


# Deep learning-enabled discovery and characterization of *HKT* genes in *Spartina alterniflora*

Maogeng Yang<sup>1,2,3,†</sup>, Shoukun Chen<sup>1,2,4,†</sup>, Zhangping Huang<sup>1,2,†</sup>, Shang Gao<sup>1,2</sup>, Tingxi Yu<sup>1,2</sup>, Tingting Du<sup>1,2</sup>, Hao Zhang<sup>1,2</sup>, Xiang Li<sup>5</sup>, Chun-Ming Liu<sup>1,6,7,8,\*</sup>, Shihua Chen<sup>3,\*</sup> and Huihui Li<sup>1,2,\*</sup> 

<sup>1</sup>State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China,

<sup>2</sup>Nanfan Research Institute, CAAS, Sanya, Hainan, China,

<sup>3</sup>Key Laboratory of Plant Molecular & Developmental Biology, College of Life Sciences, Yantai University, Yantai, Shandong, China,

<sup>4</sup>Hainan Yazhou Bay Seed Laboratory, Sanya, Hainan, China,

<sup>5</sup>State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China,

<sup>6</sup>Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing, China,

<sup>7</sup>College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, and

<sup>8</sup>School of Advanced Agricultural Sciences, Peking University, Beijing, China

Received 10 April 2023; revised 3 July 2023; accepted 11 July 2023.

\*For correspondence (e-mail lihuihui@caas.cn; chensh@ytu.edu.cn; liuchunming@pku.edu.cn).

†These authors contributed equally.

## SUMMARY

*Spartina alterniflora* is a halophyte that can survive in high-salinity environments, and it is phylogenetically close to important cereal crops, such as maize and rice. It is of scientific interest to understand why *S. alterniflora* can live under such extremely stressful conditions. The molecular mechanism underlying its high-saline tolerance is still largely unknown. Here we investigated the possibility that high-affinity K<sup>+</sup> transporters (HKTs), which function in salt tolerance and maintenance of ion homeostasis in plants, are responsible for salt tolerance in *S. alterniflora*. To overcome the imprecision and unstable of the gene screening method caused by the conventional sequence alignment, we used a deep learning method, DeepGOPlus, to automatically extract sequence and protein characteristics from our newly assemble *S. alterniflora* genome to identify SaHKTs. Results showed that a total of 16 *HKT* genes were identified. The number of *S. alterniflora* HKTs (SaHKTs) is larger than that in all other investigated plant species except wheat. Phylogenetically related SaHKT members had similar gene structures, conserved protein domains and *cis*-elements. Expression profiling showed that most *SaHKT* genes are expressed in specific tissues and are differentially expressed under salt stress. Yeast complementation expression analysis showed that type I members SaHKT1;2, SaHKT1;3 and SaHKT1;8 and type II members SaHKT2;1, SaHKT2;3 and SaHKT2;4 had low-affinity K<sup>+</sup> uptake ability and that type II members showed stronger K<sup>+</sup> affinity than rice and *Arabidopsis* HKTs, as well as most SaHKTs showed preference for Na<sup>+</sup> transport. We believe the deep learning-based methods are powerful approaches to uncovering new functional genes, and the *SaHKT* genes identified are important resources for breeding new varieties of salt-tolerant crops.

**Keywords:** deep learning, *Spartina alterniflora*, HKT, salinity tolerance.

## INTRODUCTION

Abiotic stresses associated with the emergence of global climate change such as extreme drought, high temperature and saline-alkali soils devastate agricultural production, and salt stress is widespread and severely limits agricultural production (Silveira & Carvalho, 2016). Exposure of plants to salinity induces physiological and biochemical

reactions, including selective uptake and exclusion of ions, accumulation and synthesis of organic solutes, and a change in membrane composition (Hameed et al., 2021; Silveira & Carvalho, 2016). Salt in soils usually exists in the form of sodium ions, and ion toxicity is triggered when Na<sup>+</sup> concentrations reach a certain threshold level. Under

these conditions,  $K^+$  maintains ionic homeostasis at the cellular level by regulating the transfer and translocation of  $Na^+$  (Hauser & Horie, 2010).

High-affinity  $K^+$  transporters (HKTs) are plant ion transporters that are mainly responsible for coordinating  $Na^+$  or  $K^+$  transport in response to salt stress (Riedelsberger et al., 2021). They are part of the Trk/Ktr/HKT family and are characterized by a typical TrkH domain, which consists of four membrane-pore-membrane (MPM) motifs (Riedelsberger et al., 2021). There are two classes of HKTs with different numbers of serine or glycine residues in the first pore loop ( $P_A$ ), type I (Ser-Gly-Gly-Gly) and type II (Gly-Gly-Gly-Gly), which show class-specific differences in ion conduction (Platten et al., 2006). In general, type I HKTs are  $Na^+$  uniporters. For example, *Arabidopsis thaliana* AtHKT1;1; *Oryza sativa* L. (rice) OsHKT1;1, OsHKT1;3 and OsHKT1;5; and *Dionaea muscipula* DmHKT1 were shown to be selective  $Na^+$  transporters in a *Xenopus laevis* oocyte expression system (Garcia-deblás et al., 2003; Hauser & Horie, 2010; Jabnourne et al., 2009; Kato et al., 2001). AtHKT1;1, *Triticum aestivum* L. (wheat) TaHKT1;5-D and *Hordeum vulgare* L. (barley) HvHKT1;1 limit the accumulation of  $Na^+$  in the above-ground parts and maintain normal  $Na^+$  concentrations in the leaves (Byrt et al., 2014; Han et al., 2018; Rus et al., 2004). Type II HKTs mediate  $Na^+$  and  $K^+$  symport (Uozumi et al., 2000) and have only been found in monocotyledonous species so far (Wei et al., 2021). The activities of type II transporters are often ion concentration dependent. For example, rice OsHKT2;1 exhibits diverse transport activities depending on the  $Na^+$  and  $K^+$  concentrations, including  $Na^+$  uniport and  $Na^+$ - $K^+$  symport (Hauser & Horie, 2010; Jabnourne et al., 2009; Platten et al., 2006), and wheat TaHKT2;1 can either increase  $K^+$  or decrease  $Na^+$  uptake efficiency in yeast depending on the ion concentration (Horie et al., 2011).

HKT genes have been investigated in different plant species including one in *Arabidopsis* (Uozumi et al., 2000), nine in rice (Garcia-deblás et al., 2003), nine in barley (Huang et al., 2008), four in sorghum (*Sorghum bicolor* L.) (Anil Kumar et al., 2022), two in poplar (*Populus trichocarpa* L.) and two in tomato (*Solanum lycopersicum* L.) (Li et al., 2019). Halophytes are naturally salt-tolerant plants that grow in soils with high NaCl content and can be classified into three categories according to their uptake, storage and secretion of salts: pseudohalophytes, euhalophytes and recretohalophytes (Mishra & Tanna, 2017). Although HKT genes have been widely demonstrated to play an important role in the transport of  $Na^+$  and  $K^+$  in plants, the molecular mechanism of the HKT genes still needs to be fully resolved in plants, especially in halophytes such as *Spartina alterniflora* Loisel. *S. alterniflora* belongs to the recretohalophytes and it is the only halophyte that can survive in the highly saline environment (up to 3.5% salinity) of salt marshes, using salt glands to excrete excess salts to

the body surface. It is a highly salt-tolerant member of the Gramineae family that is closely related to many important cereal crops such as wheat, barley, maize, rice, sorghum and foxtail millet. *S. alterniflora* is an ideal model plant for functional genomics studies of Gramineae as well as for halophyte research. However, little is known about the salt tolerance mechanism in *S. alterniflora*. It has been shown that overexpression of *S. alterniflora* vacuolar  $H^+$ -ATPase subunit c1 (*SaVHAc1*), Salt Responsive Protein 3–1 (*SaSRP3-1*) and Regulator of Chromosome Condensation 1 (*SaRCC1*) in rice can improve plant salt tolerance (Baisakh et al., 2012; Biradar et al., 2018; Li et al., 2022), indicating that the utilization of *S. alterniflora* genes has great potential in plant salt tolerance. Since HKT genes have been reported to improve plant salt tolerance, it is therefore of great interest to investigate the possible roles of the *S. alterniflora* HKT genes in salt tolerance.

Deep learning (DL) is a tool that enables better and more efficient data mining and analysis. Its application in plant molecular research has been limited to a few species and a limited algorithm set. Recently, DL models have been widely used in the field of molecular biology for tasks such as gene identification, protein structure prediction and DNA element identification. For example, the DL model AlphaFold2 (Tsaban et al., 2022) can predict protein conformational changes induced upon peptide binding, which is difficult to achieve with traditional biophysical algorithms. Prediction models based on a multilayer DL framework, including Deep4mC (Xu et al., 2021), Deep4mCPred (Zeng & Liao, 2020) and DeepTorrent (Liu et al., 2021), can effectively capture key features and successfully identify DNA N4-methylcytosine modification sites. The DeepGOPlus convolutional network model proposed by Kulmanov and Hoehndorf (2020) has been used to predict cellular component ontologies by predicting functional protein sequences and combining them with the functions of similar proteins. Yan et al. (2022) developed an attention-based multi-label neural network DL model, PlantBind, to predict plant transcription factor binding sites; and Zhang et al. (2021) developed the D-AEDNet, a DL-based model, to identify the location of transcription factors binding sites in DNA sequences; Sun et al. (2021) performed RNA structural information modelling by using a deep neural network (PrismNet), which could accurately model and predict RNA binding protein targets in vivo. These studies demonstrate that DL is an effective tool for turning data into knowledge. However, DL has not yet been used to discover new functional genes.

To identify functional genes, the traditional method is based on conventional sequence alignment to establish homology. But the number of genes retained largely depended on the artificially predefined threshold, that is the similarity score. To overcome the weakness of traditional methods, here, we aimed to identify HKT genes

using a cutting-edge protein function prediction method based on a DL framework from our newly generated *S. alterniflora* genome assembly and investigate their roles in K<sup>+</sup> and Na<sup>+</sup> transport via yeast complementation expression analysis. In addition, we also investigated phylogenetic relationships, gene components, *cis*-elements and expression profiles of *SaHKT* genes. Our study shows the great potential of DL frameworks for deciphering functional genes and lays the foundation for a better understanding of the functional mechanisms of the *HKT* genes in *S. alterniflora*. In addition, the *SaHKT* genes identified are an important resource for improving salt tolerance in staple food crops.

## RESULTS

### Workflow of DeepGOPlus to identify gene families

Traditionally, gene families were identified by conventional sequence alignment, which requires manual setting of screening thresholds, resulting in the high risk of elimination of functional genes and an unstable number of genes identified. Without any artificial parameters, the superiority of DL-based methods over traditional sequence alignment is that DL could automatically extract appropriate and complicated sequence and protein features, and then classify genes accordingly. DeepGOPlus is a novel DL method for predicting protein functions from protein sequence alone which combines a deep convolutional neural network (CNN) model with sequence similarity-based predictions. Compared with the inaccuracy of gene screening caused by conventional methods based on sequence alignment, DeepGOPlus can automatically obtain complex and applicable sequence features, accurately identify the sequence functions and then efficiently classify genes. The working principle of DeepGOPlus used in our research is as follows (Figure 1a): first, all protein sequences of each species were vectored using one-click encoding to obtain a set of 21 × 2000 input matrices. This matrix was then fed into a convolutional layer and feature vectors with different dimensions were generated. Finally, a fully connected layer (classifier) was activated by a sigmoid activation function to output the classification predictions. After screening genes using DeepGOPlus, an experimental procedure was used to demonstrate the accuracy of the recognition results (Figure 1b).

