Universidade Federal do Rio Grande do Sul

Instituto de Biociências

Programa de Pós-Graduação em Genética e Biologia Molecular

Aspectos Funcionais e Evolutivos da Família WAK em *Oryza sativa*

Luiz Felipe Valter de Oliveira

Orientação: Rogerio Margis

Porto Alegre, abril de 2011

Universidade Federal do Rio Grande do Sul

Instituto de Biociências

Programa de Pós-Graduação em Genética e Biologia Molecular

Aspectos funcionais e evolutivos da família WAK em *Oryza sativa*

Luiz Felipe Valter de Oliveira

Orientação: Rogerio Margis

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do grau de Mestre em Genética e Biologia Molecular.

Porto Alegre, abril de 2011

O presente trabalho foi desenvolvido no Laboratório de Genomas e Populações de Plantas (Centro de Biotecnologia – UFRGS) e no Núcleo de Genômica Funcional de Plantas (Departamento de Genética – UFRGS), sendo este subvencionado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Aos meus pais, meu irmão e a Ana, que sempre estiveram ao meu lado, me apoiando, incondicionalmente, nas minhas decisões pessoais e profissionais, mesmo que isso significasse a minha distância.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer ao meu orientador, prof. Dr. Rogério Margis, que me aceitou como seu orientado, mesmo sem me conhecer. Que durante esses dois anos proporcionou todas as condições para que eu pudesse me aperfeiçoar como pesquisador, além de estar sempre presente e disposto ajudar. Gostaria de agradecer profundamente por todo apoio pessoal que recebi durantes estes dois, dos quais não tenho palavras para descrever, apenas posso dizer foram (e continuam sendo) um dos pilares para minha vida pessoal como profissional. Gostaria de agradecer da mesma forma a profª. Drª Márcia Margis, por estar presente no desenvolvimento do meu trabalho, pelos ensinamentos, ajudas , oportunidades e conselhos. Agradeço novamente a vocês dois por não terem se limitado em serem apenas orientadores, sim mestres. Mais uma vez, obrigado.

Ao PPGBM, principalmente a Comissão de Pós-Graduação de 2009/2010 por permitirem que nós, da representação discente, pudéssemos participar ativamente no desenvolvimento da nossa pós-graduação. Gostaria de agradecer também, em especial, ao Elmo Cardoso, por ser uma pessoa ímpar, por toda atenção e responsabilidade, além da amizade que desenvolvemos nestes dois anos.

Aos professores Dr. Gilberto Sachetto Martins; Dr. Diego Bonatto, Dr. José Artur Bogo Chies e a Prof. Drª Luciana M. P. Pasaglia por terem aceito fazer parte da banca de avaliação deste trabalho e contribuir com o mesmo.

Á todos os meus colegas do LGPP, Andreia, Cláudia, Guilherme Cordenonsi, Guilherme Loss, Fernanda, Franceli, Josiane, Júlio, Lorrayne e Maurício por proporcionarem um excelente ambiente de trabalho, pela competência, responsabilidade e seriedade com a qual toda a equipe leva o nosso ambiente de trabalho, pela amizade com o qual fui recebido no laboratório, pelas discussões e os cafezinhos de todos os dias.

À toda a equipe do LabDros (UFSM), que fizeram, e continuam fazendo parte da minha vida. Em especial ao Gabriel, Ronaldo e Paloma por toda

a amizade e o apoio que ainda recebo. A Nina Roth Mota por ter sido fundamental na minha vinda para Porto Alegre, e ao Marcos Oliveira de Carvalho, pela amizade, companheirismo e confiança nesta jornada profissional que iniciamos juntos.

Aos meus grandes amigos Juliano e Giovanni Montagner Madruga, pela amizade e apoio pessoal que recebo ao longo de todos esses anos em que nos conhecemos.

Ao meu irmão, Fernando, por todo o companheirismo e amizade, que vai além do parentesco que nos une.

À Ana, que esteve ao meu lado nestes últimos quatro anos, que conseguiu superar (quase) tranquilamente todos os problemas de um relacionamento a distância, e que na reta final foi a principal catalisadora para a conclusão desta dissertação. Te amo.

E por final, gostaria de deixar toda minha gratidão e admiração aos meus pais, que além do amor e apoio, devotaram suas vidas para proporcionar as condições necessárias para que eu e meu irmão pudéssemos seguir nas nossas escolhas pessoais e profissionais.

Sumário

RESUMO

O ambiente é um sistema dinâmico que está sempre mudando e a capacidade de reconhecimento dessas modificações é uma característica crucial para a vida. A família gênica RLK (do inglês *Receptor-Like Kinase*) engloba as proteínas receptoras do tipo quinase que estão aptas a reconhecer sinais ambientais, através de seu domínio extracelular, e ativar uma cascata de sinalização por modificações pós-traducionais em outras proteínas utilizando a atividade de fosforilação de seu domínio quinase. A subfamília WAK (do inglês The Wall-Associated Kinase) pertence a família gênica RLK, sendo que algumas proteínas desta subfamília foram identificados associados fisicamente à parede celular, sugerindo estes genes como fortes candidatos para agir como sensores, ligando diretamente o ambiente extracelular com o citoplasma e desencadeando sinais intracelulares. Os genes WAK formam uma grande subfamília em arroz, com 130 genes descritos para a subespécie *japonica*, em comparação aos 27 genes WAK descritos para arabidopsis. A ausência de dados sobre a expansão desta subfamília e suas implicações nas plantas justifica uma análise evolutiva e estrutural entre os membros da subfamília WAK de arabidopsis e arroz, com uma comparação mais extensiva entre os genomas das subespécies de arroz *indica* e *japonica*. Quando a organização e os resíduos conservados do domínio quinase das WAKs foram comparados com a superfamília de proteínas quinases de eucariotos, a identificação de dois grupos distintos de WAKs tornou-se evidente em arabidopsis e arroz. Um grupo é formado somente por OsWAKs

que provavelmente se expandiu depois da separação entre monocotiledônea-dicotiledônea, os quais derivaram para a classe de quinases não-RD. O outro grupo corresponde à classe RD-quinase, sendo esta a classe de quinase mais frequente entre os eucariotos, formado por ambos os genes AtWAK e OsWAK. Além disso, com os resultados de comparação entre os genomas das subespécies *indica* e *japonica* foi possível identificar uma grande variação com relação aos domínios das proteínas e ao padrão de expressão entre os genes OsWAK dessas subespécies. O conjunto de resultados sugere que as WAKs constituem duas subfamílias evolutivamente relacionadas, mas independentes: OsWAK-RD e WAK-nonRD. O reconhecimento desta divisão poderá contribuir para a compreensão das funções e regulação das WAKs.

ABSTRACT

The environment is a dynamic system, which is always changing and the recognition of its modifications is a crucial feature for life. The RLK is a gene family of receptor quinase in plants that are able to recognize the environment signals, through its extracellular domain, and activating a signal cascade through the post-translational modifications of others protein using the phosphorylation activity from the quinase domain. The Wall-Associated Kinase (WAK) is a subfamily from RLK, with some members identified as associated to the cell wall, suggesting these genes are strong candidates to act as sensors directly linking the extracellular environment to the citoplasm and triggering intracellular signals. The WAK is a large subfamily in rice genome, with 130 genes being described in *japonica*, compared to the twenty-seven in *A. thaliana.* The absence of data about this gene subfamily expansion and their implications for plants, justify an evolutionary and structural analysis of *A. thaliana* and *O. sativa* members of the WAK subfamily, with a more extensive comparison between genomes of japonica and indica rice subspecies. When the organization of plant WAK domains and conserved residues are compared with those of other eukaryotes protein kinase superfamilies, the identification of two distinct groups of WAKs become evident in Arabidopsis and Rice. One group corresponds to a cluster containing just OsWAK that should have expanded after the *monocot-dicot* separation, which evolved to form a non-RD kinase class. The other group corresponds to the classical RD-kinases with both AtWAK and OsWAK

representatives. Moreover, by comparing OsWAK from *indica* and *japonica* subspecies was possible to identify a large divergence in the protein domain features and gene pattern expression. We propose that plant WAKs constitutes two evolutionary related but independent subfamilies: the WAK-RD and the WAK-nonRD. The recognition of this division would contribute to the understanding of WAK function and regulation.

CAPÍTULO I – INTRODUÇÃO GERAL

Proteínas quinase em eucariotos – ePK

A capacidade de detectar estímulos do meio extracelular, e processá-los, utilizando esta informação para ativar o correto mecanismo de resposta, é uma das propriedades básicas de todos os organismos vivos. A maioria das proteínas envolvidas neste processo de transdução de sinal apresenta um domínio proteico conhecido como quinase. A superfamília gênica denominada de ePK (proteínas quinase eucarióticas – do inglês eukaryotic protein kinase) compreende todas as quinases encontradas nos eucariotos, sendo considerada uma das maiores famílias gênicas encontras nos eucariotos (Hanks and Hunter, 1995). Estas proteínas apresentam um papel essencial na modificação pós-traducional de muitas outras proteínas, através de sua atividade de fosforilação. Estas modificações são consideradas um dos mais importantes mecanismos de controle das atividades celulares (Hanks, 2003), atuando diretamente na repressão ou ativação de diversas rotas metabólicas. O processo de fosforilação pode ser divido em três etapas principais: i) ligação do ATP (ou GTP) doador do fosfato; ii) ligação da proteína (ou peptídeo) a ser utilizado como substrato; iii) transferência do fosfato para o resíduo aceptor (serina, treonina ou tirosina) da proteína substrato (Hanks and Hunter, 1995).

O domínio quinase das ePK pode ser dividido em doze subdomínios, sendo estes identificados por números romanos (Figura 1). O domínio quinase está dividido em duas principais porções, um pequeno lobo amino-terminal e um grande lobo carboxi-terminal. Os subdomínios do I ao IV

fazem parte do lobo amino terminal, sendo que este lobo está relacionado com a ligação e orientação do doador do grupo fosfato (ATP ou GTP) para a atividade de fosforilação. O lobo carboxi-terminal, formado pelos subdomínios do VIA até o XI, é responsável pelo reconhecimento, ligação e fosforilação do substrato (Hanks, 2003).

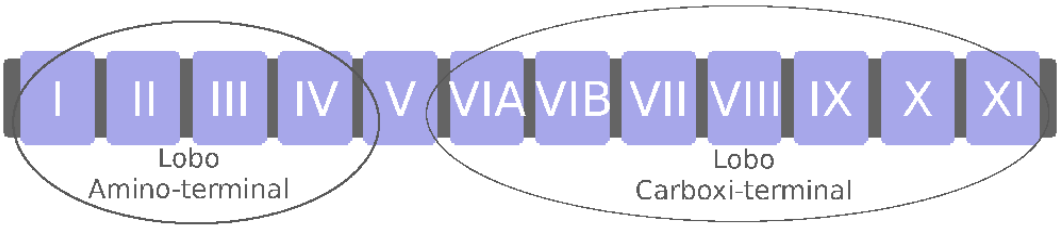

Figura 1: Esquema dos subdomínios do domínio quinase das proteínas ePK.

O subdomínio I possui o motivo conservado GxGxxG, o qual está associado com a ligação do ATP (ou GTP). No subdomínio II é encontrado um resíduo de lisina invariante, o qual é conhecido como essencial para a máxima atividade enzimática, ajudando no ancoramento e orientação do ATP. Este resíduo também forma uma ponte de sal com um resíduo de glutamina invariante no domínio III, sendo este resíduo importante para a estabilização da ligação da lisina com o ATP. O subdomínio IV parece não estar associado com atividades desempenhadas pelo domínio quinase e não apresenta resíduos conservados entre o domínio quinase dos genes da superfamília ePK. O subdomínio V faz a ligação entre os dois lobos, e parece contribuir para o ancoramento do ATP. O subdomínio VIA não parece interagir diretamente na ligação do ATP ou do

peptídeo. O subdomínio VIB possui dois resíduos invariáveis, sendo um ácido aspártico e uma asparagina, formando o motivo DxxxxN. Esses dois resíduos são catalíticos, e atuam diretamente na fosforilação do substrato. O subdomínio VII possui o motivo DFG muito conservado, sendo esta região importante para o auxílio na orientação da transferência do grupo fosfato para o subtrato. O subdomínio VIII apresenta o motivo APE altamente conservado. Esse subdomínio parece ser o mais importante para o reconhecimento e ligação do substrato. O subdomínio IX possui o motivo DxxxxG conservado, o qual parece estar relacionado com a estabilização da região catalítica do subdomínio VIB. O subdomínio X parece ser o menos conservado entre todos os subdomínios não apresentando motivos ou resíduos conservados. No subdomínio XI existe um resíduo de arginina conservado, mas sua função ainda não está definida (Hanks et al., 1988; Hanks and Hunter 1995; Hanks, 2003; Dardick and Ronald 2006; Nolen et al., 2004).

As proteínas quinases, consensualmente, utilizam um mecanismo de regulação de suas atividades baseado na sua própria fosforilação (Krupa et al. 2004). Essa fosforilação ocorre em uma região conhecida como segmento de ativação (em inglês chamado de *activation segment* ou *activation loop*), que se localiza entre o final do subdomínio VII e o motivo APE do subdomínio VIII. Cinco resíduos antes do motivo APE existe uma treonina conservada que é utilizada como sítio de fosforilação para ativação da atividade quinase sobre o substrato. Existem, dentro do domínio quinase, resíduos que são carregados positivamente. Estes resíduos se repelem dentro da estrutura terciária da proteína

desestabilizando-a, e assim, impedindo sua correta ligação e fosforilação do substrato (Johnson et al. 1996; Krupa et al., 2004) A fosforilação do resíduo treonina, do segmento de ativação, neutraliza esse excesso de cargas positivas, permitindo a correta conformação da proteína. Um dos principais resíduos positivos relacionados com essa desestabilização é uma arginina conservada que precede o motivo DxxxxN do subdomínio VIII. A presença ou ausência deste resíduo é utilizada para a classificação dos domínios quinases em RD-quinase ou não-RD, respectivamente (do inglês RD-kinase e non-RD) (Krupa et al., 2004; Dardick and Ronald 2006; Afzal., 2008). As proteínas não-RD, por não possuírem o resíduo positivo arginina, aparentemente não precisam da neutralização para tornarem-se ativas, assim apresentando a atividade de fosforilação constitutiva ou regulada por mecanismos alternativos (Johnson et al., 1996). As proteínas quinases da classe não-RD parecem estar associadas, em plantas e animais, com a resposta imune inata a patógenos, ao passo que as RD-quinase aprecem estar envolvidas em diversos processos fisiológicos (Dardick and Ronald, 2006).

Receptores quinase em plantas – RLK

Até o final da década de 80, não se compreendia muito bem como as células vegetais conseguiam realizar sua comunicação com o meio extracelular, pois era desconhecida a existência de proteínas receptoras que tivessem a capacidade de realizar essa função, como verificado no caso das proteínas receptoras do tipo quinase, que já eram conhecidas em animais.

Em 1990 foi descrita a primeira proteína receptora quinase em plantas, a qual tinha em sua porção extracelular um domínio similar a uma glicoproteína encontrada em *Brassica napus* que estava relacionada com reconhecimento e incompatibilidade do estigma pelo seu próprio pólen, (Walker and Zhang, 1990). Esta e outras proteínas posteriormente descritas, pertencem a família RLK (Receptores do tipo quinase – do inglês Receptor-Like Kinase), a qual faz parte da superfamilia ePK. Estas proteínas são apontadas como chave na percepção de sinais no meio extracelular da superfície das células vegetais, e na inicialização da cascata de uma via de transdução de sinal, para que as células possam responder adequadamente ao estímulo (Figura 2) (Morris and Walker 2003; Shiu and Bleecker 2001b).

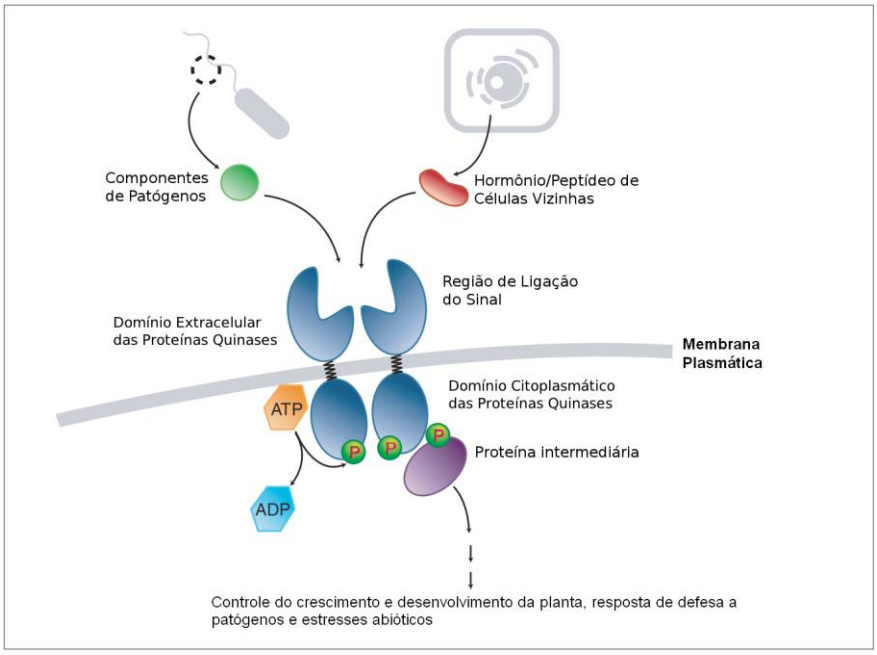

Figura 2: Esquema geral para receptores quinase. Modificado de Shiu and Bleecker, 200b.

As proteínas da família RLK formam uma das maiores e mais diversas famílias de proteínas encontradas em plantas. Os genomas de *Arabidopsis thaliana* e *Oryza sativa*, por exemplo, apresentam mais de 600 e 1.100 membros reportados, respectivamente, para esta família (Shiu and Bleecker 2001a, 2001b, 2003; Shiu et al. 2004; Lehti-Shiu et al. 2009). A familia RLK de plantas forma uma família gênica monofilética com receptores quinases de animais denominada de RLK/Pelle (Shiu and Bleecker 2001a), sendo que, em fungos, esta família não foi encontrada nos genomas até então analisados (Shiu and Bleecker 2001a; Lehti-Shiu et al. 2009). A análise comparativa entre animais e plantas sugere que uma drástica expansão da família RLK/Pelle ficou restrita às plantas, sendo assim, esta expansão deve ter ocorrido após a divergência entre estes dois reinos (Shiu and Bleecker 2001a). As proteínas da família RLK normalmente são encontradas com ambos domínios, extracelular e citoplasmático (quinase). Além disso é possível encontrar proteínas com ausência de um dos domínios. As proteínas classificadas como RLCK (do inglês *Receptor-Like Cytoplasmatic Kinase*) são aquelas que possuem apenas o domínio citoplasmático (o domínio quinase), ou seja, uma proteína que não possui domínio extracelular (Zhang et al. 2005). Também é possível encontrar genes que codificam apenas proteínas relacionadas com a porção extracelular das proteínas RLK, sendo estas chamadas de RLP (do inglês *Receptor-Like Protein*) (Zhang et al. 2005).

Existe uma diversidade muito grande dentro da família das RLK

com relação aos seus domínios extracelulares. Diversos domínios já foram descritos e em conjunto com análises filogenéticas é possível agrupar estas proteínas em diversas subfamílias (Shiu and Bleecker 2001a 2001b, 2003; Shiu et al. 2004; Lehti-Shiu et al. 2009). Como sugerido pela diversidade de domínios extracelulares, e pelo número de genes da família RLK presentes em plantas, estes estão envolvidos nos mais diversos processos celulares (Lehti-Shiu et al. 2009), tais como no desenvolvimento, controlando a regulação da proliferação dos meristemas, especificação de órgãos (Muto *et al.*, 2004; Shpak *et al.*, 2004; Torii *et al.*, 1996), na reprodução (Albrecht *et al.*, 2005; Colcombet *et al.*, 2005) e na transdução de sinal de hormônios (Osakabe et al. 2005). Outros membros da família RLK possuem funções na sinalização ambiental, tanto biótica (Endre *et al.*, 2002; Gomez-Gomez, Boller, 2000; Song *et al.*, 1995) como abiótica (Hou *et al.*, 2005; Sivaguru *et al.*, 2003).

