Research Article

Molecular Identification and Phylogenetic Analysis of Putative Senescence Associated Gene 21 in *Stevia rebaudiana* Accession MS007

Nur Farhana Mustafa[1], Siti Noor Eliana Mohamad Nazar[1], Zabirah Abdul Rahim[1], Nurul Hidayah Samsulrizal[1,2]

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

Abstract

Background: *Stevia rebaudiana* is a perennial semi-shrub plant which comes from the Asteraceae family, with an approximate height of around 30 cm. The leaves of Stevia are small, elliptic, and serrated, measuring 2 to 4 cm long. It has been used commercially as a natural sweetener in Japan due to the steviol glycosides (SGs) content in the leaves. The stevioside content is 300 times sweeter than sucrose. It has non-nutritive values, which is good for diabetes and obesity patients. The SGs content in Stevia can be improved by increasing light exposure (long day condition). The Senescence Associated Gene 21 (*SAG21*) gene is one of the interesting genes to be identified and discovered in Stevia. **Aims and Objectives:** The objectives of this research were to identify and characterise the *SAG21* gene using *in silico* analysis. **Materials and Methods:** These data analyses were obtained using ExPASy, blastP, InterPro, Pfam, TMHMM, ProtParam, and MEGA software. **Results:** Putative *SAG21* MS007 showed high homology with the *SAG21* gene in *Helianthus annuus* with a high percentage of identity, which was 80.90%. It also confirmed that the putative SAG21 MS007 protein contained the domain LEA_3. It was usually found in land plants and accumulated heavily in the last stage of seed formation. ProtParam analysis found that the putative *SAG21* protein was a stable globular protein. TMHMM analysis predicted that this protein is a hydrophilic protein and is located outside of transmembrane helices. **Conclusion:** The phylogenetic tree showed that the putative *SAG21* MS007 gene had a close relationship with the *SAG21* protein of *H. annuus,* with a bootstrap value of more than 70%.

Keywords: *Stevia rebaudiana*, *Helianthus annuus*, SGs and *SAG21*

Introduction

Sweetness is one of the taste sensations that allow humans to savor foods. It is highly related to sugar as it gives a sweet taste to the dishes. A lot of research has been done to find alternative resources for sweeteners aside from sugarcane and honey. Stevia has been one of the natural sweeteners that can replace sugar in our food intake. In Japan, Stevia has been widely used as a sweetener in food and beverages. Steviol glycosides (SGs) are the compounds present in Stevia which give the food a sweet taste.[1] The main components of SGs are stevioside and rebaudioside A. These components are highly soluble in water and have intensely sweet properties.[2] Researchers have found that stevioside is 300 times sweeter than sucrose.[3]

Furthermore, Anton *et al*. mentioned that Stevia contains zero calories, which allows people with obesity to control their calorie intake and at the same time taste the sweetness.[4] Roth *et al*. reported that obesity can lead to various serious diseases such as diabetes, cancer, and heart disease, which are increasing over the years and also have a rising death risk.[5] SGs are extracted from the leaves of *Stevia rebaudiana*, which have the highest content of these compounds, followed by flowers, stems, and seeds.[6] Stevia has been known as a short-day plant

Access this article online

Quick Response Code:

Website: www.matrixscipharma.org

DOI: 10.4103/mtsp.mtsp_6_22

Address for correspondence: Dr. Nurul Hidayah Samsulrizal, Department of Plant Science, Kulliyyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia.
Plant Productivity and Sustainable Resource Unit, Kulliyyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia.
E-mail: hidayahsamsulrizal@iium.edu.my

How to cite this article: Mustafa NF, Mohamad Nazar SN, Rahim ZA, Samsulrizal NH. Molecular identification and phylogenetic analysis of putative senescence associated gene 21 in *Stevia rebaudiana* accession MS007. Matrix Sci Pharma 2022;6:35-40.