### *HKT* genes in 54 plant species were identified by a DL framework

To explore the evolutionary conservations and diversifications of *HKT* genes in plants, we first identified the *HKT* genes from 54 plant species, namely 13 monocot plants, 24 eudicots, 2 basal angiosperm (*Amborella trichopoda* and *Nymphaea colorate*), 7 pteridophytes, 2 bryophytes (*Sphagnum fallax* and *Physcomitrium patens*) and 6

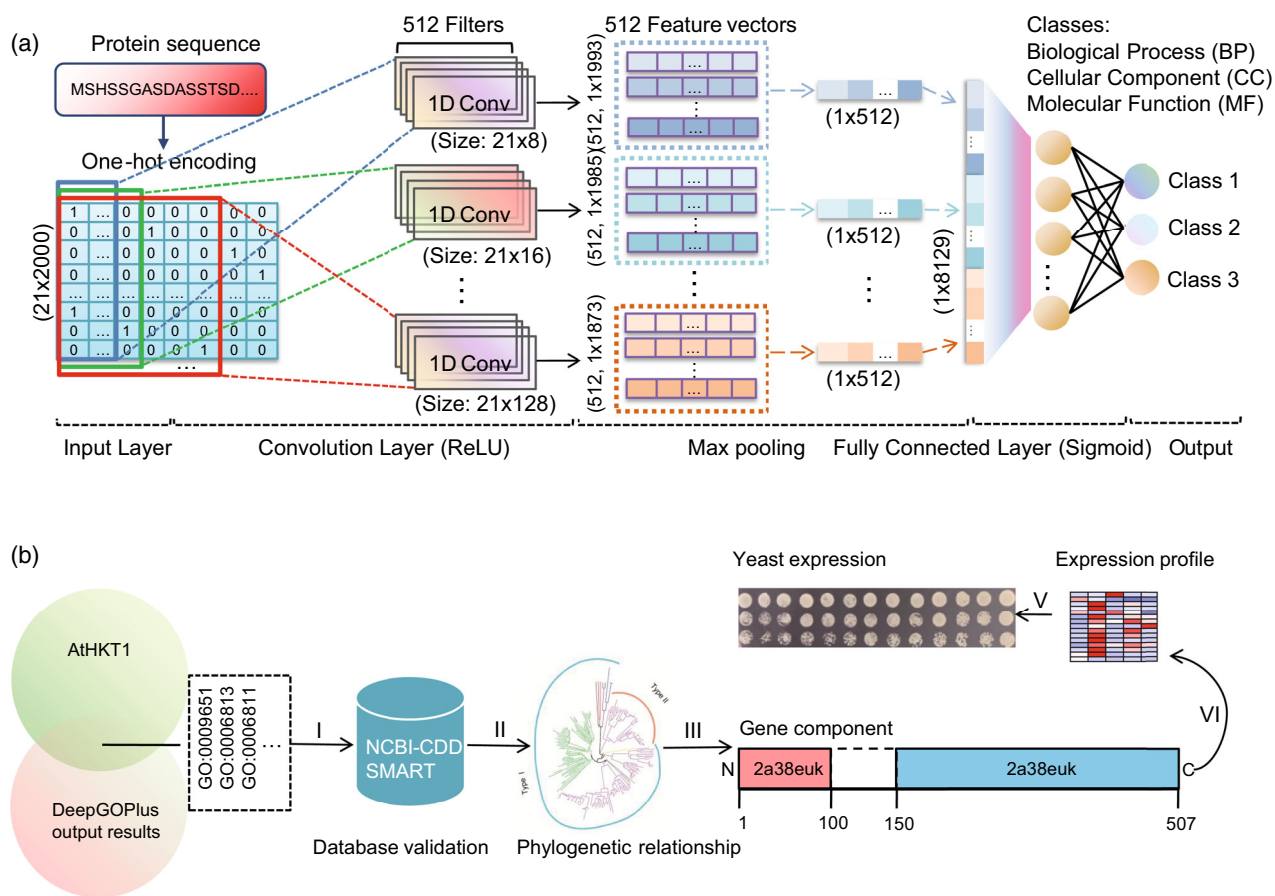
chlorophytes using the DeepGOPlus method (Kulmanov & Hoehndorf, 2020). Predictions of gene ontology (GO) terms by DeepGOPlus showed that all the identified *HKT*s were involved in Na<sup>+</sup>/K<sup>+</sup> transport and response to salt stress, and each of them showed the best hit to the *Arabidopsis* *HKT1* protein sequences in BLAST searches. Searches against the NCBI-conserved domains database and SMART database further confirmed that the *HKT* members identified by DeepGOPlus have a typical TrkH domain. According to the presence of serine or glycine residues in P<sub>A</sub> in the TrkH domain, the *HKT* proteins were classified into two types, I and II. The number of type I members was larger than that of type II, and they were found in all 46 plant species. In accordance with previous reports (Hauser & Horie, 2010; Véry et al., 2014), the type II *HKT* members were only found in monocots (Figure 2a), indicating that they may have arisen from gene duplication of type I members and may be the products of the adaptation of monocots plants to the environment.

*T. aestivum* contained the most *HKT*s (23 genes), followed by *S. alterniflora* (16) and *Triticum turgidum* (14) (Figure 2a). The reason for their greater number than other species may be due to heterozygous polyploidy and genomic complexity. For example, wheat was a heterozygous hexaploidy with 2.5 and 3.3 times more *HKT* genes than rice and barley, respectively. While genome size was not the main reason for the higher number of effects. For example, the genome of rice was smaller than that of *Aegilops tauschii* and *Triticum urartu*, but the number of rice *HKT* genes was higher than that of these two species. There were more type I *HKT*s than type II *HKT*s in the monocot species. For example, the number of type I *HKT*s in *S. alterniflora* was three times that of type II *HKT*s. *Eucalyptus grandis* contained the most *HKT*s among eudicots, and the numbers in the other eudicots ranged between one and five. Only two *HKT*s were identified in bryophyta, and no one identified in chlorophyta.

### Phylogenetic and duplication analyses of plant *HKT*s

An unrooted neighbour-joining (NJ) tree was constructed to visualize the phylogenetic relationships between *HKT*s. The two types of *HKT*s formed separated branches (Figure 2b), further supporting the classification according to P<sub>A</sub> residues. For the type I *HKT*s, they were further divided into four groups (groups A, B, C and D). Group A includes only all investigated eudicot *HKT*s, group B contains only the monocot *HKT*s, group C contains only all basal angiosperm *HKT*s and group D contains only all pteridophyte and bryophyta *HKT*s. Two or more copies were found in several eudicots, basal angiosperm, pteridophyte and bryophyta *HKT*s, indicating that a gene duplication event occurred during their evolution (Figure S1). For wheat and *S. alterniflora*, two species with a large amount

## 4 Maogeng Yang et al.



**Figure 1.** Flowchart showing the deep learning-based method and the gene validation procedure.

(a) The general workflow of the deep learning-based methods.

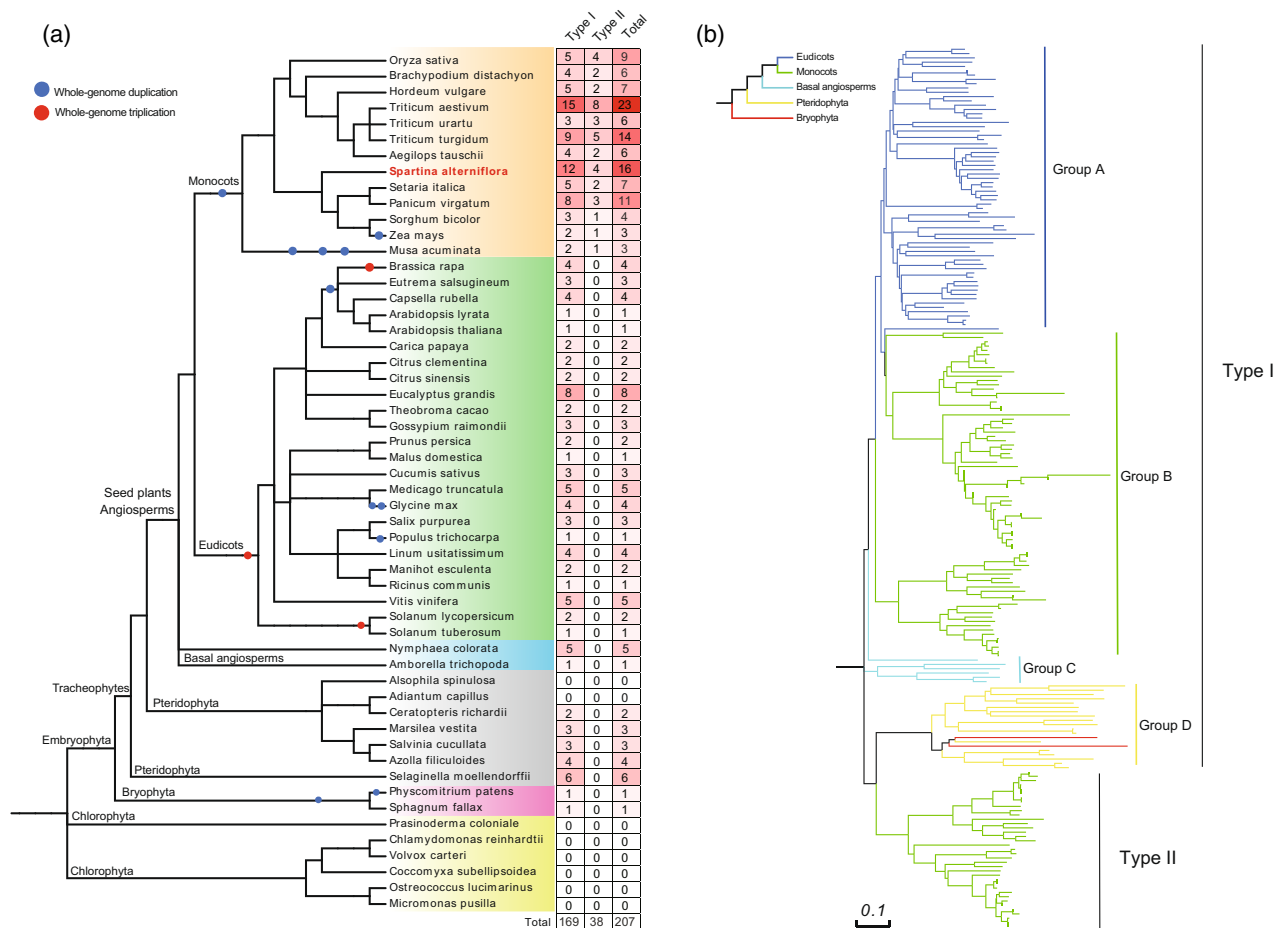
(b) The process used for experimental validation of *S. alterniflora* HKT gene based on deep learning prediction results, including (I) database validation, (II) phylogenetic relation tree construction, (III) gene component analysis, (IV) expression profiling and (V) yeast expression analysis.

of HKTs showed many gene copies within species. And many monophyletic groups were found in both groups A, C and D, four and 29 duplications were survived from segmentally and tandemly duplicated in these groups, further indicating that these groups underwent duplication during long-term diversification. In addition, type II HKTs constructed a group independently, and this group showed phylogenetically close to the group D HKTs, indicating that a recent duplication event occurred before the generation of grass and that the resulting type II HKTs were retained after duplication. To observe the differences in the protein structure of HKTs, conserved regions analysis was further analysed (Figure S2). Results found that domain 9 was conserved in all HKTs, domain 8 was only identified in seed plants and domain 15 and domain 18 were only found in the type II HKTs. Compared with the seed plants, many domains did not exist in pteridophyta and bryophyta, indicating that the seed plants' HKT genes may evolve new domains in order to obtain functions during evolution and they may be the products of adaptation to the environment.

HKT proteins typically contain four MPM domains, and the pore regions ( $P_A$ ,  $P_B$ ,  $P_C$  and  $P_D$ ) are the key structure of the MPM domains. Figure S3 illustrates a high degree of amino acid sequence conservation in the four pore domains of the distinct evolutionary groups. The  $P_D$  appears the most conserved domain. Many residues showed group specificity. For example, the 9th amino acid of the  $P_A$  is presented only in groups B, C and type II HKTs, indicating that the monocots HKTs maybe evolved from the basal angiosperms. And the 21st residue, serine, is identified only in the  $P_A$ .

