

Published in final edited form as:

*New Phytol.* 2012 July ; 195(1): 23–31.

## Update on sumoylation: defining core components of the plant SUMO conjugation system by phylogenetic comparison

Maria Novatchkova<sup>1,2</sup>, Konstantin Tomanov<sup>3</sup>, Kay Hofmann<sup>4</sup>, Hans-Peter Stuiblé<sup>5</sup>, and Andreas Bachmair<sup>3</sup>

<sup>1</sup>Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, A-1030 Vienna, Austria

<sup>2</sup>Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Dr. Bohr-Gasse 3, A-1030 Vienna, Austria

<sup>3</sup>Department of Biochemistry and Cell Biology, Max F. Perutz Laboratories, Center for Molecular Biology, University of Vienna, Dr. Bohr-Gasse 9, A-1030 Vienna, Austria

<sup>4</sup>Institute for Genetics, University of Cologne, Zùlpicher Straße 47a, D-50674 Cologne, Germany

<sup>5</sup>Physical Engineering Department, University of Applied Sciences of Gelsenkirchen, August-Schmidt-Ring 10, D-45665 Recklinghausen, Germany

### Summary

The conjugation of the small ubiquitin-related modifier, SUMO, to substrate proteins is a reversible and dynamic process, and an important response of plants to environmental challenges. Nevertheless, reliable data have so far been restricted largely to the model plant *Arabidopsis thaliana*. The increasing availability of genome information for other plant species offers the possibility to identify a core set of indispensable components, and to discover species-specific features of the sumoylation pathway. We analyzed the enzymes responsible for the conjugation of SUMO to substrates for their conservation between dicots and monocots. We thus assembled gene sets that relate the Arabidopsis SUMO conjugation system to that of the dicot species tomato, grapevine and poplar, and to four plant species from the monocot class: rice, *Brachypodium distachyon*, *Sorghum bicolor* and maize. We found that a core set of genes with clear assignment in Arabidopsis had highly conserved homologs in all tested plants. However, we also observed a variation in the copy number of homologous genes, and sequence variations that suggested monocot-specific variants. Generally, SUMO ligases and proteases showed the most pronounced differences. Finally, we identified potential SUMO chain-binding ubiquitin ligases, pointing to an *in vivo* function of SUMO chains as degradation signals in plants.

### Keywords

Arabidopsis; monocots; pathway comparison; phylogenetic comparison; protein modification; signal transduction; small ubiquitin-related modifier SUMO

### Introduction

The modification of substrate proteins by covalent linkage to the small ubiquitin-related modifier, SUMO, occurs by a dedicated set of enzymes. The process is mechanistically similar to the conjugation of ubiquitin, the most prominent representative of a family of

small protein modifiers with conserved structure (Hochstrasser, 2009). SUMO is synthesized as a pre-protein that needs to be processed by SUMO proteases to expose a carboxyl-terminal diglycine motif. In a two-step reaction, the heterodimeric SUMO activating enzyme, SAE, forms a thioester between SUMO's terminal glycine (Gly) residue and an active site cysteine (Cys) of the enzyme. The process starts with the formation of an AMP-SUMO linkage from ATP and SUMO's carboxyl-terminal Gly residue. Under release of AMP, the activated SUMO carboxyl terminus is then transferred onto the active site Cys of SAE to form a thioester. The whole process involves dramatic conformational changes of the enzyme (Olsen *et al.*, 2010). Subsequently, activated SUMO is transferred to the active site Cys of the SUMO conjugating enzyme (SCE; also called UBC9 in some animals and fungi). SCE can conjugate SUMO to substrate proteins, resulting in an isopeptide linkage formed between the carboxyl-terminal Gly of SUMO and the  $\epsilon$ -amino group of a lysine (Lys) residue within the substrate. So far, conjugation has been observed exclusively to  $\epsilon$ -amino groups of Lys residues. This differs from the more complex ubiquitin conjugation machinery, where transfer to  $\alpha$ -amino groups, or to substrate Cys residues, has also been documented (cf. Vosper *et al.*, 2009, and references therein). *In vitro*, and probably also *in vivo*, SCE can modify many substrates in the absence of substrate specificity factors (SUMO ligases). Direct substrate interaction and modification by SCE usually depend on the presence of the short sumoylation consensus motif, consisting of a hydrophobic aliphatic amino acid, followed by the Lys residue to be modified, any amino acid and an acidic residue ( $\Psi$ KxD/E in one-letter code). Nonetheless, SUMO ligases play important roles *in vivo* to determine the substrate range and extent of sumoylation. Figure 1 summarizes the reaction steps of the SUMO conjugation cycle.

Functional studies in plants, as well as the characterization of sumoylation enzymes, have so far been restricted largely to *Arabidopsis thaliana* (for recent reviews, see Miura *et al.*, 2007a; Lois, 2010; Miura & Hasegawa, 2010; H. J. Park *et al.*, 2011). More recently, the first experimental data for the monocot plant rice were published (Park *et al.*, 2010; Thangasamy *et al.*, 2011; Wang *et al.*, 2011). SUMO conjugation has been shown to be essential in *Arabidopsis* (Saracco *et al.*, 2007), and work by several groups has demonstrated its importance for the integration of environmental inputs and for adequate reaction to stress conditions (Yoo *et al.*, 2006; Catala *et al.*, 2007; Conti *et al.*, 2008; Jin *et al.*, 2008; Chen *et al.*, 2011; Miura *et al.*, 2011). A significant number of plant-specific sumoylation substrates have been identified recently (Budhiraja *et al.*, 2009; Elrouby & Coupland, 2010; Miller *et al.*, 2010). With an impressive body of information available for *Arabidopsis*, but relatively little insight into SUMO conjugation in other plants, we wanted to understand which of the components identified in *Arabidopsis* are conserved in other plants, and which genes point to more divergent features of the pathway. The increasing number of sequenced plant genomes (for review, see Feuillet *et al.*, 2011) and plant gene databases (Martinez, 2011) provide promising tools for the characterization of complete pathways. For comparison with *Arabidopsis*, we used the assembled genome data of tomato (*Solanum lycopersicum*, genome size *c.* 800 Mb), grapevine (*Vitis vinifera*, genome size 487 Mb), poplar (*Populus trichocarpa*, genome size 550 Mb), rice (*Oryza sativa*, genome size 466 Mb), *Brachypodium distachyon* (genome size 270 Mb), *Sorghum bicolor* (genome size 697 Mb) and maize (*Zea mays*, genome size *c.* 2800 Mb). In the latter case, we expected to find genes with high similarity to the *Sorghum* homologs, but, as a result of a recent genome duplication, the number should be twice that of *Sorghum*. Surprisingly, we did not find this situation in most instances, which we ascribe to incomplete sequence availability/annotation in maize. Generally, annotation of the *Arabidopsis* and rice genomes is most advanced, whereas we found some genes not annotated in their full length in other species. We are nonetheless convinced that the survey provides a valid overview over the set of SUMO conjugation and deconjugation enzymes in plants.