Received: 23-May-2022 **Accepted:** 26-May-2022 **Available Online:** 08-Nov-2022

with 13 h of critical length, which induces the reproductive organs and limits the leaf production.[7] However, flowering in Stevia has adverse effects on the SGs as it reduces the concentration of the compounds.[8]

Ullah *et al*. stated that a long day period can delay the flowering time and increase the glycoside yield. Hence, identifying the flowering genes in Stevia is beneficial for manipulation of the flowering process.[9] The Senescence Associated Gene 21 (SAG21) gene is one of the interesting genes to be identified and discovered in Stevia. This gene was found abundant in roots and flowers of *Arabidopsis thaliana* during light regulation, dehydration, and oxidation, which functions to prevent premature ageing such as flowering and senescence and to promote root development.[10] Salleh *et al*.(2012) also reported that this gene was recognized at an early stage of leaf senescence when the leaves started to change color from green to yellow and when the plant was induced by darkness or absence of sunlight. The expression of the gene is reduced once the leaves reach total senescence.[10]

Besides, gene expression occurs at the roots, stems, and leaves.[11] It was found abundant in roots and flowers during light regulation as well as during dehydration and oxidation.[10] The expression of this gene can be induced by oxidants like hydrogen peroxide and menadione and can accumulate due to the response towards dehydration and abscisic acid.[10,11] Other than that, the expression of this gene can also be induced by other stresses such as ozone, pathogen infection, and low nitrogen.[12,13] Therefore, the flowering time in Stevia could be manipulated by identifying and characterizing the functions of the SAG21 gene. This involves comparison of this gene with other species, which can contribute to the discovery of the protein interaction in the flowering mechanism of *S. rebaudiana*.

Material and Methods

Translating nucleotide

The putative SAG21 sequence was identified based on annotation from a previous study and saved in a FASTA file. The sequence was translated into protein sequence using the ExPASy Server (the Expert Protein Analysis System) World Wide Web server (https://web.expasy.org/translate/) and the longest open reading frame was used for the next steps of analysis.[14]

Homology search

Basic local alignment search tool (BLAST) is a sequence analysis tool which is available online at (https://blast.ncbi.nlm.nih.gov/BLAST/). There are several programs in BLAST, such as blastP, blastn, blastx, and tblastn. The protein sequence was inserted into the query sequence entry in the blastP suite which was specified in the protein sequence and the database selected was reference proteins (refseq_protein).[15] All the protein sequences acquired had a percent identity in the range of 45%–85%. The percent identity represents the matches between query sequences with target sequences. All the sequences were selected and saved in FASTA format.

Domain search using InterPro and Pfam

InterPro is an integrative database that classifies protein sequences into families and can predict domain sites (http://www.ebi.ac.uk/interpro/). InterPro is integrated with Pfam databases. Pfam is available online at (http://pfam.xfam.org/), which measures protein domains and families.[16] It used multiple sequence alignment and a hidden Markov model (HMM) to measure the matches of the query sequence with the protein database. The SAG21 protein in FASTA format was submitted to InterPro and Pfam tools, respectively.[16] The results acquired from both tools were compared.

Physicochemical analysis using ProtParam and transmembrane helix prediction using TMHMM

A tool called ProtParam (Protein Parameter) is a tool that calculates the physicochemical properties of a protein. It is available online at (https://web.expasy.org/protparam/). It measures the molecular weight of a protein, theoretical pI, amino acid composition, extinction coefficient (EC), instability index (II), aliphatic index (AI), and the grand average of hydropathicity (GRAVY).[17] The protein sequence of the *SAG21* gene was submitted in FASTA format to the ProtParam tool for physicochemical analysis. Next, TMHMM Server version 2.0 is a server that predicts transmembrane helices of proteins based on a HMM which is available at (http://www.cbs.dtu.dk/services/TMHMM/). It has high accuracy and high sensitivity in the prediction of transmembrane helices.[18] The SAG21 protein of Stevia was submitted to the TMHMM Server for prediction of transmembrane protein topology.

