*Journal of Integrative Plant Biology* 2007, **49** (6): 940–952

# Genome-wide Analysis of Kelch Repeat-containing F-box Family

**Yujin Sun[1], Xiaofan Zhou[2] and Hong Ma[1, 2*]**

([1]*Department of Biology and the Huck Institutes of Life Sciences*;
[2]*Intercollege Graduate Program in Cell and Developmental Biology, The Pennsylvania State University,*
University Park, PA 16802, USA)

## Abstract

**The ubiquitin-dependent protein degradation pathway plays diverse roles in eukaryotes. Previous studies indicate that both F-box and Kelch motifs are common in a variety of organisms. F-box proteins are subunits of E3 ubiquitin ligase complexes called SCFs (SKP1, Cullin1, F-box protein, and Rbx1); they have an N-terminal F-box motif that binds to SKP1 (S-phase kinase associated protein), and often have C-terminal protein-protein interaction domains, which specify the protein substrates for degradation via the ubiquitin pathway. One of the most frequently found protein interaction domains in F-box proteins is the Kelch repeat domain. Although both the F-box and Kelch repeats are ancient motifs, Kelch repeats-containing F-box proteins (KFB) have only been reported for human and *Arabidopsis* previously. The recent sequencing of the rice genome and other plant genomes provides an opportunity to examine the possible evolution history of KFB. We carried out extensive BLAST searches to identify putative KFBs in selected organisms, and analyzed their relationships phylogenetically. We also carried out the analysis of both gene duplication and gene expression of the KFBs in rice and *Arabidopsis*. Our study indicates that the origin of KFBs occurs before the divergence of animals and plants, and plant KFBs underwent rapid gene duplications.**

**Key words:** *Arabidopsis thaliana*; F-box motif; Kelch repeats; Kelch repeats-containing F-box proteins; SCF; *SKP1* like.

**Sun Y, Zhou X, Ma H** (2007). Genome-wide analysis of Kelch repeat-containing F-box family. *J. Integr. Plant Biol.* **49**(6), 940–952.
Available online at www.blackwell-synergy.com/links/toc/jipb, www.jipb.net

Selective protein degradation through the ubiquitin-dependent pathway plays essential roles in cell cycle progression, transcriptional regulation, and signal transduction (Hershko and Ciechanover 1998). The ubiquitin-activating enzyme (E1) and the ubiqutin-conjugating enzyme (E2) function with the ubiquitin ligase (E3), which specifies the protein target(s), to facilitate the degradation of the ubiqitinated substrates by the 26S proteasome (Koepp et al. 2001; Pickart 2001). In the past 10 years, studies in human and yeast have uncovered a class of cullin based ubiquitin ligases (E3) (Lyapina et al. 1998; Kobayashi et al. 2004; Willems et al. 2004; Hong et al. 2005). One of the largest and best characterized families of cullin-based ubiquitin ligases is the SCF complex, which consist of SKP1 (S-phase kinase associated protein), Cullin1/Cdc53, Rbx1, and an F-box protein. Several studies have found that Cullin1/Cdc53 interacts with SKP1 and Rbx1 through its long N-terminal stalk domain and the C-terminal globular domain, respectively (Krek 1998; Skowyra et al. 1999; Zheng et al. 2002). Rbx1 contains a ring finger domain that interacts with the E2 enzyme, while SKP1 bridges Cullin1 and an F-box protein (Kamura et al. 1999). F-box proteins have a relative conserved F-box domain near the N-terminus interacting with SKP1 and a less conserved protein-protein interaction domain at the C-terminus specifying the ubiquitylational target(s) (del Pozo and Estelle 2000).

Among the SCF subunits, Cullin1 and Rbx1 are highly

Received 9 Jan. 2007 Accepted 13 Mar. 2007
Supported by the Grant from US Department of Energy (DE-FG02-02ER15332).
Publication of this paper is supported by the National Natural Science Foundation of China (30624808).

*Author for correspondence.
Tel: +1 814 863 6414;
Fax: +1 814 863 1357;
E-mail: <hxm16@psu.edu>.

doi: 10.1111/j.1672-9072.2007.00498.x

conserved in diverse organisms and are present at low copy numbers, whereas the numbers of SKP1 homologs range from one in fungi and vertebrates to more than 20 in plants and invertebrates (Gagne et al. 2002; Risseeuw et al. 2003; Kong et al. 2004). The numbers of putative F-box proteins are even greater, particularly in plants; *Arabidopsis* and human have approximately 700 and 68 predicted F-box genes, respectively (Gagne et al. 2002; Kuroda et al. 2002; Jin et al. 2004). Genetic studies have uncovered the functions of several F-box proteins in *Arabidopsis*, including TIR1, COI1, SLY1, and EBF1/EBF2 in hormone signaling (Ruegger et al. 1998; Xie et al. 1998; McGinnis et al. 2003; Parry and Estelle 2006), SON1 in defense response (Kim and Delaney 2002), ORE9/MAX2 in controlling shoot branching and leaf senescence (Woo et al. 2001; Stirnberg et al. 2002), EID1 in photomorphogenesis (Dieterle et al. 2001), ZTL, FKF1, and LKP2 in flowering time and the circadian clock (Nelson et al. 2000; Somers et al. 2000; Yasuhara et al. 2004), and UFO in floral organ development (Ingram et al. 1995; Samach et al. 1999; Zhao et al. 2001). In addition, molecular analysis suggests that the same *Arabidopsis* F-box proteins may bind multiple SKP1 homologs, suggesting the combinatorial potential for formation of a very large set of SCF complexes (Takahashi et al. 2004).

The large plant F-box protein family can be divided into subfamilies according to the presence of additional protein-protein interaction domains near the C-terminus. These domains include the WD40 repeat, the Leucine-rich repeat, Tub, Lectin, the Kelch repeat and other motifs. The Kelch motif contains 44–56 amino acid residues and was initially identified in the *Drosophila melanogaster* KELCH protein (Xue and Cooley 1993; Bork and Doolittle 1994). Previous studies have uncovered the consensus of Kelch motif that is characterized by four highly conserved residues: two adjacent glycines (G), and a pair of tyrosine (Y) and trytophan (W) separated by about six residues (Adams et al. 2000; Prag and Adams 2003). A single Kelch motif forms four beta sheets, and multiple Kelch motifs can associate together, forming a bladed beta-propeller that interacts with other proteins (Ito et al. 1991). For example, the well-studied human Keap1 protein contains seven Kelch motifs and can form an E3 ligase together with Cul3 and Rbx1 to ubiquitinate the Srf2 protein (Li et al. 2004).

Although the Kelch motif is commonly found in many organisms, including viruses, bacteria, fungi, plants and animals, only a few Kelch motifs containing F-box proteins (KFBs) have been characterized (Xue and Cooley 1993; Bork and Doolittle 1994; Adams et al. 2000). The only well-studied KFBs are the three highly similar *Arabidopsis* proteins (ZTL, FKF, LKP2), which are involved in the flowing time and circadian control (Nelson et al. 2000; Han et al. 2004; Somers et al. 2004; Yasuhara et al. 2004; Imaizumi et al. 2005). Furthermore, little is known about the evolutionary history of the KFBs. The recent determination of the genomic sequences of several plants has allowed a thorough analysis of KFBs in plants. In this study, we carried out extensive BLAST searches for all putative KFBs in several organisms and carried out phylogenetic analyses of both animal and plant KFBs. The gene expression profiles of *Arabidopsis* KFBs were also provided by microarray data analyses. In addition, the information on the chromosome distribution and possible gene duplication events in both rice (OsKFBs) and *Arabidopsis* (AtKFBs) was presented. The existence of KFBs in both plant and animal suggests the origin(s) that predates the divergence of animals and plants, although none was detected in fungi and other kingdoms. Comparative analysis of the plant KFBs from angiosperms, a gymnosperm and a moss indicated that the KFBs form a number of subfamilies that are well conserved in plants. Moreover, one subfamily has experienced rapid gene birth primarily through tandem duplication events that occurred before the split of *Arabidopsis* and *Brassica*. Most of these recently duplicated genes are expressed at very low levels in seven *Arabidopsis* organs/structures that we analyzed. Our results indicate that the KFB family has expanded in plants, and contains both members that are highly stable and conserved, as well as members that are very dynamic and rapidly evolving.

## Results

### Plant genomes encode a large number of KFBs with different numbers of Kelch motif

We used the protein sequences of the known *Arabidopsis* KFBs as queries to carry out BLAST searches in the *Arabidopsis* genome. Ninety-seven KFBs were detected in the *Arabidopsis* genome, with the Pfam E-value cut off at 0.5 for both the F-box domain and the Kelch motif. We then used these *Arabidopsis* KFBs as queries to carry out BLAST searches against sequences of other plant genomes, including *Brassica rapa*, *Populus trichocarpa*, maize, rice, *Physcomitrella patens*, and pine ESTs. We also searched against the budding yeast genome, but no KFB was detected. The single human KFB named F-box 42 was also used as query to search for the animal, fungi, protist and prokaryote KFBs through the NCBI website. In summary, no KFB was found in single-cell organisms, and only a single copy of KFB was detected in human, zebrafish (Identity to human = 471/683 (68%), Similarity = 509/683 (74%)), *Drosophila malenogaster* (Identity = 215/706 (30%), Similarity = 321/706 (45%)), and other insects. All animal KFBs are close homologs of the human F-box 42, and each contains three Kelch motifs. In contrast, a large number of KFBs were identified in plant genomes. For example, at least 43 in *Brassica rapa* (36 partial sequences with more than 60% identity to *Arabidopsis* homologs are not included in this study), 41 in *Populus*, 28 in rice, 34 in maize, 10 in pine and 20 in

*Physcomitrella*. Furthermore, plant KFBs contain different numbers of Kelch motif, from one to five. Among the 273 plant KFBs included in this study, 37 KFBs contain a single Kelch motif, 161 have two Kelch motifs, 53 have three Kelch motifs, 10 have four Kelch motifs, and the remaining 12 have five Kelch motifs.

