Molecular Plant • Volume 1 • Number 5 • Pages 800–815 • September 2008 RESEARCH ARTICLE

Classification, Expression Pattern, and E3 Ligase Activity Assay of Rice U-Box-Containing Proteins

Li-Rong Zeng[a,b,2], Chan Ho Park[a,2], R.C. Venu[a], Julian Gough[c] and Guo-Liang Wang[a,d,1]

a Department of Plant Pathology, Plant Molecular Biology and Biotechnology Program, The Ohio State University, 2021 Coffey Rd, 201 Kottman Hall, Columbus, OH 43210, USA
b Present address: Boyce Thompson Institute for Plant Research, Tower Rd, Ithaca, NY 14853, USA
c Department of Computer Science, University of Bristol, Woodland Rd, Bristol BS8 1UB, UK
d Crop Gene Engineering Key Laboratory of Hunan Province, Hunan Agricultural University, Changsha, Hunan 410128, China

ABSTRACT Ubiquitin ligases play a central role in determining the specificity of the ubiquitination system by selecting a myriad of appropriate candidate proteins for modification. The U-box is a recently identified, ubiquitin ligase activity-related protein domain that shows greater presence in plants than in other organisms. In this study, we identified 77 putative U-box proteins from the rice genome using a battery of whole genome analysis algorithms. Most of the U-box protein genes are expressed, as supported by the identification of their corresponding expressed sequence tags (ESTs), full-length cDNAs, or massively parallel signature sequencing (MPSS) tags. Using the same algorithms, we identified 61 U-box proteins from the *Arabidopsis* genome. The rice and *Arabidopsis* U-box proteins were classified into nine major classes based on their domain compositions. Comparison between rice and *Arabidopsis* U-box proteins indicates that the majority of rice and *Arabidopsis* U-box proteins have the same domain organizations. The inferred phylogeny established the homology between rice and *Arabidopsis* U-box/ARM proteins. Cell death assay using the rice protoplast system suggests that one rice U-box gene, *OsPUB51*, might act as a negative regulator of cell death signaling. In addition, the selected U-box proteins were found to be functional E3 ubiquitin ligases. The identification and analysis of rice U-box proteins hereby at the genomic level will help functionally characterize this class of E3 ubiquitin ligase in the future.

INTRODUCTION

Ubiquitination is a major type of post-translational modification of proteins that occurs in eukaryotic cells (Ciechanover, 2005). The ubiquitination process involves in-concert catalytic activities of three types of enzymes, ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin ligase (E3). The ubiquitin molecule—a highly conserved 76-amino acid protein—is initially activated by E1 in an ATP-dependent manner, and the activated ubiquitin is then transferred to an E2. With the help of an E3, the ubiquitin molecule is covalently attached to the target protein through the ε-amino group of a lysine residue (Ciechanover and Schwartz, 1998). The attachment of the first ubiquitin moiety to the substrate protein is usually followed by the addition of more ubiquitin molecules to form a polyubiquitin chain. A polyubiquitin chain with at least four sequentially attached ubiquitins linked through the lysine 48 of the ubiquitin molecules will be sufficient to allow target protein to be recognized and degraded by the 26S proteasome (Thrower et al., 2000). Importantly, polyubiquitination that links through other lysine residues of the ubiquitin molecules as well as mono-ubiquitination also regulate non-proteolytic processes in the cell (Bray et al., 2005; Huang et al., 2006; Li et al., 2003; Terrell et al., 1998).

Of the three enzymes that are implicated in the ubiquitination process, E3 ubiquitin ligase plays a central role in determining specificity of the ubiquitination system by selecting appropriate candidate proteins (Ciechanover, 1998). Not surprisingly, given the myriad of different substrates, E3s are the most diverse components in the ubiquitination pathway. For example, over 1300 distinct E3 components are encoded by the *Arabidopsis* genome (Smalle and Vierstra, 2004). During the transfer of activated ubiquitin molecule from E2 to the substrate, the ubiquitin molecule may either be or not be covalently attached to the E3 ligase before it is attached to the substrate. Accordingly, E3 ligases can be largely classified into two classes—the HECT type and the family of RING/U-box-type ligases—based on the occurrence of such covalent linkage. The

[1] To whom correspondence should be addressed. E-mail wang.620@osu.edu, fax 614 292 4455, tel. 614 292 9280.

[2] These authors contributed equally to the project.

doi: 10.1093/mp/ssn044
Received 12 May 2008; accepted 23 June 2008

HECT-type ligases are the only known E3 ubiquitin ligases that form an E3-ubiquitin thiolester intermediate before the final attachment of ubiquitin to the substrate (Huibregtse et al., 1995; Scheffner et al., 1993). They contain an approximately 350-amino-acid in length HECT (homologous to E6-AP COOH terminus) domain that was initially identified in the viral E6-associated protein (E6-AP). A conserved active cysteine residue that is required for the ligase function is presented at the C-termini of these ligases (Scheffner et al., 1993). Depending on their subunit composition and mechanism of actions, members of the RING/U-box family can be further classified into three sub-classes—the CRL (Cullin-RING-Ligase) subgroup, including SCF type (Skp1-Cullin (CDC53)-F-box protein complex) (Bai et al., 1996; Skowyra et al., 1997) and other cullin-based types (such as ElonginC-CUL2-SOCS box type, CUL3-BTB type, etc.) (Pintard et al., 2003; Stebbins et al., 1999); RING (really interesting new gene) (Freemont et al., 1991)/U-box (Koegl et al., 1999) subgroup; and APC (anaphase promoting complex) subgroup (Irniger et al., 1995; King et al., 1995; Sudakin et al., 1995). The family of RING/U-box-type E3 ligases can be either single-subunit proteins or multi-subunit complexes. The RING-type E3 ligases belong to the class of single-polypeptide (subunit) E3s that contain a motif known as RING. The RING motif is maintained by the arrangement of eight cysteines and histidines in a cross-brace manner that chelate two zinc ions, which is distinct from other zinc finger domains (Borden, 2000). It is proposed that the RING-type E3 ligases serve as scaffolds to bring together the activated ubiquitin-E2 intermediate and the substrate protein to promote the transfer of the ubiquitin molecule to the substrate protein (Borden, 2000; Lorick et al., 1999).

The U-box is a more recently identified E3 ligase activity-related protein domain that was first shown in yeast to be involved in polyubiquitin chain assembly (Koegl et al., 1999). The U-box contains ~70 amino acids and possesses a tertiary structure resembling that of the RING domain (Aravind and Koonin, 2000; Ohi et al., 2003). The major difference between U-box and RING domains is that the U-box lacks the characteristic zinc-chelating cysteine and histidine residues. Consequently, the U-box E3s use intramolecular interactions other than zinc chelation to maintain the RING finger motif (Ohi et al., 2003). U-box proteins are present in yeast, plants, and animals (Azevedo et al., 2001; Koegl et al., 1999; Meacham et al., 2001; Stone et al., 2003). Nevertheless, only a few U-box proteins have been functionally studied to date. UFD2 (ubiquitin fusion degradation protein-2) is responsible for the assembly of polyubiquitin chains in yeast (Koegl et al., 1999). The Prp19p (precursor RNA processing 19p) is a subunit of the spliceosome and is involved in pre-mRNA splicing (Ohi and Gould, 2002). The human homolog of Prp19p, SNEV (senescence evasion factor), has been demonstrated to be a multifaceted protein playing a role in cellular senescence, cellular lifespan, pre-mRNA splicing, DNA double-strand break repair and the transport of ubiquitinated substrate to the proteasome (Loscher et al., 2005; Mahajan and Mitchell, 2003; Voglauer et al., 2006). The CHIP (C-terminal of Hsc70-interacting protein) protein acts as chaperone or co-chaperone in the cell to ensure protein homeostasis and has been implicated in stress responses and several neurodegenerative diseases (Ballinger et al., 1999; Dickey et al., 2007; Imai et al., 2002; Jiang et al., 2001; Miller et al., 2005; Rosser et al., 2007; Sahara et al., 2005). The plant proteins CMPG1 and ACRE276 are implicated in defense against pathogen attack (Gonzalez-Lamothe et al., 2006; Kirsch et al., 2001; Yang et al., 2006). SPL11 is related to cell death and other cellular processes in rice, such as flowering time control (Zeng et al., 2004) (L. Zeng and GL Wang, unpublished data). PHOR1 (photoperiod-responsive 1) from potato is a general component of the gibberellins signaling pathway (Amador et al., 2001). The protein ARC1 (arm repeat containing 1) was shown interacting with the S-locus kinase and is involved in self-incompatibility of Brassica (Gu et al., 1998; Stone et al., 2003).

The completion of rice genome sequencing allows us to study the function of the rice U-box proteins at the genome scale (Goff et al., 2002; Yu et al., 2002). By using a battery of extensive whole genome analysis algorithms, we identify 77 genes encoding U-box domain-containing proteins from the rice genome in this study. Expressed sequence tags (ESTs) or full-length cDNAs are found for most of the identified U-box proteins. These proteins are classified into eight major groups, based on their domain organizations. Phylogenetic analysis establishes the orthologies and paralogies of rice and *Arabidopsis* U-box proteins, which will help functionally characterize plant U-box protein genes in the future. In-vitro ubiquitination assay indicates that five randomly picked rice U-box proteins display E3 ligase activity in an E2-dependent manner. Interestingly, silencing one of the U-box genes causes cell death in rice protoplasts, suggesting that it might be involved in rice cell death control or defense responses.

