



Molecular evolutionary analysis of the high-affinity K^+ transporter gene family in angiosperms

P. Yang¹, C. Hua¹, F. Zhou¹, B.-J. Zhang¹, X.-N. Cai¹, Q.-Z. Chen¹ and R.-L. Wang²

¹School of Biochemical and Environmental Engineering, Nanjing Xiaozhuang University, Nanjing, China

²Jiangsu Second Normal University, Nanjing, China

Corresponding authors: Q.-Z. Chen / R.-L. Wang
E-mail: qzchen65@hotmail.com / wr135001988@163.com

Genet. Mol. Res. 15 (3): gmr.15038781

Received May 16, 2016

Accepted June 3, 2016

Published July 15, 2016

DOI <http://dx.doi.org/10.4238/gmr.15038781>

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. The high-affinity K^+ transporter (HKT) family comprises a group of multifunctional cation transporters widely distributed in organisms ranging from Bacteria to Eukarya. In angiosperms, the HKT family consists primarily of nine types, whose evolutionary relationships are not fully understood. The available sequences from 31 plant species were used to perform a comprehensive evolutionary analysis, including an examination of selection pressure and estimating phylogenetic tree and gene duplication events. Our results show that a gene duplication in the *HKT1;5/HKT1;4* cluster might have led to the divergence of the *HKT1;5* and *HKT1;4* subfamilies. Additionally, maximum likelihood analysis revealed that the HKT family has undergone a strong purifying selection. An analysis of the amino acids provided strong statistical evidence for a functional divergence between

subfamilies 1 and 2. Our study was the first to provide evidence of this functional divergence between these two subfamilies. Analysis of co-evolution in HKT identified 25 co-evolved groups. These findings expanded our understanding of the evolutionary mechanisms driving functional diversification of HKT proteins.

Key words: High-affinity K⁺ transporters; Gene duplication; Selection pressure; Functional divergence; Co-evolution

INTRODUCTION

Potassium (K⁺) is an essential nutrient for plants. In contrast, high cytosolic sodium (Na⁺) concentrations are harmful to most plants, and can result in severe agricultural losses. Osmotic stress and ion toxicity are two salinity stress factors that lead to harmful effects on plant growth and yield (Hauser and Horie, 2010). To address this challenge, a better understanding of the mechanisms underlying salt tolerance was required (Schubert and Lauchli, 1990). The process of sodium ion transport plays an important role in salinity tolerance; an increase in cytoplasmic K⁺ levels, or a reduction in Na⁺ uptake relative to that of K⁺, leads to increased Na⁺ tolerance in plants (Schubert and Lauchli, 1990).

High-affinity K⁺ transporters (HKT), crucial components in the regulation of Na⁺ homeostasis, are responsible for a primary mechanism mediating salt tolerance in plants (Munns and Tester, 2008; Horie et al., 2009). The HKT gene family is one of five families of plant-transport systems (Corratgé-Faillie et al., 2010). Recently published data suggests that the plant HKT family members include both Na⁺-selective uniporters and Na⁺-K⁺ co-transporters (Munns and Tester, 2008). The wheat *TaHKT2;1* (formerly known as *TaHKT1*), was the first cloned HKT transporter gene in plants (Schachtman and Schroeder, 1994). The product of this gene functions as a Na⁺-K⁺ co-transporter, and is preferred for Na⁺-selective low-affinity Na⁺ transport in *Xenopus laevis* oocytes and yeast expression systems (Rubio et al., 1995). In contrast, *AtHKT1;1* (formerly named *AtHKT1*) encodes a relatively specific Na⁺-selective transporter in *Arabidopsis thaliana* (Uozumi et al., 2000).

HKT genes are classified into two distinct groups: Subfamily 1 and Subfamily 2, based on the phylogenetic analysis of HKT amino acid sequences, (Amtmann, 2009). All proteins encoded by Subfamily 1 members contain a serine in the first p-loop of the protein domain structure. Proteins encoded by members of Subfamily 2, in contrast, have glycine in place of serine at this site, and function as co-transporters of Na⁺ and K⁺ (Rubio et al., 1995). Monocotyledonous plants possess a larger number of *HKT* genes than do dicotyledonous species. For example, the dicot *A. thaliana* contains a single *HKT* gene, whereas the monocot rice possesses nine (Uozumi et al., 2000). We used the rapid amplification of cDNA ends (RACE) method in this study to clone the full-length cDNA sequence of the *Theilungiella halophila* *HKT* gene. *T. halophila* was a close relative of *A. thaliana*, but, unlike *A. thaliana*, it grew well under extreme cold, salt, drought, and limited nitrogen conditions (Amtmann, 2009).

The evolutionary determinants of the functional differences among HKT gene families are poorly understood. Our study's aim was to uncover evolutionary differences among HKTs by integrating comprehensive approaches, including structural analyses and comparative sequence analysis, to estimate the functional mutation at divergent residues. Additionally, we elucidated relationships among functional constraints, purifying selec-

tion, and gene duplications. The results of our bioinformatics analyses provide a solid foundation for the evaluation of evolutionary relationships among HKT gene families in angiosperms.

MATERIAL AND METHODS

Thellungiella halophila HKT gene (ThHKT1) isolation

The PMD19-T Vector was acquired from Promega (Madison, WI, USA). The RACE kit was purchased from Takara (Dalian, China). The RNAisoPlus and a MiniBEST plasmid purification kit were purchased from Invitrogen (Carlsbad, CA, USA).

To clone the *T. halophila* HKT partial coding sequence (CDS), we designed primers 1U (5'-ATCTGGCTCCTAATCCCTCAAGTA-3') and 1D (5'-CGTCAACGGCATAAATAAA GTG-3') based on the *A. thaliana* HKT gene (*AtHKT1*: AT4G10310) sequence in The Arabidopsis Information Resource (TAIR: <http://www.arabidopsis.org/>). To clone the 5' and 3' untranslated regions (UTRs) of the *ThHKT1* gene, we designed the following rapid amplification of cDNA ends (RACE) primers: 5' RACE GSP inner primer (5'-CAAGCCCTAACACCG TCAAAGC-3'), 5' RACE GSP outer primer (5'-GATTTGTGATCTTATGAAGTCCC-3'), 3' RACE GSP inner primer (5'-TTCTTGCTTTGACGGTGTAGGGC-3'), and 3' RACE GSP outer primer (5'-TCCTACGAGAAGTTGGTTGG-3'). All primers were designed using PRIMER 6.0 (<http://www.PremierBiosoft.com>) and were synthesized by Sangon Biotech, Shanghai, China. After generating the partial CDS, the *ThHKT1* full-length mRNA sequence was obtained by splicing overlapping areas of cloned fragments together. The full-length sequence was then submitted to GenBank (<http://www.ncbi.nlm.nih.gov>) (EF025500).

