Molecular Plant • Volume 6 • Number 5 • Pages 1646–1660 • September 2013 RESEARCH ARTICLE

Diversification of SUMO-Activating Enzyme in *Arabidopsis*: Implications in SUMO Conjugation

Laura Castaño-Miquel[a,2], Josep Seguí[a,2], Silvia Manrique[a,2], Inês Teixeira[a], Lorenzo Carretero-Paulet[b], Félix Atencio[a] and **L. Maria Lois[a,1]**

a Center for Research in Agricultural Genomics CRAG (CSIC-IRTA-UAB-UB), Edifici CRAG-Campus UAB, Bellaterra (Cerdanyola del Vallés), 08193 Barcelona, Spain
b Department of Biological Sciences, SUNY-University at Buffalo, North Campus, 109 Cooke Hall, Buffalo, NY 14260, USA

ABSTRACT Sumoylation is an essential posttranslational modification that participates in many biological processes including stress responses. However, little is known about the mechanisms that control Small Ubiquitin-like MOdifier (SUMO) conjugation *in vivo*. We have evaluated the regulatory role of the heterodimeric E1 activating enzyme, which catalyzes the first step in SUMO conjugation. We have established that the E1 large SAE2 and small SAE1 subunits are encoded by one and three genes, respectively, in the *Arabidopsis* genome. The three paralogs genes *SAE1a*, *SAE1b1*, and *SAE1b2* are the result of two independent duplication events. Since *SAE1b1* and *SAE1b2* correspond to two identical copies, only two E1 small subunit isoforms are present *in vivo*: SAE1a and SAE1b. The E1 heterodimer nuclear localization is modulated by the C-terminal tail of the SAE2 subunit. *In vitro*, SUMO conjugation rate is dependent on the SAE1 isoform contained in the E1 holoenzyme and our results suggest that downstream steps to SUMO–E1 thioester bond formation are affected. *In vivo*, SAE1a isoform deletion in T-DNA insertion mutant plants conferred sumoylation defects upon abiotic stress, consistent with a sumoylation defective phenotype. Our results support previous data pointing to a regulatory role of the E1 activating enzyme during SUMO conjugation and provide a novel mechanism to control sumoylation *in vivo* by diversification of the E1 small subunit.

Key words: E1 activating enzyme; conjugation rate; subcellular localization; regulation; abiotic stress; gene duplication.

INTRODUCTION

In eukaryotic cells, posttranslational modifications by SUMO (Small Ubiquitin-like MOdifier) modulate protein activity through regulation of subcellular localization, protein activity and stability, and protein–protein interactions (Wilkinson and Henley, 2010). SUMO is synthesized as a precursor that is processed via specific proteases, ULP, releasing a SUMO mature form with a Gly–Gly motif at its C-terminus. SUMO is conjugated to protein targets through three sequential reactions. In a first step, SUMO is activated by the heterodimeric E1-activating enzyme in an ATP-dependent reaction. Activated SUMO is transferred to the E2-conjugating enzyme as a prior step to conjugation to the substrate, which can be achieved by direct transfer from the E2 or facilitated by E3 ligases. Sumoylation is a reversible modification and SUMO excision from the substrate is catalyzed by the same class of cysteine proteases involved in the maturation step (Gareau and Lima, 2010).

Many studies have addressed the biological function of SUMO in plants by using the dicot *Arabidopsis* and the monocot rice as models. In both models, SUMO has a role in response to abiotic stresses, such as heat, cold, salt, and ABA signaling, in addition to a wide array of developmental processes (Chaikam and Karlson, 2010; Lois, 2010; Miura and Hasegawa, 2010; Park et al., 2010; Thangasamy et al., 2011; Wang et al., 2011). In rice, sumoylation has also been involved in hybrid male sterility (Long et al., 2008). Moreover, defense responses (Kim et al., 2008), flowering (Murtas et al., 2003),

[1] To whom correspondence should be addressed. E-mail maria.lois@cragenomica.es, tel. +34 93 5636600 ext. 3215, fax +34 93 5636601.

2 These authors contributed equally to this work.

doi: 10.1093/mp/sst049, Advance Access publication 12 March 2013

Received 14 February 2013; accepted 3 March 2013

nitrogen metabolism (Park et al., 2011), phosphate starvation (Miura et al., 2005), drought tolerance (Catala et al., 2007), and sensitivity to copper (Chen et al., 2011) are processes affected by altered sumoylation in *Arabidopsis*.

Most of the mentioned studies used plants with mutations in E3 ligases and ULP proteases. Among them, the most studied mutants are the *siz1* E3 ligase and the *esd4* ULP protease. *siz1*-null mutant displays a reduction in endogenous SUMO conjugate accumulation while an overaccumulation of SUMO conjugates is found in *esd4* mutant. Even though they have opposite molecular effects, the physiological outcome of these mutations is very similar. An explanation to these observations is that the molecular mechanisms that mediate the severe growth defects found in those mutants might be different (Hermkes et al., 2011). More importantly, these results indicate that SUMO conjugation homeostasis is under a tight control and over- or under-accumulation of SUMO conjugates results in a misregulation of essential processes.

In the sumoylation cascade, the role of the E1 activating enzyme, which is the first control point in the conjugation pathway, is particularly interesting. The E1 is a heterodimeric enzyme consisting of a large subunit, SAE2, and a small subunit, SAE1. SAE2 is structured in four functional domains: adenylation, catalytic cysteine, ubiquitin-fold (UFD), and C-terminal domains (Lois and Lima, 2005). On the other hand, SAE1 contributes the essential Arg21 to the adenylation domain (Lee and Schindelin, 2008). The adenylation domain is responsible for SUMO recognition and SUMO C-terminus adenylation, as a prior step to thioester bond formation with the E1 catalytic cysteine. By using SUMO intermediate analogs, structural studies have established that thioester bond formation requires an active site remodeling that involves a 130 degree rotation of the catalytic cysteine domain (Olsen et al., 2010).

In contrast, much less is known about how E1 is regulated *in vivo*. In mammalian cells, SUMO E1 is localized to the nucleus by a nuclear localization signal located in the C-terminal tail of the large subunit SAE2 (Moutty et al., 2011). Also, low ROS levels result in SUMO conjugation inhibition by inducing the formation of a disulphide bridge between the catalytic Cys residues of the E1 activating and the E2 conjugating enzymes. This inhibition is reversible and could constitute an early step in ROS signaling (Bossis and Melchior, 2006).

In plants, the regulation of SUMO conjugation is far from being understood. The SUMO system is apparently more complex in terms of number and the biochemical properties of its components. In *Arabidopsis*, SUMO1 and 2 have diverged to the extent that double mutants *sumo1sumo2* cannot be complemented by SUMO3 and 5 (Saracco et al., 2007). This divergence is reflected by a preferential conjugation of SUMO1/2 versus SUMO3 and 5 through a mechanism involving SUMO isoform selection by the E1 (Castaño-Miquel et al., 2011). In addition to SUMO, other components of the SUMO conjugation system have also diverged into several isoforms, such as the E1 and E2 (Novatchkova et al., 2012). It has been proposed that *Arabidopsis* expresses two isoforms of the E1 activating enzyme that differ in the small subunit composition, SAE1a or SAE1b, whereas the large subunit, SAE2, is unique. The existence of a functional specialization of both isoforms has not been reported so far.

Here, we have analyzed the biological implications of E1 activating enzyme diversification. We showed that the E1 activating enzyme is encoded by one variant of the *SAE2* gene and three copies of the *SAE1* gene: *SAE1a*, *SAE1b1*, and *SAE1b2*. Since *SAE1b1* and *SAE1b2* are identical, only two isoforms of the E1 activating enzyme can be found: SAE2/SAE1a (E1a) and SAE2/SAE1b (E1b). The heterodimeric E1 subcellular localization is facilitated by the C-terminal tail of the SAE2 subunit. Even though most of the E1 functional domains are located in the SAE2 subunit, we found that SAE1a and SAE1b isoforms conferred distinct conjugation rates *in vitro*, suggesting a regulatory role of the SAE1 small subunit during SUMO conjugation. *sae1a* mutant plants, which do not express SAE1a isoform, displayed defects in SUMO conjugation upon heat and drought responses, which is characteristic of SUMOylation-deficient plants (Catala et al., 2007; Saracco et al., 2007). These results indicate that SAE1a is necessary for maintaining SUMO conjugation homeostasis *in vivo* and that SAE1a and SAE1b are not fully redundant activities. Diversification of the E1 enzyme is not restricted to *Arabidopsis*, suggesting that the presence of several E1 isoforms might constitute an evolutionary advantage. Taken together, we postulate that the E1 small subunit SAE1 could be modulating downstream steps during SUMO conjugation providing a novel mechanism to control sumoylation *in vivo*.

RESULTS

Genes Encoding the SUMO-Activating Enzyme

According to annotations in the Arabidopsis Information Resource (TAIR), two variants of the gene encoding the E1 large subunit SAE2 could exist (Figure 1A). The second variant would result from an alternative splicing that would generate a mature mRNA retaining the last intron. In order to identify this second variant, we extracted total RNA from several tissues and used it as a template for cDNA synthesis using oligo dT. The resulting cDNA was analyzed by PCR using three primer pair combinations (Figure 1A). We only detected PCR products when primers annealing to the *SAE2 variant 1* were present in the reaction. These products were unique and displayed an Rf consistent with the presence of *SAE2 variant 1* (SM020–SM004 and SM019–SM004). When primers annealing specifically to *SAE2 variant 2* were used (SM021–SM004), no reaction products were obtained in any of the analyzed samples, suggesting that *SAE2 variant 2* does not exist *in vivo* and that is the result of an annotation error (Figure 1B). Moreover, three gene accessions codifying for the E1 small subunit SAE1 are found: At4g24940, At5g50580, and At5g50680. The first accession codifies for the SAE1 isoform a, and the others correspond to two exact copies of a gene codifying for the SAE1 isoform b located in tandem