A Subfamília gênica *WAK*

A subfamília gênica denominada de WAK (do inglês Wall-Associated Kinase) pertence à família RLK e apresentam um papel importante na comunicação entre o citoplasma e o meio extracelular (He et al., 1996; He et al., 1998; Anderson et al., 2001; Kohorn, 2001). Os genes da subfamília codificam na porção extracelular da proteína um domínio conservado rico em Cisteínas, conhecido como EGF (figura 3). Em *A. thaliana* existem cinco isoformas (WAK1-WAK5) em tandem no cromossomo 1, que compartilham 86% de identidade no domínio quinase e 40-64% na porção extracelular (He *et al.*, 1999; He *et al.*,

1996). Existem ainda vinte dois genes, posteriormente descritos (Verica and He, 2002) denominados de WAKL1-WAKL22, os quais estão distribuídos ao longo dos outros cromossomos. Em *Oryza sativa* foram descritos 125 genes pertencentes a subfamília WAK, sendo estes denominados de OsWAK (Zhang *et al.*, 2005). Dos 125 genes encontrados, 67 foram classificados como RLK, 28 como RLCK, 13 como RLP, 12 como genes curtos e 5 como pseudo-genes (Zhang *et al.*, 2005). As análises filogenéticas realizadas nesse trabalho, comparando os genes WAKs de *A. thaliana* e *O. sativa,* não resolveram, de forma significativa, as relações evolutivas entre os genes destas duas espécies.

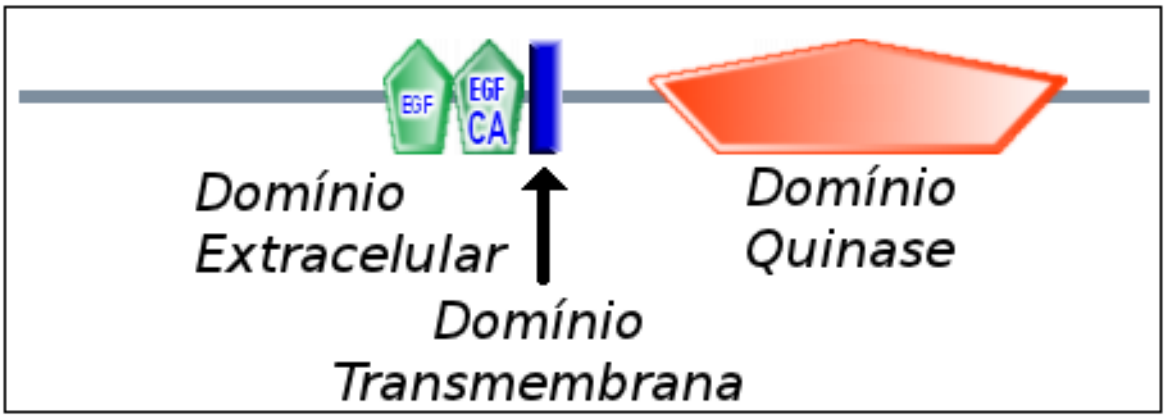

Figura 3: Esquema dos domínios encontrados nas proteínas codificadas pelos genes da subfamília WAK.

Como o próprio nome sugere, as WAK interagem com a parede celular. Isto foi demonstrado através de imunolocalização com anticorpo contra o domínio quinase da WAK1 (He *et al.*, 1996). Epítopos de WAK foram encontrados fortemente associados com a parede celular, uma vez que não foram liberados na

presença de detergentes que solubilizam a membrana celular (Wagner, Kohorn, 2001), sendo apenas liberado após a utilização de substâncias ou proteínas que degradam a parede celular (pectinases). Folhas de plantas transgênicas expressando a fita antisenso de genes WAK reduziram os níveis de proteínas destas e exibiram perda na capacidade de expansão celular (Wagner and Kohorn, 2001), sugerindo que a ligação entre WAK e parede celular tem um papel significante para o controle da expansão celular. Experimentos utilizando ELISA (*enzyme-linked immunosorbent assays*) demonstraram que a ligação do domínio extracelular da WAK1 com ácido poligalacturônico, oligogalacturonídeos e pectinas extraídos da parede celular está sujeita a modulação por cálcio. Estas ligações são impedidas com a esterificação de grupamentos metil, quelantes de cálcio e despolimerização de pectinas (Decreux and Messiaen, 2005; Decreux *et al.*, 2006). Estudos recentes indicam que WAK1 está ligada em endomembranas, e que o seu transporte para a superfície celular é necessário para a correta síntese da parede celular (Kohorn *et al.*, 2006). Em protoplastos foi demonstrado que as pectinas ativam a transcrição do gene *vacuolar invertase* de maneira dependente de WAK2. Além disso, mutantes nulos para WAK2 alteram a regulação do gene MAPK (do inglês *mitogen-activated protein kinase*) pelas pectinas. Análises com microarranjo demonstraram que a WAK2 é necessária para que as pectinas ativem diversos genes em protoplastos, dos quais muitos estão envolvidos na biogênese da parede celular (Kohorn et al., 2009). A utilização do sistema duplo-híbrido (*two-hybrid system*), e ligação *in vitro* da região de um subdomínio recombinante (178-334 aminoácido) do gene WAK1,

possibilitou a identificação da interação de WAK com proteína secretada rica em glicina (*At*GRP3). Foi também demonstrado que o complexo WAK1-GRP3 interage com uma proteína denominada KAPP (*intracellular kinase-associated protein phosphatase*), formando um complexo de 500 KDa (Park *et al.*, 2001). Um dos possíveis alvos deste complexo pode ser a proteína OEE2 (*chloroplast oxygen-envolving enhancer protein 2*), que aparentemente é dependente da interação WAK1-GRP3 para ser fosforilada (Yang *et al.*, 2003), sendo que ainda não é conhecida a significância fisiológica desta interação.

Além de apresentarem importância na formação da parede celular e na expansão celular, as WAK parecem estar envolvidas na resposta a estresses. A WAK1 é induzida em condição de infecção por *Pseudomonas syringae*, na presença de ácido salicílico, ácido 2,2 dicloro isonicotínico (He *et al.*, 1998) e tratamento com alumínio (Sivaguru *et al.*, 2003). O silenciamento de WAK1 através da expressão da sua fita anti-senso produziu plantas suscetíveis a níveis não tóxicos de ácido salicílico e ácido 2,2 dicloro isonicotínico. A expressão ectópica de WAK1 inteira, ou de apenas seu domínio quinase, em *A. thaliana* forneceu a resistência da planta a níveis letais de ácido salicílico (He *et al.*, 1998). A indução de WAK1 em *A. thaliana* por tratamento com alumínio foi verificado na raiz e na parte aérea das plantas tratadas, porém o aumento dos níveis da proteína foi observado apenas na raiz por *northern blot*, e sua super expressão resultou em plantas mais tolerantes à presença de altas concentrações de alumínio. Recentemente, foi demonstrado que um novo gene caracterizado em *O. sativa*, denominado de OsWAK1 pelos autores, parece estar envolvido na

resposta a inoculação do fungo *Magnaporthe oryzae* , pois na presença deste ocorre uma indução na transcrição de OsWAK1 e sua super expressão tornou as linhagens transgênica resistentes à infecção por *M. oryzae* (Li et al., 2009).

Objetivo geral

Este trabalho visa contribuir para o entendimento das relações filogenéticas entre os genes da subfamília WAKs nos genomas de *A. thaliana* e *O. sativa*, bem como elucidar aspectos relacionados ao processo de expansão desta subfamília no genoma do arroz, através de análises comparativas entre as subespécies *indica* e *japonica.*

Objetivos específicos

1 – Identificar e caracterizar os genes WAK no genoma da subespécie *O. sativa ssp indica;*

2 – Comparar os genes WAK entre as duas subespécies de arroz, *indica* e *japonica*, quanto a distribuição no genoma e características das proteínas, por eles condificadas;

3 – Contribuir para o esclarecimento das relações evolutivas entre os genes WAK de *A. thaliana* e *O. sativa*;

4 – Caracterizar o grupo de genes OsWAK que apresentem potencial de redundância funcional;

5 - Comparar o padrão de expressão de genes OsWAK das subespécies *indica* e *japonica* em resposta a estresse por frio.

CAPÍTULO II – EVOLUTION AND EXPRESSION OF WALL-ASSOCIATED KINASE GENE FAMILY IN RICE GENOMES

ABSTRACT

The environment is a dynamic system, which is always changing and the recognition of its modifications is a crucial requirement for life. The *RLK* is a family of receptor kinase in plants that are able to recognize the environment signals, through its extracellular domain, and to activate a signal cascade through the post-translational modifications of others proteins using the phosphorylation activity from the kinase domain. The Wall-Associated Kinase (*WAK*) is a subfamily of *RLK*, containing some members, which were identified as associated to the cell wall, suggesting that these genes are strong candidates to act as sensors linking the extracellular environment and triggering intracellular signals. The *WAK* is a large subfamily in rice genome, with 130 genes described in *japonica*, compared to the twenty-seven in *A. thaliana.* The absence of data about this gene family expansion and their implications for plants, justify an evolutionary and structural analysis between *A. thaliana* and *O. sativa* members of the *WAK* subfamily, with a more extensive comparison of genomes of japonica and indica rice subspecies. When plant *WAK*s domain organization and conserved residues are compared with those of other eukaryotes protein kinase superfamilies, the identification of two distinct groups of WAKs become evident in Arabidopsis and Rice. One group corresponds to a cluster containing just *OsWAK* that should have expanded after the *monocot-dicot* separation, which evolved to form a non-RD kinase class. The other group corresponds to the classical RD-kinases with both *AtWAK* and OsWAK representatives. Moreover, by comparing OsWAK from *indica* and *japonica* subspecies was possible to identify a large divergence in the protein domain features and gene pattern expression. We propose that plant *WAK*s

constitutes two evolutionary related but independent subfamilies: the *WAK*-RD and the *WAK*-nonRD. The recognition of this division would contribute to the understanding of WAK function and regulation.

INTRODUCTION

The environment is a dynamic system with continuous changes which recognition is an important requirement for life. Several environmental factors such as light, temperature, gravity and water are used by plants as signals in many different physiological processes like grown, flowering and dormancy. However, changes in the environment can perturb plant metabolic homeostasis, characterizing an abiotic stress [1]. The ability of recognize signals from the environment and from other adjacent cells, afterward activating correct and specific downstream signaling cascades is an important process for plant responses. Up to the late 80's it was not clear how plant cells communicated with the environment. In 1990 the first receptor kinase protein (called RLK - Receptor Like Kinase) was described in plants [2]. Since then, several different plant *RLK* gene subfamilies have been identified, making it one of the larger and most diverse gene familes found in plants, with more than 600 and 1100 genes belonging to the *RLK* in *Arabidopsis thaliana* and *Oryza sativa,* respectively [3-6].

The *RLK* family belong to the eukaryotic protein kinase (ePK) superfamily, which comprises all of the protein kinases found in the eukaryotes. The ePK superfamily has an important role in the post-translational protein modification through phosphorylation activity. These modifications are considered a major mechanism to control almost all activities of eukaryotic cells [7]. The RLK family are involved in a wide range of process in plants such as hormone signaling [8-10], tissue development [11-14], reproduction [15,16,11,17], symbiosis [18], nodulation [19], responses to biotic and abiotic stress [20-22,9]. A comparative analysis between animal and plant suggests that the large expansion of *RLK* was

restricted in plants and probably occurred after the diverge between these two kingdoms [3]. In general, RLK proteins present an extracellular (EC) and a kinase domain. However, is also possible to find two other of protein structural arrangement related to RLK: (i) genes that encode only a kinase domain, named as RLCK (***R****eceptor-****L****ike* ***C****ytoplasmatic* ***K****inase*) [23,24,6,3], and (ii) genes that encode only the extracellular domain, which are called RLP (***R****eceptor-****L****ike* ***P****rotein*)[24].

Members of the ePK superfamily were described as possessing twelve conserved subdomains. The subdomains I to IV are part the of amino-terminal lobe and are related with ATP binding. Subdomain V corresponds to a link between the amino and the carboxy-terminal lobe. The subdomains VIA, VIB, VII, VIII, IX, X and XI are part of carboxy-terminal lobe, which is related to peptide-substrate binding and phosphorylation activity [25,7,26,27]. Subdomain VIB presents the invariant catalytic motif DxxxxN, which acts directly in substrate phosphorylation and is found in all ePK kinases. One criterion to classify and to group ePKs is the presence of a conserved arginine (R) residue preceding this DxxxxN motif. Such kinases are named as RD-kinases. Conversaly, kinases without this specific residue are known as non-RD kinases. These two classes appear to have their activity regulated by different mechanisms, and seems to be involved in different processes [25,28,29].

The Wall-Associated Kinases (WAKs) [10] comprise a RLK subfamily presents in vascular plants [7]. *WAKs* were initially described in *Arabidopsis thaliana* as a cluster of five genes (*WAK1-5*) [31], sharing in the protein sequence 86% of amino acid identity in the kinase domain and 40-64% in the extracellular

domain. Afterward, twenty-two other *WAK*-like genes were described and called *WAKL1*-*WAKL22* [32]. In *A thaliana,* the *WAK* family show a diverse tissue-specific and development pattern of expression [33,31]. At the present WAK are the only know proteins that physically link cell walls to plasma membrane, being considered to be able to transmit directly extracellular signal, from their EC domain, to cytoplasm through the cytoplasmic kinase domain [30,34,35]. WAK proteins were found strongly bound to the pectin polyssaccharide homogalacturonan, suggesting that their interactions with pectins are important to cell expansion regulation [34]. It was also suggested that the WAK EC domains are noncovalently bound in a calcium-dependent way to the oligogalactorinides, polygalacturonic acid and pectins from cell wall [37]. Another study suggest that Arabidopsis WAK1 is crosslinked to the endomembranes and its transport to the cell surface requires correct cell wall synthesis [38]. Two-hybrid-system and *in vitro* binding experiments suggest that AtGRP3 (a secreted glycine rich protein) are associated specifically with the EC domain of WAK1 and KAPP (kinase-associated protein phosphatase) into a multimeric complex of 500 kDa [39]. The OEE2 (oxygen-evolving enhancer protein 2), a chloroplast protein, was found as interacting and acting as a substrate for WAK1, phosphorylated in an AtGRP3-WAK1 interaction dependent manner. Moreover, the expression of WAK1 are induced by salicylic acid and 2,2 dichloro isonicotinic acid [40], which are related with pathogen infection signaling. The silencing of *WAK1*, through its anti-sense expression, produced susceptible plants to non toxical levels of salicylic acid and 2,2 dichloro isonicotinic acid. However, the ectopic expression of *WAK1* generated plants more tolerant to toxic levels of salicylic acid [40]. An aluminum treatment

induces the *WAK1* expression in both leaf and root tissues, but increasing of protein levels was only observed in the root. The overexpression of this gene led to the increasing of tolerance to high aluminum concentrations in *A. thaliana*, suggesting that *WAK1* may have a crucial role in plant defense against Aluminum toxicity [41].

Wang *et al. 2005 describe,* in *O. sativa* genome, 125 genes belonging to the *WAK* subfamily were identified, which were designed as *OsWAKs* [24]. Sixty-seven of these genes were classified as RLK, twenty-eight as *RLCK*, thirteen as *RLP*, twelve as short genes and five as pseudogenes [24]. It was previously suggested that *OsWAK* expansion occurred after the *O. sativa* and *A. thaliana* separation [6,24,22]. Recently, *OsWAK1* was suggested to be involved in rice response against *Magnaporthe oryzae* infection, since, its expression is induced in the presence of this fungus. In addition the overexpression of this gene led to the transgenic plants more resistant to the *M. oryzae* infection [42].

In plants, the *WAK* subfamily appears to be one of the most important receptors for cell signaling, being related with cell wall synthesis and responses to biotic and abiotic stresses. Besides these major roles in cellular signal transduction and responses, questions remain about the implications of *OsWAKs* expansion and their potential gene subfunctionalization, and about the relationship between the *OsWAK* genes in rice and their orthologous in arabidopsis. In this work we investigated the phylogenetic relationship of WAK subfamily among the *A. thaliana* and two *O. sativa* subspecies: *indica* and *japonica*. Moreover, twenty-two *OsWAK* gene expression, for both *indica* and *japonica*, was investigated in response to cold stress, and the pattern of structural redundancy and divergences

of OsWAK EC and kinase protein domains was compared.

RESULTS

Reannotation of OsWAK in *japonica* genome

Currently, the 130 *japonica OsWAK* (Rice Genome Annotation project - http://rice.plantbiology.msu.edu) are described as follow: 72 *RLK*, 27 *RLCK*, 12 *RLP*, 14 short gene and 5 pseudo-genes. To improve our search of the indica genome, we perform a manually reannotation for all the *OsWAK* genes described for RGPA v7 through two reannotation procedures (see Material and Methods). The first one procedures is based only predicted protein domain and cDNA analysis. The second one procedures is based on new splicing prediction and future domain analysis of the predicted protein from this splicings, for the annotated genomic region of each *OsWAK* gene (for more detail see Material and Methods). Through this two reannotation procedures we suggest the reclassification of seventeen genes (Table 1). Using the first procedure we proposed reannotation for seven genes (*OsWAK2, OsWAK7, sWAK33, OsWAK76, OsWAK96, OsWAK109, OsWAK84*) and with second one procedure for ten genes (*OsWAK12, OsWAK18b, OsWAK21, OsWAK23, OsWAK35a, OsWAK37, OsWAK66, OsWAK83, OsWAK85 and OsWAK88*). In comparison of proposed japonica reannotation with results from indica genome analysis, we found thirteen *OsWAK* matching the classification between japonica and indica, two genes were not found in indica and two genes were found with different classification (see Table 1). All the results for *japonica* and *indica OsWAK* genomic analysis are compiled on Supplementary Table S1.

OsWAK in *indica* characterization and comparison with *japonica*

An extensive characterization of orthology between Japonica and Indica WAK genes were conducted. Hundred eleven sequences of *indica* genome closely related with *japonica OsWAK* were found. According to the protein domain features predicted for these genes, we could classify sixty-four genes as *RLK*, twenty-eight genes as *RLCK*, fourteen genes as *RLP* and three as short gene (Table 2). In the Indica genome we found two genes (*OsWAK26_ind and OsWAK52_ind*) that were impossible to be classified. The predicted protein of *OsWAK26_ind* gene does not present kinase domain and the EC domain were separated in two fragments by an insertion of 18,509 nucleotides. The *OsWAK52_ind* was found in Chromosome 4, and the predicted protein was not found the EC domain, but the gene sequence referente to kinase domain was found in the Superscaffold 2267, not mapped in genome (Table 3), suggest may this gene is placed in a low-quality genomic assembly region. Two genes, *OsWAK98_ind* and *OsWAK102_ind* (Table 3) are duplicated in the *indica* genome. The gene *OsWAK98_ind* was found duplicated in tandem on the same direction separated by 1,859 nucleotides on chromosome 10, we named as this genes as *OsWAK98a_ind* and *OsWAK98b_ind* (see Table 3). In *japonica* genome, *OsWAK102*, which is classified as *RLCK*, is located in chromosome 10. The *OsWAK102_ind* was also found duplicated, with one copy in the original position in chromosome 10 named as OsWAK102a_ind. The another one *OsWAK102_ind* copy was found in the chromosome 3, named as *OsWAK102b_ind*. We found a new *indica OsWAK*, *OsWAK103/5_ind* (see Table 3), howeve in our searches and

in phylogeny analyses this gene has *japonica's* OsWAK103 and OsWAK105 as the best neighbors, and because of this, it was not possible to define precisely its orthologue in the Japonica genome (see Figure S1). For all the others sequences we were able to determine precisely their *japonica OsWAK* orthologous (Table 3).

