Phylogenetic relationships and protein modelling revealed two distinct subfamilies of group II HKT genes between crop and model grasses.

Journal:	<i>Genome</i>
Manuscript ID	gen-2016-0035.R1
Manuscript Type:	Note
Date Submitted by the Author:	19-Mar-2016
Complete List of Authors:	Ariyaratna, Chandima; The University of Western Australia, School of Plant Biology and Institute of Agriculture; Murdoch University, State Agricultural Biotechnology Centre Francki, Michael; Department of Agriculture and Food Western Australia, ; Murdoch University, State Agricultural Biotechnology Centre
Keyword:	HKT subfamilies, gene duplication, protein structure, functional domains, positive Darwinian selection

SCHOLARONE™
Manuscripts

NOTE: *Genome*

Phylogenetic relationships and protein modelling revealed two distinct subfamilies of group II *HKT* genes between crop and model grasses.

H.A. Chandima K. Ariyaratna^{1,2} and Michael G. Francki^{2,3*}

¹School of Plant Biology and Institute of Agriculture, The University of Western Australia, Crawley WA 6009, Australia

²State Agricultural Biotechnology Centre, Murdoch University, Murdoch WA 6150, Australia

³Department of Agriculture and Food Western Australia, 3 Baron Hay Ct, South Perth WA 6151, Australia

*Address for correspondence: michael.francki@agric.wa.gov.au

Draft

Abstract: Molecular evolution of large protein families in closely related species can provide useful insights on structural functional relationships. Phylogenetic analysis of the grass specific gene family, group II HKT identified two distinct subfamilies, I and II. Subfamily I was confined to the small grain cereals and possibly originated from an ancestral gene duplication post divergence from the coarse grain cereal lineage. The core protein structures were highly analogous despite there being no more than 58% amino acid identity between members of the two subfamilies. Distinctly variable regions in known functional domains, however, indicated functional divergence of the two subfamilies. The subs-sets of codons residing external to known functional domains predicted signatures of positive Darwinian selection potentially identifying new domains of functional divergence and providing new insights on the structural functional relationships between protein members of the two subfamilies.

Key words: HKT subfamilies, gene duplication, protein structure, functional domains, positive Darwinian selection

Introduction

High affinity K^+ transporters (HKTs) have an important role in Na^+ transport in plants which makes the genes encoding these proteins potential targets for improvement of salt tolerance in crops. While both dicots and monocots have group I HKTs, grass species are unique in that they evolved a second type of HKTs encoded by the group II *HKT* genes. The group II HKT transporter proteins comprised of a hallmark glycine (G) residue in the first selectivity filter position resulting in a G-G-G-G type selectivity filter, in contrast to group I HKTs where a serine (S) substitution gave rise to a S-G-G-G type selectivity filter (Platten et al. 2006). This structural difference resulted in an important functional modification whereby group I HKTs mediated Na^+ uniport but the transport mode of group II HKT can vary between K^+ or Na^+ uniport and Na^+/K^+ symport (Corratge-Faillie et al. 2010), indicating functional divergence between HKT groups. Four group II *HKT* genes were identified from rice (Garcia-deblas et al. 2003) with orthologs of rice *OsHKT2;1* identified in wheat (*TaHKT2;1*) (Ariyaratna et al. 2014; Schachtman and Schroeder 1994) and barley (*HvHKT2;1*) (Oomen et al. 2012). Systematic BLAST searches in public databases followed by experimental analysis recently identified a second group II *HKT* gene family in wheat, *TaHKT2;2* which was orthologous to rice *OsHKT2;3* and *OsHKT2;4* (Ariyaratna et al. 2016). Although three additional group II *HKT* genes were predicted from *Brachypodium* *BiHKT2;1* and *BiHKT2;4*, and sorghum *SbHKT2;1* (Wang et al. 2014), there is a lack of detailed characterization of these genes and the predicted proteins. Therefore, the wheat and rice genes provided a basis to study the relationship between group II *HKT* genes and comparative analysis of protein structure in grasses.

The structure of membrane transporter proteins can provide inferences on their function in biological systems. Evolutionary relationships between K^+ channels and transporters supports the hypothesis that HKT and related members of the superfamily of K^+ transporters (SKT) evolved from small K^+ channel type prototypes of bacterial origin (Durell et al. 1999; Hanelt et al. 2011). As such, experimental data on the crystal structure of bacterial proteins KcsA K^+ channel from *Streptomyces lividans* (Doyle 1998; Kato et al. 2001) and TrkH transporter from *Vibrio parahaemolyticus* (Cao et al. 2011; Vieira-Pires et al. 2013), were used in protein homology modelling to predict 3-D atomic scale structures of HKT (Ariyaratna et al. 2016; Cotsaftis et al. 2012; Durell et al. 1999). The models can be used to identify functional domains based on structural analogy with known proteins in other organisms.

Evolutionary history of large protein families with functionally diverse members, such as the one encoded by *HKT*, can be used to gain insights on structural and functional relationships. The rate of nonsynonymous (dN) and synonymous substitutions (dS) referred to as the dN/dS ratio (ω) has been used as a common matrix to detect selection and divergence that shaped the

evolutionary history of genes (Kimura 1980; Yang and Bielawski 2000). When $\omega > 1$, a gene is under positive selection and, conversely, when $\omega < 1$, the gene is interpreted as evolved under purifying selection. Similarly, when normalized ($\omega \sim 1$), the gene evolved neutrally. Protein coding genes normally are under purifying selection and, therefore, $\omega < 1$ is often a rare case when the entire gene is considered (Wolfe and Sharp 1993). However, it is not uncommon that amino acid sequences coding for specific domains within the protein resulted in $\omega > 1$ and such domains lead to adaptive evolution (Yang 1998). Therefore, phylogenetic analysis of HKT families across species and genera can be used to track ω values along lineages and detect amino acid changes causing protein structure modifications and potentially functional variations.