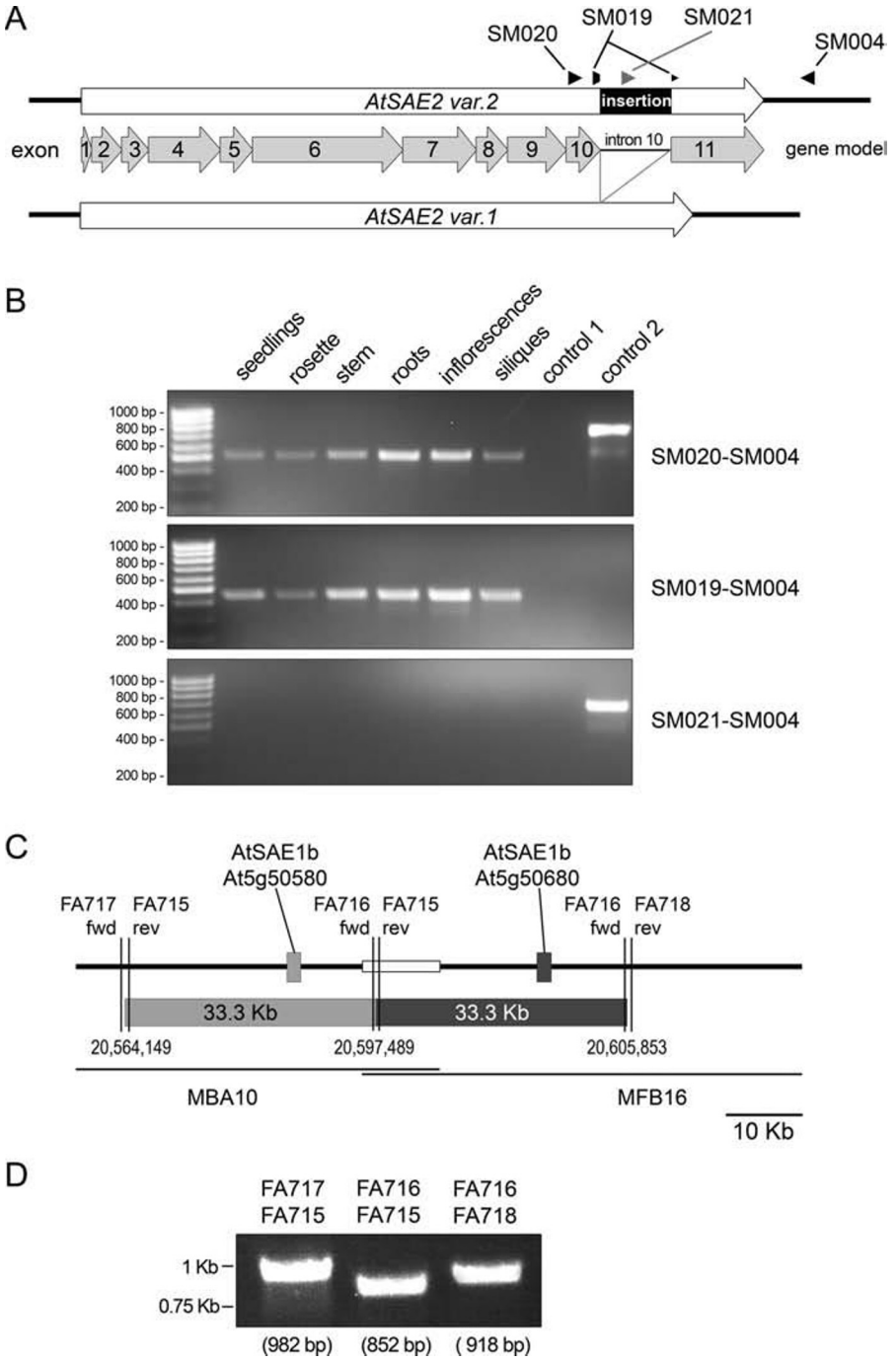

Figure 1. Gene Structure of the *Arabidopsis* E1-Activating Enzyme.

(A) Schematic representation of the *SAE2* variant 2 as annotated in Arabidopsis Information Resource (TAIR) is shown on top. The insertion in *SAE2* variant 2 corresponds to intron 10 (middle) that is excised in variant 1 (bottom). The annealing sites of the primers used to analyze the *SAE2* variant 2 expression are shown above the scheme.

(B) RNA extracted from the specified tissues was retrotranscribed using oligo dT and the resulting cDNA synthesized was analyzed by PCR using the primers pair indicated on the right. PCR products were resolved on agarose gel electrophoresis and stained with ethidium bromide. Control 1 corresponds to RT negative control and control 2 contains Col0 genomic DNA as a template.

(C) Schematic representation of the *Arabidopsis* chromosome 5 genomic region containing a 33-Kb-sequence duplication. SAE1b genomic regions are represented by light gray and dark gray boxes. The MBA10 and MFB16 overlapping region is indicated by a white rectangle (top). Duplicated regions sharing a 99% of DNA sequence identity are represented by light gray and dark gray rectangles, and their coordinates in chromosome 5 are also indicated (middle). Contigs containing the represented genomic region are indicated at the bottom, MBA10 and MFB16.

(D) PCR products from reactions performed in the presence of the indicated primers were analyzed as in (B). The expected size of each PCR product is indicated below each lane.

(Figure 1C). PCR analysis of *Arabidopsis* genomic DNA confirmed the existence of a 33.3-Kb tandem duplication that includes the gene codifying for SAE1b (Figure 1D). Sequencing of the PCR product synthesized using the primer pair FA716–FA715 confirmed that the amplified fragment corresponded to the region connecting both tandem repeats. This genomic duplication affects 11 additional genes (Supplemental Table 1). Comparative genome analysis indicated that this duplication is not present in the *Arabidopsis thaliana* ecotypes Ler-1, C24, Bur-0, and Kro-0 (Supplemental Figure 1).

Evolutionary Diversification of SUMO-Activating Enzyme among Plants

We next examined SUMO E1 evolutionary diversification in plants. Search for *SAE1* genes resulted in the identification of 39 sequences, including the SAE1 family from 31 species with fully sequenced genomes plus a gymnosperm representative from *Picea sitchensis* (Supplemental Table 2). Similarly, searches for *SAE2* genes returned 33 sequences corresponding to the SAE2 family from 27 species with fully sequenced genomes (Supplemental Table 3). Both sets of sequences included representative of the main plant evolutionary lineages. To investigate the evolutionary relationships among *SAE2* and *SAE1* genes in plants, we performed phylogenetic analyses on the basis of protein sequence alignments. Three alternative methods of phylogenetic reconstruction were used: Bayesian inference, maximum likelihood, and neighbor joining. Bayesian phylogenetic trees are shown in Figure 2. Trees obtained from the three alternative methods show almost identical topologies, except for a few internal nodes. The exon/intron structures of *SAE1* and *SAE2* genes are also displayed next to the corresponding trees. The intron positions and exon phases are remarkably well conserved, providing further support to the phylogenetic analyses. Only minor differences could be observed in intron number and length.

In both cases, clustering of the sequences in the trees reflects quite well the taxonomical relationships of the species represented. Two independent clades grouping dicot and monocot SAE1 and SAE2 sequences were retrieved with high statistical support. Many plant species show SAE1 and SAE2 families composed of a single gene. However, phylogenetic analyses reveal lineage-specific gene duplication within the SAE1 and SAE2 families. This is the case of *Arabidopsis thaliana*, displaying three *SAE1* genes resulting from two duplication events. One likely resulted from a polyploidization event predating the emergence of the Brassicaceae lineage. Consistently, *A. thaliana At5G50580* and *At4G24940* genes map in genomic regions of chromosomes 5 and 4 resulting from a whole-genome duplication event estimated to have occurred in the last 25–40 million years (Blanc et al., 2003), matching the time of origin and diversification of the Brassicaceae family (Couvreur et al., 2010). A second involved a more recent tandem duplication likely occurring after divergence of *A. thaliana* Col-0 from other ecotypes (Supplemental Figure 1). Besides *A. thaliana* Col-0, the remaining Brassicaceae species display two *SAE1* genes, with the exception of only *Brassica rapa*, which displays a single *SAE1* gene. Interestingly, searches of the *B. rapa* genome identify two additional sequences with shorter lengths showing similarity with SAE1 (Bra023205 and Bra03552), maybe corresponding to *SAE1* pseudogenes. Additional lineage-specific *SAE1* gene duplications could be observed in *Glycine max*, *Populus trichocarpa*, and *Manihot esculenta*. These three species, together with *Malus domestica*, *B. Rapa*, and the basal land plant species *Physcomitrella patens* also show two *SAE2* genes resulting from gene duplication events occurring recently in the corresponding lineages.

Conservation and Structure Prediction of SUMO-Activating Enzyme Isoforms

In yeast and mammals, there is a single form of the SUMO E1 enzyme. In contrast, *Arabidopsis* expresses two forms of the E1 that differ in the small subunit composition. The *Arabidopsis* E1 large subunit SAE2 shares a 36% of amino acid sequence identity with its human ortholog HsSAE2, although this conservation is not evenly distributed between functional domains. The adenylation domain (residues 1–153 + 379–432) is the most conserved, with a 55% of sequence identity, the C-terminal tail is the less conserved region (16% identity; residues 546–625), and the catalytic cysteine (residues 154–378) and the UFD domains (residues 433–545) share a 31% of sequence identity with their human counterparts (Figure 3A and 3B). Similarly, when we analyzed the conservation degree among plant SAE2 paralogs, the SAE2 adenylation domain and the C-terminal tail presented the highest and the lowest conservation degree, respectively.

The two *Arabidopsis* isoforms of the E1 activating enzyme small subunit, SAE1a and SAE1b, are highly conserved. The analysis of the E1 structure model showed that the SAE1a/b regions involved in the contacts with the large subunit have a larger conservation degree with its human ortholog (41–44% of sequence identity; residues 6–106 + 271–317), whereas, sequences located in distant regions from the heterodimer interface are more divergent (21–20% of sequence identity; residues 107–270; Supplemental Figure 3). The most conserved regions correspond to those domains in the human SAE1 ortholog that undergo conformational changes and/or that contain residues that establish different contacts between the E1 open and closed conformations during SUMO activation. These residues are also conserved between *Arabidopsis* SAE1 paralogs. On the other hand, we could not identify any domain susceptible to determine functional specialization between SAE1a and SAE1b since non-conserved residues are evenly distributed among these paralogs (Figure 3C).

