



## **INFORME FINAL DE INVESTIGACIÓN**

# **“CARACTERIZACIÓN BIOINFORMÁTICA DE LA PROTEÍNA SAG29 DE *Arabidopsis thaliana* INVOLUCRADA EN LA RESPUESTA AL ESTRÉS SALINO”**

**LIMA – PERÚ**

**2015**

## Resumen

**Objetivo:** Realizar el análisis bioinformático de la proteína SAG29 de *Arabidopsis thaliana* involucrada en la respuesta al estrés salino. **Material y métodos:** Se realizó la búsqueda de proteínas relacionadas al estrés salino en plantas empleando la base de datos del GenBank. A partir de dicha información, se obtuvo la secuencia aminoacídica de la proteína SAG29 de *A. thaliana* (código de acceso AED91859.1). Dicha secuencia fue utilizada para determinar sus principales parámetros bioquímicos, dominios conservados entre proteínas cercanamente relacionadas, predicción de estructuras secundarias y modelamiento tridimensional, empleando las herramientas bioinformáticas ProtParam, Prosite, PHD Secondary Structure Prediction, SWISS-MODEL y Phyre2, respectivamente. La visualización de modelos tridimensionales se realizó empleando PyMol. **Resultados:** La proteína SAG29 posee 292 aminoácidos y un peso molecular de 32.9 kDa, así como dominios conservados correspondientes a proteínas involucradas en el transporte de azúcares a través de la membrana plasmática de células vegetales. Además, de acuerdo con la predicción de su estructura tridimensional, resulta evidente que SAG29 es una proteína transmembrana con siete dominios formados por hélices  $\alpha$  y que poseen poca similitud estructural con otras proteínas reportadas. De la red de interacción de proteínas, se tomó un score superior o igual 0.642 y se determinó la relación filogenética entre la proteína de membrana SAG29 con LTP4 y DCP1 con identidades de homología de 61.5% y 40.5% respectivamente. **Conclusión:** La proteína SAG29 de *A. thaliana* es una molécula básica de mediano peso molecular relacionada con las proteínas de la familia MtN3\_slv. Además resultan importantes en la regulación de la viabilidad celular y respuesta al estrés salino.

**Palabras claves:** *Arabidopsis thaliana*, SAG29, estrés salino, modelamiento

## Abstract

**Objective:** To perform a bioinformatics analysis of SAG29 protein from *Arabidopsis thaliana* which is involved in salt stress response. **Material and methods:** The analysis began using sequences from proteins related to salt stress that were obtained from GenBank. In this particular case, the sequence of SAG29 protein from *A. thaliana* was obtained (accession number AED91859.1). This primary sequence was used to the determination of main biochemical parameters, conserved domains between closely related proteins, secondary structure prediction and tridimensional modelling, by means of ProtParam, Prosite, PHD Secondary Structure Prediction, SWISS-MODEL and Phyre2 server bioinformatics tools. Visualization of tridimensional models was performed under PyMol. **Results:** SAG29 protein has 292 amino acids and a molecular weight of 32.9 kDa, and shows conserved domains related to proteins involved in sugar transport across plasma membrane from plant cells. According to the prediction of tridimensional structure, it is evident that SAG29 is a transmembrane protein with seven domains formed by  $\alpha$ -helix and has low structural similarity with other reported proteins. **Conclusion:** SAG29 protein from *A. thaliana* is a basic molecule with medium molecular weight that is related to different members of MtN3\_slv protein family. This protein is also important in the regulation of cell viability and response to salt stress.

**Keywords:** *Arabidopsis thaliana*, SAG29, salt stress, modelling

## Introducción

Las plantas están frecuentemente expuestas a diversas condiciones bióticas y abióticas adversas durante su ciclo de vida. Debido a esto, dichos organismos han desarrollado mecanismos de defensa versátiles para contrarrestar los efectos del estrés ambiental, los cuales están íntimamente interconectados con principales eventos tales como la germinación de la semilla, el crecimiento de las plántulas, la transición floral y la senescencia. Por esto, la modulación del crecimiento y los procesos de desarrollo constituye una estrategia adaptativa para tolerar las condiciones de estrés (1, 2).