Genome and protein data for a diverse array of grasses will enable a robust comparative analysis between species. In this study, gene and protein sequence information for group II *HKT* genes from selected crop and model grass species was integrated to construct phylogenies and compare protein structural models. The outcome of the study will provide information on specific protein domains that were targets of selection during evolution of group II *HKT* and, hence, potential functional sites can be postulated.

Methods

Database searching and protein modelling

Full length cDNA (FL-cDNA) of group II *HKT*s from rice and wheat were used as query sequence in BLASTN and TBLASTX search of rice genome browser (<http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/>), the wheat survey sequence repository (<http://wheat-urgi.versailles.inra.fr/Seq-Repository/BLAST>), and other grass genome sequences database available through nodes of Phytozome 11.0 (<https://phytozome.jgi.doe.gov/pz/portal.html>) to identify related FL-cDNA from rice, wheat, barley, *Brachypodium*, sorghum and maize. Sequences having significant similarity ($E=0.00$) were further studied for predicted translated proteins. Structural models of the proteins were generated using PHYRE 2 (Kelley and Sternberg 2009) and analysed using PyMOL viewer and Molecular Graphics System 1.5.0.4. (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.).

Phylogenetic analysis and prediction of evolutionary history

FL-cDNA alignments were done using translation alignment procedure in Geneious 6.1.2 (Kearse et al. 2012) with default parameters. Maximum likelihood phylogenetic trees were constructed using PHYML (Guindon et al. 2010) based on Jukes Cantor amino acid substitution model. Five group I *HKT* FL-cDNA from *O.sativa*, *T.aestivum* and *T.monococum* were used as an outgroup. One thousand bootstrap replications were performed to calculate the bootstrap support for

identified clades and nodes. The branch-site model in CODEML from PAML package (Yang 2007) was used to study positive Darwinian selection upon a subset of sites in given branches of the phylogenetic tree of the *HKT* gene family. The constructed phylogenetic trees were converted into Newick format and manually adjusted before using in CODEML. The alternative model for episodic evolution ($\omega > 1$) was tested against null model (neutral evolution, $\omega \approx 1$) by comparing likelihood ratio tests (LRTs) to a Chi-square distribution (df=1) (Zhang et al. 2005). When the LRTs indicated positive selection, Bayes empirical Bayes (BEB) method was applied to calculate the posterior probabilities that each codon is from the site class of positive selection under the testing model.

Results

DNA sequence based phylogeny

The phylogeny of group II *HKT* genes were inferred using FL-cDNA sequence alignments. FL-cDNA of group II *HKT* orthologs from rice *OsHKT2;1*, *OsHKT2;2*, *OsHKT2;3* and *OsHKT2;4* and wheat homeologous families *TaHKT2;1* and *TaHKT2;2* (Table I) were used as query sequences in BLASTN and TBLASTN (E=0.00) searches to retrieve orthologous FL-cDNA from barley, *Brachypodium*, sorghum and maize. Altogether, 17 group II *HKT* cDNA were identified and determined as putatively functional as they were transcribed in the selected grass species and coded for non-truncated proteins (Table I). Variable copy number of group II *HKT* genes existed within and between species, ranging from one copy in sorghum and maize to four copies in rice and seven homeologous gene copies in wheat. FL-cDNA of group I *HKT* from rice and wheat were used as an outgroup. An unrooted maximum likelihood tree was generated using PHYML procedure (Figure 1A) and was supported by similar tree topology generated by other algorithms including neighbour-joining and UPGMA (data not shown). The group II *HKT* genes identified two distinct clusters (subfamilies I and II in Figure 1A). Although, the phylogenetic analysis revealed a common origin and, therefore, supported the hypothesis that gene duplication and divergence resulted in evolution of the group II separately from group I *HKT* genes, less than 65% pairwise sequence identity existed between members of the two subfamilies. Based on the analysis of the draft or finished whole genome sequences, both subfamilies were represented in wheat, rice, barley and *Brachypodium* whereas subfamily II were only represented in the finished genome sequence of sorghum and maize, indicating that duplication and divergence within putative functional group II *HKT* genes occurred in the small grain cereals lineage post-divergence of the coarse grain cereals. Among the grass genomes analysed, species specific and within subfamily gene duplications were evident in rice and the B genome of hexaploid wheat. Therefore, the group II *HKT* evolved through both ancient and relatively recent events possibly indicating differential selective constraints across the grass species.

Predicted proteins and 3-D modelling

The full length cDNA sequences from individual species were used to predict and align the translated proteins. Most predicted proteins were of 508 to 555 amino acids in length (Table I). Although AK370026 from barley encoded a largely truncated protein of 370 amino acids due to the presence of an in-frame stop codon (data not shown) and, therefore, possibly represented a transcribed pseudogene. This gene, therefore, was not included in protein analysis and modelling. The predicted proteins encoded by the remaining group II cDNAs revealed 57.7% pairwise amino acid sequence identity. Each domain in the subfamily I proteins showed higher degree of amino acid sequence conservation compared to the same within subfamily II (Figure 1B) further supporting the view that subfamily I evolved from a relatively recent duplication event, whereas subfamily II represented an ancient gene family. The pairwise amino acid sequence identity between subfamilies was 64.9-72% in the membrane-pore-membrane (MPM) domains. In contrast, the amino terminus, cytoplasmic domains and the carboxyl terminus were highly variable (Figure 1B). The between group pairwise amino acid identity was considerably less than that of the within group for the amino terminus, cytoplasmic domains 1, 2 and 3 (CyD1, CyD2 and CyD3) and the carboxyl-terminus (Figure 1B), indicating diversification of the two subfamilies in these domains.