In Vitro SUMO Conjugation Rate Is Dependent on SUMO-Activating Enzyme Isoforms

Arabidopsis SAE1 paralogs display a 20% of sequence divergence and we analyzed a possible effect of these differences

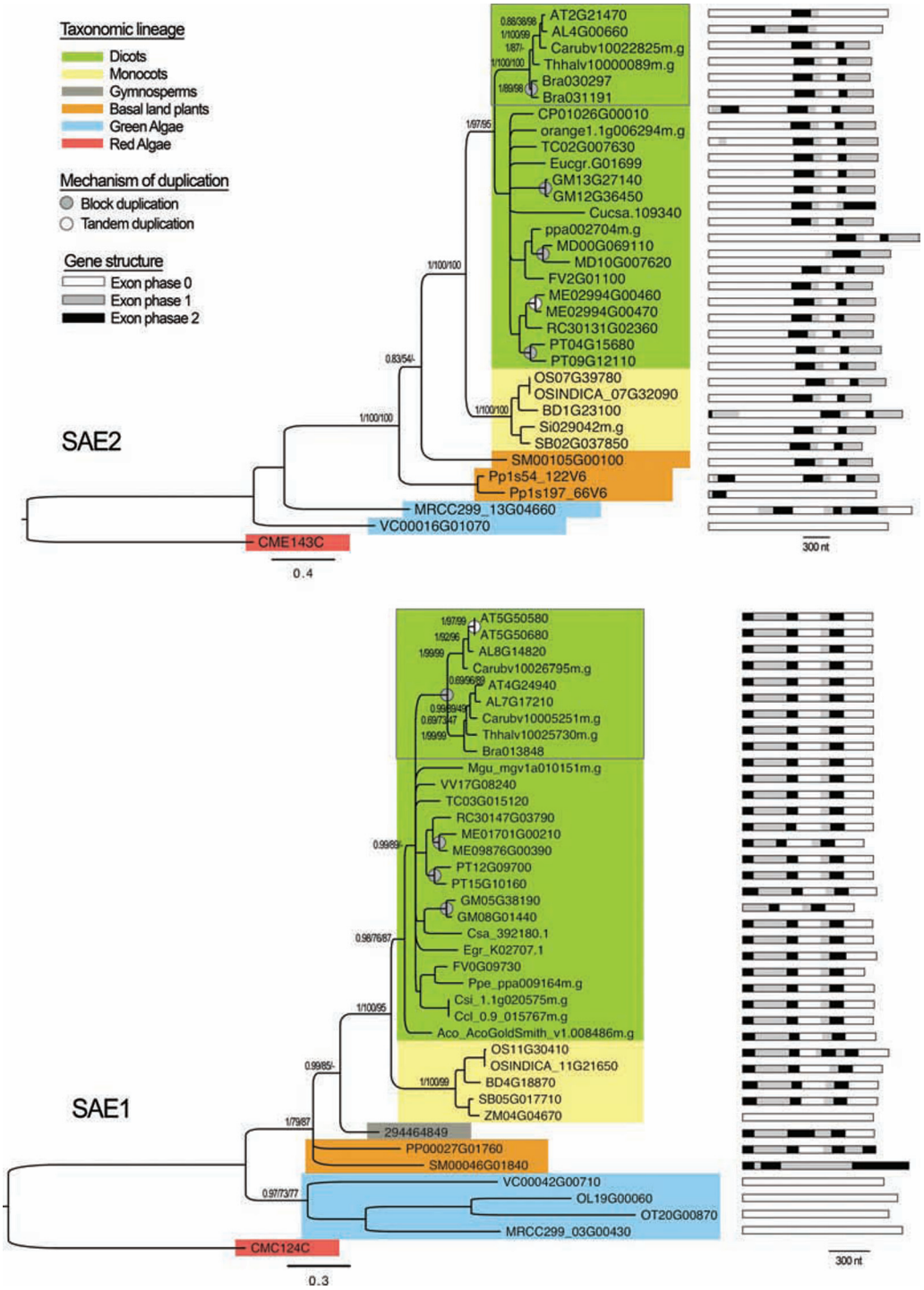

Figure 2. Bayesian Phylogeny and Exon–Intron Structure of Plant SUMO-Activating Enzyme (*SAE*) Genes.
Bayesian phylogenetic tree depicting the evolutionary relationships among 39 *SAE1* protein sequences from 32 plant species **(A)**, and 33 *SAE2* protein sequences from 27 plant species **(B)**. Both trees were rooted using the corresponding ortholog from the red algae *Cyanidioschyzon merolae*. Values next to the nodes indicate statistical support on relevant clades (Bayesian posterior probabilities/maximum-likelihood a LRT support values/neighbor-joining bootstrap values). The tree is drawn to scale, with branch lengths proportional to evolutionary distances between nodes. The scale bar indicates the estimated number of amino acid substitutions per site. The origin of gene duplication (tandem versus block duplication) is indicated at the corresponding nodes. Clades clustering Brassicaceae representatives are highlighted within a dashed box. Exon organization is shown on the right, with boxes colored according to phases, except for the *P. sitchensis* SAE1 representative, for which the genomic sequence was not available.

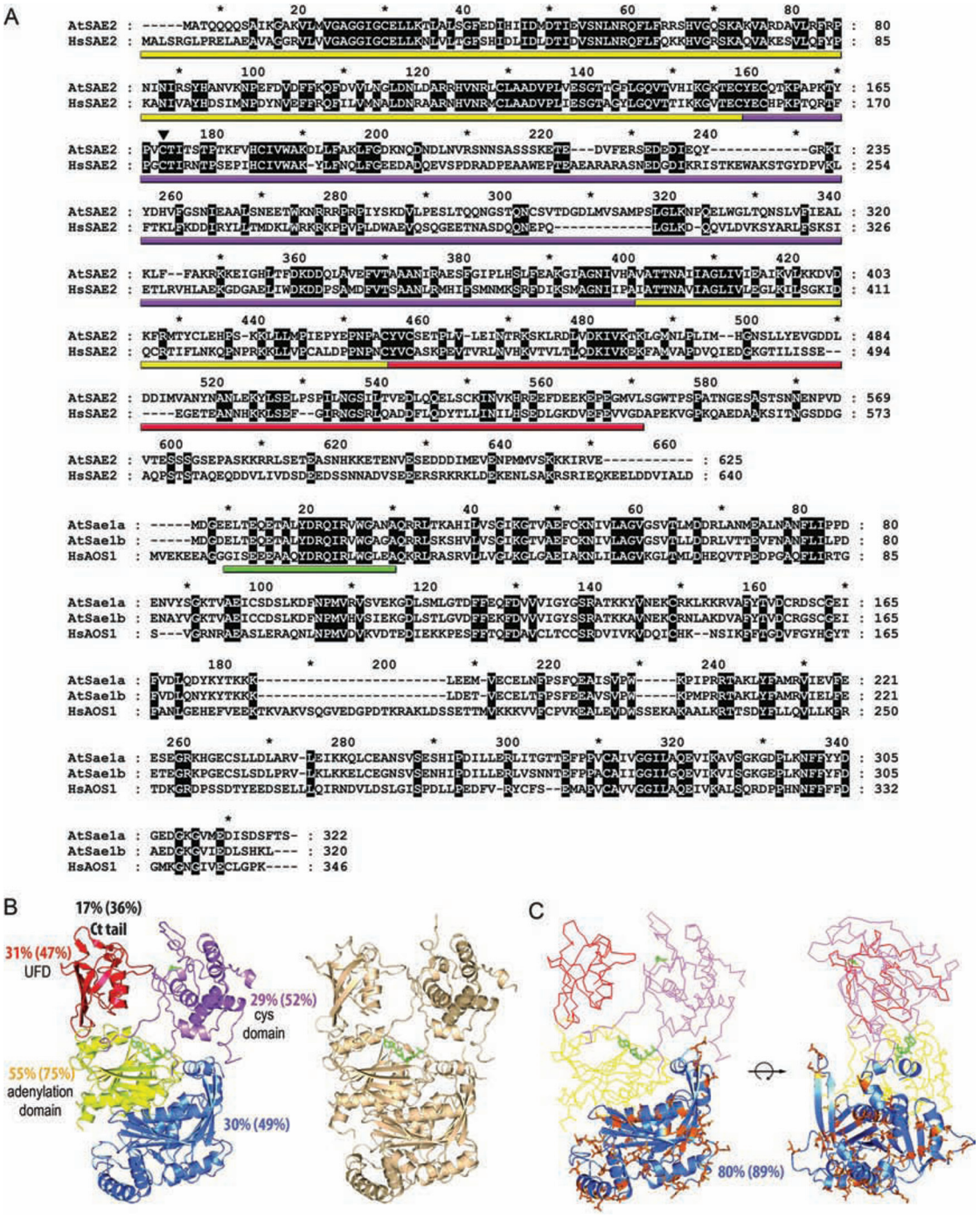

Figure 3. SUMO-Activating Enzyme Isoforms Conservation.
(A) Amino acid sequence alignment of *Arabidopsis* (At) and human (Hs) orthologs of E1 SUMO-activating enzyme large subunit (SAE2; top) and small subunit (SAE1; bottom). Black background and white letters correspond to 100% sequence identity. Distribution of SAE2 functional domains, adenylation domain (yellow; 1–153 + 379–432), cysteine domain (violet; 154–378), and the ubiquitin-fold domain (UFD, red; 433–545) is shown below the sequence. The C-terminal tail (546–625) is not underlined. The catalytic cysteine residue is indicated by an arrowhead. The SAE1 region containing residues that undergo conformational changes during SUMO activation is underlined (green).
(B) *Arabidopsis* SUMO E1 structure, as predicted by the SWISS-MODEL comparative protein modeling server, and the human SUMO E1 structure used as a template (1Y8Q) are shown as a ribbon diagram on the right and the left, respectively. SAE2 functional domains are colored as in (A), the catalytic cysteine and ATP are in green, and the SAE1 subunit is in blue. Sequence identity percentage is shown by domain and similarity is indicated in parentheses. The *Arabidopsis* SAE2 C-terminal tail is not represented, since is not resolved in the original structure template.
(C) Non-conserved amino acids between SAE1a and SAE1b paralogs are shown in brown and side chains are represented as sticks. Sequence identity and similarity are shown as in (B).

in the heterodimer activity. We tested SUMO conjugation efficiency in reactions containing the SUMO-activating enzyme isoform a (E1a: SAE2/SAE1a) or isoform b (E1b: SAE2/SAE1b). Time-course sumoylation reactions incubated at 22°C, 37°C, and 42°C showed that SUMO conjugation increased with temperature (Figure 4A–4C). Quantification of relative reaction efficiency among E1a and E1b isoforms indicated that reactions containing the E1a isoform conjugated SUMO significantly more efficiently at 37°C and 42°C, although these differences were not significant at 22°C (Figure 4E). We investigated whether these differences were related to SUMO–E1 thioester formation. Our results showed that the E1 holoenzyme containing the isoform SAE1a was also slightly more efficient at establishing SUMO–E1 thioester bonds but this effect was restricted to high temperatures (42°C and 48°C) (Figure 4D and 4F). These results suggest that E1 small subunit isoform could have an effect in downstream steps to SUMO–E1 thioester bond formation.

