The Plant Journal (2007) **50**, 873–885 doi: 10.1111/j.1365-313X.2007.03097.x

Patterns of gene duplication in the plant *SKP1* gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth

Hongzhi Kong[1,*], Lena L. Landherr[2], Michael W. Frohlich[3], Jim Leebens-Mack[2,†], Hong Ma[2] and Claude W. dePamphilis[2,*]
[1]*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, The Chinese Academy of Sciences, Beijing 100093, China,*
[2]*Department of Biology, The Huck Institutes of the Life Sciences, and Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, PA 16802, USA, and*
[3]*Department of Botany, Natural History Museum, London, SW7 5BD, UK*

Received 22 November 2006; revised 7 February 2007; accepted 14 February 2007.
[*]For correspondence (fax +86 10 6259 0843; email hzkong@ibcas.ac.cn or fax +1 814 863 1357; email cwd3@psu.edu).
[†]Present address: Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

Summary

Gene duplication plays important roles in organismal evolution, because duplicate genes provide raw materials for the evolution of mechanisms controlling physiological and/or morphological novelties. Gene duplication can occur via several mechanisms, including segmental duplication, tandem duplication and retroposition. Although segmental and tandem duplications have been found to be important for the expansion of a number of multigene families, the contribution of retroposition is not clear. Here we show that plant *SKP1* genes have evolved by multiple duplication events from a single ancestral copy in the most recent common ancestor (MRCA) of eudicots and monocots, resulting in 19 *ASK* (Arabidopsis *SKP1-like*) and 28 *OSK* (*Oryza SKP1-like*) genes. The estimated birth rates are more than ten times the average rate of gene duplication, and are even higher than that of other rapidly duplicating plant genes, such as type I MADS box genes, R genes, and genes encoding receptor-like kinases. Further analyses suggest that a relatively large proportion of the duplication events may be explained by tandem duplication, but few, if any, are likely to be due to segmental duplication. In addition, by mapping the gain/loss of a specific intron on gene phylogenies, and by searching for the features that characterize retrogenes/retrosequences, we show that retroposition is an important mechanism for expansion of the plant *SKP1* gene family. Specifically, we propose that two and three ancient retroposition events occurred in lineages leading to Arabidopsis and rice, respectively, followed by repeated tandem duplications and chromosome rearrangements. Our study represents a thorough investigation showing that retroposition can play an important role in the evolution of a plant gene family whose members do not encode mobile elements.

Keywords: plant *SKP1* genes, birth-and-death evolution, gene duplication, tandem duplication, segmental duplication, retroposition.

Introduction

Gene duplication events are important for gene family evolution, because duplicate genes provide the raw materials for the evolution of new gene functions, which in turn facilitate the generation of structural and functional novelties. It is already known that, at least in eukaryotes, a large proportion of genes are members of multi-gene families generated by gene duplications (Horan *et al.*, 2005; Maere *et al.*, 2005; Zhang, 2003). Gene duplications may arise through three principal mechanisms: (i) segmental duplication (of the whole genome, of one to a few chromosomes, or of large parts of a chromosome), (ii) tandem duplications (of one to a few adjacent genes), or (iii) retroposition (and other transposition events). Of the three main types, segmental duplication occurs most frequently in plants because most plant

species are diploidized polyploids and retain numerous duplicated chromosomal blocks within their genomes (Adams and Wendel, 2005). These duplicate blocks, although highly degenerate and distributed on different chromosomes or different parts of the same chromosome, are of great importance for the elucidation of the polyploid past of a diploid genome (Blanc and Wolfe, 2004; Blanc *et al.*, 2000, 2003; Bowers *et al.*, 2003; Raes *et al.*, 2003; Simillion *et al.*, 2002; Vandepoele *et al.*, 2003; Vision *et al.*, 2000). Unlike segmental duplication, tandem (or local) duplication often results from unequal crossing-over and usually generates tandemly arrayed gene copies. In many cases, these tandemly arrayed copies are very similar to each other due to their recent origins and/or frequent gene conversion(s) among them. Retroposition, which occurs when an mRNA of an expressed gene is reverse-transcribed to a cDNA and then inserted into the genome, usually creates an unlinked gene, because the insertion of cDNA into the genome is more or less random (Graur and Li, 2000; Zhang, 2003).

Several studies have evaluated the relative contributions of the three main types of gene duplication in the generation of new members in nuclear gene families (Cannon *et al.*, 2004; Leister, 2004; Nakano *et al.*, 2006; Sampedro *et al.*, 2005; Tian *et al.*, 2004). By analyzing the complete genomes of model species, much has been learned about the relative contributions of tandem and segmental duplications. In *Arabidopsis thaliana*, for example, Cannon *et al.* (2004) reported the analyses of 50 gene families and found that tandem duplication is most prominent in some gene families, whereas segmental duplications occurred more frequently in others. Interestingly, these two modes of duplication were not observed to be simultaneously important in the same gene family (Cannon *et al.*, 2004). However, retroposition was not observed as a major mechanism for any of these gene families (Cannon *et al.*, 2004), although a considerable number of retroposed genes have been identified in plant and animal genomes (Benovoy and Drouin, 2006; Choo *et al.*, 2007; Javaud *et al.*, 2003; Marques *et al.*, 2005; Nicholson *et al.*, 2005; Pavlicek *et al.*, 2006; Strichman-Almashanu *et al.*, 2003; Vinckenbosch *et al.*, 2006; Wang *et al.*, 2006). In addition, because retroposed genes usually lack regulatory regions, it is generally believed that most retrogenes are non-functional, and that retroposition plays a very minor role in the expansion of gene families (Graur and Li, 2000).

Skp1 (S-phase kinase-associated protein 1) is a small protein of approximately 160 amino acids. As a core component of the SCF-type E3 ubiquitin ligases that mediate protein degradation by the 26S proteosome, Skp1 plays key roles in cell-cycle progression, transcriptional regulation, signal transduction, and many other cellular processes in eukaryotes (Hellmann and Estelle, 2002). Mutations of the *SKP1* gene in the budding yeast *Saccharomyces cerevisiae* cause defects in cell-cycle progression at both G_1/S and G_2/M transitions (Bai *et al.*, 1996; Connelly and Hieter, 1996). In *Caenorhabditis elegans*, the *SKP1-Related* genes (*SKR*s) are involved in posterior body morphogenesis, embryonic and larval development, and cell proliferation (Nayak *et al.*, 2002; Yamanaka *et al.*, 2002). In plants, *SKP1*-like genes have been shown to be important for auxin, gibberellin (GA), ethylene, jasmonate and light responses (Gray *et al.*, 1999; Guo and Ecker, 2003; Hobbie, 2005; Parry and Estelle, 2006; Xu *et al.*, 2002; Zhou *et al.*, 2002), vegetative and flower development (Liu *et al.*, 2004; Ni *et al.*, 2004; Porat *et al.*, 1998; Zhao *et al.*, 1999, 2001, 2003b), and male meiosis (Drouaud *et al.*, 2000; Wang and Yang, 2005; Yang *et al.*, 1999, 2006; Zhao *et al.*, 2003a, 2006). In most of these cases, Skp1 functions by forming one of the numerous SCF complexes with Cullin, Rbx1 and an F-box protein. Within the SCF complexes, the scaffold-like Cullin forms a core catalytic complex with Rbx1, an F-box protein functions to recognize target proteins, and Skp1 serves to link the variable F-box protein to Cullin (Zheng *et al.*, 2002).

There is only one *SKP1* gene in protists , algae, fungi and vertebrate animals (Kong *et al.*, 2004). Nevertheless, the single Skp1 protein can interact with many F-box proteins to ubiquitinate different substrates (Schulman *et al.*, 2000). In invertebrate animals and vascular plants, however, there are multiple *SKP1* genes that have evolved at highly heterogeneous rates (Farras *et al.*, 2001; Kong *et al.*, 2004; Nayak *et al.*, 2002; Yamanaka *et al.*, 2002). The slowly evolving *SKP1* genes are highly similar in sequence and are expressed widely and at high levels, suggesting that they perform critical functions. Where tested, mutation of these genes causes severe defects in vegetative and/or reproductive development (Nayak *et al.*, 2002; Yamanaka *et al.*, 2002; Yang *et al.*, 1999; Zhao *et al.*, 2001, 2003a). The rapidly evolving *SKP1* genes, which are less conserved in sequence, usually have very limited expression patterns and may serve minor or redundant functions. Some rapidly evolving *SKP1* paralogs may have lost the ability to form SCF complexes because key residues for them to interact with other SCF components have been replaced by amino acids with quite different properties (Kong *et al.*, 2004; Takahashi *et al.*, 2004; Zhao *et al.*, 2003b).

