DOI: 10.31276/VJSTE.62(3).62-67

Identification, structural characterization, and *in silico* expression analysis of the sucrose transporter 'SWEET' gene family in peanut (*Arachis hypogaea*)

Ha Duc Chu[1*], Quynh Thi Ngoc Le[2], Yen Hai Thi Hoang[3], Thu Phuong Pham[3], Hong Viet La[3], Nguyet Minh Thi Nguyen[1], Anh Xuan Duong[4], Thao Duc Le[1], Linh Hung Le[1], Hoi Xuan Pham[1]

[1]Agricultural Genetics Institute, Vietnam Academy of Agricultural Sciences
[2]Faculty of Chemistry and Environment, Thuyloi University
[3]Faculty of Biology, Agricultural Technology, Hanoi Pedagogical University 2
[4]Padworth College, United Kingdom

Received 2 March 2020; accepted 29 May 2020

Abstract:

SWEET (Sugars Will Eventually be Exported Transporter) proteins are well known to play pivotal roles in the growth and development of plants. Here, we report the presence of 43 members of the AhSWEET family in peanut (*Arachis hypogaea*) and determine their general characteristics including chromosomal localization, gene structure, and numerous physical and chemical features of the proteins. We found that the *AhSWEET* genes were unevenly distributed among the peanut's 20 chromosomes. The AhSWEET proteins were hydrophobic with a grand average hydropathicity >0 while a majority of the proteins were basic with isoelectric points >7.0. Additionally, most of the *AhSWEET* genes contained 6 exons and 5 introns. The expression profiles of the *AhSWEET* genes were explored based on the previous transcriptome atlas. Interestingly, we found that the *AhSWEET* genes exhibited differential expression patterns across various organs and tissues during the growth and development of peanut plants. Our study provides a solid foundation of the *AhSWEET* gene family for further functional characterization of *AhSWEET* genes in the regulation of peanut growth and development.

Keywords: **bioinformatics, expression profiles, genome-wide, peanut, sucrose transporter, SWEET.**

Classification number: **3.1**

Introduction

Peanut (*Arachis hypogaea*) is considered to be one of the most important legume crops and is mainly cultivated in tropical and subtropical areas. These legumes provide a good source of protein, monounsaturated fats, and antioxidants [1]. Several peanut by-products such as peanut meal, peanut skin, peanut hull, and peanut vine can be used by the food processing industry and consequently play an essential role in food security [2]. However, peanut production and quality are severely affected by abiotic stress [3].

It has been confirmed that the concentration of soluble sugars (predominantly sucrose) can be boosted when plants are exposed to abiotic stress [4, 5]. Of our interest, a group of sucrose transporters, so-called "Sugars Will Eventually be Exported Transporters" or SWEETs, are reported as the functional proteins involved in the translocation of sucrose [6, 7]. Thus, SWEETs regulate numerous biological processes in the growth and development of plants such as nectar secretion, phloem loading and development, and seed filling [7, 8]. Previously, some studies have identified and characterized *SWEET* genes in many main crops such as rice (*Oryza sativa*) [9], soybean (*Glycine max*) [10], sorghum (*Sorghum bicolor*) [11], rapeseed (*Brassica napus*) [12], cotton (*Gossypium* spp.) [13], wheat (*Triticum aestivum*) [14, 15], and litchi (*Litchi chinensis*) [16]. Meanwhile, information from the *SWEET* gene family in peanut is lacking.

Therefore, in this study, the *SWEET* gene family in peanut is identified and characterized based on a bioinformatics approach. Specifically, a comprehensive survey of all putative *SWEET* genes was conducted in the

*Corresponding author: Email: hachu_amser@yahoo.com

peanut genome. Subsequently, the expression profiles of the *SWEET* genes in various organs were generated based on a previous transcriptome atlas.

Materials and methods

Materials

The latest reference genome, proteome, and transcriptome of the peanut ('Tifrunner' cultivar) [17, 18] from the Legume Information System [19] and PeanutBase [18] were used as the platforms for our *in silico* analyses.