The 3D atomic scale models of proteins encoded by group II *HKT* were predicted by homology based protein structure modelling using PHYRE and analysed by PHYMOL. The tertiary structures of the group II HKT were structurally conserved in MPM domains (Figure 2A and B), despite no more than 72% amino acid residue identity was revealed between the two subfamilies in these domains. As expected, all these proteins had a conserved G-G-G-G type selectivity filter motif (Figure 2A and B), however, protein encoded by rice *OsHKT2;1* was an exception having S-G-G-G filter motif (data not shown). However, distinctly different motifs between the two group II *HKT* subfamilies were identified in the cytoplasmic domain 1 and 3 (CyD1 and CyD3), carboxyl terminus of the sixth transmembrane domain (M_{2C}), external domains 1 and 2 (ExD1 and ExD2) and in carboxyl terminus, (Figure 2A and B). Although the NH_2 - terminus of M_{2C} domain in proteins of both subfamily I and II identified an intermembrane loop, the carboxyl terminus of M_{2C} was variable whereby proteins of subfamily II predicted an α -helix but all proteins of subfamily I, except the rice protein *OsHKT2;1*, predicted a coil in this domain causing notable structural differences between the subfamilies (Figure 2A). CyD1 in all proteins of subfamily II except orthologs from sorghum and maize identified a 13 amino acid deletion and, therefore, this domain was truncated in most members of subfamily II (Figure 2C). Similarly, the subfamily II proteins had a truncated carboxyl terminus in all proteins except in rice proteins *OsHKT2;3* and *OsHKT2;4* and sorghum protein *SbHKT2;1* (Figure 2A). Therefore, superimposing models for each of the 16 proteins revealed structural analogy in the core protein structure but variable regions in the cytoplasmic and external domains.

Divergence following gene duplication may indicate relaxed constraints and an evolutionary role for specialized adaptation or accumulation of neutral substitutions. These possibilities were investigated by comparing neutral ($\omega \approx 1$) against selective models ($\omega > 1$) for group II *HKT* evolution based on ω values in the CODEML branch-site model of the PAML package. The model ($\omega > 1$) fitted with the phylogenetic tree of the group II *HKT* cDNAs ($P < 0.0004$) showed that positive selection affected a subsets of amino acids in proteins of subfamily I during a specific evolutionary time. The amino acid residues that were fixed faster than the neutral substitution rate were identified by Bayes Empirical Bayes (BEB) method in CODEML. Residue S93 and F288 in the protein encoded by *TaHKT2;1 7AL-1* and the corresponding residues in the remaining proteins of subfamily I had a high probability ($> 96\%$) of undergoing positive selection (Figure 1B and 2C). The amino acids within proteins of subfamily II corresponding to residues 93 and 288 were substituted with alanine and phenylalanine, respectively (Figure 2C). Therefore, it is likely that these sites were signature sequences for positive selection and, hence, identified sites of potential functional divergence between the two protein subfamilies.

Discussion

Phylogenetic analysis of group II *HKT* provided evidence for gene duplication and divergence resulting in two distinct subfamilies. In each species, members of the two subfamilies were located adjacently on the same chromosome (Huang et al. 2008; Very et al. 2014) and supported the view that these genes resulted from duplication of a common ancestral gene. Members of the subfamily II were identified from all grass species studied; therefore, members of this subfamily were present in the common ancestral grass species. In contrast, the absence from sorghum and maize indicated that subfamily I evolved in the small grains lineage possibly 60 to 32 million years ago, based on the molecular clock for grass genome evolution (International *Brachypodium* 2010). Accelerated gene evolution caused diversification of subfamily I and resulted in a significantly low degree of sequence homology between members of the two subfamilies. Rice and wheat provided evidence for further expansion of the subfamilies identifying at least one additional duplication event that resulted in related but not identical gene copies in rice but lack of divergence resulted in the evolution of identical gene duplicates in B genome of wheat. However, not all duplication events were taken into account in this analysis because some of the duplicated genes were either not transcribed or expressed but contained a pre-mature stop codon resulting in largely truncated and presumably non-functional proteins. Transcribed but non-functional group II *HKT* gene copies were identified in barley in this study and non-transcribed pseudogenes were previously reported in wheat (Ariyaratna et al. 2014; Huang et al. 2006) and rice (Garcia-deblas et al. 2003) but the significance of duplicated genes resulting in pseudogenes are yet to be ascertained. Gene duplication and divergence is a key mechanisms of molecular evolution of stress tolerance in plants, with evidence found from

different salt transporter families such as *NHX* and *SOS1* (Pires et al. 2013). Similarly this study showed sustained gene duplication events and divergence in members of group II *HKT* during speciation and adaptive evolution of grass species.