### Plant KFB family has expanded dramatically via multiple duplication events

To investigate the evolutionary relationships of KFBs, multiple protein sequences alignment of KFBs were carried out as described in the **Materials and Methods**. Both neighbor joint (NJ) and maximum likelihood (ML) methods generated trees with similar topology (Figure 1; a larger phylogenetic tree with 284 sequences is available upon request). Interestingly, all plant KFBs form a separate clade from animal KFBs, with 100% bootstrap support, suggesting that both plant and animal KFBs could be derived from as few as a single gene in the common ancestor of animals and plants. Furthermore, the well-studied *Arabidopsis* ZTL subfamily (G6, see below) occupies the basal position within the plant KFB lineage, suggesting that the ZTL members might resemble the ancestral KFB in plants, consistent with the fact that, among plant KFBs, the ZTL proteins are most closely related to the human F-box 42 and its vertebrate orthologs (25% identity and 40% similarity to the human KFB).

→

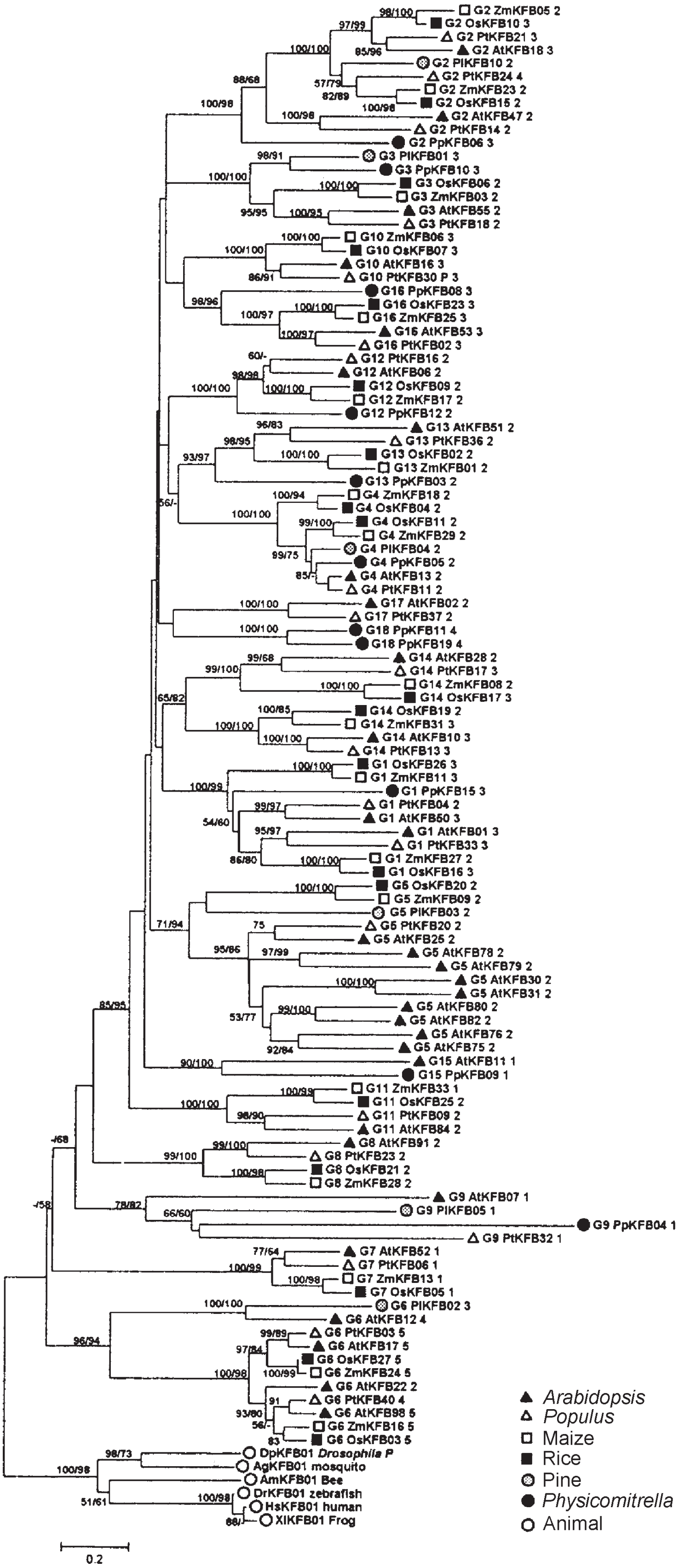


**Figure 1.** Phylogenetic tree of 113 representative Kelch repeats-containing F-box proteins (KFBs).

The KFBs were selected based on the phylogenetic analysis of all 284 KFBs identified in this study (available upon request). The tree was constructed by the neighbor-joining method with Poisson correction, pairwise deletion and bootstrap of 1 000 replicates. The bootstrap values of both neighbor-joining (NJ) tree (first number; 1 000 replicates) and maximum likelihood (ML) tree (second number; 100 replicates) higher than 50 are shown for each clade. We divided the plant KFBs into 18 subfamilies named as G1 to G18. Animal KFBs form a single clade. The KFB name in the tree combines subfamily, species name, and Kelch motif information. For example, G2 ZmKFB05 2 means "*Zea Mays* KFB05 with two Kelch motifs, belonging to the G2 subfamily". Ag, *Anopheles gambiae*; Am, *Apis mellifera*; At, *Arabidopsis thaliana*; Dp, *Drosophila pseudoobscura*; Dr, *Danio rerio*; Hs, *Homo sapiens*; Os, *Oryza sativa*; Pl, *Pinus taeda*; Pp, *Physcomitrella patens*; Pt, *Populus trichocarpa*; Xl, *Xenopus laevis*; Zm, *Zea mays*.

While the KFBs in animals remained single copy, plant KFBs have increased dramatically in number and could be grouped into 18 highly supported subfamilies, named G1 to G18, for small to moderately-sized clades that have good bootstrap support (at least 65/82 for NJ/ML) from the phylogentic analysis (Figure 1). The subfamilies are further supported by the presence of additional conserved motifs that are shared by members of a subfamily. One large clade with 85/95 bootstrap values was not considered as a single subfamily because it was too large and complex. Eleven subfamilies were found to have members from at least one angiosperm species analyzed here, and from pine and/or *Physcomitrella*, six subfamilies were only detected in the angiosperm, one subfamily was only detected in *Physcomitrella*, indicating that the majority of the subfamilies were generated by duplications that occurred before the split between gymnosperms and angiosperms. Because we have examined only a few species and the complete pine genomic sequence is not available, the absence of some subfamily members in either pine or the angiosperm taxa is inconclusive. Fourteen subfamilies have at least one member from each of rice, maize, *Arabidopsis*, and poplar; among them, three subfamilies (G2, G6, and G14), each have two well-supported clades, each with members from these four species. Therefore, the ancestor of angiosperms likely had at least 17 copies of KFBs. The existence of at least six other angiosperm KFBs (G1, G2, G4, G9, G15, G17) outside the above 17 clades suggests that the number of ancestral angiosperm KFBs might be as many as 23. The increase from possibly a single gene in the common ancestor of plants and animals to about 20 before the separation of angiosperms and gymnosperms indicates that a number of gene duplications had occurred before the emergence of flowering plants.