Seventy-four sequences characterized as *indica OsWAK* had the same classification, chromosome localization and orientation as their respective *japonica OsWAK* orthologous (Supplementary Table 2). Twenty-two *japonica OsWAK* genes were not found in *indica* genome (Supplementary Table 3), among them, three genes do not have any EGF or kinase domain in its *japonica* protein sequence (*OsWAK35b, OsWAK62* and *OsWAK67*). Four genes were found not already mapped in *indica* genome (*OsWAK30/31_ind, OsWAK39_ind, OsWAK57_ind* in Superscaffold 336, Superscaffold 17, Superscaffold 9488, respectively).

Fourteen *indica OsWAK* presented differences in the classification, chromosome localization and/or sense compared with their *japonica OsWAK* orthologs (see Table 4). Five genes does not show EC domain in *indica* protein sequence, suggesting the classification as *RLCK* (*OsWAK8, OsWAK57, OsWAK66, OsWAK95* and *OsWAK101*). Three genes not show on their protein sequences, kinase domain in *indica* suggesting their classification as RLP (*OsWAK34, OsWAK44 and OsWAK120*). We found three genes located in different chromosomes in indica than yours japonica's orthologous (*OsWAK93, OsWAK96, OsWAK100* and *OsWAK123*). The genes *OsWAK35a* and *OsWAK35d* were found with inverted sense in *indica* genome compared with *japonica* (Table 4).

Phylogenetic analysis

Phylogenetic analyses have been performed for ninety-four OsWAK protein sequences from *japonica* genome, eighty-five *indica* OsWAK and twenty-three AtWAK (from *A. thaliana*). In addition, one protein from RLK family with a corresponding orthologous in both *O. sativa* (Os01g02560) and *A. thaliana* (AT1G67000) species, but not belonging to WAK subfamily, was used as an out group. The phylogenetic analysis, using the amino acid sequences of kinase domain, was performed using four different methods: Bayesian phylogenetics analysis, Maximum Likelihood, Neighbor-Joining and Maximum Parsimony. The complete phylogenies are shown in the Supplementary Material (Figure S1, S2, S3 and S4).

Independently of the method used, twelve clusters were always formed harboring the same genes. These clusters were named using an alpha-numeric code to assist the presentation and discussion of phylogenies as well as to further comparison with others approaches performed in this work. However, some incongruities were verified on the topology that corresponds to the relationship among the clusters. The Bayesian phylogenetic analysis was one that showed the best statistical support, provided by posteriori probability (Figure 1 and supplementary Material 1- Figure S1). The Maximum Likelihood, Neighbor-Joining and Maximum Parsimony methods showed lower statistical support, with the majority of the bootstrap value below 0.90 (Supplementary Material 1 - Figure S2, S3 and S4, respectively). Because of the statistical support, we assume the Bayesian phylogeny as the best representation of the phylogenetic relationship for our data.

The cluster A is formed by thirty-three genes, fifteen from *indica* and eighteen from *japonica*. The cluster B has twenty-nine genes (fourteen from *indica* and fifteen from *japonica*). The cluster C is formed by OsWAK50, OsWAK50_ind, OsWAK53b and OsWAK53b_ind. The proteins OsWAK52 and OsWAK52_ind changed their positions between the group C and D1_At, depending on the reconstruction method, making impossible their classification in one definitive cluster. The groups D1_At (fourteen proteins) and D2_At (six proteins) have only WAK from *A. thaliana*. Cluster D2_At is formed by AtWAK1, AtWAK2, AtWAK3, AtWAK4 and AtWAK5, corresponding to the first WAK cluster described in the literature, plus AtWAKL16. The cluster E is formed by twelve proteins and was the only one cluster that showed proteins from rice and Arabidopsis (OsWAK1, OsWAK2, OsWAK10d and OsWAK25 from *indica* and *japonica* genomes, and AtWAK14, AtWAK15, AtWAK20 and AtWAK21). The cluster F is formed by fifteen proteins (seven from *indica* and eight from *japonica*), cluster G consist of twelve proteins (six from *indica* and six from *japonica*), the cluster H has twenty-eight proteins (thirteen from *indica* and fifteen from *japonica*), the cluster I by thirteen proteins (six from from *indica* and seven from *japonica*) and the cluster J by thirty-one proteins (fifteen from *indica* and sixteen from *japonica*). The genes from group x did not form a concise cluster, since these genes have different positions in the four phylogenetic methods used. For this reason this group is named as X (*OsWAK11, OsWAK11_ind, OsWAK61* and *OsWAK61_ind*).

The topology of Bayesian phylogeny suggests that there are two main groups (Figure 1).The first group is formed by the clusters A, B, C, D1_At, D2_At and OsWAK52, being this grouping seems to be congruent among the four

phylogenies. The second main group presents only OsWAK proteins and is formed by the clusters F, G, H, I, J and X. The topology of the tree strongly suggests cluster E as the ancestor group of *A. thaliana and O. sativa* WAKs, as this group share proteins from both species

Clustering results

Kinase domain clustering

After kinase similarity clustering analysis, a total of 1211 local alignments were found for protein sequences from 108 genes (Figure 2A). When only the best local alignment for each sequence is plotted (Figure 2C), it becomes clear that the results for kinase similarity clustering analysis are closely related with the phylogenetic clusters. The best hit for OsWAK11 (cluster X) was OsWAK13 (cluster H), being this gene the only one that did not show the best hit inside the same phylogeny subgroup. However, when all local alignments were plotted (Figure 2A), a lot of similarities were found among the sequences from clusters A, B and C. Curiously, cluster D2_At showed more local alignments with clusters A and C than with cluster D1_At. The proteins from clusters G, H and OsWAK11 shared similarity in kinase domain. The genes OsWAK36 (cluster A), OsWAK86 (cluster B), OsWAK52, OsWAK47 (cluster C), OsWAK66 (cluster D), OsWAK79 (cluster H), OsWAK106, OsWAK123, OsWAK121, OsWAK116 (the last four sequences from cluster J) and the out group genes did not shown local alignment with any protein using a cutoff value of 1E-90.

Extracellular domain clustering

When EC domain was used for similarity clustering analysis, 74 proteins showed 242 local alignments for 1E-90 E-value cutoff (Figure 2B). The similarities among sequences in EC domain, plotting for the best hit (Figure 2D) were related to the phylogenetic group as verified for the kinase domain. However, OsWAK20 and OsWAK89b, from cluster A, presented their best similarity with OsWAK22, OsWAK23 and OsWAK24, from cluster B on EC domain, despite not presenting similarity at kinase domain for 1E-90 cutoff values. None protein from cluster E had shown similarity to other WAKs with an E-value of 1E-90 for the EC domain.

kinase and EC domain similarity clustering

To integrate the results of kinase and EC domain, two sequence similarity networks were created using the presence of local alignments with cutoff of 1E-90 among gene regions encoding kinase and EC domains. Two approaches were performed. The first one was an intersection between the kinase and EC domains. The result was a network that showed the genes that have similarity in both; kinase and EC domain (Figure 3). In this analysis clusters also correspond to genes that are closely related in the phylogenetic studies. We found two clusters for group A (with 9 and 3 genes in each cluster, respectively), three clusters for group B (4, 2 and 2 proteins), one clusters for C group (2 proteins), two clusters for group D1_At (7 and 6 proteins), one cluster for groups D2_At (5 proteins), F (5 proteins) and G (3 proteins), three clusters for group H (2 proteins for each cluster), one cluster for group I (7 proteins) and three clusters for group J (4, 2 and 2 proteins). The clusters A1, B1, C1, F1, I1, J1, D1 and D2 have almost all proteins

highly connected intra cluster. These clusters represent a set of genes encoding proteins that could be implicated in the perception of similar stimulus and in the activation of related cascades.

The second approach consisted on the searching for genes presenting only EC domain similarity, to determine what proteins could have the potential to recognize similar extracellular stimuli, and possibly to activate different cascades of signal transduction (see Figure 4). Five proteins sets were found: 1) OsWAK27, OsWAK126, OsWAK128b, OsWAK129b and OsWAK129c (from cluster I); 2) OsWAK73-OsWAK86 (from cluster B); 3) OsWAK80-OsWAK83 (from cluster H); 4) OsWAK99-OsWAK108 (from cluster J) and 5) OsWAK121-OsWAK123 (from cluster J) (Figure 4A).

Two proteins set from different clusters showed EC domain similarity. Both proteins OsWAK20 and OsWAK89b (from cluster B) showed similarity with OsWAK23 and OsWAK24 (from cluster G). The protein OsWAK13 (from cluster H) showed similarity with OsWAK45 (from cluster F) (see Figure 4B).

Kinase domain motifs description

After the analysis of arabidopsis and rice kinase domain motifs, the same twelve subdomains previously described for ePK superfamily were identified [27]. Besides that we also found eight additional important motifs described for eukaryotic kinase domain (see Figure 5). Among these eight motifs, we found the region related to ATP binding, in the amino-terminal lobe, and five others motifs in the carboxy-terminal lobe, responsible for substrate binding and phosphorylation. The three conserved amino-terminal motifs that are highly conserved in WAK

kinases are: **G**x**G**xx**G**x**V** (subdomain I), where x represents any amino acid; the conserved residues **K** (the lysine in subdomain II) and **E** (the glutamic acid insubdomain III). For the carboxy-terminal lobe, the motifs **K**xx**DFG** (subdomain VII), **D**xxxx**G** (subdomain IX) and the residue **R** (subdomain XI) were found as highly conserved in the WAK kinase domain.

Nevertheless, two motifs presented differences among ePK and the WAK kinase domains. The conserved motif **APE** (subdomain VIII) was found in the WAK proteins as **DPE**, where the hydrophobic residue of alanine was replaced by the Aspartic acid polar residue. A conserved threonine residue, corresponding to the fifth amino acid before the **APE** motif, in the activation segment, is present in almost all WAK. There are some exceptions for this conserved residue. The *indica* and *japonica* OsWAK36, from the cluster A, have a threonine to proline replacement. All proteins encoded by genes from cluster J do not have the conserved threonine, which was substituted by a serine, aspartic acid, asparagine or methionine.

For the invariant catalytic motifs to R**D**xxxx**N** (from the subdomain VIB, where R represent the conserved Arginine used for kinase classification) we found two main groups with different motifs. In the WAK clusters A, B, C, D, D1_At, D2_At and E, with one exception, we found proteins with a conserved Arginine, showing the conserved R**D**xxxx**N** motif. The *japonica* OsWAK9 , from cluster B, was the only one in the first large group that do not belongs to the RD-kinase class, and shows a H**D**xxxx**N** motif. However, its orthologous in *indica*, OsWAK9_ind, kept the arginine residue belonging to the RD-class (R**D**xxxx**N**).

In genes from clusters F, G, H, I, J and X a new consensus G**D**xxxx**N** (see

Figure 5) was found, and members of these clusters were classified as non-RD kinases. Two genes from cluster J showed differences in the catalytic conserved motif **D**xxxx**N.** The protein OsWAK116, from both *indica* and *japonica* subspecies, present the G**P**xxxx**D** motif. The *japonica* OsWAK108 presented the motif G**N**xxxx**N**, however the OsWAK108_ind contain G**D**xxxx**N** motif.

Cold stress affect OsWAK expression pattern

In attempt to initiate the characterization and comparison of the expression pattern of *OsWAK* genes between Indica em Japonica we determined the profile expression of twenty-two genes in response to cold stress. Eleven genes presented the same expression profile after cold treatment in both rice subspecies (Figure 6). Six genes were up regulated (*OsWAK6, OsWAK22, OsWAK50, OsWAK55, OsWAK73* and *OsWAK75*), while four genes were down regulated (*OsWAK2, OsWAK25, OsWAK32* and *OsWAK56*) and *OsWAK87* does not show a significant variation by the cold treatment. Eleven genes presented a contrasting expression pattern: *OsWAK98b* was down regulated in *japonica*, but up regulated in *indica;* Seven genes were found down regulated in *japonica*, but remained unchanged in *indica* (*OsWAK11, OsWAK16, OsWAK38, OsWAK53b* and *OsWAK76*); the genes OsWAK28 and OsWAK92 were down regulated in *japonica* and up regulated in *indica; OsWAK20* was up regulated in *japonica* but does not changed significantly its expression in *indica.* The genes *OsWAK14* and *OsWAK74* did not vary significantly in *japonica* but were found up and down regulated, respectively, in *indica* subspecie.

DISCUSSION

As expected from their recent evolutionary divergence, the genomic distribution and characteristics of *WAK* genes found in rice subspecies *indica* and *japonica* are well conserved. Notwithstanding, among the 130 gene from *japonica*, we were unable to identify twenty-two *WAK* orthologs in *indica* genome (Table S3). One new gene and two gene duplications from *indica* had no correspondence in japonica genome (Table 3). These results could be attributed to differences on the quality of both genomic annotations and assembly, but similar differences in the number of genes and in their evolutionary pattern have already been described in other gene family comparisons between *indica* and *japonica* genomes [43,44].
These differences found in the genes from the same subfamily among these subspecies is in agreement to the dynamic gene duplication and losses process that been ongoing after the separation of this subspecies, being these phenomena discussed in previously genomic comparative analysis of *japonica* and *indica* [45].

The phylogenetic analysis supports the indication of cluster E, which is the only one having *OsWAK* and *AtWAK*, as the ancestor group among *A. thaliana* and the two subspecies of *Oryza sativa.* The clusters D1_At and D2_At were identified as a specific *A. thaliana* gene cluster. The cluster E, D1_At and D2_At had been already identified in previous phylogenetics analysis encompassing *AtWAK* and japonica *OsWAK* sequences [24]. However, the relationship among these three clusters and the other clusters from *O. sativa* were not very well understood so far. Other previous work comparing all RLK family from these species suggested that there is a group of WAK (called as rice WAK-like or WAKL-OS) that is a specific *O. sativa* subfamily [6], since in the protein phylogeny,

this group was found separated from another AtWAK/OsWAK, and grouped with other rice kinase subfamily. More recently, the same group showed that this rice-specific WAK subfamily did not clearly resolved into a separate subfamily from the AtWAK/OsWAK when the complete repertoire of all RLK proteins found in poplar and moss genome [22] were added in the phylogenetic analysis. These WAKL-OS subfamily is composed by the same proteins found in our cluster J, which showed high divergence among these sequences in our analisys (Figure 1), suggesting, that this group maybe evolving faster than the others. It is important to consider that this divergence could be an artifact that place this group as an outgroup [46] of *WAK* subfamily, like previous results found by Shiu et al [6] using the NJ method.

Our phylogenetic result suggest the formation of two main groups of WAKs (Figure 1, represented for 1 and 2, respectively). The first group shares AtWAK (D1_At and D2_At) and OsWAK (A, B and C) proteins, while the second group is composed only by OsWAK (F, G, H, I, J and X). Detailed analysis of conserved motifs from kinase domain and phylogeny results, revealed that the first main cluster belongs to RD-kinase class (see Figure 5). All genes from the second main group fit in the non-RD class of kinases, but without WAK representative from *A. thaliana* genome. It suggests that the second group expanded from a non-RD WAK after the divergence of the *monocot-dicot*. In order to test the hypothesis of the restriction of WAK-*non*RD in monocots, we expanded the phylogenetics and residues analysis for WAKs from more two *monocots* species (*Sorghum bicolor and Zea mays*) and two *dicots* species (*Glycine max and Vitis vinifera*). This expended phylogeny and residues analysis reinforced the hypothesis of the WAK-

non-RD are restricted to *monocots* (figure Sx). We not found WAK-nonRD in *dicots* genome, only in *monocots* one. The phylogenetics analysis clustered WAK-RD from monocots and dicots in a restricted group, like showed on Figure 1, separated from WAK-nonRD group. The cluster E (WAK-RD) remained as basal group, reinforcing the hypothesis that WAK-RD are ancestor and after the split of monocots and dicots, the *W*AK-*non*RD group emerged on *monocots.* Interestingly, the *japonica* OsWAK9 (from cluster B) changed from RD to non-RD kinase after the separation of *japonica* and *indica* subspecies, since its *indica* orthologous, OsWAK9_ind, belongs to RD-kinase class, and these genes remain together in the phylogeny in the first main cluster. The existence of OsWAK non-RD was already been reported, but none detailed analysis or extensive commentary about this have been done so far [28]. It was also found that about half of WAK sequences, used in our analysis, corresponds to each main group: 50% as RD-kinase and 50% as non-RD kinase. However, in general, sequences belonging to the non-RD group are less frequent and correspond from 9 to16%of the kinases present in an organism. Both rice subspecies, with about 29% [28], have more non-RD class than any other organism already studied. This difference in the expected and identified proportion of WAK non-RD and their maintenance in rice genome rise issues about the importance of particularities on this organims.

Differently from the classic RD-kinase, that have been found involved in many different processes, the non-RD kinase seems to be much more involved with innate immunity [28]. In addition, the kinase activity of these two classes appears to be regulated by different processes. The RD-kinase usually uses the phosphorylation/autophosphorylation of the conserved Ser/Thr residues from the

activation segment, a small region between the VI and VII subdomain (see Figure 5). The phosphorylation of these residues is used as a nucleation center to neutralize a cluster of positive residues and stabilize the protein structure in order to enable the phosphorylation activity on the peptide-substrate [29,25,28]. On the other hand it is not very clear how the non-RD kinases are regulated. It has been shown that the phosphorylation of the activation segment is not essential for their activity. Since the arginine, from the RD motif, is one of the more important residue from this positive cluster, the absence of a positive residue in this position may resulting in an apparently constitutive activity or yet a regulation through alternative mechanisms [29,25,28]. The cluster J is the only one that does not show the conserved threonine that corresponds to the target residue used to phosphorylation in the activation segment. The lack of this residue could indicate that the phosphorylation of the activation segment would not be necessary for the non-RD full activity. Besides, some works have shown that for certain kinase proteins the phosphorylation of the activation segment is not necessary for the substrate phosphorylation activity, but may lead to maximal catalytic activity. Taken together, our dates indicate that the non-RD can use this conserved threonine to improve the kinase activity. The absence of the threonine residue in the cluster J may have changed the activity of these proteins and the selective pressure that they were submitted, more than in other non-RD kinases. Then, as suggested by phylogenetic analysis, members from this cluster are diverging more rapidly in *O. sativa* genome. The congruency among results of comparative analysis from phylogeny and conserved motifs lead us to propose a new subclassification of WAK subfamily. In this new classification WAKs would be categorized as WAK-RD

and WAK-nonRD, according to the presence or absence of conserved arginine in the Catalytic motif RDxxxN.

The similarity clustering analysis of the EC domain, four genes share homology in EC domain, but belongs to different cluster and different kinase (OsWAK24 and OsWAK89b from cluster B - RD class; OsWAK20 and OsWAK23 from cluster G - non-RD class). It is interesting because the EC domain seems to have evolved with a diversity of selective pressure to recognize multiple signals, while the kinase domains were submitted to a purifying selection, due to restriction of the conserved structural conformation to the functional activity [9].

In agreement of the ongoing diversification of *indica* and *japonica* genome, some works demonstrated that *indica* and *japonica* subspecies respond different to the same abiotic stress [47-49]. We found twelve genes, from twenty-two genes tested that showed different expression in *indica* and *japonica* in response to low temperature (Figure 6). Taking together the results from intersection of similarity network between kinase and EC domain (Figure 3) and the japonica expression profile (Figure 6), we found three sets of genes that have potential redundant function in signal reception and substrate phosphorylation (based in similarity), but show distinct profile expression to cold stress: i) *OsWAK14, OsWAK16*, *OsWAK28* and *OsWAK32*; ii) *OsWAK50* and *OsWAK53b*; iii) *OsWAK20* and *OsWAK98b* and iv) *OsWAK74* and *OsWAK76*. As the WAK subfamily is very large in rice genome, this diversification in gene expression may be a product of the subfunctionalization of these genes in a independent way among these two subspecies. Some works in different organism has shown that the gene duplication can led to this subfunctionalization process [50-52].