The core structure and the selectivity filter domain of the proteins in the two group II *HKT* subfamilies were highly conserved, specifically in tertiary core structures. Furthermore, it was found that most functional domains previously studied for *HKT* transporters (Ali et al. 2012; Diatloff et al. 1998; Kato et al. 2001; Maser et al. 2002; Rubio et al. 1995; Rubio et al. 1999) were conserved between the two subfamilies (data not shown). However, the M_{2C} domain that identified an important regulatory structure in the SKT super family including *HKT* (Levin and Zhou 2014) was variable between the two subfamilies. The frequent presence of charged residues in M_{2C} of *HKT* formed an intermembrane loop directly below the selectivity filter and a tilted α -helix in the carboxyl terminus that prevented ion diffusion facilitating K^+ transport by these transporters (Kato et al. 2014). Compared to other members of the SKT family such as *TrkH* (Cao et al. 2011) or group I *HKT* (data not shown), the α -helix in the M_{2C} carboxyl terminus was not well formed in proteins of both group II *HKT* subfamilies except those encoded by sorghum and maize genes and the rice *OsHKT2;1*. However, the absence of the α -helix in subfamily I, except in rice *OsHKT2;1* proteins in contrast to partial α -helix in subfamily II proteins could have different impacts on the K^+ transport mechanisms by the proteins in the two subfamilies. Similarly, sequence diversity resulting in structural differences was detected within a known functional domain in the *CyD1* in members of the subfamily II. The *CyD1* in group II *HKT* was similar to *Trk1,2*, but unusually large compared to other members in SKT and K^+ channels (Durell et al. 1999). Functional significance was revealed when a 30bp deletion in *CyD1* (E150-L166) improved K^+ selectivity of the transporter encoded by *TaHKT2;1 7BL-1/2* which resulted in an enhanced salt tolerance when expressed in yeast cells (Liu et al. 2000). Therefore, it appears that truncation of *CyD1* increases selectivity for K^+ over Na^+ in the *HKT* transporters. The subfamily II proteins except in sorghum and maize were distinguished from subfamily I in having a truncated carboxyl terminus. Carboxy termini have an essential functional role in ion transport by *HKT* (Kato et al. 2001) and the truncation of this domain could result in largely modified functional properties in group II *HKT*. However, rice *OsHKT2;1* was an exception although clustered with subfamily I, this protein had a S-G-G-G type selectivity filter, a distinct α -helix in M_{2C} and in the $-COOH$ and was structurally analogous to functional domains group I *HKT* proteins. Diversification in these known functional domains clearly indicated functional divergence between the group II subfamilies.

Interestingly, a significant level of structural variability was also revealed in the cytoplasmic and external domains between the two group II *HKT* subfamilies. Mechanistic models based on protein chemistry and kinetics for ion transport by *HKT* have recognized ion binding events within the transmembrane segments containing the selectivity filter (Benito et al.

2014; Corratge-Faillie et al. 2010; Durell and Guy 1999; Gassmann et al. 1996; Haro and Rodríguez-Navarro 2002; Rubio et al. 1999; Tholema et al. 2005). These models, however, did not discriminate whether external or cytoplasmic domains were involved in the transport process. On the other hand cytoplasmic and external segments in other ion transporter systems from plants and animals played a crucial role in metal ion binding and ion selectivity or in essential conformational changes associated with ion transport (Cohen et al. 2003; Smicun et al. 1999; Viganoa et al. 2002; Weinglass and Kaback 2000). These highly variable sites could be evolving as adaptive mechanisms specific to each subfamily or simply represented an accumulation of neutral mutations due to relaxed selective constrain. Evaluating this possibility by evolutionary trace analysis for $\omega > 1$ identified a subset of codons that showed signatures of episodic adaptive evolution on the first p-loop adjacent to the selectivity filter and also on CyD1 indicating diversification between the group II HKT subfamily I and II in structural domains external to known functional domains. Although a single glycine molecule was necessary for K^+ / Na^+ selectivity, models proposed for HKT function required the involvement of a filter domain with multiple residues (Benito et al. 2014). In the related TrkH protein, the selectivity filter in each P-loop domain (A-D) contained 5 consecutive but variable amino acid residues (Domene and Furini 2012). The position 93 in the $M_{A1}P_A M_{A2}$ filter domain of TaHKT2;1 7AL-1 and corresponding residues in other subfamily II proteins was structurally analogous to the 6th Tyrocinine in the $M_{A1}P M_{A2}$ selectivity filter of TrkH. The fact that these residues were variable between the two protein subfamilies and was subjected to episodic positive selection implicated differential functional role between the two subfamilies. Further studies in these domains will provide opportunities to understand functional sites and direct specific biological experiments to validate the role of group II HKTs in selective ion transport mechanisms.

In conclusion, the group II HKT genes in crop and model grasses are classified into two distinct sub-families. Although rice genes *OsHKT2;1*, *OsHKT2;2* and *OsHKT2;4* are variable for ion selectivity and functionality including maintaining Na^+/K^+ homoeostasis, selective cation and transporters and channels (Horie et al. 2007; 2011; Lan et al. 2010; Sassi et al. 2012; Yao et al. 2010), the within and between sub-family variation indicated that orthologs from other grasses may have similar or, indeed, different ion selectivity and transport properties. Nevertheless, this study demonstrated that presuming gene function based on classification within a sub-family needs to be treated with some caution and that gene functionality needs to be determined experimentally for each member of the group II *HKT* genes in crop and model grasses.

Acknowledgments

This work was partly funded by a Scholarship for International Research Fees (SIRF), University International Stipend (UIS) and Underwood PhD Scholarship, UWA awarded to HACKA.

