2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) | 979-8-3503-9961-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICITISEE57756.2022.10057821

# An Improved Method for Prioritizing Polymerase Chain Reaction (PCR) Primer Design in Sanger Sequencing

Abdullah Faqih Septiyanto
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025221022@mhs.its.ac.id

Yuni Fitri Cahyaningsih Cahyono
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
yunifitri.cc@gmail.com

Riyanarto Sarno
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

Fadlilatul Taufany
Department of Chemical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
f_taufany@chem-eng.its.ac.id

Siti Halimah Larekeng
Faculty of Vocational
Hasanuddin University
Makassar, Indonesia
sitih5h.82@gmail.com

Kelly Rossa Sungkono
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
kelly@its.ac.id

Karlina Syahruddin
Research Center for Food Crops
National Research and Innovation Agency
Bogor, Indonesia
karlinasyahruddin@gmail.com

Endang Gati Lestari
Research Center for Food Crops
National Research and Innovation Agency
Bogor, Indonesia
endanggatilestari@gmail.com

Sholiq
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
sholiq@is.its.ac.id

***Abstract*—Designing primers is a challenge for researchers to design primers that can amplify sequence genes. Primer design has several parameters considered, namely primer length, thermal melting temperature (TM), and guanine-cytosine content (GC-content). Therefore, bioinformatics is needed to help design primers more informatively by showing the parameter values of the primer design. One of the programs used to design primers is Primer3Plus. Primer3Plus can create several primer designs. Thus, Primer3Plus requires an expert to choose a primer pair that fits with the primer length, TM, and GC content parameters. This research proposed an improvised method for prioritizing design primer pairs from the Primer3Plus results to obtain the most suitable primer pairs according to the specified parameters. Before being processed into Primer3Plus, the gene sequence is split into several sequences with a length of around 1000bp to suit Sanger Sequencing (SGS). Therefore, the primers designed from each sequence can amplify all parts of the gene sequence. The experiment used the sequence gene Sorghum bicolor phytochrome A (PHYA) gene from the National Center for Biotechnology Information (NCBI) and implemented using Python with the Biopython library. The results get the best pair primers from each sequence based on primer length, TM, and GC content parameters with a product size of 900-1000bp for PCR and sequencing using SGS.**

***Keywords—Bioinformatics, Polymerase Chain Reaction, Primer Design, Primer3Plus, Sanger Sequencing***

## I. INTRODUCTION

The development of sequencing machines continues to grow as many species emerge. One of the machines that are often used to perform sequencing deoxyribonucleic acid (DNA) is Sanger Sequencing (SGS). SGS is the first-generation sequencing machine that has become the gold standard for detecting mutations of DNA [1]. However, the presence of SGS has been replaced by Next Generation Sequencing (NGS). SGS is still used because the price is lower for sequencing up to 20 samples [2], and many researchers still use SGS to amplify Polymerase Chain Reaction (PCR) results.

PCR is a widespread technique needed in the biological and medical fields. An essential component of the PCR process is the primer [3]. Designing primers is crucial to amplify sequence DNA in the PCR process. One of the programs that are often used to design primers are Primer3, Primer3Plus [4], BatchPrimer3 [5], and openPrimeR [6]. Problems regarding the primer design are still being researched. Thus, the primer design can be maximized by the DNA sequence that will be carried out by PCR and readings using a sequencing machine.

In previous research, BatchPrimer3 is a development method of web-based Primer3 which is used to design several types of primers in a high-throughput manner [5]. openPrimeR was used to design multiplex primers for multiplex PCR [6]. openPrimeR includes the development of Primer3 but uses the R language. In addition to the development of Primer3 for primer design, the use of SGS is still being carried out in several studies. Dorlass et al. [7] conducted a study on a primer design of SARS-CoV-2 to detect variants using SGS. Matsubara et al. [8] used long-distance Reverse Transcription Polymerase Chain Reaction (RT-PCR) with SGS. Some of these studies show that primer design is crucial to support the success of PCR, and SGS is still widely used for research and new case studies. However, specific knowledge is needed to make a primer

design able to amplify the whole sequence in the DNA sequence template.

