

In-silico Exploration of Channel Type and Efflux Silicon Transporters and Silicification Proteins in 80 Sequenced Viridiplantae Genomes

Muhammad Amjad Nawaz ¹, Farrukh Azeem ², Alexander Mikhailovich Zakharenko ³, Xiao Lin ⁴, Rana Muhammad Atif ⁵, Faheem Shehzad Baloch ⁶, Ting-Fung Chan ⁴, Gyuhwa Chung ⁷, Junghee Ham ⁸, Sangmi Sun ^{7,*} and Kirill S. Golokhvast ^{3,9,10,*}

¹ Laboratory of Bio-Economics and Biotechnology, Department of Bio-Economics and Food Safety, School of Economics and Management, Far Eastern Federal University, 690950 Vladivostok, Russia; amjad_ucauos@yahoo.com

² Department of Bioinformatics and Biotechnology, Government College University, Faisalabad 38000, Pakistan; farrukh@gcuf.edu.pk

³ N.I. Vavilov All-Russian Institute of Plant Genetic Resources, 42, 44 Bolshaya Morskaya Street, 190000 St. Petersburg, Russia; rarf@yandex.ru

⁴ Center for Soybean Research, State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong SAR, Hong Kong, China; alanlamsiu@gmail.com (X.L.); tf.chan@cuhk.edu.hk (T.-F.C.)

⁵ US-Pakistan Centre for Advanced Studies in Agriculture and Food Security, University of Agriculture, Faisalabad 38040, Pakistan; dratif@uaf.edu.pk

⁶ Faculty of Agricultural Sciences and Technologies, Sivas University of Science and Technology, 58140 Sivas, Turkey; balochfaheem13@gmail.com

⁷ Department of Biotechnology, Chonnam National University, Chonnam 59626, Korea; chung@chonnam.ac.kr

⁸ Department of Health Policy and Management, Wonkwang University, Iksan, Jeonbuk 54538, Korea; hamssine@hamssine.co.kr

⁹ Education and Scientific Center of Nanotechnology, Far Eastern Federal University, 690950 Vladivostok, Russia

¹⁰ Pacific Geographical Institute, FEB RAS, 7 Radio street, 690014 Vladivostok, Russia

* Corresponding authors: smsun@chonnam.ac.kr (S.S.); droopy@mail.ru (K.S.G.)

Received: 17 October 2020; Accepted: 18 November 2020; Published: 20 November 2020

Abstract: Silicon (Si) accumulation protects plants from biotic and abiotic stresses. It is transported and distributed within the plant body through a cooperative system of channel type (e.g., *OsLsi1*) and efflux (*Lsi2s* e.g., *OsLsi2*) Si transporters (SITs) that belong to Noduline-26 like intrinsic protein family of aquaporins and an uncharacterized anion transporter family, respectively. Si is deposited in plant tissues as phytoliths and the process is known as biosilicification but the knowledge about the proteins involved in this process is limited. In the present study, we explored channel type SITs and *Lsi2s*, and *siplant1* protein (*Slp1*) in 80 green plant species. We found 80 channel type SITs and 133 *Lsi2s*. The channel type SITs characterized by the presence of two NPA motifs, GSGR or STAR selectivity filter, and 108 amino acids between two NPA motifs were absent from Chlorophytes, while Streptophytes evolved two different types of channel type SITs with different selectivity filters. Both channel type SITs and *Lsi2s* evolved two types of gene structures each, however, *Lsi2s* are ancient and were also found in Chlorophyta. Homologs of *Slp1* (225) were present in almost all Streptophytes regardless of their Si accumulation capacity. In Si accumulator plant species, the *Slp1s* were characterized by the presence of H, D-rich domain, P, K, E-rich domain, and P, T, Y-rich domain, while moderate Si accumulators lacked H, D-rich domain and P, T, Y-rich domains. The digital expression analysis and coexpression networks highlighted the role of channel type and *Lsi2s*, and how *Slp1* homologs were ameliorating plants' ability to withstand different stresses by

co-expressing with genes related to structural integrity and signaling. Together, the in-silico exploration made in this study increases our knowledge of the process of biosilicification in plants.

Keywords: biosilicification; channel type Si transporter; influx transporter; NIPs; silica accumulation; Siliplant 1; Si efflux; Viridiplantae

1. Introduction

Silicon (Si) exerts beneficial effects on the growth and productivity of plants and has been recognized as a beneficial element by the International Plant Nutrition Institute [1]. Decades of research on Si have authenticated its importance towards plant's ability to withstand biotic and abiotic stresses [2]. Si accumulation in plants is regarded as beneficial regardless of their Si accumulation capacity; when deprived of Si, plants are susceptible to stresses as compared to the ones grown in Si. In soil, Si is present as monosilicic acid $[\text{Si}(\text{OH})_4]$ at pH 9 with variable concentrations based on soil type [3]. Silicic acid is taken up through a cooperative system of channel type Si transporters (SITs) and efflux silicon transporters (homologs of *OsLsi2*, from now on Lsi2s) [4]. The identification of these SITs in rice i.e., *OsLsi1* and *OsLsi2* was a breakthrough, which led to a substantial understanding of Si transport in plants [5,6]. The first identified channel type SIT (*OsLsi1*) is an influx SIT and a member of Noduline-26-like intrinsic proteins (NIPs) [5]. Functional characterization of *OsLsi1* homologs in different plant species have enabled us to understand that the Si specificity is associated with two highly conserved NPA motifs (Asn-Pro-Al) and an aromatic/arginine (ar/R) region formed by four residues which function as selectivity filter (SF). The Lsi1s have characteristic SFs i.e., Gly-Ser-Gly-Arg (GSGR) or Ser-Thr-Ala-Arg (STAR) [7,8]. In addition, a precise spacing of 108 amino acids (AAs) (or 109 AAs according to [9]) between the two NPA motifs is essential for Si permeability in plants [10]. These channel type SITs facilitate the transport of Si across the plasma membrane between the external solution (i.e., apoplast) and the plant cells. After entering in the cell, Si moves further across the same cell (endodermis) or different cell types (epidermal, hypodermal, and cortical cells) by Lsi2, depending on the plant species [4,6]. The rice Lsi2 (*OsLsi2*) has 11 predicted transmembrane domains (TMDs) and it belongs to an uncharacterized anion transporter family. It lacks homology with channel type SIT and limited studies have described any conserved characteristics within the characterized as well as predicted Lsi2s [2,11]. The cooperative system of both channel type SITs and Lsi2s enable plants to translocate Si to the aerial parts of the plants where it is deposited as amorphous $\text{Si}(\text{SiO}_2)$ [12].

The amorphous Si deposits are defined as phytoliths [13] and the process of deposition is called biosilicification [14]. Si is deposited in different locations within plants i.e., root endodermis, leaf epidermis, and abaxial epidermis of inflorescence bracts where it can accumulate as phytoliths within the cell lumen or directly laid down on the cell wall matrix; cell wall matrix provides structural template on which Si can be laid down [12,15]. The process of biosilicification crosses kingdoms of living organisms and can be found in diverse eukaryotic lineages [16]. Recent studies have demonstrated that SITs appeared in plants due to the selectivity shift allowing subfunctionalization of ancestral NIPs [17]. Furthermore, the bacterial arsenic Lsi2s are the source of modern-day seed plant nutrient transporters. Therefore, the process of biosilicification is ancient [16] and has evolved in diverse kingdoms and lower phylogenetic clads differently. Although many studies explored the evolution of biosilicification in relevance to plants [18], limited studies attempted to explore the proteins that are responsible for biosilicification upon the arrival of Si at depositing sites within plant [19]. However, contrary to plants, the process of biosilicification is well understood in diatoms [20]. The functional characterization of biosilicification-related proteins in diatoms did not help much in the identification of proteins involved in this process because of no sequence homology between plant and diatom silicification proteins [21,22]. So far, the research in nonplant lineage biosilicification has demonstrated that the silicification related proteins must have certain features such as proline, lysine, and glutamine-acid-rich (P, K, E-rich) domain(s), RXL domain [23,24], and several proline

residues [25]. These characteristics and the developments in RNA sequencing have recently aided the characterization of a Siliplant1 (Slp1) protein in sorghum [19].

Developments in genomics and next generation sequencing are rapidly increasing the available plant genomes in public repositories [26]. These developments allow us to identify homologs and fill the missing links in different pathways such as biosilicification. With the confirmation that NIPs and arsenite transporters having essential signatures can transport Si across membranes, we report the discovery of SITs in 80 sequenced viridiplantae genomes. Additionally, we present the first large scale dataset of putative Slps in plants. This effort will allow a precise understanding of the biosilicification process in plants and will greatly supplement the studies related to evolution of exploration of biosilicification.

2. Material and Methods

2.1. Data Retrieval

Oryza sativa channel type SIT *OsLsi1* (LOC_Os02g51110), efflux Si transporter *OsLsi2* (LOC_Os03g01700), and *Sorghum bicolor* siliplant1 (*SbSlp1*, *Sobic.001G266300*) were used as query sequences for BlastP 2.2.28+ searches using the default parameters against 80 Viridiplantae genomes available from Phytozome v12.1; Joint Genome Institute (JGI; <https://www.phytozome.net>) [27]. The 80 species belonged to all major phylogenetic clads of Viridiplantae (Table 1). All the retrieved results were collected and manually curated to produce a nonredundant dataset that was then subjected to subsequent analysis.

Table 1. The 80 Viridiplantae species and respective genome versions used in this study.

