Sumoylation in Arabidopsis: stress response, signaling & evolution

Hammoudi, V.

Citation for published version (APA):
Genetic characterization of T-DNA insertions in the genome of the *sumo1/2* knock-down line

Valentin Hammoudi, Georgios Vlachakis, Ronnie de Jonge¹, Timo M. Breit² and Harrold A. van den Burg

¹: VIB Department of Plant Systems Biology, Ghent University, Belgium
²: RNA Biology and Applied Bioinformatics, University of Amsterdam, The Netherlands

ABSTRACT

Sumoylation is a post-translation modification that affects a plethora of processes in the model plant *Arabidopsis thaliana*. It consists in the conjugation of SUMO proteins onto lysines of protein targets. The Arabidopsis genome contains eight SUMO genes. *SUMO1* and *SUMO2* code for the main substrates used for sumoylation. Their essential role is highlighted by the embryo-lethality of the *sumo1 sumo2* double knockout mutant. Consequently, a line where *SUMO1* is knocked out and *SUMO2* is knocked down (*sumo1/2KD*) via the insertion of an artificial microRNA construct (*amiR-SUMO2*) was created that constitutes a valuable tool to investigate the role of SUMO *in planta* for genetic approaches. However, the number of *amiR-SUMO2* construct insertions and their localization in the genome remained undetermined. Thus, the genotyping of the *sumo1/2KD* mutant partially relies on the segregation of the kanamycin resistance trait linked with the *amiR-SUMO2* construct, which makes the genetic works tedious. Here, we sequenced the genome of *sumo1/2KD* with Ion Torrent Technology and used a rapid method to identify both the number of insertions and their localization. We identified two insertions, whose presences were eventually confirmed by PCR, making genotyping possible using PCRs. These results facilitate genetic studies with the *sumo1/2KD* line.
SUMOylation is a post-translational modification resulting in the conjugation of SUMO (Small Ubiquitin-like Modifier) proteins onto targets through the side chain of lysine residues. SUMO is encoded by a single copy gene in many eukaryotes like budding yeast (*Saccharomyces cerevisiae*), *Caenorhabditis elegans* and fruit fly (*Drosophila melanogaster*) (Floaho & Melchior, 2013). In contrast, the genome of Arabidopsis (*Arabidopsis thaliana*) contains eight SUMO genes. SUMO1 and SUMO2 are the main isoforms used for sumoylation (Kurepa *et al.*, 2003; Lois *et al.*, 2003; Saracco *et al.*, 2007; Hammoudi *et al.*, 2016). They act redundantly and are essential in Arabidopsis, as both the sumo1 and sumo2 single mutants do not display any aberrant development phenotype, while the corresponding double mutant is embryo-lethal (Saracco *et al.*, 2007). To understand the function of sumoylation in planta, we created transgenic lines (hereafter called sumo1/2KD) where SUMO1 is knocked out (KO) and SUMO2 knocked down (KD; van den Burg *et al.*, 2010). These lines were obtained by crossing sumo1-1 with the SUMO2 KD line B, a line silenced for SUMO2 using an artificial micro-RNA (amiR) targeting SUMO2 transcripts (amiR-SUMO2); this amiR-SUMO2 was engineered according to the instructions of WMD MicroRNA Designer (http://wmd3.weigelworld.org/cgi-bin/webapp.cgi; Schwab *et al.*, 2006). The sumo1/2KD mutant displays a strong phenotype characterized by enhanced accumulation of salicylic acid (SA), accumulation of the Pathogenesis-Related proteins 1 and 2 (PR1/2), spontaneous cell death in leaves, early flowering, partial sterility and dwarf plants (Suppl. Fig. 1) (van den Burg *et al.*, 2010). Apparently, the low levels of SUMO2 protein that remain in the cells are sufficient to maintain plant viability, while the SUMO2 conjugation levels are strongly suppressed.

As the insertion site of the amiR-SUMO2 is unknown, genotyping of the SUMO2KD allele is consequently based on the assessment of presence of the amiR-SUMO2 by PCR and on the segregation of seedlings for the kanamycin-resistance trait (the plant selection marker that was co-integrated with the amiR-SUMO2 construct). Genetics with sumo1/2KD is therefore tedious: homozygous lines can only be found by examining the segregation for kanamycin resistance in the next generation.

