Volume 282, number 1, 98–102 FEBS 09635 April 1991

ADONIS 001457939100342B

Diversity of the protein kinase gene family in rice

Xin-Hua Feng and Shain-dow Kung

Center for Agricultural Biotechnology, Maryland Biotechnology Institute, Department of Botany, University of Maryland, 1206 H.J. Patterson Hall, College Park, MD20742, USA

Received 15 January 1991; revised version received 22 February 1991

Multiple genes have been found to encode families of protein kinases in animals and yeasts. Little is known of the diversity of protein kinase families in plants. We have used the polymerase chain reaction to identify members of protein kinase gene family in rice. We have cloned eight partial cDNA sequences from which deduced amino acid sequences contained conserved sequences or amino acid residues characteristic of catalytic domains of eukaryotic protein serine/threonine kinases. Our results suggest that there is great complexity in the protein kinase gene family in plants and that protein phophorylation may play an as important role in plants as in other eukaryotes.

Protein kinase; Rice seedling; Gene multiplicity; Signal transduction; Polymerase chain reaction

1. INTRODUCTION

Protein kinases are important components of eukaryotic signal transducing systems. They control diverse cellular processes by phosphorylating specific proteins. Various protein kinases respond to different signals and serve in different regulatory roles in cells. They share the analogous function of controlling substrate proteins by transferring the terminal phosphate of ATP to specific serine, threonine, or tyrosine residues. They also share a basic structure for their analogies in function; the catalytic domains, including ATP-binding and phosphorylation sites, are generally conserved among protein kinases studied to date [1,2].

The availability of amino acid sequence data for a number of protein kinases leads to several molecular genetic approaches to the identification and characterization of additional protein kinase family members. Degenerate oligonucleotide sequences encoding the conserved regions can serve as specific hybridization probes to identify some protein kinase members from a DNA library. By using this method, DNA clones containing protein kinases have been identified in mammals [3], yeast [4] and plants [5,6]. Alternatively, the conserved regions can be used to generate oligonucleotide primers to amplify a particular target DNA sequence in a polymerase chain reaction (PCR) [7–9]. In this study, we used PCR to obtain eight sequences encoding putative protein kinases from rice (*Oryza sativa* L.) (designated OSPKs). Seven of these have not been reported before. Here we report the cloning, sequencing and characterization of these eight rice protein kinase sequences.

Correspondence address: S.D. Kung, Center for Agricultural Biotechnology, University of Maryland, College Park, MD 20742, USA

2. EXPERIMENTAL

Rice (*Oryza sativa* L.) IR36 was grown at 25°C under 16 h photoperiod in a growth chamber. Entire 7-day-old seedlings were harvested for poly(A) RNA isolation. Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer. Chemicals for sequencing gels were obtained from BioRad (NY, USA).

One μg of rice poly(A) RNA was reverse transcribed at 42°C for 60 min in a 20 μl reaction mixture containing 1 × PCR buffer (20 mM Tris-Cl, pH 8.4, 50 mM KCl, 1.5 mM KCl, 1.5 mM $MgCl_2$), 1 mM dNTPs, 0.5 μg of oligo$(dT)_{12-18}$, 20 units of RNAs in ribonuclease inhibitor (Promega, WI, USA), and 200 units of M-MLV reverse transcriptase (Life Technologies, MD, USA). This reaction mixture was brought to a volume of 100 μl with an 80 μl PCR mixture containing 1 × PCR buffer, 100 μM dNTPs, 10 μg of each phosphorylated oligonucleotide primer and 2 units of AmpliTaq DNA polymerase (Perkin-Elmer/Cetus, CT, USA), then overlaid with 100 μl of mineral oil. PCR was performed for 30 cycles on a thermal cycler (Perkin-Elmer/Cetus) with temperature cycles as follows: 94°C for 1 min, 53°C for 2 min, and 72°C for 2 min. Fifteen μl of the amplified products were fractionated on a 2% NuSieve/1% SeaPlaque agarose (FMC, Bioproducts, ME, USA) gel. DNA ranging from 120 to 370 bp in agarose was pooled, phenol-extracted and ethanol-precipitated. Pooled DNA was then cloned in the SmaI site of M13mp10 vector. Sequencing of recombinant M13 clones was performed using Sequenase (United States Biochemicals, CA, USA).

