Identification of Potential Genes Encoding Protein Transporters in *Arabidopsis thaliana* Glucosinolate (GSL) Metabolism

Sarahani Harun [1,*], **Nor Afiqah-Aleng** [2], **Fatin Izzati Abdul Hadi** [3], **Su Datt Lam** [4] and **Zeti-Azura Mohamed-Hussein** [1,4]

1 Centre for Bioinformatics Research, Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia; zeti.hussein@ukm.edu.my

2 Institute of Marine Biotechnology, Universiti Malaysia Terengganu, Kuala Nerus 21030, Terengganu, Malaysia; afiqahaleng@umt.edu.my

3 Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia; p105328@siswa.ukm.edu.my

4 Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia; sudatt@ukm.edu.my

* Correspondence: sarahani@ukm.edu.my

Abstract: Several species in *Brassicaceae* produce glucosinolates (GSLs) to protect themselves against pests. As demonstrated in *A. thaliana*, the reallocation of defence compounds, of which GSLs are a major part, is highly dependent on transport processes and serves to protect high-value tissues such as reproductive tissues. This study aimed to identify potential GSL-transporter proteins (TPs) using a network-biology approach. The known *A. thaliana* GSL genes were retrieved from the literature and pathway databases and searched against several co-expression databases to generate a gene network consisting of 1267 nodes and 14,308 edges. In addition, 1151 co-expressed genes were annotated, integrated, and visualised using relevant bioinformatic tools. Based on three criteria, 21 potential GSL genes encoding TPs were selected. The AST68 and ABCG40 potential GSL TPs were chosen for further investigation because their subcellular localisation is similar to that of known GSL TPs (SULTR1;1 and SULTR1;2) and ABCG36, respectively. However, AST68 was selected for a molecular-docking analysis using AutoDOCK Vina and AutoDOCK 4.2 with the generated 3D model, showing that both domains were well superimposed on the homologs. Both molecular-docking tools calculated good binding-energy values between the sulphate ion and Ser419 and Val172, with the formation of hydrogen bonds and van der Waals interactions, respectively, suggesting that AST68 was one of the sulphate transporters involved in GSL biosynthesis. This finding illustrates the ability to use computational analysis on gene co-expression data to screen and characterise plant TPs on a large scale to comprehensively elucidate GSL metabolism in *A. thaliana*. Most importantly, newly identified potential GSL transporters can serve as molecular tools in improving the nutritional value of crops.

Keywords: glucosinolate; co-expression network; molecular docking; transporter proteins

1. Introduction

Plants are sessile organisms that are regularly subjected to a variety of biotic and abiotic stresses, resulting in biochemical and physiological changes that have a significant impact on plant development and survival. In general, plants have two basic defence mechanisms to overcome these challenges: structural responses and metabolic changes [1,2]. The production of secondary metabolites is one of the metabolic changes that occur in response to both biotic and abiotic stresses [3]. Glucosinolates (GSLs) are an extensively studied group of secondary metabolites [4] due to their role as major defence compounds in plants, protecting against herbivores and pathogens [5]. GSLs are unique to the *Brassicaceae* family, and they are found in plants such as *Arabidopsis thaliana* and many cultivated vegetables

(broccoli, cauliflower, cabbage, wasabi, horseradish, and mustard) [4,6,7]. The genotype, climate, and cultivation conditions, such as fertilisation and harvest time, all influence the composition and content of GSLs, and they are very diverse amongst the GSL containing plants [8]. GSLs are characterised by the existence of nitrogen and at least two sulphur atoms in the GSL core structure, suggesting that sulphur metabolism is essential in GSL biosynthesis [9].

GSLs are grouped based on their precursors (having different side chains). There are three GSL groups—aliphatic GSLs, which are produced from methionine, alanine, leucine, isoleucine, or valine; indolic GSLs, which are synthesised from tryptophan; and benzyl GSLs, which are produced from phenylalanine or tyrosine [10–12]—and more than 130 different GSLs in GSL-containing plants have been identified [13]. GSLs are water-soluble compounds that are stable when stored in plant cells [10]. Plant-cell disruption from insect feeding or mechanistic disruption leads to the GSL-myrosinase (GM) activation of nitriles, epithionitriles, isothiocyanates, and/or thiocyanates, which are converted from unstable aglycones, to protect plants against biotic and abiotic stresses [14,15]. This process is catalysed by myrosinases (TGG, EC 3.2.1.147), which also contribute to the variation of GSL products and GSL side-chain compositions [4,16]. The GM mechanism is known to release a toxic "mustard oil bomb" for repelling pathogens and insects [4,17]. The isothiocyanates, which appear to be universally toxic, have been attributed to the GM system's defence role. When utilised in bioassays with insects in pure form, their toxicity is comparable to that of commercial insecticides [18].

GSLs are stored in several locations, such as the laticifer-like S-cells and along the leaf margin and seeds [19–23]. The composition of GSLs differs quantitatively and qualitatively in different GSL-containing plant organs. The GSL concentrations are generally higher in roots than shoots [24]. Previous studies suggested that the distribution of GSLs across various plant organs can be explained by the optimal defence theory [25]. Compounds involved in a defence mechanism are preferentially distributed to organs that are more attractive to pests [26,27]. Thus, the reproductive organs, such as the flowers and seeds, store the highest concentrations of GSL, while the concentrations of GSL in the tissues below ground are the highest in the tap and lateral roots [28].

In GSL metabolism, sulphated GSLs are produced in the cytoplasm and stored in the vacuoles or S-cells in the periphery of the phloem [29]. The GSL concentration in S-cells is up to 20 times higher than that in the surrounding tissues [30]. However, the transport mechanism for GSL storage in both vacuoles and S-cells is still unclear [31]. A proteomic analysis on S-cell cytoplasm conducted by Koroleva et al. [20] failed to identify the existence of GSL-biosynthetic enzymes, hence suggesting the involvement of transporters in the accumulation of GSLs in the cells [20]. Transporter proteins (TPs) play an essential role in various mechanistic properties in plants, such as signalling, metabolism, and physiology involving the translocation of different molecules (hormones, amino acids, sugars, inorganic ions, water, and solutes) through plant membranes [32]. At present, five TPs have been experimentally validated as being involved in various GSL metabolisms, i.e., GSL transporters (GTR1 and GTR2), sulphate transporters (SULTR1;1 and SULTR1;2), and sodium symporter family protein 5 (BAT5) [12].

The long-distance transportation of a GSL from its source (leaf, root, and silique) to where it accumulates (leaf, root, silique, and embryo) is regulated by GSL TPs (GTRs) [33–35]. Nour-Eldin et al. [33] found that the *Arabidopsis* nitrate/peptide group GSL transporters AtNPF2.10 (AtGTR1) and AtNPF2.11 (AtGTR2) were responsible for the long-distance transport of short- and long-chain aliphatic GSLs from source tissues to target tissues. GTR1 and GTR2 facilitate long-distance GSL transport through phloem and xylem tissues [33,34,36,37]. Andersen et al. [34] conducted a micro-grafting experiment in *Arabidopsis* and found a specific GSL transporter that facilitates the transport of indole GSLs between the rosette leaves and roots [37]. Meanwhile, Madsen et al. [38] found that leaves of *Arabidopsis gtr1gtr2* mutants reduced the fitness of green peach aphids (*Myzus persicae*) by reducing the availability of GSL in phloem sap and increasing GSL in the tissues around the phloem. This observa-

tion suggests a potential application for these transporters in novel resistance against the insect. Detailed understanding of the defence mechanism involving GSLs can facilitate the development of crops that are more resistant to pests [35,38].

The sulphur uptake in GSL-containing plants suggests a significant role for sulphur in GSL biosynthesis. Metabolomic and transcriptomic studies by Koprivova and Kopriva [39] and Morikawa-Ichinose et al. [40] found that GSL accumulation was significantly reduced in a sulphur-deficient environment, suggesting a role for sulphur in GSL biosynthesis [41–43]. SULTR1;1 and SULTR1;2 are sulphate transporters found in *Arabidopsis* roots, and their expression is increased during sulphur limitation [44]. The sulphate transportation and GSL-transport machinery mechanism is more complex in *Brassica* crops. Thus, elucidating this mechanism is essential, as its knowledge can be used to design transporters as molecular tools in crop improvement. Another known GSL TP, BAT5, functions specifically in the side-chain elongation of aliphatic GSL biosynthesis by directing 2-oxo acids into the chloroplast. The involvement of the chloroplastic TP in the GSL mechanism was shown by depleting the function of *BAT5* in a *bat5*-knockout *Arabidopsis* mutant, in which the level of aliphatic GSL was reduced [45,46]. Thus, GSL biosynthesis is known to occur within plastids and the cytoplasm.

Over the last few decades, genome sequencing, combined with rapid advancement in bioinformatics has led to comprehensive molecular studies on the model plants as well as non-model plants with economic importance. Furthermore, the application of bioinformatics can be used to elucidate the complex biological processes generated from molecular datasets that would unravel hypotheses of the gene's functions, protein interactions, and other molecular mechanisms efficiently [47–49]. However, public databases host inaccurate information on the putative roles of TPs due to various limitations in the gene- and protein-sequence annotation processes and erroneous mismatching between genomic and functional data on protein function. Therefore, we propose a 'guilt-by-association' (GBA) approach to identify and characterise possible GSL TPs involved in GSL metabolism [32]. The GBA principle has been used to identify regulators [50–55] and enzymes [56–59] involved in GSL biosynthesis. Detailed information on most of the molecular components related to GSL biosynthesis and metabolism can be found in SuCComBase, which is accessible at http://plant-scc.org (accessed on 16 December 2021) [60].

