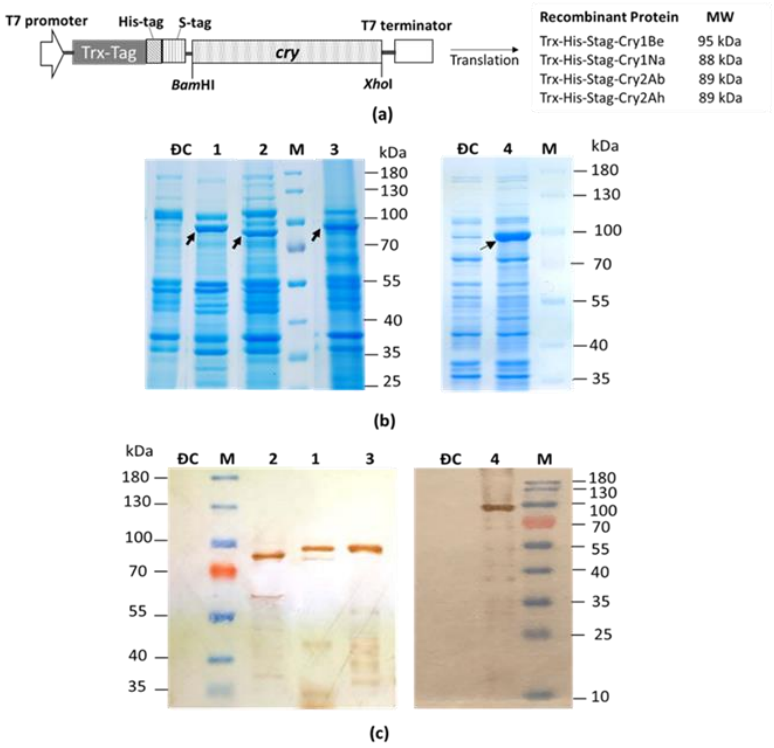


Figure 3.14. Expression results of *cry1Be*, *cry1Na*, *cry2Ab*, and *cry2Ah* genes in *E. coli* BL21.

Evaluation of insecticidal activities against soybean pod borer of recombinant proteins

Based on the mortality rate after 7 days, Cry1Be had the weakest activity with a mortality rate of 66.67%. Cry1Na and Cry2Ah showed higher toxicity with mortality rates of 86.67% and 80%, respectively. Notably, Cry2Ab had the highest mortality rate (96.67%), demonstrating its superior efficacy in controlling soybean pod borer, making it a prime candidate for further analysis to determine the median lethal dose (LC50).

Optimization of *cry2Ab* expression in *E. coli* BL21 to increase soluble recombinant protein yield

Temperature optimization: SDS-PAGE analysis (Figure 3.17) showed that at 28°C and 15°C, the Trx-His-Stag-Cry2Ab fusion protein was mostly insoluble. Lowering the induction temperature to 4°C significantly increased the soluble protein yield, although more than 50% remained in inclusion bodies. Therefore, a 4°C induction temperature was chosen to maximize soluble Trx-His-Stag-Cry2Ab yield.

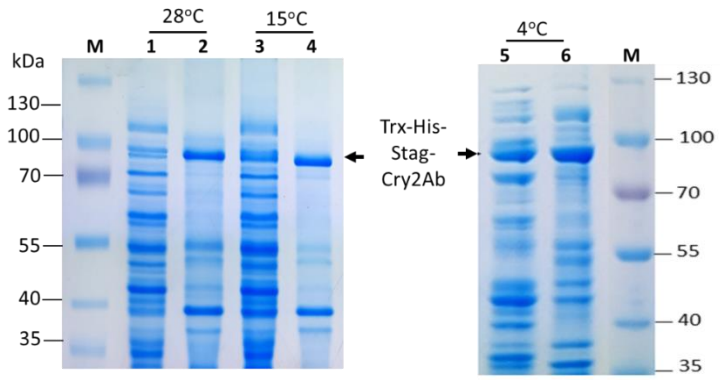


Figure 3.17. Expression analysis of Trx-His-Stag-*Cry2Ab* at different temperatures

IPTG Concentration Optimization: The effects of different concentration of inducer IPTG (0.01 mM; 0.05 mM; 0.1 mM; 0.5 mM, 1 mM) on the biosynthesis of the Trx-His-Stag-Cry2Ab fusion protein in soluble form were examined. SDS-PAGE analysis (Figure 3.18) indicated that 0.5 mM IPTG yielded the highest amount of soluble recombinant protein.