Species	Genome Version	Species	Genome Version
<i>Amaranthus hypochondriacus</i>	V2.1	<i>Kalanchoe laxiflora</i>	V1.1
<i>Amborella trichopoda</i>	V1.0	<i>Lactuca sativa</i>	V8
<i>Anacardium occidentale</i>	V0.9	<i>Linum usitatissimum</i>	V1.0
<i>Ananas comosus</i>	V3	<i>Malus domestica</i>	V1.0
<i>Aquilegia coerulea</i>	V3.1	<i>Manihot esculenta</i>	V6.1
<i>Arabidopsis halleri</i>	V1.1	<i>Marchantia polymorpha</i>	V3.1
<i>Arabidopsis lyrata</i>	V2.1	<i>Medicago truncatula</i>	4.0V1
<i>Arabidopsis thaliana</i>	TAIR10	<i>Micromonas pusilla</i> CCMP1545	V3.0
<i>Asparagus officinalis</i>	V1.1	<i>Mimulus guttatus</i>	V2.0
<i>Boechera stricta</i>	V1.2	<i>Miscanthus sinensis</i>	V7.1
<i>Botryococcus braunii</i>	V2.1	<i>Musa acuminata</i>	V1
<i>Brachypodium distachyon</i>	V3.1	<i>Olea europaea</i> Var. <i>sylvestris</i>	V1.0
<i>Brachypodium hybridum</i>	V1.1	<i>Oropetium thomacum</i>	V1.0
<i>Brachypodium stacei</i>	V1.1	<i>Oryza sativa</i>	V7_JGI
<i>Brachypodium sylvaticum</i>	V1.1	<i>Ostreococcus lucimarinus</i>	V2.0
<i>Brassica oleracea capitata</i>	V1.0	<i>Panicum hallii</i>	V3.1
<i>Brassica rapa FPsc</i>	V1.3	<i>Panicum Virgatum</i>	V4.1
<i>Capsella grandiflora</i>	V1.1	<i>Phaseolus Vulgaris</i>	V2.1
<i>Capsella rubella</i>	V1.0	<i>Physcomitrella patens</i>	V3.3
<i>Carica papaya</i>	V0.4	<i>Populus deltoides</i> WV94	V2.1
<i>Chenopodium quinoa</i>	V1.0	<i>Populus trichocarpa</i>	V3.1
<i>Chlamydomonas reinhardtii</i>	V5.5	<i>Prunus persica</i>	V2.1
<i>Chromochloris zofingiensis</i>	V5.2.3.2	<i>Ricinus communis</i>	V0.1
<i>Cicer arietinum</i>	V1.0	<i>Salix purpurea</i>	V1.0
<i>Citrus clementina</i>	V1.0	<i>Selaginella moellendorffii</i>	V1.0

<i>Citrus sinensis</i>	V1.1	<i>Setaria italica</i>	V2.2
<i>Coccomyxa subellipsoidea</i> C-169	V2.0	<i>Setaria Viridis</i>	V2.1
<i>Coffea arabica</i>	UCDV0.5	<i>Solanum lycopersicum</i> iTAG	2.4
<i>Cucumis sativus</i>	V1.0	<i>Solanum tuberosum</i>	V4.03
<i>Daucus carota</i>	V2.0	<i>Sorghum bicolor</i>	V3.1.1
<i>Dunaliella salina</i>	V1.0	<i>Sphagnum fallax</i>	V0.5
<i>Eucalyptus grandis</i>	V2.0	<i>Spirodela polyrhiza</i>	V2
<i>Eutrema salsugineum</i>	V1.0	<i>Theobroma cacao</i>	V1.1
<i>Fragaria Vesca</i>	V1.1	<i>Trifolium pratense</i>	V2
<i>Glycine max</i> Wm82	a2.V1	<i>Triticum aestivum</i>	V2.2
<i>Gossypium hirsutum</i>	V1.1	<i>Vigna unguiculata</i>	V1.1
<i>Gossypium raimondii</i>	V2.1	<i>Vitis Vinifera</i>	12X
<i>Helianthus annuus</i>	r1.2	<i>Volvox carteri</i>	V2.1
<i>Hordeum Vulgare</i>	r1	<i>Zea mays</i>	Ensembl-18
<i>Kalanchoe fedtschenkoi</i>	V1.1	<i>Zostera marina</i>	V2.2

Available at Phytozome v12.1; Joint Genome Institute (JGI; <https://www.phytozome.net>) [27].

2.2. Sequence Analyses

The retrieved sequences were aligned in MEGA X using built-in MUSCLE. For channel type SITs, the SF and NPA motif amino acids (AAs) were manually searched against *OsLsi1*. The distance between the 2 NPA motifs was calculated. Transmembrane domains (TMDs) were predicted by TMHMM server (cbs.dtu.dk/services/TMHMM/; [28]). Protein molecular weight (Mw) and theoretical isoelectric point (pI) were computed in protein identification and analysis tools on the ExPASy Server (https://web.expasy.org/compute_pi/) [29]. Amino acid frequencies were calculated in TargetP 1.1 Server [30]. Protein subcellular localization was predicted in TargetP-2.0 Server [30]. Repeats in peptide sequences were predicted in RADAR (ebi.ac.uk/Tools/pfa/radar/) [31]. Conserved motifs were found in the peptide sequences using MEME tool (meme.nbcr.net/meme/tools/meme) [32]. SignalP was used to predict subcellular localization and cleavage sites [33].

2.3. Phylogenetic Analyses

Separate alignments for channel type SITs, Lsi2s, and Slp1s were generated using MUSCLE in MEGA X [34]. The alignments were submitted to IQ-TREE (<http://www.cibiv.at/software/iqtree>) with default parameters for tree inference [35]. ModelFinder [36] was used to find the best fit model according to the Bayesian information criterion with ultrafast bootstrap [37]. The trees were visualized and edited in iTOL (<http://itol.embl.de/>) [38]. Estimates of evolutionary divergence between sequences were computed in MEGA X using the Poisson correction model [39].

2.4. Prediction of Silica Precipitation Potential of Selected Peptide Sequences

Selected peptide sequences (repeats) were submitted to protein secondary structure ROSETTA server which uses consensus results from 3 programs, including HHpred, RaptorX, and SPARKS [40]. The generated peptide models were used to study alanine scanning and molecular interactions with $\text{Si}(\text{OH})_3\text{O}^-$. Ligplot v.4.5.3 software implemented in MOLECULAR OPERATING ENVIRONMENT (MOE) were used for prediction of this interactions [41]. The figures were generated in Icn3D v 2.19.0 and ChemWindow V6.0 (Bio-Rad Laboratories).

2.5. Digital Expression and Co-Expression Network Analysis

The digital expression comparison and graphical presentation for selected genes in rice, soybean, and tomato was performed in ‘eFP browser’ of BAR tools (The Bio-Analytic Resource for Plant Biology, <http://bar.utoronto.ca/>) [42,43]. Coexpression network comparison was performed with Network comparer tool in PlaNet (<http://aranet.mpimp-golm.mpg.de/index.html>; [44]).

3. Results

3.1. Identification of SITs and Slps

3.1.1. Channel Type SITs in Viridiplantae

Using the characterized *OsLsi1* as a query sequence, BLAST searches against 80 plant genomes resulted in 1158 homologs belonging to 78 species (Supplementary Table S1). The homologs were screened on the basis of BlastP 2.2.28+ searches using the default parameters in Phytozome. The criteria used to screen the database for homologs/candidate genes may have limited the detection of poorly annotated, or truncated sequence reads. These putative channel type SITs belonged to aquaporin NIP gene family or uncharacterized proteins. The screening for STAR or GSGR SF and two NPA motifs resulted in 80 genes [10]. In total, 3 genes had 109 AAs between both NPA motifs while the rest had 108 AAs. The channel type SITs were characterized by the presence of five or six TMDs similar to *OsLsi1* [5]. One *F. vesca* gene (*gene03812-v1.0-hybrid*) had 10 TMDs. A very small number of genes had variation in their second NPA motif where Ala (A) was replaced by Val (V) (Figure 1a). In total, 1 of 80 genes had STAR SF (*AHYPO_007318*). It was reported earlier that a P-to-L mutation in pumpkin causes miss-localization of channel type SITs into endoplasmic reticulum [45] but this mutation was not found in the 80 channel type SITs.

The phylogenetic tree showed that the monocots formed a separate clad, which was divided into two subclads. Members of the two subclads were grouped with *OsLsi1* and *OsLsi6*, respectively. The monocot specific clad contained four banana homologs that grouped as a parent branch in this clad. The gene containing STAR SF did not group with the rest of the channel type SITs except one papaya gene (*gene03812-v1.0-hybrid*) but both genes had different SF. The homologs in Fabales (soybean, red clover, and chickpea) clustered together. Similarly, genes belonging to Chenopodiaceae and Malvaceae grouped together. Interestingly, our study involved two Caryophyllales i.e., *C. quinoa* (Chenopodiaceae) and *A. hypochondriacus* (Amaranthaceae), which resulted in one candidate gene from each species but had different SF i.e., GSGR and STAR, respectively (Figure 1a). Other homologs from *C. sativus* (Cucurbitales), *R. communis*, *P. deltoids*, *P. trichocarpa* (Malpighiales), *S. lycopersicum* (Solanales), *M. domestica*, *P. persica* (Rosales), *V. vinifera* (vitales), *A. coerulea* (Ranunculales), *M. guttatus* (Lamiales), *C. papaya* (Brassicales), *D. carota* (Apiales), *S. purpurea* (Ericales) *M. esculanta* (*mushroom*) formed a separate clad. We identified three highly conserved motifs in 80 channel type SITs (Figure 1b). Furthermore, previous studies reported that there could be a correlation between the sequence similarity and Si transport capacity [11]. In this regard, we computed the estimates of evolutionary divergence between sequences. Interestingly, it was found that the sequences from known Si accumulators had higher similarity with *OsLsi1* (Figure 1c).