We applied several different gene comparison algorithms in order to obtain robust results in the search for homologs of the Arabidopsis SUMO conjugation apparatus. Proteome sequences were obtained for the above-mentioned species from Phytozome v8.0 (Goodstein *et al.*, 2012; [phytozome.net](http://phytozome.net)) and analyzed using OrthoMCL v2 (Li *et al.*, 2003; <http://www.orthomcl.org>), Inparanoid v4.1 (Ostlund *et al.*, 2010; <http://inparanoid.sbc.su.se>), OMA v0.99 (Altenhoff *et al.*, 2010; <http://www.cbrg.ethz.ch/research/orthologous>), Roundup (Deluca *et al.*, 2006; [roundup.hms.harvard.edu](http://roundup.hms.harvard.edu)) and Phytozome v7.0. Table 1 lists categorizations and gene identifiers. Protein alignments of Table 1 entries are provided as Supporting information (Fig. S1 to Fig. S11).

## SUMO activating enzyme SAE

The SUMO activating enzyme is a heterodimer. The smaller subunit, SAE1, is represented by two genes in Arabidopsis: SAE1a and SAE1b (see Table 1). There is considerable difference in amino acid sequence, but both are competent for SUMO activation (Budhiraja *et al.*, 2009), and no functional differentiation has been reported so far. The larger subunit SAE2, which contains the active site Cys, is encoded by a single gene in Arabidopsis. Taken together, we found that all plants contained SAE genes in low copy numbers, suggesting that single copies of SAE1 and SAE2 were present in the common ancestor of monocots and dicots.

## SUMO conjugating enzyme SCE

Arabidopsis encodes a single SCE gene, SCE1 (Kurepa *et al.*, 2003; Novatchkova *et al.*, 2004). Another entry of previous surveys, annotated as a possible gene with similarity to SCE1 (At5g02240 in Kurepa *et al.*, 2003; Novatchkova *et al.*, 2004), probably consists of two distinct open reading frames. One is a presumed pseudogene with identifier At5g02244 in the most recent TAIR release; the other is an abscisic acid (ABA)-responsive steroid dehydrogenase (At5g02240). Neither of these loci is listed in Table 1. In contrast with Arabidopsis, all other plants of Table 1 encode at least two SCE genes. Monocots have additional SCE genes with a slightly different sequence (cf. SCE alignment provided as Supporting Information Fig. S3), which may be considered as a monocot-specific subgroup.

## SUMO

SUMO genes encode precursor proteins with carboxyl-terminal extensions. After extension cleavage by SUMO-specific proteases, the exposed, conserved carboxyl terminus is linked to enzyme active site Cys residues, and eventually to substrates. Some ‘non-canonical’ SUMO proteins have mutations in conserved residues of the carboxyl-terminal region. The functional implications of these changes are not yet fully understood, but decreased cleavage by SUMO-specific proteases is one of the consequences (Budhiraja *et al.*, 2009).

Arabidopsis SUMO1 and SUMO2 are each other’s paralogs, representing the most highly expressed, ‘canonical’ isoforms. Over-expression of either SUMO1 or SUMO2 is correlated with an attenuation of ABA-mediated growth inhibition, and combined mutation of both SUMO isoforms is lethal (Lois *et al.*, 2003; Saracco *et al.*, 2007). Arabidopsis contains six additional SUMO genes, SUMO3–8, and one pseudogene (Novatchkova *et al.*, 2004).

Among this group, there is good evidence for the participation of SUMO3 (At5g55170) in SUMO conjugation, whereas evidence for the conjugation of the other isoforms is scarce or nonexistent. SUMO3 is nonessential, and its expression level is lower than that of SUMO1 and SUMO2. SUMO1 and SUMO2, but not SUMO3, can form SUMO chains, and the isoforms also differ in their characteristics as substrates of desumoylating enzymes (Chosed *et al.*, 2006; Colby *et al.*, 2006). SUMO3 is elicitor inducible, and over-expression activates plant defense (van den Burg *et al.*, 2010).

Table 1 and Fig. S4 list the sequences identified as orthologs of Arabidopsis SUMO1 and SUMO2. Although other genes homologous to Arabidopsis SUMO genes exist in flowering plants, their relationship to the *Arabidopsis* genes has not been resolved clearly using orthology tools, and they were not included in Table 1.

## SUMO ligases

Ligases are proteins that increase the rate of SUMO conjugation to substrates and influence the substrate specificity of the SUMO conjugation system. This can occur if the ligase brings substrate and SCE into close proximity, by providing binding interfaces for both. Interestingly, as SCE itself often binds to substrates, this is not the only method of catalytic enhancement. It has also been suggested that certain SUMO ligases may only interact with SCE, and enhance SUMO transfer by imposing conformational constraints on SUMO-loaded SCE (Gareau & Lima, 2010). Likewise, domains specifying a particular subcellular localization, plus SCE interaction, can increase local SCE concentration to promote sumoylation of certain substrates. A known SCE interaction domain, present in most (but not all) identified SUMO ligases, is the MIZ-type zinc finger (zf-MIZ, also called SP-RING; Hochstrasser, 2001), a domain in which two zinc ions are coordinated via a set of conserved Cys and histidine (His) residues. zf-MIZ is part of all three SUMO ligase types known in Arabidopsis.

### SIZ1

The best characterized SUMO ligase of Arabidopsis is SIZ1, a nuclear protein encoded by At5g60410. It contains the SCE1 binding zf-MIZ RING finger, a plant homeodomain (PHD), which usually mediates chromatin association and seems to contribute to SCE1 binding, and a part similar to a DNA binding domain (SAP) (Garcia-Dominguez *et al.*, 2008; Suzuki *et al.*, 2009). Null mutants exhibit a severe and highly pleiotropic phenotype, with most changes related to stress responses (Miura *et al.*, 2007a; Lois, 2010; Miura & Hasegawa, 2010; B. S. Park *et al.*, 2011). AtSIZ1 facilitates the sumoylation of regulatory proteins, such as the Myc transcription factor ICE1, the putative histone demethylase FLD and the bZIP transcription factor ABI5 (Catala *et al.*, 2007; Miura *et al.*, 2007b, 2009; Jin *et al.*, 2008), and of nitrate reductase (B. S. Park *et al.*, 2011). All plants listed in Table 1 encode at least one predicted ortholog. Rice has two paralogs, which have been described recently (Park *et al.*, 2010; Thangasamy *et al.*, 2011; Wang *et al.*, 2011).