La senescencia vegetal es una etapa del desarrollo altamente regulada que participa en la redistribución de metabolitos y la maduración reproductiva, dando lugar posteriormente a la desintegración celular programada del organismo. Este proceso es regulado coordinadamente a través de redes reguladoras genéticas, entre las cuales se han identificado genes asociados a la senescencia (del inglés, SAG) (3). Asimismo, sus productos proteicos han sido identificados y caracterizados funcionalmente en diversas especies vegetales (4). A partir de estos estudios, se ha determinado que dichos componentes juegan un rol importante en la degradación de macromoléculas, reciclaje y transporte de nutrientes, metabolismo de azúcares, detoxificación de metabolitos oxidativos, y el establecimiento de la tolerancia al estrés biótico y abiótico, lo cual indica una comunicación estrecha entre los procesos de senescencia y respuesta al estrés (4).

Dentro de las principales proteínas SAG, destacan las proteínas MtN3/saliva/SWEET, las cuales poseen siete motivos potenciales transmembrana y participan activamente en los procesos de desarrollo embrionario y activación de genes (5). En plantas, las proteínas MtN3/saliva/SWEET están involucradas en el desarrollo de tejidos reproductivos y la adaptación ambiental. Es claro que las proteínas MtN3/saliva están involucradas en diversas etapas del desarrollo vegetal, así como en la senescencia y la respuesta a diversos tipos de estrés biótico y abiótico, sobre todo estrés salino. En el caso particular de la proteína SAG29 de *Arabidopsis thaliana*, esta molécula es una proteína de membrana plasmática con una organización similar a las proteínas MtN3/saliva/SWEET (6), además de estar vinculada con la regulación del estrés osmótico, alta

salinidad, baja temperatura y sequía.

Por este motivo, en el presente trabajo el análisis bioinformático de la proteína SAG29 de *Arabidopsis thaliana* involucrada con la respuesta al estrés salino a fin de explorar los principales mecanismos moleculares subyacentes a su función.

## **Materiales y métodos**

### *Búsqueda de secuencias proteicas en bases de datos*

Las secuencias proteicas fueron obtenidas a partir de la búsqueda de proteínas relacionadas al estrés salino en *Arabidopsis thaliana* en la base de datos del GenBank y del UniProt. En este sentido se analizó la secuencia de la proteína SAG29 (senescence-associated protein 29) (código de acceso: AED91859.1) proveniente del estudio de secuencia y análisis del cromosoma 5 de dicha planta (7). Dicha secuencia fue utilizada para los diversos ensayos de predicción bioinformática.

### *Determinación de parámetros bioquímicos y dominios conservados*

Para determinar los principales parámetros bioquímicos de la proteína SAG29 se empleó la herramienta ProtParam del ExPASy (<http://web.expasy.org/protparam/>) (8), mientras que los dominios conservados presentes en dicha proteína fueron analizados mediante la herramienta Prosite del ExPASy (<http://prosite.expasy.org/>) (9). La búsqueda y obtención de secuencias relacionadas fue realizada empleando la herramienta BLAST del NCBI (10).

### *Predicción de estructuras secundarias y alineamiento estructural*

El análisis de predicción de estructuras secundarias presentes en la proteína SAG29 fue realizada empleando la herramienta PHD Secondary Structure Prediction Method (<http://npsa-prabi.ibcp.fr>) (11). Además, el alineamiento estructural se realizó en la plataforma Cn3D (versión 4.3.1) utilizando como molde diversas proteínas estructuralmente similares (12).

### *Modelamiento tridimensional de la proteína SAG29*

El modelo tridimensional de la proteína SAG29 de *Arabidopsis thaliana* se obtuvo empleando la herramienta SWISS-MODEL (<http://swissmodel.expasy.org>) (13), a partir de la información depositada en la base de datos del Protein Data Bank (PDB). Opcionalmente se empleó la herramienta Phyre2 Protein Homology/analogy Recognition Engine (<http://www.sbg.bio.ic.ac.uk/phyre2/>) (14) para la búsqueda de patrones de plegamiento conservados y modelamiento por *threading*. La visualización de los modelos obtenidos se realizó empleando el programa PyMol Molecular Graphics System (versión 1.3) (Schrödinger,

LLC).