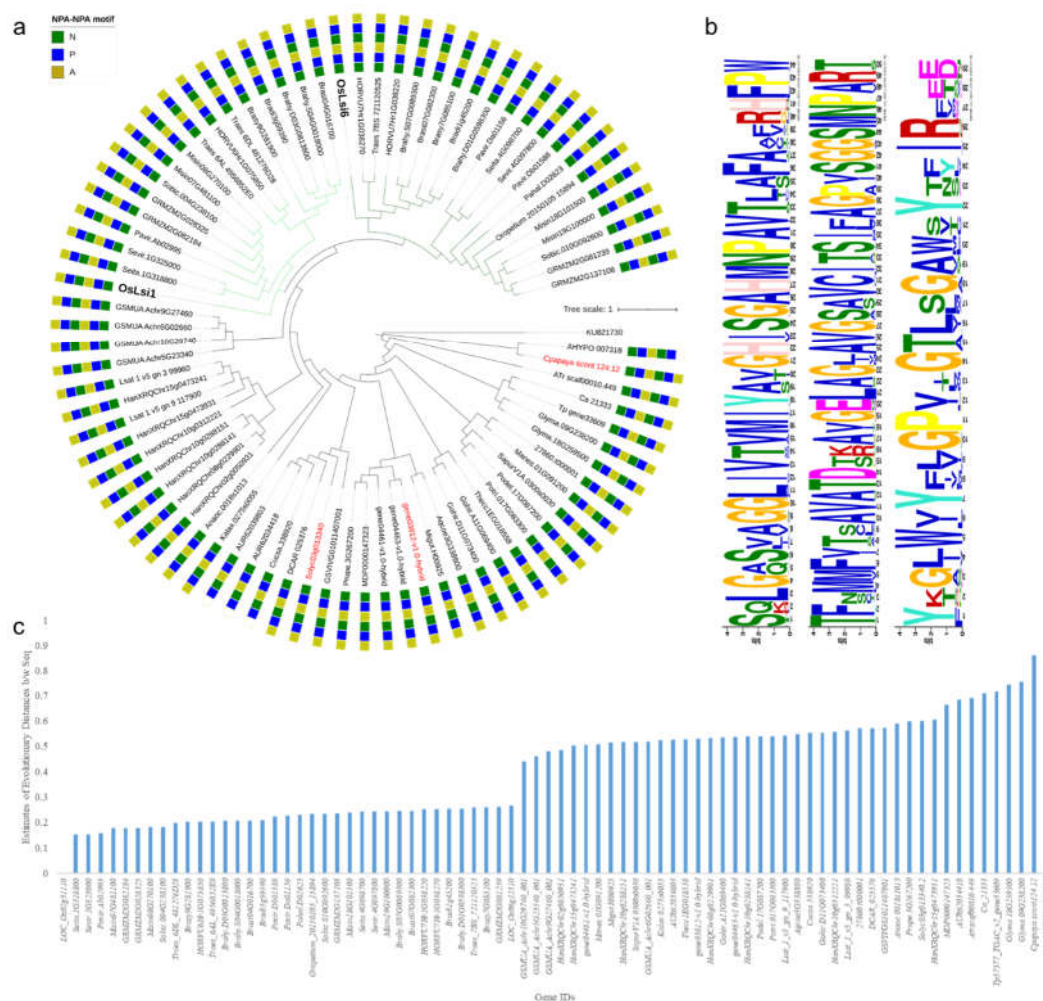


Figure 1. (a) Maximum-likelihood tree of *OsLsi1* homologs (80) in studied viridiplantae genomes. KU821730 (*Spongosphaera streptacantha* SIT-L gene) was used as an outgroup. The sequences were then aligned by MUSCLE in MEGA X and exported to IQ-Tree. The tree was generated using substitution model JTT + I + G4 as a model of rate heterogeneity and Ultrafast Bootstrap with 1000 replicates. The red colored genes have 109 AAs between the NPA motifs. The empty golden square represents the presence of AA other than Ala i.e., Val. The light green clad color shows *OsLsi1* subclad, the dark green color shows *OsLsi6* subclad. (b) Conserved motifs present in plant channel type SITs. (c) Estimation of evolutionary divergence between *OsLsi1* and the identified genes.

3.1.2. Lsi2s in Viridiplantae

The BlastP search using *OsLsi2* as a query against 80 viridiplantae genomes resulted in the identification of 255 genes. To screen the *OsLsi2* gene homologs, first we searched the known Lsi2s in rice (*OsLsi2* and *OsLsi3*), maize (*ZmLsi2*), cucumber (*CsLsi2*), barley (*HVLsi2*), and pumpkin (*CmLsi2-1* and *CmLsi2-2*) to visualize any conserved sequences in them [46] (Figure 2a). The alignment showed that the *OsLsi2* residues between 212 to 343 were highly variable and the rest were conserved regions (Supplementary Figure S1a). By using MEME tool, we found three highly conserved motifs (Supplementary Figure S1b). According to a previous report, the Lsi2s contain 9–11 TMDs [46] and in the current study we found 133 genes having at least 9 TMDs. These 133 genes contained the three conserved motifs present in known Lsi2s (Figure 2b). The pI ranged from 4.87 to 9.41 and the Mw ranged from 8457.98 kDa to 101310.2 kDa. One gene i.e., *AHYPO_010712* had the highest Mw (114798.26 kDa) and the largest number of TMDs i.e., 18 (Supplementary Table S2).

The maximum-likelihood (ML) phylogenetic tree formed one major clad rooted on one gene i.e., *OsLsi3*, which further divided the genes into clads (Figure 2c). The first two clads (shown in green) consisted of members of Poales (*Poaceae*), while, all other genes formed a separate clad (shown in

orange). The genes belonging to early plant lineages i.e., *P. patens*, *S. moellendorffii*, *M. polymorpha*, and *S. fallax* were grouped together (represented by a green circle). The remaining genes grouped in a major clad forming two subclads i.e., monocot specific (represented by a dark yellow circle) and dicot specific (represented by a blue circle) (Figure 2c). Since some monocot genes formed a separate subclad with dicots, therefore, we wanted to know if there is any distinct motif present within the genes that formed clad with known rice Lsi2s (shown in green). We found that these genes had a much-conserved motif i.e., XXTKHXWFXXCXXXRX (Figure 2d). This highly conserved region is located in rice Lsi2s exactly before the sixth TMD. This region was highly variable in all early lineage, dicot, and remaining monocot homologs. Since, the Equisetum (*Equisetum arvense*) Lsi2s i.e., *EaLsi2s*, and *CmLsi2s* did not contain this motif, therefore, based on the studies conducted in homologs of Lsi2 in different plant species, it could be suggested that it might have no functional importance. A further characterization study of this region could increase our understanding in this regard. It suggested a possibility of differential evolution for these Lsi2s. For further confirmation, we introduced the peptide sequences of *EaLsi2-1* and *EaLsi2-2* and prepared the ML tree where *EaLsi2s* grouped with ancestral plant lineages instead of rice or other dicots (Supplementary Figure S2).

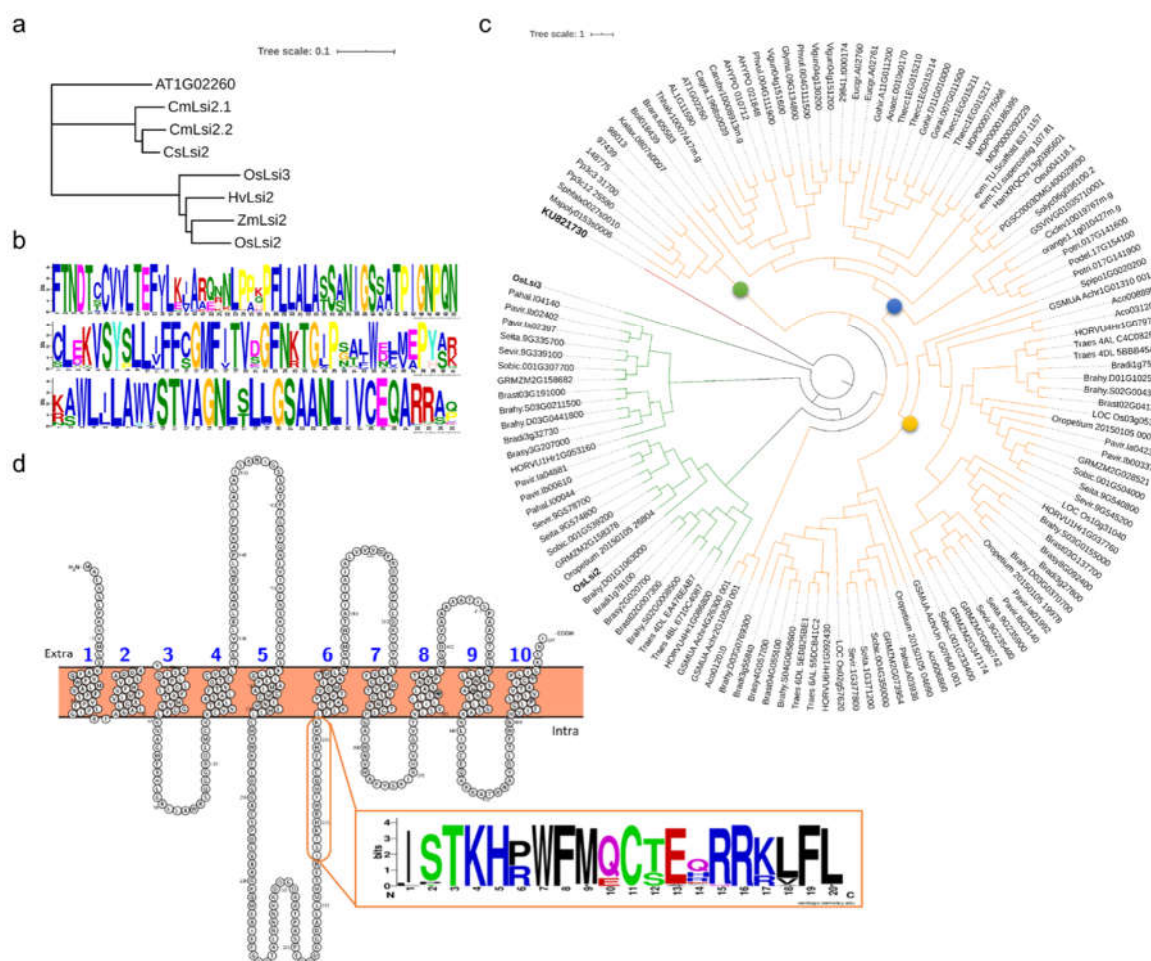


Figure 2. (a) Neighbor-joining tree of Lsi2s that have been characterized. (b) Conserved motifs found in all *OsLsi2* homologs, and (c) ML tree of Lsi2s in studied viridiplantae genomes. *KU821730* (*Spongosphaera streptacantha* SIT-L gene) was used as an outgroup. The sequences were then aligned by MUSCLE in MEGA X and exported to IQ-Tree. The tree was generated using substitution model VI + I + G4 (Invar + Gamma with 4 categories) as a model of rate heterogeneity and Ultrafast Bootstrap with 1000 replicates. (d) *OsLsi3* gene showing TMDs (prepared in PROTER V 1.1 with default settings) [47]. The orange circle highlights the highly conserved region. The conserved region is represented by a sequence logo (prepared in WebLogo [48]).

3.1.3. Siliplant1 Homologs in Viridiplantae

A search employing *SbSlp1* as a query resulted in 342 protein sequences. We calculated the pI (7.03–10.34) and Mw (7135–74,345.9) along with the prediction of cleavage site and subcellular localization. Since positively charged amino acids are involved in biological silicification, we discarded all the proteins with a predicted pI value less than seven [21]. We also rejected the proteins that lacked signal peptide, as a silicifying protein must be secreted outside the silica cell membrane for biosilicification to take place in the paramural space. We also removed those peptide sequences for which we did not find the cleavage site. This resulted in the selection of 225 protein sequences. These sequences belonged to different gene families i.e., pistil-specific extension-like proteins, pollen Ole e 1 allergen and extensin family proteins, proline rich proteins (PRP), and some uncharacterized proteins. The ML tree was divided into three clads i.e., 1, 2, and 3. The first two clads had genes belonging to pollen Ole e 1 gene family while the third clad had mostly genes belonging to PRP family and some pollen ole e 1 members. Interestingly, all the genes which formed a clad with the *SbSlp1*, were pollen Ole e 1s. Monocot and dicot genes grouped in separate clads i.e., 1 and 2, respectively (Figure 3).

