JOIV : Int. J. Inform. Visualization, 5(2) - June 2021 139-143

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Bioinformatic Analysis in Designing Mega-primer in Overlap Extension PCR Cloning (OEPC) Technique

Mardalisa[a,*], Sony Suhandono[b], Novi Yanti[c], Fazrol Rozi[d], Fitri Nova[d], Primawati[e]

[a]*Department of Marine Science, Faculty of Fisheries and Marine, Riau University, 28293, Indonesia*
[b]*Research Expertise Group of Genetics and Molecular Biotechnology, School of Life Science and Technology, Institut Teknologi Bandung, 40116, Indonesia*
[c]*Department of Information Technology, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau, 28293, Indonesia*
[d]*Department of Information Technology, Politeknik Negeri Padang, 21562, Indonesia*
[e]*Department of Mechanical Engineering, Faculty of Engineering, Universitas Negeri Padang, 25131, Indonesia*
Corresponding author: [*]mardalisa@lecturer.unri.ac.id

***Abstract*— Bioinformatics has developed into an application tool for basic and applied research in the biomedical and biotechnology field. Polymerase Chain Reaction (PCR) is a common technique in the molecular area that has always involved bioinformatics science. PCR cloning techniques such as TA cloning and PCR-mediated cloning exhibit complex processes with low success rates. One easy, effective, and practical solution is to use a mega-primer with the Overlap Extension PCR Cloning (OEPC) technique. The success of PCR cloning using the mega-primer design in the OEPC technique is strongly influenced by the characteristics of the mega-primer used. Knowledge of mega-primer characteristics is one of the important factors in the success of PCR cloning. The design process for the mega-primer str promoter was characterized based on the principle of a genetic algorithm using the web-based bioinformatics tools such as ClustalW, NetPrimer, and BLAST. The success of the mega-primer construction in producing recombinant pSB1C3 vector has been confirmed by the sequencing method and the function of the reporting protein (AmilCP). DNA analysis shows a 100 % homologous sequence on the str promoter, while *E. coli* colonies successfully express the purplish-blue color. Mega-primer characters can save costs and time of the research by maintaining the primer parameters that provide optimal values and increase the success value of PCR cloning via bioinformatics software. Hence, implications on biological problems, especially using DNA and amino acid sequences, could solve rapidly.**

***Keywords*— Bioinformatics; mega-primer design; Overlap Extension PCR Cloning (OEPC); genetic algorithm.**

Manuscript received 20 Nov. 2020; revised 26 Dec. 2020; accepted 5 Mar. 2021. Date of publication 30 Jun. 2021.

I. INTRODUCTION

The world today shows progress in the field of computerization and molecular biology. Bioinformatics is a multidisciplinary field of biology, computer science, informatics, and mathematics that supports the development of future biotechnology. Bioinformatics applies methods for solving biological problems, especially using DNA and amino acid sequences and related information. Currently, the most commonly applied molecular analysis technique is the PCR (Polymerase Chain Reaction) technique. PCR is a diagnostic technique that requires very high specificity and sensitivity [1]. PCR is commonly used for a variety of purposes, such as gene cloning, hereditary diseases detection, fingerprints DNA, diseases diagnosis, paternity testing, and DNA computation [2], [3].

The Overlap Extension PCR Cloning (OEPC) technique is a variant of PCR. Overlap Extension PCR Cloning (OEPC) consists of two stages: overlap-extension PCR and PCR cloning [4-6]. Overlap-extension PCR functions to produce DNA fragments (mega-primer) which have homologous tail sequences at the 5 'and 3' ends of the plasmid vector. Meanwhile, PCR cloning was used to enter the mega-primer results of the overlap-extension PCR into the plasmid vector. Mega-primer contains homologous gene sequences at the 3 'end and homologous vector sequences at the 5' end in both forward and reverses primers [7]. The mega-primer attaches to the complement sequence and starts the cloned DNA synthesis process in the overlapping area.