Methods

Identification and annotation of the SWEET genes: to identify SWEET proteins in the peanut, we conducted a study in which the domain of the plant's SWEET, namely 'PF03083' [6, 7], obtained from the Pfam server [20], was acquired to search against the recent peanut assembly (BioProject: PRJNA419393) [17] published in NCBI and the Legume Information System [19]. The identified protein sequences were then subjected to a BlastP search against the proteome of the peanut [17] to obtain their necessary annotated information, which includes coding DNA sequence (CDS), genomic DNA sequence (gDNA), and chromosomal localization.

Analysis of characteristics of SWEET proteins: the full-length protein sequences of SWEET proteins were searched against the ExPASy Protparam to obtain general features such as molecular mass, length, instability index, isoelectric point, and grand average of hydropathicity (GRAVY) [21]. An instability index score of <40 and >40 indicates potential stability and instability, respectively. GRAVY values of <0 and >0 suggest hydrophilic and hydrophobic characteristics, respectively [21].

Phylogenetic analysis and gene organization of SWEET genes: a neighbour-joining phylogenetic tree comprised of the full-length amino acid sequences of all identified SWEET proteins was constructed with the aid of MEGA (Molecular Evolutionary Genetics Analysis) 7.0 [22] using the following essential criteria: a gap extension penalty of 0.2 and a gap open penalty of 10 [23]. Bootstrapping was performed with 1,000 replications. The exon/intron structure of each *SWEET* gene was analysed by subjecting the CDS and corresponding gDNA to the GSDS (Gene Structure Display Server) 2.0 tool [24].

Expression profiles of SWEET genes: the PeanutBase database was explored to provide a previous transcriptome atlas of different tissues/organs in the peanut [18]. Particularly, data from seven vegetative plant parts, including vegetative shoot tips, reproductive shoot tips, primary stem leaves, seedling leaves, lateral stem leaves, roots, and nodules were collected. The cluster heatmap for the relative expression of the *SWEET* gene was visualized in *R* software with the gplots package [25].

Results and discussion

Identification and annotation of the SWEET gene family in peanut

To identify all potential members of the SWEET family in the peanut, a comprehensive search of a well-established conserved domain of SWEETs [6, 7] against the newest peanut database [17] was performed. As a result, a total of 43 members of the SWEET family were found in the peanut. The annotation of these identified proteins, including protein identifiers and locus name, are subsequently explored and listed in Table 1. Previously, the *SWEET* gene family has been reported in several plant species. More specifically, 21 members of the *OsSWEET* family have been investigated in rice [9], while 52 and 23 *SWEET* genes have been found in soybean and sorghum, respectively [10, 11]. Recently, it has been reported that the *SWEET* gene family in rapeseed contained 68 members [12]. In the cotton species, the members of the *SWEET* gene family varied from 22 to 60 [13]. Our results indicated that the number of *SWEET* genes in plant species is highly variable.

Next, to annotate the chromosomal localization of the *SWEET* genes, we matched their corresponding gDNA sequence to the peanut genome [17]. We found that all members of the *SWEET* genes were randomly distributed among the 20 chromosomes of the peanut genome and no *SWEET* gene was localized in unplaced scaffolds (Fig. 1). Among them, chromosome Arahy.13 and Arahy.03 share the highest members of the *SWEET* family by 6 and 5 genes, respectively (Fig. 1). Additionally, there are 4 *SWEET* genes found in chromosome Arahy.08, while the chromosomes Arahy.14, 15, 16, 17, and 18 have 3 *SWEET* genes (Fig. 1). We also found that 2 *SWEET* genes were mapped on each of chromosomes Arahy.05, 06, and 20, while only 1 *SWEET* gene was reported in chromosomes Arahy.01, 04, 07, 09, 10, 11, and 19 (Fig. 1). It is also noted that no *SWEET* gene was localized in chromosomes Arahy.02 and 12 (Fig. 1). The entire 43 *SWEET* genes set was based on the order of the occurrences on the chromosomes (Table 1, Fig. 1).