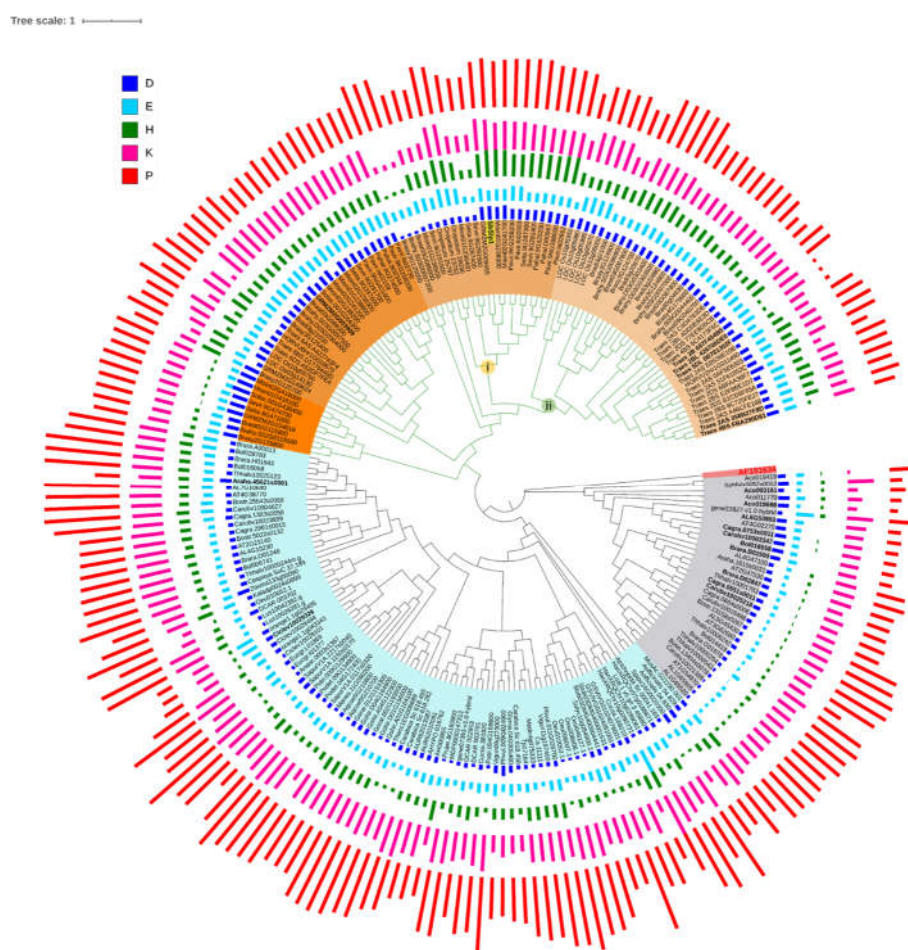


Figure 3. Maximum-likelihood tree of *SbSlp1* homologs in viridiplantae. *Cylinthotheca fusiformis* silaffin precursor protein (*sil1p*, AF191634) gene was used as an outgroup. The sequences were aligned by MUSCLE in MEGA X and exported to IQ-Tree. The tree was generated using substitution model VT + F + G4 as a model of rate heterogeneity and Ultrafast Bootstrap with 1000 replicates. The bold IDs show those genes which have no repeat sequences. The bars on the tree nodes represent the frequencies of the amino acids. Orange = Clad 1; subclad (i) = monocot subclad containing H, D-rich, P, K, E-rich, and P, T, Y-rich domains, subclad (ii) = *SbSlp1* homologs in known Si accumulators, light blue = Clad 2, and grey = Clad 3.

We then looked for the presence of repeats and found 17 genes without any repeats (bold in the tree) (Supplementary Table S3). The reported peptide sequence for in vitro silica precipitation is KEKPVKPPKKHPPP [21], therefore we search for the presence of this signature sequence in the identified homologs. A total of five genes in *SbSlp1* subclad i.e., *Brasy3G125000*, *Brahy.S03G0097200*, *Brast03G088000*, *Misin01G238500*, and *SeVir.9G291600* did not contain the H or D-rich regions or the consensus sequences KKPXPXKPKXPXKPKXPX. The remaining genes in this subclad (i) had the H, D-rich, P, K, E-rich, and P, T, Y-rich domains, which were repeated more than once (Supplementary Table S3; Supplementary Figure S3).

Interestingly, we did not find the rice and other known high Si accumulators such as wheat that grouped closely with *SbSlp1*. Consequently, we checked wheat and rice homologs for the presence of consensus domains/sequences. Particularly, the rice *SbSlp1* homologs formed a separate subclad (ii) in Clad 1. The homologs of *B. staci*, *B. sylvaticum*, *B. hybridum*, were grouped together, while the wheat homologs were grouped together along with the one *H. vulgare* gene. Therefore, we aligned the genes in subclads (i and ii) to see the conserved/consensus sequences (Supplementary Figure S3a). The genes from clad-1 (ii) showed the presence of the P, K, E-rich region and H, D-rich domains. A total of five members of clad 1 (ii) belonging to *B. hybridum* and *B. staci*, and *B. distachyon* had three repeats of these regions. We noticed that the repeated sequences were followed by a G rich motif consisting of two or three Gs. One common feature of all the homologs containing these motifs was that before the consensus region, an H, D-rich domain was present and immediately after the consensus region a P, T, Y-rich domain was present similar to the work on *SbSlp1* [21]. In clad 1 (i), this arrangement was repeated almost five times in the homologs with some modifications (Supplementary Figure S3b). While for clad 1 (ii) members, the arrangement of the above-mentioned domains was repeated only once with the exception of the five members of the clad 1(ii) mentioned above.

Now that we knew the conserved features of *SbSlp1* protein, we wanted to know if this feature was also present in the moderate-accumulators and low(non)-accumulators. For this, we selected known Si accumulators (rice, wheat, and sorghum), Si-intermediate types (soybean, cucumber) and low Si-accumulator (tomato) homologs. These homologs belonged to PRP proteins and did not contain the three domains that were found in Si-accumulators. Since, pollen Ole 1 allergens are present in other plants too (i.e., low-accumulators and moderately accumulators) [49], therefore, we searched for Arabidopsis pollen Ole 1s against *SbSlp1* but did not find the conserved domains. However, in a separate study against soybean genome (data not shown), an extension-like protein repeat containing soybean gene *Glyma.09G092700* was found. A search for repeats showed the presence of a domain (KPPIYKPPVYTPPVYKPPVEKPP) that was repeated eight times. By using *Glyma.09G092700* as a query gene to search against the known low (non) Si accumulators and moderate accumulators, the tomato genome did not result in any homologs even though its genome has PRP genes [50]. We used *TPRP1-F1* (a tomato PRP gene) to search for any homologs in the tomato genome. The results showed the members of seed storage/bifunctional inhibitor/lipid transfer proteins and hydrophobic seed proteins. However, none of the genes showed the domains of interest. As far as soybean and Arabidopsis PRPs are concerned, all the genes had repeats. However, two genes *Glyma.09G092700* and *At4g38770* had P, K, E-rich repeats (Supplementary Figure S3c). These observations suggest that moderate Si accumulators have the P, K, E-rich domain but they lack the other two domains present in Si-accumulators. Based on these observations, it could be proposed that at least these three sequences are required for biosilicification in Si-accumulators.

3.2. Silica Precipitation Potential of Repeats

We studied the molecular interactions and performed in-silico alanine scanning on the repeats present in *SbSlp1*, *Glyma.09G092700*, *Cucsa.381820*, and *At4g38770*. The computation tool Ligplot implemented in MOE allowed us to understand the silica polymerization potential of the selected AA repeats. When K is replaced with A, the energy barrier to $\text{Si}(\text{OH})_3\text{O}^-$ tetramer formation increases (Figure 4a). With this increase in the energy barrier, the predicted rate of silica precipitation decreases (Figure 4b). The prediction suggested that the *SbSlp1* AA repeat, when mutated K to A, still was able to precipitate silica. This prediction is in accordance with the observation by Kumar et al. [19], that

the same peptide was able to precipitate silica in a gel-like material (Figure 4c). Similarly, the changes in *Cucsa.381820* repeat showed decrease in the predicted rate of polymerization. *Glyma.09G092700* and *At4g38770* repeats showed limited efficiency to form $\text{Si}(\text{OH})_3\text{O}^-$ tetramer (Figure 4d,e).

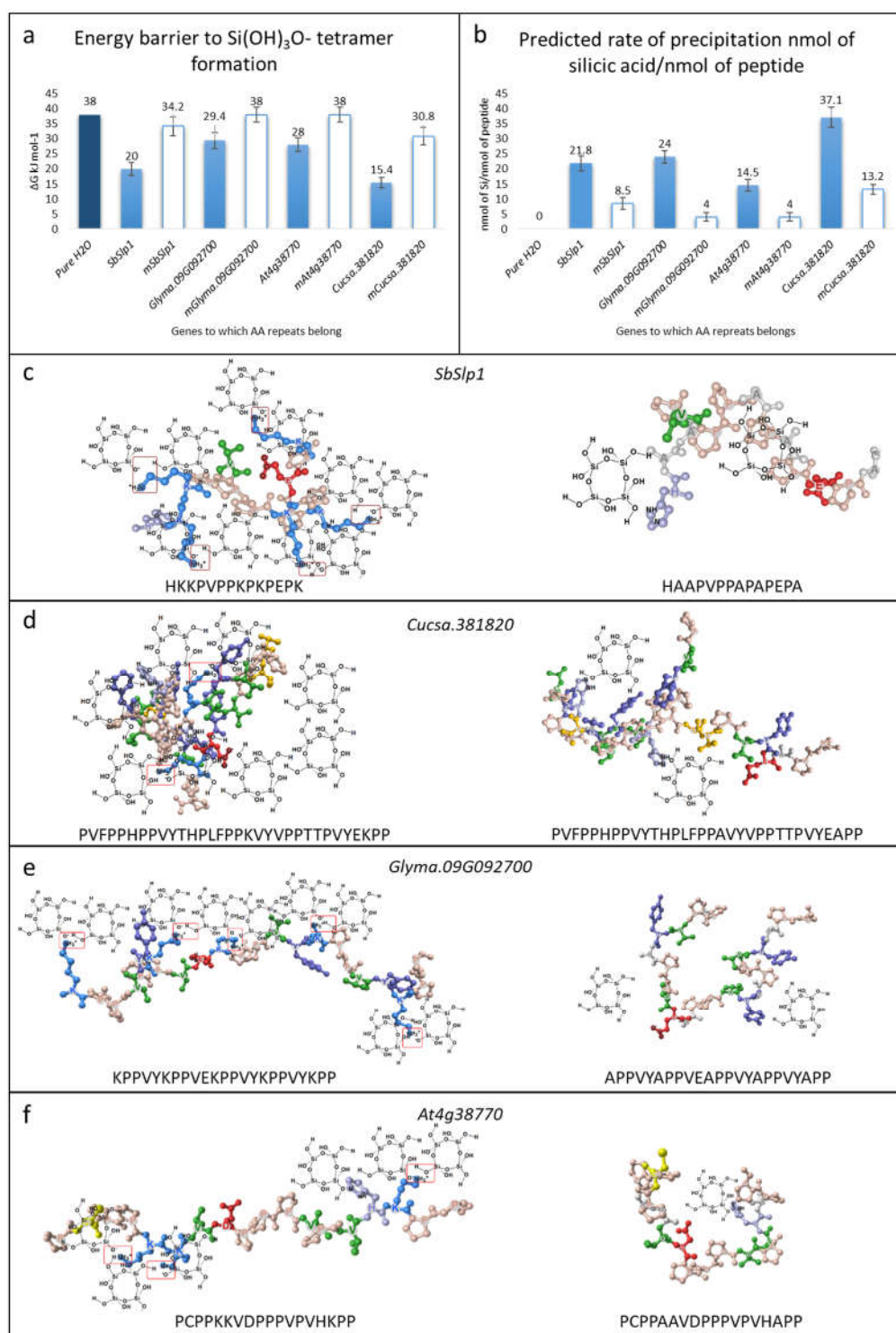


Figure 4. Silica precipitation potential of AA repeats in *SbSlp1*, *Cucsa.381820*, *Glyma.09G092700*, and *At4g38770*. (a) Energy barrier for $\text{Si}(\text{OH})_3\text{O}^-$ tetramer formation (ΔG kJ mol⁻¹), (b) predicted rate of precipitation nmol of Si/nmol of peptide, and the interaction of (c) *SbSlp1*, (d) *Cucsa.381820*, *Glyma.09G092700*, and *At4g38770* peptides with silicic acid to form $\text{Si}(\text{OH})_3\text{O}^-$ tetramer. The first panel of the figures c–f shows normal sequences and the second panels show the peptides where K was replaced with A. The letter “m” before the genes names in panels a,b represents the modified peptides.