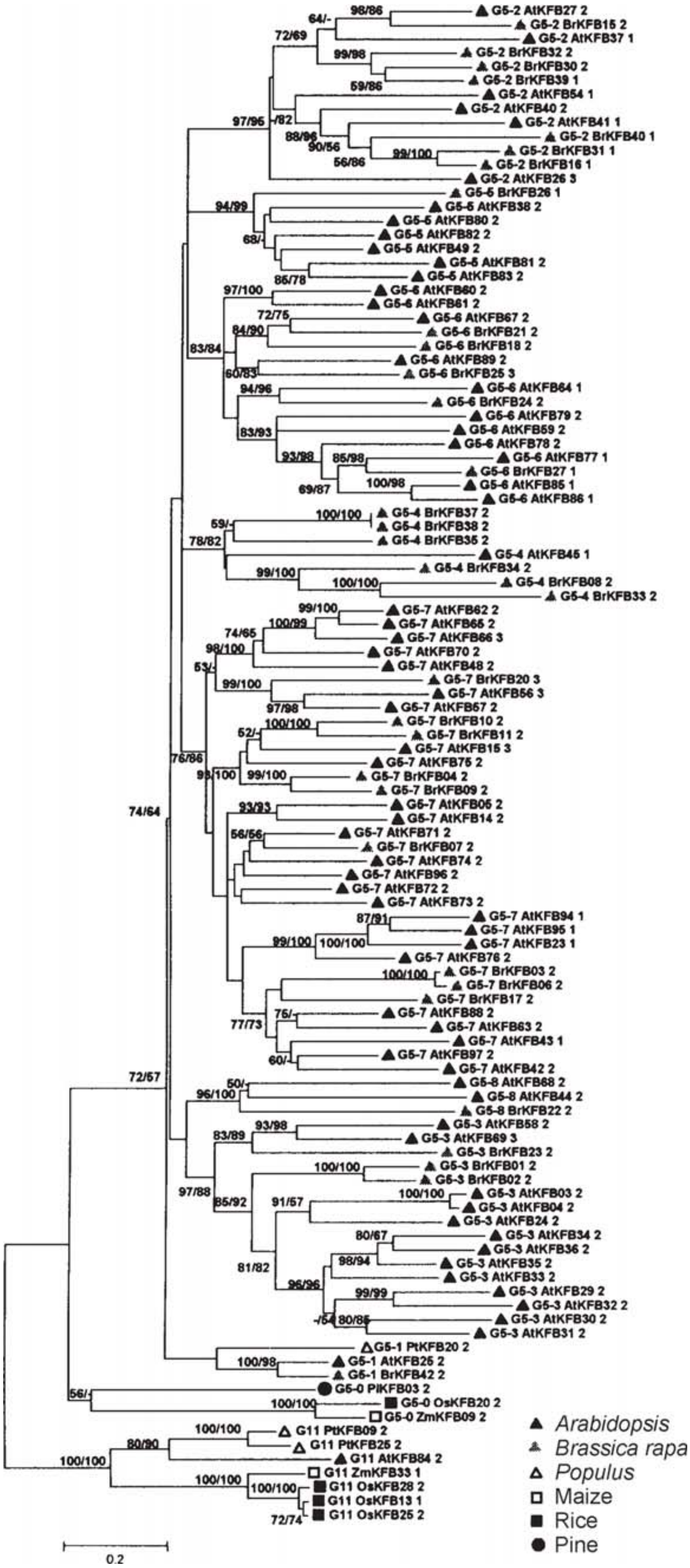
The phylogeny of the KFBs also provides evidence for more recent duplications within the specific lineages of flowering plants. Among well-supported clades with both rice and maize sequences, 17 have one from each species, five have one rice KFB and two or three maize KFBs, one has one maize KFB and three rice KFBs, and four have only rice or maize genes. Maize is a recent tetraploid and has a much larger genome than rice; it is possible that additional maize KFBs will be identified as more maize genome sequences become available. In clades with *Arabidopsis* and poplar members, nine have one *Arabidopsis* KFB and two close poplar paralogs, five have one from each species, three have a pair of paralogs from each species, and one has one *Arabidopsis* gene and three close paralogs from poplar. The frequent detection of two poplar paralogs corresponding to one *Arabidopsis* gene is consistent with the fact that poplar is a recent tetraploid.

### G5 KFBs underwent multiple recent gene duplications in *Arabidopsis* and *Brassica*

Our initial analysis (not shown) indicated that the G5 subfamily of KFBs has a large number of members in *Arabidopsis*, but only a single copy in rice, poplar, and pine (Figure 1), suggesting that there have been recent gene duplication events in the lineage leading to *Arabidopsis* since the divergence from poplar. To further study the evolutionary history of the G5 subfamily of KFBs, we carried out analysis with the addition of KFBs from *B. rapa*, which is in the same family (Brassicaceae) as *Arabidopsis* and also has the most abundant genomic sequence information available among the *Brassica* species. As shown in Figure 2, multiple KFBs closely related to *Arabidopsis* G5 members were identified in *B. rapa*, suggesting that many of the duplication events predated the split of *Brassica* and *Arabidopsis*. Members of the G5 subfamily form nine highly supported clades, named as G5-0 to G5-8. Using the G11 subfamily sequences (closest to the G5 subfamily) as the outgroup, G5-0 is the basal clade and contains the single copy sequences from pine, rice, and maize, suggesting that they represent the ancestral state of the G5 subfamily in the early seed plants. G5-1 includes AtKFB25, BrKFB42 and the single-copy PtKFB20 (poplar) forming a sister clade to all remaining G5 members; it is possible that G5-1 is the closest to the eudicot origin of the G5 KFB. Each of the remaining seven clades contains sequences from both *Arabidopsis* and *Brassica*, suggesting that they originated between the time of separation from poplar and the time of split of *Arabidopsis* and *Brassica* due to several rounds of duplications. Furthermore, within these clades, small clades of only BrKFBs or only AtKFBs provide evidence for gene duplication events after the divergence of the *Arabidopsis* and *Brassica*.

### Most of the G5 *AtKFBs* are present as tandem repeats in the *Arabidopsis* genome

To investigate the gene duplication events of plant *KFBs*, we carried out the analysis of chromosome distributions of plant *KFBs* in both *Arabidopsis* and rice. Since the G5-*KFBs* have expanded greatly in *Arabidopsis* and *Brassica*, not in the other plants that we analyzed, we investigated them separately (Figure 3). As shown in Figure 3, the 66 *AtKFBs* in the G5 subfamily distribute unevenly on the chromosomes with high densities on the lower arm of chromosomes II and IV. Among them, multiple groups of closely related *AtKFBs* form tandem arrays on the same chromosomes, strongly suggesting that they were generated by tandem duplications. Specifically, *AtKFB29* to *AtKFB36* form a clade of tandemly arrayed genes on chromosome II. Similarly, *AtKFB03* and *AtKFB04*, *AtKFB56* and *AtKFB57*, *AtKFB60* and *AtKFB61*, *AtKFB77* to *AtKFB79*, *AtKFB85* and *AtKFB86*, *AtKFB94* and *AtKFB95* are members of the same clades, respectively, and are adjacently located on



the same chromosomes. On the other hand, although *AtKFB71* to *AtKFB76* also form a tandem array on chromosome IV, they form a large clade with other *AtKFBs*, suggesting that other mechanism(s) of gene duplication might also be involved. Similar situations are found with *AtKFB26* and *AtKFB27*, *AtKFB80* to *AtKFB83*, *AtKFB87* and *AtKFB88*, *AtKFB94* (*AtKFB95*) and *AtKFB96*. In summary, at least 38 of 66 G5 *AtKFBs* seem to have been generated by the tandem duplication events.

The other groups of *AtKFBs* and *OsKFBs* also seem to distribute unevenly in both *Arabidopsis* and rice (Figures 4, 5), with high density on chromosome I in *Arabidopsis* and on chromosome II in rice. Only *AtKFB91* and *AtKFB92* are adjacent and possibly generated by tandem duplication, no tandem duplication is obvious in rice *KFBs*.

## Gene expression profiles of *AtKFBs* and *OsKFBs*

Because gene expression patterns often provide important clues for gene functions, we examined microarray data to learn about expression profiles of *AtKFBs*. Among 97 *AtKFBs*, 67 genes were included in the Affymetrix chips (Figure 6); 41 of them belong to the G5 subfamily and 26 to the other subfamilies. As shown in Figure 6, 15 of the G5 *AtKFBs* were expressed in one or more organs/structures, whereas the expression intensity of the remaining 26 G5 genes were below 50 in all of the

←

**Figure 2.** Phylogenetic tree of 102 Kelch repeats-containing F-box proteins (KFBs) in the plant G5 subfamily.

The tree was constructed by the neighbor-joining (NJ) method with Poisson correction, pairwise deletion and bootstrap of 1 000 replicates. The bootstrap values of both NJ (1 000 replicates) and maximum likelihood (ML) trees (100 replicates) higher than 50 are shown for each clade with the first number from the NJ tree and second number from the ML tree. The G11 subfamily was used as the out-group. A large number of G5 KFBs were identified in *Arabidopsis* and *Brassica*, and only one in rice, maize, *Populus* and pine. The G5 subfamily are further divided into nine clades and named as G5-0 to G5-8. Ag, *Anopheles gambiae*; Am, *Apis mellifera*; At, *Arabidopsis thaliana*; Dp, *Drosophila pseudoobscura*; Dr, *Danio rerio*; Hs, *Homo sapiens*; Os, *Oryza sativa*; Pl, *Pinus taeda*; Pp, *Physcomitrella patens*; Pt, *Populus trichocarpa*; Xl, *Xenopus laevis*; Zm, *Zea mays*.