3. RESULTS

The sequences of fully degenerate oligonucleotides corresponding to the two conserved regions used in PCR are shown in Fig. 1A and B. After 30 cycles of PCR amplification using rice cDNA as the template, a very intense band of approximately 370 bp could be detected in ethidium bromide-stained agarose gel (Fig.

 Published by Elsevier Science Publishers B.V.

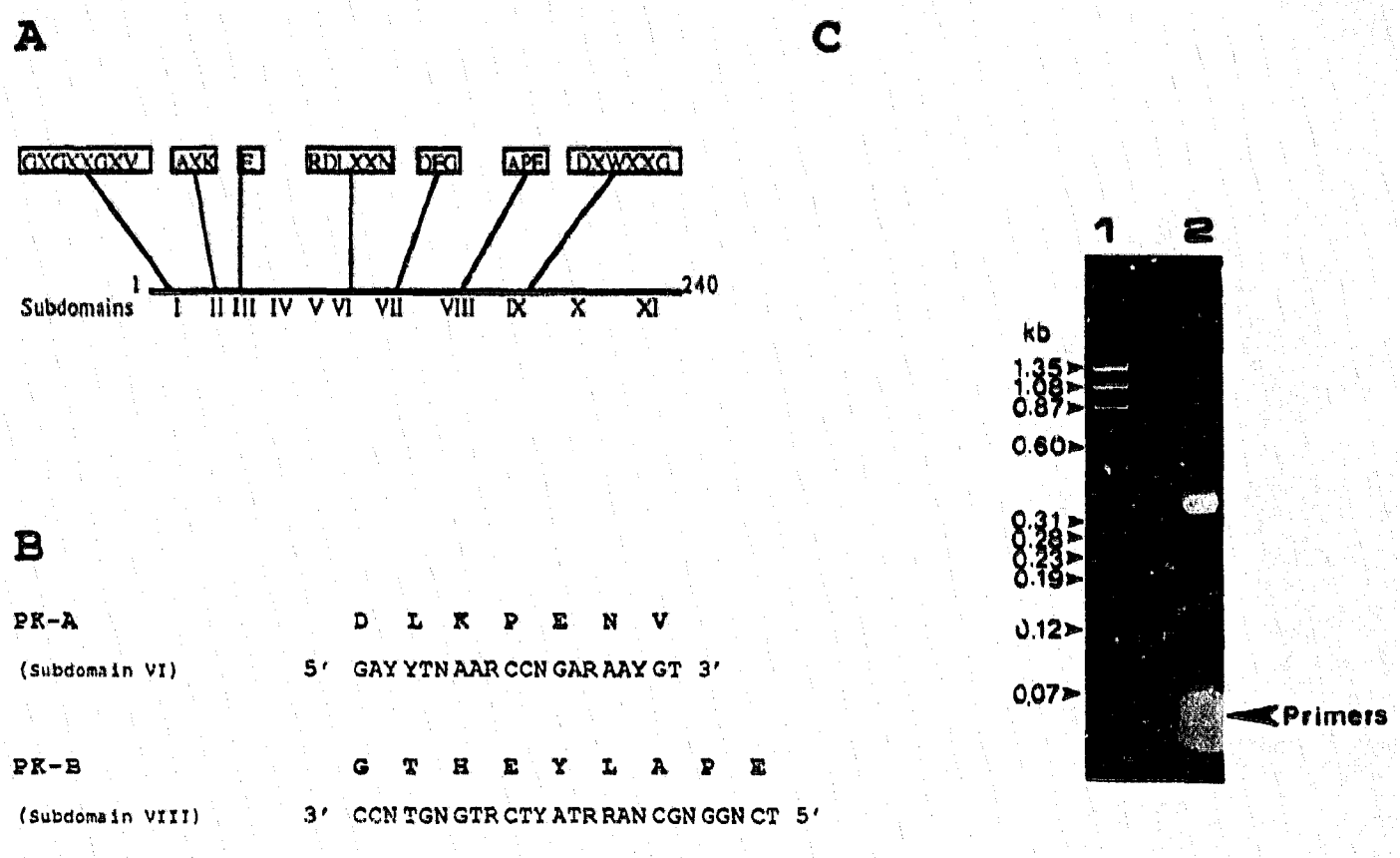

Fig. 1. Polymerase chain reaction. (A) Highly conserved subdomains in eukaryotic protein kinases. The positions of these subdomains are approximately within 240-amino-acid protein kinase domain, as indicated by numbers. One-letter code is used for amino acid residue, of which X represents any amino acid residue. (B) Sequences of the oligonucleotide used in PCR. The primer PK-A is the sequence that specified those amino acids DLKPENV and the primer PK-B is the complement of the coding sequence specifying the amino acids GTHEYLAPE. (C) Fractionation of PCR products by electrophoresis on a 2.0% NuSieve/1.0% SeaPlaque agarose gel and visualization by staining with ethidium bromide. (Lane 1) molecular weight standards; (lane 2) amplified PCR products.

1C). Other minor products were also present (Fig. 1C). Based on the knowledge that every DNA fragment ranging from 120 bp to 370 bp between the subdomains DLKPENV and GTHEYLAPE (single-letter codes for amino acids) may represent protein kinases in plants, these fragments were pooled and cloned except that the 370 bp fragment was treated separately. PCR was independently carried out three times to ensure no errors in cloned sequences (Table I).