Table 1. General information on *SWEET* gene family in the peanut.

#	Gene name	Protein code	Locus code	Size	MM	pI	II	GRAVY
1	*AhSWEET01*	XP_025675869.1	LOC112776072	227	25.81	8.82	43.35	0.96
2	*AhSWEET02*	XP_025689387.1	LOC112790966	279	31.48	8.99	38.23	0.46
3	*AhSWEET03*	XP_025689047.1	LOC112790726	253	27.92	9.37	46.63	0.56
4	*AhSWEET04*	XP_025676850.1	LOC112776807	159	18.23	9.74	33.58	0.87
5	*AhSWEET05*	XP_025691262.1	LOC112792298	293	32.86	8.49	31.46	0.48
6	*AhSWEET06*	XP_025691265.1	LOC112792300	220	25.04	9.68	28.98	1.05
7	*AhSWEET07*	XP_029153458.1	LOC112795203	175	20.21	9.83	38.61	0.82
8	*AhSWEET08*	XP_025697629.1	LOC112799834	250	27.59	9.19	32.87	0.72
9	*AhSWEET09*	XP_025697627.1	LOC112799832	242	26.85	8.68	37.49	0.80
10	*AhSWEET10*	XP_025606390.1	LOC112697429	244	26.70	8.91	32.98	0.78
11	*AhSWEET11*	XP_025603714.1	LOC112695553	312	34.23	9.15	21.70	0.41
12	*AhSWEET12*	XP_025610780.1	LOC112703520	235	26.20	8.38	44.42	0.84
13	*AhSWEET13*	XP_025612828.1	LOC112705985	246	27.12	9.30	32.96	0.64
14	*AhSWEET14*	XP_025612772.1	LOC112705945	262	29.03	9.02	27.45	0.51
15	*AhSWEET15*	XP_025616058.1	LOC112708127	301	34.32	8.84	51.86	0.10
16	*AhSWEET16*	XP_025616059.1	LOC112708128	285	32.77	9.05	39.49	0.40
17	*AhSWEET17*	XP_025616659.1	LOC112708960	320	35.26	8.34	34.39	0.70
18	*AhSWEET18*	XP_025622395.1	LOC112714914	292	32.46	8.82	33.82	0.66
19	*AhSWEET19*	XP_025629145.1	LOC112722363	226	25.74	8.82	43.49	0.95
20	*AhSWEET20*	XP_025641347.1	LOC112736207	261	29.73	6.99	41.81	0.53
21	*AhSWEET21*	XP_025637471.1	LOC112732876	278	31.37	9.09	38.33	0.48
22	*AhSWEET22*	XP_025637469.1	LOC112732875	249	28.29	9.71	33.13	0.84
23	*AhSWEET23*	XP_025636904.1	LOC112732406	253	27.90	9.37	46.29	0.56
24	*AhSWEET24*	XP_025639618.1	LOC112734493	293	32.90	8.97	32.73	0.48
25	*AhSWEET25*	XP_025639626.1	LOC112734496	225	25.30	9.66	31.43	1.00
26	*AhSWEET26*	XP_025650623.1	LOC112745021	274	30.31	8.95	35.74	0.60
27	*AhSWEET27*	XP_025646602.1	LOC112741728	200	21.71	8.46	30.60	0.58
28	*AhSWEET28*	XP_025650514.1	LOC112744946	261	29.73	6.99	41.08	0.54
29	*AhSWEET29*	XP_025655604.1	LOC112750900	248	27.61	8.72	36.50	0.68
30	*AhSWEET30*	XP_025655423.1	LOC112750786	250	27.76	9.18	33.60	0.71
31	*AhSWEET31*	XP_025650947.1	LOC112747164	242	26.89	8.84	38.20	0.80
32	*AhSWEET32*	XP_025659195.1	LOC112755367	268	29.33	9.16	24.33	0.44
33	*AhSWEET33*	XP_025662459.1	LOC112758096	310	34.17	9.22	25.22	0.36
34	*AhSWEET34*	XP_025659173.1	LOC112755353	244	26.62	9.22	32.26	0.77
35	*AhSWEET35*	XP_025667967.1	LOC112766275	262	28.88	9.01	28.76	0.54
36	*AhSWEET36*	XP_025664711.1	LOC112763193	104	11.33	6.51	38.82	0.65
37	*AhSWEET37*	XP_025666957.1	LOC112765256	246	26.95	9.30	34.20	0.64
38	*AhSWEET38*	XP_025672359.1	LOC112771761	235	26.10	7.62	43.28	0.86
39	*AhSWEET39*	XP_025673457.1	LOC112772695	301	34.38	8.71	57.28	0.10
40	*AhSWEET40*	XP_025672677.1	LOC112772012	287	33.01	9.05	39.29	0.38
41	*AhSWEET41*	XP_025679981.1	LOC112779842	281	31.26	8.47	40.94	0.78
42	*AhSWEET42*	XP_025685723.1	LOC112786568	248	27.49	8.84	36.64	0.81
43	*AhSWEET43*	XP_025683874.1	LOC112784769	292	32.52	8.81	35.55	0.65

