Whole-Genome Analysis of *Oryza sativa* Reveals Similar Architecture of Two-Component Signaling Machinery with Arabidopsis [1][W]

Ashwani Pareek2*, Anupama Singh2, Manoj Kumar, Hemant R. Kushwaha, Andrew M. Lynn, and Sneh L. Singla-Pareek

Stress Physiology and Molecular Biology Laboratory, School of Life Sciences (A.P., M.K.), and Centre for Computational Biology and Bioinformatics (A.S., H.R.K., A.M.L.), Jawaharlal Nehru University, New Delhi 110067, India; and Plant Molecular Biology, International Centre for Genetic Engineering and Biotechnology, New Delhi 110 067, India (S.L.S-P.)

The two-component system (TCS), which works on the principle of histidine-aspartate phosphorelay signaling, is known to play an important role in diverse physiological processes in lower organisms and has recently emerged as an important signaling system in plants. Employing the tools of bioinformatics, we have characterized TCS signaling candidate genes in the genome of *Oryza sativa* L. subsp. japonica. We present a complete overview of TCS gene families in *O. sativa*, including gene structures, conserved motifs, chromosome locations, and phylogeny. Our analysis indicates a total of 51 genes encoding 73 putative TCS proteins. Fourteen genes encode 22 putative histidine kinases with a conserved histidine and other typical histidine kinase signature sequences, five phosphotransfer genes encoding seven phosphotransfer proteins, and 32 response regulator genes encoding 44 proteins. The variations seen between gene and protein numbers are assumed to result from alternative splicing. These putative proteins have high homology with TCS members that have been shown experimentally to participate in several important physiological phenomena in plants, such as ethylene and cytokinin signaling and phytochrome-mediated responses to light. We conclude that the overall architecture of the TCS machinery in *O. sativa* and *Arabidopsis thaliana* is similar, and our analysis provides insights into the conservation and divergence of this important signaling machinery in higher plants.

---

1 This work was supported by the Government of India (Department of Science and Technology and Department of Biotechnology), by the International Foundation for Science (Sweden), by the International Atomic Energy Agency (Austria), and by the Council of Scientific and Industrial Research, India (Junior Research Fellowship to A.S.).

2 These authors contributed equally to the paper.

* Corresponding author; e-mail ashwanip@mail.jnu.ac.in; fax 91–11–26704504.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Ashwani Pareek (ashwanip@mail.jnu.ac.in).

[W] The online version of this article contains Web-only data. www.plantphysiol.org/cgi/doi/10.1104/pp.106.086371
carried out comprehensive analysis of the available similar system of Arabidopsis. Additionally, we have

Figure 1. Basic machinery of TCS. A, A simple TCS member. Sensing of an extracellular signal is initiated by the input domain of sensory HK, phosphorylating the conserved His in its transmitter domain. In the next step, the conserved Asp in the RD of the RR is autophosphorylated, resulting in the signal output. B, A hybrid-type TCS in which the conserved His and Asp are found in the same protein, which serves as the sensory HK and is usually membrane bound. The nonmembrane-bound Hpt acts as a mediator for the transfer of the phosphoryl group between the HK and the RR. The TM domains are depicted as vertical bars.

cerevisiae), which controls the HOG pathways in response to osmotic stress (Wurglers-Murphy and Saito, 1997). Such a multistep phosphorelay reaction is believed to have the advantage of providing multiple regulatory checkpoints for signal cross talk or negative regulation by certain phosphatases. The existence of multiple phosphorelay reactions in both prokaryotes and eukaryotes suggests that similar mechanisms might be more widely used (Urso et al., 2000a).

Recently, the TCS signaling machinery has been shown to be present in several plant species, including Arabidopsis (Arabidopsis thaliana; Hwang and Sheen, 2001; Grefen and Harter, 2004). In Arabidopsis, TCS members have been documented to be involved in cytokinin signaling, ethylene signaling, and light perception (Hwang et al., 2002). A member of the family, Arabidopsis HK1 (AtHK1), has been shown to work as a putative osmosensor in the Arabidopsis (Urso et al., 1999). However, the physiological function of most TCS members remains to be elucidated.

With the availability of whole-genome sequences of Oryza sativa and Arabidopsis, we decided to use the tools of bioinformatics to examine the key components of the TCS pathways that are operative in these two model systems. Our analysis is based on the high-quality, finished (The Institute for Genomic Research [TIGR] version 4.0) O. sativa genome sequence (International Rice Genome Sequencing Project, 2005). Genome-wide analyses for TCS signal transduction genes have previously been carried out for fungal pathogens (Catlett et al., 2003) and Arabidopsis (Hwang et al., 2002). In this article, we provide information on the TCS machineries operative in O. sativa and compare them with the similar system of Arabidopsis. Additionally, we have carried out comprehensive analysis of the available massively parallel signature sequencing (MPSS) data as well as the Tos17 database to comment on the expression profile and possible role(s) of the O. sativa TCS members, respectively.

RESULTS AND DISCUSSION

To avoid any ambiguity in nomenclature, we have deliberately opted for naming the various TCS proteins as 1, 2, and 3. In cases where prior information is already available in literature about any of the O. sativa TCS members, the nomenclature has been maintained. However, in case a changed name has been suggested for uniformity purpose, the existing name is also mentioned in parentheses. Further, in some cases, the postscript a, b, or c has been added to represent the products of alternative splicing, as reported by TIGR 4.0 protein pseudo molecule database. Nomenclature of genes has been based on the longest predicted open reading frame (ORF) followed by the number added to the protein. Since the products of alternate spliced genes may also be functionally diverse, we have considered them to represent a separate entity in our multiple sequence alignments; however, for phylogenetic analysis, we have considered only the ones that encode the longest reading frame.

Two-Component Analysis in O. sativa

Our analysis has been carried out with the high-quality, finished O. sativa genome sequence using the HMMER software (for details, see “Materials and Methods”). All the significant hits with positive scores were selected for analysis based on conserved domain search. Initially, we validated the sensitivity and specificity of our methodology using the Arabidopsis genome and compared the output with published reports. We found our results to be in total conformity with those reported by Hwang et al. (2002). Arabidopsis has at least 54 distinct TCS member genes (Hwang et al., 2002). Interrogation of the O. sativa genome identified 51 genes, encoding 73 putative proteins (Fig. 2; Tables I–III), having strong homology with HKs (OsHKs), Hpt proteins (OsHpts), and RRss (OsRRs).

The HK Protein Family

O. sativa has at least 14 distinct HKs compared to 16 such genes in Arabidopsis (Hwang et al., 2002). Further, these putative HK homologs in O. sativa belong to three distinct families: the typical OsHK family (that includes cytokinin receptor homologs), an ethylene receptor (ETR) homolog family (that includes OsETRs and OsERRs), and the phytochromes (OsPHYs). Table I summarizes the relevant genomic information for these genes and their products. The predicted proteins range in size from 359 to 1,186 amino acids, thus representing great variations in structure and possible functions. However, the OsHK members share an overall 42% to 77% similarity in amino acid sequence and 30% to 66%
identity, excluding the PHY proteins. The identity of PHY proteins with other OsHKs varies from 25% to 27%, while similarity varies between 35% and 41%.