In summary, the phylogenetic analysis revealed the existence of two main cluster of the *WAK* gene subfamily. The first group, represented by *WAK* genes from *O. sativa* and *A. thaliana* species, belongs to the RD-kinase class. The second group is restricted to *O. sativa*, and their kinase domain corresponds to non-RD kinase class. So far, we propose a new subclassification of *WAK* gene family as: *WAK-RD* and *WAK-nonRD*. The expression pattern of *WAKs* in response to cold suggests that there are differences in the regulation of the *OsWAK* genes in *indica* and *japonica* genome. The present results reinforce that should exists a functional diversification between *A. thaliana* and O*. sativa.* This diversification process apparently happened too intensively and independently in *indica* and *japonica* genomes, through the structural variations, differential gene expression and kinase domain modifications. Additional works looking for functional issues will be helpful to understand the implications of this diversification in rice genome.

MATERIAL AND METHODS

OsWAK identification, gene prediction and domain characterization

OsWAK searches were made using all genes previously described for this subfamily in *O. sativa* ssp *japonica* in the *Rice Genome Annotation Project [*http://rice.plantbiology.msu.edu/*]*. To perform the searches we used tBLASTx [53] in Gramene database (http://www.gramene.org) against *O. sativa* ssp *indica* genome. The first round of searches was performed using *OsWAK japonica* genomic sequence as query. This approach was used because some *japonica OsWAK* gene could have been misannotated and the lacking of some predicted

protein domain could undermine the searching. All the hits confirmed as WAK genes from the *indica* genome were used to a new round of tBLASTx in *indica* genome.

The sequences recovered from *indica* genome were submitted to a gene splicing prediction through GENSCAN ([54]) using the *A. thaliana* matrix. All proteins deduced from all predicted CDS were analyzed with SMART (Simple Modular Architecture Research Tool) [55] and Pfam [56] to define their protein domains. All sequences that presented in the predicted protein sequence a domain related to a kinase and/or EGF were selected as putative *OsWAK* gene. Thereafter, we performed BLASTp for all putative *indica* OsWAK against the Rice *japonica* pseudomolecules protein database (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1/all.dir/all.pep). Finally, the best hit for the putative *indica OsWAK* was annotated as its orthologous and named as *indica OsWAK* (*OsWAK_ind*).

OsWAK reannotation procedures.

We used two procedures to reannotate the *japonica OsWAK* genes. For the first reannotation procedure, we performed the protein domain analysis through SMART for all annotated splicing for all *japonica OsWAK*. The second reannotation procedure, we did for the genes that were found in *indica* genome with different protein feature than that observed for *japonica OsWAK* previously annotated or from our reannotation. This procedure was used to determined the potential of *japonica* OsWAK being transcribe as alternative splicing forms. For this procedure we perform the splicing prediction using GENSCAN for the *OsWAK*

genomic sequence followed by the protein domain analysis, through SMART for all predicted CDS.

Multiple sequence alignment

The multiple sequence alignment, using the predicted protein sequences, was performed using MUSCLE software ([57]), revised manually and individually adjusted the *indels*, from the multiple alignments matrix, through the inclusion of "*?*" character in all except one gap position. The consensus of kinase domain was performed for all OsWAK and AtWAK used on the phylogenetics analysis.

Domain characterization

The graphics of consensus domains were performed using the amino acids alignment from phylogeny analysis through the WebLOGO 3.0 tool ([58]) using probability as the unit of the figure scale, and their lateral chain physic-chemical nature as a color scheme. It was performed a consensus for each clusters found in the phylogenetic and sequence similarity clustering analysis. A consensus for all OsWAK and all AtWAK proteins was also made in separated, in order to verify residue features of each cluster and on each species, respectively.

Clustering analysis

Similarity clustering analysis was done using Circoletto [59], a pipeline in perl to create a similarity visualization through Circos [60] using results from BLASTp. The E-value was set as the parameter to create the edge information. To perform the BLASTp, both EC domain database and kinase database were prepared independently, from the same original sequences. These two databases were created with ninety-four *japonica* OsWAK proteins, twenty-four AtWAK and the two outgroups that were used in phylogenetic analysis. As indica and japonica

OsWAK orthologous are very similar, as suggested by phylogeny, and as similarity relationship among genes from different clusters is the main goal, sequences from indica OsWAK were removed from this analysis. The cutoff of BLASTp was set as E-value 1E-90 for both databases. All local alignments were computed for each gene to create the similar sequence network, and the best local alignments for each gene, to compare to phylogenetic results.

The similarity sequence network was created using the presence of local alignments from BLASTp as information of interaction among WAKs, for each database independently, with the cutoff E-value 1E-90. Both networks were imported to Cytoscape [61] and the plug-in GraphMerge was used to perform the comparison between the two networks and to recover proteins that have homology in both EC and kinase domain or sequences with similarity only in the EC domain.

Phylogenetic analysis

A set of four independent phylogenetic approaches were performed using the WAK amino acid sequences, in order to determine the statistically more significant and robust approach to infer the relationship among WAK studied in this work. All phylogenies were done using the same protein alignment. The Bayesian phylogenetic analysis was performed using MrBayes [62] setting the model as “mixed”. The Maximum Likelihood phylogeny was performed using PhyML [63]. The MEGA4 software [64] was employed for the Neighbor-Joining and Maximum Parsimony, using the bootstrap with 1000 replicates. For Neighbor-Joining phylogeny the poison-correction substitution model was used.

Plant material, cold stress and qPCR Analysis

Cold stress experiments were made with *O. sativa* spp *japonica* cv

nipponbare and *O. sativa* spp *indica* cv. Embrapa Taim plants. The seedlings were grown in hydroponic culture [65] at 28ºC. The cold treatment consisted on keeping plants at 4ºC for 24 hours, while control plants remained at 28ºC. The experiment was conducted with quadruplicate for each sample condition, and each sample consisted of four seedlings. The total RNA was extracted by TRIZOL® reagent (Invitrogen, CA, USA). The reverse transcription was performed using M-MLV Reverse Transcriptase (PROMEGA) and a T24V-poli-T primer. The Real Time quantitative PCR (qPCR) was performed in a StepOnePlus™ Real-Time PCR System (Applied Biosystems) using SYBR Green I (Invitrogen) to detect double-stranded cDNA synthesis. The reactions were completed in a volume of 24 µL containing 12 µL of cDNA, 1X SYBR Green I (Invitrogen), 0.025 mM dNTP, 1X PCR Buffer, 3 mM $MgCl_2$, 0.25 U Platinum Taq DNA Polymerase (Invitrogen) and 200 nM of each reverse and forward primer. The amplification consisted on the following steps: 94ºC for 5 minutes, 40 cycles of 10 second at 94ºC, 15 seconds at 60ºC and 15 seconds at 72ºC. For all reaction were performed the dissociation curve, obtained on the increase of 0,1ºC for reading from 60ºC to 99ºC, to confirm the specificity of amplification. For statistical support, each sample was also quantified as the average of four technical replicates, and used as sample value. The mathematical methods to determine the differential gene expression was $2^{-\Delta\Delta Ct}$ [66] and the Student T-test was performed to assess the statistical significance, where was considered the confidence interval of $p<0.05$. OseFa1(LOC_Os03g08010) and OsFDH(LOC_Os02g57040) were used as housekeeping genes for OsWAK gene normalization. Twenty-two *OsWAK* genes (*OsWAK2, OsWAK6, OsWAK11, OsWAK14, OsWAK16, OsWAK20, OsWAK22,*

OsWAK25, OsWAK28, OsWAK32, OsWAK38, OsWAK50, OsWAK53b, OsWAK55, OsWAK56, OsWAK73, OsWAK74, OsWAK75, OsWAK76, OsWAK87, OsWAK92, OsWAK98b) were evaluated by RT-qPCR. The set of primers used for these genes are shown on Table S4.

REFERENCE

1. Shulaev V, Cortes D, Miller G, Mittler R (2008) Metabolomics for plant stress response. Physiologia plantarum 132: 199-208.

2. Walker JC, Zhang R (1990) Relationship of a putative receptor protein kinase from maize to the S-locus glycoproteins of Brassica. Nature 345: 743-6.

3. Shiu SH, Bleecker AB (2001) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proc Nat Acad Sci U S A 98: 10763-8.

4. Shiu SH, Bleecker AB (2001) Plant receptor-like kinase gene family: diversity, function, and signaling. Sci STKE 2001: re22.

5. Shiu SH, Bleecker AB (2003) Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. Plant Physiology 132: 530-43.

6. Shiu S-H, Karlowski WM, Pan R, Tzeng Y-H, Mayer KFX, et al. (2004) Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. The Plant cell 16: 1220-34.

7. Hanks S (2003) Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. Genome Biology 4: 111.

8. Osakabe Y, Maruyama K, Seki M, Satou M, Shinozaki K, et al. (2005) Leucine-rich repeat receptor-like kinase1 is a key membrane-bound regulator of abscisic acid early signaling in Arabidopsis. The Plant cell 17: 1105-19.

9. Afzal AJ, Wood AJ, Lightfoot DA (2008) Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Molecular plant-microbe interactions 21: 507-17.

10. Chae L, Sudat S, Dudoit S, Zhu T, Luan S (2009) Diverse Transcriptional Programs Associated with Environmental Stress and Hormones in the Arabidopsis Receptor-Like Kinase Gene Family. Molecular Plant 2: 84-107.

11. Mizuno S, Osakabe Y, Maruyama K, Ito T, Osakabe K, et al. (2007) Receptor-like protein kinase 2 (RPK 2) is a novel factor controlling anther development in Arabidopsis thaliana. The Plant Journal 50: 751-66.

12. De Smet I, Vassileva V, De Rybel B, Levesque MP, Grunewald W, et al. (2008) Receptor-like kinase ACR4 restricts formative cell divisions in the Arabidopsis root. Science 322: 594-7.

13. Germain H, Gray-Mitsumune M, Lafleur E, Matton DP (2008) ScORK17, a transmembrane receptor-like kinase predominantly expressed in ovules is involved in seed development. Planta 228: 851-62. doi:10.1007/s00425-008-0787-0

14. Clark SE, Williams RW, Meyerowitz EM (1997) The CLAVATA1Gene Encodes a Putative Receptor Kinase That Controls Shoot and Floral Meristem Size inv rabidopsis. Cell 89: 575-585.

15. Shpak ED, Berthiaume CT, Hill EJ, Torii KU (2004) Synergistic interaction of

three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation. Development 131: 1491-501.

16. Colcombet J, Boisson-Dernier A, Ros-Palau R, Vera CE, Schroeder JI (2005) Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASES1 and 2 are essential for tapetum development and microspore maturation. Plant Cell 17: 3350-61.

17. Fulton L, Vaddepalli P, Yadav RKK, Batoux M, Schneitz K (2010) Inter-cell-layer signalling during Arabidopsis ovule development mediated by the receptor-like kinase STRUBBELIG. Biochemical Society transactions 38: 583-587.

18. Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, et al. (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature 417: 959-62.

19. Indrasumunar A, Searle I, Lin M-H, Kereszt A, Men A, et al. (2010) Nodulation Factor Receptor Kinase 1α Controls Nodule Organ Number in Soybean (Glycine max L. Merr.). The Plant Journal 65:39-50

20. Asano T, Hakata M, Nakamura H, Aoki N, Komatsu S, et al. (2010) Functional characterisation of OsCPK21, a calcium-dependent protein kinase that confers salt tolerance in rice. Plant molecular biology: 1-13-13.

21. Coca M, Segundo BS (2010) AtCPK1 calcium-dependent protein kinase mediates pathogen resistance in Arabidopsis. The Plant Journal 63: 526-540

22. Lehti-Shiu MD, Zou C, Hanada K, Shiu S-H (2009) Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. Plant physiology 150: 12-26.

23. Vij S, Giri J, Dansana PK, Kapoor S, Tyagi AK (2008) The receptor-like cytoplasmic kinase (OsRLCK) gene family in rice: organization, phylogenetic relationship, and expression during development and stress. Molecular plant 1: 732-50.

24. Zhang S, Chen C, Li L, Meng L, Singh J, et al. (2005) Evolutionary expansion, gene structure, and expression of the rice wall-associated kinase gene family. Plant Physiol 139: 1107-24.

25. Krupa A, Preethi G, Srinivasan N (2004) Structural modes of stabilization of permissive phosphorylation sites in protein kinases: distinct strategies in Ser/Thr and Tyr kinases. Journal of molecular biology 339: 1025-39.

26. Hanks SK, Quinn a M, Hunter T (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241: 42-52.

27. Hanks SK, Hunter T (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. The FASEB journal : official publication of the Federation of American Societies for Experimental Biology 9: 576-96.

28. Dardick C, Ronald P (2006) Plant and animal pathogen recognition receptors

signal through non-RD kinases. PLoS pathogens 2: e2..

29. Johnson LN, Noble MEM, Owen DJ (1996) Active and Inactive Protein Kinases: Structural Basis for Regulation. Cell 85: 149-158.

30. He ZH, Fujiki M, Kohorn BD (1996) A cell wall-associated, receptor-like protein kinase. J Biol Chem 271: 19789-93.

31. He ZH, Cheeseman I, He D, Kohorn BD (1999) A cluster of five cell wall-associated receptor kinase genes, Wak1-5, are expressed in specific organs of Arabidopsis. Plant Mol Biol 39: 1189-96..

32. Verica JA, He ZH (2002) The cell wall-associated kinase (WAK) and WAK-like kinase gene family. Plant Physiology 129: 455-9.

33. Verica JA, Chae L, Tong H, Ingmire P, He ZH (2003) Tissue-specific and developmentally regulated expression of a cluster of tandemly arrayed cell wall-associated kinase-like kinase genes in Arabidopsis. Plant Physiology 133: 1732-46.

34. Kohorn BD (2001) WAKs; cell wall associated kinases. Current Opinion in Cell Biology 13: 529-33.

35. Anderson CM, Wagner TA, Perret M, He ZH, He D, et al. (2001) WAKs: cell wall-associated kinases linking the cytoplasm to the extracellular matrix. Plant Mol Biol 47: 197-206.

36. Wagner TA, Kohorn BD (2001) Wall-associated kinases are expressed throughout plant development and are required for cell expansion. Plant Cell 13: 303-18.

37. Decreux A, Messiaen J (2005) Wall-associated kinase WAK1 interacts with cell wall pectins in a calcium-induced conformation. Plant Cell Physiology 46: 268-78.

38. Kohorn BD, Kobayashi M, Johansen S, Friedman HP, Fischer A, et al. (2006) Wall-associated kinase 1 (WAK1) is crosslinked in endomembranes, and transport to the cell surface requires correct cell-wall synthesis. Journal of Cell Science 119: 2282-90.

39. Park AR, Cho SK, Yun UJ, Jin MY, Lee SH, et al. (2001) Interaction of the Arabidopsis receptor protein kinase Wak1 with a glycine-rich protein, AtGRP-3. J Biol Chem 276: 26688-93.

40. He ZH, He D, Kohorn BD (1998) Requirement for the induced expression of a cell wall associated receptor kinase for survival during the pathogen response. The Plant Journal 14: 55-63.

41. Sivaguru M, Ezaki B, He ZH, Tong H, Osawa H, et al. (2003) Aluminum-induced gene expression and protein localization of a cell wall-associated receptor kinase in Arabidopsis. Plant Physiology 132: 2256-66.

42. Li H, Zhou SY, Zhao WS, Su SC, Peng YL (2009) A novel wall-associated receptor-like protein kinase gene, OsWAK1, plays important roles in rice blast disease resistance. Plant Molecular Biology 69: 337-46.

43. Jia L, Clegg MT, Jiang T (2004) Evolutionary dynamics of the DNA-binding

domains in putative R2R3-MYB genes identified from rice subspecies indica and japonica genomes. Plant physiology 134: 575-85.

44. Ross CA, Liu Y, Shen QJ (2007) The WRKY Gene Family in Rice (Oryza sativa). Journal of Integrative Plant Biology 49: 827-842

45. Yu J, Wang J, Lin W, Li S, Li H, et al. (2005) The Genomes of Oryza sativa: a history of duplications. PLoS biology 3: e38.

46. Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, et al. (2011) Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. PLoS Biology 9: e1000602.

47. Weng J-H, Chen C-Y (1987) Differences between Indica and Japonica rice varieties in CO2 exchange rates in response to leaf nitrogen and temperature. Photosynthesis Research 14: 171-178.

48. Baruah AR, Ishigo-Oka N, Adachi M, Oguma Y, Tokizono Y, et al. (2008) Cold tolerance at the early growth stage in wild and cultivated rice. Euphytica 165: 459-470.

49. Liu F, Xu W, Wei Q, Zhang Z, Xing Z, et al. (2010) Gene expression profiles deciphering rice phenotypic variation between Nipponbare (Japonica) and 93-11 (Indica) during oxidative stress. PloS one 5: e8632.

50. Nielsen MG, Gadagkar SR, Gutzwiller L (2010) Tubulin evolution in insects: gene duplication and subfunctionalization provide specialized isoforms in a functionally constrained gene family. BMC evolutionary biology 10: 113.

51. Kleinjan D a, Bancewicz RM, Gautier P, Dahm R, Schonthaler HB, et al. (2008) Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence. PLoS genetics 4: e29.

52. Geng S, Zhao Y, Tang L, Zhang R, Sun M, et al. (2011) Molecular evolution of two duplicated CDPK genes CPK7 and CPK12 in grass species: A case study in wheat (Triticum aestivum L.). Gene 475: 94-103.

53. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of molecular biology 215: 403-10.

54. Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. Journal of molecular biology 268: 78-94.

55. Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. Nucleic acids research 37: D229-32.

56. Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. (2010) The Pfam protein families database. Nucleic acids research 38: D211-22.

57. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32: 1792-7.

58. Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. Genome research 14: 1188-90.

59. Darzentas N (2010) Circoletto: visualizing sequence similarity with Circos. Bioinformatics (Oxford, England) 26: 2620-2621.

60. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. (2009) Circos: an information aesthetic for comparative genomics. Genome research 19: 1639-45.

61. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research 13: 2498-504.

62. Huelsenbeck J, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

63. Guindon S, Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Systematic Biology 52: 696-704.

64. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-1599.

65. Baier AC, Somers DJ, Gusiafson JP (1995) Aluminium tolerance in wheat: correlating hydroponic evaluations with field and soil performances*. Plant Breeding 114: 291-296.

66. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif.) 25: 402-8.

Figure 1: Bayesian phylogeny of AtWAK and OsWAK subfamily.

Representation of arabidopsis and rice Bayesian phylogeny of WAK proteins, where each gene cluster are shown collapsed, with the branch lengths representing the divergence inside the cluster. The height of collapsed clusters is illustrative and does not represent the number of genes present in each cluster. The number **1** indicated the first large group that contain AtWAK and OsWAK and encompasses almost all genes from RD-kinase class. The number **2** represent the second large group restricted to *O. sativa*, corresponding to non-RD kinases.

Figure 2: Visualization of WAK domain similarities.

The Circos visualization for BLASTp results for kinase and EC domains with e-values below to 1E-90. Each cluster from phylogenetic result are plotted in the green semicircle and identified with the same name as in the phylogeny. A: All results of the BLASTp for kinase domain. B: All results of the EC domain BLASTp. C: The best hit for each gene for the kinase BLASTp results. D: The best hit for each gene for the EC domain BLASTp results.

Figure 3: Intersection of sequence similarity network from EC and kinase domains.

Similarity sequence networks based in the intersection between BLASTp results for EC and kinase domain. These networks represent the WAK proteins that shared similarity on EC and kinase domain.

Figure 4: Exclusive similarity on EC domain.

Similarity sequence networks based in the differences between BLASTp results for EC and kinase domain. These networks represent proteins that shared similarity only on the EC domain. A: Similarity sequence networks for proteins from same

cluster from phylogeny. B: Similarity sequence networks for genes from different clusters in the phylogeny, where genes with the same color in the network are from the same cluster.

Figure 5: Subdomains and motifs from kinase domain.

Schematic representations of the kinase domains and subdomains with emphasis in the conserved motifs. The subdomain VIB and VII are shown for each cluster from phylogeny, through the WebLogo Consensus figure. Amino acid were represented according to their chemical nature: polar G,S,T,Y,C,Q,N (green); basic: K,R,H (blue); acidic: D,E (red) and hydrophobic: A,V,L,I,P,W,F,M (black). The red rectangles represent the amino acids consensus for each species. The red and blue arrows represent, respectively, the conserved arginine in subdomain VIB and the conserved threonine in subdomain VIII.

