



# Insights into a key sulfite scavenger enzyme sulfite oxidase (SOX) gene in plants

Ertugrul Filiz<sup>1</sup> · Recep Vatansever<sup>2</sup> · Ibrahim Ilker Ozyigit<sup>2</sup>

Received: 28 September 2016 / Revised: 7 March 2017 / Accepted: 16 March 2017 / Published online: 22 March 2017  
© Prof. H.S. Srivastava Foundation for Science and Society 2017

**Abstract** Sulfite oxidase (SOX) is a crucial molybdenum cofactor-containing enzyme in plants that re-oxidizes the sulfite back to sulfate in sulfite assimilation pathway. However, studies of this crucial enzyme are quite limited hence this work was attempted to understand the SOXs in four plant species namely, *Arabidopsis thaliana*, *Solanum lycopersicum*, *Populus trichocarpa* and *Brachypodium distachyon*. Herein studied SOX enzyme was characterized with both oxidoreductase molybdopterin binding and Mo-co oxidoreductase dimerization domains. The alignment and motif analyses revealed the highly conserved primary structure of SOXs. The phylogeny constructed with additional species demonstrated a clear divergence of monocots, dicots and lower plants. In addition, to further understand the phylogenetic relationship and make a functional inference, a structure-based phylogeny was constructed using normalized RMSD values in five superposed models from four modelled plant SOXs herein and one previously characterized chicken SOX structure. The plant and animal SOXs showed a clear divergence and also implicated their functional divergences. Based on tree topology, monocot *B. distachyon* appeared to be diverged from other dicots, pointing out a possible monocot–dicot

split. The expression patterns of sulfite scavengers including SOX were differentially modulated under cold, heat, salt and high light stresses. Particularly, they tend to be up-regulated under high light and heat while being down-regulated under cold and salt stresses. The presence of *cis*-regulatory motifs associated with different stresses in upstream regions of SOX genes was thus justified. The protein–protein interaction network of AtSOX and network enrichment with gene ontology (GO) terms showed that most predicted proteins, including sulfite reductase, ATP sulfurylases and APS reductases were among prime enzymes involved in sulfite pathway. Finally, SOX–sulfite docked structures indicated that arginine residues particularly Arg374 is crucial for SOX–sulfite binding and additional two other residues such as Arg51 and Arg103 may be important for SOX–sulfite bindings in plants.

**Keywords** Macroelement · Sulfite · *cis*-Element · Stress · Modeling · PPI network · Docking

## Introduction

Sulfur (S) is an important macroelement in plants and plays very crucial roles in plant growth and productivity. Many critical oligopeptides are produced using S containing components such as glutathione (GSH), phytochelatins, prosthetic groups, vitamins and cofactors (biotin, molybdenum cofactor, thiamine, coenzyme A, and S-adenosyl-methionine), secondary metabolites and lipids (Saito 2000; Leustek 2002). Sulfur is usually available in nature as oxidized anion sulfate ( $\text{SO}_4^{2-}$ ) which is reduced to sulfite ( $\text{SO}_3^{2-}$ ) and incorporated into cysteine and methionine. In sulfite turnover for incorporation into metabolites and detoxification, its usage is performed by four major

**Electronic supplementary material** The online version of this article (doi:10.1007/s12298-017-0433-z) contains supplementary material, which is available to authorized users.

✉ Ertugrul Filiz  
ertugrulfiliz@gmail.com

<sup>1</sup> Department of Crop and Animal Production, Cilimli Vocational School, Duzce University, 81750 Cilimli, Duzce, Turkey

<sup>2</sup> Department of Biology, Faculty of Science and Arts, Marmara University, 34722 Goztepe, Istanbul, Turkey

enzymes (Fig. 1) such as chloroplast-localized sulfite reductase (SiR; EC 1.8.7.1) (Khan et al. 2010), chloroplast-localized UDP-sulfoquinovose synthase (SQD1; EC 3.13.1.1; Sanda et al. 2001), thiosulfate sulfurtransferase (EC 2.8.1.1) (Papenbrock and Schmidt 2000) and sulfite oxidase (SOX; EC 1.8.3.1) (Eilers et al. 2001). Of these, SOX is the molybdenum cofactor-containing enzyme located in the peroxisomes and catalyzes the reaction of  $\text{SO}_3^{2-} + \text{H}_2\text{O} \rightarrow \text{SO}_4^{2-} + 2\text{H}^+ + 2\text{e}^-$ , which is the crucial step for degradation of cysteine, methionine and membrane components such as sulfatides (Eilers et al. 2001; Nowak et al. 2004).

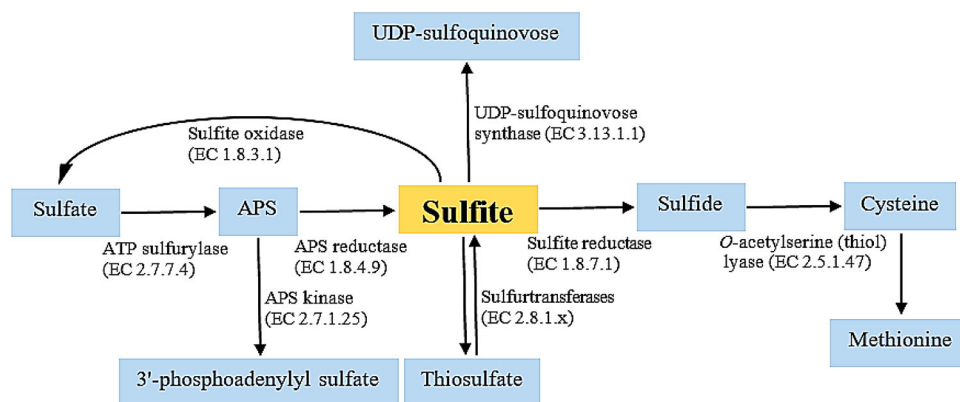
A crystallographic structure from *Arabidopsis* sulfite oxidase showed that enzyme includes only a single Moco domain without an additional redox center. In addition, plant SOXs were reported to be homodimeric and homologous to animal SOXs (Schrader et al. 2003). In *Arabidopsis*, an in vitro study of enzymatic and non-enzymatic reaction between catalase and sulfite was reported to occur in peroxisome (Hänsch et al. 2006). In *Arabidopsis*, SOX was constitutively expressed but not significantly induced by  $\text{SO}_2$  application. However, a chloroplast-localized enzyme sulfite reductase was induced quickly by  $\text{SO}_2$  exposure. In addition, other sulfite-requiring enzymes like mercaptopyruvate sulfur transferases and UDP-sulfoquinovose synthase 1 were induced subsequently (Brychkova et al. 2007). In tomato, SOX was upregulated in plant leaves and other sulfite network components were down-regulated under extended dark condition (Brychkova et al. 2013). Herein, taking into account the literature limitation about this crucial enzyme, this study has aimed to further improve our understanding of SOX orthologs in tree dicots of *Arabidopsis thaliana* (model for flowering plants), *Solanum lycopersicum* (important crop) and *Populus*

*trichocarpa* (model for woods), and in one monocot species *Brachypodium distachyon* (model for grasses).