Reference

- Ali, Z., Park, H.C., Ali, A., Oh, D.H., Aman, R., Kropornicka, A., Hong, H., Choi, W., Chung, W.S., Kim, W.Y., Bressan, R.A., Bohnert, H.J., Lee, S.Y., and Yun, D.J. 2012. TsHKT1;2, a *HKT1* homolog from the extremophile *Arabidopsis* relative *Thellungiella salsuginea*, shows K⁺ specificity in the presence of NaCl. *Plant physiol.*, **158**(3): 1463-1474. doi: 10.1104/pp.111.193110.
- Ariyaratna, H.C., Ul-Haq, T., Colmer, T.D., and Francki, M.G. 2014. Characterization of the multigene family *TaHKT 2; 1* in bread wheat and the role of gene members in plant Na⁺ and K⁺ status. *BMC Plant Biol.*, **14**(1): 159.
- Ariyaratna, H.C.K., Oldach, K.H., and Francki, M.G. 2016. A comparative gene analysis with rice identified orthologous group II HKT genes and their association with Na⁺ concentration in bread wheat. *BMC Plant Biol.*, **16**(1): 1.
- Benito, B., Haro, R., Amtmann, A., Cuin, T.A., and Dreyer, I. 2014. The twins K⁺ and Na⁺ in plants. *J. Plant. Physiol.*, **171**(9): 723-731. doi: 10.1016/j.jplph.2013.10.014.
- Cao, Y., Jin, X., Huang, H., Derebe, M.G., Levin, E.J., Kabaleeswaran, V., Pan, Y., Punta, M., Love, J., Weng, J., Quick, M., Ye, S., Kloss, B., Bruni, R., Martinez-Hackert, E., Hendrickson, W.A., Rost, B., Javitch, J.A., Rajashankar, K.R., Jiang, Y., and Zhou, M. 2011. Crystal structure of a potassium ion transporter, TrkH- supplementary. *Nature*, **471**(7338): 336-340. doi: 10.1038/nature09731.
- Cohen, A., Nevo, Y., and Nelson, N. 2003. The first external loop of the metal ion transporter DCT1 is involved in metal ion binding and specificity. *Proc. Natl. Acad. Sci. U S A.*, **100**(19): 10694-10699. doi: 10.1073/pnas.1934572100.
- Corratge-Faillie, C., Jabnune, M., Zimmermann, S., Very, A.A., Fizames, C., and Sentenac, H. 2010. Potassium and sodium transport in non-animal cells: the *Trk/Ktr/HKT* transporter family. *Cell Mol. Life Sci.*, **67**(15): 2511-2532. doi: 10.1007/s00018-010-0317-7.
- Cotsaftis, O., Plett, D., Shirley, N., Tester, M., and Hrmova, M. 2012. A Two-Staged Model of Na⁺ Exclusion in Rice Explained by 3D Modeling of *HKT* Transporters and Alternative Splicing. *PLoS One*, **7**(7). doi: 10.1371/journal.pone.0039865.g001.
- Diatloff, E., Kumar, R., and Schachtman, D.P. 1998. Site directed mutagenesis reduces the Na⁺ affinity of *HKT1*, an Na⁺ energized high affinity K⁺ transporter. *FEBS Lett.*, **432**(1): 31-36.
- Domene, C., and Furini, S. 2012. Molecular dynamics simulations of the TrkH membrane protein. *Biochemistry* **51**(8): 1559-1565. doi: 10.1021/bi201586n.
- Doyle, D.A. 1998. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science*, **280** (5360): 69-77. doi: 10.1126/science.280.5360.69.

- Durell, R.D., Hao, Y., Nakamura, T., Bakker, E.P., and Guy, R.H. 1999. Evolutionary Relationship between K^+ Channels and Symporters. *Biophys. J.*, **77**: 775-788.
- Durell, S.R., and Guy, H.R. 1999. Structural models of the KtrB, TrkH, and Trk1, 2 symporters based on the structure of the KcsA K^+ channel. *Biophys. J.*, **77**(2): 789-807.
- Garciadeblas, B., Senn, M.E., Banuelos, M.A., and Rodríguez-Navarro, A. 2003. Sodium transport and *HKT* transporters: the rice model. *Plant J.*, **34**(6): 788-801.
- Gassmann, W., Rubio, F., and Schroeder, J.I. 1996. Alkali cation selectivity of the wheat root high-affinity potassium transporter *HKT1*. *Plant J.*, **10**(5): 869-882. doi: 10.1046/j.1365-313X.1996.10050869.x.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**(3): 307-321.
- Hanelt, I., Tholema, N., Kroning, N., Vor der Bruggen, M., Wunnicke, D., and Bakker, E.P. 2011. KtrB, a member of the superfamily of K^+ transporters. *EJCB.*, **90**(9): 696-704. doi: 10.1016/j.ejcb.2011.04.010.
- Haro, R., Banuelos, M.A., Senn, M.E., Barrero-Gil, J., and Rodríguez-Navarro, A. 2005. *HKT1* mediates sodium uniport in roots. Pitfalls in the expression of *HKT1* in yeast. *Plant Physiol.*, **139**(3): 1495-1506. doi: 10.1104/pp.105.067553.
- Haro, R., and Rodríguez-Navarro, A. 2002. Molecular analysis of the mechanism of potassium uptake through the *TRK1* transporter of *Saccharomyces cerevisiae*. *Biochim. Biophys.*, **1564**(1): 114-122.
- Horie, T., Yoshida, K., Nakayama, H., Yamada, K., Oiki, S., and Shinmyo, A. 2001. Two types of *HKT* transporters with different properties of Na^+ and K^+ transport in *Oryza sativa*. *Plant J.*, **27**(2): 129-138.
- 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^+ influx component into K^+ starved roots for growth. *The EMBO J.* **26**(12):3003-14.
- Horie, T., Brodsky, D.E., Costa, A., Kaneko, T., Lo Schiavo, F., Katsuhara, M., et al. K^+ transport by the *OsHKT2;4* transporter from rice with atypical Na^+ transport properties and competition in permeation of K^+ over Mg^{2+} and Ca^{2+} ions. *Plant Physiol.* **156**(3):1493-507.
- Huang, S., Spielmeyer, W., Lagudah, E.S., James, R.A., Platten, J.D., Dennis, E.S., and Munns, R. 2006. A sodium transporter (*HKT7*) is a candidate for *Nax1*, a gene for salt tolerance in durum wheat. *Plant Physiol.*, **142**(4): 1718-1727.
- Huang, S., Spielmeyer, W., Lagudah, E.S., and Munns, R. 2008. Comparative mapping of *HKT* genes in wheat, barley, and rice, key determinants of Na^+ transport, and salt tolerance. *J. Exp. Bot.*, **59**(4): 927-937. doi: 10.1093/jxb/ern033.

