

Technical Notes

Biopython: Comparing the DNA Polymerase I (polA) Gene of Thermophilic, Mesophilic, and Psychrophilic Bacteria

Chaidir Adam^{1*}

¹Department of Biology Education, Faculty of Teacher Training and Education, University of Palangka Raya, Palangka Raya, Indonesia

*email: chaidir.adam03@gmail.com

Keywords:

Biopython

Python

Bioinformatics

Computational Biology

GC-Contents

Submitted: 04/05/2021

Revised: 06/06/2021

Accepted: 08/06/2021

Abstract. Biopython is a specialized Python tool for computational molecular biology. Various computational molecular analysis that can be performed using Biopython, such as reconstructing phylogenetic trees, multiple sequence analysis, generating complementary sequences, counting amino acids, etc. This technical notes paper describes in detail the procedures computational DNA sequence analysis using Biopython. The DNA polymerase I (polA) gene sequences of bacteria were used in this study to compare the differences between Thermophilic, Mesophilic, and Psychrophilic bacteria.



This is an open access article distributed under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2021 by author.

1. INTRODUCTION

Computational biology in the 21st century is growing rapidly, reaching almost all areas of biological sciences. (Markowitz, 2017) said that “*all biology is computational biology*”, biological knowledge nowadays is analyzed, defined, organized, and accessed through computation. Computational biology can be used to solve problem in bioinformatics, such as analyzing the large data of genetic code. Analysis in bioinformatics mainly focuses on three types of big data sets available in molecular biology, e.g. Genome Sequences (Bayat, 2002; Luscombe et al., 2001; Markowitz, 2017).

There are several programming languages that are commonly used in

bioinformatics, such Python, R, Java, C++, and Bash (Fourment & Gillings, 2008). Python is the most popular among them as there is a Python package that has been developed specifically for biologist, namely Biopython (Chapman & Chang, 2000).

Biopython is a specialized Python tool for computational molecular biology. Various computational molecular analysis that can be performed using Biopython, such as reconstructing phylogenetic trees, multiple sequence analysis, generating complementary sequences, counting amino acids, finding specific sequences, etc (Chapman & Chang, 2000; Cock et al., 2009).

Simple example of DNA sequence analysis using Biopython is Comparing the

DNA polymerase I (polA) gene sequences of several bacterial species from different groups in order to understand the characteristic differences between them genetically, as there are several bacterial species with capability to survive in extremes of temperature and pressure. Based on their ability to survive in an environment with a certain temperature range, bacteria can be classified into 3 three groups, namely Thermophilic, Mesophilic, and Psychrophilic (Chen & Berns, 1980; Shing et al., 1975). These three bacterial groups are different genetically.

DNA polymerases are enzymes needed for DNA replication every time cell division. These enzymes form two identical DNA strands by assembling nucleotides (Building Blocks of DNA) based on one original DNA molecule (Garcia-Diaz & Bebenek, 2007; Lehman & Uyemura, 1976; Loeb & Monnat, 2008; Yoon et al., 2014). The essential role of DNA polymerases is to precisely and effectively recreate the genome to ensure the maintenance of the inherited genetic information through generations (Allen et al., 2011; Garcia-Diaz & Bebenek, 2007).

Specific DNA polymerases such as *Taq* Polymerase play a key role in the PCR (Polymerase Chain Reaction) process (Chien et al., 1976; Ishino & Ishino, 2014). *Taq* polymerase is a heat-resistant and

thermostable DNA polymerase that was originally isolated from thermophilic bacteria, *Thermus aquaticus*, in 1976 (Chien et al., 1976). *T. aquaticus* is a bacterium that lives in hot springs and hydrothermal vents, this is the reason that *Taq* polymerase is an enzyme that is tolerant of high temperatures.

This technical notes paper describes in detail the procedures computational DNA sequence analysis using Biopython. The DNA polymerase I (polA) gene sequences of bacteria were used in this study to compare the differences between Thermophilic, Mesophilic, and Psychrophilic bacteria.

2. DNA SEQUENCE ANALYSIS USING BIOPYTHON

2.1. DNA Polymerase I (polA) Gene Sequences

The DNA polymerase I (PolA) gene sequences of Thermophilic, Mesophilic and Psychrophilic bacteria are downloaded from NCBI directories. The detailed information about the DNA polymerase I gene sequences used in this paper are presented in Table 1.