## Materials and methods

### Sequence analysis

Studied SOX sequences were retrieved from UniProtKB/Swiss-Prot database under NCBI ([ncbi.nlm.nih.gov/](http://ncbi.nlm.nih.gov/); Romiti 2010). Physico-chemical features of SOX proteins were analyzed using ProtParam tool ([web.expasy.org/protparam/](http://web.expasy.org/protparam/); Gasteiger et al. 2005). Protein domain families were searched in Pfam database ([pfam.xfam.org/](http://pfam.xfam.org/); Sonnhammer et al. 1997). Conserved motifs in SOX sequences were investigated by using MEME tools ([meme.nbcrl.net/meme/](http://meme.nbcrl.net/meme/); Bailey et al. 2009) with parameters; maximum motif number to find, 5 and motif width, 6–50. Exon/intron organization was analyzed using GSDS server ([gsds.cbi.pku.edu.cn/](http://gsds.cbi.pku.edu.cn/); Hu et al. 2015). 1000 bp upstream regions of SOX genes from transcription start site were retrieved from phytozome and supplied to PlantCARE database for promoter analysis ([bioinformatics.psb.ugent.be/webtools/plantcare/html/](http://bioinformatics.psb.ugent.be/webtools/plantcare/html/); Lescot et al. 2002). Interaction partners of SOX proteins were predicted by STRING server ([string-db.org/](http://string-db.org/); Franceschini et al. 2013). Phylogenetic tree was constructed by MEGA 6 (Tamura et al. 2011) with parameters of maximum likelihood (ML), poisson correction, pairwise deletion and 1000 bootstraps, using herein four studied sequences along with 16 additional SOX proteins from *Chlamydomonas reinhardtii*, *Oryza brachyantha*, *Citrus sinensis*, *Solanum pennellii*, *Setaria italica*, *Brassica napus*, *Cicer arietinum*, *Jatropha curcas*, *Populus euphratica*, *Vitis vinifera*, *Morus notabilis*,



**Fig. 1** The schematic representation of sulfite pathway in plants (modified from Brychkova et al. 2013). Sulfite is produced from sulfate by two consecutive enzymatic reactions of ATP sulfurylase and APS reductase. The produced sulfite is further reduced to cysteine

by sulfite reductase and *O*-acetylserine (thiol) lyase respectively, to UDP-sulfoquinovose by UDP-sulfoquinovose synthase and to thio-sulfate by sulfurtransferases. Sulfite is also oxidized to sulfate by a molybdenum cofactor-containing enzyme, sulfite oxidase

*Nicotiana tomentosiformis*, *N. sylvestris*, *Vigna angularis*, *Capsicum annuum* and *Arachis ipaensis*.

## Gene expression analysis

Microarray data with GSE41935 record was retrieved from NCBI's GEO datasets ([ncbi.nlm.nih.gov/geo/](http://ncbi.nlm.nih.gov/geo/)). Ten natural *Arabidopsis* ecotypes such as An-1 (Antwerpen/Belgium), Cvi (Cape Verdia Islands), Col-0 (Columbia/United States), C24 (Coimbra/Portugal), Eri (Erigsboda/Sweden), Kas-1 (Kashmir/India), Kond (Kondara/Tajikistan), Kyo-2 (Kyoto/Japan), Ler (Landsberg/Poland), and Sha (Shak-dara/Tajikistan) were exposed to salt (100 mM NaCl), cold (10 °C), heat (38 °C) and light (800  $\mu\text{m photons m}^{-2} \text{s}^{-1}$ ) stresses. The expression profiles of four sulfite scavengers such as sulfite oxidase (EC 1.8.3.1; SOX), sulfite reductase (EC 1.8.7.1; SiR), UDP-sulfoquinovose synthase (EC 3.13.1.1; SQD) and thiosulfate sulfurtransferase (EC 2.8.1.1; STR) were evaluated. Data were processed using NCBI's GEO2R tool and heatmap was generated using CIMminer tool ([discover.nci.nih.gov/cimminer/home.do](http://discover.nci.nih.gov/cimminer/home.do)).

## Structure analysis

Secondary structure of SOX proteins was evaluated using SOPMA server (Geourjon and Deleage 1995). 3D models were constructed employing Phyre2 server ([sbj.bio.ic.ac.uk/phyre2/](http://sbj.bio.ic.ac.uk/phyre2/); Kelley and Sternber 2009). Predicted binding sites of SOXs were evaluated by using CASTp server ([sts.bioengr.uic.edu/castp/](http://sts.bioengr.uic.edu/castp/); Dundas et al. 2006). Model validation/reliability was performed with VADAR 1.8 ([vadar.wishartlab.com/](http://vadar.wishartlab.com/); Willard et al. 2003) using Ramachandran plot. Structure-based phylogeny was constructed by FITCH program (PHYLP package 3.6) based on Fitch and Margoliash methods using normalized RMSD (root mean-square deviation) values in superposed models (Felsenstein 2005). RMSDs values were calculated using CLICK server based on alpha carbon superposition of models ([mspc.bii-a-star.edu.sg/minhn/](http://mspc.bii-a-star.edu.sg/minhn/); Nguyen et al. 2011). Tree was visualized by FigTree v1.4.2 ([tree.bio.ed.ac.uk/software/figtree/](http://tree.bio.ed.ac.uk/software/figtree/)).

## Molecular docking

PDB files were energy-minimized by Swiss-PDB Viewer to release the internal molecular constraints using 100 steps of Steepest Descent and Conjugate Gradients algorithms, respectively (Kaplan and Littlejohn, 2001). Ligand sulfite was downloaded from ChEBI database ([ebi.ac.uk/chebi/](http://ebi.ac.uk/chebi/)) in Molfile format and converted into PDB file using Open Babel toolbox (O'Boyle et al. 2011). SOX proteins were docked with sulfite using Autodock Vina for 500 exhaustiveness. A grid box size (60  $\times$  70  $\times$  56 points) almost covering whole protein with 1 Å search spacing was used.

Addition of polar hydrogens and merging the non-polar hydrogens were performed using AutoDockTools 1.5.6 (Morris et al. 2009). Other parameters were adapted as default. Binding conformations were visualized using Pymol viewer ([pymol.org](http://pymol.org)).