In this paper, we describe the process for searching for potential GSL genes that may be involved in the transportation of GSL-related components in *A. thaliana*. Firstly, a GSL co-expression network was constructed in search of the co-expressed genes, which was followed by GO enrichment analysis to infer the function of the identified potential GSL genes. Three criteria were designed to facilitate the selection of potential GSL TPs: (i) involvement in transport and localisation, (ii) sharing similar expression patterns with known GSL genes, and (iii) having similar subcellular localisation with known GSL TPs.

2. Results

2.1. Data Collection and Establishment

A total of 188 known GSL genes were identified from the literature (55 genes), KEGG (23 genes), and AraCyc (110 genes); however, only 116 were used in this analysis after redundancies were removed. The GSL genes were identified using "glucosinolate" and "GSL" as keywords in the search tab of each database and specifically selecting *Arabidopsis thaliana* datasets (Figure 1). The complete list of known GSL genes is shown in Supplementary Table S1.

2.2. Gene Co-Expression Network Analysis

All 116 known GSL genes were used as queries against transcriptomic data from the four specified co-expression network tools, including ATTED-II (http://atted.jp/ (accessed on 16 February 2021)) [61], AraNet v2 (https://www.inetbio.org/aranet/ (accessed on 16 February 2021)) [62], GeneMANIA (https://genemania.org/ (accessed on 16 February 2021)) [63,64], and STRING (https://string-db.org/ (accessed on 16 February 2021)) [65–67]. All the iden-

tified interactions were combined to form a single gene co-expression network using Cytoscape 3.8.2 [68]. Figure 2 shows the interaction between 116 known GSL genes with 1151 potential GSL genes linked with 14,308 edges. The potential GSL genes are defined as the identified co-expressed genes in the gene network. The integrated co-expression network was generated from 293 nodes and 2265 edges from AraNet; 932 nodes and 2894 edges from ATTED; 213 nodes and 9910 edges from GeneMANIA; and 211 nodes and 4470 edges from STRING. These networks of the individual genes were merged using Cytoscape 3.8.2, resulting in an integrated co-expression network consisting of 1267 nodes and 14,308 edges linking the genes (Figure 2).
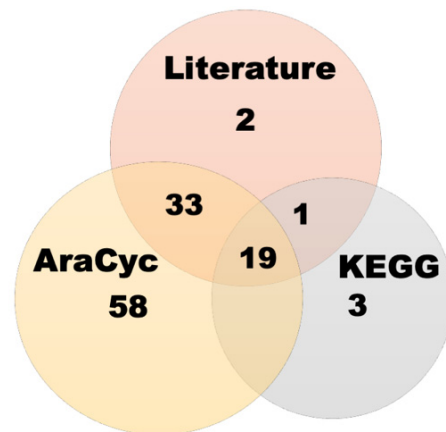

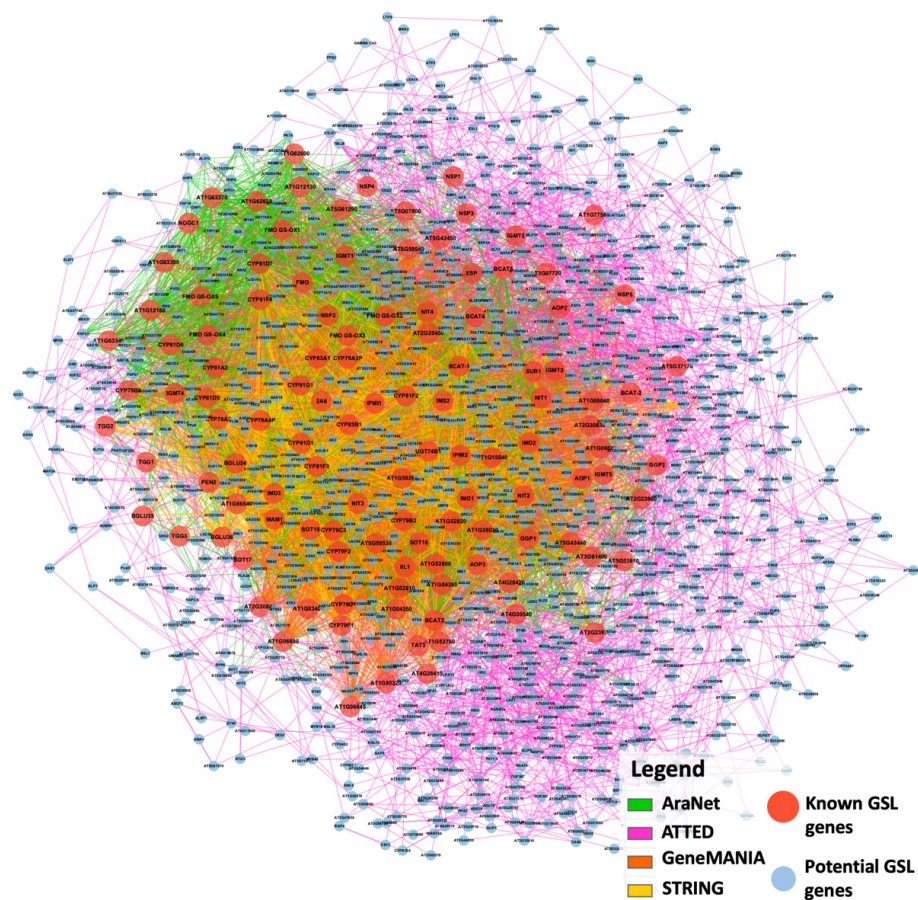**Figure 1.** Identification of GSL genes from the literature, KEGG, and AraCyc.

Figure 2. Integrated co-expression network of GSL genes consisting of 1267 nodes and 14,308 edges.

2.3. GO Enrichment

BINGO was used to analyse the GO enrichment on the constructed gene network. We used the overrepresented GO biological processes (Figure 3) as a guide to search for potential genes that encoded transporter proteins involved in GSL metabolism. The GO enrichment analysis showed that the nodes on localisation and transport were among the overrepresented biological processes.
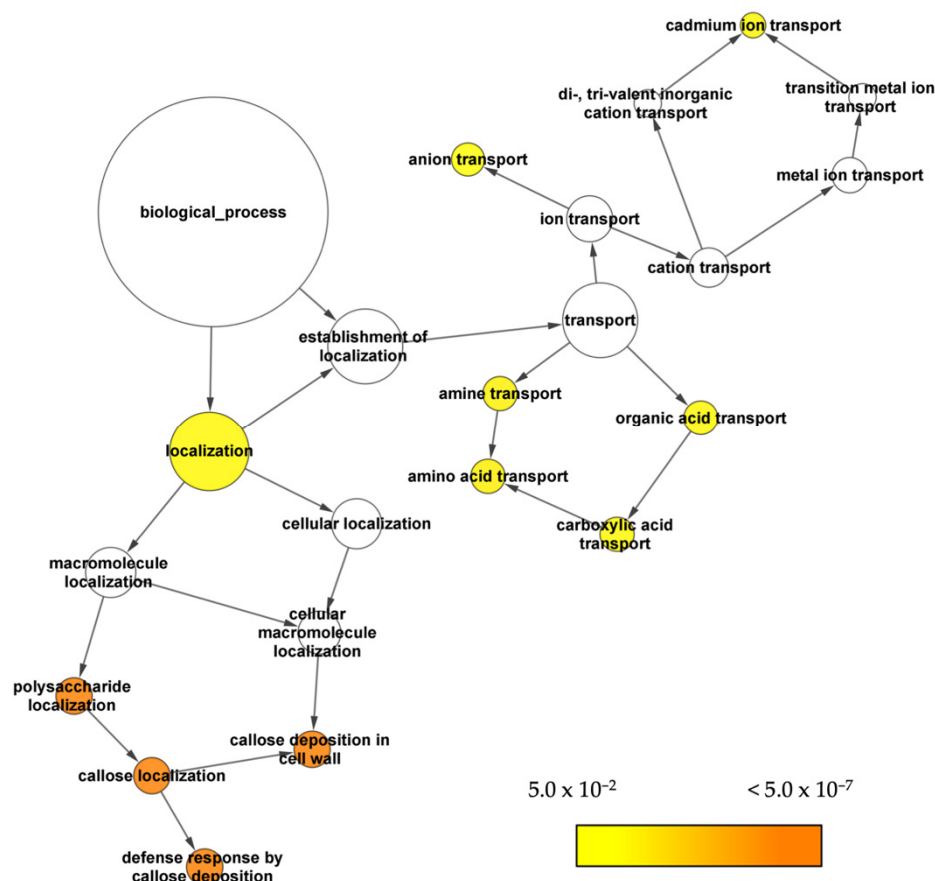

Figure 3. The overrepresented localisation biological processes of GO in the co-expressed GSL genes from the gene network. The significance levels of the overrepresented GO terms are shown using a heatmap, where darker nodes mean more significant ontologies.

