Journal of Experimental Botany, Vol. 69, No. 9 pp. 2367–2378, 2018
doi:10.1093/jxb/ery097 Advance Access publication 10 March 2018

RESEARCH PAPER

Molecular evolution and diversification of the SMXL gene family

Taraka Ramji Moturu[1], Sravankumar Thula[1], Ravi Kumar Singh[2], Tomasz Nodzyński[1], Radka Svobodová Vařeková[1,3], Jiří Friml[4] and Sibu Simon[1,*]

[1] Mendel Centre for Plant Genomics and Proteomics, Central European Institute of Technology (CEITEC), Masaryk University, Kamenice 5, CZ-625 00 Brno, Czech Republic
[2] Department of Bioinformatics, SRM University, Haryana, Sonepat 131029, India
[3] National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno-Bohunice, Czech Republic
[4] Institute of Science and Technology (IST) Austria, Am Campus 1, 3400 Klosterneuburg, Austria

* Correspondence: sibu.simon@ceitec.muni.cz

Received 22 January 2018; Editorial decision 3 March 2018; Accepted 5 March 2018

Editor: Richard Napier, University of Warwick, UK

Abstract

Strigolactones (SLs) are a relatively recent addition to the list of plant hormones that control different aspects of plant development. SL signalling is perceived by an α/β hydrolase, DWARF 14 (D14). A close homolog of D14, KARRIKIN INSENSTIVE2 (KAI2), is involved in perception of an uncharacterized molecule called karrikin (KAR). Recent studies in Arabidopsis identified the SUPPRESSOR OF MAX2 1 (SMAX1) and SMAX1-LIKE 7 (SMXL7) to be potential SCF–MAX2 complex-mediated proteasome targets of KAI2 and D14, respectively. Genetic studies on SMXL7 and SMAX1 demonstrated distinct developmental roles for each, but very little is known about these repressors in terms of their sequence features. In this study, we performed an extensive comparative analysis of SMXLs and determined their phylogenetic and evolutionary history in the plant lineage. Our results show that SMXL family members can be sub-divided into four distinct phylogenetic clades/classes, with an ancient SMAX1. Further, we identified the clade-specific motifs that have evolved and that might act as determinants of SL-KAR signalling specificity. These specificities resulted from functional diversities among the clades. Our results suggest that a gradual co-evolution of SMXL members with their upstream receptors D14/KAI2 provided an increased specificity to both the SL perception and response in land plants.

Keywords: D14, DLK2, evolution, KAI2, karrikin, MAX2, SMAX1-like, strigolactone.

Introduction

Plant growth and development are fine-tuned by various endogenous and exogenous signals. Plant hormones are one of the major endogenous signals that act with a predetermined framework to help the plant attain optimized growth and to respond rapidly to environmental stimuli (Gray, 2004). Strigolactones (SLs) are a family of terpenoid lactone hormones that control multiple developmental events, from Charophyte algae to higher-plant monocots and dicots (Yoneyama *et al.*, 2007*a*, 2007*b*, 2009; Umehara *et al.*, 2008; Xie *et al.*, 2010; Proust *et al.*, 2011; Ruyter-Spira *et al.*, 2011, 2013; Delaux *et al.*, 2012; Brewer *et al.*, 2013). SLs secreted from host plants promote hyphal branching of arbuscular

mycorrhizal (AM) fungi to facilitate symbiotic interactions between the host and fungus that lead to the exchange of inorganic nutrients, especially phosphate to the host plants (Akiyama *et al.*, 2005; Bouwmeester *et al.*, 2007; Chen and Xiong, 2009). SLs also stimulate the germination of seeds of parasitic weeds from the Orbanchaceae family (order Lamiales), such as *Striga hermonthica*, leading to attachment to the host and eliciting devastating effects on global crop production (Cook *et al.*, 1966; Spallek *et al.*, 2013). *Striga* infects major food crops including maize, sorghum, rice, and pearl millet (Humphrey and Beale, 2006; Parker *et al.*, 2009; Westwood *et al.*, 2010).

SLs are carotenoid-derived molecules that are mainly synthesized in plastids of root cells and are transported acropetally (Kohlen *et al.*, 2011; Seto *et al.*, 2012). They regulate various stages of plant development. In the roots, SLs promote increases in cell numbers in the primary root meristem, repress lateral and adventitious root formation, and promote root-hair elongation (Agusti *et al.*, 2011; Koltai *et al.*, 2011; Ruyter-Spira *et al.*, 2011; Rasmussen *et al.*, 2012). In the shoot, SLs regulate shoot branching, leaf senescence, secondary thickening, photomorphogenesis, and stem elongation (Gomez-Roldan *et al.*, 2008; Umehara *et al.*, 2008; Waters *et al.*, 2017). SL biosynthesis is strongly affected by phosphate deficiency in the rhizosphere (Ruyter-Spira *et al.*, 2011; Brewer *et al.*, 2013; Bennett and Leyser, 2014; Smith and Li, 2014; Waldie *et al.*, 2014; Al-Babili and Bouwmeester, 2015).

In eukaryotes, ubiquitination followed by proteasome-mediated protein degradation is a well-established mechanism for the removal of abnormal proteins. However, in plants this machinery is also an integral part of many hormonal signalling pathways. Ubiquitination of repressors of hormone-response target genes, a multi-step process involving interaction with the E3/E2 (ubiquitin ligase/conjugating enzymes) complex in response to the phytohormones, is crucial for plant development (Stone and Callis, 2007; Hua and Vierstra, 2011; Kelley and Estelle, 2012). In common with other hormones such as auxins, gibberellins, and jasmonates, SL signalling is also based on proteasome-mediated repressor degradation (Waters *et al.*, 2017). An F-box leucine-rich protein, MAX2, which is well-conserved among land plants, has been shown to be a part of the SKP1-CUL1-F-box protein (SCF)-type ubiquitin ligase complex and acts as a common factor in the signalling pathways of strigolactones and karrikins (KARs)/karrikin-like (KLs) (Delaux *et al.*, 2012; Waters *et al.*, 2012; Zhou *et al.*, 2013, 2016). KARs are found in smoke deposits formed when plant material is burned. DWARF 14 (D14) and KARRIKIN INSENSTIVE2 (KAI2) are classes of α/β hydrolases that are signalling receptors for SL and KL, respectively (Nelson *et al.*, 2011; Delaux *et al.*, 2012; Waters *et al.*, 2012; Zhou *et al.*, 2013, 2016). The DWARF 14-LIKE 2 (DLK2) protein, while structurally very similar to D14 and KAI2, has been identified as functioning independently of either of them (Végh *et al.*, 2017; Wallner *et al.*, 2017). SLs bind to the D14 receptor, producing a covalently linked intermediate molecule (CLIM) that in turn interacts with the MAX2–SCF complex. This interaction triggers the 26S proteasome-mediated degradation of target proteins such as SUPPRESSOR OF MAX2 1-LIKE 7 (SMXL7) and DWARF53 ((D53) in Arabidopsis and rice, respectively, leading to de-repression of the target genes (Wang *et al.*, 2015; Yao *et al.*, 2016). The perception of SL and the subsequent degradation of SMXL7/D53 requires the RGKT functional motif domains, which when mutated or deleted affect the degradation of SMXL repressors (Nelson *et al.*, 2011; Jiang *et al.*, 2013; Zhou *et al.*, 2013, 2016; Soundappan *et al.*, 2015; Liang *et al.*, 2016). Although D14 and KAI2 perceive SL and KAR signalling in an identical way, their perception leads to distinct developmental events in plants (Wang *et al.*, 2015; Bennett *et al.*, 2016; Yao *et al.*, 2016, 2018). Among the eight SMXL genes identified in Arabidopsis, SMAX1 responds to KARs and regulates seed germination and hypocotyl length (Stanga *et al.*, 2016). In contrast, SMXL3, SMXL4, and SMXL5 do not respond to either KARs or SLs, and are involved in phloem formation and primary root growth in *Arabidopsis thaliana* (Wallner *et al.*, 2017; Wu *et al.*, 2017). SMXL4 and SMXL5 mutants exhibit a leaf pigmentation phenotype with over-accumulation of starch and anthocyanin (Wu *et al.*, 2017). SMXL6, SMXL7, and SMXL8 regulate shoot branching, leaf shape, and promotion of lateral roots in Arabidopsis and are reported to be the degradation targets of D14 (Soundappan *et al.*, 2015; Wang *et al.*, 2015; Bennett *et al.*, 2016).

Recent evolutionary studies have demonstrated the early appearance of KAI2 and the divergence of DDK proteins (D14/DLK2/KAI2) in plants (Delaux *et al.*, 2012; Bythell-Douglas *et al.*, 2017). The KAR and SL signalling pathways gradually evolved within this DDK lineage (Bythell-Douglas *et al.*, 2017). SUPPRESSOR OF MAX2 1 (SMAX1)-like homologs have weak similarity to ClpB heat chaperonins, with the double Clp-N and P-loop motifs that are characteristic of the nucleoside triphosphate hydrolases superfamily (Jiang *et al.*, 2013; Stanga *et al.*, 2013; Zhou *et al.*, 2013, 2016). They are induced by heat stress and are known to be involved in the proteolytic machinery. It remains to be determined whether SMXLs have simply just retained the domain organization or if they also possess ClpB-like activity.

