NJB, Volume 32(1), June, 2019

FUNCTIONAL CHARACTERIZATION OF *GENLISEA AUREA* (S8E8K3) PROTEIN

***Anakaa, T. M., Bem, A. A. and Hafiz, S.S.**
Department of Biological Sciences, Federal University, Dutsin-Ma, Nigeria

Received 4th April, 2019; accepted 24th May, 2019

ABSTRACT

This work predicts the functions of *Genlisea aurea* (S8E8K3) protein. Identification of corresponding proteins, conserved domains, functions and pedigree tree of target and corresponding proteins was obtained. Uniprot database, Basic Local Alignment Search Tool (BLAST) and Clustal Omega were used in this study. Results indicated that proteins from *Sesamum indicum* (XP011073982.1), *Erythranthe guttate* (XP012836557.1), *Handroanthus impetiginosus* (PIN05468.1) and *Olea europaea var. sylvestris* (XP022874946.1) showed 91%, 90%, 89% and 87% similarity, respectively, to S8E8K3. Model proteins all possessed WD40 domain. All model proteins functioned as ribonucleoprotein and phylogenetic tree showed that all proteins had eukaryotic origin. Therefore, S8E8K3 is and performs the role of a ribonucleoprotein.

Keywords: Functional, Characterisation, S8E8K3, Protein.

INTRODUCTION

S8E8K3 is a protein encoded by M569_05966 gene of *Genlisea aurea* (Corkscrew plant) that is carnivorous in nature (Marchler-Bauer *et al.,* 2004, 2009, 2011, 2015; Leushkin *et al.,* 2013).The plant comprises of a small genome and belongs to the family Lentibulariaceae. Taxonomically, S8E8K3 transcends from eukaryotes through several lineages that include viridiplantae, embryophyta, tracheophyta, magnoliophyta and genlisea .The protein has a molecular mass of 37,437 Da and consists of 339 amino acid sequences dominated by WD40 repeats (Marchler-Bauer *et al.,* 2004, 2009, 2011, 2015). The name WD40 repeats is derived from the WD conserved dipeptide and 40 amino acids in the sequence (Xu and Min, 2011). This domain consists of 11- 24 GH dipeptide residues and a WD40 domain at the N and C-terminus, respectively. These residues aid in constant and reversible binding to proteins which is credited to seven circular propeller-like blades of antiparallel beta sheets (Marchler-Bauer *et al.,* 2004, 2009, 2011, 2015; Hellsten *et al.,* 2013; Finn *et al.,* 2016). The several functions of this domain include post-translational transcription, protein-protein interaction, apoptosis and signal transduction (Anakaa *et al.*,2017).

Functionally, S8E8K3 is uncharacterised (Marchler-Bauer *et al.,* 2004, 2009, 2011, 2015). In this analysis, functional characterization of some proteins using bioinformatics tools was carried out. Plant proteins corresponding to S8E8K3; *Sesamum indicum* (XP011073982.1), *Erythranthe guttate* (XP012836557.1), *Handroanthus impetiginosus* (PIN05468.1) and *Olea europaea var. sylvestris* (XP022874946.1) were checked and used as model proteins; the conserved domains of model proteins was accessed. Functional similarity between corresponding proteins was done as an insight to that of target protein and pedigree tree of all proteins was performed to check ancestral origin.

*Author for correspondence

MATERIALS AND METHODS

Uniprot database (https://www.uniprot.org/uniprot/S8E8K3) was logged in to obtain the amino acid sequence (Fasta format) of *Genlisea aurea* (S8E8K3) hypothetical protein. The obtained sequence was blasted in Basic Local Alignment Search Tool (BLAST)(https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=ncbiblast_I20190402-140123-0940-75505038-p2m]. The four closest plant proteins were used as the model proteins for the study. NCBI conserved domain database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) was used to access the conserved domain of all proteins. Amino acid sequences of various proteins were analysed to arrive at the various domains. Uniprot (Marchler-Bauer *et al.*, 2004, 2009, 2011, 2015) was further checked to obtain the predicted functions of model proteins as an insight to that of S8E8K3. To obtain this, the accession numbers of all model proteins were entered in uniprot and reviewed independently. Sequence alignment of obtained model sequences together with S8E8K3 was analysed in Clustal Omega Database (https://www.ebi.ac.uk/Tools/msa/clustalo/) to acquire pedigree tree. Amino acid sequences of model and target proteins were aligned to arrive at sequence similarity and pedigree data.

