Article

Investigation and computational analysis of sulfotransferase (*SOT*) gene family in potato (*Solanum tuberosum*): Insights into sulfur adjustment for proper development and stimuli responses

Sahar Faraji[1], Parviz Heidari[2,*], Hoorieh Amouei[1], Ertugrul Filiz[3], Abdullah[4], Peter Poczai [5,6,*]

[1] Department of Plant Breeding, Faculty of Crop Science, Sari Agricultural Sciences and Natural Resources University (SANRU), Sari, Iran; sahar.faraji@rocketmail.com (S.F.); h.amooei65@gmail.com (H.A.)
[2] Faculty of Agriculture, Shahrood University of Technology, Shahrood 3619995161, Iran
[3] Department of Crop and Animal Production, Cilimli Vocational School, Duzce University, Duzce 81750, Turkey; ertugrulfiliz@gmail.com
[4] Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan; abd.ullah@bs.qau.edu.pk
[5] Finnish Museum of Natural History, P.O. Box 7, FI-00014 University of Helsinki, Finland
[6] Faculty of Biological and Environmental Sciences, PO Box 65 FI-00065 University of Helsinki, Finland
* Correspondence: heidarip@shahroodut.ac.ir; Tel.: +98-912-0734-034.(P.H.); (peter.poczai@helsinki.fi (P.P.)

Abstract: Various kinds of primary metabolisms in plants are modulated through sulfate assimilation that the uptake of this inorganic compound can be regulated via the sulfate transporters, such as sulfotransferases (SOTs), engaged in the sulfur metabolism. In the current study a genome-wide approach has been utilized for recognition and characterization of *SOT* family genes in the significant nutritional crop potato (*Solanum tuberosum* L.). As a result, 29 *StSOT* genes were identified in the potato genome, which were mapped onto the nine *S. tuberosum* chromosomes. The protein motifs structure demonstrated two highly conserved 5' PSB region and 3' PB motif that are essential for sulfotransferase and catalytic activities. The protein-protein interaction networks also significantly demonstrated an interesting collaboration between SOTs and the other genes, such as PRTase, APS-kinase, protein phosphatase and APRs, in sulfur compounds biosynthesis and regulation of the flavonoid and brassinosteroid metabolic processes, which clearly detected the importance of sulfotransferases for potato proper growth/development and stress dealing. Notably, the homology modeling of StSOT proteins and docking analysis of their ligand-binding sites revealed the presence of some stress-responsive residues, such as proline, glycine, serine and lysine, in their active sites. The expression assay of *StSOT* genes via the potato RNA-seq data clearly suggested the engagements of these gene family members in plants growth and extension as well as responses to various hormones and biotic/abiotic stimulus circumstances. Our predictions can be informative for the functional characterization of the *SOT* genes in potato and may the other nutritional crops.

Keywords: Sulfur; Sulfotransferase; Potato; Bioinformatics; Proteins structure; Stimuli coping.

1. Introduction

The chemical element sulfur (S) as the necessary factor during life cycle are frequently found in the amino acid components of proteins such as cysteine (Cys) and methionine (Met), some vitamins like thiamin and biotin, multiple co-enzymes (like S-adenosyl methionine), iron-sulfur complex, prosthetic substances, glutathione (GSH) antioxidant as well as in various natural secondary metabolites [1]. This important element is engaged in a vast collection of biological functions in organisms, because of its particular capability in changing its oxidation status that has caused it to be abundant in different compounds. Although the oxyanion sulfate is considered as the significant form of inorganic sulfur in the living world, a large number of metabolites encompass the organic style of sulfide, i.e., reduced sulfur [2]; so, organisms should expend energy for production of the reduced form of the sulfate content, while all organisms have not this ability to assimilate sulfate. The plants, algae, fungi and bacteria taxa have been specialized in oxidizing reduced

forms of sulfur in order to produce the required energy [3], thus can be introduced as the important regulators of the sulfur cycle in food chain.

After the proper cultivar selection and preparation of a persistent water supply, an appropriate nutrient management can be significantly considered as an agronomic parameter in crop production. The nutrients nitrogen, potassium, phosphorous, magnesium, calcium and sulfur (with a significant balance between N and S to attain plants full potential) importantly regulate various physiological functions in the plant metabolisms and participate in tuber yield generation in the tuber crops. With a sufficient content of sulfur, plants are able to impressively use of N and the other nutrients, which this critical element can be gained from the growing medium as sulfate or SO_4 [4]. Sulfur, as a regulatory nutrient in plants yield and as a component of the significant methionine and cysteine amino acids in the structures of proteins, is critical for multiple cellular metabolites [5]. Sulfur constructs an equilibrium with phosphorus and nitrogen that functions an important role in plants nourishment [6], so that the inadequate magnitudes of sulfur in the soil can postpone the plant growth and development as well as disturb its resistance to various biotic and abiotic conditions [7]. At the global scale, sulfur deficiency in crops is discussed as a critical concern, because stimulates multiple detrimental alterations in plant physiology resulting in the yield quantity and quality damages via some disorders in the lipid and protein combinations [4]. A sulfur content of less than 0.25% in any plant tissue can be associated with severe S deficiency and plants enduring a sulfur shortage reveal an overall chlorosis and yellowish phenotype that can be confounded with conditions resulted by nitrogen deficiency, which both are due to a lack of chlorophyll [8]. These observations can be generally found during the early stages of crop development, but over the time, the upper leaves may demonstrate a pallid phenotype with some necrotic spots in the central parts of the leaflets, in comparison with the down-stream leaves. Although the atmospheric S (in the form of SO2) can be absorbed by the higher plants, the significant level of the required sulfur is prepared by absorption of the sulfur-containing fertilizers through the root system [9], for example, potatoes are usually fertilized with K2SO4 to meet the need for S [10]. Furthermore, the excessive levels of sulfur nutrient may lead to sulfur toxicity in plants that is very rare and will probably not arise even if the sulfuric acid is extremely entered into the water supply to counteract its alkalinity [6]. With increasing accumulation of sulfate in the plant cells, its reduction also promotes, because of the storage of excess sulfur in a metabolically inactive compartment, vacuole, which appears to occur in most plants. Also, it was found that an extra and toxic magnitude of sulfur in the growing medium may emulate with nitrogen and eventually lead to N deficiency [10]. Hence, the sulfur regulating pathways have evolved in the higher plants to assimilate the required S contents as well as cope with excess sulfur in their environments.

Sulfotransferases (SOTs) (EC 2.8.2.-) have been discovered as the momentous sulfate regulating genes in various organisms. The sulfated substances in plants function as the significant secondary metabolites and hormones in coping with stimulus situations and can be considered as an important S storage during the life cycle [11]. The plant SOTs are directly engaged in the sulfation process of desulpho-glucosinolate compounds (ds-Gl), which have been introduced as the important secondary metabolites causing resistance against multiple biotic/abiotic stimuli in the brassicales plants [5]. All SOT proteins can be identified through a histidine residue in their PAPS-binding region as well as a specific SOT domain (Pfam: PF00685) [12]. The SOT family members have been specified by four conserved regions (I to IV) in their protein sequences [13], which the I and IV regions were detected as the highly conserved sections [11]. Three *AtSOT16*, *AtSOT17* and *AtSOT18* genes in the *Arabidopsis thaliana* (At) genome have been reported to be responsible for transferring of a sulfuryl group to various ds-Gl compounds [11,14]. The various substances including brassinosteroids, gibberellic acids, glucosinolates, flavonoids, coumarins and phenolic acids can be sulfated by SOT proteins in various plant species [15,16].

According to the multiple previous studies, *SOT* genes can regulate the plant stimuli responses, stress sensing and signaling mechanisms and developmental processes. For instance, the expression profiling of the 35 *SOT* genes from *Oryza sativa* demonstrated the low transcript magnitudes in the apical meristem and young leaves, while some increments were found in the *SOTs* mRNA expression in root, stigma and ovary tissues [17]. The transcription rate of some *Brassica rapa ds-Gl SOTs* also illustrated the strong expression of *BrSOT16* in all tissues except for stamen and *BrSOT18* in the carpel and stamen [18]. Additionally, all the studied *SOTs* from different organisms revealed the induction by various stressor compounds. Interestingly, it was suggested that some *ds-Gl AtSOTs*, such as *AtSOT15*, are responsible for the circadian control [15]. Furthermore, the expression level of eleven *OsSOTs* demonstrated some up and down regulation in coping with dehydration, high/low temperatures and hormone stresses in various tissues [17]. A northern blot experiment in *AtSOT12* showed that the deduced protein employs the flavonoids, brassinosteroids and salicylic acid compounds as its substrates and can be expressed in leaves, flowers and roots as well as in dealing with abiotic stimulus salt, sorbitol, cold, hormones and interactions with biotic pathogens [19,20]. The same reports were also conducted about the relative homologous genes from *B. napus*, which the *BNST3* and *BNST4* transcript magnitudes revealed the significant increments under hormones, low oxygen, xenobiotics and herbicide stimulus [16,21], indicating the momentous role of these genes in stress coping and detoxification.

The importance of sulfur nutrition during the plants life cycle and their biological and chemical processes has been widely known and the knowledge regarding the effects of sulfur shortage in crop production has improved. Potato is considered as the other important food crop after wheat, maize and rice worldwide. Adequate sulfur content in potato plants, helps with multiple nutrients uptake, carbohydrate formation and vitamin synthesis, chlorophyll production, seed development and stress and pest resistance [7,22]. The upward curving of the leaves and their lighter green to yellow color can be mentioned as the important signs of sulfur deficiency in potato plants. Defective sulfur contents also can lead to poor growth of plants, a prolate form of them and their postponed maturity [6]. In most potato production the sandy low organic soils are used and the S content is most available in soils with pH >6.0 that a slight need for sulfur nutrition can be met through the irrigation water in these soils, which may provide some or all of the requirement. There is evidence declaring that higher magnitudes of the applied sulfur in cultivation can remarkably diminish the infection of the plant tuber by common scab and black scurf, as the significant diseases in potato plants [7]. In the previous studies it has been reported that the sufficient sulfur element elevated the yield of potato tubers and their quality as well as potato tolerance against various pathogens through the sulfur-induced resistance (SIR) mechanism [7]. According to an examination on the metabolite pools of the S-assimilatory pathway in the roots of potato plants exposed to S deficiency it was also detected that the sulfate, Cys, glutathione, leaf tissue metabolite pools and total S content significantly declined after S limitation [23]. Such environmental issues necessitate the understanding of plant sulfur biology and adjustment of sulfur nutrition in agricultural programs. Therefore, identification of important sulfate transporters in the S metabolism may elucidate the S-mediated proper growth and resistance mechanisms in potato. Regarding the development of genomes sequencing, the significant sulfate transporters SOTs, have been distinguished in *A. thaliana* with 22 members [11], *O. sativa* [17] and *B. rapa* [18]. However, the identification and characterization of SOT proteins in the potato (*Solanum tuberosum*) genome had been limited so far. Hence, in the current study various bioinformatics approaches have been utilized to distinguish an important cluster of sulfur regulating genes, *SOTs*, and their expression patterns in multiple tissues and during different biotic/abiotic stimuli in the potato genome. Our predictions can prepare helpful insights into the functional evaluation of the *SOT* gene family members in potato and the relative crop species.

2. Results

2.1. Identification of StSOT genes

The deduced amino acid sequence of sulfotransferase domain (PF00685) was searched against the HMM program and Phytozome database that led to the identification of 29 putative StSOT proteins, including 26 cytosolic sulfotransferases, two flavonol sulfotransferases and one steroid sulfotransferase all containing the Sulfotransfer_1 domain and were named according to their chromosomal order (Table 1).

The distinguished StSOT proteins revealed a diversification in their length, ranged from 101 (encoded by *StSOT07* and *StSOT08*) to 359 (encoded by *StSOT21*), with the molecular weights (MW) ranging from 11.83 kDa (encoded by *StSOT07*) to 41.56 kDa (encoded by *StSOT21*). The majority of the identified StSOT proteins (~65.5%) illustrated an acidic nature, because of the theoretical isoelectric points (*pI*) lower than 7.0, ranging from 4.95 (in cytosolic StSOT28) to 6.83 (in cytosolic StSOT13). The subcellular location of proteins indicated that the majority of StSOTs (~76%) can be considered as the cytoplasmic proteins with no putative transmembrane domains (TMDs), while StSOT07, StSOT08 and StSOT28 were predicted to be located in the nucleus in addition to cytoplasm (Table 1). The proteins StSOT01 and StSOT22 were predicted to be localized in the nucleus and extracellular region as well. Two StSOT proteins, namely StSOT23 and StSOT29, could also be found in the mitochondria. Not all StSOT proteins contained any putative TMDs in both cytosolic N- and C-terminal regions that can suggest their specific function during the other cellular pathways apart from membrane transport. The StSOT proteins post-transcriptional phosphorylation analysis illustrated a wide variety of phosphorylated serine (S) residues along with some changed threonine (T) and tyrosine (Y) sites (Figure 1). The proteins StSOT02, StSOT05, StSOT07, StSOT08, StSOT23, StSOT28 and StSOT29 were predicted to contain a limited phosphorylated regions (in 1-2 residues) in their amino acid sequences, while some StSOTs, such as StSOT01, StSOT04, StSOT06, StSOT12, StSOT14, StSOT22 and StSOT26, were predicted as the possible highly phosphorylated sulfotransferase proteins in potato.