Primers (5'-GTATTCCACAATCTCCTCGG-3' and 5'-GGGCCACTAGTTTGTTTTT C-3'), based on the *ThHKT1* mRNA sequence, were designed to clone the *ThHKT1* complete CDS. PCR amplifications were performed in 20 μ L total reaction volumes containing 2 μ L cDNA, 1 μ L each forward and reverse primers, 2 μ L 10X PCR buffer, 1.2 μ L MgCl₂ (25 mM), 2 μ L dNTP mixture (2 mM), and 0.2 μ L *Taq* DNA polymerase. PCR conditions were as follows: initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 55°C for 50 s and 72°C for 30 s, and a final extension of 72°C for 10 min. Amplification products were detected by electrophoresis on 1% agarose gel.

Sequence collection

To identify HKT genes in angiosperms, BLAST, BLASTP, TBLASTN, and PSI-BLAST (Altschul et al., 1997) searches were used to against the non-redundant sequence databases at NCBI, TAIR, UniProt (<http://www.uniprot.org/>), and Plant GDB (<http://www.plantgdb.org/>), using the published *A. thaliana* and *Oryza sativa* HKT (Uozumiet al., 2000; Horie et al., 2001) sequences as queries. HKT homologs were identified as those matches over a threshold value of 30% protein sequence similarity, with expectation values $<10^{-10}$, and containing consecutively ordered TrkH protein domains (Pfam02386) (Calero et al., 2000). Moreover, only HKT or HKT-like genes from plants with complete start and stop codons, and sequences longer than 470 amino acids, were included in our analyses. A total of 74 sequences from 31 species were used in the final analyses ([Table S1](#)).

Sequence alignment and phylogenetic analysis

Multiple amino acid sequences were aligned using ClustalX v.1.83 (Thompson et al., 1997) and the corresponding multiple nucleotide alignment was obtained with PAL2NAL (Suyama et al., 2006).

Best-fit amino acid and nucleotide substitution models were selected based on the Bayesian information criterion using ProtTest v.1.4 (Abascal et al., 2005) and jModelTest (Posada, 2008), respectively. Phylogenetic trees were then obtained using Bayesian inference under these models in MrBayesv.3.2 (Ronquist and Huelsenbeck, 2003). Neighbor-joining (NJ) trees were generated by MEGA v.6.0 (Tamura et al., 2013), with a 95% partial deletion of gaps and missing data under the P-distance model, and with statistical support for the resulting tree topology assessed using 1,000 bootstrap replications (Figure S1). Clade A contained HKT homologs from the lycophyte *Selaginella moellendorffii* (XP_002972315, XP_002963006, XP_002980327, and XP_002976863) and the bryophyte *Physcomitrella patens* (XP_001763760 and CAZ65436), which were used as outgroup sequences to root the tree (Table 1 and Figure 1). FigTree v.1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>) was used to display phylogenetic trees.

Table 1. Species and number of high-affinity K⁺ transporter (HKT) sequences.

	Species	N	
Moss	<i>Physcomitrella patens</i>	2	
Spikemoss	<i>Selaginella moellendorffii</i>	4	
Angiosperms (Monocots)	<i>Aegilops tauschii</i>	2	
	<i>Brachypodium distachyon</i>	5	
	<i>Hordeum brevisubulatum</i>	1	
	<i>Hordeum vulgare</i> subsp <i>spontaneum</i>	3	
	<i>Oryza glaberrima</i>	1	
	<i>Oryza rufipogon</i>	2	
	<i>Oryza sativa</i>	15	
	<i>Phragmites australis</i>	2	
	<i>Puccinellia tenuiflora</i>	1	
	<i>Sorghum bicolor</i>	3	
	<i>Triticum aestivum</i>	5	
	<i>Triticum durum</i>	1	
	<i>Triticum monococcum</i>	1	
	<i>Triticum urartu</i>	1	
	Angiosperms (Dicots)	<i>Arabidopsis lyrata</i> subsp <i>Lyrata</i>	1
		<i>Arabidopsis thaliana</i>	1
<i>Cochlearia anglica</i>		1	
<i>Eucalyptus camaldulensis</i>		2	
<i>Eutrema salsugineum</i>		1	
<i>Glycine max</i>		4	
<i>Medicago truncatula</i>		1	
<i>Mesembryanthemum crystallinum</i>		2	
<i>Populus trichocarpa</i>		1	
<i>Salicornia bigelovii</i>		1	
<i>Solanum cheesmanii</i>		2	
<i>Solanum lycopersicum</i> var. <i>cerasiforme</i>		1	
<i>Suaeda salsa</i>		1	
<i>Vitis vinifera</i>		5	
<i>Thellungiella halophila</i>	1 ^a		
Total	31	74	

^aSequences that were experimentally determined in the present study.

