ISSN: 1410-8917
e-ISSN: 2597-9914

Jurnal Kimia Sains dan Aplikasi 23 (5) (2020): 183–188

Jurnal Kimia Sains dan Aplikasi
Journal of Scientific and Applied Chemistry

Journal homepage: http://ejournal.undip.ac.id/index.php/ksa

Polymerization Domain Translated from 0.9 kb Gene Fragment of DNA Polymerase I from a Thermo-Halophilic PLS A Strain

Teuku Mohamad Iqbalsyah [a,*], Muhammad Akbar Velayati [a], Hira Helwati [a], Febriani [a]

[a] Biomolecules Application Research Group, Chemistry Department, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh, indonesia

*Corresponding author: t.iqbalsyah@unsyiah.ac.id

https://doi.org/10.14710/jksa.23.5.183-188

Article Info

Article history:

Received: 24th February 2020
Revised: 7th April 2020
Accepted: 30th April 2020
Online: 31st May 2020

Keywords:
DNA polymerase; *Geobacillus*; halophilic; thermophilic; Sabang

Abstract

The search for novel DNA Polymerases I, with higher fidelity and better polymerization rate, is essential to improve the Polymerase Chain Reaction method. A thermo-halophilic bacterium has been isolated from an undersea hot spring, dubbed Pria Laot Sabang (PLS) A strain. The 0.9 kb DNA Polymerase I gene fragments from the isolate were amplified, sequenced, and identified. The fragments were part of the polymerization domain of the enzyme. Homological analysis of the gene sequence showed that the PLS A strain was closely related to *Bacillus caldolyticus* strain XM. However, Swissprot structural analysis reveals that PLS A strain had high homology to *Geobacillus stearathermophilus*. Full sequence analysis is still needed to identify the species and evaluate the intact enzyme structure.

1. Introduction

The ability to artificially synthesis DNA by the Polymerase Chain Reaction (PCR) method [1] has changed the landscape of biotechnological research, particularly in utilizing DNA amplification and sequencing [2]. The elongation of the DNA primer during PCR procedure is enzymatically catalyzed by a protein called DNA polymerases.

Based on the conserved region in the gene, DNA Polymerases are grouped into six families: A, B, C, D, X, and Y. Family A and B members are found in eukaryotes and prokaryotes. The former has two exonuclease domains (3'–5' and 5'–3'), while the latter contains 3'–5' exonuclease (proofreading) domain. Family C polymerases are only found in prokaryotes and function as the proofreading enzyme. Family D is replicative polymerases and found only in eukaryotes. Family X members are found in bacteria, viruses, and mammals and functions to fill the gap in a newly synthesized DNA sequence. Family Y polymerase is the most recent family of small size and has no exonuclease activity. They are involved in DNA lesion tolerance pathways [3, 4].

In the early period, DNA polymerases were produced by a mesophilic *Escherichia coli*. As primers extension is conducted at a temperature of around 70-75 °C, the use of the enzyme from *E. coli* was ineffective. The availability of thermostable DNA polymerases from thermophilic *Thermus aquaticus* has solved this problem [5]. However, *Taq* polymerase exhibits 5'–3' exonuclease and 5'–3' polymerase activities but not a 3'–5' exonuclease activity. The fidelity is low as the enzyme is unable to proofread nucleotides insertion [6].

While the PCR technique has been evolving rapidly, the lack of fidelity and robustness of DNA polymerases remains a serious challenge. This problem is sometimes overcome by mixing *Taq* polymerase with other polymerases with high fidelity [7] in the PCR amplification process. Native DNA polymerases are often improved by mutagenesis to accommodate new techniques that require extreme conditions or in the presence of inhibitors [8]. Another approach is to find new variants of DNA polymerases from thermostable bacteria and archaea sources [9, 10].

