FUNCTIONAL CHARACTERIZATION OF PUTATIVE *LATE ELONGATED HYPOCOTYL* (*LHY*) GENE IN *Stevia rebaudiana* MS007 via *IN SILICO* ANALYSIS

NURUL HIDAYAH SAMSULRIZAL [1,2*], ZABIRAH ABDUL RAHIM [1], SITI NOOR ELIANA MOHAMAD NAZAR [1], NUR FARHANA MUSTAFA [1]

[1]*Department of Plant Science, Kulliyyah of Science, International Islamic University Malaysia, 25200 Kuantan, Pahang, Malaysia*
[2]*Plant Productivity and Sustainable Resource Unit, Kulliyyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia*

**Corresponding author: hidayahsamsulrizal@iium.edu.my*

ABSTRACT: *Stevia rebaudiana* is a medicinal herb that has been known as a low-calorie sweetener. It has been recognised as an artificial sweetener that is 300 times sweeter than sucrose. It is a commercially important sweetener that has been consumed as food and medicine, due to its ability to reduce blood sugar levels. Stevioside is one of the important compounds that give food a sweet taste. Previous studies showed that high amount of stevioside could be produced by delaying the flowering process of stevia. The aim of this study is to characterise the putative *Late Elongated Hypocotyl* (*LHY*) gene that is involved in the flowering process by using bioinformatics tools. Both analyses by using BLAST and domain search found that *LHY* gene is involved in the flowering process. These domains are SANT/Myb domain (IPR001005/SM000717), TEA domain (SM00426), Myb domain (IPR017930), and Myb domain plants (IPR006447). The phylogenetic tree was built using 20 out of 100 protein sequences from BLAST. The physico-chemical properties of putative LHY protein had been conducted through ProtParam and TMHMM, which showed that the protein is a globular protein. The phylogenetic tree construction indicated the common ancestor for the putative LHY protein, which is from the same family, i.e., Asteraceae. In summary, this study improves our knowledge of Stevia MS007 *LHY* gene by *in-silico* analysis. Therefore, future research should focus on determining the precise function of the protein in regulating the blooming stage of the stevia plant.

KEY WORDS: *Stevia rebaudiana, Late Elongated Hypocotyl, LHY, steviol glycosides.*

1. INTRODUCTION

One of the alarming illnesses that had been around since ancient Egypt around 3,000 years ago is diabetes mellitus. In 2019, according to The Star Online, the Health of Minister Malaysia stated that there were 3.6 million people suffering from diabetes, and approximately 31.3% of adults aged 18 years and older will have the disease by 2025 [37]. In addition, the World Health Organization (WHO) estimates that 108 million people worldwide have diabetes, and the number will continue to rise in 2024. Nonetheless, according to the International Diabetes Federation (IDF), the number of individuals with diabetes is rising steadily [1]. Due to its reputation as a low-calorie sweetener, *Stevia rebaudiana* offers several options to control the regulation of sugar. *S. rebaudiana* is a member of the Asteraceae family and is a popular artificial sweetener that is roughly 300

times sweeter than sucrose [2]. Besides that, the leaves of *S. rebaudiana* were found to contain diterpene glycosides like stevioside, rebaudioside A-F, dulcoside, and steviolbioside [3]. These substances also have diuretic, antitumor, anti-inflammatory, anti-hypertensive, anti-hyperglycemic, and anti-tumor effects [4].

Steviol glycosides are sweetener compounds that can be extracted from the leaves and are considered to contain zero calories. However, the production of steviol glycosides in *S. rebaudiana* is influenced by many factors, such as environmental conditions, geographic localities, physiological properties of plants, altitude, nutrient deficiency in the soil, and pathogenic diseases [5,6]. Besides, it was mentioned that the highest steviol glycoside content is found in the leaves by delaying the flowering of stevia, allowing more time for glycoside accumulation [7]. Stevia is a short-day plant that requires 13 hours of critical light for the growth process [8]. Research shows that the highest accumulation of stevia glycosides occurs during the budding phase, followed by an initial flowering stage with less than 10% flowers [8]. In addition, it was mentioned that when stevia plants grow under short daylight, the production of steviol glycosides is lower and the growth of the plants will be retarded [5].