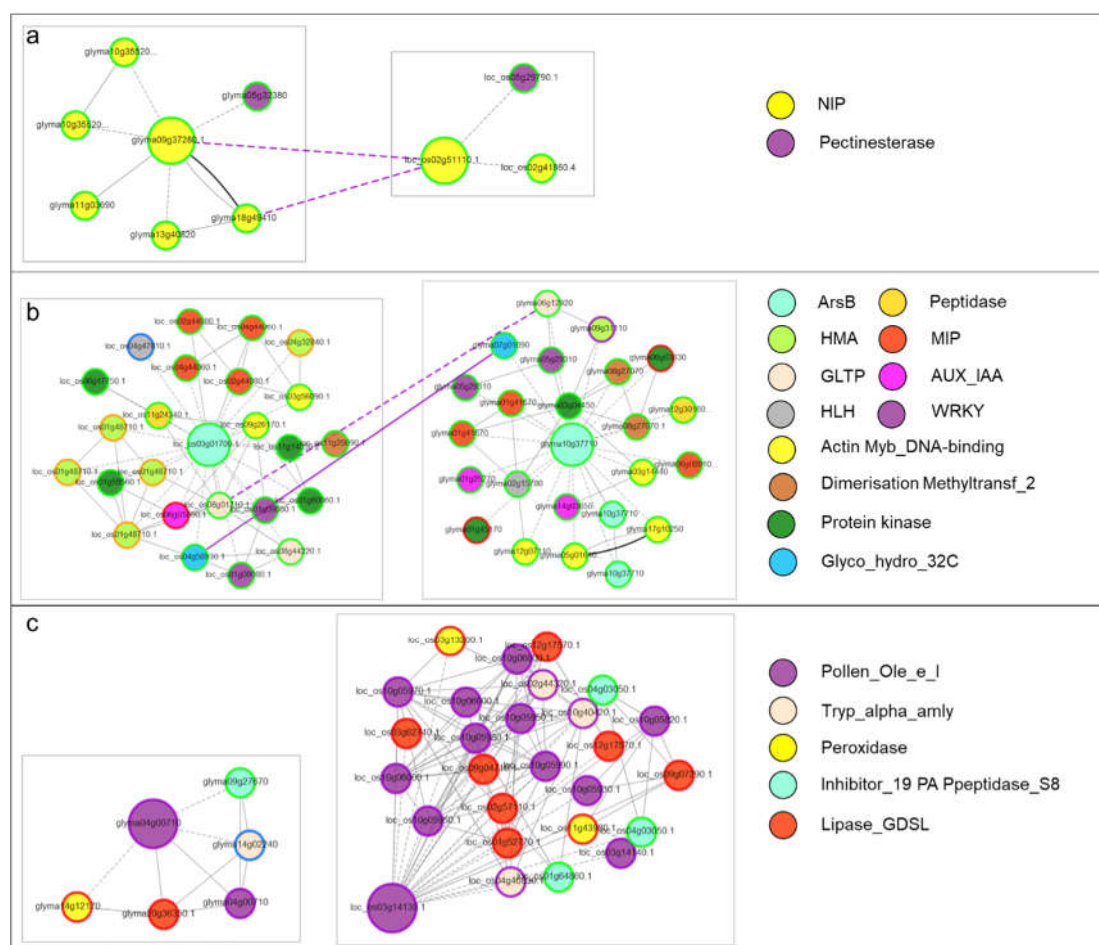


Figure 6. Monocot (rice) and dicot (soybean) gene coexpression networks of (a) channel type SITs, (b) Lsi2s, and (c) putative *Slp1s*. The solid edges show duplication while the dotted edges show speciation. The genes with the same shape and color belong to the same gene family and/or have pfam domains in common. Node borders indicate the phylostratum of the gene i.e., green (green plants), red (land plants), light blue (vascular plants), orange (monocot/dicots), brown (rosids/brassicales/malvids), and black (genus-specific phylostratum). The key on the right shows the gene family. The detail on the genes in the network is given in Supplementary Table S4.

4. Discussion

Channel type SITs that belong to NIP-III subfamily of aquaporins are thought to be evolutionarily restricted to monocots and sedges (known to be high Si accumulators), and early-diverging lineages (e.g., Sellaginellaceae and Equisetaceae) [18]. Furthermore, it is known that only a limited number of species from angiosperms have NIP-IIIs, and those which do have, show variation in the spacing between NPA motifs [1]. No detection of GSGR/STARG channel type SITs in Chlorophyta in our results suggests that the species explored in this research are either low Si-accumulators or their genomes contain other types of Si transporters. Previously, it is known that some Chlorophyte taxa i.e., *Tetraselmis pediastrum*, *Hydrodictyon* and some golden-algae accumulate Si in their cell walls [52–54]. Additionally, a recent study reported the presence of NIP family members in green algae however, the reported species did not contain STAR/GSGR SF, instead the authors found FAAR SF containing NIPs. Furthermore, the FAAR SF containing Chlorophyte NIP6:1 (*Klebsormidium nitens*) was unable to transport Germanium [17]. Therefore, the absence of channel type SITs from Chlorophyta may suggest that the NIPs evolved the Si transport-related features after the split of viridiplantae into Chlorophyta and Streptophyta or latter during the evolution of seed plants [55]. However, the observation that TIPs, PIPs, and MIPs were present in these species suggests this gene superfamily is omnipresent in these viridiplantae members. These observations are in agreement with the recent work by Pommering et al. [17] that the ability to transport Si could have

arisen (due to subfunctionalization and neofunctionalization of arsenic efflux transporters into essential and beneficial plant nutrient transporters) later during the evolution of seed plants. Considering the focus of our study i.e., Lsi1 and Lsi2 homologs, the possibility of the presence of other types of transporters is not discussed here. Additionally, more specific studies on the Si concentrations could give clues about the classification of Chlorophyte taxa or species as Si accumulators, moderate Si-accumulators, and low Si-accumulators. All of channel type SITs in our study were found in Mangoliopsida (Supplementary Table S1). Current results are clearly in agreement with earlier reports that channel type SITs are widespread in monocots, since we found a higher number of channel type SITs in all studied members of Poaceae (Table 2). However, not all monocots included in our study have two NPA motifs with 108/109 AAs between the two motifs, for example, *Z. marina*, *A. comosus*, and *A. officinalis* channel type SITs homologs did not show these characteristic features known for the ability to transport Si (Supplementary Table S1). In dicots species belonging to *Brassicaceae* and *Solanaceae* families, we did not find channel type SITs having GSGR/STAR SF and defined NPA to NPA spacing. These observations are relevant to the fact that these species show very limited Si accumulation in planta [56]. However, a recent report in tomato suggested that a spacing of 109 AAs can still be functional when expressed in rice plant and *Xenopus* oocytes [9]. Other studies have also reported the absence of NIP-III members in this plant family [1,17]. Nonetheless, we did find NIP-III members (e.g., in tomato and papaya) but in both cases, the NPA to NPA spacing was 109 AAs (Figure 1). Another member of *Solanaceae* i.e., *Nicotiana tabaccum* has also been reported to have a functional Lsi1 homolog with two NPA motifs and a precise spacing of 108 AAs between the two NPA motifs [57]. Other dicot families i.e., *Myrtaceae*, *Rutaceae*, *Linaceae*, and *Oleaceae* also showed the absence of characteristic channel type SITs (Table 2), suggesting that the studied species are probably low Si-accumulators. Nonetheless, Si-transport inability is not universal in dicots and some species might have evolved reduced Si transport and absorption ability due to neofunctionalization as suggested in a recent study [17,58]. The interspecific variation of presence or absence and a variable number of channel type SITs and SF is interesting. Particularly, *H. annuus* which accumulates >1% (252–10,909 mg/Kg of dry weight; [59]) is possibly due to the presence of a higher number of channel type SITs (Table 2). The presence of two different types of SFs in two different families of Caryophyllales i.e., *C. quinoa* (*Chenopodiaceae*) and *A. hypochondriacus* (*Amaranthaceae*) suggest that modern plants (with an ability to accumulate Si) have evolved two different types of channel type SIT i.e., with STAR and GSGR SFs (Figure 1a). We state this because of the observation that STAR SF is present in species not restricted to few genera. For example, the earlier reports presented that only Equisetales (e.g., *E. arvense*) have STAR SF [46]. Our statement is consistent with a recent study by Deshmukh et al. [1], where authors also reported the presence of STAR SF in fern (*Mapania palustris*) and monocots (*Dipteris conjugata* and *Lepidosperma gibsonii* from Cyperaceae). Nevertheless, the presence of STAR SF in fern, monocot, and dicot clearly explain the presence and evolution of two SF in Lsi1s. This would be an interesting question for future studies targeting the evolution of aquaporin SFs in plants.

Table 2. Summary of Lsi1, Lsi2, and Slp1 homologs found in 80 studied Viridiplantae species.