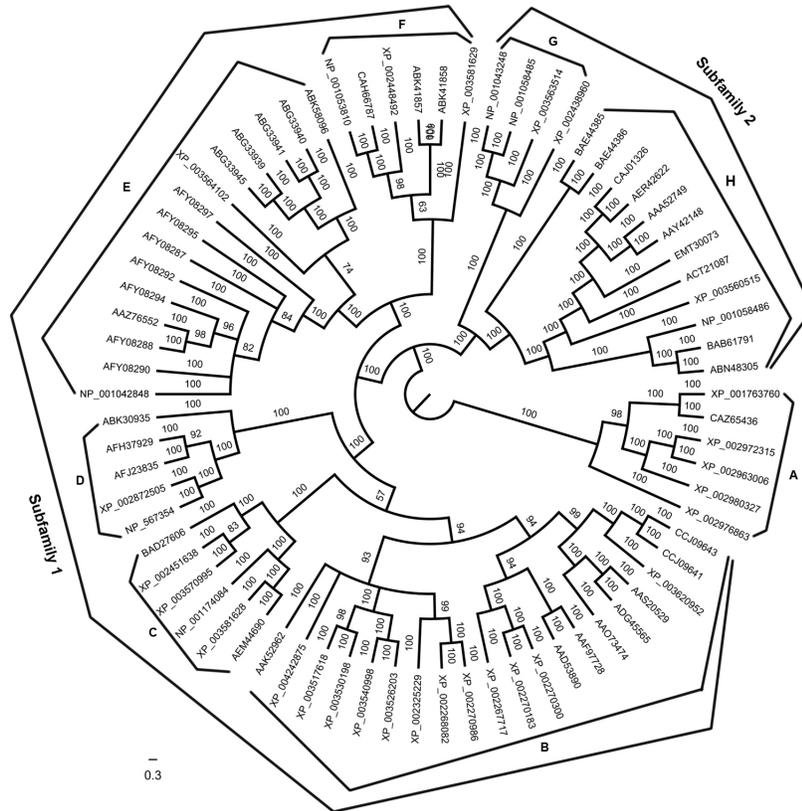


Figure 1. Phylogenetic tree derived by Bayesian inference from plant high-affinity K^+ transporter (HKT) nucleotide sequences. **A.** Bryophyta (XP_001763760, CAZ65436) and ferns (XP_002972315, XP_002963006, XP_002980327 and XP_002976863). **B.** HKT from dicots. **C.** *HKT1;1* (AEM44690, XP_003581628 and NP_001174084) and *HKT1;3* (XP_003570995, XP_002451638 and BAD27606) from monocots. **D.** HKT1 from cruciferous dicots. **E.** *HKT1;5* from monocots. **F.** *HKT1;4* from monocots. **G.** *HKT2;3* (NP_001043248 and XP_002438960) and *HKT2;4* (NP_001058485 and XP_003563514) from monocots. **H.** HKT from monocots. Aet = *Aegilops tauschii*; At = *Arabidopsis thaliana*; Al = *Arabidopsis lyrata*; Bd = *Brachypodium distachyon*; Ca = *Cochlearia anglica*; Ec = *Eucalyptus camaldulensis*; Th = *Thellungiella halophila*; Es = *Eutrema salsugineum*; Gm = *Glycine max*; Hb = *Hordeum brevisubulatum*; Hv = *Hordeum vulgare*; Mt = *Medicago truncatula*; Mc = *Mesembryanthemum crystallinum*; Og = *Oryza glaberrima*; Or = *Oryza rufipogon*; Os = *Oryza sativa*; Pp = *Physcomitrella patens*; Pha = *Phragmites australis*; Put = *Puccinellia tenuiflora*; Pt = *Populus trichocarpa*; Sab = *Salicornia bigelovii*; Sm = *Selaginella moellendorffii*; Sc = *Solanum cheesmaniae*; Sl = *Solanum lycopersicum*; Sb = *Sorghum bicolor*; Ss = *Suaeda salsa*; Tt = *Triticum turgidum*; Tm = *Triticum monococcum*; Ta = *Triticum aestivum*; Tu = *Triticum urartu*; Vv = *Vitis vinifera*. Node numbers are posterior probability values.

Estimation of duplication history

We constructed a cladogram of the species tree of our representative taxa using the NCBI taxonomy tool (Figure S2). A reconciliation between this species tree and the consensus HKT gene tree, as well as inferences of gene duplication and loss scenarios, were estimated by NOTUNG v.2.6 (Chen et al., 2000) (edge weight threshold = 1.0, duplication = 1.5, loss = 1.0). The consensus of the NJ gene tree was inferred by MEGA v.6.0 (Tamura et al., 2013).

Tests for positive selection

The selection analysis of HKT family genes was performed with the Codeml program as implemented in the PAML v.4.5 package (Yang, 1997). The basis of this test was the nonsynonymous to synonymous substitution rate ratio ($\omega = d_N/d_S$), which can be used as an indicator of selective pressure (Hurst, 2002). Simply stated, values of $\omega < 1$, $\omega = 1$, and $\omega > 1$ correspond to negative purifying selection, neutral evolution, and positive selection, respectively (Bielawski and Yang, 2003). The ratio ω was calculated based on full-length *HKT* coding sequences. To test positive selection, we used the following random-site models: M0 (one ratio), M1a (neutral), M2a (selection), M3 (discrete) with three site classes, M7 (beta), M8a (beta and $\omega = 1$), and M8 (beta and ω) (Nielsen and Yang, 1998).

Branch-site models were used to identify the genes containing positively selected sites in the monocot and dicot lineages. The modified model A test was used to reduce the false positive results as advised in the PAML manual. The parameters for the null model were set as “model =2, NSsites =2, fix_kappa =0, fix_omega =1, omega =1”, while the parameters for the alternative model were set as “model =2, NSsites =2, fix_kappa =0, fix_omega =0”.

We detected variation in ω among sites using the likelihood ratio test (LRT) between M1a vs M2a, M7 vs M8, M8a vs M8, and alternative test (w_2 estimated) vs null test ($w_2 = 1$, fixed). In the LRT, two times the log-likelihood difference between the two models was compared with the chi-squared statistic.

Functional divergence analysis

DIVERGE v.3.0 (Wang and Gu, 2001; Gu and Vander Velden, 2002; Gu et al., 2013) was used to evaluate functional divergence between HKT subfamilies, and to predict critical amino acid residues associated with these functional differences. Type-I functional divergence occurred shortly after gene duplication. Type-II functional divergence occurred in the late phase, after gene duplication. As coefficients of type-I and type-II functional divergence (θI and θII) increased from 0 to 1, functional divergence increased from very weak to extremely strong (Gu, 2006). A large posterior probabilities value (Q_k) indicated a high probability that the functional constraint, or the evolutionary rate of a site, was different between two clusters (Gu, 2006).

Structural analysis

The three-dimensional structure of *AtHKT1* (NP_567354) was inferred by Homology modeling. We used the KtrAB potassium transporter from *Bacillus subtilis* (PDB id 4j7c) as our template (Vieira-Pires et al., 2013), which was obtained from the Protein Data Bank (<http://www.pdb.org/pdb/home/home.do>). Amino acid sequences were aligned with ClustalX v.1.83 (Thompson et al., 1997). Homology modeling was carried out using the alignment mode of the SWISS-MODEL server (Arnold et al., 2006). The quality assessment of the *AtHKT1* 3D model ([Figure S3](#)) was estimated by Structure Analysis and Verification Server (<http://services.mbi.ucla.edu/SAVES/>). The *AtHKT1* model was visualized using the Pymol software (<http://www.pymol.org/>). All images were rendered and resized using Adobe Photoshop v8.0.

Co-evolution analysis

A Co-evolution Analysis, using the Protein Sequences (CAPS) program, was performed to identify co-evolution between amino acid sites (Fares and McNally, 2006). Phylogenetic sequence relationships were used to remove phylogenetic and stochastic dependencies between sites (Fares and McNally, 2006).

