In-Silico Analysis for *cry*I Gene Amplification from *Bacillus thuringiensis*

Febriana Dwi Wahyuni[1*], Henny Saraswati[1], Kartika Sari Dewi[2]
[1]Faculty of Health Sciences, Universitas Esa Unggul, Jakarta 11510, Indonesia
[2]Research Center for Biotechnology, Indonesian Institute of Science, Cibinong 16911, Indonesia
email: hennysaraswati@esaunggul.ac.id

Abstract

Bacillus thuringiensis is one type of bacteria that has been used as a microbiological control agent for pests and a vector of plant disease. The presence of *Cry* proteins inside the *B. thuringiensis* can be acted as a specific insect repellent that only toxic to certain insects. The *Cry*I protein is toxic to Lepidoptera insects, which can attack various types of plants. Polymerase Chain Reaction (PCR) is a standard method that can be used to amplify the gene encoding *Cry*I proteins from *B. thuringiensis*. This research aimed to design a primer candidate for *cry*I gene amplification from *B. thuringiensis*. *In silico* analysis for designing *cry*I primer was carried out using some software, such as BLAST for searching *cry*I gene sequence, Bioedit for sequences alignment, and DINAmelt for analyzing dimer structure of the primers. Ten primer candidates were successfully obtained based on the result of the primer3 software. A pair of primer was selected to amplify the *cry*I gene, with a forward primer 5'- CGGTGAATGCCCTGTTTACT -3' and reverse primer 5'- CGGTCTGGTTGCCTATTGAT -3'. Amplification of the *cry*I gene by PCR method using selected primer resulting in a PCR product with a length of approximately 200 bp.

Keywords: *Bioinformatics, cry gene, PCR, primer, in silico analysis, Bacillus thuringiensis*

1. INTRODUCTION

One of many barriers in the production of a commodity in tropical and humid countries is the attack of plant-disturbing organisms such as insect pests and plant pathogens. Even in the plant such as rice, insects are still a significant obstacle and become a severe problem, for example, brown plant hopper and stem borer (Herman, 2002).

The control effort done by farmers is spraying insecticides. The use of this chemical insecticides has allegedly harmed the environment and caused resistance to target insects. Concerns about the negative impact of synthetic pesticides have increased public attention to bioinsecticides as an alternative for reducing pest populations (Bahagiawati, 2003).

Bacillus thuringiensis (Bt) produces crystalline proteins that are toxic to insects and nematodes during sporulation (Saraswati, 2007). *B. thuringiensis* bioinsecticides constitute 90-95% of the commercialized biopesticides used by farmers in various countries. The advantage of using *B. thuringiensis* is a specific buffer called a *Cry* protein (from the word *cry*stal) or also called δ-endotoxin.

Cry protein is only toxic to certain types of insects and is not toxic to useful insects or other organisms (Saraswati, 2007). A kind of *Cry* protein, *Cry*I, is known to kill insects from the Lepidoptera group (Suryanto, 2017). Lepidoptera is a group of pests that attack various plants.

The development of recombinant DNA technology is now possible to carry out the transformation of *cry* genes into plant cells. The introduction of *cry* genes into plant genomes is expected to produce plants capable of expressing its endotoxins. Therefore, it will reduce the use of harmful insecticides.

Advances in technology in the field of molecular biology have encouraged scientists to isolate chromosomes or plasmids contain *cry* genes. The *cry* gene is a DNA sequence that codes for the expression of *Cry* proteins (Saraswati, 2007). Several studies have been done for *cry* gene amplification (Ceron *et al.*, 1995; Malik *et al.*, 2013). Polymerase Chain Reaction (PCR) is one method for *cry* gene amplification.

Bioinformatics is an interdisciplinary field broadly defined as a combination of biology (molecular biology) and computing using computer and software assistance. One of the most significant functions of bioinformatics is to design a primer sequence.