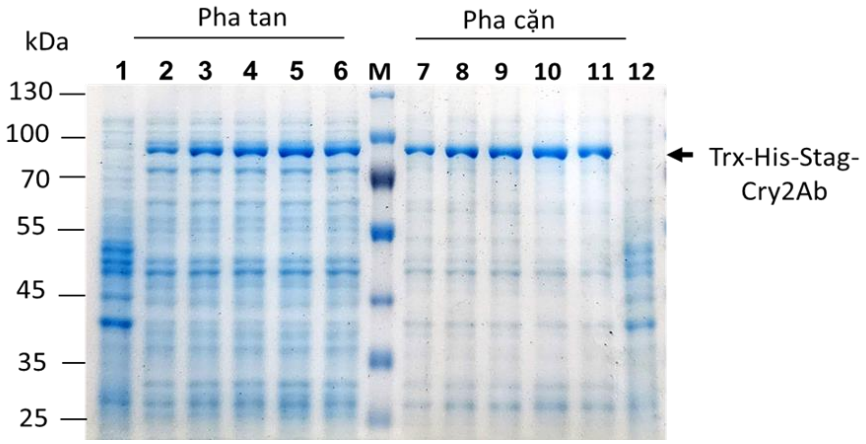
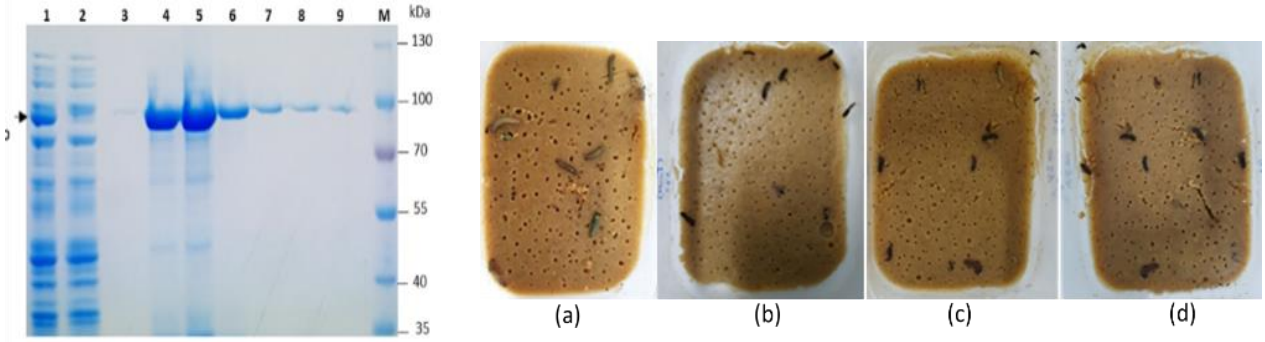


Figure 3.18. Expression analysis of Trx-His-Stag-*Cry2Ab* at different IPTG concentrations.

Purification of Trx-His-Stag-Cry2Ab

Recombinant protein was purified using affinity chromatography with Ni^{2+}-attached columns. The purified product was approximately 90 kDa, matching the expected size of Trx-His-Stag-Cry2Ab (Figure 3.19). The purification yield was estimated at 3.2 mg of purified protein per 100 ml of bacterial fermentation broth.



Figures 3.19/3.20. SDS-PAGE analysis of purified Trx-His-Stag-*Cry2Ab* and bioassay results against soybean pod borer *E. zinckenella*.

Determination of LC50 value of recombinant Cry2Ab protein against *E. zinckenella* larvae

Probit regression analysis estimated the LC50 of Trx-His-Stag-Cry2Ab for second-instar *E. zinckenella* larvae to be 1.74 µg protein/g diet (Figure 3.21, table 3.12). From these analyses, Cry2Ab from strain SP14.2 is confirmed as a novel toxin with high efficacy against *E. zinckenella*. Given its 99.21% amino acid sequence similarity with known insecticidal Bt proteins, Cry2Ab is ranked as novelty level 4. The nucleotide and deduced amino acid sequences of this toxin gene have been registered in GenBank with accession numbers MN319700.1 and QIQ19560.1, respectively. The new toxin has been named Cry2Ab39 by the Bt endotoxin nomenclature committee. Additionally, the *Bt* strain SP14.2 producing this toxin has been patented by Vietnamsese Intellectual Property Office (patent no. 37733).