Class	Order	Family	Species	Lsi1	Lsi2	Slp1
Trebouxiophyceae	Trebouxiophyceae	Coccomyaceae	<i>Coccomyxa subellipsoidea</i>			
	Trebouxiales	Botryococcaceae	<i>Botryococcus braunii</i>			
Mamiellophyceae	Mamiellales	Mamiellaceae	<i>Micromonas PUSILLA ccmp1545</i>		1	
		Bathycoccaceae	<i>Osteococcus lucimarinus</i>		1	
Chlorophyceae	Sphaeropleales	Chromochloridaceae	<i>Chromochloris zofingiensis</i>		2	
		Volvocaceae	<i>Volvox carteri</i>		2	
	Chlamydomonadales	Chlamydomonadaceae	<i>Chlamydomonas reinhardtii</i>		1	
		Dunaliellaceae	<i>Dunaliella salina</i>		1	
Lycopodiopsida	Selaginellales	Selaginellaceae	<i>Selaginella moellendorffii</i>		3	
Sphagnopsida	Sphagnales	Sphagnaceae	<i>Sphagnum fallax</i>		1	1

Bryopsida	Funariales	Funariaceae	<i>Physcomitrella patens</i>	2	
Marchantiopsida	Marchantiales	Marchantiaceae	<i>Marchantia polymorpha</i>	1	
Liliopsida	Alismatales	Araceae	<i>Spirodela polyrhiza</i>	1	1
		Zosteraceae	<i>Zostera marina</i>		1
		Asparagales	Asparagaceae	1	2
		Zingiberales	Musaceae	4	4
	Poales	Poaceae	<i>Musa acuminata</i>	4	1
			<i>Ananas comosus</i>	4	4
			<i>Oropetium thomaeum</i>	1	4
			<i>Panicum halli</i>	1	3
			<i>Panicum virgatum</i>	3	8
			<i>Brachypodium stacei</i>	2	5
			<i>Brachypodium hybridum</i>	4	10
			<i>Brachypodium sylvaticum</i>	2	4
			<i>Brachypodium distachyon</i>	2	5
			<i>Zea mays</i>	4	5
			<i>Triticum aestivum</i>	3	6
			<i>Sorghum bicolor</i>	2	5
			<i>Setaria viridis</i>	2	5
			<i>Setaria italica</i>	2	5
			<i>Oryza sativa</i>	2	4
			<i>Hordeum vulgare</i>	3	5
			<i>Miscanthus sinensis</i>	4	10
Magnoliopsida	Ranunculales	Ranunculaceae	<i>Aquilegia coerulea</i>	1	1
	Myrtales	Myrtaceae	<i>Eucalyptus grandis</i>	2	2
	Saxifragales	Crassulaceae	<i>Kalanchoe laxiflora</i>	1	2
			<i>Kalanchoe fedtschenkoi</i>	1	1
	Vitales	Vitaceae	<i>Vitis vinifera</i>	1	1
	Caryophyllales	Chenopodiaceae	<i>Chenopodium quinoa</i>	2	2
		Amaranthaceae	<i>Amaranthus hypochondriacus</i>	1	2
	Gentianales	Rubiaceae	<i>Coffea arabica</i>	1	3
	Amborellales	Amborellaceae	<i>Amborella trichopoda</i>	1	4
	Asterales	Asteraceae	<i>Lactuca sativa</i>	2	1
			<i>Helianthus annuus</i>	7	6
		Lamiaceae	<i>Capsella grandiflora</i>	1	5
		Phrymaceae	<i>Mimulus guttatus</i>	1	1
	Solanales	Solanaceae	<i>Olea europaea</i>	1	5
			<i>Solanum tuberosum</i>	1	2
			<i>Solanum lycopersicum</i>	1	1
	Apiales	Apiaceae	<i>Daucus carota</i>	1	1
			<i>Trifolium pratense</i>	1	3
			<i>Vigna unguiculata</i>	3	2
			<i>Phaseolus vulgaris</i>	2	2
	Fabales	Fabaceae	<i>Medicago truncatula</i>	2	1
			<i>Glycine max</i>	2	1
			<i>Cicer arietinum</i>	1	1
	Rosales	Rosaceae	<i>Fragaria vesca</i>	3	2
			<i>Prunus persica</i>	1	1

		<i>Malus domestica</i>	1	2	1
	Linaceae	<i>Linum usitatissimum</i>		2	2
	Euphorbiaceae	<i>Ricinus communis</i>	1	1	
		<i>Mahoea esculenta</i>	1	1	1
Malpighiales		<i>Salix purpurea</i>	1	2	3
	Salicaceae	<i>Populus deltoides</i>	1	1	2
		<i>Populus trichocarpa</i>	1	2	2
Cucurbitales	Cucurbitaceae	<i>Cucumis sativus</i>	1	1	2
		<i>Capsella rubella</i>		1	6
		<i>Eutrema salugineum</i>		1	6
		<i>Boechera stricta</i>		1	4
Brassicales	Brassicaceae	<i>Brassica oleracea</i>		1	5
		<i>Brassica rapa</i>		1	6
		<i>Arabidopsis halleri</i>		1	2
		<i>Arabidopsis lyrata</i>		1	7
		<i>Arabidopsis thaliana</i>		1	6
	Caricaceae	<i>Carica papaya</i>	1	1	1
		<i>Theobroma cacao</i>	1	4	1
Malvales	Malvaceae	<i>Gossypium raimondii</i>		3	2
		<i>Gossypium hirsutum</i>	2	3	4
	Anacardiaceae	<i>Anacardium occidentale</i>	1	1	
Sapindales		<i>Citrus clementia</i>		1	3
	Rutaceae	<i>Citrus sinensis</i>		1	2

A very limited knowledge is available on the evolution of Lsi2s mainly because they do not show any sequence homology with the channel type SITs [4]. However, they are similar to arsenite Lsi2s that are present in bacteria and Archea [6]. Considering this i.e., a generalized metalloid transport capacity because of molecular mimicry from the evolutionary perspective, it is very much clear that these genes are widespread in Eukaryotes [18,60]. Our results showed that *OsLsi2* homologs were found in 71 viridiplantae species both from Streptophyta and Chlorophyta which confirms that Si transport is widespread in Viridiplantae. Moreover, together with previous reports (Table 2) [4,16], it could be stated that Si transport is an ancient feature. This is consistent with the findings of Marron et al. [16], where they reported that plant Lsi2s formed a separate clad with other Eukaryotic Lsi2s or Lsi2-like genes. As we said earlier, no sequence homology has been reported with channel type SITs, therefore, we did not find any conserved features of these genes to principally differentiate if a candidate gene could be dedicated to Lsi2s only or might have both Si and arsenite efflux capability like reported in rice. However, we could at least find a conserved motif in reported monocot Lsi2 homologs and the ones we report (Figure 2D) but its functional importance for monocots needs further investigation. Based on our findings, we could state that Lsi2s evolved distinctly i.e., two types of Lsi2s after the emergence of Polypodiopsida [16]. This assumption is based on the fact that all the Lsi2 homologs from Polypodiopsida, Lycopodiopsida, Sphanopsida, Bryopsida, and Marchantiopsida are grouped together (Supplementary Figure S2) [8].

The presence of Slps across Eukaryotic lineages is known mainly from the origin of the Si requirement in diatoms [20]. However, scarcity of the known Slps in plants has not enabled us to fully understand this process and the putative proteins in plants that take part in silica deposition [2]. Our knowledge of the plant lineages having the potential to accumulate Si has increased to a greater extent, thanks to the discovery of SITs by JF Ma and others [4–6,10,56]. However, we still lack the basic understanding of the gene/proteins involved in the biosilicification process in planta. In this work, we found the homologs of *SbSlp1* in almost all Streptophyta species (Table 2). Despite the inability of many plant species to accumulate large concentrations of silica in planta, where low Si-

accumulator plants have shown extremely low biosilicification (and accumulation of phytoliths) in their cells, the homologs of *SbSlp1* are present in all Streptophyta species [2,4]. Nevertheless, the detailed investigations on the types of proteins for biosilicification in plants are scarce, and a complete picture of the evolution of these putative proteins will remain an open subject. However, considering the fact that *SbSlp1* is a member of pollen allergen Ole e 1 gene family, studies have confirmed that pollen allergens are present in green algae to angiosperms, where they expanded in angiosperms through multiple rounds of duplication and changes in polyploidy levels [49]. This is also true regarding gene structure evolution since, we found that in so-called Si accumulators, the H, D-rich, P, K, E-rich, and P, T, Y-rich domains are present in multiple repeats (Supplementary Figure S3). Meanwhile, so far, two studies have demonstrated the ability of a cationic PRP protein from cucumber and a Slp1 from sorghum to deposit silica in the cell wall and paramural space, respectively [19,25]. Therefore, based on the observations on putative Slps in our study and structures reported earlier, it cannot be necessarily said that the process of biosilicification needs the presence of H, D-rich domain and/or P, T, Y-rich domain. Regardless of the fact that biosilicification has been reported in soybean (i.e., the presence of phytoliths) [61], we did not find H, D-rich domain and P, T, Y-rich domain in the soybean *SbSlp1* homologs found in our BlastP search (Supplementary Table S3; Supplementary Figure S3c). This further strengthens the less important role of both domains in biosilicification. The lack of P, K, E-rich domain from non-accumulators, and its presence in moderate-accumulators and Si-accumulators suggests the essential role of this domain in biosilicification process. However, based on the study by Kauss et al. (2003), due to the presence of a high content of positively charged AAs in three peptides tested by their team, it can be concluded that positioning of K along with R as clusters or partially adjacent positioning did not clearly disturb the ability of peptide to precipitate silica. Our results showing the changes in the energy barrier and the predicted rate of precipitation of silicic acid/nmol of peptide also suggest that replacing K with A increases the energy barrier and reduced the rate of precipitation (Figure 4). The authors concluded that silica deposition is mainly dependent on a sufficiently high density of positively charged amino acids than on a peptide's primary sequence.

The Si deposition benefits plants by increasing plants' ability to withstand biotic and abiotic stresses; hence considered as biostimulant in agriculture [62]. However, to benefit from Si, plants must uptake transport, and accumulate/deposit it in different plant tissues at physiological conditions with temperatures ranging from 0 to 40 °C, neutral pH, and ambient pressure [14]. This marvelous process is completed by the above discussed SITs i.e., channel type SITs and Lsi2s and many uncharacterized Slps as well as many unknown players in this process [21]. To complete the process of biosilicification, plants may mobilize Si from roots through the expression of channel type SITs and Lsi2s in roots and stems [4]. The expression of these SITs varies in different species [2]. Our digital expression results in rice, soybean, and tomato greatly complement that channel type SITs are expressed in roots (Figure 5a). However, the homologs of Lsi1 in rice, and maize i.e., Lsi6s have been reported to be expressed in stem as well [63]. The Lsi2s are mainly expressed in rice in roots, where these genes make a cooperative network to move Si within roots [6]. Interestingly, we also found that the putative Lsi2s in three species differing in Si accumulation capacity express in roots (Figure 5b). The understanding of the mechanism of Si transport has been now well developed, however, emerging data of the interactome helping researchers understand how Si interacts with other genes and proteins to help plants against biotic and abiotic factors. The coexpression of channel-type SITs with pectinesterases is interesting because the level and pattern of pectin esterification play a role in constitutive resistance to many fungal and bacterial pathogens influencing the susceptibility of plant cell wall to microbial pectin-degrading enzymes [64,65]. Apart from this, it was recently suggested that a *B. distachyon* mutant exhibits a range of alterations in the composition of non-cellulosic polysaccharides i.e., monosaccharides associated with pectins (rhamnose, galacturonic acid, and galactose) [66]. This study and our in-silico coexpression results suggest that transport of Si via Lsi1, when disrupted, could affect the cell wall composition. There is an increasing body of literature on the role of cell wall modification related genes' involvement in multiple abiotic and biotic stress related processes [67–69]. Furthermore, the coexpression of Lsi2s with known transcription factors,

i.e., WRKYs, is quite understandable since WRKYs are known for their ability to help plants against abiotic stresses [70]. However, so far, the combined role of Si and WRKYs has not been explored in plants and would be an interesting topic for future research. In addition to WRKYs, the coexpression of peptidase, protein kinases and AUX-IAA is possibly due to the fact that silicic acid is involved in signal transduction pathways in plant defenses against microbes [71–73]. Previously, Vatansever et al. [11] also reported such coexpression between SITCs and resistance signaling related genes. Finally, the expression of putative rice and tomato Slp1 in seed and inflorescence, and fruit suggests that these proteins have additional roles (Figure 5c) i.e., important physiological roles in pollen, especially the pollination process [49,74]. This is consistent with the findings of Kumar et al. [19] i.e., the expression of *SbSlp1* was detected in immature leaves and inflorescence. Additionally, the PRP in cucumber was expressed in cell walls [25], and it is known that biosilicification in plants is associated with cell wall polymers [75], particularly in plants having resistance to diseases [69]. Our results also confirmed the coexpression of arabinogalactan proteins (Figure 6c). Hence, in view of these reports, Si seems to increase plants' ability to withstand against different stresses by coexpressing with structural integrity and signaling related genes. Together, these results enhance our understanding about the mechanism of biosilicification in plants and lead us towards many interesting questions to be answered.