## Results and discussion

### Sequence analysis of SOXs

Having insights about protein conservancy, four SOX protein sequences were aligned using ClustalW. The identical and similar residues were shaded as black and grey, respectively with 100% shading threshold to reveal the full conserved residues (Fig. 2). The alignment analysis showed that studied SOX proteins have highly conserved protein sequences. The similarity search also showed that *Arabidopsis* SOX has 80, 78 and 76% identities to *P. trichocarpa*, *S. lycopersicum* and *B. distachyon*, respectively. All SOX proteins were characterized with two domain structures such as an oxidoreductase molybdopterin binding domain (PF00174) and a Mo-co oxidoreductase dimerization domain (PF03404) (Table 1). They were distributed in a range of 392–397 amino acid residues with 43.21–43.93 kDa molecular weights and with slightly basic nature (7.64–8.80 *pI*). In addition, with exception of poplar (11 exons), all SOXs included 12 exons with various lengths at different positions, implicating those flexible splicing variants may be associated with the dynamic cellular environment of sulfite scavengers. In literature, studies of plant SOXs were rather scarce thereby a database search was conducted to have insight about the sequence features of SOX proteins from different species. The search revealed that protein features of herein studied SOXs are mainly shared by many other plants including *O. brachyantha* (NCBI access. XP\_006659588), *C. sinensis* (XP\_006470055), *S. pennellii* (XP\_015065665), *S. italica* (XP\_004973996), *B. napus* (XP\_013696421), *C. arietinum* (XP\_004489611), *J. curcas* (XP\_012070455), *P. euphratica* (XP\_011003919), *V. vinifera* (XP\_002269273), *M. notabilis* (XP\_010112361), *N. tomentosiformis* (XP\_009620347), *N. sylvestris* (XP\_009784618), *V. angularis* (XP\_017439833), *C. annuum* (XP\_016561725) and *A. ipaensis* (XP\_016162374).

Moreover, to further investigate the protein conservancy, the most conserved five motif sequences in four SOX proteins were identified (Suppl. Fig. 1). All motifs with about 50 residues were present in all SOX sequences. Motif 1 comprised of residues with LDSINIIAEECQ GFFMQKDYKMFPTVNWDNINWSTRPQMDFPVQS AIC, motif 2 with HVEFVSIDKCKEENGOPYKASIPLS QATNPEADVLLAYEMNGEPLNRDHG, motif 3 with





**Fig. 2** Multiple alignment of four SOX protein sequences from *A. thaliana*, *S. lycopersicum*, *B. distachyon* and *P. trichocarpa*. Alignment was shaded with 100% threshold to figure out the strictly conserved residues. Below yellow and blue lines show the

approximate location of oxidoreductase molybdopterin binding (PF00174) and Mo-co oxidoreductase dimerization (PF03404) domains, respectively (color figure online)

**Table 1** Gene and protein features of SOX sequences from *A. thaliana*, *S. lycopersicum*, *P. trichocarpa* and *B. distachyon*

| Species                | UniProtKB/Swiss-Prot accession | Phytozome transcript ID | Exon number | Protein length (aa) | Molecular weight (kDa) | pI   | Pfam domain family |
|------------------------|--------------------------------|-------------------------|-------------|---------------------|------------------------|------|--------------------|
| <i>A. thaliana</i>     | Q9S850                         | AT3G01910.1             | 12          | 393                 | 43.33                  | 8.80 | PF00174, PF03404   |
| <i>S. lycopersicum</i> | A5H1Q7                         | Solyc02g094120.2.1      | 12          | 392                 | 43.21                  | 8.43 | PF00174, PF03404   |
| <i>P. trichocarpa</i>  | B9GJT3                         | Potri.001G330900.1      | 11          | 393                 | 43.40                  | 8.18 | PF00174, PF03404   |
| <i>B. distachyon</i>   | I118W3                         | Bradi3g41170.1          | 12          | 397                 | 43.93                  | 7.64 | PF00174, PF03404   |

PF00174 oxidoreductase molybdopterin binding domain (oxidored\_molyb), PF03404 Mo-co oxidoreductase dimerisation domain (Mo-co\_dimer)

PSDYSQEPPRHPSLKINAKEPFNAEPPRSALISSYVTPVDFFYKRNHGPI, motif 4 with KYNVTATLQCA GNRRTAMSKVVRTVKGVGWDVSAIGNAVWGGAKLADVLEL and motif 5 with WAWVLFEATIDIPQSTEIVA KAVDSAANVQPENVEEIVNLRGILNTSWHR. Domain analysis also revealed that motif 2 and 4 were related with oxidoreductase molybdopterin binding domain (PF00174) while motif 5 was associated with Mo-co oxidoreductase dimerization domain (PF03404). However, motifs 1 and 3 did not demonstrate any relation to the known protein

families. In addition, these five motifs were evenly distributed on SOX sequences almost covering all protein. Thus, all these indicated that SOXs of different plants could have well conserved primary structure.

### Phylogenetic distribution of SOX proteins

Using a total of 20 different plant SOX proteins from four herein worked species along with 16 additional sequences, phylogenetic tree was constructed by MEGA 6 with ML

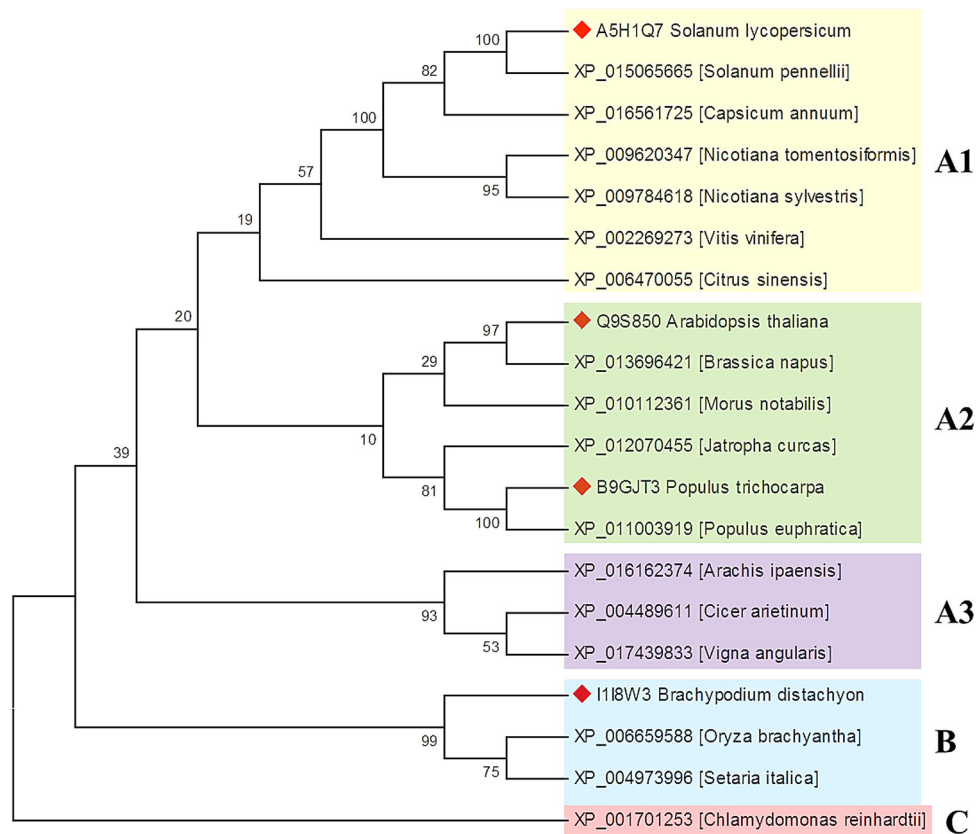
method for 1000 bootstraps (Fig. 3). The tree demonstrated the three major clusters namely group A, B and C with a clear divergence of monocot, dicot and lower plant. The group A included only dicot members, group B contained solely monocots and group C had just lower plant. Based on the clustering topology, group A was further subdivided into three sub-groups such as A1, A2 and A3. In subgroup A1, Solanaceous species such as *S. lycopersicum* (studied herein), *S. pennellii*, *C. annuum*, *N. tomentosiformis* and *N. sylvestris* diverged from the Vitaceae family member *V. vinifera* and Rutaceae family member *C. sinensis*. In subgroup A2, two studied species of *A. thaliana* and *P. trichocarpa* were included herein. The Brassicaceae family member *A. thaliana* was closely clustered with the same family member *B. napus* and slightly diverged from *M. notabilis* (Moraceae family). Besides, in same subgroup, *P. trichocarpa* and *P. euphratica* from the Salicaceae family were closely clustered and slightly separated from Euphorbiaceae family member *J. curcas*. However, no studied species were included in subgroup A3 but *A. ipaensis*, *C. arietinum* and *V. angularis* from Fabaceae family grouped together herein. In group B, studied species *B. distachyon* from Poaceae family was distributed with the same family members of *O. brachyantha* and *S. italica*. As of group C, it only included a lower plant *C. reinhardtii*. Overall phylogenetic distribution demonstrated a clear