OsHK Family

This family contains at least seven typical hybrid type HKs [OsHK1, 2, 3 (a and b), 4, 5, and 6]. To verify the number of HK sequences, an exhaustive TBLASTN search of full-length O. sativa cDNA clones available at Knowledge-based Oryza Molecular biological Encyclopedia (KOME) and a TBLASTX search of the National Center for Biotechnology Information (NCBI) O. sativa genome dataset were also performed, using all seven OsHKs as the query. No additional HK member was found in these searches. The structural details of the OsHKs are presented in Figure 2A, while the sequence details are compared in Figure 3, A (for transmitter domain) and B (for RD). The primary domain analysis of these proteins confirmed them to have a typical hybrid HK-type structure with a conserved His-containing kinase domain (KD) and a complete RD with a highly conserved Asp as the phospho-acceptor, although the number and position of the transmembrane (TM)
domain varied (Fig. 2A). All OsHK family members (except OsHK1 and OsHK2) have the extracellular domain (ED), which may be involved in sensing signals. Further, in the case of OsHK1 and OsHK3b, no putative TM domain could be predicted with any significance, while the rest of the members have one or two TM domains, possibly indicating the diverse roles of these members in membrane-associated or nonmembrane-associated sensing operations in plants. A similar variable number of TM domains have been reported in Arabidopsis (Hwang et al., 2002).

The five conserved signature motifs, such as H, N, G1, F, and G2, also feature in OsHKs (Fig. 3A). The transmitter domain of HKs has an H box that contains the conserved His residue, which is reversibly phosphorylated in the signaling process. The highly invariable residue Pro, after five residues of His, is conserved across all OsHKs except OsHK2. Further, His has not been found to be conserved in OsETR2, ETR3a, ETR3b, OsERS3, OsERS4, OsPHYB, OsPHYCa, and OsPHYCb. However, this His residue is found to be displaced further down at position 7 on all these members except the OsPHY. Experimental evidence for the participation of this displaced His in phosphorylation is needed. In most cases, a conserved D is missing in the G1 box when the His is missing in the H box.

It is now known that members of the TCS (such as AHK2, 3, and CRE1/Ahk4/WOL) are responsible for

**Table I. Putative HK and related proteins in *O. sativa***

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein Notation</th>
<th>TIGR ID</th>
<th>BAC</th>
<th>Accession No.</th>
<th>Gene Locus</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>OsHK1</td>
<td>OsHK1</td>
<td>11976.m08969</td>
<td>P0709F06</td>
<td>AP003579</td>
<td>LOC_Os06g4410</td>
<td>968</td>
</tr>
<tr>
<td>OsHK2</td>
<td>OsHK2</td>
<td>11976.m05564</td>
<td>P0592C05</td>
<td>AP004672</td>
<td>LOC_Os01g69920</td>
<td>866</td>
</tr>
<tr>
<td>OsHK3</td>
<td>OsHK3a</td>
<td>11971.m13041</td>
<td>P0592C05</td>
<td>AP004672</td>
<td>LOC_Os01g69920</td>
<td>866</td>
</tr>
<tr>
<td>OsHK3b</td>
<td>OsHK3b</td>
<td>11971.m42798</td>
<td>P0592C05</td>
<td>AP004672</td>
<td>LOC_Os01g69920</td>
<td>866</td>
</tr>
<tr>
<td>OsHK4</td>
<td>OsHK4</td>
<td>11973.m10071</td>
<td>OSJNBa0078A17</td>
<td>AC091532</td>
<td>LOC_Os03g50860</td>
<td>1,013</td>
</tr>
<tr>
<td>OsHK5</td>
<td>OsHK5</td>
<td>11980.m05192</td>
<td>OSJNBa0073L01</td>
<td>AC092548</td>
<td>LOC_Os10g21810</td>
<td>1,186</td>
</tr>
<tr>
<td>OsHK6</td>
<td>OsHK6</td>
<td>11972.m10082</td>
<td>P0684F11</td>
<td>AP005112</td>
<td>LOC_Os02g50480</td>
<td>997</td>
</tr>
<tr>
<td>OsETR2</td>
<td>OsETR2</td>
<td>11974.m06161</td>
<td>OSJNBb0089K06</td>
<td>AL663013</td>
<td>LOC_Os04g08740</td>
<td>763</td>
</tr>
<tr>
<td>OsETR3</td>
<td>OsETR3a</td>
<td>11972.m10774</td>
<td>QJ1119_A01</td>
<td>AP004200</td>
<td>LOC_Os02g57530</td>
<td>770</td>
</tr>
<tr>
<td>OsETR3b</td>
<td>OsETR3b</td>
<td>11972.m33481</td>
<td>QJ1119_A01</td>
<td>AP004200</td>
<td>LOC_Os02g57530</td>
<td>836</td>
</tr>
<tr>
<td>OsERS1</td>
<td>OsERS1a</td>
<td>11973.m35115</td>
<td>OSJNBb0089K06</td>
<td>AL663013</td>
<td>LOC_Os04g08740</td>
<td>763</td>
</tr>
<tr>
<td>OsERS1b</td>
<td>OsERS1b</td>
<td>11973.m35115</td>
<td>QJ1119_A01</td>
<td>AP004200</td>
<td>LOC_Os02g57530</td>
<td>836</td>
</tr>
<tr>
<td>OsERS2</td>
<td>OsERS2a</td>
<td>11975.m05159</td>
<td>P0431G05</td>
<td>AC087551</td>
<td>LOC_Os05g06320</td>
<td>635</td>
</tr>
<tr>
<td>OsERS2b</td>
<td>OsERS2b</td>
<td>11975.m27546</td>
<td>P0431G05</td>
<td>AC087551</td>
<td>LOC_Os05g06320</td>
<td>359</td>
</tr>
<tr>
<td>OsERS2c</td>
<td>OsERS2c</td>
<td>11975.m27546</td>
<td>P0431G05</td>
<td>AC087551</td>
<td>LOC_Os05g06320</td>
<td>518</td>
</tr>
<tr>
<td>OsPHYA</td>
<td>OsPHYA</td>
<td>11973.m10085</td>
<td>B1377B10</td>
<td>AC147426</td>
<td>LOC_Os03g51030</td>
<td>1,128</td>
</tr>
<tr>
<td>OsPHYB</td>
<td>OsPHYB</td>
<td>11973.m07375</td>
<td>OSJNBb0042B10</td>
<td>AC137071</td>
<td>LOC_Os03g19590</td>
<td>1,171</td>
</tr>
<tr>
<td>OsPHYC</td>
<td>OsPHYC</td>
<td>11973.m10374</td>
<td>QJ1112_G08</td>
<td>AC133225</td>
<td>LOC_Os03g54084</td>
<td>1,137</td>
</tr>
<tr>
<td>OsPHYCb</td>
<td>OsPHYCb</td>
<td>11973.m78893</td>
<td>QJ1112_G08</td>
<td>AC133225</td>
<td>LOC_Os03g54084</td>
<td>957</td>
</tr>
</tbody>
</table>

*Notation of putative proteins based on conserved domains.  

**Table II. Hpt proteins in *O. sativa***

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein Notation</th>
<th>TIGR ID</th>
<th>BAC</th>
<th>Accession No.</th>
<th>Gene Locus</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>OsHpt1</td>
<td>OsHpt1</td>
<td>11971.m11569</td>
<td>OJ1150_A11</td>
<td>AP003928</td>
<td>LOC_Os01g54050</td>
<td>208</td>
</tr>
<tr>
<td>OsHpt2</td>
<td>OsHpt2</td>
<td>11978.m08408</td>
<td>P0439E07</td>
<td>AP003768</td>
<td>LOC_Os08g44350</td>
<td>147</td>
</tr>
<tr>
<td>OsHpt3</td>
<td>OsHpt3</td>
<td>11979.m06878</td>
<td>OJ1155_H10</td>
<td>AP005558</td>
<td>LOC_Os09g34000</td>
<td>149</td>
</tr>
<tr>
<td>OsHpt4</td>
<td>OsHpt4</td>
<td>11975.m05454</td>
<td>QJ1725_E07</td>
<td>AC093920</td>
<td>LOC_Os09g94140</td>
<td>150</td>
</tr>
<tr>
<td>OsHpt5</td>
<td>OsHpt5a</td>
<td>11975.m64188</td>
<td>P0483D07</td>
<td>AC130611</td>
<td>LOC_Os05g44570</td>
<td>152</td>
</tr>
<tr>
<td>OsHpt5b</td>
<td>OsHpt5b</td>
<td>11975.m08583</td>
<td>P0483D07</td>
<td>AC130611</td>
<td>LOC_Os05g44570</td>
<td>104</td>
</tr>
<tr>
<td>OsHpt5c</td>
<td>OsHpt5c</td>
<td>11975.m64253</td>
<td>P0483D07</td>
<td>AC130611</td>
<td>LOC_Os05g44570</td>
<td>96</td>
</tr>
</tbody>
</table>

*Notation of putative proteins based on conserved domains.  