Primer is a component that plays an essential role in the PCR process (Septiari *et al.*, 2015). A good primer recognizes specific sequences on a genome or template and is restricted only to a target sequence. Unspecified primers cause amplification of other regions in the genome that are not targeted, or otherwise, there are no amplified genome regions. To get a primer that fulfills the right criteria for amplification, *in silico* design is carried out (Suryadi *et al.*, 2014). This study aimed to analyze the combination of primer pairs used in the PCR process to amplify the *cry*I gene from *B. thuringiensis*. The *cry*I gene that will be amplified is a domain of the whole *cry*I gene sequences that already exist in the genebank.

2. RESEARCH METHOD

2.1. *Bacillus thuringiensis* Isolate

The culture of *B. thuringiensis* was obtained from the Indonesian Culture Collection (InaCC) in the Indonesian Institute of Sciences with code number B432.

2.2. Bacterial DNA Extraction

B. thuringiensis DNA extraction method was used the gSYNC DNA extraction kit from Geneaid.

2.3. Target Gene Sequences

The target gene used in this research is the *cry*I gene from *B. thuringiensis* (accession number X56144). These sequences were obtained from the NCBI website (National Center for Biotechnology Information, www.ncbi.nlm.nih.gov).

2.4. BLAST Analysis

The gene then was analyzed with Basic Local Alignment Search Tools (BLAST) software with other *B. thuringiensis* strains so that identity values from the highest to the lowest were obtained. Gene sequences that have high identity values are used in the multiple alignment method to obtain conserved sequences.

2.5. Multiple Alignments

The determination of sustainable areas is determined using the Bioedit program. The sequence of bases obtained is aligned with the ClustalW program. Then the one which has the most similarity in its constituent bases is similar to 90-100% as a basis for tracking primary attachments (Saraswati *et al.*, 2019).

2.6. Design of Primer

Determination analysis of DNA primer was performed with Primer3 software and manually selected the most sustainable areas by considering the area is the cipher sequence, with general requirements, including the number of nucleotides, GC content. There is no possibility of complementarity between bases in one primary chain or between one primer with another.

2.7. Analysis of Primer Dimer

Sometimes the designed primers recognize the sequence of themselves, bind to one another to form a structure called a dimer. It can be a problem because instead of binding with the target gene, Dimer will tend to stick together, and thus reduce DNA concentration. For this reason, it is necessary to do a dimer prediction analysis using available software to predict the presence of dimers in primary candidates, namely DINAmelt (http://unafold.rna.albany.edu/?q=DINAMelt) made by NM Markham and Michael Zucker of Rensselaer Polytechnic Institute (Zuker, 2003; Markham and Zuker, 2005).

2.8. Analysis of Restriction Sites in the *cry*I Gene

An analysis to find out the restriction sites contained in the *cry*I was performed using *Snapgene* software. The purpose of knowing the restriction site is so that genes can be cut with one of the desired endonuclease restriction enzymes.

2.9. Amplification of the *cry*I gene by PCR

Amplification of the *cry*I gene using specific primer *cry*1 forward 5'-CGGTGAATGCCCTGTTTACT-3' and

reverse 5'-CGGTCTGGTTGCCTATTGAT-3' with a total volume of 25 µl containing 1 µl DNA genome, 12.5 µl PCR HS redmix master mix (Bioline), 1 µl for each primer and 9.5 ddH2O. PCR amplification was carried out as many as 35 cycles using Sensoquest. One cycle consists of 3 stages, namely denaturation, attachment, and elongation. The pre-denaturation stage is carried out for 3 minutes at a temperature of 95°C once, the denaturation stage is 1 minute, and the primer attachment stage is at 50-60°C for 1 minute, the extension of the DNA chain at a temperature of 72°C for 1 minute. In the last cycle, the chain was extended longer at 72°C for 5 minutes.

2.10.PCR Electrophoresis

The PCR products were migrated into 1% agarose gel under 100 volt for 30 minutes. A 1 kb DNA ladder is used as a marker. Fluorosafe DNA stain was used for gel staining, which is directly added to the agarose gel. A Gel containing DNA fragments was visualized using UV Trans Illuminators and documented using the Digibox Camera Documentation System gel.

3. RESULTS

3.1. The *Cry*I Gene Sequences

Based on the search results of the *cry*I gene sequences in genebank with accession number X56144, *cry*I gene sequences were obtained from *B. thuringiensis* consisting of 3516 bp.