Primers are DNA sequences that complement the target DNA sequence and one of the determinant factors in PCR. The primer design can utilize sequence codes that are readily

available on the internet on several sites, like GenBank (www.ncbi.nlm.nih.gov) which includes sequence data collection in America, European Molecular Biology Laboratory - European Bioinformatics Institute / EMBL-EBI (www.ebi.ac.uk) covers Europe [8], and DDBJ (www.ddbj.nig.ac.jp) in Japan. Bioinformatic analysis using software (software) such as ClustalW, NetPrimer, and BLAST can characterize primers based on their genetic algorithms [9, 10].

The characteristics of the mega-primer designed influence the success in designing mega-primer in OEPC reactions. The good mega-primer characters are determined by the primer length, primer melting temperature (Tm), primer annealing temperature (Ta), the difference between primer melting temperature (ΔTm) and GC content, GC clamp, secondary structure, repeats and run, specificity and uniqueness, and product length [11, 12]. PCR cloning such as TA cloning and PCR-mediated cloning shows a complicated process and low success rates. The presence of mega-primer with the OEPC technique is an easy, effective, and practical solution [13].

The mega-primer character of the str promoter that was characterized was constructed into the pSB1C3 - Bba_K592025 vector. The pSB1C3 - Bba_K592025 vector part has a pMBI ori which is responsible for the replication process, including the replication of the CAT gene which causes recombinant *Escherichia coli* to become resistant to chloramphenicol antibiotics (Fig. 1). The success of the recombinant pSB1C3 vector was confirmed by restriction analysis and reporting protein function (AmilCP).

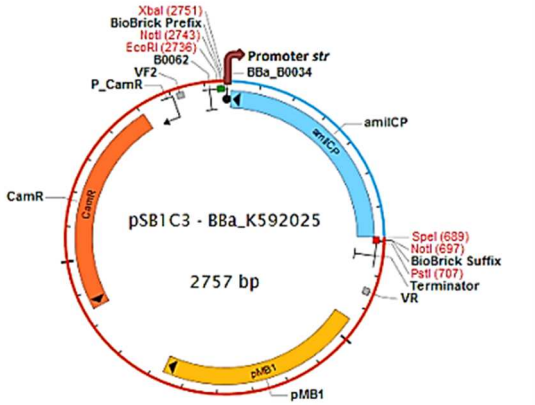

Fig. 1 Location of the mega-primer insertion to pSB1C3 - Bba_K592025) vector

II. Material and Method

A. Primer Design via Bioinformatic Software

The str promoter sequences of *E. coli* were obtained from the NCBI GenBank website. Sixteen sequences of the str promoter [14] obtained from various *E. coli* strains were downloaded in Fasta format and then carried out the alignment process using the CluswalW software (http://www.ebi.ac.uk /clustalw /index.html) [15]. The results of the analysis show that the * sign is the conserved sequence area in all strains of *E. coli* bacteria. Forward and reverse primer candidates were selected in the conserved portion of 20 nucleotides. The eligibility of primer candidates was tested with NetPrimer software (http://www.premierbiosoft.com/ NetPrimer/AnalyzePrimer.jsp) to see the characteristics of the possible occurrence of hairpin, dimer, cross dimer, palindrome, and repeat and run structures [16, 17]. The best primer selected will be added to the tail of 20 nucleotides as an overlapping area in the pSB1C3 - Bba_K592025 vector.

B. Mega-primer Production for Overlap-Extension PCR

The mega-primer design contains a homologous sequence of the str promoter at the 3 'end and a vector homologous sequence at the 5' end in both the forward and reverses primers. Mega-primer is nucleotide sequences with forward and reverse directions that overlap with the str promoter sequence and pSB1C3 - Bba_K592025 vector [18]. Overlapping sequences were used to produce overlap extension PCR using the KAPA Hifi HotStart PCR Kit (KAPABiosystems, USA) (Fig. 2).

C. Mega-primer Construct into The pSB1C3 - Bba_K592025 Vector

The mega-primer construction into the pSB1C3 Bba_K592025 vector was generated by the PCR cloning method. The PCR cloning method used the KAPA Hifi HotStart PCR Kit (KAPABiosystems, USA) with a mega-primer concentration of 6.5 ul in a total mixture of 20 ul for 10 PCR cycles. The results of PCR cloning were added with the DpnI enzyme, which functions to destroy vectors that do not contain mega-primer based on the presence of the methyl group. *E. coli* bacteria are transformed with the recombinant plasmid and the success of the process will result in the appearance of a purple-blue bacterial colony (AmilCP).