#### *S. alterniflora* HKT genes

Membrane transporter proteins are embedded in membrane-bound organelles and are prime targets for improving the efficiency of water and nutrient transport. In *S. alterniflora*, 559 members were involved in ion homeostasis that was identified using DeepGOPlus, and this number was larger than that in *Arabidopsis* (226), rice (249) and maize (280) (Figure S4 and Table S1). The HKT protein was one of the ion transport proteins involved in the transport



**Figure 2.** Phylogenetic relationships between *HKT* genes.

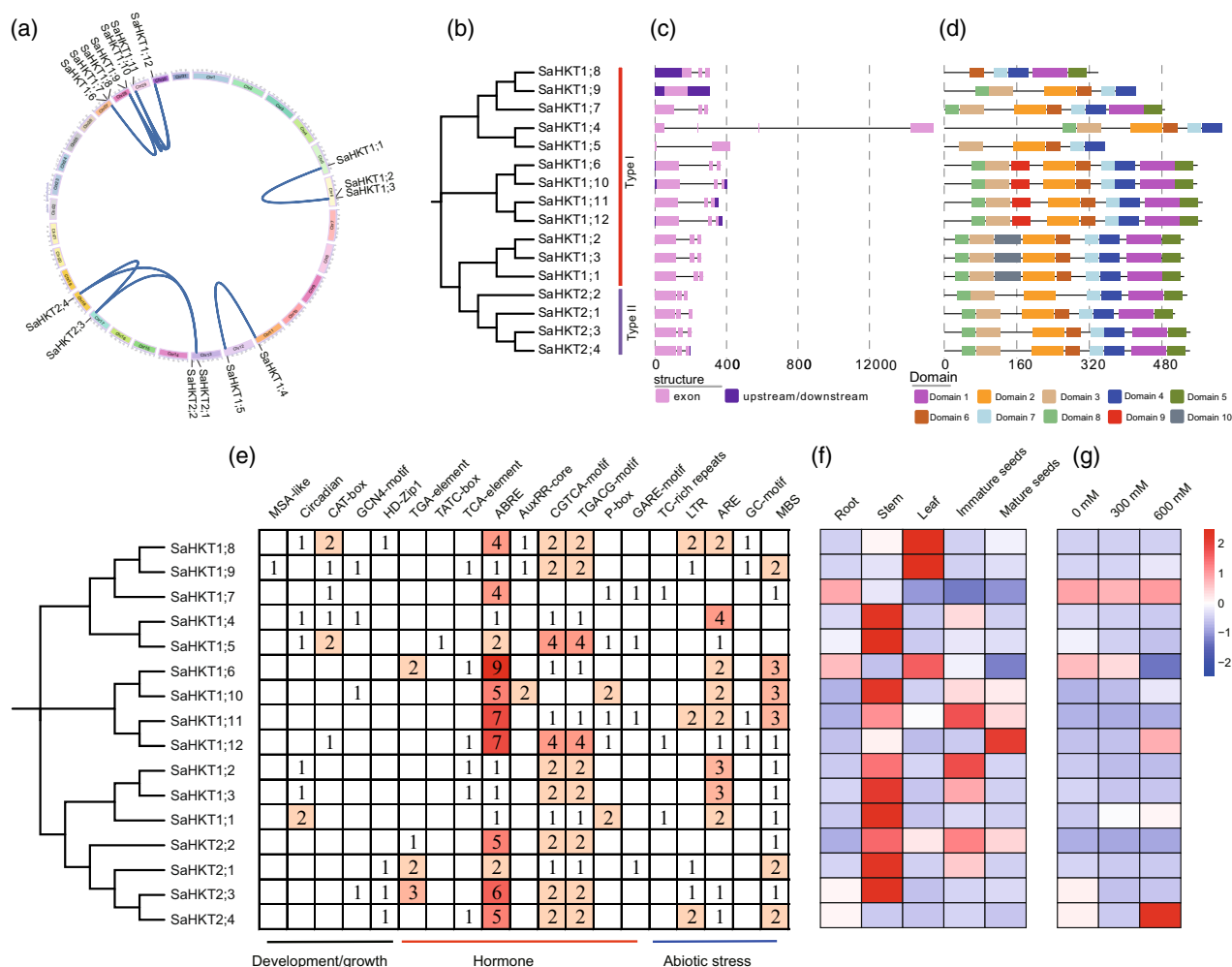
(a) Number of *HKT* genes identified in 54 plant species. Deep red indicates a larger number of *HKT* genes. Red and blue circles indicate whole-genome duplication and triplication, respectively.

(b) Neighbor-joining tree of the *HKT* gene family. Four groups, including groups A, B, C and D, were classified into type I HKTs.

of  $K^+$ . A total of 16 *HKT* genes were identified in the *S. alterniflora* genome using DeepGOPlus (Figure 2a), namely 12 type I and four type II SaHKTs. They were named *SaHKT1;1* to *SaHKT2;4* according to the chromosome locations and homology with *OsHKT* genes (Table S2). The 16 *SaHKT* genes were unevenly distributed across the 11 *S. alterniflora* chromosomes (Figure 3a and Table S2) and had lengths ranging from 339 (*SaHKT1;8*) to 612 (*SaHKT1;4*) amino acids and predicted molecular weights ranging from 38.12 (*SaHKT1;8*) to 67.86 (*SaHKT1;4*) kDa (Table S2). Subcellular localization prediction showed that all members were predicted to be localized only in the cytoplasmic membranes. To verify this, we took *SaHKT1;4* and *SaHKT2;4* as examples for subcellular localization analyses. Two constructs, *SaHKT1;4-eGFP* and *SaHKT2;4-eGFP*, were made in which *SaHKT1;4* and *SaHKT2;4* were fused in-frame with *eGFP*, expressed under the control of the  $2 \times CaMV35S$  promoter (Chen, Wang, et al., 2021), and transiently expressed in rice protoplasts. Examination of

protoplasts at 48 h after the transfection showed that both *SaHKT1;4-eGFP* and *SaHKT2;4-eGFP* were located in cytoplasmic membranes (Figure S5), in contrast to the cytoplasmic localization of the control construct of *pAN580-eGFP*. These results are consistent with those reported for HKTs in other plant species (Cao et al., 2019; Horie et al., 2011; Huang et al., 2020; Imran et al., 2022; Wu et al., 2020). We also used homology comparisons and hidden Markov models to predict the *S. alterniflora* *HKT* genes, respectively. The results showed that the genes identified by hidden Markov models were consistent with those identified by DeepGOPlus. Although there are also 16 SaHKTs were identified by homology comparisons, the percentages of identity/similarity between *Arabidopsis*/rice HKTs and SaHKTs were less than 50% (Table S3).

In an unrooted NJ phylogenetic tree, all *SaHKT* genes clustered into two types (Figure 3b), consistent with the classification in Figure 2a and the NJ tree of HKT proteins from multiple plant species (Figure 2b). Analysis of gene



**Figure 3.** Phylogenetic relationships, gene structures, sequences and *cis*-elements of SaHKTs.

(a) Chromosomal locations of *S. alterniflora* HKTs. Different chromosomes are shown in different colours. Lines indicate pairs of segmentally duplicated genes.

(b) NJ phylogenetic tree of full-length SaHKT proteins was constructed using MEGA 7 software with 1000 bootstrap replications.

(c) Gene structure of SaHKT genes. Exons and introns are shown in boxes and lines, respectively.

(d) SaHKT protein domains. Different boxes represent different domains.

(e) Three types of *cis*-elements were identified in the promoter regions of SaHKT genes.

(f) Expression heat map of SaHKT genes in different tissues.

(g) Expression patterns of SaHKT genes under different NaCl concentrations.

structure showed that the number of exons ranged from one (*SaHKT1;9*) to four (*SaHKT1;4*) (Figure 3c). In addition, 10 domains were identified (Figure 3d), with domains 1, 2, 4, 5, 6, 7 and 10 constituting the TrkH structural domain; of these domains, only domain 4 was found in all SaHKT proteins (Tables S4 and S5). All SaHKT proteins consisted of four MPM motifs, just like other known OshKT proteins (Figure S6). The serine or glycine residues at the P<sub>A</sub>, P<sub>B</sub>, P<sub>C</sub> and P<sub>D</sub> regions were highly conserved in all SaHKT proteins except for SaHKT1;8, which lacks the P<sub>A</sub> region but was highly conserved with type I HKT members (Figure S6). Furthermore, the HKT proteins were highly conserved in the four amino acid residues (GGGG in type I and SGGG in type II) that have been shown to confer salt

tolerance in point mutation studies (Platten et al., 2006). Phylogenetically related genes shared similar gene structures and conserved domains (Table S5).

#### Analysis of *cis*-element analysis in SaHKT promoters

We identified three types of *cis*-elements, namely growth and development elements, hormone elements and stress-responsive elements, in the 2-kb promoter regions of the SaHKT genes (Figure 3e). *Cis*-elements related to growth and development included a cell cycle regulatory element (MSA-like, TCCAACGGT), circadian rhythm regulatory element (circadian, CAAAGATATC), meristematic tissue expression element (CAT-box, GCCACT), endosperm expression element (GCN4-motif, TGAGTCA) and

fenestrated chloroplast differentiation element (HD-Zip1, CAAT(A/T)ATTG). Hormone-related elements included methyl jasmonate (MeJA)-related elements (CGTCA-motif and TGACG-motif), abscisic acid-related element (ABRE, ACGTG), gibberellin-related element (GARE-motif) and salicylic acid-related element (TCA-element, CCATCTTTT and TCAGAAGAGG). *Cis*-elements associated with abiotic stresses included the hypoxia-specific inducible element (GC-motif, CCCCCG), anaerobic inducible element (ARE, AAACA), drought-inducible *cis*-element (MBS, CAACTG) and low-temperature responsive *cis*-element (LTR, CCGAAA). The ABRE was found in the promoter regions of all *SaHKT* genes. Compared with the regulatory elements of *HKT* genes in rice and *Arabidopsis*, regulatory elements in most *HKT* genes in *S. alterniflora* have more ABRE elements (Figure S7).

### The duplications and synteny of *SaHKTs*

Gene duplication analysis detected 12 segmentally duplicated genes and nine segmental duplicated gene pairs (Figure 3a and Table S6). Three type II genes were duplicated from each other, suggesting that they may have evolved from a common ancestor. No tandem duplicate *SaHKT* genes were found. To elucidate the origin and evolutionary relationships of the *SaHKT* genes, we analyzed the synteny between *S. alterniflora* and five monocot plants, namely rice, sorghum, barley, millet and maize (Figure S8). We identified 8, 8, 7, 10 and 7 *SaHKT* genes that were syntenic to rice, sorghum, barley, millet and maize genes, respectively. However, there were no genes with synteny to *Arabidopsis*. These results indicated that *HKT* genes show a high level of synteny within monocot plants.

To detect selection pressure after duplication of *HKT* genes, we calculated the ratio of non-synonymous (Ka) to synonymous (Ks) substitution (i.e., Ka/Ks) for duplicate gene pairs. The Ka/Ks values for segmentally duplicated *SaHKT* gene pairs ranged from 0.28 to 0.76, with an average of 0.40, suggesting that these *SaHKT* genes underwent strong purifying selection during evolution. In addition, Ka/Ks ratios were calculated for gene pairs between *S. alterniflora* and five monocot plants (Table S7). The average Ka/Ks values between *S. alterniflora* and barley, maize, sorghum, foxtail millet and rice *HKTs* were 0.41, 0.38, 0.33, 0.30 and 0.24, respectively, suggesting that duplication events extended the evolution of *HKT* genes.

### Expression analyses of *SaHKT* genes

The expression levels of *SaHKT* genes in different tissues of *S. alterniflora* at the heading stage were determined utilizing quantitative reverse transcription PCR (qRT-PCR). As shown in Figures 3f and S9a, *SaHKT* genes exhibited tissue-specific expression. Ten *SaHKT* genes (*SaHKT1*;1, -1;2, -1;3, -1;4, -1;5, -1;10, -1;11, -2;1, -2;2, -2;3) were

expressed highly in stems; three genes (*SaHKT1*;6, *SaHKT1*;8 and *SaHKT1*;9) were expressed more highly in leaves than in other tissues; and *SaHKT1*;12 and *SaHKT2*;4 were specifically expressed in mature seeds and roots, respectively.