In the present study, we attempted to retrieve, annotate, and classify the proteasome targets of the SL/KL signalling pathways in order to get a better insight into the evolution of the SL/KAR response in plants. We also consider the origins and functional diversification of the SMXL family, with emphasis on the conserved motifs.

Materials and methods

The methodology used for the analysis is depicted in Supplementary Fig. S1 at *JXB* online, which contains the steps outlined below.

Assembly of dataset of SMXL homologs

To annotate the SMXL members in plant genomes, we used the homology search tool blastp to scan the Phytozome v12.1.5 database (Goodstein *et al.*, 2012) using full-length sequences of Arabidopsis SMXLs as the query. A hidden Markov model (HMM) was built using JACKHMMER (Finn *et al.*, 2015) with Arabidopsis SMXLs as reference sequences to scan the reference proteome (KW-0181)

with a cut-off *e*-value of 0.01. Homologs from parasitic, non-angiosperm species were obtained from the 1000 Plants (1KP) project (www.onekp.com, accessed 23rd November 2017) (Johnson *et al.*, 2012; Matasci *et al.*, 2014; Wickett *et al.*, 2014; Xie *et al.*, 2014). Only full-length peptides of homologs were retained for further annotation. To annotate, each predicted homolog was compared with Arabidopsis SMXL peptide sequences using the 'reciprocal best hits' method, as implemented in Proteinortho v5.16 (Lechner *et al.*, 2011). Characteristic domains (double Clp-N and P-Loop NTPase) in the homologs were confirmed using the Structural Classification Of Proteins (SCOP) database (Wilson *et al.*, 2009). To enable better phylogenetic inferences, homologs with 90% or more sequence identity were trimmed out using the CD-HIT suite (Huang *et al.*, 2010).

Multiple sequence alignment (MSA) was performed using the GUIDANCE2 tool (Sela *et al.*, 2015) with the option of MAFFT alignment (Yamada *et al.*, 2016) and 100 bootstrap replicates, with sequence cut-off at 0.6 and column cut-off at 0.93. Sequences below the threshold were removed and re-run to increase the GUIDANCE confidence score to 0.73. Poorly aligned columns below the threshold were removed and the resulting alignment was used for further phylogenetic analysis. ProtTest v3.4.2 (Darriba *et al.*, 2011) was used to test the best-fitted amino acid substitution model and the parameter values for the dataset. The Jones–Taylor–Thornton (JTT) model with an estimated gamma-distribution parameter, an estimated proportion of invariant sites (I), and pre-defined base frequencies (F) was the best-fit model according to the Akaike Information Criterion (AIC) framework (Posada and Buckley, 2004).

Phylogenetic analysis

Evolutionary relationships among SMXL family members were determined using methods based on discrete data (Maximum Likelihood, ML), probabilistic models (Bayesian Inference, BI), and a distance matrix (Neighbour-Joining, NJ). ML analysis was performed using RAxML v8.2.11 (Stamatakis, 2014) with the best-fit evolutionary model selected from ProtTest. Statistical support for each clade was calculated using 100 bootstrap replicates. The Bayesian phylogeny was reconstructed using MrBayes v3.2.6 (Ronquist *et al.*, 2012) implemented in the CIPRES cluster (Miller *et al.*, 2015) with the default parameters and using a discrete gamma model for one million chains of generations. Chain convergence was assessed by standard deviation <0.01 sampled every 1000 generations.

We also reconstructed the phylogenetic tree using IQTree v1.5.6 (Nguyen *et al.*, 2015; Trifinopoulos *et al.*, 2016) implemented on the webserver (http://www.iqtree.org/), which was shown to improve the accuracy and robustness of the phylogenetic tree. Clade support was calculated using 1000 bootstrap alignments and single-branch tests, such as SH-aLRT and the Bayes test. The tree was visualized and modified using the iTOL v3 online server (Letunic and Bork, 2016).

Evolutionary expansion of SMXLs

To understand and infer the evolutionary expansion history of the SMXL family, we used the homologs only from the 58 completely sequenced genomes available in the Phytozome v12.1.5 database (Goodstein *et al.*, 2012). The SMXL gene family tree (GFT) was reconciled with the species tree of the 58 genomes, generated with the NCBI Taxonomy Browser (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi), using Notung v2.9 (Chen *et al.*, 2000; Stolzer *et al.*, 2012) to calculate the gain and loss events of SMXL genes at each successive stage during evolution (see Supplementary Fig. S7). However, the whole-genome duplication and triplication histories were adapted from the Plant Genome Duplication Database (PGDD) (Lee *et al.*, 2013, 2017) and from Li *et al.* (2016). To test the robustness of the duplication/loss events that could affect the evolutionary inferences, the 'column score' of GUIDANCE2 was changed to 0.95, based on the sensitivity and specificity percentage of the tool (Sela *et al.*, 2015).

Molecular clock

A molecular clock test was performed to estimate the divergence of SMXLs (see Supplementary Fig. S8). Comparisons were made with ML values for given tree topologies with and without molecular clock constraints under the JTT (+G+I) model (Jones *et al.*, 1992) using MEGA 7 (Kumar *et al.*, 2016). A null hypothesis of equal evolutionary rate throughout the tree was rejected at a significance level of 5%.

Structural modelling and comparisons of SMXLs

The I-TASSER webserver (Yang and Zhang, 2015*a*, 2015*b*) was used to model the structures of Arabidopsis SMAX1, SMXL7, SMXL3, and SMXL4 as representatives from each phylogenetic clade, and also AtClpB. I-TASSER generates structural conformations called decoys and clusters them based on the pair-wise sequence similarity. Based on the C-score (see Supplementary Table S4), the top models were selected for representation. Structures were compared using the TM-Align server (Zhang and Skolnick, 2005) (Supplementary Tables S5, S6). The global homology iterative refinement method (MAFFT G-INS-i) was used to align SMXLs with ClpB proteins.

Sequence motif analysis

Motif analysis was performed using MEME suite v4.12.0 (Bailey *et al.*, 2009, 2015), which scans for motifs recurring in the set of sequences. Motif analysis was carried out on the SMXL homologs using the MEME server (http://meme-suite.org/), keeping the minimum motif length at 4–10 amino acids and present at least once per sequence. Selected significant motifs were further screened for conservation across the plant kingdom. All SMXL members and clade-specific homologs were aligned using MAFFT (Auto mode) and sequence logos were generated using WebLogo3 (Crooks *et al.*, 2004). All motifs were represented as per PROSITE patterns (https://prosite.expasy.org/) (see Supplementary Table S3). Cavity prediction was done using the Computed Atlas of Surface Topology of proteins (CASTp) server (Dundas *et al.*, 2006) with the default probe radius of 1.4 Å. Class-specific motifs together with predicted cavities were visualized in PyMOL (https://pymol.org/2/).

Deposition of phylogenetic tree

The phylogenetic tree of SMXLs generated using RAxML is shared on the iTOL server (Letunic and Bork, 2016), and is available at https://itol.embl.de/tree/1472519069309921519489548. It is also presented in a Nexus format as Supplementary Dataset S1.

Results

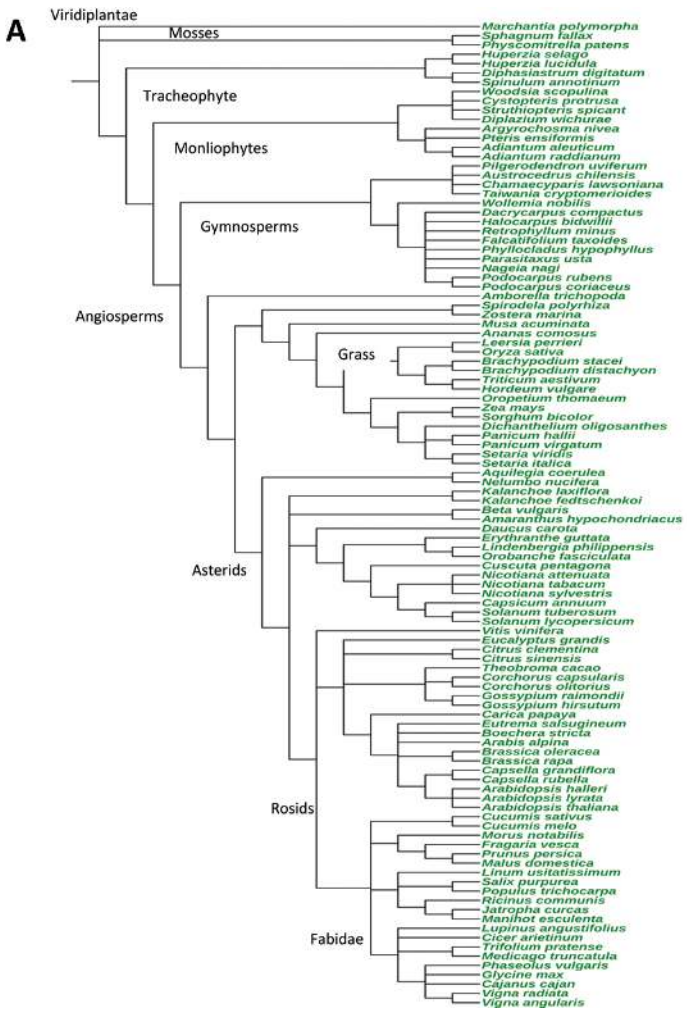
Identification and annotation of SMXLs in the plant kingdom