divergence between monocots, dicots and lower plants, and also a family specific clustering pattern was present in SOX sequences.

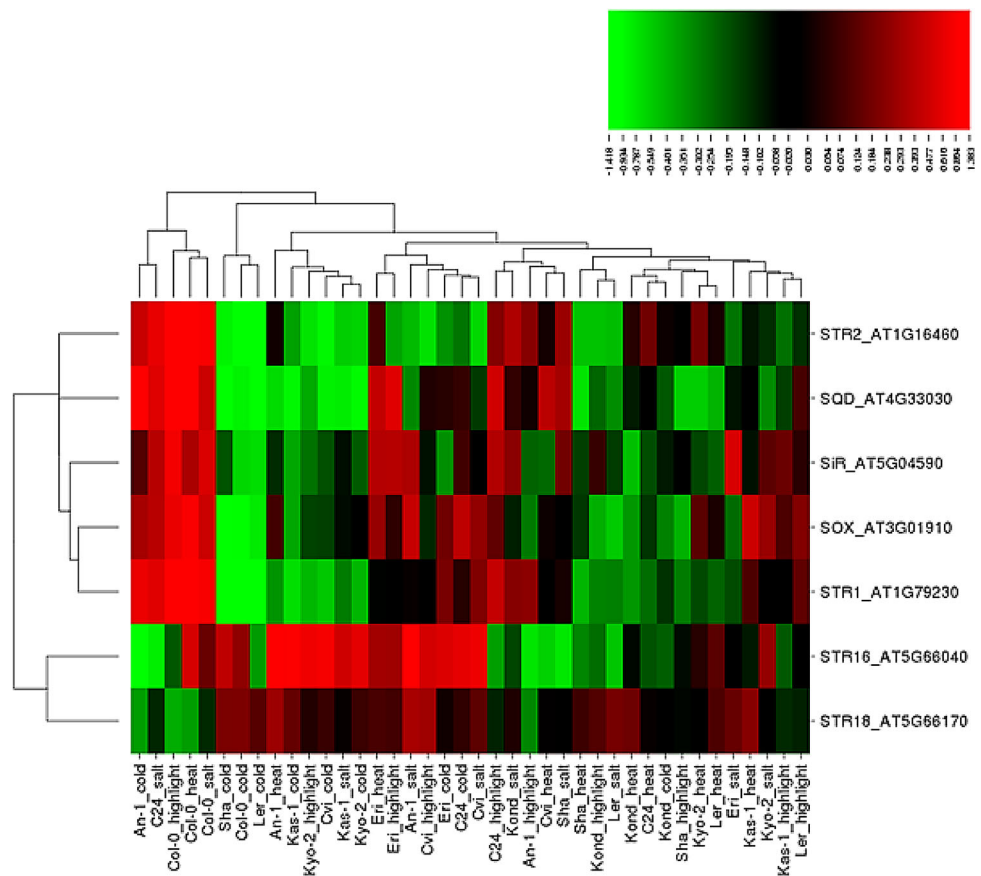
### Cis-regulatory elements in SOX genes

The characterization of upstream regions of genes is essential step to have insights about the transcriptional regulation mechanisms (Suzuki et al. 2001). In this sense, *cis*-regulatory element distribution was searched within a 1000 bp upstream region of four SOX genes from the transcription start site (TSS). In *A. thaliana*, *S. lycopersicum*, *P. trichocarpa* and *B. distachyon* respectively, 19, 12, 17 and 15 *cis*-regulatory elements were identified (Supp. Table 1). Of these, three types of *cis*-elements were common in all SOX genes such as 5 UTR Py-rich stretch, CAAT and TATA boxes. Py-rich stretch confers the high transcription levels without other *cis*-element requirements except for TATA-box (Fang et al. 2013). CAAT-box is a common *cis*-acting element in promoters of many eukaryotic genes (Laloum et al. 2013). TATA-box is considered as core promoter element in eukaryotic genes for binding either general TFs or histones (Bae et al. 2015). Although many other *cis*-elements have been identified in promoter regions of SOX genes, some elements were uniquely present in certain species. For example, ACE and

**Fig. 3** Phylogenetic distribution of 20 plant SOX proteins. Tree was constructed by MEGA 6 with ML method for 1000 bootstraps using four studied sequences (marked with red diamond) such as *A. thaliana*, *S. lycopersicum*, *P. trichocarpa* and *B. distachyon* along with 16 additional SOXs from *C. reinhardtii*, *O. brachyantha*, *C. sinensis*, *S. pennellii*, *S. italica*, *B. napus*, *C. arietinum*, *J. curcas*, *P. euphratica*, *V. vinifera*, *M. notabilis*, *N. tomentosiformis*, *N. sylvestris*, *V. angularis*, *C. annuum* and *A. ipaensis*. SOXs were annotated with NCBI accession number and full species name (color figure online)



**Fig. 4** The expression profiles of seven sulfite scavengers such as sulfite oxidase (SOX), sulfite reductase (SiR), sulfurtransferases (STR1, 2, 16 and 18) and UDP-sulfoquinovose synthase (SQD) in ten natural *Arabidopsis* ecotypes such as An-1, Cvi, Col-0, C24, Eri, Kas-1, Kond, Kyo-2, Ler and Sha. Ecotypes were exposed to four different stresses such as salt (100 mM NaCl), cold (10 °C), heat (38 °C) and high light (800  $\mu\text{m photons m}^{-2} \text{s}^{-1}$ ). Green and red colors, respectively indicate the down and upregulated genes. The similar expression profiles of conditions (up) and genes (left) were hierarchically clustered using Pearson correlation (color figure online)



Chs-CMA2b (light responsiveness), Box-W1 (fungal elicitor responsiveness), TATCCAT/C motif (sugar repression responsiveness), TC-rich repeats (stress-inducible) and W-box (wounding and pathogen responsiveness) were only present in *Arabidopsis*. The unique *cis*-regulatory elements in poplar included the Box s (wounding and pathogen responsiveness), O<sub>2</sub>-site (zein metabolism regulation), I- and L-boxes and BOX I (light responsiveness), A-box (*cis*-acting regulatory element) and ERE (ethylene-responsive element). Besides, a MBS motif (drought-inducibility) was only identified in *B. distachyon*. Taking into account the various sulfur metabolites are involved in plant stress tolerance, the availability of *cis*-regulatory motifs associated with different stresses is justifiable.