Table 1. The identified *StSOT* gene family members and their characteristics in the potato genome

Gene ID	Gene symbol	Protein length	MW (KDa)	Isoelectric point	Subcellular localization
PGSC0003DMG400000144	*StSOT01*	296	34.38	6.54	Nuclear, Cyt., Extra.
PGSC0003DMG400027779	*StSOT02*	345	40.01	7.12	Cyt.
PGSC0003DMG400003287	*StSOT03*	337	38.80	5.73	Cyt.
PGSC0003DMG400031776	*StSOT04*	344	40.10	5.4	Cyt.
PGSC0003DMG400024622	*StSOT05*	350	40.15	6.54	Cyt.
PGSC0003DMG400018798	*StSOT06*	326	37.56	5.62	Cyt.
PGSC0003DMG400026753	*StSOT07*	101	11.83	5.74	Nuclear, Cyt.
PGSC0003DMG400026752	*StSOT08*	101	11.98	7.68	Nuclear, Cyt.
PGSC0003DMG400039363	*StSOT09*	313	36.15	6.27	Cyt.
PGSC0003DMG400005584	*StSOT10*	330	38.49	6.6	Cyt.

PGSC0003DMG400028349	*StSOT11*	335	39.05	6.8	Cyt.
PGSC0003DMG400028301	*StSOT12*	335	39.17	7.11	Cyt.
PGSC0003DMG400025717	*StSOT13*	308	35.90	6.83	Cyt.
PGSC0003DMG400036271	*StSOT14*	329	38.38	6.42	Cyt.
PGSC0003DMG400046427	*StSOT15*	330	38.58	7.13	Cyt.
PGSC0003DMG400028302	*StSOT16*	332	38.66	6.72	Cyt.
PGSC0003DMG400028350	*StSOT17*	240	28.31	6.31	Cyt.
PGSC0003DMG400015051	*StSOT18*	269	31.41	7.71	Cyt.
PGSC0003DMG400028341	*StSOT19*	268	31.24	7.72	Cyt.
PGSC0003DMG403028340	*StSOT20*	209	24.68	7.67	Cyt.
PGSC0003DMG400002358	*StSOT21*	359	41.56	7.03	Cyt.
PGSC0003DMG400014962	*StSOT22*	226	26.06	6.59	Nuclear, Extra.
PGSC0003DMG400029882	*StSOT23*	118	13.63	6.5	Cyt., Mitochondrial
PGSC0003DMG400020968	*StSOT24*	316	36.90	7.16	Cyt.
PGSC0003DMG400039919	*StSOT25*	244	28.49	5.51	Cyt.
PGSC0003DMG400046295	*StSOT26*	329	38.25	5.83	Cyt.
PGSC0003DMG400046521	*StSOT27*	161	19.20	5.76	Cyt.
PGSC0003DMG400014947	*StSOT28*	105	12.24	4.95	Cyt., Nuclear
PGSC0003DMG400009660	*StSOT29*	106	12.10	8.99	Cyt., Mitochondrial,Nuclear

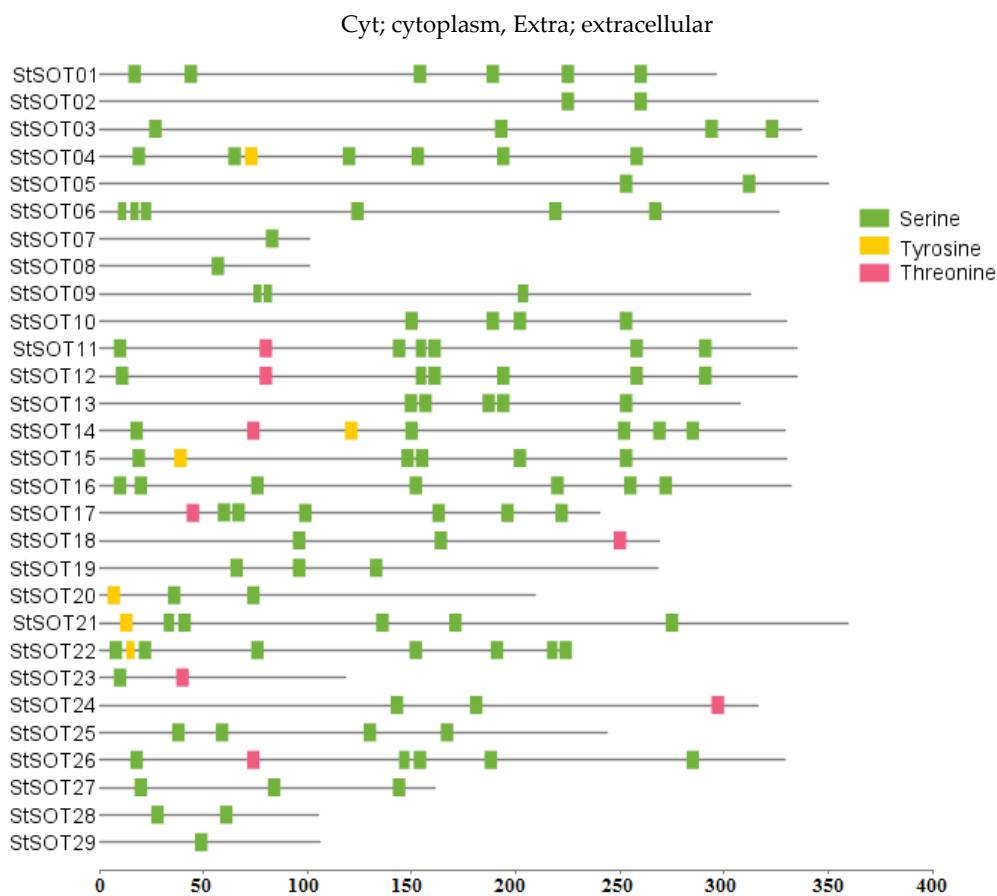Cyt; cytoplasm, Extra; extracellular

Figure 1.   Phosphorylation prediction with scores higher than 0.90 in StTOT proteins based on three amino acids, serine, threonine, and tyrosine, using NetPhos 3.1 server.

2.2. Phylogenetic relationships, conserved motifs/residues and gene structure of StSOTs

The sulfotransferase proteins from potato, *Arabidopsis thaliana*, *Solanum lycopersicum*, and *Sorghum bicolor* were used to generate a phylogenetic tree to classify the SOT proteins into subfamilies (Figure 2). The phylogenetic tree grouped SOTs into the four main clades according to the tree topology and classification of the sulfotransferases in Arabidopsis. Four SOTs of tomato along StSOT09 were classified in the group I and they showed a high genetic distance. Six StSOTs and five SOTs of tomato were located in the group II and all sorghum SOT proteins were grouped with StSOT01, StSOT02, StSOT04, StSOT05 and StSOT25 from potato and AtSOT16, AtSOT17 and AtSOT18 from Arabidopsis and four tomato SOTs in the clade III. Interestingly, all sorghum SOT proteins were obviously separated from dicots SOTs. In the group IV as the largest group, most SOTs of potato, Arabidopsis and tomato were located in this clade (Figure 2).

The eight conserved motifs have been predicted in the StSOT protein sequences via MEME program (Figure 3a and Table S1). The StSOT proteins belonging to the same phylogenetic group shared an approximately similar conserved motif composition. Five out of the eight predicted motifs, including motif 1, motif 2, motif 3, motif 4 and motif 6, were detected to be the Sulfotransfer_1 domain section (Table S1). The motif 1 and motif 6 possessed the critical N-terminal PSB loop and C-terminal PB region, respectively which are extremely momentous for the sulfotransferase activity of SOT proteins (Figure S1). As is demonstrated in Figure S1, the sequences related to these two important motifs are significantly conserved; it was reported that this high conservation can be found in both cytosolic and membrane sulfotransferases.

The N-terminal region 5' PSB in motif 1 are related to the PSB-loop and helix 3 sections in the sulfotransferase protein structure that encompasses five successive residues engaged in an interaction with the PAPS compound 5'-phosphate region. In this study, the amino acid residues in this motif that are engaged in sulfotransferase catalytic activity include completely conserved Lys-103 and relatively conserved Thr-106 that can be substituted by the functionally similar residues Ser and Cys (Figure 3a and Figure S1). Our results detected that genes within each subfamily have a significant similarity in exon/intron numbers. For instance, all the *StSOT* genes demonstrated an intronless structure except for *StSOT18*, *StSOT19*, *StSOT23* and *StSOT24* that contained two exons and one intron and classified into the phylogenetic clade II (Figure 3b).
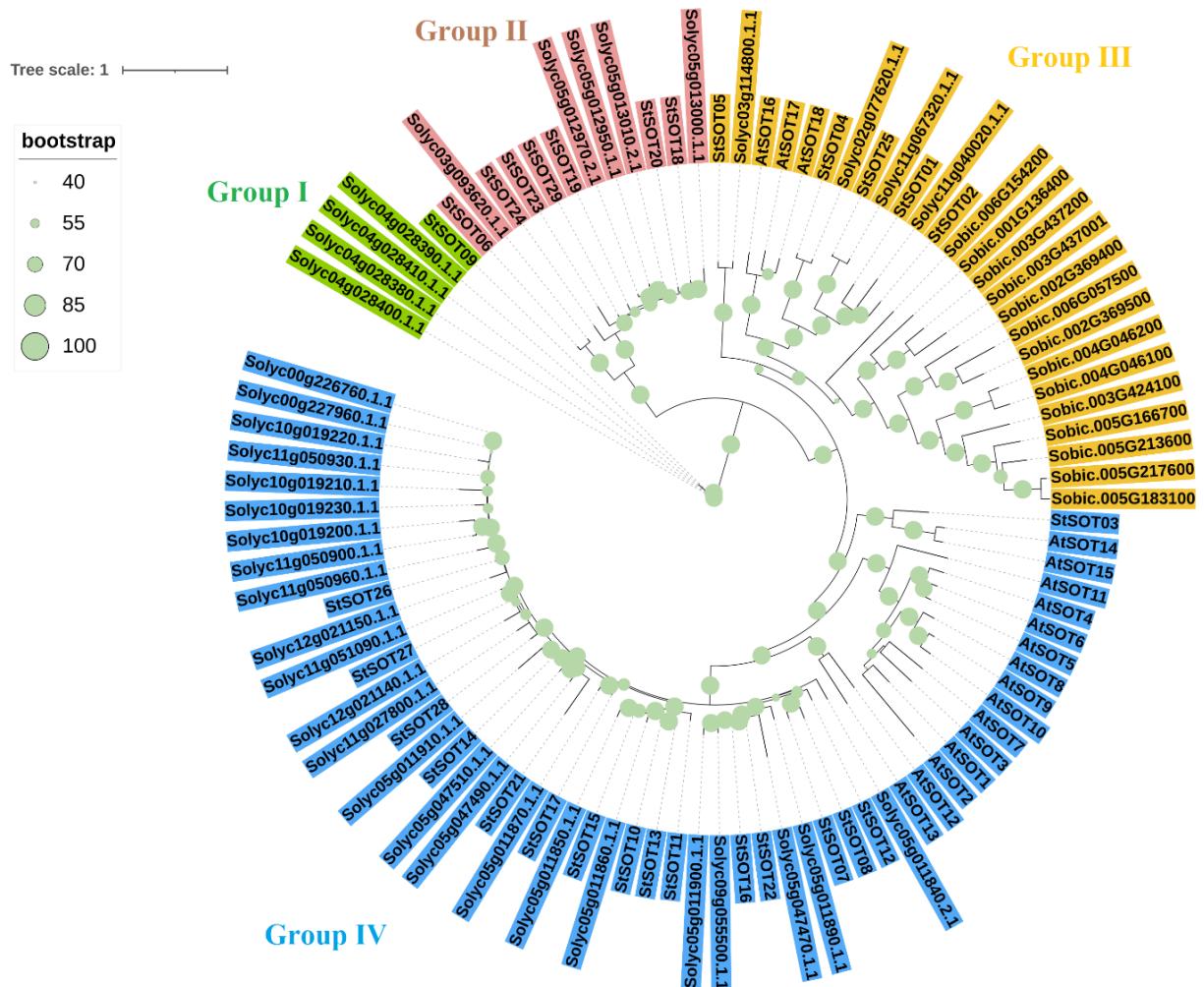
Figure 2. Phylogenetic relationships of SOT proteins from the potato, tomato, *Arabidopsis* and sorghum. The four main clusters have been detected based on the ML method in the phylogenetic tree. Symptoms: St: *Solanum tuberosum*; Solyc: *Solanum lycopersicum*; Sobic: *Sorghum bicolor*; At: *Arabidopsis thaliana*
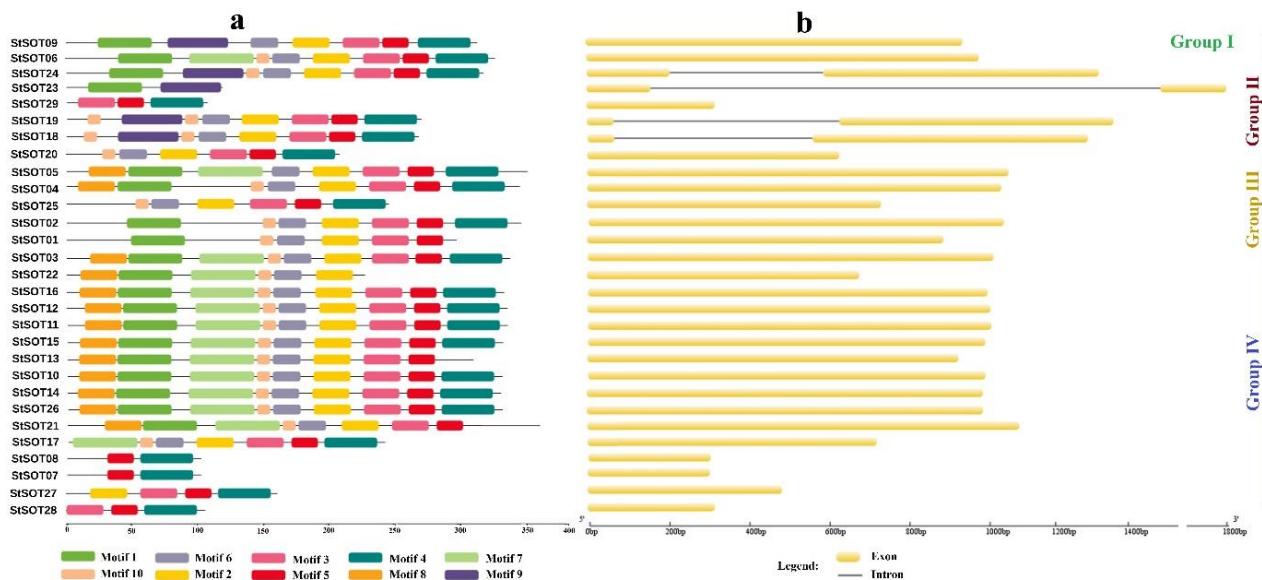