2.4. Gene-Expression Pattern Analysis and Visualisation

We used jasmonic-acid-treated *A. thaliana* gene-expression data obtained from Expression Angler (http://bar.utoronto.ca/ExpressionAngler/ (accessed on 12 April 2021)) to validate the potential genes identified in this study. Gene-expression data for *A. thaliana* wild type Col-0 were collected at 30 min, 1 h, and 3 h time points in both control and MeJA-treated conditions. The expression patterns for the selected potential GSL genes compared to those for known GSL genes were generated using ClustVis (https://biit.cs.ut.ee/clustvis/ (accessed on 12 April 2021)) [69] (Figure 4), and known GSL genes were grouped based on their function in GSL biosynthesis. The expression patterns in Expression Angler were calculated using *r*-values based on Pearson's correlation coefficient (PCC). Figure 4 shows that the expression patterns for known GSL genes (*UGT74B1*, *CYP79B2*, *CYP79B3*, *CYP83B1*, and *ABCG36*) were similar to those for 21 potential GSL genes encoding TPs in control and treated conditions (MeJA) in *A. thaliana*.
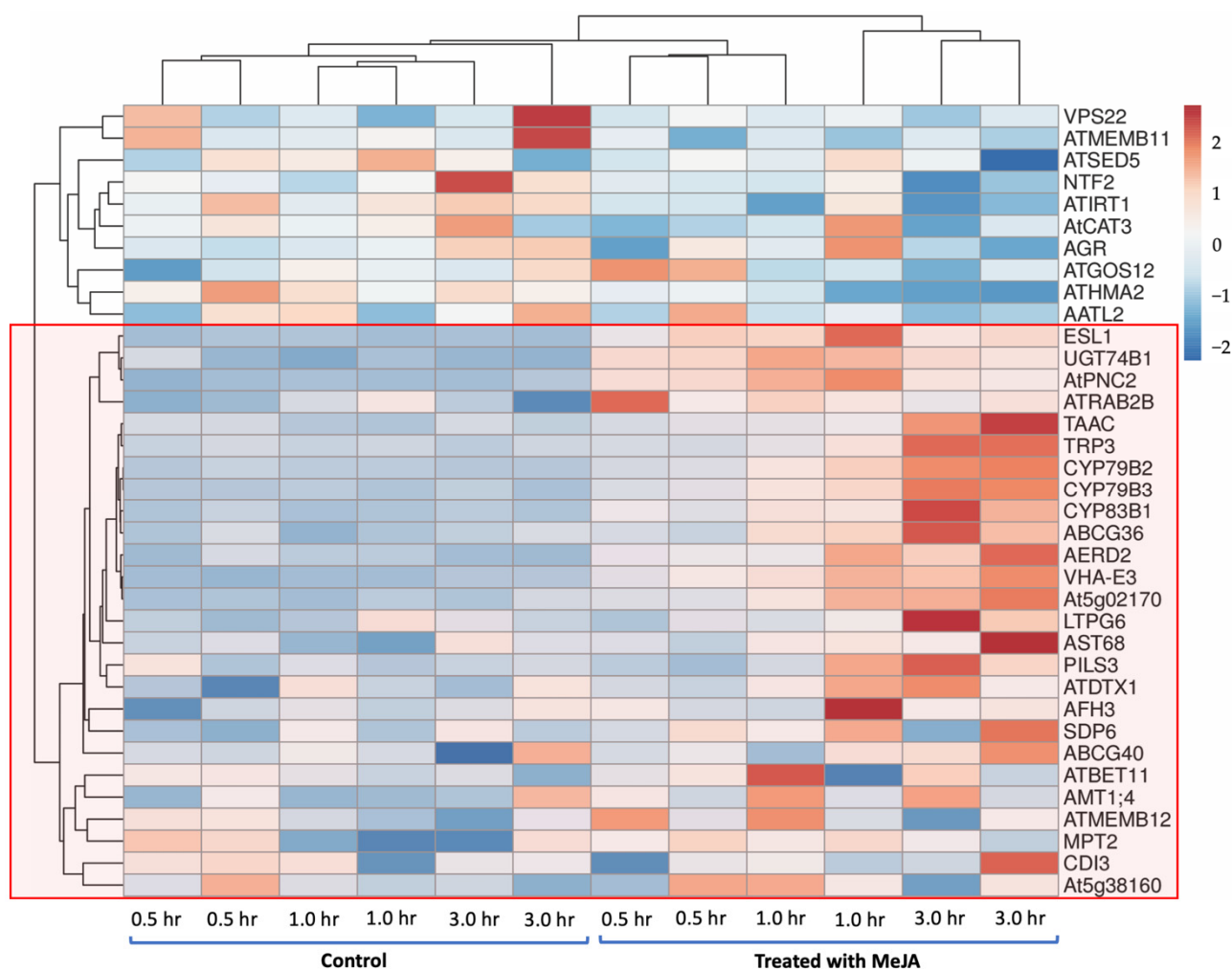
Figure 4. The expression patterns for known GSL genes (*UGT74B1*, *CYP79B2*, *CYP79B3*, *CYP83B1*, and *ABCG36*) and potential GSL genes encoding TPs in control and MeJA–treated conditions in *A. thaliana*.

2.5. Sequence Analysis of Potential GSL TPs

The subcellular location information of each potential GSL TP was retrieved from the SUBA4 database. The information extracted for the known GSL TPs was used as a reference for protein structural analysis to predict the function of those potential GSL TPs in the GSL-biosynthesis pathway. Supplementary Table S2 shows the results collected from various databases such as TAIR, UniProt, GO, and SUBA4. Twenty-one potential GSL genes encoding TPs have expression patterns similar to those of known GSL genes and are associated with localisation and transport processes, including *ESL1*, *AtPNC2*, *AtRAB2B*, *TAAC*, *AERD2*, *VHA-E3*, *At5g02170*, *AST68*, *PILS3*, *AtDTX1*, *ABCG40*, *AtBET11*, *AMT1;4*, *AtMEMB*, *MPT2*, *CDI3*, *TRP3*, *LTPG6*, *At5g38160*, *AFH3*, and *SDP6*. Then, *AST68* and *ABCG40* were selected for further analysis, as they have subcellular localisations similar to those of known GSL TPs, i.e., *SULTR1;1* and *SULTR1;2*, and *ABCG36*, respectively.

2.6. Evolutionary Relationship Analysis of the Potential GSL TPs

Two phylogenetic trees of selected potential GSL TPs (AST68 and ABCG40) and their related sequences were constructed using MEGA11 (Figure 5). AST68 was located in the same clade (clade 1) with known GSL TPs, i.e., SULTR1;1 (AST101) and SULTR1;2. Meanwhile, ABCG40 was grouped into known GSL TPs, i.e., ABCG36 or PEN3 in clade 2.
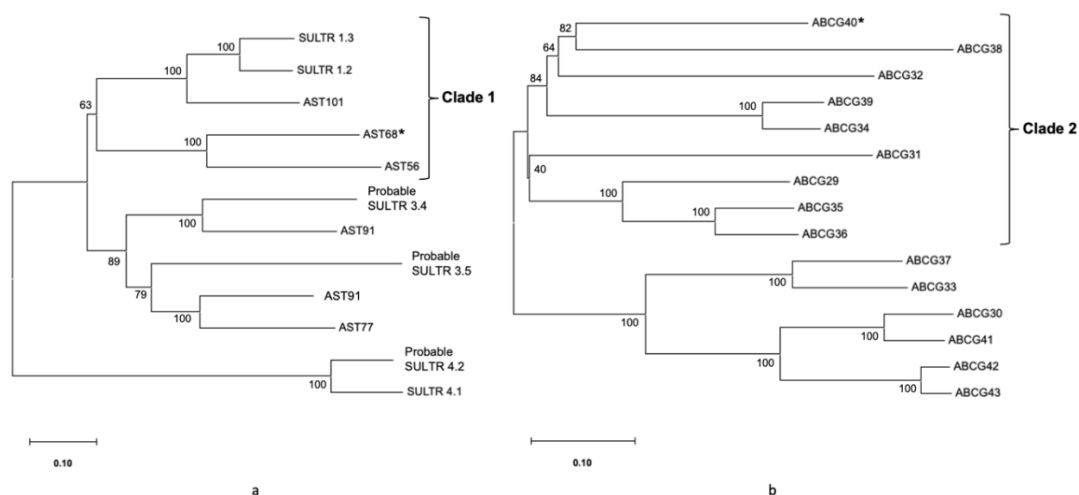
Figure 5. The phylogenetic trees of the selected potential GSL TPs indicated with the symbol "*". (**a**) Phylogenetic tree of AST68 and its related sequences. (**b**) Phylogenetic tree of ABCG40 and its homologs.

2.7. Protein Structure Prediction and Model Evaluation

AST68 and ABCG40 contain 677 and 1423 amino acids, respectively. Possible homologous structures for AST68 are solute carrier family-12 member (PDB ID 7CH1_B), solute carrier family-26 member (PDB ID 6RTC_A), and sulphate transporter (PDB ID 5DA0_A), whilst the ATP-binding cassette sub-family G members (PDB IDs 5DO7_D, 5DO7_C, and 6HZM_A) are homologs of ABCG40. The sequence identity of those homologs is within 20–30%; hence, threading and ab initio approaches were used to predict the tertiary structure of GSL TPs. However, the length of ABCG40 exceeded the maximum number of 1000 amino acids required by most servers; therefore, only the sequence numbers 506–1415 that contain transmembrane and cytoplasmic domains were retained.

trRosetta generated the best 3D models for AST68 and ABCG40, as shown from the MolProbity score, Ramachandran plot, and clashscore (Table 1). A comparison between the 3D model of AST68 (Figure 6a) and its homologs (i.e., SLC29A9 (PDB: 7CH1) and SLC26A9 (PDB: 6RTC)) showed that both domains (i.e., the STAS domain and transmembrane domain) were well superimposed on their homologs (7CH1 (Figure 6b) and 6RTC (Figure 6c)) even though the orientation of the whole structure was different, as shown by the RMSD values (3.65—PDB 7CH1; 3.69—PDB 6RTC). However, a comparison of the 3D model of ABCG40 against its homologs (PDB: 5NJ3, 6HCO, and 5DO7) showed that one of the domains did not superimpose well on its homolog. Thus, the generated 3D model of ABCG40 was eliminated from the molecular-docking analysis.

Table 1. Scores of the potential GSL TP structures predicted by three different servers.