Two-Component Signal Transduction Genes of *Oryza sativa*
cytokinin signaling (Kakimoto, 1996; Hwang and Sheen, 2001). Further, the Arabidopsis mutants (e.g. cre1-1 mutant) lacking these members have been reported to lack cytokinin responses (Inoue et al., 2001). Our analysis indicates that the putative cytokinin receptor candidates in *O. sativa* are OsHK3a, 3b, 4, 5, and 6, which show high identity (50%–62%) with the established cytokinin receptors AHK2, AHK3, and CRE1 from Arabidopsis. However, OsHK3 to 6 show a low level of similarity (28%–30%) with CKI1. This is in agreement with other reports where sequence divergence of CKI1 (only 25% identity with CRE1) has been implicated to be the possible reason for failure of cytokinin binding to CKI1 (Yamada et al., 2001).

<table>
<thead>
<tr>
<th>Table III. RR and RR-like proteins in <em>O. sativa</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Genes identified using HMMER search; nomenclature based on domain conservation. Putative alternate spliced protein products have postscript a, b, or c.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein Notation</th>
<th>TIGR ID</th>
<th>BAC</th>
<th>Accession No.</th>
<th>Gene Locus</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-type RR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OsRRA1</td>
<td>OsRRA1a</td>
<td>11981.m04670</td>
<td>OSJNBo059j06</td>
<td>AC123526</td>
<td>LOC_Os11g04720</td>
<td>201</td>
</tr>
<tr>
<td>OsRRA2</td>
<td>OsRRA1b</td>
<td>11981.m28648</td>
<td>OSJNBo059j06</td>
<td>AC123526</td>
<td>LOC_Os11g04720</td>
<td>208</td>
</tr>
<tr>
<td>OsRRA3</td>
<td>OsRRA2</td>
<td>11982.m04444</td>
<td>OSJNBo034E23</td>
<td>AL928752</td>
<td>LOC_Os12g04500</td>
<td>201</td>
</tr>
<tr>
<td>OsRRA4</td>
<td>OsRRA3</td>
<td>11971.m13268</td>
<td>P0431G06</td>
<td>AP003683</td>
<td>LOC_Os10g2330</td>
<td>232</td>
</tr>
<tr>
<td>OsRRA5</td>
<td>OsRRA4</td>
<td>11974.m09387</td>
<td>OSJNBo065J09</td>
<td>AL663010</td>
<td>LOC_Os04g4280</td>
<td>134</td>
</tr>
<tr>
<td>OsRRA6</td>
<td>OsRRA5</td>
<td>11977.m06949</td>
<td>O1104_A06</td>
<td>AP003802</td>
<td>LOC_Os07g26720</td>
<td>206</td>
</tr>
<tr>
<td>OsRRA7</td>
<td>OsRRA6</td>
<td>11974.m10657</td>
<td>OSJNBo018M05</td>
<td>AL666457</td>
<td>LOC_Os04g57720</td>
<td>170</td>
</tr>
<tr>
<td>OsRRA8</td>
<td>OsRRA7</td>
<td>11972.m10856</td>
<td>O1112_D06</td>
<td>AP004043</td>
<td>LOC_Os02g58350</td>
<td>131</td>
</tr>
<tr>
<td>OsRRA9</td>
<td>OsRRA8</td>
<td>11972.m09243</td>
<td>B1250G12</td>
<td>AP006452</td>
<td>LOC_Os02g42060</td>
<td>147</td>
</tr>
<tr>
<td>OsRRA10</td>
<td>OsRRA9</td>
<td>11974.m08696</td>
<td>P0076O17</td>
<td>BX548156</td>
<td>LOC_Os04g36070</td>
<td>231</td>
</tr>
<tr>
<td>OsRRA11</td>
<td>OsRRA10</td>
<td>11972.m08606</td>
<td>OSJNBo0038F20</td>
<td>AP005808</td>
<td>LOC_Os02g35180</td>
<td>252</td>
</tr>
<tr>
<td>OsRRA12</td>
<td>OsRRA11</td>
<td>11978.m06899</td>
<td>O11705_C03</td>
<td>AP003962</td>
<td>LOC_Os08g28950</td>
<td>167</td>
</tr>
<tr>
<td>OsRRA13</td>
<td>OsRRA12</td>
<td>11978.m06712</td>
<td>P0488B06</td>
<td>AP005523</td>
<td>LOC_Os08g26990</td>
<td>135</td>
</tr>
<tr>
<td>OsRRA14</td>
<td>OsRRA13</td>
<td>11978.m06895</td>
<td>O11705_C03</td>
<td>AP003962</td>
<td>LOC_Os08g28900</td>
<td>185</td>
</tr>
<tr>
<td>OsRRA15</td>
<td>OsRRA14</td>
<td>11973.m10284</td>
<td>O1136_D05</td>
<td>AC096855</td>
<td>LOC_Os03g3100</td>
<td>127</td>
</tr>
<tr>
<td>OsRRA16</td>
<td>OsRRA15</td>
<td>11973.m10284</td>
<td>O1136_D05</td>
<td>AC096855</td>
<td>LOC_Os03g3100</td>
<td>127</td>
</tr>
<tr>
<td>OsRRA17</td>
<td>OsRRA16</td>
<td>11974.m06587</td>
<td>OSJNBo083</td>
<td>AL662982</td>
<td>LOC_Os04g13480</td>
<td>132</td>
</tr>
<tr>
<td>OsRRA18</td>
<td>OsRRA17</td>
<td>11975.m07520</td>
<td>OSJNBo092G21</td>
<td>AC134932</td>
<td>LOC_Os05g32880</td>
<td>620</td>
</tr>
<tr>
<td>OsRRA19</td>
<td>OsRRA18</td>
<td>11974.m07929</td>
<td>OSJNBo020I02</td>
<td>AL662963</td>
<td>LOC_Os04g28120</td>
<td>252</td>
</tr>
<tr>
<td>OsRRA20</td>
<td>OsRRA19</td>
<td>11978.m10556</td>
<td>P0493A04</td>
<td>AP004586</td>
<td>LOC_Os08g35670</td>
<td>614</td>
</tr>
<tr>
<td>B-type RR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OsRRB1</td>
<td>OsRRB1a</td>
<td>11973.m06685</td>
<td>OSJNBo081P02</td>
<td>AC107226</td>
<td>LOC_Os03g12350</td>
<td>691</td>
</tr>
<tr>
<td>OsRRB2</td>
<td>OsRRB1b</td>
<td>11973.m15102</td>
<td>OSJNBo081P02</td>
<td>AC107226</td>
<td>LOC_Os03g12350</td>
<td>691</td>
</tr>
<tr>
<td>OsRRB3</td>
<td>OsRRB1c</td>
<td>11973.m15103</td>
<td>OSJNBo081P02</td>
<td>AC107226</td>
<td>LOC_Os03g12350</td>
<td>691</td>
</tr>
<tr>
<td>OsRRB4</td>
<td>OsRRB2</td>
<td>11972.m06196</td>
<td>O11297_C09</td>
<td>AP004087</td>
<td>LOC_Os02g08500</td>
<td>626</td>
</tr>
<tr>
<td>OsRRB5</td>
<td>OsRRB3</td>
<td>11976.m08921</td>
<td>OSJNBo062J02</td>
<td>AP003517</td>
<td>LOC_Os06g43910</td>
<td>694</td>
</tr>
<tr>
<td>OsRRB6</td>
<td>OsRRB4</td>
<td>11976.m05563</td>
<td>OSJNBo036B04</td>
<td>AP007226</td>
<td>LOC_Os06g08440</td>
<td>696</td>
</tr>
<tr>
<td>OsRRB7</td>
<td>OsRRB5</td>
<td>11972.m10558</td>
<td>O1104_E04</td>
<td>AP003975</td>
<td>LOC_Os02g53520</td>
<td>688</td>
</tr>
<tr>
<td>OsRPR1</td>
<td>OsRRB6</td>
<td>11971.m12881</td>
<td>OSJNCo013M08</td>
<td>AP006838</td>
<td>LOC_Os01g77770</td>
<td>518</td>
</tr>
<tr>
<td>OsRPR2</td>
<td>OsRRB7</td>
<td>11974.m07930</td>
<td>OSJNBo020I02</td>
<td>AL662963</td>
<td>LOC_Os04g28130</td>
<td>390</td>
</tr>
</tbody>
</table>