- International *Brachypodium*, I. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**(7282): 763-768. doi: 10.1038/nature08747.
- Kato, N., Akai, M., Zulkifli, L., Matsuda, N., Kato, Y., Goshima, S., Hazama, A., Yamagami, M., Guy, R.H., and Uozumi, N. 2014. Role of Positively Charged Amino Acids in the M2DTransmembrane Helix of *Ktr/Trk/HKT* Type Cation Transporters. *Channels* **1**(3): 161-171. doi: 10.4161/chan.4374.
- Kato, Y., Sakaguchi, M., Mori, Y., Saito, K., Nakamura, T., Bakker, E.P., Sato, Y., Goshima, S., and Uozumi, N. 2001. Evidence in support of a four transmembrane-pore-transmembrane topology model for the *Arabidopsis thaliana* Na⁺/K⁺ translocating *AtHKT1* protein, a member of the superfamily of K⁺ transporters. *Proc. Natl. Acad. Sci. U S A.*, **98**(11): 6488-6493. doi: 10.1073/pnas.101556598.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**(12): 1647-1649. doi: 10.1093/bioinformatics/bts199.
- Kelley, L.A., and Sternberg, M.J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.*, (3): 363-371.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, **16**(2): 111-120.
- Lan, W.Z., Wang, W., Wang, S.M., Li, L.G., Buchanan, B.B., Lin, H.X., et al 2010. A rice high affinity potassium transporter (HKT) conceals a calcium-permeable cation channel. *Proc Natl Acad Sci U S A.* 107(15):7089-94.
- Levin, E.J., and Zhou, M. 2014. Recent progress on the structure and function of the TrkH/KtrB ion channel. *Curr. Opin. Struct. Biol.*, **27**: 95-101. doi: 10.1016/j.sbi.2014.06.004.
- Liu, W., Schachtman, D.P., and Zhang, W. 2000. Partial deletion of a loop region in the high affinity K⁺ transporter HKT1 changes ionic permeability leading to increased salt tolerance. *J. Biol. Chem.*, **275**(36): 27924-27932. doi: 10.1074/jbc.M002056200.
- Maser, P., Eckelmana, B.V., R. Horiea, T., Fairbairnb, D.J., Kuboc, M., Yamagami, M., Yamaguchie, K., Nishimura, M., Uozumi, N., Robertson, W., and Sussman, M.R.S., J.I. 2002. Altered shoot/root Na⁺ distribution and bifurcating salt sensitivity in *Arabidopsis* by genetic disruption of the Na⁺ transporter *AtHKT1*. *FEBS Lett.*, **531**: 157-161.
- Matsumoto, T., Tanaka, T., Sakai, H., Amano, N., Kanamori, H., Kurita, K., Kikuta, A., Kamiya, K., Yamamoto, M., and Ikawa, H. 2011. Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. *Plant Physiol.*, **156**(1): 20-28.
- Oomen, R.J., Benito, B., Sentenac, H., Rodriguez-Navarro, A., Talon, M., Very, A.A., and Domingo, C. 2012. *HKT2;2/1*, a K⁺ permeable transporter identified in a salt-tolerant rice

- cultivar through surveys of natural genetic polymorphism. *Plant J.*, **71**(5): 750-762. doi: 10.1111/j.1365-313X.2012.05031.x.
- Pires, I.S., Negrão, S., Pentony, M.M., Abreu, I.A., Oliveira, M.M., and Purugganan, M.D. 2013. Different evolutionary histories of two cation/proton exchanger gene families in plants. *BMC Plant Biol.*, **13**(1): 1.
- Platten, J.D., Cotsaftis, O., Berthomieu, P., Bohnert, H., Davenport, R.J., Fairbairn, D.J., Horie, T., Leigh, R.A., Lin, H.X., Luan, S., Maser, P., Pantoja, O., Rodriguez-Navarro, A., Schachtman, D.P., Schroeder, J.I., Sentenac, H., Uozumi, N., Very, A.A., Zhu, J.K., Dennis, E.S., and Tester, M. 2006. Nomenclature for *HKT* transporters, key determinants of plant salinity tolerance. *Trends Plant Sci.*, **11**(8): 372-374. doi: 10.1016/j.tplants.2006.06.001.
- Ren, Z.H., Gao, J.P., Li, L.G., Cai, X.L., Huang, W., Chao, D.Y., Zhu, M.Z., Wang, Z.Y., Luan, S., and Lin, H.X. 2005. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nature Genet.*, **37**(10): 1141-1146. doi: 10.1038/ng1643.
- Rubio, F., Gassmann, W., and Schroeder, J.I. 1995. Sodium-driven potassium uptake by the plant potassium transporter *HKT1* and mutations conferring salt tolerance. *Science*, **270**(5242): 1660-1663.
- Rubio, F., Schwarz, M., Gassmann, W., and Schroeder, J.I. 1999. Genetic selection of mutations in the high affinity K⁺ transporter *HKT1* that define functions of a loop site for reduced Na⁺ permeability and increased Na⁺ tolerance. *J. Biol. Chem.*, **274**(11): 6839-6847.
- Sassi, A., Mieulet, D., Khan, I., Moreau, B., Gaillard, I., Sentenac, H., et al. 2012. The rice monovalent cation transporter *OsHKT2; 4*: revisited ionic selectivity. *Plant Physiol.* **160**(1):498-510.
- Schachtman, D.P., and Schroeder, J.I. 1994. Structure and transport mechanism of a high-affinity potassium uptake transporter from higher plants. *Nature*, **370**: 655-658.
- Smicun, Y., Campbell, S.D., Chen, M.A., Gu, H., and Rudnick, G. 1999. The role of external loop regions in serotonin transport. *J. Biol. Chem.*, **274**(17).
- Tholema, N., der Brüggen, M.V., Mäser, P., Nakamura, T., Schroeder, J.I., Kobayashi, H., Uozumi, N., and Bakker, E.P. 2005. All four putative selectivity filter glycine residues in KtrB are essential for high affinity and selective K⁺ uptake by the KtrAB system from *Vibrio alginolyticus*. *J. Biol. Chem.*, **280**(50): 41146-41154.
- Very, A.A., Nieves-Cordones, M., Daly, M., Khan, I., Fizames, C., and Sentenac, H. 2014. Molecular biology of K⁺ transport across the plant cell membrane: what do we learn from comparison between plant species? *J. Plant Physiol.*, **171**(9): 748-769. doi: 10.1016/j.jplph.2014.01.011.
- Vieira-Pires, R.S., Szollosi, A., and Morais-Cabral, J.H. 2013. The structure of the KtrAB potassium transporter. *Nature*, **496**(7445): 323-328.