III. Results And Discussion

A. Primer Characteristics

Based on multiple sequence alignment analysis using the ClustalW software, results were obtained with a sequence length of 413 bp. Determination of primer candidates forward (5'-TACCTTGAAAGTCTGCGTAA-3') and reverse (5'-AAA TAGCTCCTGGTTTTAGC-3') based on the conserved area with the symbol * (Fig. 3) and the following parameters [19-21].

1) Primer Length: The primer candidate has a length of 20 base pairs (bp) consisting of 4 combinations of DNA letters (ATGC). A good primer length generally consists of 18 - 22 bp. This is based on the consideration of random combinations that may be found in one organism's genome sequence. The probability of finding one nitrogen base (ATGC) on one base is 1/4, the probability of finding two base sequences (AG, AC, CG, etc.) is 1/16, the probability of finding four bases (ATGC, TACG, etc.) is 1/256. So that in 20 bp of primer lengths based on statistical analysis will be found once in every 1 trillion base sequences. Primer lengths with more than 30 bases are not recommended because they can produce low specificity and secondary structure formation.

2) Primer Melting Temperature (Tm): Primer Melting Temperature (Tm) is the temperature required by the primer to dissociate the bond of its DNA complement. One theory to calculate the primer melting temperature is Wallace's Formula (1). Based on Wallace's formula, it is known that the forward and reverse primers candidate calculations show a melting

temperature at 56 °C. The melting temperature 56 °C is the optimal temperature range for the DNA Taq Polimerase enzyme performance.

$$Tm = (nG+nC)*4 + (nA+nT)*2 \quad (1)$$

Tm (f) = (4+4)*4 + (6+6)*2 = 56 °C
Tm (r) = (4+4)*4 + (5+7)*2 = 56 °C

Informations:
Tm : Primer melting temperature Wallace's Formula
nG : the number of bases G in the primer
nC : the number of bases C in the primer
nA : the number of bases A in the primer
nT : the number of bases T in the primer