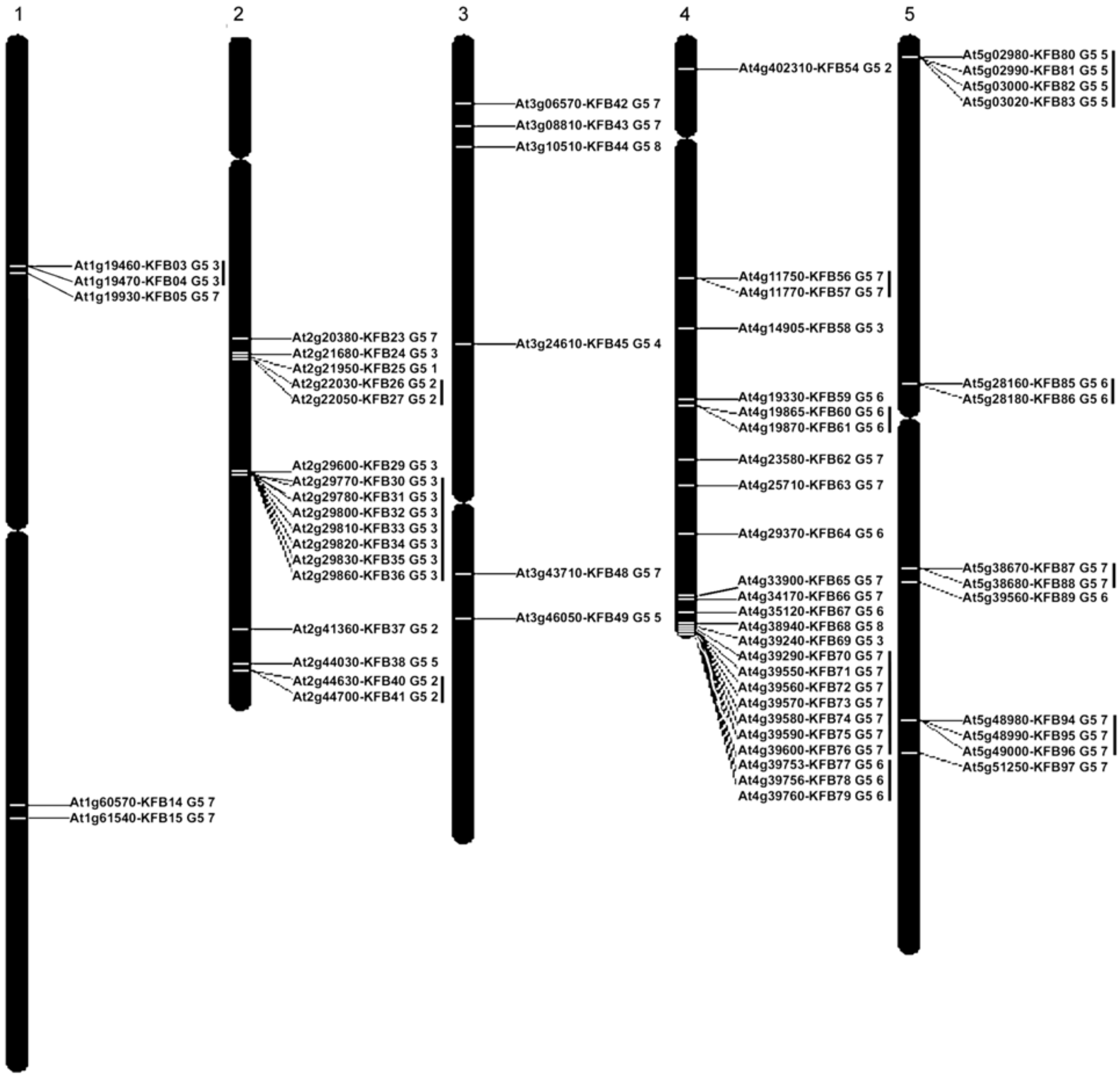


**Figure 3.** Chromosome distribution of 66 G5-Kelch repeats-containing F-box proteins (KFBs) in *Arabidopsis*.

Thirty-eight of 66 G5-*KFBs* are tandem duplicates. *KFBs* within the same subgroup were labeled with a number on the right. The tandem duplicated *KFBs* were marked with line to the right of the gene names.

seven organs/structures, indicating that they are not expressed at reliably detectable levels in these structures. These genes might be expressed at higher levels in some other tissues or under conditions different from our growth conditions; also some of them could be pseudogenes. It is worth noting that half of these 26 G5 genes with little or no expression are found in tandem arrays, as described above, such as *AtKFB30*, *AtKFB34*, and *AtKFB36*, also *AtKFB80*, *AtKFB81*, *AtKFB82* and *AtKFB83*.

Among the remaining G5 members, seven were expressed ubiquitously. For example, *AtKFB42* and *AtKFB63* are close paralogs with similar gene expression patterns, suggesting that they may share some redundant function. On the other hand, these genes might still have different functions, either because the slight sequence divergence between these two genes might be sufficient to cause functional differences, there might be

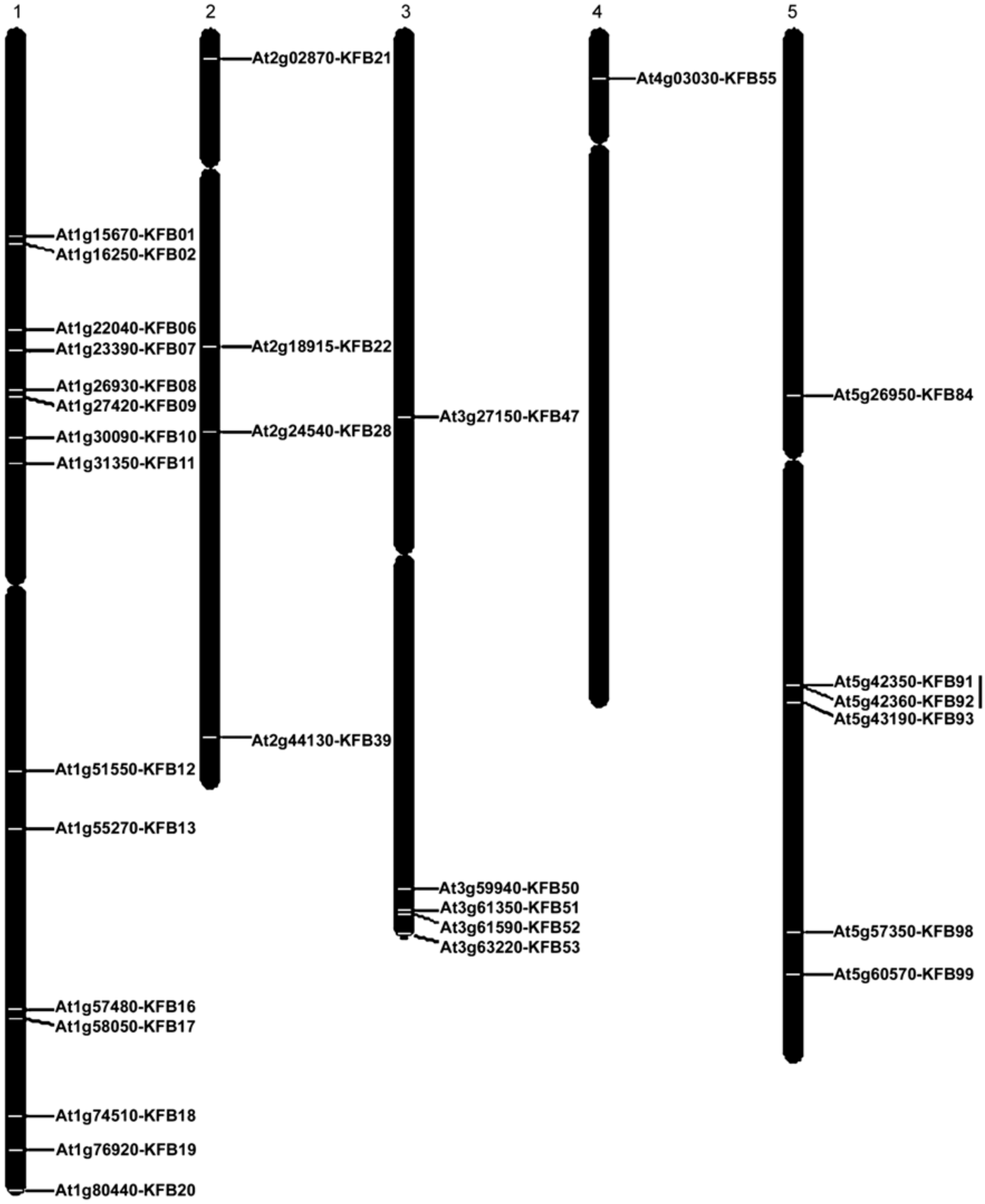


**Figure 4.** Chromosome distribution of 31 *Arabidopsis* non-G5 Kelch repeats-containing F-box proteins (KFBs).

All these *AtKFBs* except *AtKFB91* and *AtKFB92* were likely to have been generated by duplications other than tandem duplication.

expression differences that were not detected by the microarray analysis, or these genes may have different expression in other organs or conditions we did not test. *AtKFB25*, *AtKFB69*, *AtKFB73* and *AtKFB96* are expressed at relatively high levels compared with other *AtKFBs* in the G5 subfamily. They may play important roles in *Arabidopsis*. Four other genes,

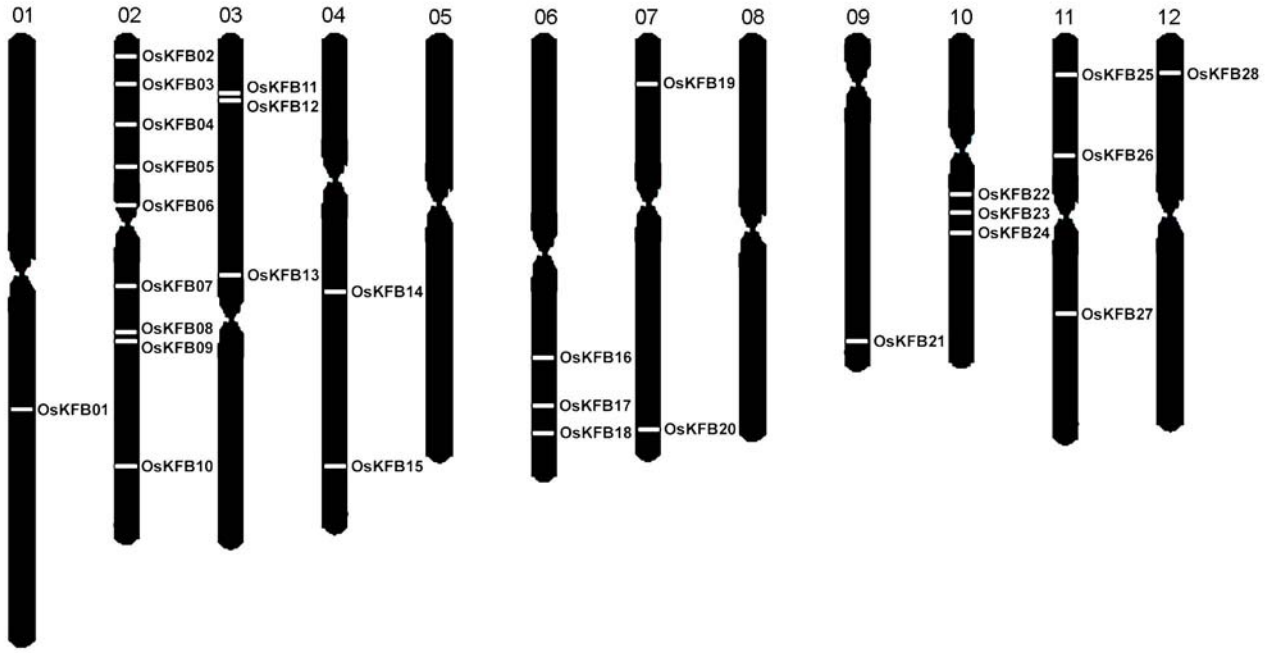


**Figure 5.** Chromosome distribution of 28 Kelch repeats-containing F-box proteins (KFBs) in rice.

There is no evidence for tandem duplication event detected for rice *KFBs*.