Figure 6: *Indica* and *japonica* OsWAK expression profile in cold stress.

The graphic corresponds to the relative expression of WAK transcripts after RT-qPCR on *indica* and *japonica* shoots submitted to 4 ^{o}C cold stress. The graphics scale are in Log2, where down regulated genes show negative values, and up regulated have positive values. The red asterisk represents results with statistical significance for p-Value<0.05 after test-T.

FIGURE 1

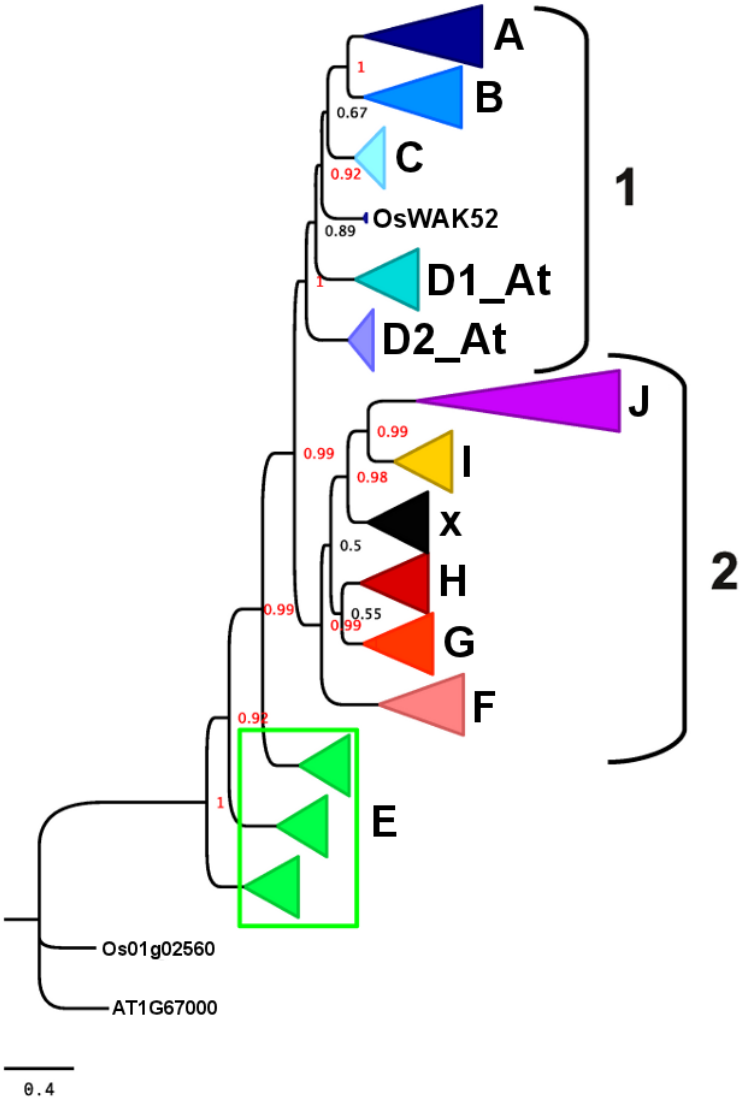
A
B
C
OsWAK52
D1_At
D2_At
J
I
X
H
G
F
E
1
2
1
0.67
0.92
0.89
1
0.99
0.98
0.99
0.5
0.99
0.55
0.99
0.92
1
Os01g02560
AT1G67000
0.4

FIGURE 2

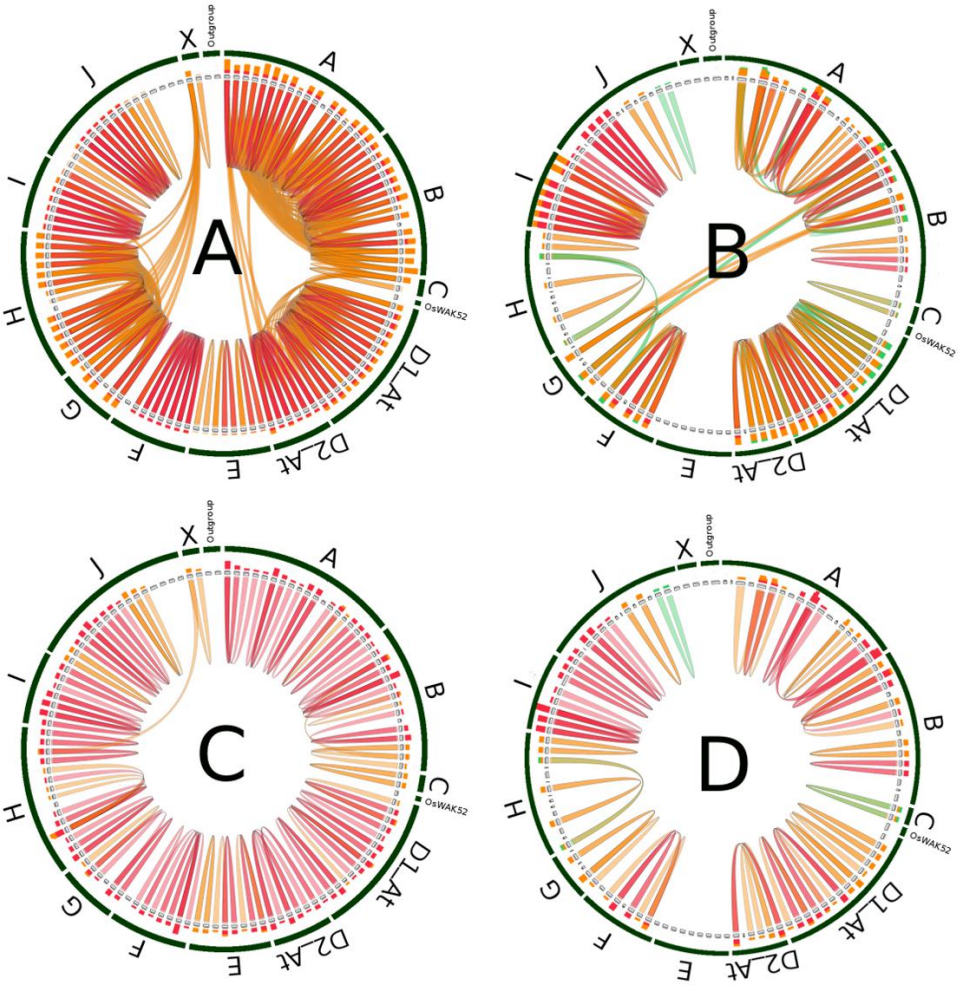

FIGURE 3

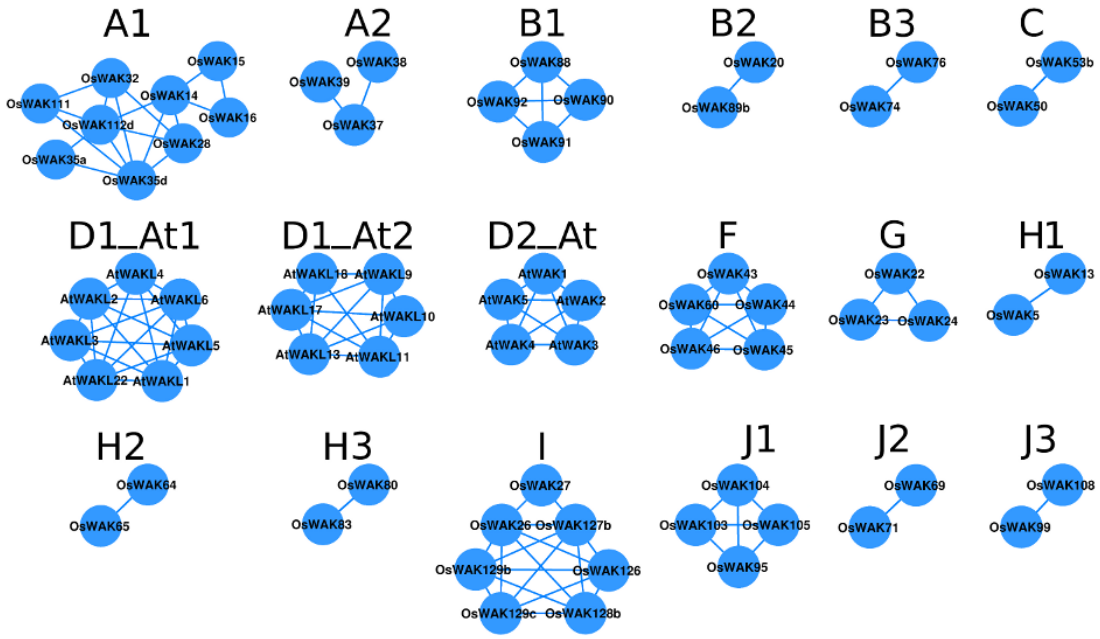
A1
OsWAK15
OsWAK32
OsWAK14
OsWAK111
OsWAK16
OsWAK112d
OsWAK28
OsWAK35a
OsWAK35d
A2
OsWAK38
OsWAK39
OsWAK37
B1
OsWAK88
OsWAK92
OsWAK90
OsWAK91
B2
OsWAK20
OsWAK89b
B3
OsWAK76
OsWAK74
C
OsWAK53b
OsWAK50
D1_At1
AtWAKL4
AtWAKL2
AtWAKL6
AtWAKL3
AtWAKL5
AtWAKL22
AtWAKL1
D1_At2
AtWAKL18
AtWAKL9
AtWAKL17
AtWAKL10
AtWAKL13
AtWAKL11
D2_At
AtWAK1
AtWAK5
AtWAK2
AtWAK4
AtWAK3
F
OsWAK43
OsWAK60
OsWAK44
OsWAK46
OsWAK45
G
OsWAK22
OsWAK23
OsWAK24
H1
OsWAK13
OsWAK5
H2
OsWAK64
OsWAK65
H3
OsWAK80
OsWAK83
I
OsWAK27
OsWAK26
OsWAK127b
OsWAK129b
OsWAK126
OsWAK129c
OsWAK128b
J1
OsWAK104
OsWAK103
OsWAK105
OsWAK95
J2
OsWAK69
OsWAK71
J3
OsWAK108
OsWAK99

FIGURE 4

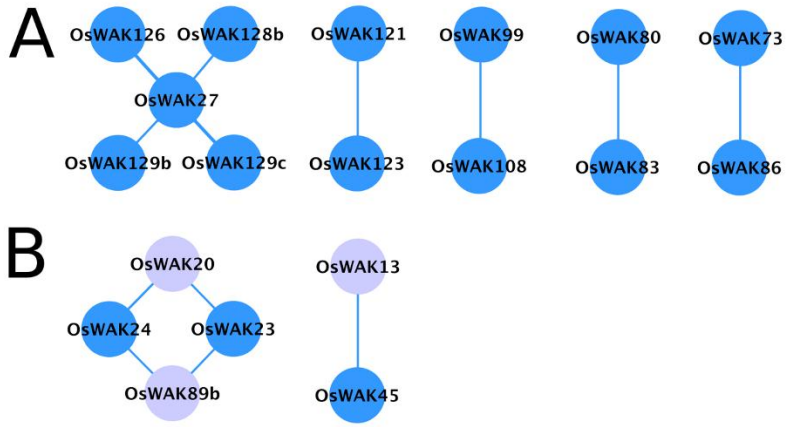
A
OsWAK126
OsWAK128b
OsWAK27
OsWAK129b
OsWAK129c
OsWAK121
OsWAK123
OsWAK99
OsWAK108
OsWAK80
OsWAK83
OsWAK73
OsWAK86
B
OsWAK20
OsWAK24
OsWAK23
OsWAK89b
OsWAK13
OsWAK45

FIGURE 5

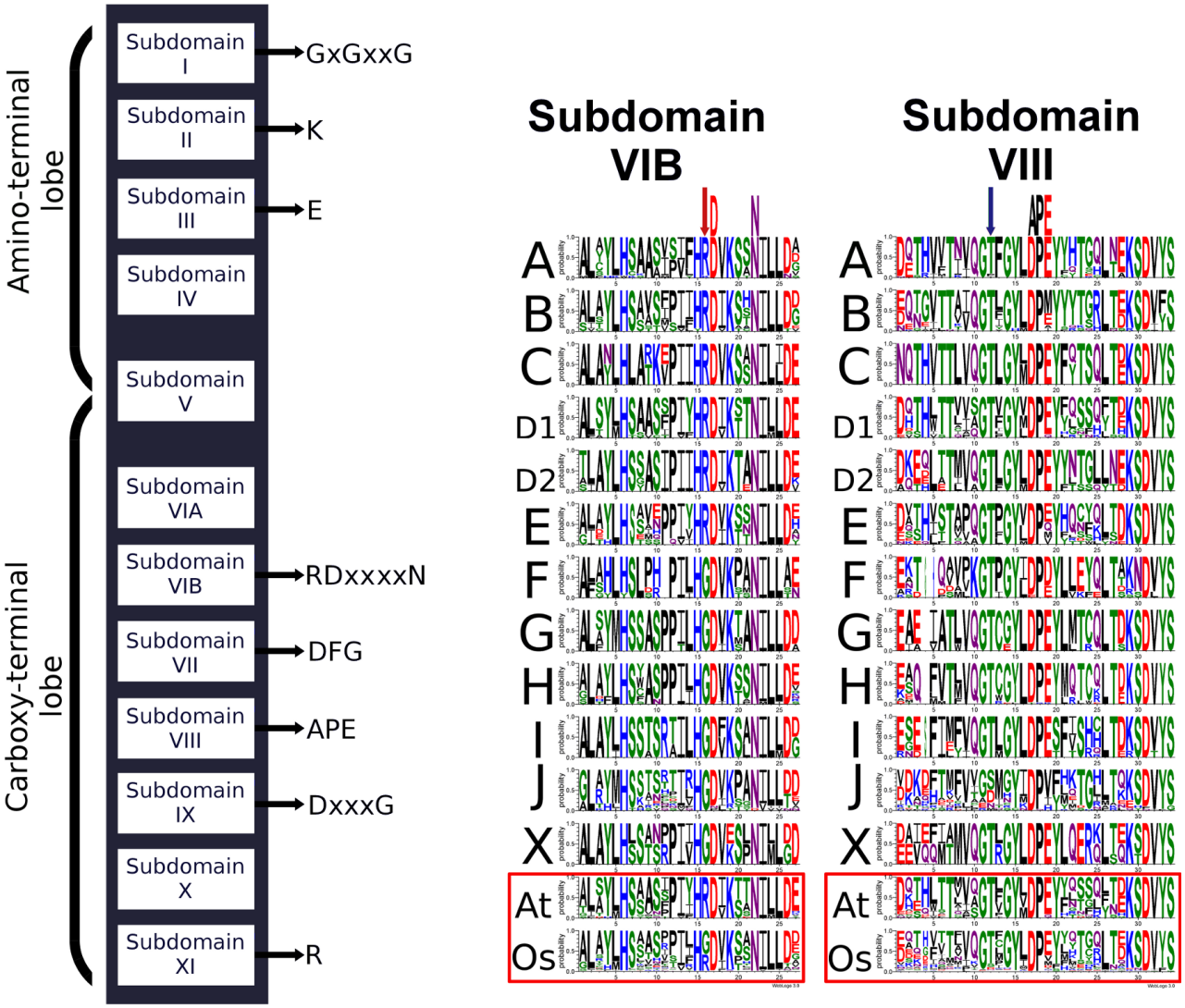
Amino-terminal lobe
Carboxy-terminal lobe
Subdomain I
Subdomain II
Subdomain III
Subdomain IV
Subdomain V
Subdomain VIA
Subdomain VIB
Subdomain VII
Subdomain VIII
Subdomain IX
Subdomain X
Subdomain XI
GxGxxG
K
E
RDxxxxN
DFG
APE
DxxxG
R
Subdomain VIB
Subdomain VIII
A
B
C
D1
D2
E
F
G
H
I
J
X
At
Os

FIGURE 6

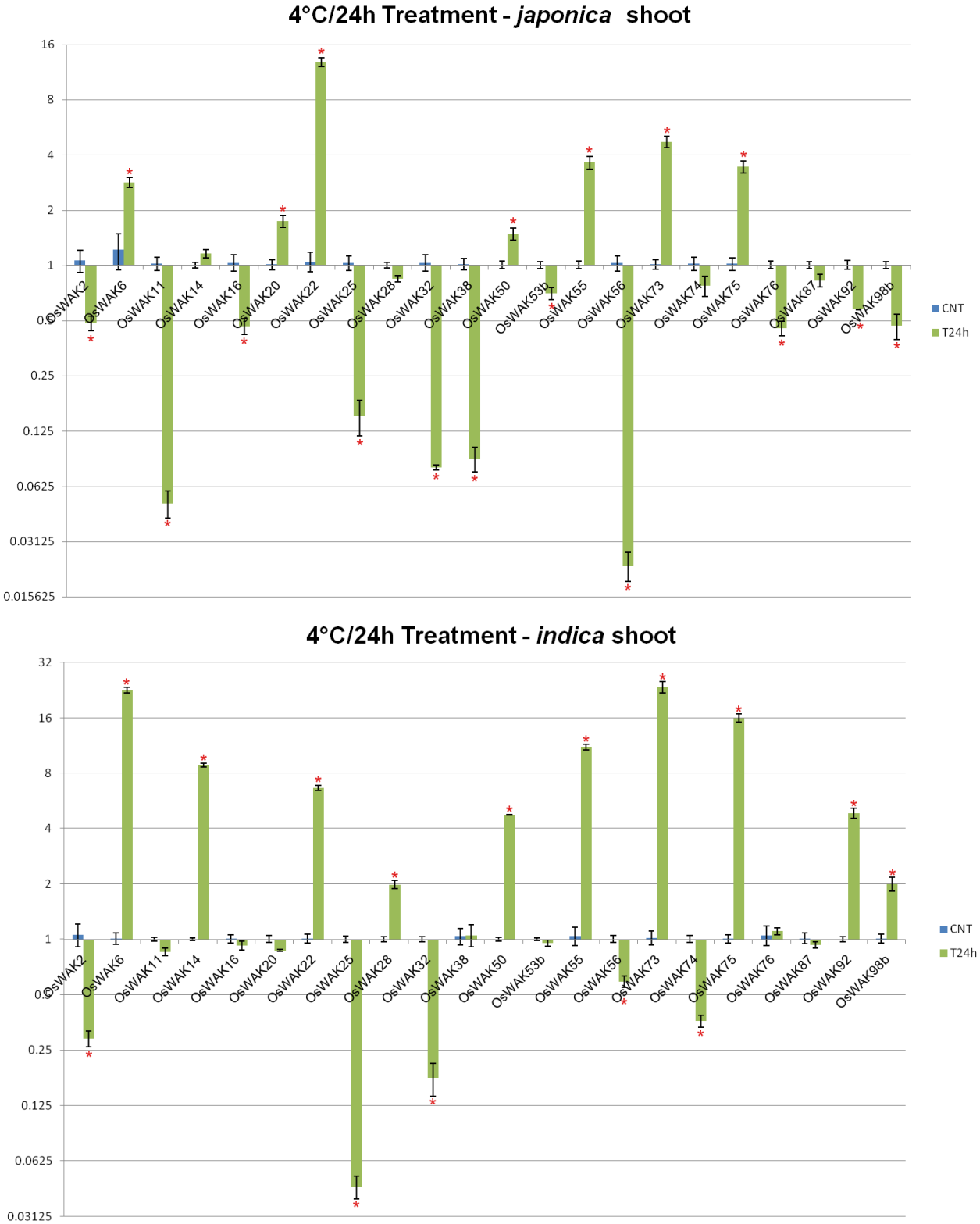
4°C/24h Treatment - japonica shoot
16
8
4
2
1
0.5
0.25
0.125
0.0625
0.03125
0.015625
OsWAK2
OsWAK6
OsWAK11
OsWAK14
OsWAK16
OsWAK20
OsWAK22
OsWAK25
OsWAK28
OsWAK32
OsWAK38
OsWAK50
OsWAK53b
OsWAK55
OsWAK56
OsWAK73
OsWAK74
OsWAK75
OsWAK76
OsWAK87
OsWAK92
OsWAK98b
CNT
T24h
4°C/24h Treatment - indica shoot
32
16
8
4
2
1
0.5
0.25
0.125
0.0625
0.03125
OsWAK2
OsWAK6
OsWAK11
OsWAK14
OsWAK16
OsWAK20
OsWAK22
OsWAK25
OsWAK28
OsWAK32
OsWAK38
OsWAK50
OsWAK53b
OsWAK55
OsWAK56
OsWAK73
OsWAK74
OsWAK75
OsWAK76
OsWAK87
OsWAK92
OsWAK98b
CNT
T24h

TABLE 1: The previous and the new classification, chromosome localization and sense of annotated japonica OsWAK and the respective indica OsWAK gene

	***O. sativa* japonica**			***O. sativa* indica**		
Name	**Chr**	**Sense**	**Class**	**Chr**	**Sense**	**Class**
OsWAK2	1	-	RLCK/**RLK***	1	-	RLK
OsWAK7	1	+	RLK/**RLP***	***		
OsWAK12	2	-	RLCK/**RLK****	2	-	RLK
OsWAK18b	2	+	short**/RLP****	2	+	RLP
OsWAK21	2	-	RLCK/**RLK****	2	-	RLK
OsWAK23	2	+	short/**RLK****	2	+	RLK
OsWAK33	4	+	RLP/**RLK***	4	+	RLP
OsWAK35a	4	+	short/**RLK****	4	-	RLK
OsWAK35d	4	-	short/**RLK****	4	+	RLK
OsWAK37	4	+	short/**RLK****	4	+	RLK
OsWAK66	7	+	pseudo/**RLK****	7	+	RLCK
OsWAK76	8	+	RLCK/**RLK***	8	+	RLK
OsWAK83	9	-	pseudo/**RLK****	9	-	RLK
OsWAK84	9	-	RLP/**RLCK***	***		
OsWAK85	9	-	RLCK/**RLK****	9	-	RLK
OsWAK88	9	+	pseudo/**RLK****	9	+	RLK
OsWAK96	10	+	RLK/**RLP***	4	-	RLP
OsWAK101	10	-	short/**RLK****	10	-	RLCK
OsWAK109	10	+	RLK/**RLP***	10	+	RLP
OsWAK127b	12	-	short/**RLK****	12	-	RLK

*Classification through domain analysis of annotated protein

** Classification through new splicing prediction and protein domain analysis of annotated genomic region

*** Not found in indica

The bold words are the new suggested classification

Table 2: Total number of known OsWAKs found in *O.sativa* japonica, the counterparts from *O.sativa* indica according to their functional classification

	Number of Genes	
Classification	*japonica*	*indica*
RLK	72	64
RLCK	27	28
RLP	12	14
Short-gene	14	3
Pseudogene	5	-
Unclassified	-	2
Total	130	111

Table 3: New genes and duplications in *O.sativa* indica genome.