This study aims to find a novel DNA polymerase I (DNA Pol I), which is commonly used in the PCR technique. The enzyme adopts a tertiary structure containing 928 amino acids with a molecular weight of 109 kDa. It was expected to have three catalytic domains, i.e., 3'–5' exonuclease, 5'–3' exonuclease, and 5'–3' polymerase [11]. Previously we have succeeded isolating microorganisms from underwater fumaroles in Pria Laot

Sabang Area in Weh Island. One of the strains (PLS 80) contains the DNA Pol I gene, and partial amplification produces gene fragments of the DNA Pol I polymerase domain [12]. In this study, we examined and amplified 0.9 kb DNA Pol I gene fragments from a different strain (PLS A) isolated from the same location. The amplified gene fragments were bioinformatically translated. The structure of the enzyme was compared to those in the Swissprot database. This study is preliminary to find a new type of DNA Pol I that may be used for the optimization of PCR techniques.

2. Methods

2.1. Bacterial Strain

A local thermo-halophilic bacterium, dubbed PLS A, was isolated from underwater fumaroles in Pria Laot Sabang, Weh Island, Aceh, Indonesia. The fumaroles have an onset temperature around 100°C, a neutral pH and a salt concentration of 3.5 M. The isolate was cultivated in a sterile solid ½T medium (0.4% bacto peptone, 0.2% yeast extract, 1% NaCl, 3% bacto agar) enriched with 0.25% glucose and diluted with seawater. The culture was incubated at 70°C for 24 hours.

The isolate was transferred to a ½T liquid medium containing 0.25% glucose. The culture was incubated at 70°C for 24 hours and 150 rpm. The cells were then isolated by centrifugation at 10000×*g* for 10 minutes. The resulting pellet was washed thrice with deionized water.

2.2. Isolation of Chromosomal DNA

The pellet was suspended in 200 µL of 10 mM Tris-HCl buffer pH 8 and incubated at 37°C for 1 hour. The suspension was mixed with lysis buffer containing 2% SDS, 0.8 mg/mL proteinase K and 200 mM EDTA pH 8.0, and incubated at 50°C for 30 min. A mixture of ice-cold potassium acetate and glacial acetic acid of the same ratio (150 µL) was added to the suspension. The mixture was separated by centrifugation at 6000×*g* for 10 minutes. The supernatant was mixed with an equal volume of chloroform: isoamyl alcohol (24:1, v/v). The aqueous phase was separated by centrifugation at 6000×*g* for 10 minutes and precipitated with 600 µL isopropanol at room temperature. Further centrifugation at 16000×*g* for 20 minutes at room temperature produced pellet containing genomic DNA. The pellet was washed twice with cold 70% ethanol and re-suspended in sterile deionized water [13].

2.3. Amplification, Sequencing, and Analysis of DNA Polymerase I Gene Fragment

The 0.9 kb gene fragment of DNA Pol I from PLS A strain was amplified with the PCR method using two sets of primers (FP1-RP1 and FP2-RP1) (Figure 1). All PCR products were sequenced with an ABI PrismR 3100 Genetic Analyzer. The gene fragment was compared to fragments of DNA pol I in the National Center for Biotechnology Information GenBank database (www.ncbi.nml.nih.gov). The phylogenetic tree of the gene fragment of DNA polymerase was constructed using Mega 6 software [14].

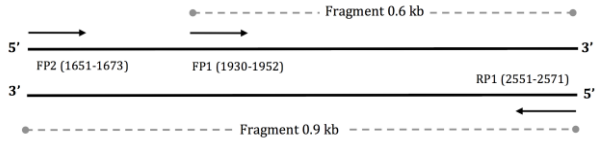

Figure 1. Profile of primers to amplify the DNA Pol I gene fragments from PLS A strain [10]

2.4. Structural Analysis

The sequence of the 0.9 kb DNA pol I gene was used to predict the structure of the protein using the SWISS-MODEL program (http://swissmodel.expasy.org/). The gene fragment was initially translated into protein. The program looked for a sequence of proteins most similar to that of this study. It was then used as a template for structural modeling in the form of the PDB file, after alignment using ClustalW. The homology was analyzed by the Visual Molecular Dynamic (VMD 1.9.1) program [15].