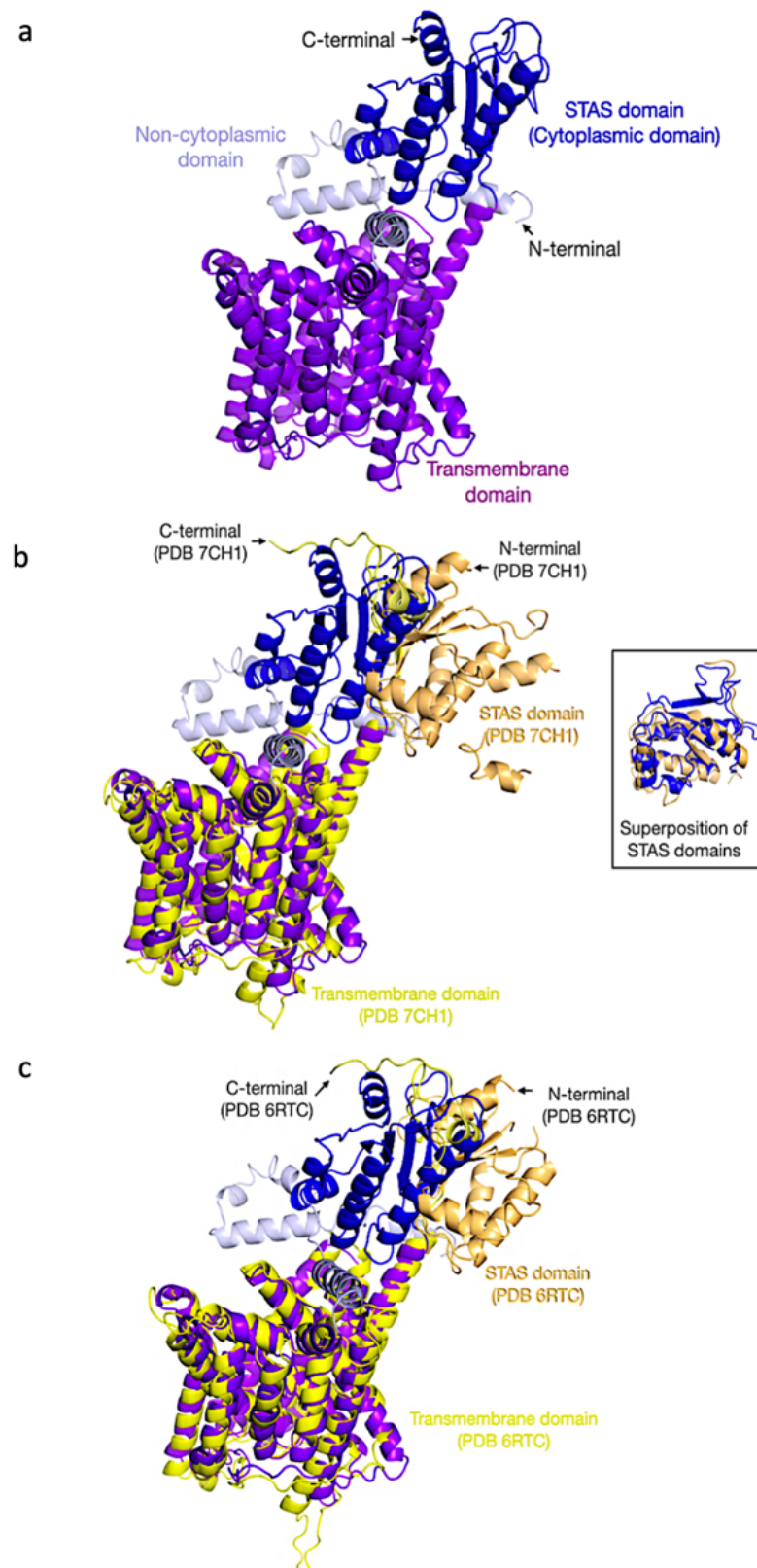| Potential GSL TP	Server	MolProbity Score	Percentage of Residues That Fall Inside Ramachandran-Favoured Regions (%)	Clashscore, All Atoms
AST68	I-TASSER	3.29	80.15	15.6
	Robetta server (trRosetta)	1.38	96.89	4.13
	Raptor-X	5.00	94.67	488.12
ABCG40 (506–1415)	I-TASSER	3.60	65.94	18.74
	Robetta server (trRosetta)	1.37	97.49	5.04
	Raptor-X	4.93	95.31	448.78

Figure 6. The 3D model of AST68 trRosetta and its homologs. (**a**) 3D structure of AST68 model. (**b**) Superimposition of AST68 3D model with SLC29A9 (PDB ID 7CH1), and superimposition of STAS domain is shown in the box. (**c**) Superimposition of AST68 model with SLC26A9 (PDB ID 6RTC).

2.8. Molecular Docking of AST68 with Sulphate Ion

The sulphate ion was docked onto AST68 using AutoDOCK Vina and AutoDOCK 4.2 and showed good binding-energy values of −3.5 kJ/mol and −4.12 kcal/mol, respectively (Figure 7a). Both tools predicted the interactions of the sulphate ion with Ser419 (forming a hydrogen bond) and Val172 (forming a van der Waals interaction) of AST68 (Figure 7b). The sulphate ion bound to the AST68 homolog at the region close to the sodium-ion (in 7CH1)- and chloride-ion-binding regions (in 6RTC) (Figure 7c).
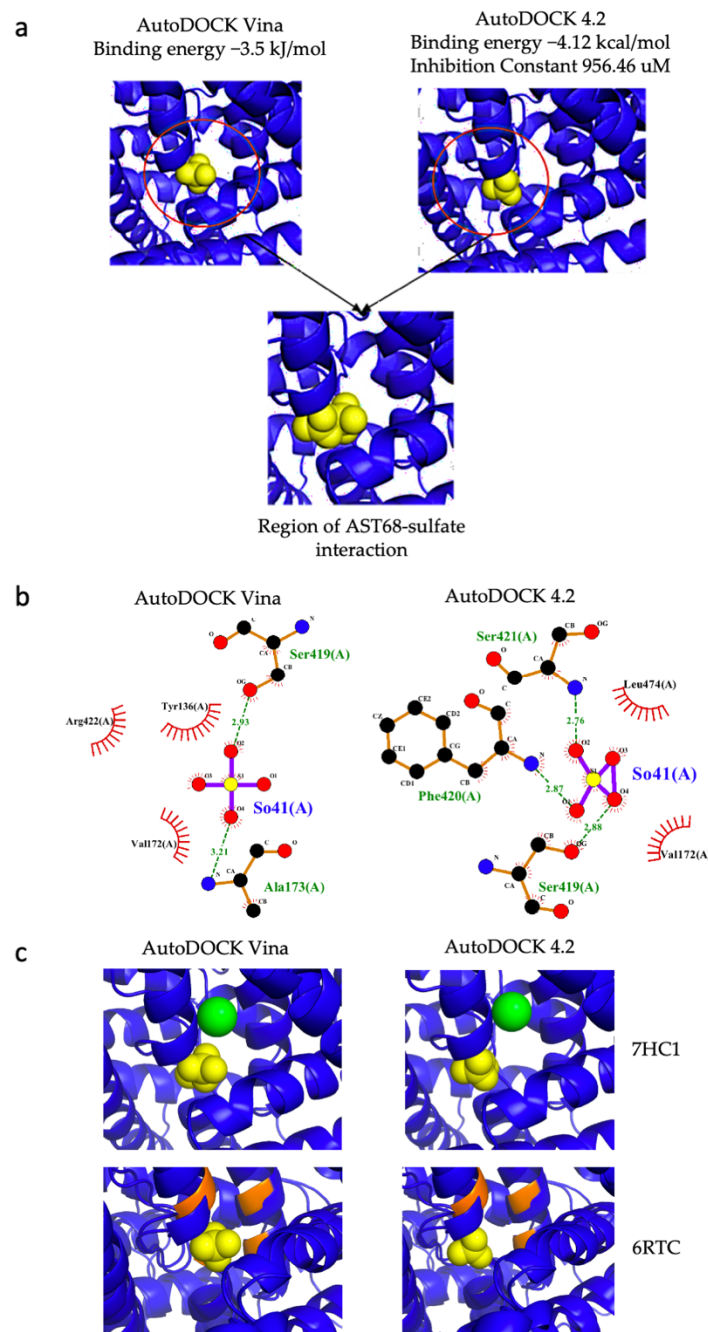
Figure 7. AST68–sulphate interaction. (**a**) Docking of sulphate ion on AST68 ion using AutoDOCK Vina and AutoDOCK 4.2 predicted similar interaction region. (**b**) LigPlot showing hydrogen-bonding interactions of AST68 and sulphate ion with Ser419 and Val172 as predicted by both approaches. (**c**) Comparison with AST68 and its homologs (7CH1 and 6RTC). In 7CH1, the sodium ion is coloured green, and the chloride-binding region in 6RTC is coloured orange.