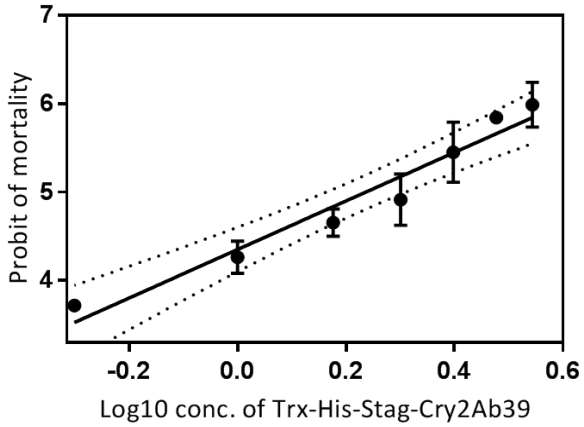


Figure 3.21. Probit regression of the concentration-mortality curve to estimate the LC_{50} of Trx-His-Stag-Cry2Ab against *E. zinckenella.* (Dotted lines represent the 95% confidence intervals of the regression line)

Table 3.12. Toxicity of Trx-His-Stag-Cry2Ab against *E. zinckenella* and the parameters of the Probit regression

LC_{50} (µg/g) (95% CI[a])	**Regression line**		**Goodness-of-fit test**		
	Slope ± SE[b]	Intercept ± SE[b]	χ^2	df	*P*[c]
1,74 (1,33-2,16)	2,74 ± 0,27	4,35 ± 0,09	2,37	6	0,88

3.5. Modification of the gene *cry2Ab39* for plant expression systems and transient expression of *Cry2Ab39* Toxin in *Nicotiana benthamiana* leaves by agroinfiltration

Optimization of *cry2Ab39*

The optimized *cry2Ab39opt* gene contains codons with usage frequencies above 50%, eliminating rare codons. The proportion of codons with high usage frequencies (91-100%) increased from 45% to 60%, while the proportion of codons with usage frequencies of 71-80% increased from 5% to 26%. Additionally, 5% of codons were in the 51-60% usage range, and 9% were in the 81-90% range (Figure 3.23). Furthermore, optimization raised the GC content of the target gene from 35.2% to 45.1%. Potential motifs or sequences that could form secondary mRNA structures or destabilize the mRNA molecule were completely eliminated (Table 3.13).

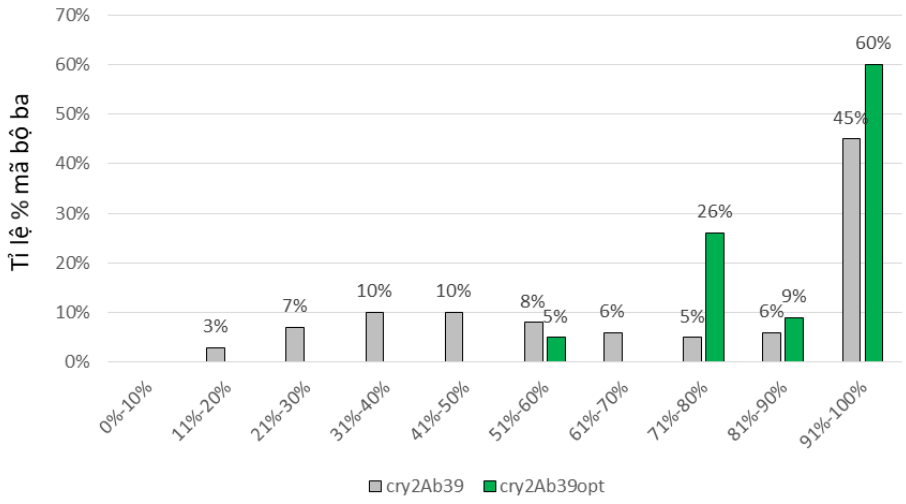


Figure 3.23. Percentage distribution of matching codons of *cry2Ab39* gene with *N. benthamiana* expression system before and after genetic modification.