**Notes:**
- a: Notation of putative proteins based on conserved domains.
- b: IDs given to pseudomolecules by TIGR (http://www.tigr.org/tdb/e2k1/osa1/) from the *O. sativa* annotation database.
- c: BAC IDs from the TIGR database.
- e: Gene locus.
- f: The number of predicted amino acid.
OsHK1 and OsHK2, all OsHK members were found to contain the conserved cyclases/HK-associated sensory extracellular (CHASE) domain. Qiu-min et al. (2004) showed that, in *O. sativa*, five proteins (OsCRL1a, OsCRL1b, OsCRL2, OsCRL3, and OsCRL4) contained the CHASE domain, but only one of these proteins (OsCRL4) contained, in addition, the putative Ser/Thr protein KD (please note that OsCRL2, OsCRL1b, OsCRL3, and OsCRL1a have been referred to as OsHK3, OsHK4, OsHK5, and OsHK6, respectively, in this study). It has been suggested that OsCRL4 represents a new AtCRE1-like gene family in *O. sativa* and was shown to be able to functionally complement the phenotype of *Atcre1* mutation, although the structure of the two were quite different (Qiu-min et al., 2004). Similarly, in the case of maize (*Zea mays*), detailed characterization of the three cytokinin-responsive HKs (ZmHK1, ZmHK2, and ZmHK3a) suggested that each has a capacity to confer cytokinin inducibility to lacZ expression in bacteria having ΔrcsC and ΔrcsC and

Two-Component Signal Transduction Genes of *Oryza sativa*
cps::lacZ. This indicates that the cytokinin signaling may be conserved machinery in monocots, though some minor differences do exist with respect to their ligand binding (Yoneyama-Sakakibara et al., 2004).

ETR Family

In Arabidopsis, ethylene is perceived by a family of five receptors (ETR1, ETR2, ERS1, ERS2, and Ein4) that display significant homology to a prokaryotic family of signal transducers (Chang et al., 1993; Hua et al., 1995, 1998; Hua and Meyerowitz, 1998; Sakai et al., 1998). In O. sativa, OsETR and OsERS are members of the putative ETR family of HKs with or without a RD (Fig. 2). There are five genes that encode three OsETRs and four OsERS. Yau et al. (2004) showed, in O. sativa L. subsp. Indica var IR36, the existence of at least five ETR genes: Os-ERS1, Os-ERS2, Os-ETR2, Os-ETR3, and Os-ETR4. In our analysis, OsERS1a represents the alternate splice product for OsERS1b, while OsERS2a represents the same for OsERS2b and OsERS2c (Table I). Similarly, OsETR3a and OsETR3b are also reported in databases as alternative spliced products (http://www.tigr.org/tdb/e2k1/osa1/expression/alt_spliced.info.shtml). The primary domain analysis of these proteins indicated that ETRs OsETR2, 3 (a and b), and OsETR4 have a hybrid HK-type structure, with their KDs having the conserved His and their RD having a highly conserved Asp as a phospho-acceptor (Fig. 2).
All the ERS members of *O. sativa* lacked the RD domain; OsERS3 lacked both the RD and receiver-like domain (RLD). All the *O. sativa* ETR members were predicted to possess two or more TM domains (except the OsERS2b) along with the GAF (cyclic GMP adenyl cyclase FlhA) domain. GAF domains bind molecules such as cAMP and cGMP. ETRs and phytochromes (PHYs) are the only plant proteins known to contain GAF and HK-related domains. Various ETRs of *O. sativa*, along with their alternate splice products, have been aligned and compared in Figure 3A and show that the G2 and F domains are missing in OsERS2c. The autophosphorylating His residue is absent at the usual conserved position in OsERS1, OsERS4, OsETR2, OsETR3a, and OsETR3b, but is shifted six amino acids towards the C terminus. It is not known if this affects activity of HKs. Arabidopsis has a similar situation where H is replaced by E (in ETR2) and D (in ERS2), and none of the PHYs except PHYC has conserved H (Hwang et al., 2002). OsETR4 presents a unique situation where H is present at both positions; which of these participates in the phosphorelay needs to be established. An analysis of the RD domain of the ETR family is presented in Figure 3B. The conserved D, which is essential for the phosphorelay, is replaced by Q in all the ETRs. Further, the second D (five positions down to the first one) is also replaced by N in two of these ETR members (OsETR3a and b). Also, the PIMT motif in the RD in all the OsETRs as well as in the OsHK5 shows deletion of all six residues. The functional significance of these structural changes is not known.

**PHY Family**

The PHY photoreceptors regulate growth and development in response to red and far-red light signals (Smith, 2000). In Arabidopsis, this family consists of PHYA, B, C, D, and E (Sharrock and Quail, 1989; Clack et al., 1994). All dicots seem to be similar, but monocots seem to have only PHYA, B, and C (Mathews and Sharrock, 1996, 1997). In Arabidopsis, *phya, b, d*, or *e* mutants have been instrumental in deciphering the specific roles of the individual proteins in photomorphogenesis (Whitelam and Devlin, 1997; Devlin et al., 1998). PHYs have been identified in tobacco (*Nicotiana tabacum*), tomato (*Solanum lycopersicum*), oat (*Avena sativa*), maize, and potato (*Solanum tuberosum*), but relatively very few PHY genes have been fully characterized from monocots (Basu et al., 2000). Our analysis identified three PHY genes (*OsPHYA, OsPHYB*, and *OsPHYC*) in *O. sativa*. Among these, *OsPHYC* encodes alternatively spliced products for *OsPHYCa* and b. The proteins encoded by the *OsPHY* members vary in length between 957 and 1,171 amino acids (Table I). The putative domain analysis of the OsPHY family members show a structure similar to that in Arabidopsis, i.e. having a GAF domain, light-sensing domain, duplicated PAS [Per (period circadian protein), Arnt (receptor nuclear translocator protein), Sim (single-minded protein)] domains, and RLD (Fig. 2). PAS domains are present in many signaling proteins in archaea, bacteria, and eukaryotes, and act as signal sensor domains. Both RLD and KD seem to be missing from OsPHYC; the functional significance of this is unknown (Fig. 2A). Analysis of OsPHYs using pfam (version 20.0) has clearly indicated toward presence of an additional PAS domain before GAF domain that has not been found in Arabidopsis. These minor structural and functional variations may have arisen through the gene duplication events and modification of gene functions as needed in evolution. However, experiments where the PHYs have been expressed in *O. sativa* from other systems, such as oat, have indicated toward the functional conservation of the transduction pathway components and the promoter sequences involved in regulation of PHY expression between the two systems (Bruce et al., 1989; Clough et al., 1995).