Previous studies have also suggested that all *SKP1* genes found in the genomes of *Arabidopsis thaliana* and *Oryza sativa* (rice) are derived from a single ancestral gene in the most recent common ancestor (MRCA) of these two species (Kong *et al.*, 2004). If this hypothesis is true, how have these two species acquired so many novel genes (18 and 27 in Arabidopsis and rice, respectively; see below) after the divergence of eudicots and monocots? Did they arise by segmental duplication, tandem duplication and/or retroposition? Here, we present evidence that evolution of the plant *SKP1* gene family is a rapid birth-and-death process. The birth rate, which is much higher than the death rate, is also

much greater than that observed in many other plant gene families. In addition, unlike the situation documented for other plant gene families, we found that tandem duplication and retroposition, rather than segmental duplication, played dominant roles in expansion of the *SKP1* gene family. Our study presents a good example showing that retroposition can be a main contributor to the expansion of a plant gene family whose members are not transposable elements.

Results and discussion

SKP1-*like genes*

The Arabidopsis genome encodes 21 *SKP1* genes (*ASK*s, *Arabidopsis thaliana SKP1*-like genes). Among these, 19 are type I and two are type II genes (*ASK20* and *ASK21*), which are much longer than type I genes and encode chimeric proteins (Figure 1). In rice, there are at least 32 full-length *SKP1* genes (*OSK*s, *Oryza sativa SKP1*-like genes), of which 28 are type I genes. Phylogenetic analyses further indicate that the evolutionary histories of type II genes are quite different from those of type I genes (data not shown). Particularly, due to the effect of long-branch attraction, the inclusion of both types usually gives unstable results. For this reason, type II genes were excluded from further studies. In addition to genes from Arabidopsis and rice, several hundreds of type I *SKP1*-like genes were identified using BLAST searches against the TIGR plant gene indices (Lee *et al.*, 2005) and PlantTribes databases (http://www.floralgenome.org/tribe.php; see Experimental procedures). In this paper, the *Glycine max* (soybean), *Populus trichocarpa* (poplar), *Brassica napus* (rape), *Zea mays* (maize) and *Triticum aestivum* (wheat) *SKP1-like* genes are named *GSK*, *PSK*, *BSK*, *ZSK* and *TSK* genes, respectively (Figure 2).

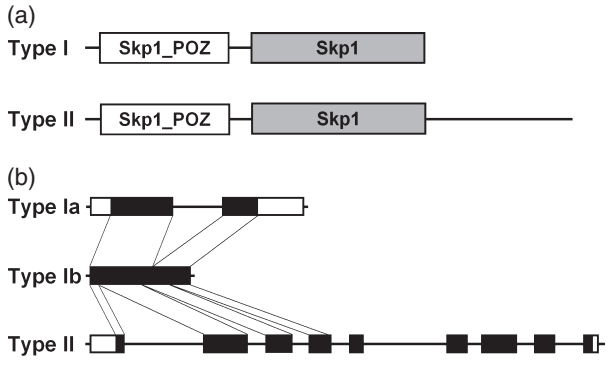

Figure 1. Schematic diagram of plant *SKP1* genes and their proteins.
(a) Two types of proteins and (b) three types of genes. Type I proteins, coded by type Ia and type Ib genes, have two conserved domains (i.e. Skp1_POZ and Skp1) and two variable regions. Type II proteins, coded by type II genes, are chimeric because an additional C-terminal region has been appended to the type I proteins. The single intron that characterizes type Ia genes is absent in type Ib genes. Type II genes, however, usually have multiple introns at various positions.

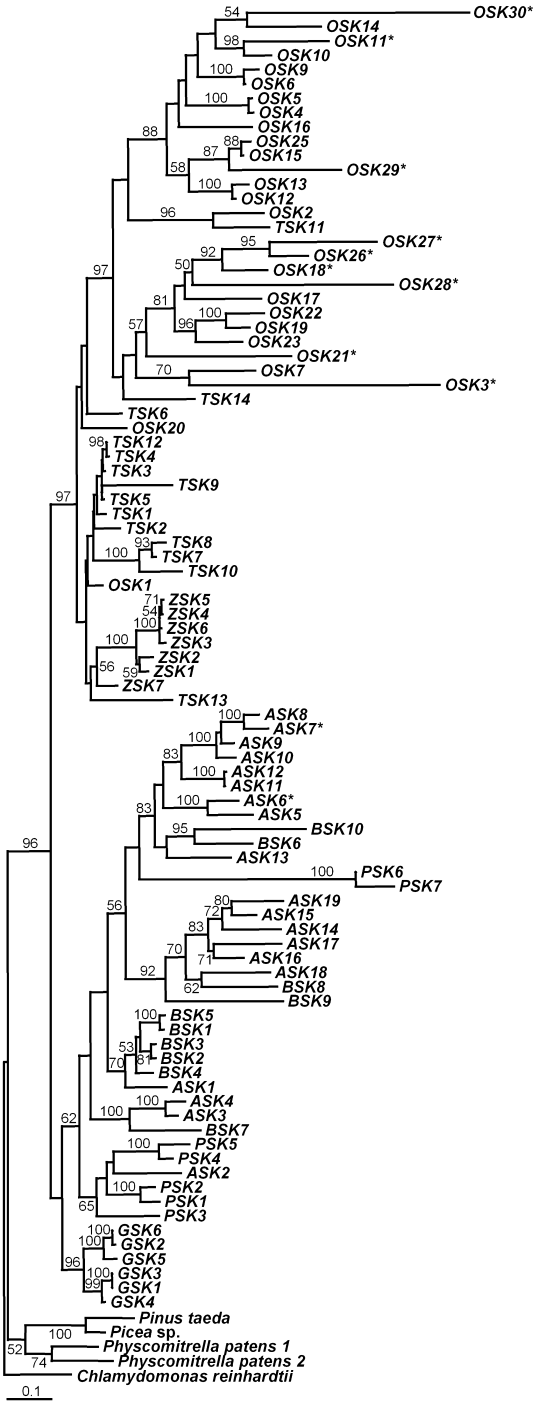

Figure 2. Phylogenetic relationships of 91 type I *SKP1* homologs from Arabidopsis, rice, soybean, rape, poplar, maize and wheat.
This ML tree was constructed by the PHYML program for the 447 bp Skp_POZ and Skp1 domain regions. The numbers for each interior branch are the percentage bootstrap values (1000 re-samplings), and only values higher than 50% are shown. Genes are abbreviated as: *ASK*, *Arabidopsis thaliana SKP1-like*; *BSK*, *Brassica rape SKP1-like*; *GSK*, *Glycine max SKP1-like*; *OSK*, *Oryza sativa SKP1-like*; *PSK*, *Populus trichocarpa SKP1-like*; *TSK*, *Triticum aestivum SKP1-like*; *ZSK*, *Zea mays SKP1-like*. Full-length genes whose protein sequences lack the structurally conserved α-helices and/or β-sheets are regarded as putative pseudogenes and are marked with asterisks.

In addition to full-length genes, partial sequences that match at least one of the aforementioned *ASK* and *OSK* genes were found in both Arabidopsis and rice genomes (Figures 3 and 4). However, as the proteins of the partial sequences lack regions corresponding to one or a few functionally important α-helices and/or β-strands, these partial sequences were regarded as pseudogenes and were excluded from further studies. Moreover, among the 19 and 28 full-length *ASK* and *OSK* genes, respectively, two and nine genes seem to have an out-of-frame insertion and/or deletion of one to a few nucleotides, making one or both of two conserved domains (Skp1_POZ and Skp1) no longer recognizable. In this study, these genes were regarded as putative pseudogenes but were not excluded from further analysis.

A typical type I *SKP1* gene contains two exons and one intron (Risseeuw *et al.*, 2003; type Ia, Figure 1b). As the position of the intron is conserved in slowly evolving genes from Arabidopsis and rice (i.e. *ASK1*, *2*, *3* and *4*, and *OSK1*, *2* and *20*; Figures 3b and 4b), as well as several *SKP1* genes from other angiosperms, gymnosperms, ferns, mosses and algae, it is likely that the presence of the intron represents a plesiomorphic or ancestral character state (Kong *et al.*, 2004). However, in plant species with multiple *SKP1* genes, some *SKP1* genes lack this intron (type Ib, Figure 1b). However, some other genes (such as *ASK15* and *OSK11*, *29* and *30*) each contain an intron but at quite different locations, suggesting that they gained the intron independently during evolution.