This research proposed an improved method for prioritizing PCR primer design in SGS. Priority is one way to get recommendations based on the parameters and weights specified [9–11]. The priority focuses on three parameters of primer3plus results, namely primer length, Thermal Melting (TM), and guanine-cytosine content (GC-content). Before being processed using Primer3Plus, the DNA sequences were split into several sequences with an 800-1000 base pairs (bp) length. Each split sequence is processed using Primer3Plus, which produces five pairs of primers. The proposed method was applied to get the best primer pair based on the primer length, TM, and GC content of the several primer pairs. Thus, the PCR process is expected to amplify the DNA sequence template and can be sequenced using SGS.

## II. Literature Review

### A. Primer Design

Primer design is one of the requirements to perform PCR. Designing primers requires DNA sequences that can be retrieved from online databases such as the National Center for Biotechnology Information (NCBI) Reference Sequence Database (RefSeq) [12]. The primer design must balance specificity and amplification efficiency [13]. The specificity of the primer design will produce unwanted PCR products if the specificity value is low. In contrast, the efficiency of the primer design shows that the proximity of primer pairs (forward and reverse) can amplify DNA sequences for each PCR cycle[13]. PCR primer designs can be designed using software or bioinformatics tools [14]. One public program used and integrated with various bioinformatics tools and software to design primers according to templates (DNA sequences) based on multiple parameters is Primer3 and Primer3Plus [15–17]. The primer design results from Primer3 balance the primer length, primer melting temperature, and product length, then the primer design results are validated in silico and in vitro. In silico validation can be analyzed using Blast [18] in vitro validation in the wet lab before synthesizing.

### B. Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) was developed in the early 1980s [19–21]. The development of PCR can be tailored to the needs such as DNA sequencing, mutation detection, in vitro mutagenesis, molecular cloning, genomic DNA, cDNA cloning, and allelotyping [22]. PCR can sequence targets in DNA of large size and high complexity. PCR can produce multiple copies by amplifying the chain reaction, which is inexpensive, and the reaction is easy to regulate. PCR can be easily applied with knowledge of the target nucleotide sequence. However, PCR is simple, sensitive, flexible, fast, and powerful [22]. There are three stages in the PCR cycle: Denaturation, Annealing, and Extension. Denaturation is the heating step of the DNA template with a temperature > 90c, Annealing is the annealing step of two primers from the denaturation of the DNA template by synthetic oligonucleotides with a length of 20-25 nucleotides, and Extension is the primary extension stage (DNA synthesis) with a temperature between 55-70c in DNA polymerase. Thermostable catalyzes enzymatic reactions starting at the 3' end of the bound primer. The PCR cycle can be repeated about 25-35 times according to the time and temperature that can be set into the device for each step in the cycle [22].

### C. Sanger Sequencing (SGS)

Sanger Sequencing (SGS) is the first-generation technology [2] to identify DNA sequences which were founded by Frederick Sanger in 1997 [23]. SGS called as "chain termination" method because it consists of polymerized DNA that selectively incorporates dideoxynucleotides resulting in fragments terminating with fluorescent labeled dideoxynucleotides at the 3' end of different fragment lengths [23]. SGS is often used to identify mutations in combination with PCR. However, SGS has the disadvantage of not being sensitive to mutations below 20% [1]. Therefore, second-generation technology was created, namely Next-Generation Sequencing (NGS). NGS is a high-throughput sequencing technology that can perform large-scale DNA sequencing (whole-genome sequencing), while SGS is a low-throughput sequencing technology with individual genomes [24]. Although SGS has its limitations, SGS still has advantages compared to the NGS. SGS was suitable for identifying DNA in up to 20 samples, while NGS was not suitable for small samples (up to 20 samples). Because NGS requires equipment and an expert laboratory with high cost [23], SGS technology can read signals using Fluorescence-based imaging with a length of 800 - 1000 base pair (bp) with high accuracy [2].