Figure 3. The conserved motifs predicted in the StSOT protein sequences (a), and exon-intron structure predicted in the *StSOT* family genes (b). Two important functional 5' PSB and 3' PB regions were detected in the motif 1 and motif 6, respectively.

2.3. *Genomic distribution, duplication assay and synteny relationships of* StSOT *genes*

All the *StSOT* gene family members were successfully mapped onto the 9 out of 12 chromosomes in the potato genome and the chromosomal map illustrated an unequal distribution of the gene family members throughout the chromosomes (Figure 4). The chromosome 5 encompassed the largest number of *StSOTs* with 13 genes, while only one *StSOT* was predicted to be localized on each chromosome 2, 4, 6 and 9. It was previously reported that many *SOT* genes in some plant species might be generated through gene duplication events. Nine segmentally-duplicated gene pairs categorized into five groups (including duplication and triplication events) were recognized in the *StSOT* gene family and each group has been demonstrated with different color, revealing the paralogous pairs (Figure 4). The highest numbers of duplicated/triplicated genes were distributed on the chromosome 5 with three duplications and three triplications clustered into the four gene groups (Table 2).

The intraspecies synteny results showed that many of the duplicated blocks were collinear, such as *StSOT07-StSOT08* and *StSOT26-StSOT27*. The *Ka/Ks* magnitudes related to the paralogous pairs covered a domain from 0.228 to 0.448 and according to these ratios the duplication events were estimated to be occurred between 0.461 to 5.769 million years ago (MYA). The Ka/Ks ratios <1 in duplicated gene pairs from StSOT family in potato suggested that the genes have been impressed by purifying selection (Table 2). The synteny analysis had been also investigated across the potato and some related plant genomes, which can detect the probable functions of the potato *StSOT* genes (Figure 5). According to the results, all the *StSOT* genes showed synteny relationships with their orthologs in the tomato (~35%) and *Arabidopsis* (~32%) genomes. The maximum orthology percentage of the *StSOT* on the potato genome was revealed with tomato. These wide synteny relations at gene level were considered as a confirmation for their close evolutionary relationships. These findings demonstrated the vast rearrangement events of the potato chromosomes during the genome evolution process.
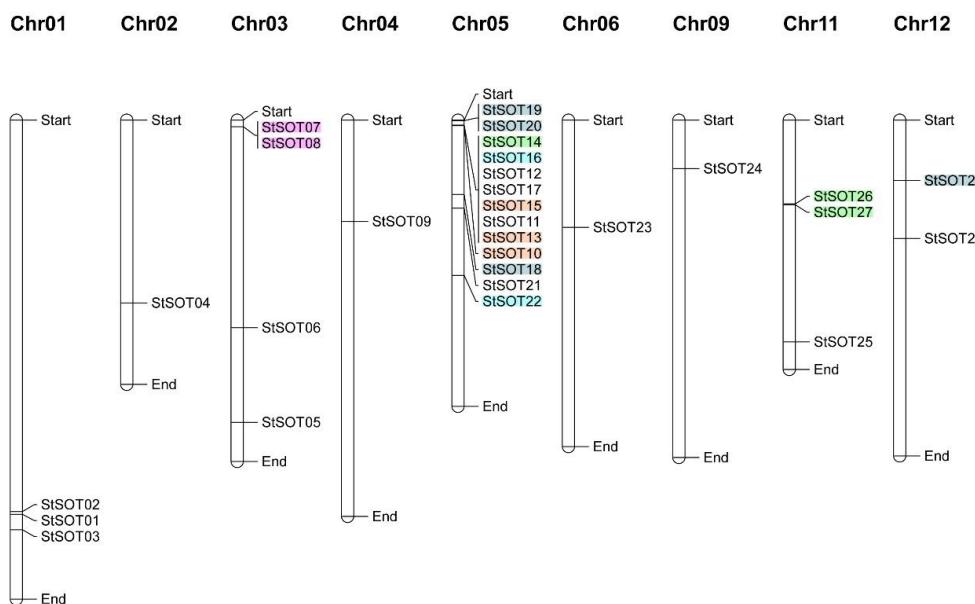

Figure 4. The chromosomal map of the *StSOT* family genes in the potato genome. Five series of duplicated/triplicated *StSOTs* were indicated in different colors

Table 2. The duplicated gene pairs in *StSOT* gene family and Ka/Ks analysis[a]. Multiple duplication/triplication events were detected in five categories (in different colors in the chromosomal map in Figure 4)

	Duplicated gene pairs	Duplication type	*Ka*	*Ks*	*Ka/Ks*	Date (million years ago)[a]
1	*StSOT07-StSOT08*	Segmental	0.0213	0.075	0.284	5.769
2	*StSOT10-StSOT13*	Segmental	0.003	0.006	0.448	0.461
	StSOT10-StSOT13-StSOT15		0.010	0.042	0.244	3.230
3	*StSOT26-StSOT27*	Segmental	0.014	0.057	0.254	4.384
	StSOT14-StSOT26-StSOT27		0.010	0.033	0.317	2.538
4	*StSOT16-StSOT22*	Segmental	0.015	0.063	0.252	4.846
5	*StSOT19-StSOT20*		0.016	0.029	0.544	2.230
	StSOT18-StSOT19-StSOT20	Segmental	0.010	0.045	0.228	3.461
	StSOT19- StSOT20-StSOT29		0.006	0.024	0.275	1.846

[a]The duplication and divergence time (million years ago) were computed based on the T= [Ks/2λ (λ = 6.5 × 10⁻⁹)] × 10⁻⁶ formula
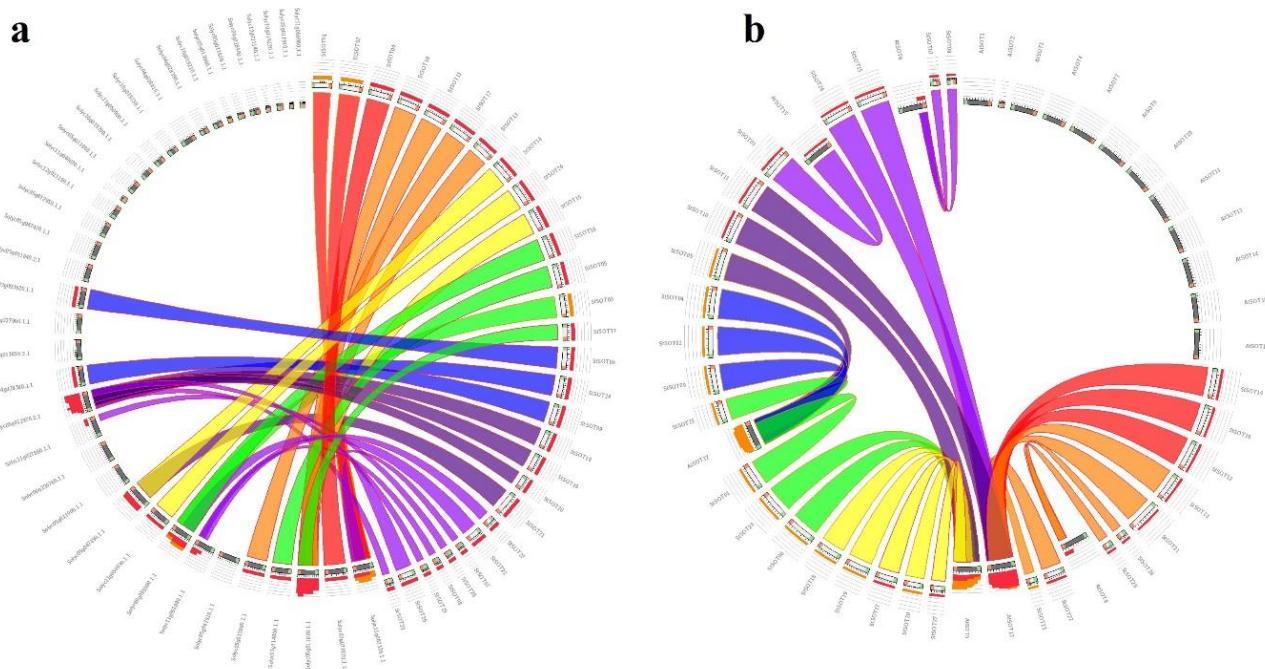

Figure 5. The synteny relationships of *StSOT* genes with the orthologs from a) tomato and b) *Arabidopsis*

2.4. Identification of the regulatory cis-elements in the StSOTs promoter

In the present study, the *StSOTs* promoter regions in the potato genome were investigated to find the putative *cis*-regulatory elements. The results demonstrated several kinds of cis-elements for dealing with various phytohormones and abiotic stimulus conditions (Table S2). The promoter common *cis*-elements, such as the core element TATA-box, CAAT-box and circadian control element were identified in all StSOT genes. The ABRE (abscisic acid responsiveness), ERE (ethylene responsiveness), and MeJA (Methyl jasmonate responsiveness) factors were predicted as the highly occurred hormone-responding *cis*-elements in approximately all *StSOT* promoters. The light

responsive G-Box and Box 4, wounding stress responsive WUN-motif, anaerobic inducible ARE, and stress responsive MYB elements were detected as the other regulatory *cis*-elements frequently distinguished in the *StSOTs* promoter areas, suggesting the important obligations of this gene family in stress dealing in potato crop. The TC-rich repeats (regulating defensive reactions), LTR (low-temperature responsive), TCA-element (salicylic acid responsive), TGA-element (auxin-responsive) and W-Box (WRKY transcription factors binding region, important for abiotic stimuli dealing) were detected as the momentous abiotic/hormone stress responsive elements significantly predicted in *StSOT08*, *StSOT11*, *StSOT13*, *StSOT16*, *StSOT22* and *StSOT26*. According to our results, multiple regulatory *cis*-elements related to phytohormones and environmental stimuli were discovered in the majority of *StSOT* genes, revealing the critical potential of these genes in potato growth and dealing with stress conditions.

