

Review

# Genome-wide comparative analyses of domain organisation of repertoires of protein kinases of *Arabidopsis thaliana* and *Oryza sativa*

A. Krupa<sup>1,2</sup>, Anamika<sup>2</sup>, N. Srinivasan<sup>\*</sup>

*Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India*

Received 24 January 2006; received in revised form 6 May 2006; accepted 18 May 2006

Available online 2 June 2006

Received by A. Bernardi

## Abstract

A comparative analysis on protein kinases encoded in the completely sequenced genomes of two plant species, namely *Arabidopsis thaliana* and *Oryza sativa* spp *japonica* cv. *Nipponbare* is reported in the current study. We have analysed 836 and 1386 kinases identified from *A. thaliana* and the *O. sativa* genomes respectively. Their classification into known subfamilies reveals selective expansions of the plant receptor kinase subfamily comprising of Ser/Thr receptor kinases. The presence of calcium dependent kinases, and potential absence of cyclic nucleotide-dependent protein kinase of the type found in other (non-plant) eukaryotes, are other notable features of the two plant kinomes described here.

An analysis on domain organisation of each of the protein kinases encoded in the plant genome has been carried out. Uncommon composition of functional domains like nuclear translocation factor domain, redox sensor domain (PAS), ACT and lectin domains are observed in few protein kinases shared between the two plant species. Biochemical functions characteristic of the domains recruited in these protein kinase gene products suggest their mode of regulation by alternate cellular localisation, oxidation potential, amino acid flux and binding of carbohydrates. Occurrence of multi-functional kinases with diverse enzymatic modules, such as Transposases and peptidases, tethered to the kinase catalytic domain is another interesting feature of the protein kinase complement of the *O. sativa* genome. Co-occurrence of diverse nucleotide and carbohydrate binding domains with catalytic kinase domain containing gene products has also been observed. Putative homologues of protein kinases of *A. thaliana* that regulate plant-specific physiological processes like ethylene hormone response, somatic embryogenesis and pathogen defence have been identified in *O. sativa* genome as well. © 2006 Elsevier B.V. All rights reserved.

**Keywords:** Protein kinases; Genome analysis; Domain organisation; Phosphorylation

**Abbreviations:** MAPK, mitogen-activated protein kinase; STYK, Ser/Thr/Tyr kinase; RPS-BLAST, Reverse PSI-BLAST; PSSMs, position specific matrices; STK, Ser/Thr kinase; PI Kinases, phosphatidyl inositol kinases; PSI-BLAST, position specific iteration BLAST; RLK, receptor like kinases; RS6K, Ribosomal protein S6 kinase; cNMP, cyclic nucleotide monophosphate; CDKs, cyclin dependent kinases; GSKs, glycogen synthase kinases; CK2, casein Kinase-2; MEK, Meiosis specific ser/thr protein kinase (MAP kinase kinase); MEKK, MEK Kinase (MAP kinase kinase kinase); NTF 2, nuclear transport factor 2; CCA1, circadian clock protein 1; CDPKs, calcium dependent protein kinases; CAMKs, calcium dependent kinases; PEPCK, phosphoenol pyruvate carboxylase kinase; Snf1, sucrose non-fermenting; CIPKs, calcineurin-B interacting protein kinases; UBA, ubiquitin association domain; KA1, kinase associated domains; IRAK, interleukin receptor associated kinase; IRK, inter leukin receptor; TLR, toll like receptor; LRR, leucine rich repeat; Usp, universal stress protein; PAC, PAS C-terminal; EF-2, elongation factor 2; TNFR, tumour necrosis factor receptor; CaMBD, calmodulin binding domain; JAK, janus kinase; GDPD, glycerophosphoryl diester phosphodiesterases; SRKs, S-domain receptor kinases; SI, self-incompatibility; PAS, Period protein-Aryl hydrocarbon receptor nuclear translocator protein-Single-minded protein; ACT, Aspartokinase, Chorismate mutase and TyrA; SCP, sperm-coating glycoprotein; PAN, amino terminal domain of Plasminogen/hepatocyte growth factor family-Apple domains of the plasma prekallikrein/coagulation factor XI family-domains of various Nematode proteins; RWD, RING finger domain-WD-repeat-DEAD (DEXD)-like helices.

\* Corresponding author. Tel.: +91 80 2293 2837; fax: +91 80 2360 0535.

E-mail address: [ns@mbu.iisc.ernet.in](mailto:ns@mbu.iisc.ernet.in) (N. Srinivasan).

<sup>1</sup> Present address: Cell Cycle Control Lab London Research Institute Cancer Research, UK Clare Hall, South Mimms Hertz EN6 3LD, UK.

<sup>2</sup> These authors have contributed equally in this work.

## 1. Introduction

Plants respond to a large number of external and endogenous stimuli, which are translated into cellular responses that enable their adaptation to changing environmental conditions. Common stress conditions that affect plant growth and crop production adversely include drought, salinity and low temperature. The cellular and molecular responses of plants to environmental stress have been studied extensively (Thomashow, 1999; Hasegawa et al., 2000).

Molecular events triggered by subsequent perception of various stress signals of plants lead to the generation of second messengers that modulate intra-cellular  $\text{Ca}^{2+}$  levels. Phosphorylation cascades targeting proteins involved in cellular protection and/or transcription are initiated in response to change in  $\text{Ca}^{2+}$  level, to control specific set of stress regulated genes. A variety of protein kinases classified into distinct families (Hardie, 1999; Wang et al., 2003) are involved in various regulatory processes of plants. Members of a large family of calcium-binding regulatory protein kinases participate in numerous aspects of plant growth and development (Harmon et al., 2000). MAPK pathways play a major role in plant signal transduction from cell division to cell death (Hirt, 2000). Receptor like protein kinases (RLKs) of plants with Ser/Thr specific kinase activity also constitutes another major family of proteins with diverse extra-cellular ligand binding domains (Strain and Muse, 2005). Their role in regulation of self-rejection, morphology of organs and pathogen resistance is well characterised (Braun and Walker, 1996; Satterlee and Sussman, 1998). Disruption in their activity leads to abnormal plant development.

Whole genome sequencing projects have enabled comprehensive comparison of the repertoires of various families of signalling proteins in different organisms. The availability of the completely sequenced genome of *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000) has set a stage to initiate genomic studies on various protein families in plants. A comprehensive survey of the two modules of reversible phosphorylation namely protein kinases and protein phosphatases in *A. thaliana* genome has been carried out (Wang et al., 2003; Kerk et al., 2002). Further the complete set of protein kinases of *A. thaliana* has been classified by a whole-protein-based hierarchical clustering approach (Kerk et al., 2002; Gribskov et al., 2001) that is largely consistent with the biochemical and functional properties of these protein kinases, reported from genetic complementation and other experimental studies. A comparison of protein kinases between *A. thaliana* and *Saccharomyces cerevisiae* (Wang et al., 2003) has previously revealed the protein kinase families that are shared and those that are unique in the two species.

In this article we review the repertoire of protein kinases encoded in the completely sequenced genomes of two plant species, namely *A. thaliana* (Arabidopsis Genome Initiative, 2000) and *Oryza sativa* spp *japonica* cv. *Nipponbare* (Goff et al., 2002). These protein kinases have been classified further into subfamilies based on their similarity in sequence in their catalytic domain (Hanks et al., 1988). A comparative analysis of the plant protein kinases with those of other eukaryotes indicates selective expansion of groups of protein kinases in the plants and those confined to the plant kingdom.

The present review has a specific focus on the domain organisation of the protein kinases encoded in the two plant genomes. It is well known that domains tethered to catalytic protein kinase domain influence the mode of regulation of kinase activity. Novel composition of functional domains in some of the protein kinases shared within and across the two plant species has been thus identified. Trans-genomic comparison of the protein kinases repertoire of *A. thaliana* and *O. sativa* has led to the identification of members that are occurring exclusively in these two plant species. We have identified protein kinases which are unique to *A. thaliana* and not identified in *O. sativa* and vice-versa. Protein kinase homologues of *A. thaliana* that regulate plant-specific physiological processes have also been identified in the *O. sativa* genome.

## 2. Materials and methods

### 2.1. Data sets

The translated ORFs of the completely sequenced genomes of *A. thaliana* (Version: v110303; March 13 release) and *O. sativa* spp *japonica* cv. *Nipponbare* (Version 3) have been obtained from MIPS (Munich Information centre for Protein Sciences; <ftp://ftpmips.gsf.de/cress/arabiprot/>) and TIGR (Version 3) (The Institute of Genomic Research; [ftp://ftp.tigr.org/pub/data/Eukaryotic\\_Projects/o\\_sativa/annotation\\_dbs/pseudomolecules/version\\_3.0/all\\_chrs/](ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_3.0/all_chrs/)) respectively.

