

Article

## Sequencing and Annotated of Goat's Milk using Bioinformatics Tool: ClustalX

Muhammad Najmi Mohammad Fauzi, Aisyah Aqilah Abu Bakar, Liyana Amalina Adnan, Tg Ainul Farha  
Tg Abdul Rahman, A'wani Aziz Nurdalila

*Halal Action Laboratory, Kolej GENIUS Insan, University Sains Islam Malaysia, Bandar Baru Nilai 71800, Nilai, Negeri Sembilan, Malaysia.*

*E-mail: najmifauzi2003@gmail.com, aisyahaqilah@gmail.com, liyanamalina@usim.edu.my, tgainulfarha@usim.edu.my, nurdalila.awani@usim.edu.my*

**Abstract**— Bioinformatics tool is a software program made to extract meaningful information from the mass of molecular biology or biological databases and carry out sequence or structural analysis. The method of determining the order of nucleotides within a deoxyribonucleic acid (DNA) molecule is known as DNA sequencing. This analysis is meant to be run to the commercialized or factory-made goat's milk (pasteurised) from various states in Malaysia to identify the milk's authenticity, either it is pure or mixed with other foreign substances from other animals. The main objective is to compare DNA sequences of commercialized and raw goat's milk (hand-milking and non-pasteurised). To achieve this, we used ClustalX to align and compare the obtained DNA from both milk samples. The sequences will be aligned using ClustalX software. ClustalX is a provider of an automated system for performing multiple alignments of sequences and profiles and evaluating the outcomes. The usage of ClustalX is helpful as it is cost-effective, user-friendly, and showing a high accuracy of the analysis.

**Keywords**— DNA sequencing; Bioinformatics tool; ClustalX; Commercialized factory-made goat's milk; Raw self-milking goat's milk

### I. INTRODUCTION

Goats generate around two percent of the world's total annual milk production. The production of goat milk is a fast-growing industry. It is now considered an important economic commodity in most countries around the world [1]. In Malaysia, the farmed goat population was estimated to increase from 429,398 to 439,667 goats between the year 2014 to 2015 [2]. Many of the goats are selectively bred for dairy purposes, such as milk. There are small and well-emulsified fat globules in goat milk, similar to natural cow milk, which implies that the fat stays trapped and suspended in the milk instead of rising to the surface [3].

In their peak, which is normally about the third or fourth lactation period, the dairy goats will be providing a daily average milk production of about 2.7 to 3.6 litres over a duration of ten-month lactation and generating more and more only after freshening. The milk then will progressively be decreasing the production by the end of the goat's lactation. Commonly, the milk measures an average of 3.5 percent natural fat or known as butterfat [4]. Goat milk plays a big role in the industry. It can be processed into products like butter and cheese. The quality of milk and its composition varies in many factors, such as animal health, breed, diet, feeding, parity, management system, and lactation stage [5].

Normally, goat milk is lighter in cholesterol than cow milk, making this option suitable for people who are observing their fat intake and cholesterol levels. Aside from that, people became highly aware that cow milk may not be suitable for certain people and have begun to search for a particularly suitable way to their body's demands. Goat milk is highly and closely similar to cow milk nutritionally than other milk and has some external features that set itself apart, which could affect health and digestibility. Those with cow milk problems could consume and drink goat milk without getting any issues or any side effects. Besides, they also claim that their symptoms like asthma, bloating, constipation, catarrh, eczema, and digestive pain are decreased and minimized or completely flee [6]. Milk and dairy products are significant for the human diet from various countries worldwide until a lot of commercialized milk is produced in the market. Goat's milk is famous for its nutrients. They are rich in quality proteins, carbohydrates, fat, vitamins, and minerals [15]. The consumption of goat's milk increases per year in Malaysia. There is an authentication problem in the commercialized pure goat's milk. In doing that, technology such as ClustalX can align and analyze the DNA from the samples. By looking at the aspects of cost-effective, easy to use, and accurate results in the analysis, the usage of ClustalX can be applied to determine the goat's milk authentication. Therefore, high demand for goat's milk might cause adultery in providing pure and authentic goat's milk to the community.