### HPY2/MMS21

Arabidopsis SUMO ligase At3g15150 was identified independently by two groups and designated as AtMMS21 and HPY2, respectively (Huang *et al.*, 2009; Ishida *et al.*, 2009). This ligase plays a role in DNA metabolism and meristem maintenance. Homologs to this protein also exist in fungi and animals. In most plants, HPY2 is apparently a single copy gene. In *Brachypodium*, two paralogs exist that are closely linked, suggesting that a species-specific tandem duplication resulted in the gene number increase. We therefore conclude that a single gene of this class belongs to the core set of plant SUMO ligases.

### PIAL1/2

Two additional proteins of Arabidopsis, At1g08910 and At5g41580, carry the zf-MIZ domain that is characteristic of many SUMO ligases (Novatchkova *et al.*, 2004). The encoded proteins, PIAS like 1 and 2 (PIAL1/2), respectively, show *in vitro* SUMO ligase activity (K. Tomanov & A. Bachmair, unpublished). Expression data (University of Toronto Arabidopsis eFP Browser <http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>) indicate that PIAL1 is stress inducible. The PIAL1/2 class of SUMO ligases has identifiable homologs in the analyzed plants. PIAL1/2 class members show a high level of sequence conservation in the

amino-terminal region that decreases in the second half of the proteins, after the zf-MIZ domain (for details, see Fig. S7). Most plants listed have one homolog, which is more similar to PIAL2 than to PIAL1.

## SUMO proteases

SUMO proteases have a dual function. They provide free SUMO by hydrolyzing peptide linkages in primary translation products of SUMO genes, which encode carboxyl-terminal extensions linked to the SUMO sequence (see Fig. 1). In addition to precursor cleavage, these proteases function as isopeptidases to release and recycle SUMO from protein conjugates (Colby *et al.*, 2006; Mukhopadhyay & Dasso, 2007). There is good evidence that SUMO proteases contribute to the regulation of flowering time, plant–pathogen interactions and adaptation to abiotic stress factors (Murtas *et al.*, 2003; Xu *et al.*, 2007; Conti *et al.*, 2008; Kim *et al.*, 2008). The specificity of SUMO proteases in animals and fungi derives, to a large extent, from their subcellular localization, and mutant enzymes with improper localization do not fully complement null phenotypes (Panse *et al.*, 2003; Mukhopadhyay & Dasso, 2007).

An exhaustive listing of plant SUMO-specific proteases is particularly challenging as explained in the following. The known SUMO proteases of *Arabidopsis* belong to the C48 clade of Cys proteases (Mukhopadhyay & Dasso, 2007; van der Hoorn, 2008). However, proteases specific for the modifier ubiquitin belong to five different structural groups, including metalloproteases (for reviews, see Routenberg Love *et al.*, 2007; Komander *et al.*, 2009), and one SUMO-specific protease of baker's yeast, Wss1, is a metalloprotease (Mullen *et al.*, 2010). Plants encode proteins with homology to Wss1 (e.g. *Arabidopsis* gene At1g55915 is a potential candidate). Likewise, *Arabidopsis* genes AtULP2a–h have been annotated as potential SUMO proteases (Kurepa *et al.*, 2003), but experimental proof has not yet been published. This suggests that the experimentally verified SUMO-specific proteases of *Arabidopsis* do not represent the complete set. We nonetheless concentrated on the C48 clade of proteases, and used Phytozome 8 gene family 318727390 as the basis for analysis. Figure 2(a) shows a phylogenetic tree of *Arabidopsis* entries (generated using MrBayes 3.2.1). It contains, in addition to branches with known SUMO proteases, two new branches that represent additional candidate genes. In a second step, we searched our set of plant genomes for presumed orthologs of the *Arabidopsis* genes, using OMA, OrthoMCL, Inparanoid and RoundUp. Figure 2(b) shows a graphic representation of gene relationships using CLANS (Frickey & Lupas, 2004). Table 1 lists all entries of Fig. 2. Footnote 10 of Table 1 lists genes that did not fall into one of the four defined classes. Figure S9 shows a protease sequence alignment. Taken together, the searched plant species encode potential orthologs of the known *Arabidopsis* SUMO proteases, but additional proteases with SUMO specificity are likely to exist in plants.

### Group A, potential SUMO protease At3g48480

This group of genes awaits experimental proof regarding its activity as SUMO-specific proteases. Its relationship to the other groups, however, is highly suggestive of the proposed activity. Most plants of Table 1 encode a single gene of this class.

### Group B1, OTS1 and OTS2

The two genes OTS1, At1g60220, and OTS2, At1g10570, have overlapping functions and have been implicated in salt stress resistance (Conti *et al.*, 2008). The copy number of this group varies from one in grapevine and poplar to six in *Sorghum* (Table 1).



### Group B2, potential SUMO proteases At1g09730 and At4g33620

These genes (called ULP2like2 and 1, respectively, in Novatchkova *et al.*, 2004) have not yet been functionally characterized. The At1g09730 gene is large and, consistent with a high expression level, cDNAs have been isolated. The second *Arabidopsis* gene of this class, At4g33620, has a significantly lower expression level than At1g09730. Mutation of At1g09730, but not of At4g33620, results in reduced growth (H-P. Stuibler & A. Bachmair, unpublished). Predicted orthologs exist in all plants of Table 1.

### Group C, ESD4 and related

ESD4 of *Arabidopsis* (At4g15880) locates to the nuclear periphery and contributes to flowering time regulation (Murtas *et al.*, 2003; Xu *et al.*, 2007). By contrast, its relative ELS1 (At3g06910) is extranuclear (Hermkes *et al.*, 2011). Both genes are functionally distinct, as is evident from the different mutant phenotypes. Because *esd4* mutants have a more severe phenotype than *els1* mutants, we hypothesize that ESD4 is the central gene of this group. A third candidate gene, At4g00690, has not been functionally characterized and may be a pseudogene. Each plant of Table 1 has at least one representative of this group.

### Proteins with a SUMO-like domain

A previously identified protein with a SUMO-like domain, At1g68185 (Novatchkova *et al.*, 2005), is conserved in fungi, animals and plants. Its fission yeast homolog, RAD60, binds to SCE (Prudden *et al.*, 2011) and is important for genome stability (Heideker *et al.*, 2011). All plants listed have one identifiable representative (Table 1, Fig. S10).