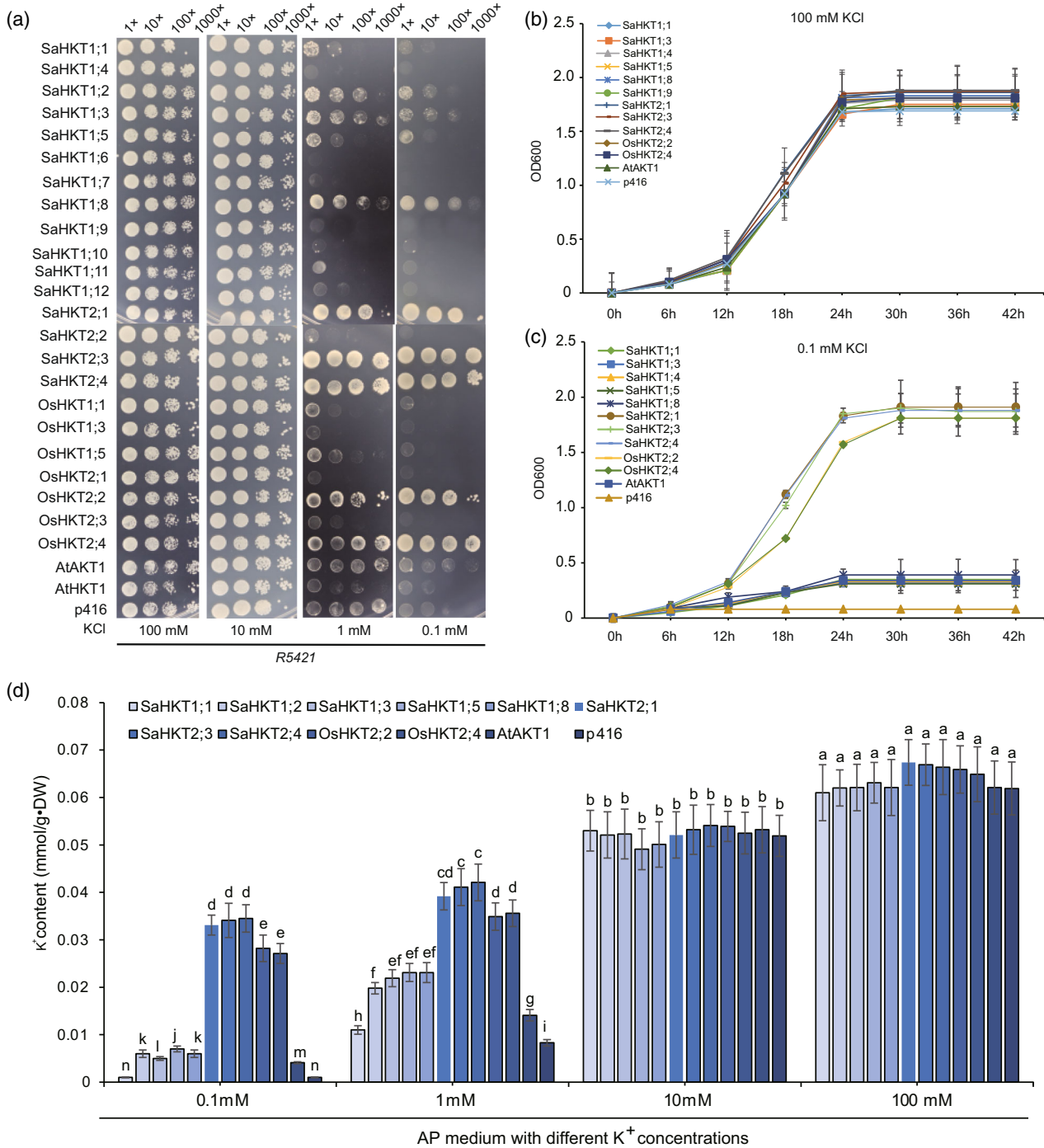
Next, we investigated the expression of *SaHKT* genes under different NaCl concentrations. All *SaHKT* genes were differentially expressed after salt treatments except for *SaHKT1*;7 and *SaHKT1*;8 (Figures 3g and S9b). *SaHKT1*;1 expression was significantly up-regulated under 300 and 600 mM NaCl compared with the control (0 mM), and *SaHKT1*;9, *SaHKT1*;10, *SaHKT1*;12 were significantly up-regulated under 600 mM NaCl. Eight *SaHKT* genes (*SaHKT1*;2, *SaHKT1*;3, *SaHKT1*;4, *SaHKT1*;5, *SaHKT1*;6, *SaHKT1*;11, *SaHKT2*;1 and *SaHKT2*;2) were down-regulated under both salt treatments, and *SaHKT2*;4 was down-regulated under 300 mM NaCl but up-regulated under 600 mM NaCl.

### Expression of *SaHKT* genes in yeast

As the *HKT* gene was encoded potassium transporters, we used a K<sup>+</sup>-transporter deficient yeast strain, *R5421*, to test the function of *SaHKT* genes in response to K<sup>+</sup>. Seven *OshKTs*, and *Arabidopsis* *AtHKT1* and *AtAKT1* were used as controls in the growth inhibition experiment. Expression of *OshKT2*;2 and *OshKT2*;4 were found to rescue the growth defects of the yeast *R5421* mutant in the presence of 0.1 mM KCl (Horie et al., 2011). *Arabidopsis* *AtHKT1* shows more Na<sup>+</sup>-selective transport activity (Rus et al., 2004; Uozumi et al., 2000), and *AtAKT1* is involved in low-affinity K<sup>+</sup> uptake at a wide range of external K<sup>+</sup> concentrations (Rubio et al., 2008; Wang et al., 2021). As shown in Figure 4a, cells expressing *SaHKT2*;1, *SaHKT2*;3 and *SaHKT2*;4 in arginine-phosphate (AP) medium containing 0.1 mM K<sup>+</sup> showed similar growth as cells expressing *OshKT2*;2 and *OshKT2*;4, and cells expressing *SaHKT1*;2, *SaHKT1*;3, *SaHKT1*;8 and *AtAKT1* accumulated significantly lower levels of K<sup>+</sup> than those expressing type II *HKT* members. However, most type I *SaHKTs* and *OshKTs*, *SaHKT2*;2, *AtHKT1*, and the empty vector p416 failed to restore growth. The *R5421* mutant strain with the empty vector p416 and all the individual *HKT* members grew normally in the AP medium with 10 and 100 mM K<sup>+</sup>.

To better understand the growth characteristics of yeast expressing these low-potassium affinity *HKT* gene, we also measured yeast growth curves and changes in intracellular and external K<sup>+</sup> contents. At 100 mM KCl, the transgenic yeast strains grew at essentially the same rate as yeast expression in the empty vector p416 (Figure 4b), whereas, at 0.1 mM KCl, only the low-potassium affinity *HKT* gene yeast strains could grow. Yeast strains expressing type II genes grew the fastest and those expression the *SaHKT* genes grew at a slightly faster rate than those expression the rice genes (Figure 4c). There was no

8 Maogeng Yang et al.



**Figure 4.** Yeast strains ectopically expressing *HKT* genes in response to K<sup>+</sup>. (a) Growth of yeast strain *R5421* cells harbouring the vector control (p416) or plasmids expressing *SaHKT*s, *OsHKT*s, *AtHKT1*, or on AP medium supplemented with 0.1, 1, 10, or 100 mM KCl. Cells were plated in 10-fold dilutions on an AP medium containing different concentrations of KCl. (b) and (c) Growth curves of the *R5421* strain transformed with the empty vector (p416) or plasmids expressing different *HKT* genes and grown in AP liquid medium with (b) 100 mM K<sup>+</sup> or (c) 0.1 mM K<sup>+</sup>. (d) K<sup>+</sup> contents of *R5421* strain transformed with empty vector (p416) or *HKT* gene plasmids in AP liquid medium with different concentrations of K<sup>+</sup>.

difference in the intracellular K<sup>+</sup> contents of the transgenic yeast strains expressing *SaHKT* genes and p416 in 10 and 100 mM KCl medium, while at 0.1 mM and 1 mM KCl, the intracellular K<sup>+</sup> contents of the yeast strains expression

*HKT* genes were higher than those of the control p416-expressing strain, and the intracellular K<sup>+</sup> contents of the yeast strains expressing *SaHKT* genes was higher than those of yeast strains expressing *OsHKT* genes (Figure 4d).



On AP medium containing 10 mM K<sup>+</sup>, cells expressing all HKTs showed no difference (Figure 4a). The K<sup>+</sup> transport activity and preference of SaHKTs over Na<sup>+</sup> were further confirmed by ectopic expression in *R5421*. *SaHKT2;1*, *SaHKT2;3*, *SaHKT2;4* and *OsHKT2;4*-expressing cells were less sensitive to high concentrations of NaCl in the presence of 1 mM K<sup>+</sup> compared with other HKTs, whereas *SaHKT2;1* and *OsHKT2;4*-expressing cells grew strongly on 10 mM K<sup>+</sup> in the presence of NaCl, which was not seen in cells expressing other HKTs investigated (Figure 5a), suggesting that *SaHKT2;1* and *OsHKT2;4* have a strong salt resistance ability.

We also used *AxT3K* to detect Na<sup>+</sup> transport in HKTs in an AP medium containing 20 mM Na<sup>+</sup>. Compared to p416, *SaHKT1;1*, *SaHKT1;2*, *SaHKT1;5*, *SaHKT1;6*, *SaHKT1;7*, *SaHKT1;11*, *SaHKT2;3* and *SaHKT2;4*-expressing *AxT3K* cells showed slightly weaker than p416, indicating that they have an ability of Na<sup>+</sup> uptake; *SaHKT1;3*, *SaHKT1;4*, *SaHKT1;9*, *SaHKT1;10*, *SaHKT1;12* and *SaHKT2;2*-expressing *AxT3K* cells grew well in high Na<sup>+</sup> medium, indicating that they have Na<sup>+</sup> efflux ability. In contrast, *SaHKT1;8* and *SaHKT2;1*-expressing *AxT3K* cells showed undetectable differences in 20 mM Na<sup>+</sup> medium (Figure 5b).

According to the results under different concentrations of K<sup>+</sup> or Na<sup>+</sup> conditions (Figures 4a and 5b), we concluded that *SaHKT1;2*, *SaHKT1;3*, *SaHKT1;5*, *SaHKT1;8*, *SaHKT2;1*, *SaHKT2;3* and *SaHKT2;4* are K<sup>+</sup>-preferring transporters, and most of the SaHKTs confer Na<sup>+</sup> transport activity except for *SaHKT1;8* and *SaHKT2;1*.

Sequence alignments illustrate a high degree of amino acid sequence conservation in the rice and *S. alterniflora* HKTs. In the four pore regions, 11 residues were found to be different between the type I and II HKTs (Figure 5c), the P<sub>D</sub> appears to be the most conserved region and nine are 100% conserved in all rice and *S. alterniflora* HKTs. The amino acid comparison between the type II HKTs in *S. alterniflora* and rice showed that the residues at the 20th (P<sub>B</sub>), 6th (P<sub>C</sub>) and 18th (P<sub>D</sub>) positions were different, indicating that these three residues were key for the functions of type II HKTs in *S. alterniflora* and rice.

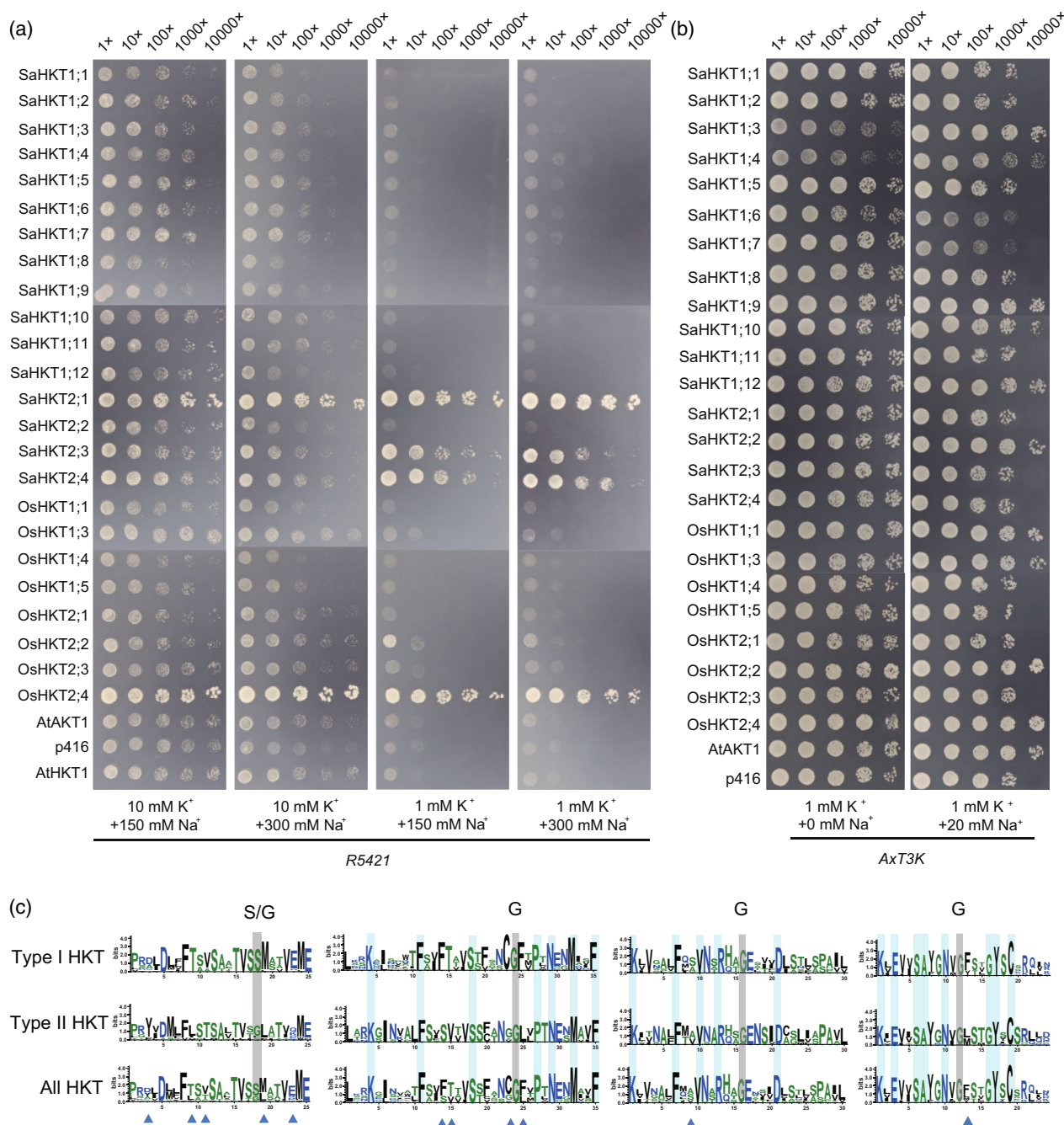
## DISCUSSION

### Deep learning is a powerful tool for functional genes identification

Accurate identification of functional genes is the first step in deciphering plant gene functions. The traditional method of identifying functional genes uses conventional sequence alignment to establish homology. However, using the same gene sequence for comparison, if the manual settings are different, the results will be different. For example, using homology comparisons, three separate studies determined that there were 241, 128 and 104 genes in the wheat genome that encode expansins, respectively