#### Gene expression profiles of sulfite scavengers

To have a global understanding about the involvement of sulfite pathway in plant stress tolerance, the expression profiles of four types of sulfite scavengers such as SOX (At3g01910), SiR (At5g04590), SQD (At4g33030), and STR1 (At1g79230), STR2 (At1g16460), STR16 (At5g66170) and STR18 (At5g66040) were comparatively analyzed using a microarray record GSE41935 from NCBI's GEO datasets (Fig. 4). The array investigated 10

natural *Arabidopsis* ecotypes such as An-1, Cvi, Col-0, C24, Eri, Kas-1, Kond, Kyo-2, Ler and Sha under cold, heat, salt and high light stresses. To identify the differentially expressed genes,  $-0.5 \leq x \leq 0.5$  threshold value based on the log<sub>2</sub>-fold change was adopted. For this threshold, SOX gene was significantly upregulated under heat (ecotype Col-0: 0.864 fold) and high light (Col-0: 0.637) while it was notably down-regulated under cold (Sha: -1.013, Col-0: -0.936 and Ler: -0.813) and salt (Ler: -0.542) stresses. In addition, seven ecotypes under heat, five ecotypes under high light and salt for each, and three ecotypes under cold stress showed an upregulated expression pattern at different levels. Besides, SiR gene was mostly up and downregulated under high light (Col-0: 0.672) and cold (Ler: -0.629) stresses respectively. For SQD gene, its highest and lowest expression values respectively were under high light (Col-0: 1.383) and cold (Sha: -1.364) stresses. From STR genes, STR1, STR2 and STR16 respectively showed the highest upregulation under high light (Col-0: 1.056) and heat (Col-0: 1.127 and An-1: 1.291), while they were mostly downregulated under cold stress respectively as Ler: -1.253, Col-0: -1.418 and An-1: -1.13. However, STR18 did not demonstrate any significant expression level based on the adopted threshold. Above expression patterns of genes indicated that sulfite



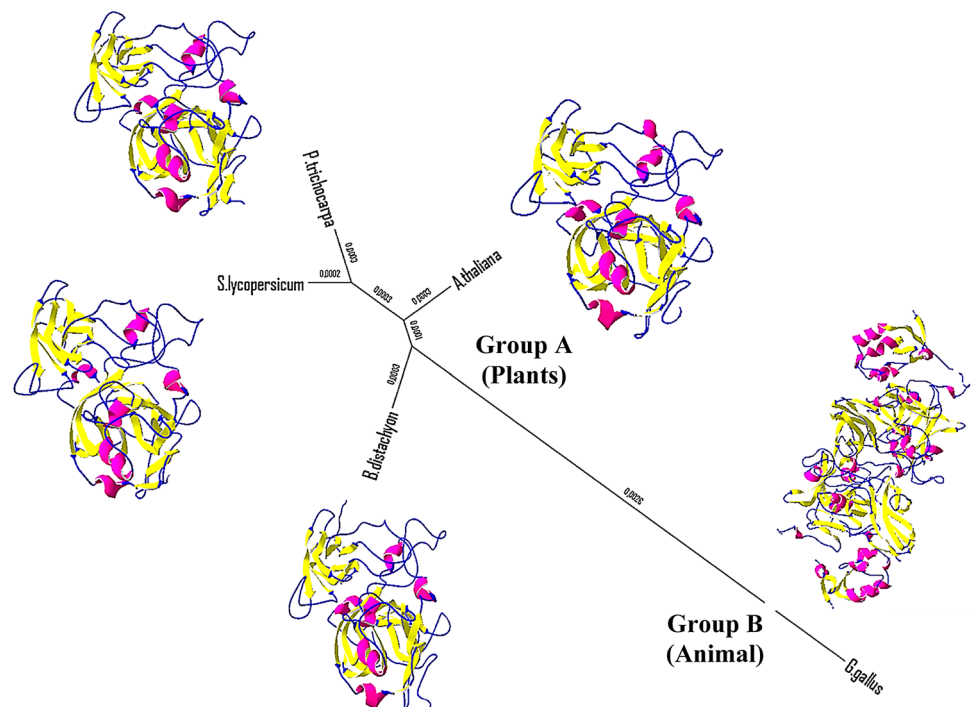
scavengers including *SOX* are differentially modulated under given abiotic stresses. Particularly, they tend to be up-regulated under high light and heat while being down-regulated under cold and salt stresses.

### 3D modeling and structure-based phylogeny

To further investigate the biologically functional state of proteins, 3D models were constructed by Phyre2 server at intensive mode using four *SOX* sequences from *A. thaliana*, *S. lycopersicum*, *P. trichocarpa* and *B. distachyon*. In modelling, template 1OGP with PDB databank entry was used to maximize the alignment coverage, percentage identity and confidence for submitted sequences. The reliability of generated models was validated by VADAR employing the Ramachandran plot. The analysis showed that  $\geq 98\%$  of residues are in core and allowed regions, demonstrating the high confidence of models for further analysis. In previous studies, the structure-based phylogeny was indicated to have noteworthy importance in understanding the protein function and evolution. Because two protein sequences with low similarity may possess a common fold or topology and could perform similar functions, therefore assessing the 3D structures of proteins can give more useful information from functional aspects (Foy, 2013; Lakshmi et al. 2015; Wolf and Grünwald 2015). Thus, to understand the phylogenetic relationship and make a functional inference in constructed models, a structure-based phylogenetic tree was constructed using normalized RMSD values in five superposed models

(Fig. 5; Suppl. Table 2), including four modeled plant *SOX*s herein and one previously characterized additional *SOX* structure from chicken (*Gallus gallus*: PDB: 1SOX). The structures were superposed based on the alpha carbon superposition of models and obtained RMSD values were normalized by dividing with the aligned residue numbers. The phylogenetic tree demonstrated two main clusters as group A and B. Group A included four plant *SOX* proteins while group B contained the animal *SOX* from chicken. This split of plant and animal *SOX*s may also implicate their functional divergences. In addition, although all plant *SOX*s were closely clustered, they are calculated to be subjected to different evolutionary changes over time. Based on tree topology, monocot *B. distachyon* appeared to be diverged from other dicots, pointing out a possible monocot–dicot split. Besides, in dicots *P. trichocarpa* and *S. lycopersicum* showed more close relationship than *A. thaliana*. The earlier search of primary sequences showed that *Arabidopsis* *SOX* has 80, 78 and 76% identities to *P. trichocarpa*, *S. lycopersicum* and *B. distachyon* respectively, complying with the structure-based similarities. The analysis of secondary structures also further corroborated this showing that *A. thaliana*, *S. lycopersicum*, *P. trichocarpa* and *B. distachyon* respectively include 24.43, 25, 25.45 and 30.23%  $\alpha$ -helices; 24.43, 22.19, 21.88 and 18.39%  $\beta$ -strands; 11.70, 10.97, 12.21, 9.32%  $\beta$ -turns; and 39.44, 41.84, 40.46, and 42.07% random coils. Despite of their similar seconder structures, *B. distachyon* showed some structural differences, maybe contributable to functional diversities of *SOX*s in monocots. However, more

