The Cell Wall-Associated Kinase (*WAK*) and *WAK*-Like Kinase Gene Family[1]

Joseph A. Verica and Zheng-Hui He*

Department of Biology, San Francisco State University, 1600 Holloway Avenue, San Francisco, California 94132

We have identified a large family of genes with sequence similarity to the cell wall-associated kinase (*WAK*) genes (He et al., 1999). Like the *WAK*s, these genes exist in multiple gene clusters, and our analyses suggest that they encode functional protein kinases that are associated with the cell wall. The *WAK*s represent a unique class of receptor-like kinase genes. Each encodes a transmembrane protein with a cytoplasmic Ser/Thr kinase (STK) domain and an extracellular region with similarity to vertebrate epidermal growth factor (EGF)-like domains. WAKs are thought to physically link the extracellular matrix and the cytoplasm and to serve a signaling function between them (He et al., 1996; Kohorn, 2000). Consistent with these ideas, WAK1 is covalently bound to pectin in the cell wall (Wagner and Kohorn, 2001) and can form an approximately 500-kD protein complex via interactions with a Gly-rich extracellular protein, AtGRP-3 (Park et al., 2001) and a cytoplasmic type 2C protein phosphatase, KAPP (Anderson et al., 2001).

Previous findings indicate that WAKs are involved in the response to pathogens (He et al., 1998). Induction of *WAK1* is required for plants to survive *Pseudomonas syringae* infection. In addition, *WAK1* is induced by salicylic acid in an *NPR1*-dependent manner (*n*onexpresser of *p*athogenesis-*r*elated genes), demonstrating that it is a pathogenesis-related gene. Moreover, *WAK1* is up-regulated during systemic acquired resistance (Maleck et al., 2000) and is induced by the fungal pathogen *Alternaria brassicicola* and the defense-related signaling molecules methyl jasmonate and ethylene (Schenk et al., 2000).

WAKs have recently been shown to be required for cell expansion (Lally et al., 2001; Wagner and Kohorn, 2001). For example, Lally et al. (2001) used a glucocorticoid-inducible system to control expression of a *WAK4* antisense gene. Induction of *WAK4* antisense expression caused a decrease in WAK protein levels that could be controlled by the concentration of inducer applied. Reduction of WAK protein levels resulted in inhibition of cell elongation, and the degree of inhibition was correlated with the concentration of inducer applied (Lally et al., 2001).

Reiterative database searches (BLAST) using the *WAK1* cDNA or WAK1 protein sequences as queries identified a large family containing 22 genes in Arabidopsis similar to *WAK*s (Table I). We have called these sequences *WAK-like* genes (*WAKL*s), as suggested by Shiu and Bleecker (2001). Intron-exon junctions were verified by identifying and sequencing cDNA clones for each of the *WAKL* genes. With the exception of three sequences (see below), all the intron-exon predictions were correct.

WAKL3 (At1g16140) is predicted to have three introns; however, our analysis shows that there are only two. The first intron (17 bp), predicted to reside between nucleotides 64,385 and 64,403 on the bacteria artificial chromosome (BAC) clone (T24D18), was not present. This results in a shift in the reading frame at the 5′ end of the first exon. As such, the first "ATG" codon in the longest open reading frame (ORF) is predicted to reside at nucleotide 64,155. *WAKL6* (At1g16110) was predicted to have a stop codon at nucleotide 55,719 on BAC T24D18; however, sequencing of a *WAKL6* cDNA shows that the stop codon was not present. The stop codon for *WAKL6* resides at nucleotide 55,984, resulting in a 2,196-bp ORF. *WAKL8* (At1g16260) is predicted to have three exons and two introns, similar to other *WAKL*s. Sequencing of a *WAKL8* cDNA shows that only the first intron is present. This intron is predicted to extend from nucleotides 16,177 to 16,098 (3′–5′) on BAC F3O9; however, our analysis shows that it extends from 16,177 to 16,120. In addition, a stop codon is present at nucleotide 16,093, giving rise to a truncated ORF of 879 bp. These corrected sequences were used in all subsequent analyses.

Five of the *WAKL* genes (*WAKL7*, *WAKL8*, *WAKL12*, *WAKL16*, and *WAKL19*) are predicted to encode abbreviated WAKL proteins. The remaining 17 genes share similar intron-exon organization with the *WAK*s, with each having three exons and two introns (Fig. 1). Each of the non-truncated *WAKL*s is predicted to encode a transmembrane protein containing a cytoplasmic STK (ProSite: PS00108. ProSite—a database for protein families and domains—http://

[1] This work was supported by the National Institutes of Health, Minority Biomedical Research Support-Support of Continuous Research Excellence Program (grant no. S06 GM52588), the National Institutes of Health Research Infrastructure in Minority Institutions Grant (no. 5 P20 RR11805), and by the National Science Foundation (grant no. MCB–9985135).