### SUMO chain binding proteins

Sumoylation functions mainly by promoting the formation of new intra- and intermolecular protein contacts (Kerscher, 2007). These interactions allow the establishment of functional networks between sumoylated proteins and their noncovalent interactors (Hecker *et al.*, 2006; Kerscher, 2007). So-called SUMO-interacting motifs (SIMs) are the mediators of noncovalent interactions between SUMO and SUMO binding proteins. SIMs are characterized by a loose consensus sequence,  $\Psi\Psi x\Psi D/S/E$  or  $D/S/E\Psi x\Psi\Psi$ , where  $\Psi$  symbolizes the hydrophobic amino acids I, L, V, M or F, x can be any amino acid, and D, S and E represent single-letter amino acid abbreviations (Miteva *et al.*, 2010). The short length and variability of these sequences result in poor conservation (they may disappear and reappear in a different part of a protein, or in a different subunit of a protein complex). However, one class of animal and fungal proteins is characterized by a tandem arrangement of four SIMs, which therefore have specific affinity for binding to SUMO chains. In addition, these proteins have a RING domain and function as ubiquitin ligases, channeling proteins with a SUMO chain into the ubiquitin-proteasome dependent degradation pathway (Plechanovova *et al.*, 2011; Praefcke *et al.*, 2012). It has been shown previously that *Arabidopsis* SUMO1/2 can form chains (Colby *et al.*, 2006). We therefore gave consideration to candidate loci with the above structural hallmarks. *Arabidopsis* has two proteins with four or five SIMs and one RING domain, At3g07200 and At5g48655. The performed orthology predictions suggest that they are each other's paralogs and that potential orthologs exist in other plants. Some monocot representatives have two characteristic sequence insertions compared with the other family members, and may therefore form a monocot-specific subclass (Table 1, Fig. S11). We thus hypothesize that plants use SUMO chains as degradation signals, channeling chain-modified substrate proteins into degradation by SUMO chain binding ubiquitin ligases.

In summary, we list the predicted orthologs of known and currently unappreciated components of the SUMO conjugation system and identify candidates for monocot-specific

subgroups of enzymes. The overall structural conservation of the SUMO conjugation system in flowering plants underpins the value of Arabidopsis as a model in SUMO research, and monocot-specific features promise interesting results from experimental approaches in monocots.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Work in A.B.'s laboratory is supported by the German Research Foundation DFG (grant 1158/5-1, SPP1365) and by the Austrian Science Foundation FWF (grant P21215-B12). K.H.'s work is supported by DFG (SPP1365).

## References

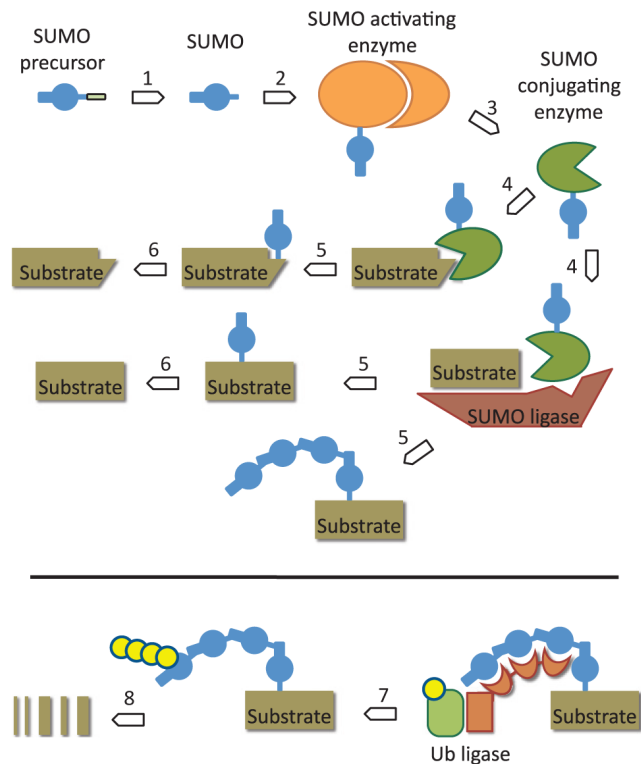
- Altenhoff AM, Schneider A, Gonnet GH, Dessimoz C. OMA2011: orthology inference among 1000 complete genomes. *Nucleic Acids Research*. 2011; 39:D289–D294. [PubMed: 21113020]
- Budhiraja R, Hermkes R, Muller S, Schmidt J, Colby T, Panigrahi K, Coupland G, Bachmair A. Substrates related to chromatin and to RNA-dependent processes are modified by Arabidopsis SUMO isoforms that differ in a conserved residue with influence on desumoylation. *Plant Physiology*. 2009; 149:1529–1540. [PubMed: 19151129]
- van den Burg HA, Kini RK, Schuurink RC, Takken FL. Arabidopsis small ubiquitin-like modifier paralogs have distinct functions in development and defense. *Plant Cell*. 2010; 22:1998–2016. [PubMed: 20525853]
- Catala R, Ouyang J, Abreu IA, Hu Y, Seo H, Zhang X, Chua NH. The Arabidopsis E3 SUMO ligase SIZ1 regulates plant growth and drought responses. *Plant Cell*. 2007; 19:2952–2966. [PubMed: 17905899]
- Chen CC, Chen YY, Tang IC, Liang HM, Lai CC, Chiou JM, Yeh KC. Arabidopsis SUMO E3 ligase SIZ1 is involved in excess copper tolerance. *Plant Physiology*. 2011; 156:2225–2234. [PubMed: 21632972]
- Chosed R, Mukherjee S, Lois LM, Orth K. Evolution of a signalling system that incorporates both redundancy and diversity: Arabidopsis SUMOylation. *Biochemical Journal*. 2006; 398:521–529. [PubMed: 16740136]
- Colby T, Matthäi A, Boeckelmann A, Stuible HP. SUMO-conjugating and SUMO-deconjugating enzymes from Arabidopsis. *Plant Physiology*. 2006; 142:318–332. [PubMed: 16920872]
- Conti L, Price G, O'Donnell E, Schwessinger B, Dominy P, Sadanandom A. Small ubiquitin-like modifier proteases OVERLY TOLERANT TO SALT1 and -2 regulate salt stress responses in Arabidopsis. *Plant Cell*. 2008; 20:2894–2908. [PubMed: 18849491]
- Deluca TF, Wu IH, Pu J, Monaghan T, Peshkin L, Singh S, Wall DP. Roundup: a multi-genome repository of orthologs and evolutionary distances. *Bioinformatics*. 2006; 22:2044–2046. [PubMed: 16777906]
- Elrouby N, Coupland G. Proteome-wide screens for small ubiquitin-like modifier (SUMO) substrates identify Arabidopsis proteins implicated in diverse biological processes. *Proceedings of the National Academy of Sciences, USA*. 2010; 107:17415–17420.
- Feuillet C, Leach JE, Rogers J, Schnable PS, Eversole K. Crop genome sequencing: lessons and rationales. *Trends Plant Science*. 2011; 16:77–88.
- Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*. 2004; 20:3702–3704. [PubMed: 15284097]
- Garcia-Dominguez M, March-Diaz R, Reyes JC. The PHD domain of plant PIAS proteins mediates sumoylation of bromodomain GTE proteins. *Journal of Biological Chemistry*. 2008; 283:21469–21477. [PubMed: 18502747]
- Gareau JR, Lima CD. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nature Reviews in Molecular Cell Biology*. 2010; 11:861–871.

- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*. 2012; 40:D1178–D1186. [PubMed: 22110026]
- Hecker CM, Rabiller M, Haglund K, Bayer P, Dikic I. Specification of SUMO1- and SUMO2-interacting motifs. *Journal of Biological Chemistry*. 2006; 281:16117–16127. [PubMed: 16524884]
- Heideker J, Prudden J, Perry JJP, Tainer JA, Boddy MN. SUMO-targeted ubiquitin ligase, Rad60, and Nse2 SUMO ligase suppress spontaneous Top1-mediated DNA damage and genome instability. *PLoS Genetics*. 2011; 7:e1001320. [PubMed: 21408210]
- Hermkes R, Fu Y-F, Nurrenberg K, Budhiraja R, Schmelzer E, Elrouby N, Dohmen RJ, Bachmair A, Coupland G. Distinct roles for Arabidopsis SUMO protease ESD4 and its closest homolog ELS1. *Planta*. 2011; 233:63–73. [PubMed: 20922545]
- Hochstrasser M. SP-RING for SUMO: new functions boom for a ubiquitin-like protein. *Cell*. 2001; 107:5–8. [PubMed: 11595179]
- Hochstrasser M. Origin and function of ubiquitin-like proteins. *Nature*. 2009; 458:422–429. [PubMed: 19325621]
- van der Hoorn RAL. Plant proteases: from phenotypes to molecular mechanisms. *Annual Review of Plant Biology*. 2008; 59:191–223.
- Huang L, Yang S, Zhang S, Liu M, Lai J, Qi Y, Shi S, Wang J, Wang Y, Xie Q, et al. The Arabidopsis SUMO E3 ligase AtMMS21, a homologue of NSE2/MMS21, regulates cell proliferation in the root. *Plant Journal*. 2009; 60:666–678. [PubMed: 19682286]
- Ishida T, Fujiwara S, Miura K, Stacey N, Yoshimura M, Schneider K, Adachi S, Minamisawa K, Umeda M, Sugimoto K. SUMO E3 ligase HIGH PLOIDY2 regulates endocycle onset and meristem maintenance in Arabidopsis. *Plant Cell*. 2009; 21:2284–2297. [PubMed: 19666737]
- Jin JB, Jin YH, Lee J, Miura K, Yoo CY, Kim WY, Van Oosten M, Hyun Y, Somers DE, Lee I, et al. The SUMO E3 ligase, AtSIZ1, regulates flowering by controlling a salicylic acid-mediated floral promotion pathway and through effects on FLC chromatin structure. *Plant Journal*. 2008; 53:530–540. [PubMed: 18069938]
- Kerscher O. SUMO junction – what’s your function? New insights through SUMO-interacting motifs. *EMBO Reports*. 2007; 8:550–555. [PubMed: 17545995]
- Kim J-G, Taylor KW, Hotson A, Keegan M, Schmelz EA, Mudgett MB. XopD SUMO protease affects host transcription, promotes pathogen growth, and delays symptom development in *Xanthomonas*-infected tomato leaves. *Plant Cell*. 2008; 20:1915–1929. [PubMed: 18664616]
- Komander D, Clague MJ, Urbé S. Breaking the chains: structure and function of the deubiquitinases. *Nature Reviews Molecular Cell Biology*. 2009; 10:550–563.
- Kurepa J, Walker JM, Smalle J, Gosink MM, Davis SJ, Durham TL, Sung D-Y, Vierstra RD. The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. *Journal of Biological Chemistry*. 2003; 278:6862–6872. [PubMed: 12482876]
- Li L, Stoekert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*. 2003; 13:2178–2189. [PubMed: 12952885]
- Lois LM. Diversity of the SUMOylation machinery in plants. *Biochemical Society Transactions*. 2010; 38:60–64. [PubMed: 20074036]
- Lois LM, Lima CD, Chua NH. Small ubiquitin-like modifier modulates abscisic acid signaling in Arabidopsis. *Plant Cell*. 2003; 15:1347–1359. [PubMed: 12782728]
- Martinez M. Plant protein-coding gene families: emerging bioinformatics approaches. *Trends in Plant Science*. 2011; 16:558–567. [PubMed: 21757395]
- Miller MJ, Barrett-Wilt GA, Hua Z, Vierstra RD. Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in Arabidopsis. *Proceedings of the National Academy of Sciences, USA*. 2010; 107:16512–16517.
- Miteva, M.; Keusekotten, K.; Hofmann, K.; Praefcke, GJK.; Dohmen, RJ. Sumoylation as a signal for polyubiquitylation and proteasomal degradation. In: Goettrup, M., editor. *Conjugation and deconjugation of ubiquitin family modifiers*. Landes Biosciences; Austin, TX, USA: 2010. p. 195-214.