3. Results and Discussion

3.1. Amplification and Sequencing of DNA Polymerase I Gene Fragment

Amplification of DNA Pol I gene in this study used two sets of primers. The FP1 and FP2 were specific forward primers for the conserved region of 1651-1673 and 1930-1952, respectively, in the overall genetic sequence of DNA Pol I. The RP1 primer was a reverse primer used to start amplification in the conserve region of 2551-2571. The sequence of FP1 primer is GATCCGAACCTGCAAAACATTCC, while FP2 primer is AACATTAACTCGCCGAAACAGCT. RP1 primer has a sequence of AATCAGCTCGTCATGCACCTG.

The amplification of the DNA Pol I gene from PLS A strain produced a single band. The genomic DNA was greater than 10 kb (Figure 1A). A combination of FP1-RP1 primers produced the DNA fragment of about 0.6 kb, while FP2-RP1 primers successfully amplified the DNA fragment to approximately 0.9 kb (Figure 1B). The gene fragments were sequenced, and the results of the 0.6 kb and 0.9 kb were superimposed to get a more accurate sequence.

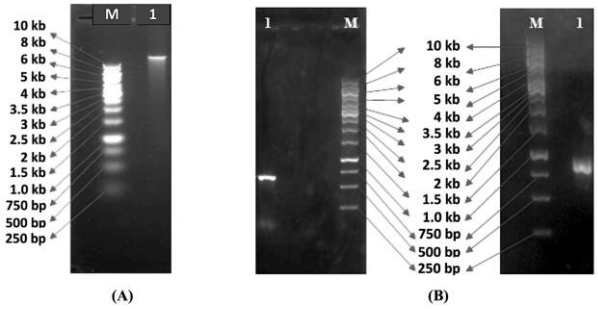

Figure 2. A) Chromosomal DNA from PLS A strain, M = DNA Ladder, 1 = Genomic DNA, (B) Results of DNA Pol I gene amplification. Left panel: primer FP1-RP1 (0.6 kb) and right panel: primer FP2-RP1 (0.9 kb), M = DNA Ladder, 1 = DNA fragment

The two sets of primers used in this study amplified roughly 900 bp, translated to 282 residues. The residues were part of the 5'–3' polymerase located at the C-terminus region. In prokaryotes, the gene encoding DNA Pol I are around 3000 base-pairs in size that encodes approximately 1000 amino acid residues [16]. It generally contains two domains with different catalytic activities, namely polymerase 5'–3' domain catalyzing DNA polymerization reactions, and exonuclease 3'–5' domain catalyzing the removal of the Okazaki fragments and filing DNA gap. Some prokaryotes, such as *E. coli*, also possess the third domain of 3'–5' exonuclease, which serves as a proofreading enzyme during DNA replication [11].

3.2. Phylogenetic Analysis

The homology of the polymerization domain of DNA Pol I from PLS A was compared to sequences of similar genes in the NCBI GenBank using the BLAST search tool. Table 2 shows that the gene from the PLS A strain had a close homology to that of six bacteria, i.e. (1) *Geobacillus thermoleovorans*, (2) *Bacillus caldolyticus*, (3) *Geobacillus stearothermophilus*, (4) *Geobacillus* sp., (5) *Bacillus caldotenax,* and (6) *Geobacillus kaustophilus* (Table 2).

Table 1. Best ten homologs with DNA Pol I gene fragment from PLS A strain

Microorganism	Accession Number	Similarity
Geobacillus thermoleovorans CCB_US3_UF5	CP003125.1	99%
Geobacillus thermoleovorans strain KCTC 3570	CP014335.1	99%
Bacillus caldolyticus strain XM	EF488810.1	99%
Bacillus caldolyticus strain EA.1	AY247636.1	99%
Bacillus sp.	EF198253.1	99%
Bacillus caldotenax	D12982.1	99%
Geobacillus stearothermophilus 10	CP008934.1	99%
Geobacillus sp. 777	KP993175.1	99%
Geobacillus sp. GHH01	CP004008.1	99%
Geobacillus kaustophilus HTA426	BA000043.1	99%

Phylogenetic analysis shows that the polymerization domain of DNA Pol I from PLS A had a close homology to *Bacillus caldolyticus* strain XM as they were clustered in the same branch (Figure 3). We just recently sequenced the 16S rRNA genes of PLS A (GenBank accession number MK606068). The analysis shows that it has a close relationship with *Geobacillus thermoleovorans* strain SGAir0734 [17]. The difference might be due to the variation of parameters in the algorithms, evolutionary rates of the genes, and the occurrence of horizontal gene transfer [18]. To alleviate the incongruency of the phylogenetic trees, concatenated data could be used. The method uses multiple genes to examine the overlap between individual and concatenated alignments [19]. However, the method was not applied in this study.