Table 1. Bacterial DNA Polymerase I (polA) Gene

| No. | Gene Name | Gene Type | Descriptions | NCBI Gene ID |
|---------------------|-----------|----------------|---|--------------------------|
| Thermophilic | | | | |
| 1 | polA | Protein Coding | DNA Polymerase I [<i>Thermus thermophilus</i> HB8] | 3169068 |
| 2 | polA | Protein Coding | DNA Polymerase I [<i>Geobacillus stearothermophilus</i>] | 58572640 |
| Mesophilic | | | | |
| 3 | polA | Protein Coding | DNA Polymerase I [<i>Streptococcus pneumoniae</i> NCTC11032] | 61430887 |
| 4 | polA | Protein Coding | DNA Polymerase I [<i>Escherichia coli</i> str. K-12 substr. MG1655] | 948356 |

| Psychrophilic | | | | |
|---------------|------|----------------|---|----------|
| 5 | polA | Protein Coding | DNA Polymerase I [<i>Arcobacter butzleri</i> ED-1] | 56463489 |
| 6 | polA | Protein Coding | DNA Polymerase I [<i>Shewanella putrefaciens</i> CN-32] | 45044408 |

2.2. DNA Sequences Analysis

The analysis of DNA polymerase I gene sequences are performed using Biopython computationally. Genetic properties analyzed in this technical notes paper are: GC-Content and Amino Acids. The detailed information regarding to the software used is presented in Table 2.

Table 2. Software List

| No. | Software | Version | Desc. |
|-----|-----------|-----------------------|--|
| 1 | MacOS | 10.9.5 (Mavericks) | Operating System |
| 2 | Python | 3.9.2 | Main Programming Language |
| 3 | IDLE | 3.9.2 | IDLE (Integrated Development and Learning Environment) is an integrated development environment (IDE) for Python |
| 4 | Biopython | 1.78 | tools for biological computation written in Python |

The following procedures of DNA sequences analysis consisting of Python scripts that are adopted from official *Biopython Tutorial and Cookbook* (Chang et al., 2020).

Calculate GC-Content

GC-content is the percentage of nitrogenous bases in DNA or RNA consisting of guanine (G) and cytosine (C) and is one of basic parameters commonly used to describe genomes (Gao & Zhang, 2006; Piovesan et al., 2019). There are 2 (two) methods of using Biopython scripts that can be used to calculate the GC-content of the gene sequences.

GC-content: First Method (1)

```

>>> import Bio.Seq
>>> from Bio.Seq import Seq
>>> my_seq = Seq("DNA Sequence")
>>> l = len(my_seq)
>>> g = my_seq.count("G")
>>> c = my_seq.count("C")
>>> 100 * float(g + c) / l

```

GC-content: Second Method (2)

```

>>> import Bio.Seq
>>> from Bio.Seq import Seq
>>> from Bio.SeqUtils import GC
>>> my_seq = Seq("DNA Sequence")
>>> GC(my_seq)

```

Notes: replace **DNA Sequence** with the strings of the DNA sequence

Figure 1. Biopython Scripts: Calculate GC-content

Bio.Seq and Bio.SeqUtils are the Biopython modules used to calculate the GC-content of a DNA sequence. The following example shows the application of the scripts above to calculate the GC-content.

```

>>> import Bio.Seq
>>> import Bio.SeqUtils
>>> from Bio.Seq import Seq
>>> from Bio.SeqUtils import GC
>>> polA_T1 = Seq("ATG....TAG")
>>> l = len(polA_T1)
>>> l
2505
>>> GC(polA_T1)
67.86427145708583

```

Figure 2. GC-content Calculation using Biopython Scripts

Figure 2 shows the results of GC-content analysis using Biopython scripts. The DNA sequence is a sequence of DNA polymerase I (polA) gene of *Thermus thermophilus*. The sequence **ATG....TAG** that presented above is the shortened of the complete sequence in order to simplify the example. Based on the results, the polA gene of *T. thermophilus* has 67.86% GC-content of a whole sequence.