2.5. Predicted miRNAs for StSOT genes

Six *StSOT* transcripts were predicted to be regulated by various miRNAs, for example, the transcripts *StSOT06*, *StSOT17*, *StSOT20* and *StSOT21* were targeted by the stu-miR8029, stu-miR8043, stu-miR8040-3p and stu-miR8051-3p, respectively (Table 3). Interestingly, four miRNAs, including stu-miR7993a-d, were predicted to target both of *StSOT11* and *StSOT15* transcripts for inhibiting with a translation (Table 3, Fig 6). Furthermore, the targeted regions of *StSOT*s by these miRNAs were predicted into the Sulfotransfer_1 domain region that can significantly indicate the *StSOT* genes regulation by the identified miRNAs. Regarding to the StSOTs phylogenetic relationships, the distinguished miRNAs remarkably targeted the *StSOT* genes in the clades IV, illustrating some important similarity in their cellular functions during the potato growth/development and degradation. Moreover, targeting of *StSOT* genes by various isoforms of a miRNA may demonstrate an important role of these genes during various biological processes in the cells, in addition to their sulfur assimilation activity.

Table 3. The predicted miRNA-targeted *StSOT* transcripts in the potato genome

| miRNA access.	Target gene	miRNA aligned fragment	Inhibition type
stu-miR8029	*StSOT06*	CGAGGUUUUGUUUCUUUUUACCGA	Translation
stu-miR7993a	*StSOT11*	UCAAUUCAAUUGGUGUAUUUUAUA	Translation
stu-miR7993b-3p	*StSOT11*	UCAAUUCAAUUGGUGUAUUUUAUA	Translation
stu-miR7993c	*StSOT11*	UCAAUUCAAUUGGUGUAUUUUAUA	Translation
stu-miR7993d	*StSOT11*	UCAAUUCAAUUGGUGUAUUUUAUA	Translation
stu-miR7993d	*StSOT15*	UCAAUUCAAUUGGUGUAUUUUAUA	Translation
stu-miR7993c	*StSOT15*	UCAAUUCAAUUGGUGUAUUUUAUA	Translation
stu-miR7993a	*StSOT15*	UCAAUUCAAUUGGUGUAUUUUAUA	Translation
stu-miR7993b-3p	*StSOT15*	UCAAUUCAAUUGGUGUAUUUUAUA	Translation
stu-miR8040-3p	*StSOT20*	CUAGUAUUAAUGUUAAUAUUC	Cleavage
stu-miR8043	*StSOT17*	CCGGUUUCAGGUUAAUAUAGU	Cleavage
stu-miR8051-3p	*StSOT21*	UUAUCAUACCAUCUUCUUUAU	Cleavage
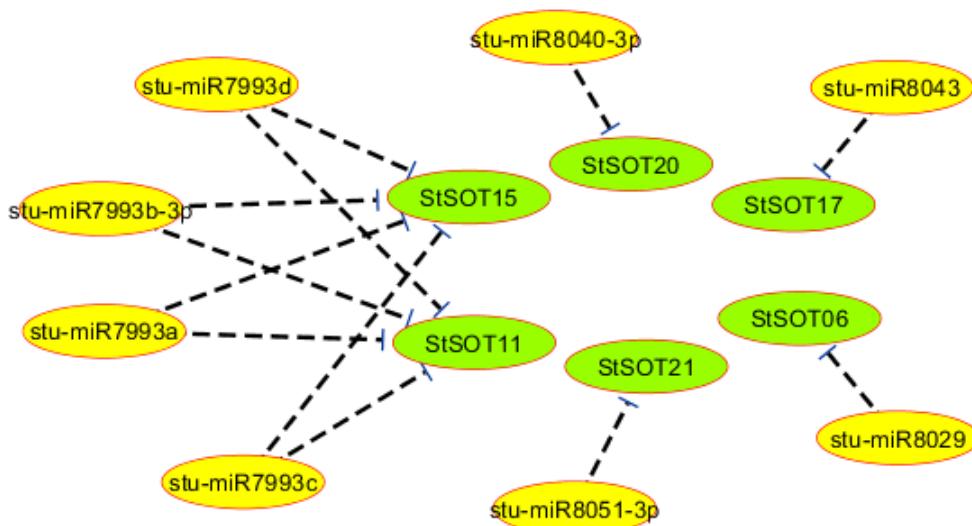
Figure 6. The interaction network between micro-RNAs and *StSOT* genes

2.6. Protein-protein interactions

The interactome data related to some important sulfotransferases in potato illustrated some sub-networks that in them SOT proteins interact with the genes engaged in transmembrane transporter, heme-binding, iron-sulfur cluster-binding and transition of phosphate group (Figure 7) (Table S3). The SOT16, SOT17 and SOT18, that were reported to regulate sulfur compounds and secondary metabolites biosynthetic processes, revealed a highly confident interaction network with a glucosyl transferase protein that contains transmembrane transporter activity and may respond to stimuli through the ion homeostasis. The APS (Pseudouridine synthase/archaeosine transglycosylase-like family protein), APR (Adenine phosphoribosyl reductase), APK (Adenylyl-sulfate kinase) and MET3-1 precorrin methyl transferase were detected as the other transferases contributing with StSOTs in sulfur compounds and secondary metabolites biosynthesis (Table S3), which can eventually adjust the potato growth and stimuli resistance. The interaction of StSOTs with the critical proteins Adenylyl-sulfate kinases can control sulfate assimilation and regulation of sulfur-containing amino acids metabolic process that are essential for plant reproduction and viability. The APR proteins in the network that constructed the significant relations with StSOTs can adjust the iron-sulfur complexes and reduces sulfate for Cys biosynthesis as well as can be induced by sulfate starvation. The annotation of the SUR, CYP and AKN proteins interacted with StSOTs demonstrated an important involvement of these collaborations in secondary metabolite biosynthetic process and sulfate assimilation, which significantly modulate plant growth and development and responses to diverse stimulus circumstances. Also, SIR protein was predicted to be engaged in metal ion transition and secondary metabolite biosynthetic process that both of them finally can regulate the potato cellular response to stresses and sulfate starvation (Table S3).
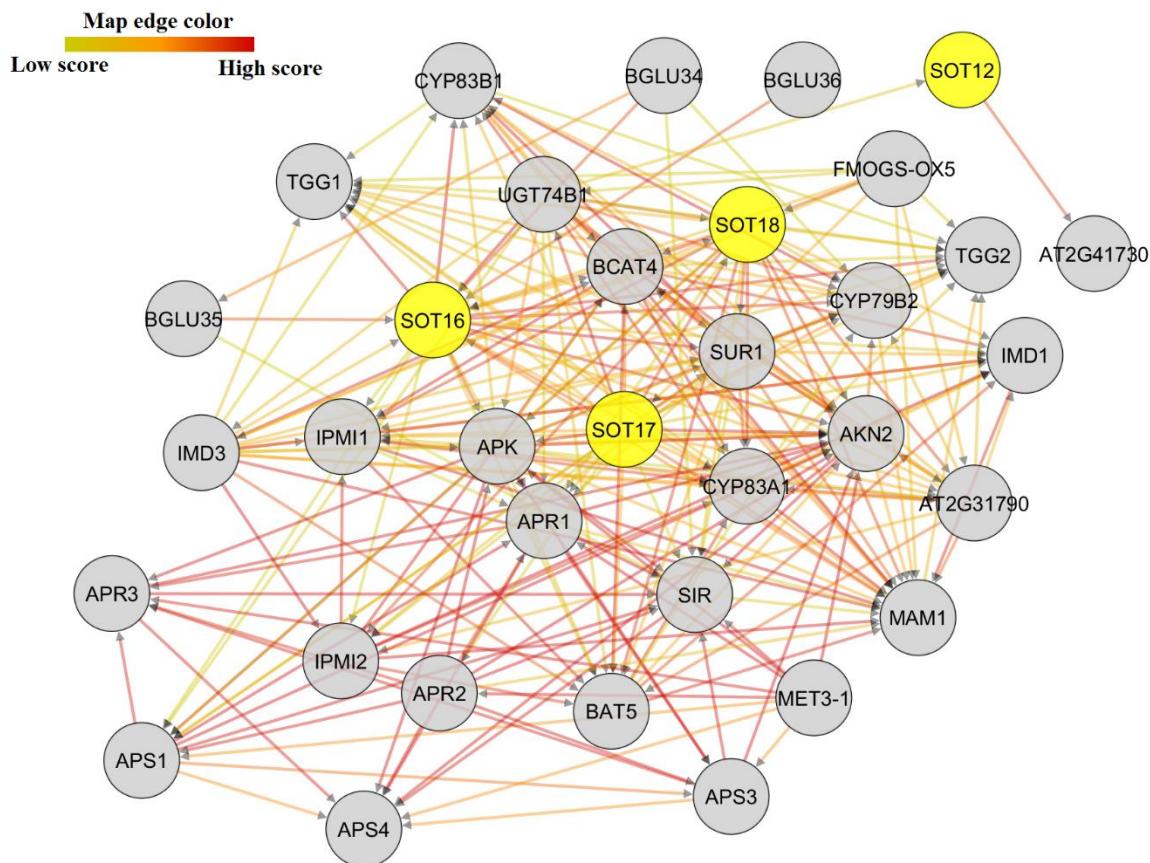
Figure 7. The protein-protein interaction network of SOT proteins using the Arabidopsis interactome data through STRING server v11 and improved using Cytoscape

2.7. Predicted 3D modeling, binding sites and validations of StSOT proteins

The 3D models of StSOT proteins were prepared through Phyre2 program under >90% confidence according to the templates 5mek, as a cytosolic sulfotransferase, and 1q44 and 1fmj, as the P-loop containing PAPS sulfotransferases in Arabidopsis. The 3D structure of StSOTs demonstrated the conserved typical frames comprising of β3-α8, as the PSB loop in the proteins 5' region and β8-α6, as the 3'PB motif (Figure 8 and S2). In models validation, the Ramachandran plot analysis demonstrated that the qualities of the StSOT proteins model varied from 80% to 95%, suggesting the good quality of the predicted 3D models and reliability of them (Table 4). To further verify, ProSA server was utilized for evaluation of the probable errors within the protein models, indicating the existence of the negative z-values in a conformation zone for the predicted models that can be experimentally distinguished through both of X-ray and NMR spectroscopy (Table 4). Regarding to ProSA results a remarkable rate of residues in each protein model was included in lowest energy regions, illustrating decreasing energies in various parts of these putative StSOT proteins .

The highest numbers of protein channels were predicted in StSOT05, StSOT06, StSOT11, StSOT12, StSOT13, StSOT16, StSOT17, StSOT19, StSOT20 and StSOT22 with 11 to 13 channel number (Table 4). Interestingly, some StSOT proteins with a considerable

similarity in their channel regions, such as StSOT05-StSOT06 and StSOT10-StSOT21, were also included in the same phylogenetic clade. Accordingly, it may suggest that the StSOTs evolutionary divergence can modulate the genes characteristics to function during various molecular pathways.