To explore and increase our understanding of the evolution of strigolactone signalling in the plant kingdom, we examined the target-repressor SMXLs in the green lineage. We mined the reference proteomes of completely sequenced plant genomes and all the species available in Phytozome v12.1.5 using the homology-based method BLAST (the data-mining scheme is presented in Supplementary Fig. S1). Homologs were also mined and identified from parasitic plants (*Orobanche*, *Cuscuta* spp.) of order Lamiales with the data available in the 1KP project. In total, 510 homologous sequences were mined

from 106 species (Fig. 1). Partial sequences from lycophytes were included with at least >50% coverage (details in Fig. 1, Supplementary Table S1). No significant hits were identified in chlorophyte members or in *Selaginella moellendorffii*. We found that the number of SMXL genes varied considerably (Supplementary Table S2), from one in the liverwort *Merchantia polymorpha* to 11 in *Zea mays* (monocot), with the highest number (14) found in *Brassica rapa* (eudicot). A common nomenclature was adopted while annotating of SMXLs in all the plant members, as follows. Arabidopsis and rice SMXL genes were designated as previously annotated in the published literature and used as a reference for annotation of homologs in other species. The orthologs of the SMXL genes from other species were named with three letter acronyms of the taxa followed by annotation (Supplementary Table S1)

Phylogenetic classification and evolution of plant SMXL homologs

The evolutionary history of the SMXL gene family was inferred using maximum-likelihood (ML) and Bayesian inference (BI). The results from both methods pointed to strong support for independent expansion of SMXLs in angiosperms (Fig 2, Supplementary Figs S2, 3). The robustness of defined clades was strongly supported by multiple statistical support measures. Only full-length homologs (505) were used for phylogenetic analyses to avoid alignment bias. We obtained similar topologies for both the ML and BI methods (see Supplementary Fig. S3) with high support values. The phylogenetic tree from neighbour-joining (NJ) analysis had many unresolved polytomies of SMAX1 in angiosperm species. The topology of the ML tree was not affected by including the partial sequences from lycophytes when constructing the phylogenetic trees (Supplementary Fig. S4). According to the ML phylogenetic tree, the SMAX1 homologs from liverworts and mosses could be grouped together as an ancient SMAX1 clade, and angiosperm SMXLs homologs were distributed into four clades/classes (Fig. 2). Increased availability of more sequenced genomes from non-angiosperm species will help to mine more SMXL members in the future. Gymnosperm SMAX1 family members were grouped together with the angiosperm SMAX1 clade I, where monilophyte clades may have emerged before angiosperm clades I and II. The lycophyte SMAX1 clade emerged before further division into sub-clades, which is consistent with the upstream receptor D14/KAI2 (Bythell-Douglas *et al.*, 2017).



B

Division	No. of SMXLs
Liverworts	1
Bryophyta	3
Lycophyta	5 (Partial)
Monilophytes	13
Gymnosperms	14
Basal Angiosperms	3
Monocots	102
Dicots	374
Total	510

Fig. 1. Sampling of SMXL homologs from the plant lineage. (A) Species tree of the sampled lineages generated using the NCBI common tree. (B) Number of homologs mined in different divisions of the plant kingdom. (This figure is available in colour at *JXB* online.)

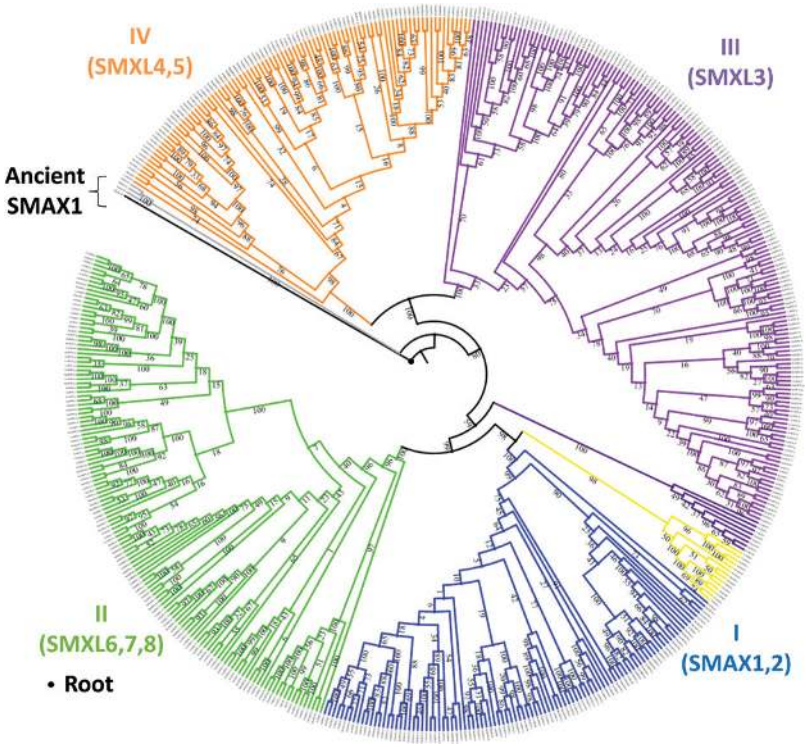


Fig. 2. Maximum-likelihood phylogeny of the SMXL family showing four robust phylogenetic clades. A total of 505 SMXL sequences from 101 species were used to infer the phylogenetic relatedness among the homologs. Clade I contains representatives of SMAX1 and 2, clade II contains representatives of SMXL6, 7, and 8, clade III contains representatives of SMXL3, and clade IV contains representatives of SMXL4 and 5. Sub-clades of clade I (dark blue and yellow) represent monilophytes and gymnosperm SMAX1s, respectively. RAxML was used to reconstruct the phylogeny of SMXLs. The tree is rooted with liverwort SMAX1.

To gain further insight into the evolutionary expansion of the SMXL family in the different plant genomes during course of evolution, we focussed our analysis on 58 completely sequenced genomes (see Supplementary Fig. S5). The reconciliation of the SMXL gene tree with the species tree indicated that the SMXL family might have evolved through 128 duplications and between 276–284 loss events. However, the whole-genome duplication and whole-genome triplication may have played an important role in the expansion of this family (Supplementary Fig. S5). We also tested the robustness of our analysis with altered MSA with different guidance scores to remove the poorly aligned regions. The ancient SMAX1 underwent at least three major duplications before the divergence of basal angiosperms. The expansion of SMXLs in lower non-angiosperm species will become clearer when more genomes have been sequenced. Reconciliation of the SMXL gene tree with the species tree indicated about 18 duplication events in monocots and >30 rounds of duplication events in eudicot lineages.

The robust phylogenetic clustering pattern from all methods (Fig 2, Supplementary Figs S2, 3) suggested that the angiosperm SMXLs can be classified into four major clades comprising of homologs of SMAX1,2 as clade I, SMXL6,7,8 as clade II, SMXL3 as clade III, and SMXL4,5 as clade IV (Fig. 2). From this analysis, we can suggest that the diversification in the SMXL family occurred either in vascular or seed plants in a similar way to that of the signalling receptor D14/KAI2 (Delaux *et al.*, 2012; Bythell-Douglas *et al.*, 2017; Waters *et al.*, 2015, 2017). The phylogenetic tree revealed that clades I, II and clades III, IV diverged independently from the ancient SMAX1 clade. The nodes of divergence (Fig. 2, Supplementary Fig. S6) were especially apparent with respect to their taxonomic/speciation events in clades II, III and IV. For example, *Amborella trichopoda* (lower angiosperms) emerged separately from the other species of monocots and eudicots to form a separate clade, and further diverging resulted in two separate sub-clades of monocots and eudicots. Homologs from Brassicales form a monophyletic sub-clade in clades/classes I–IV, and they retained the largest number of homologs (see Supplementary Table S2). In clade II (SMXL6,7,8), eudicots further formed sub-groups into two separate clades separating SMXL6,7 and SMXL8 (Supplementary Fig. S6). Interestingly, we also observed two separate sub-clades in clade III in both monocots and eudicots (Supplementary Fig. S6). Using the molecular clock test (Supplementary Fig. S8), we found that all four clades have evolved at different evolutionary rates (the null hypothesis of equal evolutionary rate among the clades was rejected, P=0.000E+000). This analysis revealed that after expansion into four phylogenetic clades, the different evolutionary rates may have contributed significantly to the changes in sequences so that they are functionally diverged, and hence neo-functionalized.

Overall, we may infer that SMXL first appeared in liverworts, underwent subsequent duplication in the taxonomic clade of green plants, and evolved into four distinct angiosperm phylogenetic clades/classes.