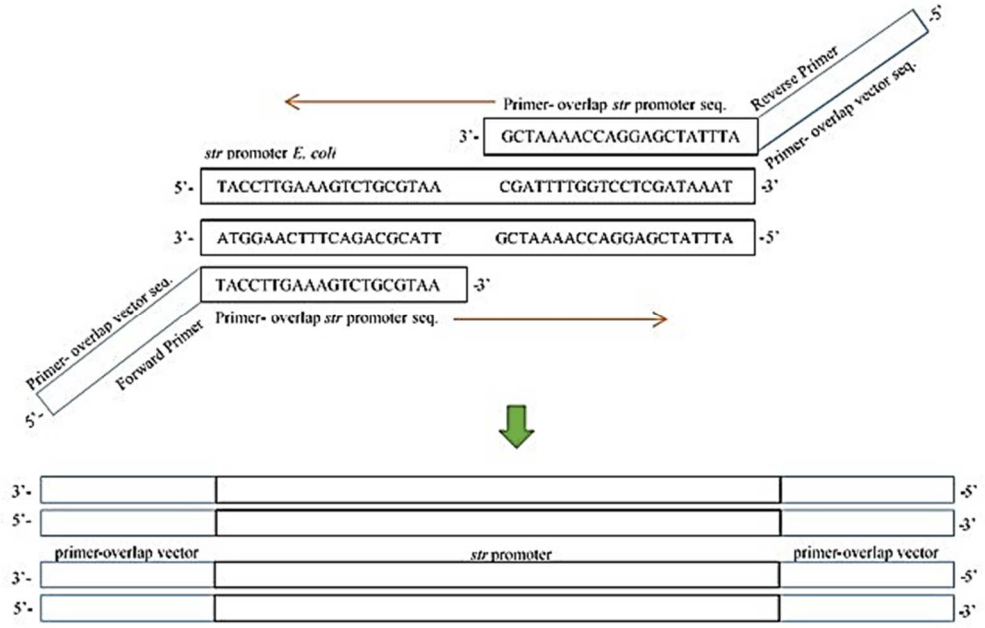

Fig. 2 The Overlap-Extension PCR technique in producing the mega-primer

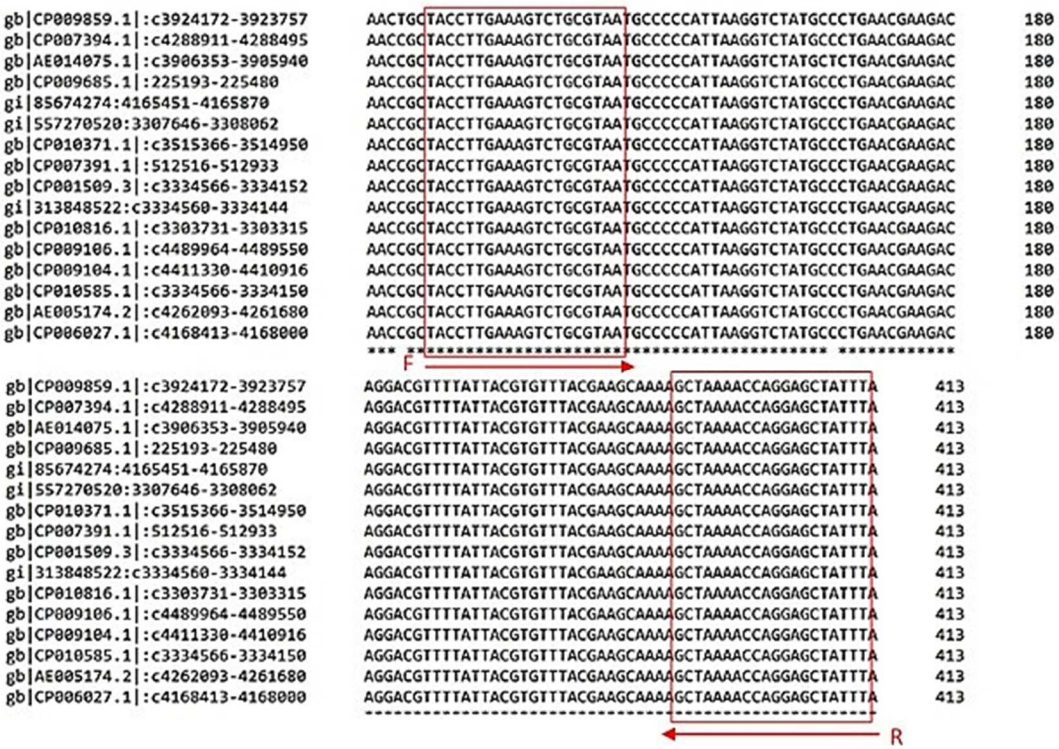

Fig. 3 Multiple sequence alignment analysis of str promoter by using ClustalW (Bioinformatic software)

3) The Difference in Primer Melting Temperature (ΔTm): A good primer pair is the one that does not have a melting temperature difference of more than 5 °C. This is based on the possibility of a failure of the amplification process. Based on Wallace's Formula calculation, it is known that there is no difference between the primer pair melting temperatures.

4) Primer Annealing Temperature (Ta): The annealing temperature is the temperature estimated when the primer can stably attach to the DNA template. The annealing temperature

value which is proportional to the melting temperature causes the annealing temperature not to be included in the calculation of the primer design optimization.

5) GC Content: The general rule followed by most primer programs is to use a base G and C percent between 40 % and 60 %.

$$\text{GC Content} = (nG+nC)/([p]) \times 100 \quad (2)$$

Informations:
nG : the number of bases G in the primer
nC : the number of bases C in the primer
[p] : Primer length

Based on this formula, it is known that the primer forward and reverse calculations show the same GC content 40% with the following calculations.

$$\text{GC Content (f \& r)} = \frac{4+4}{20} \times 100 = 40\ \%$$

6) GC Clamp: Some programs require the primer pairing with a GC base at the 3' end of the primer. This arrangement makes hybridization more stable. However, it is necessary to avoid more than 3 bases G or C in the last 5 bases of the 3′ end to prevent secondary structure formation.