3. Discussion

There are several limitations in characterising potential GSL TPs responsible for GSL metabolism. First, there remains some inaccurate information in biological databases regarding the roles of TPs. Thus, the annotation of genes and proteins with putative roles in TPs appears to face erroneous matching between genomic and functional data on protein function. Second, this limitation also affects the capability of using traditional homology-based approaches to categorise the TP features and assign the TP substrate specificity information to the physiological details of plants [32]. Therefore, several criteria have been used to identify and select potential GSL TPs: (1) potential TPs that are involved in transport and localisation from the GO analysis; (2) potential GSL genes encoding TPs that share similar expression patterns with known GSL genes in control and treated conditions (MeJA) in *A. thaliana*; and (3) potential TPs that have subcellular localisation similar to that of known GSL TPs. These criteria have been described by Larsen et al. [70]. It also highlighted the in silico-based approaches that employed the 'guilt-by-association' (GBA) principle in identifying transporters in plant specialised metabolism. In relation to GSL biosynthesis, the GBA approach has been used to identify regulators [50–55] and enzymes [56–59]. Identifying TPs using co-expressed genes successfully defined a boron transporter candidate in *A. thaliana* [71]. A similar approach was employed in the non-model plant *Catharanthus roseus*, wherein *CrNPF2.9* was co-expressed in the mono-indole alkaloid (MIA) pathway [72]. To our best knowledge, this is the first study reporting the application of co-expressed genes to identify potential TPs in GSL metabolism. The abundance of publicly available *Arabidopsis* microarray and RNAseq data facilitates the development of in silico techniques to identify candidate genes based on their co-expression with other known genes involved in similar biological processes of interest [70]. Supplementary Figure S1 shows a complete step-by-step procedure for identifying potential genes encoding TPs involved in GSL metabolism.

Twenty-one potential TPs related to transport and localisation have been retrieved, and gene-expression pattern analysis was conducted. We used an expression-based approach to search for genes that encode the GSL transporter in the GSL mechanism. An expression-based approach is usually used to identify transporters from differential-expression patterns in the specialised metabolism of the plant under various conditions or stresses [73]. Bioinformatic analyses were conducted on the genes before and after treatment with methyl jasmonate (MeJA) to observe their expression or response profiles relative to those for known GSLs. JA stimulation causes a mechanism response (movement, secretion, the production of enzymes, and gene expression). In addition, the JA exposure of plants can stimulate secondary-metabolite production. These metabolites play an essential role in plants' responses and adaptation to their natural environment [73]. Based on Figure 4, 21 potential GSL genes encoding TPs shared similar expression patterns with the known GSL genes (*UGT74B1*, *CYP79B2*, *CYP79B3*, *CYP83B1*, and *ABCG36*) shown in the red box. For additional protein characterisation, the subcellular location of each putative GSL TP was collected from the SUBA4, TAIR, UniProt, and GO databases.

Next, we selected two potential GSL TPs with the possible substrates in the GSL-biosynthesis mechanism that fulfilled the three criteria in this study: (1) AST68 and (2) ABCG40 (Table 2). The two genes are involved in both transport and localisation, based on GO analysis. They share a similar expression pattern with known GSL genes in control and MeJA-treated conditions in *A. thaliana*. They also have similar subcellular localisation to known GSL TPs. AST68, known as sulphate transporter 2;1, is located in the plasma membrane, similar to SULTR1;1 and SULTR1;2. These two TPs are involved in the GSL sulphur-assimilation process that transports sulphate to the *Arabidopsis* roots [39,40]. The gene expression of *SULTR1;1* and *SULTR1;2* is significantly increased in *Arabidopsis sdi1sdi-* knockout lines. The sulphur-deficiency-induced genes *SD1* and *SD2* are major repressors that control GSL biosynthesis during sulphur deficiency [43]. In the phylogenetic tree, the potential GSL TP was positioned in the same clade (clade 1) as SULTR1;1 (AST101) and SULTR1;2 (Figure 5a), suggesting its possible involvement in GSL sulphur assimilation.

Table 2. Selected potential GSL TPs with possible substrates in the GSL-biosynthesis mechanism.

Known GSL TP (TAIR ID/UniProt ID)	Potential GSL TP (TAIR ID/UniProt ID)	Localisation	Possible Substrate
SULTR1;1 (At4g08620/Q9SAY1) and SULTR1;2 (At1g78000/Q9MAX3)	AST68 (At5g10180/O04722)	Plasma membrane; sulphate transporter 2;1	Sulphate
ABCG36/PEN3 (At1g59870/Q9XIE2)	ABCG40 (At1g15520/Q9M9E1)	Plasma membrane; ABC transporter G family member 40	4OH-X

Another potential GSL TP is the ABC transporter G family member 40 (ABCG40), which is located in the plasma membrane. ABCG40 belongs to the same subfamily as ABCG36. ABC transporters are located in most membranes (e.g., the plasma membrane) and found in all living organisms [74]. There are several types of substrates for this transporter group: small molecules (heavy metals, inorganic acids, and peptides), large molecules (lipids, polysaccharides, and steroids), and intact proteins [75,76]. In plants, these transporters are involved in diverse biological processes, such as responses to pathogens, diffusion-barrier formation, and phytohormone transport [76]. Meanwhile, ABCG36 or PEN3 was proposed to transport distinct indole-derived metabolites once the plant was attacked by pests in the indolic GSL-biosynthetic pathway. In the study, 4-O-β-D-glucosyl-indol-3-yl formamide (4OGlcI3F) was found to be abundant in *pen3 Arabidopsis* leaf, known as pathogen-inducible compounds. Thus, the PEN3 substrate was suggested to be the precursor of 4OGlcI3F for resistance against pests in *Arabidopsis* [77]. However, the underlying mechanism in transporting small molecules across the plasma membrane remains unknown [78]. Figure 5b shows the location of ABCG40 in clade 2, relative to its homologs. A known GSL TP, ABCG36, was found in the same clade as ABCG40, suggesting its possible role as a GSL TP in indolic GSL metabolism.

The 3D protein models of both potential GSL TPs were constructed using trRosetta. The structural analysis of the models against their homologs suggested further analysis of the AST68 model due to the well-superimposed domains of STAS and the transmembrane on the known structures, i.e., 7CH1 (Figure 6b) and 6RTC (Figure 6c). In addition, results from the ModFOLD8 analysis (significant confidence e-value of 1.255×10^{-4}) also suggested the suitability of the docking of sulphate ions on AST68. Furthermore, Chi et al. [79] and Walter et al. [80] demonstrated the ligand's tendency to bind to the transmembrane domain. Thus, we docked the sulphate ion onto the transmembrane domain. Both molecular-docking tools (AutoDOCK Vina and AutoDOCK 4.2) calculated good values of binding energy between the sulphate ion and Ser419 and Val172, with the formation of a hydrogen bond and van der Waals interaction, respectively.

Our proposed in silico-based approaches facilitated the discovery of several potential GSL TPs, which can be experimentally validated. However, due to the limited capability of identifying possible substrates for potential GSL TPs, we selected proteins with similar protein families to the known GSL TPs, including the sulphate transporters (SULTR1;1 and SULTR1;2) and the ABC transporter G family member 36 (ABCG36) or PENETRATION 3 (PEN3). These potential GSL TPs should be validated further using targeted mutation techniques conducted on the model plant, *A. thaliana*. As a result, this knowledge can be applied in other GSL-containing plants producing better yields and showing greater stress tolerance against pests for crop improvement.

4. Materials and Methods

4.1. Data Collection and Construction of the Gene-Co-Expression Network

A comprehensive literature search was performed using relevant literature databases, including PubMed, Google Scholar, and Science Direct. Several relevant keywords (e.g., "glucosinolate" and "glucosinolate pathway") were queried to find known GSL genes. Pathway databases, including Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/ (accessed on 11 February 2021)) [81,82] and AraCyc (https://

www.arabidopsis.org/biocyc/(accessed on 11 February 2021)) [83], were used in querying those databases for the known GSL genes used in this study, searching with the keywords search "glucosinolate" and "GSL". These known GSL genes were used as queries for four co-expression tools—ATTED [61], AraNet v2 [62], GeneMANIA [64,65], and STRING [61]—to identify "additional" co-expressed genes. ATTED is a dedicated co-expression database exclusively for plants for unravelling functionally related genes [61]. AraNet v2, GeneMANIA, and STRING interactions are based on integration from experiments and computational predictions that include co-expression data [62,63,66]. "Additional" genes are defined as potential GSL genes based on the 'guilt-by-association' principle. An integrated gene network was constructed using Cytoscape 3.8.2 [68].

4.2. GO Enrichment Analysis

Gene ontology (GO) analysis was conducted using Cytoscape 3.8.2 with the Biological Network Gene Ontology (BiNGO) plugin [84] to determine the overrepresented GO categories. In addition, a hypergeometric test with a Benjamini and Hochberg false-discovery rate (FDR) was performed using the default parameters for adjusted *p*-values [85].

4.3. Expression-Pattern Analysis

Expression Angler (http://bar.utoronto.ca/ExpressionAngler/ (accessed on 12 April 2021)) was used to obtain relevant information on the genes of interest with similar expression or response profiles in specific conditions or treatments [86]. The expression profiles were extracted from Expression Angler, and the heatmap was generated using ClustVis (https://biit.cs.ut.ee/clustvis/ (accessed on 12 April 2021)) [69].

4.4. Characterisation of Potential GSL TPs

Protein sequences were retrieved from the TAIR10 (The Arabidopsis Information Resource) and UniProt [87] databases for protein-sequence analysis and characterisation. In addition, the SUBA database (The Subcellular Localization of Proteins in Arabidopsis Database) was used to predict the cellular localisation of TPs [88]. Different locations of TPs are presumed to carry different types of GSL derivatives. For example, one known GSL TP, BAT5, is found in chloroplasts and facilitates the localisation of 2-oxo acids from cytosol chloroplasts [45,46].

4.5. Sequence Analysis of GSL TPs

The following analysis was conducted for (1) potential TPs associated with transport and localisation from the GO analysis, (2) potential GSL genes encoding TPs that had expression patterns similar to those of known GSL genes in control and treated conditions with methyl jasmonate (MeJA) in *A. thaliana*, and (3) potential TPs that had subcellular localisation similar to that for known GSL TPs.