RESULTS

Full-length cDNA sequence cloning of *ThHKT1*

The HKT partial CDS was successfully cloned by PCR using our 1U and 1D primers. The RACE method was used to amplify the 3' and 5' UTR sequences to obtain the *ThHKT1* full-length cDNA sequence; resulting fragments were detected by electrophoresis. Lengths of the partial CDS, 3' RACE, and 5' RACE products were 362, 860, and 955 bp, respectively. The 1,817-bp long *ThHKT1* full-length cDNA sequence encoded 500 amino acids and included a 1,503bp open reading frame, a 131-bp 5' UTR, and a 160-bp 3' UTR.

Phylogenetic analysis

A total of 74 sequences from 31 species were used in our analyses. The distribution of HKT genes among the various species is shown in Table 1. Among these genes, 68 were found in 30 angiosperm species, and six in two non-seed plant taxa. Phylogenetic trees generated using Bayesian inference and NJ display similar topologies, although branches in the Bayesian tree were more strongly supported (Figure 1). In both trees, HKT genes clustered according to their evolutionary relationships into three independent clades: A, Subfamily 1, and Subfamily 2 (Figure 1). All genes falling into Subfamily 1 encoded Na⁺ selective transporters, whereas members of Subfamily 2 function as co-transporters of Na⁺ and K⁺ (Mäser et al., 2002).

Within Subfamily 1, subclade C (*HKT1;1* and *HKT1;3*), E (*HKT1;5*), and F (*HKT1;4*), which derived from monocotyledonous plants, were more closely related to dicotyledonous subclades B (*HKT1*) and D (*HKT1*), and diverged farther from the HKTs in the monocotyledonous Subfamily 2. A gene duplication during the evolutionary history resulted in the branching of E and F subclades (*HKT1;5* and *HKT1;4*) and might have contributed to their divergence. Subfamily 2 further divided into two subclades with 100% Bayesian posterior probability: G and H (Figure 1). The HKT genes in both subclades were all derived from monocotyledonous plants.

According to the reconciled phylogeny (Figure 2), we estimated 39 independent gene duplications and 74 gene loss events. These gene loss events were not shown on the phylogeny; rather, they might reflect gene sequences that had yet to be identified and/or sequenced.

Selective pressure on HKT genes in angiosperms

Purifying selection was essential in protein evolution. HKT genes are thought to be undergoing strong purifying selection, as the primary proportion of the sites ($p_0 > 95\%$, $w_0 < 0.1$) in the HKT family are under purifying selection (Table 2).

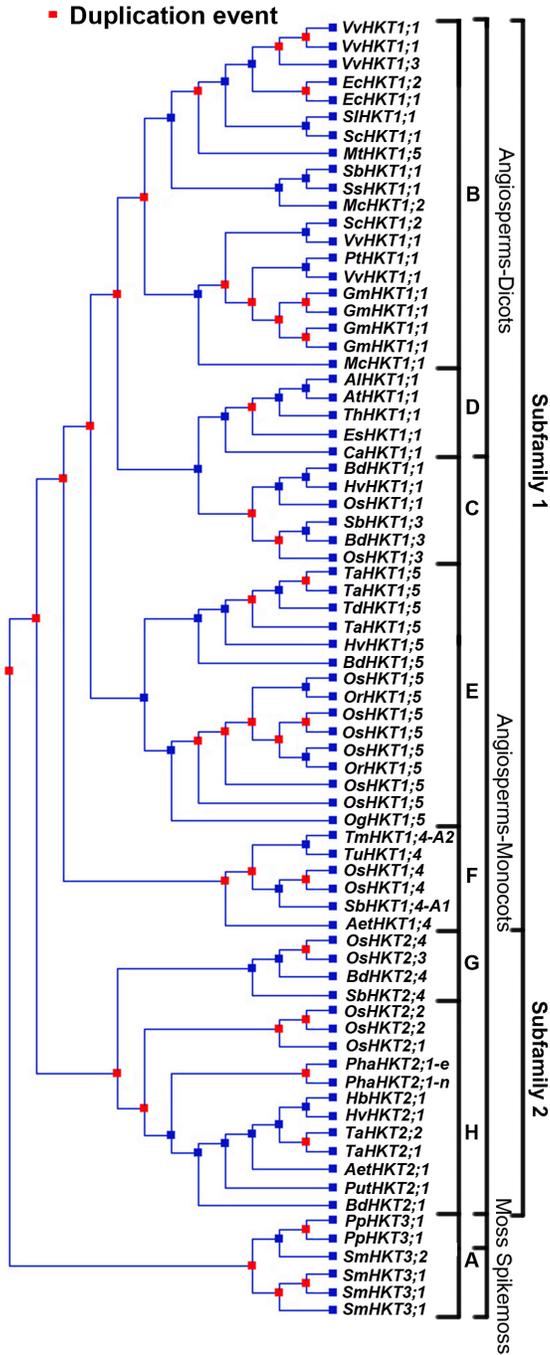


Figure 2. Phylogeny of the HKT family indicating duplication events. The reconciled HKT tree has a 132.5 duplication/loss score and shows 39 gene duplications.

Table 2. Log-likelihood values and parameter estimates obtained by Bayes Empirical Bayes (BEB) analysis under random-site models for HKT genes in angiosperms.

Model code	<i>L</i>	Parameters	Positive selection sites*
M0 (one ratio)	-41,121.01	$\omega = 0.292$	None
M1a (neutral)	-39,300.84	$P_0 = 0.952, P_1 = 0.048, \omega_0 = 0.109, \omega_1 = 1$	Not allowed
M2a (selection)	-39,300.84	$P_0 = 0.952, P_1 = 0.015, P_2 = 0.032, \omega_0 = 0.109, \omega_1 = 1, \omega_2 = 1$	278S**, 370T*, 484M*
M7 (beta)	-38,941.30	$P = 0.096, q = 0.827$	Not allowed
M8 (beta& ω)	-38,802.95	$p_0 = 0.987, p = 0.14, q = 1.737, (p_1 = 0.012) \omega = 1.124$	52H*, 278S**, 287V*, 370T*, 459G*, 474M**
M8a (beta& $\omega=1$)	-38,902.19	$p_0 = 0.977, p = 0.208, q = 1.85, (p_1 = 0.022) \omega = 1$	Not allowed
Branch-site models			
Subfamily 1			
Model A: $\omega_2 = 1$, fixed	-39,300.3	$p_0 = 0.899, \omega_0 = 0.109; p_1 = 0.045, \omega_1 = 1; p_{2a} + p_{2b} = 0.055, \omega_2 = 1$	Not allowed
Model A: ω_2 estimated	-39,297.0	$p_0 = 0.912, \omega_0 = 0.109; p_1 = 0.045, \omega_1 = 1; p_{2a} + p_{2b} = 0.042, \omega_2 = 3.36$	164Y, 189V, 269L, 286C, 311S, 366F, 459G, 474M
Subfamily 2			
Model A: $\omega_2 = 1$, fixed	-39,298.3	$p_0 = 0.884, \omega_0 = 0.109; p_1 = 0.044, \omega_1 = 1; p_{2a} + p_{2b} = 0.07, \omega_2 = 1$	Not allowed
Model A: ω_2 estimated	-39,300.3	$p_0 = 0.899, \omega_0 = 0.109; p_1 = 0.045, \omega_1 = 1; p_{2a} + p_{2b} = 0.055, \omega_2 = 1$	405I, 406C, 419Q

*Codon (amino acid) positions are based on the *Arabidopsis thaliana* AtHKT1 gene (NP_567354). Significance at 5 and 1% levels is indicated by one and two asterisks, respectively. The listed sites are those identified by BEB analysis under M2a, M8, and model A (ω_2 estimated) models.