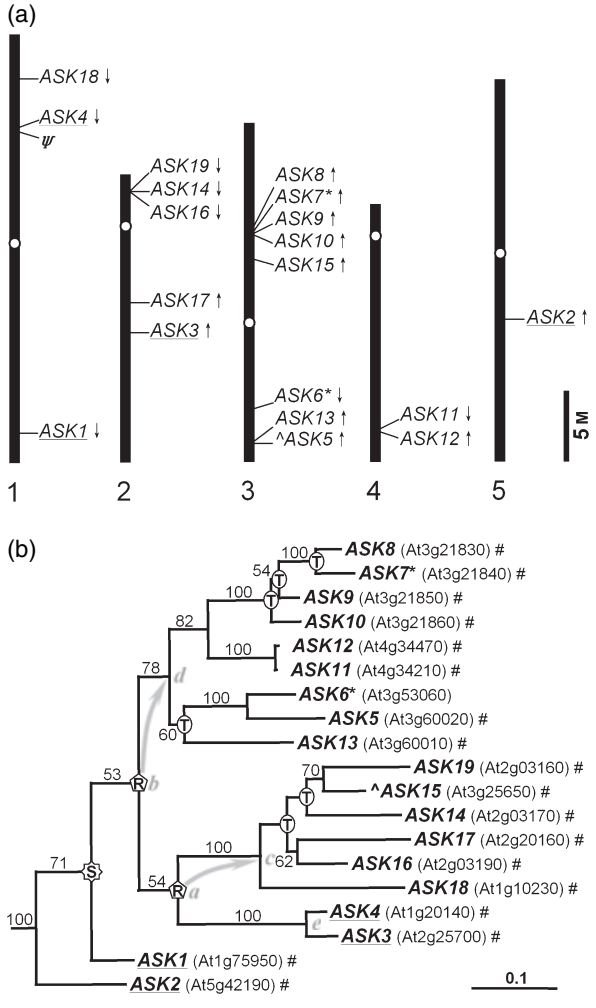

Figure 3. Evolution of the Arabidopsis *SKP1-like* genes (*ASK*s).
(a) Chromosomal locations and (b) phylogenetic relationships. The orientations of each gene are shown by arrows. Locus names are shown in parentheses. Genes with the specific intron are underlined, while those with different, late-gained introns are marked with a '^' symbol. Expressed genes are labeled with a '#' symbol. Real pseudogenes (i.e. genes with partial sequences) are highlighted with a 'ψ' symbol, and putative pseudogenes are labeled with an asterisk. The letters T, S and R on the nodes of the phylogenetic tree indicate the positions where tandem duplication, segmental duplication and retroposition have occurred, respectively. Hypothesized donor genes and retrogenes during retroposition are connected by curved arrows.

Rapid birth-and-death evolution of plant SKP1 *genes*

Our previous results (Kong *et al.*, 2004) suggest that all of the Arabidopsis and rice genes are derived from a single ancestral gene in the MRCA of the two species. To provide further support for this hypothesis, we performed additional analysis and found that the *ASK* and *OSK* genes form two species-specific clades (Figure S1). Therefore, Arabidopsis and rice are likely to have gained at least 18 and 27 novel *SKP1* genes, respectively, since their divergence 140–145 million years ago (MYA; Anderson *et al.*, 2005; Davies *et al.*, 2004). If this is true, then the birth rate of *SKP1* genes (counting only surviving copies) is at least 12.4 and 18.6 genes per 100 million years (MY) per ancestral gene in lineages leading to Arabidopsis and rice, respectively. Both of these rates are more than ten times the average rate of gene duplication. Lynch and Conery (2000) estimated that gene duplication arises at an approximate rate of 1 gene per 100 MY per ancestral gene in eukaryotes such as *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae* and *Arabidopsis thaliana*. In fact, the birth rate of *SKP1* genes is also greater than that observed for many other plant genes, such as MADS box (Nam *et al.*, 2004), disease resistance (R) loci (Michelmore and Meyers, 1998), and receptor-like kinases (Shiu *et al.*, 2004). There are 101 and 30 type I (both functional and non-functional) and 47 and 48 type II MADS box genes in the Arabidopsis and rice genomes, respectively, and the numbers of type I and type II genes in the MRCA of the two species were estimated to be 15–20 and 4–8, respectively (Nam *et al.*, 2004). Accordingly, the birth rate of type I MADS box genes is about 3.5–4.6 genes per 100 MY per ancestral gene in Arabidopsis and 1.0–1.4 genes per 100 MY per ancestral gene in rice, whereas that of type II genes is about 4.1–8.1 genes per 100 MY per ancestral gene in Arabidopsis and 4.1–8.3 genes per 100 MY per ancestral gene in rice.

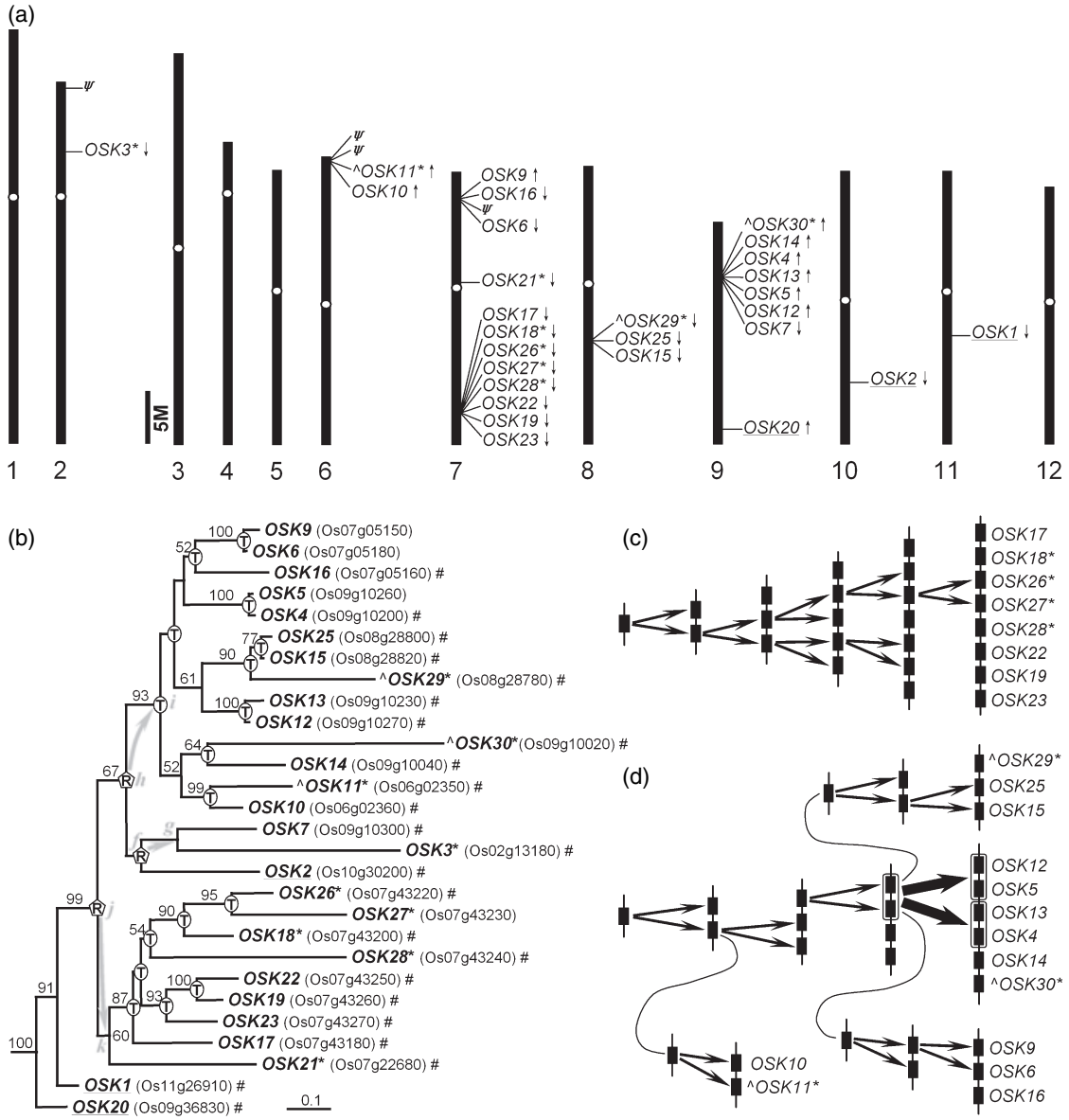

Figure 4. Evolution of the *Oryza SKP1-like* genes (*OSK*s).
(a) Chromosomal locations, (b) phylogenetic relationships, (c) hypothetical origins of eight *OSK* genes by tandem duplication, and (d) hypothetical origins of 14 *OSK* genes by tandem duplication and, most likely, retroposition. Symbols are the same as in Figure 3.

To determine whether *SKP1* genes from the same species always cluster together, genes from several other model species were added to the phylogenetic analysis (Figure 2 and Figure S2). In addition, because the split between Arabidopsis and rice is very ancient, addition of the sequences from such species as the eudicots soybean, poplar and rape, and the monocots wheat and maize may allow a step-by-step comparison and reduce the effects (if any) of long-branch attraction. From Figure 2, it is clear that genes from eudicots and monocots also form two separate clades, suggesting that they are indeed derived from a single ancestral gene in the MRCA of eudicots and monocots. The consistency in relationships among the *ASK* and *OSK* genes further implies that long-branch attraction was not a significant problem in the phylogenetic analysis of the Arabidopsis and rice genes. In addition, we found that the newly added genes, although small in number from each species due to incomplete genomic information, tend to form species-specific clusters, particularly when distantly related species are compared. For example, the six soybean *GSK* genes form the most basal cluster within the eudicot clade, followed by a cluster formed by *ASK2* and five poplar *PSK* genes (Figure 2). In the clade containing all Brassicaceae genes, there is some intermingling of Arabidopsis and rape

genes, but most genes form species-specific clusters (Figure 2 and Figure S2). This suggests that while orthologous relationships can be recognized for some genes from poplar, rape and Arabidopsis, the six soybean genes do not have strict orthologs in Arabidopsis. In other words, the soybean genes may all have been derived from a series of gene duplication events that occurred after the split of the eurosid I and eurosid II lineages some 110 MYA (Davies *et al.*, 2004). A similar situation was found in the well-supported clade of the grass genes, in which all the maize genes, as well as most wheat genes, form a species-specific clade (Figure 2 and Figure S2), suggesting that many *SKP1* gene duplications occurred after the divergence of lineages leading to these two grass genera some 55 MYA (Kellogg, 2001).