Photoperiod plays a major role in floral induction. Phototropism is when the plants align their photosynthetic organs toward the direction of sunlight. The alignment of plants' organs was regulated by the circadian clock. For example, in *Helianthus annuus*, the circadian clock guides solar tracking in the plants by controlling the movement of the stem and shoot apices either facing east at dawn or facing west at dusk [9]. The researchers reported that circadian oscillators enhanced the fitness of plants by coordinating physiological processes with changes in the environment. Studies of molecular genetics and biochemical studies have shown that the *Late Elongated Hypocotyl* (*LHY*) gene regulates photoperiodic flowering and expression of photoperiodic flowering genes *via* the circadian clock in Arabidopsis mutant plants [10]. The initial studies focused on isolating and characterizing the *LHY* gene in *Arabidopsis thaliana*, a model plant. This involved identifying its sequence, understanding its genomic organization, and examining its expression patterns over a 24-hour period to establish its involvement in circadian rhythms [35]. Besides, research by Niwa (2007) [36] demonstrated that *LHY* interacts with other clock components, such as *Circadian Clock Associated 1* (*CCA1*) and *Timing of Cab Expression 1* (*TOC1*) genes. These interactions form a complex regulatory network that contributes to the precise timing of circadian rhythms. The central components of circadian oscillators consist of the *LHY* gene and the *Circadian Clock Associated 1* (*CCA1*) gene, which help plants exhibit early flowering, even in unsuitable environments [10, 11]. The circadian clock provides information and photoreceptors to integrate light signals for plants to sense photoperiodic changes [10].

Hence, identification of putative *LHY* gene that is involved in flowering process will improve stevia plant production to not only produce high yield of leaves but also steviol glycosides content. Hence, the aim of this study is to characterise the putative *LHY* gene, which was predicted to be involved in the formation of Stevia's flowers. By modifying the flowering process, the anticipated outcome of this study will assist in raising the output of steviol glycosides.

2. MATERIALS AND METHODS

2.1 Data Collection

In this study, bioinformatics studies were carried out on the transcriptome data of Stevia MS007, where the dataset consists of raw sequence data derived through transcriptome sequencing [12, 13]. Seventeen query IDs from the dataset were verified, by referring to the *LHY* gene of Stevia MS007. Only one protein Cluster-31069.11078, out of the 17 *LHY* genes, was chosen to a larger extent.

2.2 Translation of nucleotides using Expasy

In order to obtain the protein sequence, the nucleotide sequence for Cluster 31069.11078 from the stevia dataset was sent to Expert Protein Analysis System (Expasy) (https://web.expasy.org/translate/). It enables accurate translation of DNA sequences into protein sequences, providing essential information for gene identification, protein prediction, comparative analysis, and functional annotation.The longest protein sequence was chosen from the raw data of Stevia's database, which was saved from *.txt into *.fasta. This conversion facilitates compatibility, metadata inclusion, analysis, and data sharing across various research areas, and accessible way to work with biological sequence data.

2.3 Homology search by BLAST

The protein sequence corresponding to Cluster-31069.11078 was then analysed to find similarity using Basic Local Alignment Searching Tools (BLAST), which are available at (https://blast.ncbi.nlm.nih.gov). The BLAST programme is an invaluable resource for researchers in various fields of biology. It provides a user-friendly interface to perform sequence similarity searches, aiding in functional annotation, evolutionary analysis, disease research, and others [14]. There are more than 100 sequences that have been aligned with the query sequence, and only 20 sequences with high percent identities from different species were selected. Then, the 20 sequences were downloaded as FASTA complete sequences for further use in phylogenetic analysis.

2.4 Protein Domain search by InterPro

The InterPro database (http://www.ebi.ac.uk/interpro/) was then used to perform a domain search for the LHY protein sequence. The database can predict and integrate the sequence by representing protein domains, families, and functional sites. Besides, it also aids researchers to uncover information about protein domains, motifs, and functional sites, leading to insights into protein function, evolution, and structure [15]. There are other sources of databases that are joined together in the InterPro, such as Protein Families (Pfam) (https://pfam.xfam.org) and Simple Modular Architecture Research Tool (SMART) (http://smart.embl.de/). Pfam is a database of protein domain families and allows comparison between query sequences with the Pfam database. Meanwhile, SMART is intended to allow automatic identification and annotation of protein domains.