Totally 47 recombinant M13 clones were obtained and their insert sequences were determined (TAble I). A total of 8 independent putative rice protein kinase sequences (OSPKs) were identified. The nucleotide sequences of all these putative protein kinase clones (OSPK clones) are shown in Fig. 2. The amino acid sequences deduced from OSPK clones were divided into two groups: Group I was compared for their similarities with published putative bean protein kinase PVPK.1 and rice protein kinase G11A [5] and maize protein kinase sequence 90.7 [6], and Group II with bovine protein kinase C-β form [10] and cAMP-dependent protein kinase β-form [11], rat Ca^{2+}/calmodulin-dependent protein kinase II α-form [12], yeast $p34^{cdc2}$ kinase [13] and a putative maize receptor type protein kinase ZmPK1 [14] (Fig. 3). For all the sequences compared, we found a consensus region of DFD/G, flanked by hydrophobic or neutral residues. Of group I kinases that had a conserved stretch of DFD, OSPK1.1 was

Table I

Number of M13 clones sequenced from independent PCRs

	Number of M13 clones		
PCR clone	PCR1	PCR2	PCR3
OSPK1.1	1+	2+, 1−	5+, 1−
OSPK1.2	2+	2−	3+, 2−
OSPK1.3	0	1+, 1−	1−
OSPK2.1	1−	0	2+
OSPK4.4	2+	1+	2+, 1−
OSPK4.6	1+	1−	1−
OSPK5.1	0	1−	3+
OSPK10.1	1+	2+, 1−	2+, 3−

Numerical figures under each individual PCR represent the number of M13 clones sequenced. The symbol '+' or '−' stands for the sense or antisense strand cloned in single stranded M13.