Amino Acids

Amino acids are organic compounds that combine to form proteins as the product of the gene expression. Amino acids have a role as the nitrogenous backbones for compounds like enzymes and hormones (LaPelusa & Kaushik, 2021; Lopez & Mohiuddin, 2021; Wu, 2009). There are 20 to 22 amino acids that comprise proteins including Alanine (A), Arginine (R), Asparagine (N), Aspartic Acid (D), Cystein (C), Glutamic Acid (E), Glutamine (Q), Glycine (G), Histidine (H), Isoleucine (I), Leucine (L), Lysine (K), Methionine (M), Phenylalanine (F), Proline (P), Serine (S), Threonine (T), Tryptophan (W), Tyrosine (Y), and Valine (V). The 21st and 22nd amino acids are Selenocystein and Pyrrolysine, both of them encoded by UGA and UAG that normally function as stop signals

(Rother & Krzycki, 2010; Wu, 2009; Yuan et al., 2010; Zhang et al., 2005). The following Biopython scripts can be used to analyze and calculate amino acids from the DNA sequence.

Transcription (1)

```

>>> from Bio.Seq import Seq
>>> sense = Seq("DNA Sequence")
>>> anti_sense = sense.reverse_complement()
>>> mRNA = sense.transcribe()

```

Translation (2)

```

>>> AA = mRNA.translate()

```

Calculate the Percentage of Amino Acids (3)

```

>>> l = len(AA)
>>> Ala = (AA.count("A") / l) * 100
>>> Arg = (AA.count("R") / l) * 100
>>> Asn = (AA.count("N") / l) * 100
>>> Asp = (AA.count("D") / l) * 100
>>> Cys = (AA.count("C") / l) * 100
>>> Glu = (AA.count("E") / l) * 100
>>> Gln = (AA.count("Q") / l) * 100
>>> Gly = (AA.count("G") / l) * 100
>>> His = (AA.count("H") / l) * 100
>>> Ile = (AA.count("I") / l) * 100
>>> Leu = (AA.count("L") / l) * 100
>>> Lys = (AA.count("K") / l) * 100
>>> Met = (AA.count("M") / l) * 100
>>> Phe = (AA.count("F") / l) * 100
>>> Pro = (AA.count("P") / l) * 100
>>> Ser = (AA.count("S") / l) * 100
>>> Thr = (AA.count("T") / l) * 100
>>> Trp = (AA.count("W") / l) * 100
>>> Tyr = (AA.count("Y") / l) * 100
>>> Val = (AA.count("V") / l) * 100

```

Notes: replace **DNA Sequence** with the strings of the DNA sequence; sense = coding strand; anti_sense = template strand; AA = Amino Acids

Figure 3. Biopython Scripts: Transcription and Translation

The scripts above represent the common process of Protein Synthesis which consists of 2 (two) main phases, namely

Transcription and Translation. Transcription is the process of copying DNA segments into mRNA by the enzyme RNA polymerase. mRNA (messenger RNA) consists of genetic codes called codons, sequences of trinucleotides that correspond to specific amino acids. Translation is the process of translating mRNA codons into amino acids to form proteins (Pánek et al., 2017; Smith, 2020; Taylor, 2006). The following example shows the application of the Transcription and Translation Biopython scripts to generate mRNA and protein sequences.

```
>>> import Bio.Seq
>>> import Bio.SeqUtils
>>> from Bio.Seq import Seq
>>> from Bio.SeqUtils import GC
>>> polA_T1 = Seq("ATG...TAG")
>>> sense = polA_T1
>>> anti_sense = sense.reverse_complement()
>>> mRNA = sense.transcribe()
>>> AA = mRNA.translate()
>>> sense
Seq('ATGGAGGCGATGCTTCCGCTCTTTGAACCCAAAGGCCG
GGTCCTCCTGGTGGAC...TAG')
>>> anti_sense
Seq('CTAACCTTGGCGGAAAGCCAGTCCTCCCCATCCCCA
CCTCCACCTCCAGGGG...CAT')
>>> mRNA
Seq('AUGGAGGCGAUGCUCUCCGCUCUUUGAACCCAAAGGCCG
GGUCCUCCUGGUGGAC...UAG')
>>> AA
Seq('MEAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRG
EPVQAVYGFAKSLLKA...KG*')

continue....
```