Multiple sequence alignment

MUSCLE is a computer program used to create multiple alignments of nucleotide and protein sequences, which is available at (https://www.ebi.ac.uk/Tools/msa/muscle/). The multiple sequence alignments can be used to create phylogenetic trees and predict the close relationship of target sequences with other species.[19] Sequence alignments were performed on 21 protein sequences, including the putative SAG21 sequence, using the MUSCLE program.

Phylogenetic analysis

The SAG21 protein sequence with 20 sequences of protein was aligned using MUSCLE in MEGA software. After the alignment, the data was saved in a*.meg format file for the phylogenetic tree. The tree was constructed by selecting the maximum likelihood method with 1000 numbers of bootstrap replication. The bootstrap method was necessary to ensure the reliability of the estimated taxa.[20] The target protein's relationship to other species was clearly conveyed and presented in a phylogenetic tree.

Results and Discussion

The objectives of this study were to identify and characterize the *SAG21* gene in Stevia by using some bioinformatics tools.

The transcriptome database from prior work was used to further analyze the gene (Cluster-31069.36919) in order to identify and define it.[14]

Homology search using basic local alignment search tool

The protein sequence obtained from the ExPASy program was then submitted to the blastP server to search for all possible homology sequences of the putative SAG21 gene. Based on the percentage of identity and E value (expected value), 20 protein sequences were chosen from the blastP results. Based on the result, all the protein sequences contain E values in the range of $9e^{-23}$ to 8e-48, with a percentage identity of around 45%–80%. Lobo (2008) reported that high percentage identity and low E value act as parameters to measure the most matched sequences between query and target sequences.[15] Table 1 showed the E value and the percentage identity of each species related to the putative SAG21 protein. It indicated that these sequences had a homolog relationship with the putative SAG21 protein. The putative SAG21 protein of Stevia had high homology with the SAG21 gene in *Helianthus annuus*. It had a high percentage identity (80.90%) and a low E value (8e-48). Based on Figure 1, the result illustrated the conserved region of the putative SAG21 sequence when compared with 20 other sequences. The alignment scores of the sequences were 80–200. The specific hit of the putative SAG21 protein was LEA_3 and came from the LEA_3 superfamily.

Domain search by using InterPro and Pfam

The data from the putative SAG21 protein sequence was used to obtain analysis at InterPro and the Pfam Server. It showed that most of the regions in the gene were identified as late embryogenesis abundant proteins. The LEA_3 subgroup (position 2–90 amino acids) was hydrophilic and usually found in land plants.[21] It was highly accumulated during natural desiccation of seed tissue during the final stage of seed formation.[22] It is also expressed during water deficit periods in vegetative organs.[23] According to Candat *et al*., the subcellular localization of SAG21 protein in *A. thaliana* was predicted to be at the chloroplast and expressed in all parts of the plant.[24]

Table 1: Top hit from homology search using BlastP in National Center for Biotechnology Information database

Description	*E*-Value	Percentage identification
Protein *SAG21*, mitochondrial-like (*Helianthus annuus)*	8e-48	80.90
Protein *SAG21*, mitochondrial-like (*Lactuca sativa)*	5e-45	78.26
Protein *SAG21*, mitochondrial-like (*Cynara cardunculus var. scolymus)*	1e-44	77.42
Protein *SAG21*, mitochondrial (*Camellia sinensis)*	5e-29	53.26
Protein *SAG21*, mitochondrial-like (*Olea europaea var. sylvestris)*	1e-27	56.99
Protein *SAG21*, mitochondrial-like (*Ipomoea triloba)*	2e-26	54.84
Protein *SAG21*, mitochondrial-like (*Solanum pennellii)*	2e-26	57.30

SAG21: Senescence Associated Gene 21

Physicochemical analysis using ProtParam and transmembrane helix prediction using TMHMM