5. Conclusions

Together, our results demonstrate that early Viridiplantae lineages, i.e., green algae, lack characteristic channel type SITs and likely have other mechanisms facilitating Si influx. Furthermore, these genes evolved two different types of SFs i.e., STAR and GSGR. Additionally, these genes are not strictly restricted in monocot species and can be found in a wide range of dicot species. Two types of Lsi2s evolved after the emergence of Polypodiopsida. However, considering the presence of Lsi2 outside Streptophyta i.e., Chlorophyta, it could be stated that Lsi2s are ancient. Similar to channel type SITs, the Slp1 homologs were not found in Chlorophyta. High number of repeats of P, K, E-rich domain in monocots (or Si accumulators) in addition to two other conserved domains could be the key feature behind the presence of relatively larger Si quantities in these plants. The presence of P, K, E-rich domain within Si-accumulators and moderate-accumulators suggests its functional importance for the biosilicification process.

Supplementary Materials: The following are available online at www.mdpi.com/2223-7747/9/11/1612/s1, Table S1. List of homologs of rice channel type SITs in the studied plant genomes. Table S2. List of homologs of rice *Lsi2* gene in the studied plant species. Table S3. Siliplant1 homologs in the studied plant species. Table S4. Details on the genes in the coexpression networks of Lsi2s and Siliplant1s. Figure S1. (a) Sequence alignment of known efflux Si transporters from rice, maize, barley, cucumber, pumpkin, and an Arabidopsis gene. (b) The highly conserved motif in known Si efflux transporters. Figure S2. Maximum-likelihood tree of Si efflux transporters. ML tree of *OsLsi2* homologs in studied viridiplantae genomes (plus two Equisetum Lsi2s). *KU821730* (*Spongosphaera streptacantha* SIT-L gene) was used as an outgroup. The sequences were then aligned by MUSCLE in MEGA X and exported to IQ-Tree. The tree was generated using substitution model VI+I+G4 (Invar+Gamma with 4 categories) as a model of rate heterogeneity and Ultrafast Bootstrap with 1000 replicates. Figure S3. (a) Alignment of Slp1 homologs (sub-clad 1Cii) showing the conserved regions. (b) Alignment of *SbSlp1* homologs that clustered together (sub-clad 1Ci). (c) Repeats found in soybean and Arabidopsis PRP genes. Red bars = H, D-rich domain, Orange bars = P, K, E-rich domain, Green bars = P, T, Y-rich domain, and Pink = G-rich region. The sequences above the alignment in (a) represent the additional sequences present within genes 8–13 in alignment. The colors of the borders match with those inserted in the alignment. The position of the inserted boxes represents the position of the respective sequences.

Author Contributions: Conceptualizing, M.A.N. and K.S.G.; methodology, M.A.N., A.M.Z., and F.A.; software, X.L., T.-F.C., and F.A.; validation, M.A.N., G.C., J.H., R.M.A., and S.S.; formal analysis, M.A.N., R.M.A., S.S.; investigation, M.A.N., S.S., K.S.G., J.H., and G.C.; resources, M.A.N., A.M.Z. and S.S.; data curation, X.L., F.S.B., J.H., T.-F.C.; writing—original draft preparation, M.A.N. and F.A.; writing—review and editing, all authors; supervision, K.S.G. and G.C.; project administration, M.A.N.; funding acquisition, S.S., T.-F.C., and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: S.S. was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1B03035600), Republic of Korea. X.L. and T.F.C. are supported by the Innovation and Technology Commission, Hong Kong Special Administrative Region Government to the State Key Laboratory of Agrobiotechnology (CUHK), CUHK Direct Grant 4053430. Any opinions, findings, conclusions, or recommendations expressed in this publication do not reflect the views of the Government of the Hong Kong Special Administrative Region or the Innovation and Technology Commission.

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

References

1. Deshmukh, R.; Sonah, H.; Belanger, R. New evidence defining the evolutionary path of aquaporins regulating silicon uptake in land plants. *J. Exp. Bot.* **2020**, *eraa342*, doi:10.1093/jxb/eraa342.
2. Nawaz, M.A.; Zakharenko, A.M.; Zemchenko, I.V.; Haider, M.S.; Ali, M.A.; Imtiaz, M.; Chung, G.; Tsatsakis, A.; Sun, S.; Golokhvast, K.S. Phytolith Formation in Plants: From Soil to Cell. *Plants* **2019**, *8*, 249.
3. Haynes, R.J. A contemporary overview of silicon availability in agricultural soils. *J. Plant Nutr. Soil Sci.* **2014**, *177*, 831–844, doi:10.1002/jpln.201400202.
4. Ma, J.F.; Yamaji, N. A cooperative system of silicon transport in plants. *Trends Plant Sci.* **2015**, *20*, 435–442.
5. Ma, J.F.; Tamai, K.; Yamaji, N.; Mitani, N.; Konishi, S.; Katsuhara, M.; Ishiguro, M.; Murata, Y.; Yano, M. A silicon transporter in rice. *Nature* **2006**, *440*, 688.
6. Ma, J.F.; Yamaji, N.; Mitani, N.; Tamai, K.; Konishi, S.; Fujiwara, T.; Katsuhara, M.; Yano, M. An efflux transporter of silicon in rice. *Nature* **2007**, *448*, 209–212, doi:10.1038/nature05964.
7. Mitani-Ueno, N.; Yamaji, N.; Zhao, F.J.; Ma, J.F. The aromatic/arginine selectivity filter of NIP aquaporins plays a critical role in substrate selectivity for silicon, boron, and arsenic. *J. Exp. Bot.* **2011**, *62*, 4391–4398, doi:10.1093/jxb/err158.
8. Gregoire, C.; Remus-Borel, W.; Vivancos, J.; Labbe, C.; Belzile, F.; Belanger, R.R. Discovery of a multigene family of aquaporin silicon transporters in the primitive plant *Equisetum arvense*. *Plant J.* **2012**, *72*, 320–330, doi:10.1111/j.1365-3113X.2012.05082.x.
9. Sun, H.; Duan, Y.; Mitani-Ueno, N.; Che, J.; Jia, J.; Liu, J.; Guo, J.; Ma, J.F.; Gong, H. Tomato roots have a functional silicon influx transporter but not a functional silicon efflux transporter. *Plant Cell Environ.* **2020**, *43*, 732–744.
10. Deshmukh, R.K.; Vivancos, J.; Ramakrishnan, G.; Guerin, V.; Carpentier, G.; Sonah, H.; Labbe, C.; Isenring, P.; Belzile, F.J.; Belanger, R.R. A precise spacing between the NPA domains of aquaporins is essential for silicon permeability in plants. *Plant J.* **2015**, *83*, 489–500, doi:10.1111/tpj.12904.
11. Vatansever, R.; Ozyigit, I.I.; Filiz, E.; Gozukara, N. Genome-wide exploration of silicon (Si) transporter genes, Lsi1 and Lsi2 in plants; insights into Si-accumulation status/capacity of plants. *BioMetals* **2017**, *30*, 185–200.
12. Kumar, S.; Soukup, M.; Elbaum, R. Silicification in grasses: Variation between different cell types. *Front. Plant Sci.* **2017**, *8*, 438.
13. Piperno, D.R. *Phytoliths: A Comprehensive Guide for Archaeologists and Paleoecologists*; Altamira Press: Lanham, MD, USA, 2006.
14. Exley, C. A possible mechanism of biological silicification in plants. *Front. Plant Sci.* **2015**, *6*, 853.
15. Guerriero, G.; Hausman, J.F.; Legay, S. Silicon and the plant extracellular matrix. *Front. Plant Science* **2016**, *7*, 463.
16. Marron, A.O.; Ratcliffe, S.; Wheeler, G.L.; Goldstein, R.E.; King, N.; Not, F.; de Vargas, C.; Richter, D.J. The evolution of silicon transport in eukaryotes. *Mol. Biol. Evol.* **2016**, *33*, 3226–3248, doi:10.1093/molbev/msw209.
17. Pommerrenig, B.; Diehn, T.A.; Bernhardt, N.; Bienert, M.D.; Mitani-Ueno, N.; Fuge, J.; Bieber, A.; Spitzer, C.; Bräutigam, A.; Ma, J.F. Functional evolution of nodulin 26-like intrinsic proteins: From bacterial arsenic detoxification to plant nutrient transport. *New Phytol.* **2020**, *225*, 1383–1396.
18. Trembath-Reichert, E.; Wilson, J.P.; McGlynn, S.E.; Fischer, W.W. Four hundred million years of silica biomineralization in land plants. *Proc. Natl. Acad. Sci.* **2015**, *112*, 5449–5454.
19. Kumar, S.; Adiram-Filiba, N.; Blum, S.; Sanchez-Lopez, J.A.; Tzfadia, O.; Omid, A.; Volpin, H.; Heifetz, Y.; Goobes, G.; Elbaum, R. Siliplant1 (Slp1) protein precipitates silica in sorghum silica cells. *J. Exp. Bot.* **2019**, *eraa258*, doi:10.1093/jxb/eraa258.