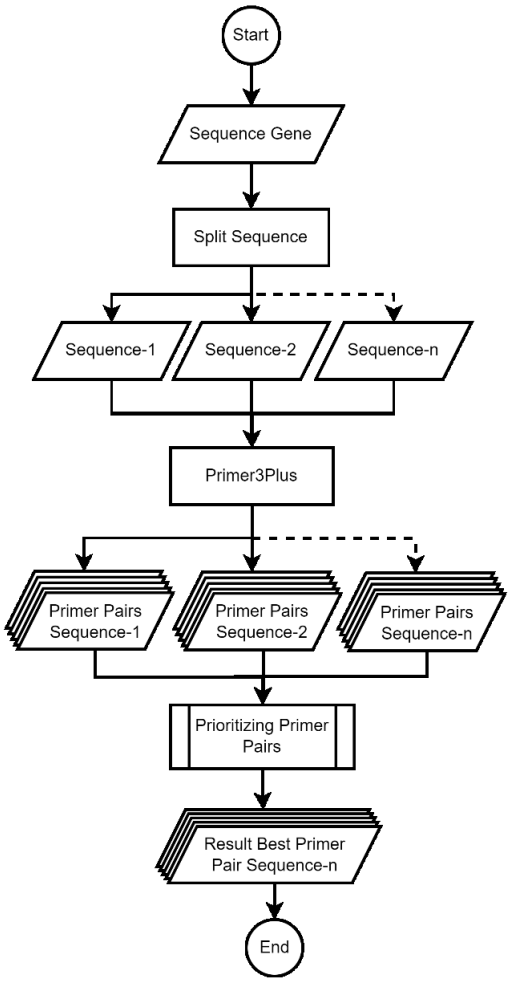


Fig. 1. Proposed Method

## III. PROPOSED METHOD

This research proposed an improved method for prioritizing PCR primer design in SGS. The flow of the proposed method is depicted in Fig. 1.

### A. Split Sequence

At this stage, split-sequence DNA into multiple sequences is applied based on the length specified by the user. The sequenced DNA of Sorghum bicolor phytochrome A (PHYA) used the FASTA or GenBank file from NCBI [25]. Sequence DNA processed in the system are complete nucleotide CDs [26]. The pseudocode for split-sequence DNA is in Table I.

In Table I, based on the length of the DNA sequence, the number of input parameters consists of 2 types, namely addSetting and minusSetting. addSetting to specify the location where the length of the sequence ends, and minusSetting to specify the location where the sequence will start based on the DNA sequence. Each parameter is filled in based on the user's need to allocate sequence length. Each sequence has a length of approximately 800-1000 bp according to the capabilities of the SGS machine [2]. The results of the pseudocode in Table I will provide a list of sequences which are then processed to obtain the primer design using Primer3Plus.

TABLE I. PSEUDOCODE SPLIT SEQUENCE

**Pseudocode Split Sequence**

```
sequenceDNA ← sequence DNA from file fasta or genbank
totalBp ← total bp of sequenceDNA
numberOfSetting ← number of add and minus setting
totalSplitSequence ← total split sequence from sequence DNA
start ← location start split sequence (start from 0)
end ← location end split sequence (start from 0)
location ← location start split sequence (start from 1)
if totalBp < 2000
    numberOfSetting = 1
elseif totalBp < 3000
    numberOfSetting = 2
elseif totalBp < 4000
    numberOfSetting = 3
elseif totalBp < 5000
    numberOfSetting = 4
endif
totalSplitSequence = numberOfSetting + 1
addListLocation ← list of location
listAddSetting ← list of value add setting
listMinusSetting ← list of value minus setting
i ← number of increments totalSplitSequence
for totalSplitSequence
    if i = 0
        location = 1
        end = listAddSetting[i]
        addListLocation() ← add location value
    elseif i = (totalSplitSequence – 1)
        start = end – listMinusSetting[i-1]
        location = start + 1
        addListLocation() ← add location value
    else
        start = end – listMinusSetting[i-1]
        location = start + 1
        end = start + listAddSetting[i]
        addListLocation() ← add location value
    endif
endfor
```

### B. Prioritizing Primer Pairs

The prioritizing primer is applied to get the best primer pair from each sequence. In primer design, several conditions must be considered: primer length, primer pair length difference, thermal melting temperature (Tm), Tm difference from a pair of primers, GC content primer, GC difference from a pair of primers, PCR buffer reagent concentration, and product size. However, this study focuses on the primer length, Tm, and GC content. Based on the primer length, TM, and GC content, the gap between optimal value and value from the Primer3Plus results for each pair of primers (forward and reverse) from each sequence is calculated using Eq. 1, Eq.2, and Eq.3.

