Available online at www.sciencedirect.com
ScienceDirect
Plant Science 172 (2007) 708–721
PLANT SCIENCE
www.elsevier.com/locate/plantsci

Genome-wide analysis and identification of genes related to potassium transporter families in rice (*Oryza sativa* L.)

R. Naga Amrutha [a], P. Nataraj Sekhar [a], Rajeev K. Varshney [b], P.B. Kavi Kishor [a,*]

[a] *Department of Genetics, Osmania University, Hyderabad 500007, India*
[b] *International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502325, India*

Received 27 August 2006; received in revised form 14 November 2006; accepted 22 November 2006
Available online 26 December 2006

Abstract

Potassium (K^+) is an important macronutrient and the most abundant cation in higher plants which plays a key role in various cellular processes. Its accumulation from soil and its distribution throughout diverse plant tissues is mediated by transporter proteins. In plants, different transport systems are known to be involved in the uptake and release of K^+ from the cells. Though most of the information about the putative K^+ transporters and their phylogenetic relationships is available in *Arabidopsis*, it is not the best model for plants with agronomic applications. Recent completion of rice genome sequencing project offered the opportunity to make an inventory of all putative K^+ transporter proteins. More than 5% of the rice genome appears to encode membrane transport proteins. Unfortunately, several hundreds of putative transporter proteins have not yet been assigned to any families or subfamilies or functions. Therefore, phylogenetic relationships of many K^+ transporters in rice are analyzed since rice is considered as a model plant because of its high degree of co-linearity with other cereals. Phylogenetic analysis of all K^+ transporters in rice revealed that they fall into five major branches. Phylogenetic trees of each family define the evolutionary relationships of the members to each other. In each family, closely related isoforms and separate subfamilies existed, indicating possible redundancies and specialized functions. The HAK family is represented by 26 genes and formed the tightest and most distinct branch in the phylogenetic tree. Around 14 genes with conserved P-loop were found in K^+ channel family out of which 11 genes belong to 1P/6TM (Shaker-type), and three genes to the 2P/4TM (ORK-type). On searching rice genome, it was found that nine genes belonged to Trk family. In rice, K^+/H^+ antiporter family is represented by a single gene. Comparative analysis of rice K^+ channels with that of *Arabidopsis*, wheat and maize revealed that while cereals are closely related, *Arabidopsis* appeared quite distant from rice.

Keywords: Potassium; Ion transporters; Salt stress; *Oryza sativa*; Genome analysis

1. Introduction

K^+ is an important macronutrient and the most abundant cation in higher plants. This cation plays a key role in various cellular processes such as: (1) charge balancing in the cytoplasm, (2) activation of crucial enzymatic reactions and (3) a substantial contribution to the osmotic pressure of the vacuole and hence to cell turgor [1,2]. Furthermore, K^+ is necessary for phloem solute transport and maintenance of cation:anion balance in the cytosol as well as in the vacuole. Apart from these, K^+ is also involved in processes such as cell elongation, stomatal movements and regulation of gaseous exchange and the transduction of various signals [3–5]. Accumulation of K^+ from soil and its distribution throughout diverse plant tissues is mediated by transporter proteins. The ratio of K^+/Na^+ in plant cells will depend on the concerted action of transport systems located at plasma and vacuolar membranes and probably involves K^+ selective, Na^+ selective and non-selective pathways. In plants, different transport systems are known to be involved in the uptake and release of K^+ from the cells. Currently, plant genomics allow the fast identification of molecular components related to mineral homeostasis. The recent completion of Arabidopsis genome sequencing project offered the opportunity to make an inventory of all of plants putative cation transporter proteins [6]. This was possible through database searches for sequences homologous to proteins already characterized. The IRGSP, formally established in 1998, pooled the resources of sequencing groups in 10 nations to obtain a complete finished quality sequence of the rice genome (*Oryza sativa* L. ssp.

* Corresponding author. Tel.: +91 40 2768 2335; fax: +91 40 2709 5178.
E-mail address: pbkavi@yahoo.com (P.B.K. Kishor).

0168-9452/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.plantsci.2006.11.019