Conserved motifs in SMXL family members

Phylogenetic analysis of the SMXL homologs yielded four different clades/classes. All SMXLs were characterized by a domain architecture of a double Clp-N motif and a P-loop containing a nucleoside triphosphate hydrolase (Jiang *et al.*, 2013; Stanga *et al.*, 2013; Zhou *et al.*, 2013, 2016). Recent advances in the field of strigolactones have highlighted the importance of the ethylene-response factor-associated amphiphilic repression (EAR) and RGKT motifs in signalling. In Fig. 3, we show the phylogenetic tree clades together with the retention of known functionally characterized motifs. The RGKT motif is present in all angiosperm SMXLs except SMXL3,4,5. We also looked for the conservation of the RGKT and EAR motifs in other non-angiosperm species (Fig. 3C), and found that they were highly conserved in the ancient clade (liverwort and mosses MAX1) and in other groups such as lycophtes, monilophytes, and gymnosperms. The conserved RGKT motif that was prominent in lower plant lineage was completely absent in SMXL3,4,5, where the EAR motif was highly conserved. Three extra EAR motifs in clade/class II were identified specifically in monocots (see Supplementary Fig. S9). We also looked for clade-specific motifs outside the double Clp-N and P-loop NTPase domain region (Supplementary Table S3). Motif analysis was performed using MEME (sequence motif analysis; see details in Methods section) and gave us the ability to identify the specific motifs for different classes. MEME analysis of the Arabidopsis SMXL homologs revealed an extra motif in SMXL6,7,8, but not in other homologs (Supplementary Fig. S10). This motif (M2) was found to be highly conserved in class/clade II of flowering plants (Fig. 4). Further investigation of class-specific motifs using MEME and WebLogo3 on the alignment showed important conserved regions shared by members of different clades, and we identified the following motifs: M1 (clade I), M3 (clade III), and M4 and M5 (clade IV) (Fig. 4). M1 and M2, specific for clades I and II, respectively, were completely absent in ancient SMAX1, but M1 was found to be conserved in other tracheophytes (Supplementary Fig. S11). The presence of the M1 motif was not so clear in lycophytes due to lack of availability of complete sequences. M4, which was specific for clade IV, was present in mosses, lycophytes, and gymnosperms only; it was completely absent in monilophytes (Supplementary Fig. S12). None of these motifs were found to be annotated in the Pfam database (https://pfam.xfam.org/). These highly conserved motifs, specific for different clades, further supported the classification

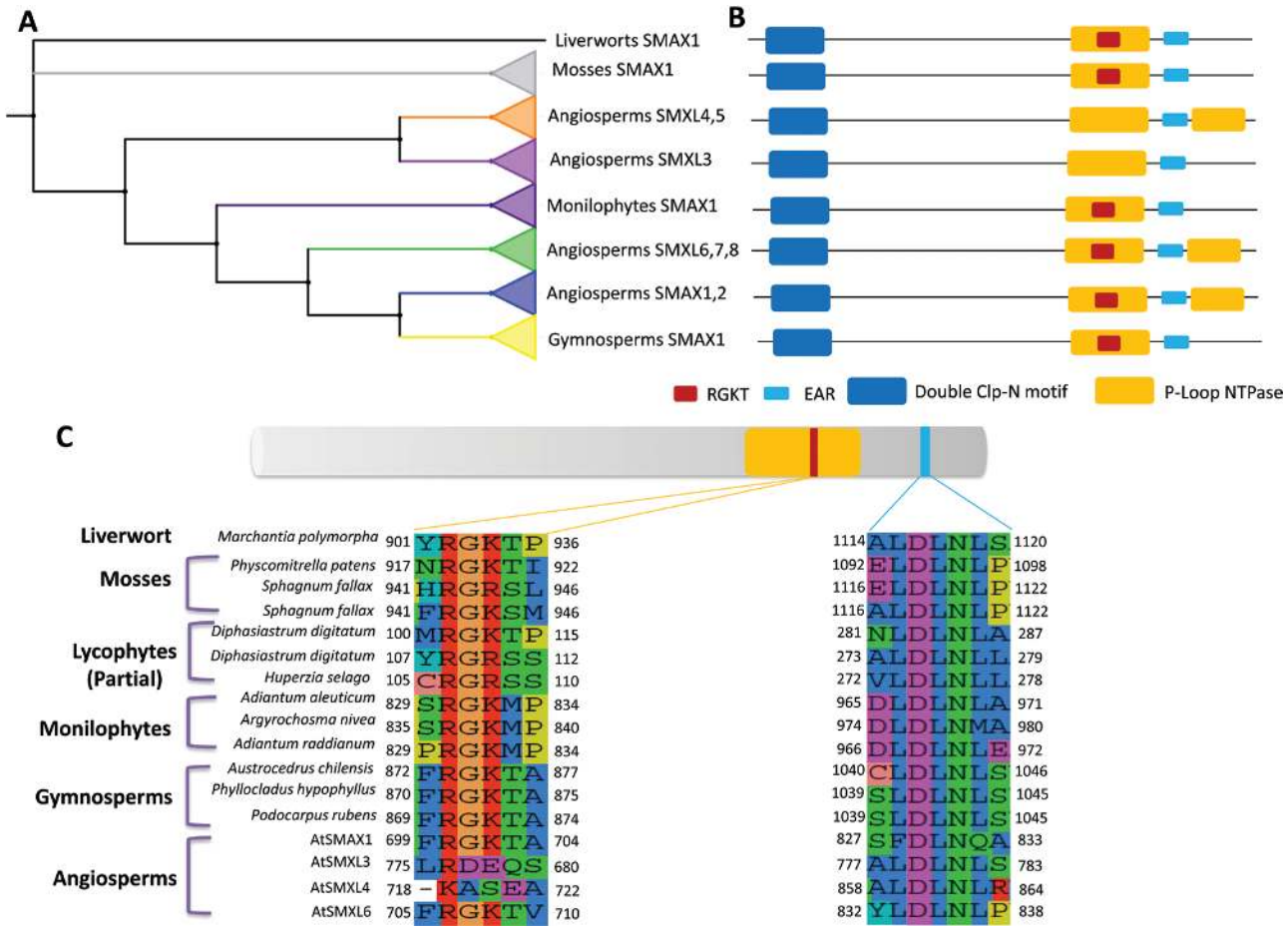


Fig. 3. SMXL domain classification and conservation of functional motifs. (A) Phylogram showing the classification of the SMXL gene family with the major clades labelled. (B) Typical SMXL consists of a double Clp-N motif and p-loop NTPase. The approximate positions of the function motifs (RGKT and EAR motif) are indicated. The RGKT motif is absent in the angiosperm SMXL3 and SMXL4,5 clades. (C) Conservation of the RGKT and EAR functional motifs in non-angiosperm species. The representation of domains is approximate and not to scale.

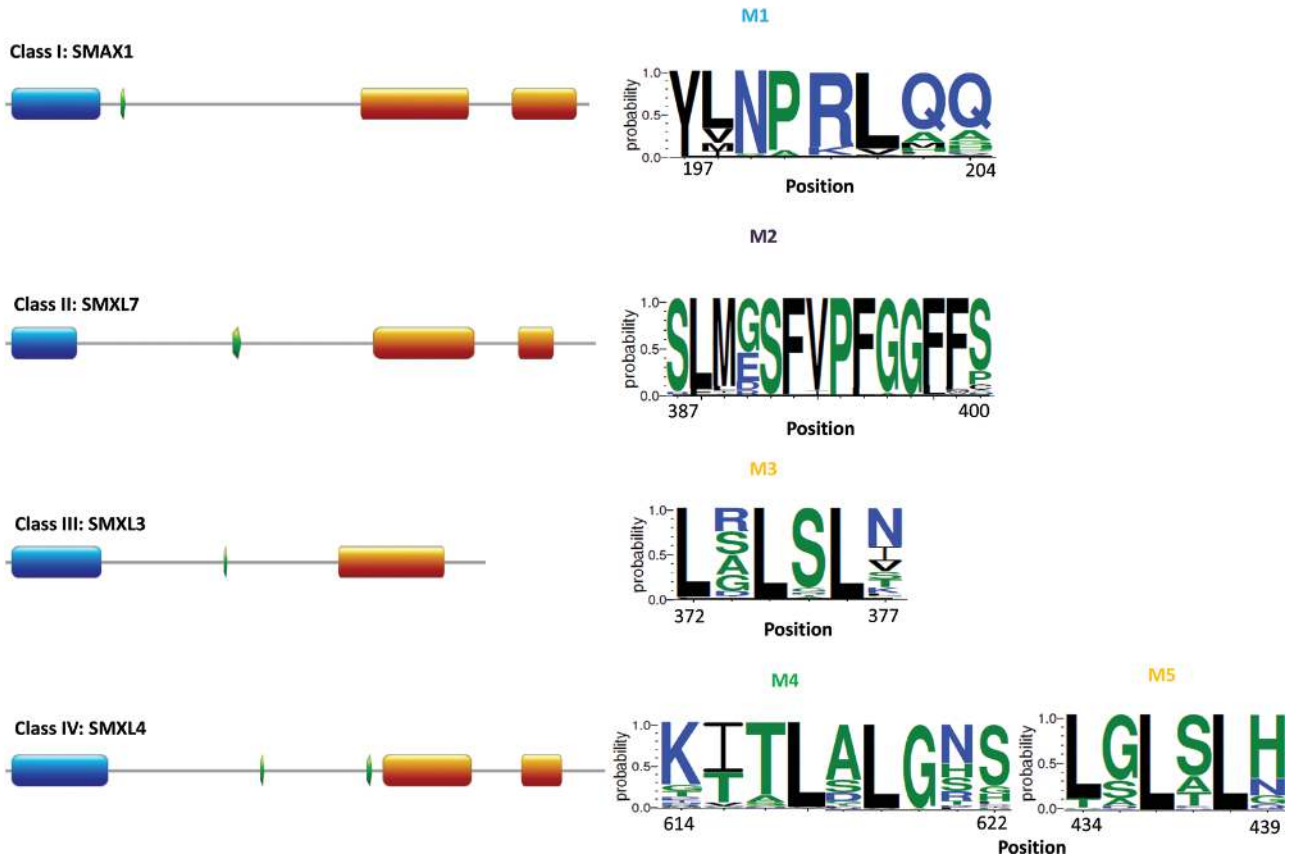


Fig. 4. Conservation of functional motifs in clades I, II, III, and IV of SMXLs in the plant kingdom. Signalling-specific motifs of strigolactone and karrikin were identified and are well-conserved in all the species of the respective clades. M1, M2, and M4 motifs are specific for class I, II, IV, respectively, and EAR-like motif (M3, M5). Positions shown are with respect to SMAX1 (class I), SMXL7 (class II), SMXL3 (class III), SMXL4,5 (class IV) of *Arabidopsis thaliana*.