3.2. Basic Local Alignment Search Tools (BLAST)

The BLAST process was carried out using *B. thuringiensis cry*I gene for insecticidal *cry*stal protein as a determinant strain with other *B. thuringiensis* strains (Figure 1). It is because it has a long DNA sequence. BLAST results obtained an identity value of a 100%-96% *cry*I gene against several strains of *B. thuringiensis* (Figure 2).

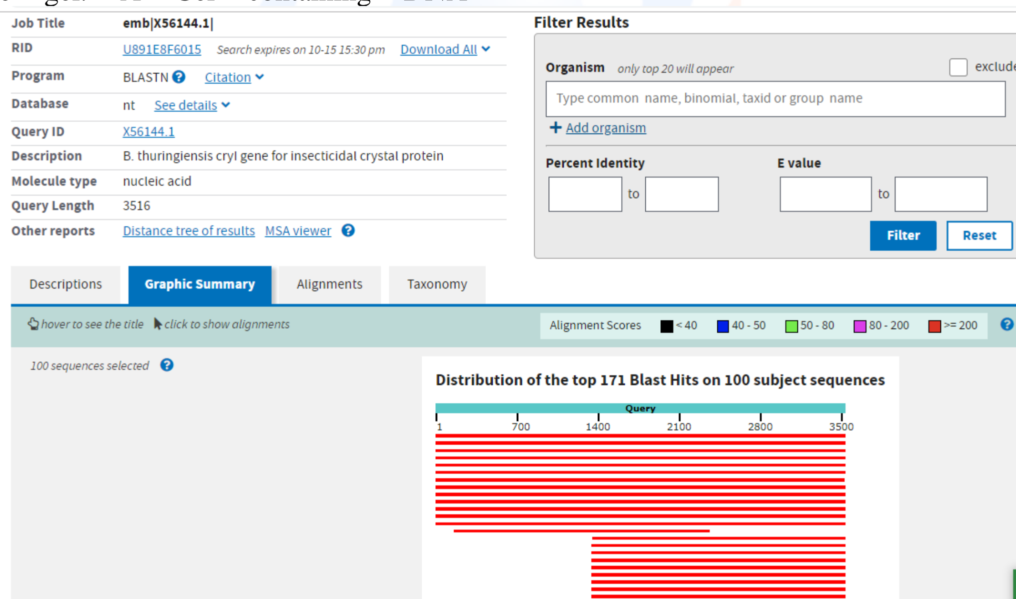
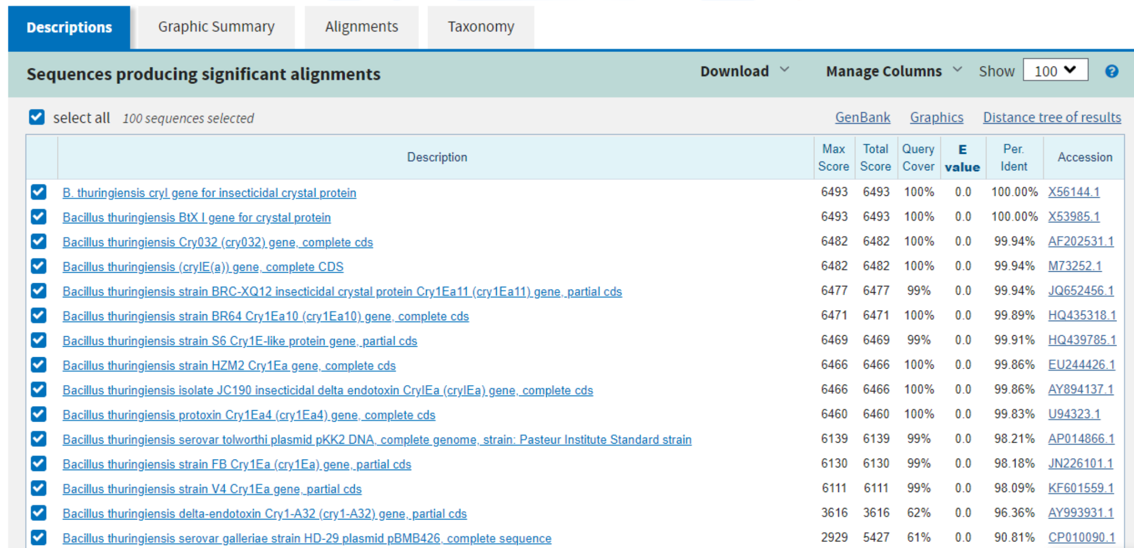