Name	Chr	Sense	Start	End	Classification
OsWAK26	3	-	24795883	24805379	RLK
OsWAK52	4	-	30010793	30015576	short/**RLK***
OsWAK98	10	-			RLCK
OsWAK102	10	+			RLCK
**					

Name	Chr	Sense	Start	End	Classification
OsWAK26_ind[1]	3	-	28716174	28722414	split
	3	-	28694607	28697665	split
OsWAK52_ind[1]	4	-	29042876	29048124	split
	SC 2267	-	2219	3889	split
OsWAK98_ind	10	-	780921	783189	RLCK
[2]	10	-	776801	779062	RLCK
OsWAK102_ind	10	+	2395753	2397702	RLCK
[2]	3	-	32787125	32789077	RLCK
OsWAK103/5_ind*	6	-	30737383	30742176	RLK

[1] Gene with a split

[2] Gene associated to duplication

* New classification by domain analysis of annotated protein

** Gene not found in japonica genome

SC Super Scaffold

Table 4: Genes with differences between *indica* and *japonica*.

	O. sativa japonica			O. sativa indica		
Name	**Chr**	**Sense**	**Classification**	**Chr**	**Sense**	**Classification**
OsWAK8[1]	1	+	RLK	1	+	RLCK
OsWAK34[1]	4	+	RLK	4	+	RLP
OsWAK44[1]	4	-	RLK	4	-	RLP
OsWAK57[1]	4	+	RLK	SC 9488	-	RLCK
OsWAK101[1]	10	-	short/**RLK****	10	-	RLCK
OsWAK95[1]	10	-	RLK	10	-	RLCK
OsWAK66[1]	7	+	pseudo/**RLK***	7	+	RLCK
OsWAK120[1]	11	-	RLK	11	-	RLP
OsWAK35a[2]	4	+	short/RLK**	4	-	RLK
OsWAK35d[2]	4	-	short/RLK**	4	+	RLK
OsWAK93[2]	10	-	RLCK	7	+	RLCK
OsWAK96[2]	10	+	RLK/**RLP***	4	-	RLP
OsWAK100[2]	10	+	RLCK	3	-	RLCK
OsWAK123[2]	11	+	RLK	10	+	RLK

[1] Different classification between OsWAK japonica and indica

[2] Different chromosome localization and/or sense between OsWAK japonica and indica

SC: Super Scaffold

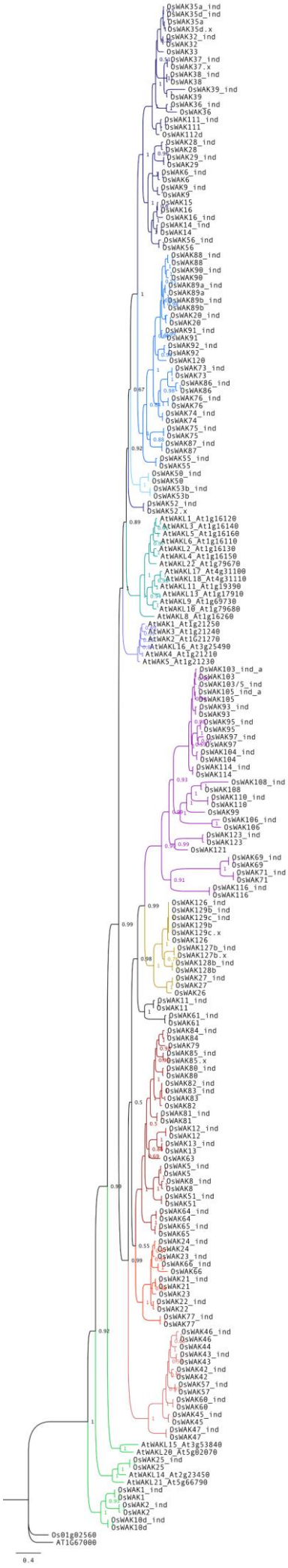

Figure S1: Complete Bayesian phylogeny. Bayesian phylogeny, where the colors in the branches represent the same cluster shown on figure 1.

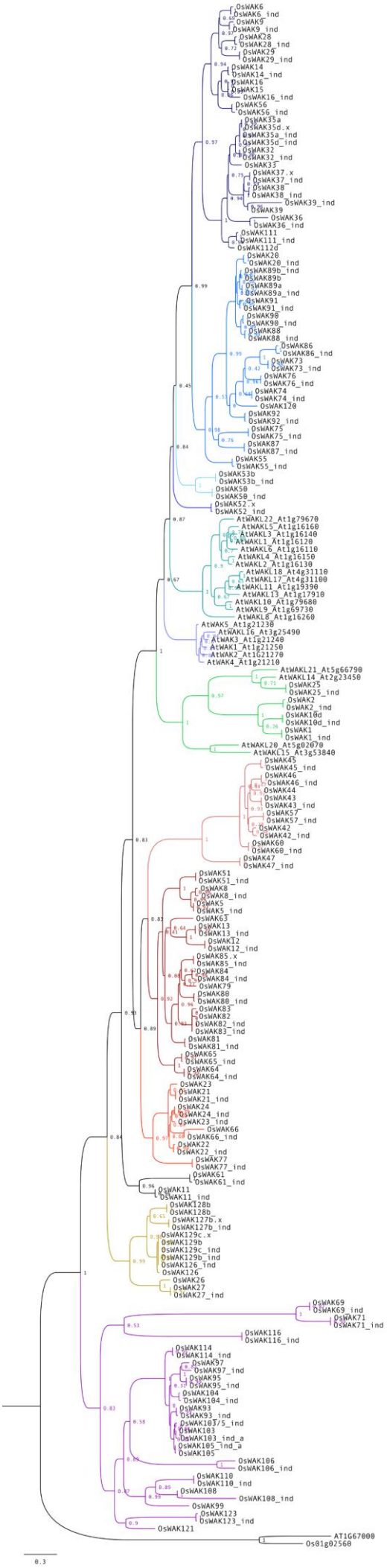

Figure S2- Maximum Likelihood phylogeny. The colors in the branches represent the same cluster shown on figure 1.

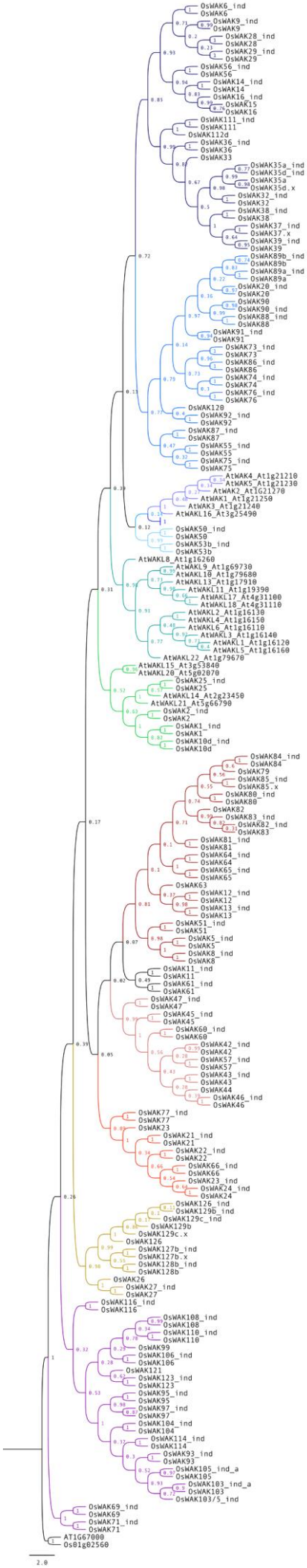

Figure 3 - Maximum Parsimony phylogeny. The colors in the branches represent the same cluster shown on figure 1.

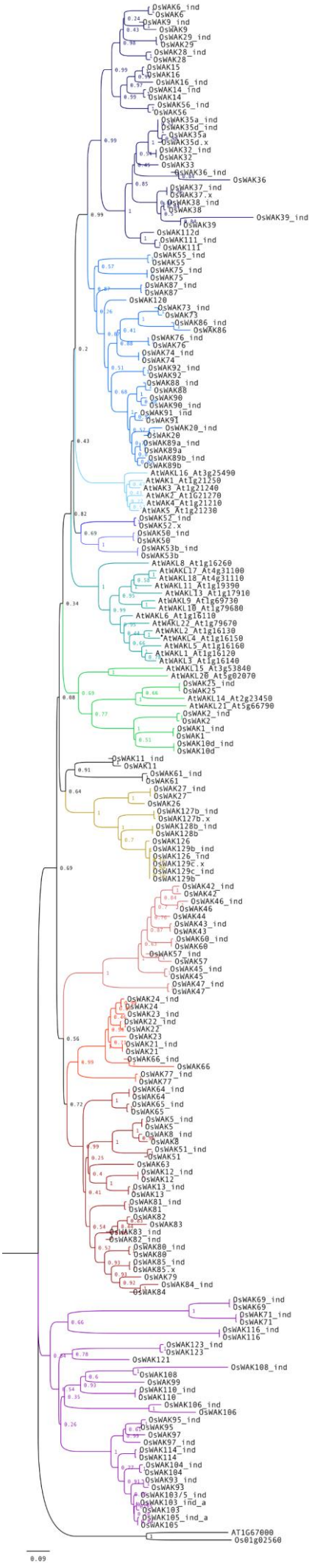

Figure 4: Neighbor-Joining phylogeny. The colors in the branches represent the same cluster shown on figure 1D.

Table S1: Complete table with indica and japonica OsWAK.

	O. sativa japonica						*O. sativa indica*					
Name	**Locus**	**Chr**	**Sense**	**Start**	**End**	**Classification**	**Name**	**Chr**	**Sense**	**start**	**end**	**Classification**
OsWAK1	LOC_Os01g04409	1	-	1956777	1966451	RLCK	OsWAK1_ind	1	-	2312092	2322889	RLCK
OsWAK2	LOC_Os01g04450	1	-	1976423	1981484	RLCK/**RLK***	OsWAK2_ind	1	-	2331441	2336500	RLK
OsWAK3	LOC_Os01g20880	1	+	11654112	11658869	RLCK	OsWAK3_ind	1	+	12700909	12705668	RLCK
OsWAK4	LOC_Os01g20900	1	+	11676457	11684281	RLCK	OsWAK4_ind	1	+	12723592	12730428	RLCK
OsWAK5	LOC_Os01g26174	1	+	14824118	14827241	RLK	OsWAK5_ind	1	+	16416095	16417637	RLCK
OsWAK6	LOC_Os01g26210	1	-	14855072	14857244	RLK	OsWAK6_ind	1	-	16423999	16426223	RLK
OsWAK7	LOC_Os01g26270	1	+	14891850	14892917	RLK/**RLP***	***					
OsWAK8	LOC_Os01g26280	1	+	14895659	14898152	RLK	OsWAK8_ind	1	+	16473149	16475488	RLCK
OsWAK9	LOC_Os01g26300	1	-	14902859	14904395	RLCK	OsWAK9_ind	1	-	16479505	16481032	RLCK
OsWAK10d	LOC_Os01g49529	1	-	28480966	28495222	RLCK	OsWAK10d_ind	1	-	31786631	31800354	RLCK
OsWAK11	LOC_Os02g02120	2	-	627342	630655	RLK	OsWAK11_ind	2	-	664488	667758	RLK
OsWAK12	LOC_Os02g41480	2	-	24856334	24858930	RLCK/**RLK****	OsWAK12_ind	2	-	26615425	26618021	RLK
OsWAK13	LOC_Os02g41500	2	-	24865877	24868342	RLK	OsWAK13_ind	2	-	26627307	26629782	RLK
OsWAK14	LOC_Os02g42150	2	-	25342352	25345873	RLK	OsWAK14_ind	2	-	27048464	27051967	RLK
OsWAK15	LOC_Os02g42160	2	-	25350877	25354142	RLK	***					
OsWAK16	LOC_Os02g42190	2	-	25368823	25372024	RLK	OsWAK16_ind	2	-	27089908	27094326	RLK
OsWAK17	LOC_Os02g47160	2	+	28783401	28789056	short	OsWAK17_ind	2	+	30642327	30647970	nc
OsWAK18b	LOC_Os02g56334	2	+	34478127	34480226	short**/RLP****	OsWAK18_ind	2	+	36568850	36572233	RLP
OsWAK19	LOC_Os02g56350	2	-	34486691	34487005	short	OsWAK19_ind	2	-	36570884	36571159	nc
OsWAK20	LOC_Os02g56370	2	-	34503654	34506805	RLK	OsWAK20_ind	2	-	36577089	36579354	RLK
OsWAK21	LOC_Os02g56380	2	-	34510753	34513605	RLCK/**RLK****	OsWAK21_ind	2	-	36590578	6591375	RLK
OsWAK22	LOC_Os02g56400	2	-	34525687	34528410	RLK	OsWAK22_ind	2	-	36596114	36598861	RLK
OsWAK23	LOC_Os02g56420	2	+	34539408	34547740	short/**RLK****	OsWAK23_ind	2	+	36608835	36620998	RLK
OsWAK24	LOC_Os02g56630	2	+	34714997	34717490	RLK	OsWAK24_ind	2	+	36840531	36843023	RLK
OsWAK25	LOC_Os03g12470	3	+	6588671	6593037	RLK	OsWAK25_ind	3	+	7058866	7063224	RLK
OsWAK26	LOC_Os03g44140	3	-	24795883	24805379	RLK	OsWAK26_ind	3	-	28716174	28722414	RLP
								3	-	28694607	28697665	
OsWAK27	LOC_Os03g44050	3	-	24757338	24761574	RLK	OsWAK27_ind	3	-	28628034	28632280	RLK

OsWAK28	LOC_Os03g62430	3	-	35353602	35358035	RLK	OsWAK28_ind	3	-	39724403	39728802	RLK
OsWAK29	LOC_Os04g03830	4	+	1718766	1722846	RLK	OsWAK29_ind	4	+	1703881	1707963	RLK
OsWAK30/31	LOC_Os04g14304	4	+	8004292	8013812	RLP	OsWAK30/31_ind	Sc336	+	6243	12560	RLP
OsWAK32	LOC_Os04g24220	4	+	13853940	13858978	RLK	OsWAK32_ind	4	+	11976909	11982585	RLK
OsWAK33	LOC_Os04g21820	4	+	12355146	12360363	RLP/**RLK***	OsWAK33_ind	4	+	10196511	10201248	RLP
OsWAK34	LOC_Os04g21790	4	+	12334041	12339359	RLK	OsWAK34_ind	4	+	10196542	10201062	RLP
OsWAK35a	LOC_Os04g24300	4	+	13910617	13915832	short/**RLK****	OsWAK35a_ind	4	-	10825669	10830881	RLK
OsWAK35b	LOC_Os04g24304	4	-	13917025	13919084	short	***					
OsWAK35d	LOC_Os04g24294	4	-	13909271	13916746	short/**RLK****	OsWAK35d_ind	4	+	10824955	10832433	RLK
OsWAK36	LOC_Os04g24510	4	+	14047115	14052003	RLK	OsWAK36_ind	4	+	12130246	12134025	RLK
OsWAK37	LOC_Os04g29580	4	+	17413031	17417795	short/**RLK****	OsWAK37_ind	4	+	15010522	15015287	RLK
OsWAK38	LOC_Os04g29680	4	-	17495896	17499452	RLK	OsWAK38_ind	4	-	15104011	15107569	RLK
OsWAK39	LOC_Os04g29740	4	-	17549872	17555538	RLK	OsWAK39_ind	Sc17	-	2374538	2377901	RLK
OsWAK40	LOC_Os04g29790	4	-	17580278	17584546	RLP	***					
OsWAK41	LOC_Os04g29880	4	-	17644044	17646621	RLP	***					
OsWAK42	LOC_Os04g29930	4	-	17674090	17678257	RLK	OsWAK42_ind	4	-	15276502	15278628	RLK
OsWAK43	LOC_Os04g29960	4	-	17702355	17705322	RLK	OsWAK43_ind	4	-	15285186	15288157	RLK
OsWAK44	LOC_Os04g29990	4	-	17716187	17719615	RLK	OsWAK44_ind	4	-	15305033	15308459	RLP
OsWAK45	LOC_Os04g30010	4	-	17734138	17737843	RLK	OsWAK45_ind	4	-	15333095	15336144	RLK
OsWAK46	LOC_Os04g30160	4	-	17829060	17833437	RLK	OsWAK46_ind	4	-	15368941	15375179	RLK
OsWAK47	LOC_Os04g30260	4	-	17889490	17892841	RLK	OsWAK47_ind	4	-	15496021	15499376	RLK
OsWAK48	LOC_Os04g30280	4	-	17906165	17912097	RLP	OsWAK48_ind	4	-	15541737	15543562	RLP
OsWAK50	LOC_Os04g51040	4	-	30029935	30032564	RLK	OsWAK50_ind	4	-	29063538	29066112	RLK
OsWAK51	LOC_Os04g43730	4	-	25699618	25703158	RLK	OsWAK51_ind	4	-	24587644	24591213	RLK
OsWAK52	LOC_Os04g51009	4	-	30010793	30015576	short/**RLK****	OsWAK52_ind	4	-	29042876	29048124	RLP
								Sc2267	-	2219	3889	RLCK
OsWAK53b	LOC_Os04g51050	4	-	30035161	30038861	RLK	OsWAK53b_ind	4	-	29068710	29072246	RLK
OsWAK54	LOC_Os04g55750	4	+	33000701	33001777	short	OsWAK54_ind	4	+	32192011	32193087	nc
OsWAK55	LOC_Os04g55760	4	+	33007511	33011019	RLK	OsWAK55_ind	4	+	32204479	32208012	RLK
OsWAK56	LOC_Os05g04460	5	-	2053456	2057572	RLK	OsWAK56_ind	5	-	2254091	2258132	RLK
OsWAK57	LOC_Os04g30370	4	+	17949471	17954182	RLK	OsWAK57_ind	Sc9488	-	1	2221	RLCK
OsWAK58	LOC_Os04g30360	4	+	17944048	17945345	RLK/**RLP***	***					
OsWAK59	LOC_Os04g30330	4	-	17927189	17930069	RLCK	***					