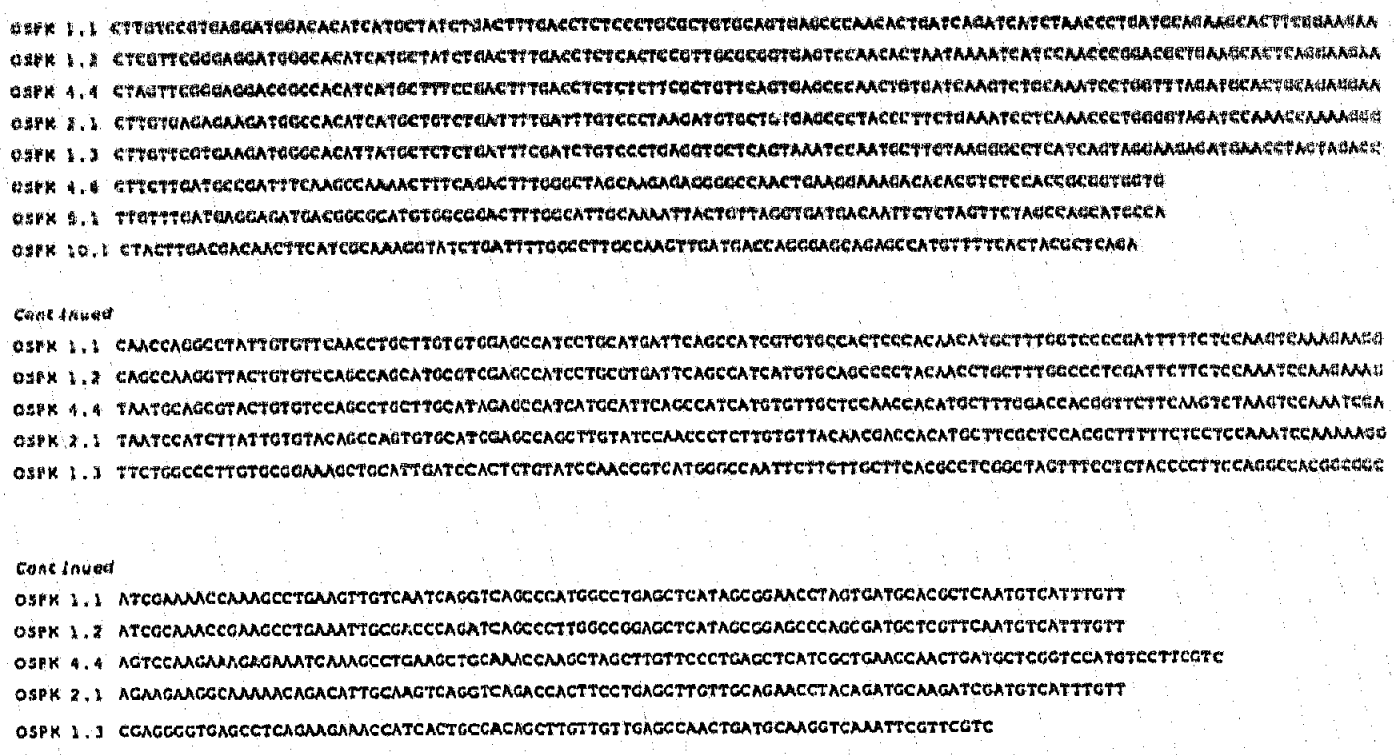

Fig. 2. Nucleotide sequences of rice protein kinase cDNA clones. PCR primer sequences are not included.

GROUP I

```
PVPK.1      LVREDGHIMLSDFDLSLRCSVSPTLVKSSN....NLQTKSSGYCVQPSCIEPTCVMQPDCIKP.SCFTPRFLSG.KSKKDKKSKPKNDMHNQVTPLPELIAEPTNARSMSFV
90.7        LVREDGHIMLSDFDLSLRCAVSPTLLRSSNPSGDN.QKGNPAYCVQPVCIEPAC.MQPSCVTTTTCFSPRFFSS.KSKEKKDKKAKADWANQVRPLPELVAEPTDAKSMSFV
OSPK 1.1    LVREDGHIMLSDFDLSLRCAVSPTLIRSSNPDAEALRKNNQAYCVQPACVEPSCMIQPSCATPTTCFGPRFFS..KSKKDRKPKP..EVVNQVSPWPELIAEPSDARSMSFV
OSPK 1.2    LVREDGHIMLSDFDLSLRCAVSPTLIKSSNPDAEALRKNSQGYCVQPACVEPSCVIQPSCAAPTTCFGPRFFS..KSKKDRKPKP..EIATQISPWPELIAEPSDARSMSFV
OSPK 4.4    LVREDGHIMLSDFDLSLRCSVSPTVIKSANPGLDALQRNNAAYCVQPACIEPSC.IQPSCVAPTTCFGPRFFKSKSKSKSKKEKSKPEAANQASLFPELIAEPTDARSMSFV
OSPK 2.1    LVREDGHIMLSDFDLSLRCAVSPTLLKSSNPGVDPNQKGNPSYCVQPVCIEPAC.IQPSCVTTTTCFAPRFFSS.KSKKEKKAKT..DIASQVRPLPELVAEPTDARSMSFV
OSPK 1.3    LVREDGHIMLSDFDLSLRCSVNPMLVRASSVGRDEPSR.PSGPCAE.SCIDPLC.IQPSHAN.SSCFTPRLVSS.TPSRPRRPRG..EPQKK.PSLPQLVVEPTDARSNSFV
Consensus   LVREDGHIMLSDFDLSLRC VSPTL  SSN           YCVQP C EP C  QPSC     CF PRF S          K       Q    PEL AEP DARSMSFV
```