Figure 4. Transcription and Translation using Biopython Scripts

The amino acid composition calculation is performed after the protein sequence has been generated. The basic principle of calculating the amino acid composition is the ratio of each particular amino acid to the total number of amino acids in a protein sequence, and then transform it into Biopython scripts as presented in the following example.

```
....
>>> l = len(AA)
>>> Ala = (AA.count("A") / l) * 100
>>> Arg = (AA.count("R") / l) * 100
>>> Asn = (AA.count("N") / l) * 100
>>> Asp = (AA.count("D") / l) * 100
>>> Cys = (AA.count("C") / l) * 100
>>> Glu = (AA.count("E") / l) * 100
>>> Gln = (AA.count("Q") / l) * 100
>>> Gly = (AA.count("G") / l) * 100
>>> His = (AA.count("H") / l) * 100
>>> Ile = (AA.count("I") / l) * 100
>>> Leu = (AA.count("L") / l) * 100
>>> Lys = (AA.count("K") / l) * 100
>>> Met = (AA.count("M") / l) * 100
>>> Phe = (AA.count("F") / l) * 100
>>> Pro = (AA.count("P") / l) * 100
>>> Ser = (AA.count("S") / l) * 100
>>> Thr = (AA.count("T") / l) * 100
>>> Trp = (AA.count("W") / l) * 100
>>> Tyr = (AA.count("Y") / l) * 100
>>> Val = (AA.count("V") / l) * 100
>>> Ala
10.778443113772456
>>> Arg
8.502994011976048
>>> Asn
1.9161676646706587
>>> Asp
4.790419161676647
>>> Cys
0.11976047904191617
>>> Glu
10.179640718562874
>>> Gly
6.347305389221557
>>> His
2.5149700598802394
>>> Ile
1.9161676646706587
>>> Leu
15.209580838323353
>>> Lys
5.389221556886228
>>> Met
1.9161676646706587
>>> Phe
3.473053892215569
>>> Pro
5.9880239520958085
>>> Ser
3.592814371257485
>>> Thr
3.7125748502994016
>>> Trp
1.437125748502994
>>> Tyr
2.5149700598802394
>>> val
7.18562874251497
>>> Gln
2.3952095808383236
```

Figure 5. Amino Acid Composition Analysis using Biopython Scripts

2.3. DNA Comparison

GC-content

The comparison of DNA polymerase I gene of Thermophilic, Mesophilic, and Psychrophilic bacteria is based on results of the sequence analysis using Biopython scripts. The results (Table 2 and Figure 6) show that the thermophilic bacterial group consisting of *Thermus thermophilus* and *Geobacillus stearothermophilus* have the highest GC-content of DNA polymerase I (polA) gene with 67.86% and 55.21% respectively.

Table 2. DNA Polymerase (polA) Gene's GC-content Comparison of Thermophilic, Mesophilic, and Psychrophilic Bacteria

| No. | Species | polA Gene Properties | |
|----------------------|---------------------------------------|----------------------|----------------|
| | | Length (bp) | GC-content (%) |
| Thermophilic | | | |
| 1 | <i>Thermus thermophilus</i> | 2505 | 67.86 |
| 2 | <i>Geobacillus stearothermophilus</i> | 2637 | 55.21 |
| Mesophilic | | | |
| 3 | <i>Streptococcus pneumoniae</i> | 2634 | 43.96 |
| 4 | <i>Escherichia coli</i> | 2787 | 51.96 |
| Psychrophilic | | | |
| 5 | <i>Arcobacter butzleri</i> | 2691 | 25.86 |
| 6 | <i>Shewanella putrefaciens</i> | 2769 | 47.02 |

The average GC-content of thermophilic bacteria polA gene is 61.54% which is greater than the mean GC-content of mesophilic (47.96%) and psychrophilic (36.44%) bacteria. The thermophilic bacteria demonstrate a tendency (as presented in Figure 6) to high GC-content due to their adaptation to high temperatures (Kagawa et al., 1984; Musto et al., 2004; Saunders et al., 2003). *Thermus thermophilus* and *Geobacillus stearothermophilus* are thermophiles organisms can grow at temperatures ranging from 50-82°C (Oshima & Imahori, 1971, 1974) and 65–75°C

(Kotzekidou, 1999) respectively. Mesophilic and psychrophilic bacteria have lower growth temperature range. Mesophilic bacteria can grow at moderate temperatures ranging from 20-54°C and can grow optimally at temperatures ranging from 30-39°C (Schiraldi & De Rosa, 2016), whereas psychrophilic bacteria can grow at temperatures ranging from 0-20°C (Cavicchioli, 2016; Irwin, 2020).

These results also suggest that the DNA polymerase I enzymes of thermophilic bacteria are more thermally stable and are ideal for use in the PCR cycle, since there are more GC-content in the sequences. Polymerase Chain Reaction (PCR) is a technique in molecular biology uses to amplify the DNA target sequences from a DNA template by using thermal cycling process (Erlich, 1989; Wages, 2005). PCR requires a temperature-dependent DNA polymerase to enzymatically replicate the desired target sequences in the process (Caetano-Anollés, 2013).