Figure 1. BLAST *cry*I gene with several strains of *Bacillus thuringiensis*



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download — Manage Columns — Show 100

select all 100 sequences selected — GenBank — Graphics — Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
B. thuringiensis cryI gene for insecticidal crystal protein	6493	6493	100%	0.0	100.00%	X56144.1
Bacillus thuringiensis BtX I gene for crystal protein	6493	6493	100%	0.0	100.00%	X53985.1
Bacillus thuringiensis Cry032 (cry032) gene, complete cds	6482	6482	100%	0.0	99.94%	AF202531.1
Bacillus thuringiensis (cryIE(a)) gene, complete CDS	6482	6482	100%	0.0	99.94%	M73252.1
Bacillus thuringiensis strain BRC-XQ12 insecticidal crystal protein Cry1Ea11 (cry1Ea11) gene, partial cds	6477	6477	99%	0.0	99.94%	JQ652456.1
Bacillus thuringiensis strain BR64 Cry1Ea10 (cry1Ea10) gene, complete cds	6471	6471	100%	0.0	99.89%	HQ435318.1
Bacillus thuringiensis strain S6 Cry1E-like protein gene, partial cds	6469	6469	99%	0.0	99.91%	HQ439785.1
Bacillus thuringiensis strain HZM2 Cry1Ea gene, complete cds	6466	6466	100%	0.0	99.86%	EU244426.1
Bacillus thuringiensis isolate JC190 insecticidal delta endotoxin CryIEa (cryIEa) gene, complete cds	6466	6466	100%	0.0	99.86%	AY894137.1
Bacillus thuringiensis protoxin Cry1Ea4 (cry1Ea4) gene, complete cds	6460	6460	100%	0.0	99.83%	U94323.1
Bacillus thuringiensis serovar tolworthi plasmid pKK2 DNA, complete genome, strain: Pasteur Institute Standard strain	6139	6139	99%	0.0	98.21%	AP014866.1
Bacillus thuringiensis strain FB Cry1Ea (cry1Ea) gene, partial cds	6130	6130	99%	0.0	98.18%	JN226101.1
Bacillus thuringiensis strain V4 Cry1Ea gene, partial cds	6111	6111	99%	0.0	98.09%	KF601559.1
Bacillus thuringiensis delta-endotoxin Cry1-A32 (cry1-A32) gene, partial cds	3616	3616	62%	0.0	96.36%	AY993931.1
Bacillus thuringiensis serovar galleriae strain HD-29 plasmid pBMB426, complete sequence	2929	5427	61%	0.0	90.81%	CP010090.1

Figure 2. Analysis of the *cry*I gene identity with the *cry*I gene from another *Bacillus thuringiensis* strain

3.3. Multiple Alignment

From several strains of BLAST results that have 100-96% DNA sequence identity values, multiple alignments are performed, so that several species with DNA sequences of *cry*I genes that have parallel DNA sequences are obtained (Figure 3). Sequences selected parallel to the CDS region that corresponds to the DNA primer sequence that has been designed with Primer3 software.