(Chen, Ren, et al., 2021; Han et al., 2019; Zhang et al., 2018); two studies identified a total of 321 and 531 AP2/ERF transcription factors in the *Brassica napus* genome (Ghorbani et al., 2020; Owji et al., 2017), respectively; and two studies classified 112 and 116 genes in the *Gossypium raimondii* genome as WRKYs, respectively (Ding et al., 2015; Dou et al., 2014). These results demonstrate that homology comparisons can lead to vastly different findings; the requirement for manual selection can cause inaccurate elimination of many functional genes.

There have been several key advances in artificial intelligence technologies recently. Notably, DL has enabled unprecedented advances in the areas of protein structural and functional predictions (Sapoval et al., 2022). For example, DeepSF is a 30-layer CNN that classifies protein sequences based on a set of 1195 protein folding types (Hou et al., 2018); DeepFam is a DL model that can accurately classify protein functions (Seo et al., 2018); and DeepHiFam categorizes proteins into functional families with higher accuracy than previous methods (Sandaruwan & Wannige, 2021). These DL tools have shown outstanding performance and can provide excellent technical support for biological research. Here, we used DeepGOPlus, a deep CNN that learns sequence and protein features and combines those with predictions based on sequence similarity (Kulmanov & Hoehndorf, 2020), to identify *S. alterniflora* HKT genes. The functions of the 16 HKT genes identified with DeepGOPlus were validated with yeast complementation assays, demonstrating the utility of DL models for accurately predicting gene families and functional genes. Although we also identified 16 *SaHKT* genes through homology comparisons using *Arabidopsis* and rice HKT genes, most of the similarity scores between the *Arabidopsis*/rice HKTs and the *SaHKTs* were < 50% (Table S3). For example, in the results of the homology comparison using the *Arabidopsis* AtHKT1 protein to BLAST against the SaHKTs and none had a similarity score greater than 50%; using rice *OsHKT1;4* for homology comparison, only four SaHKTs had similarity scores greater than 70% (between 70% and 74%), and the other 12 SaHKTs were all less than 50% including six in the range from 40% to 50% and six in the range of 30%–40%; and using the other *OsHKTs* to BLAST against the SaHKTs, the results showed that the similarity score was almost less than 50%. Therefore, it indicated that the higher the similarity score threshold (e.g. >80%), the higher risk to eliminate the functional genes. In other words, the number of genes declared depended on the threshold of the similarity score we predefined. Compared to traditional methods, DeepGOPlus can yield more accurate and stable results than conventional sequence alignment. The application of DL methods to plant biology enables functional interpretation of sequence feature predictions, accelerating the discovery of biologically meaningful results from large sequencing datasets.



**Figure 5.** Yeast strains ectopically expressing *HKT* genes in response to Na<sup>+</sup>.

(a) Growth of yeast strain *R5421* cells expressing *HKT*s under different concentrations of KCl and NaCl. The control represents 100 mM KCl. All plates and liquid medium cultures were incubated at 30°C, and photographed after 3–5 days.

(b) Growth of yeast strain *Axt3K* cells expressing the *HKT* genes under different concentrations of NaCl. All plates and liquid medium cultures were incubated at 30°C, and photographs were taken after 3–5 days.

(c) Sequence logos of the four pore regions specified for type I and II *HKT*s in rice and *S. alterniflora*. The triangle indicates the difference in residues between the *HKT*s. S/G and G indicate the selectivity filter residues. A motif switch in P<sub>A</sub>, including the selectivity filter residue, is shaded, as well as the selectivity filter residues in P<sub>B</sub>, P<sub>C</sub> and P<sub>D</sub>.

### Gene duplication promoted the expansion of *HKT*s

Gene duplication is a major mechanism for generating new genes (Magadum et al., 2013). Here we found that the

number of type I *HKT*s in each plant species investigated was greater than that of type II *HKT*s, suggesting that type I *HKT* genes might have emerged first, and that type II *HKT*

genes might have arisen from gene duplication. One example is wheat type II HKTs. There are eight type II members in hexaploidy wheat, which is not the expected multiple of three, while the number of in type I HKT members, 15, is a multiple of three. Given these observations combined with the finding results of no type II member in dicotyledons, we speculate that type II members might have arisen from the duplication of type I members.

Duplicated genes often lose their original functions and/or acquire new functions that enhance the adaptability of plants (Magadum et al., 2013). Divergence in expression in different tissues or in response to stress might be the reason for the retention of duplicated genes in plant genomes (Gu et al., 2002). Among the 16 *SaHKT* genes, 12 were duplicated, forming nine segmentally duplicated pairs. Some of these gene pairs showed different expression patterns under salt stress. For example, *SaHKT1;1* exhibited high expression after salt stress, whereas expression of its duplicate *SaHKT1;2* was down-regulated; the pairs of *SaHKT1;10-SaHKT1;11* and *SaHKT1;10-SaHKT1;12* also showed different expression levels under salt stress. In general, gene duplication occurred during the evolution of plants, and these events resulted in the current size of the plant HKT gene family.

#### The identified *SaHKT* proteins show stronger K<sup>+</sup> affinity than those from plant species

We identified 16 *SaHKT* genes in the *S. alterniflora* genome. This number is larger than that in species that have been studied to date, including *Arabidopsis*, rice, barley and sea barley (*Hordeum marinum*) (Anil Kumar et al., 2022; Byrt et al., 2014; Horie et al., 2009; Kuang et al., 2022; Mian et al., 2011; Tada, 2019). Of these species, sea barley is also a halophyte belonging to grass plants, and it has a larger genome than *S. alterniflora*. However, only four type I and three type II HKT transporters were identified in the sea barley genome (Kuang et al., 2022). Therefore, a larger number of *SaHKT* genes indicates that the *SaHKT* may play important roles in the adaption of *S. alterniflora* to saline environments.

The P-loop hypothesis has been proposed that the selective-pore-forming region, referred to as the MPM motif (Horie et al., 2001, 2009; Uozumi et al., 2000). The conserved Gly residue in each P-loop structural domain was reported to play a key role in the K<sup>+</sup> selectivity, and it is homologous to the first Gly in the highly conserved GYG sequence, which determines K<sup>+</sup> selectivity (Mäser et al., 2002; Tholema et al., 2005). In this study, yeast complementation experiments showed that *SaHKT2;1*, *SaHKT2;3* and *SaHKT2;4* exhibited stronger K<sup>+</sup>-affinity than rice *OsHKT2;2* and *OsHKT2;4*. Although *SaHKT1;2*, *SaHKT1;3* and *SaHKT1;8* did not exhibit strong K<sup>+</sup> affinity, they were also able to complement *R5421* at low K<sup>+</sup> concentrations. In general, type I HKTs are Na<sup>+</sup> uniporters,

whereas type II HKTs mediate Na<sup>+</sup> and K<sup>+</sup> symport (Horie et al., 2007; Rus et al., 2004), thus we believe that *SaHKT1;2*, *SaHKT1;3* and *SaHKT1;8* proteins are unique to *S. alterniflora*. Compared to other *SaHKT* proteins, *SaHKT1;2* and *SaHKT1;3* lack the domain 9, and *SaHKT1;8* has a shorter amino acid length and fewer domains (Figure 3d). The *cis*-elements in the promoters of these three genes are not significantly different from other *SaHKT* genes (Figure 3e). Different from other *SaHKTs*, *SaHKT1;2* and *SaHKT1;3* have conserved Met amino acid in the 28th of the P<sub>B</sub> and the 9th of P<sub>D</sub> regions, indicating that the Met in these positions may be key for the low K<sup>+</sup> affinity. *SaHKT1;8* lacks the Gly residue in the P<sub>A</sub> region. It was still able to rescue the growth defects of *R5421* yeast cells, indicating that the loss of the first membrane-pore-membrane motif conferred *SaHKT1;8* with the ability for K<sup>+</sup> uptake. In addition, the *SaHKT2;2* protein lacks domain 6; this domain maybe is essential for K<sup>+</sup> selectivity; however, molecular experiments are still needed to test this. Our finding indicates that *S. alterniflora* type I HKT members also have a distinct K<sup>+</sup> affinity and believe that *SaHKT1;2*, *SaHKT1;3* and *SaHKT1;8* proteins are unique to *S. alterniflora*.

#### HKTs mediate K<sup>+</sup>/Na<sup>+</sup> transport in *S. alterniflora*

HKT transporters participate in Na<sup>+</sup> and K<sup>+</sup> transport and homeostasis in plants (Horie et al., 2009), and studies have shown that the ion transport properties of HKTs are dependent on the Na<sup>+</sup> and K<sup>+</sup> concentrations in the external environment. For example, wheat *TaHKT2;1* has been shown to be a Na<sup>+</sup>-K<sup>+</sup> symporter in wheat when the external Na<sup>+</sup> and K<sup>+</sup> concentrations are balanced; however, when the external Na<sup>+</sup> is in excess, *TaHKT2;1* acts as a Na<sup>+</sup>-selective uniporter (Laurie et al., 2002). *SbHKT1;4* mediates Na<sup>+</sup> transport under high external Na<sup>+</sup> and K<sup>+</sup> deficiency, but it performs K<sup>+</sup> uptake when K<sup>+</sup> levels are sufficient (Wang et al., 2014). Our yeast complementation experiment results showed that type II *SaHKTs* have a preference for K<sup>+</sup> transport under high external Na<sup>+</sup> and K<sup>+</sup> deficiency; and type I *SaHKTs*, except for *SaHKT1;8*, have a preference for Na<sup>+</sup> uptake or efflux under high Na<sup>+</sup> concentrations, suggesting that *SaHKTs* mediate K<sup>+</sup> or Na<sup>+</sup> transport in *S. alterniflora*.

It seems that maintaining high K<sup>+</sup> uptake under Na<sup>+</sup> stress is essential to achieve Na<sup>+</sup>/K<sup>+</sup> ratio balance and salt tolerance in halophytes. For example, *PutHKT2;1*, a type II member in the halophyte plant *Puccinellia tenuiflora*, acts as a high-affinity K<sup>+</sup>-Na<sup>+</sup> symporter with the ability to take up K<sup>+</sup> at low K<sup>+</sup> concentrations and under NaCl stress, thus ensuring that a physiologically suitable Na<sup>+</sup>/K<sup>+</sup> ratio is maintained (Ardie et al., 2009). *Thellungiella salsuginea* *TsHKT1;2* takes up K<sup>+</sup> in the presence of NaCl, which maintains a high cytoplasmic K<sup>+</sup>/Na<sup>+</sup> ratio and therefore contributes to salt tolerance of this halophytic species

(Ali et al., 2012). Our study showed that SaHKT2;1, SaHKT2;3 and SaHKT2;4 displayed a strong effect on the  $K^+$  uptake, as demonstrated by their ability to complement the  $K^+$  deficient phenotypes of yeast *R5421* under low  $K^+$  and high  $Na^+$  stresses (Figure 5a), indicating that they are able to maintain  $Na^+$  and  $K^+$  balance under high salt conditions.