* Corresponding author; e-mail zhe@sfsu.edu; fax 415–338–2295.

www.plantphysiol.org/cgi/doi/10.1104/pp.011028.

Table I. *WAK and WAKL genes in* Arabidopsis

Each of the *WAK* and *WAKL* genes is indicated, along with its location within the *Arabidopsis* genome. Accession nos. are given for corresponding ESTs. *WAKL19* was not assigned a locus number because it was not identified as a transcriptional unit in the database. *WAKL19* was uncovered via a TBLASTN search and verified by the identification and sequencing of a *WAKL19* cDNA clone.

Gene	Location	ESTs
WAK1	At1g21250	T04358, AI998376, AW004557, AV442022, AV442007, AV439996, AV562748, AV520441, AV561465, AV526755
WAK2	At1g21270	N65506, N65248, AI994173, AV526813
WAK3	At1g21240	
WAK4	At1g21210	
WAK5	At1g21230	
WAKL1	At1g16120	
WAKL2	At1g16130	
WAKL3	At1g16140	
WAKL4	At1g16150	
WAKL5	At1g16160	
WAKL6	At1g16110	AI999346, AV524272
WAKL7	At1g16090	
WAKL8	At1g16260	
WAKL9	At1g69730	AI997196, Z48380, F13735
WAKL10	At1g79680	
WAKL11	At1g19390	
WAKL12	At1g22720	
WAKL13	At1g17910	
WAKL14	At2g23450	AI999307, AV544096, BE524984, AV552564
WAKL15	At3g53840	
WAKL16	At3g25490	
WAKL17	At4g31100	BE845240
WAKL18	At4g31110	AA067421, AV527054, BE527556
WAKL19	BAC F3I3	
WAKL20	At5g02070	
WAKL21	At5g66790	AA605380, H37747
WAKL22	At1g79670	AV440728, AV442794

www.expasy.ch/prosite/) and an extracellular region containing a calcium-binding EGF-like domain (EGF-Ca^{2+}; PS01187) and/or an EGF2-like domain (EGF2; PS01186). In some cases (see below), the EGF-like domains were slightly degenerate. In these cases, all the Cys residues predicted to be involved in the formation of disulfide bridges were conserved; however, the spacing between them was off by one to two residues or other hallmark residues characteristic of EGF-like domains were changed. Because EGF domains have not been well characterized in plants, it is not clear what effect, if any, this degeneracy has on their potential function.

We have divided the *WAK*s and *WAKL*s into four groups based on pairwise comparisons of their predicted protein sequences (Fig. 2). Group I contains five members, *WAK1* through *WAK5* (He et al., 1999). WAK1 through WAK5 all have an EGF-Ca^{2+} domain and an overlapping Asn hydroxylation site (PS00010). Immediately following the EGF-Ca^{2+} domain is an EGF2 domain, which is degenerate in both WAK4 and WAK5. These domains are completely

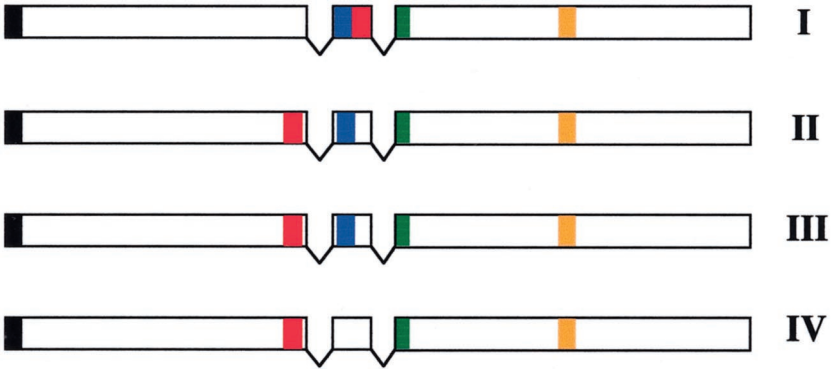

Figure 1. Intron-exon structure is conserved between the *WAK* and *WAKL* genes. The diagram shows a standardized depiction of a *WAK* or *WAKL* gene from each of the four groups (I-IV). Exons are represented by boxes. Introns are represented as 'V's. Regions of each gene encoding functional domains are indicated with shaded boxes: N-terminal signal sequence (black), EGF2-like domain (red), calcium-binding EGF domain (blue), transmembrane domain (green), and Ser/Thr protein kinase active site (orange).