*Predicción de la interacción proteica*

La red de interacción de proteínas se obtuvo usando el programa STRING (15)

## Resultados

La proteína SAG29 posee 292 aminoácidos con un peso molecular de 32.9 kDa. Así mismo, SAG29 tiene un punto isoeléctrico de 8.20 y una proporción de 20 aminoácidos ácidos versus 22 aminoácidos básicos. Empleando la herramienta BLAST se encontró que la proteína en estudio presenta porcentajes de similitud del 94% con la proteína nodulina de *Arabidopsis lyrata* (XP\_002873602.1) y 91% con el transportador de azúcares SWEET15 de *Camelina sativa* (XP\_010453373.1). Además se logró determinar la posición de dos regiones conservadas comprendidas entre los aminoácidos 12-99 (E-value: 3.36e-25) y 134-218 (E-value: 6.34e-27) correspondientes al dominio MtN3\_slv (Pfam: 03083) (Fig. 1). Posteriormente se realizó la predicción de estructuras secundarias, encontrándose que la proteína SAG29 poseen 38.7% de hélices  $\alpha$ , 31.5% de láminas  $\beta$ , y 29.8% de loop internos (Fig. 2). Finalmente, el modelamiento de la proteína en estudio mostró la presencia de un core compuesto de forma predominante por siete  $\alpha$ -hélices transmembrana (Fig. 3).

De la red de patrones de interacción proteica, con fines prácticos se tomó un score superior o igual 0.642 solo de aquellas proteínas que tiene una estructura resuelta tridimensionalmente y se determinó con una identidad del 61.5% y 40.5%, la relación filogenética entre la proteína de membrana SAG29 con LTP4 y DCP1 respectivamente; así como también se obtuvo la coexpresión de SAG29 y LTP4 en nuestro organismo *Arabidopsis thaliana* (Fig. 4).



```

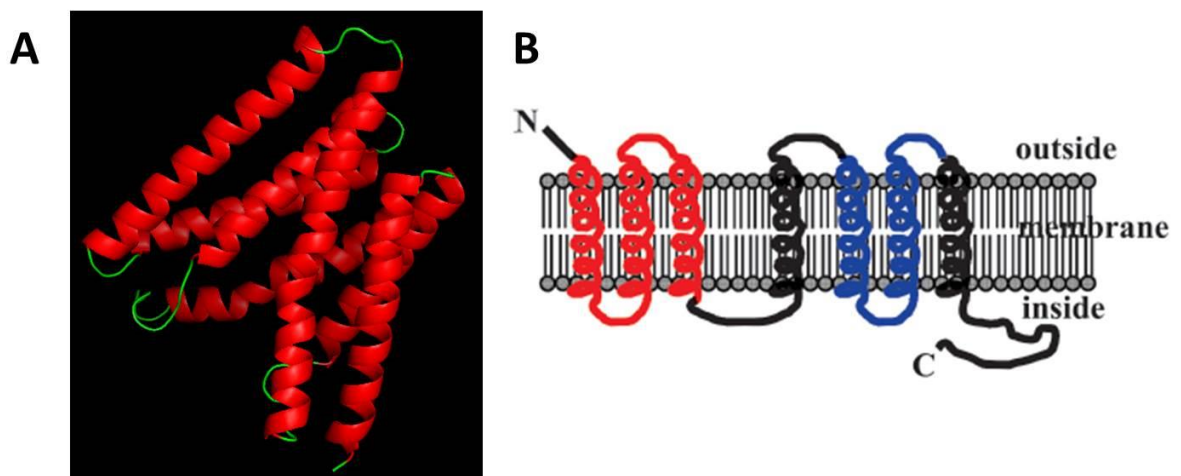
1  MGVMINHHFL AFIFGILGNV I SFLVFLAPV PTFYRIYKRK STESFQSLPY QVSLFSCMLW
61 LYYALIKKDA FLLITINSFG CVVETLYIAM FFAYATREKR ISAMKLFIAM NVAFFSLILM
121 VTHFVVKTPP LQV SVLGWIC VAISVSVFAA PLMIVARVIK TKSVEYMPFT LSFFLTISAV
181 MWFAYGLFLN DICIAIPNVV GFVLGLLQMV LYLVYRNSNE KPEKINSSEQ QLKSIVVMSP
241 LGVSEVHPVV TESVDPLSEA VHEDLSKVT KVEEPISENG KCYVEATRPE TV 292