### 2.2. Identification of protein kinases

Sequence search and profile search methods such as PSI-BLAST (Altschul et al., 1997), RPS-BLAST (Marchler-Bauer et al., 2003) and HMMer (Eddy, 1998) which is based on Hidden Markov models have been used to extract the gene products containing sequences similar to that of STYK catalytic domain from the complete set of ORFs of *A. thaliana* and *O. sativa*. Potential kinase sequences identified by various search procedures, using stringent search parameters as described in our earlier studies (Krupa and Srinivasan, 2002, 2005; Anamika et al., 2005) have been merged to obtain a data set of putative catalytically active protein kinases. These parameters include an alignment coverage of greater than 70% over the kinase domain region of the putative kinases with the query kinase domain. Further the exhaustive coverage of all the protein kinases of *A. thaliana* in our current data set has been ensured by comparing the list with the recently compiled list of protein kinases of *A. thaliana* at PlantsP (Gribskov et al., 2001). Similarly PKs have been identified in *O. sativa* genome and the rice PKs have been referred with the gene product identifiers corresponding the Release 3 of rice genome data. In order to eliminate sequences that are likely to be fragments of larger gene products in the data set, shorter sequences within the data set sharing an identity of 85% with any of the sequences have not been included in the current analysis. This data set is thus ensured to be a repertoire of protein kinases to study their classification and diverse domain organisation. The conservation of the functional residues in the kinase catalytic domain of STYKs has been inferred by generating the multiple

sequence alignment of the catalytic domain using ClustalW (Thompson et al., 1994). The final data set chosen for the analysis comprised of only those sequences with the catalytic residue which acts as base (Asp) that serves as a nucleophile during phospho-transfer (Krupa et al., 2004a). The current rice genome data set (Release 3) with 98% completeness and improved prediction of gene model has enabled the assessment of accuracy of domain organisation. The protein kinases with atypical domain composition have been particularly searched against the transcript sequence databases in order to ensure the consistency with respect to unusual domain combinations.

### 2.3. Hanks and Hunter scheme of classification

Complete sets of protein kinases obtained from the genomes of *A. thaliana* and *O. sativa* have been further classified based on sequence similarity in their kinase catalytic domains, as proposed by Hanks and Hunter (Hanks et al., 1988) using RPS-BLAST.

RPS-BLAST was used to search each of the protein kinase-like sequences identified in the plant genomes as a query against the database containing 55 PSSMs created for the various subgroups of protein kinases in each of the subfamilies. The query kinase sequence was associated to their nearest subfamily based on the extent of sequence similarity as described in our previous analysis (Krupa and Srinivasan, 2002, 2005; Anamika et al., 2005).

Protein kinases of *A. thaliana* have also been classified by whole-protein-based clustering method by Gribskov and co-workers (Wang et al., 2003). In the current work the discussion on the *A. thaliana* however is confined to the Hanks and Hunter scheme of classification. In addition, the occurrence of non-kinase domain families that are tethered to the catalytic kinase domains has been considered in further characterization of Hanks and Hunter classified kinases of the two plant species.

### 2.4. Assignment of other functional domains to the plant protein kinases

The domain organisation of each plant STK has been analysed for the occurrences of other functional domains by using HMMER based search on a profile data set comprised of domains PFAM (Bateman et al., 2004). Putative trans-membrane spanning segments have been identified using TMHMM (Krogh et al., 2001).

### 2.5. Identification of closest homologues of protein kinases

An attempt has been made to identify the closest homologue of each STK of *A. thaliana* in the *O. sativa* genome and vice-versa. Searching every *A. thaliana* STK sequence in the *O. sativa* kinome data set and every *O. sativa* STK in the kinome data set of *A. thaliana*, has enabled identification of putative orthologues across the two genomes. The sequences of closest homologues thus obtained from the cross-species comparison share highest identity with each other, when compared to other STKs across the two species. Other features such as Hanks and Hunter subfamily, whole-protein length and domain organisation of the STKs forming a potential set of closest homologues have also been used as additional guidelines for the identification of nearest relatives.

A list of 49 protein kinases of *A. thaliana* with well defined functional roles, mostly based on prior experimental studies, has been obtained from the Swissprot database (Boeckmann et al., 2003). Each of these protein kinases has been queried in the rice genome database to search for their counterparts in the rice genome using BLAST. The decision on the putative rice kinases identified as orthologues of well characterized *A. thaliana* kinases has been based on the sequence similarity over the whole length of the protein, domain organisation and the subfamily to which they are associated.

## 3. Results

The completed genomes of *A. thaliana* and *O. sativa* spp *japonica* cv. *Nipponbare* provide ideal systems for trans-genomic comparison of protein kinases among the plant species. While *A. thaliana* is a model for plant molecular and cellular studies, the availability of another plant genome provides a platform to understand the conserved themes in the signalling events involving plant protein kinases. In order to carry out the cross-species comparison of the protein kinase repertoire, a comprehensive list of protein kinases encoded in the two plant species have been obtained by various sequence search methods as described in the 'Materials and methods' section.

### 3.1. Protein kinases encoded in the *A. thaliana* and *O. sativa* genomes

A total of 1052 gene products with STK catalytic domains have been identified in *A. thaliana*. A non-redundant data set was further obtained as described in the 'Materials and methods' section by excluding 106 sequences. Sequences lacking catalytic residue (Asp) required for phospho-transfer (110), were further eliminated to obtain a data set containing 836 PKs for further analysis. A set of 2286 STKs has been identified in the *O. sativa* genome. This data set was further filtered to exclude redundant sequences (688). Among the non-redundant sequences (1598), sequences with incomplete kinase domain (48) and those lacking the catalytic residue (164) have been removed to derive a set of 1386 protein kinases that have been considered for further analysis. Domain organisation of PK-like sequences encoded in the two genomes, that lack catalytic residue have also been analysed and any atypical domain combination observed, has been discussed in the further sections.

Validation of the observations of the genomic surveys has been carried out by cross-referencing with other publicly available databases such as PlantsP for *A. thaliana* and current release of rice genome data at TIGR (Release 3). PKs with atypical domain combination have been further searched against the currently available cDNA data set at the TIGR.

The protein kinase encoding genes of most of the completely sequenced eukaryotic species constitute 2% of their genomes (Hunter and Plowman, 1997; Plowman et al., 1999; Morrisson et al., 2000; Manning et al., 2002; Kostich et al., 2002; Krupa and Srinivasan, 2002; Krupa et al., 2004b; Anamika et al., 2005), while a significantly large proportion (3.2%) of protein kinase coding genes occur in the *A. thaliana* genome as suggested from

the translated ORFs bearing similarity to protein kinases. The investigations have been focused on Ser/Thr and Tyr kinases. Atypical protein kinases, histidine kinases and PI Kinases have been excluded from the current analysis. The homologues of Ser/Thr and Tyr kinases have been classified into various subfamilies as proposed by Hanks and Hunter (Hanks et al., 1988).

Detection of relationships between proteins and identification of catalytic protein kinase domain and other associated domains discussed in the following sections are based on the highly reliable similarity as suggested by the parameters of sequence and profile search methods. An *E*-value cut-off of  $10^{-4}$  for PSI-BLAST and RPS-BLAST searches and an *E*-value cut-off of 0.01 for searches using Hidden Markov Model have been used. Various programs, as described in the Materials and methods section and in our earlier publications (Krupa and Srinivasan, 2002, 2005; Krupa et al., 2004b; Anamika et al., 2005), coupled with manual inspection have been employed to prune the data set for protein kinases with catalytic potential. The similarities and relationships discussed in the subsequent sections pertain to a data set qualifying the stringent cut-off parameters (see Materials and methods section).

The detailed listing of the classification and domain organisation of the *A. thaliana* and rice protein kinases is provided in our KinG database (Krupa et al., 2004b) at the URL: <http://hodgkin.mbu.iisc.ernet.in/~king>.

A recent study on trans-genomic comparison of protein kinases of *S. cerevisiae* and *A. thaliana* (Wang et al., 2003) reveals selective expansion of certain subfamilies and unique plant protein kinases. More recent studies (Shiu and Bleecker, 2003; Shiu et al., 2004) on the phylogenetic analysis of RLK of *A. thaliana* and rice have indicated selective expansion of (RLK) subfamily and high divergence of the extra-cellular domain of the receptor kinases in plants.

Genome-wide surveys carried out in the recent past confine to specific plants STK families such as receptor kinase families (Shiu and Bleecker, 2003; Shiu et al., 2004; Shiu and Bleecker, 2001) and cell cycle related kinases (Vandepoele et al., 2002). The current work aims to provide a comprehensive and systematic comparison of all subfamilies of STKs encoded in the *A. thaliana* and *O. sativa* genomes. We carried out further comparisons of the plant protein kinases with those from other eukaryotes whose entire genome has been sequenced to understand their common and diverse domain organisation.

### 3.2. Comparison of the protein kinase subfamily repertoire of *A. thaliana* and *O. sativa* protein kinases

STKs encoded in the two plant genomes have been grouped into distinct eukaryotic protein kinase subfamilies based on their similarity in the kinase catalytic domain. The distribution of STKs in various subfamilies has been compared between the two plant species and the details are summarised in Table 1. As seen from the Table the most highly populated subfamily corresponds to the plant receptor kinases. The receptor kinase family is further expanded in the rice genome. The calcium dependent protein kinase family also has a significantly large representation compared to other protein kinase subfamilies. Biochemical characterisation of

Table 1  
Number of *A. thaliana* and *O. sativa* Ser/Thr kinases in various subfamilies

Subfamily/Subgroup	<i>A. thaliana</i>	<i>O. sativa</i>
AGC	30	24
Cyclin dependent kinases	10	11
MAP kinases	18	14
Glycogen synthase kinases	10	7
Casein kinase 2	4	2
MAP kinase kinase (MEK)	10	2
MAP kinase kinase kinase	22	4
Raf	40	8
Calcium dependent kinases		
Calcium dependent protein kinase (EF-hand)	28	30
Calcium dependent protein kinase (non-EF hand)	22	18
SNF1 family	30	27
Casein kinase 1	13	12
NimA	10	3
Translation regulatory kinase	1	3
Plant receptor kinases	491	807

protein kinases of distinct subfamilies by various groups has suggested their roles in various aspects of plant signalling (Wrzaczek and Hirt, 2001; Laurie and Halford, 2001; Clark et al., 2001; Becraft, 1998).