Note: MM: molecular mass (kDa), pI: isoelectric point, II: instability index, GRAVY: grand average of hydropathicity.

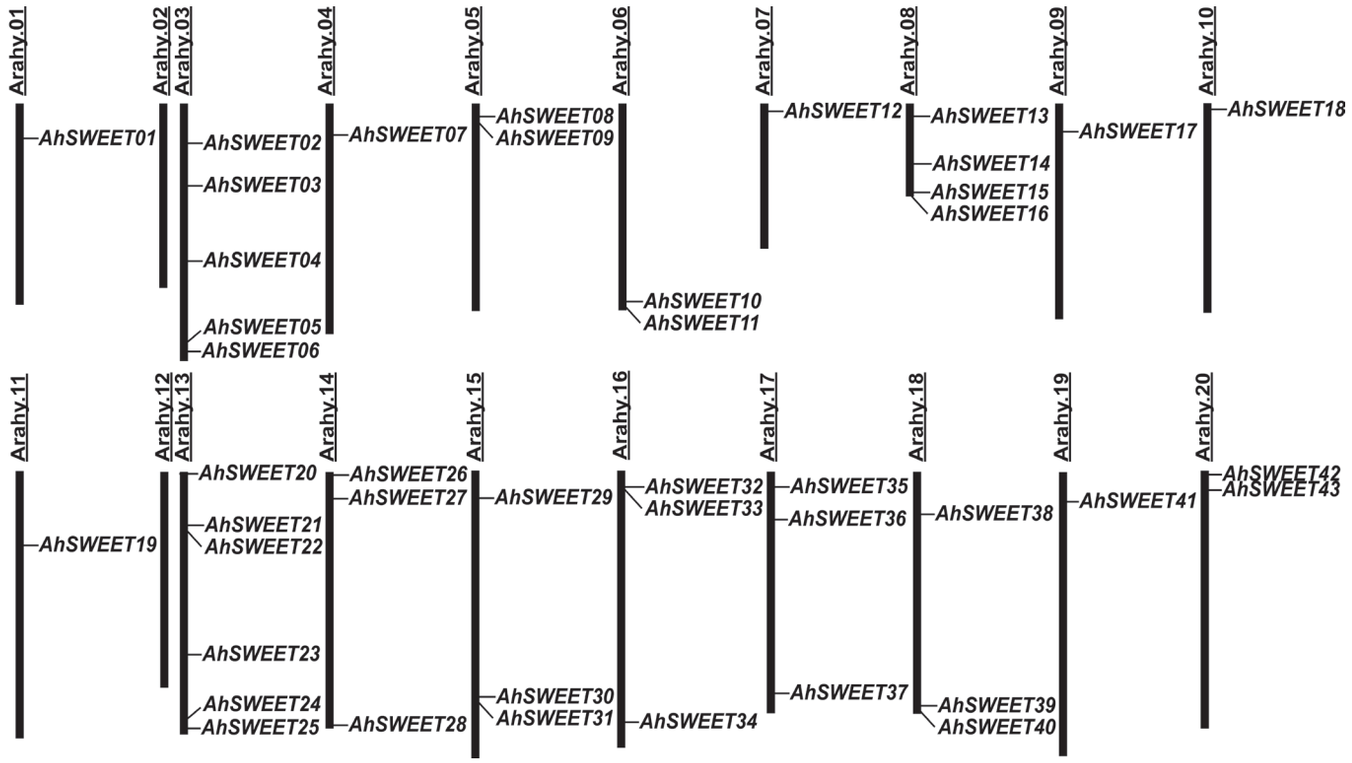

Fig. 1. Chromosomal distribution of *AhSWEET* genes in peanut. Chromosomal localization of 43 *AhSWEET* genes was based on the latest physical map described in NCBI and the Legume Information System.

Structural analysis of the SWEET gene family in peanut

The exon/intron organization of each *AhSWEET* gene was first analysed in order to gain insight into the *AhSWEET* gene family. As shown in Fig. 2, the most common motif of the gene structure of the *AhSWEET* family was 6 exons/5 introns. Only *AhSWEET41* and *AhSWEET36* contained 2 exons/1 intron and 3 exons/2 introns, respectively, while 3 genes, including *AhSWEET04*, *07*, *17* had 4 exons/3 introns (Fig. 2). Our findings were also confirmed by previous studies [10, 12-15, 