```

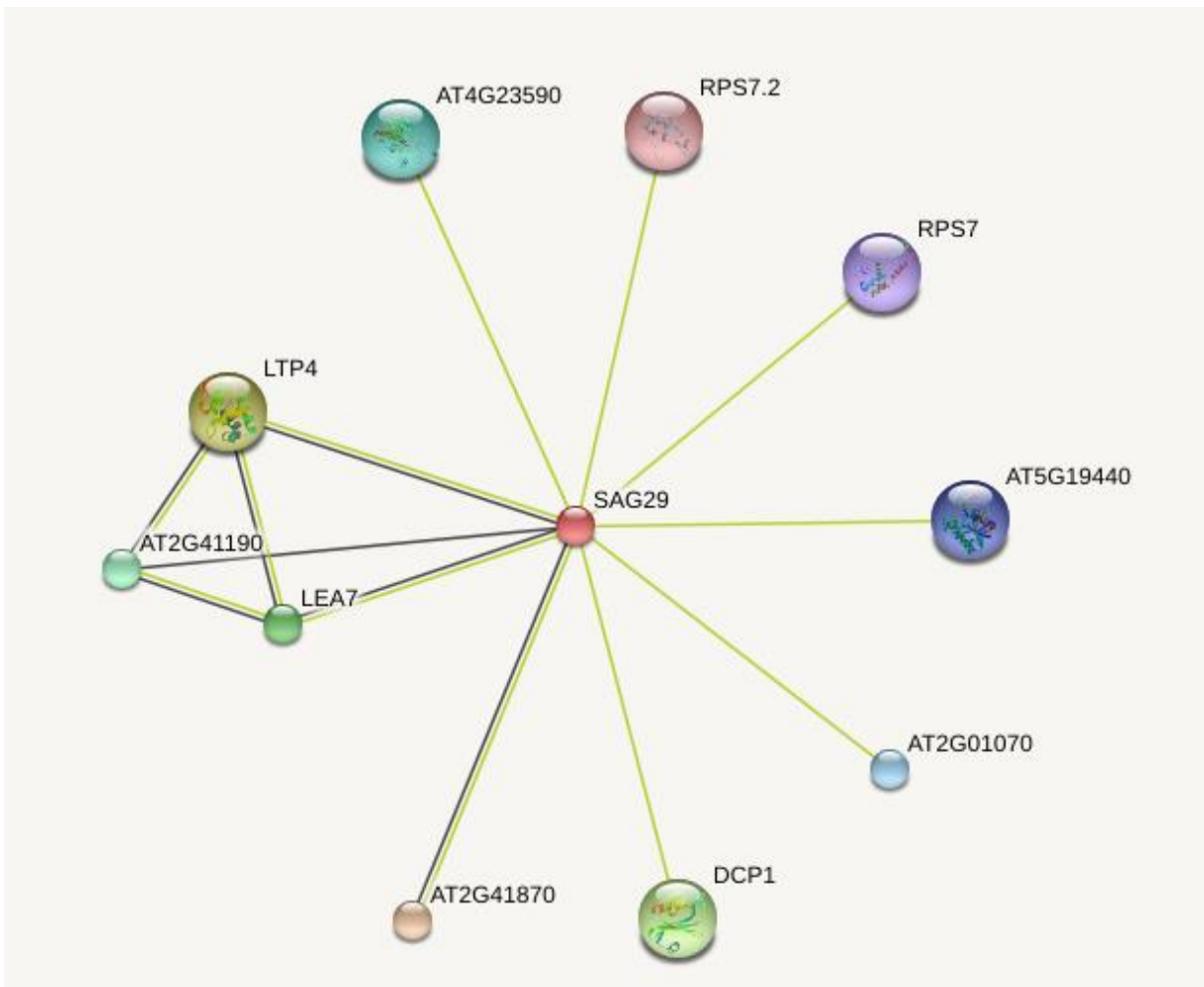
**Fig. 1** Distribución de los dominios conservados MtN3\_slv de la proteína SAG29 de *A. thaliana*. Se resaltan los dominios conservados comprendidos entre los aminoácidos 12-99 y 134-218 a lo largo de la secuencia primaria de la proteína.