It should be mentioned that, although *SKP1* genes from the Arabidopsis and rice genomes were descendants of a single ancestral gene, the MRCA of the two species may have had more than one gene; however, due to frequent gene death, descendents of all but one gene were deleted from each genome. This can be seen from the observations that several moss and gymnosperm species (such as *Physcomitrella patens* and *Pinus taeda*) have multiple, closely related *SKP1* genes (Kong *et al.*, 2004; Soltis *et al.*, 2005). In addition, our estimate for the rate of gene birth in the plant *SKP1* gene family is still conservative for two reasons. First, the actual numbers of duplicate genes that once existed in the Arabidopsis and rice genomes could be greater than we can observe nowadays, as some duplicate copies might have already been deleted from the genome. The presence of recognizable partial Arabidopsis and rice *SKP1*-like sequences (Figures 3a and 4a) hints that they are divergent copies and that some duplicated copies might be too divergent to be recognized now. Second, because genes from each species form a single clade in the phylogenetic trees, the MRCAs of the genes from each species could be much younger than the MRCA of the genes from the two species. Indeed, the rice genes seem to have more recent origins than the Arabidopsis genes, because the largest d_S value (synonymous changes per synonymous site) for rice genes (0.640, between *OSK1* and *OSK29*) is only about two-thirds of the largest d_S value estimated for Arabidopsis genes (0.936, between *ASK2* and *ASK11*).

Are there any biological needs or advantages for plant *SKP1* genes to evolve this way? To answer this question, first we need to consider the function of the Skp1 proteins. Within the SCF complexes, Skp1 acts as an important adaptor that links the variable F-box proteins to Cullin and Rbx1 (Schulman *et al.*, 2000; Zheng *et al.*, 2002). While Cullin and Rbx1 are relatively non-specific, the F-box proteins are very diverse and act to specify the substrates for ubiquitination (Cardozo and Pagano, 2004). In the Arabidopsis genome, the number of F-box proteins has been estimated as approximately 700, and several recent studies have shown that the F-box domains of F-box proteins have accumulated considerable differences during evolution so that several distinct groups are recognizable (Gagne *et al.*, 2002; Kuroda *et al.*, 2002; Risseeuw *et al.*, 2003). This implies that some Skp1 genes might have also evolved accordingly, otherwise the interactions between Skp1 proteins and the F-box domains of F-box proteins could not be maintained. In fact, yeast hybrid assays have already indicated that the ability of Skp1 proteins to interact with F-box proteins varies considerably. Some ASK proteins (e.g. ASK1, 2, 11, 12 and 13) are able to interact with a wide spectrum of F-box proteins, while others (e.g. ASK3, 4 and 16) only interact with a few types of F-box proteins (Gagne *et al.*, 2002; Risseeuw *et al.*, 2003; Takahashi *et al.*, 2004). There are a few ASK proteins (e.g. ASK6, 10 and 17) that can not interact with any F-box proteins so far examined. These results, although quite preliminary, indicate that plant *SKP1* genes may have diversified considerably to interact with different types of F-box proteins. Interestingly, the F-box protein family also undergoes a rapid birth-and-death evolution in plants (Thomas, 2006).

Relatively large contribution of tandem duplication to the birth of new genes

It has been suggested that the Arabidopsis genome experienced three duplication events (referred to as 1R, 2R and 3R) within the past 250 MY, with 2R and 3R having occurred after the eudicot–monocot divergence (Arabidopsis Genome Initiative, 2000; Blanc *et al.*, 2003; Bowers *et al.*, 2003; Simillion *et al.*, 2002; Vision *et al.*, 2000). Similarly, the rice genome is believed to have experienced a genome-wide duplication approximately 70 MYA, before the diversification of the Poales but after the divergence of the Poales from the Liliales and Zingiberales (Goff *et al.*, 2002; Paterson *et al.*, 2004; Vandepoele *et al.*, 2003; Yu *et al.*, 2002). Therefore, if there were only one ancestral gene in the MRCA of Arabidopsis and rice, genome-scale segmental duplication could at most account for four and two novel genes in Arabidopsis and rice, respectively. However, the observation that both species have much greater numbers of *SKP1* genes suggests that other mechanisms, such as small-scale segmental duplication, tandem duplication and/or retroposition must have also contributed to the expansion of the *SKP1* gene family.

To search for evidence of additional duplication mechanisms for *SKP1* genes, we examined the genomic distribution of the *ASK* and *OSK* genes. As expected, we found that some *ASK* and *OSK* genes are clustered together at the same chromosome location (Figures 3a and 4a), suggesting that they were the results of tandem duplications, and that tandem duplication probably accounts for the generation of at least six and 17 genes in Arabidopsis and rice, respectively (Figures 3b and 4b). In Arabidopsis, the most obvious

example is the tandem repeat formed by *ASK8*, *7*, *9* and *10* on chromosome 3 (Figure 3a). Phylogenetically, these genes form a well-supported terminal clade (Figure 3b), suggesting that they are the results of three recent tandem duplication events. Other examples that may be the results of tandem duplications include *ASK5* and *13*, and *ASK19*, *14* and *16*. However, the evolutionary histories of these genes are still uncertain, because the clade formed by them also contains genes from other locations (e.g. *ASK6*, *15* and *17*) (Figure 3). In rice, the largest *OSK* gene cluster is located on chromosome 7 and contains eight tandemly arrayed members, i.e. *OSK17*, *18*, *26*, *27*, *28*, *22*, *19* and *23* (Figure 4a). Phylogenetically, these eight genes form a single clade (Figure 4b,c), suggesting that they also resulted from recent tandem duplications. The second largest *OSK* gene cluster is on chromosome 9 and contains six tandemly arrayed genes, i.e. *OSK12*, *5*, *13*, *4*, *14* and *30* (Figure 4a). However, these genes may be results of more ancient tandem duplication events, because the clade formed by them also contains genes from other locations: *OSK29*, *25* and *15* are located on chromosome 8, whereas *OSK9*, *16* and *6* are on chromosome 7 (Figure 4b). Interestingly, we found that the *OSK12*/*5*/*13*/*4* cluster may have been generated from a two-gene cluster through a single tandem duplication event (Figure 4d).

It is worth noting that there seems to be a correlation between tandem gene duplication and pseudogenization, because many of the tandemly arrayed genes possess characteristics of pseudogenes. In the aforementioned eight-gene cluster in rice, for example, there are four putative pseudogenes, *OSK18*, *26*, *27* and *28*. In fact, among the three (two full-length and one partial) and 13 (nine full-length and four partial) pseudogenes in Arabidopsis and rice, respectively, two (66.7%) and ten (76.9%) are in tandemly arrayed gene clusters with their relatives. Sequence comparison and phylogenetic analysis further suggest that these (putative) pseudogenes were derived from their upstream or downstream functional genes by tandem duplications: *ASK7* was from *ASK8*, and *OSK11*, *OSK18*/*26*/*27*/*28*, *OSK29* and *OSK30* were from *OSK10*, *OSK22*/*19*/*23*, *OSK15*/*25* and *OSK14*, respectively (Figures 3 and 4).

Limited role played by segmental duplication in the generation of new genes

To understand the contribution of segmental duplication in the increase of gene number, we searched in the Arabidopsis and rice genomes for chromosomal segments (or duplicate blocks) that contain *SKP1*-like genes. For this purpose, tandemly arrayed genes were treated as a single gene copy. We found that, in Arabidopsis, 11 previously identified duplicate blocks contain *ASK* genes (Table 1). One block (At1g72180–At1g78270) contains *ASK1* (At1g75950), while its duplicate block (At1g17230–At1g22340) includes *ASK4* (At1g20140) at the same position (Figure 5a). This suggests that *ASK1* and *ASK4* might be the results of a

Table 1 Duplicate blocks in the Arabidopsis and rice genomes that contain SKP1 genes