2.5 Physico-chemical Properties of Proteins using ProtParam and TMHMM

ProtParam (https://web.expasy.org/protparam/) was used to identify the molecular weight, theoretical pl and amino acid composition of a protein sequence. On the other hand, the Transmembrane Helices Hidden Markov Model (TMHMM) (http://www.cbs.dtu.dk/services/TMHMM-2.0/) is an essential resource for predicting transmembrane helices within protein sequences. This prediction provides insights into the structure, topology, and function of membrane proteins, impacting various areas of research.

2.6 Phylogenetic analysis

There are two main steps to build phylogenetic tree, which are Multiple Sequence Alignment (MSA) and tree construction using Molecular Evolutionary Genetics Analysis (MEGA) [16]. The Multiple Sequence Alignment (MSA) will be generated using the MUSCLE software, which is available at (https://www.ebi.ac.uk/Tools/msa/muscle/) that is used for aligning multiple biological sequences, such as DNA, RNA, or protein sequences. This is essential for understanding sequence relationships, identifying functional domains, and performing comparative analyses. Then, Molecular Evolutionary Genetics Analysis (MEGA) software is designed for comparative analysis of the query sequence with the reference sequence. MEGA software can be downloaded from (https://www.megasoftware.net). The software is available for installation on Microsoft systems. Then, a phylogenetic tree was constructed by using 20 sequences from different plant species with 1000 replicates of bootstrap.

3. RESULTS AND DISCUSSION

3.1 Homology search

The BLAST programme was used to evaluate the information that was taken from the dataset of stevia MS007. There are 100 sequences matching hits with the putative *LHY* gene according to the homology search done using blastP. Nonetheless, only 20 sequences from various species with high percentages of identity were chosen. The maximum E-value in the study is between 10^{-30}. In order to establish the sequences, the E-value computation was important. Sequences were considered homologous if they have low E values with both protein sequences and sequences from the protein database [17]. According to study data presented in Table 1, all protein sequences have scores larger than 115 and percentage identities greater than 80%. *Helianthus annuus* has the highest total score and percentage identity, with scores of 134 and 94.12%, respectively. Hence, it is concluded that all proteins were closely related to this stevia protein sequences.However, protein LHY from [*Cynara cardunculus* var. *scolymus*] shows high identity but lower score value. In BLAST analysis, both "percentage identity" and "total score" are important metrics used to assess the similarity between a query sequence and a database sequence. However, they capture different aspects of the alignment and should be interpreted differently. A high percentage identity indicates a strong sequence similarity in the aligned regions, but the total score considers various factors including gaps and mismatches. If it shows high identity but a lower score, it suggests that while there are matching segments, there are also differences or gaps in the alignment that affect the overall alignment quality.

3.2 Domain search analysis

A crucial component that is stable and encapsulates the entirety of a protein's structure is a domain [18]. One or more domains make up the majority of the protein. Also, these conserved areas can be found in various types of proteins, which aids in the activity of the protein and are not just unique to one family of genes [18]. Thus, InterPro programme was used in this study's analysis of domain searches. The outcome of the domain search investigation was displayed in Table 2.

In addition, the prediction from the InterPro analysis showed that there were three types of domains present. The length of the query protein sequence is 67 amino acids, and the location of those domains is between 21 to 67 amino acids.

Table 1: The top ten hit sequences from homology search using blastP.

Accession	Description	Percentage Identity (%)	Total Score	E-value
XP_021983251.1	protein LHY-like [*Helianthus annuus*]	94.12	134	5.3e-35
XP_016735595.1	PREDICTED: protein CCA1-like [*Gossypium hirsutum*]	86.57	119	1.6e-33
XP_023753679.1	protein LHY isoform X3 [*Lactuca sativa*]	92.54	129	2.3e-33
XP_007139035.1	hypothetical protein PHAVU_009G259600g [*Phaseolus vulgaris*]	85.07	118	2.5e-33
XP_021896768.1	protein CCA1-like [*Carica papaya*]	85.07	115	3.3e-32
XP_024996248.1	protein LHY [*Cynara cardunculus var. scolymus*]	98.28	125	5.3e-32
XP_030535347.1	protein LHY isoform X1 [*Rhodamnia argentea*]	88.06	124	2.3e-31
NP_001296649.1	protein LHY [*Cicer arietinum*]	91.80	123	3.7e-31
XP_016463073.1	PREDICTED: protein LHY-like [*Nicotiana tabacum*]	89.55	123	4e-31
XP_020211115.1	protein LHY isoform X2 [*Cajanus cajan*]	90.16	122	6.1e-31