Table 3.13: Comparison of the original and optimized *cry2Ab39opt gene*

Parameter	***cry2Ab39***			***cry2Ab39opt***		
	Base	No.	Percentage (%)	Base	No.	Percentage (%)
Hàm lượng các base	A	661	32.45	A	573	28.13
	T	659	32.35	T	626	30.73
	G	368	18.07	G	414	20.32
	C	349	17.13	C	424	20.81
GC%	35.2%			45.1%		
mRNA instability motifs	76			0		
mRNA secondary structures	12			0		

Construction of pFGC/35S_cry2Ab39opt and pFGC/rbcS1A-pro_cry2Ab39

The vectors pUC57/cry2Ab39opt and pFGC5941 were both digested with the restriction enzymes *NcoI* and *XmaI*. The *cry2Ab39opt* gene fragment and the vector backbone were then ligated to form the pFGC/35S_cry2Ab39opt vector.

Specific primers rbcS1A-pro_F/R were designed to amplify the promoter region of the *rbcS1A* gene, which encodes the small subunit of ribulose-1,5-bisphosphate carboxylase in *Arabidopsis thaliana*. This promoter was used to replace the 35S sequence in the pFGC/35S_cry2Ab39opt vector using *EcoRI* and *XhoI* digestion and ligation, generating the pFGC/rbcS1A-pro_cry2Ab39opt vector.

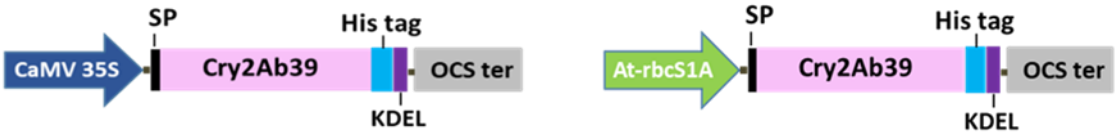


Figure 3.24/3.25. Results of constructing vectors pFGC/35S_cry2Ab39opt and pFGC/rbcS1A-pro_cry2Ab39opt

Transient expression of recombinant *Cry2Ab39* protein in tobacco leaves

Impact of promoter on transient expression of cry2Ab39opt

The efficiency of two promoters, CaMV 35S and *rbcS1A-pro*, in driving the expression of *cry2Ab39* in tobacco leaves was analyzed and compared. Western blot results showed that the *cry2Ab39opt* gene was active under the control of both promoters. However, the expression level of this gene in tobacco leaves differed significantly between the two promoters. Specifically, protein bands in samples transformed with the pFGC/rbcS1A-pro_cry2Ab39 construct were thicker and darker compared to those in samples transformed with the pFGC/35S_cry2Ab39 construct (Figure 3.27). This indicates that the *rbcS1A* promoter is more effective than the 35S promoter in transiently expressing the *cry2Ab39* gene in *Nicotiana benthamiana* leaves.

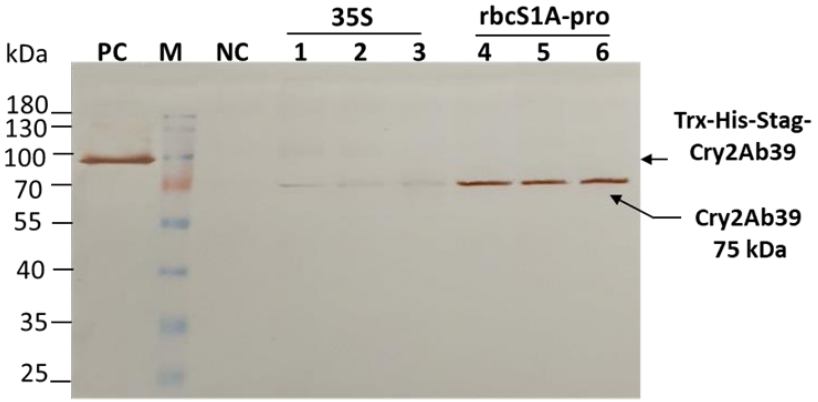


Figure 3.27. Analysis of the expression of the *cry2Ab39opt* gene under the control of the 35S and rbcS1A-pro promoters by Western blot using anti-6X His antibody.

Impact of bacterial density for infiltration on transient expression of cry2Ab39opt

Agroinfiltration experiments were conducted with bacterial suspensions carrying the pFGC/rbcS1A-pro_cry2Ab39 construct, diluted to OD600 values of 0.1, 0.3, 0.5, and 0.8. The results indicated that the synthesis of *Cry2Ab39* protein was most effective in samples infected with a bacterial suspension at OD600 = 0.5. In contrast, in samples infected with a bacterial suspension at OD600 = 0.1, the Western blot method was not sensitive enough to detect the expression of the target protein (Figure 3.28).