Table 4. Properties of the secondary and tertiary structures of StSOT proteins, their validation and channel numbers

Protein name	α-Helixes (%)		β-Sheets (%)		Coils (%)		Turns (%)		Channel number	Ramachandran plot (%)	z-values
StSOT01	132	(44%)	50	(16%)	114	(38%)	76	(25%)	7	93.50%	-8.4
StSOT02	161	(46%)	41	(11%)	143	(41%)	92	(26%)	9	93.90%	-8.73
StSOT03	141	(41%)	50	(14%)	146	(43%)	84	(24%)	8	90.10%	-8.15
StSOT04	148	(43%)	46	(13%)	150	(43%)	88	(25%)	7	93.90%	-8.15
StSOT05	142	(40%)	39	(11%)	169	(48%)	68	(19%)	12	92.80%	-8.61
StSOT06	152	(46%)	39	(11%)	135	(41%)	80	(24%)	12	94.10%	-8.16
StSOT07	47	(46%)	0	(0%)	54	(53%)	32	(31%)	5	90.90%	-1.85
StSOT08	50	(49%)	3	(2%)	48	(47%)	20	(19%)	4	92.90%	-2.01
StSOT09	148	(47%)	44	(14%)	121	(38%)	76	(24%)	10	93.20%	-8.71
StSOT10	151	(45%)	47	(14%)	132	(40%)	72	(21%)	10	94.20%	-8.45
StSOT11	140	(41%)	40	(11%)	155	(46%)	84	(25%)	11	94.00%	-8.52
StSOT12	146	(43%)	42	(12%)	147	(43%)	84	(25%)	12	92.50%	-8.66
StSOT13	120	(38%)	36	(11%)	152	(49%)	96	(31%)	13	81.70%	-7.64
StSOT14	145	(44%)	46	(13%)	138	(41%)	80	(24%)	5	94.50%	-8.6
StSOT15	152	(46%)	50	(15%)	128	(38%)	88	(26%)	3	95.10%	-7.93
StSOT16	148	(44%)	42	(12%)	142	(42%)	76	(22%)	12	91.50%	-9.05
StSOT17	115	(47%)	20	(8%)	105	(43%)	64	(26%)	12	95.40%	-6.17
StSOT18	128	(47%)	30	(11%)	111	(41%)	44	(16%)	7	93.60%	-7.99
StSOT19	132	(49%)	31	(11%)	105	(39%)	64	(23%)	11	95.90%	-7.92
StSOT20	103	(49%)	12	(5%)	94	(44%)	44	(21%)	12	94.20%	-6.67
StSOT21	143	(39%)	43	(11%)	173	(48%)	76	(21%)	10	91.30%	-7.93
StSOT22	94	(41%)	29	(12%)	103	(45%)	72	(31%)	13	79.90%	-5.5
StSOT23	37	(31%)	25	(21%)	56	(47%)	36	(30%)	5	92.20%	-4.01
StSOT24	146	(46%)	35	(11%)	135	(42%)	68	(21%)	9	89.80%	-8.12
StSOT25	113	(46%)	21	(8%)	110	(45%)	60	(24%)	5	93.00%	-5.86
StSOT26	154	(46%)	45	(13%)	130	(39%)	96	(29%)	6	93.30%	-8.77
StSOT27	83	(51%)	3	(1%)	74	(46%)	32	(20%)	4	94.30%	-4.56
StSOT28	49	(46%)	0	(0%)	56	(53%)	24	(22%)	5	80.60%	-3.48
StSOT29	49	(46%)	0	(0%)	57	(53%)	24	(22%)	5	94.20%	-2.78

The various numbers of ligand and ligand-binding amino acid residues were detected in the StSOT proteins structure (Table S4). Some metallic as well as non-metal heterogenes were predicted in the center of the binding region in all the candidate protein models (Figure 8). The Ser, Pro, Gly, Lys, Tyr and Arg amino acids were predicted as the binding residues in approximately all candidate StSOT proteins ligand-binding regions, which may manifest the importance of these residues in positioning on the DNA molecule

and finally the cellular function performance. The Ca, Zn and Mg ions were identified as the metallic heterogenes in the StSOTs functional domains. Based on our docking assay, although some binding residues were predicted outside of the specific domain, most of these functional regions were included in the Sulfotransfer_1 domain. The binding residues and their metallic/non-metalic interacting heterogenes demonstrated some variations that may illustrate some functional specificity of StSOT genes in addition to their common functions under stimuli exposure and in dealing to variations in cell metabolisms.
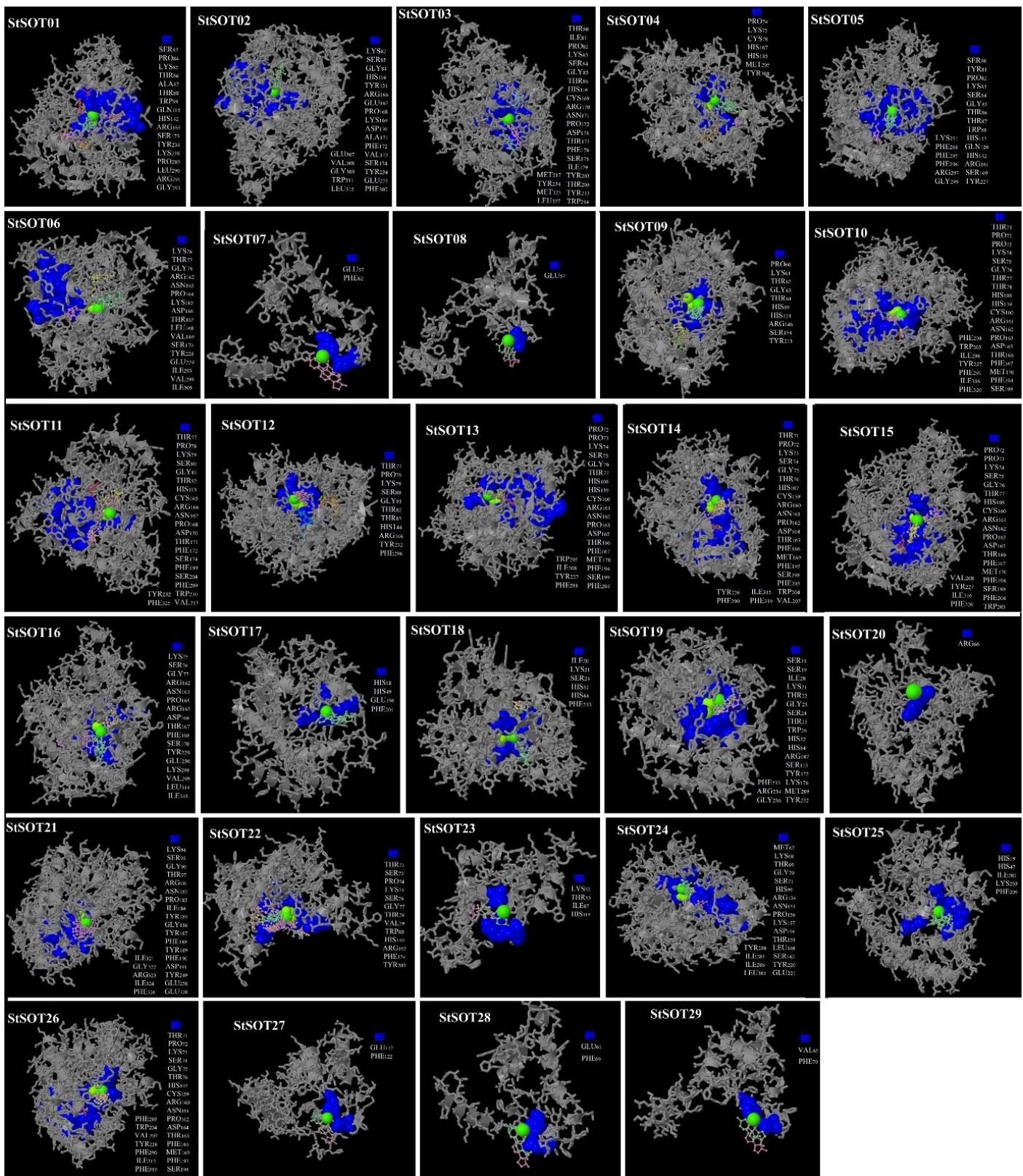
Figure 8. 3D docking analysis of StSOT proteins ligand-binding sites. The binding residues, metalicheterogen and non-metalicheterogen were detected in blue spacefill, green spacefill and colorful wireframe, respectively

2.8. Digital expression analyses of StSOT genes

The high throughput sequencing technologies, such as RNA-seq data obtained from the next generation sequencing approach, were introduced as the strong modern tools for genome studies that remarkably paved the way for the complete and simultaneous assay of a wide range of genomic transcripts and investigation of their features. In the present study, to comprehend the potato *SOT* genes functions the normalized FPKM magnitudes obtained from the RNA-seq data sets were employed to survey the mRNA transcription patterns of these genes in various tissues (Figure 9a). All of the *StSOT* family genes demonstrated the significant transcript magnitudes at least in one of the potato tissues, except for *StSOT29*, which may contain a regulation role in another cellular pathway. Some *StSOT*s, including *StSOT04*, *StSOT11*, *StSOT12*, *StSOT13*, *StSOT15*, *StSOT17,* and *StSOT24*, showed the substantial expression levels in all the potato candidate tissues, which can reveal the fundamental functions of these sulfotransferases during the potato growth and expansion. The developmental functions of these genes may be modulated via the ABRE/ERE hormones-related and light responsive Box 4 *cis*-elements present in these genes promoter regions (Table S2). Some of the *StSOT* genes also illustrated a tissue-specific expression pattern, as some instances, *StSOT09* and *StSOT25* had an approximately similar mRNA transcript rates only in the stem and tuber tissues, respectively. The sulfotransferase gene *StSOT27* was strongly expressed in the tuber pith and root tissues, while *StSOT28* had the remarkable FPKM values in the leaf and petiole samples. The other *StSOT*s also showed the various transcription levels in two, three or more tissues from potato, proposing the engagement of these sulfotransferases in a wide variety of cellular functions in these tissues during multiple developmental stages.

The expression patterns of the potato *SOT* family-related genes were also inspected under various biotic/abiotic and hormones exposure (Figure 9b). Among the biotic stimuli-induced *StSOTs* more induction responses were observed under BABA and phytophthora exposures with remarkable transcription rates in 19 and 14 *StSOT* genes, respectively (Figure 9b). Eight out of 29 StSOTs, such as the significantly induced *StSOT10*, *StSOT06*, *StSOT15* and *StSOT11* were also up-regulated in response to the BTH treatment. Amongst the biotic stresses-induced genes six *StSOT*s, including *StSOT05*, *StSOT06*, *StSOT12*, *StSOT21* and *StSOT25* revealed the significant mRNA transcription rates in coping with all stimuli treatments, suggesting their important engagements during biotic pathogens dealing in potato. Regarding to our results, 13, 9 and 7 *StSOT*s had been recognized as the highly expressed genes during abiotic stimulus NaCl, manitol and high temperature, respectively, which three *StSOT02*, *StSOT05* and *StSOT11* genes demonstrated the remarkable transcription rates in response to all abiotic stimuli (Figure 9b). In addition, in dealing with multiple hormones exposure approximately 59%, 55%, 34% and 24% of the *StSOT*s were substantially up-regulated in coping with the BAP, ABA, GA3 and IAA hormones, respectively. Based on our expression assay, *StSOT02* and *StSOT29* can be considered as the multiple hormones dealing sulfotransferases in the potato genome, because of their significant transcript magnitudes under all the candidate hormones situations. These remarkable transcription levels in different *StSOT*s can be significantly associated with their stress-coping *cis*-regulatory elements predicted in the promoter areas. Also, most of these high regulated *StSOT*s under these stimulus

circumstances demonstrated the significant involvement in the secondary metabolites biosynthetic process. These predictions can remarkably clarify the *StSOT* family-related genes critical roles in defensive responses of potato under various stimulus conditions and may ascertain the potential genes for further functional assay in order to improve potato and its relative crops endurance to various biotic/abiotic stresses.
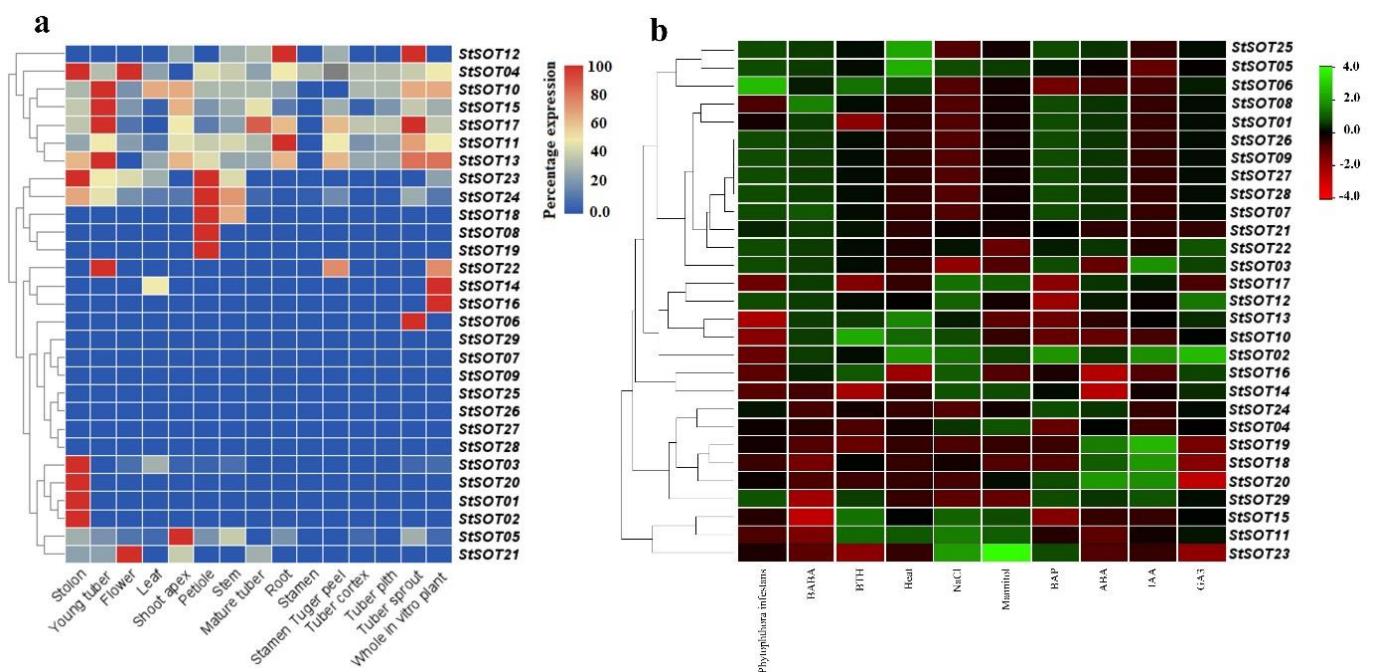