10            20            30            40            50            60            70  
 |            |            |            |            |            |            |  
 MGVMINHHFLAFIFGILGNVISFLVFLAPVPTFYRIYKRKSTESFQSLPYQVSLFSCMLWLYYALIKKDA  
 CCeeeHHHHHHHHHHHHHHHhheeEeCCCceEEEEcCCCccccCCcCEeHHHHHHHHHHHHHhcc  
 FLLITINSFGCVVETLYIAMFFAYATREKRISAMKLFIAMNVAFFSLILMVTHFVVKTPPLQVSVLQWIC  
 eEEEEcCCchHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHhEEEEEEcCCCEEEEEEE  
 VAISVSVFAAPLMIVARVIKTKSVEYMPFTLSFFLTISAVMWFAYGLFLNDICIAIPNVVGFVLGLLQMV  
 EEEEEEcchHhHhHhheecceEEEEcCEeEehhhHHHHHHHHHHHhEEEEcChHHHHHHHHH  
 LYLVYRNSNEKPEKINSSEQQLKSIVVMSPLGVSEVHPVVTESVDPLSEAVHHEDLSKVTKVEEESIENG  
 HHheeEcCCCCcCCChHHhhceEEEEcCCceeeeEEeEEeCCcccccccccccccecccCCCCC  
 KCYVEATRPETV  
 ceEEEEcCCCC

**Fig. 2.** Predicción de estructura secundarias para la proteína SAG29 de *A. thaliana*. Se observan las regiones correspondientes a las hélices  $\alpha$  (Hh), láminas  $\beta$  (Ee) y loops internos (Cc) distribuidos a lo largo de la secuencia de aminoácidos de la proteína.



**Fig. 3.** (A) Modelo tridimensional de la proteína SAG29 de *A. thaliana* empleando modelamiento por *threading*. Las  $\alpha$  hélices se muestran de color rojo mientras que los loops internos se muestran de color verde. (B) Representación esquemática de la distribución de  $\alpha$  hélices transmembrana para la proteína en estudio (5)



**Fig. 4.** Red de patrón de interacción de la proteína SAG29 de *Arabidopsis thaliana* con las proteínas LTP4 y DCP1 con identidades de homología de 61.5% y 40.5% respectivamente.

## Discusión

En el presente artículo se ha realizado el análisis bioinformático de la proteína SAG29 de *Arabidopsis thaliana* involucrada en la respuesta al estrés salino. Del análisis de la estructura primaria y búsqueda de dominios conservados, se observó la presencia de dos regiones comprendidas entre los aminoácidos 12-99 y 134-218. Estas porciones corresponden a las proteínas involucradas en el transporte de azúcares a través de la membrana plasmática de las células vegetales (17). Si bien no se conoce al detalle la función molecular de esta familia de proteínas, se ha determinado su importancia en el desarrollo de las plantas, en la viabilidad del polen, y en la susceptibilidad a patógenos. En el caso particular de *A. thaliana*, SAG29 sería clave en la regulación de la viabilidad celular mediante la modulación de la integridad de la membrana durante la senescencia vegetal y la respuesta al estrés (6). Con respecto a la predicción de la estructura tridimensional de esta proteína, la búsqueda en bases de datos de estructuras proteicas sólo permitió obtener proteínas con una identidad de hasta 37%. Eso se debió principalmente a diferencias en la porción C-terminal de la secuencia de aminoácidos. Por esta razón, se utilizó la predicción de estructuras utilizando el servidor Phyre2, el cual realiza la búsqueda de proteína en base a la conservación de patrones de plegamiento (14). Para este fin se empleó como plantilla la estructura homotrimérica del transportador SWEET procedente de *Oryza sativa* (resolución: 3.69 Å) (17). A partir del análisis realizado, se determinó el modelo tridimensional de la proteína SAG29, el cual está compuesto de 7 hélices  $\alpha$  con posiciones alternadas de loops internos a lo largo de toda la estructura de la proteína. Dichas estructuras fueron predecidas inicialmente y muestra la congruencia entre los resultados obtenidos por herramientas complementarias.