While out-crossing the sumo1/2KD line B21 to different mutant backgrounds, we noted that the dwarf phenotype segregates in the resulting F2 generation, although these plants were genotyped as homozygous for the sumo1-1 and amiR-SUMO2 alleles. As stable transformation of Arabidopsis can result in multiple T-DNA insertions (Castle *et al.*, 1993), we reasoned that the sumo1/2KD line B21 contained more than one amiR-SUMO2 integration site. Variation in the number of insertions might lead to different SUMO2 silencing levels and
explain the heterogeneous phenotype. The identification of the insertions is crucial for reverse genetics with this line. To facilitate genetics with this genetic tool for SUMO research in Arabidopsis, we here report on the amiR-SUMO2 insertion locations in its genome.

Traditionally, Southern-blotting is used to reveal the number of T-DNA insertions, while TAIL-PCR (Thermal asymmetric interlaced-PCR) is used to identify the integration sites (Liu et al., 1995). However, TAIL-PCR does not guarantee identification of all integration sites. We here used a rapid method to identify both the number of insertions and their localization based on whole genome re-sequencing of Arabidopsis using Ion Torrent technology. We mapped the reads generated and identified two insertion events. We eventually confirmed the presence of these two insertion sites by PCR.

MATERIALS AND METHODS

Genomic DNA extraction, re-sequencing and short read mapping
We isolated gDNA from pools of seedlings of sumo1/2<sup>KD</sup> plants (progeny of sumo1/2<sup>KD</sup> line B21) using the Nucleospin II plant kit (Macherey-Nagel). The gDNA isolation yielded to 38.8 ng/uL, with A260/280 ratios of 1.87, and A260/230 ratios of 2.47. The gDNA was sequenced according to the manufacturer instructions for the Ion Torrent (ThermoFischer). The obtained short reads were then mapped in the CLC workbench v6.5 software to both Arabidopsis genomic sequence (TAIR10) and the amiR-SUM2 plasmid using the strategy outlined in Fig. 2A. The parameters used for mapping were: miss match cost = 2; insertion cost= 3; deletion cost= 3.

Primer design and PCR genotyping
PCR genotyping was carried out on sumo1-1, SUMO2<sup>KD</sup> line B (i.e. parental lines), sumo1/2<sup>KD</sup> line B21, and sumo1/2<sup>KD</sup> line B22; B21 and B22 are two sister F2 lines obtained from the same cross between sumo1-1 and SUMO2<sup>KD</sup>. SUMO1 genotyping was done with the primers 3039 (5’ TCTGCAAACCAGGAGGAAG 3’) and 6541 (5’ TAGGATCCGATACAAACGAACAA) for the wild type SUMO1 allele, 6541 (5’ TAGGATCCGATACAAACGAACAA 3’) and 3249 (5’ TGGTTCACGATAGGGCCATCG 3’) for sumo1-1. ProCYP98A3 genotyping was done with the primers 5733 (5’ CAGCAGACGAAACCAACAACACT 3’) and 5578 (5’ CACCGCTATTAGAAACCACGAC 3’) for ProCYP98A3 wild type, 5733 and 4714 (5’ CATTAATGAATCGGCCAACGCGCG 3’) for amiR-SUMO2 in ProCYP98A3; PFK7 genotyping was done with primers 4904 (5’ AGTTTCTTGGGGCCCTAAGGATACA 3’) and 4980 (5’ AGTGTGAAACCAACATATACAAGAAC 3’) for PFK7 wild type, 4904 and 4719 (5’ TCGCCTTCTTGGACGAGTCTTCTGAA 3’) for amiR-SUMO2 in PFK7. The fragments were amplified using a touch-down PCR (35 cycles): (i) a melting temperature of 95°C for
30s, (ii) an annealing temp of 60°C with −1°C each cycle for 10 cycles and then 25 additional cycles at 50°C, (iii) an elongation time of 1m 15s at 72°C, and back to (i).