- Viganoa, C., Grimarda, V., Margollesb, A., Goormaghtigha, E., Veenc, H.W.V., Koningsb, W.N., and Ruysschaerta, J.-M. 2002. A new experimental approach to detect long-range conformational changes transmitted between the membrane and cytosolic domains of LmrA, a bacterial multidrug transporter. *FEBS Lett.*, **530**: 197-203.
- Wang, T.T., Ren, Z.J., Liu, Z.Q., Feng, X., Guo, R.Q., Li, B.G., Li, L.G., and Jing, H.C. 2014. SbHKT1; 4, a member of the high-affinity potassium transporter gene family from *Sorghum bicolor*, functions to maintain optimal Na⁺/K⁺ balance under Na⁺ stress. *J. Integr. Plant Biol.*, **56**(3): 315-332.
- Weinglass, A.B., and Kaback, H.R. 2000. The central cytoplasmic loop of the major facilitator superfamily of transport proteins governs efficient membrane insertion. *Proc. Natl. Acad. Sci. U S A.*, **97**(16): 8938-8943. doi: 10.1073/pnas.140224497.
- Wolfe, K.H., and Sharp, P.M. 1993. Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. *J. Mol. Evol.*, **37**(4): 441-456.
- Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.*, **15**(5): 568-573.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, **24**(8): 1586-1591.
- Yang, Z., and Bielawski, J.P. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.*, **15**(12): 496-503.
- Yao, X., Horie, T., Xue, S., Leung, H.Y., Katsuhara, M., Brodsky, D.E., et al. 2010. Differential sodium and potassium transport selectivities of the rice *OsHKT2;1* and *OsHKT2;2* transporters in plant cells. *Plant Physiol.* **152**(1):341-55.
- Zhang, J., Nielsen, R., and Yang, Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.*, **22**(12): 2472-2479.

Table I. Group II *HKT* transcripts identified from grass genome browsers and group I transcripts used as outgroups for phylogenetic analysis.

			Predicted protein			Reference
	Accession number	Gene	length a.a.	Species	chromosome	
Group II	XM_003560467.3	<i>BraHKT2;1</i>	532	<i>B. distachyon</i>	1	Unpublished
	XM_003563466.2	<i>BraHKT2;4</i>	508	<i>B. distachyon</i>	1	Unpublished
	AK370026	Not determined	370	<i>H.vulgare</i>	7L	Matsumoto et al., 2011
	AM000056.1	<i>HvHKT2;1</i>	531	<i>H.vulgare</i>	7L	Haro et al., 2005
	AB061311	<i>OsHKT2;1</i>	530	<i>O. sativa</i>	6	Horie et al.,2001
	EF373552	<i>OsHKT2;2</i>	530	<i>O. sativa</i>	6	Horie et al. 2001; Mäser et al. 2002; Oomen et al., 2012
	AJ491820	<i>OsHKT2;3</i>	509	<i>O. sativa</i>	3	Garciadeblas et al., 2003
	AJ491855	<i>OsHKT2;4</i>	509	<i>O. sativa</i>	6	Garciadeblas et al., 2003
	Sb10g029000	<i>SbHKT2;1</i>	545	<i>S.bicolor</i>	10	Wang et al.2014
	KJ540616	<i>TaHKT2;1 7AL-1</i>	533	<i>T.aestivum</i>	7AL	Ariyaratna et al, 2014
	U16709	<i>TaHKT2;1 7BL-1/2*</i>	533	<i>T.aestivum</i>	7BL	Schachtman, and Schroeder, 1994
	KJ540618	<i>TaHKT2;1 7DL-1</i>	531	<i>T.aestivum</i>	7DL	Ariyaratna et al, 2014
	KR422354	<i>TaHKT2;2 7AL-1</i>	508	<i>T. aestivum</i>	7AL	Ariyaratna et al.2016
	KR422355	<i>TaHKT2;2 7BL-1</i>	508	<i>T. aestivum</i>	7BL	Ariyaratna et al.2016
	KR422356	<i>TaHKT2;2 7DL-1</i>	508	<i>T. aestivum</i>	7DL	Ariyaratna et al.2016
	GRMZM2G135674	Not determined	555	<i>Z.mays</i>	5	Schnable et al., 2009
Group I	AJ491816.1	<i>OsHKT1;1</i>	552	<i>O. sativa</i>	4	Garciadeblas et al.,2003
	AJ491818.1	<i>OsHKT1;3</i>	531	<i>O. sativa</i>	2	Garciadeblas et al.,2003
	JF920142.1	<i>OsHKT1;4</i>	554	<i>O. sativa</i>	4	Cotsaftis et al.,2012
	DQ148410.1	<i>OsHKT1;5</i>	500	<i>O. sativa</i>	1	Ren et al.,2005
	EF062819.1	<i>TmHKT1;4-A2</i>	554	<i>T.monococum</i>	Unknown	Huang et al., 2006