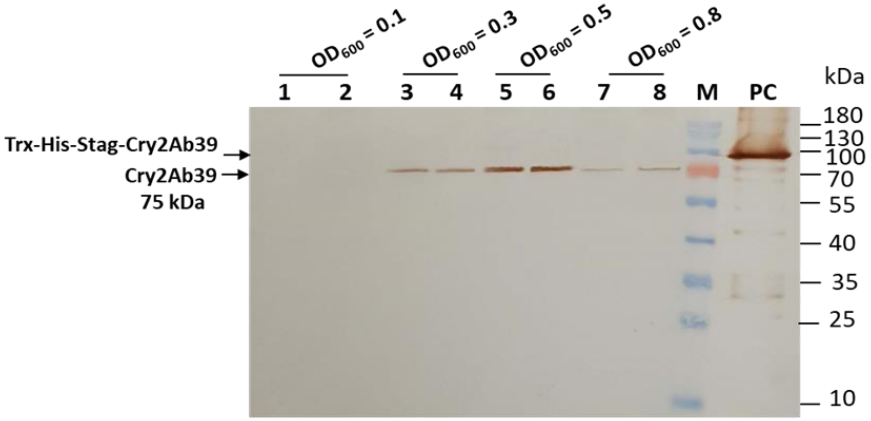


Figure 3.28. Analysis of the effect of *Agrobacterium* infection density on the transient expression of the *cry2Ab39opt* gene in *N. benthamiana* by Western blot using anti-6X His antibody

Optimization of harvest time post-transformation

Leaf samples were collected at seven different time points: 1, 2, 3, 4, 5, 6, and 7 days post-infection. The expression of recombinant *Cry2Ab39* was detectable starting from the second day post-transformation. The expression level of this protein increased gradually over the following days, peaking on the sixth day. However, by the seventh day post-transformation, the amount of recombinant protein had slightly decreased (Figure 3.29). Therefore, the sixth day post-transformation was identified as the optimal harvest time to obtain the highest yield of *Cry2Ab39* protein.

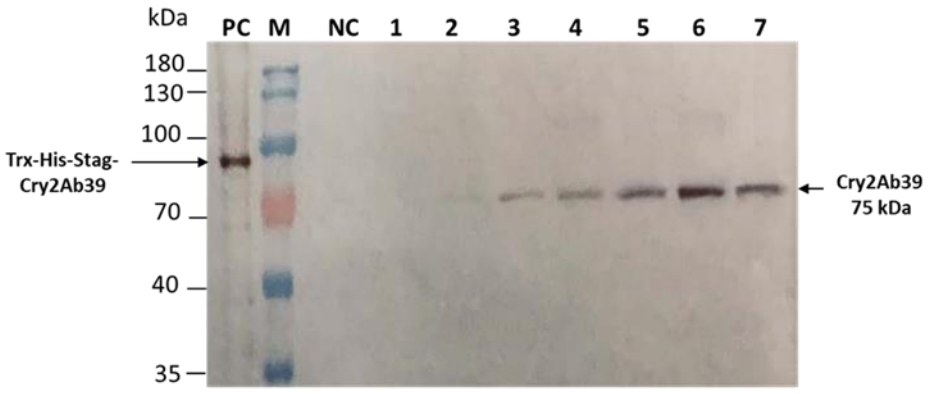


Figure 3.29. Analysis of Cry2Ab39 protein expression in *N. benthamiana* leaves after transformation days by Western blot using anti-6X His antibody

Purification of recombinant *Cry2Ab39* protein from tobacco leaves

Cry2Ab39 was purified using affinity chromatography with a Ni2+-charged column. The analysis (Figure 3.30) showed successful purification of *Cry2Ab39* protein, with an approximate size of 75 kDa.