Subcellular Localization of SUMO-Activating Enzyme Isoforms

SAE2 and SAE1a/b amino acid sequences were analyzed using the software for prediction of protein localization PSORT (http://wolfpsort.org/). As a result, five putative nuclear localization signals were identified in SAE2 and none in SAE1a/b. Four of them belong to the NLS pat4 type, located at the amino acid positions 257, 258, 326, and 582, and the fifth belongs to the NLS pat7 type and it is located at the amino acid position 579 (Figure 5A). To perform the functional analysis of the predicted NLS motifs, the enhanced yellow fluorescent protein (EYFP) was fused to the C-terminus of full-length SAE2 (residues 1–625) or the C-terminal deletion mutant SAE2ΔCt

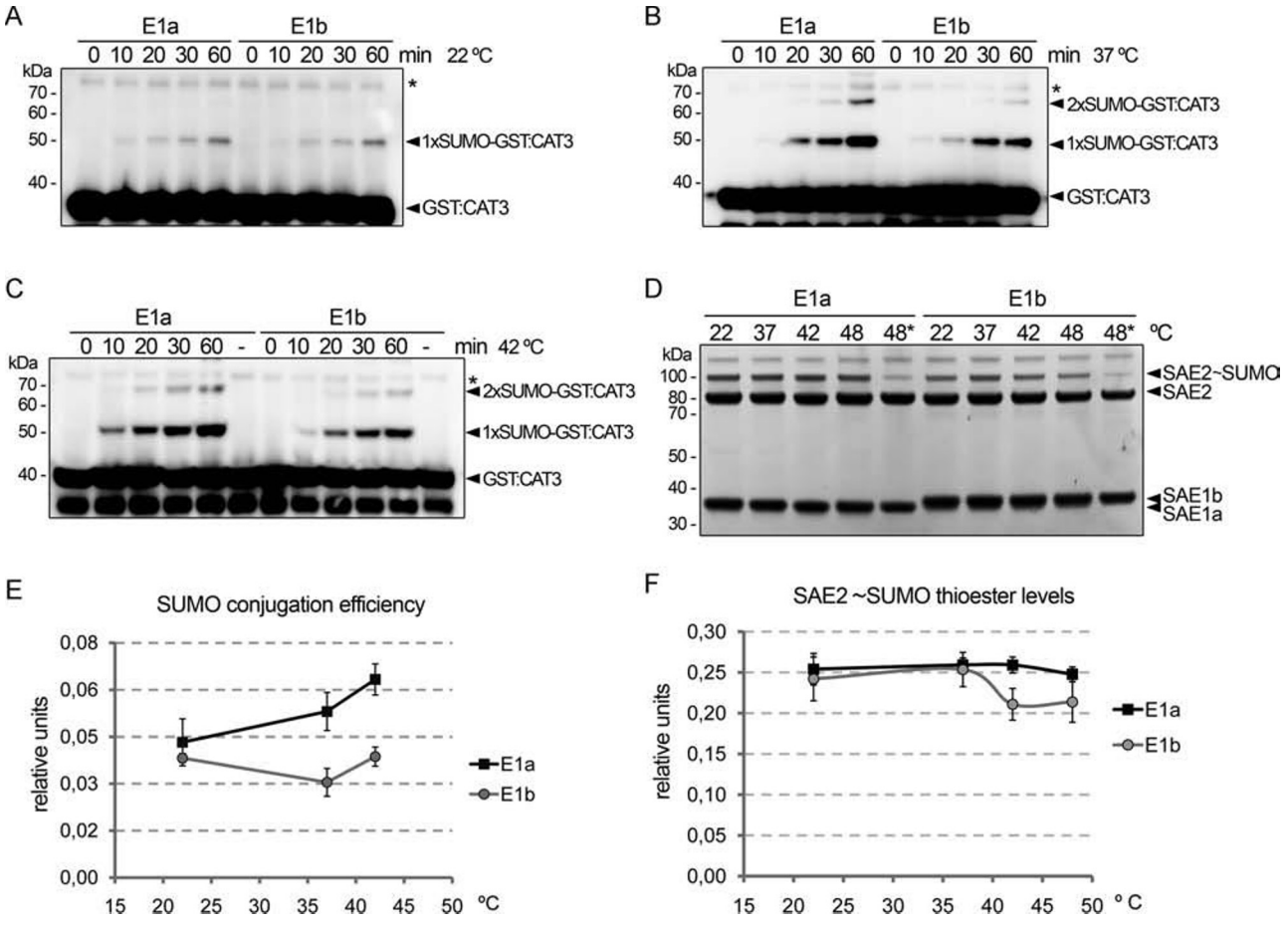

Figure 4. *In Vitro* Characterization of *Arabidopsis* SUMO-Activating Enzyme Isoforms.
(A–C) *In vitro* sumoylation assays were performed at 22°C, 37°C, and 42°C in the presence of E1a or E1b, SUMO2, SCE1, and GST:CAT3Ct as a substrate. Reaction mixtures were stopped at the indicated time, and products were resolved by SDS–PAGE and examined by immunoblot analysis with anti-GST antibodies. The asterisk indicates a contaminating protein.
(D) SAE2~SUMO thioester formation assays were performed using E1a (SAE2:SAE1a) or E1b (SAE2:SAE1b) and SUMO1. Reaction mixtures were incubated at the indicated temperatures and stopped after 2 min. As a control for thioester bond formation, an aliquot of the reaction incubated at 48°C was treated with DTT (48*). Reaction products were separated by SDS–PAGE and stained with Coomassie fluor orange.
(E, F) Reactions were performed at least in triplicate and GST:CAT3Ct sumoylation efficiency or SAE2~SUMO thioester bond formation level quantified. Average values and standard deviation bars were plotted on the graphs.

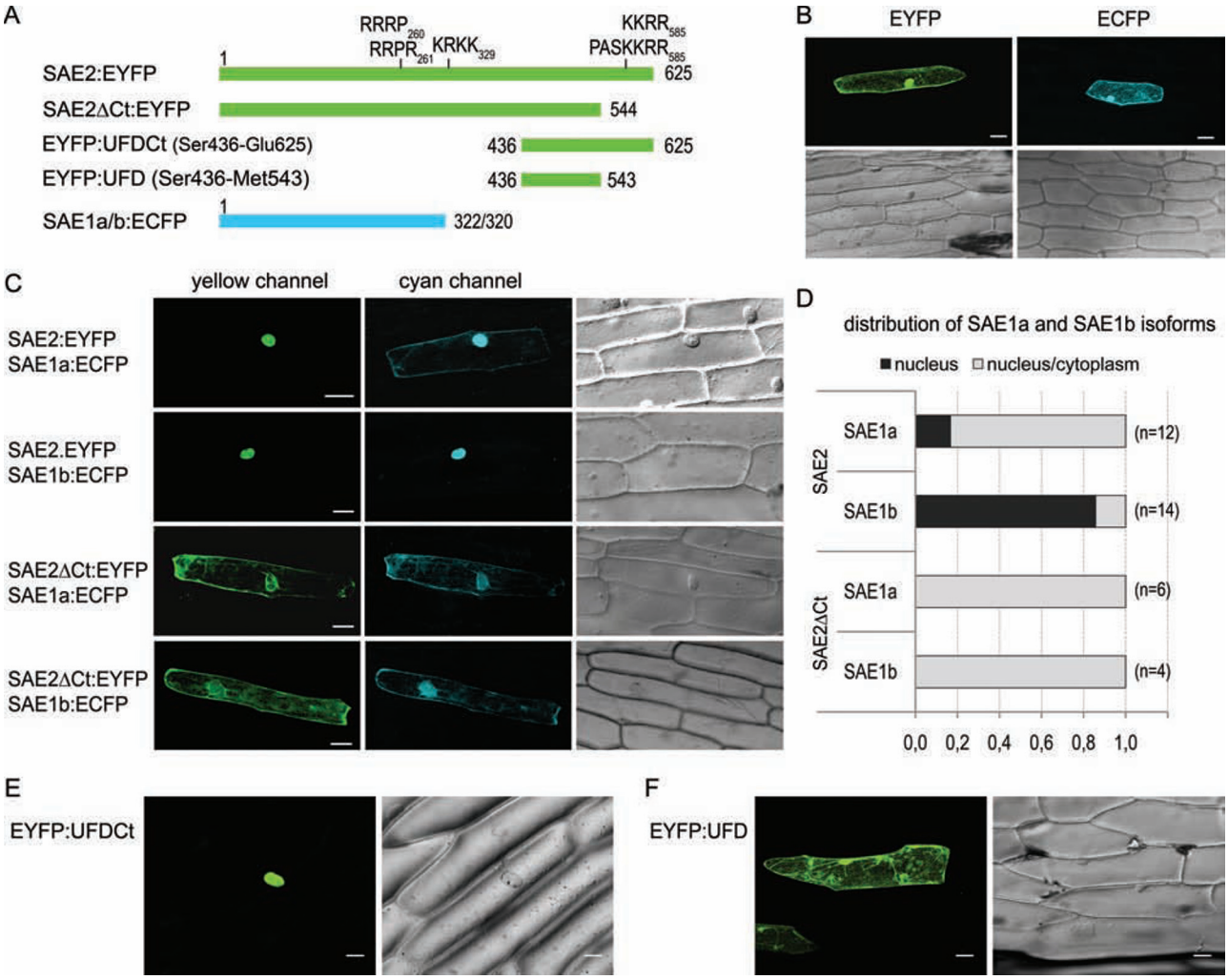

Figure 5. Subcellular Localization of *Arabidopsis* SUMO E1-Activating Enzyme Isoforms.
(A) Full-length SAE2, SAE2 with a C-terminal tail deletion, and SAE2 domain S436–E625 were fused to the EYFP (enhanced yellow fluorescent protein) and transiently expressed in onion epidermal cells. For co-localization studies, SAE1a/b were fused to the ECFP (enhanced cyan fluorescent protein). Nuclear localization signals as predicted by the PSORT software (http://psort.ims.u-tokyo.ac.jp/) are shown above the SAE2 representation and their position on the protein sequence indicated.
(B) Cells expressing EYFP or ECFP were used as control.
(C) E1a and E1b localization experiments were performed by co-expression of SAE2:EYFP and SAE1a:ECFP, SAE2ΔCt:EYFP and SAE1a:ECFP, and EYFP:SAE2 and SAE1b:ECFP. Light transmission images of the onion epidermal cells are also shown. Bars = 50 μm.
(D) Subcellular distribution of SAE2, SAE2ΔCt, SAE1a, and SAE1b was scored and represented on the plot (*n* indicates the number of cells analyzed).
(E) EYFP:SAE2 Ser436–Glu625 and **(F)** EYFP:SAE2 Ser436–Met543 localization. Bars = 50 μm.