26]. More specifically, a total of 34 (out of 52) *GmSWEET*s was recorded to contain 6 exons/5 introns [10], while the majority of *BnSWEET*s (51 out of 68) also had 6 exons/5 introns [12]. This phenomenon was also reported in other plant species such as cotton [13], wheat [14, 15], and litchi [16]. Taken together, it would be a reliable assumption that the general structure of *SWEET* genes in higher plant species is 6 exons/5 introns.

Next, the full-length protein sequence of each SWEET was used for retrieval from the ExPASY Protparam [21] in order to analyse the general features of the SWEET family in the peanut. The length of the SWEET proteins varied

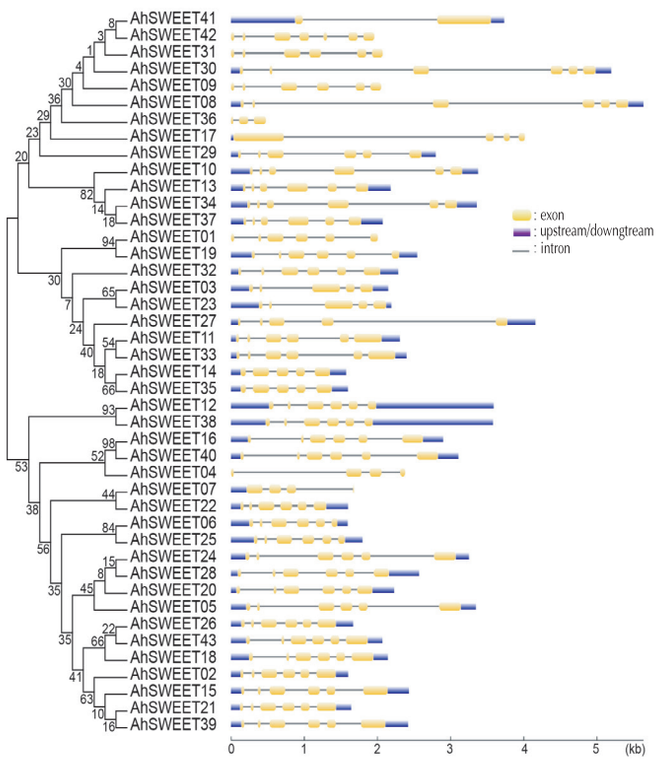

Fig. 2. Gene structure of *AhSWEET* gene family. An unrooted neighbour-joining tree was derived from the full-length AhSWEET sequences (left) and exon/intron organization analysis (right).