The following sections describe the extent of occurrence of protein kinases of major protein kinase subfamilies in *A. thaliana* and *O. sativa* and their well characterised functional roles in plant signalling.

#### 3.2.1. AGC subfamily

Among the different groups of AGC kinases, homologues of the members of AGC\_6 group that includes the RS6K and the flowering plant protein kinases group (PVPK/AGC\_8) have only been identified in the two plant species. These protein kinases have further been suggested to play critical roles in the regulation of plant growth (Bogre et al., 2003). An unusual feature of a set of protein kinases of this group observed during the current analysis, is the occurrence of the two tandem sets of PAS and PAC domains in the N-terminal of the gene products (At5g58140, At3g45780 in *A. thaliana*; 11670.m02224, 11686.m00014 in *O. sativa*) followed by the C-terminal kinase domains. These domains are implicated as sensors for oxygen and the redox potential in the cells. A search for proteins kinases with similar domain combination in other eukaryotic species in the KinG database (Krupa et al., 2004b) has revealed CAMK group of kinases from *S. cerevisiae*, *D. melanogaster* and *Homo sapiens*. Raf family kinases of the two plant species also contain a single set of PAS and PAC domains. The identification of such sensors in AGC kinases recruits a new function to the subfamily.

The other main groups of AGC kinases comprising Akt kinase,  $\beta$ -adrenergic receptor kinase and the cyclic nucleotide-regulated kinases like those encoded by other eukaryotic species are not detected in the two plant species. A single gene product (11668.m01706|Os02g17970) in the rice genome possesses a catalytic domain bearing high similarity with the kinase domains of AGC protein kinases. N-terminal phosphatase 2C domain and two central cNMP binding domains precede the C-terminal kinase domain of the gene product. However the catalytic kinase domain

lacks the catalytic aspartate and hence is unlikely to carry out phosphorylation.

### 3.2.2. CMGC subfamily

CMGC subfamily of protein kinases includes members of four distinct group of kinases namely CDKs, GSKs, MAPK and CK2.

The role of CDKs in the regulation of cell cycle in *A. thaliana* and *O. sativa* has been extensively studied (Clark et al., 2001; Joubes et al., 2000; Umeda et al., 1999). The cognate cyclins for B-type CDKs have been identified as cyclin D2 in *A. thaliana* and B2-cyclin in rice (Kono et al., 2003; Lee et al., 2003). A recent genome-wide survey (Vandepoele et al., 2002) in *A. thaliana* has identified various cell cycle regulators. In the current work a total of 10 and 11 CDKs have been identified in the *A. thaliana* and rice genomes respectively.

In *Nicotiana tabacum*, phosphorylation of Ser/Arg rich splicing factors by PK12, a member of the LAMMER kinase family has been shown to alter their splicing activity on target mRNA (Savaldi-Goldstein et al., 2000). The LAMMER kinases bearing similarity to CDKs in their catalytic domain have been identified in the current analysis in *A. thaliana* (15 LAMMER-like kinases) and rice (11 LAMMER-like kinases) genomes are likely to be localised in the nucleus (as observed in the case of PK12) to bring about the regulation of various splicing factors.

The GSK family members have been associated with various processes during floral meristem patterning, brassinosteroid signalling, salt stress and wound healing (Choe et al., 2002; Jonak and Hirt, 2002; Dornelas et al., 2000) in *A. thaliana*. *O. sativa* and *A. thaliana* genomes encode nearly equal number of GSKs (Table 1). A high sequence similarity (>60% identity) over the whole-protein length shared across the GSKs of the two species suggests analogous functions of the GSKs encoded in the rice genome.

MAP Kinases and other components of MAP Kinase signalling cascades, MEK, MEKK and Raf kinases have been identified in both the genomes (Table 1). The role of plant MAPKs in innate immunity, hormonal responses, cell cycle regulation and abiotic stress signalling and defense mechanisms has been well documented in *A. thaliana* and rice genomes (Tena et al., 2001; Ichimura et al., 2002; Agrawal et al., 2003).

A novel class of MEKs conserved across the two plant species has been identified in the current analysis. This unusual MEK is characterised by a C-terminal NTF 2 domain following the N-terminal kinase catalytic domain. The NTF domains occurring in various proteins mediate the interactions with nuclear porins and Ran GDP to facilitate the translocation of the associated protein to the nucleus (Stewart et al., 1998). MEKs phosphorylate MAPKs which are targeted to the nucleus. Recognition of NTF domains in MEKs suggests an alternate mode of regulation of MAPKs with respect to their subcellular localisation.

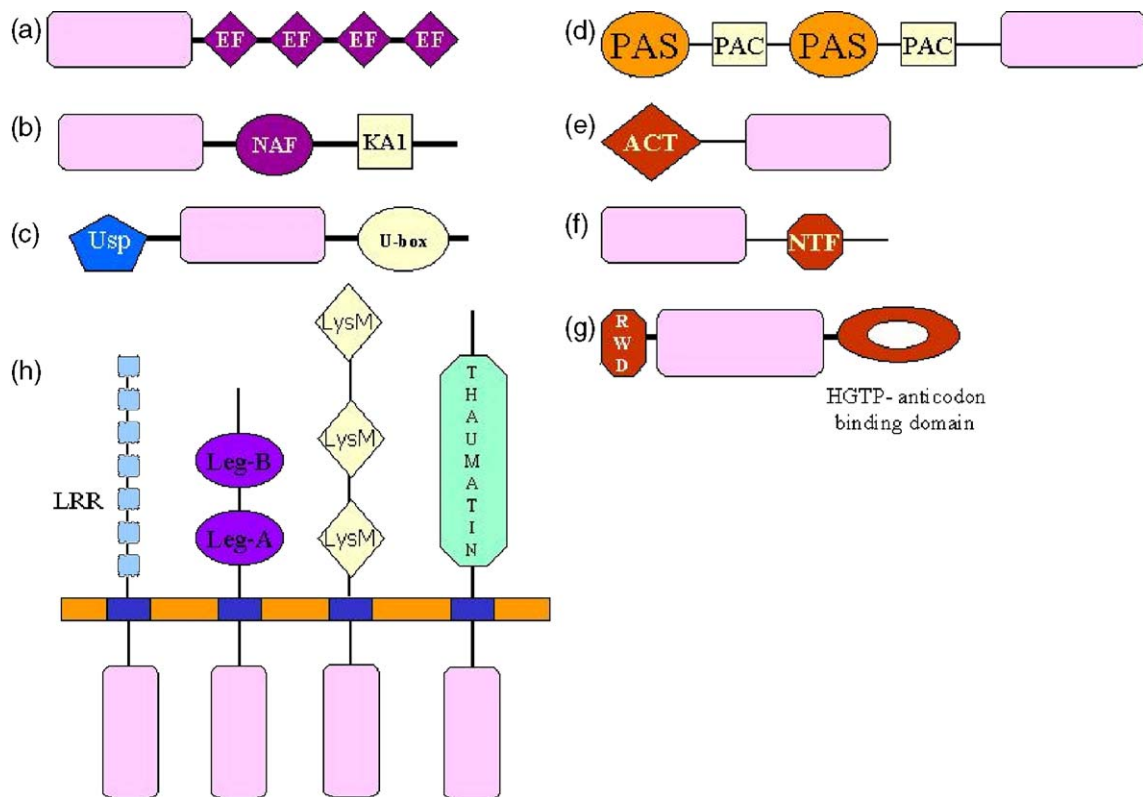


Fig. 1. Representative domain organisation of multi-domain protein kinases shared by *A. thaliana* and *O. sativa*. Protein kinase catalytic domain, cell membrane and transmembrane regions are shown in pink. LegA, LegB: legume lectin like; EF: EF-hand motifs; Usp: universal stress protein; KAI: kinase associated domain; NTF: nuclear transport factor. Gene product identifiers of the corresponding *A. thaliana* and rice PKs respectively: (a) At1g18890; 11687.m00641Os11g07040. (b) At2g38490; 11681.m02270Os09g25100 (c) At5g61560; 1668.m011631Os02g12670 (d) At5g5814; 11670.m02224. (e) At4g38470; 1668.m001931Os02g02780. (f) At5g40440; 11680.m02660Os06g27890; (g) At3g59410; 11670.m039881Os12g41090. h At1g14390, At5g10530, At1g51940, At5g38250; 11668.m012471Os02g13510, 11673.m037151Os07g38250, 1681.m030431Os09g33630, 11667.m001411Os01g02310.

CK2 family kinases are known to phosphorylate a wide number of proteins. The role of CK2 in the phosphorylation of CCA1 of the circadian clock that controls numerous physiological proteins has been recently established in *A. thaliana* (Daniel et al., 2004). *O. sativa* and *A. thaliana* genomes encode 4 isoforms of CK2.