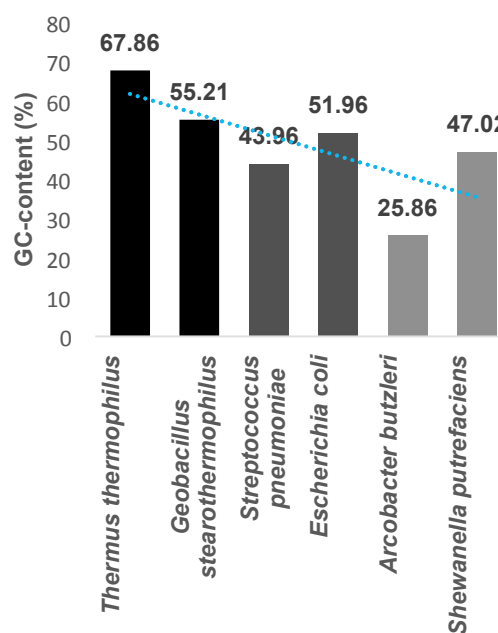


Figure 6. DNA Polymerase (polA) Gene's GC-content Comparison: Thermophilic, Mesophilic, and Psychrophilic Bacteria

Amino Acids

The length of generated protein sequences from the complete sequence of DNA polymerase I gene of thermophilic, mesophilic, and psychrophilic bacteria ranging from 835-929 amino acids (Table 3).

Table 3. The Length of Generated Protein Sequences from the Complete Sequence of DNA Polymerase I (polA) Gene

| No. | Species | Sequence Length (amino acids) |
|----------------------|---------------------------------------|-------------------------------|
| Thermophilic | | |
| 1 | <i>Thermus thermophilus</i> | 835 |
| 2 | <i>Geobacillus stearothermophilus</i> | 879 |
| Mesophilic | | |
| 3 | <i>Streptococcus pneumoniae</i> | 878 |
| 4 | <i>Escherichia coli</i> | 929 |
| Psychrophilic | | |
| 5 | <i>Arcobacter butzleri</i> | 897 |
| 6 | <i>Shewanella putrefaciens</i> | 923 |

Amino acid composition analysis results using Biopython scripts show that the DNA polymerase I (polA) gene of thermophilic, mesophilic and psychrophilic bacteria have relatively similar amino acid composition. The DNA polymerase genes from the three groups of bacteria are rich in leucine (Leu; L), Alanine (Ala; A), and glutamic acid (Glu; E), and low in cysteine (Cys; C) as presented in Figure 7. A sequence characterization study of DNA polymerase I (Brown et al., 1982), show a comparison of amino acid composition of *E. coli* DNA polymerase I and its derived fragments that are rich in leucine and glutamic acid, and low in cysteine. Leucine is one of nine essential amino acids and important for many metabolic functions, whereas glutamic acid and cysterin are non-essential amino acids (Lopez & Mohiuddin, 2021; Tessari et al., 2016).