Arabidopsis thaliana			*Oryza sativa* ssp. *japonica*	
Duplicate block I *SKP1-like* gene	Duplicate block II *SKP1-like* gene	Estimated age (MY)	Duplicate block I *SKP1-like* gene	Duplicate block II *SKP1-like* gene
At4g34410-At4g34530 *ASK12* (At4g34470)	At1g25320-At1g25470 –	236.44	8358.m00291-m03565 *OSK1* (8358.m01943; Os11g26910)	8359.m00273-m03539 –
At1g75280-At1g75840 *ASK1* (At1g75950)	At4g34540-At1g35020 –	204.54	8362.m01574-m02528 *OSK2* (8362.m02270; Os10g30200)	8356.m00804-m01463 –
At1g75250-At1g76350 *ASK1* (At1g75950)	At4g38340-At4g39250 –	204.54	8362.m00908-m02520 *OSK2* (8362.m02270; Os10g30200)	8359.m00965-m04145 –
At2g02370-At2g04310 *ASK19/14/16* (At3g03160/70/90)	At1g25784-At1g27000 –	197.00	8351.m00406-m03778 *OSK3* (8531.m01149; Os0213180)	8352.m00504-m03896 –
At3g59540-At3g62870 *ASK5/13* (At3g60010/20)	At2g43460-At2g47800 –	70.08	8351.m00835-m04007 *OSK3* (8531.m01149; Os0213180)	8352.m03517-m04141 –
At2g19810-At2g20900 *ASK17* (At2g20160)	At4g28120-At4g29190 –	69.91	8351.m00843-m03401 *OSK3* (8531.m01149; Os0213180)	8354.m03963-m03978 –
At3g21465-At3g23870 *ASK8/7/9/10* (At3g21830/40/50/60)	At4g13800-At4g15640 –	69.42	8355.m00673-m03956 *OSK21* (8355.m02007; Os07g22680)	8360.m02697-m05209 –
At1g72180-At1g78270 *ASK1* (At1g75950)	At1g17230-At1g22340 *ASK4* (At1g20140)	68.82	8355.m01169-m03875 *OSK21* (8355.m02007; Os07g22680)	8360.m05420-m05462 –
At2g01090-At2g04038 *ASK19/14/16* (At3g03160/70/90	At1g13600-At1g15120 –	67.55	8356.m02270-m04230 *OSK29/25/15* (8356.m02725/27/29; Os08g28780/800/820)	8357.m01943-m02960 –
At1g08970-At1g10570 *ASK18* (At1g10230)	At1g56170-At1g60220 –	64.94		
At3g52340-At3g54100 *ASK6* (At3g53060)	At2g35840-At2g37980 –	63.15		

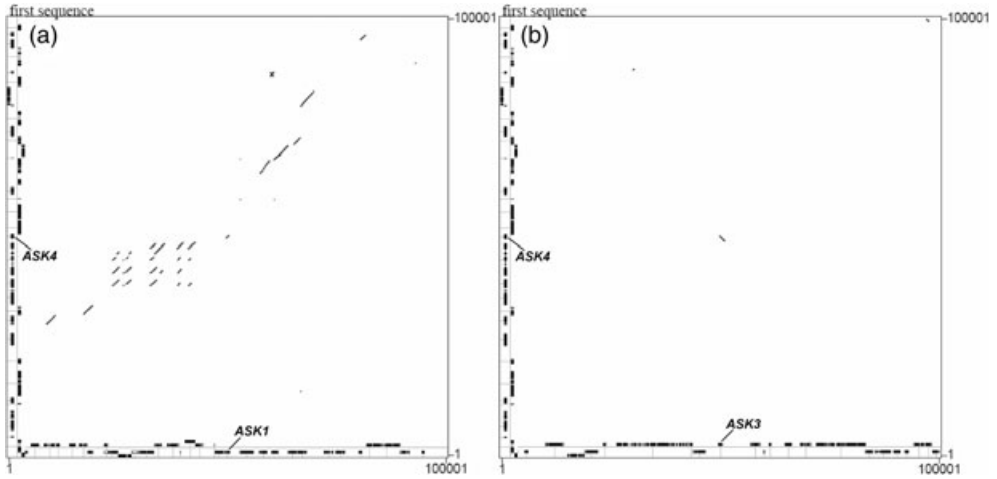

Figure 5. Comparisons of the flanking genomic regions (50 kb at both sides) of selected *ASK* gene pairs.
(a) *ASK1* versus *ASK4*, and (b) *ASK3* versus *ASK4*. *ASK1* and *ASK4* can be explained by a segmental duplication event, whereas *ASK3* and *ASK4* cannot. For more information, see Figure S3.

single large-scale segmental duplication event. However, none of the other *ASK* gene-containing blocks has an *ASK* gene in its duplicate block. This suggests that, after segmental duplications, either these *ASK* gene-containing blocks recruited *ASK* genes from elsewhere, or their duplicate blocks lost the *ASK* genes. In this paper, we call these two alternative possibilities the 'gain' and 'loss' hypotheses, respectively.

To test these two alternatives, we searched for remnants of *ASK* genes in the corresponding regions of the duplicate blocks that would be expected to contain *ASK* genes based on the 'loss' hypothesis. Our extensive search did not find any putative *ASK* pseudogenes or partial sequences in those regions (data not shown), suggesting that the 'loss' hypothesis could not be supported. However, we noticed that the estimated age for the block pair containing *ASK1* and *ASK4* is 68.82 MYA, overlapping the most recent (3R, or α in Bowers *et al.*, 2003) genome-wide duplication event in Arabidopsis that occurred 75 ± 22 MYA (Simillion *et al.*, 2002). This suggests that, if there were any *ASK* genes generated after the separation of *ASK1* and *ASK4*, those genes were not created by genome-scale segmental duplication events. More interestingly, as shown in Figure 3, it is evident that the split between *ASK4* and *ASK1* represents the second oldest duplication event, post-dating the origin of *ASK2* but pre-dating the diversification of all other *ASK* genes. This implies that most surviving *ASK* genes were not created by genome-wide duplication events, and that the corresponding *ASK* gene-containing blocks must have acquired these genes after the 3R genome duplication.

Although the investigation of previously identified duplicate blocks tends to favor the 'gain' hypothesis, there is still the possibility that some *ASK* genes have been generated by small-scale segmental duplication. For this reason, we have also compared the flanking regions of several candidate gene pairs to look for clues of recent, small-scale segmental duplication. We started with the five gene pairs that probably resulted from very recent duplication events (e.g. *ASK3* and *ASK4*, and *ASK5* and *ASK6*), and then expanded to consider the two gene pairs that were probably derived from earlier duplications. Genomic regions containing *ASK* gene pairs were regarded as arising from small-scale segmental duplication if at least one additional gene pair could be found in corresponding positions near *ASK* genes. Even with this weak criterion, we failed to find any evidence for small-scale segmental duplication (Figure 5 and Figure S3).

The situation in rice is very similar to that in Arabidopsis. Following the same procedures, we found nine duplicated blocks that contain *OSK* gene(s) (Table 1). However, none of their corresponding blocks contains *OSK* genes. Comparison of the upstream and downstream regions of many paired genes, such as *OSK3* and *OSK7*, *OSK4*/*5* and *OSK6*/*16*/*9*, *OSK10*/*11* and *OSK14*/*30*, and *OSK21* and *OSK17*/*18*/*26*/*27*/*28*/*22*/*29*/*23*, also failed to find evidence for segmental duplication (Figure S3). This suggests that segmental duplication has played a very limited role in the expansion of the plant *SKP1* gene family.

Significant contribution of retroposition to the birth of new genes

As tandem and segmental duplications cannot explain all the duplication events, we wondered whether some surviving *SKP1* genes were actually generated by transposition-related mechanisms. In particular, retroposition may have occurred, because many *ASK* and *OSK* genes lack the intron that characterizes the slowly evolving *ASK* and *OSK* genes. In fact, Figures 3(b) and 4(b) show that genes with the con-

served intron usually occupy basal positions in the phylogenetic trees, whereas genes without the intron often form terminal clades. This suggests that the intronless genes may have been derived from intron-containing genes by retroposition.

To test this hypothesis, first we tried to identify the candidate donor genes. We know that, after retroposition, the descendent retrogene will initially have an identical sequence (except for the intron) to the donor gene, so the donor gene and retrogene will cluster together in the phylogenetic tree. If two or more retrogenes are derived from the same donor gene, the donor gene will cluster with the most recent retrogene, with all earlier retrogenes arranged paraphyletically as outgroups. Thus, judged solely from the presence/absence of the specific intron, the most probable locations of retroposition events in Arabidopsis are nodes *a* and *b* of the tree shown in Figure 3(b), with the donor gene lying on the lineage leading towards *ASK3* and *ASK4*, and generating the lineages leading to nodes *c* and *d*, respectively. Both *ASK3* and *ASK4* contain the conserved intron, and it is very likely that their ancestors, nodes *e* and *a*, also contained the intron. In contrast, because none of the *ASK18*/*16*/*17*/*14*/*15*/*19* clade members possesses the intron, it is reasonable to hypothesize that their MRCA, node *c*, was intronless, and that the lineage leading to node *c* may have been generated from its ancestral gene, node *a*, through retroposition. In addition, because node *a* contained the intron, it is reasonable to hypothesize that its ancestral gene, node *b*, was also an intron-containing gene, from which the intronless MRCA of the *ASK8*/*7*/*9*/*10*/*12*/*11*/*6*/*5*/*13* clade, node *d*, was generated. Similarly, in rice, the most probable donor genes are the ancestral genes represented by nodes *f*, *h* and *j* in Figure 4(b), with nodes *g*, *i* and *k* as their retroposed copies, respectively. Note that under this hypothesis, nodes *a*, *b* and *e* represent the three different evolutionary stages of the same donor gene in Arabidopsis, and nodes *j*, *h* and *f*, as well as the *OSK2* gene, stand for the four different stages of the same donor gene in rice.