$$G_{of} = C(olp - lp_f)w + C(otm - tm_f)w + C(ogc - gc_f) \quad (1)$$

$$G_{or} = C(olp - lp_r)w + C(otm - tm_r)w + C(ogc - gc_r) \quad (2)$$

$$TG_{ofr} = G_{of} + G_{or} \quad (3)$$

where:

$G_{of/r}$ : Gap optimal forward/reverse
$w$ : Weight parameter
$C$ : Convert if get negative value to positive value
$olp$ : Optimal length primer
$otm$ : Optimal TM
$ogc$ : Optimal GC
$lp_{f/r}$ : Length primer forward/reverse
$tm_{f/r}$ : TM forward/reverse
$gc_{f/r}$ : GC forward/reverse

In Eq. 1 and Eq. 2, the optimal values of primer length, TM, and GC content are subtracted from the primer length, TM, and GC content results for each primer (forward and reverse) from Primer3Plus. If the result of the optimal value minus the value of the Primer3Plus result gets a negative value, then it is converted to a positive value and equal with $w$ for weighted TM and GC content. After obtaining $G_{of}$ and $G_{of}$ values, Eq. 3 is used to calculate $TG_{ofr}$. After getting the $TG_{ofr}$ value, then calculate the gap between the primer pairs (forward and reverse) for each primer length, TM, and GC content value using Eq. 4 and continue to calculate the score of each primer pair to determine the priority of the best primer pair using Eq. 5.

$$TG_{fr} = C(lp_f - lp_r) + C(tm_f - tm_r) \quad (4)$$

$$TS = TG_{ofr} + TG_{fr} \quad (5)$$

Where:

$TG_{fr}$ : Total gap forward and reverse
$TS$ : Total scores for prioritized pair primer

In Eq. 4, the primer length, TM, and GC content values are subtracted between the forward and reverse primers. In this equation, converting negative to positive values is also carried out. After getting the total value of the gap between the primer pairs, $TG_{ofr}$ and$TG_{fr}$ were added to obtain the total score used to determine the best priority $(TS)$. The primer pair with the lowest score will be the top priority. Implementation of the proposed method for prioritizing primer pairs uses Python with the library Biopython [27] in pseudocode in Table II.

TABLE II. PSEUDOCODE PRIORITIZING PRIMER

**Pseudocode Prioritizing Primer**

```
recommendationList ← list of recommendation primer of sequence
optLength = 20 ← optimal value of length primer
optTm = 60.0 ← optimal value of termal melting
optGc = 50.0 ← optimal value of GC content
w = 100 ← Weight parameter
forwardGapOpt ← list of total range optimal with result forward
reverseGapOpt ← list of total range optimal with result reverse
forwardList ← list of forward primer sequence from result primer3
reverseList ← list of reverse primer sequence from result primer3
j = 1
positive() ← convert negative to positive value
minIndex ← location index minValue
minValue minimal value total gap of pair primer
for i in primerList
    code ← code forward or reverse primer
    Seq ← sequence forward or reverse primer
    startPrimer ← start location value of primer from primerList[i]
    length ← length vaue of primerList[i]
    Tm ← termal melting value of primerList[i]
    Gc ← GC content value of primerList[i]
    gapOptLength = optLength – length
    gapOptTm = optTm – Tm
    gapOptGc = optGc – Gc
    if gapOptLength < 0
        gapOptLength = positive(gapOptLength) * w
    else
        gapOptLength = 0
    endif
    if gapOptTm < 0
        gapOptTm = positive (gapOptTm) * w
    else
        gapOptTm = 0
    endif
    if gapOptGc < 0
        gapOptGc = positive (gapOptGc)
    Else
        gapOptGc = 0
    endif
    totalGapOpt = sum(gapOptLength, gapOptTm, gapOptGc)
    if j modulus 2 = 1
        forwardList() ← add code, startPrimer, seq, length, Tm, Gc
        forwardGapOpt() ← add totalGapOpt
    elseif j modulus 2 = 0
        reverseList() ← add code, startPrimer, seq, length, Tm, Gc
        reverseGapOpt() ← add totalGapOpt
    endif
    j = j + 1
endfor
totalList ← list of summarize totalGapOpt and totalGap
for i range(5)
    totalGapOpt = forwardGapOpt[i] + reverseGapOpt[i]
    gapLength = positive(forwardList[i][3] – reverseList[i][3])
    gapTm = positive(forwardList[i][4] – reverseList[i][4])
    gapGc = positive(forwardList[i][5] – reverseList[i][5])
    totalGap = gapLength + gapTm
    total = totalGapOpt + totalGap
    totalList() ← add total
endfor
for i range(5)
    if i = 0
        minIndex = i
        minValue = totalList[i]
    else
        total = totalList[i]
        if total < minValue
            minIndex = i
            minValue = total
        endif
    endif
endfor
recommendationList() ← add forwardList[miniIndex] and
                       reverseList[miniIndex]
```