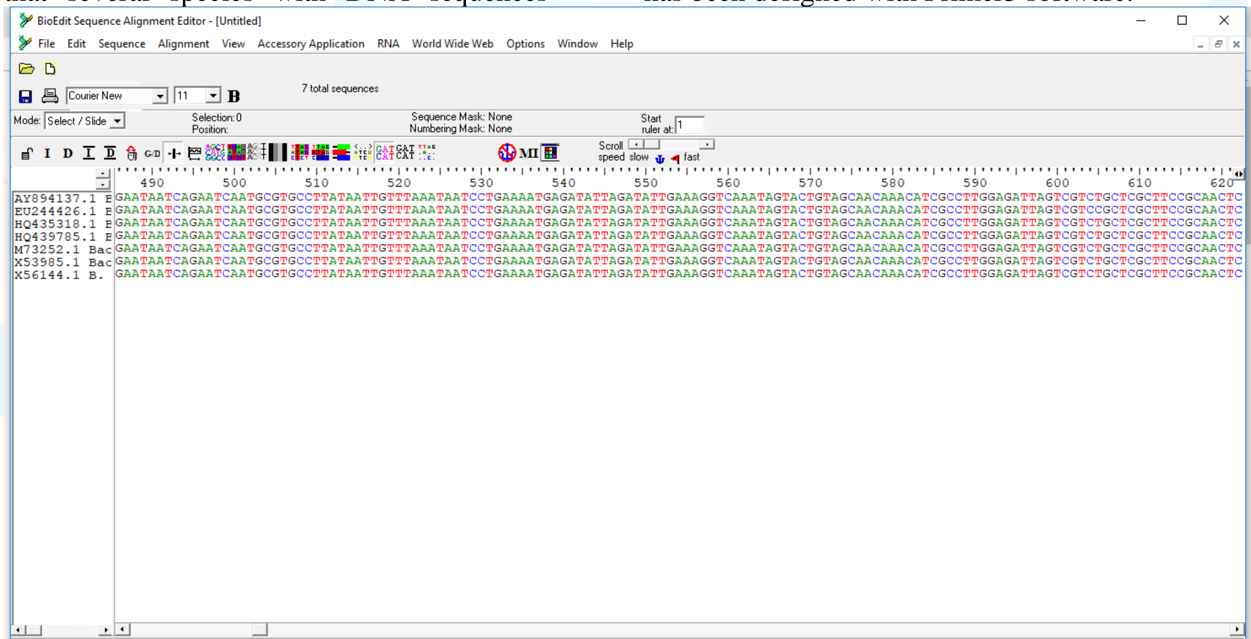


Figure 3. The multiple alignment process of *cry*I gene from *B. thuringiensis* with other strain that has 100-96% DNA sequence identity using Bioedit software

Table 1. *cry*1 Primer design from *B. thuringiensis*

Primer (*cry*I)	Sequence (5'-3')	Tm (^{0}C)	GC%	Self 3' complementarity
Forward	CGGTGAATGCCCTGTTTACT	60	50	1
Reverse	CGGTCTGGTTGCCTATTGAT	60	50	2

3.4. DNA primer *cry*I gene

The design of the DNA primer of the *cry*I gene from Bacillus thuringiensis uses primer3 software that has passed the BLAST process and multiple alignments (Table 1).

3.5. Dimer Structure in Primer Candidate

The results of the analysis with DINAmelt software shown that the designed primers only have 2 G-C bonds (Figure 4). It is advantageous because it can increase the concentration of PCR products. On the other hand, the possibility for the primer to bond to each other is minimal.

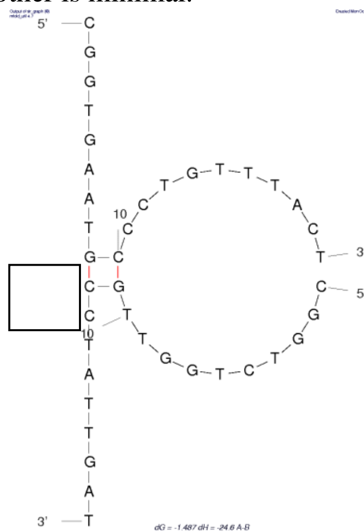


Figure 4. This figure shows DINAmelt analysis results to predict dimer structure on primer pair previously designed

3.6. Map of *cry*I gene

The results of the snapgene analysis showed that the amplified *cry*I gene had eight restriction sites (Figure 5). This information can be used as a basis for cloning.