OsWAK60	LOC_Os04g30240	4	-	17870170	17877479	RLK	OsWAK60_ind	4	-	15486263	15493528	RLK
OsWAK61	LOC_Os06g05050	6	+	2228207	2232268	RLK	OsWAK61_ind	6	+	2390381	2393895	RLK
OsWAK62	LOC_Os06g07260	6	-	3476114	3476404	short	***					
OsWAK63	LOC_Os06g07330	6	-	3514131	3517029	RLK	***					
OsWAK64	LOC_Os06g49170	6	-	29790165	29794255	RLK	OsWAK64_ind	6	-	31645654	31649707	RLK
OsWAK65	LOC_Os06g49260	6	-	29849304	29854923	RLK	OsWAK65_ind	6	-	31708287	31713457	RLK
OsWAK66	LOC_Os07g07390	7	+	3689168	3691323	pseudo/**RLK***	OsWAK66_ind	7	+	3583782	3586148	RLCK
OsWAK67	LOC_Os07g14470	7	+	8247400	8248023	short	***					
OsWAK68	LOC_Os07g14490	7	+	8256202	8257257	pseudo/**RLCK***	***					
OsWAK69	LOC_Os07g31250	7	+	18497006	18501820	RLCK	OsWAK69_ind	7	+	16699247	16701980	RLCK
OsWAK70	LOC_Os07g31210	7	+	18478050	18480059	RLCK	OsWAK70_ind	7	+	16690209	16692812	RLCK
OsWAK71	LOC_Os07g31190	7	+	18452662	18461463	RLCK	OsWAK71_ind	7	+	16671500	16673532	RLCK
OsWAK72	LOC_Os07g31130	7	+	18415304	18419427	RLCK	***					
OsWAK73	LOC_Os08g39180	8	-	24747340	24752621	RLK	OsWAK73_ind	8	-	26387258	26392571	RLK
OsWAK74	LOC_Os08g39210	8	-	24764629	24768176	RLK	OsWAK74_ind	8	-	26404357	26407904	RLK
OsWAK75	LOC_Os08g39220	8	+	24770247	24773031	RLK	OsWAK75_ind	8	+	26410171	26412955	RLK
OsWAK76	LOC_Os08g39240	8	+	24777596	24783376	RLCK/**RLK***	OsWAK76_ind	8	+	26417492	26423289	RLK
OsWAK77	LOC_Os08g27780	8	-	16916342	16917379	RLCK	OsWAK77_ind	8	-	17986233	17987264	RLCK
OsWAK79	LOC_Os09g20740	9	-	12493568	12497571	RLK	***					
OsWAK80	LOC_Os09g29510	9	-	17948935	17953034	RLK	OsWAK80_ind	9	-	16625087	16629187	RLK
OsWAK81	LOC_Os09g29520	9	-	17960168	17961784	RLCK	OsWAK81_ind	9	-	16637777	16639576	RLCK
OsWAK82	LOC_Os09g29540	9	-	17967401	17970303	RLCK	***					
OsWAK83	LOC_Os09g29560	9	-	17975510	17978417	pseudo/**RLK***	OsWAK83_ind	9	-	16643600	16646489	RLK
OsWAK84	LOC_Os09g29584	9	-	17986173	17995612	RLP/**RLCK***	***					
OsWAK85	LOC_Os09g29600	9	-	18001497	18007035	RLCK/**RLK****	OsWAK85_ind	9	-	16679179	16684719	RLK
OsWAK86	LOC_Os09g16980	9	-	10377531	10380849	RLCK	OsWAK86_ind	9	-	9140055	9141566	RLCK
OsWAK87	LOC_Os09g30454	9	+	18549670	18552900	RLK	OsWAK87_ind	9	+	17201596	17204829	RLK
OsWAK88	LOC_Os09g38800	9	+	22291868	22295105	pseudo/**RLK***	OsWAK88_ind	9	+	21128388	21131640	RLK
OsWAK89a	LOC_Os09g38830	9	+	22302503	22305042	RLK	OsWAK89a_ind	9	+	21135072	21137608	RLK
OsWAK89b	LOC_Os09g38834	9	+	22307021	22309699	RLK	OsWAK98b_ind	9	+	21139588	21142278	RLK
OsWAK90	LOC_Os09g38840	9	+	22311116	22314168	RLK	OsWAK90_ind	9	+	21156023	21159244	RLK
OsWAK91	LOC_Os09g38850	9	+	22315245	22318384	RLK	OsWAK91_ind	9	+	15704045	15707785	RLK
OsWAK92	LOC_Os09g38910	9	+	22348258	22351014	RLK	OsWAK92_ind	9	+	21191333	21194407	RLK

OsWAK93	LOC_Os10g01390	10	-	238156	241301	RLCK	OsWAK93_ind	7	+	8298241	8301393	RLCK
OsWAK94	LOC_Os10g01410	10	-	248310	250408	RLP	OsWAK94_ind	10	-	88379	89905	RLP
OsWAK95	LOC_Os10g02250	10	-	791735	798344	RLK	OsWAK95_ind	10	-	688611	682784	RLCK
OsWAK96	LOC_Os10g02276	10	+	815226	816179	RLK/**RLP***	OsWAK96_ind	4	-	6818331	6819282	RLP
OsWAK97	LOC_Os10g02284	10	+	817676	820860	RLK	***					
OsWAK98	LOC_Os10g02360	10	-	847146	849408	RLCK	OsWAK98_ind	10	-	780921	783189	RLCK
								10	-	776801	779062	RLCK
OsWAK99	LOC_Os10g02720	10	-	1065852	1071048	RLK	***					
OsWAK100	LOC_Os10g05170	10	+	2522948	2523430	RLCK	OsWAK100_ind	3	-	32964682	32965164	RLCK
OsWAK101	LOC_Os10g05259	10	-	2593690	2595888	short/**RLK****	OsWAK101_ind	10	-	2396119	2398305	RLCK
OsWAK102	LOC_Os10g05400	10	+	2661288	2663240	RLCK	OsWAK102_ind	10	+	2395753	2397702	RLCK
								3	-	32787125	32789077	RLCK
OsWAK103	LOC_Os10g06030	10	+	3059913	3064524	RLK	OsWAK103_ind	10	+	2850707	2855312	RLK
OsWAK104	LOC_Os10g06090	10	+	3089573	3096422	RLK	OsWAK104_ind	10	+	2913199	2920596	RLK
OsWAK105	LOC_Os10g06140	10	-	3120996	3126249	RLK	OsWAK105_ind	10	-	2936589	2941832	RLK
							OsWAK103/5_ind	6	-	30737383	30742176	RLK
OsWAK106	LOC_Os10g09550	10	-	5174173	5177495	RLK	***					
OsWAK107	LOC_Os10g09570	10	-	5179982	5181470	RLP	OsWAK107_ind	10	-	4385679	4386748	RLP
OsWAK108	LOC_Os10g09620	10	-	5215480	5221183	RLK	OsWAK108_ind	10	-	4455581	4463656	RLK
OsWAK109	LOC_Os10g09690	10	+	5254116	5255240	RLK/**RLP***	OsWAK109_ind	10	+	4507536	4508660	RLP
OsWAK110	LOC_Os10g09700	10	+	5258692	5261784	RLK	OsWAK110_ind	10	+	4512103	4515169	RLK
OsWAK111	LOC_Os10g10030	10	-	5458681	5465611	RLK	OsWAK111_ind	10	-	4662887	4668053	RLK
OsWAK112d	LOC_Os10g10130	10	-	5554559	5561117	RLK	***					
OsWAK113	LOC_Os10g17890	10	+	8970821	8971732	RLP	OsWAK113_ind	10	+	7419007	7419918	RLP
OsWAK114	LOC_Os10g17910	10	+	8985162	8986337	RLCK	OsWAK114_ind	10	+	7422545	7423664	RLCK
OsWAK115	LOC_Os08g27810	8	-	16936882	16939072	RLP	OsWAK115_ind	8	-	18025188	18027206	RLP
OsWAK116	LOC_Os11g35120	11	-	20122967	20125006	RLCK	OsWAK116_ind	11	-	17303303	17305342	RLCK
OsWAK117	LOC_Os11g35220	11	-	20184951	20188025	RLCK	OsWAK117_ind	11	-	17355381	17358220	RLCK
OsWAK118	LOC_Os11g35260	11	-	20201483	20202160	RLCK	OsWAK118_ind	11	-	17368010	17368687	RLCK
OsWAK119	LOC_Os11g35290	11	-	20217631	20219616	RLCK	OsWAK119_ind	11	-	17393532	17395517	RLCK

OsWAK120	LOC_Os11g35860	11	-	20580361	20582361	RLK	OsWAK120_ind	11	-	17852448	17854278	RLP
OsWAK121	LOC_Os11g47110	11	+	27798505	27804757	RLK	***					
OsWAK122	LOC_Os11g47140	11	+	27842362	27846663	RLK	***					
OsWAK124	LOC_Os12g16540	12	-	9473783	9476303	RLP	***					
OsWAK125	LOC_Os12g29430	12	+	17471960	17473667	RLP	OsWAK125_ind	12	+	13844436	13846147	RLP
OsWAK126	LOC_Os12g42040	12	+	26039751	26042595	RLK	OsWAK126_ind	12	+	21716743	21719715	RLK
OsWAK127b	LOC_Os12g42044	12	-	26048058	26053235	short/**RLK****	OsWAK127b_ind	12	-	21704378	21709557	RLK
OsWAK128b	LOC_Os12g42060	12	+	26053463	26057050	RLK	OsWAK128b_ind	12	+	21709755	21713342	RLK
OsWAK129b	LOC_Os12g42064	12	+	26060046	26064120	RLK	OsWAK129b_ind	12	+	21715819	21719893	RLK
OsWAK129c	LOC_Os12g42070	12	-	26059665	26064338	RLK	OsWAK129c_ind	12	-	21715438	21720111	RLK

* Classification through domain analysis of annotated protein

** Classification through new splicing prediction and protein domain analysis from the annotated genomic region

*** Not found in *O. sativa* indica

New WAK classifications are in bold

Table S2: Genes without differences between indica and japonica.

	O. sativa japonica					*O. sativa* indica				
Name	**Chr**	**Sense**	**Start**	**End**	**Classification**	**Chr**	**Sense**	**start**	**end**	**Classification**
OsWAK1	1	-	1956777	1966451	RLCK	1	-	2312092	2322889	RLCK
OsWAK3	1	+	11654112	11658869	RLCK	1	+	12700909	12705668	RLCK
OsWAK4	1	+	11676457	11684281	RLCK	1	+	12723592	12730428	RLCK
OsWAK5	1	+	14824118	14827241	RLK	1	+	16416095	16417637	RLCK
OsWAK6	1	-	14855072	14857244	RLK	1	-	16423999	16426223	RLK
OsWAK9	1	-	14902859	14904395	RLCK	1	-	16479505	16481032	RLCK
OsWAK10d	1	-	28480966	28495222	RLCK	1	-	31786631	31800354	RLCK
OsWAK11	2	-	627342	630655	RLK	2	-	664488	667758	RLK
OsWAK13	2	-	24865877	24868342	RLK	2	-	26627307	26629782	RLK
OsWAK14	2	-	25342352	25345873	RLK	2	-	27048464	27051967	RLK
OsWAK16	2	-	25368823	25372024	RLK	2	-	27089908	27094326	RLK
OsWAK20	2	-	34503654	34506805	RLK	2	-	36577089	36579354	RLK
OsWAK22	2	-	34525687	34528410	RLK	2	-	36596114	36598861	RLK
OsWAK24	2	+	34714997	34717490	RLK	2	+	36840531	36843023	RLK
OsWAK25	3	+	6588671	6593037	RLK	3	+	7058866	7063224	RLK
OsWAK27	3	-	24757338	24761574	RLK	3	-	28628034	28632280	RLK
OsWAK28	3	-	35353602	35358035	RLK	3	-	39724403	39728802	RLK
OsWAK29	4	+	1718766	1722846	RLK	4	+	1703881	1707963	RLK
OsWAK32	4	+	13853940	13858978	RLK	4	+	11976909	11982585	RLK
OsWAK36	4	+	14047115	14052003	RLK	4	+	12130246	12134025	RLK
OsWAK38	4	-	17495896	17499452	RLK	4	-	15104011	15107569	RLK
OsWAK42	4	-	17674090	17678257	RLK	4	-	15276502	15278628	RLK
OsWAK43	4	-	17702355	17705322	RLK	4	-	15285186	15288157	RLK
OsWAK45	4	-	17734138	17737843	RLK	4	-	15333095	15336144	RLK
OsWAK46	4	-	17829060	17833437	RLK	4	-	15368941	15375179	RLK
OsWAK47	4	-	17889490	17892841	RLK	4	-	15496021	15499376	RLK
OsWAK48	4	-	17906165	17912097	RLP	4	-	15541737	15543562	RLP
OsWAK50	4	-	30029935	30032564	RLK	4	-	29063538	29066112	RLK
OsWAK51	4	-	25699618	25703158	RLK	4	-	24587644	24591213	RLK
OsWAK53b	4	-	30035161	30038861	RLK	4	-	29068710	29072246	RLK
OsWAK55	4	+	33007511	33011019	RLK	4	+	32204479	32208012	RLK
OsWAK56	5	-	2053456	2057572	RLK	5	-	2254091	2258132	RLK
OsWAK60	4	-	17870170	17877479	RLK	4	-	15486263	15493528	RLK
OsWAK61	6	+	2228207	2232268	RLK	6	+	2390381	2393895	RLK
OsWAK64	6	-	29790165	29794255	RLK	6	-	31645654	31649707	RLK
OsWAK65	6	-	29849304	29854923	RLK	6	-	31708287	31713457	RLK
OsWAK69	7	+	18497006	18501820	RLCK	7	+	16699247	16701980	RLCK
OsWAK70	7	+	18478050	18480059	RLCK	7	+	16690209	16692812	RLCK
OsWAK71	7	+	18452662	18461463	RLCK	7	+	16671500	16673532	RLCK
OsWAK73	8	-	24747340	24752621	RLK	8	-	26387258	26392571	RLK
OsWAK74	8	-	24764629	24768176	RLK	8	-	26404357	26407904	RLK
OsWAK75	8	+	24770247	24773031	RLK	8	+	26410171	26412955	RLK
OsWAK77	8	-	16916342	16917379	RLCK	8	-	17986233	17987264	RLCK
OsWAK80	9	-	17948935	17953034	RLK	9	-	16625087	16629187	RLK
OsWAK81	9	-	17960168	17961784	RLCK	9	-	16637777	16639576	RLCK
OsWAK86	9	-	10377531	10380849	RLCK	9	-	9140055	9141566	RLCK
OsWAK87	9	+	18549670	18552900	RLK	9	+	17201596	17204829	RLK
OsWAK89a	9	+	22302503	22305042	RLK	9	+	21135072	21137608	RLK
OsWAK89b	9	+	22307021	22309699	RLK	9	+	21139588	21142278	RLK
OsWAK90	9	+	22311116	22314168	RLK	9	+	21156023	21159244	RLK
OsWAK91	9	+	22315245	22318384	RLK	9	+	15704045	15707785	RLK
OsWAK92	9	+	22348258	22351014	RLK	9	+	21191333	21194407	RLK
OsWAK94	10	-	248310	250408	RLP	10	-	88379	89905	RLP
OsWAK98	10	-	847146	849408	RLCK	10	-	780921	783189	RLCK
OsWAK102	10	+	2661288	2663240	RLCK	10	+	2395753	2397702	RLCK
OsWAK103	10	+	3059913	3064524	RLK	10	+	2850707	2855312	RLK
OsWAK104	10	+	3089573	3096422	RLK	10	+	2913199	2920596	RLK
OsWAK105	10	-	3120996	3126249	RLK	10	-	2936589	2941832	RLK
OsWAK107	10	-	5179982	5181470	RLP	10	-	4385679	4386748	RLP
OsWAK108	10	-	5215480	5221183	RLK	10	-	4455581	4463656	RLK
OsWAK110	10	+	5258692	5261784	RLK	10	+	4512103	4515169	RLK
OsWAK111	10	-	5458681	5465611	RLK	10	-	4662887	4668053	RLK
OsWAK113	10	+	8970821	8971732	RLP	10	+	7419007	7419918	RLP
OsWAK114	10	+	8985162	8986337	RLCK	10	+	7422545	7423664	RLCK
OsWAK115	8	-	16936882	16939072	RLP	8	-	18025188	18027206	RLP
OsWAK116	11	-	20122967	20125006	RLCK	11	-	17303303	17305342	RLCK
OsWAK117	11	-	20184951	20188025	RLCK	11	-	17355381	17358220	RLCK
OsWAK118	11	-	20201483	20202160	RLCK	11	-	17368010	17368687	RLCK
OsWAK119	11	-	20217631	20219616	RLCK	11	-	17393532	17395517	RLCK
OsWAK125	12	+	17471960	17473667	RLP	12	+	13844436	13846147	RLP
OsWAK126	12	+	26039751	26042595	RLK	12	+	21716743	21719715	RLK
OsWAK128b	12	+	26053463	26057050	RLK	12	+	21709755	21713342	RLK
OsWAK129b	12	+	26060046	26064120	RLK	12	+	21715819	21719893	RLK
OsWAK129c	12	-	26059665	26064338	RLK	12	-	21715438	21720111	RLK

Table S3:WAK genes not found in indica genome.