Table 2: Result of domain search analysis through InterPro, SMART and Pfam

Accession	Domain	Function of Domain
IPR001005 or SM000717		SANT domain • Present in chromatin-remodeling enzymes and histone acetyltransferase. • It has a strong structural similarity to the DNA-binding domain of Myb-related proteins. • It plays role in helix packing due to the presence of three α helices in the domain [19]. • The SANT domain is stabilizing the histone binding conformation by directly interacting to histone tail [19]. • SANT domain is functionally divergent from the canonical Myb DNA-binding domain. • Mostly found in eukaryotes, bacteria, viruses and archaea kingdom.
SM00426		TEA domain • The domain is a DNA-binding region of about 66 to 68 amino acids that has been named after the two proteins that originally defined the domain: TEF-1 and ABAA [20]. • TEF-1 can control transcription of genes while ABAA can regulate gene in asexual spore differentiation [20]. • This domain appears to be conserved throughout evolution and it is predicted to contain three alpha-helices which are involved in DNA binding [20]. • Mostly found in eukaryotes kingdom.
IPR006447	Myb domain, plants (24-67 amino acids)	• DNA-binding domain is restricted to plant proteins and contain a response regular domain [21]. • The domain related to Myb-like DNA-binding domain.
IPR017930	Myb domain (21-67 amino acids)	• The myb-type helix-turn-helix (HTH) domain approximately 55 amino acids occur in a tandem repeat in eukaryotic transcription factors [22]. • The 3-D structure of domain forms three alpha-helices which Helix 3 is a recognition helix that binds DNA major groove.

3.3 Prediction of Physico-chemical Properties of Putative LHY Proteins

The presence of the transmembrane helices is important and useful for functional annotation analysis. The prediction of transmembrane helices is much easier than that of helices in globular domains, where the prediction of transmembrane helices is based on the hydrophobicity properties of the transmembrane towards the query protein [23].

Furthermore, it was mentioned that the "Hidden Markov Model" (HMM), which provides information on the overall topology of the protein, can be used to predict the

topology of proteins [23]. In addition, TMHMM can predict all membrane proteins in full sequenced genomes and present statistics on the frequency of proteins with different topologies. Fig. 1 illustrates the result of TMHMM for accession of *LHY* gene for Cluster-31069.11078.

The graph shows three different lines with different colours: the bold pink line means the residues are outside the cell, the pink line shows whether the residues are inside the cytoplasm or not, and the bold blue line means the residues are inside the cell. Thus, it can be concluded that the protein is a globular protein and does not have transmembrane helices because the protein is located outside of the transmembrane.

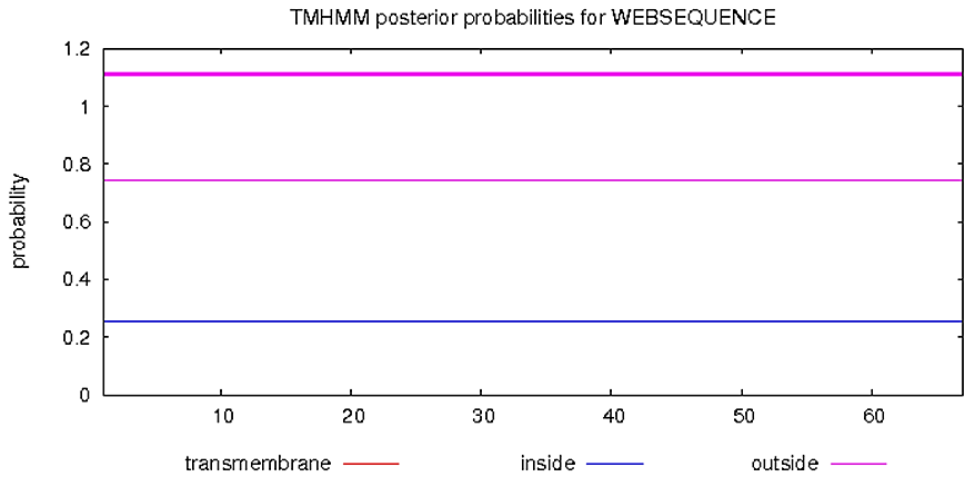

Fig. 1. Result of TMHMM for accession Cluster-31069.11078. The above bold pink line depicts that this protein is a globular protein without the presence of transmembrane.