GROUP II

Bovine PKC-β	MLDSEG...HIKIADFGMCKE..NIWDGVTTKTTC
Bovine cAPK-β	LIDHQG...YIQVTDFGFAKR....VKGR.TWTLC
Rat CaMII-α	LLASKLKGAAVKLADFGLAIE..VEGEQQAWFGFA
Yeast $p34^{cdc2}$	LIDKEG...NLKLADFGLARS..FGVPLRNYTHEI
ZmPK1	LLDQAF...EPKITDFGLVKLLNRGGSTQNVSHVR
OSPK 4.6	LLDADF...KPKLSDFGLAREGPTEGKTHVSTAVV
OSPK 5.1	LFDEEM...TAHVADFGIAKLLLGDDNSLVLASMP
OSPK 10.1	LLDDNF...IAKVSDFGLAKL.MTREQSHVFTTLR
Consensus	L D K DFG A

Fig. 3. Amino acid sequence comparison of known protein kinases [5,6,10-14] and the newly-identified sequences. Consensus sequence indicated identities or strong similarities at the corresponding positions. Boldtype indicated the highly conserved short stretch **DFD** or **DFG**.

Table II

Sequence similarity among individual members of protein kinases

OSPK	1.1	1.2	4.4	2.1	1.3	PVPK.1
1.1	-	82.1	75.7	70.4	56.4	61.5
1.2	91.7	-	72.9	63.0	53.5	64.1
4.4	75.2	75.2	-	68.2	56.1	61.0
2.1	71.6	71.6	71.2	-	58.7	69.2
1.3	44.4	41.4	49.5	50.0	-	56.5
PVPK.1	65.8	67.7	65.2	64.5	49.5	-

For each pair of aligned sequences, the number of identical amino acid residues (below the diagonal) or nucleotide sequences (above the diagonal) was divided by the total number of positions in the aligned sequence. In either case, sequences in PCR primer regions were excluded. Similarity value was expressed as a percent. PVPK.1 represents partial sequences of catalytic domains of bean protein kinases [5].

identical to the corresponding region of the rice putative kinase known as G11A reported by Lawton et al. [5]. These putative kinases contained identical or nearly identical residues around the DFD stretch. OSPK1.1 and OSPK1.2 are more closely related to each other than either to others (Table II): OSPK1.1 and OSPK1.2 show 91.7% amino acid homology, whereas OSPK1.1 and OSPK1.3 exhibit a 44.4% homology and OSPK1.2 and OSPK1.3 exhibit a 41.4% homology. The high level of homology between OSPK1.1 and OSPK1.2 raises the possibility that one may have been duplicated from the other during evolution. When they are compared to the animal homologs, there are insertional sequences including lysine-rich region and 4-6 tandem repeats of $CX_{2-3}P$ (X for any amino acid residue) between the two primer regions in the domain of all Group I products. These features have not yet been observed in other eukaryotes, suggesting that these insertional sequences may be unique to plants; otherwise, they should be identified in other organisms by using PCR. The Group II kinases, OSPK4.6, OSPK5.1 and OSPK10.1 have the size (between the two primer regions) of typical protein kinases in most eukaryotic organisms. They possess a stretch consisting of DFG, which is also common in typical protein kinases in animals and lower eukaryotes. These products may represent functional homologs of PKC and cAPK subfamilies of animal protein kinases [10,11].

Southern hybridization showed that these OSPK clones did not cross-hybridize to each other and each individual clone hybridized to different restriction fragments in the rice genome (data not shown). This analysis as well as sequence data indicate that these OSPK clones represent 8 independent genes encoding putative protein kinases in rice.