The unrooted relationship tree of the HK and the related proteins in *O. sativa* and Arabidopsis indicates the presence of subgroups within the protein subfamilies (Fig. 3C). This analysis indicates that OsHK3 to 6, AtHK2 and 3, as well as CRE1 have a close relationship with established cytokinin receptors, while the putative ETR proteins of *O. sativa* and Arabidopsis form a distinct cluster. In Arabidopsis, Hwang et al. (2002) have reported that CRE is close to the HK genes. Within the ETR family, two separate branches have evolved, one comprising ERS proteins and the other ETRs. However, ETR1 and ERS2 seem to be clear exceptions in this regard. It seems, from this tree, that the ERSs and ETRs of *O. sativa* and Arabidopsis have remained more heterogeneous within the ETR subfamily unlike the PHY and the cytokinin receptor families. Hence, the “in group” heterogeneity is higher in the ETR subgroup when compared to the other two. In the tree, all the PHYs from both *O. sativa* and Arabidopsis form a distinct group. The Arabidopsis PHYE remains in a distinct branch because *O. sativa* has no reported PHYE. OsPHYB appears to be more distant in evolution from Arabidopsis PHYB and is closer to the Arabidopsis PHYD. Perhaps the OsPHYB evolved before the evolution of PHYD in dicots as in Arabidopsis.

In Arabidopsis, a specific member of the HK group has been documented to play an important role as an osmosensor. The gene *AtHK1* (or *AHK1*), showing high structural similarity with the yeast osmosensing HK *SLN1* (synthetic lethal of N-end rule), is able to functionally complement the yeast double mutant lacking its two osmosensors, i.e. *shl1* and *sho* (Urao et al., 1999), indicating the functional conservation of the osmosensing machinery between the two systems. Further, employing yeast two-hybrid analysis, *AtHK1* has also been shown to interact physically with AHP2 (Urao et al., 2000b), suggesting that the transduction of the stress signal could occur via a multistep phosphorelay. Moreover, the *AtHK1* transcript has been found to be most abundant in roots and is up-regulated by external stresses.
osmolarity changes. Nonetheless, the precise mechanism of the functioning of the osmosensor in Arabidopsis is yet to be deciphered. Our analysis of the *O. sativa* genome has indicated that the overall percent identity of OsHKs with AtHK1 ranges from 25% to 29% with OsHK2 being the closest relative of AtHK1 (Fig. 3C). It remains to be proven experimentally which of the *O. sativa* HKs functions as an osmosensor.

Genes encoding TCS members were distributed over the 12 different chromosomes of *O. sativa* (Fig. 6). The largest number (nine) of the TCS members is located on chromosome II, followed by eight each on chromosomes III and IV, while there is only one on chromosomes X and XII. Genes for OsHK family members show the following distribution: chromosome I (OsHK3), II (OsHK6, OsETR3), III (OsPHYB, OsERS1, OsHK4, OsPHYA, OsPHYC), IV (OsETR2), V (OsERS2), VI (OsHK1 and 2), VII (OsETR4), and X (OsHK5). These distributions highlight some interesting aspects. All the PHY genes are clustered on chromosome III; two of them, *OsPHYA* and *OsPHYC*, are placed within 1 Mb of each other. In a previous observation, *PHYA* and *PHYC* were located on long arm, while *PHYB* had been located on the short arm of chromosome III of *O. sativa* var IR36 (Basu et al., 2000). This is in synteny with the three PHY genes from sorghum (*Sorghum bicolor*) that are located on the same linkage group (referred to as a linkage group C by Paterson et al. [1995] and as a linkage group A by Childs et al. [1997]). In contrast, PHY genes are distributed on chromosomes I, V, VI, and IX in maize and on chromosomes I, II, and V in Arabidopsis (Hwang et al., 2002).

The Hpt Protein Family

In Arabidopsis, five putative Hpt proteins (AHP1–5) might mediate the transfer of phosphoryl groups from membrane-localized HKs to nuclear-localized RRs (Hwang et al., 2002). Hwang and Sheen (2001) showed the movement of an Hpt into the nucleus by green fluorescent protein tag fusion. The analysis of the OsHpts in *O. sativa* indicated five genes (OsHpt1–5) coding for seven putative Hpts (OsHpt5 has three alternate splice products: OsHpt5a, b, and c) ranging in length from 96 to 208 amino acids (Table II). Different members of the OsHpt family lack overall sequence similarity but share a common four-helix bundle motif without any TM domain indicating their cytoplasmic location (Fig. 2B). All Hpt proteins of Arabidopsis contain a consensus sequence of XHQXKGSSXS with the position of H conserved. However, in *O. sativa*, only OsHpt2 and OsHpt13 show this consensus; all others lack it (Fig. 4A). The OsHpt5a, 5b, and 5c have a conserved H one position before the...
conserved H in OsHpts 2 and 3. These putative OsHpts lacking H but showing conserved Hpt domain may have some alternative degenerated phosphorelay role in cell signaling. The unrooted relationship tree for these different OsHpts is shown in Figure 4B, indicating the absence of any clustering among them. However, it can be seen that all AHP members (except AHP4 and 6) make a group that is separate from OsHpts, possibly indicating their diverse evolutionary pattern. The overall percentage identity within OsHpt family ranges from 37% to 71% and similarity varies from 64% to 84%. These putative OsHpts (1–5) are located on chromosomes I, V, VIII, and IX (Fig. 6).

The RR Family

In most systems, RR represent the terminal component of the TCS pathway, functioning as phosphorylation-activated switches that catalyze the phosphoryl transfer from the phospho-His of the HK to a conserved His in its own regulatory domain. RR can catalyze phosphoryl transfer independent of assistance from an HK because small molecules such as acetyl phosphate, carbamoyl phosphate, imidazole phosphate, and phosphoramide can serve as phosphodonorsto them (Lukat et al., 1992). The autodephosphorylation of RR is also known, thus limiting the lifetime of their activated state.

Genome-wide analysis in Arabidopsis has indicated the existence of 32 genes encoding putative RR and related proteins that are not fused to the His protein KD (Hwang et al., 2002; Schaller et al., 2002). The majority of these RR have two main domains: a conserved N-terminal regulatory domain and a variable C-terminal effector domain. *O. sativa* has a total of 32 genes that code for 44 putative OsRR proteins (Table III). Based on the predicted structural features shown in Figure 2C, the OsRR proteins in *O. sativa* can be divided into three distinct categories similar to those in Arabidopsis: OsRRAs, which possess only a RD; OsRRBs, containing RD fused to a DNA-binding domain; and OsPRRs, which are RR-like proteins or pseudo-RRs (PRRs; sometimes also referred to as RRLs).

*O. sativa* possesses 20 genes coding for the A-type RR (Table III, OsRR1 through OsRR19 and OsRR22; note that OsRR20 and 21 are encoded by OsPRR3 and OsPRR1, respectively) in contrast to 11 such genes in Arabidopsis. The identity among the A-type RR is between 26% to 90% and similarity 48% to 93%. All OsRRA-type RR have conserved DDK or EDK motifs as in Arabidopsis. A-type RR are involved in cytokinin signaling. The expression of some RR has been reported to be sensitive to environmental stresses such as drought, salinity, and low temperature (Urao et al., 1998), indicating a possible link between stresses and cytokinin signaling.