7) Secondary Structure: The PCR reaction should preferably not contain secondary structures in the form of hairpins or dimers. The stability of the secondary structure is determined by free energy (ΔG) and its melting temperature. Secondary structure formation causes primer unable to attach to the template DNA. The hairpin formation is caused by the occurrence of complement in the base sequence of the primer sequence. Hairpin with free energy (ΔG) not more than -3 kcal mol^{-1} can still be tolerated. Primer sequences that bind to other sequences are called self-dimers. Primers that bind to their other primer (forward-reverse) are called Cross-Dimers. Self-dimers and cross-dimers with free energy (ΔG) below -7 kcal mol^{-1} can still be tolerated.

8) Repeat and Run: Repetition of the same base sequence must be avoided because it can cause breathing to the primer so that the process of attaching the primer to the target sequence becomes difficult.

9) Specificity and Uniqueness: A good primer consists of sequences that are unique to the DNA target so that they are not attached to other DNA sequences' locations. This character can be analyzed through BLAST-NCBI software to find out that the primers used are truly unique and do not stick to other organisms [22].

10) Product Length: The distance between the 5' ends of the two primers is known as the amplicon or product length. In general, the product length by PCR product is less than 2000 bp.

NetPrimer is a web-based analysis software that allows users to perform primer analysis. The characteristics of the primer candidates based on NetPrimer analysis can be seen in Table 1. Based on NetPrimer analysis, the melting temperature calculation is different from Wallace's formula (56 °C). The Tm calculation from NetPrimer software is based on the Nearest Neighbor Thermodynamic Formula (3) [23, 24].

$$Tm = (\Delta H\ (P)/((\Delta S(P)+R.\ln \{\gamma/4\}]) + T_{\circ} + t \quad (3)$$

Informations:
Tm : Primer melting temperature The Nearest Neighbor Thermodynamic's Formula.
P : Primer length
R : 1,987 (cal/°)
γ : 50 x 10^{-9}, molar concentration
$T_{\circ}$: -273.15 °C
t : -21,6 °C
ΔH : Enthalpy
ΔS : Entropy

TABLE I
PRIMER CANDIDATE ANALYSIS

Parameter	Primer Forward	Primer Reverse
Tm	51.21 °C	51.46 °C
ΔTm	0.25 °C	0.25 °C
GC %	40.0	40.0
GC Clamp	2	2
Hairpin (ΔG)	-	-
Self-dimer (ΔG)	-3.3 kcal mol^{-1}	-6.34 kcal mol^{-1}
Cross-dimer (ΔG)	-4.89 kcal mol^{-1}	-4.89 kcal mol^{-1}
Repeats	-	-
Run	3 bp	4 bp

The result shows that the difference in melting temperature is almost the same between the forward and backward primers. GC Clamp, repeats and run, and secondary structures (hairpin, self-dimer, cross-dimer) show good results on the other hand, although this can occur but is still within tolerable limits. The results of BLAST-NCBI analysis on the genome of *E. coli* cells showed the uniqueness of the 100 % primer-specific attachment to the str promoter region. Where the length of the resulting product measures about 289 bp, this proves that the primer candidate str already has primer characters that are suitable to be used in OEPC techniques. Then the primer candidate is given additional sequences of 20 bp to form a mega-primer.

B. Overlapping-Extension Product Confirmation

The mega-primer amplification of the str promoter has been successfully carried out with a product length of about 289 bp. The electropherogram resulting from the amplification of the mega-primer promoter str was visualized on 1 % agarose gel as shown in Fig. 4.

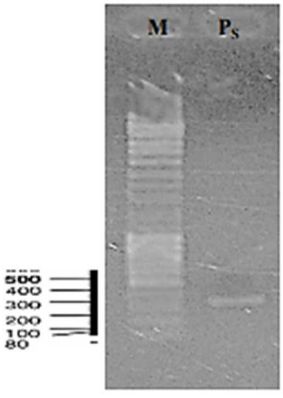

Fig. 4 The electropherogram of the mega-primer str promoter. M: massruler DNA ladder, Ps: str promoter of *E. coli*

C. PCR Cloning Product Confirmation

Confirmation of recombinant pSB1C3 vector by PCR cloning method has been successfully carried out. The PCR cloned electropherogram was visualized on 1 % agarose gel (Fig. 5).