from 104 (AhSWEET36) to 320 residues (AhSWEET17) with their molecular masses ranging from 11.33 to 35.26 kDa, respectively (Table 1). The pI values of a majority of the SWEET proteins were >7, which revealed that these proteins were basic whereas only AhSWEET36 was acidic (pI=6.51) (Table 1). The two remaining SWEET proteins, AhSWEET20 and 28, were neutral (pI≈7) (Table 1). We also found that 32 SWEET proteins were stable (instability score <40) (Table 1). Furthermore, all 43 SWEET proteins were hydrophobic with a GRAVY value >0 (Table 1).

Previously, the characteristics of SWEET proteins have also been investigated in other plant species. For example, the SWEET proteins in rapeseed varied from 56 to 303 residues, while their molecular weight ranged from 6.5 to 33.45 kDa [12]. A total of 63 members (out of 68) of SWEET proteins were basic [12]. Additionally, most of the identified cotton's SWEET proteins ranged between 180 and 311 residues, while the molecular masses and isoelectric values of these proteins varied from 9.93 to 38.04 kDa and from 5.47 to 10.08, respectively [13]. In wheat, the molecular weights of SWEET proteins ranged from 10.93 to 33.86 kDa, while a majority of members in the SWEET family exhibited pI values >7 (basic) [14, 15]. Recently, the sizes and molecular weights of the LcSWEET proteins have been found to vary from 229 to 300 residues and from 25.6 to 33.6 kDa, respectively, while the pI values ranged from 7.66 to 9.81 [16]. Our findings suggest a diversity of molecular features of SWEETs in the peanut and perhaps in the plant species.

Expression profiles of AhSWEET genes in various tissues

To understand the expression patterns of the *AhSWEET* gene family, we visualized the transcriptome data obtained from 7 tissues respectively taken from vegetative shoot tip, reproductive shoot tip, main stem leaf, seedling leaf, lateral stem leaf, root, and nodule [18] by R programming with the gplots package [25]. We found that 17 genes, including *AhSWEET03*, *04*, *07*, *10*, *14*, *17*, *18*, *20*, *21*, *23*, *31*, *34*, *35*, *36*, *39*, *41*, and *42*, had no information on the expression profiles. The expressions of the remaining *AhSWEET* genes are displayed in Fig. 3.

Among them, 11 *AhSWEET* genes had no changes in the transcriptional levels of the 7 collected tissues (Fig. 3). Interestingly, *AhSWEET02* was noted to exclusively express in 3 samples of leaves and the reproductive shoot tip, while *AhSWEET15* was also strongly induced in lateral stem leaves, seeding leaves, and main stem leaves (Fig. 3). *AhSWEET27* was found to be strongly up-regulated in both