* *TaHKT2;1 7BL-1/2* indicate two identical genes (Ariyaratna et al., 2014)

Figure captions

Figure 1. (A). The maximum likelihood tree based on FL-cDNA of group II *HKT* in crop and model grass species where the group I *HKT* genes were used as an outgroup. The bootstrap consensus tree was inferred from 1000 replications with bootstrap values indicated for each branch. The scale bar indicates number of substitutions per site. (B)-Structurally variable domains in the group II HKT subfamilies identified as a schematic representation of membrane topology of HKT transporters. Membrane-P-loop-membrane domains (MPM-black square), cytoplasmic domains (CyD1-3- black lines) amino (NH₄- black lines) and carboxy (COOH black lines) terminal ends are indicated. Lipid bilayer is indicated as a grey box. The pairwise amino acid sequence identity is indicated as a percentage below each domain including between subfamilies (top), within subfamily 1 (middle) and within sub-family II (bottom). The residues at S93 and F288 in proteins relative to that encoded by *TaHKT2;1 7AL-1* showed a 96% probability of undergoing positive selection.

Figure 2 (A) Side view of the 3D-atomic scale structural models of proteins encoded by group II *HKT* genes. Predicted proteins of subfamily II were superimposed on subfamily I where blue represents conserved structure between subfamily I and II and variability in domains are represented in yellow specific to subfamily I and green is specific to subfamily II. The conserved filter glycine residues were indicated in red with variable regions in M_{2C}, CyD1 and ExD1 indicated. The enlarged view of M_{2C} domain and -COOH terminus of proteins encoded by *TaHKT2;1 7AL-1* and *TaHKT2;2 7AL-1* is shown to the right of the figure representing examples for variable domains between subfamilies. (B). Top view of the 3D model where filter glycine molecules (red) and external domains (ExD) are indicated. An enlarged view of stick representation of S93 in *TaHKT2;1 7AL-1* and corresponding residue in *TaHKT2;2 7AL-1*, A82 are shown to the right of the figure representing examples of signature residues for episodic positive selection. (c). Variable segments of multiple sequence alignments between group II HKT proteins with respect to the position in *TaHKT2;1 7AL-1*; signature residues of episodic positive selection S93 and F288 are highlighted for subfamily I (orange letters) and subfamily II (blue letters) and 13 amino acid deletion for the variable domain in CyD1 indicated by the red box. Rice protein *OshKT2;1* was an exception to the general observations made on the remaining group II HKT and was not included in any of these figures for clarity.

A

B

A

B

C

	89	99	117	172
KR422356	VTGLATIKME		EMFVSLGLGHIHELKQNKHDPEDSRVTSVTVC	DESQIEEAI
KR422354	VTGLATIKME		EMFVSLGLGHIHELKQNKHDPEDSRVRSVTVC	DESQIEEAI
KR422355	VTGLATIKME		EMFVSLGLGHIHELKQNKHDPEDSRVRSVTVC	DESQIEEAI
XM_003560467.3	VTGLATIKME		EMFVSLGLVHELKQNKHDPEDSRVRSVTVI	NESQVEEAI
AJ491855	VIGLATIQME		KMFLSFLGLVLESSKQNKHDPENRRVSSVTVC	KQSQLEEAT
AJ491820	VTGLATTQME		EMFLSFLGLVLESSKQNKHDPENRRVSSVTVC	EQSHLEEAI
GRMZM2G135674	VTGLATVEME		EVFVSVLGLVLESSRQQRQRRQDHAGRVIR-PSSVTA AVRDEPDLEEAG	
Sb10g029000	VTGLATVEME		EMFVSLGLVLESSRNRQQ--QRDHQDHSRVM-----AAA VRDEPDLEEAN	
KJ540618	VSGLS TVKME		EIFVSLGLMLRVNHQDMQDLPVKISSVPVELEVLDLANSMALCDESQLEEDAS	
KJ540616	VSGLS TVKME		EIFVSLGLMLRVNHQDMQDLPVKISSVPVELEVLDLANSMALCDESQLEEEAA	
AM000056.1	VSGLS TITME		EIFVSLGLMLRVNHQDMQDLPVKISSVPVELEEIDLANSMALSDSQLEEAT	
U16709	VSGLS TITME		EIFVSLGLMLRVNHQDMQDLPVKISSVPVELEEIDLANSMALCDESQLEEEAA	
XM_003563466.2	VSGLS TVKME		EIFVSLGLMLRVNHQDMQDLPVKISSVPVELEAIDSASSVTFCDESKLEEAT	
EF373552	VSGLS TIEME		EVFVSVLGLMLRLKHKHNPEFSGDRVSSVPIELDTIEPTRTVMSSEELQIEAAV	
AB061311	LSSLITIME		EVFVSVLGLMLRLNKHKNPEFSGDKVSSVPIELDTINSASTVISCEELQLEAAI	