The physicochemical analysis of the putative SAG21 protein was carried out using ProtParam. It measured protein stability and the nature of the protein. The putative SAG21 protein contained 92 amino acids and had a theoretical isoelectric point of 9.91. Enany mentioned that pI measured the pH of particular molecules or surfaces that carried no net electrical charge.[25] This measurement was useful for understanding the stability of protein charges. The EC refers to the quantity of light absorbed by the protein at certain wavelengths. The EC of SAG21 protein [Table 2] was $8480M^{-1}\ cm^{-1}$, at 280 nm measured in water. Besides, the II indicated protein stability. Gasteiger *et al*. stated an II of more than 40 was considered not stable. The II of the SAG21 protein was 39.82, which is classified as stable.[17]

GRAVY for a protein was calculated as the sum of hydropathy values for all amino acids over a number of residues in a sequence. The putative SAG21 protein had a GRAVY value of −0.386, indicating that it was a globular protein (hydrophilic). The negative values of GRAVY were predicted as globular proteins (hydrophilic) while positive values were membranous proteins (hydrophobic).[25,26] The AI was 73.15 which is defined as the relative volume occupied by side protein chains (alanine, valine, and leucine).

Besides, transmembrane helix in the putative SAG21 protein sequence of Stevia were predicted using TMHMM 2.0 which showed the prediction of the most probable location and orientation of transmembrane helices in the sequence. Krogh *et al*. stated that hydrophilic regions of proteins are located outside or inside of the membrane. The TMHMM 2.0 analyses revealed that the entire putative SAG21 protein sequence was labeled as "outside," which is defined as a hydrophilic protein with no membrane helices.[18] The putative SAG21 protein was expected to be a hydrophilic and stable globular protein with no membrane helices in general.

Table 2: The analysis result of physico-analysis properties of putative Senescence Associated Gene 21 protein

Description	Value
Number of amino acids	92
Molecular weight	10053.61
Theoretical pI	9.91
II	39.82
AI	73.15
GRAVY	−0.386

pI: Isoelectric point, II: Instability index, AI: Aliphatic index, GRAVY: Grad average of hydropathicity

Phylogenetic tree construction

A total of 21 sequences, including a putative SAG21 protein, were submitted to the MUSCLE program for the global alignment [Figure 1]. This analysis indicated that Stevia had a high relationship with XP_022010509.1 protein SENESCENCEASSOCIATED GENE 21, mitochondrial-like (*H. annuus*), where most of them have the same conserved region and most of the conserved sequences exist around the domain coding area.

Figure 1: CLUSTAL multiple sequence alignment by MUSCLE (3.8). The alignments of putative SAG21 sequence with 20 protein sequence. An asterisk (*) indicated a single, fully conserved residue positions; a colon (:) indicated conservation between strongly similar properties groups which scored above 0.5 in Gonnet PAM 250 matrix; a period (.) indicated conservation between weakly similar properties groups which scored ≤0.5 and >0 in the Gonnet PAM 250 matrix

The phylogenetic tree [Figure 2] was divided into four groups. As for Group 1, all of the plant species come from the family Asteraceae. The species in Group 1 were *S. rebaudiana*, *H. annuus*, *Lactuca sativa*, and *Cynara cardunculus* var. *scolymus*. These species were predicted to have close relations with Stevia compared to the other 17 species. The putative *SAG21* gene in *S. rebaudiana* was more relatable to the *SAG21* gene in *H. annuus* compared to the other species, with 80% of the bootstrap values. According to Hall (2013), the bootstrap percentage must be estimated more than 70% to be considered a reliable taxon.[20] According to Dwivedi and Sharma, *H. annuus* is an erect annual plant that grows to a height of 1–3 m.[27]

The root of *H. annuus* is tap rooted at an early age of this plant. As the sunflower plant matured, the roots became fibrous with lateral root spread. When comparing Stevia and sunflower, both of these plants came from the same family, upright stems and also tap rooted plants. *L. sativa* or also known as lettuce is an annual glabrous herb and has a thin tap root with spirally arranged leaves which formed a dense rosette.[28] This species belongs to the leafy vegetable family. Moreover, *C. cardunculus* var. *scolymus* is known as vegetables crops because of its fleshy stem and leaf stalks.[29] Hence, *L. sativa* and *C. cardunculus* var. *scolymus* was classified as a vegetables crop and had low relationship with Stevia.