To further test our hypothesis that nodes *c*, *d*, *g*, *i* and *k* in Figures 3(b) and 4(b) were retrogenes, we checked whether these genes possess features of retrosequences such as poly(A) strings or short directed repeats associated with known transposons, in addition to the absence of the intron. However, because all these genes are at 'internal nodes', it is very likely that such features have been obscured by gene duplications. In other words, even if these nodes arose as retrogenes and possessed the hallmarks of retrosequences, their surviving descendents may have lost these attributes. Particularly, the poly(A) tail and short direct repeats may no longer be recognizable, because they can be easily masked by base substitutions and/or insertions and deletions (Betran *et al.*, 2002). However, very luckily, we found that *OSK3*, one of the two descendents of node *g*, is intronless and has a 14 bp poly(A) tail at its 3′ end and two putative direct repeats (Figure 6a). This confirms that *OSK3* is indeed a retrogene and that retroposition played an important role in the diversification of plant *SKP1* genes.

To determine the relative contribution of retroposition to the expansion of the plant *SKP1* gene family, we have re-examined all the aforementioned *ASK* and *OSK* gene pairs to see whether at least one of the two paired genes possesses diagnostic features of retrosequences. We found that, in a few cases, the upstream gene of an *ASK* or *OSK*

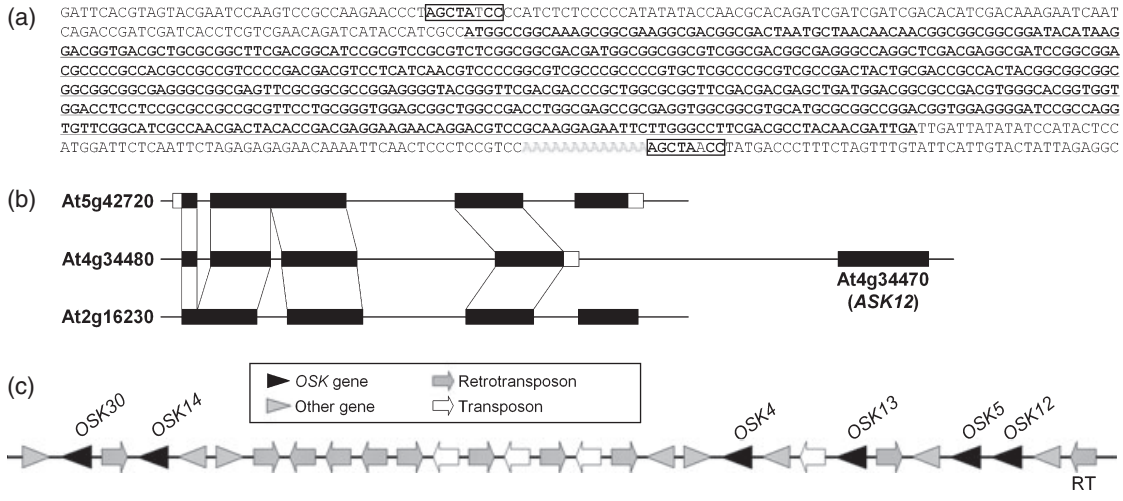

Figure 6. *SKP1*-like retrogenes and the distribution of *SKP1*-like genes in a mobile element-rich region.
(a) The intronless *OSK3* gene (underlined and in bold) possesses a poly(A) tail (in gray) and two possible direct repeats (in boxes).
(b) The upstream gene of *ASK12*, At4g34480, which encodes a glocosyl hydrolase family protein, is truncated compared with its closest paralogs, At2g16230 and At5g42720, with its protein lacking approximately 100 amino acids at the C-terminus.
(c) Six *OSK* genes located in a chromosomal region that contains 11 retrotransposons, four transposons, and eight other genes.

gene does encode a truncated protein, compared with their closest relatives (Table S1). For example, the upstream gene of *ASK12*, At4g34480, which encodes a glycosyl hydrolase family protein, has lost its last exon compared with its closest relatives, At2g16230 and At5g42720 (Figure 6b). The fact that some intronless *OSK* genes are located within a retrotransposon-rich region (Figure 6c) further suggests that retroposition, or at least retroposition-related mechanisms, may have contributed to the increase in the number of plant *SKP1* genes.

It is generally believed that most retroposed sequences will become non-functional because they lack the regulatory elements required for their expression (Graur and Li, 2000). However, several recent studies have indicated that some retrosequences are transcribed and have evolved under constraint (Benovoy and Drouin, 2006; Betran *et al.*, 2002; Marques *et al.*, 2005; Vinckenbosch *et al.*, 2006; Wang *et al.*, 2006). In the *SKP1* family, transcripts of many intronless *ASK* and *OSK* genes have been detected in RT-PCR and/or microarray analyses (Kong *et al.*, 2004; Takahashi *et al.*, 2004; Zhao *et al.*, 2003b). Yeast two-hybrid assays further reveal that the protein products of several intronless *ASK* genes (i.e. *ASK9, 11-14, 16, 18* and *19*) are able to interact with certain F-box proteins (Gagne *et al.*, 2002; Risseeuw *et al.*, 2003; Takahashi *et al.*, 2004). This, together with the fact that many intronless *ASK* and *OSK* genes have evolved under strong purifying selection (as indicated by the *small* d_N/d_S values; see Kong *et al.*, 2004 and Figure S4) suggests that some retroposed *SKP1* sequences can gain regulatory elements by some unknown mechanisms and become functional. It is also not known why some gene families are subject to more retroposition than others.

Experimental procedures

Data retrieval

The Arabidopsis *SKP1-like* (*ASK*) genes were retrieved as described previously (Kong *et al.*, 2004), and the *Oryza SKP1-like* (*OSK*) genes were obtained by BLASTP searches against the database 'Genes in TIGR rice pseudomolecules: protein sequences' at http://tigrblast.tigr.org/euk-blast/index.cgi?project=osa1, using multiple Skp1 proteins as queries. To better understand the evolutionary history of plant *SKP1* genes, we also included many other plant species, i.e. *Chlamydomonas reinhardtii*, *Physcomitrella patens*, *Picea* sp. (spruce), *Pinus taeda*, *Amborella trichopoda*, *Nuphar advena* (water lily), *Acorus americanus* (sweet flag), *Allium cepa* (onion), *Liriodendron tulipifera* (tulip poplar), *Persea americana* (avocado), *Mesembryanthemum crystallinum* (ice plant), *Medicago truncatula*, *Glycine max* (soybean), *Populus trichocar- pa* (poplar), *Brassica napus* (rape), *Solanum tuberosum* (potato), *Antirrhinum majus* (snapdragon), *Helianthus annuus* (sunflower), *Vitis vinifera* (grape), *Zea mays* (maize) and *Triticum aestivum* (wheat). *SKP1* genes from these species were either retrieved by TBLASTN searches against the TIGR EST databases (Lee *et al.*, 2005) or obtained from the PlantTribes database at http://www.floralgenome.org/tribe.php. ESTs or unigenes from the same species with >95% identity in the coding region were considered alleles; only one such allele was included in the analysis.

A typical Skp1 protein contains a variable N-terminus, two highly conserved domains, Skp1_POZ and Skp1, and an intervening region between these two domains (Schulman *et al.*, 2000; Figure 1a). The Skp1_POZ domain contains three β-strands (S1–S3) and three α-helices (H1–H3), and is important for the interaction with Cullin, whereas the Skp1 domain, which includes five α-helices (H4–H8), contains many key residues by which the F-box domain of an F-box protein is bound (Schulman *et al.*, 2000; Zheng *et al.*, 2002). Because deletion of these β-strands or α-helices usually causes loss of function (Risseeuw *et al.*, 2003), we regarded any *SKP1* genes whose protein sequences lack these structurally conserved regions as pseudogenes. In addition, it has been shown that some Skp1 proteins may have accumulated so many mutations at these key residues that protein–protein interactions will not occur (Gagne *et al.*, 2002; Kong *et al.*, 2004; Risseeuw *et al.*, 2003; Takahashi *et al.*, 2004). For this reason, we conducted domain structure analyses using the Pfam and SMART platforms. Full-length Skp1 homologs with less conserved Skp1_POZ and Skp1 domains were treated as putative pseudogenes.

Alignment and phylogenetic analysis

Protein sequence alignment was generated by using the CLUSTALX 1.81 program , with manual adjustment. The alignment for the Skp1_POZ and Skp1 domain regions was straightforward and required few gaps. However, for the two variable regions, the alignment was very difficult. Despite this, we noticed that sequences from the same, or closely related, species tend to have similar features in these variable regions (constituting species- or lineage-specific features), suggesting that gene duplications may have occurred after the divergence of organisms. However, to avoid possible error in the estimation of phylogenetic relationships, these less-conserved regions were excluded from further analyses, leaving 149 residues from the conserved Skp1_POZ and Skp1 domain regions. A DNA version of this alignment was then generated using the publicly available software aa2dna (http://www.bio.psu.edu/People/Faculty/Nei/Lab/software.htm).