RESULTS

Identification of U-Box-Containing Proteins from the Rice Genome

To have a comprehensive assessment of the rice U-box proteins, genome sequences of two rice subspecies were explored in our survey (Goff et al., 2002; Yu et al., 2002). To have the latest genome information, rice genome pseudomolecules release five sequences from The Institute of Genome Research (TIGR; the same database is now moved to Michigan State University) were used in the identification of U-box genes in *japonica* subspecies (http://rice.plantbiology.msu.edu). U-box proteins from the two annotated genomes were identified with high confidence using the same battery of algorithms (see Methods). The two sets of identified U-box proteins were then compared and combined. Feng et al. (2002) and Han and Xue (2003) reported that there is, on average, approximately 1% polymorphism in amino acid sequence of the same protein encoded by the *japonica* and

indica genomes. However, proteins from the same loci of the *japonica* and *indica* genome may have lower protein sequence identity due to the outcome of different annotation procedures. Therefore we considered U-box proteins from the two genomes that show at least 95% protein sequence identity are likely encoded by the same loci and the corresponding genomic DNA sequences of these proteins were then identified and compared manually. The putative U-box proteins identified in the *indica* genome (see Supplemental Doc. 1) were eliminated from the combined list after their counterparts are confirmed to exist in the annotated *japonica* genome. Such analyses indicate that all U-box proteins identified from the *indica* genome have their corresponding counterparts in the japonica genome. Seventy-seven rice U-box proteins were identified after such extensive comparisons (Table 1, Supplemental Figure 1, and Supplemental Doc. 2). Using the same algorithms, we identified 61 U-box proteins from the *Arabidopsis* genome. Among them, sixty are common to those identified by other groups (http://www.arabidopsis.org/info/genefamily/pub.html), corroborating the efficacy of the algorithms we used in the survey (see Supplemental Table 1 and Supplemental Doc. 3).

Domain Organization of the Rice U-Box Proteins

The amino acid sequences of rice U-box proteins were used to search against the PFAM database (Bateman et al., 2004) and the National Center for Biotechnology Information (NCBI) protein database to identify other domains and motifs. Besides the U-box, there are various other protein domains/motifs present in these proteins (Table 2). The rice U-box proteins are grouped into eight classes, based on the presence of common motifs/domains other than the U-box. The numbering of these classes is mandated so that the previously described classes in *Arabidopsis* are incorporated (Azevedo et al., 2001). It is noteworthy that, in addition to U-box and the domain used for classifying, proteins in the same class could also possess other domains. For example, the U-box/ARM-type protein SPL11 was shown containing a coiled-coil motif in the central region of the protein (Zeng et al., 2004). In general, E3 ubiquitin ligases have a protein–protein interaction domain to interact with their substrates for ubiquitination (Patterson, 2002). Consistently, most rice U-box proteins contain domains or motifs that were implicated in protein–protein interactions.

Similarly to *Arabidopsis*, class I consists of one rice gene, *OsUFD2*. OsUFD2 is highly homologous to AtUFD2 (Azevedo et al., 2001), with 62% identity at amino acid level (data not shown). Previous study indicated that yeast UFD2 interacted with CDC48, a member of AAA-type ATPase family that is implicated in multiple cellular processes (Koegl et al., 1999). The region of yeast UFD2 that is proposed to interact with CDC48 and its human ortholog p97 is conserved in OsUFD2.

The largest class of rice U-box proteins contains the armadillo (ARM)/HEAT repeats. The ARM repeat is an approximately 40-amino-acid-long tandemly repeated sequence motif first identified in the *Drosophila melanogaster* segment polarity gene armadillo (Riggleman et al., 1989). It was shown to be involved in protein–protein interactions (Huber and Weis, 2001). HEAT repeats derive their name from four diverse eukaryotic proteins in which they were first identified: *h*untingtin, *e*longation factor 3, PR65/*A* subunit of protein phosphatase A, and the *T*OR (target of rapamycin) (Andrade and Bork, 1995). ARM and HEAT repeats are grouped into the same class in this study due to their structural similarity (Andrade et al., 2001). It is noteworthy that the ARM/HEAT repeats found in these proteins are quite divergent, similar to what was described previously (Mudgil et al., 2004).

The second largest class of U-box proteins in rice and *Arabidopsis* show no significant domain or motif hits other than U-box in the PFAM or NCBI databases. Nevertheless, sequence alignments do detect a conserved domain containing ~100 amino acid residues that is located close to the C-terminus of these proteins (Figure 1). In addition to the high percentage of leucine residues described before (Azevedo et al., 2001), a high percentage of homology and several highly conserved residues were detected. We named this putative domain GKL domain after the conserved Glycine (G), Lysine (K)/Arginine (R) residues and its leucine-rich feature. Sequence alignments of the *Arabidopsis* proteins from the same class gave a highly similar pattern (see Supplemental Figure 2). NtCMPG1, the *Nicotiana benthamiana* homolog to the *Arabidopsis* U-box proteins AtPUB20 and AtPUB21 from this class, was shown to be essential for plant defense and disease resistance (Gonzalez-Lamothe et al., 2006). It remains unknown, however, if members of the corresponding rice class have similar functions.

The third largest class of rice U-box-containing proteins has a kinase domain at the N-terminal region of the proteins. Given the involvement of phosphorylation modification in most, if not all, cellular processes, we speculate that members of this class of rice U-box proteins could play a broad range of roles in the cell.

In addition to the three classes mentioned above, there are another two classes that have more than one protein from each genome. These two classes contain tetratrico peptide repeat (TPR) motif and WD40 repeats domain, respectively. Both TPR and WD40 domains have been shown to be involved in protein–protein interactions (Das et al., 1998; Li and Roberts, 2001). The rice and *Arabidopsis* U-box proteins that contain the WD40 repeats are homologous to the animal and human Prp19p protein, which has been shown to be involved in pre-mRNA splicing and other biological processes (Loscher et al., 2005; Mahajan and Mitchell, 2003; Ohi and Gould, 2002; Voglauer et al., 2006).

Difference of Rice and *Arabidopsis* U-Box Proteins in Domain Contents

The domain contents of *Arabidopsis* U-box proteins were also investigated using the same method as described above

Table 1. Rice U-Box Gene Names, their Corresponding Groups, Rice Genome Locus, GenBank GIs, Protein IDs, Accessions for their cDNAs and Cognate ESTs and the Presence of their Corresponding MPSS Tags.

Class	Name	Genome locus[1]	GenBank GI	GenBank Protein ID	cDNA[2]	EST	MPSS tags[3]
I	OsPUB1 (OsUFD2)	LOC_Os03g31400	50916773	XP_468783	AK242886	CB639670, CB640908, CB619295, CB619296, AU078410, CB643846, CB643847, CB000450, CA764512, AU078411, CR281540, CF328714, CF328715	Yes
II	OsPUB2	LOC_Os05g39930	46575997		AK099529	D23908, AU164828, AI096152, C73275	Yes
II	OsPUB3	LOC_Os01g60860	21104594	BAB93187	AK065307	BI809145, AU070503, CB631190, D48112, AU165954	Yes
II	OsPUB4	LOC_Os02g13960	50251218	BAD27662	AK062592	CB630469, CB630470, CB643247, CB643248, AU032329, CF325064, CK738435, BQ908529, AU173816	Yes
II	OsPUB5	LOC_Os08g32060	7573049	BAC98577	–	–	Yes
II	OsPUB6	LOC_Os09g21120	49389067	BAD26307	–	–	Yes
II	OsPUB7	LOC_Os04g28100	38346501	CAE02102.2	AK243053	CB650396, CB650395	–
II	OsPUB8	LOC_Os02g28720	47848077	BBAD21861	AK242710	AU225089, CF310095, BI801229, AA754456, BI807368	Yes
II	OsPUB9	LOC_Os02g49950	46390655	BAD16137	AK102091	CB672091, CB672092, CB634863, CB634864, AU097467, AU033197	Yes
II	OsPUB10	LOC_Os03g16824	25815287	–	AK121978, AK060190	BE040196, CA758916, CB636819	Yes
II	OsPUB11 (SPL11)	LOC_Os12g38210	51038703	AAT94161	AK105835	CB680406, BF430451, CF305067, AT003699, AT003719, AA749599, AU173870, AU197209, BI812454	Yes
II	OsPUB12	LOC_Os06g01304	55296754	BAD67946	AK071080, AK067611	CB629073, CB629075, CB661839, CB661840, CB647707, CB647708, CB663113, CB663114, CB661157, CB661158, CB640944, BM421775, AU070995	Yes
II	OsPUB13	LOC_Os06g51130	54291136	BAD61809	–	–	Yes
II	OsPUB14	LOC_Os08g37570	42409028	BAD10281	AK072890	CA762420	Yes
II	OsPUB15	LOC_Os08g01900	42408388	BAD09539	AK106557, AK102080	AU069544, CB675602, CB675603, CB637298, CF304665, C26426, CB631051, AU092305, CA760274, CR281099	Yes
II	OsPUB16	LOC_Os01g66130	56784489	BAD82582	AK072566	CF326076	Yes
II	OsPUB17	LOC_Os08g32610	25553724	BAC24957	AK240942	–	Yes
II	OsPUB18	LOC_Os09g21740	49387716	BAD26106	AK243643	CB655398, CA756899, CB655399, D49059, D49048, AU070443, AU032876	Yes
II	OsPUB19	LOC_Os06g13090	51535151	BAD37863	–	–	Yes
II	OsPUB20	LOC_Os06g13080	51535813	BAD37898	–	–	Yes