Following that, Table 3 provided examples of predictions for the physical and chemical characteristics of putative LHY proteins. The analysis provided information on the quantity, molecular weight, and isoelectric point of the amino acids (pI). In addition, the instability index could also be obtained from this analysis, which indicates whether the protein is stable or unstable. If the instability index is lower than 40, the protein is stable, but if the index value is higher than 40, the protein is known as an unstable protein. LHY proteins are unstable proteins, so they cannot be easily extracted in the laboratory. All information from ProtParam analysis is handy, such as the molecular weight of amino acid, which is needed during electrophoresis analysis. In addition, incorporating analysis of physico-chemical properties adds depth and context to LHY proteins. The changes in properties across different species can provide insights into the adaptive evolution of proteins [38].

Table 3: Analysis of physico-chemical properties of putative LHY proteins.

Accession	Number of amino acids	Molecular weight (Da)	Isoelectric point (pI)	Instability index
Cluster-31069.11078	67	7829.75	9.52	46.73

3.4 Multiple Sequence Alignment

A vital initial phase in creating phylogenetic trees is multiple protein sequence alignments. In 2018, it was said that protein alignment is essential for identifying a protein's family, domain, and functional site as well as for predicting the protein's structure and function [24]. Local and global sequence alignment are the two different types [25]. Multiple sequence alignment was performed to align each sequence in the query set. In this study, MUSCLE was applied to align protein sequences. Fig. 2 presents the analysis of multiple sequence alignments. The different colours of proteins depict the different physico-chemical properties of proteins [26]. From the figure, it can be seen that at the C terminal end of proteins, there is not much variation based on the colour of the protein. Based on Edgar (2004), he said the conserved residues denoted with an asterisk (*). With the N- and C-termini of the sequences removed, a solid phylogenetic tree was created using the alignment of the sequences utilised in the analysis [26].

Fig. 2. Analysis of multiple sequence alignment using MUSCLE. In the protein sequences, the domain region is shown by a red box, and identical amino acid residues are marked by an asterisk (*).

3.5 Phylogenetic analysis

The putative LHY protein in this study is predicted to be involved in regulating the circadian clock of plants based on phylogram tree that contained bootstrapping values greater than 50% [10]. The maximum likelihood algorithm was chosen to construct this phylogenetic tree. These algorithms used a statistical approach to infer the probabilities of sequences in a model tree. The phylogenetic tree generated using the maximum likelihood tree approach is depicted in Fig. 3. The phylogenetic tree can be separated into three subgroups based on the "bootstrapping value," which is more than 70% with 1000 replicates of the tree.