3.3. Amino Acid Sequence Alignment

The amino acid residues in the C-terminus region of PLS A were compared to those of PLS 80, *Bacillus caldolyticus strain XM,* and *Geobacillus stearothermophilus.* Alignment with DNA Pol I from PLS 80 was conducted because they were isolated from the same sampling location. *B. caldolyticus* strain XM was chosen because it was closely related to the phylogenetic tree (Figure 3). At the same time, *G. stearothermophilus* was selected because its protein structure has the closest similarity to PLS A (Figure 5).

Previous studies show that sequence alignments indicate five conserved regions in the C-terminal of DNA Pol I [3, 20]. In general, there were only marginal variances of the amino acid sequences of DNA Pol I in this study. In Region 1, except for PLS 80, all DNA Pol I contained a sequence of FNQALTQTGRLSSTEPNL (Figure 4). Region 1 is located at the tip of the thumb subdomain and forms a helix-loop that binds with the minor groove of the double-stranded DNA (see Figure 5). The sequence in Region 2, Motif A and Motif B were perfectly identical. However, there was a slight alteration in Motif C, particularly downstream of the conserved HDE sequence (Figure 4). Region 2 is located within the palm subdomain and interacts with the template strand. Motifs A, B, and C facilitate the dNTP binding cleft within the DNA polymerase active site. Motifs A and C are located within the palm sub-domain (see Figure 5) and are conserved within families A, B, C, and X polymerases. In contrast, motif B is located in the fingers subdomain [3].

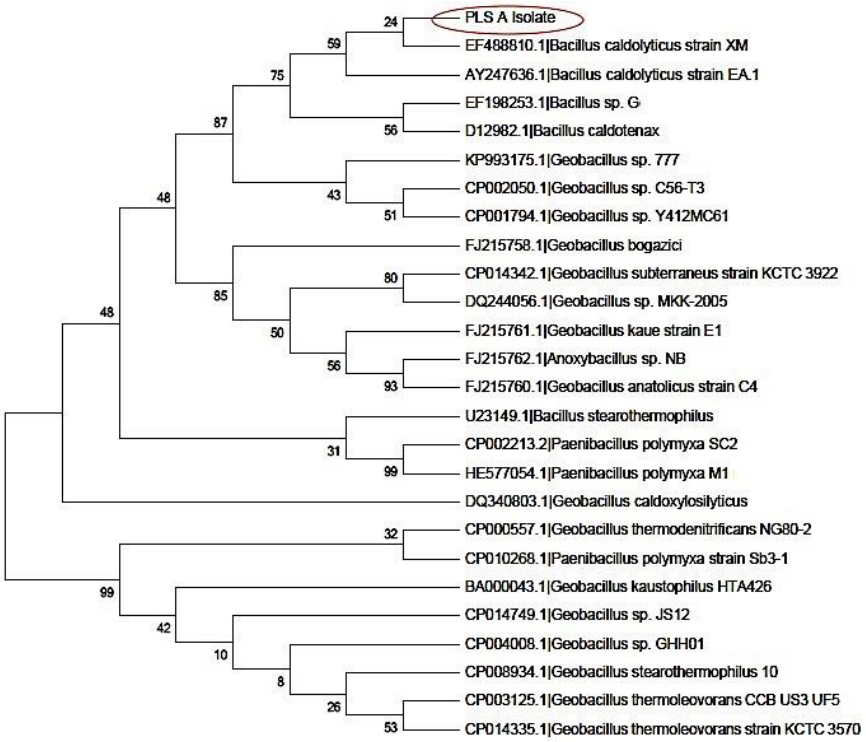

Figure 3. Phylogenetic tree of DNA Pol I gene fragments of the PLS A strain with 26 sequences using the neighbor-joining method in MEGA 6 with 1000 bootstrap replicates.