In theoretical and applied studies, DNA has become an interesting subject for the past years. DNA is a nucleotide polymer, while the basic components of DNA are nucleotides. There are four distinct nucleotide bases in DNA, consisting of cytosine (C), guanine (G), adenine (A), and thymine (T).

The method of determining the order of nucleotides within a molecule of DNA is known as DNA sequencing. Learning DNA sequencing has now become a prime method in general biological studies. Also, it has implementations in biological systematics, biotechnology, forensic biology, and pathology, as DNA is the source of genetic details.

In many areas, the techniques of DNA sequencing are the main instruments. The advantages of using these methods are gained by various science studies, spanning from anthropology, psychology, anatomy, biotechnology, microbiology, forensics, and many more. In many contexts, a quiet and spectacular revolution is underway, whereas DNA sequencing develops new revelation that fundamentally changes the conceptual foundations of many disciplines. Nonetheless, new and critical problems are arising with these changes, like health-related queries and bioethical concerns.

As an important method for the medical field, such as diagnosis, DNA sequencing has been evaluated. Main DNA sequencing purposes identify and comprehend the inner structure of molecules in genes, knowing which pattern codes for which type of proteins, detecting disease-related mutated genes, handling proteins due to their structural information, and treating illness based on the structural and DNA sequencing understanding [9].

DNA sequencing analysis is among the most significant bioinformatics components and has been evolving rapidly in recent years. The origin of any occurrence of existence is believed to be revealed by this analysis. In an attempt to comprehend the genetic variability among species, sequencing analysis is essential and needed. Various methods and ways of analyzing and comparing genetic sequences are used. Originally, many of these methods are the alignment-based approaches that are regularly used. In each of these approaches, based on specified outcome measures, molecular sequences are optimally matched. The alignment-based methods mainly offer a result with good and better accuracy. These techniques also have the potential to reveal the relationships among DNA sequences. Some algorithms have been arranged and established into the bioinformatics tool for sequence arrangement. Although these approaches generally help in biology, some of the biggest disadvantages are that they consume time and are very expensive to buy and purchase [13].

There is a verse in Al-Quran related to milk and somehow to this study, obtained from Surah Al-Mukminun verse 21.

وَإِنَّ لَكُمْ فِي الْأَنْعَامِ لَعِبْرَةً ۚ نُسْقِيكُمْ مِمَّا فِي بُطُونِهَا وَلَكُمْ فِيهَا مَنَافِعُ  
كَثِيرَةٌ وَمِنْهَا تَأْكُلُونَ

"And indeed, for you in livestock is a lesson. We give you drink from that which is in their bellies, and for you in them are numerous benefits, and from them, you eat."

The milk mentioned in the verse above is fresh milk that is 100% milk without any additional substances based on tafsir Al-Muyassar. This verse also stated that there are many benefits that we can get from the milk where the milk needs to be fresh and pure. According to the food act (1985), milk is defined as pure and fresh milk from cows, goats, buffalos, or sheep [14], which explains the verse above. This verse is somehow related to this study. The goat's milk samples will be identified using sequencing to see whether the samples are 100% pure milk or contain another mixture of other species such as bovine milk that can cause adultery to the milk.

To analyze the goat's milk's authenticity, one method involved in the process is running a sequencing of DNAs [7] obtained from the milk and comparing it between the self-milking and commercialized milk found in the market. In this research, the DNA from the goat's milk will be used in the DNA sequencing step.

## II. MATERIALS AND METHODOLOGY

### A. Collection and Preservation

Fresh goat milk was individually taken from local farm representing two species, *Capra aegagrus hircus* (Jamnapari) and *Capra aegagrus hircus* (Saanen) from the total of 14 states of Malaysia (Perlis, Pulau Pinang, Kedah, Perak, Kelantan, Pahang, Negeri Sembilan, Selangor, Kuala Lumpur, Melaka, Johor, Sabah, Sarawak). Comparing that, each commercialized milk from brand Nubian and brand UK Farm was collected from Malaysia's North, South, East, West, and Klang Valley. Nubian is centralized in Johor Bahru and under Farm Fresh, Malaysia, while UK Farm is centralized in Kluang, Johor.