Figure 9. The tissue-specific (a) and stimuli-induced gene expression analysis (b) of the *StSOT* family members in the potato genome based on the RNA-seq data reported by the potato genome sequencing consortium

3. Discussion

The deduced amino acid sequence of sulfotransferase domain searched against the HMM program and Phytozome database led to the identification of 29 putative StSOT proteins, which revealed the extensive variations in the physicochemical properties that can demonstrate an effective role of the genomic duplication and integration events during the evolution of this gene family in potato. The majority of the identified StSOT proteins (~65.5%) illustrated an acidic nature; because of the theoretical isoelectric points (*pI*) lower than 7.0, suggesting the probable correlation of these StSOTs with the secretory pathways-related proteins. The significant diversity predicted in the *StSOT* genes features can refer to the potato genome evolutionary changes during the prior times. The StSOT proteins post-transcriptional phosphorylation analysis illustrated a wide variety of phosphorylated serine (S) residues along with some changed threonine (T) and tyrosine (Y) sites. Some StSOTs, such as StSOT01, StSOT04, StSOT06, StSOT12, StSOT14, StSOT22 and StSOT26, were predicted as the possible highly phosphorylated sulfotransferase proteins in potato. Protein phosphorylation can significantly adjust multiple biological processes such as plants development and stimuli dealing [24,25], suggesting the importance of these highly phosphorylated StSOTs during the potato life cycle. The

distinguished post-transcriptional phosphorylation changes were reported to illustrate the dynamic modulation of plants proteins [26].

According to the conserved motifs predicted in StSOT proteins, the N-terminal region 5' PSB in motif 1 is related to the PSB-loop and helix 3 sections in the sulfotransferase protein structure that encompasses five successive residues engaged in an interaction with the PAPS compound 5'-phosphate region   [27]. In this study, the amino acid residues in this motif that are engaged in sulfotransferase catalytic activity include completely conserved Lys-103 and relatively conserved Thr-106 that can be substituted by the functionally similar residues Ser and Cys (Figure 3 and Figure S1). The conserved 3' PB motif in the C-terminal part of the StSOTs encompassed β-sheet 8 and α-helix 6 that contained Arg-199 and Ser-207 as the interacting sides with the PAPS 3'-phosphate group and modulate its selectively binding [28]. Our findings indicated a remarkable structural similarity among these motifs as well as their fixed number of separating residues in all StSOT proteins, suggesting that *SOT* genes probably derived from a common ancestral gene. The similarities in the gene structures may also refer to a significant resemblance in expression patterns and regulatory functions in the cell [29]. Moreover, a highly similar distribution of exonic regions may refer to the evolutionary variations that were significantly occurred in the potato genome. The findings may propose that the exon/intron pattern can importantly prepare valuable insights into the evolutionary relationships amongst gene family members.

It was previously reported that many *SOT* genes in some plant species might be generated through gene duplication events. At least two whole genome duplication events have been also reported in the *S. tuberosum* genome [30,31], revealing a paleopolyploid origination for this important nutritional crop. Furthermore, the non-synonymous (*Ka*) and synonymous (*Ks*) substitution rates amongst the duplicated pairs can be considered as an important index to assay the selection pressure and approximate time related to the occurrence of duplications [32]. Because of the Ka/Ks ratios <1 in duplicated gene pairs from *StSOT* gene family in potato, it can be suggested that the genes have been impressed by purifying selection [33]. It was significantly suggested that the genes having conserved functions and/or pseudogenization may be generated via purifying selection [30]. Regarding to the predicted motifs in StSOT proteins, it was found that the genes within a duplicated gene group might be functionally conserved. It can be attributed to one of more periods of primeval polyploidy occurrence in multiple angiosperm plant lineages [31]. Therefore, these gene duplications in the potato genome can be propounded as a substantial potency for the evolutionary novelties appearance.

The wide synteny predicted amongst potato-tomato and potato-*Arabidopsis* at the gene levels may suggest their close evolutionary relationships. These revealed the chromosomal duplication and inversion rearrangement events, organizing the *SOT* genes in these genomes [34,35]. The results proposed that most of the *StSOT* genes shared a common ancestor and functions with the *SOT* counterparts from tomato and *Arabidopsis*. Despite these close evolutionary relations between potato and its relatives, some *SOT* genes from *Arabidopsis* and tomato were not mapped on any colinear blocks compared to potato genes. As a conceivable elucidation for this can refer to the rearrangements and fusions, which can extensively occur on the chromosomes in plants [36,37], which may

lead to the selective gene loss caused by the environmental situations [38]. The prepared information from the comparative synteny can propose a necessary prelude for comprehension of the evolution among crops.

The various stimuli responses are controlled via the genes transcriptional adjustment that can be modulated by *cis*-elements present in the genes promoter area [32,39]. According to our results, multiple regulatory *cis*-elements related to phytohormones and environmental stimuli were discovered in the majority of *StSOT* genes, revealing the critical potential of these genes in potato growth and dealing with stress conditions. Presence of the light responsive elements, especially G-Box, may illustrate that light signals can significantly adjust the *StSOT* genes transcription that eventually regulates the genes engaged in defensive lines such as flavonoid biosynthesis pathways [40,41]. The small non-coding RNAs, known as microRNAs (miRNA), also can be identified in most organisms including plants and animals, and are engaged in various cellular processes, such as stress responses, RNA silencing, protein degradation and genes post-transcriptional adjustment [42,43]. Due to the significant role of transcription factors and ion transferases in regulation of growth and stress coping in plants, these genes can be propounded as the important clades of miRNAs targets [39,40]. Therefore, potential miRNAs targeting six *StSOT* transcripts were considered to evaluate the post-transcriptional regulation of potato *SOT* genes. With the potato genome sequencing, miRNAs were proved to interact with multiple genes and contain an integral role in determining the tuberization rates [44]. Regarding to the StSOTs phylogenetic relationships, the distinguished miRNAs remarkably targeted the *StSOT* genes in the clades IV, illustrating some important similarity in their cellular functions during the potato growth/development and degradation. Moreover, targeting of *StSOT* genes by various isoforms of a miRNA may demonstrate an important role of these genes during various biological processes in the cells, in addition to their sulfur assimilation activity [1].

Protein-protein interactions can significantly modulate various cellular functions, such as replication and transcriptional adjustment of DNA, growth and development, signaling process and coordinating of multiple metabolic systems [45–47]. The engagement of StSOT proteins in adjustment of secondary metabolites biosynthetic process can detect their significant functions during the potato proper growth and tuberization as well as stress coping through the signaling pathways [45,46]. Moreover, our findings demonstrated the involvement of some StSOTs in hormones metabolic process that is importantly critical for the guard cells ABA responses and eventually plant resistance against various herbivores and pathogens. Therefore, it can be suggested that StSOT proteins significantly collaborated with the proteins from iron-sulfur complexes and amino acid metabolism which can regulate plant responses to external stimuli [41,45]. Moreover, collaboration of StSOTs with various development-related proteins can effectively module the potato growth and tuberization. As demonstrated in the *StSOT* genes interaction network, APS-kinase, protein phosphatases, ATP-sulfurylase, protein methyltransferase and NIR, can significantly modulate the defensive amino acids metabolic pathways in potato. It was reported that the amino acid catabolic system can adjust the seedlings tolerance against pathogenic infection importantly through the over-production of multiple toxic metabolites like serotonin [48]. The construction of these

defensive compounds as well as various sulfur-containing biologically active phytochemicals derived from amino acids, such as tryptophan, is associated with glutathione [48]. It was found that the glutathione and tryptophan metabolisms can be mentioned as two essential systems for plants hypersensitive immunity responses under various pathogens exposure [48,49]. Furthermore, our interaction network showed that biosynthesis of amino acid-derived compounds under stimuli situations is also regulated through the SOT-interacting genes, which were previously proved to be necessary for resistance to pathogens. Hence, these interacting proteins contain the indispensable roles during the life cycle of the potato cells and it can be declared that sulfotransferases have the dynamic gene network during the momentous metabolisms in plants species.

According the 3D structure of StSOTs, it can be suggested that the β-turn and random coil regions in the proteins structure are effective in tolerance to unfavorable circumstances [45,50]. Generally, our predicted 3D models were in good accordance with the parameters related to the typical SOT proteins and can be significantly utilized for the peptides ligand as well as docking assay. In the protein structures the channels and cavities are significantly engaged in protein functions adjustment and can detect their binding specificity [46,51]. The highest numbers of protein channels were predicted in StSOT05, StSOT06, StSOT11, StSOT12, StSOT13, StSOT16, StSOT17, StSOT19, StSOT20 and StSOT22 with 11 to 13 channel number (Table 4). It was suggested that the sulfotransferase proteins with similar structures in channels and cavity regions may also function similarly in the cells and during various environmental conditions [37,50][45,46]. Interestingly, some StSOT proteins with a considerable similarity in their channel regions, such as StSOT05-StSOT06 and StSOT10-StSOT21, were also included in the same phylogenetic clade. Accordingly, it may suggest that the StSOTs evolutionary divergence can modulate the genes characteristics to function during various molecular pathways. Based on our docking assay, although some binding residues were predicted outside of the specific domain, most of these functional regions were included in the Sulfotransfer_1 domain. The binding residues and their metallic/non-metalic interacting heterogenes demonstrated some variations that may illustrate some functional specificity of *StSOT* genes in addition to their common functions under stimuli exposure and in dealing to variations in cell metabolisms [29].

A wide range of studies have promoted the knowledge about plant growth and development through preparing some valuable evidence also regarding the roles of flavonoid and brassinosteroid metabolites in developmental processes [2]. Flavonoids, usually ascribed as the phytochemical secondary metabolites, and the significant steroid hormones brassinosteroids, can adjust various kinds of physiological processes in plants such as growth and enlargement as well as immunity systems through modulation of the various cells division, elongation and differentiation [52]. Based on promoter site analysis and expression profile of *StSOT* genes, it seems that *StSOT*s are involved in the potato growth and development and response to phytohormones such as brassinosteroids. According to the previously conducted studies, the induced mutations and disorders in the genes encoding for the main compounds constructing brassinosteroids and flavonoids disturbed the signaling systems that led to severe growth failure and devastated organ development, which eventually resulted in an impaired productivity and yielding in

plants [52]. The remarkable expression level of *StSOT01*, *StSOT3*, *StSOT21*, *StSOT26* and *StSOT28* in the potato leaf tissue also can be ascribed to the multiple light responsive G-Box and Box 4 *cis*-regulatory elements present in the promoter areas of these sulfotransferases that can collaborate with the flavonoid producer genes and eventually regulate growth process and tuberization in potato [40]. Also the presence of various hormones responsive elements in the multiple *StSOT*s may further justify these genes importance for the potato optimal growth/development and tuberizations [53]. Therefore, more functional investigation of *SOT* genes in potato may significantly influence the impressive production of some varieties with larger tubers and more nutritional values.

The remarkable transcription levels in different *StSOT*s can be significantly associated with their stress-coping *cis*-regulatory elements predicted in the promoter areas [54]. Also, most of these high regulated *StSOT*s under these stimulus circumstances demonstrated the significant involvement in the secondary metabolites biosynthetic process. Secondary metabolites are biologically active and genetically variable compounds found in various plant species that function as the momentous natural pesticides and can annihilate insect herbivores [45,46]. The strong defensive responses of *StSOT02*, *StSOT05* and *StSOT11* during the abiotic stress circumstances may be related to their regulatory functions during the secondary metabolites biosynthetic pathway and salicylic acid signaling [45,46]. Furthermore, the potato resistance mechanisms in coping with the multiple kinds of stimuli can be modulated through the interaction and co-expression relationships of sulfotransferases with the other stress-responsive genes. In this regard, a wide range of plant genes, that the most important of which are the interacting *APS-kinase*, *protein phosphatase*s, *APR* and *MET3-2* genes for sulfate assimilation and sulfur compound biosynthesis [45,46], as well as the co-expressed *methyltransferases* and *SIR* genes engaged in the stress-responsive amino acid metabolism, are involved in a significant collaboration with sulfotransferases in order to module the stress responses in potato. These predictions can remarkably clarify the *StSOT* family-related genes critical roles in defensive responses of potato under various stimulus conditions and may ascertain the potential genes for further functional assay in order to improve potato and its relative crops endurance to various biotic/abiotic stresses.