RESULTS

Table 1 shows that XP011073982.1(91%), XP012836557.1(90%), PIN05468.1(89%) and XP022874946.1(87%) proteins are consistent with S8E8K3. Table 2 indicates that WD40 domain is central among the four corresponding proteins as well as query protein. Table 4 shows that the four proteins function as ribonucleoproteins which is an insight to that of S8E8K3. Figure 1 indicates that all corresponding proteins including S8E8K3 share a common ancestral origin.

Table 1: Sequence similarity of model proteins

Sequence Identity	Name of plant	Identity to S8E8K3 protein
XP011073982.1	*Sesamum indicum*	91%
XP012836557.1	*Erythranthe guttate*	90%
PIN05468.1	*Handroanthus impetiginosus*	89%
XP022874946.1	*Olea europaea var. sylvestris*	87%

Table 2: Conserved domain of all proteins

Sequence Identity	Name of plant	Conserved Domain
S8E8K3	*Genliseaaurea*	WD40
XP011073982.1	*Sesamum indicum*	WD40
XP012836557.1	*Erythranthe guttate*	WD40
PIN05468.1	*Handroanthus impetiginosus*	WD40
XP022874946.1	*Olea europaea var. sylvestris*	WD40

Table 3: Functional similarity of model proteins

Sequence Identity	Name of plant	Functional Identity
XP011073982.1	*Sesamum indicum*	Ribonucleoprotein
XP012836557.1	*Erythranthe guttate*	Ribonucleoprotein
PIN05468.1	*Handroanthus impetiginosus*	Ribonucleoprotein
XP022874946.1	*Olea europaea var. sylvestris*	Ribonucleoprotein

Figure 1: Phylogenetic tree of proteins

DISCUSSION

High similarity (Table 1) in sequence between model proteins to that of S8E8K3 is an indication that several characteristics that include functional similarity are common among corresponding and target protein. This suggests that all the plants including *Genlisea aurea* are from the same family. This is in line with Sharma and Kaur (2017) who suggested the classification of plants based on similar features including phytochemicals. Close similarity in sequence between proteins as an indication of the same function is consistent with Esposti (2002) who corresponded similar functions to proteins based on similarity in sequence.

The five proteins contained WD40 domain in line with earlier results (Table 2). This finding clearly confirms the present and distant homology between target and model proteins as described by Lee and Lee (2009), who used domain architecture level for the comparison of proteins by domain-limited techniques. Having checked the conserved domain of the various proteins, functional prediction of corresponding proteins was done (Table 3). The same functional outcome among model proteins suggests that S8E8K3 is a ribonucleoprotein. This implies that the changes did not occur in the 'conserved domains; which is why there was no difference. This is in line with Walbot (1996) who suggested that these changes are preserved in the offspring. Results from Table 3 indicated that the changes that led to variations from the pedigree (Figure 1) were not of nuclear origin (Moore and Haber, 1996; Kirk *et al.*, 2000) and so did not affect the function of this protein nor lead to deletion of M569_05966 gene in common descents.

Pedigree tree (Figure 1) indicates that all five proteins emanated from eukaryotic ancestral origin. This phylogeny further gives the reason for the close similarity in sequence (Table 1). It also explains the similarity in function. Mount (2004) noted that comparative sequence analysis is the best evidence of common descent. Similar evolutionary origin has been reported to be a guide to the same function between proteins. Petrov *et al*. (2000) noted that RNA, DNA and amino acid configuration is preserved through generations. This suggests that proteins of the same nature will carry out the same function in dissimilar organisms. Deviations in the family tree could be attributed to mutation that can be caused by different factors that include junk DNA such as pseudogenes thus diversifying the pedigree lineage (Petrov *et al.*, 2000). Other factors that could be responsible for the diversions include errors in the repair mechanisms during breaks of double-stranded DNA's that can cause insertions and other re-arrangements in the genome (Pipiras *et al*., 1998). Deletion slippages during replication and unequal crossover could also cause these changes along the pedigree (Capy, 2000; Kirk *et al.*, 2000). Duplication of the entire genome (Grant *et al*., 2000) of various plants can also be enough to cause divisions in the phylogeny.