In *O. sativa*, B-type RR (OsRRBs) share 27% to 94% identity and all are more than 48% similar. They show weak relationships with A-type RR. The B-type RR have a highly conserved C-terminal stretch of around 70 amino acids. This stretch encodes for protein with high structural similarity to the DNA-binding domain of the c-MYB (Figs. 2C and 5B). There are seven B-type RR genes in *O. sativa*, compared to 12 in Arabidopsis. *OsRRB1* encodes three proteins, OsRRB1a, b, and c. The alignment of the putative RD and DNA-binding motifs of the B-type RR is shown in Figure 5, A and B, respectively. Detailed investigations on a B-type RR, *Ehd1 (Early heading date 1)* from *O. sativa*, have shown that it can function as a floral inducer under short-day conditions (Doi et al., 2004). Further, Ehd1 activates downstream genes such as FT-like genes and MADSbox genes, which are closely related to floral transition (Doi et al., 2004). The acute expression of *Ehd1* in response to “light-on” signals suggests that its expression is regulated by PHYs. B-type RR are also involved in cytokinin and ethylene signaling in Arabidopsis (Hwang et al., 2002), while the *O. sativa* Ehd1 is possibly also involved in developmental and environmental signals mediated by light, cytokinin, and ethylene (Doi et al., 2004).

In Arabidopsis, nine genes encoding for PRRs have been identified (Hwang et al., 2002). There are five such genes in the *O. sativa* genome that encode 10 OsPRRs lacking a typical RD (Table III; Fig. 2C). These five PRRs have previously been reported (Murakami et al., 2003). In this analysis, we have followed the nomenclature as OsPRR1, OsPRR2, OsPRR3, and OsPRR4 representing reported OSPRR1, OSPRR95, OSPRR73, and OSPRR37, respectively. However, the sequence for reported OSPRR59 is not available for analysis. The RD of the OsPRRs shows about 40% similarity with the RDs of the OsRRBs. The identity within OsPRRs ranges from 40% to 70%. Further, the three amino acid residues (D-D-K) in typical RR are not conserved in PRRs (Fig. 2C), and the phospho-accepting D of the RD is replaced by the E, N, or Q.

Arabidopsis is one of the five members of the circadian-associated PRRs, including APRR1/TOC1. Each has a common structural design containing the pseudo-RD of about 120 amino acids at its N terminus and a short CCT motif of about 50 amino acids at the C-terminal end. Based on these criteria, *O. sativa* has at least five genes homologous to APRRs, both in their amino acid sequence structural motifs. Recently, the analysis of the PRR in japonica *O. sativa* (var Nipponbare) has given a new insight in the control of circadian rhythm (Murakami et al., 2003). The OsPRRs are nuclear localized and all have a pseudo-RD, which is similar to the phosphoaccepting RD of the authentic RR (Mizuno, 1998; Imamura et al., 1999). The expression of OSPRR genes is under circadian control. They are expressed in diurnal and sequential manner in *O. sativa*, which is reminiscent of the circadian oscillations of the APRR1/TOC1 in Arabidopsis. The extension of this analysis in an indica variety (Kasalath) of *O. sativa* by Murakami et al. (2005) showed that the indica genome also has homologous members of the OSPRR family. Interestingly, among them, OSPRR37 (OsPRR4 in this article) has been mapped very close to the photoperiod sensitivity Hd2 (Heading date2) quantitative trait locus,

---

**Two-Component Signal Transduction Genes of *Oryza sativa***

---

*Plant Physiol.* Vol. 142, 2006 389
which is the major locus that enhances the photoperiod sensitivity of flowering in Nipponbare. Further, Kasalath has a severe mutational lesion in this OSPRR37 coding sequence (i.e. there is premature termination codon in its CCT motif) as compared to that of Nipponbare (Murakami et al., 2005).

The unrooted tree of all members of the PRR family from *O. sativa* and Arabidopsis shows three distinct groups, each comprised of RRA-type, RRB-type, and PRR members; this suggests independent coevolution of these distinct subspecies in these two plant genomes (Fig. 5C). The OsRRs are distributed over all chromosomes except X (Fig. 6). The maximum numbers of OsRRs are present on contigs of chromosome IV. Further, overlapping ORFs also occurs in the *O. sativa* TCS genes. For example, *OsPRR1* is predicted to encode for OsRRA21 in addition to OsPRR1, and the cluster is located on the longer arm of chromosome II. This association or coupling-in of two different types of RRs (i.e. *OsPRR1* and *OsRRA21*) may throw some light on their close interlinks or cross talks at the signaling level. Likewise, the gene encoding the *OsPHYC* is predicted to encode for a Type II DNA topoisomerase, indicating the possible coregulation of
some light-induced process at the nuclear level. There are several other cases where TCS members are in close proximity. These include the TCS clustering of OsRRB6, OsRRA3, and OsHK3 at the distal arm of chromosome I, which may allow correlated expression and regulation of these members. A similar situation occurs for the lower arm of chromosome II. On chromosome V, OsRRA16 and OsRRA18 are very close to each other. On chromosome IV, the OsRRA17, OsRRB7, and OsRRA22 are predicted to be arranged back-to-back and there is low identity among them. On chromosome VI, OsRRB4 and OsRRB3 are close to OsHK2 and OsHK1, respectively (but on the opposite coding strands in the case of OsHK2 and OsRRB4). Another interesting feature of this genome distribution is the strong clustering or condensing of the TCS member genes within a region of chromosomes (for example, see the short and long arms of chromosome III). In contrast, in Arabidopsis there is a more uniform distribution of the HK genes between chromosomes (Hwang et al., 2002).

Analysis of Expression Profiles for TCS Members in *O. sativa* Employing MPSS Database

MPSS is a valuable tool to have an insight into gene expression (Brenner et al., 2000). MPSS involves the cloning of a cDNA library on beads and the acquisition of 17- to 20-nt signatures (tags) from these cDNAs using an unconventional sequencing method. Signatures of this length have a high specificity and may match to just one position in a complex genome. MPSS has been used previously for genome-level expression analysis in several systems including Arabidopsis (Meyers et al., 2004). To extract information about the relative abundance of transcripts of *O. sativa* TCS members, we have carried out the analysis in the available MPSS database (http://mpss.udel.edu/).
rice), and the results are presented as Supplemental Table S2. This database is derived from the TIGR *O. sativa* genome sequence version 3.0, and our search has been performed employing the 20-nt-long signatures. On an overall basis, we found high variability in abundance of TCS-specific mRNA tags (measured as transcript per million; TPM) in different libraries. For example, tags corresponding to *OsHK4*, *OsHK6*, *OsETR2*, *OsPHYA*, and *B* are relatively more abundant, while the rest of the *OsHK* family members showed very insignificant presence. Similarly, among the various Hpt proteins, *OsHpt2* has been found to be most abundant, being represented in almost all libraries analyzed here. We found that *OsRRA1* and *OsRRA2* (located on chromosomes 11 and 12, respectively), though regulated differentially, show a hit score >1 in MPSS analysis, reconfirming their gene duplication status. In addition, we found *OsRRB3*, *4*, *5*, and *7* are relatively insignificant in expression, and *OsRRB1*, *2*, and *6* are more abundant. Further, all PRR (except *OsPRR2*) are abundant in these expression libraries. These observations indicate that the TCS-related transcriptome of *O. sativa* is very complex, showing a tissue-specific differential regulation for most of the members.