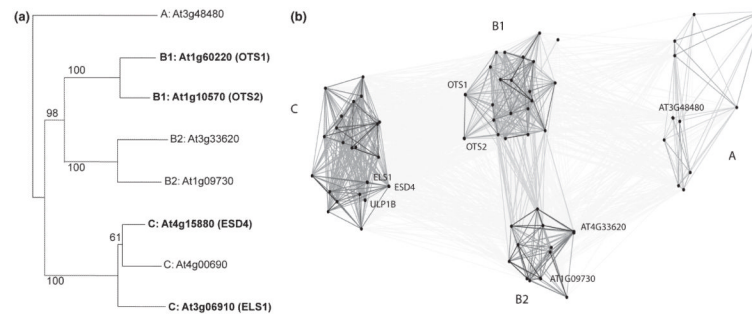
- Miura K, Hasegawa PM. Sumoylation and other ubiquitin-like post-translational modifications in plants. *Trends in Cell Biology*. 2010; 20:223–232. [PubMed: 20189809]
- Miura K, Jin JB, Hasegawa PM. Sumoylation, a post-translational regulatory process in plants. *Current Opinion in Plant Biology*. 2007a; 10:495–502. [PubMed: 17720613]
- Miura K, Jin JB, Lee J, Yoo CY, Stirm V, Miura T, Ashworth EN, Bressan RA, Jun D-J, Hasegawa PM. SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1A expression and freezing tolerance in Arabidopsis. *Plant Cell*. 2007b; 19:1403–1414. [PubMed: 17416732]
- Miura K, Lee J, Gong Q, Ma S, Jin JB, Yoo CY, Miura T, Sato A, Bohnert HJ, Hasegawa PM. SIZ1 regulation of phosphate starvation-induced root architecture remodeling involves the control of auxin accumulation. *Plant Physiology*. 2011; 155:1000–1012. [PubMed: 21156857]
- Miura K, Lee J, Jin JB, Yoo CY, Miura T, Hasegawa PM. Sumoylation of ABI5 by the Arabidopsis SUMO E3 ligase SIZ1 negatively regulates abscisic acid signaling. *Proceedings of the National Academy of Sciences, USA*. 2009; 106:5418–5423.
- Mukhopadhyay D, Dasso M. Modification in reverse: the SUMO proteases. *Trends in Biochemical Sciences*. 2007; 32:286–295. [PubMed: 17499995]
- Mullen JR, Chen CF, Brill SJ. Wss1 is a SUMO-dependent isopeptidase that interacts genetically with the Slx5-Slx8 SUMO-targeted ubiquitin ligase. *Molecular and Cellular Biology*. 2010; 30:3737–3748. [PubMed: 20516210]
- Murtas G, Reeves PH, Fu YF, Bancroft I, Dean C, Coupland G. A nuclear protease required for flowering-time regulation in Arabidopsis reduces the abundance of SMALL UBIQUITIN-RELATED MODIFIER conjugates. *Plant Cell*. 2003; 15:2308–2319. [PubMed: 14507998]
- Nigam N, Singh A, Sahi C, Chandramouli A, Grover A. SUMO-conjugating enzyme (Sce) and FK506-binding protein (FKBP) encoding rice (*Oryza sativa* L.) genes: genome-wide analysis, expression studies and evidence for their involvement in abiotic stress response. *Molecular Genetics and Genomics*. 2008; 279:371–383. [PubMed: 18219493]
- Novatchkova M, Bachmair A, Eisenhaber B, Eisenhaber F. Proteins with two SUMO-like domains in chromatin-associated complexes: the RENi (Rad60-Esc2-NIP45) family. *BMC Bioinformatics*. 2005; 6:22. [PubMed: 15698469]
- Novatchkova M, Budhiraja R, Coupland G, Eisenhaber F, Bachmair A. SUMO conjugation in plants. *Planta*. 2004; 220:1–8. [PubMed: 15449058]
- Olsen SK, Capili AD, Lu X, Tan DS, Lima CD. Active site remodelling accompanies thioester bond formation in the SUMO E1. *Nature*. 2010; 463:906–912. [PubMed: 20164921]
- Ostlund G, Schmitt T, Forslund K, Köstler T, Messina DN, Roopra S, Frings O, Sonnhammer EL. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Research*. 2010; 38:D196–D203. [PubMed: 19892828]
- Panse VG, Kuster B, Gerstberger T, Hurt E. Unconventional tethering of Ulp1 to the transport channel of the nuclear pore complex by karyopherins. *Nature Cell Biology*. 2003; 5:21–27.
- Park BS, Song JT, Seo HS. Arabidopsis nitrate reductase activity is stimulated by the E3 SUMO ligase AtSIZ1. *Nature Communications*. 2011; 2:400.
- Park HC, Kim H, Koo SC, Park HJ, Cheong MS, Hong H, Baek D, Chung WS, Kim DH, Bressan RA, et al. Functional characterization of the SIZ/PIAS-type SUMO E3 ligases, OsSIZ1 and OsSIZ2 in rice. *Plant, Cell & Environment*. 2010; 33:1923–1934.
- Park HJ, Kim WY, Park HC, Lee SY, Bohnert HJ, Yun DJ. SUMO and SUMOylation in plants. *Molecules and Cells*. 2011; 32:305–316. [PubMed: 21912873]
- Plechanovova A, Jaffray EG, McMahon SA, Johnson KA, Navratilova I, Naismith JH, Hay RT. Mechanism of ubiquitylation by dimeric RING ligase RNF4. *Nature Structural and Molecular Biology*. 2011; 18:1052–1059.
- Praefcke GJ, Hofmann K, Dohmen RJ. SUMO playing tag with ubiquitin. *Trends in Biochemical Sciences*. 2012; 37:23–31. [PubMed: 22018829]
- Prudden J, Perry JJP, Nie M, Vashisht AA, Arvai AS, Hitomi C, Guenther G, Wohlschlegel JA, Tainer JA, Boddy NM. DNA repair and global sumoylation are regulated by distinct Ubc9 noncovalent complexes. *Molecular Cellular Biology*. 2011; 31:2299–2310.
- Routenberg Love K, Catic A, Schlieker C, Ploegh HL. Mechanisms, biology and inhibitors of deubiquitinating enzymes. *Nature Chemical Biology*. 2007; 3:697–705.

- Saracco SA, Miller MJ, Kurepa J, Vierstra RD. Genetic analysis of SUMOylation in Arabidopsis: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. *Plant Physiology*. 2007; 145:119–134. [PubMed: 17644626]
- Suzuki R, Shindo H, Tase A, Kikuchi Y, Shimizu M, Yamazaki T. Solution structures and DNA binding properties of the N-terminal SAP domains of SUMO E3 ligases from *Saccharomyces cerevisiae* and *Oryza sativa*. *Proteins*. 2009; 75:336–347. [PubMed: 18831036]
- Thangasamy S, Guo CL, Chuang MH, Lai MH, Chen J, Jauh GY. Rice SIZ1, a SUMO E3 ligase, controls spikelet fertility through regulation of anther dehiscence. *New Phytologist*. 2011; 189:869–882. [PubMed: 21083564]
- Vosper JMD, McDowell GS, Hindley CJ, Fiore-Herich CS, Kucerova R, Horan I, Philpott A. Ubiquitylation on canonical and non-canonical sites targets the transcription factor neurogenin for ubiquitin-mediated proteolysis. *Journal of Biological Chemistry*. 2009; 284:15458–15468. [PubMed: 19336407]
- Wang H, Makeen K, Yan Y, Cao Y, Sun S, Xu G. OsSIZ1 regulates the vegetative growth and reproductive development in rice. *Plant Molecular Biology Reporter*. 2011; 29:411–417.
- Xu XM, Rose A, Muthuswamy S, Jeong SY, Venkatakrishnan S, Zhao Q, Meier I. NUCLEAR PORE ANCHOR, the Arabidopsis homolog of Tpr/Mlp1/Mlp2/megator, is involved in mRNA export and SUMO homeostasis and affects diverse aspects of plant development. *Plant Cell*. 2007; 19:1537–1548. [PubMed: 17513499]
- Yoo CY, Miura K, Jin JB, Lee J, Park HC, Salt DE, Yun D-J, Bressan RA, Hasegawa PM. SIZ1 small ubiquitin-like modifier E3 ligase facilitates basal thermotolerance in Arabidopsis independent of salicylic acid. *Plant Physiology*. 2006; 142:1548–1558. [PubMed: 17041025]