(residues 1–545), which retains the three predicted NLS motifs located at positions 257, 258, and 326 but not the two motifs located at positions 579 and 582. The resulting fusion proteins are indicated as SAE2:EYFP and SAE2ΔCt:EYFP, respectively. In addition, enhanced cyan fluorescent protein (ECFP) was fused to the C-terminus of SAE1a or SAE1b generating the fusion proteins SAE1a:ECFP and SAE1b:ECFP, respectively. In these experiments, free EYFP and ECFP were used as controls (Figure 5B).

When SAE2:EYFP was co-expressed with SAE1a:ECFP, the fluorescence signal from EYFP was only detected in the nucleus, indicating that the E1 large subunit SAE2 localizes to the nucleus. On the contrary, also in the presence of SAE1a:ECFP, SAE2ΔCt:EYFP localized to the cytoplasm and the nuclear envelope, indicating that the SAE2 C-terminal region is required for SAE2 nuclear targeting. Identical results were obtained when SAE1b:ECFP was used instead of SAE1a:ECFP in co-expression experiments, suggesting that SAE2 subcellular localization is not affected by the small subunit SAE1 isoform (Figure 5C, left column). The same pattern was observed in all analyzed cells.

To further evaluate whether the SAE2 region containing the predicted nuclear localization signals at positions 579 and 582 was sufficient to determine nuclear localization, the SAE2 domain S426–E625 was fused to EYFP C-terminus

generating the construct EYFP:UFDCt. In onion epidermal cells, EYFP:UFDCt localized exclusively in the nucleus (Figure 5E). Consistently with a function of the SAE2 C-terminal tail in nuclear targeting, the fusion protein in which the C-terminal tail was removed, EYFP:UFD, displayed the same localization as the control EYFP (Figure 5F and 5B). In the last case, as in the control, signal was also detected in the nucleus since EYFP:UFD molecular weight, 40.5 kDa, is below the nuclear size exclusion limit.

Since the active E1 enzyme consists of the two subunits SAE2 and SAE1, it would be expected that both subunits display the same subcellular localization. When co-expressed with the nuclear-localized SAE2:EYFP, fluorescence from SAE1b:ECFP was observed exclusively in the nucleus in the 86% of analyzed cells. In addition to the nucleus, fluorescence from SAE1b:ECFP could also be detected in the cytoplasm in the other 14% of the cells. In contrast, in the presence of the nuclear-localized SAE2:EYFP, fluorescence from SAE1a:ECFP was detected in the nucleus and the cytoplasm in 83% of the analyzed cells. Exclusive nuclear localization of SAE1a:ECFP was restricted to the other 17% of the cells (Figure 5C, middle, and 5D). We excluded that these distinct distributions were the result of differential expression levels since SAE1a:ECFP and SAE1b:ECFP displayed comparable expression levels and localization, nuclear and cytoplasmic in all the analyzed cells, when co-expressed with SAE2ΔCt:EYFP (Figure 5C, middle, and 5D).

SAE1a Depletion Results in SUMO Conjugation Defects in Response to Stress

Since both isoforms of the SUMO-activating enzyme small subunit confer distinct conjugation efficiency *in vitro*, we explored the effect of SAE1a and SAE1b deletion in SUMO conjugation during stress responses. As *SAE1b* is duplicated in tandem in the genome, it is virtually impossible to obtain null T-DNA insertion *atsae1b* mutant plants. Thus, we focus on the analysis of the effect of SAE1a deletion in SUMO conjugation *in vivo*. We identified an *atsae1a* mutant line with a T-DNA insertion in the ninth exon of the *SAE1a* genomic sequence (Figure 6A). The T-DNA insertion site was corroborated by PCR combining primers complementary to *SAE1a* and T-DNA

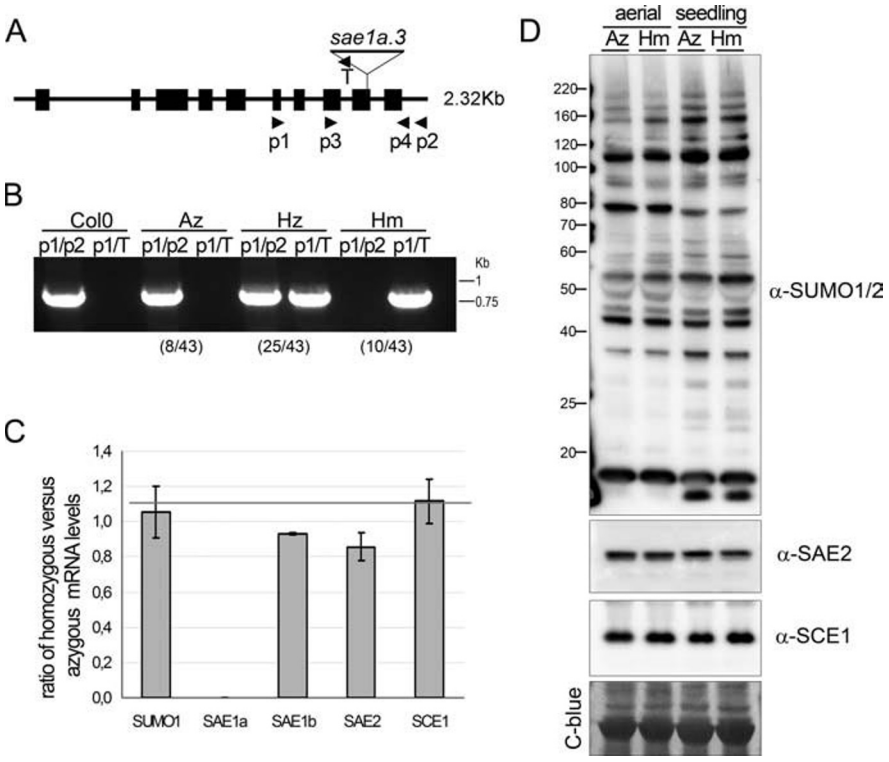

Figure 6. SUMO Conjugation in *atsae1a* Mutant Plants.

(A) *SAE1a* genomic region showing T-DNA insertion in the ninth exon. Annealing regions of primers used in (B) are represented by arrowheads.

(B) PCR products from plant genotype analysis were resolved in DNA agarose gel. Primers annealing to SAE1a genomic region (p1 and p2) and the T-DNA region (T) were used to identify azygous, heterozygous, and homozygous plants. Segregation results with respect to 43 individuals are indicated below the DNA agarose gel.

(C) mRNA levels corresponding to SUMO1, E1-activating enzyme (SAE1a, SAE1b, and SAE2), and E2-conjugating enzyme (SCE1) were determined in *atsae1a* azygous and homozygous lines. Collected data were normalized by using PR65 as a reference gene. Average values and standard deviation bars correspond to three biological replicates. Results are expressed as a ratio of the expression levels in homozygous plants versus azygous plants.

(D) Relative protein levels of free SUMO, SUMO conjugates, the E1-activating enzyme large subunit SAE2, and the E2-conjugating enzyme SCE1 in azygous and homozygous *sae1a* plants. 18 μg of total protein extracts from aerial parts and seedlings were resolved by SDS–PAGE and examined by immunoblot analysis with specific antibodies. A portion of the Coomassie blue (C-Blue)-stained membrane is shown as a loading control.

genomic sequences. The segregation analysis showed that *atsae1a* mutant was viable in homozygosis (Figure 6B). From the original segregating population, we selected azygous (#3.7) and homozygous lines (#3.9) for further characterization. Quantification of mRNA levels of other members of the sumoylation machinery showed that there were no major differences between the homozygous *atsae1a* and the corresponding azygous lines, excluding the presence of a compensatory mechanism affecting mRNA levels (Figure 6C). Also, no major differences could be observed among SUMO conjugates accumulation and SAE2 and SCE1 levels in *atsae1a* plants (Figure 6D). Under standard growth conditions, no significant developmental defects were observed in *atsae1a* plants (Supplemental Figure 5).

It is well established that heat and drought stresses induce a massive accumulation of high-molecular-weight SUMO conjugates (Kurepa et al., 2003), which presumably requires a highly active sumoylation machinery. Upon heat and drought stresses, *atsae1a* homozygous plants accumulated SUMO conjugates to a lower extent than azygous plants, which is consistent with the presence of defects in SUMO conjugation machinery (Figure 7A and 7B).

DISCUSSION

The E1-activating enzyme catalyzes the first step in SUMO conjugation and it has been proposed to have a role in SUMO isoform selection (Castaño-Miquel et al., 2011). In *Arabidopsis*, the E1-activating enzyme is present as two isoforms that differ in small subunit composition. The diversification of the small subunit into two conserved isoforms, SAE1a and SAE1b, is intriguing, since most of the E1 functional domains are located in the large subunit, SAE2. Structural analyses allowed the mapping of all functional domains in SAE2 as described for its human ortholog, although their conservation degree is not equally distributed among them. The most conserved is the adenylation domain, which is the only domain that depends on heterodimer formation. On the other hand, the cysteine and UFD domains are more divergent. It is remarkable that both the cysteine and the UFD domains establish non-covalent interactions with their cognate E2 conjugating enzyme, SCE1 (Wang et al., 2007, 2009, 2010), suggesting that sequence divergence in these regions between paralogs could be the result of protein–protein interaction surface optimization in each species. Other SAE2 regions containing regulatory elements display also a high divergence degree between paralogs, such as the Cys domain loop containing SUMO attachment sites and the C-terminal tail containing nuclear localization signals. These specific SAE2 regulatory region divergences suggest that the molecular mechanisms controlling SUMO conjugation *in vivo* might have diverged during evolution. Whether these differences are related to the distinct biological roles that SUMO plays in different organisms remains to be elucidated.