of SMXLs that we have presented here. Interestingly, we also found extra EAR motifs in other clade/classes. EAR-like motifs (LXLXLX) were absent in clades I and II (except for monocots), but were present in clades III, IV, and in other clades of gymnosperms, monilophytes and mosses (Supplementary Fig. S12). They could potentially be important for protein binding, similar to that of monocot D53 (EAR motifs) (Ma *et al.*, 2017). Intriguingly, these conserved regions were either part of or near to the cavities predicted by CASTp (Supplementary Fig. S13). Overall, the conservation of these additional motifs in their respective clades may imply importance for their functional specificity.

Early ancestors of SMXLs

Phylogenetic and sequence motif analyses gave deep insights into the possible evolution and functional diversification of SMXL homologs from ancient SMAX1. However, they indicated that SMXL homologs were completely absent in the lower plant groups. Hence the exact origin may not be predicted confidently. Previous studies have found that SMXLs have weak homology with ClpB chaperonins (Jiang *et al.*, 2013; Stanga *et al.*, 2013). The structure of representative AtSMXLs of each class was modelled using I-TASSER (see Supplementary Table S4). Sequence and structural homology searches were performed in the NCBI and RCSB PDB (http://www.rcsb.org/) databases, respectively. The closest hits included the ClpB proteins (Zhou *et al.*, 2013, 2016) from Arabidopsis, bacteria, and yeast. These proteins are characterized by a double Clp-N motif and a P-loop, which are present in the nucleoside triphosphate hydrolase superfamily and in the molecular machinery involved in protein quality control. They share a similar domain architecture with SMXLs (although the domain size is variable) (Fig. 5A, Supplementary Table S7). However, the significance of evolutionary conservation of those domains (i.e. double Clp-N and P-loop NTPase) for the strigolactone signalling function of SMXLs has yet to be clarified. Their sequence identity was low (20–30%) but they showed high structural homology between them (Fig. 5B, C, Supplementary Tables S5, 6). The phylogenetic (NJ) tree based on the sequence distance matrix (Fig. 5B) indicated that AtClpB was the closest to all SMXLs compared to the other heat-shock chaperonins. Structural comparisons were also made between all the proteins using TM-Align, and root-mean-square deviations from atomic positions (RMSD, Å) were visualized in the form of a heat map (Fig. 5C). Representatives from all classes were closer to ClpB of *Thermus thermophilus* (*THET8*) than compared to each other, except for class III (SMXL3). Classes I, II, and IV shared the same fold in the protein structure, with template modelling (TM) scores 0.97, 0.95, 0.97, respectively (Supplementary Table S6), normalized against the length of

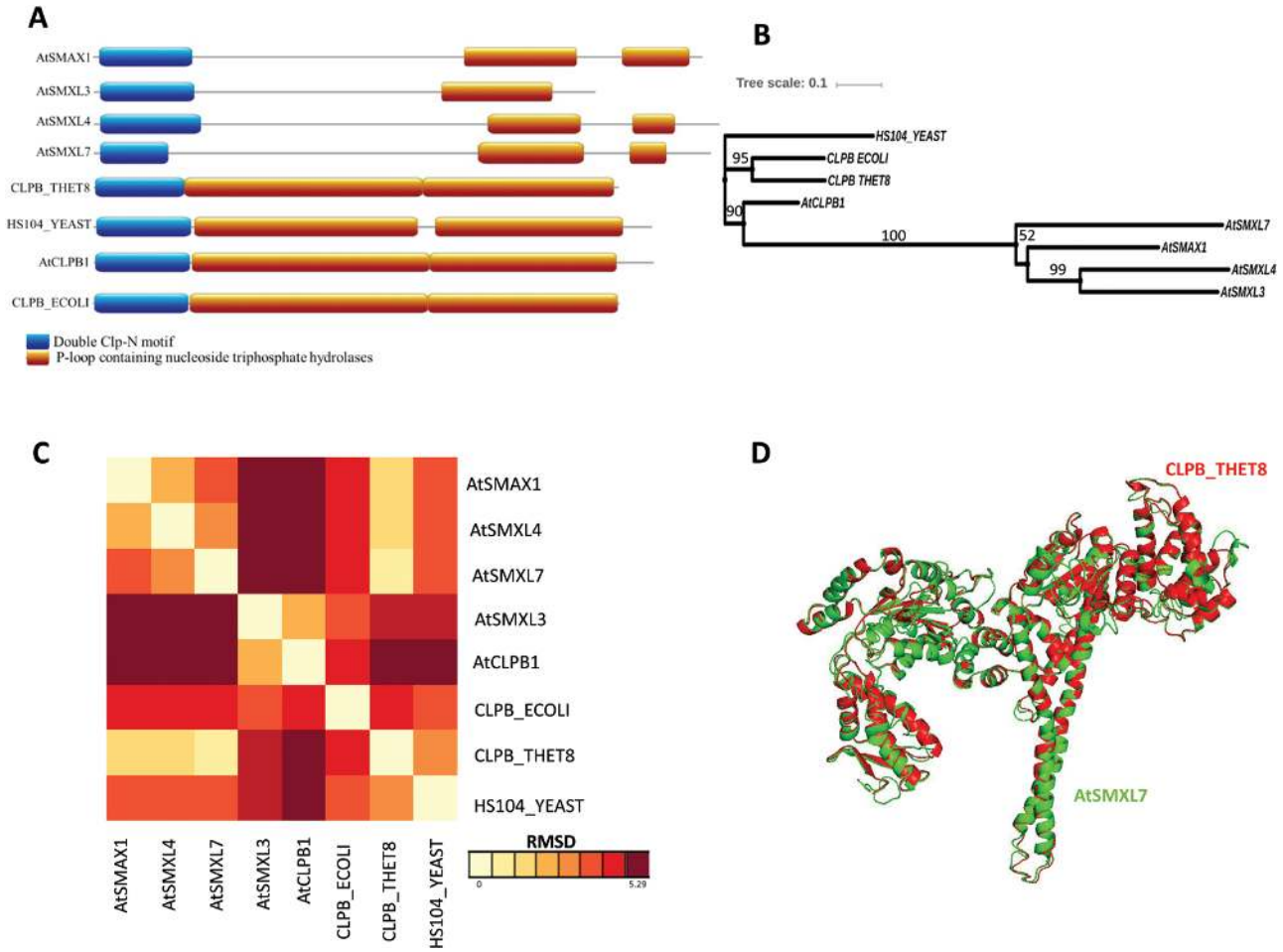


Fig. 5. Comparative sequence and structural analysis of representative SMXLs from each class relative to *E. coli*, *Tthermophilus*, *Arabidopsis*, and HS104 of yeast. (A) Comparison of primary structures of SMXLs with ClpB proteins. Domains were identified using the SCOP database and scaled with respective to their length. (B) Neighbour-Joining (NJ) clustering of SMXLs with ClpB proteins showing the sequence relatedness. (C) Heatmap of root-mean-square deviations from atomic positions (RMSD, Å) values showing the structural diversity among the representative SMXLs according to the colour scale. (D) Structural similarity between AtSMXL7 and ClpB of *Thermococcus* with RMASD score of 1.08Å.

ClpB of THET8 (see Fig. 5D for a superimposition of SMXL7 and ClpB of THET8). Unlike the other classes, class III shared the same fold with AtClpB, with a TM-score of 0.79 normalized against AtClpB (Supplementary Table S6). The class III (AtSMXL3) structure was different to other SMXLs (Supplementary Table S5), which suggests that it may be functionally different from the other classes. These sequence and structural similarity comparisons indicated a divergent evolution of SMXLs.