In Table II, pseudocode is used to get the best priority primer pair (forward and reverse) based on the DNA sequence from the split sequence results using the split sequence pseudocode in Table I. Each sequence gets one best primer pair for the PCR process. The results of the priority of the primer pair in each sequence will be shown from the results of the pseudocode process in Table II.

## IV. RESULT AND DISCUSSION

### A. Result

The DNA sequence of the PHYA with a sequence length of 3834 was used to obtain primer pairs for the PCR process shown in Table III. Using pseudocode split sequence, we get four sequences with lengths ranging from 900 to 1000 bp. After the primer design process using Primer3Plus, the primer prioritizing pair is applied using the primer prioritizing pseudocode. Calculation results for Eq. 1 - 5 are shown in Table IV, and the results of prioritizing primer in each sequence are shown in Table V.

Table IV shows the list of primer pairs from the Primer3Plus results obtained in sequence 1 with the calculation results for each Eq. 1-5. Each primer pair is calculated to get a TS value to get the best primer pair priority results from the five primer pairs from the Primer3Plus results. From the calculation results in Table III, the primer pair with code (F1 and R1) gets the smallest value. Therefore, primer pairs become the top priority of sequence 1.

Table V show priority results for each sequence obtained with detailed information on the primer location on the template sequence (PHYA sequence gene), primer length, TM, GC content, and product size, indicating the length of the sequence generated from the primer pairs.

### B. Discussion

Based on the results of the priority of the primer pair, each sequence gets the value of length, TM, and GC close to the optimal value. In the results of the calculation of prioritizing primer pairs, the primer length, TM, and GC that exceed the optimal value have a higher score, thus the primer pair with the primer length, TM, and GC content below the optimal value is better than above the optimal value. The location of each primer can cover all parts of the template sequence based on the location of the primer (forward and reverse). Therefore, the primer can amplify the DNA template. Each sequence gets a product size with a sequence length ranging from 900-1000 bp. Thus, the primer pairs can be sequenced using SGS. From the priority results obtained, primer designs are ready to be used in the PCR and sequencing process using SGS.

## V. CONCLUSION

An improved method for prioritizing PCR primers was created to design primers before the PCR process is carried out, which will be sequenced using the SGS machine. SGS machine has limitations in reading DNA sequences. Researchers proposed a method for designing primer by prioritizing the best pair primer from the result Primer3Plus to prepare for sequencing using SGS. Before using Primer3Plus, the sequence gene was split into several sequences based on the sequence length in bp units. In this researcu, the gene sequence used was PHYA.