Besides, the plant organisms in Group 4 were belonged to Brassicaceae family which were *Arabidopsis lyrata* subsp. *Lyrata*, *Eutrema salsugineum*, *Brassica rapa*, *Brassica napus* and *Capsella rubella*. All of these species had a distant relationship with Stevia. *A. lyrata subsp. Lyrata* and *E. salsugineum* were closely related to *A. thaliana*.[30] The genome size of *A. lyrata* is 50% bigger than *A. thaliana*.[31] *E. salsugineum* which can also be known as *Arabidopsis salsuginea* had a small nuclear genome.[32] Moreover, the *SAG21* gene of *B. rapa* was relatable to the *SAG21* gene in *B. napus* with 100% bootstrap value. The seeds of *B. rapa* produced more oil content than *B. napus*.[33]

Meanwhile, plant species in Group 2 and Group 3 had different classification of families except *Solanum lycopersicum* and

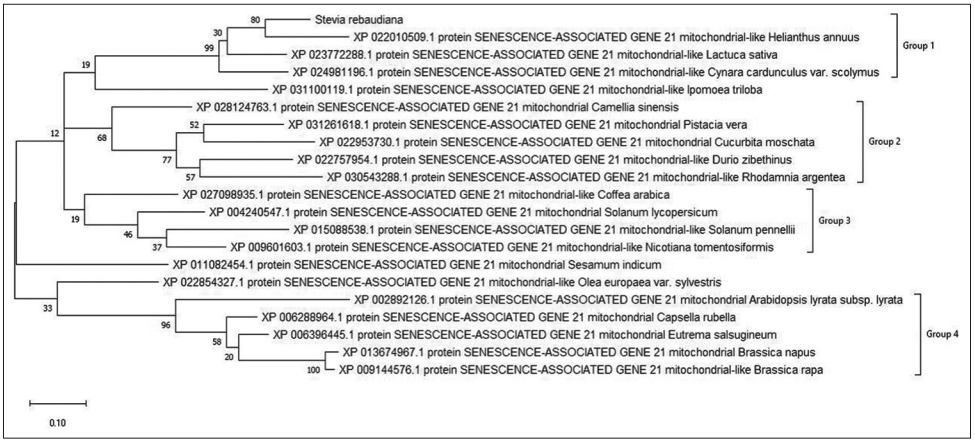

Figure 2: The original phylogenetic tree of 21 protein sequences including putative SAG21 protein by maximum likelihood method using MEGA software

Solanum pennellii. Both of these species come from the same family, Solanaceae. *S. lycopersicum* is an edible vegetable plant while *S. pennellii* is a wild tomato which can withstand stress environments.[34] Other than that, the examples of plant species in Group 2 were *Pistacia vera* and *Cucurbita moschata*. *P. vera* is known as pistachio belongs to Anacardiaceae family which is a deciduous herbaceous plant with height up to 10 m tall. *C. moschata* or its common name is pumpkin comes from the Cucurbitaceae family with long crawling and climbing stems.[35] In short, Group 2 and Group 3 were not relatable to Stevia as the family of these species were different.

Apart from that, the analysis of the phylogenetic tree showed the relationship of orthologs and paralogs between 21 species. Most of the species that had orthologous relationships came from the same ancestor as Stevia. One of the orthologous relationships can be seen in Group 1. *H. annuus* and Stevia were two different species from the Asteraceae family and had a high relationship between SAG21 proteins. It had the same protein domain as the late embryogenesis abundant_3 protein. The *SAG21* protein exists as a result of a speciation event in the Stevia and *H. annuus* genomes. Fang *et al*. stated that orthologs are homologous genes related by a common ancestor with the same function in different species.[36] It evolved from speciation events, while paralogs are homologous genes that come from duplication events that had different functions coded for the same protein.[37] Paralogs relationship can be found in Group 4 between similar genus Brassica (*B. rapa* and *B. napus*). These species had a distant relationship with the putative SAG21 protein in Stevia. It was predicted that the SAG21 protein in *Brassica* sp. was unrelated to Stevia. This protein probably existed as a result of a duplication event.[37]