All samples were kept in the icebox with a temperature below 0°C (in frozen condition) to ensure that every milk samples' freshness was preserved.

### B. DNA Sequencing

After every milk samples' DNA was extracted and underwent PCR, the PCR products were purified by innuPREP DOUBLEpure Kit (Analytik Jena, Germany). According to the manufacturer's protocol, the samples were then directly sequenced using Applied BioSystem3730xl DNA Analyzer (ABI, USA). The DNA sequencing was sent and done by MATRIOUX (M) Sdn. Bhd. [16].

### C. Alignment of DNA

When the DNA sequencing for every sample was successfully done, DNA sequences were aligned using ClustalX (ver. 2.1) [17].

This program is friendly-user as it is freely available and can be downloaded from the website (<http://www.clustal.org/clustal2/>). Next, open the File and choose Load Sequences. It is then followed by choosing the DNA sequences to be aligned. The format of the file must be in (.txt). After all the DNA sequencing data have been placed in

the program, open the Alignment and select Do Complete Alignment to perform the progressive alignment of DNA [18]. Multiple Sequence Alignment was used as the parameter since multiple DNA sequences were needed to be aligned, and the consensus is more accurate. The manual of ClustalX is as shown in Fig. 2.

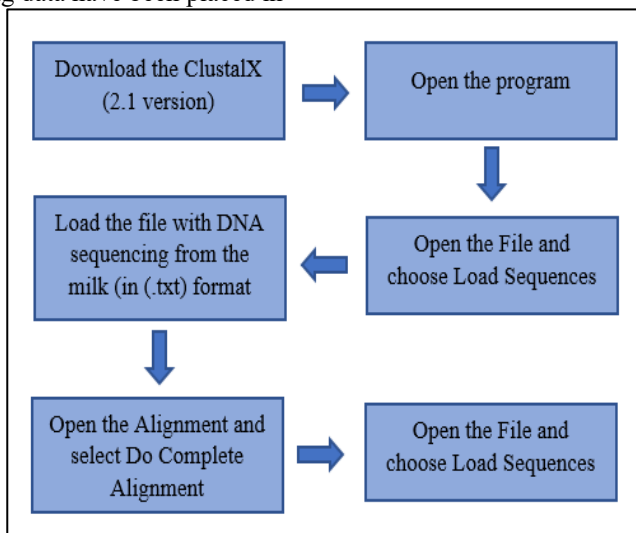


Fig. 1 The steps to align DNA using ClustalX (Version 2.1)