4. Materials and Methods

4.1. *Recognition of the StSOT family members*

The HMM profile related to the SOT domain (PF00685) was first retrieved through Pfam database (http://pfam.xfam.org) [12] and the HMM search (HMMER3.0) was conducted to detect the putative SOT proteins in the *S. tuberosum* genome, with an expected value of E-10. Also, the protein HMM profile was subjected to the Phytozome v12.1 database (https://phytozome.jgi.doe.gov/pz/portal.html) [55] in order to detect the SOT proteins in potato. The recognized non-redundant putative SOT proteins were manually checked for the SOT domain (PF00685) by employing Pfam. The corresponding cDNA and genomic sequences of the distinguished *SOTs* have been obtained from Phytozome and the genes nomenclature were prepared as *StSOT01* to *StSOT29* according to the genes order on the potato chromosomes.

The physicochemical properties of StSOT proteins including molecular weights, isoelectric points (pI) and amino acid compositions were distinguished through ProtParam

program (http://web.expasy.org/protparam) [56]. The possible transmembrane domains and post-transcriptional phosphorylation changes have been predicted in StSOTs using SCAMPI program (http://scampi.cbr.su.se/) [57] and NetPhos 3.1 server (http://www.cbs.dtu.dk/services/NetPhos/) [58], respectively. The StSOT proteins location in the cell were also detected via CELLO program (http://cello.life.nctu.edu.tw) [59].

4.2. *StSOT proteins alignment, phylogenetic relationships and identification of the conserved residues*

The sequence alignment of StSOT proteins was conducted by using T-COFFEE multiple sequence alignment package [60]. The phylogenetic relationships was assessed by constructing the maximum likelihood (ML) phylogenetic tree via MEGAX software according to the protein sequences of SOTs from potato, *Solanum lycopersicum*, *Sorghum bicolor* and *Arabidopsis thaliana* with 1000 bootstrap replicates [61]. The MEME (Multiple Em for Motif Elicitation) server (http://meme-suite.org/tools/meme) was also employed to discover the conserved protein motifs in StSOT members [62].

4.3. *StSOT genes structure and their chromosomal map*

The potato *StSOT* genes exon/intron organizations were predicted through Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/) [63]. The chromosomal localization of *StSOT* gene family members was also detected on the 12 chromosomes (Chr) of potato by using S. tuberosum genome info from the Potato Genome Sequencing Consortium database (PGSC) [31]. The MapChart software (https://www.wur.nl/en/show/Mapchart.htm) was eventually employed to generate a graphical chromosomal map for *StSOT* genes into the potato genome [64].

4.4. *Gene duplication and synonymous and non-synonymous substitution rates of StSOTs*

The identified *StSOT* genes were evaluated for the gene duplication events through alignment of their cDNA sequences by ClustalX v.21 program [65]. An identity matrix between the aligned CDSs was prepared and the duplicated gene pairs were detected as the genes sharing ≥90% identity in their nucleotide sequences. The duplicated *StSOT* gene pairs were subjected to a codon alignment by using ClustalW codon alignment tool in MEGA software and the synonymous (Ka) and non-synonymous (Ks) substitution values were estimated utilizing the Ka/Ks Calculator tool (http://code.google.com/p/kaks calculator/wiki/KaKs_Calculator) [33]. The time of duplication and divergence (million years ago) were also estimated through a synonymous mutation rate of λ substitutions per synonymous site per year as T= [Ks/2λ (λ = 6.5 × 10−9)] × 10-6 [66]. The comparative synteny relationships of *SOT* genes among the orthologous pairs between potato-tomato and potato-Arabidopsis at gene levels were finally visualized through Circos software [67].

4.5. *Promoter analysis, miRNA-targets, and proteins interaction assay*

The conserved cis-elements existing in the promoter area of *StSOT* genes were predicted by subjecting the 1500 bp upstream region of the start codon ATG in each putative *StSOTs* into the PlantCARE server (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [68]. The targeting miRNAs for the *StSOT* transcripts were identified by searching the genes coding sequences against the published miRNAs in the *S. tuberosum* genome in psRNATarget database [69] and finally visualized via Cytoscape [70]. The key *StSOTs* in the sulfotransferase family and sulfur compound and secondary metabolites biosynthetic processes were identified according to their gene ontology annotations and their protein-protein interaction network was eventually predicted via STRING v11 program (http://stringdb.org) [71].

4.6. *Protein 3D modeling, validation and docking analysis of the ligand site*

The three-dimensional structures of StSOT proteins were predicted through Protein Homology/analogy Recognition Engine V 2.0 (Phyre2) server [72]. The predicted protein models validation have been assessed through Ramachandran Plot Analysis [73] and Vadar server [74]. The protein secondary structures related to StSOTs were also distinguished by utilizing Vadar program. The proteins molecular voids and pocket/channel numbers were estimated via BetaCavity Web server [75]. The ProSA server was employed for calculation of errors and plots in proteins structure and validation of the 3D models [76]. Docking analysis of the ligand-binding regions in the predicted protein models was also performed via 3DLigandSite program [77].

4.7. *Digital expression profiling of StSOT genes*

The RNA-seq data published by the Potato Genome Sequencing Consortium [31] was employed for expression assay of the *StSOT* family members in multiple tissues and during various biotic/abiotic stimuli exposure. The biotic stimulus situations comprised of infection with *Phytophthora infestans*, DL-b-amino-n-butyricacid (BABA) and elicitors acibenzolar-S-methyl (BTH) in mixed samples after 24, 36 and 72 h exposure. The in vitro grown whole plants (after 24 h) were also subjected to three main abiotic stresses including heat (35℃), salinity (150 mM NaCl) and drought (mannitol 260 μM). Furthermore, the treatments with four significant hormones, including 6-benzylaminopurine (BAP; 10 μM), abscisic acid (ABA; 50 μM), indole-3-acetic acid (IAA; 10 μM) and gibberellic acid (GA3; 50 μM), were also considered for hormone stress-induced expression assay of *StSOT* genes. The expression levels of each *StSOT* gene in various tissues and multiple stimuli conditions were identified based on the transcripts ID search in the potato genome sequencing consortium RNA-seq dataset [31] and the transcripts magnitudes were determined in fragments per kilo base of exon model per million mapped reads (FPKM) and evaluated by using Cufflinks [78]. The heatmap related to the *StSOT* genes expression was then provided via Heatmapper program (http://www.heatmapper.ca/) [79].

5. Conclusions

The plant various primary metabolisms are reported to be depended on sulfate assimilation. The uptake of inorganic sulfate through the sulfate transporters existed in the plant cells plasma membranes is the first stage of plant sulfur metabolism. The transportation of this important sulfur element into the hydroxyl containing substrates has been identified as the sulfation reaction that is catalyzed by sulfotransferase genes. The *SOT* genes can regulate the plant stimuli responses, stress signaling mechanisms and developmental processes. The tuberization process in the significant nutritional crop potato can be disturbed by stimuli circumstances which may diminish the transportation of the photosynthetic products into the tubers, resulting in remarkable production damage. Regarding to the whole-genome sequencing of potato the comprehensive characterization study of the *SOT* gene family can provide valuable insights into the various developmental and resistance mechanisms that may lead to detect novel sulfotransferases as well as their interacting/co-expressed genes. In the present study, we demonstrated how this important crop effectively employ the numerous strategies like the secondary metabolites biosynthesis, sulfur compounds generations, transferase activity and iron-sulfur complexes production to modulate the various developmental and stimuli resistance processes. Our systematic study of *SOT* gene family can be useful for better understanding the complexity of these genes and support valuable insights into their regulation roles during growth and expansion as well as stimuli coping in the economically important crop species.

Supplementary Materials:

Figure S1. Multiple sequence alignments of the SOT family proteins in potato. The crucial 5' PSB loop and 3' PB regions required for sulfotransferase activity were indicated as black rectangles

Figure S2. The forecasted three-dimensional models of StSOT proteins in potato by using Phyre2 server

Table S1. The conserved motifs predicted in StSOT protein sequences

Table S2. The important cis-regulatory elements predicted in the promoter region of *StSOT* genes in Solanum tuberosum

Table S3. The interaction relationships between Sulfotransferases and the other genes during multiple cellular functions

Table S4. The docking analysis of the Ligand binding site present in the StSOT family proteins. The binding residues, metalic and non-metalic heterogenes were detected in blue spacefill, green spacefill and colorful wireframe, respectively, in the related Figure 8.

References

1.   Takahashi, H.; Buchner, P.; Yoshimoto, N.; Hawkesford, M.J.; Shiu, S.-H. Evolutionary relationships and functional diversity of plant sulfate transporters. *Front. Plant Sci.* **2012**, *2*, 119.

2.   Mazid, M.; Khan, T.A.; Mohammad, F. Role of secondary metabolites in defense mechanisms of plants. *Biol. Med.* **2011**, *3*, 232–249.

3.   Patron, N.J.; Durnford, D.G.; Kopriva, S. Sulfate assimilation in eukaryotes: fusions, relocations and lateral transfers. *BMC Evol. Biol.* **2008**, *8*, 1–14.

4.   d'Hooghe, P.; Dubousset, L.; Gallardo, K.; Kopriva, S.; Avice, J.-C.; Trouverie, J. Evidence for proteomic and metabolic adaptations associated with alterations of seed yield and quality in sulfur-limited Brassica napus L. *Mol. Cell. Proteomics* **2014**, *13*, 1165–1183.

5.   Rausch, T.; Wachter, A. Sulfur metabolism: a versatile platform for launching defence operations. *Trends Plant Sci.* **2005**, *10*, 503–509.

6.   Barczak, B.; Nowak, K. Effect of sulphur fertilisation on the content of macroelements and their ionic ratios in potato tubers. *J. Elem.* **2015**, *20*.

7.   Klikocka, H.; Haneklaus, S.; Bloem, E.; Schnug, E. Influence of sulfur fertilization on infection of potato tubers with Rhizoctonia solani and Streptomyces scabies. *J. Plant Nutr.* **2005**, *28*, 819–833.

8.   Gupta, U.C.; Sanderson, J.B. Effect of sulfur, calcium, and boron on tissue nutrient concentration and potato yield. *J. Plant Nutr.* **1993**, *16*, 1013–1023.

9.   De Kok, L.J.; Castro, A.; Durenkamp, M.; Koralewska, A.; Posthumus, F.S.; Stuiver, C.E.E.; Yang, L.; Stulen, I. Pathways of plant sulfur uptake and metabolism-an overview. *Landbauforsch. Völkenrode, Spec. Issue* **2005**, *283*, 5–13.

10.   Norici, A.; Hell, R.; Giordano, M. Sulfur and primary production in aquatic environments: an ecological perspective. *Photosynth. Res.* **2005**, *86*, 409–417.

11.   Klein, M.; Papenbrock, J. The multi-protein family of Arabidopsis sulphotransferases and their relatives in other plant species. *J. Exp. Bot.* **2004**, *55*, 1809–1820.

12.   Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J. Pfam: the protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230.

13.   Varin, L.; DeLuca, V.; Ibrahim, R.K.; Brisson, N. Molecular characterization of two plant flavonol sulfotransferases. *Proc. Natl. Acad. Sci.* **1992**, *89*, 1286–1290.

14.   Klein, M.; Reichelt, M.; Gershenzon, J.; Papenbrock, J. The three desulfoglucosinolate sulfotransferase proteins in Arabidopsis have different substrate specificities and are differentially expressed. *FEBS J.* **2006**, *273*, 122–136.

15.   Komori, R.; Amano, Y.; Ogawa-Ohnishi, M.; Matsubayashi, Y. Identification of tyrosylprotein sulfotransferase in Arabidopsis. *Proc. Natl. Acad. Sci.* **2009**, *106*, 15067–15072.

16.   Marsolais, F.; Sebastià, C.H.; Rousseau, A.; Varin, L. Molecular and biochemical characterization of BNST4, an ethanol-inducible steroid sulfotransferase from Brassica napus, and regulation of BNST genes by chemical stress and during development. *Plant Sci.* **2004**, *166*, 1359–1370.

17.   Chen, R.; Jiang, Y.; Dong, J.; Zhang, X.; Xiao, H.; Xu, Z.; Gao, X. Genome-wide analysis and environmental response profiling of SOT family genes in rice (Oryza sativa). *Genes Genomics* **2012**, *34*, 549–560.