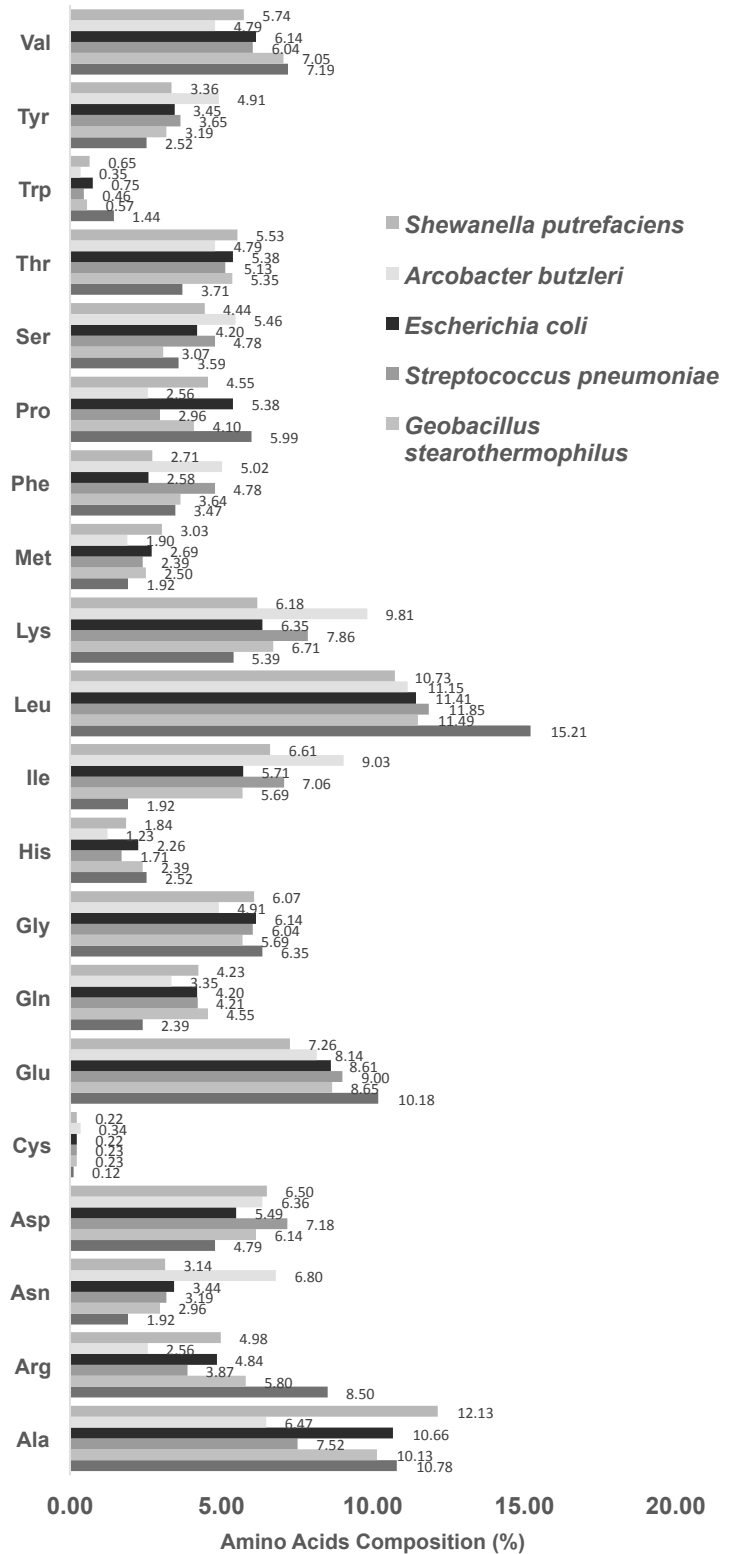


Figure 7. Amino Acid Composition (%) of The Generated Protein Sequences of DNA Polymerase I (polA) Gene

3. CONCLUSION

The DNA sequence analysis can be done using Biopython including GC-content and amino acid composition analysis. Bio.Seq and Bio.SeqUtils are the main modules of Biopython that used in this paper to compare the GC-content and amino acid composition of the DNA polymerase I (polA) gene of thermophilic, mesophilic, and psychrophilic bacteria. The sequence analysis shows the following results:

- 1) Thermophilic bacteria have the highest GC-content of the DNA polymerase I (polA) gene compare to mesophilic and psychrophilic bacteria;
- 2) DNA polymerase I (polA) gene of thermophilic, mesophilic and psychrophilic bacteria have relatively similar amino acid composition.