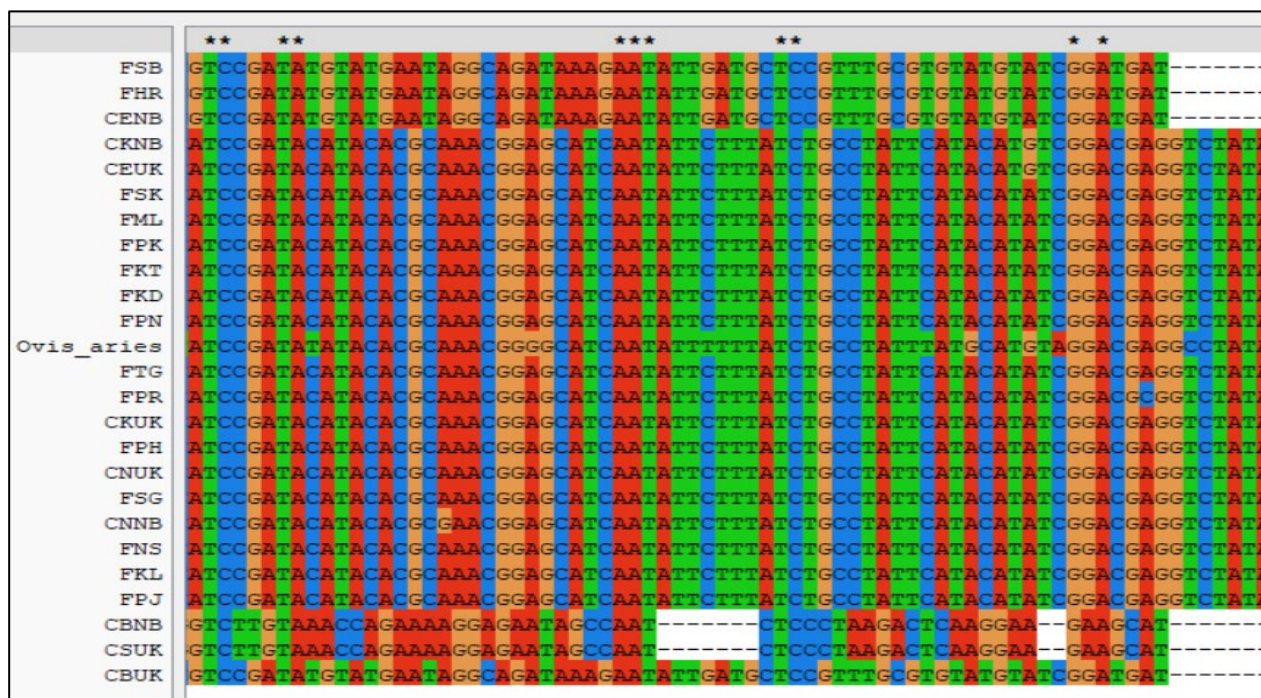


Fig. 2 The results of DNA alignments for every sample of milk from every state in Malaysia sample, self-milking goat's milk, and commercialized goat's milk from

### III. RESULTS

The alignments of DNA from every sample are shown in Fig. 2. It shows the sequence of *Ovis Aries* sp. as the control

different locations. There is a total of 24 samples of goat milk that have been compared with *Ovis aries* sp. Every color

presented in Fig. II represents the DNA bases, whereas red is for adenine, green is for thymine, blue is for cytosine, and orange is for guanine. Every representative of fresh goat milk from Fig. 2 is as shown in Table I.

TABLE I  
THE NAME SAMPLES OF FRESH GOAT MILK FROM EVERY STATE IN MALAYSIA

<i>SAMPLES</i>	<i>STATES</i>
<b>FPR</b>	FRESH PERLIS
<b>FPN</b>	FRESH PULAU PINANG
<b>FKD</b>	FRESH KEDAH
<b>FPK</b>	FRESH PERAK
<b>FKT</b>	FRESH KELANTAN
<b>FTG</b>	FRESH TERENGGANU
<b>FPH</b>	FRESH PAHANG
<b>FSG</b>	FRESH SELANGOR
<b>FKL</b>	FRESH KUALA LUMPUR
<b>FNS</b>	FRESH NEGERI SEMBILAN
<b>FML</b>	FRESH MELAKA
<b>FHR</b>	FRESH JOHOR
<b>FSK</b>	FRESH SARAWAK
<b>FSB</b>	FRESH SABAH

Every representative of commercialized fresh goat milk of brand Nubian from Fig. 2 is shown in Table II.

TABLE II  
THE NAME SAMPLES OF COMMERCIALIZED FRESH GOAT MILK OF BRAND NUBIAN AROUND MALAYSIA

<i>SAMPLES</i>	<i>AREA</i>	<i>STATES</i>
<b>CNNB</b>	<b>NORTH</b>	KEDAH PENANG PERAK PERLIS
<b>CENB</b>	<b>EAST</b>	KELANTAN PAHANG TERENGGANU
<b>CSNB</b>	<b>SOUTH</b>	MELAKA NEGERI SEMBILAN JOHOR
<b>CBNB</b>	<b>WEST</b>	SARAWAK SABAH
<b>CKNB</b>	<b>KLANG VALLEY</b>	KUALA LUMPUR SELANGOR

Every commercialized fresh goat milk of brand UK Farm from Fig. 2 is shown in Table III.