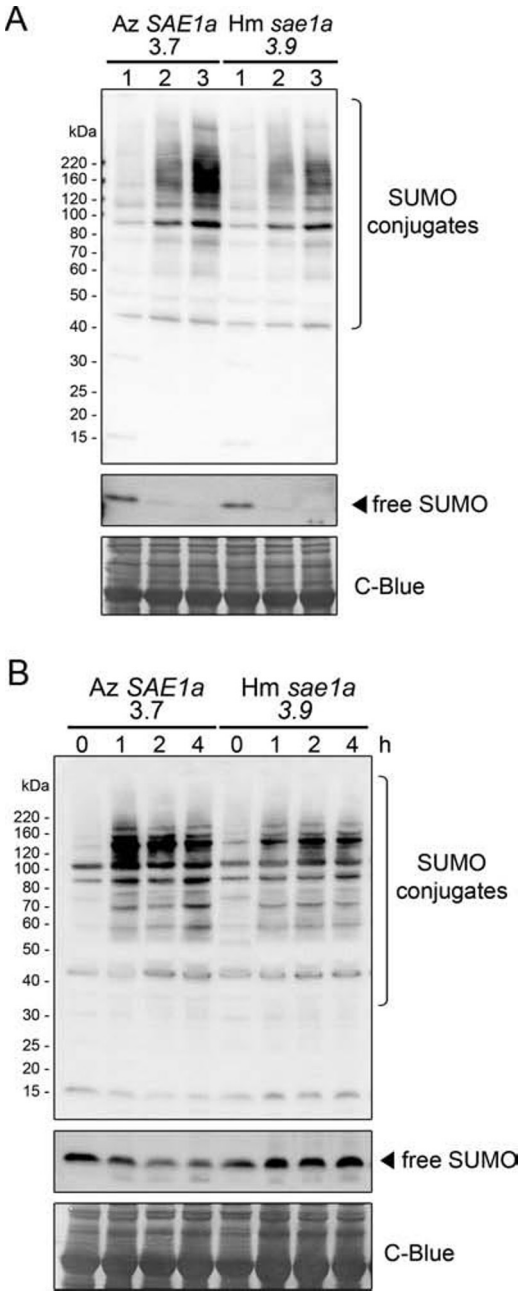

Figure 7. *atsae1a* Plants Display Defects in SUMO Conjugate Accumulation upon Abiotic Stress.

(A) SUMO conjugation upon heat-shock treatment. Azygous *SAE1a* (line #3.7) and homozygous *atsae1a* (lines #3.9) plants grown in liquid culture for 10 d were exposed to 42°C for 30 min (lane 2) and then returned at 22°C for 30 min for recovery (lane 3). Control plants were kept at 22°C (lane 1).

(B) SUMO conjugation upon dehydration treatment. Rosettes from 2-week-old plants grown in MS were detached from their roots and placed in a laminar flow hood. Samples were collected at the specified times. Total protein extracts were resolved by SDS–PAGE and SUMO conjugate levels examined by immunoblot analysis with anti-SUMO1 antibodies.

The E1 small subunit complexity is even higher at the gene organization level. *Arabidopsis thaliana* displays three *SAE1* genes (*SAE1a*, *SAE1b1*, and *SAE1b2*) resulting from two

Table 1. *Arabidopsis* SUMO Conjugation Machinery Components Analyzed in this Study.

Component	Name	TAIR accession	Protein MW
SUMO isoform 1	SUMO1	At4g26840	10.9 kDa
SUMO-activating enzyme subunit 2	SAE2	At2g21470	69.7 kDa
SUMO-activating enzyme subunit 1a	SAE1a	At4g24940	36.1 kDa
SUMO-activating enzyme subunit 1b	SAE1b	At5g50580 At5g50680	35.6 kDa
SUMO-activating enzyme	SCE1	At3g57870	18 kDa

duplications events, one preceding the emergence of the *Brassicacea* as indicated by the presence of two *SAE1* genes (*SAE1a* and *SAE1b*) in all family representatives, with the exception of *B. rapa*, and a more recent tandem duplication, which would have generated *SAE1b1* and *SAE1b2* genes. The fact that *SAE1b1* (At5g50580) and *SAE1b2* (At5g50680) codify for two identical proteins brought speculation about a possible annotation error (Novatchkova et al., 2012). Our results demonstrated the existence of this tandem duplication that affects 13 genes codifying for functionally unrelated proteins. Surprisingly, this duplication is not present in other *Arabidopsis thaliana* ecotypes analyzed, suggesting that tandem duplication giving raise to At5g50580 and At5g50680 occurred recently in the Col-0 lineage, after divergence from the closely related accessions/ecotypes Ler-1, C-24, Bur-0, and Kro-0. In addition to the evolutionary questions raised by this finding, it also opens the possibility to address the *in vivo* functional divergence of SAE1a/b isoforms in other *Arabidopsis thaliana* ecotypes containing a single gene codifying for each isoform.

The analysis of SAE1a and SAE1b primary sequences did not reveal the presence of any divergent domain that could suggest a functional specialization. Residues relevant to adenylation and those that undergo conformational changes during SUMO–SAE2 thioester bond formation are conserved in both isoforms. Nonetheless, SAE1a conferred higher conjugation efficiency relative to SAE1b under several incubation temperatures. In the case of SAE2 thioester bond formation, differences were restricted to reactions incubated at high temperatures, suggesting that SAE1 subunit composition would have a larger effect in events occurring downstream of SUMO–E1 thioester bond formation. SUMO activation and transfer to E2 involve dramatic conformational changes consisting of SAE2 rotations and unfolding and folding of specific regions in SAE2/SAE1 (Schulman, 2011). Considering that conformational flexibility is critical to biological activity (Freire, 2001), it is tempting to speculate that divergences in atomic interactions defining SAE1a and SAE1b three-dimensional structure might confer different stability properties to the E1 activating enzyme, which would affect SUMO conjugation.

The functional E1 nuclear localization signals have been analyzed using two approaches. In the first one, the full E1 heterodimer containing the native SAE2 or the SAE2 truncated form in which the C-terminal tail was removed were used. In the second, individual expression of the SAE2 C-terminal region not containing the predicted NLS located in the cysteine domain and including or not the C-terminal tail were analyzed. In both cases, the results indicated the crucial role of the SAE2 C-terminal tail targeting SAE2 to the nucleus, which is consistent with the nuclear localization of most SUMO conjugates (Saracco et al., 2007). Similarly, recent studies have shown that a nuclear localization signal located in the mammalian E1 large subunit is crucial for E1 nuclear targeting (Moutty et al., 2011). Even more recent is the finding that sumoylation of human SAE2 C-terminal tail is involved in SAE2 nuclear targeting (Truong et al., 2012b), suggesting that SAE2 nuclear localization is a highly regulated process involving different molecular strategies. All together, the emerging data support that the essential nuclear role of SUMO has been conserved through evolution. In contrast, the E1 small subunit SAE1 could be detected in the nucleus and the cytoplasm. This distribution was particularly evident for the isoform SAE1a, suggesting that a cytosolic reservoir of the SAE1 apoenzyme could exist. Previous studies showed that SAE2 transcript levels are significantly lower relative to SAE1a and SAE1b and it was proposed that a posttranslational mechanism would adjust the stoichiometry of E1 heterodimer subunit levels (Saracco et al., 2007). On the contrary, our results suggest that SAE1, and predominantly the SAE1a isoform, could exist as a free form. In this case and under SAE2-limiting concentrations, as suggested by transcriptional analysis, the most abundant SAE1 isoform would be preferentially assembled into the E1 holoenzyme, which could constitute a mechanism to regulate the SUMO conjugation rate *in vivo*.

In order to explore the relative contributions of SAE1a and SAE1b to SUMO conjugation *in vivo*, we searched for null *sae1a* and *sae1b* mutant plants. Under standard growth chamber conditions, the deletion of SAE1a did not result in obvious developmental alterations, consistently with a normal basal SUMO conjugate pattern displayed by *sae1a* homozygous plants relative to azygous plants. On the contrary, when sumoylation was analyzed in plants under heat and drought stresses, which induce a massive accumulation of SUMO conjugates, *sae1a* homozygous plants showed a reduction in SUMO conjugates accumulation. Overall, these results suggest that SAE1a is required for a complete functional sumoylation system under stress. Unfortunately, equivalent analysis could not be performed with SAE1b-null plants since *SAE1b* is duplicated in tandem in the *A. thaliana* Col0 genome and *sae1b* mutant plants cannot be isolated. As duplications within the SAE1 family are not restricted to *A. thaliana*, one hypothesis could be that duplications of the sumoylation system could constitute an advantage for its functionality.

There is increasing evidence supporting a regulatory role of the SUMO E1-activating enzyme in SUMO conjugation. In

mammals, low ROS levels inhibit SUMO conjugation by inducing a cross-linking between E1 and E2 enzymes active site (Bossis and Melchior, 2006). Also, the E1 large subunit SAE2 is modified by SUMO in two sites: one provides an inactive E1 pool that becomes activated by desumoylation under heat shock stress and the other contributes to SAE2 nuclear localization (Truong et al., 2012a, 2012b). In *Arabidopsis*, SUMO E1 was shown to participate in SUMO paralog selection (Castaño-Miquel et al., 2011). Overall, our results contribute to the notion that SUMO activation could constitute a regulatory step in the SUMO conjugation pathway. In addition to the conserved nuclear localization of the E1-activating enzyme, we have shown that the SAE1 isoform contained in the E1 holoenzyme has an effect in SUMO conjugation efficiency through a mechanism occurring most probably downstream of SUMO activation. *In vivo*, although we cannot discriminate between a gene dosage effect or the fact that SAE1a isoform confers a higher conjugation efficiency or both, our results show that SAE1a is necessary to maintain *in vivo* sumoylation homeostasis, pointing to a rate-limiting role of SUMO activation. E1 diversification is not restricted to *Arabidopsis* and it remains to be elucidated whether other regulatory mechanisms, such as SAE1 posttranslational modifications, may contribute to bringing functional diversity to other systems having a single E1 isoform.