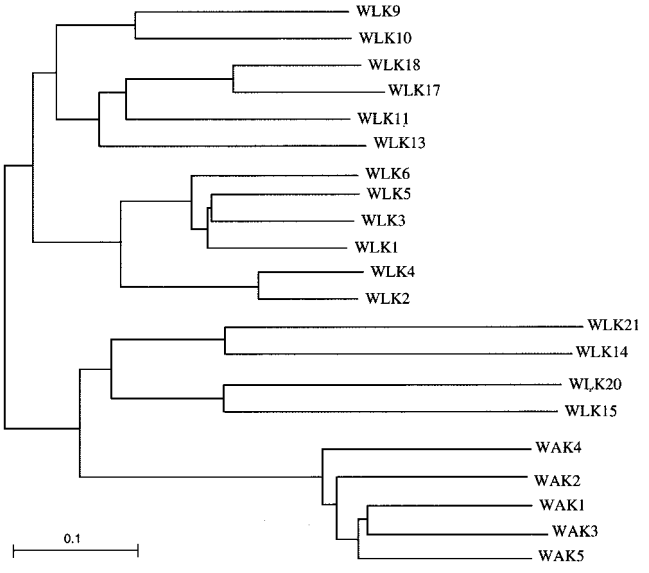

Figure 2. Dendrogram showing the relationships between the various WAKs and WAKLs. The tree was generated by ClustalW analysis with the corrected full-length WAK/WAKL protein sequences using the MacVector software package (version 7.0; Accelrys, San Diego). The bar beneath the dendrogram represents a distance of 0.1 change per amino acid.

encoded by the second exon (Fig. 1). Group II includes seven members, *WAKL1* through *WAKL6* and *WAKL22*. Their predicted proteins all contain EGF-Ca^{2+} and EGF2 domains (both of which are degenerate in WAKL1, WAK6, and WAKL22). In addition, the domains are separated by a short gap of 15 to 18 amino acids and are in reversed order relative to Group I. The regions of the genes encoding them are split by the first intron (Fig. 1). Group III contains six members: *WAKL9*, *WAKL10*, *WAKL11*, *WAKL13*, *WAKL17*, and *WAKL18*. Their corresponding proteins all contain EGF-Ca^{2+} and EGF2 domains, and they are structurally similar to the Group II WAKLs. In WAKL13, the EGF-Ca^{2+} domain is degenerate. With the exception of WAKL17, all have degenerate EGF2 domains (Fig. 1). Group IV contains four members: *WAKL14*, *WAKL15*, *WAKL20*, and *WAKL21*. Each has an EGF2 domain encoded by the first exon. This domain is degenerate in both WAKL20 and WAKL21. All four members lack the EGF-Ca^{2+} domain (Fig. 1). In addition, each has a cytoplasmic protein kinase ATP-binding domain (PS00107).

The remaining sequences (*WAKL7*, *WAKL8*, *WAKL12*, *WAKL16*, and *WAKL19*) are predicted to encode abbreviated WAKL proteins. WAKL7, WAKL8 and WAKL19 are similar to various other WAKLs in their extracellular regions, and lack a transmembrane domain. WAKL8 and WAKL9 both contain an EGF-Ca^{2+} domain and WAKL19 contains a degenerate EGF2 domain. Neither of these domains is present in WAKL7. WAKL12 also contains an EGF-Ca^{2+} domain, but unlike WAKL8, it contains a transmembrane domain. WAKL16 contains a transmembrane domain, an STK domain that is most similar to WAK3, and a short extracellular domain of eight amino acids that lacks both of the EGF-like domains. Recent evidence form our laboratory shows that at least three of these genes (*WAKL7*, *WAKL8*, and *WAKL19*) are expressed (J.A. Verica and Z.H. He, unpublished data). Their protein products are likely secreted from the cell. This raises the possibility that they may play a role in the formation of an active WAKL receptor complex. A similar role has been proposed for SCR/SP11 in the SRK receptor complex in self-incompatible *Brassica* spp. (Schopfer et al., 1999; Takayama et al., 2000). Alternatively, these sequences may encode nonfunctional WAKL isoforms. Potential roles for the remaining two sequences (*WAKL12* and *WAKL16*) remain to be determined.