TABLE III  
THE NAME SAMPLES OF COMMERCIALIZED FRESH GOAT MILK OF BRAND UK FARM AROUND MALAYSIA

<i>SAMPLES</i>	<i>AREA</i>	<i>STATES</i>
<b>CNUK</b>	<b>NORTH</b>	KEDAH PENANG PERAK PERLIS
<b>CEUK</b>	<b>EAST</b>	KELANTAN PAHANG TERENGGANU
<b>CSUK</b>	<b>SOUTH</b>	MELAKA NEGERI SEMBILAN JOHOR
<b>CBUK</b>	<b>WEST</b>	SARAWAK SABAH
<b>CKUK</b>	<b>KLANG VALLEY</b>	KUALA LUMPUR SELANGOR

#### IV. DISCUSSIONS

Mitochondrial DNA is a crucial marker that allows the researcher to recognize and identify the *Capra aegagrus hircus* sp. for its benefits. Modern molecular biology has enabled comparison between the amino acid and nucleotide sequences of different populations for evaluating phylogeny and genetic diversity. In this world, sequencing technologies have played an important part in the study of gene sequences. Files that contain sequences of DNA are produced by a DNA sequencer [18]. These sequences are named readings on an alphabet consisting of five letters: A, C, G, T, or N. In 1977, Sanger was the first developer of sequencing technology and was awarded a Nobel Prize. His discovery has introduced the study of the genetic code of living life and the development of faster and efficient technology. Until now, there are three generations in the development of sequencing technologies [20].

A segment of the *cytb* genome of 358bp was examined to identify species throughout this analysis. This sample has the largest taxonomic fragmentation that symbolized the collection of data of nucleotides. A sequence record of an unknown sample, or at least of a similar species, is likely to be identified. In The European Bioinformatics Institute (EMBL) or GenBank or The DNA Data Bank of Japan (DDJB), there seem to be currently over 8000 *cytb* genome sequences of animal species accessible, and this set of data is steadily expanding and growing.

In this research, the DNA from the goat's milk will be used to undergo DNA sequencing. The milk goat has been obtained

from samples of commercialized fresh goat milk of brand Nubian and UK Farm around Malaysia. Milk and dairy products are significant for the human diet from various countries worldwide, leading to the production of many commercialized pure milks and commercialized milk produced in the market. Goat milk is famous for its nutrients. They are rich in quality proteins, carbohydrates, fat, vitamins, and minerals [21]. The consumption of goat's milk increases per year in Malaysia. In this research, the DNA from the goat milk will be used to undergo DNA sequencing.

By implementing effective DNA sequencing methods, the DNA sequences of the whole genomes from 5 eukaryotic species, ten eukaryotic chromosomes, and 55 prokaryotes have been established. This results in a large number of available data from DNA sequencing, and much more will indeed be available soon in other years. Bioinformatic tools are needed to analyze this enormous quantity of information to identify the genes encoding ribonucleic acid (RNA) or functional protein [22].

In this study, the main problem is the authentication problem in the commercialized (pasteurised) goat milk, either the milk is authentic or not. In doing that, many technologies can be used to analyze the DNA from the commercialized milk (pasteurised). By looking at the aspects of cost-effective, easy-to-use, and accurate results in the analysis, ClustalX is chosen as the tool to perform multiple sequence alignment of goat's milk samples.

The application of software, computing, and arithmetic to the management and monitoring of biological information to resolve biological issues is known as bioinformatics [23]. Moreover, a bioinformatics tool is a type of software program designed to extract valuable data from cell genetics or biological datasets. It is also used to conduct sequence or analysis of DNA structure in an organism [24].

Based on this research, the bioinformatics tool chosen to analyze the DNA of commercialized goat milk samples is ClustalX. We have used the ClustalX program to carry out multiple automatic alignments of amino acid sets or nucleotide sequences. ClustalX has been chosen because it uses an easy text menu system portable to, more or less, all computer systems. ClustalX features a few powerful graphical utilities to aid the interpretation of alignments and a graphical computer program. The program is used to require a collection of homologous genes from sequences. Not only that, it provides multiple alignments in the simplest use [25]. This encompasses the vast majority of the ClustalX usage and can be adequate for certain situations. Nevertheless, ClustalX even has detailed services to analyze basic phylogenetics, add sequences to current alignments, combine current alignments, identify and correct the alignment errors, and realign alignments parts.