the *AtKFB04*, *AtKFB29*, *AtKFB32*, and *AtKFB75*, are expressed specifically in the immature anther, suggesting functions in the developing anther. Although *AtKFB29* and *AtKFB32* are highly similar in sequence, *AtKFB29* is expressed at much higher levels than *AtKFB32*, suggesting that both genes may be involved in the same pathway or share some redundant function, and *AtKFB29* may play a major role, compared with that of *AtKFB32*. Other possibilities also exist similar to what were suggested above for *AtKFB42* and *AtKFB63*. The *AtKFB14* gene is specifically expressed in stage 12 flowers, and *AtKFB31* is preferentially expressed in both anther and leaf tissues. In contrast to the G5 *AtKFBs*, the other *KFBs* are mostly expressed ubiquitously at high levels, suggesting that they play important general roles. Interestingly, the gene expression pattern of *AtKFB17* is also different from the members of *ZTLs*, further supporting the idea that its function may be different from its homologs in the ZTL family.

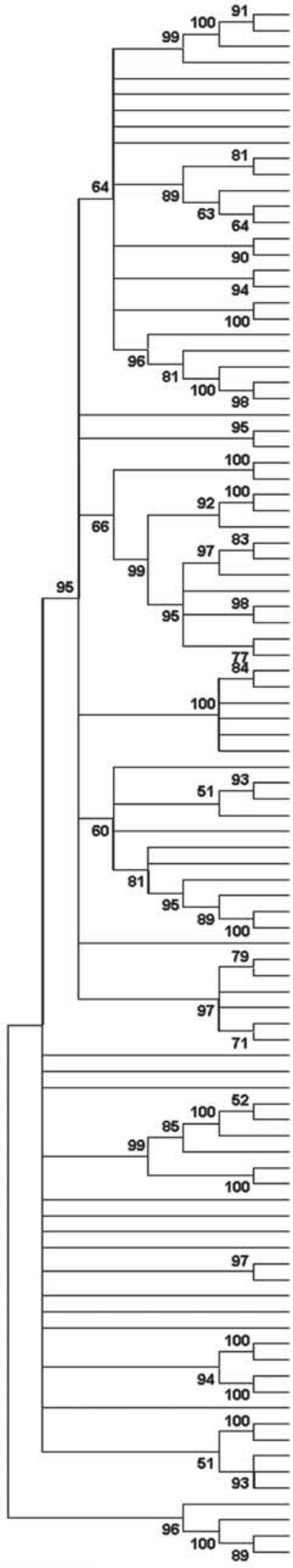
To obtain information about the expression of the rice KFB genes, we examined the publicly available EST data. We found that 24 of the 28 rice *KFBs* have corresponding EST information (not shown), indicating that they are expressed, but *OsKFB08*, *OsKFB1*3, *OsKFB21*, and *OsKFB28* did not have an EST. Either these four genes are not expressed, are expressed at low levels, or are expressed under certain conditions that were different from those used to grow the plants for the EST analysis. Of three closely related genes, *OsKFB13*, *OsKFB25* and *OsKFB28,* only *OsKFB25* has EST information, suggesting that it may play an important role in plants grown in common conditions.

## Discussion

### Rapid gene birth evolution of plant KFBs

Protein degradation through the ubiquitin-mediated pathway is a key process in regulating cell cycle progression, transcription and signal transduction in eukaryotic organisms. Previous studies have found that both F-box protein and Kelch repeat-containing protein are ancient and widely distributed, and that both can interact with other proteins that participate in protein degradation processes (Xue and Cooley 1993; Bork and Doolittle 1994; Adams et al. 2000; del Pozo and Estelle 2000; Li et al. 2004; Lechner et al. 2006). But the Kelch repeat-containing F-box proteins were only reported in animals and plants (Andrade et al. 2001; Jin et al. 2004). Our results suggest that the F-box and the Kelch motifs are present together in the same proteins only in eukaryotes. The fact that KFBs are only detected in multi-cellular organisms suggests that the combination of the F-box and Kelch motifs might have contributed to the evolutionary success of multi-cellular organisms, which probably needed more complicated mechanisms of protein degradation to regulate complex biological processes.

Although only a single copy of KFB is highly conserved in human and other animals, dozens of KFBs were found in plants, suggesting that rapid gene birth events have occurred in plants. Among the 18 subfamilies of plant KFBs, 11 subfamilies were well conserved in both angiosperms and a gymnosperm (pine) or a moss, suggesting that their functions had diversified in early land plants and have been conserved during seed plants evolution. Further analysis of these KFBs found that members of most of the subfamilies were expressed ubiquitously at relatively high levels, supporting the idea that they may play important roles in plants. Since the divergence of gymnosperms and angiosperms, while most subfamilies have been relatively stable or only expanded slightly, a dramatic example of rapid gene birth is the G5 subfamily, which experienced numerous gene duplication events in the lineage leading to *Arabidopsis* and *Brassica*. Although the G5 members form a large subfamily, the functions of most members are not clear, since they are expressed at very low levels and might be pseudogenes. Nevertheless, we found that some of the G5 members are expressed more specifically, suggesting that they may have evolved more specialized functions following gene duplication.