4.6. Construction of Phylogenetic Tree

The protein sequences of the selected GSL TPs were used as queries for a sequence-similarity search using BLASTP at https://blast.ncbi.nlm.nih.gov/Blast.cgi (accessed on 30 December 2021) [89]. The UniProt database and *Arabidopsis thaliana* were selected against the annotated protein sets and between the *A. thaliana* paralogs as queries using default parameters. The retrieved sequences were subject to multiple sequence alignments using MAFFT at https://www.ebi.ac.uk/Tools/msa/mafft/ (accessed on 31 December 2021) [90]. The aligned sequences were used to construct phylogenetic trees using the neighbour-joining method in the MEGA software (version 11) [91,92]. One thousand replicates were used to obtain bootstrapping values in the constructed phylogenetic trees.

4.7. Protein-Structure Prediction and Model Evaluations

The tertiary-structure prediction of GSL TPs was conducted using threading and ab initio methods. The models were predicted by I-TASSER [93], Robetta (trRosetta) [94],

and Raptor-X [95]. The quality check for each model was evaluated using the MolProbity score [96], Ramachandran plot, and Clashscore. ModFOLD8 [97] was used to calculate the best scoring model.

4.8. Molecular Docking of Potential GSL Transporters

The structure and relevant information of the potential substrate for GSL TP were obtained from the PubChem database [98]. Molecular docking between GSL TP and its substrate was performed using AutoDOCK Vina [99] and AutoDOCK 4.2 [100] to obtain a consensus prediction of the binding-site region.

5. Conclusions

This study demonstrated the use of a computational approach to identify potential GSL TPs from co-expression data. The selected genes coding for TPs (*AST68* and *ABCG40*) were identified using three criteria that were used in the selection process: (a) involvement in transport and localisation biological processes, (b) sharing similar expression patterns with known GSL genes, and (c) having subcellular localisation similar to that of known GSL TPs. The application of these criteria was based on the 'guilt-by-association' (GBA) principle to identify and characterise possible GSL TPs efficiently. Two 3D models were generated, and further analysis was conducted on AST68 due to the well-superimposed essential domains of the homologs. The molecular-docking study was conducted on the 3D model of AST68 to determine its interaction with the sulphate ion to support its function as a sulphate transporter in GSL metabolism. The results from this study could be experimentally validated in the targeted verification of gene expression and metabolite data in *A. thaliana*. Furthermore, applying this bioinformatics approach will increase the ability to screen and characterise plant TPs on a large-scale basis to understand the mechanical properties of GSL metabolism in *A. thaliana*.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/life12030326/s1. Figure S1: The step-by-step procedure to identify potential genes encoding GSL TPs; Table S1: List of known GSL genes identified from the literature, KEGG, and AraCyc; Table S2: List of potential GSL transporters based on gene-expression patterns with known GSL genes and biological processes.

Author Contributions: Conceptualisation, S.H. and Z.-A.M.-H.; methodology, S.H., N.A.-A. and Z.-A.M.-H.; formal analysis, S.H., F.I.A.H. and S.D.L.; data curation, S.H. and F.I.A.H.; writing—original draft preparation, S.H.; writing—review and editing, S.H., S.D.L., N.A.-A. and Z.-A.M.-H.; visualisation, S.H., N.A.-A., F.I.A.H. and S.D.L.; supervision, S.H., S.D.L. and Z.-A.M.-H.; funding acquisition, Z.-A.M.-H. All the authors have read and agreed to the published version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Manghwar, H.; Hussain, A.; Ullah, A.; Gul, S.; Shaban, M.; Khan, A.H.; Ali, M.; Sani, S.G.A.S.; Chaudhary, H.J.; Munis, M.F.H. Expression analysis of defense related genes in wheat and maize against Bipolaris sorokiniana. *Physiol. Mol. Plant Pathol.* **2018**, *103*, 36–46. [CrossRef]
2. Manghwar, H.; Hussain, A. Mechanism of tobacco osmotin gene in plant responses to biotic and abiotic stress tolerance: A brief history. *Biocell* **2022**, *46*, 623–632. [CrossRef]
3. Isah, T. Stress and defense responses in plant secondary metabolites production. *Biol. Res.* **2019**, *52*, 39. [CrossRef]
4. Chhajed, S.; Misra, B.B.; Tello, N.; Chen, S. Chemodiversity of the Glucosinolate-Myrosinase System at the Single Cell Type Resolution. *Front. Plant Sci.* **2019**, *10*, 618. [CrossRef] [PubMed]
5. Clay, N.K.; Adio, A.M.; Denoux, C.; Jander, G.; Ausubel, F.M. Glucosinolate Metabolites Required for an Arabidopsis Innate Immune Response. *Science* **2009**, *323*, 95–101. [CrossRef] [PubMed]
6. Fahey, J.W.; Zalcmann, A.T.; Talalay, P. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* **2001**, *56*, 5–51. [CrossRef]
7. Reichelt, M.; Brown, P.D.; Schneider, B.; Oldham, N.; Stauber, E.; Tokuhisa, J.; Kliebenstein, D.; Mitchell-Olds, T.; Gershenzon, J. Benzoic acid glucosinolate esters and other glucosinolates from Arabidopsis thaliana. *Phytochemistry* **2002**, *59*, 663–671. [CrossRef]
8. Ishida, M.; Hara, M.; Fukino, N.; Kakizaki, T.; Morimitsu, Y. Glucosinolate metabolism, functionality and breeding for the improvement of Brassicaceae vegetables. *Breed. Sci.* **2014**, *64*, 48–59. [CrossRef]
9. Falk, K.L.; Tokuhisa, J.; Gershenzon, J. The Effect of Sulfur Nutrition on Plant Glucosinolate Content: Physiology and Molecular Mechanisms. *Plant Biol.* **2007**, *9*, 573–581. [CrossRef]
10. Barba, F.J.; Nikmaram, N.; Roohinejad, S.; Khelfa, A.; Zhu, Z.; Koubaa, M. Bioavailability of Glucosinolates and Their Breakdown Products: Impact of Processing. *Front. Nutr.* **2016**, *3*, 24. [CrossRef]
11. Seo, M.-S.; Kim, J.S. Understanding of MYB Transcription Factors Involved in Glucosinolate Biosynthesis in Brassicaceae. *Molecules* **2017**, *22*, 1549. [CrossRef] [PubMed]
12. Harun, S.; Abdullah-Zawawi, M.-R.; Goh, H.-H.; Mohamed-Hussein, Z.-A. A Comprehensive Gene Inventory for Glucosinolate Biosynthetic Pathway in Arabidopsis thaliana. *J. Agric. Food Chem.* **2020**, *68*, 7281–7297. [CrossRef] [PubMed]
13. Blažević, I.; Montaut, S.; Burčul, F.; Olsen, C.E.; Burow, M.; Rollin, P.; Agerbirk, N. Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry* **2019**, *169*, 112100. [CrossRef]
14. Halkier, B.A.; Gershenzon, J. Biology and Biochemistry of Glucosinolates. *Annu. Rev. Plant Biol.* **2006**, *57*, 303–333. [CrossRef] [PubMed]
15. Liu, Y.; Rossi, M.; Liang, X.; Zhang, H.; Zou, L.; Ong, C.N. An Integrated Metabolomics Study of Glucosinolate Metabolism in Different Brassicaceae Genera. *Metabolites* **2020**, *10*, 313. [CrossRef]
16. Wittstock, U.; Meier, K.; Dörr, F.; Ravindran, B.M. NSP-Dependent Simple Nitrile Formation Dominates upon Breakdown of Major Aliphatic Glucosinolates in Roots, Seeds, and Seedlings of Arabidopsis thaliana Columbia-0. *Front. Plant Sci.* **2016**, *7*, 1821. [CrossRef] [PubMed]
17. Chhajed, S.; Mostafa, I.; He, Y.; Abou-Hashem, M.; El-Domiaty, M.; Chen, S. Glucosinolate Biosynthesis and the Glucosinolate–Myrosinase System in Plant Defense. *Agronomy* **2020**, *10*, 1786. [CrossRef]
18. Winde, I.; Wittstock, U. Insect herbivore counteradaptations to the plant glucosinolate–myrosinase system. *Phytochemistry* **2011**, *72*, 1566–1575. [CrossRef]
19. Koroleva, O.A.; Davies, A.; Deeken, R.; Thorpe, M.R.; Tomos, A.D.; Hedrich, R. Identification of a New Glucosinolate-Rich Cell Type in Arabidopsis Flower Stalk. *Plant Physiol.* **2000**, *124*, 599–608. [CrossRef]
20. Koroleva, O.A.; Cramer, R. Single-cell proteomic analysis of glucosinolate-rich S-cells in Arabidopsis thaliana. *Methods* **2011**, *54*, 413–423. [CrossRef]
21. Kissen, R.; Bones, A.M. Nitrile-specifier Proteins Involved in Glucosinolate Hydrolysis in Arabidopsis thaliana. *J. Biol. Chem.* **2009**, *284*, 12057–12070. [CrossRef] [PubMed]
22. Petersen, B.L.; Chen, S.; Hansen, C.H.; Olsen, C.E.; Halkier, B. Composition and content of glucosinolates in developing Arabidopsis thaliana. *Planta* **2002**, *214*, 562–571. [CrossRef] [PubMed]
23. Magrath, R.; Mithen, R. Maternal Effects on the Expression of Individual Aliphatic Glucosinolates in Seeds and Seedlings of Brassica napus. *Plant Breed.* **1993**, *111*, 249–252. [CrossRef]
24. van Dam, N.M.; Tytgat, T.O.G.; Kirkegaard, J.A. Root and shoot glucosinolates: A comparison of their diversity, function and interactions in natural and managed ecosystems. *Phytochem. Rev.* **2008**, *8*, 171–186. [CrossRef]
25. Tsunoda, T.; Grosser, K.; Van Dam, N.M. Locally and systemically induced glucosinolates follow optimal defence allocation theory upon root herbivory. *Funct. Ecol.* **2018**, *32*, 2127–2137. [CrossRef]
26. Meldau, S.; Erb, M.; Baldwin, I.T. Defence on demand: Mechanisms behind optimal defence patterns. *Ann. Bot.* **2012**, *110*, 1503–1514. [CrossRef]
27. Touw, A.J.; Mogena, A.V.; Maedicke, A.; Sontowski, R.; van Dam, N.M.; Tsunoda, T. Both Biosynthesis and Transport Are Involved in Glucosinolate Accumulation During Root-Herbivory in Brassica rapa. *Front. Plant Sci.* **2020**, *10*, 1653. [CrossRef]
28. Tsunoda, T.; Krosse, S.; Van Dam, N.M. Root and shoot glucosinolate allocation patterns follow optimal defence allocation theory. *J. Ecol.* **2017**, *105*, 1256–1266. [CrossRef]