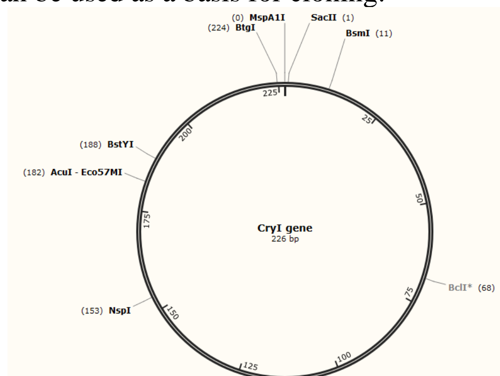


Figure 5. Mapping Restriction Enzymes in *cry*I Genes with Snapgene software

3.7. PCR Electrophoresis

A DNA visualization results of PCR product that amplified using *cry*1 forward and *cry*1 reverse primers showed DNA bands with a size of approximately 200 bp (Figure 6).

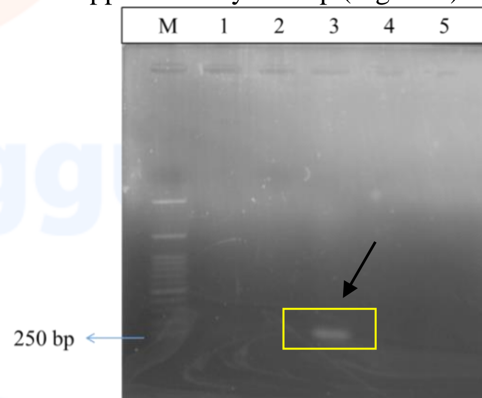


Figure 6. Amplification of *cry*I gene; M: 1 kb DNA ladder, 1: PCR with annealing 50°C; 2: PCR with annealing 52°C; 3: PCR with annealing 55°C, 4: PCR with annealing 60°C, 5: negative control.

4. Discussion

The design of primers is the first step that determines the success of DNA amplification by the PCR method (Suparman *et al.*, 2016). Criteria need to be considered in the selection of primers include the length of the primer, temperature melting (Tm), GC content, and bonds at the end of 3'. A good primer has a range from 18-30 base pairs. Primers that have a length of more than 30 base pairs will cause the unspecific attachment. A Pair of selected primers from the design in this research has a length of 20 base pairs for each primer, with a forward primer 5'- CGGTGAATGCCCTGTTTACT-3' and reverse primer 5'- CGGTCTGGTTGC CTATTGAT-3'. The primer can amplify the nucleotide bases with a PCR product size of 200 base pairs.

The second characteristic to consider in primer selection is Tm. A pair of suitable primers have a Tm difference of around 5°C between forward and reverse. It is intended to prevent a decrease in the amplification process. The percentage between bases G and C also needs to be considered because the content of the number of bases G and C can affect the Tm of a primer (Dewi *et al.*, 2018). A good primer has a percentage of G and C

around 40-60%. Another criterion for suitable primers is having a low self 3' complementarity so that there is no attachment between primer pairs and forms a structure called a hairpin (Sasmito *et al.*, 2014).

Dimers are structures that formed between primer pairs. They are united because they have a complementary base. This process occurs at the appropriate attachment temperature, usually at low temperatures. By looking at the stages of the PCR process, it can be seen that the primer attachment to the template DNA occurs at optimal annealing temperatures. This process can coincide with the formation of dimers. The problem that may arise is that primers tend to anneal to one another, and not stick to template DNA. If this dimer bond is too strong, it will interfere with the DNA extension process and will result in low DNA concentration. From the analysis using DINAmelt, it can be seen that the primer forms a dimer with two GC bonds. The bond between bases G and C is strong because it consists of 3 hydrogen bonds. That will make the primer easier to put together.

In silico analysis is an essential computational prediction in design primer. Primers must also be tested through a series of optimizations in the laboratory. Primary candidate optimization involves optimization in annealing temperature (Ta) using the PCR gradient and optimization of primary concentration. In addition to primers, optimization of the PCR reaction is also carried out to check the minimum detection and quantification of nucleic acids in the reaction, and this requires work in the laboratory to produce a good PCR test.

5. CONCLUSION

The primers designed with a forward primer 5'- CGGTGAATGCCCTGTTTACT - 3' and reverse primer 5'-CGGTCTGGTTGC CTATTGAT -3' amplified the *cry*I gene with molecular size approximately 200 base pairs.