References

- Allen, J. M., Simcha, D. M., Ericson, N. G., Alexander, D. L., Marquette, J. T., Van Biber, B. P., Troll, C. J., Karchin, R., Bielas, J. H., Loeb, L. A., & Camps, M. (2011). Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. *Nucleic Acids Research*, 39(16), 7020–7033. <https://doi.org/10.1093/nar/gkr157>
- Bayat, A. (2002). Science, medicine, and the future: Bioinformatics. *BMJ (Clinical Research Ed.)*, 324(7344), 1018–1022. PubMed. <https://doi.org/10.1136/bmj.324.7344.1018>
- Brown, W. E., Stump, K. H., & Kelley, W. S. (1982). Escherichia coli DNA polymerase I. Sequence characterization and secondary structure prediction. *The Journal of Biological Chemistry*, 257(4), 1965–1972.
- Caetano-Anollés, D. (2013). Polymerase Chain Reaction. In S. Maloy & K. Hughes (Eds.), *Brenner's Encyclopedia of Genetics (Second Edition)* (pp. 392–395). Academic Press. <https://doi.org/10.1016/B978-0-12-374984-0.01186-4>
- Cavicchioli, R. (2016). On the concept of a psychrophile. *The ISME Journal*, 10(4), 793–795. PubMed. <https://doi.org/10.1038/ismej.2015.160>
- Chang, J., Chapman, B., Friedberg, I., Hamelryck, T., de Hoon, M., Cock, P., Antao, T., Talevich, E., & Wilczynski, B. (2020). *Biopython Tutorial and Cookbook*. <https://biopython.org/>
- Chapman, B., & Chang, J. (2000). Biopython: Python Tools for Computational Biology. *SIGBIO Newsl.*, 20(2), 15–19. <https://doi.org/10.1145/360262.360268>
- Chen, C.-H., & Berns, D. S. (1980). Thermotropic Properties of Thermophilic, Mesophilic, and Psychrophilic Blue-green Algae. *Plant Physiology*, 66(4), 596–599. <https://doi.org/10.1104/pp.66.4.596>
- Chien, A., Edgar, D. B., & Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *Journal of Bacteriology*, 127(3), 1550–1557. PubMed. <https://doi.org/10.1128/JB.127.3.1550-1557.1976>
- 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*, 25(11), 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>
- Erlich, H. A. (1989). Polymerase chain reaction. *Journal of Clinical Immunology*, 9(6), 437–447. <https://doi.org/10.1007/BF00918012>
- Fourment, M., & Gillings, M. R. (2008). A comparison of common programming languages used in bioinformatics. *BMC Bioinformatics*, 9(1), 82. <https://doi.org/10.1186/1471-2105-9-82>
- Gao, F., & Zhang, C.-T. (2006). GC-Profile: A web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. *Nucleic Acids Research*, 34(suppl_2), W686–W691. <https://doi.org/10.1093/nar/gkl040>
- Garcia-Diaz, M., & Bebenek, K. (2007). Multiple functions of DNA polymerases. *Critical Reviews in Plant Sciences*, 26(2), 105–122. PubMed. <https://doi.org/10.1080/07352680701252817>
- Irwin, J. A. (2020). Chapter 6—Overview of extremophiles and their food and medical applications. In R. Salwan & V. Sharma (Eds.), *Physiological and Biotechnological Aspects of Extremophiles* (pp. 65–87). Academic Press. <https://doi.org/10.1016/B978-0-12-818322-9.00006-X>
- Ishino, S., & Ishino, Y. (2014). DNA polymerases as useful reagents for biotechnology—The history of developmental research in the field. *Frontiers in Microbiology*, 5, 465–465. PubMed. <https://doi.org/10.3389/fmicb.2014.00465>
- Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T., & Oshima, T. (1984). High guanine plus cytosine content in the third letter of codons of an extreme thermophile. DNA sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. *The Journal of Biological Chemistry*, 259(5), 2956–2960.
- Kotzekidou, P. (1999). BACILLUS | *Bacillus Stearothermophilus*. In R. K. Robinson (Ed.), *Encyclopedia of Food Microbiology* (pp. 124–129). Elsevier. <https://doi.org/10.1006/rwfm.1999.0110>
- LaPelusa, A., & Kaushik, R. (2021). Physiology, Proteins. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK555990/>
- Lehman, I., & Uyemura, D. (1976). DNA polymerase I: essential replication enzyme. *Science*, 193(4257), 963. <https://doi.org/10.1126/science.781842>
- Loeb, L. A., & Monnat, R. J. (2008). DNA polymerases and human disease. *Nature Reviews Genetics*, 9(8), 594–604. <https://doi.org/10.1038/nrg2345>
- Lopez, M. J., & Mohiuddin, S. S. (2021). Biochemistry, Essential Amino Acids. In *StatPearls*. StatPearls Publishing.