METHODS

Plant Material and Growth Conditions

The *A. thaliana* T-DNA insertion mutant *atsae1a.3* in the Col-0 ecotype (Salk_060834) was identified at the TAIR database and obtained from the Arabidopsis Biological Resource Center (ABRC) (Alonso et al., 2003) as segregating T3 seeds. Homozygous *atsae1a.3* plants were identified by genomic PCR using the primers P1 and P2 for SAE1a genomic DNA amplification and P1 and LBb1 for T-DNA insertion detection. For *in vitro* cultures, seeds were stratified for 3 d, plated on Murashige and Skoog salts (Murashige and Skoog, 1962) (Duchefa), pH 5.7, supplemented with 0.8% BactoAgar (Difco), and transferred to a tissue culture room in a LD photoperiod (16 h light/8 h dark) at 22°C. For dehydration stress treatments, 16-day-old plants were cut near the stem–root junction and detached rosettes placed in a flow laminar hood for 1, 2, and 4 h. After treatments, plants were immediately frozen in liquid N_2 and stored at –80°C. For heat shock stress treatments, plants were grown in liquid medium Gamborg B5 (B5 vitamins and salts Duchefa, pH 5.7, glucose 20 g L^{-1}, MES 0.5 g L^{-1}) for 10 d at constant agitation (120 rpm). Plants were transferred to a water bath at 42°C for 30 min and then transferred to 22°C growth chamber for an additional 30 min of recovery. Plant samples were collected and frozen in liquid N_2. Plant tissues were collected from plants grown in soil except for young rosette and root tissues that were collected from 11-day-old seedlings grown on half-strength Murashige and Skoog salts supplemented with 1.5% BactoAgar.

Protein Extraction and Immunoblot

Anti-SUMO1/2, anti-SAE2, and anti-SCE1 antisera were generated in rabbits using full-length recombinant proteins (Cocalico). Plant tissue was ground in liquid nitrogen and proteins extracted with 50 mM Tris-HCl, pH 8, 150 mM NaCl, 0.2% Triton X-100, 1 mM PMSF, 1 μg ml^{-1} pepstatin, 1 μg ml^{-1} leupeptin, 2 mM *N*-ethylmaleimide, 10 mM iodoacetamide, 5 mM EDTA. 18 μg of total protein were resolved under reducing conditions by using SDS polyacrylamide gels and NuPage Novex 4–12% Bis/Tris Gels (Invitrogen). Proteins were transferred onto polyvinylidene difluoride (PVDF) membranes (Millipore), incubated with primary antibody overnight and secondary antibody, peroxidase-conjugated anti-rabbit (GE Healthcare), for 1 h at room temperature in TBST (20 mM Tris-HCl, pH 7.6, 20 mM NaCl, 0.1% (v/v) Tween20) supplemented with 3% non-fat dry milk. Peroxidase activity was developed in ECL Plus reagent (GE Healthcare) and chemiluminescence signal captured with the LAS-3000 imaging system (Fujifilm).

RNA Extraction and Quantitative Real-Time RT–PCR

Total RNA was extracted using the RNeasy Plant Mini kit (Qiagen) according to the manufacturer's recommendations. In the case of siliques, RNA was first extracted with TRIzol reagent (Invitrogen) and then cleaned up with a RNeasy spin column (Qiagen). Total RNA was treated with RNase-free DNase I (Promega) prior to retrotranscription. 1 μg of purified total RNA was retrotranscribed using oligodT as a primer with a Transcriptor High Fidelity cDNA Synthesis kit (Roche). Real-Time qPCR was carried out using Lightcycler® 480 SYBR Green I Master (Roche) in a Lightcycler® 480 (Roche) detection system following manufacturer's recommended amplification conditions. Reaction products were confirmed by melting curve analysis and by DNA agarose gel electrophoresis. PR65 (At1g13320) gene was chosen as a reference gene (Czechowski et al., 2005) to normalize the data from the different samples.

Transient Expression of Fluorescent Protein Fusions in Onion Cells

SAE2, *SAE2ΔCt SAE1a*, and *SAE1b* were fused in frame to the 5′ end of the coding sequences of yellow fluorescent protein (YFP) or cyan fluorescent protein (CFP) downstream of the 35S constitutive promoter. Onion epidermal cells were bombarded with 5 μg of each DNA construct using a helium biolistic gun (BIO-RAD). Treated epidermal cells were kept in the dark at room temperature for 16 h before analysis by confocal microscopy (Confocal Olympus FV 1000). YFP was excited with a 515-nm argon laser and images collected with a 550–630-nm range. CFP was excited with a 405-nm argon laser and images collected in the 460–500-nm range. Imaging of YFP and CFP imaging and transmissible light images collection were performed sequentially. Samples were scanned with the Z-stack mode and image stacks projection was calculated with ImageJ software (Rasband, 1997–2009).

In Vitro SUMO Conjugation and E1-Thioester Assays

Recombinant proteins were purified as previously described (Castaño-Miquel et al., 2011). In conjugation assays, we used the C-terminal tail of the *Arabidopsis* catalase 3 (419–472) fused to GST, GST:AtCAT3Ct. Reactions were carried out at the indicated temperatures in 25-µl reaction mixtures containing 1 mM ATP, 50 mM NaCl, 20 mM Hepes, pH 7.5, 0.1% Tween 20, 5 mM $MgCl_2$, 0.1 mM DTT, 2 µM SUMO, 0.5 µM AtSAE2/AtSAE1a, 0.5 µM AtSCE1, and 5 µM GST–AtCAT3Ct. After the specified incubation time, reactions were stopped by the addition of protein-loading buffer, incubated at 70°C for 10 min, and 10 µl aliquots were resolved by SDS–PAGE. Reaction products were detected by immunoblot analysis with anti-GST polyclonal antibodies (SIGMA, G7781). E1-thioester assays were performed at the specified temperatures in 25 µl reaction mixtures containing 1 mM ATP, 50 mM NaCl, 20 mM Hepes, pH 7.5, 0.1% Tween 20, 5 mM $MgCl_2$, 0.1 mM DTT, 2 µM SUMO, and 1 µM E1a or E1b. After 2 min, 15 µl aliquots were removed and analyzed by SDS–PAGE followed by Coomasie Fluor Orange staining according to the manufacturer's indications (Molecular Probes C-33250). As a thioester bond formation control, an aliquot of each reaction was treated with 100 mM DTT previously to loading into polyacrylamide gel. For SUMO conjugation efficiency quantifications, time-course reactions incubated at the same temperature and containing SAE1a or SAE1b isoforms were resolved in the same protein gel. After blotting to a PVDF (Millipore) membrane and incubated with anti-GST polyclonal antibodies (Sigma, G7781), luminescence signal generated by ECL Prime assay (GE Healthcare) was captured with a CCD camera (LAS4000, Fujifilm) and quantified with Multigauge software (Fujifilm). Each data point was normalized to the average of all data points obtained from each analyzed membrane in order to remove variability resulting from antibodies incubations and time exposure differences. The normalized values were used to calculate the corresponding slopes (relative luminescence signal versus time). The average slope from at least three independent experiments is shown in Figure 4. Relative SAE2~SUMO thioester levels indicate the amount of SAE2 in complex with SUMO relative to total SAE2 present in each data point. The average of three independent experiments is shown in Figure 4.

Structure Modeling, Sequence Identification, and Phylogenetic Analysis

Protein structure models were generated by using the SWISS-MODEL workspace (Arnold et al., 2006) on automated mode and specific structure template selection. SAE1a/b and SAE2 structures were predicted by using human E1 1Y8Q (2.25 Å) as a template. As a result of its low conservation degree, SAE2 Cys domain had to be modeled using the high-resolution human SAE2 Cys domain structure 2PX9 (1 Å). Models were assembled and images generated using PyMOL (DeLano, 2002).

Search for SAE1 and SAE2 sequences was performed throughout the whole genome of representative plant species using different BLAST-based programs (Altschul et al., 1997) in selected databases, including PLAZA 2.5 (http://bioinformatics.psb.ugent.be/plaza/) (Proost et al., 2009), phytozome v8.0 (www.phytozome.net/) (Goodstein et al., 2012), Cyanidioschyzon merolae Genome Project (http://merolae.biol.s.u-tokyo.ac.jp/), and NCBI (www.ncbi.nlm.nih.gov/). All hits were combined and redundant, unfinished, or non-full-length sequences were discarded. To confirm our data set of sequences as SAE1 and SAE2 sequences, we examined their fit to SAE1 and SAE2 key functional regions (Lois and Lima, 2005). Exon/intron location, distribution, and phases at the genomic sequences encoding for SAE1 and SAE2 sequences were examined through comparisons with the predicted encoded protein using GENEWISE (Birney et al., 2004).

Phylogenetic analyses were performed on the basis of multiple alignments of amino acid sequences obtained using MUSCLE (Edgar, 2004) and alignments edited with GeneDoc software (Nicholas and Nicholas, 1997). For SAE1 phylogenetic reconstruction, maximum-likelihood and Bayesian analyses were carried out using the JTT protein evolution model (Jones et al., 1992), heterogeneity of amino acid substitution rates corrected using a γ-distribution (G) with eight categories plus the proportion of invariant sites estimated by the data (I), selected by ProtTest v2.4 as the best-fitting amino acid substitution model according to the Akaike information criterion (Abascal et al., 2005). Similarly, maximum-likelihood and Bayesian analyses were run on SAE2 sequences using the JTT model plus G with eight categories. Bayesian analysis was implemented in MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001). Searches were run with four Markov chains for one million generations and sampling every 100th tree. After the stationary phase was reached (determined by the average standard deviation of split sequences approaching 0, which reflects the fact that independent tree samples became increasingly similar), the first 2500 trees were discarded as burn-in and a consensus tree was then constructed to evaluate Bayesian posterior probabilities on clades. Maximum-likelihood trees were constructed using PhyML v3.0 (Guindon and Gascuel, 2003; Guindon et al., 2010). Tree topology searching was optimized using the subtree pruning and regrafting option. The statistical support of the retrieved topology was assessed using the Shimodaira-Hasegawa-like approximate likelihood ratio test (Anisimova and Gascuel, 2006). Neighbor-joining phylogenetic analyses were conducted in MEGA 5.0 (Tamura et al., 2007). The evolutionary distances for neighbor-joining phylogenetic reconstruction were computed using the Poisson correction method. To obtain statistical support on the resulting clades, a bootstrap analysis with 1000 replicates was performed. Resulting trees were represented and edited using FigTree v1.3.1.