6. ACKNOWLEDGMENT

The authors would like to thank to Yayasan Pendidikan Kemala Bangsa, Universitas Esa Unggul and Directorate General of Strengthening Research and Development, Ministry of Research, Technology and Higher Education, the Republic of Indonesia through the Higher Education Applied Research Grant Scheme.

7. REFERENCES

Bahagiawati (2003) 'Penggunaan Bacillus Thuringiensis sebagai biolarvasida', *Buletin AgroBio*, 5(1), pp. 21–28.

Ceron, J. *et al.* (1995) 'Specific PCR primers directed to identify cryI and cryIII genes within a Bacillus thuringiensis strain collection', *Applied and Environmental Microbiology*, 61(11), pp. 3826–3831.

Dewi, R. W. *et al.* (2018) 'Desain Primer Untuk Amplifikasi Regio Promoter Gen inh A Isolat P016 Multidrug Resistance Mycobacterium tuberculosis Dengan Metode Polymerase Chain Reaction', *Jurnal Farmasi Udayana*, 7(1), pp. 34–39.

Herman, M. (2002) 'Perakitan Tanaman Tahan Serangga Hama melalui Teknik Rekayasa Genetik', *Balai Penelitian Bioteknologi dan Sumberdaya Genetik Pertanian, Bogor*, 5(1), pp. 1–13.

Malik, K., Muhammad, M. and Talpur, A. (2013) 'Cloning and Expression of a cry III Gene Isolated from the Local Habitat into a Modified Strain of Bacillus thuringiensis,' *IOSR Journal of Pharmacy and Biological Sciences*, 7(5), pp. 87–95.

Markham, N. R., and Zuker, M. (2005) 'DINAMelt web server for nucleic acid melting prediction,' *Nucleic Acid Research*, 33, pp. 577–581. DOI: 10.1093/nar/gki591.

Saraswati, H. and Wahyuni, F. D. (2019) 'Desain Primer Secara In Silico untuk Amplifikasi Gen cryIII dari Bacillus thuringiensis Isolat Lokal', *Indonesian Journal of Biotechnology and Biodiversity*, 3(1), pp. 33–38.

Saraswati, N. P. A. (2007) 'Deteksi dan Identifikasi Gen Cry4 pada Isolat Bacillus thuringiensis Daerah Bogor dan Sekitarnya', *Institut Pertanian Bogor, Bogor*.

Sasmito, D. E. K., Kurniawan, R. and Muhimmah, I. (2014) 'Karakteristik Primer

pada Polymerase Chain Reaction (PCR) untuk Sekuensing DNA : Mini Review', *Seminar Nasional Informatika Medis (SNIMed)*, pp. 93–102.

Septiari, I. G. A. A., Yustiantara, P. S. and Yowani, S. C. (2015) 'Analisis Primer untuk Amplifikasi Promoter inhA Multidrug Resistance Tuberculosis (MDR-TB) dengan Metode Polymerase Chain Reaction (PCR)', *Jurnal Kimia*, 9(1), pp. 117–123.

Suparman, Ahmad, H. and Ahmad, Z. (2016) 'Desain Primer PCR Secara in silico untuk Amplifikasi Gen COI pada Kupu-kupu Papilio ulysses Linnaeus dari Pulau Bacan', *Jurnal Pendidikan Matematika dan IPA*, 7(1), pp. 14–25.

Suryadi, P. T., Ratnayani, K. and Yowani, S. C. (2014) 'Desain Primer untuk Amplifikasi Gen katG Multidrug Resistance Tuberculosis (MDR-TB) dengan Metode Polymerase Chain Reaction (PCR)', *Jurnal Kimia*, 8(1), pp. 77–82.

Suryanto, D. (2017) 'Amplifikasi gen Cry1 dan Analisis Genom isolate Bacillus thuringiensis Lokal', *Journal of Biological Researches*, 15(1), pp. 1–4. DOI: 10.23869/bphjbr.15.1.20091.

Zuker, M. (2003) 'Mfold web server for nucleic acid folding and hybridization prediction,' *Nucleic Acids Research*, 31(13), pp. 3406–3415. DOI: 10.1093/nar/gkg595.