| | An | In | S12 | Si | St | Lf | Rt |
|---|---|---|---|---|---|---|---|
| G5-7 AtKFB94 1 | | | | | | | |
| G5-7 AtKFB95 1 | 70 | 87 | 49 | 59 | 46 | 37 | 43 |
| G5-7 AtKFB23 1 | 21 | 22 | 19 | 28 | 34 | 34 | 20 |
| G5-7 AtKFB76 2 | | | | | | | |
| G5-7 AtKFB96 2 | 355 | 275 | 187 | 260 | 131 | 186 | 152 |
| G5-7 AtKFB71 2 | 120 | 191 | 111 | 116 | 95 | 63 | 95 |
| G5-7 AtKFB74 2 | | | | | | | |
| G5-7 AtKFB72 2 | 72 | 74 | 53 | 49 | 42 | 74 | 55 |
| G5-7 AtKFB73 2 | 182 | 220 | 132 | 199 | 149 | 149 | 182 |
| G5-7 AtKFB63 2 | 109 | 137 | 92 | 87 | 98 | 103 | 141 |
| G5-7 AtKFB88 2 | | | | | | | |
| G5-7 AtKFB43 1 | 38 | 23 | 20 | 29 | 27 | 26 | 23 |
| G5-7 AtKFB42 2 | 146 | 104 | 79 | 120 | 135 | 131 | 80 |
| G5-7 AtKFB97 2 | 33 | 24 | 25 | 20 | 33 | 31 | 16 |
| G5-7 AtKFB05 2 | 29 | 27 | 20 | 33 | 33 | 29 | 22 |
| G5-7 AtKFB14 2 | 33 | 18 | 86 | 43 | 19 | 27 | 26 |
| G5-7 AtKFB15 3 | | | | | | | |
| G5-7 AtKFB75 2 | 59 | 11 | 40 | 16 | 7 | 8 | 25 |
| G5-7 AtKFB56 3 | 14 | 13 | 12 | 20 | 28 | 28 | 13 |
| G5-7 AtKFB57 2 | | | | | | | |
| G5-7 AtKFB48 2 | | | | | | | |
| G5-7 AtKFB70 2 | | | | | | | |
| G5-7 AtKFB66 3 | | | | | | | |
| G5-7 AtKFB62 2 | 12 | 7 | 41 | 6 | 9 | 9 | 8 |
| G5-7 AtKFB65 2 | 31 | 4 | 32 | 21 | 2 | 4 | 21 |
| G5-1 AtKFB25 2 | 190 | 197 | 166 | 196 | 209 | 340 | 278 |
| G5-3 AtKFB44 2 | | | | | | | |
| G5-3 AtKFB68 2 | 106 | 50 | 56 | 81 | 71 | 60 | 104 |
| G5-3 AtKFB58 2 | | | | | | | |
| G5-3 AtKFB69 3 | 303 | 362 | 292 | 344 | 302 | 260 | 254 |
| G5-3 AtKFB03 2 | 78 | 1 | 2 | 1 | 8 | 1 | 3 |
| G5-3 AtKFB04 2 | 78 | 1 | 2 | 1 | 8 | 1 | 3 |
| G5-3 AtKFB24 2 | 16 | 10 | 4 | 13 | 15 | 19 | 6 |
| G5-3 AtKFB34 2 | 47 | 23 | 12 | 25 | 25 | 21 | 18 |
| G5-3 AtKFB36 2 | 31 | 21 | 20 | 22 | 16 | 36 | 24 |
| G5-3 AtKFB35 2 | 54 | 8 | 1 | 2 | 2 | 2 | 2 |
| G5-3 AtKFB33 2 | | | | | | | |
| G5-3 AtKFB29 2 | 183 | 28 | 22 | 23 | 24 | 33 | 25 |
| G5-3 AtKFB32 2 | 75 | 15 | 6 | 18 | 17 | 30 | 7 |
| G5-3 AtKFB30 2 | 32 | 6 | 3 | 12 | 3 | 3 | 8 |
| G5-3 AtKFB31 2 | 60 | 25 | 38 | 30 | 33 | 51 | 19 |
| G5-5 AtKFB81 2 | 17 | 25 | 11 | 16 | 9 | 23 | 19 |
| G5-5 AtKFB83 2 | 25 | 17 | 18 | 19 | 28 | 29 | 39 |
| G5-5 AtKFB49 2 | | | | | | | |
| G5-5 AtKFB82 2 | 27 | 20 | 32 | 14 | 22 | 29 | 19 |
| G5-5 AtKFB38 2 | 28 | 23 | 18 | 18 | 23 | 34 | 10 |
| G5-5 AtKFB80 2 | 18 | 24 | 16 | 27 | 23 | 25 | 19 |
| G5-6 AtKFB89 2 | 30 | 33 | 16 | 40 | 22 | 35 | 24 |
| G5-6 AtKFB60 2 | 96 | 118 | 64 | 89 | 67 | 54 | 29 |
| G5-6 AtKFB61 2 | | | | | | | |
| G5-6 AtKFB67 2 | 26 | 17 | 15 | 24 | 31 | 30 | 29 |
| G5-6 AtKFB64 1 | 14 | 14 | 10 | 35 | 11 | 16 | 10 |
| G5-6 AtKFB79 2 | | | | | | | |
| G5-6 AtKFB59 2 | 2 | 2 | 11 | 3 | 1 | 3 | 3 |
| G5-6 AtKFB78 2 | 17 | 6 | 8 | 5 | 7 | 14 | 7 |
| G5-6 AtKFB77 1 | 24 | 4 | 49 | 21 | 17 | 24 | 6 |
| G5-6 AtKFB85 1 | 2 | 6 | 1 | 1 | 2 | 3 | 3 |
| G5-6 AtKFB86 1 | 2 | 6 | 1 | 1 | 2 | 3 | 3 |
| G5-4 AtKFB45 1 | | | | | | | |
| G5-2 AtKFB40 2 | 50 | 48 | 38 | 44 | 41 | 29 | 25 |
| G5-2 AtKFB41 1 | | | | | | | |
| G5-2 AtKFB26 3 | 15 | 9 | 45 | 8 | 7 | 11 | 7 |
| G5-2 AtKFB54 1 | 2 | 4 | 4 | 6 | 8 | 11 | 4 |
| G5-2 AtKFB27 2 | 2 | 3 | 2 | 2 | 3 | 2 | 2 |
| G5-2 AtKFB37 1 | | | | | | | |
| G12 AtKFB06 2 | 100 | 107 | 102 | 131 | 106 | 134 | 135 |
| G11 AtKFB84 2 | 128 | 182 | 183 | 108 | 105 | 148 | 85 |
| G15 AtKFB11 1 | 64 | 66 | 81 | 76 | 86 | 112 | 231 |
| G2 AtKFB18 3 | 115 | 173 | 150 | 137 | 165 | 100 | 208 |
| G2 AtKFB21 3 | 130 | 92 | 255 | 252 | 219 | 169 | 340 |
| G2 AtKFB08 4 | 38 | 33 | 64 | 139 | 45 | 112 | 175 |
| G2 AtKFB99 2 | 269 | 260 | 228 | 305 | 324 | 197 | 455 |
| G2 AtKFB47 2 | 11 | 23 | 4 | 16 | 14 | 9 | 96 |
| G2 AtKFB90 2 | 25 | 21 | 14 | 234 | 28 | 27 | 29 |
| G14 AtKFB28 2 | 29 | 45 | 90 | 163 | 370 | 190 | 85 |
| G3 AtKFB55 2 | 299 | 197 | 275 | 194 | 169 | 532 | 289 |
| G16 AtKFB53 3 | 273 | 146 | 147 | 147 | 126 | 147 | 161 |
| G13 AtKFB51 2 | 83 | 85 | 51 | 49 | 50 | 52 | 58 |
| G10 AtKFB09 3 | | | | | | | |
| G10 AtKFB16 3 | 203 | 260 | 160 | 147 | 292 | 585 | 1250 |
| G4 AtKFB13 2 | 225 | 203 | 186 | 256 | 161 | 197 | 176 |
| G14 AtKFB10 3 | 127 | 141 | 143 | 143 | 130 | 145 | 128 |
| G17 AtKFB02 2 | | | | | | | |
| G1 AtKFB01 3 | 329 | 122 | 294 | 146 | 99 | 877 | 1434 |
| G1 AtKFB20 3 | 539 | 75 | 180 | 196 | 104 | 1688 | 638 |
| G1 AtKFB39 2 | 509 | 73 | 644 | 415 | 719 | 267 | 146 |
| G1 AtKFB50 3 | 1688 | 240 | 430 | 556 | 824 | 1427 | 100 |
| G9 AtKFB07 1 | 364 | 102 | 286 | 493 | 347 | 2273 | 333 |
| G8 AtKFB91 2 | 239 | 221 | 168 | 220 | 165 | 146 | 163 |
| G8 AtKFB92 2 | 239 | 221 | 168 | 220 | 165 | 146 | 163 |
| G7 AtKFB93 1 | 276 | 167 | 141 | 240 | 114 | 605 | 259 |
| G7 AtKFB19 2 | 368 | 322 | 234 | 297 | 286 | 479 | 465 |
| G7 AtKFB52 1 | 109 | 142 | 194 | 108 | 105 | 99 | 164 |
| G6 AtKFB12 4 | 51 | 66 | 42 | 45 | 60 | 77 | 48 |
| G6 AtKFB17 5 | 395 | 314 | 280 | 145 | 54 | 83 | 117 |
| G6 AtKFB22 2 | 114 | 85 | 111 | 122 | 139 | 186 | 138 |
| G6 AtKFB98 5 | 190 | 145 | 150 | 180 | 219 | 317 | 145 |

←

**Figure 6.** Expression patterns of Kelch repeats-containing F-box proteins (KFBs) in *Arabidopsis*.

**(A)** A phylogenetic tree of 97 KFBs in *Arabidopsis*. The tree was constructed by the neighbor-joining (NJ) method with Poisson correction, pairwise deletion and bootstrap of 1 000 replicates. Only bootstrap values higher than 50 are shown.

**(B)** Gene expression profile of *KFBs* in *Arabidopsis*. 67 *AtKFBs* were analyzed previously by micorarray and labeled with the intensity value; the other 30 shown as blank were not included in the Affymetrix microarray slide. The intensity value of 50 is regarded as the cut off for reliable detection of gene expression. The expression data for *AtKFB03/04* were from the same probe set, and the same as those for *AtKFB85/86* and *AtKFB91/92* gene pairs. All tissues were from wild type *Arabidopsis* Landsberg *erecta* plants. An, anther; In, young inflorescence; Lf, leaf; Rt, root; S12, flower at flower stage 12; Si, silique; St, stem.

### Mechanism for controlling flower timing and circadian oscillator may be conserved in flowering plants

The well-supported subfamily G6 family contains four *Arabidopsis* members (AtKFB12, AtKFB17, AtKFB22, AtKFB98), with three of them (LKP1/ZTL/ADO1/AtKFB98, LKP2/ADO2/FKL2/AtKFB22, FKF1/ADO3/AtKFB17) having been shown genetically to be important for the timing of normal flowering and the circadian clock. LKP1 and LKP2 are recent duplicates in *Arabidopsis* and share some redundant function (Nelson et al. 2000; Somers et al. 2004; Yasuhara et al. 2004). Our phylogenetic results indicate that the duplication of LKP1/LKP2 and FKF1 is likely to have occurred before the divergence of eudicots and monocots. The existence of the LKP1/LKP2 and FKF1 (co-) orthologs in both eudicots and monocots strongly suggests that the function of these genes in controlling the circadian clock and flowering time is highly conserved in angiosperms. Unlike the three well-characterized members, AtKFB12 (At1g51550) lacks the light-absorbing LOV domain and has a distinct gene expression pattern, suggesting that its function might have diverged from the other members in the G6/ZTL subfamily. In addition, the absence of orthologs of AtKFB12 in rice and other angiosperm species suggests further evolutionary and possible functional differences between AtKFB12 and other G6 members.

### Contribution of tandem duplication to the KFBs in *Arabidopsis*

We showed that multiple AtKFBs, particularly G5 members, are tandemly located on the same chromosome, suggesting their generation by tandem duplication. In contrast, no tandem arrays of KFBs were found in rice. Furthermore, most of the tandem arrayed G5 members are expressed either at very low levels or below reliable detection levels in the organs/structure that we tested. In addition, many of them have degenerate Kelch motifs, suggesting that they might be pseudogenes or their functions may be divergent. Indeed, some of the G5 members exhibit preferential expression in some organs, supporting the idea of recently evolved functions for these members. It is possible that some of the novel functions provide selective advantages, allowing the duplicated copies to persist in the genomes of *Arabidopsis* and *Brassica*. Although rice and poplar do not have similar rapid gene births to those seen in the G5 subfamily in Brassicaseae, it is possible other plant genomics efforts may reveal additional expansions of the KFBs in the near future.