Table 2 displayed the codon-based maximum likelihood analysis dataset generated under the random-site models. In the M0 model, a single ω (0.291) was assumed for all codons in the sequence. Using this value, no evidence was found for positive selection. Compared with other models, M0 was a better fit for the data because of its much lower log-likelihood value ($L = -41,121.01$) (Table 2). The M8 model fitted the data significantly better than models not permitting positive selection (M7 and M8a) (Table 3). To test for positive selection, three groups of models were compared using the LRT. Models M1a and M2a showed no difference ($2D\ln L = 0$), whereas the M8 model shown a better fit to the data than M7 and M8a (Table 3). Under the M8 (beta and ω) model, 1.2% of codons were estimated to fall into the $\omega = 1.124$ category (Table 2). According to the Bayesian posterior probabilities, under the M8 model, six site candidates were identified (Table 2).

Table 3. Likelihood ratio test statistics for the evaluation of model fit.

LRT pairs	d.f.	$2\Delta\ln L$	<i>LRT p</i>
M1a vs M2a	2	0	1
M7 vs M8	2	276.7	<0.0001
M8a vs M8	1	198.48	<0.0001
A-A1 test			
Subfamily1	1	6.48	0.0109
Subfamily2	1	3.96	0.0466

Moreover, the branch-site model A (ω_2 estimated) fitted monocot and dicot branches (Subfamily 1 branch and Subfamily 2 branch) significantly better than its null model ($\omega_2 = 1$ fixed) (Table 2). Subfamily1 and Subfamily 2 were also identified to be under positive selection (Table 3). Eight sites (164Y, 189V, 269L, 286C, 311S, 366F, 459G, and 474M) were detected under adaptive evolution along the branch of Subfamily 1 (Table 2). In addition, three sites (405I, 406C, and 419Q) were found to be positively selected on the Subfamily 2 branch (Table 2).

Functional divergence analysis of HKT

Type-I functional divergence resulted in altered functional constraints (e.g., different

evolutionary rates) between duplicate genes. Type-II functional divergence, in contrast, did not lead to altered functional constraints, but instead to a radical change in amino acid properties (e.g., charge and hydrophobicity) (Wang and Gu, 2001). To elucidate the relationship between gene evolution and functional divergence, we evaluated HKT genes for type-I and type-II functional divergence using the Gu method as implemented in DIVERGEv.3.0 (Wang and Gu, 2001; Gu and Vander Velden, 2002; Gu et al., 2013). As shown in Table 4, coefficients of type-I functional divergence (θ_I) between HKT clades (other than the F/B pair) were greater than 0 and statistically significant at the 1% level, according to the LRT (Table 4), with θ_I values ranging from 0.196 to 0.676. These results showed that there were significant site-specific altered functional constraints on most HKT members. Similarly, coefficients of type-II functional divergence (θ_{II}) were statistically significant ($P < 0.01$; Table 4) for six compared pairs (E/F, E/H, E/D, F/H, F/D, and H/D), providing solid evidence of type-II functional divergence between these HKT gene pairs.

Table 4. Analysis of functional divergence between clusters of HKTs in angiosperms.

Comparison	$\theta_I \pm SE^a$	P^b	Sites ^c	$\theta_{II} \pm SE^a$	P^b	Sites ^c
Subfamily1/Subfamily2	0.258 ± 0.045	<0.01	10	-0.223 ± 0.17	>0.05	0
E/F	0.416 ± 0.114	<0.01	4	0.172 ± 0.054	<0.01	56
E/H	0.362 ± 0.102	<0.01	2	0.231 ± 0.058	<0.01	107
E/B	0.196 ± 0.064	<0.01	1	-0.19 ± 0.101	>0.05	0
E/D	0.642 ± 0.118	<0.01	38	0.263 ± 0.05	<0.01	124
F/H	0.555 ± 0.117	<0.01	24	0.295 ± 0.06	<0.01	141
F/B	0.148 ± 0.079	>0.05	0	-0.221 ± 0.10	>0.05	0
F/D	0.676 ± 0.134	<0.01	42	0.177 ± 0.058	<0.01	61
H/B	0.463 ± 0.071	<0.01	22	-0.035 ± 0.10	>0.05	0
H/D	0.666 ± 0.112	<0.01	51	0.323 ± 0.057	<0.01	154
B/D	0.4 ± 0.08	<0.01	5	-0.127 ± 0.10	>0.05	0

^a $\theta_{I \text{ or } II} \pm SE$: coefficient and standard error of type-I or type-II functional divergence between clusters. ^bSignificance level (P value) computed using Fisher's transformation. ^cNumber of critical amino acid sites with the highest posterior values (type-I $Q_k > 0.75$; type-II $Q_k > 0.95$).

According to the posterior probability (Q_k), some important amino acid residues potentially involved in functional divergence were predicted. A $Q_k > 0.75$ cutoff was used to predicted type-I functional divergence sites, and $Q_k > 0.95$ was used in type-II. A total of 112 potential type-I sites were detected between all pairs except for the F/B pair. Under the statistical value between Subfamily 1 and 2, no evidence was found for type-II functional divergence ($P > 0.05$; Table 4). Nevertheless, we performed an additional analysis to identify potential type-II sites. By assuming that an absolute posterior probability greater than 1.3 ($|Q_k| > 1.3$) indicated a potential type-II site, we detected 12 potential type-II sites between Subfamily1/ Subfamily 2 pairs.