RESULTS AND DISCUSSION

The sumo1/2<sup>KD</sup> (line B21) carries two amiR-SUMO2 insertions. After extraction of the gDNA from the sumo1/2<sup>KD</sup> line B21, samples were sequenced using Ion Torrent technology. We obtained 36,576,076 reads of 177 bp median length. In order to localize the insertion sites of the amiR-SUMO2 T-DNA, we identified the reads that partially (i.e. 30 %) mapped to the Arabidopsis genome and partially to the amiR-SUMO2 construct (Fig. 1). Briefly, we first removed the reads that align either for >98% with the Arabidopsis genome or with the amiR-SUMO2 construct. From the remaining reads, we then selected the reads that partially map onto the amiR-SUMO2 construct (min. 30%). These reads were then mapped onto the Arabidopsis genome allowing a min. 30% match. With this method 1,012 reads remained in the final bin (Fig. 1A). Some of these reads partially mapped onto the amiR-SUMO2 construct at the Left (LB) or Right Border (RB). Sequences corresponding to the non-mapping part of these 1,012 reads were then blasted in NCBI in order to identify their location in the Arabidopsis genome. Reads that mapped onto the LB of the amiR-SUMO2 construct led to the identification of the 5’UTR of PFK7, a gene encoding for PHOSPHOFRUCTOKINASE 7 that is located on chromosome 5 (Fig. 1B). Reads that mapped onto the RB of the amiR-SUMO2 construct led to the identification of the promoter region of CYP98A3 (proCYP98A3) located on chromosome 2 (Fig. 1C). Our sumo1/2<sup>KD</sup> (line B) contains, therefore, two different insertions located on two different chromosomes. The given insertion sites correspond to the last nucleotide of the selected reads, which does not map onto the silencing construct (black arrows, Fig. 1B and 1C). We concluded that two amiR-SUMO2 integrations are present: one at 587 bp upstream of the start codon of CYP98A3 and one at 2531 bp downstream of the start codon of PFK7.

Figure 1. Identification of the two amiR-SUMO2 insertions.
A. Pipeline used to identify the T-DNA insertion sites. (Step 1) Reads were mapped onto the Arabidopsis genome assembly (TAIR10), with 98% similarity and a length cut-off of 100%. (Step 2) To remove the reads that fully matched to the amiR-SUMO2 construct, the unmapped reads were then mapped to the amiR-SUMO2 construct, with 98% similarity and a length cut-off of 98%. (Step 3) We then selected in the remaining unmapped reads for reads that partially mapped to the Arabidopsis genome, using 98% similarity and a length cut-off of 30%. (Step 4) These selected reads (from Step 3) were then mapped to the amiR-SUMO2 construct, with 98% similarity and a length cut-off of 30% to obtain the read that cover individual integration sites. We found 1,012 reads that mapped to two insertions sites. The insertions were identified by blast searches with these latter reads. Visualization of the mapped reads of Step 3 (A) on the SUMO2 silencing construct sequence at the Left Border (B) and Right Border (C). Black arrows represent breaking points of the SUMO2 silencing construct sequence at Left and Right Borders.
Genetic characterization of the *sumo1/2*<sup>KO</sup> line

**A**

36,576,076 row reads

- mapping onto TAIR10 Arabidopsis genome
  - Similarity 98%
  - Length cut off 100%

22,904,382 mapped reads

13,671,694 unmapped reads

- mapping onto silencing construct sequence
  - Similarity 98%
  - Length cut-off 98%

630 mapped reads

13,671,064 unmapped reads

- mapping onto TAIR10 Arabidopsis genome
  - Similarity 98%
  - Length cut-off 30%

12,113,453 mapped reads

1,557,611 unmapped reads

- mapping onto silencing construct sequence
  - Similarity 98%
  - Length cut-off 30%

1,012 mapped reads

12,112,437 unmapped reads

**B**

Left Border

- *amir-SUMO2* sequence consensus

<table>
<thead>
<tr>
<th>Blasted sequence</th>
<th>Mapped reads of Step 4</th>
</tr>
</thead>
</table>

**C**

Right Border

- *amir-SUMO2* sequence consensus

<table>
<thead>
<tr>
<th>Blasted sequence</th>
<th>Mapped reads of Step 4</th>
</tr>
</thead>
</table>
Figure 2. The SUMO2 silencing construct is homozygously inserted in both PFK7 (AT5G56630) and proCYP98A3 (AT2G40890) in the sumo1/2<sup>KD</sup> line.

The short sequencing reads were mapped onto the TAIR10 Arabidopsis genome assembly, with a similarity match of 98% and length cut-off of 100%. Visualization of the reads on (A) PFK7 genomic and (B) CYP98A3 promoter (proCYP98A3) sequences shows the gap in read coverage in the two identified insertion sites (black arrows).

The *amiR-SUMO2* insertions in *PFK7* and in *ProCYP98A3* are both homozygous

The 36,576,076 reads were then mapped onto the TAIR10 Arabidopsis genome assembly with a similarity of 98% and on a length cut-off of 98%. When we graphically depicted the reads onto Arabidopsis genome in CLC workbench (Fig. 2A and 2B), we observed a read coverage gap (black arrows) in both *PFK7* and *proCYP98A3* exactly at the expected insertion locations, in contrast to surrounding areas. Both insertions were, therefore, homozygous in the *sumo1/2<sup>KD</sup>* line used (van den Burg *et al.*, 2010).