Because of their  $Na^+$  selective channel-line activity, type II HKTs can mediate  $Na^+$  entry across the plasma membrane of the plant root epidermis and cortex under salt stress (Shi et al., 2000). In rice, to prevent massive  $Na^+$  influx, the ion transport activity of OsHKT2;1 in roots is repressed in the presence of high concentrations of  $Na^+$  (Horie et al., 2007). In the wheat *hkt2;1* mutant, root  $Na^+$  uptake is decreased under salt stress, suggesting that *TaHKT2;1* can mediate  $Na^+$  influx into roots under a salt environment (Laurie et al., 2002). In our study, the expressions of most *SaHKT* genes were suppressed by severe salinity, indicating that  $Na^+$  influx via SaHKTs may be reduced in *S. alterniflora*.

As a halophyte that can grow under high salinity conditions, *S. alterniflora* can actively excrete salt from the body. SaHKTs function as membrane transporters that transport salt out of the body under high salinity conditions. Among the SaHKTs, SaHKT2;1, SaHKT2;3 and SaHKT2;4 exhibited the same phenotype as OsHKT2;4 and showed the ability to be salt resistant under  $K^+$  deficiency and high  $Na^+$  (Figure 5a). A previous study showed that overexpression of *SaHKT1* (*SaHKT1;9* in this study) with an alternative 3' UTR in rice protoplasts increased the mRNA accumulation of salt tolerance genes, including *SOS1*, *NHX1*, *COR15A* and *RD22*, in an AU-rich element-dependent manner, which ultimately enhanced the ability of the plant to resist high salt stress (Wang et al., 2023). In addition, regulatory elements in most *HKT* genes in *S. alterniflora* have more ABRE elements than in rice and *Arabidopsis*. Previous studies have shown that the ABRE element is involved in the salt stress response of plants (Liu et al., 2022; Narusaka et al., 2003; Zhang et al., 2022). It is plausible that the increased number of ABREs confers elevated salt tolerance to *S. alterniflora*. Therefore, the analysis allows us to infer the reasons why *S. alterniflora* is able to live under high salt concentrations. One is the unique salt-secreting structure of *S. alterniflora* itself; another one is the ability to improve the salt resistance by altering the expression of *SaHKTs* and other salt tolerance genes and mediating ion transport capacity under salt stress conditions.

## EXPERIMENTAL PROCEDURES

### Plant materials and growth conditions

Seeds of *S. alterniflora* were collected from a saline inland environment (N37°26'; E118°34') in Dongying City, Shandong

Province, China. Seedlings of *S. alterniflora* were watered with Hoagland nutrient solution and grown in an artificial climate under a 16 h light (24°C; 08:00–00:00)/8 h dark (22°C; 00:00–08:00) cycle. All experimental procedures, including planting and seed collection, were in accordance with local and national regulations.

### Identification of the *HKT* genes using a DL method

A DL method, DeepGOPlus (Kulmanov & Hoehndorf, 2020), was used to capture protein molecular features from amino acid sequences with a prediction threshold of 0.3 from 54 plants at the genome-wide level. The target HKTs were annotated to 15 GO categories: GO:0051179, GO:0051234, GO:0006810, GO:0006811, GO:0006812, GO:0050896, GO:0030001, GO:0006813, GO:0006950, GO:0009987, GO:0009987, GO:0009628, GO:0006970, GO:0009651 and GO:0006814. The genes with the highest match to *Arabidopsis* HKT1 were identified as candidate *HKT* genes. The prediction architecture of the DeepGOPlus model is shown in Figure 1a; the model consists of an input layer, a convolutional layer, a pooling layer, and a fully connected layer. For the input data, we assume that the input data of DeepGOPlus is denoted as  $S$  and consists of  $N$  whole protein sequences, each sequence  $s \in S$  is an amino acid region with a length of  $L_{max} = 2000$ . Each amino acid corresponds to a sequence-and-structure vector of  $D = 21$  dimensions. Therefore, we use the one-hot method to encode  $S$  as a tensor  $X \in R^{N \times L \times D}$ . The input data, including genome protein sequences of the 54 investigated species in this study, is passed through a set of CNN layers with different filter sizes, each CNN layer containing 512 filters to learn a particular sequence of a particular size. We use the *ReLU* activation function to activate the intermediate layers in this process. Next, the MaxPooling layer filters out the maximum feature values from the features learned by each filter. The purpose of this process is to force the CNN filter to learn a set of similar motifs, and when the filter finds a motif in the sequence, it returns a higher value which is pooled by the Max-Pooling layer. With the above setup, we obtained an output containing 8192 ( $16 \times 512$ ) convolution filters as sequence features. Finally, a sigmoid activation function combined with a fully connected layer is used for the final prediction and classification. The final classification prediction of DeepGOPlus can be expressed as:

$$Y = \text{DeepGOPlus}(X)$$

The functional expression of the *DeepGOPlus* model can be defined as:

$$\text{DeepGOPlus}(X) = \sigma(\text{FC}(f_{crP}(f_{crC}(X))))$$

where  $\sigma$  denotes a *sigmoid*( $x$ ) =  $\frac{1}{1+e^{-x}}$  function, and *FC* is a fully connected layer;  $f_p$  and  $f_c$  represent the pooling layer and the fully connected layer, respectively, and the function is defined as:

$$f_p(x) = \text{MaxPool}(x)$$

$$f_{crC}(x) = \text{ReLU}(\text{BN}(\text{Conv}_{21,m}(x)))$$

in  $f_p$ ,  $L_{max}$  denotes the maximum length of the input protein sequence and  $m$  (8, 16, 24, ..., 128) indicates the size of the convolution kernel, and  $\text{Pool}_{size} = L_{max} - m + 1$ . In  $f_c$ , *BN* denotes a batch normalization. *ReLU* is a rectifier linear unit activation function, and  $\text{ReLU}(x) = \max(0, x)$ .

Next, the NCBI-conserved domains database (Marchler-Bauer et al., 2015) and SMART database (Letunic & Bork, 2018) were used to confirm the predicted HKT proteins; proteins without a TrkH domain were removed, and the putative HKTs were obtained. The physicochemical properties of the HKT proteins were predicted using the ExPASy web server (Artimo et al., 2012).

The subcellular localization prediction was performed using the Plant-mPloc and Euk-mPloc 2.0, two modules of Cell-Ploc 2.0 webserver (Chou & Shen, 2008), and the CELLO web server (Yu et al., 2006). The protein sequences, coding sequences (CDSs), DNA sequences, and 2-kb upstream DNA sequences of *S. alterniflora* were obtained from our genome sequencing data (data not shown). Protein sequences of the other plant species were downloaded from the Ensembl Plants (Bolser et al., 2017) and Phytosome databases (Goodstein et al., 2012), and the Pteridophyta species were downloaded from Fernbase (<https://www.fernbase.org>). Homology comparison and hidden Markov model were also used to predict the HKTs in the *S. alterniflora* genome. For homology comparison, *Arabidopsis* and rice HKT proteins were used to BLAST against the *S. alterniflora* genome proteins with  $E$ -value <  $e^{-5}$ . For hidden Markov model (HMM) search, two methods were used. One is to use the rice and *Arabidopsis* HKTs to build an HMM and then search against the *S. alterniflora* genome proteins with  $E$ -value <  $e^{-5}$ ; the other one is to use the feature domain of HKT proteins, TrkH (PF02386), from the Pfam database (Mistry et al., 2021) to search against the *S. alterniflora* genome proteins with  $E$ -value <  $e^{-5}$ .

### Gene duplication, synteny and chromosome location analyses

The gene duplicates and synteny between *S. alterniflora* and other plants were analyzed using the MCScanX program (Wang et al., 2012). Gene location information was obtained from the *S. alterniflora* genome and annotation files. Gene location, duplication and synteny relationship were visualized using TBtools software (Chen et al., 2020). The Ka/Ks ratios were used to detect the selection pressure on a given gene after duplication. Ka/Ks ratios of = 1, > 1 and < 1 indicate neutral, positive and purifying selection, respectively (Lynch & Conery, 2000).

### Phylogenetic tree construction and gene components analysis

Alignment of full-length of SaHKT proteins was performed using the Clustal X2 (Larkin et al., 2007) with default parameters. MEGA 7.0 (Kumar et al., 2016) was used to construct unrooted NJ trees with 1000 bootstrap replications. *Cis*-acting elements were predicted using PlantCARE (Lescot et al., 2002) by uploading the 2-kb upstream DNA sequences. Exon-intron structures of SaHKT genes were displayed using the Gene Structure Display Server (GSDS 2.0) (Hu et al., 2015). Conserved domains were predicted using the MEME Suite web server (Bailey et al., 2009) with default parameters. The phylogenetic tree, gene structure and domains were visualized using Evolview v3 (Subramanian et al., 2019).

### RNA extraction and RT-qPCR

Roots, stems, leaves and immature seeds were collected 10 days after grouting, and mature seeds were collected at the seed ripening stage for different tissue analyses. Four-leaf-stage seedlings of *S. alterniflora* were treated with 0, 300, or 600 mM NaCl for 48 h in hydroponic culture, then whole seedlings were collected for total RNA extraction. Total RNA isolation and first-strand cDNA synthesis were performed according to the manufacturer's instructions (TIANGEN, Beijing, China). Next, RT-qPCR analysis was performed in triplicate. The RT-qPCR conditions and the reaction systems were as described by the product instructions (TIANGEN, Beijing, China). Data acquisition was performed using the LightCycler<sup>®</sup> 96 system (Roche, Switzerland). Relative expression levels of each genes were calculated using the  $2^{-\Delta\Delta Ct}$  method (Schmittgen &

Livak, 2008) and normalized to the *S. alterniflora* GAPDH gene. One-way ANOVA was carried out using SPSS 19.0 (SPSS Inc., Chicago, IL, USA). The values in all graphs are reported as means  $\pm$  standard errors, and significant differences are represented by different letters ( $P < 0.05$ ). Each experiment was repeated three times. All primers used in this study are listed in Table S8.

### Subcellular location assay

For the infusion of the eGFP at the C-terminus of SaHKTs, the coding sequences of SaHKTs were amplified by PCR and were inserted between the  $2 \times CaMV35S$  promoters and eGFP of pAN580-eGFP vector (Chen, Wang, et al., 2021). Protoplasts were isolated from the leaf sheath of 7-day-old rice seedlings and then transformed as described in the previous study. Subcellular localization was observed and image analysis was using confocal microscopy (Zeiss LSM880, Germany).

### Vector construction, yeast transformation and K<sup>+</sup> measurement

The CDSs of all SaHKT genes were amplified using the cDNA of *S. alterniflora* seedlings and were cloned into the p416-GDP vector. HKT genes of rice and *Arabidopsis* were cloned into the p416-GDP vector according to previous studies (Horie et al., 2011; Huang et al., 2008; Rubio et al., 2008; Tada, 2019). All the recombinant plasmids were transformed into the *Saccharomyces cerevisiae* strains R5421 (*MAT $\alpha$  ura3-52 leu2 trk1 $\Delta$  his3 $\Delta$ 200 his4-15 trk2 $\Delta$ 1::pCK64*) and AxT3K ( *$\Delta$ ena1::HIS3::ena4,  $\Delta$ nha1::LEU2,  $\Delta$ nhx1::KanMX4*), respectively. Yeast transformation of the recombinant plasmids was performed using LiCl following methods described in a previous study (Wang et al., 2020). Next, positive transformants were selected on a medium containing uracil. Yeast growth experiments were performed using AP medium supplemented with different concentrations of K<sup>+</sup> or combinations of K<sup>+</sup> and Na<sup>+</sup>. Yeast cells were plated on AP medium in 10-fold serial dilutions with OD<sub>600</sub> values ranging from 0.6 to  $0.6 \times 10^{-4}$ . Photographs were taken after incubation for 3–5 days at 30°C.

Yeast cells were expanded using an AP liquid medium containing 0.1 or 100 mM KCl. Cell culture was stopped when the OD<sub>600</sub> was 0.3, and cells were collected by centrifuging at  $5000 \times g$  for 2 min at room temperature. The collected cells were washed twice using ice-cold buffers (10 mM MgCl<sub>2</sub>, 10 mM CaCl<sub>2</sub> and 1 mM HEPES), and then the cells were digested using a Super Microwave Digester (EXPEC 790S). Finally, the K<sup>+</sup> content was quantified using an inductively coupled plasma mass spectrometer (ISUPEC 7000 series). Each experiment was repeated six times.