Conclusion

This project aims to identify and characterize putative *SAG21* gene from *S. rebaudiana* MS007 using bioinformatics approaches. The results from these analyses showed that the putative *SAG21* gene from Stevia MS007 contains a conserved region that was identified as a late embryogenesis abundant protein, LEA_3 group. It prevents premature ageing and promotes root development. From the previous study, this gene was identified during the last stage of seed formation and water deficit in vegetative organs. The SAG21 protein was predicted to be a stable globular protein and located outside of transmembrane helices. Phylogenetic analysis was done to confirm the function of our putative *SAG21* gene from *S. rebaudiana* MS007 based on evolutionary relationships. From this analysis, it showed a very high relationship with *SAG21* genes from *H. annuus*, with 80% of the bootstrap values.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Ashwell M. Stevia, nature's zero-calorie sustainable sweetener: A new player in the fight against obesity. Nutr Today 2015;50:129-34.
2. Samsulrizal NH, Zainuddin Z, Noh AL, Sundram TC. A review of approaches in steviol glycosides synthesis. Int J Life Sci Biotechnol 2019;2:145-57.
3. Momtazi-Borojeni AA, Esmaeili SA, Abdollahi E, Sahebkar A. A review on the pharmacology and toxicology of steviol glycosides extracted from *Stevia rebaudiana*. Curr Pharm Des 2017;23:1616-22.
4. Anton SD, Martin CK, Han H, Coulon S, Cefalu WT, Geiselman P, *et al.* Effects of stevia, aspartame, and sucrose on food intake, satiety, and postprandial glucose and insulin levels. Appetite 2010;55:37-43.
5. Roth J, Qiang X, Marbán SL, Redelt H, Lowell BC. The obesity pandemic: Where have we been and where are we going? Obes Res 2004;12 Suppl 2:88S-101S.
6. Guerrero AB, San Emeterio L, Domeño I, Irigoyen I, Muro J. Steviol glycoside content dynamics during the growth cycle of *Stevia rebaudiana* Bert. Am J Plant Sci 2018;9:892.
7. Serfaty M, Ibdah M, Fischer R, Chaimovitsh D, Saranga Y, Dudai N. Dynamics of yield components and stevioside production in *Stevia rebaudiana* grown under different planting times, plant stands and harvest regime. Ind Crops Prod 2013;50:731-6.
8. Tavarini S, Passera B, Angelini LG. Ch. 1: Crop and Steviol Glycoside Improvement in Stevia by Breeding , in Steviol Glycosides: Cultivation, Processing, Analysis and Applications in Food; 2018. p. 1-31 DOI: 10.1039/9781788010559-00001. eISBN: 978-1-78801-055-9
9. Ullah A, Munir S, Mabkhot Y, Badshah SL. Bioactivity profile of the diterpene isosteviol and its derivatives. Molecules 2019;24:678.
10. Salleh FM, Evans K, Goodall B, Machin H, Mowla SB, Mur LA, *et al.* A novel function for a redox-related LEA protein (SAG21/AtLEA5) in root development and biotic stress responses. Plant Cell Environ 2012;35:418-29.
11. Mowla SB, Cuypers A, Driscoll SP, Kiddle G, Thomson J, Foyer CH, *et al.* Yeast complementation reveals a role for an *Arabidopsis thaliana* late embryogenesis abundant (LEA)-like protein in oxidative stress tolerance. Plant J 2006;48:743-56.
12. Samsulrizal NH, Khadzran KS, Meenakshi Sundram TC, Zainuddin Z, Shaarani SH, Azmi NS, *et al.* Transcriptome profiling of *Stevia rebaudiana* MS007 revealed genes involved in flower development. Turk J Biol 2021;45:314-22.
13. Tang X, Wang H, Chu L, Shao H. KvLEA, a new isolated late embryogenesis abundant protein gene from *Kosteletzkya virginica* responding to multiabiotic stresses. Biomed Res Int 2016;2016:9823697.
14. Samsulrizal NH, Khadzran KS, Shaarani SH, Noh AL, Sundram TC, Naim MA, *et al. De novo* transcriptome dataset of *Stevia rebaudiana* accession MS007. Data Brief 2020;28:104811.
15. Stephen AF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology 1990;215:403-10.
16. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, *et al.* Pfam: The protein families database. Nucleic Acids Res 2014;42:D222-30.
17. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The Proteomics Protocols Handbook; 2005. p. 571-607.
18. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 2001;305:567-80.
19. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792-7.
20. Hall BG. Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol 2013;30:1229-35.
21. Garay-Arroyo A, Colmenero-Flores JM, Garciarrubio A, Covarrubias AA. Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. J Biol Chem 2000;275:5668-74.
22. Hundertmark M, Hincha DK. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics 2008;9:118.
23. Hand SC, Menze MA, Toner M, Boswell L, Moore D. LEA proteins during