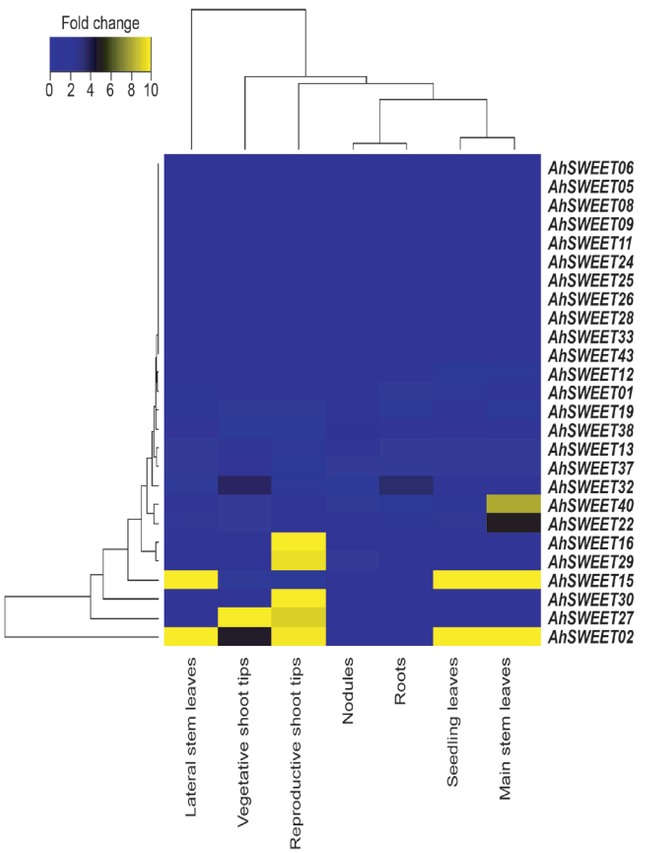

Fig. 3. Expression profiles of the *AhSWEET* genes in various tissues. The heat map was generated using R software with the gplots package. The detailed microarray data were obtained from the peanut gene atlas database.

reproductive and vegetative shoot tips (Fig. 3). In some cases, the *AhSWEET* genes were down-regulated in organs/tissues during the growth and development of the peanut plants. For example, *AhSWEET13* and *37* were recorded to be strongly reduced in lateral stem leaves, seeding leaves, and main stem leaves (Fig. 3). Taken together, the *AhSWEET* genes displayed differential transcription patterns in the investigated organs. Our results suggest that AhSWEET proteins might have diverse functions in controlling the development of various organs in peanut plants.

Conclusions

In this study, 43 *AhSWEET* genes have been identified in the peanut genome. Structural analyses revealed that the AhSWEET proteins were highly variable. Our expression re-analysis showed that the *AhSWEET* genes displayed differential expression levels in various organs. Two genes, *AhSWEET02* and *15*, were noted to strongly express in leaves and *AhSWEET27* was strongly induced in shoot tips, which indicate that these genes might play crucial roles

in these organs during the growth and development of the peanut.

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] O.T. Toomer (2018), "Nutritional chemistry of the peanut (*Arachis hypogaea*)", *Crit. Rev. Food Sci. Nutr.*, **58(17)**, pp.3042-3053.

[2] X. Zhao, J. Chen, F. Du (2012), "Potential use of peanut by-products in food processing: a review", *J. Food Sci. Technol.*, **49(5)**, pp.521-529.

[3] D.M. Kambiranda, et al. (2011), "Impact of drought stress on peanut (*Arachis hypogaea* L.) productivity and food safety", *Plants Environ*., Tech Publisher, pp.249-272.

[4] P.V. Minorsky (2003), "Raffinose oligosaccharides", *Plant Physiol.*, **131(3)**, pp.1159-1160.

[5] M.R. Morsy, L. Jouve, J.F. Hausman, L. Hoffmann, J.M. Stewart (2007), "Alteration of oxidative and carbohydrate metabolism under abiotic stress in two rice (*Oryza sativa* L.) genotypes contrasting in chilling tolerance", *J. Plant Physiol.*, **164(2)**, pp.157-167.

[6] B.T. Julius, K.A. Leach, T.M. Tran, R.A. Mertz, D.M. Braun (2017), "Sugar transporters in plants: new insights and discoveries", *Plant Cell Physiol.*, **58(9)**, pp.1442-1460.

[7] R.F. Baker, K.A. Leach, D.M. Braun (2012), "SWEET as sugar: new sucrose effluxers in plants", *Mol. Plant*, **5(4)**, pp.766-768.

[8] L.Q. Chen (2014), "SWEET sugar transporters for phloem transport and pathogen nutrition", *New Phytol.*, **201(4)**, pp.1150-1155.

[9] M. Yuan, S. Wang (2013), "Rice MtN3/saliva/SWEET family genes and their homologs in cellular organisms", *Mol. Plant*, **6(3)**, pp.665-674.