Discussion

Diversification and origin of strigolactone signalling

SMXL repressors play important roles in various aspects of plant physiology. Evidence suggests distinct physiological roles for different SMXL orthologs (Stanga *et al.*, 2013, 2016; Soundappan *et al.*, 2015; Wang *et al.*, 2015; Bennett *et al.*, 2016; Wallner *et al.*, 2017; Wu *et al.*, 2017); however, the basis of their functional diversification has been largely unknown. Here, we performed a genome-wide comparative analysis among different lineages and found that evolution of specific SL signalling targets occurred in angiosperm SMXLs. The angiosperm SMXL gene family can be divided into four distinct phylogenetic clades/classes. Furthermore, we found an ancient SMAX1 clade of liverwort and mosses, together with sub-clades of gymnosperms, monilophytes, and lycophytes (Fig. 2, Supplementary Figs S3, 4). However, we found that the root of SMXL evolution lies in the prokaryotic lineages. Using comparative sequence and structure analyses of ClpB of Arabidopsis, bacteria (*Escherichia coli* and *T. thermophilus*), and yeast (*Saccharomyces cerevisiae*), we found weak sequence homology of SMXLs with the ClpB chaperonins, but all the representative AtSMXL (1,3,4,7) members had a close structural similarity with the bacterial ClpB (Fig. 5C, D) except for AtSMXL3, which was closer to AtClpB. Thus, AtSMXL3 may be a somewhat divergent relative of SMXL members, and the possibility of SMXLs retaining the functions of ClpB remains to be determined. Earlier studies on the upstream receptor of SMXLs, D14/KAI2, demonstrated neo-functionalization and structural features required for SL perception (Bythell-Douglas *et al.*, 2017). We identified an early SMXL in non-vascular plants, liverworts, and mosses (Fig. 2, Supplementary Figs S3, 4). It had high identity with

Arabidopsis SMAX1 (60–65%), and we named it as the ancient form of SMAX1. Since we could not identify any SMXL family members in the charophytes, the first appearance of SMXLs would be in liverworts and mosses, which is consistent with developmental responses (Proust *et al.*, 2011; Delaux *et al.*, 2012).

Ancient SMXLs have undergone three major gene duplications (see Supplementary Fig. S5), a primary source of new genes with novel or altered functions. SLs in *Physcomitrella patens* regulate protonema branching and 'quorum sensing', and function as hormones (Proust *et al.*, 2011). There is some mixed evidence regarding SL receptors (DDK proteins) that suggests mosses use SLs as developmental regulators (Lopez-Obando *et al.*, 2016; Bythell-Douglas *et al.*, 2017) but not for rhizosphere communication. Our phylogenetic classification of the angiosperm SMXL gene family turned out to be in agreement with the diversity in the molecular functions among the members of the four clades as evidenced from various experimental data, except for clade/class III which is yet to be characterized completely (Soundappan *et al.*, 2015; Bennett *et al.*, 2016; Liang *et al.*, 2016; Stanga *et al.*, 2016; Wu *et al.*, 2017). Members of clade I (SMXL 1,2) have been shown to be involved in the karrikin pathway (Stanga *et al.*, 2016) and members of clade II (AtSMXL 6,7,8 and D53 in rice) in the strigolactone pathway (Zhou *et al.*, 2013, 2016; Soundappan *et al.*, 2015; Liang *et al.*, 2016), whilst members of clades III and IV are independent of these two pathways and are involved in phloem formation (Végh *et al.*, 2017; Wallner *et al.*, 2017). SMXL 4,5 repression in *dcl4* mutants resulted in over-accumulation of starch and anthocyanin together with defective phloem transport (Wu *et al.*, 2017). Although a proto-KAI2 protein was identified in charophyte algae, we did not identify any SMXLs, which raises the question of the existence of alternative target proteins other than SMXLs in charophyte algae. SMXLs were identified throughout the land plants and a single SMAX1-like protein was found in lower organisms (liverworts, mosses, lycophytes, and monilophytes). This pattern of evolution is very much like that of the interacting protein partners (DDK lineage) arising from neo-functionalization of the KAI2-family of receptors. The fact that these signalling components (DDK and SMXLs) shared similar evolutionary patterns probably contributed to the evolution of regulatory networks from lower to higher plants. The expansion pattern of the SMXL gene family reflects its importance during the evolution of plants and also its complexity. The expansion of SMXLs near the origin of the angiosperms indicates the necessity of expanding the hormonal system and the importance of its diversification to cope with the development of increasingly complex architecture. The evolution of angiosperms led to organised architecture in the anatomy of plants with differentiation into the system of tissues. Gene duplication followed by accumulation of diverse residues at important positions may lead to functional diversity among the duplicates. These neo-functionalizations in SMXL members might have co-evolved with the necessity for diverse mechanistic roles in the regulation of shoot branching and root architecture.

Conservation of functional motifs

Although SMXL members share the same domain architecture, they are involved in different signalling processes. The subtle mechanisms of differential recognition and their functional diversification remain areas that need to be addressed. Recent studies have highlighted the importance of the functional motifs RGKT and EAR in the peptide sequences (Soundappan *et al.*, 2015; Liang *et al.*, 2016; Ma *et al.*, 2017). Both focused and extensive analyses of the sequence motifs of individual clades/classes have suggested clues with regards to possible functional diversification. The organization of signature sequences plays a major role in recognition of the target signalling processes. Three EAR motifs were present in class/clade II (monocots), of which two have been found to be functionally important for interaction with TOPLESS domains (TPDs) (Ma *et al.*, 2017), mediating TPD oligomerization and nucleosome interaction. Class II SMXL proteins have been demonstrated to be involved in SL perception via the D14-SCF^{MAX2} pathway (Zhou *et al.*, 2013, 2016; Soundappan *et al.*, 2015; Liang *et al.*, 2016; Waters *et al.*, 2017), whereas class I SMXLs are presumed to physically interact with KAI2-SCF^{MAX2} in the karrikin-like (KL) signalling pathway (Stanga *et al.*, 2016). Given the comprehensive experimental evidence for the evolution of MAX2, D14/KA2, and D53/SMXL7 (Jiang *et al.*, 2013; Zhou *et al.*, 2013, 2016; Liang *et al.*, 2016), the RGKT motif in classes I and II may be necessary for SCF^{MAX2}-mediated degradation. Presuming that this is true for KL signalling pathway as well, we observed that the RGKT motif was well conserved from the ancient SMAX1 that was identified in liverworts and mosses (Fig. 3). Our analysis suggests that ancient SMAX1 from non-vascular plants was the first in the lineage and that it was a proteasome target of eu-KAI2-SCF^{MAX2}-mediated signalling. The motif similarity study of the ancient SMAX1 clade with different Arabidopsis SMXLs showed that, apart from the EAR motif, ancient SMAX1 not only retained the RGKT motif but also the M4, EAR-like, motif in classes III and IV (see Supplementary Fig. S12). Interestingly, the extra EAR motif was absent in liverworts whereas other non-angiosperm plants carried the extra EAR-like (LxLxL) motif (Supplementary Fig. S12). Recently, the angiosperm DDK lineage has been subdivided into D14, KAI2, and DLK23 (with sub-clades DLK2,3) (Bythell-Douglas *et al.*, 2017). An assumption can be made that clades/classes III and IV are probable targets of the DLK2 subclade in the DDK lineage in a MAX2-independent pathway (Végh *et al.*, 2017) with the RGKT motif being absent. Indeed, the functional specificity of SMXLs has been observed in physiological studies of Arabidopsis and rice (Stanga *et al.*, 2013, 2016; Bennett and Leyser, 2014; Soundappan *et al.*, 2015; Wang *et al.*, 2015; Liang *et al.*, 2016). The perception of SLs and KARs evolved independently and diversified in the later stages of land plants (Fig. 6). The early clade/class I-specific signature was observed in gymnosperms, which is consistent with the partial rescue of the Arabidopsis *kai2* mutant by the SmKAI2A receptor from *Selaginella moellendorffii* (Waters *et al.*, 2015). Incomplete sequences from lycophytes limited the determination of the precise point of the appearance of the clade/class-specific signature motifs that were identified

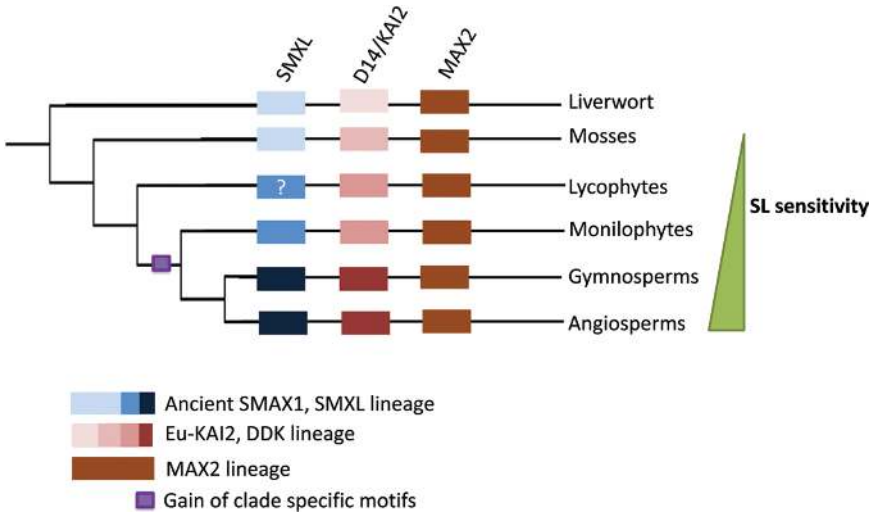


Fig. 6. Model for evolution of the signalling genes of the strigolactone pathway. SMXLs might have evolved in parallel with D14/KAI2 and MAX-2 across the plant kingdom. The increasing diversity of the protein families is represented in the form of a colour gradient (light→dark): SMXL, blue; D14/KAI2, pink; MAX2, brown. Response to strigolactone (SL) is first observed in mosses, and a gain of specificity for the SMAX1 motif is first observed in gymnosperms.