water stress: Not just for plants anymore. Annu Rev Physiol 2011;73:115-34.
24. Candat A, Paszkiewicz G, Neveu M, Gautier R, Logan DC, Avelange-Macherel MH, *et al.* The ubiquitous distribution of late embryogenesis abundant proteins across cell compartments in *Arabidopsis* offers tailored protection against abiotic stress. Plant Cell 2014;26:3148-66.
25. Enany S. Structural and functional analysis of hypothetical and conserved proteins of *Clostridium tetani*. J Infect Public Health 2014;7:296-307.
26. Chang KY, Yang JR. Analysis and prediction of highly effective antiviral peptides based on random forests. PLoS One 2013;8:e70166.
27. Dwivedi A, Sharma GN. A review on *Heliotropism* plant: *Helianthus annuus* L. J Pharmacol 2014;3:149-55.
28. Křístková E, Doležalová I, Lebeda A, Vinter V, Novotná A. Description of morphological characters of lettuce (*Lactuca sativa* L.) genetic resources. Hortic Sci 2008;35:113-29.
29. Falistocco E. Cytogenetic characterization of cultivated globe artichoke (*Cynara cardunculus* var. scolymus) and cardoon (*C. cardunculus* var. altilis). Caryologia 2016;69:1-4.
30. Schmickl R, Jørgensen MH, Brysting AK, Koch MA. The evolutionary history of the *Arabidopsis lyrata* complex: A hybrid in the amphi-Beringian area closes a large distribution gap and builds up a genetic barrier. BMC Evol Biol 2010;10:98.
31. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, *et al.* The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet 2011;43:476-81.
32. Li Y, Sun W, Liu F, Cheng J, Zhang X, Zhang H, *et al.* Methods for grafting *Arabidopsis thaliana* and *Eutrema salsugineum*. Plant Methods 2019;15:93.
33. Cartea E, De Haro-Bailón A, Padilla G, Obregón-Cano S, Del Rio-Celestino M, Ordás A. Seed oil quality of *Brassica napus* and *Brassica rapa* germplasm from northwestern Spain. Foods 2019;8:292.
34. Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, *et al.* The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nat Genet 2014;46:1034-8.
35. Suresh S, Sisodia SS. Phytochemical and pharmacological aspects of *Cucurbita moschata* and *Moringa oleifera*. Pharm Biosci J 2018;6:45-53.
36. Fang G, Bhardwaj N, Robilotto R, Gerstein MB. Getting started in gene orthology and functional analysis. PLoS Comput Biol 2010;6:e1000703.
37. Jensen RA. Orthologs and paralogs-we need to get it right. Genome Biol 2001;2:1-3.