**Fig. 5** Structure-based phylogeny of four modelled plant and one previously characterized chicken (*G. gallus*) *SOX*s. Tree was constructed by FITCH program (PHYLIP package 3.6) based on the Fitch and Margoliash methods using normalized RMSD values in superposed models. The branch length distance and species names were indicated on tree. The purple, yellow and blue colors represent alpha helix, beta strand and loop structures, respectively (color figure online)

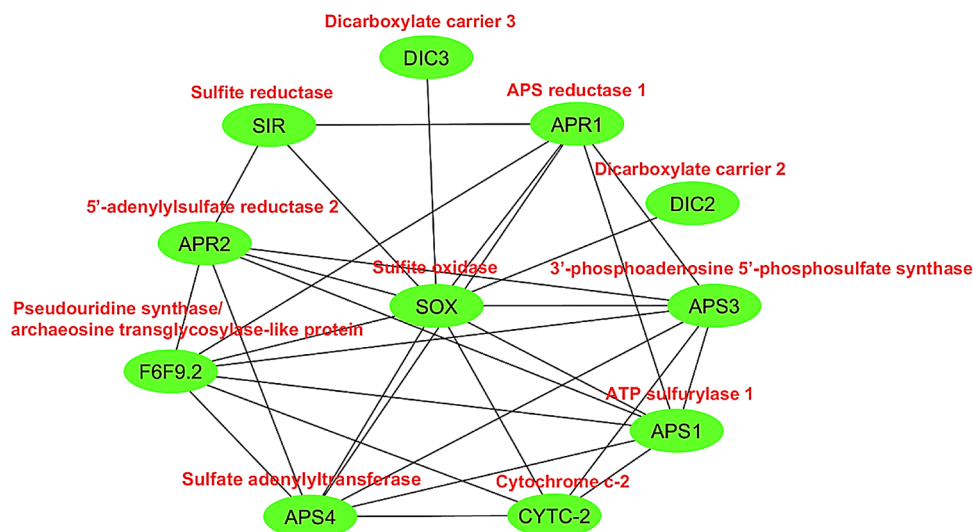


related clustering of poplar, tomato, *Arabidopsis* and *Brachypodium* SOXs compared to chicken could make inference to their common topologies thereby functional relationships. Overall, it was implicated that structure-based phylogenies could produce more realistic results from functional aspects especially for sequences with low similarities but similar functionalities.

### Protein–protein interaction (PPI) network of AtSOX

PPI network was constructed for *A. thaliana* SOX (AT3G01910.1) protein using Cytoscape for STRING data (Fig. 6). The protein–protein associations were mainly predicted from the curated databases with high confidence score. Predicted 10 potential interaction partners of AtSOX included the sulfite reductase (SiR), ATP sulfurylase 1 (APS1), 3'-phosphoadenosine 5'-phosphosulfate synthase (APS3), sulfate adenylyltransferase (APS4), APS reductase 1 (APR1), 5'-adenylylsulfate reductase 2 (APR2), dicarboxylate carrier 3 (DIC3), dicarboxylate carrier 2 (DIC2), pseudouridine synthase/archaeosine transglycosylase-like protein (F6F9.2) and cytochrome c-2 (CYTC-2) proteins. Most proteins predicted herein were among the main enzymes involved in the sulfite pathway. For example, sulfite reductase (SiR; EC 1.8.7.1) is a prime enzyme in sulfate assimilation pathway in which catalyzes the sulfite to sulfide (Nakayama et al. 2000; Khan et al. 2010). ATP sulfurylases (ATPS; EC 2.7.7.4) synthesize the adenosine 5'-phosphosulfate (APS) from sulfate and ATP (Khan et al.

2010). APS reductases (APR; EC 1.8.4.9) catalyze the two-electron reduction of APS to sulfite and AMP (Brychkova et al. 2012). DICs (dicarboxylate carriers) are belonged to the mitochondrial carrier protein family and they transport the dicarboxylic acids including malate, oxaloacetate and succinate as well as phosphate, sulfate and thiosulfate at high rates (Palmieri et al. 2008). Moreover, this PPI network was functionally enriched with gene ontology (GO) terms to have further insights about the molecular, cellular and biological functions of these potential interactors. The terms enriched in network included as “Biological Process”: sulfate assimilation (GO:0000103), sulfur compound metabolic process (GO:0006790), hydrogen sulfide biosynthetic process (GO:0070814), sulfur compound biosynthetic process (GO:0044272) and sulfate reduction (GO:0019419). “Molecular Function”: sulfate adenylyl-transferase (ATP) activity (GO:0004781), adenylylsulfate kinase activity (GO:0004020), oxidoreductase activity, acting on a sulfur group of donors (GO:0016667), adenylyl-sulfate reductase activity (GO:0009973) and adenylyl-sulfate reductase (glutathione) activity (GO:0033741). “Cellular Component”: chloroplast stroma (GO:0009570), cytoplasmic part (GO:0044444), intracellular organelle part (GO:0044446), cytoplasm (GO:0005737), intracellular membrane-bounded organelle (GO:0043231), chloroplast (GO:0009507), intracellular (GO:0005622) and cell (GO:0005623). Network enrichment also further supported that these proteins/enzymes are mainly involved in the sulfate assimilation pathway in plants.