Table 1. Continued

Class	Name	Genome locus[1]	GenBank GI	GenBank Protein ID	cDNA[2]	EST	MPSS tags[3]
II	OsPUB21	LOC_Os02g49520	46390590	BAD16089	AK065834, AK120552, AK062565, AK068514, AK060172	–	Yes
II	OsPUB22	LOC_Os06g16410	55295990	BAD68030	AK110058	CB656843, CB656844, CB638682, CB638681, CB638683, CB654780, CB654781	Yes
II	OsPUB23	LOC_Os12g17900	31414506	–	AK102429	CB644678, CB644679, BE228945, CA765839	Yes
II	OsPUB24	LOC_Os03g45420	31415948	AAP50990	AK120547, AK066133	AU101622, AU164585	Yes
II	OsPUB25	LOC_Os12g17880	31414506	–	AK099846, AK066847	CB672899, CB672900, CF306503, CB640118, CB675872, CB675873, AU056255, CB632057, AU056256, BI809202	Yes
II	OsPUB27	LOC_Os04g41250	38345231	CAD41127.2	AK100423	C72885, AU082230	Yes
II	OsPUB28	LOC_Os01g67500	56784865	BAD82105	AK101205, AK065658	CB679893, CB679894, AU094854, BX899859, C91848	Yes
II	OsPUB29	LOC_Os02g47670	41053097	BAD08040	AK101760, AK065645, AK100727	CB659631, CB659632, AU056349, CA759692, AU056350	Yes
II	OsPUB30	LOC_Os08g04470	38636778	BAD03021	–	AU069933	Yes
III	OsPUB31	LOC_Os01g64570	56785190	BAD81908	–	–	Yes
III	OsPUB32	LOC_Os04g58920	21741126	CAD41926	AK108418	CB641925, CB641926, CB641964, CB641965	Yes
III	OsPUB33	LOC_Os02g33590	50252669	BAD28838	AK108494	BI808859	Yes
III	OsPUB34	LOC_Os04g34030	39545718	CAD40926.3	–	–	Yes
III	OsPUB35	LOC_Os04g49970	38346713	CAE04863.2	AK070506, AK110959	–	–
III	OsPUB36	LOC_Os02g46500	41052814	BAD07682	AK110831	–	Yes
III	OsPUB37	LOC_Os12g06410	–	–	–	–	Yes
III	OsPUB38	LOC_Os04g35680	38344067	CAD40819.2	AK121647	–	Yes
III	OsPUB39	LOC_Os06g13870	52076769	BAD45713	–	AU174669	Yes
III	OsPUB40	LOC_Os02g50460	46390551	BAD16037	AK243317	AU100984, AU100985, AU063504, AU172336	Yes
III	OsPUB41	LOC_Os03g13740	–	–	AK109161	BQ907174, C27098, AU166815	Yes
III	OsPUB42	LOC_Os10g03440	22711531	AAN04506	–	–	Yes
III	OsPUB43	LOC_Os02g34410	46390762	BAD16270		AU184395	Yes
III	OsPUB44	LOC_Os05g36360	46981302	AAT07620	AK121082	CB669445, CB669446, CB650659, CB650660, CB645255, CB645256, CB650137, CB650138, CB649839, CB647557, BM421109, BI807625, BI808006, BQ908725, AI096168	Yes
III	OsPUB45	LOC_Os02g33680	50251283	BAD28063	–	–	Yes
III	OsPUB46	LOC_Os04g34140	38605834	CAE02914.3	AK121255	AU089778, AU089779, D49041	Yes
IV	OsPUB47	LOC_Os06g37620	54291179	BAD61851	–	CB627135, CB668863, TC257914	Yes
IV	OsPUB48	LOC_Os02g12670	46805847	BAD17181	AK289262	C73105, CB096600, CR280370, CB647037, AU075377	Yes

Table 1. Continued

Class	Name	Genome locus[1]	GenBank GI	GenBank Protein ID	cDNA[2]	EST	MPSS tags[3]
IV	OsPUB49	LOC_Os10g41220	13569983	AAK31267	AK099675	AU094827, CB672905	Yes
IV	OsPUB50	LOC_Os03g01160	AC125411	–	–	AU222811	Yes
IV	OsPUB51	LOC_Os06g04880	–	–	AK065683, AK105657	CB634275, CB634276, CB663207, CB663208, CB661735, CB661736, CB663242, CB676242	Yes
IV	OsPUB52	LOC_Os09g39620	52077180	BAD46225	–	C73201	Yes
IV	OsPUB53	LOC_Os10g40060	3876523	AAK43499	AK064032, AK073764	D48628, CB666071, CB666072, BP184834	Yes
IV	OsPUB54	LOC_Os10g40100	13876536	AAK43512	AK066841, AK065488	CB674756, BM421191	Yes
IV	OsPUB55	LOC_Os02g44610	50251841	BAD27770	–	–	Yes
IV	OsPUB56	LOC_Os09g39640	52077183	BAD46228	AK058797	CB631963, CF316548, CF316549, CA882843	Yes
IV	OsPUB57	LOC_Os03g31070	50916707	XP_468750	AK069682 AK067154	CB674493	Yes
IV	OsPUB58	LOC_Os10g40120		AAK43506	–	–	Yes
IV	OsPUB59	LOC_Os03g31000	30089735	AAP20839	–	–	–
IV	OsPUB60	LOC_Os02g44599			AK109040	–	Yes
IV	OsPUB61	LOC_Os10g01060	18481962	AAL73560	AK102556, AK061113	CB648412	Yes
IV	OsPUB62	LOC_Os02g57700	48716305	BAD22918	AK243658	BQ908820	Yes
V	OsPUB26	LOC_Os03g60140	28876015	AAO60024		CI636688, CI660771, AC133007, CK047093, CI419583, CI436441, CI441137, CI440634, CI483651	Yes
V	OsPUB69	LOC_Os08g13780	38175575	BAD01285	–	–	Yes
V	OsPUB73	LOC_Os02g28870	47848102	BAD21885	AK106789	–	–
V	OsPUB74	LOC_Os06g28590	51090767	BAD35246	AK066627	CB633820, CB633821, CA305715	Yes
V	OsPUB75	LOC_Os03g13010	20330772	AAM19135	AK068218	CB636145, CB636146, C91823, CB684588, CB684587, AU094784, BI807109	Yes
V	OsPUB76	LOC_Os04g30470	21742147	CAD40573	–	–	Yes
V	OsPUB77	LOC_Os04g49500	38346817	CAE04155.2	AK099968	–	Yes
VI	OsPUB71	LOC_Os01g12930	56783887	BAD81324	–	–	Yes
VI	OsPUB72	LOC_Os10g32880	37535206	NP_921905	AK064778	CB64800, CB64800, CB65460, CB65461, C28936, CB64479, CB64479, CB64050, CB63621, CB63622, CB63994, CB64625, CB64625, CB630643, CB630644, CB633728, CB633729, CA753148, CB684534, CB684533, CB684504, AU063371, AU063355, AU101014, CR288259, AU068591, BI81131, CR288238, AU166674, AU222544, BQ908380, AU092951, CF308939, CF308938	Yes

Table 1. Continued

Class	Name	Genome locus[1]	GenBank GI	GenBank Protein ID	cDNA[2]	EST	MPSS tags[3]
VII	OsPUB63	LOC_Os06g06490	34898307	BAD69206	AK063418	CA758128, CF294635, CF294636, CF306389, CF305746, CF303031, CF303481, C73147, CA765281, AU075414	Yes
VII	OsPUB64	LOC_Os06g06470	55296606	BAD69204	AK068486	CB669618, CB669619, CB666007, CB666008, CB655998, CB655999, CB656035, CB656036, AU068890, CB651921, CB651922, CB662541, CB662542, CB659153, CB659154, CB673778, CB649572, CB672550, CA762292, CA763973, AU068891, BI807157, AU078129, CA999626, C20397, C50849	Yes
VII	OsPUB65	LOC_Os06g06450	55296605	BAD69203	AK101286	–	Yes
VII	OsPUB66	LOC_Os05g01460			AK069824	–	Yes
VII	OsPUB67	LOC_Os10g40490	31433414	–	AK121815	CA759007, CA759008	Yes
VII	OsPUB68	LOC_Os08g02140	42409284	BAD10546	AK106523	AU081544	Yes
VIII	OsPUB70	LOC_Os06g06760	55296082	BAD67644	AK069245, AK069675, AK072504	CB684359, CB684360, CB661779, BI812595, BI811980	Yes

1 Based on Rice Psedomolecules Release 5 information.
2 KOME (http://cdna01.dna.affrc.go.jp/cDNA/) database was used.
3 Rice MPSS database (http://mpss.udel.edu/rice/) was used for blast.

Table 2. Domain Organizations of Rice and *Arabidopsis* U-Box Proteins.

Group	Domain organization (N-terminus → C-terminus)	Number of proteins included	
		Rice	*Arabidopsis*
I	UFD2 specific motif + U-box	1	1
II	U-box + ARM/HEAT	28	29
III	U-box + GKL-box	16	12
IV	Kinase + U-box	16	9
V	U-box only	8	7
VI	U-box + WD40	2	2
VII	TPR + U-box	6	1
VIII	TPR + Kinase + U-box	1	0
IX	MIF4G + U-box	0	2

(Table 2). One significant difference between rice and *Arabidopsis* U-box proteins is that there is only one TPR-U-box-type protein in *Arabidopsis*, while six are identified in the rice genome. The TPR is a motif present in a wide range of proteins (Lamb et al., 1995). It mediates protein–protein interactions and the assembly of multi-protein complexes (D'Andrea and Regan, 2003). The only *Arabidopsis* TPR-U-box protein is homologous to the mammalian carboxyl terminus of Hsc70-interacting protein (CHIP) (AtCHIP) and it is involved in abiotic stress responses and in the control of chloroplast protein turnover (Luo et al., 2006; Shen et al., 2007a, 2007b; Yan et al., 2003). In human and animals, CHIP interacts with the protein chaperones such as Hsp70 and Hsp90 and acts as chaperone or co-chaperone in the cell to ensure protein homeostasis. CHIP has been implicated in stress responses and several neurodegenerative diseases (Ballinger et al., 1999; Dickey et al., 2007; Imai et al., 2002; Jiang et al., 2001; Miller et al., 2005; Rosser et al., 2007; Sahara et al., 2005). Nevertheless, no study regarding the biological functions of rice TPR-U-box-type protein has been reported. Interestingly, we identified an expressed protein (OsPUB70) that contains both TPR and kinase domains in addition to the U-box domain. It will be intriguing to find out whether this protein is a newly evolved protein that acquired either the TPR or kinase domain relatively recently.