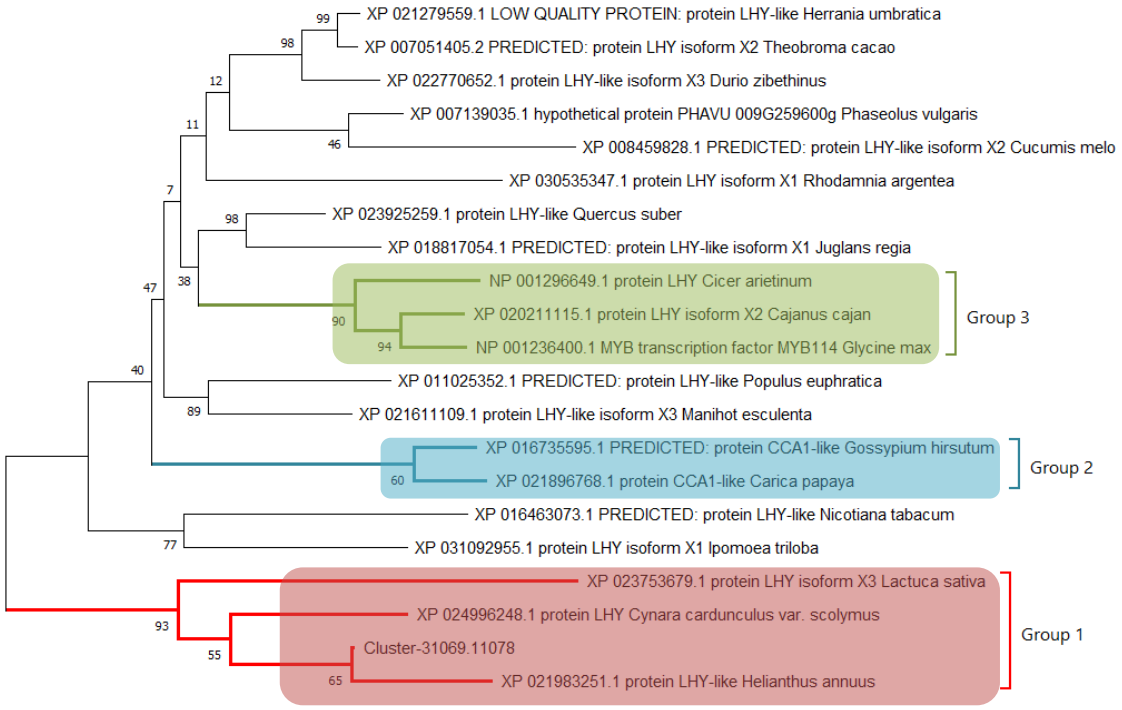

Fig. 3. Phylogenetic tree construction by the maximum likelihood tree method. Parameter: JTT model with amino acid changes per site, the branch lengths are measured in the same units as the developmental distances used to infer the evolution of the phylogenetic tree.

According to the phylogenetic tree, sub-group 1 includes *Lactuca sativa*, *Cynara cardunculus*, and *H. annuus* from the Asteraceae family, as well as putative LHY protein. It can be assumed that the putative LHY protein also comes from the same family, i.e., Asteraceae. Based on a study in 2016, it was stated that asterid species are known as ancestral genome duplications that contribute to the evolution of the Asteraceae family [27]. These species shared approximately the same ancestor, with bootstrap value higher than 90%. The rapid diversification of the Asteraceae is associated with the polyploidization in this family [27].

CCA1 and LHY have been shown to support physiological processes in plants, including plant growth and biotic and abiotic stress response [28]. Besides, it was identified that the LHY protein acts as a central player in the circadian clock of plants that controls the gene expression rhythm and also the photoperiodic induction in plants [29]. In addition, *H. annuus* depicts that *LHY/CCA1* expression can suppress the expression of PRR proteins in the morning [30].

On the other hand, sub-group 2 consists of different plant families, i.e., *Gossypium hirsutum* and *Carica papaya*, which came from the Malvaceae and Caricaceae families with 'bootstrap' values equal to 60%. Mainly, this group contained CCA1-like proteins with a conserved domain. Gossypium species appear to have divergent protein sequences due to duplication events [31]. The substitution of an amino acid leads to paralog, which appears in the linkage over a long period of time. When paralogous occurs, the composition and function of sequences or genes will be changed. In addition, families of Caricaceae undergo

duplication events in their chromosome numbers and appear to have stable genome sizes [32].

The sub-group 3 of the Fabaceae family contains the HTH (helix-turn-helix) myb-domain, which plays a role in DNA binding during the transcription process. This subgroup has a high 'bootstrap' value of more than 90%. According to the study, MYB is one of the largest gene families for plant transcription factors and has a very conservative DNA binding domain [33]. They play an important role in the regulation of hormones, organogenesis, leaf morphogenesis, and seed germination [34]. LHY with CCA1 and TOC1 located in the centre of the circadian clock is a MYB protein transcription. Additionally, the amino acid sequences of LHY and CCA1 are remarkably similar and control the circadian rhythms in plants [33].

4. CONCLUSIONS

From the analysis that has been carried out, the putative *LHY* gene of Stevia MS007 was successfully identified and characterised through an *in-silico* method. The LHY protein shows high percentages of identity with *H. annus*, *Gossypium hirsutum*, and *Cynara cardunculus* var. *scolymus*. Besides, the discovery demonstrates that the SANT domain, TEA domain, and Myb-like DNA-binding domain are present in the *LHY* gene and are important in stabilising the histone binding conformation by directly interacting with the histone tail during the transcription process. Nonetheless, based on a physicochemical property investigation, it was shown that LHY proteins are unstable proteins that cannot be easily isolated in the laboratory. The relationship between Cluster-31069.11078 of Stevia MS007 LHY and the other homologous proteins was determined using phylogenetic tree analysis. The evolutionary tree shows that the Stevia MS007 LHY protein shares a common ancestor with individuals from the Asteraceae family. *S. rebaudiana* and *H. annuus* were closely linked since the clade is well supported and the bootstrap values were greater than 50%. While they share the same plant family, they are distinct species with their own characteristics and attributes. It would be a huge benefit to be able to modify this information, in order to manufacture artificial sweeteners.