- <http://www.ncbi.nlm.nih.gov/books/NBK557845/>
- Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*, 40(4), 346–358.
- Markowitz, F. (2017). All biology is computational biology. *PLOS Biology*, 15(3), e2002050. <https://doi.org/10.1371/journal.pbio.2002050>
- Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F., & Bernardi, G. (2004). Correlations between genomic GC levels and optimal growth temperatures in prokaryotes. *FEBS Letters*, 573(1–3), 73–77. <https://doi.org/10.1016/j.febslet.2004.07.056>
- Oshima, T., & Imahori, K. (1971). Isolation of an Extreme Thermophile and Thermostability of Its Transfer Ribonucleic Acid and Ribosomes. *The Journal of General and Applied Microbiology*, 17(6), 513–517. <https://doi.org/10.2323/jgam.17.513>
- Oshima, T., & Imahori, K. (1974). Description of *Thermus thermophilus* (Yoshida and Oshima) comb. Nov., a Nonsporulating Thermophilic Bacterium from a Japanese Thermal Spa. *International Journal of Systematic and Evolutionary Microbiology*, 24(1), 102–112. <https://doi.org/10.1099/00207713-24-1-102>
- Pánek, T., Žihala, D., Sokol, M., Derelle, R., Klimeš, V., Hradilová, M., Zadrobílková, E., Susko, E., Roger, A. J., Čepička, I., & Eliáš, M. (2017). Nuclear genetic codes with a different meaning of the UAG and the UAA codon. *BMC Biology*, 15(1), 8. <https://doi.org/10.1186/s12915-017-0353-y>
- Piovesan, A., Pelleri, M. C., Antonaros, F., Strippoli, P., Caracausi, M., & Vitale, L. (2019). On the length, weight and GC content of the human genome. *BMC Research Notes*, 12(1), 106. <https://doi.org/10.1186/s13104-019-4137-z>
- Rother, M., & Krzycki, J. A. (2010). Selenocysteine, Pyrrolysine, and the Unique Energy Metabolism of Methanogenic Archaea. *Archaea*, 2010, 453642. <https://doi.org/10.1155/2010/453642>
- Saunders, N. F. W., Thomas, T., Curmi, P. M. G., Mattick, J. S., Kuczek, E., Slade, R., Davis, J., Franzmann, P. D., Boone, D., Rusterholtz, K., Feldman, R., Gates, C., Bench, S., Sowers, K., Kadner, K., Aerts, A., Dehal, P., Detter, C., Glavina, T., ... Cavicchioli, R. (2003). Mechanisms of thermal adaptation revealed from the genomes of the Antarctic Archaea *Methanogenium frigidum* and *Methanococcoides burtonii*. *Genome Research*, 13(7), 1580–1588. <https://doi.org/10.1101/gr.1180903>
- Schiraldi, C., & De Rosa, M. (2016). Mesophilic Organisms. In E. Drioli & L. Giorno (Eds.), *Encyclopedia of Membranes* (pp. 1–2). Springer. https://doi.org/10.1007/978-3-642-40872-4_1610-2
- Shing, Y. W., Akagi, J. M., & Himes, R. H. (1975). Psychrophilic, mesophilic, and thermophilic triosephosphate isomerases from three clostridial species. *Journal of Bacteriology*, 122(1), 177–184.

- Smith, M. (2020). mRNA Transcription, Translation, and Defects in Developmental Cognitive and Behavioral Disorders. *Frontiers in Molecular Biosciences*, 7. <https://doi.org/10.3389/fmolb.2020.577710>
- Taylor, W. R. (2006). Transcription and translation in an RNA world. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 361(1474), 1751–1760. PubMed. <https://doi.org/10.1098/rstb.2006.1910>
- Tessari, P., Lante, A., & Mosca, G. (2016). Essential amino acids: Master regulators of nutrition and environmental footprint? *Scientific Reports*, 6(1), 26074. <https://doi.org/10.1038/srep26074>
- Wages, J. M. (2005). POLYMERASE CHAIN REACTION. In P. Worsfold, A. Townshend, & C. Poole (Eds.), *Encyclopedia of Analytical Science (Second Edition)* (pp. 243–250). Elsevier. <https://doi.org/10.1016/B012-369397-7/00475-1>
- Wu, G. (2009). Amino acids: Metabolism, functions, and nutrition. *Amino Acids*, 37(1), 1–17. <https://doi.org/10.1007/s00726-009-0269-0>
- Yoon, J.-H., Roy Choudhury, J., Park, J., Prakash, S., & Prakash, L. (2014). A role for DNA polymerase θ in promoting replication through oxidative DNA lesion, thymine glycol, in human cells. *The Journal of Biological Chemistry*, 289(19), 13177–13185. <https://doi.org/10.1074/jbc.M114.556977>
- Yuan, J., O'Donoghue, P., Ambrogelly, A., Gundllapalli, S., Sherrer, R. L., Palioura, S., Simonović, M., & Söll, D. (2010). Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. *FEBS Letters*, 584(2), 342–349. PubMed. <https://doi.org/10.1016/j.febslet.2009.11.005>
- Zhang, Y., Baranov, P. V., Atkins, J. F., & Gladyshev, V. N. (2005). Pyrrolysine and selenocysteine use dissimilar decoding strategies. *The Journal of Biological Chemistry*, 280(21), 20740–20751. <https://doi.org/10.1074/jbc.M501458200>