### 3.2.3. Calcium dependent kinases

Calcium dependent kinases are central to the plant calcium-signalling network. Calcium dependent kinases in plants vary significantly in terms of their regulation and domain organisation (Fig. 1a,b) compared to metazoan and fungal protein kinases. They are classified based on their modes of regulation. The binding of  $\text{Ca}^{2+}$  to an intrinsic calmodulin-like module with four EF hand motifs regulates the CDPKs.  $\text{Ca}^{2+}$  or calmodulin is involved in the regulation of another class of CAMKs and protein kinases activated by both, constitute CAMKs, which are also found in the other metazoan and fungal protein kinases. The CAMK group of kinases in plants includes plant-specific PEPCCK and related kinases that are involved in various regulatory processes (Nimmo et al., 1990; Outlaw et al., 2002). A set of protein kinases referred to as Snf 1 are also identified in the plants and they share similarity in the catalytic domain with other calcium dependent kinases. They include the CIPKs with C-terminal NAF domains that is likely to interact with calcineurin like calcium sensors (Albrecht et al., 2001). Another class of protein kinases sharing sequence similarity with Snf1 kinase domains contains C-terminal UBA and KA1, located at the C-terminal of the kinase catalytic domain.

The functional roles of various members of CDPKs have been well characterised. They are implicated in seed development, responses to light, cold stress and gibberelin in rice (Abo-el-Saad and Wu, 1995; Martin and Busconi, 2001; Lu and Hrabak, 2002). In *A. thaliana* a specific set of CDPKs are associated with endoplasmic reticulum and stress response (Hwang et al., 2000).

### 3.2.4. Casein kinase 1

Casein kinase family is well represented in two plant species (Table 1). It has been previously shown that a protein kinase of this family, OsCKII in rice has a regulatory role in the root development and plant hormone sensitivity (Liu et al., 2003). Like in many other eukaryotic species, multiple isoforms has been identified in both *A. thaliana* and *O. sativa*. These isoforms share high sequence similarity (45%).

### 3.2.5. Receptor like kinases in plants

The receptor kinase family in plants includes protein kinases with single trans-membrane segments (RLKs) and those kinases without membrane spanning regions but bearing high sequence similarity in the catalytic domain to the trans-membrane kinases (RLCKs). Trans-membrane Ser/Thr kinases of distinct types occur abundantly in the plants unlike other eukaryotic species where the TGF- $\beta$  receptor kinases constitute the only Ser/Thr receptor kinase family. The human IRAK and pelle kinases of *D. melanogaster* bear high similarity in their catalytic domain to plant receptor kinases. However trans-membrane region is not present in human IRAK and Dm Pelle. They associate with IRK/

TLR through adaptor protein and act as co-receptor. Such a role of co-receptor is similar to non-membrane spanning members of receptor like kinase. The plant receptor kinases are classified further based on the extra-cellular ligand binding domains into various types (Laurie and Halford, 2001; Shiu and Bleecker, 2001; Morris and Walker, 2003).

A major fraction of the protein kinases encoded in the genomes of *A. thaliana* (>58%) and *O. sativa* (>70%) comprises of plant receptor kinases (Table 1). Further the plant receptor kinase family is highly proliferated in the rice genome showing high divergence with respect to their extra-cellular ligand binding domains while their kinase catalytic domains share significant sequence similarity (>40%).

Plant receptor kinases play key roles in the cell–cell recognition process during development (Becraft, 1998, 2002). The relationship between plant receptor kinases has been analysed extensively (Shiu and Bleecker, 2001, 2003; Shiu et al., 2004). Many of these receptor kinases are the targets of various bacterial elicitors that trigger signalling events such as those involved during symbiosis (Radutoiu et al., 2003; Madsen et al., 2003) and response against pathogens (Chen et al., 2003; Gaspero and Cipriani, 2003; Montesano et al., 2003; Du and Chen, 2000; Czernic et al., 1999; Wang et al., 1998). The regulatory roles of a few plant receptor kinases have been implicated in light dependent processes like photo periodism and energy transduction (Bassett et al., 2000; Deeken and Kaldenhoff, 1997). The recognition of self-related pollen by stigma, inducing self-incompatibility, that promotes out-crossing in the flowering plants, is governed by S-locus receptor kinase and related kinases (Cock, 2000; Muschietti et al., 1998; Murase et al., 2004). Another group of plant receptor kinases, wall associated kinases (WAKs) with extra-cellular EGF like domains, link the extra-cellular matrix to the cytoplasm (He et al., 1999; Verica and He, 2002). WAKs are up regulated in response to pathogens conferring resistance to the plants. Receptor kinases of *A. thaliana* that control the meristem growth (CLAVATA), organ shape (ERECTA) and floral abscission (HASEA) and Brassinosteroid Insensitive 1 (BRI1) contain LRR domain in their extra-cellular regions (Torii, 2000). The receptor kinases with extra-cellular lectin domains have also been identified in *A. thaliana* (Herve et al., 1999) and their role in wound healing, senescence and oligogalacturonic acid has been demonstrated (Riou et al., 2002). In the current analysis we identify a large number of receptor kinases in the two plant species which include representatives of these well characterised receptor kinases described above and novel receptor kinases with diverse extra-cellular domains and intra-cellular domains such as Jacalin-like domains and enzymatic domains like DNA-methylases. The comparisons of domain organisation of receptor kinases and their conservation across the two plant species have been discussed in the following sections.

## 3.3. Comparison of domain organisation of plant protein kinases with other eukaryotic protein kinases

The comprehensive sets of plant kinases analysed have been compared with the protein kinase complement in the other eukaryotic species to study the extent of conservation and divergence of

various subfamilies of protein kinases and their domain organisation. The occurrences of GSKs, RS6K, CDKs, MAPKs, MEK, MEKKs, Raf Kinases, CK1 and CK2 subfamilies of protein kinases in most of the eukaryotic species suggest they constitute the core components of eukaryotic signal transduction machinery. The common regulatory processes associated with these kinases like control of cell cycle, transcription, general metabolism and response to stress are fundamental to the sustenance of eukaryotic cells.

The metazoan specific protein kinase subfamilies including activin/TGF- $\beta$  receptor kinases, tyrosine kinases and polo family kinases have not been detected in the two plant genomes. Interestingly, *Plasmodium falciparum* (a protozoan) which lacks typical MEKs has a TGF $\beta$  receptor like kinase (Ward et al., 2004). While cyclic nucleotide-dependent phosphorylation events have been reported in plants (Friedrich et al., 1999) the type of fungal and metazoan protein kinases regulated by cyclic nucleotides have not been detected in the genomic survey on the two plant genomes. The role of putative kinase-like (without the catalytic base hence inactive as a kinase) gene product (11668.m017061 Os02g17970) with cyclic nucleotide binding domains and phosphatase 2 C domain identified in the rice genome is unclear.

The plant receptor kinase family and the calcium dependent kinase families of protein kinases identified in *A. thaliana* and *O. sativa* are highly divergent and uniquely conserved among the plants. Protein kinases of the CDPK family with intrinsic calmodulin-like modules however are explicitly found in plants and in some protozoa and are not detected in other eukaryotic species.

#### 3.4. Comparison of protein kinases of *A. thaliana* and *O. sativa*

In order to identify the putative counterparts of *A. thaliana* protein kinases in the rice genome and putative counterparts of *O. sativa* protein kinases in the *A. thaliana* genome, a search has been carried out with every *A. thaliana* protein kinase as query in the rice genome and *vice-versa*. Protein kinases identified in trans-genomic comparison were further categorised as nearest homologues if they share similar domain organisation, and their catalytic kinase domains shared highest similarity to same protein kinase subfamily. A set of 184 and 334 protein kinases has been identified in *A. thaliana* and *O. sativa* respectively, for which close homologues with similar domain organisation could not be detected across the two genomes.

The following sections describe the domain organisation of protein kinases conserved across the two genomes and those protein kinases detected in only one of the genomes.

#### 3.5. Domain organisation of protein kinases shared across *A. thaliana* and *O. sativa* genome

The functional diversity associated with the protein kinase containing gene products involved in numerous molecular interactions are mostly attributed to the intrinsic regulatory modules and other regulatory proteins known to interact with protein kinases. The comparison of the domain organisation of protein kinases across different species therefore highlights the unique

and common functionality associated with the key signalling processes. The following section describes the domain organisation of protein kinases that are largely confined to the plant kingdom including those protein kinases that have been newly identified in the current genome-wide analysis.

The calcium dependent protein kinases in plants contain intrinsic regulatory domains (Fig. 1a) that bind to Ca<sup>2+</sup> (EF-hand motifs of CDPKs) or domains like NAF (Fig. 1b) that binds to calcium sensor proteins like calcineurin-B (Snf1 kinase family).

A unique class of stress response proteins of the plant receptor like kinase (RLCK) family lacking trans-membrane segment are characterised by N-terminal Usp domain, central kinase catalytic domain and C-terminal U-box domain (Fig. 1c). These protein kinases are suggested to be related to the pathogen induced kinases like pto1 in tomato plants (Kerk et al., 2003). The presence of an U-box domain that is a modified ring finger domain found in ubiquitination domains suggest a link between stress-related response and protein degradation pathways to induce resistance against pathogens.