ACKNOWLEDGEMENT

We want to express our gratitude to all staff and students of the Kulliyyah of Science, IIUM, who contributed to this work.

REFERENCES

[1] Cho, NH, Shaw, JE, Karuranga, S, Huang, Y, da Rocha Fernandes, JD. Ohlrogge, AW, & Malanda, B. (2018) IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Research and Clinical Practice, 138: 271–281.

[2] Gandhi, S, Gat, Y, Arya, S, Kumar, V, Panghal, A, Kumar, A. (2018) Natural Sweeteners: Health Benefits of Stevia. Foods and Raw Materials, 6(2).

[3] Goyal, SK, Samsher RK, Goyal, RK. (2010) Stevia (Stevia rebaudiana) a bio-sweetener: a review. International Journal of Food Sciences and Nutrition, 61(1): 1–10.

[4] Chatsudthipong, V, Muanprasat, C. (2009) Stevioside and related compounds: Therapeutic benefits beyond sweetness. Pharmacology & Therapeutics, 121(1): 41–54.
[5] Kazmi, A, Khan, MA, Mohammad, S, Ali, A, Ali, H. (2019) Biotechnological production of natural calorie free steviol glycosides in Stevia rebaudiana: An update on current scenario. Current Biotechnology: 08.
[6] Samsulrizal NH, Zainuddin Z, Noh AL, Sundram TC. (2019) A review of approaches in steviol glycosides synthesis. International Journal of Life Sciences and Biotechnology, 2(3):145-57.
[7] Yadav, AK, Singh, S, Dhyani, D, Ahuja, PS. (2011) A review on the improvement of Stevia [Stevia rebaudiana (Bertoni)]. Canadian Journal of Plant Science, 91(1): 1-27.
[8] Ceunen, S, Werbrouck, S, Geuns, JMC. (2011) Accumulation pattern of steviol glycosides in Stevia rebaudiana Bertoni and its stimulation by red-light. In Proceedings of the 5 th Stevia symposium, organized by EUSTAS.
[9] Atamian, HS, Creux, NM, Brown, EA, Garner, AG, Blackman, BK, Harmer, SL. (2016) Circadian regulation of sunflower heliotropism, floral orientation and pollinator visits. Research, 353(6299): 587-590.
[10] Park, M, Kwon, Y, Gil, K, Park, C. (2016) Late Elongated Hypocotyl regulates photoperiodic flowering via the circadian clock in *Arabidopsis*. BMC Plant Biol, 16(114):1-12.
[11] Carre, IA, Kim, J. (2002). MYB transcription factors in the *Arabidopsis* circadian clock. Journal of Experimental Botany, 53(374): 1551-1557.
[12] Samsulrizal, NH, Khadzran, KS, Shaarani, SH, Noh, AL, Sundram, TC, Naim, MA, Zainuddin, Z. (2020) De novo transcriptome dataset of Stevia rebaudiana accession MS007. Data in brief, 28.
[13] Samsulrizal, NH, Khadzran, KS, Sundram, TSM, Zainuddin, Z, Shaarani, S. H. N., Azmi, NSA, Harun, S. (2021) Transcriptome profiling of Stevia rebaudiana MS007 revealed genes involved in flower development. Turkish Journal of Biology, 45(3): 314-322.
[14] Wheeler, D, Bhagwat, M. (2007) Comparative Genomics (Vols. 1-2). Humana Press Inc.
[15] Hunter, S, Apweiler, R, Attwood, T. K, Bairoch, A, Bateman, A, Binns, D, Lopez, R. (2009). InterPro: the integrative protein signature database. Nucleic Acids Research, 37: D211–D215.
[16] Kumar, S, Stecher, G, Li, M, Knyaz, C, Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution, 35(6): 1547.
[17] Kerfeld, C.A, Scott, K. M. (2011) Using BLAST to Teach "E-value-tionary" Concepts. PLoS Biol 9(2): e1001014.
[18] Basu, M. K, Poliakov, E, Rogozin, I. B. (2009) Domain mobility in proteins: functional and evolutionary implications. Briefings in Bioinformatics, 10(3): 205-216.
[19] Boyer, LA, Latek, RR, Peterson, CL. (2004) The SANT domain: a unique histone-tail-binding module? Nature reviews Molecular cell biology, 5(2): 158-163.
[20] Landin-Malt, A, Benhaddou, A, Zider, A, Flagiello, D. (2016) An evolutionary, structural and functional overview of the mammalian TEAD1 and TEAD2 transcription factors. Gene, 591(1): 292-303.
[21] Rose, A, Meier, I, Wienand, U. (1999) The tomato I-box binding factor LeMYBI is a member of a novel class of Myb-like proteins. The Plant Journal, 20(6): 641-652.
[22] Tahirov, TH, Sato, K, Ichikawa-Iwata, E, Sasaki, M, Inoue-Bungo, T, Shiina, M, Ogata, K. (2002) Mechanism of c-Myb–C/EBPβ cooperation from separated sites on a promoter. Cell, 108(1): 57-70.
[23] Krogh, A, Larsson, B, von Heijne, G, Sonnhammer, EL. (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Journal of Molecular Biology, 305(3): 567–580.
[24] Nute, M, Saleh, E, Warnow, T. (2018) Evaluating Statistical Multiple Sequence Alignment in Comparison to Other Alignment Methods on Protein Data Sets. Systematic Biologist, 68(3): 396–411.
[25] Kaghed, HN, Al-Shamery, SA, Al-Khuzaie, FEK. (2016) Multiple Sequence Alignment based on Developed Genetic Algorithm. Indian Journal of Science and Technology, 9(2).