F-box proteins are known or thought to interact with the SKP1 homologs as subunits of the SCF complexes (del Pozo and Estelle, 2000; Zheng et al. 2002; Risseeuw et al. 2003). The evolution of the SKP1 gene family has a similar pattern to that of the KFBs (Kong et al. 2004). In vertebrate animals, there is only one copy of SKP1 in each genome, whereas rice and *Arabidopsis* each have more than 20 SKP1 homologs, indicating that rapid gene birth events also happened in the SKP1 family in plants. Furthermore, the *Arabidopsis* SKP1 homologs (ASKs) also form several tandem repeats. It has been shown that different ASKs could interact with different F-box proteins, including the KFBs (Yamanaka et al. 2002; Risseeuw et al. 2003). The similarity in the patterns of evolution of SKP1s and KFBs suggests possible co-evolution between these two gene families. Further more, the analysis of protein-protein interaction between the KFBs and SKP1 homologs may provide insights into this possible co-evolution of these key regulators of protein degradation.

In this study, we showed that the plant KFBs experienced numerous gene duplication events since the divergence of animals and plants, including many that resulted in many subfamilies shared by angiosperms and gymnosperms, and even more in the Brassicaseae, forming the large G5 subfamily. In addition, during the evolution of angiosperms, most of the subfamilies have remained very stable, preserving (co-)orthologous relationships for many genes between eudicots and monocots. It is possible that the first expansion of KFBs had contributed to the evolution and success of land plants, with general conserved functions of most subfamilies in many cells and tissues of the angiosperms.  It is also possible that the more recent expansion of G5 in Brassicase*ae* has created many opportunities for further divergence and specialization of gene functions. Therefore, the KFB family exhibits both rapid expansion and stable maintenance of gene numbers, in different periods of evolution and in different subfamilies. This is a fascinating example of gene family evolution that should continue to yield insights into the evolution of gene family, gene function, and organisms.

## Materials and Methods

### Sequence retrieval and protein domain analysis

The known KFBs protein sequences from previous studies in *Arabidopsis* were downloaded from the *Arabidopsis* database (www.arabidopsis.org) (Andrade et al. 2001). Both genomic sequences and protein sequences of *Arabidopsis thaliana* (TIGR (The Institute for Genome Research) release version 5.0) and *Oryza sativa* (TIGR release version 4.0) were downloaded for local searches. The genomic sequences of *Brassica rapa* (www.arabidopsis.org), *Populus trichocarpa* (www.jgi.doe.gov, release version), *Physcomitrella patens* (www.jgi.doe.gov; access kindly granted by R. Quatrano; Quatrano et al. 2007), and the EST sequences of *Pinus taeda*

(TIGR) were also downloaded for local BLAST searches. To search for the plant KFBs, we used the protein sequences of all known KFBs (Andrade et al. 2001) as queries to carry out both the TBLASTN and BLASTP against the *Arabidopsis* genome with a cut off of E-value at 1e-5. All new sequences were then used as queries to carry out another round of BLAST searches. The process was repeated until no new sequences were obtained. The protein sequences that lack either the F-box domain or the Kelch motif based on the Pfam domain analysis (http://www.sanger.ac.uk/Software/Pfam/search.shtml) were eliminated. By choosing the cut off of E-value at 0.5 for both F-box domain and Kelch motif, we identified 97 KFBs in the *Arabidopsis* genome, five of which, At3g24610, At4g34170, At4g39560, At2g29860 and At2g20380, were modified from the prediction according to our multiple sequence alignment (see below). To search for the KFBs from several other plant species, the protein sequences of all *Arabidopsis* 97 KFBs were used as queries to carry out TBLASTN searches against the downloaded plant databases as mentioned above. We also carried out BLAST searches against the *Zea Mays* genome on the website www.plantgdb.org. The genomic sequences of the BLAST hits were then retrieved, and protein sequences were predicted based on sequence similarities.

Among the 68 human F-box proteins, only a single KFB called F-box 42 was detected previously (Jin et al. 2004). We used it as a query to carry out BLASTP, TBLASTN, and PSI-BLAST searches of both animal and fungi KFBs in the NCBI database.

All predicted KFB protein sequences collected in this study were examined using the Pfam domain analysis with the default cut off (http://www.sanger.ac.uk/Software/Pfam/). Sequences with only one kind of domain, either F-box domain or Kelch motif(s) were then analyzed individually with a cut off of E-value at 1.0. Finally, all the domain information was collected and the protein sequences with the E-values of F-box domain or the Kelch motifs of more than 0.5 were eliminated from the further analysis.

### Multiple sequences alignment

Multiple sequences alignment of all protein sequences was carried out by using Clustal X 1.83 with BLOSUM 30 as the protein weight matrix, and different values of Gap opening and Gap extension were tried. Finally, we chose the Gap opening value of 4.0 and the default value for Gap extension since they produced the best alignment results (Jeanmougin et al. 1998). The MUSCLE (version 3.52) software was also used to carry out the multiple sequence alignment to compare with the Clustal results (Edgar 2004). All sequences were then grouped into subgroups based on the preliminary NJ tree generated by MEGA 3.0 (Kumar et al. 1994). The protein sequences of each subgroup were aligned, and realigned between subgroups using the profile alignment in Clustal X. Alignments of all the protein sequences were finally adjusted manually using both alignments generated by MUSCLE and the results of Pfam domain analysis as the references. The amino acid sequences and alignment are available upon request.

### Phylogenetic analysis

Phylogenetic analyses were conducted by using both NJ and maximum likelihood (ML) methods. The NJ trees were generated by MEGA (3.0) with the "parewise deletion" option, "Poisson correction" model, and bootstrap of 1 000 replicates (Kumar et al. 1994; Guindon and Gascuel 2003). The ML trees were constructed using PHYML (version 2.4.4) with a bootstrap of 100 replicates, JTT (Jones, Taylor and Thornton) substitution model, and gamma distributed rates (determined by PHYML) (Kumar et al. 1994; Guindon and Gascuel 2003). ML tree files were then viewed and modified in MEGA. Only the NJ trees were presented in this study, with the bootstrap values from the analysis of both NJ and ML methods. Although the plant KFBs had a range of numbers of the Kelch motif and some plant KFBs had only one Kelch motif, if we used a cut off of E-value as 0.5, in most cases we saw at least another degenerated Kelch motif in the alignment. Finally, we used the sequences of the F-box domain and the first-two Kelch motifs for the phylogenetic analysis. For the subfamily of plant KFBs, additional regions may be used depending on the conservation of the protein sequence in a specific group.

### Chromosome distribution and the duplication types of AtKFBs and OsKFBs

To understand the mechanism of the gene duplication events of plant KFBs, we analyzed the chromosome distribution of the KFBs from *Arabidopsis* and rice and investigated possible duplication types of these genes. Three main types of gene duplication have been reported previously, including tandem duplication, segment duplication, and gene duplication caused by retrotransposition (Vision et al. 2000; Baumbusch et al. 2001; Cannon et al. 2004). If closely related genes are arrayed in tandem on the same chromosome, the duplication type is called tandem duplication. Large chromosomal blocks with syntenic distribution of similar genes provide evidence for segment duplication. For the retrotransposition type, the duplicated genes (also called retrogenes) normally lack intron, may have the stretches of poly(A) at the 3' end and short direct repeats at both ends, and are located on the different chromosome positions.

### Gene expression analysis of KFBs

The anthers (at anther stages 4–6) from *Arabidopsis* Landsberg *erecta* were collected under a dissection scope, and total RNA was then extracted from two biological anther samples using an RNeasy Plant Kit (Qiagen, Valencia, CA, USA). The kit was then used to carry out the microarray experiment as described previously (Zhang et al. 2005). The public microarray data of the other six tissues, including roots, stems, leaves, young inflorescences (stages 1–9), stage-12 flowers, and siliques, were provided by Zhang et al. (2005) in our lab. All the microarray data were analyzed and normalized to make the data comparable as described previously (Zhang et al. 2005). The Pearson's correlation coefficients for the two biological replicates of each of the seven tissues were all greater than 95%, indicating a very small variation between the two biological replicates. For simplicity, the average signal intensity values were used and presented for the gene expression here. The signal intensity value of 50 was used as a conservative cut off for reliable detection of gene expression as discussed in Zhang et al. (2005). To search for *OsKFB* ESTs, the genomic sequences of rice *KFB* were downloaded from TIGR and used as query sequences to search for the highly similar ESTs sequences (at least 95% identity) in the TIGR database.

## Acknowledgements

We thank Ralph Quatrano for sharing unpublished information on *Physcomitrella patens* genomic sequences, Hongzhi Kong for helpful discussion, and Guanfang Wang for comments on the manuscript.

## References

**Adams J, Kelso R, Cooley L** (2000). The kelch repeat superfamily of proteins: Propellers of cell function. *Trends Cell Biol.* **10**, 17–24.

**Andrade MA, Gonzalez-Guzman M, Serrano R, Rodriguez PL** (2001). A combination of the F-box motif and kelch repeats defines a large *Arabidopsis* family of F-box proteins. *Plant Mol. Biol.* **46**, 603–614.

**Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R et al.** (2001). The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res.* **29**, 4319–4333.