The novel domain organisation of a protein kinase of AGC group includes 2 tandem sets of PAS and PAC domains located at the N-terminal followed by the C-terminal kinase domains (Fig. 1d). The PAS Kinases of non-plant species with single set PAS and PAC domains bear close sequence similarity in their catalytic domain with the CAMK group of protein kinases. The PAS domains have been shown to be sensors of redox, oxidative and nutrient stress signals in Rim15 of yeast, which integrates nutrient and redox stress signals to elicit appropriate response (Cameroni et al., 2004). Further the PAS domains shown to restrain the kinase catalytic activity in human PAS kinases (Amezcuca et al., 2002). A single set of PAS and PAC domains have also been identified in certain Raf family kinases of *A. thaliana* and *O. sativa*. These observations suggest the recruitment of additional functionality to the AGC and Raf kinases of the plant species.

Certain Raf kinase members of plants contain N-terminal ACT domains preceding the kinase catalytic domain (Fig. 1e). The ACT domains have been noted in metabolic enzymes that are regulated by changes in the amino-acid flux dependent regulation of the associated protein domains (Schuller et al., 1995). A novel mode of regulation of Raf kinases by ACT domains that serve as allosteric binding sites to various amino acids is hence suggested.

Unusual MEKs with C-terminal NTF2 domain have also been identified in the two plant genomes (Fig. 1f). Previous studies on MAPK signalling pathways in eukaryotic cells have reported the phosphorylation of MAPK localised in the nucleus by MEKs (Mizukami et al., 1997; Kim and Kahn, 1997; Gonzalez et al., 1993). The translocation of MEKs into the nucleus appears to be a pre-requisite for subsequent phosphorylation of MAPK that eventually regulates gene expression. The identification of NTF2 domain, known to interact with Ran-GDP and nucleoporins (Stewart et al., 1998), in MEKs suggests its role in nuclear import of MEKs.

An atypical protein kinase (At3g59410) and the only recognisable protein kinase of the protein translation regulating kinase family in *A. thaliana*, whose catalytic domain bears high similarity with GCN2 member of translation kinase (supplementary

information 1), has N-terminal RWD domain, central kinase catalytic domain and C-terminal HGTP (His, Gly, Thr and Pro)-anticodon binding domains (Fig. 1g). The catalytic domain bearing similarity to other eukaryotic protein kinases involved in control of protein translation suggests a role of this protein kinase in phosphorylation of EF2-like translation factors during protein biosynthesis. The accessory domains of the protein kinase further suggests the interactions of this protein kinase with other components of the translation machinery mediated by the HGTP anticodon binding domain similar to the non-catalytic domains of t-RNA synthetases (Wolf et al., 1999). RWD domain of the protein kinase, previously identified in proteins involved in nuclear proteins interacting with nucleic acids (Doerks et al., 2002) also indicates the association of the protein kinases with diverse biomolecules involved in protein biosynthesis. The rice genome also encodes a single protein kinase of the translation regulatory subfamily which contains only the N-terminal RWD domain preceding the kinase catalytic domain and no detectable HGTP-anticodon binding domain at the C-terminal. These distinct domains of the kinases involved in protein biosynthesis could reflect in the functional differences associated with regulation of translation.

The plant receptor kinase family of *A. thaliana* and *O. sativa* include members with similar extra-cellular and intra-cellular domains that have been implicated in various signalling processes (Shiu and Bleeker, 2001; Torii, 2000). Members of previously well characterised plant receptor kinases (Fig. 1h) that are found in both the plant species include the LRR kinases with varying number of LRRs and legume–lectin family of kinases with extra-cellular legume–lectin like domains. Receptor kinases with extra-cellular LysM domains (Radutoiu et al., 2003; Madsen et al., 2003) involved in recognition of microbes and Thaumatin receptor kinases previously identified in *A. thaliana* (Shiu and Bleeker, 2001) have been identified in the rice genome. These groups of receptor kinases hence appear to be conserved across diverse plant species. Receptor kinases with TNFR-like cysteine rich repeats have also been observed in *A. thaliana* (At3g09780) and *O. sativa* (11669.m04341|Os03g43670). Such repeats have been previously identified in CRINKLY receptor (CR4) kinases of maize and have been shown to influence growth factors mediated differentiation (Jin et al., 2000; Schafer and Schmulling, 2002).

### 3.6. Distinct domain organisation of STKs in *A. thaliana*

The Ser/Thr receptor kinase family in plants contributes to the high functional diversity of the plant kinases in accordance with the numerous external stimuli perceived by plants. This feature is also apparent from the current analysis that suggests a high number of protein kinases of the group that differ between the two plants have constituted the plant receptor family in these plants. Domain organisations of few plant receptor kinases in *A. thaliana* that have not been detected in the rice genome are discussed below.

Protein kinase (At2g26290) containing C-terminal CaMBD has a kinase catalytic domain bearing high similarity to that of plant receptor kinase subfamily (Fig. 2a). This novel receptor like kinase lacking a trans-membrane segment, adds on to the existing sets of non-fungal, non-metazoan protein kinases that are reg-

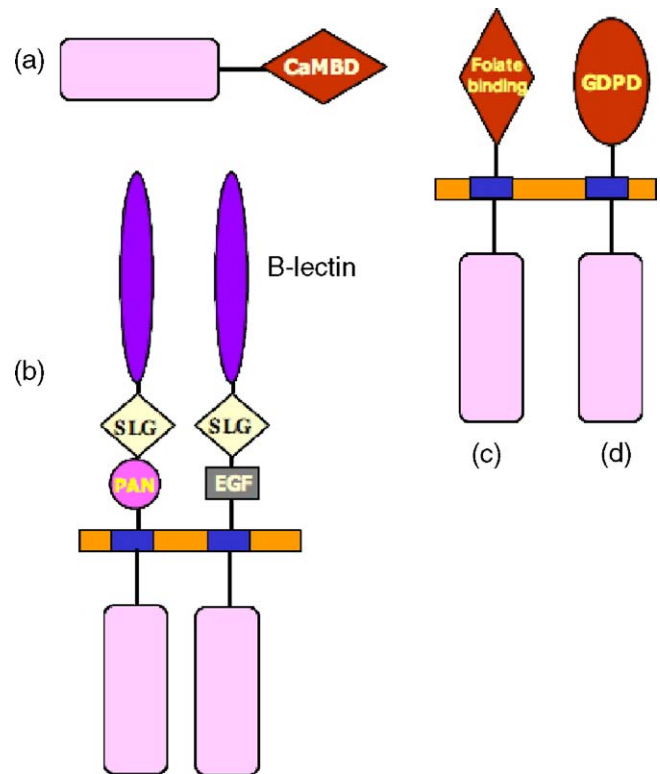


Fig. 2. Protein kinases of *A. thaliana* with domain composition distinct from those observed in protein kinases of *O. sativa*. Cell membrane and trans-membrane regions are shown in pink, blue and orange respectively. The colouring scheme is similar to Fig. 1. (a) At2g2690; (b) At1g65790, At5g35370; (c) At1g25390; (d) At1g66980.

ulated by  $\text{Ca}^{2+}$  or calmodulin described in previous sections. A single gene-product (At2g32800) with two kinase/kinase-like domains sharing high similarity to plant receptor kinase families has been identified. However the C-terminal “kinase domain” lacks the catalytic aspartate, which is also the case with pseudo kinase domain of JAK family of tyrosine kinases observed in higher (non-plant) eukaryotes.

Among the receptor kinases with trans-membrane segment (Fig. 2b,c,d) extra-cellular domains like agglutinin, folate-binding domain, GDPD have been identified in *A. thaliana* genome with no apparent counterparts in the rice genome. Periplasmic and cytoplasmic activity of GDPD is observed in *E. coli*. The occurrence of an extra-cellular catalytic domain like GDPD in the protein kinase suggests its activity on the membrane glycerophosphoryl diesters in response to signals that are currently unknown.

A large number (35) of B-lectin (previously classified as Agglutinin family in PFAM) receptor kinases (Fig. 2b) have been identified in *A. thaliana* (Shiu and Bleeker, 2003, 2001). They could be categorised into two types based on their extra-cellular domain organisation. The extra-cellular region containing N-terminal agglutinin domain, S-domain, and PAN domain comprise one group of Agglutinin receptor kinases. The PAN domain is replaced by EGF domain in the extra-cellular region of other group of agglutinin receptor kinases. The PAN and EGF domains are known to play key roles in protein–carbohydrate and protein–protein interactions respectively. The SRKs have been shown to be the determinants of SI specificity in flowering plants of



Brassicaceae (Kachroo et al., 2002). The B-lectin domain is found in lectins known to bind to D-mannose that adopt a  $\beta$ -prism II fold. This receptor kinase family also differs from the legume lectin domains (LegA and LegB) that adopt concanavalin-A/ glucanase folds identified in other receptor kinases of *A. thaliana* and *O. sativa* (Fig. 1h).

An unusual S-domain receptor kinase (At1g11300) with two kinase domains has been identified in *A. thaliana*. The extra-cellular kinase catalytic domain is sandwiched between the S-domain and the second agglutinin domain while the intra-cellular kinase domain follows the juxta-membrane region.

Dual kinase domains have also been observed in another gene product (At1g56140) in LRR-receptor kinase family where the extra-cellular kinase domain is flanked by LRR repeats followed by trans-membrane segment and an intra-cellular kinase domain. The kinase catalytic domains in the two gene products (At1g11300 and At1g56140) have well conserved catalytic aspartate.

### 3.7. Domain organisation of distinct protein kinases in *O. sativa*

Extensive proliferation and diversification of the plant receptor kinase family is observed in the rice genome. This is apparent from the domain organisation of the various plant receptor kinases of rice that are likely to have resulted from multiple events of domain duplication, shuffling and recruitment.