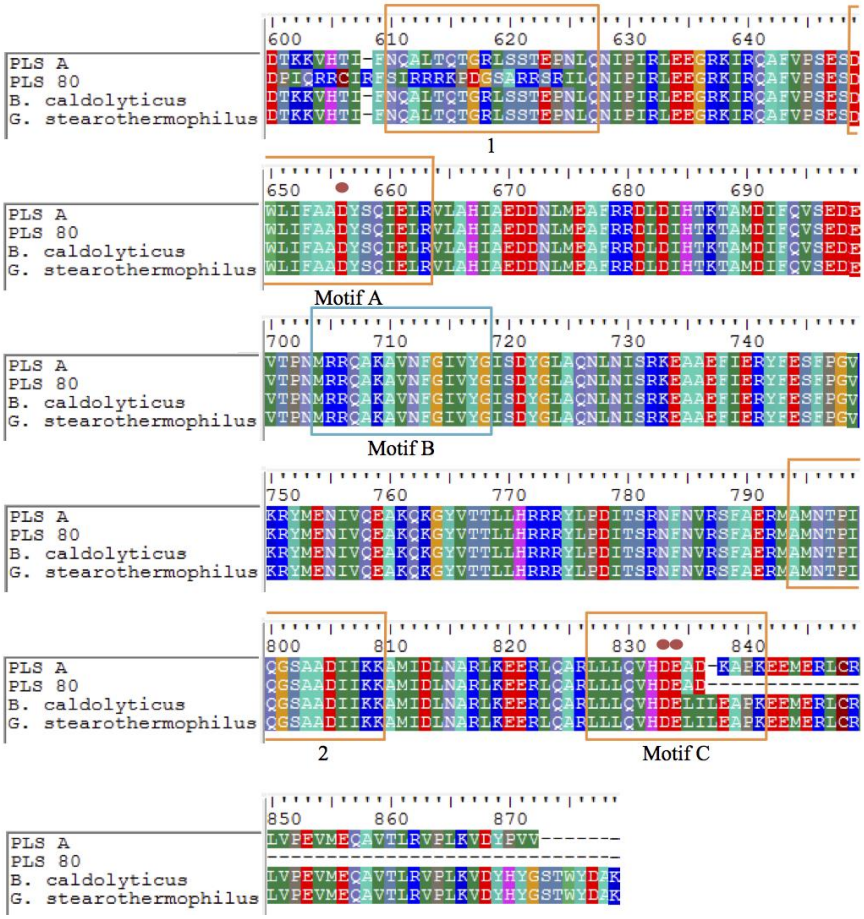

Figure 4. Sequence alignment of the amino acids translated from the 0.9 kb DNA Pol I gene fragment from PLS A using the ClustalW program. The amino acids were aligned with those of PLS 80, *Bacillus caldolyticus*, and *Geobacillus stearathermophilus.* The sequences in boxes represent the conserved regions; (1), (2), (3) are located in the palm sub-domain area, Motif A and Motif C are in the palm subdomain, while Motif B is in the fingers sub-domain. (●) shows amino acid residues that bind to divalent Mg^{2+} ion. All amino acid numbering was based on the PLSA sequence.

3.4. Structural Analysis

The structure of the DNA Pol I polymerization domain from PLS A was predicted by translating the DNA sequence to amino acids. The amino acid sequence was then compared to those of another DNA Pol I structures in the PDB. The highest similarity was shown by DNA Pol I from *G. stearothermophilus*, which was then used as a template for protein structure comparison. Before performing the modeling, the amino acid sequence was first aligned using the ClustalW program.