**Fig. 1.**

The small ubiquitin-related modifier (SUMO) conjugation cycle. Each plant has several SUMO isoforms, all of which are translated as precursor proteins. SUMO-specific proteases cleave off a carboxyl-terminal peptide to expose a glycine (Gly) residue at the carboxyl terminus, which extends from the globular body of SUMO (step 1). SUMO's terminal Gly is activated by forming a thioester with the active site cysteine (Cys) residue of SUMO activating enzyme SAE (step 2). From there, SUMO is transferred to a Cys residue of SUMO conjugating enzyme SCE (step 3). SCE can transfer SUMO directly to substrates (step 4), provided that these contain a binding motif, usually the 'sumoylation consensus sequence' (see text). Proteins devoid of an SCE interaction motif require SUMO ligase assistance for modification. Substrate release (step 5) results usually in a monosumoylated protein. In an unspecified number of cases, however, a SUMO chain is attached to the substrate. The reversibility of SUMO conjugation results from the hydrolysis of the isopeptide bond by SUMO-specific proteases to release SUMO for further conjugation cycles (step 6). In a separate modification cascade, SUMO chains serve as the signal for the attachment of a ubiquitin chain (yellow dots) by a dedicated ubiquitin ligase complex (step 7), which results in proteasomal degradation of the substrate (steps summarized as arrow number 8).



**Fig. 2.** Sequence similarity of known and suggested small ubiquitin-related modifier (SUMO) proteases. (a) Phylogenetic tree of experimentally confirmed Arabidopsis SUMO proteases (bold print) and of closely related SUMO protease candidates, defining four subgroups (A, B1, B2, C). Numbers at the branches indicate per cent Bayesian posterior probabilities. (b) BLAST similarity network of SUMO proteases listed in Table 1. Each dot represents one entry, dark connecting lines indicate high similarity, decreasing darkness of grey lines symbolizes decreasing similarity. The four clusters correspond to subgroups A, B1, B2 and C.

Table 1

Enzymes from the sumoylation pathway in *Arabidopsis thaliana* (Arabidopsis), *Solanum lycopersicum* (tomato), *Vitis vinifera* (grapevine), *Populus trichocarpa* (poplar), *Oryza sativa* (rice), *Brachypodium distachyon* (Brachypodium), *Sorghum bicolor* (Sorghum) and *Zea mays* (maize)

Enzyme category				
SUMO activating enzyme subunit 1 (SAE1)	Arabidopsis	Tomato	Grapevine	Poplar
	At4g24940 (SAE1a)	Solyc03g019730	GSVIVT01008554001	POPTR_0015s11110
	At5g50580 (SAE1b) <sup>1</sup>	Solyc06g072080		POPTR_0012s10270
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os11g30410	Bradi4g18870	Sb05g017710	GRMZM2G149108
SUMO activating enzyme subunit 2 (SAE2)	Arabidopsis	Tomato	Grapevine	Poplar
	At2g21470 (SAE2)	Solyc01g109960	GSVIVT01023831001	POPTR_0004s16630 POPTR_0009s12300
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os07g39780	Bradi1g23100	Sb02g037850	GRMZM2G129575 <sup>2</sup>
	SUMO conjugating enzyme (SCE)	Arabidopsis	Tomato	Grapevine
At3g57870 (SCE1) <sup>3</sup>		Solyc03g044260	GSVIVT01009448001	POPTR_0005s18460
		Solyc02g093110	GSVIVT01022074001	POPTR_0002s12510
		Solyc12g088680		POPTR_0014s02500
		Solyc04g078620		POPTR_0014s02480
Rice		<i>Brachypodium</i>	<i>Sorghum</i>	Maize
Os10g39120 (OsSCE1) <sup>4</sup>		Bradi1g77010	Sb01g030580	GRMZM2G063931
Os03g03130 (OsSCE2) <sup>4</sup>		Bradi3g32080	Sb01g049010	GRMZM2G070047
Os04g49130 <sup>5</sup>		Bradi5g19200 <sup>5</sup>	Sb06g026280 <sup>5</sup>	GRMZM2G312693
			Sb06g026270 <sup>5</sup>	GRMZM2G163398
			Sb06g026250 <sup>5</sup>	GRMZM2G433968 <sup>5</sup>
				GRMZM2G038851 <sup>5</sup>
			GRMZM2G341089 <sup>5</sup>	
			GRMZM2G146142 <sup>5</sup>	
SUMO (SUM1 homologs) <sup>6</sup>	Arabidopsis	Tomato	Grapevine	Poplar
	At4g26840 (SUM1)	Solyc07g064880	GSVIVT01003301001	POPTR_0014s18990
	At5g55160 (SUM2)	Solyc12g006010	GSVIVT01003307001	POPTR_0002s21680



Enzyme category				
		Solyc09g059970	GSVIVT01030502001	POPTR_0002s21690
		Solyc07g049360		POPTR_0014s15650
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os01g68940	Bradi2g58830	Sb03g043870	GRMZM2G082390
	Os01g68950			GRMZM2G053898 <sup>7</sup>
SUMO ligase SIZ1 type	Arabidopsis	Tomato	Grapevine	Poplar
	At5g60410 (SIZ1)	Solyc11g069160	GSVIVT01025151001	POPTR_0009s02040
		Solyc06g010000		POPTR_0004s21990
				POPTR_0010s22630
				POPTR_0008s04190
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os05g03430 (OsSIZ1)	Bradi2g38030	Sb09g002225	GRMZM2G155123
	Os03g50980 (OsSIZ2)	Bradi2g62697	Sb05g000360	GRMZM2G455664
		Bradi4g26720	Sb08g000380	GRMZM2G337659
		Bradi4g45080		GRMZM2G002999 <sup>8</sup>
SUMO ligase HPY2/MMS21 type	Arabidopsis	Tomato	Grapevine	Poplar
	At3g15150 (HPY2/MMS21)	Solyc07g062780	GSVIVT01014276001	POPTR_0011s14450
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os05g48880 <sup>9</sup>	Bradi2g16600	Sb09g028460	GRMZM2G022065
		Bradi2g16577		
SUMO ligase PIAS-like	Arabidopsis	Tomato	Grapevine	Poplar
	At1g08910 (PIAL1)	Solyc08g008130	GSVIVT01026971001	POPTR_0003s13280
	At5g41580 (PIAL2)		GSVIVT01026973001	
			GSVIVT01019423001	
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os06g06870	Bradi1g48426	Sb10g004470	GRMZM2G075582
SUMO protease class A <sup>10</sup>	Arabidopsis	Tomato	Grapevine	Poplar
	At3g48480	Solyc00g007260	GSVIVT01007539001	POPTR_0015s10250
		Solyc11g072220		POPTR_0010s18760
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os11g10780	Bradi4g22405	Sb05g009053	GRMZM2G545326
		Bradi4g22300		
SUMO protease class B1 OTS type	Arabidopsis	Tomato	Grapevine	Poplar