La predicción de patrones de redes de interacción permite visualizar y analizar de manera precisa la relación filogenética, co-expresión de genes y proteínas relacionadas en algunas u otras especies. En nuestro resultado solo se observa la co-expresión de SAG29 y LTP4 en nuestro organismo *Arabidopsis thaliana* mas no de los genes ortólogos que son transferidos por homología en otros organismo, esta expresión conjunta revela la interacción entre ellas. Así también se observó algunos reportes experimentales de las proteínas LTP4 y DCP1 con un alto grado de score. Además existe una semejanza funcional entre estas proteínas, ya que todas

están relacionadas de manera directa o indirecta por diferentes mecanismos que le permiten la tolerancia al estrés salino (18).

## Conclusiones

La proteína SAG29 de *A. thaliana* es una molécula básica de mediano peso molecular con una alta similitud y conservación de secuencia con proteínas de la familia MtN3\_slv. Así mismo, esta proteína resultaría ser clave en la regulación de la viabilidad celular y respuesta al estrés salino. Se requieren de estudios complementarios para dilucidar su mecanismo de acción y distribución en la membrana plasmática de las células vegetales.

## Literatura citada

1. Ward JM, Hirschi KD, Sze H. Plants pass the salt. *Trends Plant Sci.* 2003;8(5):200-1.
2. Christmann A, Moes D, Himmelbach A, Yang Y, Tang Y, Grill E. Integration of abscisic acid signalling into plant responses. *Plant Biol (Stuttg).* 2006;8(3):314-25.
3. Li Z, Peng J, Wen X, Guo H. Gene network analysis and functional studies of senescence-associated genes reveal novel regulators of Arabidopsis leaf senescence. *J Integr Plant Biol.* 2012;54(8):526-39.
4. Lim PO, Kim HJ, Nam HG. Leaf senescence. *Annu Rev Plant Biol.* 2007;58:115-36.
5. Yuan M, Wang S. Rice MtN3/saliva/SWEET family genes and their homologs in cellular organisms. *Mol Plant.* 2013 May;6(3):665-74.
6. Seo PJ, Park JM, Kang SK, Kim SG, Park CM. An Arabidopsis senescence-associated protein SAG29 regulates cell viability under high salinity. *Planta.* 2011;233(1):189-200.
7. Tabata S, Kaneko T, Nakamura Y, Kotani H, Kato T, Asamizu E, et al. Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana. *Nature.* 2000;408(6814):823-6.
8. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExpASY Server. In: Walker JM, editor. *The Proteomics Protocols Handbook.* Humana Press; 2005. p. 571-607.
9. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 2006;34(Web Server issue):W362-5.
10. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389-402.
11. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol.* 1993 Jul 20;232(2):584-99.
12. Wang Y, Geer LY, Chappey C, Kans JA, Bryant SH. Cn3D: sequence and structure views for Entrez. *Trends Biochem Sci.* 2000;25(6):300-2.
13. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 2014;42(Web Server issue):W252-8.



14. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10(6):845-58.
15. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, Simonovic, N., Roth, A., Santos, A., Tsafou, KP., Kuhn, M. Bork, P., Jensen, LJ., von Mering, C. (2014). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*, gku1003.
16. Artero RD, Terol-Alcayde J, Paricio N, Ring J, Bargues M, Torres A, et al. saliva, a new *Drosophila* gene expressed in the embryonic salivary glands with homologues in plants and vertebrates. *Mech Dev.* 1998;75(1-2):159-62.
17. Tao Y, Cheung LS, Li S, Eom JS, Chen LQ, Xu Y, et al. Structure of a eukaryotic SWEET transporter in a homotrimeric complex. *Nature.* 2015;527(7577):259-63.
18. Fujita, Y., Fujita, M., Satoh, R., Maruyama, K., Parvez, M. M., Seki, M., Hiratsu, K., Ohme-Takagi, M., Shinozaki, K & Yamaguchi-Shinozaki, K. (2005). AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in *Arabidopsis*. *The Plant Cell*, 17(12), 3470-3488.