Protein structure

Because the membrane-pore-membrane (MPM; i.e., transmembrane segment-pore-transmembrane segment) domain structure was highly conserved across HKT family members (Corratgé-Faillie et al., 2010), we were able to use homology modeling to construct a three-dimensional model of *AtHKT1* (NP_567354) (Figure 3A and B). The hydrophobic transmembrane core of Trk and Ktr (K^+ transporter in fungi and bacteria) ion-conducting subunits, as well as that of fungal Trk and plant HKT proteins, typically displayed four

successively arranged MPM domains (a-d) (Corratgé-Faillie et al., 2010). Through gene duplication and fusion, the Trk/Ktr/HKT transporter family had evolved from an ancestral K^+ channel subunit displaying a single MPM domain into transporter proteins with four MPM domains (Corratgé-Faillie et al., 2010). To gain insight into the roles of positive selection and functional divergence in the evolution of the family, we mapped positively selected as well as functionally diverged amino acid sites onto the constructed AtHKT1 model (Figure 3C-E). We determined that the 52H site is located in the MPMa domain, the 278V site is in the MPMc domain, and sites 459G and 474M are contained in the MPMd domain (Figure 3). The distribution of these sites was found to be largely concentrated in the MPM domains, which contain four positively selected sites, eight potential type-I sites, and all potential type-II sites (Figure 3D and E).

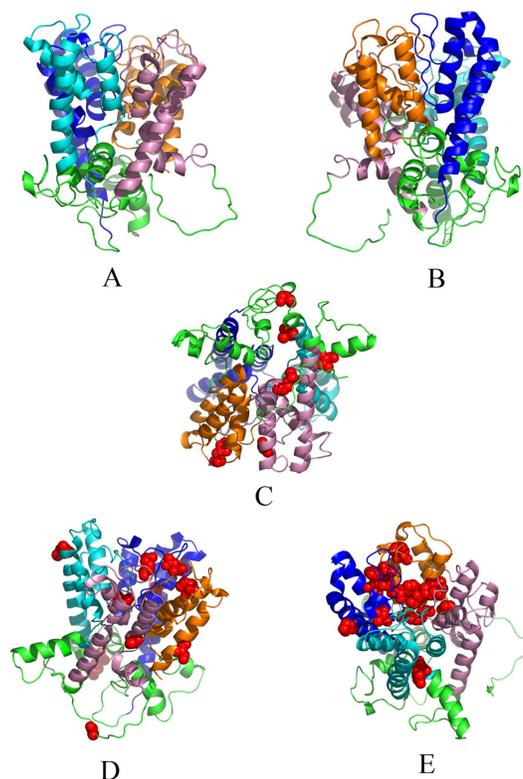


Figure 3. Three-dimensional domain architecture of *AtHKT1* (NP_567354) shown from the front (A) and rotated 90° (B). Domains are indicated as follows: MPMa, orange; MPMb, blue; MPMc, cyan; MPMd, pink. Amino acid sites that are positively selected, type-I functionally divergence-related, or type-II functionally divergence-related are highlighted in red, in C, D, and E, respectively.

Co-evolution analysis

We analyzed HKT co-evolutionary patterns using CAPS (Fares and McNally, 2006), and identified 25 co-evolutionary groups (Table 5). A co-evolutionary group comprised all

those sites that presented a co-evolutionary signal along with other sites within the same group; for example, if site A was coevolving with B, B with C, and A with C, then all three sites were included in one co-evolutionary group. The smallest group contained only three, while the largest group (Table 5) included nine members (65T, 82S, 94L, 179Y, 186A, 221M, 233L, 354I, and 488F). The residues belonging to most of the co-evolving pairs presented significant correlations in their physicochemical properties, including hydrophobicity, molecular weight, or both. For example, we detected that four pairs (65T/82L, 65T/190L, 65T/204T, and 65T/354I) were correlated by their molecular weight variance.

Table 5. Groups of co-evolving amino acids.

Group	Residue indices	Residues	Members
1	41, 56, 65, 78, 488	K, L, T, S, F	5
2	41, 65, 78, 179, 233, 354, 488	K, T, S, Y, L, I, F	7
3	59, 74, 176, 204, 216	T, M, L, T, P	5
4	59, 185, 216, 423	T, T, P, I	4
5	59, 204, 207, 236	T, T, T, P	4
6	60, 73, 221, 398	S, D, M, S	4
7	60, 251, 321	S, L, E	3
8	60, 308, 472	S, S, S	3
9	60, 398, 472	S, S, S	3
10	65, 82, 94, 179, 186, 221, 233, 354, 488	T, S, L, Y, A, M, L, I, F	9
11	65, 82, 94, 179, 186, 221, 233	T, L, E, Y, A, M, L	7
12	65, 179, 186, 190, 221, 233, 354, 488	T, Y, A, L, M, L, I, F	8
13	65, 204, 221	T, T, M	3
14	73, 94, 221, 398	D, E, M, S	4
15	74, 94, 221, 398	M, E, M, S	4
16	74, 204, 221	M, T, M	3
17	78, 79, 82, 179, 221	S, N, L, Y, M	5
18	94, 186, 190, 221, 233, 342	E, M, L, S	4
19	179, 186, 190, 221, 233, 242	Y, A, L, M, L, D	6
20	179, 190, 218, 221, 242, 488	Y, L, N, M, D, F	6
21	179, 221, 233, 332, 242, 354	Y, M, L, N, D, I	6
22	218, 356, 361	N, L, P	3
23	356, 361, 440	L, P, N	3
24	356, 414, 440	L, E, N	3
25	360, 489, 491	L, K, F	3

DISCUSSION

Proteins with similar amino acid sequences could perform very different functions. However, the reverse could also occur: proteins with very different sequences could have similar structure and function (Horie et al., 2001). This phenomenon also existed in the HKT family. Remarkable differences existed between HKT transporter subfamilies 1 and 2 with regard to the structure of putative selectivity pore-forming regions and K⁺ selectivity, as characterized by heterologous expression systems (Hauser and Horie, 2010). Usually single amino acid residues played critical roles in the structure of proteins. The two subfamilies were associated with a glycine/serine substitution of a residue predicted to be in the first pore loop of the protein (Uozumi et al., 2000). HKT transporters from Subfamily 1, including monocotyledonous *HKT1;1*, *HKT1;3*, *HKT1;4*, and *HKT1;5* and dicotyledonous HKTs, were all highly selective for Na⁺; in contrast, HKT transporters from Subfamily 2 co-transport Na⁺ and K⁺ (Corratgé-Faillie et al., 2010). Our phylogenetic analysis suggested that a functional divergence between the monocotyledonous and dicotyledonous *HKT* genes must have occurred early during the evolution of both types of plants. We inferred NJ and Bayesian trees for the HKT sequences.

NJ and Bayesian approaches yielded identical topology. We compared bootstrap proportions of the MEGA NJ tree with posterior probabilities of the Bayesian tree. Where NJ and Bayesian approaches yielded the same topology, the posterior probability values were greater than the bootstrap proportion values.