Expressed sequence tags (ESTs) have been identified for many of the WAK and WAKL genes (Table I). In addition, we have isolated cDNAs for all of the non-abbreviated members. Bacterial expression of several *WAKL* members shows that they have autophosphorylation activity, and analysis of several of the WAKL proteins shows that they localize to the cell wall (J.A. Verica and Z.H. He, unpublished data). This suggests that the WAKLs, like the WAKs, are protein kinases that are tightly associated with the cell wall.

The *WAK* and *WAKL* genes are distributed among all five chromosomes (Fig. 3), with the majority (19) being present on chromosome I. Sixteen of the genes (*WAK1-*

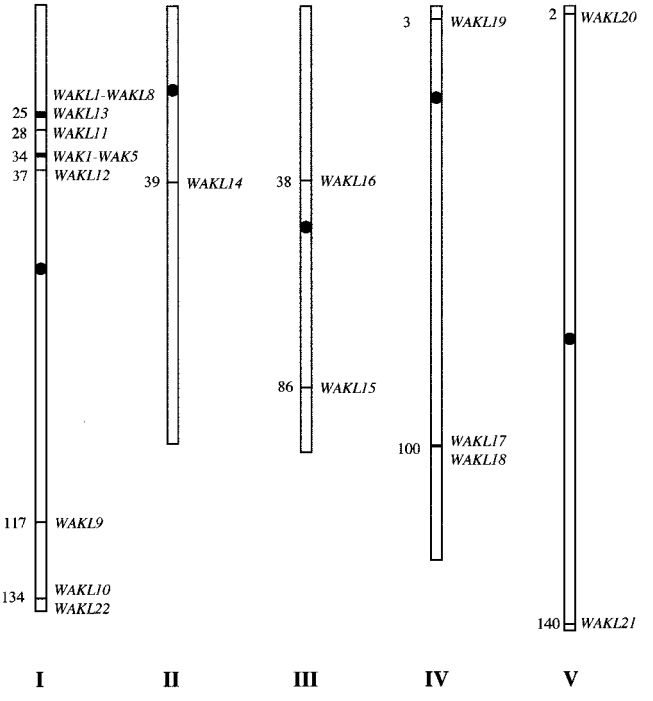

Figure 3. The *WAK* and *WAKL* genes are distributed among all five chromosomes. Chromosomes (I–V) are indicated by the vertical bars. Centromeres are indicated by the darkened circles. Horizontal bars indicate the location of each of the WAKL/WAK genes, and their physical position is given in centiMorgans.

WAK5, *WAKL1*-*WAKL8*, and *WAKL11*-*WAKL13*) are located on the upper arm within a region spanning less than 12 cM. Moreover, *WAK1* through *WAK5* and *WAKL1* through *WAKL7* are both present as two separate clusters of tandemly arrayed genes (He et al., 1999). *WAKL9*, *WAKL10*, and *WAKL22* are located on the lower arm, with *WAKL10* and *WAKL22* being adjacent to the telomere. The remaining six genes are distributed more or less evenly among chromosomes II through V. Chromosome II contains only one gene on its lower arm, *WAKL14*. Chromosome III contains two genes (*WAKL15* and *WAKL16*), one on each arm. Chromosome IV contains three genes. *WAKL19* is located on the upper arm adjacent to the telomere. *WAKL17* and *WAKL18* are located on the lower arm as two tandem genes. Chromosome V contains two genes, *WAKL20* and *WAKL21*, that are located at opposite ends of the chromosome in regions directly adjacent to the telomeres.

Closer examination of the Arabidopsis genome reveals several potential explanations for the expansion of the *WAK* gene family. For example, the arrangement of and the high sequence similarity between the genes in the *WAKL1* through *WAKL6* cluster suggest that they may have arisen via tandem duplication. Likewise, the genes in the *WAK1* through *WAK5* cluster may have independently arisen via a similar scenario, although other possibilities cannot be ruled out. For example, the regions flanking the *WAK1*, *WAK2*, *WAK4*, *WAK5*, *WAKL11*, and *WAKL16* contain sequences that are similar to the copia-like retrotransposon Hopscotch (White et al., 1994) or non-long terminal repeat retroelements (Kumar and Bennetzen, 1999). These retroelements transpose via RNA intermediates that are generated by reverse transcription. In some cases, adjacent genes (or portions of genes) may be acquired by a retrotransposon, resulting in their amplification and dispersal throughout the genome (Kumar and Bennetzen, 1999). If these retroelements are functional in Arabidopsis, this could be one mechanism for the expansion of the *WAKL* gene family. A third possibility arises from the observation that the Arabidopsis genome contains extensive duplications of large chromosomal segments (Arabidopsis Genome Initiative, 2000; Blanc et al., 2000). Several of the *WAKL* genes are contained within these regions. For example, the region overlapping *WAKL11* and *WAKL13* on the lower arm of chromosome I is duplicated on the upper arm in the region overlapping the locations of *WAKL9* and *WAKL10*, suggesting that these genes could have arisen via segmental duplication. The high degree of sequence similarity between these genes is consistent with this possibility.