We then developed for both insertions primers for genotyping. These primer pairs amplified either (i) the region surrounding the T-DNA integration site or (ii) allowed amplification of a fragment consisting in part of the *amiR-SUMO2* construct and part of the integration site (Fig. 3A and 3B). Using the primer pairs 4904+4980, 4904+4719, 5733+5578 and 5733+4714, we could, therefore, genotype for *PFK7<sup>WT</sup>*, *PFK7<sup>amiR-SUMO2</sup>* , *proCYP98A3<sup>WT</sup>* and *proCYP98A3<sup>amiR-SUMO2</sup>* alleles by PCR. Presence of both homozygous insertions was
Genetic characterization of the \textit{sumo1/2}^{KD} line

Figure 3. PCR-based genotyping of the \textit{PFK7} and \textit{proCYP98A3} alleles in the \textit{sumo1/2}^{KD} line.

\textbf{(A, B)} To scale diagrams of \textit{PFK7} and \textit{CYP98A3} genes. The \textit{amiR-SUMO2} integration sites (arrowheads) are located in \textbf{(A)} \textit{PFK7} at +2531 bp and \textbf{(B)} in \textit{CYP98A3} –587 bp, calculated from the start codon. The exons and introns are represented with boxes and broken lines, respectively. The white boxes reflect the 5' and 3' untranslated regions, while the black boxes refer to the coding regions. The primers used for genotyping are indicated by small black arrows with their ID numbers given (not to scale). Orientation of the Left Border (LB) and Right Border (RB) are indicated on the \textit{amiR-SUMO2} constructs.

\textbf{C.} Genotyping of the \textit{SUMO2}^{KD} line B and \textit{sumo1/2}^{KD} lines B21 and B22 (see also Suppl. Fig. S1) by PCRs using the primers represented in \textbf{(A)} and \textbf{(B)}.

eventually confirmed by PCRs in \textit{sumo1/2}^{KD} line B21 and in \textit{sumo1/2}^{KD} line B22, two sister lines obtained from the same cross between \textit{sumo1-1} and \textit{SUMO2}^{KD} line B (Fig. 3C).

Surprisingly, PCRs using primers that flank the RB and the LB of respectively the \textit{PFK7}^{amiR-SUMO2} and \textit{proCYP98A3}^{amiR-SUMO2} alleles failed. This coincides with the fact that the identification of these two loci were obtained thanks to reads that mapped onto Arabidopsis genome and respectively only onto the LB and the RB of the \textit{amiR-SUMO2} construct. Intriguingly, we observed that both \textit{amiR-SUMO2} constructs seem to co-segregate, while they are located on different chromosomes. Considering all of these observations, and the fact that chromosome rearrangement can be promoted by T-DNA transformation in Arabidopsis
(Naery et al., 1998), we hypothesized that chromosome rearrangement may have occurred, linking the two amiR-SUMO2 constructs. However, this remains today puzzling and does not interfere with the PCR genotyping we set up for the sumo1/2\textsuperscript{KD} line.

**CONCLUSION**

By using a rapid method based on whole genome re-sequencing, we identified the insertion sites of T-DNA constructs. This result is important since we observed that variation of the dwarf phenotype was correlated with the number of insertions present when we out-crossed this mutant to other genetic backgrounds. Genetics with this sumo1/2\textsuperscript{KD} line must consequently take this into consideration, as crossing may result in the loss of one insertion, altering the silencing levels of SUMO2.

**ACKNOWLEDGMENTS**

We would like to thank Jeffrey Ham and the RNA Biology & Applied Bioinformatics (University of Amsterdam) for help with genotyping and sequencing, respectively. We are also thankful to Peter van Dam for the valuable inputs in the bioinformatics works. This research was (partially) funded by NWO Earth and Life Sciences (ALW), project 834.12.003.

**SUPPLEMENTAL INFORMATION**

Supplemental Figure S1. sumo1/2\textsuperscript{KD} displays a strong auto-immune phenotype (van den Burg et al., 2010). Lines B21 and B22#1 are independent crosses between sumo1-1 and SUMO2\textsuperscript{KD} line B.

**REFERENCES**


Genetic characterization of the sumo1/2KD line