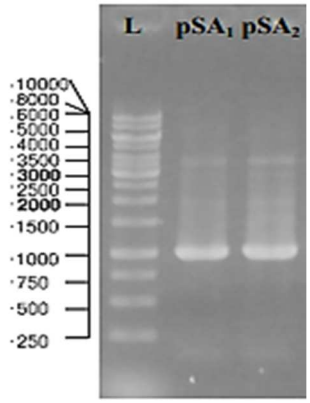

Fig. 5 The electropherogram of the PCR cloning pSB1C3 recombinant vector. L: ladder 1kb, pSA: pSB1C3 recombinant vector (str-AmilCP, 1000 bp)

The sequencing product of the str promoter and AmilCP gene was analyzed using ClustalW and BLASTN software. The results showed that the homology level was 100 % with the str promoter of *E. coli* and 99 % with the AmilCP gene of *Acropora millepora*. The ability of the str promoter to express the reporting protein AmilCP (purplish-blue color) [12] can be seen in Fig. 6.

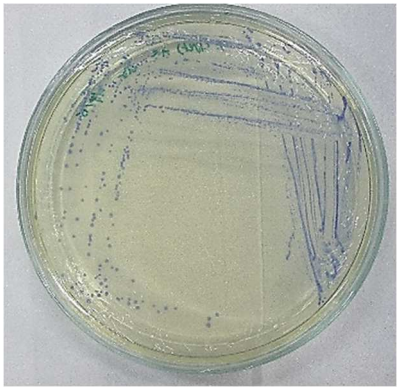

Fig. 6 The success of the PCR cloning construct

IV. Conclusion

A good mega-primer design is critical to a successful OEPC reaction. The use of web-based bioinformatic technology has been proven to save time and trial costs in the success of PCR Cloning. Mega-primer promoter str characterized by ClustalW, Net-Primer, and BLAST-NCBI software has been shown to increase the success of PCR cloning. The success of the construct is evidenced by the results of 100 % homologous sequencing and the ability of *E. coli* bacteria to express the purplish-blue protein (AmilCP).

References

[1] P. Moezi, M. Kargar, A. Doosti, and M. Khoshneviszadeh, "Multiplex touchdown PCR assay to enhance specificity and sensitivity for concurrent detection of four foodborne pathogens in raw milk," *J. of appl. microbiol.,* vol. 127, pp. 262-273, 2019.

[2] R. Nyaruaba, C. Mwaliko, K. K. Kering, and H. Wei, "Droplet digital PCR applications in the tuberculosis world," *Tuberculosis,* vol. 117, pp. 85-92, 2019.

[3] F. Feliatra, M. Mardalisa, J. Setiadi, I. Lukistyowaty, and A. Hutasoit, "Potential of Secondary Metabolite from Marine Heterotrophic Bacteria against Pathogenic Bacteria in Aquaculture," in *J. of Phys.: Conf. Series*, 2020, p. 012044.

[4] A. Bryksin and I. Matsumura, "Overlap extension PCR cloning," in *Synthetic Biology*, ed: Springer, 2013, pp. 31-42.

[5] Y. Lu, S. Xiao, M. Yuan, Y. Gao, J. Sun, and C. Xue, "Using overlap-extension PCR technique to fusing genes for constructing recombinant plasmids," *J. of basic microbiol.,* vol. 58, pp. 273-276, 2018.

[6] P.-T. Tran, C. F. Zhang, and V. Citovsky, "Rapid generation of inoculum of a plant RNA virus using overlap PCR," *Virology,* vol. 553, pp. 46-50, 2020.

[7] K. M. Goh, K. J. Liew, K. P. Chai, and R. M. Illias, "Use of mega-primer and overlapping extension PCR (OE-PCR) to mutagenize and enhance cyclodextrin glucosyltransferase (CGTase) function," in *In Vitro Mutagen.*, ed: Springer, 2017, pp. 385-396.