Japonica cv. Nipponbare). The IRGSP released a high-quality map-based draft sequence in December 2002, and the sequencing project was completed in December 2004, and the results were published in August 2005. At present, IRGSP submitted sequences to GenBank that comprise over 95% of the genome (http://www.tigr.org/tdb/e2k1/osa1/BACmapping/description.shtml). Most of the information about putative K^+ transporters that might play a role in K^+ transport came from sequence analyses and phylogenetic relationships [6]. Though functional identification and characterization of many K^+ transporters is possible in Arabidopsis, it is not the best model for plants with agronomic applications. Rice perhaps is an ideal crop having the smallest genome of the major cereals, dense genetic maps and relative ease of genetic transformation [7]. Discovery of extensive genome colinearity among the Poaceae [8] established rice as the model organism for the cereal grasses. These properties, along with the finished sequence and other tools under development, set the platform for a complete functional characterization of rice genome. A total of 37,544 protein-coding genes were identified in rice of which 71% have a putative homologue in Arabidopsis [9]. The first insight in the molecular analysis of membrane K^+ transport in plants came in 1992 with the identification of two *Arabidopsis* K^+ channels, AKT1 [10] and KAT1 [11]. Most plant K^+ channels were found to be members of Shaker family and were successfully expressed and characterized in heterologous systems. Another breakthrough in functional identification of plant potassium transporters came in the year 1994, with the cloning of HKT1 from wheat [12] that corresponds to HKT family. Members of HKT gene family act as Na^+ and Na^+/K^+ transporters in controlling Na^+ accumulation [13–17]. Based on PCR [18], *in silico* analyses [19] and functional complementation of yeast [20], one more family of plant K^+ transporters was identified. This family is named as HAK [21] or KUP/HAK/KT [6]. Further *in silico* approaches for new plant counter parts of animal K^+ channels lead to the identification of the KCO channels [22,23]. Also using *in silico* approaches, several authors [6,24–27] identified putative K^+ transporters (CNGC) and also cation transporters (LCT) that might play a role in K^+ transport in plants. The recently identified CNGC family was assumed to conduct K^+ transport in an unspecific manner [28]. Unfortunately, several of putative transporter proteins were not yet assigned to any families or subfamilies in rice. Therefore, an analysis of the genomic sequences related to K^+ transporter families in rice was carried out in the present study by searching the Japonica variety genome in public databanks. The purpose of this study is to contribute to the understanding of molecular mechanisms of K^+ transport and functional characterization of identified new K^+ transporter genes that play a major role in salt tolerance.

2. Materials and methods

Reference proteins of well established molecular function, representing each of the protein families investigated, were chosen as query sequences for searches in the rice (*O. sativa*) genome databases. These reference proteins were AtKUP1 (GenBank accession no. AAB87687), AtAKT1 (GenBank accession no. AAP21250), AtKCO1 (GenBank accession no. AAM64705), AtHKT (GenBank accession no. AAF68393) and AtKEA1 (GenBank accession no. NP_171684). Searches were made using the TBLASTN tool [29] against GenBank database non-redundant (NR), with search specifications for *O. sativa*. The other databases used were Rice Genome Research Program (RGP) (http://rgp.dna.affrc.go.jp/), The Institute of Genome Research (TIGR), rice genome annotation database (http://www.tigr.org/tdb/e2k1/osa1/index.shtml) and Universal Protein resource Uniprot (http://www.ebi.uniprot.org/uniprot-srv/protein/uniProtView). The BLAST server used was that of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/). As selection criteria of BLAST hits for genomic sequences, a cut off *e*-value of *e*-10 was previously set. The genomic sequences found were used to predict putative genes contained within them. Whenever possible, genes were predicted on the basis of sequences generated by the IRGSP, since these sequences present a higher degree of accuracy. To that end, a mixed procedure was adopted combining *ab initio* gene prediction algorithms of genomic sequence alignments with similar sequences from expressed genes (ESTs and cDNAs). The prediction algorithms were GenScan (Burge and Karlin, 1997; http:// genes.mit.edu/GENSCAN.html), GenomeScan [30]; http://genes.mit.edu/genomescan.html), FGENESH [31]; http://www.softberry.-com/berry.phtml?topic=gfind), GeneMark.hmm (Borodovsky and Lukashin, unpublished; http://opal.biology.gatech.edu/GeneMark/eukhmm.cgi) and GrailEXP [32]; http://compbio.ornl.gov/grailexp/). Such expressed sequences were found by BLAST searches against EST and NR databases of GenBank, using the genomic sequence as query. The algorithm of choice for the multiple alignments of protein sequences was ClustalW 1.8 [33], available through the BCM Search Launcher server (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html). The multiple alignments were edited with the help of GENEDOC (Free Software Foundation Inc.). All the proteins with greater than 30% identity, with at least one of the reference proteins used in the searches, were regarded as functionally similar (homologous) to the reference proteins, receiving the same name [34–37]. Those sequences that did not conform to this criterion were discarded. Only in case of OsAKT the identity degree of 26% was accepted due to the conserved functional domains between this protein and the reference proteins. Prediction of homology and signature sequences for the putative K^+ transporter proteins were carried out with PROSITE (http://www.ebi.ac.uk/InterProScan/) [38] and Pfam databases [39]. Sequences were included into families based on homology and presence of signature sequences. For topology prediction, HMMTOP [40] was used. RGI gene codes for families were obtained from http://www.gramene.org/Multi/blastview and http://tigrblast.tigr.org/euk-blast/index. Protein alignments obtained with ClustalW 1.8 [33] were used as starting points for phylogenetic analysis. Unrooted trees were prepared by the neighbor-joining method using either Clustal, PHYLIP [41], or and 1000 bootstrap replicates were performed. Bold lines on trees indicate protein