The 3D structure modeling of DNA Pol I from PLS A was performed using the SWISS-MODEL program (http://swissmodel.expasy.org/), based on the homology principle [21]. The VMD result confirmed that the structure of DNA Pol I from PLS A was the 5'–3' polymerase domain, which plays a vital role in the DNA polymerization process. There was no notable difference between the structure of the 5'–3' polymerase domain from PLS A and *G. stearathermophilus* as they were perfectly superimposed (Figure 5). This is in agreement with the explanation in Section 3.3 as they have nearly identical amino acids sequence in the conserved region. It has been well accepted that despite distantly related and show low homology (<50 %), the 3D structures of prokaryotic DNA Pol I are practically indistinguishable [3, 22]. This explains why despite distantly related in the phylogenetic tree (Figure 3), PLS A and *G. stearathermophilus* were closely similar in terms of their protein structures.

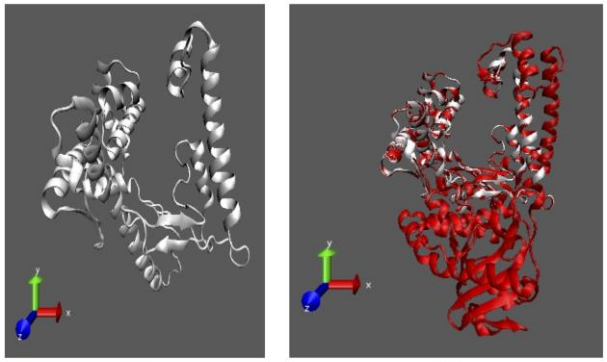

Figure 5. (A) Structural modeling of the DNA Pol I from PLS A strain. (B) Reflection of the structure of DNA pol I from PLS A strain to that of *Geobacillus stearathermophilus*

4. Conclusion

The homology analysis of the 0.9 kb gene fragment of DNA Pol I from PLS A showed that it was closely related to *B. caldolyticus*, with a sequence similarity of 99%. The results showed that the translated fragment was part of the polymerase domain and homolog to that of *G. stearothermophilus*. The results of this study may be used for further investigation and development of DNA pol I from PLSA strain.

Acknowledgments

This study is funded by Prospective Professorship Research Grant 2019 from Universitas Syiah Kuala. We thank Diana Eka Putri in refining the amino acid alignment.

References

[1] Kary B. Mullis and Fred A. Faloona, [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction, in: *Methods in Enzymology*, Academic Press, 1987, pp. 335-350 https://doi.org/10.1016/0076-6879(87)55023-6

[2] Joos Aschenbrenner and Andreas Marx, DNA polymerases and biotechnological applications, *Current Opinion in Biotechnology*, 48, (2017), 187-195 https://doi.org/10.1016/j.copbio.2017.04.005

[3] Premal H. Patel, Motoshi Suzuki, Elinor Adman, Akeo Shinkai and Lawrence A. Loeb, Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection11Edited by J. H. Miller, *Journal of Molecular Biology*, 308, 5, (2001), 823-837 https://doi.org/10.1006/jmbi.2001.4619

[4] Jonathan Filée, Patrick Forterre, Tang Sen-Lin and Jacqueline Laurent, Evolution of DNA Polymerase Families: Evidences for Multiple Gene Exchange Between Cellular and Viral Proteins, *Journal of Molecular Evolution*, 54, 6, (2002), 763-773 https://doi.org/10.1007/s00239-001-0078-x

[5] I. R. Lehman, Discovery of DNA Polymerase, *Journal of Biological Chemistry*, 278, 37, (2003), 34733-34738 http://doi.org/10.1074/jbc.X300002200

[6] Anna Bębenek and Izabela Ziuzia-Graczyk, Fidelity of DNA replication—a matter of proofreading, *Current Genetics*, 64, 5, (2018), 985-996 https://doi.org/10.1007/s00294-018-0820-1

[7] Sonoko Ishino and Yoshizumi Ishino, DNA polymerases as useful reagents for biotechnology – the history of developmental research in the field, *Frontiers in Microbiology*, 5, 465, (2014), 1-8 https://doi.org/10.3389/fmicb.2014.00465

[8] Roberto Laos, Ryan W. Shaw and Steven A. Benner, Engineered DNA Polymerases, in: K.S. Murakami, M.A. Trakselis (Eds.) *Nucleic Acid Polymerases*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 163-187 https://doi.org/10.1007/978-3-642-39796-7_7