Another difference of domain contents in rice and *Arabidopsis* U-box proteins is that two MIF4G-type U-box proteins exist in *Arabidopsis* but the same type of protein is absent in rice. The MIF4G domain is named after the *m*iddle domain of eukaryotic *i*nitiation *f*actor *4G* (eIF4G) and it is implicated in RNA metabolic processes (Ponting, 2000). The absence of such a protein in the rice genome suggests that the arising of the corresponding genes in *Arabidopsis* occurred after speciation of the two species.

Figure 1. Identification of a Putative Conserved Domain Located Close to the C-Terminus of Rice U-Box-Leucine-Rich-Type U-Box Proteins.

The 150 amino acids from the C-terminal end of each protein were used for alignment by the program Clustal_X (Larkin et al., 2007; Thompson et al., 1997). The online tool BOXSHADE (www.ch.embnet.org/software/BOX_form.html) was used to generate the final figure using the result from the Clustal_X analysis. The high percentage of leucine residues presented in the ~100 amino acid residues located at the C-terminal half of these proteins and highly conserved glycine and lysine/arginine residues are outline in the consensus. We named this putative domain GKL domain after the conserved G and K/R residues and its leucine-rich feature. The GKL domain is denoted by the empty black box. The highly conserved glycine, lysine/arginine and leucine residues are marked as gray in the alignment and bold black in the consensus line.

Relationship between Rice U-Box/ARM Proteins

The U-box/ARM-type proteins constitute the largest family of U-box-containing proteins in both rice and *Arabidopsis* genomes. Moreover, most plant U-box-containing proteins that have been functionally characterized belong to this class (Amador et al., 2001; Gu et al., 1998; Stone et al., 2003; Yang et al., 2006; Zeng et al., 2004). This prompted us to further study the relationship of members in this class. Phylogenetic analysis using full-length protein sequences indicates that rice U-box/ARM proteins largely fall into four clusters in the phylogenetic tree. Consistently, members that are evolutionally closer in the phylogenetic map (Figure 2A) also show higher similarity in their number and arrangement of ARM repeats (Figure 2B). Unlike clusters I, II, and III, in which the ARM repeats in most members occur successively, most of the ARM repeats of cluster IV members are distributed discretely. The phylogeny of rice U-box/ARM proteins are slightly different when only the U-box sequences of each protein are used to construct the phylogenetic tree (data not shown), suggesting that the U-box domain might have evolved differently from other parts of those proteins. Similar to what have been observed in *Arabidopsis* (Mudgil et al., 2004), the domain organizations (Figure 2B) indicate that the ARM repeats of rice Ubox/ARM proteins are divergent in both number and arrangement.

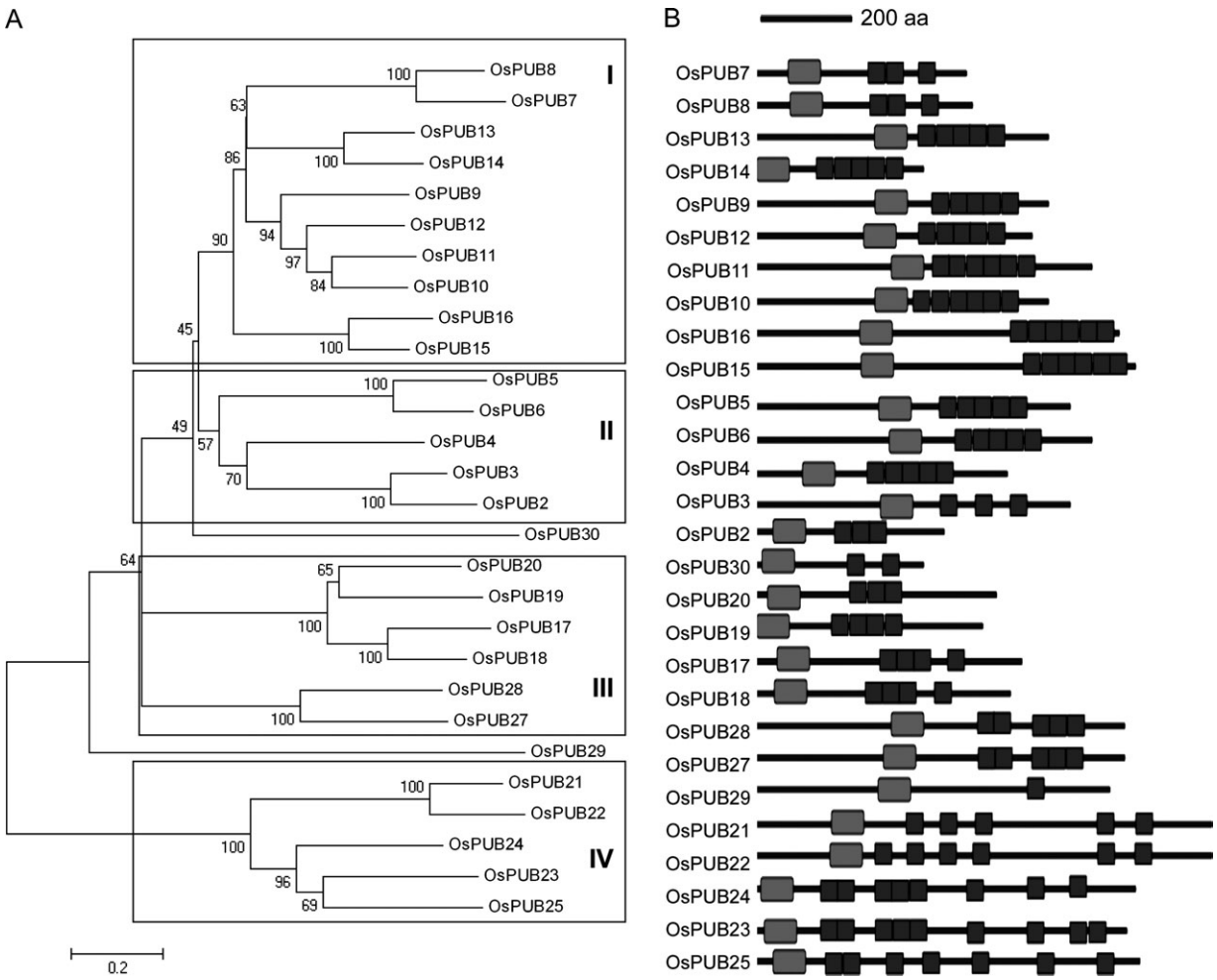

Figure 2. Phylogenies and Domain Organizations of Rice U-Box/ARM Proteins.

The phylogenies were generated in MEGA 4 (Tamura et al., 2007) and neighbor-joining with 400 bootstrap replicates and were rooted at midpoint based on full-length protein sequence analysis. The bootstrap values are shown as percentages.

(A) Phylogeny was generated using the full-length protein sequences from the rice U-box/ARM proteins. The four clusters are indicated with empty boxes.

(B) Schematic representation of domain organizations in the rice U-box/ARM proteins, ordered according to (A). The ARM repeats were identified by searching the Pfam database (Bateman et al., 2004) with cut-off E-value set at 10.0. The light-gray box indicates the U-box domain, and the individual ARM repeat of the ARM domain is indicated by a black box. The scale bar represents length of protein in amino acids. The relative position of the U-box and armadillo (ARM) repeat on the bold midnight black line indicates their position in primary structure of the protein.

Relationship between Rice and *Arabidopsis* U-Box/ARM Proteins

The identification of the corresponding set of U-box-containing proteins from both rice and *Arabidopsis* genomes allowed us to evaluate the evolutionary relationships within the U-box gene families of both species. Therefore, we performed a combined phylogenetic analysis of both rice and *Arabidopsis* U-box/ARM proteins to obtain a joint tree. As shown in Figure 3, most rice U-box/ARM proteins have their corresponding putative orthologs in *Arabidopsis*. Moreover, most rice U-box/ARM proteins show closer phylogenetic distance to their putative rice homologs than to their corresponding putative *Arabidopsis* orthologs. Nevertheless, rice proteins OsPUB4, OsPUB12, OsPUB29, and OsPUB30 show closer phylogenetic relationship to the *Arabidopsis* proteins than to their rice paralogs, suggesting that these rice proteins and their corresponding *Arabidopsis* orthologs have evolved from a common ancestor before the speciation of the two species.