		O. sativa japonica				
Name	**Locus**	**Chr**	**Sense**	**Start**	**End**	**Classification**
OsWAK7	LOC_Os01g26270	1	+	14891850	14892917	RLK/**RLP***
OsWAK15	LOC_Os02g42160	2	-	25350877	25354142	RLK
OsWAK35b	LOC_Os04g24304	4	-	13917025	13919084	short
OsWAK40	LOC_Os04g29790	4	-	17580278	17584546	RLP
OsWAK41	LOC_Os04g29880	4	-	17644044	17646621	RLP
OsWAK58	LOC_Os04g30360	4	+	17944048	17945345	RLK/**RLP***
OsWAK59	LOC_Os04g30330	4	-	17927189	17930069	RLCK
OsWAK62	LOC_Os06g07260	6	-	3476114	3476404	short
OsWAK63	LOC_Os06g07330	6	-	3514131	3517029	RLK
OsWAK67	LOC_Os07g14470	7	+	8247400	8248023	short
OsWAK68	LOC_Os07g14490	7	+	8256202	8257257	pseudo/**RLCK***
OsWAK72	LOC_Os07g31130	7	+	18415304	18419427	RLCK
OsWAK79	LOC_Os09g20740	9	-	12493568	12497571	RLK
OsWAK82	LOC_Os09g29540	9	-	17967401	17970303	RLCK
OsWAK84	LOC_Os09g29584	9	-	17986173	17995612	RLP/**RLCK***
OsWAK97	LOC_Os10g02284	10	+	817676	820860	RLK
OsWAK99	LOC_Os10g02720	10	-	1065852	1071048	RLK
OsWAK106	LOC_Os10g09550	10	-	5174173	5177495	RLK
OsWAK112d	LOC_Os10g10130	10	-	5554559	5561117	RLK
OsWAK121	LOC_Os11g47110	11	+	27798505	27804757	RLK
OsWAK122	LOC_Os11g47140	11	+	27842362	27846663	RLK
OsWAK124	LOC_Os12g16540	12	-	9473783	9476303	RLP

* Classification through domain analysis of annotated protein

New WAK classifications are in bold

Table S4: Primer pairs used on RTqPCR experiments

Gene	Forward Primer	Reverse Primer
OsWAK2	ATGGGTGCAGAAGGAGAGG	CAACGCCAAGTATACCGAGA
OsWAK6	CTGGATTGACGAATTCAACG	TTGTGGCACATTCCCTTACA
OsWAK11	GGCCTGGGACTCAAAGTACA	GCTTTCTTTTCTGGCATTCG
OsWAK14	GCTACTGCCTGAACGTCACC	AATTTTGACACGTGCCATTG
OsWAK16	CCGCAGTATCCACAGCAAT	TGGAACCAATGCCACAACTA
OsWAK20	AGTGCACCAACACGCTAGG	GAATGCCATGGCCAAAAG
OsWAK22	GGCTACCAGGGTAACCCCTA	GCAAACATTCTTCTCGTGGA
OsWAK25	CATGTGGCAAGTTGGTTCAG	TCGCGGTACGTGTAGAAGG
OsWAK28	TGGTTATCGTTGCAAGTGCT	TGCACTGCCTTCTTACTCGAT
OsWAK32	ATGCATGGATGGCTATTTCG	CACCCTTCTTCTAGGCCAAA
OsWAK38	CCTTGCTATGGGAAGTGCAT	TTACGGACAAGGAATGCTACG
OsWAK50	GACAGGAGATGGCAAGAAGC	GCTGCTGCAATAACAAACCA
OsWAK53b	ACCATTAAGAGGCCATGCAC	GCCAATGGGTCCAGAATATG
OsWAK55	GCCAGCAGCCAGATGTTTAT	TGATTTTGGCAACACCAAGA
OsWAK56	ATTACTGTTCCTGCCGCAAT	TGTGCCCTTCGAATTCTCTT
OsWAK73	AATGGAACCTACGGGGATG	ATCTTGCGCAGCTTCATTTT
OsWAK74	GCAATTCGACATGTGGTACG	CGCCACTGACTACAAGACCA
OsWAK75	GAGAGGCTATGCGTGCAAGT	GTGCCACCTGGACAGATACA
OsWAK76	TACGCCAGGAGCTTTTGTTT	GCGAACTATGAAGGGAGCAG
OsWAK87	AAGGAGGATGTGAACCAACG	CAACAAGTTGCAGCAGCAGT
OsWAK92	CAACGCAATGGATGTTTCAG	CCCACGATTCTGCTTGAAGT
OsWAK98b	ATGTTGCACCATCCAAGACA	GCGTCAAGACAAGAACACCA
OseFa1	GACTTCCTTCACGATTTCATCGTAA	TTTCACTCTTGGTGTGAAGCAGAT
OsFDH	TTCCAATGCATTCAAAGCTG	CAAAATCAGCTGGTGCTTCTC

CAPÍTULO III – CONSIDERAÇÕES FINAIS

Discussão e Conclusão

Até o presente momento, as relações filogenéticas entre os genes da subfamília WAK de arroz e arabidopsis permaneciam obscuras, pois as filogenias ou não apresentavam uma análise criteriosa com suporte estatístico adequado (Zhang et al. 2005), ou analisavam um número muito grande de genes de subfamílias de quinases distintas, sendo que a relação entre os membros pertencentes a subfamília WAK poderiam ser afetados por artefatos na filogenia (Lehti-Shiu et al., 2009; Shiu et al., 2004; Philippe et al., 2011). As análises realizadas neste trabalho permitiram melhorar a compreensão sobre as relações filogenéticas entre os genes da subfamília WAK de arroz e arabidopsis. Com a comparação das análises dos resíduos de aminoácidos do domínio quinase e a filogenia destes genes foi possível visualizar que a expansão dos genes WAK no genoma do arroz está, de alguma forma, associada com o surgimento de um novo grupo de WAKs pertencentes às quinases não-RD. A partir desses resultados poi possível sugerir a divisão das WAKs em duas subfamílias evolutivamente relacionadas mas independentes: OsWAK-RD e OsWAK-non-RD. Esta nova classificação é baseada nas relações filogenéticas e na análise do resíduo de arginina conservado no motivo RDxxxxN do subdomínio VIB, podendo contribuir para um melhor entendimento e organização de aspectos funcionais destas duas subfamílias de WAK, principalmente para monocotiledôneas.

Em arabidopsis já foi relatado que a WAK1 parece estar envolvida na resposta tanto a bactéria *Pseudomonas syringae* (He et al., 1998) quanto ao

fungo *Botrytis cinerea* (Brutus et al., 2010). Em arroz, a única quinase funcionalmente descrita foi a então denominada OsWAK1(Li et al. 2009), que não é a mesma descrita por Zhang et al., 2005 (LOC_Os01g04409), e aparentemente está envolvida na resposta e tolerância do arroz ao fungo *Magnaporthe oryzae*. Analisando os resíduos presentes na sequência da proteína, apresentada neste trabalho (Li et al. 2009), observamos que a mesma pertence à classe de quinases não-RD. Porém, no trabalho os autores não fornecem o código do locus do presente gene. Buscamos o locus este gene no genoma do arroz, utilizando a ferramenta BLASTP contra o banco de pseudo-moléculas de proteínas de arroz, e a ferramenta TBLASTN contra o genoma do arroz, em ambos os métodos, utilizando como isca a sequência protéica apresentada no trabalho. Não conseguimos determinar o locus deste gene, pois encontramos genes subfamília WAK com 62% de identidade com o presente gene, porém estas sequências resgatados pelo BLAST pertenciam a OsWAKs não-RD. Em arabidopsis encontramos apenas WAKs RD-quinases, logo o gene WAK1, que é responsivo a fungo, também pertence as WAKs RD-quinase. A superexpressão de um gene WAK não-RD em arroz (Li et al. 2009) tornou a planta resistente a infecção por fungo. Este resultado pode indicar que o surgimento de um subgrupo da subfamília WAK como quinases não-RD, poderia estar associada a uma especialização dos mesmos a respostas contra patógenos, uma vez que as quinases não-RD parecem estar associadas à resposta contra patógenos (Dardick and Ronald 2006; Afzal et al., 2008). Porém, mais estudos experimentais são necessários para elucidar melhor as implicações funcionais destas duas

subclasses em arroz.

A utilização de métodos de agrupamento de genes baseados na homologia de sequências, através alinhamentos locais, vem sendo apontado como uma forma eficiente de analisar um grande volume de sequências proteicas onde não é possível a utilização de análises convencionais tais como métodos filogenéticos, devido ao volume e heterogeneidade dos dados (Atkinson et al., 2009; Goldstein et al., 2009). Neste trabalho, além de empregar métodos filogenéticos para compreender a relação evolutiva das WAKs, utilizamos métodos de agrupamentos baseados em alinhamentos locais, realizados através da ferramenta BLASTp. Com esta abordagem foi possível identificar grupos de genes WAK que teriam o potencial de serem redundantes (Figura 3, do Capítulo 2), tanto na percepção do sinal extracelular quanto na ativação de uma cascata de transdução de sinal. Também encontramos dois grupos de genes OsWAK (Figura 4B, do Capítulo 2) que teriam o potencial de perceber o mesmo sinal, porém ativar cascatas de sinalizações distintas. A comparação desses resultados, para o domínio quinase, em conjunto à análise filogenética apresentou alta congruência entre os agrupamentos obtidos em ambos os métodos. Já no domínio extracelular, onde existe uma grande heterogeneidade de sequências, não é possível a realização de análises filogenéticas, porém a utilização de métodos de agrupamento por similaridade de sequências demonstra a potencialidade desta estratégia para situações similares ao do presente trabalho.

A análise comparativa entre os genomas de arroz ,das subespécies *indica* e *japonica,* revelaram a existência de uma ampla diversidade entre suas

WAKs, em aspectos estruturais dos genes, e no padrão de expressão em resposta a tratamento com baixas temperaturas. A espécie *O. sativa*, bem como suas subespécies *indica* e *japonica*, surgiram através do melhoramento de uma espécie selvagem, onde *O. rufipogon* tem sido apontada como a mais provável ancestral destas (London et al., 2006). A domesticação do arroz ocorreu na china, sendo que, evidências arqueológicas estimam que a domesticação de culturas agrícolas nesta região iniciou-se a ~9.000-11.000 anos atrás (MacNeish, 1992). Este seria o tempo estimado para a separação das duas subespécies de arroz.,Do ponto de vista evolutivo, essa separação ocorreu recentemente. O evento de expansão dos genes OsWAK, bem como o processo de seleção artificial sobre estas subespécies podem ser dois fatores importantes para que estas diversificações estruturais e de regulação da expressão dos genes OsWAK, entre as subespécies *indica* e *japonica*, tenham ocorrido em tão pouco tempo. O genoma de *A. thaliana*, tem aproximadamente 157 milhões de pares de base (Bennett et al., 2003), já o genoma de arroz possui aproximadamente 389 milhões de pares de bases (Shiguo Zhou et al., 2007). Sendo assim, o genoma do arroz é aproximadamente 2.45 vezes maior que o genoma de *A. thaliana*. E, o genoma de arroz possui 4.8 vezes mais genes WAKs que arabidopsis. Esse aumento desproporcional das WAKs no genoma de arroz ocasionou uma potencial redundância funcional entre os vários genes OsWAK com redução da pressão de seleção nos genes, com consequente aumento dos eventos de subfuncionalização dos mesmos (Kleinjan et al., 2008; Nielsen et al., 2010).

Este trabalho contribuiu para melhorar o entendimento das relações

entre a subfamília gênica WAK de arroz e arabidopsis, através da proposição de uma nova classificação para as WAKs, WAK-RD e WAK-nonRD, tendo como base análise filogenética e características de resíduos conservados. Através das análises comparativas entre as subespécies de arroz *indica* e *japonica*, foi possível descrever variações estruturais e diferenças no padrão de expressão das WAKs nestas duas subespécies. Os resultados presentes neste trabalho, além de abrir novas perspectivas, podem servir como base para o direcionamento de estudos futuros que visem tanto a caracterização funcional como evolutiva das WAKs.

Perspectivas

Futuros trabalhos serão necessários para uma melhor compreensão sobre os papéis exercidos por esta família no genoma do arroz. Estudos prévios já apontaram para o envolvimento das WAKs tanto na resposta de estresses bióticos, quanto abióticos, em arabidopsis. Para arroz, foi reportado anteriormente o envolvimento da OsWAKs na resposta a fungo, e neste reportamos genes que respondem a estresse ao frio. Seria conveniente que outros experimentos envolvendo, tanto resposta a outros estresses bióticos e abióticos quanto a expressão tecido-específica e ao longo do desenvolvimento fossem avaliados para o direcionamento de estudos funcionais. A obtenção de mutantes nulos para estes genes seria interessante para a demonstração do potencial de redundância

funcional destes genes. Outro ponto interessante seria a utilização de sistemas que permitissem a detecção da interação com outras proteínas, tanto no domínio extracelular, como no domínio quinase, para tentar compreender melhor a maneira de atuação das proteínas das OsWAK na transdução de sinal.

REFERÊNCIAS

Afzal AJ, Wood AJ, and Lightfoot DA (2008). Plant receptor-like serine threonine kinases: roles in signaling and plant defense. *Mol Plant Microbe Interact* 21, 5: 507-17.

Albrecht C, Russinova E, Hecht V, Baaijens E, de Vries S (2005) The Arabidopsis thaliana SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1 and 2 control male sporogenesis. *Plant Cell* 17:3337-3349.

Anderson, CM, Wagner TA, Perret M, He ZH, He D and Kohorn BD (2001). WAKs: cell wall-associated kinases linking the cytoplasm to the extracellular matrix. *Plant Mol Biol* 47, 1-2: 197-206.

Atkinson HJ, Morris JH, Ferrin TE and Babbitt, PC (2009). Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PloS one* 4: e4345.

Bennett MD, Leitch IJ, Price HJ and Johnston JS (2003). Comparisons with Caenorhabditis (~100 Mb) and Drosophila (~175 Mb) Using Flow Cytometry Show Genome Size in Arabidopsis to be ~157 Mb and thus 25% Larger than the Arabidopsis Genome Initiative Estimate of ~125 Mb. Ann Bot, 91:547-557

Brutus A, Sicilia F, Macone A, Cervone F and De Lorenzo G (2010). A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. Proc Natl Acad Sci U S A 107: 9452-9457.

Colcombet J, Boisson-Dernier A, Ros-Palau R, Vera CE, Schroeder JI (2005) Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASES1 and 2 are essential for tapetum development and microspore maturation. *Plant Cell* 17:3350-3361.

Dardick C and Ronald P (2006). Plant and animal pathogen recognition receptors signal through non-RD kinases. *PLoS pathog* 2:e2.

Decreux A, Messiaen J (2005) Wall-associated kinase WAK1 interacts with cell wall pectins in a calcium-induced conformation. *Plant Cell Physiol* 46:268-278.

Decreux A, Thomas A, Spies B, Brasseur R, Van Cutsem P, Messiaen J (2006) In vitro characterization of the homogalacturonan-binding domain of the wall-associated kinase WAK1 using site-directed mutagenesis. *Phytochemistry* 67:1068-1079.

Endre G, Kereszt A, Kevei Z, Mihacea S, Kalo P, Kiss GB (2002) A receptor kinase gene regulating symbiotic nodule development. *Nature* 417:962-966.

Goldstein P, Zucko J, Vujaklija D, Krisko A, Hranueli D, Long PF, Etchebest C, Basrak B and Cullum J (2009). Clustering of protein domains for functional and evolutionary studies. *BMC bioinformatics* 10: 335.

Gomez-Gomez L, Boller T (2000) FLS2: an LRR receptor-like kinase involved in

the perception of the bacterial elicitor flagellin in Arabidopsis. *Mol Cell* 5:1003-1011.

Hanks SK and Hunter T (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* 9:576-96.

Hanks SK, Quinn M and Hunter T (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* 241: 42-52.

Hanks S (2003). Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biology* 4: 111.

He ZH, Fujiki M and Kohorn BD (1996). A cell wall-associated, receptor-like protein kinase. *J Biol Chem* 271: 19789-93.

He ZH, He D, and Kohorn BD (1998). Requirement for the induced expression of a cell wall associated receptor kinase for survival during the pathogen response. *Plant J* 14:55-63.

Hou X, Tong H, Selby J, Dewitt J, Peng X, He ZH (2005) Involvement of a cell wall-associated kinase, WAKL4, in Arabidopsis mineral responses. *Plant Physiol* 139:1704-1716.

Johnson LN, Noble EM and Owen DJ (1996). Active and Inactive Protein Kinases: Structural Basis for Regulation. *Cell* 85: 149-158.

Kleinjan DA, Bancewicz RM, Gautier P, Dahm R, Schonthaler HB, et al. (2008) Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence. PLoS genetics 4: e29

Kohorn BD (2001). WAKs; cell wall associated kinases. *Curr Opin Cell Biol* 13: 529-33.

Kohorn BD, Johansen S, Shishido A, Todorova T, Martinez R, Defeo E, and Obregon P (2009). Pectin activation of MAP kinase and gene expression is WAK2 dependent. *Plant J* 60: 974-82

Kohorn BD, Kobayashi M, Johansen S, Friedman HP, Fischer A, Byers N (2006) Wall-associated kinase 1 (WAK1) is crosslinked in endomembranes, and transport to the cell surface requires correct cell-wall synthesis. *J Cell Sci* 119:2282-2290.

Krupa A, Preethi G and Srinivasan N (2004). Structural modes of stabilization of permissive phosphorylation sites in protein kinases: distinct strategies in Ser/Thr and Tyr kinases. *J Mol Biol* 339: 1025-39.

Lehti-Shiu MD, Zou C, Hanada K, and Shiu SH (2009). Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiol* 150: 12-26.

Li H, Zhou SY, Zhao WS, Su SC, and Peng YL (2009). A novel wall-associated receptor-like protein kinase gene, OsWAK1, plays important roles in rice blast

disease resistance. *Plant Mol Biol* 69: 337-46.

Londo JP, Chiang YC, Hung KH, Chiang TY; Schaal BA (2006). Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple independent domestications of cultivated rice, Oryza sativa. PNAS, 103: 9578-9583.

MacNeish RS (1992) The Origins of Agriculture and Settled Life (Univ. of Oklahoma Press, Norman), pp. 150–156.

Morris E and Walker JC (2003). Receptor-like protein kinases: the keys to response. *Curr Opin Plant Biol* 6: 339-342.

Muto H, Yabe N, Asami T, Hasunuma K, Yamamoto KT (2004) Overexpression of constitutive differential growth 1 gene, which encodes a RLCKVII-subfamily protein kinase, causes abnormal differential and elongation growth after organ differentiation in Arabidopsis. *Plant Physiol* 136:3124-3133.

Nielsen MG, Gadagkar SR and Gutzwiller L (2010) Tubulin evolution in insects: gene duplication and subfunctionalization provide specialized isoforms in a functionally constrained gene family. BMC Evol Biol 10: 113

Nolen B, Taylor S and Ghosh G. (2004). Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell* 15: 661-75.

Osakabe Y, Maruyama K, Seki M, Satou M, Shinozaki K and Yamaguchi-Shinozaki K (2005). Leucine-rich repeat receptor-like kinase1 is a key membrane-bound regulator of abscisic acid early signaling in Arabidopsis. *Plant Cell* 17: 1105-19.

Park AR, Cho SK, Yun UJ, Jin MY, Lee SH, Sachetto-Martins G, Park OK (2001) Interaction of the Arabidopsis receptor protein kinase Wak1 with a glycine-rich protein, AtGRP-3. *J Biol Chem* 276:26688-26693.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G and Baurain D (2011). Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. Ed. David Penny. *PLoS Biol* 9: e1000602.

Shiguo Z, Bechner MC, Place M, Churas CP, Pape L, Leong SA, Runnheim R, Forrest DK, Goldstein S, Livny M and Schwartz DC (2007). Validation of rice genome sequence by optical mapping. BMC Genomics 2007, 8:278

Shiu SH and Bleecker AB (2001a). Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *PNAS* 98: 10763-8.

Shiu SH and Bleecker AB (2001b). Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci STKE* 2001: re22.

Shiu SH and Bleecker AB (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. *Plant Physiol* 132: 530-43.

Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KFX, and Li WH (2004). Comparative analysis of the receptor-like kinase family in Arabidopsis and

rice. *Plant Cell* 16: 1220-34.

Shpak ED, Berthiaume CT, Hill EJ, Torii KU (2004) Synergistic interaction of three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation. *Development* 131:1491-1501.

Sivaguru M, Ezaki B, He ZH, Tong H, Osawa H, Baluska F, Volkmann D, Matsumoto H (2003) Aluminum-induced gene expression and protein localization of a cell wall-associated receptor kinase in Arabidopsis. *Plant Physiol* 132:2256-2266.

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH*, et al.* (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270:1804-1806.

Verica JA, He ZH (2002) The cell wall-associated kinase (WAK) and WAK-like kinase gene family. *Plant Physiol* 129:455-459.

Torii KU, Mitsukawa N, Oosumi T, Matsuura Y, Yokoyama R, Whittier RF, Komeda Y (1996) The Arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell* 8:735-746.

Wagner TA, Kohorn BD (2001) Wall-associated kinases are expressed throughout plant development and are required for cell expansion. *Plant Cell* 13:303-318.

Walker JC and Zhang R (1990). Relationship of a putative receptor protein kinase from maize to the S-locus glycoproteins of Brassica. *Nature* 345: 743-6.

Yang EJ, Oh YA, Lee ES, Park AR, Cho SK, Yoo YJ, Park OK (2003) Oxygen-evolving enhancer protein 2 is phosphorylated by glycine-rich protein 3/wall-associated kinase 1 in Arabidopsis. *Biochem Biophys Res Commun* 305:862-868.

Zhang S, Chen C, Li L, Meng L, Singh J, Jiang N, Deng XW, He ZH, and Lemaux PG (2005). Evolutionary expansion, gene structure, and expression of the rice wall-associated kinase gene family. *Plant Physiol* 139: 1107-24.