Recently studies found that the rice *OsHKT1;4* gene from Subfamily 1 worked as a Na⁺ transport when *OsHKT1;4* was expressed in *X. laevis* oocytes and *Saccharomyces cerevisiae*. Unexpectedly, under salinity stress conditions, there was no obvious difference on the growth and ion accumulation between transgenic plants and control plants in the nutrient growth stage (Suzuki et al., 2016). However, under a 200-mM NaCl stress, overexpression of the GmHKT1 gene from Subfamily 1 in transgenic tobacco plants allowed normal growth, whereas the growth of control plants was severely retarded (Chen et al., 2011). Overexpression of *HvHKT2;1* from Subfamily 2 in transgenic barley plants increased tolerance to salt stress. When transgenic plants were grown in 50 or 100 mM NaCl, overexpression of *HvHKT2;1* increased Na⁺ concentrations in the xylem sap (Mian et al., 2011). We theorized that the considerable differences in HKT evolution between halophyte species and glycophyte species would be associated with the salinity tolerance in plants. Although plants with HKT family genes live in different environments, further research is needed to support the association of environmental differences with gene duplication and functional divergence with specific adaptations to the environment.

Maximum likelihood analysis of the *HKT* gene family in angiosperms revealed considerable variation in selective pressure among sites. Some potential selection sites were detected, although the signal of the positive selection was not very strong in these data. Protein structure determined protein function. Therefore, the three-dimensional structure of a protein provided vital information as to how the protein functions (Hunter and Fromme, 2011). After projecting potential selection sites onto the simulated three-dimensional structure, we found that many of the positively selected sites existed in the MPM domain. Positive selection in these domains might have been a factor during the evolution of the HKT family. Gene duplication might result in a new functional gene, or lead to gene loss during evolution (Lynch and Conery, 2000). Positive selection had given rise to functional diversity within HKT Subfamily members, such as the transport of Na⁺ in Subfamily 1 and the co-transport of Na⁺ and K⁺ in Subfamily 2 (Corratgé-Faillie et al., 2010). Despite the polymorphisms observed among *HKT* genes, the overall structure of the protein with respect to the MPM domain remained highly conserved. Purifying selection might have played an essential role during the evolution of this protein.

In this study, 10 of the 11 gene pairs showed statistically significant type-I divergence, and 6 displayed statistically significant type-II divergence. The type-I divergence signal was stronger than that of type-II in our system. This strong type-I signal implied that the early phase of evolution continued to occur long after the initial duplication event that gave rise to these pairs, and suggested that site-specific evolutionary rate changes occurred quite frequently in this evolutionary trajectory. Further evaluation of type-I and type-II functional divergence coefficients between subfamilies 1 and 2 allowed us to strongly reject the null hypothesis of the absence of any functional divergence, because the coefficient (θ_1) value was statistically significant ($\theta_1 = 0.258 \pm 0.045$, $P < 0.01$). These results demonstrated that different functional constraints must work on various amino acid sites in dicotyledon and monocotyledon HKTs. Furthermore, there were 10 critical amino acid sites that were predicted to be highly related to type-I functional divergence, including sites 33S, 73D, and 87I in the MPMa domain, 205T in the MPMb domain, 305C in the MPMc domain, and 441V, 450E, and 468A in the MPMd

domain. In contrast, there was no statistical evidence for type-II functional divergence between Subfamilies 1 and 2 ($p > 0.05$). Nevertheless, using a $|Q_k| > 1.3$ criterion for each site, we identified 12 sites that have a high probability of being related to type-II functional divergence. All of these sites were located in MPM domains (MPMa: 61V/T, 68S/G, 69M/L, 77F/L, and 82L/I; MPMb: 203F/S, 208T/S, 214F/L, and 236P/G; MPMc: 363Y/S; MPMd: 443F/L and 466G/S). Finally, functional analysis outcomes with *TaHKT2;1*, *AtHKT1;1*, and rice genes suggested that site 68S/G of the MPMa domain might play a central role in determining the Na^+ selectivity of the transporter (Horie et al., 2001; Vieira-Pires et al., 2013).

After gene duplication, paralogs went through one of four processes: in the first (non-functionalization) one copy initially became a pseudogene and gradually became extinct (Petrov and Hartl, 2000). This is the most common outcome of a duplication event. In the second process, each copy retained the original function. The others, neo-functionalization and sub-functionalization, are less common but still crucial evolutionary scenarios (Shakhnovich and Koonin, 2006). A gene used in our study, *OsHKT1;2*, was a pseudogene in rice; internal stop codons were overridden to create a full-length amino acid sequence.

ACKNOWLEDGMENTS

Research supported by the National High Technology Research and Development Program of China (“863” Program; #2012AA021701), the National Sciences Foundation of China (#21376112), the Natural Science Foundation of Jiangsu Province Fund (#BK20131344 and #BK20141081), and the Natural Science Foundation of Jiangsu Province Youth Fund (#BK2012073).

REFERENCES

- Abascal F, Zardoya R and Posada D (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104-2105. <http://dx.doi.org/10.1093/bioinformatics/bti263>
- Altschul SF, Madden TL, Schäffer AA, Zhang J, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402. <http://dx.doi.org/10.1093/nar/25.17.3389>
- Amtmann A (2009). Learning from evolution: Thellungiella generates new knowledge on essential and critical components of abiotic stress tolerance in plants. *Mol. Plant* 2: 3-12. <http://dx.doi.org/10.1093/mp/ssn094>
- Arnold K, Bordoli L, Kopp J and Schwede T (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201. <http://dx.doi.org/10.1093/bioinformatics/bti770>
- Bielawski JP and Yang Z (2003). Maximum likelihood methods for detecting adaptive evolution after gene duplication. *J. Struct. Funct. Genomics* 3: 201-212. <http://dx.doi.org/10.1023/A:1022642807731>
- Calero F, Gómez N, Ariño J and Ramos J (2000). Trk1 and Trk2 define the major K(+) transport system in fission yeast. *J. Bacteriol.* 182: 394-399. <http://dx.doi.org/10.1128/JB.182.2.394-399.2000>
- Chen H, He H and Yu D (2011). Overexpression of a novel soybean gene modulating Na^+ and K^+ transport enhances salt tolerance in transgenic tobacco plants. *Physiol. Plant.* 141: 11-18. <http://dx.doi.org/10.1111/j.1399-3054.2010.01412.x>
- Chen K, Durand D and Farach-Colton M (2000). NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* 7: 429-447.
- Corratgé-Faillie C, Jabnoute M, Zimmermann S, Véry A-A, et al. (2010). Potassium and sodium transport in non-animal cells: the Trk/Ktr/HKT transporter family. *Cell. Mol. Life Sci.* 67: 2511-2532. <http://dx.doi.org/10.1007/s00018-010-0317-7>
- Fares MA and McNally D (2006). CAPS: coevolution analysis using protein sequences. *Bioinformatics* 22: 2821-2822. <http://dx.doi.org/10.1093/bioinformatics/btl493>
- Gu X (2006). A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Mol. Biol. Evol.* 23: 1937-1945. <http://dx.doi.org/10.1093/molbev/msl056>