Enzyme category				
	At1g60220 (OTS1/ULP1d)	Solyc04g026200	GSVIVT01020235001	POPTR_0010s04980
	At1g10570 (OTS2/ULP1c)	Solyc05g005630		
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os01g53630	Bradi1g42280	Sb03g034030	GRMZM2G432931
	Os06g29310	Bradi4g02360	Sb08g020827	GRMZM2G174667
	Os12g41380		Sb08g020823	GRMZM2G351786
			Sb08g020821	GRMZM2G072939
			Sb04g001916	GRMZM2G142553
			Sb10g016910	
SUMO protease class B2	Arabidopsis	Tomato	Grapevine	Poplar
	At1g09730	Solyc01g105830	GSVIVT01009173001	POPTR_0005s17970
	At4g33620	Solyc11g017040		POPTR_0002s10590
				POPTR_0007s03240
				POPTR_0017s07140
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os05g11770	Bradi3g60530	Sb03g040230 <sup>11</sup>	GRMZM2G177324 <sup>11</sup>
SUMO protease class C ESD4 type	Arabidopsis	Tomato	Grapevine	Poplar
	At4g15880 (ESD4)	Solyc01g066830	GSVIVT01017729001	POPTR_0010s01730
	At3g06910 (ELS1/ULP1a)	Solyc12g099530		POPTR_0008s22250
	At4g00690 (ULP1b)			POPTR_0082s00230
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os03g29630	Bradi1g18100	Sb01g035640	GRMZM5G849959
	Os03g22400	Bradi1g62760	Sb02g042620	GRMZM2G010505
	Os01g25370	Bradi2g12720	Sb03g013930	GRMZM2G012601
				GRMZM2G088653
SUMO domain containing protein	Arabidopsis	Tomato	Grapevine	Poplar
	At1g68185	Solyc03g013030	GSVIVT01011822001	POPTR_0010s12880
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	Maize
	Os01g42250	Bradi2g43070	Sb03g027375	GRMZM2G073404
SUMO chain binding protein (ubiquitin ligase)	Arabidopsis	Tomato	Grapevine	Poplar
	At3g07200	Solyc06g007870	GSVIVT01007471001	POPTR_0002s24690
	At5g48655	Solyc09g059450	GSVIVT01020997001	POPTR_0014s18890
				POPTR_0003s14810

Enzyme category				
	Rice	<i>Brachypodium</i>	<i>Sorghum</i>	POPTR_0001s11540
	Os09g32690	Bradi4g34450	Sb02g029150	Maize GRMZM2G013892
	Os01g69040 <sup>12</sup>	Bradi3g40940	Sb03g043910 <sup>12</sup>	GRMZM2G039968
		Bradi2g58870 <sup>12</sup>		GRMZM2G359505 <sup>12</sup>

<sup>1</sup> Previous annotations of the *Arabidopsis* Col-0 genome had, in addition, Gene At5g50680 listed with identical sequence to At5g50580. This entry was removed from the most recent update.

<sup>2</sup> Hypothetical open reading frame GRMZM2G129575 is significantly shorter than other SAE2 reading frames. However, *de novo* prediction allows this open reading frame to be extended, generating a gene that is as long as and highly similar to *Sorghum* SAE2 (see Supporting Information Fig. S2 on alignment).

<sup>3</sup> A gene previously annotated as a potential SCE1 pseudogene, At5g02240 (SCE1b), encompasses a conserved gene currently annotated as a steroid dehydrogenase. It was therefore excluded from the table. See text for further details.

<sup>4</sup> Gene designation from Nigam *et al.* (2008).

<sup>5</sup> Genes Os04g49130, Bradi5g19200, Sb06g026280, Sb06g026270, Sb06g026250, GRMZM2G433968, GRMZM2G038851, GRMZM2G341089 and GRMZM2G146142 apparently form a monocot-specific subgroup (see Fig. S3 on alignment).

<sup>6</sup> For the complete set of *Arabidopsis* SUMO genes, see Novatchkova *et al.* (2004). Orthologs to the additional *Arabidopsis* SUMO genes are difficult to find in other plants, although all plants encode 'noncanonical' SUMO genes, with unknown function.

<sup>7</sup> GRMZM2G053898 does not contain the conserved carboxyl-terminal residues of SUMO. It may therefore be nonfunctional, or the gene may be incompletely annotated.

<sup>8</sup> According to current annotation, GRMZM2G002999 is shorter, aligning only to the carboxyl-terminal part of other genes listed in this group. This could be a result of incompleteness of the annotation, or to gene truncation.

<sup>9</sup> The gene model currently representing Os05g48880 in data bases is shorter than other HPY2 genes, lacking part of the zf-MIZ domain. However, *de novo* prediction for this gene allows the construction of an extended reading frame that contains all parts expected for an HPY2 ortholog (see Fig. S6 on alignment).

<sup>10</sup> In addition to the genes listed, which fall into classes A, B1, B2 and C, we identified SUMO protease candidates that do not fit into any of these classes: rice genes Os01g33530; Os03g42960; Os04g25110; Os09g08450; Os09g11860; Os09g23240; Os10g33450; Os11g12500; *Brachypodium* genes Bradi2g26350; Bradi2g33410; Bradi3g28630; Bradi5g15320; maize genes GRMZM2G312375; GRMZM2G321795; GRMZM2G332829; *Sorghum* gene Sb03g029665; grapevine gene GSVIVT01007609001; and poplar gene POPTR\_0004s06880.

<sup>11</sup> Current annotations suggest that *Sorghum* gene Sb03g040230 and maize GRMZM2G177324 are significantly shorter than other homologs, aligning to only a portion of the reading frame of *Arabidopsis* protease At1g09730. However, these hypothetical open reading frames could be part of longer genes with more extended homology.

<sup>12</sup> Genes Os01g69040, Bradi2g58870, Sb03g043910 and GRMZM2G359505 have two short amino acid insertions in common, and may therefore form a monocot-specific subgroup (see Fig. S11 on alignment).