Phylogenetic analyses were conducted using both protein sequences and their corresponding DNA sequences. A previous study had shown that, due to rate heterogeneity, long-branch attraction could cause strongly distorted results in the phylogenetic analyses of *SKP1* genes (Kong *et al.*, 2004). For this reason, we used maximum likelihood (ML), in addition to neighbor-joining (NJ) or maximum parsimony (MP), methods because ML worked best for *SKP1* genes (Kong *et al.*, 2004). Furthermore, we noticed that phylogenetic trees based on amino acid sequences tend to be sensitive to taxon sampling. The reason for this may be that protein sequences are less informative, compared with DNA sequences; some *SKP1* genes have evolved under very strong purifying selection so that only very few non-synonymous substitutions could be observed. However, synonymous substitutions are not rare, and thus should be taken into account during phylogenetic estimate. For this reason, we based our conclusion mainly on the analyses of nucleotide sequences. We utilized the newly developed likelihood tree-building program, PHYML version 2.4.3 (Guindon and Gascuel, 2003), for our ML analyses. A similar ML analysis was also performed using PAUP* version 4.10b (Swofford, 2001) to determine the reliability of PHYML. NJ and MP trees were built using MEGA 3.1 (Kumar *et al.*, 2004) and PAUP, respectively. Ratios of non-synonymous to

synonymous nucleotide substitutions ($\omega = d_N/d_S$) for *ASK* and *OSK* genes were estimated using the codeml program of PAML version 3.13 (Yang, 1997).

Determination of the duplication types

To elucidate the mechanism by which new *SKP1* genes were born so frequently in the Arabidopsis and rice genomes, we investigated the relative contributions of the three main types of gene duplication. We recognized tandem duplications by a close phylogenetic relationship among tandemly arrayed genes at the same chromosomal location. We looked for segmental duplicates by comparing positions of *SKP1* genes in known duplicated chromosomal blocks in the Arabidopsis and rice genomes at http://bioinformatics.psb.ugent.be/supplementary_data/cesim/simillion_pnas02/ and http://bioinformatics.psb.ugent.be/supplementary_data/klpoe/vandepoele_ricedup/ (Simillion *et al.*, 2002; Vandepoele *et al.*, 2003). However, as these investigations only dealt with relatively ancient duplication events, we also compared the flanking regions of the *ASK* and *OSK* gene pairs that were possibly the results of recent segmental duplications. To do this, we compared 50 kb regions both upstream and downstream of the gene pairs concerned using the MacVector software (Rastogi, 2000) and the DotPlot function of the PipMaker program (Schwartz *et al.*, 2000) at http://pipmaker.bx.psu.edu/pipmaker/. Where this comparison failed to find evidence that two paired genes were the results of a segmental duplication event, we examined the genes to determine whether one or both genes were actually retrogene(s), or processed gene(s), generated by retroposition. A gene is regarded as a retrogene if it has such features as: (i) lack of the intron, (ii) stretches of poly(A) at the 3′ end, (iii) short direct repeats at both ends, or (iv) chromosomal positions different from the locus of the donor gene from which the mRNA was transcribed. In addition, if a cDNA is inserted into the coding region of another (target) gene, the target gene becomes interrupted, resulting in a truncated gene. For this reason, special attention was paid to the identification of truncated genes near *SKP1* genes.

Acknowledgements

We thank Drs Masatoshi Nei, Manyuan Long, Yoshihito Niimura, Yoshiyuki Suzuki, Masafumi Nozawa, Hongyan Shan and two anonymous reviewers for critical reading of the manuscript and valuable comments. This work was supported by National Nature Science Foundation of China (grant numbers 30530090, 30470116 and 30121003), the National Science Foundation of the United States (The Floral Genome Project, Plant Genome Award DBI-0115684), and an IBCAS Grant for Scientific Frontiers.

Supplementary Material

The following supplementary material is available for this article online:

Table S1. Sequence information of *ASK* and *OSK* genes

Figure S1. Phylogenetic relationships of the 19 *ASK* genes and 28 *OSK* genes

Figure S2. Phylogenetic relationships of 120 type I *SKP1* homologs from plants

Figure S3. Comparisons of the flanking genomic regions (50 kb at both sides) of selected *ASK* and *OSK* gene pairs

Figure S4. Phylogenetic relationships of the 28 *OSK* genes, with d_N/d_S values indicated for each branch

Table S1. Sequence information of *ASK* and *OSK* genes

This material is available as part of the online article from http://www.blackwell-synergy.com.

References

Adams, K.L. and Wendel, J.F. (2005) Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* **8**, 135–141.

Anderson, C.L., Bremer, K. and Friis, E.M. (2005) Dating phylogenetically basal eudicots using *rbcL* sequences and multiple fossil reference points. *Am. J. Bot.* **92**, 1737–1748.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

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.

Benovoy, D. and Drouin, G. (2006) Processed pseudogenes, processed genes, and spontaneous mutations in the *Arabidopsis* genome. *J. Mol. Evol.* **62**, 511–522.

Betran, E., Thornton, K. and Long, M. (2002) Retroposed new genes out of the X in *Drosophila*. *Genome Res.* **12**, 1854–1859.

Blanc, G. and Wolfe, K.H. (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*, **16**, 1667–1678.

Blanc, G., Barakat, A., Guyot, R., Cooke, R. and Delseny, M. (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell*, **12**, 1093–1101.

Blanc, G., Hokamp, K. and Wolfe, K.H. (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* **13**, 137–144.

Bowers, J.E., Chapman, B.A., Rong, J. and Paterson, A.H. (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, **422**, 433–438.

Cannon, S.B., Mitra, A., Baumgarten, A., Young, N.D. and 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.

Cardozo, T. and Pagano, M. (2004) The SCF ubiquitin ligase: insights into a molecular machine. *Nat. Rev. Mol. Cell Biol.* **5**, 739–751.

Choo, K.B., Hsu, M.C., Chong, K.Y. and Huang, C.J. (2007) Testis-specific expression and genomic multiplicity of the rat Rtdpoz genes that encode bipartite TRAF- and POZ/BTB-domain proteins. *Gene*, **387**, 141–149.

Connelly, C. and Hieter, P. (1996) Budding yeast *SKP1* encodes an evolutionarily conserved kinetochore protein required for cell cycle progression. *Cell*, **86**, 275–285.

Davies, T.J., Barraclough, T.G., Chase, M.W., Soltis, P.S., Soltis, D.E. and Savolainen, V. (2004) Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proc. Natl Acad. Sci. USA*, **101**, 1904–1909.

Drouaud, J., Marrocco, K., Ridel, C., Pelletier, G. and Guerche, P. (2000) A *Brassica napus skp1*-like gene promoter drives GUS expression in *Arabidopsis thaliana* male and female gametophytes. *Sexual Plant Reprod.* **13**, 29–35.

Farras, R., Ferrando, A., Jasik, J., Kleinow, T., Okresz, L., Tiburcio, A., Salchert, K., del Pozo, C., Schell, J. and Koncz, C. (2001) SKP1–SnRK protein kinase interactions mediate proteasomal binding of a plant SCF ubiquitin ligase. *EMBO J.* **20**, 2742–2756.

Gagne, J.M., Downes, B.P., Shiu, S.H., Durski, A.M. and Vierstra, R.D. (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.

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

Graur, D. and Li, W. (2000) *Fundamentals of Molecular Evolution*. Sunderland, MA: Sinauer Associates.

Gray, W.M., del Pozo, J.C., Walker, L., Hobbie, L., Risseeuw, E., Banks, T., Crosby, W.L., Yang, M., Ma, H. and Estelle, M. (1999) Identification of an SCF ubiquitin-ligase complex required for auxin response in *Arabidopsis thaliana*. *Genes Dev.* **13**, 1678–1691.

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

Guo, H. and Ecker, J.R. (2003) Plant responses to ethylene gas are mediated by $SCF^{EBF1/EBF2}$-dependent proteolysis of EIN3 transcription factor. *Cell*, **115**, 667–677.

Hellmann, H. and Estelle, M. (2002) Plant development: regulation by protein degradation. *Science*, **297**, 793–797.

Hobbie, L. (2005) Seek and Ye shall [eventually] find: the end of the search for the auxin receptor. *J. Integr. Plant Biol.* **47**, 1412–1417.

Horan, K., Lauricha, J., Bailey-Serres, J., Raihhel, N. and Girke, T. (2005) Genome cluster database: a sequence family analysis platform for *Arabidopsis* and rice. *Plant Physiol.* **138**, 47–54.

Javaud, C., Dupuy, F., Maftah, A., Julien, R. and Petit, J.M. (2003) The fucosyltransferase gene family: an amazing summary of the underlying mechanisms of gene evolution. *Genetica*, **118**, 157–170.

Kellogg, E.A. (2001) Evolutionary history of the grasses. *Plant Physiol.* **125**, 1198–1205.