18.   Zang, Y.; Kim, H.U.; Kim, J.A.; Lim, M.; Jin, M.; Lee, S.C.; Kwon, S.; Lee, S.; Hong, J.K.; Park, T. Genome-wide identification of glucosinolate synthesis genes in Brassica rapa. *FEBS J.* **2009**, *276*, 3559–3574.

19.   Baek, D.; Pathange, P.; CHUNG, J.; Jiang, J.; Gao, L.; Oikawa, A.; Hirai, M.Y.; Saito, K.; Pare, P.W.; Shi, H. A stress-inducible sulphotransferase sulphonates salicylic acid and confers pathogen resistance in Arabidopsis. *Plant. Cell Environ.* **2010**, *33*, 1383–1392.

20.   Lacomme, C.; Roby, D. Molecular cloning of a sulfotransferase in Arabidopsis thaliana and regulation during development and in response to infection with pathogenic bacteria. *Plant Mol. Biol.* **1996**, *30*, 995–1008.

21.   Rouleau, M.; Marsolais, F.; Richard, M.; Nicolle, L.; Voigt, B.; Adam, G.; Varin, L. Inactivation of brassinosteroid biological activity by a salicylate-inducible steroid sulfotransferase from Brassica napus. *J. Biol. Chem.* **1999**, *274*, 20925–20930.

22.   Bednarek, P. Sulfur-containing secondary metabolites from Arabidopsis thaliana and other Brassicaceae with function in plant immunity. *ChemBioChem* **2012**, *13*, 1846.

23.   Hopkins, L.; Parmar, S.; Błaszczyk, A.; Hesse, H.; Hoefgen, R.; Hawkesford, M.J. O-acetylserine and the regulation of expression of genes encoding components for sulfate uptake and assimilation in potato. *Plant Physiol.* **2005**, *138*, 433–440.

24.   Heidari, P.; Ahmadizadeh, M.; Izanlo, F.; Nussbaumer, T. In silico study of the CESA and CSL gene family in Arabidopsis thaliana and Oryza sativa: Focus on post-translation modifications. *Plant Gene* **2019**, *19*.

25.   Rezaee, S.; Ahmadizadeh, M.; Heidari, P. Genome-wide characterization, expression profiling, and post- transcriptional study of GASA gene family. *Gene Reports* **2020**, *20*, 100795.

26.   Faraji, S.; Hasanzadeh, S.; Heidari, P. Comparative in silico analysis of Phosphate transporter gene family, PHT, in Camelina sativa gemome. *Gene Reports* **2021**, 101351.

27.   Hell, R.; Dahl, C.; Knaff, D.; Leustek, T. Sulfur metabolism in phototrophic organisms. **2008**.

28.   Klaassen, C.D.; Boles, J.W. The importance of 3 '-phosphoadenosine 5 '-phosphosulfate (PAPS) in the regulation of sulfation. *FASEB J.* **1997**, *11*, 404–418.

29.   Kakuta, Y.; Pedersen, L.G.; Pedersen, L.C.; Negishi, M. Conserved structural motifs in the sulfotransferase family. *Trends Biochem. Sci.* **1998**, *23*, 129–130.

30.   Visser, R.G.F.; Bachem, C.W.B.; de Boer, J.M.; Bryan, G.J.; Chakrabati, S.K.; Feingold, S.; Gromadka, R.; van Ham, R.C.H.J.; Huang, S.; Jacobs, J.M.E. Sequencing the potato genome: outline and first results to come from the elucidation of the sequence of the world's third most important food crop. *Am. J. Potato Res.* **2009**, *86*, 417–429.

31.   Diambra, L.A. Genome sequence and analysis of the tuber crop potato. *Nature* **2011**, *475*.

32.   Sheshadri, S.A.; Nishanth, M.J.; Simon, B. Stress-mediated cis-element transcription factor interactions interconnecting primary and specialized metabolism in planta. *Front. Plant Sci.* **2016**, *7*, 1725.

33.   Zhang, Z.; Li, J.; Zhao, X.-Q.; Wang, J.; Wong, G.K.-S.; Yu, J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics. Proteomics Bioinformatics* **2006**, *4*, 259–263.

34.   Abdullah; Faraji, S.; Mehmood, F.; Malik, H.M.T.; Ahmed, I.; Heidari, P.; Poczai, P. The GASA Gene Family in Theobroma cacao: Genome Wide Identification and Expression Analyses. *Agronomy* **2021**, *11*.

35.   Heidari, P.; Faraji, S.; Ahmadizadeh, M.; Ahmar, S.; Mora-Poblete, F. New insights into structure and function of TIFY genes

in Zea mays and Solanum lycopersicum: a genome-wide comprehensive analysis. *Front. Genet.* **2021**, *12*, 534.

36.   Fujii, S.; Kazama, T.; Yamada, M.; Toriyama, K. Discovery of global genomic re-organization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes. *BMC Genomics* **2010**, *11*, 1–15.

37.   Musavizadeh, Z.; Najafi-Zarrini, H.; Kazemitabar, S.K.; Hashemi, S.H.; Faraji, S.; Barcaccia, G.; Heidari, P. Genome-Wide Analysis of Potassium Channel Genes in Rice: Expression of the OsAKT and OsKAT Genes under Salt Stress. *Genes (Basel).* **2021**, *12*, 784.

38.   Xuan, Y.H.; Piao, H.L.; Je, B. Il; Park, S.J.; Park, S.H.; Huang, J.; Zhang, J.B.; Peterson, T.; Han, C. Transposon Ac/Ds-induced chromosomal rearrangements at the rice OsRLG5 locus. *Nucleic Acids Res.* **2011**, *39*, e149–e149.

39.   Ahmadizadeh, M.; Chen, J.-T.; Hasanzadeh, S.; Ahmar, S.; Heidari, P. Insights into the genes involved in the ethylene biosynthesis pathway in Arabidopsis thaliana and Oryza sativa. *J. Genet. Eng. Biotechnol.* **2020**, *18*, 1–20.

40.   Biłas, R.; Szafran, K.; Hnatuszko-Konka, K.; Kononowicz, A.K. Cis-regulatory elements used to control gene expression in plants. *Plant Cell, Tissue Organ Cult.* **2016**, *127*, 269–287.

41.   Faraji, S.; Ahmadizadeh, M.; Heidari, P. Genome-wide comparative analysis of Mg transporter gene family between Triticum turgidum and Camelina sativa. *BioMetals* **2021**, *4*.

42.   Cui, Q.; Yu, Z.; Purisima, E.O.; Wang, E. Principles of microRNA regulation of a human cellular signaling network. *Mol. Syst. Biol.* **2006**, *2*, 46.

43.   Heidari, P.; Mazloomi, F.; Nussbaumer, T.; Barcaccia, G. Insights into the SAM Synthetase Gene Family and Its Roles in Tomato Seedlings under Abiotic Stresses and Hormone Treatments. *Plants* **2020**, *9*, 586.

44.   Amrutha, R.N.; Sekhar, P.N.; Varshney, R.K.; Kishor, P.B.K. Genome-wide analysis and identification of genes related to potassium transporter families in rice (Oryza sativa L.). *Plant Sci.* **2007**, *172*, 708–721.

45.   Braun, P.; Aubourg, S.; Van Leene, J.; De Jaeger, G.; Lurin, C. Plant protein interactomes. *Annu. Rev. Plant Biol.* **2013**, *64*, 161–187.

46.   Fukao, Y. Protein-protein interactions in plants. *Plant Cell Physiol.* **2012**, *53*, 617–625.

47.   Kazemi, E.; Zargooshi, J.; Kaboudi, M.; Heidari, P.; Kahrizi, D.; Mahaki, B.; Mohammadian, Y.; Khazaei, H.; Ahmed, K. A genome-wide association study to identify candidate genes for erectile dysfunction. *Brief. Bioinform.* **2021**, *22*, bbaa338.

48.   Hiruma, K.; Fukunaga, S.; Bednarek, P.; Piślewska-Bednarek, M.; Watanabe, S.; Narusaka, Y.; Shirasu, K.; Takano, Y. Glutathione and tryptophan metabolism are required for Arabidopsis immunity during the hypersensitive response to hemibiotrophs. *Proc. Natl. Acad. Sci.* **2013**, *110*, 9589–9594.

49.   Ishihara, A.; Hashimoto, Y.; Tanaka, C.; Dubouzet, J.G.; Nakao, T.; Matsuda, F.; Nishioka, T.; Miyagawa, H.; Wakasa, K. The tryptophan pathway is involved in the defense responses of rice against pathogenic infection via serotonin production. *Plant J.* **2008**, *54*, 481–495.

50.   Faraji, S.; Filiz, E.; Kazemitabar, S.K.; Vannozzi, A.; Palumbo, F.; Barcaccia, G.; Heidari, P. The AP2/ERF Gene Family in Triticum durum: Genome-Wide Identification and Expression Analysis under Drought and Salinity Stresses. *Genes (Basel).* **2020**, *11*, 1464.

51.   Heidari, P.; Abdullah; Faraji, S.; Poczai, P. Magnesium transporter Gene Family: Genome-Wide Identification and Characterization in Theobroma cacao, Corchorus capsularis and Gossypium hirsutum of Family Malvaceae. *Agronomy* **2021**, *11*, 1651.

52.   Jain, M. Next-generation sequencing technologies for gene expression profiling in plants. *Brief. Funct. Genomics* **2012**, *11*, 63–70.

53.   Ghelis, T. Signal processing by protein tyrosine phosphorylation in plants. *Plant Signal. Behav.* **2011**, *6*, 942–951.

54.   Ahmadizadeh, M.; Heidari, P. Bioinformatics study of transcription factors involved in cold stress. *Biharean Biol.* **2014**, *8*.

55.   Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186.

56.   Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and

analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*; Humana Press: Totowa, NJ, 2005; pp. 571–607.

57.   Bernsel, A.; Viklund, H.; Falk, J.; Lindahl, E.; von Heijne, G.; Elofsson, A. Prediction of membrane-protein topology from first principles. *Proc. Natl. Acad. Sci.* **2008**, *105*, 7177–7181.

58.   Blom, N.; Sicheritz-Pontén, T.; Gupta, R.; Gammeltoft, S.; Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **2004**, *4*, 1633–1649.

59.   Yu, C.-S.; Cheng, C.-W.; Su, W.-C.; Chang, K.-C.; Huang, S.-W.; Hwang, J.-K.; Lu, C.-H. CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. *PLoS One* **2014**, *9*, e99368.

60.   Notredame, C.; Higgins, D.G.; Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000**, *302*, 205–217.

61.   Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549.

62.   Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208.

63.   Hu, B.; Jin, J.; Guo, A.-Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297.

64.   Voorrips, R.E. MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78.

65.   Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R. Clustal W and Clustal X version 2.0. *bioinformatics* **2007**, *23*, 2947–2948.

66.   Yang, Z.; Gu, S.; Wang, X.; Li, W.; Tang, Z.; Xu, C. Molecular evolution of the CPP-like gene family in plants: insights from comparative genomics of Arabidopsis and rice. *J. Mol. Evol.* **2008**, *67*, 266–277.

67.   Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: an information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645.

68.   Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327.

69.   Dai, X.; Zhuang, Z.; Zhao, P.X. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **2018**, *46*, W49–W54.

70.   Franz, M.; Lopes, C.T.; Huck, G.; Dong, Y.; Sumer, O.; Bader, G.D. Cytoscape. js: a graph theory library for visualisation and analysis. *Bioinformatics* **2016**, *32*, 309–311.

71.   Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613.

72.   Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858.

73.   Lovell, S.C.; Davis, I.W.; Arendall III, W.B.; De Bakker, P.I.W.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins Struct. Funct. Bioinforma.* **2003**, *50*, 437–450.

74.   Willard, L.; Ranjan, A.; Zhang, H.; Monzavi, H.; Boyko, R.F.; Sykes, B.D.; Wishart, D.S. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **2003**, *31*, 3316–3319.

75.   Kim, J.-K.; Cho, Y.; Lee, M.; Laskowski, R.A.; Ryu, S.E.; Sugihara, K.; Kim, D.-S. BetaCavityWeb: a webserver for molecular voids and channels. *Nucleic Acids Res.* **2015**, *43*, W413–W418.

76.   Wiederstein, M.; Sippl, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, *35*, W407–W410.

77.   Wass, M.N.; Kelley, L.A.; Sternberg, M.J.E. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res.* **2010**, *38*, W469–W473.

78.   Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; Van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515.

79.   Babicki, S.; Arndt, D.; Marcu, A.; Liang, Y.; Grant, J.R.; Maciejewski, A.; Wishart, D.S. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* **2016**, *44*, W147–W153.