TABLE III. DATASET SEQUENCE DNA PHYA

| Sorghum bicolor phytochrome A (PHYA) gene, complete CDs | | | | | | |
|---|---|---|---|---|---|---|
| 1 | atgtcttcct | cgaggcctgc | ccactcttcc | agttcatcca | gtaggactcg | ccagagctcc |
| 61 | caggcaagga | tattagcaca | aacaaccctt | gatgctgaac | tcaatgcaga | gtatgaagaa |
| … | … | … | … | … | … | … |
| 3781 | accttcatcc | tcactgctga | acttgctgct | gctccttcag | cagttggaca | atga |

TABLE IV. LIST PAIR PRIMER IN SEQUENCE 1

| Code | Primer | Length | TM | GC | $G_{of}$ | $G_{or}$ | $TG_{ofr}$ | $TG_{fr}$ | TS |
|---|---|---|---|---|---|---|---|---|---|
| F1 | AGAGCTCCCAGGCAAGGATA | 20 | 60.0 | 55.0 | 5.0 | 0 | 5.0 | 0 | 5.0 |
| R1 | AGCCATGACAAGGGATGCAA | 20 | 60.0 | 50.0 | | | | | |
| F2 | AGAGCTCCCAGGCAAGGATA | 20 | 60.0 | 55.0 | 5.0 | 0 | 5.0 | 0.5 | 5.5 |
| R2 | ACAACAGCCATGACAAGGGA | 20 | 59.5 | 50.0 | | | | | |
| F3 | AGAGCTCCCAGGCAAGGATA | 20 | 60.0 | 55.0 | 5.0 | 5.0 | 10.0 | 0.6 | 10.6 |
| R3 | CACAACAGCCATGACAAGGG | 20 | 59.4 | 55.0 | | | | | |
| F4 | AGAGCTCCCAGGCAAGGATA | 20 | 60.0 | 55.0 | 5.0 | 5.0 | 10.0 | 0.6 | 10.6 |
| R4 | CCACAACAGCCATGACAAGG | 20 | 59.4 | 55.0 | | | | | |
| F5 | AGAGCTCCCAGGCAAGGATA | 20 | 60.0 | 55.0 | 5.0 | 75.0 | 80.0 | 0.7 | 80.7 |
| R5 | CAGCCATGACAAGGGATGCA | 20 | 60.7 | 55.0 | | | | | |

TABLE V. RESULT PRIORITIZING PRIMER

| Sequence | Code | Primer | Start | Length | TM | GC | Product Size |
|---|---|---|---|---|---|---|---|
| Sequence 1 | F1 | AGAGCTCCCAGGCAAGGATA | 53 | 20 | 60.0 | 55.0 | 971 |
| | R1 | AGCCATGACAAGGGATGCAA | 1023 | 20 | 60.0 | 50.0 | |
| Sequence 2 | F1 | TTGCATCCCTTGTCATGGCT | 1004 | 20 | 60.0 | 50.0 | 966 |
| | R1 | ACCCTGACAATTCTGCTACCT | 1969 | 21 | 58.7 | 47.6 | |
| Sequence 3 | F1 | TGGTTCGCCTCATGGAAACA | 1871 | 20 | 59.9 | 50.0 | 963 |
| | R1 | TTGTGCTGTCTGCTCAGAGG | 2833 | 20 | 60.0 | 55.0 | |
| Sequence 4 | F2 | CGTACATGCGACATGCCATC | 2853 | 20 | 59.8 | 55.0 | 959 |
| | R2 | CAGCAGCAAGTTCAGCAGTG | 3811 | 20 | 60.0 | 55.0 | |

Therefore, a total of 4 sequences were obtained. Each sequence is processed using Primer3Plus to get the list of primer pairs. Based on the primer list from each sequence, an improvised method was applied to prioritize the best primer pairs that focused on the primer length, TM, and GC content. Therefore, we get one pair of primers for each sequence. Each primer pair from each sequence will be used as a base material for the PCR process, which can amplify the entire gene sequence and is ready to proceed to the sequencing process using SGS. The advantage of the proposed method is that it can provide recommendations for the best primer pair based on the parameters of primary length, TM, and GC content compared to Primer3Plus, and disadvantageous there is no web-based platform like Primer3Plus.

Future work of this research is to change the wet lab testing to test the success of PCR from primers that have been made with a platform that can estimate the success rate of the PCR process. Next, improvement can be developed, and add several integrations with other tools to build complex bioinformatic platforms.

## ACKNOWLEDGMENT

This research was funded by the Directorate General of Higher Education, Research, and Technology, and Indonesian Ministry of Education, Culture, Research and Technology under Matching Fund (MF) Kedaireka Program; Lembaga Pengelola Dana Pendidikan (LPDP) under Riset Inovatif Produktif (RISPRO) Invitation Program; and Institut Teknologi Sepuluh Nopember (ITS) under project scheme of the Publication Writing and IPR Incentive Program (PPHKI).