[8] S. Chojnacki, A. Cowley, J. Lee, A. Foix, and R. Lopez, "Programmatic access to bioinformatics tools from EMBL-EBI update: 2017," *Nucl. acids researc,* vol. 45, pp. W550-W553, 2017.

[9] F. M. Putra, F. K. Surado, and G. I. Sampurno, "Feature Selection Techniques for Selecting Proteins that Influence Mouse Down Syndrome Using Genetic Algorithms and Random Forests," *JOIV: Int. . on Informat. Visualiz.* vol. 4, pp. 162-165, 2020.

[10] M. Mardalisa, F. Feliatra, and N. Nursyirwani, "Multiple Antibiotic Resistance Index of Escherichia coli Isolates from Dumai Sea Waters Riau Province," *Berkala Perikanan Terubuk,* vol. 49, pp. 734-739, 2021.

[11] H. Christensen and J. E. Olsen, "Primer Design," in *Introduction to Bioinfo. in Microbiol.*, ed: Springer, 2018, pp. 81-102.

[12] A. V. Koehler, P. K. Korhonen, R. S. Hall, N. D. Young, T. Wang, S. R. Haydon, and R. B. Gasser, "Use of a bioinformatic-assisted primer design strategy to establish a new nested PCR-based method for Cryptosporidium," *Parasit. & vect.,* vol. 10, p. 509, 2017.

[13] S. Ahsan and D. Summers, "Identification of a toxin coding fragment in pBSSB1, a linear plasmid from Salmonella enterica serovar Typhi that can stabilize a multicopy plasmid," *Asian Pacific J. of Tropical Biomed.,* vol. 8, p. 365, 2018.

[14] L. E. Post, A. E. Arfsten, F. Reusser, and M. Nomura, "DNA sequences of promoter regions for the str and spc ribosomal protein operons in E. coli," *Cell,* vol. 15, pp. 215-229, 1978.

[15] C. Yi, J. Sjöberg, and D. Johansson, "Numerical modelling for blast-induced fragmentation in sublevel caving mines," *Tunnell. and Undergr. Space Techn.,* vol. 68, pp. 167-173, 2017.

[16] J. Zuber, B. J. Cabral, I. McFadyen, D. M. Mauger, and D. H. Mathews, "Analysis of RNA nearest neighbor parameters reveals interdependencies and quantifies the uncertainty in RNA secondary structure prediction," *Rna,* vol. 24, pp. 1568-1582, 2018.

[17] J. Guo, D. Starr, and H. Guo, "Classification and review of free PCR primer design software," *Bioinformatics,* 2020.

[18] M. Mardalisa, S. Suhandono, and M. Ramdhani, "Isolation and Characterization of str Promoter from Bacteria Escherichia coli DH5α using Reporter Gene AmilCP (Acropora millepora)," in *IOP Conf. Series: Earth and Environ. Sci.*, 2020, p. 012014.

[19] T.-W. Xue and Z.-Y. Guo, "What Is the Real Clausius Statement of the Second Law of Thermodynamics?," *Entropy,* vol. 21, p. 926, 2019.

[20] L. Du, C. Zhang, Q. Liu, X. Zhang, and B. Yue, "Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design," *Bioinformatics,* vol. 34, pp. 681-683, 2018.

[21] L. Mancabelli, C. Milani, G. A. Lugli, F. Fontana, F. Turroni, D. van Sinderen, and M. Ventura, "The Impact of Primer Design on Amplicon-Based Metagenomic Profiling Accuracy: Detailed Insights into Bifidobacterial Community Structure," *Microorganisms,* vol. 8, p. 131, 2020.

[22] A. C. Panda and M. Gorospe, "Detection and analysis of circular RNAs by RT-PCR," *Bio Protoc,* vol. 8, p. e2775, 2018.

[23] G. A. Hudson, R. J. Bloomingdale, and B. M. Znosko, "Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides," *Rna,* vol. 19, pp. 1474-1482, 2013.

[24] M. F. Sloma, M. Zuker, and D. H. Mathews, "Predictive Methods Using RNA Sequences," *Bioinformatics,* p. 155, 2020.