[26] Edgar, RC. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics, 5: 113.
[27] Scaglione, D, Reyes-Chin-Wo, S, Acquadro, A, Froenicke, L, Portis, E, Beitel, C, Lanteri, S. (2016) The genome sequence of the outbreeding globe artichoke constructed de novo incorporating a phase-aware low-pass sequencing strategy of F1 progeny. Scientific Reports, 6(1): 19427.
[28] Higashi, T, Murase, H, Fukuda, H. (2013) Characterization of Circadian Rhythms Through A Bioluminescence Reporter Assay in Lactuca Sativa L. Acta Horticulturae, (1011): 459–464.
[29] Spensley, M, Kim, JY, Picot, E, Reid, J, Ott, S, Helliwell, C, Carre, IA. (2009) Evolutionarily Conserved Regulatory Motifs in the Promoter of the Arabidopsis Clock Gene Late Elongated Hypocotyl. The Plant Cell, 21(9): 2606-2623.
[30] Pokhilko, A, Mas, P, Millar, AJ. (2013). Modelling the widespread effects of TOC1 signalling on the plant circadian clock and its outputs. BMC Systems Biology, 7(23).
[31] Pfeil, BE, Brubaker, CL, Craven, LA, Crisp, MD. (2004) Paralogy and Orthology in the Malvaceae rpb2 Gene Family: Investigation of Gene Duplication in Hibiscus. Molecular Biology and Evolution, 21(7): 1428-1437.
[32] Rockinger, A, Sousa, A, Carvalho, FA, Renner, SS. (2016) Chromosome number reduction in the sister clade of Carica papaya with concomitant genome size doubling. American Journal of Botany, 103(6).
[33] Zhang, Z, Chen, J, Su, Y, Liu, H, Chen, Y, Luo, P, Zhang, H. (2015) TaLHY, a 1R-MYB Transcription Factor, Plays an Important Role in Disease Resistance against Stripe Rust Fungus and Ear Heading in Wheat. PLoS One, 10(5).
[34] Bian, S, Li, R, Xia, S, Liu, Y, Jin, D, Xie, X, Li, X. (2018) Soybean CCA1-like MYB transcription factor GmMYB133 modulates isoflavonoid biosynthesis. Biochemical and Biophysical Research Communications, 507(1-4): 324-329.
[35] Mas, P. (2005). Circadian clock signaling in Arabidopsis thaliana: from gene expression to physiology and development. The International journal of developmental biology, 49(5-6), 491-500.
[36] Niwa, Y., Ito, S., Nakamichi, N., Mizoguchi, T., Niinuma, K., Yamashino, T., & Mizuno, T. (2007). Genetic linkages of the circadian clock-associated genes, TOC1, CCA1 and LHY, in the photoperiodic control of flowering time in Arabidopsis thaliana. Plant and Cell Physiology, 48(7), 925-937.
[37] The Star Online. Available: https://www.thestar.com.my/news/nation/2020/05/29/health-ministry-nearly-one-in-five-malaysians-has-diabetes.
[38] Puchalski, B. L. (2018). Study of C-repeat Binding Factors (CBF) structural elements conferring stability under adverse conditions (Doctoral dissertation, University of Saskatchewan).