in flowering plants. The idea that there is functional specialization of SMXL members that derives from their phylogenetic classification and their retention of motifs in each clade/class member would provide the rationale for a future detailed expression study of each member in Arabidopsis. It is noteworthy that although multi-member SMXLs and their interacting protein partners D14/KAI2 have been identified in land plants, knowledge of their physiological roles is very limited. A comprehensive evolutionary history (Bythell-Douglas *et al.*, 2017) and present experimental evidence regarding SMXL and DDK proteins might suggest that SL signalling was initially involved in developmental roles and that it soon developed into a rhizosphere communication role (alongside its developmental role) in the land plants (Bouwmeester *et al.*, 2007; Yoneyama *et al.*, 2007*a*; Proust *et al.*, 2011; Rasmussen *et al.*, 2012; Stanga *et al.*, 2013, 2016; Zhou *et al.*, 2013, 2016; Wang *et al.*, 2015; Lopez-Obando *et al.*, 2016). In a tissue specific-expression study discussed previously (Stanga *et al.*, 2013), SMAX1 expression prevailed over the other SMXLs in most of the tissues (seeds, seedlings, roots, green leaves, senescent leaves, and axillary stems). SMXL3 was predominant in roots while other SMXLs had relatively low expression. SMXL7 was highly expressed in axillary branches and SMXL4,5 had low expression in all the tissues. These tissue-specificities might also correspond with the phylogenetic classification. However, biochemical studies need to be conducted for clade/class-specific signature motifs in order to confirm the specificity of their physiological functions.

With the increase in angiosperm species diversity, it is conceivable that expansion of both SMXLs and their receptor DDK lineage provided the greater complexity of hormonal signalling that was necessary to co-ordinate the more complex developmental program of flowering plants. Rapid advances in technologies should enable us to address the functional specificity and biochemical similarity of these proteins. These future studies will improve our understanding of the evolution of the strigolactone signalling pathway and its significance in plant development.

Supplementary data

Supplementary data are available at *JXB* online.

Table S1. List of SMXLs used in the study and their accessions.

Table S2. List of plant species and number of SMXLs found in them, and number of species from each plant division in which SMXLs were identified.

Table S3. Motifs detected in all SMXLs and their sub-clades.

Table S4. Confidence (C) score of the modelled proteins in I-TASSER.

Table S5. Structural similarity between AtSMXLs and their homologous ClpB proteins represented in RMSD.

Table S6. TM-scores of AtSMXLs with their homologous ClpB proteins, normalized to the length of SMXL proteins.

Table S7. Comparisons of conserved Clp-N and P-Loop NTPase domain length for selected species.

Fig. S1. Methodology used for the mining of SMXL sequences and their phylogenetic analysis.

Fig. S2. ML tree generated using IQ-Tree with clade support.

Fig. S3. Phylogenetic classification of SMXLs using ML and BI methods.

Fig. S4. ML tree generated using IQTree for SMXLs members together with partial sequences of lycophytes.

Fig. S5. Expansion of SMXLs during plant evolution.

Fig. S6. Phylogram showing pruned trees of all angiosperm clades.

Fig. S7. SMXL gene tree reconciled with the species tree to identify the duplication and loss events at each branch using Notung v2.9.

Fig. S8. Log-likelihood parameter of the molecular clock test.

Fig. S9. Three EAR motifs specific to the monocot subclade of class II.

Fig. S10. Motif analysis using the MEME server.

Fig. S11. Origin of the clade/class I signalling-specific motif.

Fig. S12. Distribution of the M4 motif specific for class IV.

Fig. S13. Arabidopsis SMXLs from each clade/class with cavities predicted using CASTp server.

Supplementary Dataset S1. Phylogenetic tree of the SMXL gene family generated by RAxML in Nexus tree format.

Acknowledgements

This project received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions and it is co-financed by the South Moravian Region under grant agreement No. 665860 (SS). Access to computing and storage facilities owned by parties and projects contributing to the national grid infrastructure, MetaCentrum, provided under the program 'Projects of Large Infrastructure for Research, Development, and Innovations' (LM2010005) was greatly appreciated (RSV). The project was funded by The Ministry of Education, Youth and Sports/MES of the Czech Republic under the project CEITEC 2020 (LQ1601) (TN, TRM). JF was supported by the European Research Council (project ERC-2011-StG 20101109-PSDP) and the Czech Science Foundation GAČR (GA13-40637S). We thank Dr Kamel Chibani for active discussions on the evolutionary analysis and Nandan Mysore Vardarajan for his critical comments on the manuscript. This article reflects only the authors' views, and the EU is not responsible for any use that may be made of the information it contains.

References

Agusti J, Herold S, Schwarz M, *et al*. 2011. Strigolactone signaling is required for auxin-dependent stimulation of secondary growth in plants. Proceedings of the National Academy of Sciences, USA **108,** 20242–20247.

Akiyama K, Matsuzaki K, Hayashi H. 2005. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. Nature **435,** 824–827.

Al-Babili S, Bouwmeester HJ. 2015. Strigolactones, a novel carotenoid-derived plant hormone. Annual Review of Plant Biology **66,** 161–186.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME Suite: tools for motif discovery and searching. Nucleic Acids Research **37,** W202–W208.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. Nucleic Acids Research **43,** W39–W49.

Bennett T, Leyser O. 2014. Strigolactone signalling: standing on the shoulders of DWARFs. Current Opinion in Plant Biology **22,** 7–13.

Bennett T, Liang Y, Seale M, Ward S, Müller D, Leyser O. 2016. Strigolactone regulates shoot development through a core signalling pathway. Biology Open **5,** 1806–1820.

Bouwmeester HJ, Roux C, Lopez-Raez JA, Bécard G. 2007. Rhizosphere communication of plants, parasitic plants and AM fungi. Trends in Plant Science **12,** 224–230.

Brewer PB, Koltai H, Beveridge CA. 2013. Diverse roles of strigolactones in plant development. Molecular Plant **6,** 18–28.

Bythell-Douglas R, Rothfels CJ, Stevenson DWD, Graham SW, Wong GK, Nelson DC, Bennett T. 2017. Evolution of strigolactone receptors by gradual neo-functionalization of KAI2 paralogues. BMC Biology **15,** 52.

Chen H, Xiong L. 2009. Localized auxin biosynthesis and postembryonic root development in Arabidopsis. Plant Signaling & Behavior **4,** 752–754.

Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. Journal of Computational Biology **7,** 429–447.

Cook CE, Whichard LP, Turner B, Wall ME, Egley GH. 1966. Germination of witchweed (*Striga lutea* Lour.): isolation and properties of a potent stimulant. Science **154,** 1189–1190.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Research **14,** 1188–1190.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics **27,** 1164–1165.

Delaux PM, Xie X, Timme RE, *et al*. 2012. Origin of strigolactones in the green lineage. New Phytologist **195,** 857–871.

Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Research **34,** W116–W118.

Finn RD, Clements J, Arndt W, Mille BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015. HMMER web server: 2015 update. Nucleic Acids Research **43,** W30–W38.

Gomez-Roldan V, Fermas S, Brewer PB, *et al*. 2008. Strigolactone inhibition of shoot branching. Nature **455,** 189–194.

Goodstein DM, Shu S, Howson R, *et al*. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40,** D1178–D1186.

Gray WM. 2004. Hormonal regulation of plant growth and development. PLoS Biology **2,** E311.

Hua Z, Vierstra RD. 2011. The cullin-RING ubiquitin-protein ligases. Annual Review of Plant Biology **62,** 299–334.

Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics **26,** 680–682.

Humphrey AJ, Beale MH. 2006. Strigol: biogenesis and physiological activity. Phytochemistry **67,** 636–640.

Jiang L, Liu X, Xiong G, *et al*. 2013. DWARF 53 acts as a repressor of strigolactone signalling in rice. Nature **504,** 401–405.

Johnson MT, Carpenter EJ, Tian Z, *et al*. 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PloS ONE **7,** e50226.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences **8,** 275–282.

Kelley DR, Estelle M. 2012. Ubiquitin-mediated control of plant hormone signaling. Plant Physiology **160,** 47–55.

Kohlen W, Charnikhova T, Liu Q, *et al*. 2011. Strigolactones are transported through the xylem and play a key role in shoot architectural response to phosphate deficiency in nonarbuscular mycorrhizal host Arabidopsis. Plant Physiology **155,** 974–987.

Koltai H, Cohen M, Chesin O, *et al*. 2011. Light is a positive regulator of strigolactone levels in tomato roots. Journal of Plant Physiology **168,** 1993–1996.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution **33,** 1870–1874.

Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics **12,** 124.

Lee TH, Kim J, Robertson JS, Paterson AH. 2017. Plant Genome Duplication Database. Methods in Molecular Biology **1533,** 267–277.

Lee TH, Tang H, Wang X, Paterson AH. 2013. PGDD: a database of gene and genome duplication in plants. Nucleic Acids Research **41,** D1152–D1158.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research **44,** W242–W245.