## REFERENCES

[1] Arsenic, R.; Treue, D.; Lehmann, A.; Hummel, M.; Dietel, M.; Denkert, C.; Budczies, J. (2015). Comparison of targeted next-generation sequencing and Sanger sequencing for the detection of PIK3CA mutations in breast cancer, *BMC Clinical Pathology*, Vol. 15, No. 1. doi:10.1186/s12907-015-0020-6

[2] MacConaill, L. E. (2013, May 20). Existing and emerging technologies for tumor genomic profiling, *Journal of Clinical Oncology*, American Society of Clinical Oncology, 1815–1824. doi:10.1200/JCO.2012.46.5948

[3] Chuang, L.-Y.; Cheng, Y.-H.; Yang, C.-H. (2013). Specific primer design for the polymerase chain reaction, *Biotechnology Letters*, Vol. 35, No. 10, 1541–1549. doi:10.1007/s10529-013-1249-8

[4] Untergasser, A.; Nijveen, H.; Rao, X.; Bisseling, T.; Geurts, R.; Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3, *Nucleic Acids Research*, Vol. 35, No. SUPPL.2. doi:10.1093/nar/gkm306

[5] Ali, S. K.; Al-Koofee, D. A. F. (2019). BatchPrimer3: A free web application for allele specific (SBE and allele flanking) primer design for SNPs genotyping in molecular diagnostics: A bioinformatics study, *Gene Reports*, Vol. 17. doi:10.1016/j.genrep.2019.100524

[6] Kreer, C.; Döring, M.; Lehnen, N.; Ercanoglu, M. S.; Gieselmann, L.; Luca, D.; Jain, K.; Schommers, P.; Pfeifer, N.; Klein, F. (2020). openPrimeR for multiplex amplification of highly diverse templates,

*Journal of Immunological Methods*, Vol. 480, 112752. doi:10.1016/j.jim.2020.112752

[7] Dorlass, E. G.; Lourenço, K. L.; Magalhães, R. D. M.; Sato, H.; Fiorini, A.; Peixoto, R.; Coelho, H. P.; Telezynski, B. L.; Scagion, G. P.; Ometto, T.; Thomazelli, L. M.; Oliveira, D. B. L.; Fernandes, A. P.; Durigon, E. L.; Fonseca, F. G.; Teixeira, S. M. R. (2021). Survey of SARS-CoV-2 genetic diversity in two major Brazilian cities using a fast and affordable Sanger sequencing strategy, *Genomics*, Vol. 113, No. 6, 4109–4115. doi:10.1016/j.ygeno.2021.10.015

[8] Matsubara, M.; Imaizumi, Y.; Fujikawa, T.; Ishige, T.; Nishimura, M.; Miyabe, A.; Murata, S.; Kawasaki, K.; Taniguchi, T.; Igari, H.; Matsushita, K. (2022). Tracking SARS-CoV-2 variants by entire S-gene analysis using long-range RT-PCR and Sanger sequencing, *Clinica Chimica Acta*, Vol. 530, 94–98. doi:10.1016/j.cca.2022.03.014

[9] Adnan, M.; Sarno, R.; Sungkono, K. R. (2019). Sentiment Analysis of Restaurant Review with Classification Approach in the Decision Tree-J48 Algorithm, *2019 International Seminar on Application for Technology of Information and Communication (ISemantic)*, IEEE, 121–126. doi:10.1109/ISEMANTIC.2019.8884282

[10] Fajar, A.; Sarno, R. (2019). Optimization of Time and Cost of Multi Organization Business Processes in A Port Container Terminal, *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 6, 259–271. doi:10.22266/ijies2019.1231.25

[11] Setyono, R. P.; Sarno, R. (2020). Comparative method of moora, copras and topsis based on weighting of best worst method in supplier selection at ABC mining companies in Indonesia, *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 19, No. 2, 890. doi:10.11591/ijeecs.v19.i2.pp890-899