**Bork P, Doolittle RF** (1994). *Drosophila* kelch motif is derived from a common enzyme fold. *J. Mol. Biol.* **236**, 1277–1282.

**Cannon SB, Mitra A, Baumgarten A, Young ND, May G** (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* **4**, 10.

**del Pozo JC, Estelle M** (2000). F-box proteins and protein degradation: An emerging theme in cellular regulation. *Plant Mol. Biol.* **44**, 123–128.

**Dieterle M, Zhou YC, Schafer E, Funk M, Kretsch T** (2001). EID1, an F-box protein involved in phytochrome A-specific light signaling. *Genes Dev.* **15**, 939–944.

**Edgar RC** (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **5**, 113.

**Gagne JM, Downes BP, Shiu SH, Durski AM, Vierstra RD** (2002). The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **99**, 11519–11524.

**Guindon S, Gascuel O** (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704.

**Han L, Mason M, Risseeuw EP, Crosby WL, Somers DE** (2004). Formation of an SCF(ZTL) complex is required for proper regulation of circadian timing. *Plant J.* **40**, 291–301.

**Hershko A, Ciechanover A** (1998). The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479.

**Hong EJ, Villen J, Gerace EL, Gygi SP, Moazed D** (2005). A Cullin E3 ubiquitin ligase complex associates with Rik1 and the Clr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation. *RNA Biol.* **2**.

**Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA** (2005). FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in *Arabidopsis*. *Science* **309**, 293–297.

**Ingram GC, Goodrich J, Wilkinson MD, Simon R, Haughn GW, Coen ES** (1995). Parallels between *UNUSUAL FLORAL ORGANS* and *FIMBRIATA*, genes controlling flower development in Arabidopsis and *Antirrhinum*. *Plant Cell* **7**, 1501–1510.

**Ito N, Phillips SE, Stevens C, Ogel ZB, McPherson MJ, Keen JN et al.** (1991). Novel thioether bond revealed by a 1.7 A crystal structure of galactose oxidase. *Nature* **350**, 87–90.

**Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ** (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**, 403–405.

**Jin J, Cardozo T, Lovering RC, Elledge SJ, Pagano M, Harper JW** (2004). Systematic analysis and nomenclature of mammalian F-box proteins. *Genes Dev.* **18**, 2573–2580.

**Kamura T, Conrad MN, Yan Q, Conaway RC, Conaway JW** (1999). The Rbx1 subunit of SCF and VHL E3 ubiquitin ligase activates Rub1 modification of cullins Cdc53 and Cul2. *Genes Dev.* **13**, 2928–2933.

**Kim HS, Delaney TP** (2002). Arabidopsis SON1 is an F-box protein that regulates a novel induced defense response independent of both salicylic acid and systemic acquired resistance. *Plant Cell* **14**, 1469–1482.

**Kobayashi A, Kang MI, Okawa H, Ohtsuji M, Zenke Y, Chiba T et al.** (2004). Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2. *Mol. Cell Biol.* **24**, 7130–7139.

**Koepp DM, Schaefer LK, Ye X, Keyomarsi K, Chu C, Harper JW et**

**al.** (2001). Phosphorylation-dependent ubiquitination of cyclin E by the SCFFbw7 ubiquitin ligase. *Science* **294**, 173–177.

**Kong H, Leebens-Mack J, Ni W, dePamphilis CW, Ma H** (2004). Highly heterogeneous rates of evolution in the *SKP1* gene family in plants and animals: Functional and evolutionary implications. *Mol. Biol. Evol.* **21**, 117–128.

**Krek W** (1998). Proteolysis and the G1-S transition: The SCF connection. *Curr. Opin. Genet. Dev.* **8**, 36–42.

**Kumar S, Tamura K, Nei M** (1994). MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput. Appl. Biosci.* **10**, 189–191.

**Kuroda H, Takahashi N, Shimada H, Seki M, Shinozaki K, Matsui M** (2002). Classification and expression analysis of Arabidopsis F-box-containing protein genes. *Plant Cell Physiol.* **43**, 1073–1085.

**Lechner E, Achard P, Vansiri A, Potuschak T, Genschik P** (2006). F-box proteins everywhere. *Curr. Opin. Plant Biol.* **9**, 631–638.

**Li X, Zhang D, Hannink M, Beamer LJ** (2004). Crystal structure of the kelch domain of human Keap1. *J. Biol. Chem.* **279**, 54750–54758.

**Lyapina SA, Correll CC, Kipreos ET, Deshaies RJ** (1998). Human CUL1 forms an evolutionarily conserved ubiquitin ligase complex (SCF) with SKP1 and an F-box protein. *Proc. Natl. Acad. Sci. USA* **95**, 7451–7456.

**McGinnis KM, Thomas SG, Soule JD, Strader LC, Zale JM, Sun TP et al.** (2003). The Arabidopsis *SLEEPY1* gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. *Plant Cell* **15**, 1120–1130.

**Nelson DC, Lasswell J, Rogg LE, Cohen MA, Bartel B** (2000). *FKF1*, a clock-controlled gene that regulates the transition to flowering in *Arabidopsis*. *Cell* **101**, 331–340.

**Parry G, Estelle M** (2006). Auxin receptors: A new role for F-box proteins. *Curr. Opin. Cell Biol.* **18**, 152–156.

**Pickart CM** (2001). Ubiquitin enters the new millennium. *Mol. Cell* **8**, 499–504.

**Prag S, Adams JC** (2003). Molecular phylogeny of the kelch-repeat superfamily reveals an expansion of BTB/kelch proteins in animals. *BMC Bioinform.* **4**, 42.

**Quatrano RS, McDaniel SF, Khandelwal A, Perroud PF, Cove DJ** (2007). Physcomitrella patens: Mosses enter the genomic age. *Curr. Opin. Plant Biol.* **10**, 182–189.

**Risseeuw EP, Daskalchuk TE, Banks TW, Liu E, Cotelesage J, Hellmann H et al.** (2003). Protein interaction analysis of SCF ubiquitin E3 ligase subunits from *Arabidopsis*. *Plant J.* **34**, 753–767.

**Ruegger M, Dewey E, Gray WM, Hobbie L, Turner J, Estelle M** (1998). The TIR1 protein of *Arabidopsis* functions in auxin response and is related to human SKP2 and yeast grr1p. *Genes Dev.* **12**, 198–207.

**Samach A, Klenz JE, Kohalmi SE, Risseeuw E, Haughn GW, Crosby WL** (1999). The *UNUSUAL FLORAL ORGANS* gene of *Arabidopsis thaliana* is an F-box protein required for normal patterning and growth in the floral meristem. *Plant J.* **20**, 433–445.

**Skowyra D, Koepp DM, Kamura T, Conrad MN, Conaway RC, Conaway JW et al.** (1999). Reconstitution of G1 cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1. *Science* **284**, 662–665.

**Somers DE, Kim WY, Geng R** (2004). The F-box protein ZEITLUPE confers dosage-dependent control on the circadian clock, photomorphogenesis, and flowering time. *Plant Cell* **16**, 769–782.

**Somers DE, Schultz TF, Milnamow M, Kay SA** (2000). *ZEITLUPE* encodes a novel clock-associated PAS protein from *Arabidopsis*. *Cell* **101**, 319–329.

**Stirnberg P, van de Sande K, Leyser HM** (2002). MAX1 and MAX2 control shoot lateral branching in *Arabidopsis*. *Development* **129**, 1131–1141.

**Takahashi N, Kuroda H, Kuromori T, Hirayama T, Seki M, Shinozaki K et al.** (2004). Expression and interaction analysis of *Arabidopsis Skp1*-related genes. *Plant Cell Physiol.* **45**, 83–91.

**Vision TJ, Brown DG, Tanksley SD** (2000). The origins of genomic duplications in *Arabidopsis*. *Science* **290**, 2114–2117.

**Willems AR, Schwab M, Tyers M** (2004). A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. *Biochim. Biophys. Acta* **1695**, 133–170.

**Woo HR, Chung KM, Park JH, Oh SA, Ahn T, Hong SH et al.** (2001). ORE9, an F-box protein that regulates leaf senescence in Arabidopsis. *Plant Cell* **13**, 1779–1790.

**Xie DX, Feys BF, James S, Nieto-Rostro M, Turner JG** (1998). *COI1*: An *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* **280**, 1091–1094.

**Xue F, Cooley L** (1993). Kelch encodes a component of intercellular bridges in *Drosophila* egg chambers. *Cell* **72**, 681–693.

**Yamanaka A, Yada M, Imaki H, Koga M, Ohshima Y, Nakayama K** (2002). Multiple Skp1-related proteins in *Caenorhabditis elegans*: Diverse patterns of interaction with Cullins and F-box proteins. *Curr. Biol.* **12**, 267–275.

**Yasuhara M, Mitsui S, Hirano H, Takanabe R, Tokioka Y, Ihara N et al.** (2004). Identification of ASK and clock-associated proteins as molecular partners of LKP2 (LOV kelch protein 2) in *Arabidopsis*. *J. Exp. Bot.* **55**, 2015–2027.

**Zhang X, Feng B, Zhang Q, Zhang D, Altman N, Ma H** (2005). Genome-wide expression profiling and identification of gene activities during early flower development in Arabidopsis. *Plant Mol. Biol.* **58**, 401–419.

**Zhao D, Yu Q, Chen M, Ma H** (2001). The *ASK1* gene regulates B function gene expression in cooperation with *UFO* and *LEAFY* in Arabidopsis. *Development* **128**, 2735–2746.

**Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P et al.** (2002). Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416**, 703–709.

(Handling editor: Bin Han)