Figure 6. Graphical (scaled) representation of location of putative genes for TCS on *O. sativa* chromosomes. The centromeres on chromosomes have been marked in ovals. The position of first exon of genes (in Mb) has been marked in the parentheses along with their names at same location on chromosomes. Arrow marks the direction of the ORF specific to the TCS gene. Where there is a case of alternative splicing, the genes located on chromosomes represent the longest ORF encoding loci.

Pareek et al.

Search into Possible Role(s) of *O. sativa* TCS Members Employing the Tos17 Mutant Database

One of the important goals of functional genomics is to construct saturation mutant libraries for *O. sativa*. Random mutant population has been generated in *O. sativa* employing endogenous transposon Tos17 tag (http://tos.nias.affrc.go.jp/). Studies have clearly established that Tos17 prefer to integrate in genic regions and hotspots for integration are distributed throughout the genome, making them more suitable for functional genomic studies as compared to other tagged lines (Miyao et al., 2003). Thus, these insertion-tagged lines serve as a useful tool to get an insight into possible functions of novel genes. Because the multiple copy nature of Tos17 as well as background mutations that accompany prolonged tissue culture are known, it is often difficult to establish the relationship between the disrupted gene and the corresponding mutant phenotype. However, it has also been suggested that if two or more lines have independent insertion in the same gene and the same phenotype is obtained, it is likely that the observed phenotype results from disruption of that gene (Miyao et al., 2003). We have carried out in silico screening for *O. sativa* TCS members by BLAST search against flanking sequences from the mutant lines (with cutoff value e<sup>−30</sup>). The results are presented as Supplemental Table S3. We found that phenotypes such as dwarfism, tillering, and heading (late or no heading) are the most common responses as generated by mutation in *OsPHY*s and *OsRR*s, while specific phenotypes are associated with some of the other members. Examples of the later category include phenotypes such as clump morphology (mutation in *OsETR4* and *OsERS1*), stenophyllous and rolled leaves (mutations in *OsHK1*, *OsHpt3*, *OsRRA16*, and *OsRRA18*), vivipary (mutations in *OsRRA1* and *OsRRA2*), and chlorina (mutations in *OsHK1*, *OsHK4*, *OsPHYA*, and *OsRRB3*). *OsPHY*s appears to have a role in yield, grain size, kernel, and leaf morphology, while mutations in *OsHpt*, *OsHK1*, and *OsHK4* exhibit lethality. Mutant lines of *OsPHY*s, *OsPRR*s, *OsRRB4*, *OsRRB6*, *OsRRA1*, *OsRRA2*, and *OsHpt5* show low fertility and sterility. Similarly, all *OsRRB*s and *OsPHY*s, *OsRRA1*, *OsRRA2*, and *OsPRR5* show lesion mimic. Though Tos17 has been reported to be biased for insertion in kinase genes (Miyao et al., 2003), no mutants could be found for TCS members such as *OsHK6*, *OsRTR2*, *OsETR3*, *OsRRA3*, *OsRRA10*, and *OsRRA15*. However, approaches such as specific gene silencing employing RNAi may prove to be helpful in assigning the precise role(s) to these TCS members.

### CONCLUSION

In view of the involvement of TCS-based signaling in diverse life processes, several genomic studies have been attempted. For example, in *E. coli*, 30 HKs and 32 RRs have been shown to be present (Mizuno, 1997). This is in contrast to yeast, where only one HK system (SLN1/YPD1/SSK1) has been shown to play a definite role in osmosensing and adaptation to the environment (Maeda et al., 1994). In this work, screening of the high-quality, finished whole-genome sequence of *O. sativa* has been carried out for the presence of genes that code for putative TCS members. This analysis has revealed that the *O. sativa* genome has 51 genes that code for 73 putative TCS family members (encoding typical *OsHK*s, *OsHpt*s, and *OsRR*s). In contrast, the Arabidopsis genome reportedly has 54 genes coding for 54 putative TCS members (Hwang et al., 2002). However, our analysis of Arabidopsis genome sequence (TIGR version 5.0) has suggested a total of 63 TCS proteins encoded by 54 genes (Table IV). Based on structural similarities of TCS members in Arabidopsis and *O. sativa*, we propose that *O. sativa* TCS members participate in diverse physiological phenomenon ranging from light perception to osmotic adjustment as well as regulation of flowering and hormone signaling. Our analysis of the *O. sativa* genome shows the presence of at least 22 putative *OsHK*s and related proteins, seven putative *OsHpt*s, and 44 putative RRs.

Our analysis of the TCS machinery in the *O. sativa* genome and its comparison with Arabidopsis has revealed that the number of putative TCS genes in the two evolutionarily diverse model systems is very similar in both species (54 versus 51). The overall structure of the TCS machinery is very similar in both species, i.e. maximum number of components belonging to the RR family and minimum number of components belonging to the Hpt family, indicating the architectural conservation of the TCS machinery (Table IV). Apart from the numbers, the individual components from *O. sativa* also show significant sequence conservation with related members of Arabidopsis (Figs. 3–5). However, with respect to TCS

<table>
<thead>
<tr>
<th>Gene Family</th>
<th>No. of Genes</th>
<th>Arabidopsis No. of Proteins</th>
<th><em>O. sativa</em> (japonica)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TIGR Version 5.0</td>
<td>Hwang et al. (2002)</td>
<td>No. of Genes (TIGR Version 4.0)</td>
</tr>
<tr>
<td>TCS proteins</td>
<td>54</td>
<td>63</td>
<td>54</td>
</tr>
<tr>
<td>HKs and related protein</td>
<td>16</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>Hpts</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>RR</td>
<td>32</td>
<td>37</td>
<td>32</td>
</tr>
</tbody>
</table>

Plant Physiol. Vol. 142, 2006 393

Downloaded from on August 21, 2017 - Published by www.plantphysiol.org
Copyright © 2006 American Society of Plant Biologists. All rights reserved.
members, the genomes of these two model plants possess some unique features, which are of great importance from the evolutionary point of view. Although the total number of putative genes in O. sativa is much greater in Arabidopsis, the portion of the genome that encodes TCS components has not expanded in O. sativa. This economy is presumably compensated for by more prevalent cases of alternative splicing of TCS members in O. sativa, where several protein variants are translated by single ORFs with differential editing of pre-mRNA. In prokaryotes, the number of TCS members has been associated with the varied environments they live in; pathogens that live in very narrow environment fluctuations possess few HKs. Does it mean that the same observation is true for eukaryotes as well? In contrast to this observation, Jain et al. (2006) found that the auxin-responsive GH3 gene family is enlarged in Arabidopsis as compared to O. sativa. Zhang et al. (2005) have reported lineage-specific expansion of certain gene families involved in pathogen resistance. However, the genome-wide analysis of these gene families in other plant genomes will help to test this hypothesis. Nonetheless, the O. sativa and Arabidopsis genomes offer interesting research materials that can be exploited to answer some basic questions related to molecular evolution of genomes.