[12] Kumar, A.; Chordia, N. (2015). In silico PCR primer designing and validation, *Methods in Molecular Biology*, Vol. 1275, 143–151. doi:10.1007/978-1-4939-2365-6_10

[13] Dieffenbach, C. W.; Lowe, T. M.; Dveksler, G. S. (1993). General concepts for PCR primer design., *Genome Research*, Vol. 3, No. 3, S30–S37. doi:10.1101/gr.3.3.S30

[14] Kumar, A.; Chordia, N. (2015). In Silico PCR Primer Designing and Validation, C. Basu (Ed.), *PCR Primer Design*, Springer New York, New York, NY, 143–151. doi:10.1007/978-1-4939-2365-6_10

[15] Koressaar, T.; Remm, M. (2007). Enhancements and modifications of primer design program Primer3, *Bioinformatics*, Vol. 23, No. 10, 1289–1291. doi:10.1093/bioinformatics/btm091

[16] Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B. C.; Remm, M.; Rozen, S. G. (2012). Primer3—new capabilities and interfaces, *Nucleic Acids Research*, Vol. 40, No. 15, e115–e115. doi:10.1093/nar/gks596

[17] Kõressaar, T.; Lepamets, M.; Kaplinski, L.; Raime, K.; Andreson, R.; Remm, M. (2018). Primer3_masker: integrating masking of template sequence with primer design software, *Bioinformatics*, Vol. 34, No. 11, 1937–1938. doi:10.1093/bioinformatics/bty036

[18] Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction, *BMC Bioinformatics*, Vol. 13, No. 1, 134. doi:10.1186/1471-2105-13-134

[19] Saiki, R. K.; Scharf, S.; Faloona, F.; Mullis, K. B.; Horn, G. T.; Erlich, H. A.; Arnheim, N. (1985). Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia, *Science*, Vol. 230, No. 4732, 1350–1354. doi:10.1126/science.2999980

[20] Mullis, K. B.; Faloona, F. A. (1987). [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction, *Methods in Enzymology* (Vol. 155), Academic Press, 335–350. doi:https://doi.org/10.1016/0076-6879(87)55023-6

[21] Mullis, K. B. (1994). The Polymerase Chain Reaction (Nobel Lecture), *Angewandte Chemie International Edition in English*, Vol. 33, No. 12, 1209–1213. doi:10.1002/anie.199412091

[22] Green, M. R.; Sambrook, J. (2019). Polymerase chain reaction, *Cold Spring Harbor Protocols*, Vol. 2019, No. 6, 436–456. doi:10.1101/pdb.top095109

[23] Dorlass, E. G.; Lourenço, K. L.; Magalhães, R. D. M.; Sato, H.; Fiorini, A.; Peixoto, R.; Coelho, H. P.; Telezynski, B. L.; Scagion, G. P.; Ometto, T.; Thomazelli, L. M.; Oliveira, D. B. L.; Fernandes, A. P.; Durigon, E. L.; Fonseca, F. G.; Teixeira, S. M. R. (2021). Survey of SARS-CoV-2 genetic diversity in two major Brazilian cities using a fast and affordable Sanger sequencing strategy, *Genomics*, Vol. 113, No. 6, 4109–4115. doi:10.1016/j.ygeno.2021.10.015

[24] Young, A. D.; Gillung, J. P. (2020, April 1). Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics, *Systematic Entomology*, Blackwell Publishing Ltd, 225–247. doi:10.1111/syen.12406

[25] White, G. M. (2004). Molecular Evolution of the Phytochrome Gene Family in Sorghum: Changing Rates of Synonymous and Replacement Evolution, *Molecular Biology and Evolution*, Vol. 21, No. 4, 716–723. doi:10.1093/molbev/msh067

[26] Furuno, M.; Kasukawa, T.; Saito, R.; Adachi, J.; Suzuki, H.; Baldarelli, R.; Hayashizaki, Y.; Okazaki, Y. (2003). CDS Annotation in Full-Length cDNA Sequence, *Genome Research*, Vol. 13, No. 6b, 1478–1487. doi:10.1101/gr.1060303

[27] Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics, *Bioinformatics*, Vol. 25, No. 11, 1422–1423. doi:10.1093/bioinformatics/btp163