Accession Numbers

Assigned accession numbers for the studied genes are as follows: At5g55160 (SUMO2), At2g21470 (SAE2), At4g24940 (SAE1a), At5g50580 (SAE1b), At3g57870 (SCE1).

SUPPLEMENTARY DATA

Supplementary Data are available at *Molecular Plant Online.*

FUNDING

This work was supported by the European Research Council (grant ERC-2007-StG-205927) and the Spanish Ministry of Education and Science (grant BIO2008-01495 and CONSOLIDER CSD 2007-00036). L.C.M., S.M., I.T., and F.A. were supported by research contracts trough the CRAG. J.S. was supported by a pre-doctoral fellowship from the Ministry of Education and Science (BES-2005-6843). We also thank the Departament d'Innovació, Universitats i Empresa from the Generalitat de Catalunya (Xarxa de Referència en Biotecnologia and 2009SGR 09626) for substantial support.

ACKNOWLEDGMENTS

The technical support from members of the Greenhouse, Microscopy, and Genomics facilities at CRAG is greatly appreciated. We greatly thank Cristina Cañadas for technical support at LML lab and Dr Javier Forment Millet for bioinformatics technical support at IBMCP. No conflict of interest declared.

REFERENCES

Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. Bioinformatics. **21**, 2104–2105.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science. **301**, 653–657.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389–3402.

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. **55**, 539–552.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. **22**, 195–201.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res. **14**, 988–995.

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

Bossis, G., and Melchior, F. (2006). Regulation of SUMOylation by reversible oxidation of SUMO conjugating enzymes. Mol. Cell. **21**, 349–357.

Castaño-Miquel, L., Seguí, J., and Lois, L.M. (2011). Distinctive properties of *Arabidopsis* SUMO paralogues support the *in vivo* predominant role of AtSUMO1/2 isoforms. Biochem. J. **436**, 581–590.

Catala, R., Ouyang, J., Abreu, I.A., Hu, Y., Seo, H., Zhang, X., and Chua, N.H. (2007). The *Arabidopsis* E3 SUMO ligase SIZ1 regulates plant growth and drought responses. Plant Cell. **19**, 2952–2966.

Chaikam, V., and Karlson, D.T. (2010). Response and transcriptional regulation of rice SUMOylation system during development and stress conditions. BMB Rep. **43**, 103–109.

Chen, C.-C., Chen, Y.-Y., Tang, I.C., Liang, H.-M., Lai, C.-C., Chiou, J.-M., and Yeh, K.-C. (2011). *Arabidopsis* SUMO E3 ligase SIZ1 is involved in excess copper tolerance. Plant Physiol.

Couvreur, T.L., Franzke, A., Al-Shehbaz, I.A., Bakker, F.T., Koch, M.A., and Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). Mol. Biol. Evol. **27**, 55–71.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiol. **139**, 5–17.

DeLano, W.L. (2002). The PyMOL molecular graphics system. DeLano Scientific, San Carlos, CA, USA, available online at www.pymol.org.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**, 1792–1797.

Freire, E. (2001). The thermodynamic linkage between protein structure, stability, and function. Methods Mol. Biol. **168**, 37–68.

Gareau, J.R., and Lima, C.D. (2010). The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. Nat. Rev. Mol. Cell Biol. **11**, 861–871.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., et al. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. **40**, D1178–D1186.

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

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59**, 307–321.

Hermkes, R., Fu, Y.F., Nurrenberg, K., Budhiraja, R., Schmelzer, E., Elrouby, N., Dohmen, R.J., Bachmair, A., and Coupland, G. (2011). Distinct roles for *Arabidopsis* SUMO protease ESD4 and its closest homolog ELS1. Planta. **233**, 63–73.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. **17**, 754–755.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8**, 275–282.

Kim, J.G., Taylor, K.W., Hotson, A., Keegan, M., Schmelz, E.A., and Mudgett, M.B. (2008). XopD SUMO protease affects host transcription, promotes pathogen growth, and delays symptom development in Xanthomonas-infected tomato leaves. Plant Cell. **20**, 1915–1929.

Kurepa, J., Walker, J.M., Smalle, J., Gosink, M.M., Davis, S.J., Durham, T.L., Sung, D.Y., and Vierstra, R.D. (2003). The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis: accumulation of SUMO1 and –2 conjugates is increased by stress. J. Biol. Chem. **278**, 6862–6872.

Lee, I., and Schindelin, H. (2008). Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. Cell. **134**, 268–278.

Lois, L.M. (2010). Diversity of the SUMOylation machinery in plants. Biochem. Soc. Trans. **38**, 60–64.

Lois, L.M., and Lima, C.D. (2005). Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. EMBO J. **24**, 439–451.

Long, Y., Zhao, L., Niu, B., Su, J., Wu, H., Chen, Y., Zhang, Q., Guo, J., Zhuang, C., Mei, M., et al. (2008). Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. Proc. Natl Acad. Sci. U S A. **105**, 18871–18876.

Miura, K., and Hasegawa, P.M. (2010). Sumoylation and other ubiquitin-like post-translational modifications in plants. Trends Cell Biol. **20**, 223–232.

Miura, K., Rus, A., Sharkhuu, A., Yokoi, S., Karthikeyan, A.S., Raghothama, K.G., Baek, D., Koo, Y.D., Jin, J.B., Bressan, R.A., et al. (2005). The *Arabidopsis* SUMO E3 ligase SIZ1 controls phosphate deficiency responses. Proc. Natl Acad. Sci. U S A. **102**, 7760–7765.

Moutty, M.C., Sakin, V., and Melchior, F. (2011). Importin alpha/beta mediates nuclear import of individual SUMO E1 subunits and of the holo-enzyme. Mol. Biol. Cell. **22**, 652–660.

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture. Physiol. Plant. **15**, 473–497.

Murtas, G., Reeves, P.H., Fu, Y.F., Bancroft, I., Dean, C., and Coupland, G. (2003). A nuclear protease required for flowering-time regulation in *Arabidopsis* reduces the abundance of SMALL UBIQUITIN-RELATED MODIFIER conjugates. Plant Cell. **15**, 2308–2319.

Nicholas, K.B., and Nicholas, H.B., Jr. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments. Available online at www.psc.edu/biomed/genedoc.

Novatchkova, M., Tomanov, K., Hofmann, K., Stuible, H.-P., and Bachmair, A. (2012). Update on sumoylation: defining core components of the plant SUMO conjugation system by phylogenetic comparison. New Phytol. **195**, 23–31.

Olsen, S.K., Capili, A.D., Lu, X., Tan, D.S., and Lima, C.D. (2010). Active site remodelling accompanies thioester bond formation in the SUMO E1. Nature. **463**, 906–912.

Park, B.S., Song, J.T., and Seo, H.S. (2011). *Arabidopsis* nitrate reductase activity is stimulated by the E3 SUMO ligase AtSIZ1. Nature Commun. **2**, 400.

Park, H.C., Kim, H.U.N., Koo, S.C., Park, H.J., Cheong, M.S., Hong, H., Baek, D., Chung, W.S., Kim, D.H., Bressan, R.A., et al. (2010). Functional characterization of the SIZ/PIAS-type SUMO E3 ligases, OsSIZ1 and OsSIZ2 in rice. Plant, Cell Environ. **33**, 1923–1934.

Proost, S., Van Bel, M., Sterck, L., Billiau, K., Van Parys, T., Van de Peer, Y., and Vandepoele, K. (2009). PLAZA: a comparative genomics resource to study gene and genome evolution in plants. Plant Cell. **21**, 3718–3731.

Rasband, W.S. (1997–2009). ImageJ. US National Institutes of Health, Bethesda, MD, USA, available online at http://rsb.info.nih.gov/ij/.

Saracco, S.A., Miller, M.J., Kurepa, J., and Vierstra, R.D. (2007). Genetic analysis of SUMOylation in *Arabidopsis*: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. Plant Physiol. **145**, 119–134.

Schulman, B.A. (2011). Twists and turns in ubiquitin-like protein conjugation cascades. Protein Sci. **20**, 1941–1954.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol.

Thangasamy, S., Guo, C.L., Chuang, M.H., Lai, M.H., Chen, J.C., and Jauh, G.Y. (2011). Rice SIZ1, a SUMO E3 ligase, controls spikelet fertility through regulation of anther dehiscence. New Phytol. **189**, 869–882.

Truong, K., Lee, T., and Chen, Y. (2012a). SUMO modification of the E1 Cys domain inhibits its enzymatic activity. J. Biol. Chem.

Truong, K., Lee, T.D., Li, B., and Chen, Y. (2012b). Sumoylation of SAE2 C terminus regulates SAE nuclear localization. J. Biol. Chem. **287**, 42611–42619.

Wang, H.D., Makeen, K., Yan, Y., Cao, Y., Sun, S.B., and Xu, G.H. (2011). OsSIZ1 regulates the vegetative growth and reproductive development in rice. Plant Mol. Biol. Rep. **29**, 411–417.

Wang, J., Hu, W., Cai, S., Lee, B., Song, J., and Chen, Y. (2007). The intrinsic affinity between E2 and the Cys domain of E1 in ubiquitin-like modifications. Mol. Cell. **27**, 228–237.

Wang, J., Lee, B., Cai, S., Fukui, L., Hu, W., and Chen, Y. (2009). Conformational transition associated with E1-E2 interaction in small ubiquitin-like modifications. J. Biol. Chem. **284**, 20340–20348.

Wang, J., Taherbhoy, A.M., Hunt, H.W., Seyedin, S.N., Miller, D.W., Miller, D.J., Huang, D.T., and Schulman, B.A. (2010). Crystal structure of UBA2ufd–Ubc9: insights into E1–E2 interactions in SUMO pathways. PloS One. **5**, e15805.

Wilkinson, K.A., and Henley, J.M. (2010). Mechanisms, regulation and consequences of protein SUMOylation. Biochem. J. **428**, 133–145.