4. DISCUSSION

In this study we have identified 8 protein serine/threonine kinase genes in rice. However, the use of genomic DNA as a template and different sets of primers (even corresponding to the same subdomains as ours) at different annealing temperatures for PCR may reveal a larger set of members of the rice protein kinase family, including tyrosine kinases. Nonetheless, this phenomenon of protein kinase gene multiplicity agrees with the current view that protein kinase genes in mammals, *Drosophila* and yeast are complex families containing multiple genes of like classes of enzymes [15]. The comparison of partial catalytic domains among OSPKs suggests that protein kinases encoded by OSPK genes might be involved in various and possible multiple signal transduction pathways. These OSPKs may have evolved to allow differential control of gene expression in different cell or tissue types. It may also be possible that the similar catalytic domains are attached to diverse regulatory domains for differential binding of specific effector ligands. With the availability of full sequence for these putative protein kinase genes, the structure-function relationships in this family may become clearer.

Until recently, only few putative protein kinase genes were isolated from plants [5,6,14,16]. The recently-developed PCR technology has proven to be very useful not only to identify conserved genes in diverse species but also to isolate additional members of gene families in single species [7-9]. By using PCR we have identified seven novel members of what appear to be the rice protein serine/threonine kinase gene family. As the number of genes isolated from plants coding for regulatory proteins such as protein kinases and GTP-binding proteins [17-19] begins to increase, the establishment and elucidation of the interrelationship between these proteins in signal transduction during plant growth and development will become possible in the future.

REFERENCES

[1] Hanks, S.K., Quinn, A.M. and Hunter, T. (1988) Science 241, 42-52.
[2] Soderling, T.R. (1990) J. Biol. Chem. 265, 1823-1826.
[3] Hanks, S.K. (1987) Proc. Natl. Acad. Sci. USA 84, 388-392.
[4] Levin, D.E., Hammond, C.I., Ralston, R.D. and Bishop, J.M. (1987) Proc. Natl. Acad. Sci. USA 84, 6035-6039.
[5] Lawton, M.A., Yamamoto, R.T., Hanks, S.K. and Lamb, C.J. (1989) Proc. Natl. Acad. Sci. USA 86, 3140-3144.
[6] Biermann, B., Johnson, E.M. and Feldman, L.J. (1990) Plant Physiol. 94, 1609-1615.
[7] Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) Science 230, 1350-1354.
[8] Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science 239, 487-491.
[9] Gould, S.J., Subramani, S. and Scheffler, I.E. (1989) Proc. Natl. Acad. Sci. USA 86, 1934-1938.
[10] Coussens, L., Parker, P., Rhee, L., Yang-Feng, T., Chen, E., Waterfield, M., Francke, U. and Ullrich, A. (1986) Science 233, 859-866.
[11] Shoji, S., Ericsson, L.H., Walsh, K.A., Fischer, E.H. and Titani, K. (1983) Biochemistry 22, 3702-3709.

[12] Hanley, R.M., Means, A.R., Ono, T., Kemp, B.E., Burgin, K.E., Waxham, N. and Kelly, P.T. (1987) Science 237, 293–297.
[13] Hindley, J. and Phear, G.A. (1984) Gene 31, 129–134.
[14] Walker, J.C. and Zhang, R. (1990) Nature 345, 743–746.
[15] Hunter, T. (1987) Cell 50, 823–829.
[16] Feller, H.S. and Jacobs, T.W. (1990) Proc. Natl. Acad. Sci. USA 87, 5397–5401.
[17] Ma, H., Yanofsky, M.F. and Meyerowitz, E.M. (1990) Proc. Natl. Acad. Sci. USA 87, 3821–3825.
[18] Matsui, M., Sasamoto, S., Kunieda, T., Nomura, N. and Ishizaki, R. (1989) Gene 76, 313–319.
[19] Palme, K., Diefenthal, T., Sander, C., Vingron, M. and Schell, J. (1989) in: The Guanine-Nucleotide Binding Proteins (Bosch, L., Kraal, B. and Parmeggiani, A. eds) pp. 273–284, Plenum, New York.