In-Silico Analysis of Rice and *Arabidopsis* U-Box Protein Gene Expressions

Expression of the identified U-box protein genes was investigated by searching the rice cDNA (http://cdna01.dna.affrc.go.jp/cDNA/), EST (http://compbio.dfci.harvard.edu/tgi/ and www.ncbi.nlm.nih.gov/), and massively parallel signature sequencing (MPSS) (http://mpss.udel.edu/rice/) databases. Evidence

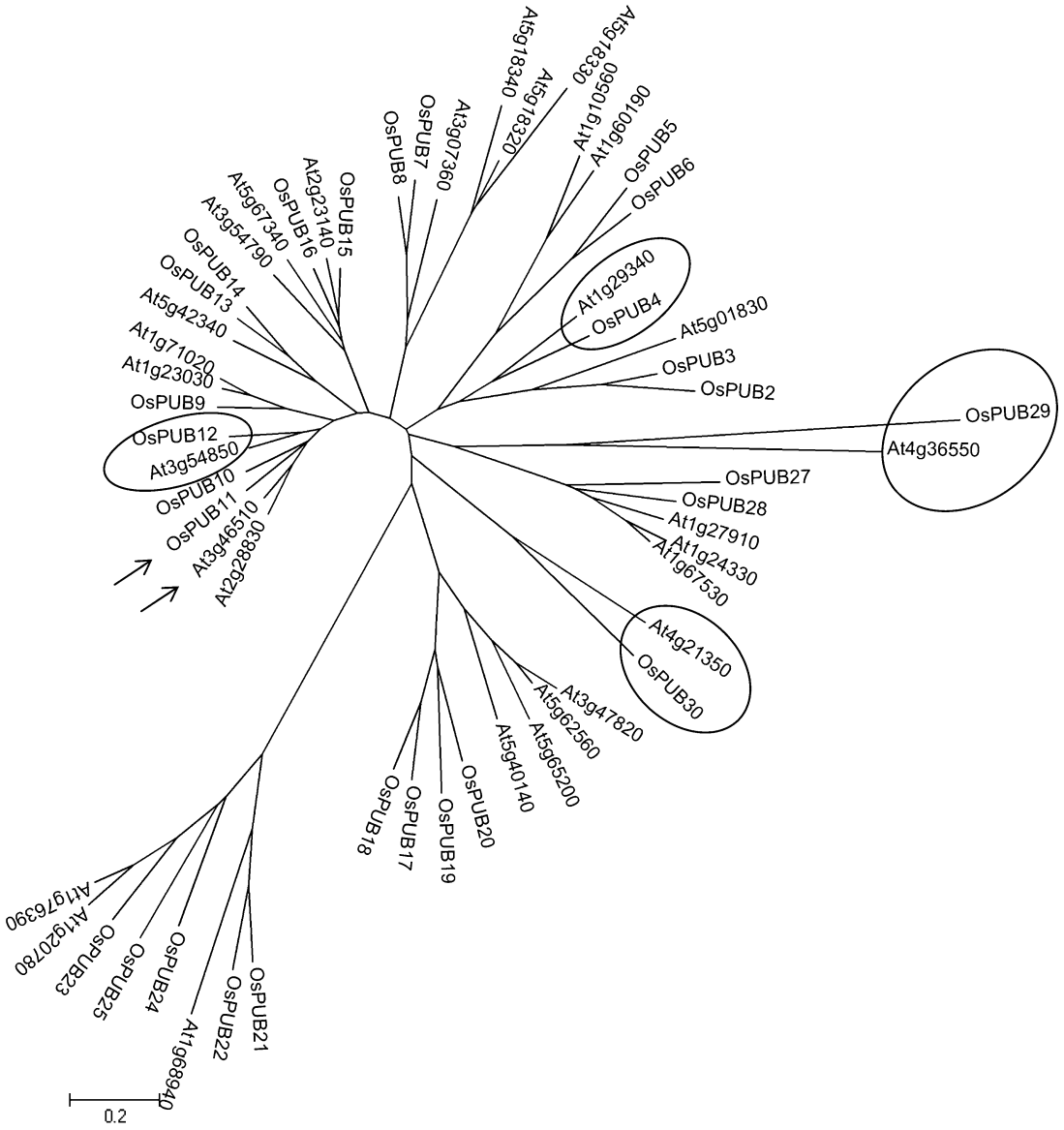

Figure 3. Phylogenetic Relationship of Rice and *Arabidopsis* U-Box/ARM Proteins.

The amino acid sequences of U-box/ARM proteins from the two genomes were used for analysis. The same method used for generating Figure 2A was used to construct the phylogeny tree. The four rice proteins that have closer phylogenetic relationship to their putative *Arabidopsis* homologs than their rice ones are marked with an oval. Rice SPL11 protein and its closest *Arabidopsis* homolog AtPUB13 are marked with black arrows.

for expression was found for the majority of rice U-box protein genes (Table 1). Corresponding ESTs or full-length cDNAs or both could be identified for 61 out of the 77 rice U-box genes, amongst which 55 have full-length cDNAs (Table 1). Based on the origin of the ESTs, the rice U-box protein genes were expressed in a variety of tissues, such as root, leaf, flower, calli, etc., suggesting that they are likely involved in a broad range of biological functions.

To have a comprehensive idea on the expression pattern of rice U-box genes, we used the same method as described previously (Nobuta et al., 2007) to analyze their MPSS tags in 61 MPSS libraries that were prepared with different rice tissues or rice leaves challenged by various abiotic or biotic stresses (see Supplemental Table 2). Of the 77 rice U-box genes, 64 of them showed significant expression in one or more libraries and only four of the rice U-box genes do not have corresponding MPSS tags presented in the libraries (see Supplemental Table 3). Some U-box genes only expressed at low level under specific conditions. For example, *OsPUB58* has low abundance of transcript only at salt-challenged young root and *OsPUB65* only express in calli and leaves challenged by incompatible fungal pathogen (see Supplemental Table 3). This may explain why no EST was identified for these genes.

By utilizing the collection of microarray data available on the Genevestigator database (Zimmermann et al., 2004), we also investigate the expression of 52 *Arabidopsis* U-box genes on the Affymetrix array chips under various experimental conditions (see Supplemental Table 4). Interestingly, lots of the *Arabidopsis* U-box genes have over two-fold change in their expression when the plant is challenged by various abiotic (such as chitin treatment or salt stress) or biotic stresses (such as *Pseudomonas syringae* infection).

Role of Rice U-Box Gene in Disease Resistance and Cell-Death Signaling

To date, a few plant U-box proteins have been implicated in defense against pathogen attack (Gonzalez-Lamothe et al., 2006; Kirsch et al., 2001; Yang et al., 2006; Zeng et al., 2004). To assess the expression of rice U-box genes during pathogen infection, we chose nine rice U-box genes that showed differential expression in our microarray hybridizations (C.H. Park and G.L. Wang, unpublished) and examined their expression in rice leaf tissue challenged by the rice fungal pathogen *Magnaporthe oryzae*. As shown in Figure 4, seven of them showed differential expression patterns after inoculation with the fungal isolate PO6-6. Specifically, OsPUB4, 12, and 23 had a similar expression in both resistant and susceptible plants. OsPUB 51, 64, and 73 showed stronger expression in the susceptible plants. Nevertheless, OsPUB57 showed a stronger expression only in the resistant plants carrying the *Pi9*-resistant gene (Qu et al., 2006). These results suggest some of the genes might play a role in the signaling or regulation of rice defense against *M. grisea*.

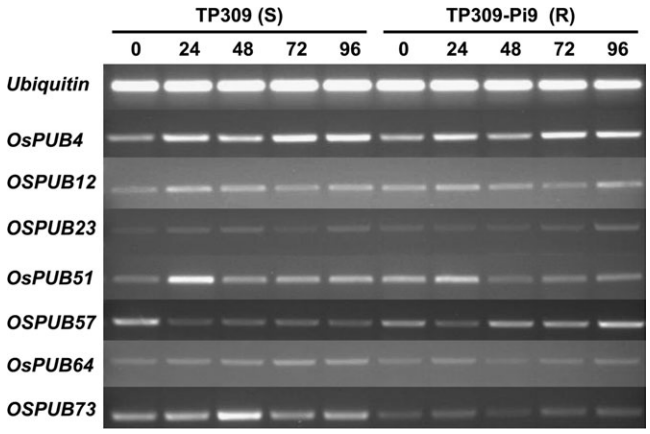

Figure 4. RT–PCR Analysis of Rice U-Box Protein Genes.

Three-week-old seedlings of susceptible variety TP309 and resistant transgenic line harboring the resistance gene *Pi9* (Qu et al., 2006) were inoculated with *M. oryzae* isolate PO6-6. RNA samples from infected plant leaves at indicated time points were used for RT–PCR analysis using gene-specific primers designed to amplify the corresponding U-box protein genes. PCR reaction for each gene was performed using 25 cycles of amplification, and the PCR products were separated on 1.4% agarose gel. The amplification of the ubiquitin gene was used as control for an equal amount of first-strand cDNA being used for each PCR reaction. This experiment was repeated for twice, with similar results.

To further confirm their function in defense responses, a rice protoplast system was used (Chen et al., 2006). The rice protoplasts were transiently co-transformed with a U-box gene-silencing construct and a beta-glucuronidase (GUS) gene expression construct. The strength of GUS activity was monitored as an indicator of viability of the protoplasts. Among the seven genes tested, OsPUB57 causes cell death when it is knocked down in the cell, which suggests that it might act as a negative regulator in rice cell death signaling (Figure 5).

E3 Ubiquitin Ligase Activities of Rice U-Box Proteins

The U-box-containing proteins are predicted to possess E3 ubiquitin ligase activity. To determine if rice U-box proteins do encode functional E3 ligases, in-vitro ubiquitination assays were performed. Five rice U-box proteins (OsPUB4, OsPUB12, OsPUB57, OsPUB73, and OsPUB77) were randomly picked for test. These proteins were expressed with GST-tags in *Escherichia coli* and tested for E3 ligase activity using an in-vitro ubiquitination assay as described (Zeng et al., 2004). The self-ubiquitination of the U-box proteins presented in the reaction, as shown by a ladder of higher molecular bands, was used as an indicator of E3 liagse activity (Hatakeyama et al., 2001).

As shown in Figure 6, the five rice U-box proteins were found to possess E3 ligase activity. Importantly, like the observation with mammalian U-box proteins (Hatakeyama et al., 2001), the ligase activity of the rice U-box proteins also depends on the E2 enzyme used for the assay. The preference for a particular E2 enzyme did not show any correlation with the presence of a particular domain in the rice U-box proteins. In the control experiment, with the omission of either E1, E2, or ubiquitin, no ubiquitination was observed (data not shown). Therefore, the ubiquitination observed was due to the E2-dependent E3 ligase activity of the rice U-box proteins.