Protein kinases associated with the plant receptor like kinases family lacking trans-membrane regions (RLCKs) contain distinct nucleic acid interacting enzymatic DNA-methylase (11669.m01219|Os03g12570), Transposase and peptidase domains (11687.m04185|Os11g44390), MuDR transposase and Zn-finger like (11687.m04170|Os11g44240) domains (Fig. 3a,b,c). In addition to the dual kinase domain containing gene products like those identified in *A. thaliana*, the rice genome encodes protein kinase of RLCK family that have three and four kinase catalytic domains. They belong to the group of SRK family

associated with self-incompatibility during plant development. Catalytic residue (Asp) is completely conserved in the triple (11687.m00660|Os11g07230) and quadruple (11681.m03382|Os09g37840) kinase catalytic domain containing gene products.

The rice RLCKs and RLKs also contain distinct carbohydrate binding domains. Like SRKs of *A. thaliana*, the rice SRK related kinases have a D-mannose binding B-lectin like domain preceding the S-domain and the PAN domain (Eg: 11681.m03382|Os09g37840, Fig. 3d). The B-lectin domain adopts a  $\beta$ -prism II scaffold differing from the 3-D fold of legume lectins.

Another protein kinase of RLCK family has comprises of N-terminal protein kinase domain and C-terminal mannose binding Jacalin-like domains (Eg: 11670.m02849|Os04g30040, Fig. 3e) which adopts a  $\beta$ -prism I fold.

The RLK family of protein kinases further shows extreme divergence in terms of their extra-cellular and intra-cellular domains giving rise to multifunctional kinases (Fig. 3f).

A plant receptor kinase (Fig. 3f) has been identified in rice (11668.m02570|Os02g27310) that has an extra-cellular SCP domain, which is previously identified in plant proteins induced during pathogenesis (Fernandez et al., 1997).

### 3.8. Inferences on functions of protein kinases in rice

A large number of plant protein kinases have been extensively characterised and their roles in regulation of plant growth and development is well documented (Laurie and Halford, 2001; Morris and Walker, 2003; Stone and Walker, 1995). However the knowledge about the functions of protein kinases confined to individual species of plants is very limited. By cross-genomic comparisons, it has been possible to relate the protein kinases of a given genome with experimentally characterised orthologues in the other genomes, thus providing clues to the biological roles of less well characterised or uncharacterised plant kinases.

In the current study, *A. thaliana* protein kinases and their functional roles available in the Swiss-Prot database have been

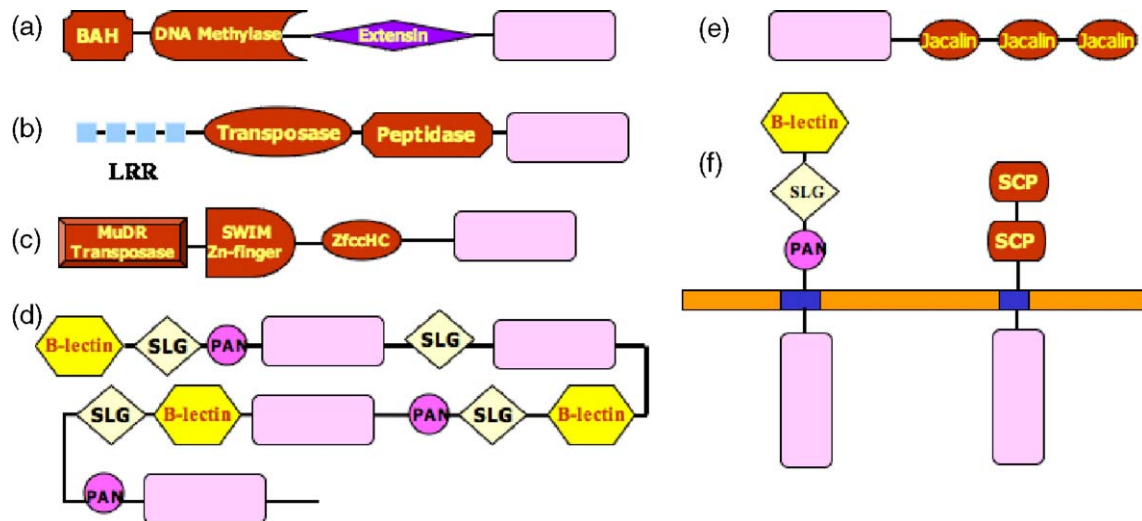


Fig. 3. Protein kinases of *O. sativa* with distinct domain organisation compared to the protein kinase in *A. thaliana*. SLG: S-domain of S-locus glycoproteins; Zf: zinc finger. BAH: bromo adjacent homology domain, a protein–protein interaction domain found in many gene-silencing proteins (Callebaut et al., 1999). (a) 11669.m01219|Os03g12570; (b) 11687.m04185|Os11g44390. (c) 11687.m04170|Os11g44240; (d) 11681.m03382|Os09g37840; (e) 11670.m02849|Os04g30040; (f) 11668.m02570|Os02g27310. The colouring scheme is similar to Figs. 1 and 2.

obtained and a search for their closest homologues in the rice genome has been carried out. The inferences on functions of the putative counterparts in rice genome have been drawn from such comparisons. Swiss-Prot database contains only small proportion of *A. thaliana* protein kinases which are well documented. Nevertheless, the functions of these protein kinases seem to be critical to the basic physiological processes related to the growth and development of plants. Protein kinases of *O. sativa*, sharing highest sequence similarity to the well characterized PKs of *A. thaliana* have been identified in this study.

Apart from various protein kinases, playing key roles in cell cycle control (CDKs), regulation of gene expression (MAPKs) and glycogen synthase kinases that are conserved across other eukaryotic species, closest homologues of *A. thaliana* protein kinases related to plant-specific physiological processes have been identified (Table 2) that share sequence identity greater than 50% (50%–80%) over the entire length of the corresponding *A. thaliana* PKs. Further, pairwise alignment between orthologs has been made using CLUSTALW in order to check conservation of important catalytic residues like Glycine in Glycine-loop and Lysine, Glutamic acid in subdomain II and III respectively which make salt bridge and is important for proper orientation of phosphate group of ATP. Our analyses have shown that all these important residues mentioned above are conserved in all the orthologous pairs. The sequence alignments obtained for various sets of orthologous proteins are presented in Supplementary information 2. Tree diagram containing each ortholog pair with bootstrap values has been shown in Supplementary information 3. These rice PKs may share similar physiological functions as that of nearest kinases of *A. thaliana*. Regulatory functions in signalling pathways related to ethylene response, host defense, brassinosteroid signalling, self-incompatibility, organogenesis, somatic embryogenesis and other aspects of growth and development has been assigned to these protein kinases.

Further experimental characterisation of these putative rice kinases associated with basic functions that coordinate various physiological processes in plants would enhance our current understanding on the underlying principles of regulation.

#### 4. Discussion

*A. thaliana* and *O. sativa* encode large number of Ser/Thr protein kinases that have diverged extensively compared to the protein kinases of other eukaryotic species. Plant specific sub-families such as the plant receptor kinase, plant receptor like kinases and calcium dependent kinases are well conserved in both the species.

Atypical domain organisation observed in few protein kinases of AGC, MEK, MEKKs, Raf kinases, and the translation regulatory kinase families has expanded the spectrum of signals and regulatory mechanisms that influence their activity.

Plant receptor kinase family comprises a major fraction of the kinase complement in the two genomes and consists of receptors with large variations in the composition of intra-cellular and extra-cellular domains. An analysis of the domain organisation of this family has revealed the occurrences of diverse families of lectin like extra-cellular domains, which are shared and specifically conserved within the species. These lectin receptor kinases are therefore suggested to play a key role in diverse glycoproteins or oligosaccharide interactions enabling their cell–cell communication and recognition of various microbes.

Multifunctional kinases of RLCK with nucleotide specific enzymatic modules and carbohydrate binding domains have been identified in rice. Few receptor kinases in rice are further associated with ras like domains. The roles of such multifunctional protein kinases, undetected so far in any other eukaryotic species, in the plant signalling remains to be elucidated and hence are priority targets for experimental studies.

Comparison of the *A. thaliana* protein kinases of known functions with those encoded in the rice genome has enabled the identification of rice protein kinases that are likely to be involved in various pathways associated with growth and development of plants.

The systematic analysis carried out on the protein kinase repertoires in the two plant genomes has thus provided clues to the

Table 2  
List of *A. thaliana* protein kinases associated with plant specific functions and their putative orthologues in the *O. sativa* genome

Protein kinases of <i>A. thaliana</i>	Biological function	Gene product codes of putative orthologous protein kinase in <i>O. sativa</i>
AFC1_ARATH	LAMMER kinase family regulated by the hormone ethylene.	11670.m03452 Os04g35700 (73%)
BAK1_ARATH BRI1_ARATH	Involved in Brassinosteroid signalling employed for self-rejection	11670.m03646 Os04g38480 (78%) 11667.m05092 Os01g52050 (52%)
CC2A_ARATH CC2B_ARATH	Cell division control proteins.	11669.m00099 Os03g01850 (82%) 11667.m06760 Os01g67160 (69%)
CLV1_ARATH	Receptor kinase that controls shoot and floral meristem size.	11680.m05016 Os06g50340 (57%)
CTR1_ARATH	Negative regulator of ethylene response.	11668.m03069 Os02g32610 (53%)
KII0_ARATH	Role in signal transduction mechanism regulating carbohydrate metabolism in plants.	11682.m04350 Os05g45420 (73%)
KP19_ARATH	Ribosomal protein S6 Kinase induced during cold/salinity stress.	11673.m04788 Os07g48290 (65%)
PBS1_ARATH	Interacts with virulence protein of <i>P. syringae</i> and regulates host defense mechanism during infection.	11668.m02942 Os02g30900 (71%) 11667.m07118 Os01g71000 (68%)
PSKR_ARATH	Receptor kinase activated by binding of phytosulphokine regulating organogenesis and somatic embryogenesis	11670.m05697g (51%)

The extent of similarity of the putative orthologues is indicated by the percentage identity (shown in parentheses) shared between the sequences over their full length amino acid sequences.

functions of previously uncharacterised plant protein kinases, which can be further investigated to understand their roles in regulation of cellular and physiological processes of plants.