The ability to cut and paste sequences to adjust the alignment pattern, the choice of sequence sections to be realigned, the selection of the sequence sub-range to be realigned, and putting it back to its original alignment are among the additional and updated features. This program can

also enhance and simplify complicated alignments, detect and catch the mistakes in input data. In addition, ClustalX has been created in many other software. Some of the software are IRIX5.3 on Silicon Graphics, SUN Solaris, Microsoft Windows (32 bit) for PCs, Linux ELF for x86 PCs, Macintosh, and Digital UNIX on DECstation [27].

Results obtained from using ClustalX have been discussed in the result section. The sequence of the commercialized goat milk samples from every state chosen has been compared and shown in the result. The control variable that has been used in this study is *Ovis aries sp.* DNA sequences were obtained from the species' milk sample. The result obtained is the same as expected except for the FSB sample, FHR sample, CENB sample, CBNB sample, CSUK sample, and CBUK sample. The result shows that most of the DNA was the same as the control variable except for the samples mentioned above. This means that the samples contained the DNA, which are almost the same as species like *Ovis aries sp.* It may slightly differ due to different parents from the DNA.

As we can see from the result shown above, there are some empty spaces between the DNA of the FSB sample, FHR sample, CENB sample, CBNB sample, CSUK sample, and CBUK sample. These empty spaces mean that the DNA from the samples cannot be read by the technology used. The results from some samples have empty spaces because sequencing is a one-time binding event. During the first-time sequence, several results cannot be read. This happened because the DNA strand that was taken to be sequenced contains very short strands. In order to avoid this problem, longer strands of DNA are needed to do sequencing [28][29].

## V. CONCLUSIONS

This study of sequencing and DNA alignment has shown that every sample of goat's milk has almost the same DNA composition. The DNA annotation can be used from this sequencing result in the next step of this study, the BLAST analysis. In the other aspect, ClustalX is also proven to be one of the best bioinformatics tools for aligning DNA sequences.

## ACKNOWLEDGEMENT

This work was supported by Headstart Grant, Universiti Sains Islam Malaysia (PPPI/HSP0118/KPI/051013/27318).

## REFERENCES

- [1] García, V., Rovira, S., Boutoial, K., & López, M. B. (2014). Improvements in goat milk quality: A review. *Small Ruminant Research*, 121(1), 51-57.
- [2] Jaafar, S. H. S., Hashim, R., Hassan, Z., & Arifin, N. (2018). A comparative study on physicochemical characteristics of raw goat milk collected from different farms in Malaysia. *Tropical life sciences research*, 29(1), 195.
- [3] Walsh, H. (2013). Starting right with milk goats. Read Books Ltd.