The existence of the WAKs and WAKLs is not unique to Arabidopsis. For example, protein gel blots using the WAK1 antibody as a probe revealed immunologically related proteins in pea (*Pisum sativum*), tobacco (*Nicotiana tabacum*), and maize (*Zea mays*; He et al., 1996; Gens et al., 2000). Expressed sequence tags for *WAK*-like genes have been identified in tomato (*Lycopersicon esculentum*; GenBank accession nos. AW220490, AW455238, and AW622503) and soybean (*Glycine max*; GenBank accession no. BE473800). In addition, GenBank searches have revealed that there are at least 10 *WAK*-like genes in rice (*Oryza sativa*; accession nos. AC079685, AF327447, AL442007, and AP003021). Like the Arabidopsis *WAK*s, many of the rice *WAK*s are clustered. Moreover, they are also flanked by transposon-like elements (MITEs, miniature inverted-repeat transposable elements; Bureau et al., 1996). These genes are more similar to *WAK* genes than they are to any of the *WAKL* genes, suggesting that the *WAKL*s may be unique to Arabidopsis. Completion of the rice genome and further sequencing of other plant genomes may help to resolve this issue.

The large number of genes in this family may provide Arabidopsis with the potential to recognize and respond to a diverse array of ligands. The observations that WAKs play roles in both the pathogen response and cell elongation suggest that they function in some manner that is common to both processes. For example, both processes involve alteration of the cell wall (Cosgrove, 1997; Grant and Mansfield, 1999). As such, the tight association of WAKs and WAKLs to the extracellular matrix could allow them to function by responding to architectural changes that occur within the wall.

Received December 11, 2001; returned for revision January 13, 2002; accepted February 23, 2002.

LITERATURE CITED

Anderson CM, Wagner TA, Perret M, He ZH, He D, Kohorn BD (2001) Plant Mol Biol **47:** 197–206

Arabidopsis Genome Initiative (2000) Nature **408:** 796–815

Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000) Plant Cell **12:** 1093–1102

Bureau TE, Ronald PC, Wessler SR (1996) Proc Natl Acad Sci USA **93:** 8524–8529

Cosgrove DJ (1997) Annu Rev Cell Dev Biol **13:** 171–201

Gens JS, Fujiki M, Pickard BG (2000) Protoplasma **212:** 115–134

Grant M, Mansfield J (1999) Curr Opin Plant Biol **2:** 312–319

He ZH, Cheeseman I, He D, Kohorn BD (1999) Plant Mol Biol **39:** 1189–1196

He ZH, Fujiki M, Kohorn BD (1996) J Biol Chem **271:** 19789–19793

He ZH, He D, Kohorn BD (1998) Plant J **14:** 55–63

Kohorn B (2000) Plant Physiol **124:** 31–38

Kumar A, Bennetzen JL (1999) Annu Rev Genet **33:** 479–532

Lally D, Ingmire P, Tong HY, He ZH (2001) Plant Cell **13:** 1317–1331

Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangl JL, Dietrich RA (2000) Nat Genet **26:** 403–410

Park AR, Cho SK, Yun UJ, Jin MY, Lee SH, Sachetto-Martins G, Park OK (2001) J Biol Chem **276:** 26688–26693

Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM (2000) Proc Natl Acad Sci USA **97:** 11655–11660

Schopfer CR, Nasrallah ME, Nasrallah JB (1999) Science **286:** 1697–1700

Shiu SH, Bleecker AB (2001) Proc Natl Acad Sci USA **98:** 10763–10768

Takayama S, Shiba H, Iwano M, Shimosata H, Che FS, Kai N, Watanabe M, Suzuki G, Hinata K, Isogai A (2000) Proc Natl Acad Sci USA **97:** 1920–1925

Wagner T, Kohorn B (2001) Plant Cell **13:** 303–318

White SE, Habera L, Wessler SR (1994) Proc Natl Acad Sci USA **91:** 11792–11796