**Fig. 6** Ten potential interaction partners of AtSOX (AT3G01910.1) protein. Interactome was generated by using cytoscape with STRING data. Potential interactors of AtSOX included sulfite reductase (SiR), sulfate adenylyltransferase (APS4), 3'-phosphoadenosine 5'-phosphosulfate synthase (APS3), APS reductase 1 (APR1), ATP sulfurylase 1

(APS1), 5'-adenylylsulfate reductase 2 (APR2), pseudouridine synthase/archaeosine transglycosylase-like protein (F6F9.2), dicarboxylate carrier 3 (DIC3), dicarboxylate carrier 2 (DIC2) and cytochrome c-2 (CYTC-2) proteins



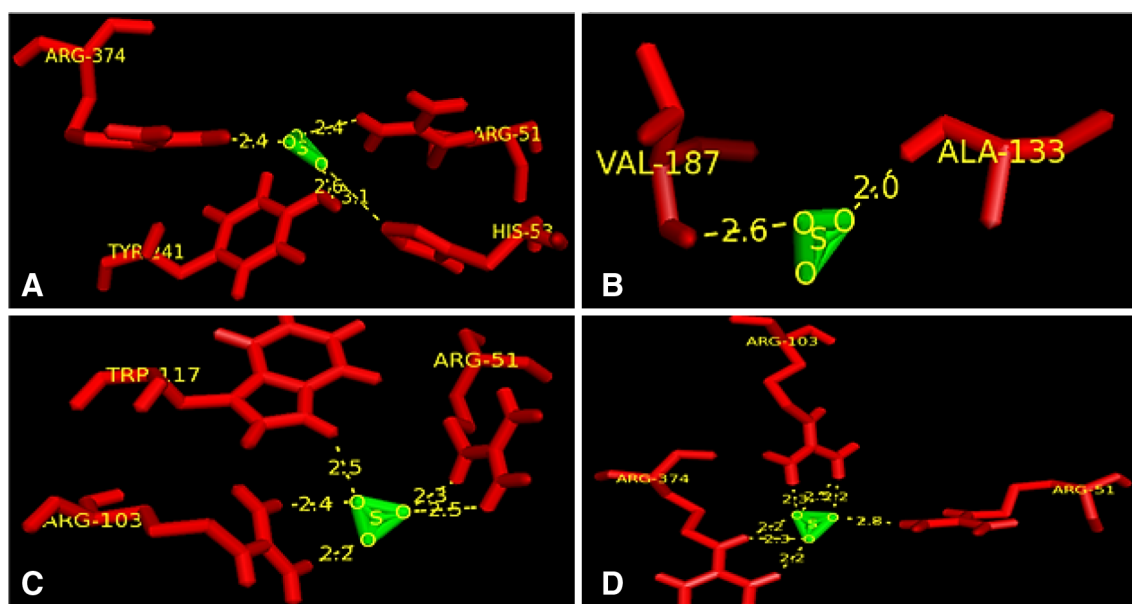
### Docking of SOXs with ligand sulfite

Finally, all studied plant SOX proteins were docked with ligand sulfite to investigate the potential binding sites and affinities in SOX–sulfite complexes (Fig. 7; Table 2). In docked complexes, each of the generated nine potential conformations with lowest binding energies was comparatively investigated with the available crystal structures of plant sulfite oxidase to explore the best/native-like binding mode. Schrader et al. (2003) reported that 164–169 residues are well conserved in active sites of plant SOXs and they may function as specific electron acceptor. Particularly Arg374 was identified as a crucial substrate binding residue and two other residues such as Trp117 and Tyr241 were also reported as important in substrate binding of plant SOXs (Schrader et al. 2003). Herein, in *A. thaliana* SOX–sulfite complex, the conformation with  $-2.9$  kcal/mol binding energy included the early reported binding residues Arg374 and Tyr241 as well as two others such as His53 and Arg51. So, this binding conformation posed to be possibly the one like native binding mode. However, in *B. distachyon* SOX–sulfite complex, no reported residues were identified but the conformation with  $-3.5$  kcal/mol lowest binding energy had polar contacts with Ala133 and Val187. These variations in binding residues tempted to speculate that SOX–sulfite binding mode and interacting residues may differ based on the species-dependent way. In case of *P. trichocarpa* SOX–sulfite complex, the top six conformations with slightly different modes possessing  $-3.5$  kcal/mol lowest binding energy made contacts with

residues Trp117, Arg51 and Arg103. From these, Trp117 was earlier reported as important in substrate binding of plant SOXs (Schrader et al. 2003). Besides, the availability of Arginine residues particularly Arg51, which was also identified in *Arabidopsis*, made an indication that this residue may also involve in SOX–sulfite binding in poplar. However, variations in binding residues compared to other complexes once more implicated that binding mode and residues could vary at species-specific way. Finally, in *S. lycopersicum* SOX–sulfite complex, the conformation with  $-3.3$  kcal/mol binding energy possessed the polar contacts with residues Arg374, Arg51 and Arg103. Interestingly, all contact residues were the arginine and importantly Arg374 was earlier characterized experimentally as a crucial substrate binding residue (Schrader et al. 2003). Besides, both Arg51 and Arg103 were also available in poplar complex while Arg51 was present in *Arabidopsis*. Apparently, all these indicate that arginine residues particularly Arg374 is crucial for SOX–sulfite binding and two other residues Arg51 and Arg103 were also implicated to be important for SOX–sulfite bindings in plants.

### Conclusion

The studies of sulfite oxidase (SOX) which is an essential enzyme in sulfate assimilation pathway in plants were quite limited. In this sense, to fill the literature gap about this crucial enzyme, the present work aimed to improve our understanding of SOXs in four plant species such as *A.*



**Fig. 7** Docked structures of *A. thaliana* SOX–sulfite (a), *B. distachyon* SOX–sulfite (b), *P. trichocarpa* SOX–sulfite (c) and *S. lycopersicum* SOX–sulfite (d). Docking was performed by using

AutoDock Vina. SOX and sulfite were displayed with red and green colors, respectively. Polar contacts, interacting residues and their atomic distances were indicated with yellow tags (color figure online)

**Table 2** The docking features of ligand sulfite and SOX proteins from *A. thaliana*, *B. distachyon*, *P. trichocarpa* and *S. lycopersicum*. Below are tabulated the binding affinities, interacting residues and atoms, and their relative distances

| Species                | PDB   | Binding energy (kcal/mol) | Interacting amino acids/atoms |      |              | Distance (Å) |
|------------------------|-------|---------------------------|-------------------------------|------|--------------|--------------|
|                        |       |                           | Amino acids                   | Atom | Sulfite atom |              |
| <i>A. thaliana</i>     | Model | −2.9                      | ARG374                        | 2HH2 | O            | 2.4          |
|                        |       |                           | TYR241                        | HH   | O            | 2.6          |
|                        |       |                           | HIS53                         | NE2  | O            | 3.1          |
|                        |       |                           | ARG51                         | 1HH1 | O            | 2.4          |
| <i>B. distachyon</i>   | Model | −3.5                      | ALA133                        | H    | O            | 2.0          |
|                        |       |                           | VAL187                        | H    | O            | 2.6          |
| <i>P. trichocarpa</i>  | Model | −3.5                      | TRP117                        | H    | O            | 2.5          |
|                        |       |                           | ARG51                         | HE   | O            | 2.3          |
|                        |       |                           |                               | 2HH2 | O            | 2.5          |
|                        |       |                           | ARG103                        | 1HH1 | O            | 2.4          |
| <i>S. lycopersicum</i> | Model | −3.3                      |                               | 1HH2 | O            | 2.2          |
|                        |       |                           | ARG374                        | 2HH1 | O            | 2.2          |
|                        |       |                           |                               | HE   | O            | 2.2          |
|                        |       |                           |                               | HE   | O            | 2.3          |
|                        |       |                           | ARG51                         | 2HH2 | O            | 2.8          |
|                        |       |                           | ARG103                        | 1HH1 | O            | 2.3          |
|                        |       |                           |                               | 1HH2 | O            | 2.2          |
|                        |       |                           |                               | 1HH2 | O            | 2.5          |

*thaliana*, *S. lycopersicum*, *P. trichocarpa* and *B. distachyon*. The general protein features of studied SOXs herein were mainly shared by many other plants. However, phylogeny constructed with additional species showed a clear divergence of monocots, dicots and lower plants. The given abiotic stresses significantly regulated the sulfite scavengers including *SOX* and the presence of stress-related *cis*-regulatory motifs in upstream regions of studied SOXs was thus reasonable. The structure-based phylogenies were implicated able to produce more realistic results for protein function and evolution particularly for sequences with low similarities but similar functionalities. In addition, enriched PPI networks could provide useful information about functional aspects. Finally, docking analysis indicated that interacting residues of SOX–sulfite complexes may show variations at species-dependent way.