CONCLUSION

This analysis was designed to set the basis for further characterisation of S8E8K3 protein. Though establishing this work with wet laboratory experiments is yet to be ascertained, the study reveals that XP011073982.1 (91%), XP012836557.1(90%), PIN05468.1(89%) and XP022874946.1(87%) are homologous to *Genlisea aurea* (S8E8K3) hypothetical protein; S8E8K3 contains WD40 domain and functions as a ribonucleoprotein and the phylogenetic tree shows that all proteins have eukaryotic origin thus, S8E8K3 is and performs the role of a ribonucleoprotein.

REFERENCES

Anakaa, T.M., Bem, A. A. and Musa, D.D. (2017). Characterisation and Functional Analysis of A0A022RDQ0 Protein. *International Journal of Science and Technology*, 5(11):1-3.

Basic Local Alignment Search Tool (2016). Protein blast. Accessed 05 October, 2018]. Available from: http://blast.ncbi.nlm.gov.

Capy, P. (2000). Perspectives: evolution. Is bigger better in cricket? *Science*, 287: 985–986.

EMBI-EBI. (2016). Clustal Omega Database. Accessed 05 October, 2018]. Available from: http://www.ebi.ac.uk.

Esposti, D.M. (2000). Sequence and functional similarities between pro-apoptotic bid and plant lipid transfer proteins. *BiochimBiophys Acta*, 15(3): 331-340.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J. and Bateman, A. (2016). *Nucleic Acids Research*, 44: 279-285. Available from: https://pfam.xfam.org.

Grant, D., Cregan, P. and Shoemaker, R.C. (2000). Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. *Proc. Natl Acad. Sci. USA*, 97: 4168–4173.

Hellsten, U., Wright, K. M., Jenkins, J., Shu, S., Yuan, Y., Wessler, S. R., Schmutz, J. and Rokhsar, D.S. (2013). Fine-scale variation in meiotic recombination in Mimulus infrared from population shotgun sequencing. *Proceedings of the National Academy of Sciences of USA*. 110:19478-19482. Available from: http://www.uniprot.org.

Lee, B. and Lee, D. (2009). Protein comparison at the domain architecture level. *BMC Bioinformatics*, 10(15): S5.

Leushkin, E.V., Sutormin, R.A., Nabieva, E.R., Penin, A.A., Kondrashov, A.S. and Logacheva, M.D. (2013). The miniature genome of a carnivorous plant *Genliseaaurea* contains a low number of genes and short non-coding sequences**.** *BMC Genomics*, 14:476-476.

Kirk, A., Salomon, S. and Puchta, H. (2000). Species-specific double-stranded break repair and genome evolution in plants. *EMBO J*., 19(20): 5562-5566.

Marchler-Bauer, A. and Bryant, S. H. (2004). Protein domain annotations on the fly. CD: NCBI's conserved domain database. *Nucleic Acids Research*, 32:327- 331.

Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Lui, C., Lu., F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N. and Bryant, S.H. (2009). Specific functional annotation with conserved domain database. CDD: NCBI's conserved domain database. *Nucleic Acids Research*, 37:205-210.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Zheng, C. and Bryant, S.H. (2011). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, 39:225-229.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G. H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C and Bryant, S.H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, 43:222-226.

Moore, J.K. and Haber, J.E. (1996). Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature*, 383: 644–646.

Mount, D. M. (2004) Bioinformatics: Sequence and Genome Analysis (2nd ed.). Cold Spring Harbour Laboratory Press, NY. 600p.

Petrov, D.A. and Hartl, D.L. (2000). Pseudogene evolution and natural selection for a compact genome. *The Journal of Heredity*, 91(3):221–227.

Petrov, D.A., Sangster,T.A., Johnston, J.S., Hartl, D.L. and Shaw, K. L. (2000). Evidence for DNA loss as a determinant of genome size. *Science*, 287: 1060–1062.

Pipiras, E., Coquelle, A., Bieth, A. and Debatisse, M. (1998). Interstitial deletions and intrachromosomal amplification initiated from a double-strand break targeted to a mammalian chromosome. *EMBO J.*, 17: 325–333.

Sharma, R. and Kaur, S. (2017). Antimicrobial and phytochemical screening of Trikuta-traditional food of western Rajasthan. *Indian Journal of Traditional Knowledge*, 16(2): 270-276.

Walbot, V. (1996). Sources and consequences of phenotypic and genotypic plasticity in flowering plants. *Trends in Plant Sciences*,1:27–33.

Xu, C. and Min, J. (2011). Structure and function of WD40 domain proteins. *Protein and Cell*, 2(3): 202–214.