20. Raven, J.; Waite, A. The evolution of silicification in diatoms: Inescapable sinking and sinking as escape? *New Phytol.* **2004**, *162*, 45–61.
21. Otzen, D. The role of proteins in biosilicification. *Scientifica* **2012**, *2012*, doi:10.6064/2012/867562.
22. Pamirsky, I.E.; Golokhvast, K.S. Origin and status of homologous proteins of biomineralization (Biosilicification) in the taxonomy of phylogenetic domains. *BioMed Res. Int.* **2013**, *2013*, doi:10.1155/2013/397278.
23. Kröger, N.; Deutzmann, R.; Sumper, M. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. *Science* **1999**, *286*, 1129–1132.
24. Kotzsch, A.; Pawolski, D.; Milentyev, A.; Shevchenko, A.; Scheffel, A.; Poulsen, N.; Shevchenko, A.; Kröger, N. Biochemical composition and assembly of biosilica-associated insoluble organic matrices from the diatom *Thalassiosira pseudonana*. *J. Biol. Chem.* **2016**, *291*, 4982–4997.
25. Kauss, H.; Seehaus, K.; Franke, R.; Gilbert, S.; Dietrich, R.A.; Kröger, N. Silica deposition by a strongly cationic proline-rich protein from systemically resistant cucumber plants. *Plant J.* **2003**, *33*, 87–95.
26. Kersey, P.J. Plant genome sequences: Past, present, future. *Curr. Opin. Plant Biol.* **2019**, *48*, 1–8.
27. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2011**, *40*, D1178–D1186.
28. Krogh, A.; Larsson, B.; Von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580.
29. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press Inc: Totowa, New Jersey, United States of America, 2005; pp. 571–607.
30. Emanuelsson, O.; Brunak, S.; Von Heijne, G.; Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2007**, *2*, 953.
31. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.; Potter, S.C.; Finn, R.D. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641.
32. Bailey, T.L.; Williams, N.; Misleh, C.; Li, W.W. MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **2006**, *34*, W369–W373.
33. Armenteros, J.J.A.; Tsirigos, K.D.; Sønderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnol.* **2019**, *37*, 420–423.
34. 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.
35. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2014**, *32*, 268–274.
36. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587.
37. Hoang, D.T.; Chernomor, O.; Von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **2017**, *35*, 518–522.
38. Letunic, I.; Bork, P. Interactive tree of life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **2006**, *23*, 127–128.
39. Zuckerkandl, E.; Pauling, L. Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins*; Bryson, V., Vogel, H., Eds.; Academic Press: New York, NY, USA, 1965.
40. Leaver-Fay, A.; Tyka, M.; Lewis, S.M.; Lange, O.F.; Thompson, J.; Jacak, R.; Kaufman, K.W.; Renfrew, P.D.; Smith, C.A.; Sheffler, W. ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **2011**, *487*, 545–574.
41. Wallace, A.C.; Laskowski, R.A.; Thornton, J.M. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* **1995**, *8*, 127–134.
42. Waese, J.; Provart, N.J. The bio-analytic resource: Data visualization and analytic tools for multiple levels of plant biology. *Curr. Plant Biol.* **2016**, *7*, 2–5.
43. Waese, J.; Fan, J.; Pasha, A.; Yu, H.; Fucile, G.; Shi, R.; Cumming, M.; Kelley, L.A.; Sternberg, M.J.; Krishnakumar, V. ePlant: Visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell* **2017**, *29*, 1806–1821.

44. Mutwil, M.; Klie, S.; Tohge, T.; Giorgi, F.M.; Wilkins, O.; Campbell, M.M.; Fernie, A.R.; Usadel, B.; Nikoloski, Z.; Persson, S. PlaNet: Combined sequence and expression comparisons across plant networks derived from seven species. *Plant Cell* **2011**, *23*, 895–910.
45. Mitani-Ueno, N.; Yamaji, N.; Ma, J.F. Silicon efflux transporters isolated from two pumpkin cultivars contrasting in Si uptake. *Plant Signal. Behav.* **2011**, *6*, 991–994.
46. Vivancos, J.; Deshmukh, R.; Gregoire, C.; Remus-Borel, W.; Belzile, F.; Belanger, R.R. Identification and characterization of silicon efflux transporters in horsetail (*Equisetum arvense*). *J. Plant Physiol.* **2016**, *200*, 82–89, doi:10.1016/j.jplph.2016.06.011.
47. Omasits, U.; Ahrens, C.H.; Müller, S.; Wollscheid, B. Protter: Interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* **2014**, *30*, 884–886.
48. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190.
49. Chen, M.; Xu, J.; Devis, D.; Shi, J.; Ren, K.; Searle, I.; Zhang, D. Origin and functional prediction of pollen allergens in plants. *Plant Physiol.* **2016**, *172*, 341–357.
50. Santino, C.G.; Stanford, G.L.; Conner, T.W. Developmental and transgenic analysis of two tomato fruit enhanced genes. *Plant Mol. Biol.* **1997**, *33*, 405–416.
51. Hodson, M.J.; White, P.J.; Mead, A.; Broadley, M.R. Phylogenetic variation in the silicon composition of plants. *Ann. Bot.* **2005**, *96*, 1027–1046, doi:10.1093/aob/mci255.
52. Finkel, Z.V. Silicification in the microalgae. In *The Physiology of Microalgae*; Borowitzka, M.A., Beardall, J., Raven, J.A., Eds.; Springer: Cham, Switzerland, 2016; pp. 289–300.
53. Epstein, E. Silicon. *Annu. Rev. Plant Biol.* **1999**, *50*, 641–664.
54. Liang, Y.; Nikolic, M.; Bélanger, R.; Gong, H.; Song, A. (Eds.) *Silicon in Agriculture*; Springer: Amsterdam, The Netherlands, 2015.
55. Becker, B. Snow ball earth and the split of Streptophyta and Chlorophyta. *Trends Plant Sci.* **2013**, *18*, 180–183.
56. Ma, J.F. Silicon transporters in higher plants. In *MIPs and Their Role in the Exchange of Metalloids*; Jahn T.P., Bienert G.P., Eds.; Springer: New York, NY, USA, 2010; pp. 99–109.
57. Zellner, W.; Lutz, L.; Khandekar, S.; Leisner, S. Identification of NtNIP2; 1: An Lsi1 silicon transporter in *Nicotiana tabacum*. *J. Plant Nutr.* **2019**, *42*, 1028–1035.
58. Deshmukh, R.K.; Vivancos, J.; Guerin, V.; Sonah, H.; Labbe, C.; Belzile, F.; Belanger, R.R. Identification and functional characterization of silicon transporters in soybean using comparative genomics of major intrinsic proteins in *Arabidopsis* and rice. *Plant Mol. Biol.* **2013**, *83*, 303–315, doi:10.1007/s11103-013-0087-3.
59. Boldt, J.K.; Locke, J.C.; Altland, J.E. Silicon accumulation and distribution in petunia and sunflower grown in a rice hull-amended substrate. *HortScience* **2018**, *53*, 698–703.
60. Bienert, G.P.; Schüssler, M.D.; Jahn, T.P. Metalloids: Essential, beneficial or toxic? Major intrinsic proteins sort it out. *Trends Biochem. Sci.* **2008**, *33*, 20–26.
61. Benvenuto, M.L.; Osterrieth, M.L. Silicophytoliths from soybean plants in different growth stages of the Argentine Pampas. *Braz. J. Bot.* **2016**, *39*, 337–347.
62. Zellner, W.; Lawrence, D. Silicon as a biostimulant in agriculture. *Burleigh Dodds Ser. Agric. Sci.* **2020**, *1*:1–38.
63. Mitani, N.; Yamaji, N.; Ma, J.F. Identification of maize silicon influx transporters. *Plant Cell Physiol.* **2009**, *50*, 5–12, doi:10.1093/pcp/pcn110.
64. Raiola, A.; Lionetti, V.; Elmaghraby, I.; Immerzeel, P.; Mellerowicz, E.J.; Salvi, G.; Cervone, F.; Bellincampi, D. Pectin methylesterase is induced in *Arabidopsis* upon infection and is necessary for a successful colonization by necrotrophic pathogens. *Mol. Plant Microbe Interact.* **2011**, *24*, 432–440.
65. He, C.; Wang, L.; Liu, J.; Liu, X.; Li, X.; Ma, J.; Lin, Y.; Xu, F. Evidence for ‘silicon’ within the cell walls of suspension-cultured rice cells. *New Phytol.* **2013**, *200*, 700–709.
66. Głazowska, S.; Baldwin, L.; Mravec, J.; Bukh, C.; Hansen, T.H.; Jensen, M.M.; Fangel, J.U.; Willats, W.G.; Glasius, M.; Felby, C. The impact of silicon on cell wall composition and enzymatic saccharification of *Brachypodium distachyon*. *Biotechnol. Biofuels* **2018**, *11*, 171.
67. Novaković, L.; Guo, T.; Bacic, A.; Sampathkumar, A.; Johnson, K.L. Hitting the wall—Sensing and signaling pathways involved in plant cell wall remodeling in response to abiotic stress. *Plants* **2018**, *7*, 89.
68. Rui, Y.; Dinnen, J.R. A wall with integrity: Surveillance and maintenance of the plant cell wall under stress. *New Phytol.* **2020**, *225*, 1428–1439.

69. Wang, M.; Gao, L.; Dong, S.; Sun, Y.; Shen, Q.; Guo, S. Role of silicon on plant–pathogen interactions. *Front. Plant Sci.* **2017**, *8*, 701.
70. Ali, M.A.; Azeem, F.; Nawaz, M.A.; Acet, T.; Abbas, A.; Imran, Q.M.; Shah, K.H.; Rehman, H.M.; Chung, G.; Yang, S.H. Transcription factors WRKY11 and WRKY17 are involved in abiotic stress responses in Arabidopsis. *J. Plant Physiol.* **2018**, *226*, 12–21.
71. Feng, J.; Shi, Q.; Wang, X.; Wei, M.; Yang, F.; Xu, H. Silicon supplementation ameliorated the inhibition of photosynthesis and nitrate metabolism by cadmium (Cd) toxicity in Cucumis sativus L. *Sci. Hortic.* **2010**, *123*, 521–530.
72. Jia-Wen, W.; Yu, S.; Yong-Xing, Z.; Yi-Chao, W.; Hai-Jun, G. Mechanisms of enhanced heavy metal tolerance in plants by silicon: A review. *Pedosphere* **2013**, *23*, 815–825.
73. Kim, Y.H.; Khan, A.L.; Kim, D.-H.; Lee, S.Y.; Kim, K.M.; Waqas, M.; Jung, H.Y.; Shin, J.H.; Kim, J.G.; Lee, I.J. Silicon mitigates heavy metal stress by regulating P-type heavy metal ATPases, Oryza sativa low silicon genes, and endogenous phytohormones. *BMC Plant Biology* **2014**, *14*, 1–13.
74. Songnuan, W. Wind-pollination and the roles of pollen allergenic proteins. *Asian Pac. J. Allergy Immunol.* **2013**, *31*, 261.
75. Brugiére, T.; Exley, C. Callose-associated silica deposition in Arabidopsis. *J. Trace Elem. Med. Biol.* **2017**, *39*, 86–90.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).