## References

- Bae SH, Han HW, Moon J (2015) Functional analysis of the molecular interactions of TATA box-containing genes and essential genes. *PLoS ONE* 10:e0120848
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L et al (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. doi:10.1093/nar/gkp335
- Brychkova G, Xia Z, Yang G, Yesbergenova Z, Zhang Z, Davydov O, Fluhr R, Sagi M (2007) Sulfite oxidase protects plants against sulfur dioxide toxicity. *Plant J* 50:696–709
- Brychkova G, Yarmolinsky D, Sagi M (2012) Kinetic assays for determining in vitro APS reductase activity in plants without the use of radioactive substances. *Plant Cell Physiol* 53:1648–1658
- Brychkova G, Grishkevich V, Fluhr R, Sagi M (2013) An essential role for tomato sulfite oxidase and enzymes of the sulfite network in maintaining leaf sulfite homeostasis. *Plant Physiol* 161:148–164
- Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acid Res* 34:W116–W118
- Eilers T, Schwarz G, Brinkmann H, Witt C, Richter T, Nieder J, Koch B, Hille R, Hänsch R, Mendel RR (2001) Identification and biochemical characterization of *Arabidopsis thaliana* sulfite oxidase: a new player in plant sulfur metabolism. *J Biol Chem* 276:46989–46994
- Fang W, Zhang Y, Zhou L, Wang W, Li X (2013) Isolation and characterization of Histone1 gene and its promoter from tea plant (*Camellia sinensis*). *Mol Biol Rep* 40:3641–3648
- Felsenstein J (2005) PHYLIP (phylogeny inference package) version 3.6
- Foy SG (2013) Calculating the structure-based phylogenetic relationship of distantly related homologous proteins utilizing maximum likelihood structural alignment combinatorics and a novel structural molecular clock hypothesis (Doctoral dissertation, University of Missouri-Kansas City)
- Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A et al (2013) STRING v9. 1: protein–protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41(Database issue):D808–D815

- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD et al (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The proteomics protocols handbook. Humana, Louisville, pp 571–607
- Geourjon C, Deleage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11:681–684
- Hänsch R, Lang C, Riebeseel E, Lindigkeit R, Gessler A, Rennenberg H, Mendel RR (2006) Plant sulfite oxidase as novel producer of H<sub>2</sub>O<sub>2</sub> combination of enzyme catalysis with a subsequent non-enzymatic reaction step. J Biol Chem 281:6884–6888
- Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31(8):1296–1297
- Kaplan W, Littlejohn TG (2001) Swiss-PDB viewer (deep view). Brief Bioinform 2:195–197
- Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4:363–371
- Khan MS, Haas FH, Samami AA, Gholami AM, Bauer A, Fellenberg K, Reichelt M, Hänsch R, Mendel RR, Meyer AJ et al (2010) Sulfite reductase defines a newly discovered bottleneck for assimilatory sulfate reduction and is essential for growth and development in *Arabidopsis thaliana*. Plant Cell 22:1216–1231
- Lakshmi B, Mishra M, Srinivasan N, Archunan G (2015) Structure-based phylogenetic analysis of the Lipocalin Superfamily. PLoS ONE 10:e0135507
- Laloum T, De Mita S, Gamas P, Baudin M, Niebel A (2013) CCAATbox binding transcription factors in plants: y so many? Trends Plant Sci 18:157–166
- Lescot M, Déhais P, Moreau Y, De Moor B, Rouzé P, Rombauts S (2002) PlantCARE: a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30:325–327
- Leustek T (2002) Sulfate metabolism. In: Somerville CR, Meyerowitz EM (eds) The Arabidopsis Book. American Society of Plant Biologists, Rockville
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30(16):2785–2791
- Nakayama M, Akashi T, Hase T (2000) Plant sulfite reductase: molecular structure, catalytic function and interaction with ferredoxin. J Inorg Biochem 82:27–32
- Nguyen MN, Tan KP, Madhusudhan MS (2011) CLICK—topology-independent comparison of biomolecular 3D structures. Nucleic Acids Res 39:W24–W28
- Nowak K, Luniak N, Witt C, Wüstefeld Y, Wachter A, Mendel RR, Hänsch R (2004) Peroxisomal localization of sulfite oxidase separates it from chloroplast-based sulfur assimilation. Plant Cell Physiol 45:1889–1894
- O’Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. J Cheminform 3:1
- Palmieri L, Picault N, Arrigoni R, Besin E, Palmieri F, Hodges M (2008) Molecular identification of three *Arabidopsis thaliana* mitochondrial dicarboxylate carrier isoforms: organ distribution, bacterial expression, reconstitution into liposomes and functional characterization. Biochem J 410:621–629
- Papenbrock J, Schmidt A (2000) Characterization of a sulfurtransferase from *Arabidopsis thaliana*. Eur J Biochem 267:145–154
- Romiti M (2010) Entrez nucleotide and entrez protein FAQs. Gene 1:270
- Saito K (2000) Regulation of sulfate transport and synthesis of sulfur containing amino acids. Curr Opin Plant Biol 3:188–195
- Sanda S, Leustek T, Theisen MJ, Garavito RM, Benning C (2001) Recombinant Arabidopsis SQD1 converts UDP-glucose and sulfite to the sulfolipid head group precursor UDP-sulfoquinovose in vitro. J Biol Chem 276:3941–3946
- Schrader N, Fischer K, Theis K, Mendel RR, Schwarz G, Kisker C (2003) The crystal structure of plant sulfite oxidase provides insights into sulfite oxidation in plants and animals. Structure 11:1251–1263
- Sonnhammer EL, Eddy SR, Durbin R (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins 28:405–420
- Suzuki Y, Tsunoda T, Sese J, Taira H, Mizushima-Sugano J, Hata H et al (2001) Identification and characterization of the potential promoter regions of 1031 kinds of human genes. Genome Res 11:677–684
- Tamura K, Stecher G, Peterson D, Peterson N, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739
- Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS (2003) VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res 31:3316–3319
- Wolf S, Grünewald S (2015) Sequence, structure and ligand binding evolution of rhodopsin-like G protein-coupled receptors: a crystal structure-based phylogenetic analysis. PloS ONE 10(4):e0123533