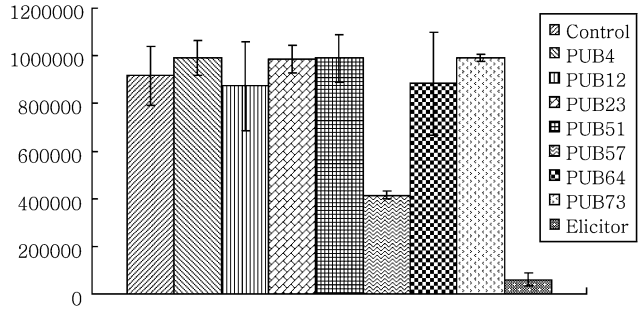

Figure 5. Analysis of the Role of Rice U-Box Genes in Cell Death Signaling Using Rice Protoplast System.

The candidate rice U-box gene was transiently knocked down in rice protoplasts using the RNAi method. The beta-glucuronidase (GUS) gene under the control of maize ubiquitin promoter was co-transiently expressed in rice protoplasts. The strength of GUS activity was monitored as an indicator of viability of the protoplasts. Protoplasts treated with fungal cell wall elicitor isolated from *Magnaporthe oryzae* known to cause rice cell death was used as the positive control (G.L. Wang, unpublished). Cells maintain high GUS activity denoting a low rate of cell death.

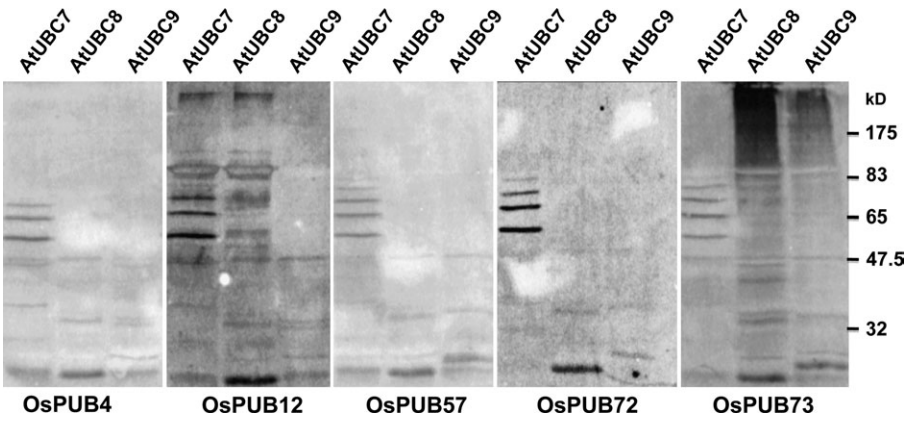

Figure 6. Rice U-Box Proteins Possess E3 Ligase Activity.

Affinity-purified recombinant rice U-box proteins were tested for E3 ligase activity in an in-vitro ubiquitination assay. The wheat E1 enzyme, E2 enzymes from *Arabidopsis* (*At*UBC), and ubiquitin were used in the reactions. The E2 enzyme used in each reaction is indicated above each lane, and the rice protein tested is listed below each panel.

DISCUSSION

The distribution of U-box proteins among species of different kingdoms is uneven. There are a larger number of genes encoding such proteins in the plant genomes. Three U-box genes were identified among the 6300 annotated genes in the yeast genome. Similarly, fewer than 20 U-box proteins were predicted in the human genome (Patterson, 2002). Only two putative U-box proteins were identified in the rice fungal pathogen *Magnaporthe oryzae*'s genome (L. Zeng, unpublished). In contrast, 61 U-box proteins were identified in *Arabidopsis* when a series of analyses were performed (Azevedo et al., 2001) (www.*Arabidopsis*.org/info/genefamily/pub.html). The U-box domain is structurally analogous to the RING domain. The large number of RING-type proteins constitutes a super-family in plants (Stone et al., 2005). The high similarity between U-box and other sub-family members of RING-type proteins makes it a daunting challenge to correctly identify U-box sequences from the RING-type proteins pool. In this study, 77 U-box-containing proteins were identified from the annotated rice genome by a six-step prodedure that involves a series of algorithms. Of the 77 U-box protein genes, either cDNA or EST were detected for 61 of them and only four of them do not have matching MPSS tags, corroborating the effectiveness of the algorithms used in the survey of this study. Sixty-one U-box proteins were identified from *Arabidopsis* when data from this study and other resources (www.*Arabidopsis*.org/info/genefamily/pub.html) were combined. The identification of two nearly complete sets of U-box-containing proteins from the genomes of the model monocot and dicot plants provided the basis for studying the phylogeny of U-box proteins in plants.

Sequence alignment indicates that there exists a class of U-box proteins with highly conserved sequences at their C-termini in both rice and *Arabidopsis* genomes (Figure 1 and supplementary Figure 2). We named this region as GKL domain, based on the presence of highly conserved Glycine, Lysine/Arginine residues and leucine-rich features. To date, however, no biological function has been associated with such a domain. It is generally considered that the U-box domain serves as the E2 recruitment domain, while another distinct domain in U-box proteins, such as ARM repeats or TPR domain, serve to recognize and recruit the substrate. However, no other known domain/motif was identified in these proteins. Therefore, we propose that the GKL domain might be involved in protein–protein interaction. Recently, one tobacco homolog of U-box/GKL-type protein, NtCMPG1, was implicated in plant immunity (Gonzalez-Lamothe et al., 2006). Identification of the substrate for NtCMPG1 and testing whether mutation in the GKL domain will compromise the interaction of NtCMPG1 to the substrate will help to verify our hypothesis.

The ARM repeat was first identified in the *Drosophila melanogaster* segment polarity gene armadillo that is involved in signal transduction (Riggleman et al., 1989). Structural characteristics of the ARM motif suggest its involvement in protein–protein interactions, which has been demonstrated in several cases (Huber et al., 1997). In a few cases, HEAT repeats were detected in proximity to the ARM repeats. Phylogenetic analysis indicated most rice U-box proteins are evolutionally closer to their homologs from the rice genome than those from the *Arabidopsis* genome (Figure 3). The establishment of such a relationship between rice and *Arabidopsis* U-box proteins could help in characterizing these genes. For example, the identification of the putative *Arabidopsis* ortholog of rice *Spl11* gene (*AtSpl11*) based on the combined phylogenetic tree led us to the identification of a loss-of-function mutant of *AtSpl11* that displayed similar cell death and altered flowering phenotype to the rice *spl11* mutant (L. Zeng and G.L. Wang, unpublished). Given the advantage of *Arabidopsis* over rice in terms of plant growth and genetic resources, we expect that an in-depth characterization of the *Atsp11* mutant will help to elucidate the connection of immunity and reproductive

development in rice. Similarly, it would be interesting to investigate whether the putative rice ortholog of AtPUB17 (ACRE276), OsPUB4, has similar functions in rice cell death and defense (Yang et al., 2006).

Why plants require more U-box proteins is unclear at present. One speculation, given the lack of animal-like immune system and the static life of plants, is that U-box proteins may play a role in plants' responding to various environmental stresses (Patterson, 2002). Domain organization analysis indicates that the U-box family could be divided into at least nine classes. A variety of other motifs/domains, such as protein–protein interaction domains, protein phosphorylation-related domain, and transcription initiation-related domain, were found to be associated with the U-box domain in rice and *Arabidopsis* U-box proteins (Table 2). U-box proteins with similar domain organization to other organisms have been demonstrated to function in different cellular processes (Amador et al., 2001; Kirsch et al., 2001; Stone et al., 2003). These data suggest that the biological functions of U-box proteins might be diverse rather than limited to stress response signaling. Of the rice and *Arabidopsis* U-box proteins identified hereby, only a few of them have been functionally characterized. In addition, the existence of multiple members in each class of U-box proteins raised the possibility of function redundancy among the members. Such function redundancy could make the study of biological roles for plant U-box genes a daunting challenge. Among the seven rice U-box genes analyzed for their role in cell death signaling, only one was found to cause cell death when it is silenced in rice cells (Figure 5). The failure of the other six genes to display cell death phenotype could be due to such function redundancy as well. Therefore, a systemic approach, using knockout mutants or transgenic RNAi lines, to modify the expression of most or all members in the same class might be essential to functionally characterize these genes on a large scale.

U-box-containing proteins are predicted to possess E3 ubiquitin ligase activity. In *Arabidopsis*, several U-box proteins were revealed to possess E3 ligase activity (Mudgil et al., 2004; Yan et al., 2003; Yang et al., 2006). Of the 77 rice U-box proteins, only SPL11 has been demonstrated to possess E3 activity in previous studies. In this study, we tested five randomly chosen rice U-box proteins for their E3 ligase activity in vitro. Such ubiquitination assay indicates that all the five proteins have E3 ligase activity and they demonstrate different preferences for E2 enzymes in these assays (Figure 6). Unlike the E3s in the HECT family, RING finger and U-box E3s facilitate the transfer of ubiquitin molecule by precise spatial orientation of the E2 and the substrate (Wu et al., 2003). The U-box or RING-finger domain serves as the E2 recruitment domain, while another distinct domain, such as ARM repeats or TPR domain, serve to recognize and recruit the substrate. The crystal structure of the U-box protein Prp19 indicated that, for those U-box E3s that oligmerize to function, the U-box and the substrate-recruiting domain are arranged in such an architecture that it brings the E2 recruitment domain into close proximity with a substrate recognition domain. Therefore, such E2 preference may reflect the minor difference in the U-box domains of these proteins. Nevertheless, such preference could make it difficult to test their in-vitro E3 ligase activity for some U-box proteins due to the availability of the best E2s they required for the assay.

Although the majority of the genes that encode the 77 identified rice U-box proteins are expressed, only the function of *Spl11* has been reported (Zeng et al., 2006). The diverse motifs/domains presented in the rice U-box proteins support the notion that the rice U-box family may target a myriad of substrates. Biochemical assay of the ligase activity of the U-box proteins, identification of their substrates, and elucidation of the mechanisms under which they work in the cell will be the next study in order to fully understand the biological roles of these genes.