### Acknowledgments

This research is supported by the International Senior Fellowship in biomedical sciences to N.S. by the Wellcome Trust, UK, National Bioscience Award to N.S. by the Department of Biotechnology, India and by the computational genomics project supported by the Department of Biotechnology, India. A.K. was supported by a fellowship from the Council of Scientific and Industrial Research, India and by Wellcome Trust, UK.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.gene.2006.05.016](https://doi.org/10.1016/j.gene.2006.05.016).

### References

- Abo-el-Saad, M., Wu, R., 1995. A rice membrane calcium-dependent protein kinase is induced by gibberellin. *Plant Physiol.* 108, 787–793.
- Albrecht, V., Ritz, O., Linder, S., Harter, K., Kudla, J., 2001. The NAF domain defines a novel protein–protein interaction module conserved in Ca<sup>2+</sup>-regulated kinases. *EMBO J.* 20, 1051–1063.
- Altschul, S.F., et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* 25, 3389–3402.
- Agrawal, G.K., Iwahashi, H., Rakwal, R., 2003. Rice MAPKs. *Biochem. Biophys. Res. Commun.* 302, 171–180.
- Amezcuca, C.A., Harper, S.M., Rutter, J., Gardner, K.H., 2002. Structure and interactions of PAS kinase N-terminal PAS domain: model for intramolecular kinase regulation. *Structure (Camb.)* 10, 1349–1361.
- Anamika, Srinivasan, N., Krupa, A., 2005. A genomic perspective of protein kinases in *Plasmodium falciparum*. *Proteins* 58, 180–189.
- Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- Bassett, C.L., Nickerson, M.L., Cohen, R.A., Rajeevan, M.S., 2000. Alternative transcript initiation and novel post-transcriptional processing of a leucine-rich repeat receptor-like protein kinase gene that responds to short-day photoperiodic floral induction in morning glory (*Ipomoea nil*). *Plant Mol. Biol.* 43, 43–58.
- Bateman, A., et al., 2004. The Pfam protein families database. *Nucleic Acids Res.* 32, D138–D141.
- Becraft, P.W., 1998. Receptor kinases in plant development. *Trends Plant Sci.* 3, 384–388.
- Becraft, P.W., 2002. Receptor kinase signaling in plant development. *Annu. Rev. Cell Dev. Biol.* 18, 163–192.
- Boeckmann, B., et al., 2003. The SWISS-PROT protein knowledge base and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31, 365–370.
- Bogre, L., Okresz, L., Henriques, R., Anthony, R.G., 2003. Growth signalling pathways in *Arabidopsis* and the AGC protein kinases. *Trends Plant Sci.* 8, 424–431.
- Braun, D.M., Walker, J.C., 1996. Plant transmembrane receptors: new pieces in the signaling puzzle. *Trends Biochem. Sci.* 21, 70–73.
- Cameron, E., Hulo, N., Roosen, J., Winderickx, J., De Virgilio, C., 2004. The novel yeast PAS kinase Rim15 orchestrates G(0)-associated antioxidant defense mechanisms. *Cell Cycle* 4, 462–468.
- Chen, K., Du, L., Chen, Z., 2003. Sensitization of defense responses and activation of programmed cell death by a pathogen-induced receptor-like protein kinase in *Arabidopsis*. *Plant Mol. Biol.* 53, 61–74.
- Choe, S., et al., 2002. *Arabidopsis* brassinosteroid-insensitive dwarf12 mutants are semidominant and defective in a glycogen synthase kinase 3beta-like kinase. *Plant Physiol.* 130, 1506–1515.
- Clark, G.B., Thompson Jr., G., Roux, S.J., 2001. Signal transduction mechanisms in plants: an overview. *Curr. Sci.* 80, 170–177.
- Cock, J.M., 2000. A receptor kinase and the self-incompatibility response in *Brassica*. *Advances in botanical research incorporating advances in plant pathology*, vol. 32, pp. 269–298.
- Czernic, P., et al., 1999. Characterization of an *Arabidopsis thaliana* receptor-like protein kinase gene activated by oxidative stress and pathogen attack. *Plant J.* 18, 321–327.
- Daniel, X., Sugano, S., Tobin, E.M., 2004. CK2 phosphorylation of CCA1 is necessary for its circadian oscillator function in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3292–3297.
- Deeken, R., Kaldenhoff, R., 1997. Light-repressible receptor protein kinase: a novel photo-regulated gene from *Arabidopsis thaliana*. *Planta* 202, 479–486.
- Doerks, T., Copley, R.R., Schultz, J., Ponting, C.P., Bork, P., 2002. Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res.* 12, 47–56.
- Dornelas, M.C., Van Lammeren, A.A., Kreis, M., 2000. *Arabidopsis thaliana* SHAGGY-related protein kinases (AtSK11 and 12) function in perianth and gynoecium development. *Plant J.* 21, 419–429.
- Du, L., Chen, Z., 2000. Identification of genes encoding receptor-like protein kinases as possible targets of pathogen- and salicylic acid-induced WRKY DNA-binding proteins in *Arabidopsis*. *Plant J.* 24, 837–847.
- Eddy, S.R., 1998. Profile hidden Markov models. *Bioinformatics* 14, 755–763.
- Fernandez, C., Szyperski, T., Bruyere, T., Ramage, P., Mosinger, E., Wuthrich, K., 1997. NMR solution structure of the pathogenesis-related protein P14a. *J. Mol. Biol.* 266, 576–593.
- Friedrich, P., Curvetto, N., Giusto, N., 1999. Cyclic AMP-dependent protein phosphorylation in guard cell protoplasts of *Vicia faba* L. *Biocell* 23, 203–209.
- Gaspero, D.G., Cipriani, G., 2003. Nucleotide binding site/leucine-rich repeats, Pto-like and receptor-like kinases related to disease resistance in grapevine. *Mol. Genet. Genomics* 269, 612–623.
- Goff, S.A., et al., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100.
- Gonzalez, F.A., Seth, A., Raden, D.L., Bowman, D.S., Fay, F.S., Davis, R.J., 1993. Serum-induced translocation of mitogen-activated protein kinase to the cell surface ruffling membrane and the nucleus. *J. Cell Biol.* 122, 1089–1101.
- Gribskov, M., et al., 2001. PlantsP: a functional genomics database for plant phosphorylation. *Nucleic Acids Res.* 29, 111–113.
- Hanks, S.K., Quinn, A.M., Hunter, T., 1988. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* 241, 42–52.
- Hardie, D.G., 1999. Plant protein serine/threonine kinases: classification and functions. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 97–131.
- Harmon, A.C., Gribskov, M., Harper, J.F., 2000. CDPKs — a kinase for every Ca<sup>2+</sup> signal? *Trends Plant Sci.* 5, 154–159.
- Hasegawa, P.M., Bressan, R.A., Zhu, J.K., Bohnert, H.J., 2000. Plant cellular and molecular responses to high salinity. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51, 463–499.
- He, Z.H., Cheeseman, I., He, D., Kohorn, B.D., 1999. A cluster of five cell wall-associated receptor kinase genes, Wak1–5, are expressed in specific organs of *Arabidopsis*. *Plant Mol. Biol.* 39, 1189–1196.
- Herve, C., et al., 1999. Characterization of the *Arabidopsis* lecRK-a genes: members of a superfamily encoding putative receptors with an extracellular domain homologous to legume lectins. *Plant Mol. Biol.* 39, 671–682.
- Hirt, H. (Ed.), 2000. *Results and Problems in Cell Differentiation: MAP Kinases in Plant Signal Transduction*. Springer (Hiedelberg) press.
- Hunter, T., Plowman, G.D., 1997. The protein kinases of budding yeast: six score and more. *Trends Biochem. Sci.* 22, 18–22.
- Hwang, I., Sze, H., Harper, J.F., 2000. A calcium-dependent protein kinase can inhibit a calmodulin-stimulated Ca<sup>2+</sup> pump (ACA2) located in the endoplasmic reticulum of *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6224–6229.
- Ichimura, K., et al., 2002. Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends Plant Sci.* 7, 301–308.
- Jin, P., Guo, T., Becraft, P.W., 2000. The maize CR4 receptor-like kinase mediates a growth factor-like differentiation response. *Genesis* 27, 104–116.