Kong, H., Leebens-Mack, J., Ni, W., dePamphilis, C.W. and 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.

Kumar, S., Tamura, K. and Nei, M. (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* **5**, 150–163.

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

Lee, Y., Tsai, J., Sunkara, S., Karamycheva, S., Pertea, G., Sultana, R., Antonescu, V., Chan, A., Cheung, F. and Quackenbush, J. (2005) The TIGR gene indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res.* **33**, D71–D74.

Leister, D. (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. *Trends Genet.* **20**, 116–122.

Liu, F., Ni, W., Griffith, M.E., Huang, Z., Chang, C., Peng, W., Ma, H. and Xie, D. (2004) The *ASK1* and *ASK2* genes are essential for *Arabidopsis* early development. *Plant Cell*, **16**, 5–20.

Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, **290**, 1151–1155.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M. and Van de Peer, Y. (2005) Modeling gene and genome duplications in eukaryotes. *Proc. Natl Acad. Sci. USA*, **102**, 5454–5459.

Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A. and Kaessmann, H. (2005) Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol.* **3**, e357.

Michelmore, R.W. and Meyers, B.C. (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* **8**, 1113–1130.

Nakano, T., Suzuki, K., Fujimura, T. and Shinshi, H. (2006) Genome-wide analysis of the *ERF* gene family in *Arabidopsis* and rice. *Plant Physiol.* **140**, 411–432.

Nam, J., Kim, J., Lee, S., An, G., Ma, H. and Nei, M. (2004) Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc. Natl Acad. Sci. USA*, **101**, 1910–1915.

Nayak, S., Santiago, F.E., Jin, H., Lin, D., Schedl, T. and Kipreos, E.T. (2002) The *Caenorhabditis elegans* Skp1-related gene family: diverse functions in cell proliferation, morphogenesis, and meiosis. *Curr. Biol.* **12**, 277–287.

Ni, W., Xie, D., Hobbie, L., Feng, B., Zhao, D., Akkara, J. and Ma, H. (2004) Regulation of flower development in *Arabidopsis* by SCF complexes. *Plant Physiol.* **134**, 1574–1585.

Nicholson, A.C., Malik, S.B., Logsdon, J.M. Jr and Van Meir, E.G. (2005) Functional evolution of ADAMTS genes: evidence from analyses of phylogeny and gene organization. *BMC Evol. Biol.* **5**, 11.

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

Paterson, A.H., Bowers, J.E. and Chapman, B.A. (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl Acad. Sci. USA*, **101**, 9903–9908.

Pavlicek, A., Gentles, A.J., Paces, J., Paces, V. and Jurka, J. (2006) Retroposition of processed pseudogenes: the impact of RNA stability and translational control. *Trends Genet.* **22**, 69–73.

Porat, R., Lu, P. and O'Neill, S.D. (1998) *Arabidopsis SKP1*, a homologue of a cell cycle regulator gene, is predominantly expressed in meristematic cells. *Planta*, **204**, 345–351.

Raes, J., Vandepoele, K., Simillion, C., Saeys, Y. and Van de Peer, Y. (2003) Investigating ancient duplication events in the *Arabidopsis* genome. *J. Struct. Funct. Genomics*, **3**, 117–129.

Rastogi, P.A. (2000) MacVector: integrated sequence analysis for the Macintosh. *Meth. Mol. Biol.* **132**, 47–69.

Risseeuw, E.P., Daskalchuk, T.E., Banks, T.W., Liu, E., Cotelesage, J., Hellmann, H., Estelle, M., Somers, D.E. and Crosby, W.L. (2003) Protein interaction analysis of SCF ubiquitin E3 ligase subunits from *Arabidopsis*. *Plant J.* **34**, 753–767.

Sampedro, J., Lee, Y., Carey, R.E., dePamphilis, C. and Cosgrove, D.J. (2005) Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. *Plant J.* **44**, 409–419.

Schulman, B.A., Carrano, A.C., Jeffrey, P.D., Bowen, Z., Kinnucan, E.R., Finnin, M.S., Elledge, S.J., Harper, J.W., Pagano, M. and Pavletich, N.P. (2000) Insights into SCF ubiquitin ligases from the structure of the Skp1–Skp2 complex. *Nature*, **408**, 381–386.

Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R. and Miller, W. (2000) PipMaker – a web server for aligning two genomic DNA sequences. *Genome Res.* **10**, 577–586.

Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F. and Li, W.H. (2004) Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell*, **16**, 1220–1234.

Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M. and van de Peer, Y. (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **99**, 13627–13632.

Soltis, D.E., Albert, V.A., Kim, S. *et al.* (2005) Evolution of the flower. In *Plant Diversity and Evolution* (Henry, R., ed). Wallingford, UK: CABI Publishers, pp. 165–200.

Strichman-Almashanu, L.Z., Bustin, M. and Landsman, D. (2003) Retroposed copies of the HMG genes: a window to genome dynamics. *Genome Res.* **13**, 800–812.

Swofford, D.L. (2001) *PAUP: Phylogenetic Analysis Using Parsimony* and Other Methods*, version 4.0 beta. Sunderland, MA: Sinauer Associates.

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

Thomas, J.H. (2006) Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. *Genome Res.* **16**, 1017–1030.

Tian, C., Wan, P., Sun, S., Li, J. and Chen, M. (2004) Genome-wide analysis of the *GRAS* gene family in rice and *Arabidopsis*. *Plant Mol. Biol.* **54**, 519–532.

Vandepoele, K., Simillion, C. and Van de Peer, Y. (2003) Evidence that rice and other cereals are ancient aneuploids. *Plant Cell*, **15**, 2192–2202.

Vinckenbosch, N., Dupanloup, I. and Kaessmann, H. (2006) Evolutionary fate of retroposed gene copies in the human genome. *Proc. Natl Acad. Sci. USA*, **103**, 3220–3225.

Vision, T.J., Brown, D.G. and Tanksley, S.D. (2000) The origins of genomic duplications in *Arabidopsis*. *Science*, **290**, 2114–2117.

Wang, Y. and Yang, M. (2005) The ARABIDOPSIS SKP1-LIKE1 (ASK1) protein acts predominately from leptotene to pachytene and represses homologous recombination in male meiosis. *Planta*, **223**, 613–617.

Wang, W., Zheng, H., Fan, C. *et al.* (2006) High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell*, **18**, 1791–1802.

Xu, L., Liu, F., Lechner, E., Genschik, P., Crosby, W.L., Ma, H., Peng, W., Huang, D. and Xie, D. (2002) The SCF^{COI1} ubiquitin-ligase complexes are required for jasmonate response in *Arabidopsis*. *Plant Cell*, **14**, 1919–1935.

Yamanaka, A., Yada, M., Imaki, H., Koga, M., Ohshima, Y. and 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.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556.

Yang, M., Hu, Y., Lodhi, M., McCombie, W.R. and Ma, H. (1999) The *Arabidopsis SKP1-LIKE1* gene is essential for male meiosis and may control homologue separation. *Proc. Natl Acad. Sci. USA*, **96**, 11416–11421.

Yang, X., Timofejeva, L., Ma, H. and Makaroff, C.A. (2006) The *Arabidopsis SKP1* homolog *ASK1* controls meiotic chromosome remodeling and release of chromatin from the nuclear membrane and nucleolus. *J. Cell Sci.* **119**, 3754–3763.

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

Zhang, J. (2003) Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**, 292–298.

Zhao, D., Yang, M., Solava, J. and Ma, H. (1999) The *ASK1* gene regulates development and interacts with the *UFO* gene to control floral organ identity in *Arabidopsis*. *Dev. Genet.* **25**, 209–223.

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

Zhao, D., Han, T., Risseeuw, E., Crosby, W.L. and Ma, H. (2003a) Conservation and divergence of *ASK1* and *ASK2* gene functions during male meiosis in *Arabidopsis thaliana*. *Plant Mol. Biol.* **53**, 163–173.

Zhao, D., Ni, W., Feng, B., Han, T., Petrasek, M.G. and Ma, H. (2003b) Members of the *Arabidopsis-SKP1-like* gene family exhibit a variety of expression patterns and may play diverse roles in *Arabidopsis*. *Plant Physiol.* **133**, 203–217.

Zhao, D., Yang, X., Quan, L., Timofejeva, L., Rigel, N.W., Ma, H. and Makaroff, C.A. (2006) *ASK1*, a *SKP1* homolog, is required for nuclear reorganization, presynaptic homolog juxtaposition and the proper distribution of cohesin during meiosis in *Arabidopsis*. *Plant Mol. Biol.* **62**, 99–110.

Zheng, N., Schulman, B.A., Song, L. *et al.* (2002) Structure of the Cul1–Rbx1–Skp1–F box^{Skp2} SCF ubiquitin ligase complex. *Nature*, **416**, 703–709.

Zhou, Y.C., Dieterle, M., Buche, C. and Kretsch, T. (2002) The negatively acting factors EID1 and SPA1 have distinct functions in phytochrome A-specific light signaling. *Plant Physiol.* **128**, 1098–1108.