METHODS

Identification of U-Box-Containing Proteins in the Rice Genome

The U-box proteins in the genomes of rice *japonica* and *indica* subspecies were identified using a battery of algorithms. The complete results are available as supplementary materials at the following address: http://supfam.org/ubox/ubox.html. Taxonomically, we place the U-box family as a member of the RING domain superfamily. The U-box family shows some distant sequence homology to other member families under the RING domain superfamily. To identify them correctly, the U-box proteins must be distinguished from other members in the same superfamily. A crucial difference between the U-box proteins and other superfamily members is the presence of fewer cysteine/histidine residues. The following seven procedures were used in the identification process: (1) WU-BLAST of known U-box proteins against the genome, (2) searching PFAM models of non-U-box families within the RING domain superfamily (Bateman et al., 2004), (3) scoring a SAM hidden Markov model (HMM) built from the PFAM seed alignment of U-box domains (Sonnhammer et al., 1998), (4) a SAM T99 'family'-level model was built for U-box proteins (Karplus and Hu, 2001), (5) scoring an HMM built from the PFAM 'full' alignment, (6) a SAM T99 'superfamily'-level model was built and scored. The final results of the above seven sequential analyses were further examined with respect to the number of aligned cysteine residues and multiple sequence alignments annotated with the structural key residues highlighting the differences between U-box family members and other RING domain superfamily members. Additional evidence such as structural information was consulted in some specific cases.

Phylogenetic Analysis

Rice and *Arabidopsis* U-box proteins were used as queries to search against the PFAM or GenBank database and were classified into different classes based on their domain contents.

Rice and *Arabidopsis* proteins from the ARM class were then aligned using the clustal_X program (Thompson et al., 1997). The aligned sequence data were then input into the MEGA4 program (Kumar et al., 2001; Tamura et al., 2007) and a neighbor-joining algorithm was used to build the phylogenetic tree.

Preparation of MPSS Libraries and Data Analysis

The growth of rice plants, collection of rice tissues, preparation of MPSS libraries, and data mining are as described (Nobuta et al., 2007).

Plant Materials and Rice Blast Inoculation

Rice seeds were first sterilized by treatment in 75% ethanol for 1 min, followed by immersion in 2% sodium hypochlorite for 15–20 min. After several washes with sterile distilled water, seeds were germinated in half-MS medium for 7–10 d, and then transferred to soil. Growth chamber conditions were 12 h light, 26°C, 80% relative humidity (RH), followed by 12 h dark, 20°C, 60% RH, unless otherwise specified. Three-week-old plants of rice *japonica* variety TP309 and its isogenic transgenic lines containing the Pi9 resistance gene were inoculated with *Magnaporthe oryzae* isolate PO6-6 at spore suspension of rice blast at 1×10^5 spores ml^{-1}. The inoculated plants were placed in a the dark for 24 h at 26°C, and leaf tissues were collected every 24 h after inoculation until 96 h.

Cell Death Analysis of the Candidate Genes in Rice Protoplasts

RNAi constructs containing the candidate gene fragments and the construct containing the beta-glucuronidase (GUS) gene under the control of maize ubiquitin promoter were transiently expressed in rice protoplasts as described (Chen et al., 2006). The assay of GUS activity was carried out according to the protocol described by Chen et al. (2006).

In-Vitro Ubiquination Assay

The in-vitro ubiquitination assays were performed as described (Zeng et al., 2004) with some modifications. In brief, full-length OsPUB proteins were expressed as N-terminal GST-tagged in BL21 (DE3) cells (Stratagene) and affinity-purified with glutathione matrix (Sigma, USA). The in-vitro ubiquitination reactions were performed by adding 1 μg of protein of interest, 40 ng of wheat E1, 100 ng of *Arabidopsis* E2, 10 μg of ubiqutin and 1.5 μL of 20X reaction buffer (1 M Tris HCl, pH 7.5, 40 mM ATP, 100 mM $MgCl_2$, 40 mM DTT, 600 mM creatine phosephate and 1 mg ml^{-1} creatine phosphokinase). The reaction was incubated at 30°C for 1.5 h in 30 μl volume before being stopped with SDS sample loading buffer and heated to 100°C for 5 min. Samples of the reactions were then separated by 10% SDS–PAGE. Ubiquitinated substrates were detected by Western blotting and detected by ubiquitin antibody (Biomol, USA) followed by detection by chemiluminescence with the ECL kit.

SUPPLEMENTARY DATA

Supplementary Data are available at *Molecular Plant Online*.

FUNDING

This project is supported by the National Research Initiative (NRI) Plant Genome Program of the USDA Cooperative State Research, Education and Extension Service (CSREES) to G.L.W. (#: 2004–35604–14226 and 2007–35301–18093). No conflict of interest declared.

REFERENCES

Amador, V., Monte, E., Garcia-Martinez, J.L., and Prat, S. (2001). Gibberellins signal nuclear import of PHOR1, a photoperiod-responsive protein with homology to Drosophila armadillo. Cell. **106,** 343–354.

Andrade, M.A., and Bork, P. (1995). HEAT repeats in the Huntington's disease protein. Nat. Genet. **11,** 115–116.

Andrade, M.A., Perez-Iratxeta, C., and Ponting, C.P. (2001). Protein repeats: structures, functions, and evolution. J. Struct. Biol. **134,** 117–131.

Aravind, L., and Koonin, E.V. (2000). The U box is a modified RING finger: a common domain in ubiquitination. Curr. Biol. **10,** R132–R134.

Azevedo, C., Santos-Rosa, M.J., and Shirasu, K. (2001). The U-box protein family in plants. Trends Plant Sci. **6,** 354–358.

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J.W., and Elledge, S.J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell. **86,** 263–274.

Ballinger, C.A., Connell, P., Wu, Y., Hu, Z., Thompson, L.J., Yin, L.Y., and Patterson, C. (1999). Identification of CHIP, a novel tetratricopeptide repeat-containing protein that interacts with heat shock proteins and negatively regulates chaperone functions. Mol. Cell. Biol. **19,** 4535–4545.

Bateman, A., et al. (2004). The Pfam protein families database. Nucleic Acids Res. **32 Database Issue,** D138–D141.

Borden, K.L. (2000). RING domains: master builders of molecular scaffolds? J. Mol. Biol. **295,** 1103–1112.

Bray, S., Musisi, H., and Bienz, M. (2005). Bre1 is required for Notch signaling and histone modification. Dev. Cell. **8,** 279–286.

Chen, S., Tao, L., Zeng, L., Vega-Sanchez, M.E., Umemura, K., and Wang, G.-L. (2006). A highly efficient transient protoplast system for analyzing defence gene expression and protein-protein interactions in rice. Mol. Plant Pathol. **7,** 417–427.

Ciechanover, A. (1998). The ubiquitin–proteasome pathway: on protein death and cell life. EMBO J. **17,** 7151–7160.

Ciechanover, A. (2005). Intracellular protein degradation: from a vague idea thru the lysosome and the ubiquitin–proteasome system and onto human diseases and drug targeting. Cell Death Differ. **12,** 1178–1190.

Ciechanover, A., and Schwartz, A.L. (1998). The ubiquitin–proteasome pathway: the complexity and myriad functions of proteins death. Proc. Natl Acad. Sci. U S A. **95,** 2727–2730.

D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. Trends Biochem. Sci. **28,** 655–662.

Das, A.K., Cohen, P.W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein–protein interactions. EMBO J. **17,** 1192–1199.

Dickey, C.A., et al. (2007). The high-affinity HSP90–CHIP complex recognizes and selectively degrades phosphorylated tau client proteins. J. Clin. Invest. **117,** 648–658.

Feng, Q., et al. (2002). Sequence and analysis of rice chromosome 4. Nature. **420,** 316–320.

Freemont, P.S., Hanson, I.M., and Trowsdale, J. (1991). A novel cysteine-rich sequence motif. Cell. **64,** 483–484.

Goff, S.A., et al. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science. **296,** 92–100.

Gonzalez-Lamothe, R., Tsitsigiannis, D.I., Ludwig, A.A., Panicot, M., Shirasu, K., and Jones, J.D. (2006). The U-box protein CMPG1 is required for efficient activation of defense mechanisms triggered by multiple resistance genes in tobacco and tomato. Plant Cell. **18,** 1067–1083.

Gu, T., Mazzurco, M., Sulaman, W., Matias, D.D., and Goring, D.R. (1998). Binding of an arm repeat protein to the kinase domain of the S-locus receptor kinase. Proc. Natl Acad. Sci. U S A. **95,** 382–387.

Han, B., and Xue, Y. (2003). Genome-wide intraspecific DNA-sequence variations in rice. Curr. Opin. Plant Biol. **6,** 134–138.

Hatakeyama, S., Yada, M., Matsumoto, M., Ishida, N., and Nakayama, K.I. (2001). U box proteins as a new family of ubiquitin-protein ligases. J. Biol. Chem. **276,** 33111–33120.

Huang, F., Kirkpatrick, D., Jiang, X., Gygi, S., and Sorkin, A. (2006). Differential regulation of EGF receptor internalization and degradation by multiubiquitination within the kinase domain. Mol. Cell. **21,** 737–748.

Huber, A.H., and Weis, W.I. (2001). The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. Cell. **105,** 391–402.

Huber, A.H., Nelson, W.J., and Weis, W.I. (1997). Three-dimensional structure of the armadillo repeat region of beta-catenin. Cell. **90,** 871–882.

Huibregtse, J.M., Scheffner, M., Beaudenon, S., and Howley, P.M. (1995). A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. Proc. Natl Acad. Sci. U S A. **92,** 5249.

Imai, Y., Soda, M., Hatakeyama, S., Akagi, T., Hashikawa, T., Nakayama, K.I., and Takahashi, R. (2002). CHIP is associated with *Parkin*, a gene responsible for familial Parkinson's disease, and enhances its ubiquitin ligase activity. Mol. Cell. **10,** 55–67.