29. Hunziker, P.; Halkier, B.A.; Schulz, A. Arabidopsis glucosinolate storage cells transform into phloem fibres at late stages of development. *J. Exp. Bot.* **2019**, *70*, 4305–4317. [CrossRef]

30. Koroleva, O.A.; Gibson, T.M.; Cramer, R.; Stain, C. Glucosinolate-accumulating S-cells in Arabidopsis leaves and flower stalks undergo programmed cell death at early stages of differentiation. *Plant J.* **2010**, *64*, 456–469. [CrossRef]

31. Borpatragohain, P.; Rose, T.; King, G.J. Fire and Brimstone: Molecular Interactions between Sulfur and Glucosinolate Biosynthesis in Model and Crop Brassicaceae. *Front. Plant Sci.* **2016**, *7*, 1735. [CrossRef]

32. David, R.; Byrt, C.S.; Tyerman, S.D.; Gilliham, M.; Wege, S. Roles of membrane transporters: Connecting the dots from sequence to phenotype. *Ann. Bot.* **2019**, *124*, 201–208. [CrossRef] [PubMed]

33. Nour-Eldin, H.H.; Andersen, T.; Burow, M.; Madsen, S.R.; Jørgensen, M.E.; Olsen, C.E.; Dreyer, I.; Hedrich, R.; Geiger, D.; Halkier, B. NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature* **2012**, *488*, 531–534. [CrossRef]

34. Andersen, T.G.; Nour-Eldin, H.H.; Fuller, V.L.; Olsen, C.E.; Burow, M.; Halkier, B.A. Integration of Biosynthesis and Long-Distance Transport Establish Organ-Specific Glucosinolate Profiles in Vegetative Arabidopsis. *Plant Cell* **2013**, *25*, 3133–3145. [CrossRef]

35. Madsen, S.R.; Olsen, C.E.; Nour-Eldin, H.H.; Halkier, B.A. Elucidating the Role of Transport Processes in Leaf Glucosinolate Distribution. *Plant Physiol.* **2014**, *166*, 1450–1462. [CrossRef] [PubMed]

36. Moussaieff, A.; Rogachev, I.; Brodsky, L.; Malitsky, S.; Toal, T.W.; Belcher, H.; Yativ, M.; Brady, S.M.; Benfey, P.N.; Aharoni, A. High-resolution metabolic mapping of cell types in plant roots. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E1232–E1241. [CrossRef] [PubMed]

37. Andersen, T.G.; Halkier, B.A. Upon bolting the GTR1 and GTR2 transporters mediate transport of glucosinolates to the inflorescence rather than roots. *Plant Signal. Behav.* **2014**, *9*, e27740. [CrossRef] [PubMed]

38. Madsen, S.R.; Kunert, G.; Reichelt, M.; Gershenzon, J.; Halkier, B.A. Feeding on Leaves of the Glucosinolate Transporter Mutant gtr1gtr2 Reduces Fitness of Myzus persicae. *J. Chem. Ecol.* **2015**, *41*, 975–984. [CrossRef]

39. Koprivova, A.; Kopriva, S. Molecular mechanisms of regulation of sulfate assimilation: First steps on a long road. *Front. Plant Sci.* **2014**, *5*, 589. [CrossRef]

40. Morikawa-Ichinose, T.; Kim, S.-J.; Allahham, A.; Kawaguchi, R.; Maruyama-Nakashita, A. Glucosinolate Distribution in the Aerial Parts of sel1-10, a Disruption Mutant of the Sulfate Transporter SULTR1;2, in Mature Arabidopsis thaliana Plants. *Plants* **2019**, *8*, 95. [CrossRef]

41. Hirai, M.Y.; Fujiwara, T.; Awazuhara, M.; Kimura, T.; Noji, M.; Saito, K. Global expression profiling of sulfur-starved Arabidopsis by DNA macroarray reveals the role of O -acetyl-l -serine as a general regulator of gene expression in response to sulfur nutrition. *Plant J.* **2003**, *33*, 651–663. [CrossRef] [PubMed]

42. Nikiforova, V.; Freitag, J.; Kempa, S.; Adamik, M.; Hesse, H.; Hoefgen, R. Transcriptome analysis of sulfur depletion inArabidopsis thaliana: Interlacing of biosynthetic pathways provides response specificity. *Plant J.* **2003**, *33*, 633–650. [CrossRef]

43. Aarabi, F.; Kusajima, M.; Tohge, T.; Konishi, T.; Gigolashvili, T.; Takamune, M.; Sasazaki, Y.; Watanabe, M.; Nakashita, H.; Fernie, A.R.; et al. Sulfur deficiency–induced repressor proteins optimize glucosinolate biosynthesis in plants. *Sci. Adv.* **2016**, *2*, e1601087. [CrossRef] [PubMed]

44. Maruyama-Nakashita, A.; Nakamura, Y.; Tohge, T.; Saito, K.; Takahashi, H. Arabidopsis SLIM1 Is a Central Transcriptional Regulator of Plant Sulfur Response and Metabolism. *Plant Cell* **2006**, *18*, 3235–3251. [CrossRef] [PubMed]

45. Gigolashvili, T.; Yatusevich, R.; Rollwitz, I.; Humphry, M.; Gershenzon, J.; Flügge, U.-I. The Plastidic Bile Acid Transporter 5 Is Required for the Biosynthesis of Methionine-Derived Glucosinolates inArabidopsis thaliana. *Plant Cell* **2009**, *21*, 1813–1829. [CrossRef] [PubMed]

46. Sawada, Y.; Toyooka, K.; Kuwahara, A.; Sakata, A.; Nagano, M.; Saito, K.; Hirai, M.Y. Arabidopsis Bile Acid:Sodium Symporter Family Protein 5 is Involved in Methionine-Derived Glucosinolate Biosynthesis. *Plant Cell Physiol.* **2009**, *50*, 1579–1586. [CrossRef]

47. Pushpparaj, P.N. *Introduction to Functional Bioinformatics*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 235–254. [CrossRef]

48. Ayaz, A.; Saqib, S.; Huang, H.; Zaman, W.; Lü, S.; Zhao, H. Genome-wide comparative analysis of long-chain acyl-CoA synthetases (LACSs) gene family: A focus on identification, evolution and expression profiling related to lipid synthesis. *Plant Physiol. Biochem.* **2021**, *161*, 1–11. [CrossRef]

49. Ayaz, A.; Huang, H.; Zheng, M.; Zaman, W.; Li, D.; Saqib, S.; Zhao, H.; Lü, S. Molecular Cloning and Functional Analysis of *GmLACS2-3* Reveals Its Involvement in Cutin and Suberin Biosynthesis along with Abiotic Stress Tolerance. *Int. J. Mol. Sci.* **2021**, *22*, 9175. [CrossRef]

50. Ashari, K.-S.; Abdullah-Zawawi, M.-R.; Harun, S.; Mohamed-Hussein, Z.-A. Reconstruction of the Transcriptional Regulatory Network in Arabidopsis thaliana Aliphatic Glucosinolate Biosynthetic Pathway. *Sains Malays.* **2018**, *47*, 2993–3002. [CrossRef]

51. Hirai, M.; Sugiyama, K.; Sawada, Y.; Tohge, T.; Obayashi, T.; Suzuki, A.; Araki, R.; Sakurai, N.; Suzuki, H.; Aoki, K.; et al. Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6478–6483. [CrossRef]

52. Gigolashvili, T.; Berger, B.; Mock, H.-P.; Müller, C.; Weisshaar, B.; Flügge, U.-I. The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in Arabidopsis thaliana. *Plant J.* **2007**, *50*, 886–901. [CrossRef] [PubMed]

53. Gigolashvili, T.; Yatusevich, R.; Berger, B.; Müller, C.; Flügge, U.-I. The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis inArabidopsis thaliana. *Plant J.* **2007**, *51*, 247–261. [CrossRef] [PubMed]

54. Sønderby, I.E.; Hansen, B.G.; Bjarnholt, N.; Ticconi, C.; Halkier, B.A.; Kliebenstein, D.J. A Systems Biology Approach Identifies a R2R3 MYB Gene Subfamily with Distinct and Overlapping Functions in Regulation of Aliphatic Glucosinolates. *PLoS ONE* **2007**, *2*, e1322. [CrossRef]

55. Harun, S.; Rohani, E.R.; Ohme-Takagi, M.; Goh, H.-H.; Mohamed-Hussein, Z.-A. ADAP is a possible negative regulator of glucosinolate biosynthesis in Arabidopsis thaliana based on clustering and gene expression analyses. *J. Plant Res.* **2021**, *134*, 327–339. [CrossRef] [PubMed]