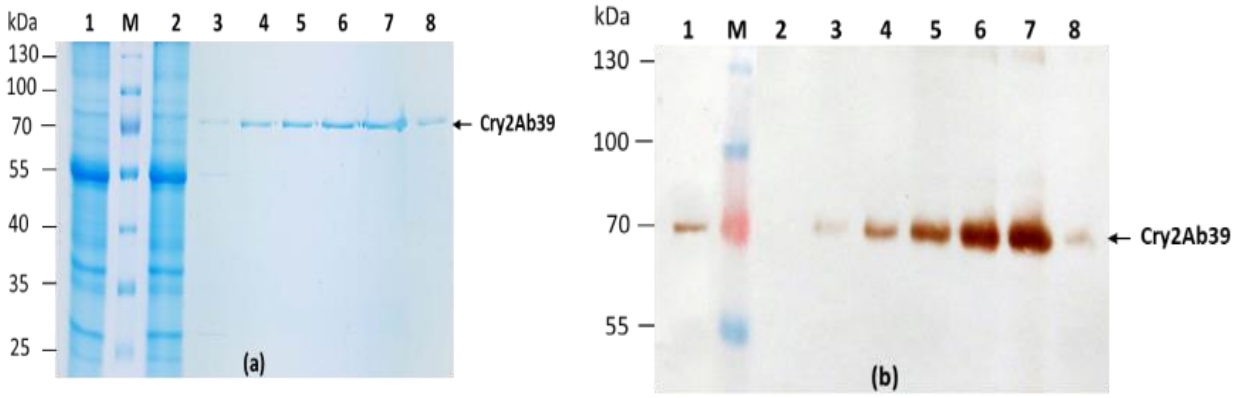


Figure 3.30. Analysis of *Cry2Ab39* purification by SDS-PAGE (a) and Western Blot (b)

The purified recombinant protein had a purity greater than 90%, with nearly all contaminating proteins from the crude leaf extract removed. Based on protein concentration measurements post-purification, the recovery efficiency of *Cry2Ab39* from *Nicotiana benthamiana* leaves was estimated to be approximately 1.5 g of protein per kg of fresh leaves

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

1. 221 *Bt* strains were activated, cultured, and screened for their activity against soybean pod borer *E. zinckenella* Treitschke. Four strains, namely SP14.2, Đ6.1, Đ6.2, BD8.2, with high insecticidal activity (over 85%) were selected.
2. The genomes of the four strains SP14.2, Đ6.1, Đ6.2, and BD8.2 were sequenced, assembled, and annotated with sizes ranging from 6.8 to 7.1 Mb. Six potential new genes involved in soybean pod borer control were identified: *cry1Aa*, *cry1Ac*, *cry1Be*, *cry1Na*, *cry2Ab*, and *cry2Ah*.
3. Four toxin protein genes *cry1Be, cry1Na, cry2Ab, and cry2Ah* were isolated and cloned. The recombined Cry proteins expressed in *E. coli* showed insecticidal activity against soybean pod borer *E. zinckenella* ranging from 66.67% to 96.67%, with Cry2Ab exhibiting the highest efficacy (96.67%).
4. The recombinant protein Cry2Ab39 was successfully expressed as a soluble form in *E. coli*. Cry2Ab39 protein exhibited toxicity against second-instar larvae with an average lethal concentration LC50 = 1.74 µg protein/g food.
5. The *cry2Ab39* gene was codon-optimized for expression in the tobacco system. A successful procedure was established for transient expression of Cry2Ab39 protein in *Nicotiana benthamiana* leaves using the agroinfiltration method under optimized conditions.

RECOMMENDATIONS

1. Further research should focus on evaluating the expression of the *cry2Ab39* gene in soybean plants as a basis for developing genetically modified soybean varieties resistant to soybean pod borers.
2. Conduct additional studies to fully assess the insecticidal spectrum and toxicity of the new toxin Cry2Ab39 against other target insect species for effective application.

NOVEL CONTRIBUTIONS OF THE THESIS

1. Deciphered the genomes of four native *Bt* strains in Vietnam. From the genomic data of these four strains, four insecticidal toxin genes predicted to have novel properties were identified, including two genes belonging to the cry1 group (*cry1Na, cry1Be*) and two genes belonging to the *cry2* group (*cry2Ab, cry2Ah*)

2. Isolated, cloned, and expressed a novel *Bt* toxin gene from the SP14.2 strain, a native strain in the Vietnam *Bt* collection. The sequence of this gene was deposited in the GenBank database under the accession number MN319700.1, and the novel toxin protein was designated as Cry2Ab39 by the International Union of Toxin Nomenclature. The *Bt* strain SP14.2 carrying the *cry2Ab39* gene has been patented by the National Office of Intellectual Property of Vietnam (patent no. 37733).

3. The recombined Cry2Ab39 protein was successfully expressed and purified from *E. coli* BL21, demonstrating its ability to control soybean pod borer E. zinckenella Treitschke larvae with an LC50 of 1.74 µg/g food after 7 days.