Li Z, Defoort J, Tasdighian S, Maere S, Van de Peer Y, De Smet R. 2016. Gene duplicability of core genes is highly consistent across all angiosperms. The Plant Cell **28,** 326–344.

Liang Y, Ward S, Li P, Bennett T, Leyser O. 2016. SMAX1-LIKE7 signals from the nucleus to regulate shoot development in arabidopsis via partially EAR motif-independent mechanisms. The Plant Cell **28,** 1581–1601.

Lopez-Obando M, Conn CE, Hoffmann B, Bythell-Douglas R, Nelson DC, Rameau C, Bonhomme S. 2016. Structural modelling and transcriptional responses highlight a clade of *PpKAI2-LIKE* genes as candidate receptors for strigolactones in *Physcomitrella patens*. Planta **243,** 1441–1453.

Ma H, Duan J, Ke J, *et al*. 2017. A D53 repression motif induces oligomerization of TOPLESS corepressors and promotes assembly of a corepressor-nucleosome complex. Science Advances **3,** e1601217.

Matasci N, Hung LH, Yan Z, *et al*. 2014. Data access for the 1,000 Plants (1KP) project. Gigascience **3,** 17.

Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M, Kaufman S, O'Leary MA. 2015. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. Evolutionary Bioinformatics Online **11,** 43–48.

Nelson DC, Scaffidi A, Dun EA, Waters MT, Flematti GR, Dixon KW, Beveridge CA, Ghisalberti EL, Smith SM. 2011. F-box protein MAX2 has dual roles in karrikin and strigolactone signaling in Arabidopsis thaliana. Proceedings of the National Academy of Sciences, USA **108,** 8897–8902.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution **32,** 268–274.

Parker D, Beckmann M, Zubair H, Enot DP, Caracuel-Rios Z, Overy DP, Snowden S, Talbot NJ, Draper J. 2009. Metabolomic analysis reveals a common pattern of metabolic re-programming during invasion of three host plant species by *Magnaporthe grisea*. Plant Journal **59,** 723–737.

Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology **53,** 793–808.

Proust H, Hoffmann B, Xie X, Yoneyama K, Schaefer DG, Yoneyama K, Nogué F, Rameau C. 2011. Strigolactones regulate protonema branching and act as a quorum sensing-like signal in the moss *Physcomitrella patens*. Development **138,** 1531–1539.

Rasmussen A, Mason MG, De Cuyper C, *et al*. 2012. Strigolactones suppress adventitious rooting in Arabidopsis and pea. Plant Physiology **158,** 1976–1987.

Ronquist F, Teslenko M, van der Mark P, *et al*. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology **61,** 539–542.

Ruyter-Spira C, Al-Babili S, van der Krol S, Bouwmeester H. 2013. The biology of strigolactones. Trends in Plant Science **18,** 72–83.

Ruyter-Spira C, Kohlen W, Charnikhova T, *et al*. 2011. Physiological effects of the synthetic strigolactone analog GR24 on root system architecture in Arabidopsis: another belowground role for strigolactones? Plant Physiology **155,** 721–734.

Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Research **43,** W7–W14.

Seto Y, Kameoka H, Yamaguchi S, Kyozuka J. 2012. Recent advances in strigolactone research: chemical and biological aspects. Plant & Cell Physiology **53,** 1843–1853.

Smith SM, Li J. 2014. Signalling and responses to strigolactones and karrikins. Current Opinion in Plant Biology **21,** 23–29.

Soundappan I, Bennett T, Morffy N, Liang Y, Stanga JP, Abbas A, Leyser O, Nelson DC. 2015. SMAX1-LIKE/D53 family members enable distinct MAX2-dependent responses to strigolactones and karrikins in Arabidopsis. The Plant Cell **27,** 3143–3159.

Spallek T, Mutuku M, Shirasu K. 2013. The genus *Striga*: a witch profile. Molecular Plant Pathology **14,** 861–869.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30,** 1312–1313.

Stanga JP, Morffy N, Nelson DC. 2016. Functional redundancy in the control of seedling growth by the karrikin signaling pathway. Planta **243,** 1397–1406.

Stanga JP, Smith SM, Briggs WR, Nelson DC. 2013. SUPPRESSOR OF MORE AXILLARY GROWTH2 1 controls seed germination and seedling development in Arabidopsis. Plant Physiology **163,** 318–330.

Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics **28,** i409–i415.

Stone SL, Callis J. 2007. Ubiquitin ligases mediate growth and development by promoting protein death. Current Opinion in Plant Biology **10,** 624–632.

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research **44,** W232–W235.

Umehara M, Hanada A, Yoshida S, *et al*. 2008. Inhibition of shoot branching by new terpenoid plant hormones. Nature **455,** 195–200.

Végh A, Incze N, Fábián A, Huo H, Bradford KJ, Balázs E, Soós V. 2017. Comprehensive analysis of DWARF14-LIKE2 (DLK2) reveals its functional divergence from strigolactone-related paralogs. Frontiers in Plant Science **8,** 1641.

Waldie T, McCulloch H, Leyser O. 2014. Strigolactones and the control of plant development: lessons from shoot branching. The Plant Journal **79,** 607–622.

Wallner ES, López-Salmerón V, Belevich I, *et al*. 2017. Strigolactone- and karrikin-independent SMXL proteins are central regulators of phloem formation. Current Biology **27,** 1241–1247.

Wang L, Wang B, Jiang L, *et al*. 2015. Strigolactone signaling in Arabidopsis regulates shoot development by targeting D53-Like SMXL repressor proteins for ubiquitination and degradation. The Plant Cell **27,** 3128–3142.

Waters MT, Gutjahr C, Bennett T, Nelson DC. 2017. Strigolactone signaling and evolution. Annual Review of Plant Biology **68,** 291–322.

Waters MT, Nelson DC, Scaffidi A, Flematti GR, Sun YK, Dixon KW, Smith SM. 2012. Specialisation within the DWARF14 protein family confers distinct responses to karrikins and strigolactones in Arabidopsis. Development **139,** 1285–1295.

Waters MT, Scaffidi A, Moulin SL, Sun YK, Flematti GR, Smith SM. 2015. A *Selaginella moellendorffii* ortholog of KARRIKIN INSENSITIVE2 functions in Arabidopsis development but cannot mediate responses to karrikins or strigolactones. The Plant Cell **27,** 1925–1944.

Westwood JH, Yoder JI, Timko MP, dePamphilis CW. 2010. The evolution of parasitism in plants. Trends in Pant Science **15,** 227–235.

Wickett NJ, Mirarab S, Nguyen N, *et al*. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proceedings of the National Academy of Sciences, USA **111,** E4859–E4868.

Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, Chothia C, Gough J. 2009. SUPERFAMILY–sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Research **37,** D380–D386.

Wu YY, Hou BH, Lee WC, Lu SH, Yang CJ, Vaucheret H, Chen HM. 2017. DCL2- and RDR6-dependent transitive silencing of SMXL4 and SMXL5 in Arabidopsis *dcl4* mutants causes defective phloem transport and carbohydrate over-accumulation. The Plant Journal **90,** 1064–1078.

Xie X, Yoneyama K, Yoneyama K. 2010. The strigolactone story. Annual Review of Phytopathology **48,** 93–117.

Xie Y, Wu G, Tang J, *et al*. 2014. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. Bioinformatics **30,** 1660–1666.

Yamada KD, Tomii K, Katoh K. 2016. Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. Bioinformatics **32,** 3246–3251.

Yang J, Zhang Y. 2015a. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Research **43,** W174–W181.

Yang J, Zhang Y. 2015b. Protein structure and function prediction using I-TASSER. Current Protocols in Bioinformatics **52,** 5.8.1–5.8.15.

Yao R, Li J, Xie D. 2018. Recent advances in molecular basis for strigolactone action. Science China. Life Sciences **61,** 277–284.

Yao R, Ming Z, Yan L, *et al*. 2016. DWARF14 is a non-canonical hormone receptor for strigolactone. Nature **536,** 469–473.

Yoneyama K, Xie X, Kusumoto D, Sekimoto H, Sugimoto Y, Takeuchi Y, Yoneyama K. 2007a. Nitrogen deficiency as well as phosphorus deficiency in sorghum promotes the production and exudation of 5-deoxystrigol, the host recognition signal for arbuscular mycorrhizal fungi and root parasites. Planta **227,** 125–132.

Yoneyama K, Xie X, Yoneyama K, Takeuchi Y. 2009. Strigolactones: structures and biological activities. Pest Management Science **65,** 467–470.

Yoneyama K, Yoneyama K, Takeuchi Y, Sekimoto H. 2007b. Phosphorus deficiency in red clover promotes exudation of orobanchol, the signal for mycorrhizal symbionts and germination stimulant for root parasites. Planta **225,** 1031–1038.

Zhang Y, Skolnick J. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Research **33,** 2302–2309.

Zhou F, Lin Q, Zhu L, *et al*. 2013. D14-SCF(D3)-dependent degradation of D53 regulates strigolactone signalling. Nature **504,** 406–410.

Zhou F, Lin Q, Zhu L, *et al*. 2016. Corrigendum: D14-SCF(D3)-dependent degradation of D53 regulates strigolactone signalling. Nature **532,** 402.