### ACKNOWLEDGEMENTS

We gratefully acknowledge Professor Weihua Wu (China Agriculture University) for providing the p416-GDP vector. This work was supported by the National Key R&D Program of China (2022YFF0711802), Nanfan special project, CAAS, Grant Nos. ZDXM2309 and YBXM2304, National Natural Science Foundation of China (32022064), Innovation Program of Chinese Academy of Agricultural Sciences, and Alibaba Foundation.

### CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

### AUTHOR CONTRIBUTIONS

HHL, CML and SHC conceived the original idea and designed the experiments. MGY, SKC, ZPH, SG, TXY, TTD

and HZ performed the experiments and data analysis. CML, XL and SHC provided valuable advice for the experimental design and optimization. MGY, SKC, TXY and HHL wrote the manuscript. HHL provided overall supervision and direction of the work. All authors discussed and commented on the manuscript.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** Phylogenetic analysis of the 206 (except OsHKT2;2) HKT protein sequences. The phylogenetic tree was reconstructed by full-length protein sequences of HKT homolog genes aligned by Clustal W, using MEGA 7 Jones–Taylor–Thornton (JTT) as the best-fitting model and 1000 replicates. Red and yellow represent tandem and segmental duplicate, respectively.

**Figure S2.** Protein domains of 206 HKT proteins. The 20 conserved domains were predicted by MEME.

**Figure S3.** Sequence logos of the four pore regions specified for all HKTs. Numbers in parenthesis next to the names of evolutionary groups indicate the number of sequences used for the generation of the logos. For the sequence logos based on all sequences, the percentage of the most frequent amino acid per position is indicated. Numbers above amino acids are percentages. S/G and G indicate the selectivity filter residues. A motif switch in P<sub>A</sub>, including the selectivity filter residue, is shaded, as well as the selectivity filter residues in P<sub>B</sub>, P<sub>C</sub> and P<sub>D</sub>.

**Figure S4.** Membrane protein families involved in ion homeostasis identified by DeepGOPlus. A total of five membrane protein families were identified in *Arabidopsis*, rice, maize and *S. alterniflora*. The membrane proteins of each family in *S. alterniflora* are larger than those in *Arabidopsis*, rice and maize.

**Figure S5.** Subcellular localization assays of SaHKT1;4 and SaHKT2;4 transiently expressed in rice protoplasts. Bars = 10  $\mu$ m.

**Figure S6.** Multiple sequences alignment of HKT proteins using Clustal X2. Amino acid sequences in the pore loop region (P<sub>A</sub>, P<sub>B</sub>, P<sub>C</sub> and P<sub>D</sub>) and the adjacent transmembrane domain (M1A, M2B, M3C and M4D) were aligned by Clustal X2. The rice and *Arabidopsis* HKTs protein sequences are included for comparison. The conserved residues are indicated in red and highlighted with boxes.

**Figure S7.** Comparative analysis of the number of *cis*-elements in the promoter regions of *Arabidopsis*, rice and *S. alterniflora* HKT genes.

**Figure S8.** Duplication HKT gene pairs between the *S. alterniflora* gene and (a) rice, (b) barley, (c) sorghum, (d) foxtail millet and (e) maize. Grey lines in the background indicate the collinear blocks between the *S. alterniflora* and rice, barley, sorghum, foxtail millet and maize genomes, while the red lines highlight the duplication gene pairs.

**Figure S9.** Expression patterns of SaHKT genes. (a) Relative expression levels of SaHKT genes in different tissues. (b) Expression patterns of SaHKT genes under different NaCl concentrations. The horizontal coordinates represent the different tissues or different concentrations of NaCl, and the vertical coordinates represent the relative expression levels. Values are means  $\pm$  SD ( $n = 3$  biological replicates). Different letters represent significant differences at  $P < 0.05$  (Duncan's multiple range test).

**Table S1.** Identification of membrane proteins in *S. alterniflora*.

**Table S2.** Characteristic features of the HKT gene family members identified in *S. alterniflora*.

**Table S3.** Identification of SaHKT genes using hidden Markov model and homology comparison methods.

**Table S4.** Summary of functional domains present in the SaHKT proteins. A TrkH (PF02386) domain was presented in all SaHKTs.

**Table S5.** Detailed information of domains identified in SaHKT proteins.

**Table S6.** Ka/Ks ratios for segmentally duplicated SaHKT gene pairs.

**Table S7.** Ka/Ks ratios for orthologous HKT genes between *S. alterniflora* and other plant species.

**Table S8.** Primers were used for RT-qPCR and yeast expression vector construction in this study. Primers used to amplify the coding sequences of genes were designed using Primer Premier 5.

## REFERENCES

- Ali, Z., Park, H.C., Ali, A., Oh, D.H., Aman, R., Kropornicka, A. et al. (2012) TsHKT1;2, a HKT1 homolog from the extremophile *Arabidopsis* relative *Thellungiella salsuginea*, shows K<sup>+</sup> specificity in the presence of NaCl. *Plant Physiology*, **158**, 1463–1474.
- Anil Kumar, S., Hima Kumari, P., Nagaraju, M., Sudhakar Reddy, P., Durga Dheeraj, T., Mack, A. et al. (2022) Genome-wide identification and multiple abiotic stress transcript profiling of potassium transport gene homologs in *Sorghum bicolor*. *Frontiers in Plant Science*, **13**, 965530.
- Ardie, S.W., Xie, L., Takahashi, R., Liu, S. & Takano, T. (2009) Cloning of a high-affinity K<sup>+</sup> transporter gene *PutHKT2;1* from *Puccinellia tenuiflora* and its functional comparison with *OsHKT2;1* from rice in yeast and *Arabidopsis*. *Journal of Experimental Botany*, **60**, 3491–3502.
- Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E. et al. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, **40**, W597–W603.
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L. et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, **37**, W202–W208.
- Baisakh, N., RamanaRao, M.V., Rajasekaran, K., Subudhi, P., Janda, J., Galbraith, D. et al. (2012) Enhanced salt stress tolerance of rice plants expressing a vacuolar H<sup>+</sup>-ATPase subunit c1 (*SaVHAc1*) gene from the halophyte grass *Spartina alterniflora* L. *Plant Biotechnology Journal*, **10**, 453–464.
- Biradar, H., Karan, R. & Subudhi, P.K. (2018) Transgene pyramiding of salt responsive protein 3-1 (*SaSRP3-1*) and *SaVHAc1* from *Spartina alterniflora* L. enhances salt tolerance in rice. *Frontiers in Plant Science*, **9**, 1304.
- Bolser, D.M., Staines, D.M., Perry, E. & Kersey, P.J. (2017) Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomic data. *Methods in Molecular Biology (Clifton, N.J.)*, **1533**, 1–31.
- Byrt, C.S., Xu, B., Krishnan, M., Lightfoot, D.J., Athman, A., Jacobs, A.K. et al. (2014) The Na<sup>+</sup> transporter, TaHKT1;5-D, limits shoot Na<sup>+</sup> accumulation in bread wheat. *The Plant Journal*, **80**, 516–526.
- Cao, Y., Liang, X., Yin, P., Zhang, M. & Jiang, C. (2019) A domestication-associated reduction in K<sup>+</sup>-preferring HKT transporter activity underlies maize shoot K<sup>+</sup> accumulation and salt tolerance. *The New Phytologist*, **222**, 301–317.
- Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y. et al. (2020) TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, **13**, 1194–1202.
- Chen, S., Ren, H., Luo, Y., Feng, C. & Li, H. (2021) Genome-wide identification of wheat (*Triticum aestivum* L.) expansin genes and functional characterization of TaEXPB1A. *Environmental and Experimental Botany*, **182**, 104307.
- Chen, W., Wang, W., Lyu, Y., Wu, Y., Huang, P., Hu, S. et al. (2021) OsVP1 activates *Sdr4* expression to control rice seed dormancy via the ABA signaling pathway. *The Crop Journal*, **9**, 68–78.
- Chou, K.C. & Shen, H.B. (2008) Cell-PLoc: a package of web servers for predicting subcellular localization of proteins in various organisms. *Nature Protocols*, **3**, 153–162.
- Ding, M., Chen, J., Jiang, Y., Lin, L., Cao, Y., Wang, M. et al. (2015) Genome-wide investigation and transcriptome analysis of the WRKY gene family in *Gossypium*. *Molecular Genetics and Genomics*, **290**, 151–171.
- Dou, L., Zhang, X., Pang, C., Song, M., Wei, H., Fan, S. et al. (2014) Genome-wide analysis of the WRKY gene family in cotton. *Molecular Genetics and Genomics*, **289**, 1103–1121.