Irniger, S., Piatti, S., Michaelis, C., and Nasmyth, K. (1995). Genes involved in sister chromatid separation are needed for B-type cyclin proteolysis in budding yeast. Cell. **81,** 269–278.

Jiang, J., Ballinger, C.A., Wu, Y., Dai, Q., Cyr, D.M., Hohfeld, J., and Patterson, C. (2001). CHIP is a U-box-dependent E3 ubiquitin ligase: identification of Hsc70 as a target for ubiquitylation. J. Biol. Chem. **276,** 42938–42944.

Karplus, K., and Hu, B. (2001). Evaluation of protein multiple alignments by SAM-T99 using the BAliBASE multiple alignment test set. Bioinformatics. **17,** 713–720.

King, R.W., Peters, J.M., Tugendreich, S., Rolfe, M., Hieter, P., and Kirschner, M.W. (1995). A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. Cell. **81,** 279–288.

Kirsch, C., Logemann, E., Lippok, B., Schmelzer, E., and Hahlbrock, K. (2001). A highly specific pathogen-responsive promoter element from the immediate-early activated *CMPG1* gene in Petroselinum crispum. Plant J. **26,** 217–227.

Koegl, M., Hoppe, T., Schlenker, S., Ulrich, H.D., Mayer, T.U., and Jentsch, S. (1999). A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. Cell. **96,** 635–644.

Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. Bioinformatics. **17,** 1244–1245.

Lamb, J.R., Tugendreich, S., and Hieter, P. (1995). Tetratrico peptide repeat interactions: to TPR or not to TPR? Trends Biochem. Sci. **20,** 257–259.

Larkin, M.A., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics. **23,** 2947–2948.

Li, D., and Roberts, R. (2001). WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cell. Mol. Life Sci. **58,** 2085–2097.

Li, M., Brooks, C.L., Wu-Baer, F., Chen, D., Baer, R., and Gu, W. (2003). Mono- versus polyubiquitination: differential control of p53 fate by Mdm2. Science. **302,** 1972–1975.

Lorick, K.L., Jensen, J.P., Fang, S., Ong, A.M., Hatakeyama, S., and Weissman, A.M. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. Proc. Natl Acad. Sci. U S A. **96,** 11364–11369.

Loscher, M., et al. (2005). Interaction of U-box E3 ligase SNEV with PSMB4, the beta7 subunit of the 20 S proteasome. Biochem. J. **388,** 593–603.

Luo, J., Shen, G., Yan, J., He, C., and Zhang, H. (2006). AtCHIP functions as an E3 ubiquitin ligase of protein phosphatase 2A subunits and alters plant response to abscisic acid treatment. Plant J. **46,** 649–657.

Mahajan, K.N., and Mitchell, B.S. (2003). Role of human Pso4 in mammalian DNA repair and association with terminal deoxynucleotidyl transferase. Proc. Natl Acad. Sci. U S A. **100,** 10746–10751.

Meacham, G.C., Patterson, C., Zhang, W., Younger, J.M., and Cyr, D.M. (2001). The Hsc70 co-chaperone CHIP targets immature CFTR for proteasomal degradation. Nat. Cell Biol. **3,** 100–105.

Miller, V.M., Nelson, R.F., Gouvion, C.M., Williams, A., Rodriguez-Lebron, E., Harper, S.Q., Davidson, B.L., Rebagliati, M.R., and Paulson, H.L. (2005). CHIP suppresses polyglutamine aggregation and toxicity *in vitro* and *in vivo*. J. Neurosci. **25,** 9152–9161.

Mudgil, Y., Shiu, S.-H., Stone, S.L., Salt, J.N., and Goring, D.R. (2004). A large complement of the predicted *Arabidopsis* ARM repeat proteins are members of the U-box E3 ubiquitin ligase family. Plant Physiol. **134,** 59–66.

Nobuta, K., et al. (2007). An expression atlas of rice mRNAs and small RNAs. Nat. Biotechnol. **25,** 473–477.

Ohi, M.D., and Gould, K.L. (2002). Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. RNA. **8,** 798–815.

Ohi, M.D., Vander Kooi, C.W., Rosenberg, J.A., Chazin, W.J., and Gould, K.L. (2003). Structural insights into the U-box, a domain associated with multi-ubiquitination. Nat. Struct. Biol. **10,** 250–255.

Patterson, C. (2002). A new gun in town: the U box is a ubiquitin ligase domain. Sci STKE. **2002,** PE4.

Pintard, L., et al. (2003). The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase. Nature. **425,** 311–316.

Ponting, C.P. (2000). Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. Trends Biochem. Sci. **25,** 423–426.

Qu, S.H., Liu, G.F., Zhou, B., Bellizzi, M., Zeng, L.R., Dai, L.Y., Han, H., and Wang, G.L. (2006). The broad-spectrum blast resistance gene *Pi9* encodes an NBS-LRR protein and is a member of a multigene family in rice. Genetics. **172,** 1901–1914.

Riggleman, B., Wieschaus, E., and Schedl, P. (1989). Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene. Genes Dev. **3,** 96–113.

Rosser, M.F., Washburn, E., Muchowski, P.J., Patterson, C., and Cyr, D.M. (2007). Chaperone functions of the E3 ubiquitin ligase CHIP. J. Biol. Chem. **282,** 22267–22277.

Sahara, N., Murayama, M., Mizoroki, T., Urushitani, M., Imai, Y., Takahashi, R., Murata, S., Tanaka, K., and Takashima, A. (2005). *In vivo* evidence of CHIP up-regulation attenuating tau aggregation. J. Neurochem. **94,** 1254–1263.

Scheffner, M., Huibregtse, J.M., Vierstra, R.D., and Howley, P.M. (1993). The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell. **75,** 495–505.

Shen, G., Adam, Z., and Zhang, H. (2007a). The E3 ligase AtCHIP ubiquitylates FtsH1, a component of the chloroplast FtsH protease, and affects protein degradation in chloroplasts. Plant J. **52,** 309–321.

Shen, G., Yan, J., Pasapula, V., Luo, J., He, C., Clarke, A.K., and Zhang, H. (2007b). The chloroplast protease subunit ClpP4 is a substrate of the E3 ligase AtCHIP and plays an important role in chloroplast function. Plant J. **49,** 228–237.

Skowyra, D., Craig, K.L., Tyers, M., Elledge, S.J., and Harper, J.W. (1997). F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. Cell. **91,** 209–219.

Smalle, J., and Vierstra, R.D. (2004). The ubiquitin 26s proteasome proteolytic pathway. Annu. Rev. Plant Biol. **55,** 555–590.

Sonnhammer, E.L., Eddy, S.R., Birney, E., Bateman, A., and Durbin, R. (1998). Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res. **26,** 320–322.

Stebbins, C.E., Kaelin, W.G., Jr, and Pavletich, N.P. (1999). Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. Science. **284,** 455–461.

Stone, S.L., Anderson, E.M., Mullen, R.T., and Goring, D.R. (2003). ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible Brassica pollen. Plant Cell. **15,** 885–898.

Stone, S.L., Hauksdóttir, H., Troy, A., Herschleb, J., Kraft, E., and Callis, J. (2005). Functional analysis of the RING-type ubiquitin ligase family of *Arabidopsis*. Plant Physiol. **137,** 13–30.

Sudakin, V., Ganoth, D., Dahan, A., Heller, H., Hershko, J., Luca, F.C., Ruderman, J.V., and Hershko, A. (1995). The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. Mol. Biol. Cell. **6,** 185–197.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24,** 1596–1599.

Terrell, J., Shih, S., Dunn, R., and Hicke, L. (1998). A function for monoubiquitination in the internalization of a G protein-coupled receptor. Mol. Cell. **1,** 193–202.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25,** 4876–4882.

Thrower, J.S., Hoffman, L., Rechsteiner, M., and Pickart, C.M. (2000). Recognition of the polyubiquitin proteolytic signal. EMBO J. **19,** 94–102.

Voglauer, R., Chang, M.W., Dampier, B., Wieser, M., Baumann, K., Sterovsky, T., Schreiber, M., Katinger, H., and Grillari, J. (2006). SNEV overexpression extends the life span of human endothelial cells. Exp. Cell Res. **312,** 746–759.

Wu, G., Xu, G., Schulman, B.A., Jeffrey, P.D., Harper, J.W., and Pavletich, N.P. (2003). Structure of a beta-TrCP1-Skp1-beta-catenin complex: destruction motif binding and lysine specificity of the SCF(beta-TrCP1) ubiquitin ligase. Mol. Cell. **11,** 1445–1456.

Yan, J., Wang, J., Li, Q., Hwang, J.R., Patterson, C., and Zhang, H. (2003). AtCHIP, a U-box-containing E3 ubiquitin ligase, plays a critical role in temperature stress tolerance in *Arabidopsis*. Plant Physiol. **132,** 861–869.

Yang, C.W., Gonzalez-Lamothe, R., Ewan, R.A., Rowland, O., Yoshioka, H., Shenton, M., Ye, H., O'Donnell, E., Jones, J.D., and Sadanandom, A. (2006). The E3 ubiquitin ligase activity of *Arabidopsis* PLANT U-BOX17 and its functional tobacco homolog ACRE276 are required for cell death and defense. Plant Cell. **18,** 1084–1098.

Yu, J., et al. (2002). A draft sequence of the rice genome (*Oryza sativa L*. ssp. *indica*). Science. **296,** 79–92.

Zeng, L.-R., Qu, S., Bordeos, A., Yang, C., Baraoidan, M., Yan, H., Xie, Q., Nahm, B.H., Leung, H., and Wang, G.-L. (2004). *Spotted leaf11*, a negative regulator of plant cell death and defense, encodes a U-box/armadillo repeat protein endowed with E3 ubiquitin ligase activity. Plant Cell. **16,** 2795–2808.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L., and Gruissem, W. (2004). GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. Plant Physiol. **136,** 2621–2632.