- Gu X and Vander Velden K (2002). DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics* 18: 500-501. <http://dx.doi.org/10.1093/bioinformatics/18.3.500>
- Gu X, Zou Y, Su Z, Huang W, et al. (2013). An update of DIVERGE software for functional divergence analysis of protein family. *Mol. Biol. Evol.* 30: 1713-1719. <http://dx.doi.org/10.1093/molbev/mst069>
- Hauser F and Horie T (2010). A conserved primary salt tolerance mechanism mediated by HKT transporters: a mechanism for sodium exclusion and maintenance of high K(+)/Na(+) ratio in leaves during salinity stress. *Plant Cell Environ.* 33: 552-565. <http://dx.doi.org/10.1111/j.1365-3040.2009.02056.x>
- Horie T, Yoshida K, Nakayama H, Yamada K, et al. (2001). Two types of HKT transporters with different properties of Na⁺ and K⁺ transport in *Oryza sativa*. *Plant J.* 27: 129-138. <http://dx.doi.org/10.1046/j.1365-313x.2001.01077.x>
- Horie T, Hauser F and Schroeder JI (2009). HKT transporter-mediated salinity resistance mechanisms in *Arabidopsis* and monocot crop plants. *Trends Plant Sci.* 14: 660-668. <http://dx.doi.org/10.1016/j.tplants.2009.08.009>
- Hunter MS and Fromme P (2011). Toward structure determination using membrane-protein nanocrystals and microcrystals. *Methods* 55: 387-404. <http://dx.doi.org/10.1016/j.ymeth.2011.12.006>
- Hurst LD (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* 18: 486-487. [http://dx.doi.org/10.1016/S0168-9525\(02\)02722-1](http://dx.doi.org/10.1016/S0168-9525(02)02722-1)
- Lynch M and Conery JS (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151-1155. <http://dx.doi.org/10.1126/science.290.5494.1151>
- Mäser P, Hosoo Y, Goshima S, Horie T, et al. (2002). Glycine residues in potassium channel-like selectivity filters determine potassium selectivity in four-loop-per-subunit HKT transporters from plants. *Proc. Natl. Acad. Sci. USA* 99: 6428-6433. <http://dx.doi.org/10.1073/pnas.082123799>
- Mian A, Oomen RJ, Isayenkov S, Sentenac H, et al. (2011). Over-expression of an Na⁺- and K⁺-permeable HKT transporter in barley improves salt tolerance. *Plant J.* 68: 468-479. <http://dx.doi.org/10.1111/j.1365-313X.2011.04701.x>
- Munns R and Tester M (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59: 651-681. <http://dx.doi.org/10.1146/annurev.arplant.59.032607.092911>
- Nielsen R and Yang Z (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929-936.
- Petrov DA and Hartl DL (2000). Pseudogene evolution and natural selection for a compact genome. *J. Hered.* 91: 221-227. <http://dx.doi.org/10.1093/jhered/91.3.221>
- Posada D (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25: 1253-1256. <http://dx.doi.org/10.1093/molbev/msn083>
- Ronquist F and Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574. <http://dx.doi.org/10.1093/bioinformatics/btg180>
- Rubio F, Gassmann W and Schroeder JI (1995). Sodium-driven potassium uptake by the plant potassium transporter HKT1 and mutations conferring salt tolerance. *Science* 270: 1660-1663. <http://dx.doi.org/10.1126/science.270.5242.1660>
- Schachtman DP and Schroeder JI (1994). Structure and transport mechanism of a high-affinity potassium uptake transporter from higher plants. *Nature* 370: 655-658. <http://dx.doi.org/10.1038/370655a0>
- Schubert S and Lauchli A (1990). Sodium exclusion mechanisms at the root surfaces of two maize cultivars. *Plant Soil* 123: 205-209. <http://dx.doi.org/10.1007/BF00011269>
- Shakhnovich BE and Koonin EV (2006). Origins and impact of constraints in evolution of gene families. *Genome Res.* 16: 1529-1536. <http://dx.doi.org/10.1101/gr.5346206>
- Suyama M, Torrents D and Bork P (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34: W609-12. <http://dx.doi.org/10.1093/nar/gkl315>
- Suzuki K, Yamaji N, Costa A, Okuma E, et al. (2016). OsHKT1;4-mediated Na⁺ transport in stems contributes to Na⁺ exclusion from leaf blades of rice at the reproductive growth stage upon salt stress. *BMC Plant Biol.* 16: 22-36. <http://dx.doi.org/10.1186/s12870-016-0709-4>
- Tamura K, Stecher G, Peterson D, FilipSKI A, et al. (2013). MEGA6: Molecular Evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729. <http://dx.doi.org/10.1093/molbev/mst197>
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882. <http://dx.doi.org/10.1093/nar/25.24.4876>
- Uozumi N, Kim EJ, Rubio F, Yamaguchi T, et al. (2000). The *Arabidopsis* HKT1 gene homolog mediates inward Na⁺ currents in xenopus laevis oocytes and Na⁺ uptake in *Saccharomyces cerevisiae*. *Plant Physiol.* 122: 1249-1259. <http://dx.doi.org/10.1104/pp.122.4.1249>
- Vieira-Pires RS, Szollosi A and Morais-Cabral JH (2013). The structure of the KtrAB potassium transporter. *Nature* 496: 323-328. <http://dx.doi.org/10.1038/nature12055>

Wang Y and Gu X (2001). Functional divergence in the caspase gene family and altered functional constraints: statistical analysis and prediction. *Genetics* 158: 1311-1320.

Yang Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13: 555-556.

Supplementary material

Figure S1. Phylogenetic Neighbor-joining (NJ) tree was generated by MEGA v.6.0 (Tamura et al., 2013), with a 95% partial deletion of gaps and missing data under the P-distance model, and with statistical support for the resulting tree topology assessed using 1000 bootstrap replications.

Figure S2. Species phylogeny used for reconciliation with the gene trees.

Figure S3. Ramachandran Plot of the AtHKT1 3D model was estimated by Structure Analysis and Verification Server (<http://services.mbi.ucla.edu/SAVES/>).

Table S1. Plant HKT sequences used in this study.