- [4] Arora, R., Bhojak, N., & Joshi, R. (2013). Comparative aspects of goat and cow milk. *Int. J. Eng. Sci. Invention Res. Dev*, 2, 7-10.
- [5] Goetsch, A. L., Zeng, S. S., & Gipson, T. A. (2011). Factors affecting goat milk production and quality. *Small Ruminant Research*, 101(1-3), 55-63.
- [6] Park, Y. W., & Haenlein, G. F. (2008). 2.5 Therapeutic and Hypoallergenic Values of Goat Milk and Implication of Food Allergy. *Handbook of milk of non-bovine mammals*, 121.
- [7] Sengupta, D. C., Hill, M. D., Benton, K. R., & Banerjee, H. N. (2020). Similarity Studies of Corona Viruses through Chaos Game Representation. *Computational molecular bioscience*, 10(3), 61
- [8] Pham, P. V. (2018). Medical Biotechnology: Techniques and Applications. In *Omics Technologies and Bio-Engineering* (pp. 449-469). Academic Press.
- [9] França, L. T., Carrilho, E., & Kist, T. B. (2002). A review of DNA sequencing techniques. *Quarterly reviews of biophysics*, 35(2), 169-200.
- [10] Whittaker, P. A. (2003). What is the relevance of bioinformatics to pharmacology? *Trends in Pharmacological Sciences*, 24(8), 434-439.
- [11] Broeckling, C. D., Reddy, I. R., Duran, A. L., Zhao, X., & Sumner, L. W. (2006). MET-IDEA: data extraction tool for mass spectrometry-based metabolomics. *Analytical Chemistry*, 78(13), 4334-4341.
- [12] Ni HM, Qi DW and Mu H (2018) Applying MSSIM Combined Chaos Game Representation to Genome Sequences Analysis. *Genomics*, 110, 180–190. 10.1016/j.ygeno.2017.09.010
- [13] Yang, Y., Adelstein, S. J., & Kassis, A. I. (2012). Target discovery from data mining approaches. *Drug discovery today*, 17, S16-S23.
- [14] Ulasan, B. U. K. U., & Alam, D. F. (2016). Kuala Lumpur: Telaga Biru Sdn. Bhd. ISBN 978-967-388-192-5. *Jurnal al-Turath*; Vol, 1(2).
- [15] Ulosoy B.H., 2015, Nutritional and Health Aspects of Goat Milk Consumption, *Akademik Gida* 13(1) (2015) 56-60.
- [16] Phylogenetic Relationships of *Epinephelus fuscoguttatus* and *Epinephelus hexagonatus* Inferred from Mitochondrial Cytochrome b Gene Sequences using Bioinformatic Tools.
- [17] Trivedi, S., Ansari, A. A., Ghosh, S. K., & Rehman, H. (2014). DNA Barcoding in Marine Perspectives. In *Aqaba International Conference on Marine and Coastal Environment, Status and Challenges in Arab World*. Aqaba, Jordan.
- [18] Barson, G., & Griffiths, E. (2016). SeqTools: visual tools for manual analysis of sequence alignments. *BMC research notes*, 9(1), 39.
- [19] Turnbull, C., Scott, R. H., Thomas, E., Jones, L., Murugaesu, N., Pretty, F. B., ... & Henderson, S. (2018). The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *Bmj*, 361.
- [20] Kchouk, M., Gibrat, J. F., & Elloumi, M. (2017). Generations of sequencing technologies: from first to next generation. *Biology and Medicine*, 9(3).
- [21] Ulosoy B.H., (2015). Nutritional and Health Aspects of Goat Milk Consumption, *Akademik Gida* 13(1) (2015) 56-60.
- [22] Roumpeka D.D., Wallace R.J., Escalettes F., Fotheringham I., Watson M., (2017). A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. *Frontiers in Genetics*, 8, 23.
- [23] Whittaker, P. A. (2003). What is the relevance of bioinformatics to pharmacology? *Trends in Pharmacological Sciences*, 24(8), 434-439.
- [24] Broeckling, C. D., Reddy, I. R., Duran, A. L., Zhao, X., & Sumner, L. W. (2006). MET-IDEA: data extraction tool for mass spectrometry-based metabolomics. *Analytical Chemistry*, 78(13), 4334-4341.
- [25] Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research*, 25(24), 4876-4882.
- [26] Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Thompson, J. D. (2007). Clustal W and Clustal X version 2.0. *bioinformatics*, 23(21), 2947-2948.
- [27] Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS: An International Journal*, 20(18), 3551-3567.
- [28] Redrejo-Rodríguez, M., Ordóñez, C. D., Berjón-Otero, M., Moreno-González, J., Aparicio-Maldonado, C., Forterre, P., ... & Krupovic, M. (2017). Primer-independent DNA synthesis by a family B DNA polymerase from self-replicating mobile genetic elements. *Cell reports*, 21(6), 1574-1587.
- [29] Krishnamurthy, Karthikeyan. (2018). Re: Please, I need help with explaining my " messy " sequencing result?. Retrieved from: [https://www.researchgate.net/post/Please\\_i\\_need\\_help\\_with\\_explaining\\_my\\_messy\\_sequencing\\_result/5b96d45211ec7302d76caa9a/citation/download](https://www.researchgate.net/post/Please_i_need_help_with_explaining_my_messy_sequencing_result/5b96d45211ec7302d76caa9a/citation/download).

Received 1<sup>st</sup> November 2020; Revised 23<sup>rd</sup> February 2021;  
Accepted 15 March 2021; Published 1 April 2021

Academic Editor: Shahrina Ismail  
USIM Press  
Malaysian Journal of Science, Health & Technology  
Vol. 7, No. 1 (2021), 6 pages  
Copyright 2021 Fakulti of Science and Technology, Universiti  
Sains Islam Malaysia.