- Jonak, C., Hirt, H., 2002. Glycogen synthase kinase 3/SHAGGY-like kinases in plants: an emerging family with novel functions. *Trends Plant Sci.* 7, 457–461.
- Joubes, J., et al., 2000. CDK-related protein kinases in plants. *Plant Mol. Biol.* 43, 607–620.
- Kachroo, A., Nasrallah, M.E., Nasrallah, J.B., 2002. Self-incompatibility in the Brassicaceae: receptor–ligand signaling and cell-to-cell communication. *Plant Cell* 14, S227–S238.
- Kerk, D., Bulgrien, J., Smith, D.W., Barsam, B., Veretnik, S., Gribskov, M., 2002. The complement of protein phosphatase catalytic subunits encoded in the genome of *Arabidopsis*. *Plant Physiol.* 129, 908–925.
- Kerk, D., Bulgrien, J., Smith, D.W., Gribskov, M., 2003. *Arabidopsis* proteins containing similarity to the universal stress protein domain of bacteria. *Plant Physiol.* 131, 1209–1219.
- Kim, S.J., Kahn, C.R., 1997. Insulin regulation of mitogen-activated protein kinase kinase (MEK), mitogen-activated protein kinase and casein kinase in the cell nucleus: a possible role in the regulation of gene expression. *Biochem. J.* 323, 621–627.
- Kono, A., Umeda-Hara, C., Lee, J., Ito, M., Uchimiya, H., Umeda, M., 2003. *Arabidopsis* D-type cyclin CYCD4;1 is a novel cyclin partner of B2-type cyclin-dependent kinase. *Plant Physiol.* 132, 1315–1321.
- Kostich, M., et al., 2002. Human members of the eukaryotic protein kinase family. *Genome Biol.* 3 (RESEARCH0043. Electronic publication 2002 Aug 22).
- Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.
- Krupa, A., Srinivasan, N., 2002. The repertoire of protein kinases encoded in the draft version of the human genome: atypical variations and uncommon domain combinations. *Genome Biol.* 3 (RESEARCH0066.1-14).
- Krupa, A., Srinivasan, N., 2005. Diversity in domain architectures of Ser/Thr kinases and their homologues in prokaryotes. *BMC Genomics* 6, 129.
- Krupa, A., Preethi, G., Srinivasan, N., 2004a. Structural modes of stabilization of permissive phosphorylation sites in protein kinases: distinct strategies in Ser/Thr and Tyr kinases. *J. Mol. Biol.* 339, 1025–1039.
- Krupa, A., Abhinandan, K.R., Srinivasan, N., 2004b. KinG: a database of protein kinases in genomes. *Nucleic Acids Res.* 32, D153–D155.
- Laurie, S., Halford, N.G., 2001. The role of protein kinases in the regulation of plant growth and development. *Plant Growth Regul.* 34, 253–265.
- Lee, J., et al., 2003. Cell cycle function of a rice B2-type cyclin interacting with a B-type cyclin-dependent kinase. *Plant J.* 34, 417–425.
- Liu, W., Xu, Z.H., Luo, D., Xue, H.W., 2003. Roles of OsCKI1, a rice casein kinase I, in root development and plant hormone sensitivity. *Plant J.* 36, 189–202.
- Lu, S.X., Hrabak, E.M., 2002. An *Arabidopsis* calcium-dependent protein kinase is associated with the endoplasmic reticulum. *Plant Physiol.* 128, 1008–1021.
- Madsen, E.B., et al., 2003. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* 425, 637–640.
- Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S., 2002. The protein kinase complement of the human genome. *Science* 298, 1912–1934.
- Marchler-Bauer, A., et al., 2003. CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res.* 31, 383–387.
- Martin, M.L., Busconi, L., 2001. A rice membrane-bound calcium-dependent protein kinase is activated in response to low temperature. *Plant Physiol.* 125, 1442–1449.
- Mizukami, Y., Yoshioka, K., Morimoto, S., Yoshida, K., 1997. A novel mechanism of JNK1 activation. Nuclear translocation and activation of JNK1 during ischemia and reperfusion. *J. Biol. Chem.* 272, 16657–16662.
- Montesano, M., Brader, G., Palva, E.T., 2003. Pathogen derived elicitors: searching for receptors in plants. *Mol. Plant Pathol.* 4, 73–79.
- Morris, E.R., Walker, J.C., 2003. Receptor-like protein kinases: the keys to response. *Curr. Opin. Plant Biol.* 6, 339–342.
- Morrison, D.K., Murakami, M.S., Cleghon, V., 2000. Protein kinases and phosphatases in *Drosophila* genome. *J. Cell Biol.* 150, F57–F62.
- Murase, K., et al., 2004. A membrane-anchored protein kinase involved in *Brassica* self-incompatibility signaling. *Science* 303, 1516–1519.
- Muschietti, J., Eyal, Y., McCormick, S., 1998. Pollen tube localization implies a role in pollen–pistil interactions for the tomato receptor-like protein kinases LePRK1 and LePRK2. *Plant Cell* 10, 319–330.
- Nimmo, H.G., Carter, P.J., Fewson, C.A., McNaughton, G.A., Nimmo, G.A., Wilkins, M.B., 1990. Regulation of phosphoenolpyruvate carboxylase: an example of signal transduction via protein phosphorylation in higher plants. *Adv. Enzyme Regul.* 30, 121–131.
- Outlaw Jr., W.H., Du, Z., Xia Meng, F., Aghoram, K., Riddle, K.A., Chollet, R., 2002. Requirements for activation of the signal-transduction network that leads to regulatory phosphorylation of leaf guard-cell phosphoenolpyruvate carboxylase during fusicoccin-stimulated stomatal opening. *Arch. Biochem. Biophys.* 407, 63–71.
- Plowman, G.D., Sudarsanam, S., Bingham, J., Whyte, D., Hunter, T., 1999. The protein kinase of *Caenorhabditis elegans*: a model for signal transduction in multicellular organisms. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13603–13610.
- Radutoiu, S., et al., 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* 425, 585–592.
- Riou, C., Herve, C., Dabos, P., Pacquit, V., Lescure, B., 2002. Expression of an *Arabidopsis* lectin kinase gene, lecRKa1 is induced during senescence, wound-healing and in response to oligogalacturonic acids. *Plant Physiol. Biochem.* 40, 431–438.
- Satterlee, J.S., Sussman, M.R., 1998. Unusual membrane-associated protein kinases in higher plants. *J. Membr. Biol.* 164, 205–213.
- Savaldi-Goldstein, S., Sessa, G., Fluhr, R., 2000. The ethylene-inducible PK12 kinase mediates the phosphorylation of SR splicing factors. *Plant J.* 21, 91–96.
- Schafer, S., Schmulling, T., 2002. The CRK1 receptor-like kinase gene of tobacco is negatively regulated by cytokinin. *Plant Mol. Biol.* 50, 155–166.
- Schuller, D.J., Grant, G.A., Banaszak, L.J., 1995. The allosteric ligand site in the Vmax-type cooperative enzyme phosphoglycerate dehydrogenase. *Nat. Struct. Biol.* 2, 69–76.
- Shiu, S.H., Bleecker, A.B., 2001. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10763–10768.
- Shiu, S.H., Bleecker, A.B., 2003. Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiol.* 132, 530–543.
- Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F., Li, W.H., 2004. Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* 16, 1220–1234.
- Stewart, M., Kent, H.M., McCoy, A.J., 1998. Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran. *J. Mol. Biol.* 277, 635–646.
- Stone, J.M., Walker, J.C., 1995. Plant protein kinase families and signal transduction. *Plant Physiol.* 108, 451–457.
- Strain, E., Muse, S.V., 2005. Positively selected sites in the *Arabidopsis* receptor-like kinase gene family. *J. Mol. Evol.* 61, 325–332.
- Tena, G., Asai, T., Chiu, W.L., Sheen, J., 2001. Plant mitogen-activated protein kinase signaling cascades. *Curr. Opin. Plant Biol.* 4, 392–400.
- Thomashow, M.F., 1999. Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 571–599.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Torii, K.U., 2000. Receptor kinase activation and signal transduction in plants: an emerging picture. *Curr. Opin. Plant Biol.* 3, 361–367.
- Umeda, M., Umeda-Hara, C., Yamaguchi, M., Hashimoto, J., Uchimiya, H., 1999. Differential expression of genes for cyclin-dependent protein kinases in rice plants. *Plant Physiol.* 119, 31–40.
- Vandepoel, K., Raes, J., De Veylder, L., Rouze, P., Rombauts, S., Inze, D., 2002. Genome-wide analysis of core cell cycle genes in *Arabidopsis*. *Plant Cell* 14, 903–916.
- Verica, J.A., He, Z.H., 2002. The cell wall-associated kinase (WAK) and WAK-like kinase gene family. *Plant Physiol.* 129, 455–459.
- Wang, D., Harper, J.F., Gribskov, M., 2003. Systematic trans-genomic comparison of protein kinases between *Arabidopsis* and *Saccharomyces cerevisiae*. *Plant Physiol.* 132, 2152–2165.

- Wang, G.L., et al., 1998. Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. *Plant Cell* 10, 765–779.
- Ward, P., Equinet, L., Packer, J., Doerig, C., 2004. Protein kinases of the human malaria parasite *Plasmodium falciparum*: the kinome of a divergent eukaryote. *BMC Genomics* 5, 79.
- Wolf, Y.I., Aravind, L., Grishin, N.V., Koonin, E.V., 1999. Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res.* 9, 689–710.
- Wrzaczek, M., Hirt, H., 2001. MAP kinase pathways: how many and what for? *Biol. Cell* 93, 81–87.