56. Harun, S.; Afiqah-Aleng, N.; Karim, M.B.; Amin, A.U.; Kanaya, S.; Mohamed-Hussein, Z.-A. Potential Arabidopsis thaliana glucosinolate genes identified from the co-expression modules using graph clustering approach. *PeerJ* **2021**, *9*, e11876. [CrossRef]

57. Knill, T.; Schuster, J.; Reichelt, M.; Gershenzon, J.; Binder, S. Arabidopsis Branched-Chain Aminotransferase 3 Functions in Both Amino Acid and Glucosinolate Biosynthesis. *Plant Physiol.* **2007**, *146*, 1028–1039. [CrossRef]

58. Sawada, Y.; Kuwahara, A.; Nagano, M.; Narisawa, T.; Sakata, A.; Saito, K.; Hirai, M.Y. Omics-Based Approaches to Methionine Side Chain Elongation in Arabidopsis: Characterization of the Genes Encoding Methylthioalkylmalate Isomerase and Methylthioalkylmalate Dehydrogenase. *Plant Cell Physiol.* **2009**, *50*, 1181–1190. [CrossRef]

59. Geu-Flores, F.; Nielsen, M.T.; Nafisi, M.; Møldrup, M.E.; Olsen, C.E.; Motawia, M.S.; Halkier, B. Glucosinolate engineering identifies a γ-glutamyl peptidase. *Nat. Chem. Biol.* **2009**, *5*, 575–577. [CrossRef]

60. Harun, S.; Abdullah-Zawawi, M.-R.; A-Rahman, M.R.A.; Muhammad, N.A.N.; Mohamed-Hussein, Z.-A. SuCComBase: A manually curated repository of plant sulfur-containing compounds. *Database* **2019**, *2019*, baz021. [CrossRef]

61. Aoki, Y.; Okamura, Y.; Tadaka, S.; Kinoshita, K.; Obayashi, T. ATTED-II in 2016: A Plant Coexpression Database Towards Lineage-Specific Coexpression. *Plant Cell Physiol.* **2015**, *57*, e5. [CrossRef]

62. Lee, T.; Yang, S.; Kim, E.; Ko, Y.; Hwang, S.; Shin, J.; Shim, J.E.; Shim, H.; Kim, H.; Kim, C.; et al. AraNet v2: An improved database of co-functional gene networks for the study of Arabidopsis thaliana and 27 other nonmodel plant species. *Nucleic Acids Res.* **2014**, *43*, D996–D1002. [CrossRef] [PubMed]

63. Montojo, J.; Zuberi, K.; Rodriguez, H.; Bader, G.D.; Morris, Q. GeneMANIA: Fast gene network construction and function prediction for Cytoscape. *F1000Research* **2014**, *3*, 153. [CrossRef] [PubMed]

64. Warde-Farley, D.; Donaldson, S.L.; Comes, O.; Zuberi, K.; Badrawi, R.; Chao, P.; Franz, M.; Grouios, C.; Kazi, F.; Lopes, C.T.; et al. The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **2010**, *38*, W214–W220. [CrossRef] [PubMed]

65. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K.P.; et al. STRING v10: Protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **2015**, *43*, D447–D452. [CrossRef]

66. 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. [CrossRef]

67. Szklarczyk, D.; Morris, J.H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N.T.; Roth, A.; Bork, P.; et al. The STRING database in 2017: Quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **2017**, *45*, D362–D368. [CrossRef]

68. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]

69. Metsalu, T.; Vilo, J. ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **2015**, *43*, W566–W570. [CrossRef]

70. Larsen, B.; Xu, D.; Halkier, B.A.; Nour-Eldin, H.H. Advances in methods for identification and characterization of plant transporter function. *J. Exp. Bot.* **2017**, *68*, 4045–4056. [CrossRef]

71. Takano, J.; Wada, M.; Ludewig, U.; Schaaf, G.; von Wirén, N.; Fujiwara, T. The Arabidopsis Major Intrinsic Protein NIP5;1 Is Essential for Efficient Boron Uptake and Plant Development under Boron Limitation. *Plant Cell* **2006**, *18*, 1498–1509. [CrossRef]

72. Payne, R.; Xu, D.; Foureau, E.; Carqueijeiro, I.; Oudin, A.; De Bernonville, T.D.; Novak, V.; Burow, M.; Olsen, C.-E.; Jones, M.; et al. An NPF transporter exports a central monoterpene indole alkaloid intermediate from the vacuole. *Nat. Plants* **2017**, *3*, 16208. [CrossRef]

73. De Geyter, N.; Gholami, A.; Goormachtig, S.; Goossens, A. Transcriptional machineries in jasmonate-elicited plant secondary metabolism. *Trends Plant Sci.* **2012**, *17*, 349–359. [CrossRef]

74. Kang, J.; Park, J.; Choi, H.; Burla, B.; Kretzschmar, T.; Lee, Y.; Martinoia, E. Plant ABC Transporters. *Arab. Book* **2011**, *9*, e0153. [CrossRef]

75. Biemans-Oldehinkel, E.; Doeven, M.K.; Poolman, B. ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett.* **2005**, *580*, 1023–1035. [CrossRef] [PubMed]

76. Gräfe, K.; Schmitt, L. The ABC transporter G subfamily in Arabidopsis thaliana. *J. Exp. Bot.* **2020**, *72*, 92–106. [CrossRef]

77. Lu, X.; Dittgen, J.; Piślewska-Bednarek, M.; Molina, A.; Schneider, B.; Svatoš, A.; Doubský, J.; Schneeberger, K.; Weigel, D.; Bednarek, P.; et al. Mutant Allele-Specific Uncoupling of PENETRATION3 Functions Reveals Engagement of the ATP-Binding Cassette Transporter in Distinct Tryptophan Metabolic Pathways. *Plant Physiol.* **2015**, *168*, 814–827. [CrossRef]

78. Bednarek, P.; Piślewska-Bednarek, M.; Svatoš, A.; Schneider, B.; Doubský, J.; Mansurova, M.; Humphry, M.; Consonni, C.; Panstruga, R.; Sanchez-Vallet, A.; et al. A Glucosinolate Metabolism Pathway in Living Plant Cells Mediates Broad-Spectrum Antifungal Defense. *Science* **2009**, *323*, 101–106. [CrossRef] [PubMed]

79. Chi, X.; Jin, X.; Chen, Y.; Lu, X.; Tu, X.; Li, X.; Zhang, Y.; Lei, J.; Huang, J.; Huang, Z.; et al. Structural insights into the gating mechanism of human SLC26A9 mediated by its C-terminal sequence. *Cell Discov.* **2020**, *6*, 55. [CrossRef]

80. Walter, J.D.; Sawicka, M.; Dutzler, R. Cryo-EM structures and functional characterization of murine Slc26a9 reveal mechanism of uncoupled chloride transport. *eLife* **2019**, *8*, e46986. [CrossRef] [PubMed]

81. Kanehisa, M.; Sato, Y.; Furumichi, M.; Morishima, K.; Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **2019**, *47*, D590–D595. [CrossRef] [PubMed]

82. Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **1999**, *27*, 29–34. [CrossRef]

83. Mueller, L.A.; Zhang, P.; Rhee, S.Y. AraCyc: A Biochemical Pathway Database for Arabidopsis. *Plant Physiol.* **2003**, *132*, 453–460. [CrossRef]

84. Maere, S.; Heymans, K.; Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **2005**, *21*, 3448–3449. [CrossRef] [PubMed]

85. Benjamini, Y.; Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **2001**, *29*, 1165–1188. [CrossRef]

86. Toufighi, K.; Brady, S.M.; Austin, R.; Ly, E.; Provart, N.J. The Botany Array Resource: E-Northerns, Expression Angling, and promoter analyses. *Plant J.* **2005**, *43*, 153–163. [CrossRef]

87. Bateman, A.; Martin, M.J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E.H.; Britto, R.; Bursteinas, B.; et al. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef]

88. Hooper, C.M.; Castleden, I.R.; Tanz, S.; Aryamanesh, N.; Millar, A.H. SUBA4: The interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Res.* **2016**, *45*, D1064–D1074. [CrossRef] [PubMed]

89. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

90. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]

91. 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. [CrossRef]

92. Stecher, G.; Tamura, K.; Kumar, S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol. Biol. Evol.* **2020**, *37*, 1237–1239. [CrossRef] [PubMed]

93. Yang, J.; Zhang, Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*, W174–W181. [CrossRef] [PubMed]

94. Yang, J.; Anishchenko, I.; Park, H.; Peng, Z.; Ovchinnikov, S.; Baker, D. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1496–1503. [CrossRef] [PubMed]

95. Xu, J.; McPartlon, M.; Li, J. Improved protein structure prediction by deep learning irrespective of co-evolution information. *Nat. Mach. Intell.* **2021**, *3*, 601–609. [CrossRef] [PubMed]

96. Williams, C.J.; Headd, J.J.; Moriarty, N.W.; Prisant, M.G.; Videau, L.L.; Deis, L.N.; Verma, V.; Keedy, D.A.; Hintze, B.J.; Chen, V.B.; et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018**, *27*, 293–315. [CrossRef]

97. McGuffin, L.J.; Aldowsari, F.M.F.; Alharbi, S.M.A.; Adiyaman, R. ModFOLD8: Accurate global and local quality estimates for 3D protein models. *Nucleic Acids Res.* **2021**, *49*, W425–W430. [CrossRef]

98. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395. [CrossRef]

99. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef]

100. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [CrossRef]