- Garcia-deblás, B., Senn, M.E., Bañuelos, M.A. & Rodríguez-Navarro, A. (2003) Sodium transport and HKT transporters: the rice model. *The Plant Journal*, **34**, 788–801.
- Ghorbani, R., Zakipour, Z., Alemzadeh, A. & Razi, H. (2020) Genome-wide analysis of AP2/ERF transcription factors family in *Brassica napus*. *Physiology and Molecular Biology of Plants*, **26**, 1463–1476.
- Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J. *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, **40**, D1178–D1186.
- Gu, Z., Nicolae, D., Lu, H.H. & Li, W.H. (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends in Genetics*, **18**, 609–613.
- Hameed, A., Ahmed, M.Z., Hussain, T., Aziz, I., Ahmad, N., Gul, B. *et al.* (2021) Effects of salinity stress on chloroplast structure and function. *Cell*, **10**, 2023.
- Han, Y., Yin, S., Huang, L., Wu, X., Zeng, J., Liu, X. *et al.* (2018) A sodium transporter HvHKT1;1 confers salt tolerance in barley via regulating tissue and cell ion homeostasis. *Plant & Cell Physiology*, **59**, 1976–1989.
- Han, Z., Liu, Y., Deng, X., Liu, D., Liu, Y., Hu, Y. *et al.* (2019) Genome-wide identification and expression analysis of expansin gene family in common wheat (*Triticum aestivum* L.). *BMC Genomics*, **20**, 101.
- Hauser, F. & Horie, T. (2010) A conserved primary salt tolerance mechanism mediated by HKT transporters: a mechanism for sodium exclusion and maintenance of high K<sup>+</sup>/Na<sup>+</sup> ratio in leaves during salinity stress. *Plant, Cell & Environment*, **33**, 552–565.
- Horie, T., Brodsky, D.E., Costa, A., Kaneko, T., Lo Schiavo, F., Katsuhara, M. *et al.* (2011) K<sup>+</sup> transport by the OshKT2;4 transporter from rice with atypical Na<sup>+</sup> transport properties and competition in permeation of K<sup>+</sup> over Mg<sup>2+</sup> and Ca<sup>2+</sup> ions. *Plant Physiology*, **156**, 1493–1507.
- Horie, T., Costa, A., Kim, T.H., Han, M.J., Horie, R., Leung, H.Y. *et al.* (2007) Rice OshKT2;1 transporter mediates large Na<sup>+</sup> influx component into K<sup>+</sup>-starved roots for growth. *The EMBO Journal*, **26**, 3003–3014.
- Horie, T., Hauser, F. & Schroeder, J.I. (2009) HKT transporter-mediated salinity resistance mechanisms in *Arabidopsis* and monocot crop plants. *Trends in Plant Science*, **14**, 660–668.
- Horie, T., Yoshida, K., Nakayama, H., Yamada, K., Oiki, S. & Shinmyo, A. (2001) Two types of HKT transporters with different properties of Na<sup>+</sup> and K<sup>+</sup> transport in *Oryza sativa*. *The Plant Journal*, **27**, 129–138.
- Hou, J., Adhikari, B. & Cheng, J. (2018) DeepSF: deep convolutional neural network for mapping protein sequences to folds. *Bioinformatics (Oxford, England)*, **34**, 1295–1303.
- Hu, B., Jin, J., Guo, A.Y., Zhang, H., Luo, J. & Gao, G. (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics (Oxford, England)*, **31**, 1296–1297.
- Huang, L., Kuang, L., Wu, L., Shen, Q., Han, Y., Jiang, L. *et al.* (2020) The HKT transporter HvHKT1;5 negatively regulates salt tolerance. *Plant Physiology*, **182**, 584–596.
- Huang, S., Spielmeier, W., Lagudah, E.S. & Munns, R. (2008) Comparative mapping of HKT genes in wheat, barley, and rice, key determinants of Na<sup>+</sup> transport, and salt tolerance. *Journal of Experimental Botany*, **59**, 927–937.
- Imran, S., Oyama, M., Horie, R., Kobayashi, N.I., Costa, A., Kumano, R. *et al.* (2022) Distinct functions of the atypical terminal hydrophilic domain of the HKT transporter in the liverwort *Marchantia polymorpha*. *Plant & Cell Physiology*, **63**, 802–816.
- Jabnoun, M., Espeout, S., Mieulet, D., Fizames, C., Verdeil, J.L., Conéjéro, G. *et al.* (2009) Diversity in expression patterns and functional properties in the rice HKT transporter family. *Plant Physiology*, **150**, 1955–1971.
- Kato, Y., Sakaguchi, M., Mori, Y., Saito, K., Nakamura, T., Bakker, E.P. *et al.* (2001) Evidence in support of a four transmembrane-pore-transmembrane topology model for the *Arabidopsis thaliana* Na<sup>+</sup>/K<sup>+</sup> translocating ATHKT1 protein, a member of the superfamily of K<sup>+</sup> transporters. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 6488–6493.
- Kuang, L., Shen, Q., Chen, L., Ye, L., Yan, T., Chen, Z.H. *et al.* (2022) The genome and gene editing system of sea barleygrass provide a novel platform for cereal domestication and stress tolerance studies. *Plant Communications*, **3**, 100333.
- Kulmanov, M. & Hoehndorf, R. (2020) DeepGOPlus: improved protein function prediction from sequence. *Bioinformatics (Oxford, England)*, **36**, 422–429.
- Kumar, S., Stecher, G. & Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**, 1870–1874.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, **23**, 2947–2948.
- Laurie, S., Feeney, K.A., Maathuis, F.J., Heard, P.J., Brown, S.J. & Leigh, R.A. (2002) A role for HKT1 in sodium uptake by wheat roots. *The Plant Journal*, **32**, 139–149.
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y. *et al.* (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, **30**, 325–327.
- Letunic, I. & Bork, P. (2018) 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*, **46**, D493–D496.
- Li, H., Xu, G., Yang, C., Yang, L. & Liang, Z. (2019) Genome-wide identification and expression analysis of HKT transcription factor under salt stress in nine plant species. *Ecotoxicology and Environmental Safety*, **171**, 435–442.
- Li, W., Wen, J., Song, Y., Yuan, H., Sun, B., Wang, R. *et al.* (2022) SaRCC1, a regulator of chromosome condensation 1 (RCC1) family protein gene from *Spartina alterniflora*, negatively regulates salinity stress tolerance in transgenic *Arabidopsis*. *International Journal of Molecular Sciences*, **23**, 8172.
- Liu, H., Cui, P., Zhang, B., Zhu, J., Liu, C. & Li, Q. (2022) Binding of the transcription factor MYC2-like to the ABRE of the *OscYP2* promoter enhances salt tolerance in *Oryza sativa*. *PLoS One*, **17**, e0276075.
- Liu, Q., Chen, J., Wang, Y., Li, S., Jia, C., Song, J. *et al.* (2021) DeepTorrent: a deep learning-based approach for predicting DNA N4-methylcytosine sites. *Briefings in Bioinformatics*, **22**, bbaa124.
- Lynch, M. & Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science (New York, N.Y.)*, **290**, 1151–1155.
- Magadum, S., Banerjee, U., Murugan, P., Gangapur, D. & Ravikesavan, R. (2013) Gene duplication as a major force in evolution. *Journal of Genetics*, **92**, 155–161.
- Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y. *et al.* (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Research*, **43**, D222–D226.
- Mäser, P., Hosoo, Y., Goshima, S., Horie, T., Eckelman, B., Yamada, K. *et al.* (2002) Glycine residues in potassium channel-like selectivity filters determine potassium selectivity in four-loop-per-subunit HKT transporters from plants. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 6428–6433.
- Mian, A., Oomen, R.J., Isayenkov, S., Sentenac, H., Maathuis, F.J. & Véry, A.A. (2011) Over-expression of an Na<sup>+</sup>- and K<sup>+</sup>-permeable HKT transporter in barley improves salt tolerance. *The Plant Journal*, **68**, 468–479.
- Mishra, A. & Tanna, B. (2017) Halophytes: potential resources for salt stress tolerance genes and promoters. *Frontiers in Plant Science*, **8**, 829.
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L. *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Research*, **49**, D412–D419.
- Narusaka, Y., Nakashima, K., Shinwari, Z.K., Sakuma, Y., Furihata, T., Abe, H. *et al.* (2003) Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis rd29A* gene in response to dehydration and high-salinity stresses. *The Plant Journal*, **34**, 137–148.
- Owji, H., Hajiebrahimi, A., Seradj, H. & Hemmati, S. (2017) Identification and functional prediction of stress responsive AP2/ERF transcription factors in *Brassica napus* by genome-wide analysis. *Computational Biology and Chemistry*, **71**, 32–56.
- Platten, J.D., Cotsaftis, O., Berthomieu, P., Bohnert, H., Davenport, R.J., Fairbairn, D.J. *et al.* (2006) Nomenclature for HKT transporters, key determinants of plant salinity tolerance. *Trends in Plant Science*, **11**, 372–374.
- Riedelsberger, J., Miller, J.K., Valdebenito-Maturana, B., Piñeros, M.A., González, W. & Dreyer, I. (2021) Plant HKT channels: An updated view on structure, function and gene regulation. *International Journal of Molecular Sciences*, **22**, 1892.
- Rubio, F., Nieves-Cordones, M., Alemán, F. & Martínez, V. (2008) Relative contribution of AtHAK5 and AtAKT1 to K<sup>+</sup> uptake in the high-affinity range of concentrations. *Physiologia Plantarum*, **134**, 598–608.

- Rus, A., Lee, B.H., Muñoz-Mayor, A., Sharkhuu, A., Miura, K., Zhu, J.K. *et al.* (2004) AtHKT1 facilitates Na<sup>+</sup> homeostasis and K<sup>+</sup> nutrition in planta. *Plant Physiology*, **136**, 2500–2511.
- Sandarawan, P.D. & Wannige, C.T. (2021) An improved deep learning model for hierarchical classification of protein families. *PLoS One*, **16**, e0258625.
- Sapoval, N., Aghazadeh, A., Nute, M.G., Antunes, D.A., Balaji, A., Baraniuk, R. *et al.* (2022) Current progress and open challenges for applying deep learning across the biosciences. *Nature Communications*, **13**, 1728.
- Schmittgen, T.D. & Livak, K.J. (2008) Analyzing real-time PCR data by the comparative C<sub>T</sub> method. *Nature Protocols*, **3**, 1101–1108.
- Seo, S., Oh, M., Park, Y. & Kim, S. (2018) DeepFam: deep learning based alignment-free method for protein family modeling and prediction. *Bioinformatics (Oxford, England)*, **34**, i254–i262.
- Shi, H., Ishitani, M., Kim, C. & Zhu, J.K. (2000) The *Arabidopsis thaliana* salt tolerance gene *SOS1* encodes a putative Na<sup>+</sup>/H<sup>+</sup> antiporter. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 6896–6901.
- Silveira, J.A.G. & Carvalho, F.E.L. (2016) Proteomics, photosynthesis and salt resistance in crops: An integrative view. *Journal of Proteomics*, **143**, 24–35.
- Subramanian, B., Gao, S., Lercher, M.J., Hu, S. & Chen, W.H. (2019) Evolve v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Research*, **47**, W270–W275.
- Sun, L., Xu, K., Huang, W., Yang, Y.T., Li, P., Tang, L. *et al.* (2021) Predicting dynamic cellular protein-RNA interactions by deep learning using in vivo RNA structures. *Cell Research*, **31**, 495–516.
- Tada, Y. (2019) The HKT transporter gene from *Arabidopsis*, *AtHKT1;1*, is dominantly expressed in shoot vascular tissue and root tips and is mild salt stress-responsive. *Plants (Basel, Switzerland)*, **8**, 204.
- Tholema, N., Vor der Brüggen, M., Mäser, P., Nakamura, T., Schroeder, J.I., Kobayashi, H. *et al.* (2005) All four putative selectivity filter glycine residues in KtrB are essential for high affinity and selective K<sup>+</sup> uptake by the KtrAB system from *Vibrio alginolyticus*. *The Journal of Biological Chemistry*, **280**, 41146–41154.
- Tsaban, T., Varga, J.K., Avraham, O., Ben-Aharon, Z., Khrumushin, A. & Schueler-Furman, O. (2022) Harnessing protein folding neural networks for peptide-protein docking. *Nature Communications*, **13**, 176.
- Uozumi, N., Kim, E.J., Rubio, F., Yamaguchi, T., Muto, S., Tsuboi, A. *et al.* (2000) The *Arabidopsis HKT1* gene homolog mediates inward Na<sup>+</sup> currents in xenopus laevis oocytes and Na<sup>+</sup> uptake in *Saccharomyces cerevisiae*. *Plant Physiology*, **122**, 1249–1259.
- Véry, A.A., Nieves-Cordones, M., Daly, M., Khan, I., Fizames, C. & Sentenac, H. (2014) Molecular biology of K<sup>+</sup> transport across the plant cell membrane: what do we learn from comparison between plant species? *Journal of Plant Physiology*, **171**, 748–769.
- Wang, T., Ye, W., Zhang, J., Li, H., Zeng, W., Zhu, S. *et al.* (2023) Alternative 3'-untranslated regions regulate high salt tolerance of *Spartina alterniflora*. *Plant Physiology*, **23**, kiad030.
- Wang, T.T., Ren, Z.J., Liu, Z.Q., Feng, X., Guo, R.Q., Li, B.G. *et al.* (2014) *SbHKT1;4*, a member of the high-affinity potassium transporter gene family from *Sorghum bicolor*, functions to maintain optimal Na<sup>+</sup>/K<sup>+</sup> balance under Na<sup>+</sup> stress. *Journal of Integrative Plant Biology*, **56**, 315–332.
- Wang, W.Y., Liu, Y.Q., Duan, H.R., Yin, X.X., Cui, Y.N., Chai, W.W. *et al.* (2020) SsHKT1;1 is coordinated with SsSOS1 and SsNHX1 to regulate Na<sup>+</sup> homeostasis in *Suaeda salsa* under saline conditions. *Plant and Soil*, **449**, 117–131.
- Wang, Y., Chen, Y.F. & Wu, W.H. (2021) Potassium and phosphorus transport and signaling in plants. *Journal of Integrative Plant Biology*, **63**, 34–52.
- Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X. *et al.* (2012) MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**, e49.
- Wei, H., Wang, X., He, Y., Xu, H. & Wang, L. (2021) Clock component OsPRR73 positively regulates rice salt tolerance by modulating *OsHKT2;1*-mediated sodium homeostasis. *The EMBO Journal*, **40**, e105086.
- Wu, Y., Henderson, S.W., Wege, S., Zheng, F., Walker, A.R., Walker, R.R. *et al.* (2020) The grapevine NaE sodium exclusion locus encodes sodium transporters with diverse transport properties and localisation. *Journal of Plant Physiology*, **246–247**, 153113.
- Xu, H., Jia, P. & Zhao, Z. (2021) Deep4mC: systematic assessment and computational prediction for DNA N4-methylcytosine sites by deep learning. *Briefings in Bioinformatics*, **22**, bbaa099.
- Yan, W., Li, Z., Pian, C. & Wu, Y. (2022) PlantBind: an attention-based multi-label neural network for predicting plant transcription factor binding sites. *Briefings in Bioinformatics*, **23**, bbac425.
- Yu, C.S., Chen, Y.C., Lu, C.H. & Hwang, J.K. (2006) Prediction of protein subcellular localization. *Proteins*, **64**, 643–651.
- Zeng, R. & Liao, M. (2020) Developing a multi-layer deep learning based predictive model to identify DNA N4-methylcytosine modifications. *Frontiers in Bioengineering and Biotechnology*, **8**, 274.
- Zhang, H., Mao, L., Xin, M., Xing, H., Zhang, Y., Wu, J. *et al.* (2022) Overexpression of *GhABF3* increases cotton (*Gossypium hirsutum* L.) tolerance to salt and drought. *BMC Plant Biology*, **22**, 313.
- Zhang, J.F., Xu, Y.Q., Dong, J.M., Peng, L.N., Feng, X., Wang, X. *et al.* (2018) Genome-wide identification of wheat (*Triticum aestivum*) expansins and expansin expression analysis in cold-tolerant and cold-sensitive wheat cultivars. *PLoS One*, **13**, e0195138.
- Zhang, Y., Wang, Z., Zeng, Y., Zhou, J. & Zou, Q. (2021) High-resolution transcription factor binding sites prediction improved performance and interpretability by deep learning method. *Briefings in Bioinformatics*, **22**, bbac425.