[10] G. Patil, et al. (2015), "Soybean (*Glycine max*) *SWEET* gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis", *BMC Genomics*, **16**, DOI: 10.1186/s12864-015-1730-y.

[11] H. Mizuno, S. Kasuga, H. Kawahigashi (2016), "The sorghum SWEET gene family: stem sucrose accumulation as revealed through transcriptome profiling", *Biotechnol. Biofuels*, **9**, DOI: 10.1186/s13068-016-0546-6.

[12] H. Jian, et al. (2016), "Genome-wide analysis and expression profiling of the SUC and SWEET gene families of sucrose transporters in oilseed rape (*Brassica napus* L.)", *Front. Plant Sci.*, **7**, DOI: 10.3389/fpls.2016.01464.

[13] L. Zhao, et al. (2018), "A genome-wide analysis of *SWEET* gene family in cotton and their expressions under different stresses", *J. Cotton Res.*, **1(1)**, DOI: 10.1186/s42397-018-0007-9.

[14] Y. Gao, et al. (2018), "Genome-wide identification of the *SWEET* gene family in wheat", *Gene*, **642**, pp.284-292.

[15] T. Gautam, et al. (2019), "Further studies on sugar transporter (*SWEET*) genes in wheat (*Triticum aestivum* L.)", *Mol. Biol. Reps.*, **46(2)**, pp.2327-2353.

[16] H. Xie, et al. (2019), "Genome-wide identification and expression analysis of *SWEET* gene family in *Litchi chinensis* reveal the involvement of LcSWEET2a/3b in early seed development", *BMC Plant Biol.*, **19(1)**, DOI: 10.1186/s12870-019-2120-4.

[17] D.J. Bertioli, et al. (2019), "The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*", *Nat. Genet.*, **51(5)**, pp.877-884.

[18] J. Clevenger, Y. Chu, B. Scheffler, P. Ozias-Akins (2016), "A developmental transcriptome map for allotetraploid *Arachis hypogaea*", *Front. Plant Sci.*, **7**, DOI: 10.3389/fpls.2016.01446.

[19] S. Dash, et al. (2016), "Legume information system (LegumeInfo.org): a key component of a set of federated data resources for the legume family", *Nucleic Acids Res.*, **44(D1)**, pp.1181-1188.

[20] S. El-Gebali, et al. (2019), "The pfam protein families database in 2019", *Nucleic Acids Res.*, **47(D1)**, pp.427-432.

[21] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch (2003), "ExPASy: the proteomics server for in-depth protein knowledge and analysis", *Nucleic Acids Res.*, **31(13)**, pp.3784-3788.

[22] S. Kumar, G. Stecher, K. Tamura (2016), "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets", *Mol. Biol. Evol.*, **33(7)**, pp.1870-1874.

[23] H.D. Chu, K.H. Nguyen, Y. Watanabe, D.T. Le, T.L.T. Pham, K. Mochida, L.P. Tran (2018), "Identification, structural characterization and gene expression analysis of members of the nuclear factor-Y family in chickpea (*Cicer arietinum* L.) under dehydration and abscisic acid treatments", *Int. J. Mol. Sci.*, **19(11)**, DOI: 10.3390/ijms19113290.

[24] B. Hu, J. Jin, A.Y. Guo, H. Zhang, J. Luo, G. Gao (2015), "GSDS 2.0: an upgraded gene feature visualization server", *Bioinformatics*, **31(8)**, pp.1296-1297.

[25] Y. Liao, G.K. Smyth, W. Shi (2019), "The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads", *Nucleic Acids Res.*, **47(8)**, DOI: 10.1093/nar/gkz114.

[26] H. Miao, et al. (2017), "Genome-wide analyses of SWEET family proteins reveal involvement in fruit development and abiotic/biotic stress responses in banana", *Sci. Rep.*, **7(1)**, DOI: 10.1038/s41598-017-03872-w.