[9] G. D. Haki and S. K. Rakshit, Developments in industrially important thermostable enzymes: a review, *Bioresource Technology*, 89, 1, (2003), 17-34 https://doi.org/10.1016/S0960-8524(03)00033-6

[10] Akhmaloka, Hendro Pramono, Laksmi Ambarsari, Evi Susanti, Santi Nurbaiti and Fida Madayanti, Cloning, homological analysis, and expression of DNA Pol I from Geobacillus thermoleovorans, *International Journal of Integrative Biology*, 1, 3, (2007), 206-215

[11] N. V. Bhagavan and Chung-Eun Ha, Chapter 22 - DNA Replication, Repair, and Mutagenesis, in: N.V. Bhagavan, C.-E. Ha (Eds.) *Essentials of Medical Biochemistry*, Academic Press, San Diego, 2011, pp. 287-299 https://doi.org/10.1016/B978-0-12-095461-2.00022-9

[12] Febriani Febriani, Hira Helwati, Muhammad Velayati and Teuku Iqbalsyah, Identification of a DNA polymerase i gene fragment from a local isolate (PLS 80) from an underwater hot spring, *Research*

Journal of Chemistry and Environment, 22, Special issue II, (2018), 189-192

[13] J. Zhou, M. A. Bruns and J. M. Tiedje, DNA recovery from soils of diverse composition, *Applied and Environmental Microbiology*, 62, 2, (1996), 316-322

[14] Koichiro Tamura, Glen Stecher, Daniel Peterson, Alan Filipski and Sudhir Kumar, MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, *Molecular Biology and Evolution*, 30, 12, (2013), 2725-2729 https://doi.org/10.1093/molbev/mst197

[15] William Humphrey, Andrew Dalke and Klaus Schulten, VMD: Visual molecular dynamics, *Journal of Molecular Graphics*, 14, 1, (1996), 33-38 https://doi.org/10.1016/0263-7855(96)00018-5

[16] P. D. Gutman, P. Fuchs, L. Ouyang and K. W. Minton, Identification, sequencing, and targeted mutagenesis of a DNA polymerase gene required for the extreme radioresistance of Deinococcus radiodurans, *Journal of Bacteriology*, 175, 11, (1993), 3581 http://doi.org/10.1128/jb.175.11.3581-3590.1993

[17] Teuku M. Iqbalsyah, Malahayati, Atikah and Febriani, Purification and partial characterization of a thermo-halostable protease produced by Geobacillus sp. strain PLS A isolated from undersea fumaroles, *Journal of Taibah University for Science*, 13, 1, (2019), 850-857 https://doi.org/10.1080/16583655.2019.1650489

[18] Jose Castresana, Topological variation in single-gene phylogenetic trees, *Genome Biology*, 8, 6, (2007), 216 https://doi.org/10.1186/gb-2007-8-6-216

[19] Thorsten Thiergart, Giddy Landan and William F. Martin, Concatenated alignments and the case of the disappearing tree, *BMC Evolutionary Biology*, 14, 1, (2014), 266 https://doi.org/10.1186/s12862-014-0266-0

[20] Marc Delarue, Olivier Poch, Noel Tordo, Dino Moras and Patrick Argos, An attempt to unify the structure of polymerases, *Protein Engineering, Design and Selection*, 3, 6, (1990), 461-467 https://doi.org/10.1093/protein/3.6.461

[21] Marco Biasini, Stefan Bienert, Andrew Waterhouse, Konstantin Arnold, Gabriel Studer, Tobias Schmidt, Florian Kiefer, Tiziano Gallo Cassarino, Martino Bertoni, Lorenza Bordoli and Torsten Schwede, SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Research*, 42, W1, (2014), W252-W258 https://doi.org/10.1093/nar/gku340

[22] F. C. Lawyer, S. Stoffel, R. K. Saiki, K. Myambo, R. Drummond and D. H. Gelfand, Isolation, characterization, and expression in Escherichia coli of the DNA polymerase gene from Thermus aquaticus, *Journal of Biological Chemistry*, 264, 11, (1989), 6427-6437