Barakat et al. (1998) have reported that related Arabidopsis genes are fairly evenly distributed on all the chromosomes, but the O. sativa genes are clustered, indicating that organization of O. sativa and Arabidopsis genomes are substantially different. Our results are in agreement with these observations because we found distinct clustering (in terms of physical proximity) of TCS members on chromosomes (see Fig. 6; for example, chromosomes I, II, III, and VIII). However, Jain et al. (2006), who analyzed the auxin-responsive GH3 gene family in O. sativa, found no evident clusters on the chromosomes. In the O. sativa genome, chromosomal segment duplication had occurred several million years ago and there has been a massive ongoing duplication of individual genes, which provides a continuous source of raw material for the gene genesis (Guyot and Keller, 2004; Yu et al., 2005). It has been established that almost 65% of the O. sativa genome is segmentally duplicated as 18 distinct pairs spread across the genome. Most of these TCS members analyzed in this study are part of the duplicated set of segments and show 55% or more identity as paralogous genes. The OsRRA1 and OsRRA2 that are part of recent segmental duplication of chromosome 11 and 12, respectively, show 96% identity in their encoded proteins. Similarly, OsHpt1 and OsHpt5 show 82% identity, and OsHpt2 and OsHpt3 show 76% identity. OsPRR3 and OsPRR4 show 63% identity in sequence. We have observed that genes encoding OsRKB2 and OsRKB3, OsHK6 and OsHK2, though mapped within segmentally duplicated regions, show low identity of 40% and 25%, respectively. Such variation in sequence identity among duplicated genes may be due to gene evolution and accumulation of insertion and deletion events (Taylor and Raes, 2004). Besides segmental duplications, there are clear indications for the occurrence of tandem duplications within TCS members in O. sativa. For example, the lower arm of chromosome VIII shows the presence of three members of OsRRA family having a minimum 84% identity with each other (Fig. 6).

The possibility of close interaction of the TCS pathway with other signaling pathways in plants is being explored. In one such study, comparative microarray analysis carried out on loss-of-function mutant att2 in Arabidopsis revealed more than 600 genes having altered expression, including genes related to auxin, ethylene, biotic and abiotic signal transduction, photomorphogenesis, protein folding, degradation, and development (Hass et al., 2004). In another study, transgenic Arabidopsis plants ectopically expressing the constitutively active B-type ARR showed alterations in the activity of several plant signaling pathways (Tajima et al., 2004). Similarly, combinatorial microarray analysis performed with wild-type Arabidopsis and those having altered cytokinin-mediated His-Asp phosphor- shy signaling circuitry revealed the involvement of 214 cytokinin-up-regulated genes (including those belonging to the AP2/EREBP, MYB, and GATA families), indicating the elaborate and interconnected nature of this TCS machinery in cytokinin signaling (Kiba et al., 2005). These observations support a model where a TCS-dependent cross talk integrates light (phyB) input and hormone (cytokinin)-regulated homeostasis, enabling the plant to react most efficiently to environmental cues (Grefen and Harter, 2004).

In the cyanobacterium Synechocystis, at least five HKs have been shown to be involved in the perception of osmotic stress, which in turn regulate a distinct set of downstream genes (Paithoonrangsarid et al., 2004). This allows the system to have a complex network of Hik-Rre components where salt and hyperosmotic stress signals could be differentially perceived (Shoumskaya et al., 2005). The O. sativa genome contains at least seven members of the OsHK family that show close relationship (25%–29% overall identity) with reported osmosensor AtHK1 from Arabidopsis. However, it remains to be established whether there is a functional homolog of AtHK1 in O. sativa, or if O. sativa is similar to Synechocystis with each OsHK member sensing a different signal. Future investigations employing novel molecular and genetic tools and techniques (such as RNAi) will shed light on the functioning of this important signaling machinery and will pave the way to engineer O. sativa plants for better adaptation to environmental cues.

MATERIALS AND METHODS

Search of Proteins Involved in the TCS in Oryza sativa

The domain structure of the TCS, and the related proteins, was used to identify and classify the TCS proteins using the TIGR O. sativa genome sequence version 4.0. Profiles unique to HKs were used to screen all predicted proteins using the HMMER software (version 3.3.2; http://hmmmer.wustl.edu/).
These unique profiles are for Pfam HMM of HK-A (phosphohexacorrespondent) domain (accession no. PF00512), Hpt (accession no. PF01627), RR RD (accession no. PF0072), CHASE domain (accession no. PF03924), and HAPTase (accession no. PF02518). We have used these profiles as default parameters in the hmmsearch program of the HMMER package. All the significant hits with positive scores were selected for classification and were then examined individually for accessory domains that are usually present. This was accomplished by searching the sequences against the Pfam database (version 20.0) to map the known domains, such as HK-A, Hpt, and the RR myb and CCT (see Fig. 2 for a schematic representation). This step, besides mapping accessory domains to the TCS and providing the basis of classifying the proteins as members of the TCS, also helped to identify false positives, such as retrotransposon or transposable elements containing fragments of TCS sequences. TM regions were also mapped using the PSORT program (Nakai and Horton, 1999) to provide specificity and sensitivity of the methodology. Similar analyses were performed on the datasets having all known and well-characterized HKs and related proteins of the TCS in Arabidopsis (Arabidopsis thaliana; http://mpids.gsi.de/proj/thal/db/index.html). The percentage values of the sequence similarity and identity within the groups were obtained using BLAST (Altschul et al., 1990).

The protein sequences of all the HKs, RRs, and Hpts from *O. sativa* were used to search in KOME (28,469 full-length cDNA clones) using the TBLASTN program (Altschul et al., 1997). Details of these analyses are presented in Supplemental Table S1. The protein sequences of the TCS in Arabidopsis were retrieved from the Munich Information Center for Protein Sequences (http://mpids.gsi.de/proj/thal/ database). TIGR has used FGENESH, an HMM-based tool for ab initio gene prediction using known properties of coding and noncoding sequence, to predict genes and gene-related information, such as the reading frame and the coding statistics (codon usage and GC contents). FGENESH has proved to be the most sensitive and specific method when compared with other ab initio gene prediction methods (Matthe et al., 2002), especially for the monocots (Yuan et al., 2005). We have also used an HMM-based tool in an unsupervised manner (i.e. no prior selection criterion) to find out the TCS in pseudomolecule protein dataset. We have used an extrinsic approach, based on similarity searches using BLAST and TBLASTN in other data sets, such as KOME and NCBI databases, to exclude the possibility of any additional member of TCS that might have been left unreported in TIGR.

**Multiple Sequence Alignments and Phylogenetic Analysis**

Multiple sequence alignments of related proteins belonging to each class from both Arabidopsis and *Oryza* were performed using T-Coffee (Notredame et al., 2000) with editing of the subsequent alignments using Seaview (Galtier et al., 1996). Alignments of the protein domain sequences were used as input for phylogenetic analysis using protpars, a maximum parsimony routine from the PHYLIP package (version 3.6; Felsenstein, 2004). Unrooted trees were plotted using “draw tree” from the same package.

**Analysis of MPSS Database for Expression Profiles**

To gain insight into expression profiles of TCS members in *O. sativa*, the MPSS database (http://mpss.udel.edu/oric/) was searched (opting 20-on-signature sequences) using the locus ID given in the TIGR database. The data thus obtained have been analyzed and grouped based on tissue specificity (Supplemental Table S2).

**Analysis of Tos17 Mutant Population**

Protein sequences of individual TCS members were used as a query to search the Tos17-tagged lines (http://tos.niias.afl.go.jp/). For better precision in picking up the phenotypes, only a high negative exponential E score value for FST hits has been the selection criteria. The information thus obtained has been presented in Supplemental Table S3.

**Supplemental Data**

The following materials are available in the online version of this article.

**Supplemental Table S1.** ESTs for TCS members in *O. sativa* available at KOME database.

**Supplemental Table S2.** Analysis of expression profiles of *O. sativa* TCS members using MPSS database.

---

**Supplemental Table S3.** Tos17-tagged lines available for TCS members in *O. sativa*.

---

**Acknowledgments**

We thank Professor Roger A. Leigh and Professor Sudhir Sepproy for a critical reading of this manuscript; we are indebted to Professor Govindjee for his suggestions to improve the language of this manuscript.

Received July 5, 2006; accepted July 24, 2006; published August 4, 2006.

**Literature Cited**


---

Two-Component Signal Transduction Genes of *Oryza sativa*


receptor that transduces cytokinin signals across the membrane. Plant Cell Physiol 42: 1017–1023


