Suppression of the Arabidopsis cinnamoyl-CoA reductase 1-6 intronic T-DNA mutation by epigenetic modification

Héloïse Bastiaanse, Stephane Rombauts, Lisanne de Vries, Tim de Pooter, Mojca Strazisar, Godfrey Neutelings, Ruben Vanholme, Wout Boerjan

Short title: *trans* T-DNA suppression of the *ccr1-6* allele

Corresponding author:

Wout Boerjan

VIB Center for Plant Systems Biology

Ghent University, Department of Plant Biotechnology

Technologiepark 71

B-9052 Ghent (Belgium)

Tel.: +32 9 3313881; Fax: +32 9 3313809; E-mail: wout.boerjan@psb.vib-ugent.be

Suppression of the Arabidopsis *cinnamoyl-CoA reductase 1-6* intronic T-DNA mutation by epigenetic modification

Marlies Wouters,[1,2] Héloïse Bastiaanse,[1,2] Stéphane Rombauts,[1,2] Lisanne de Vries,[1,2] Tim De Pooter,[3,4] Mojca Strazisar,[3,4] Godfrey Neutelings,[5] Ruben Vanholme,[1,2,†] and Wout Boerjan[1,2,*,†]

[1] VIB Center for Plants Systems Biology, VIB, Ghent, Belgium

[2] Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

[3] Neuromics Support Facility, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium

[4] Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

[5] Unité de Glycobiologie Structurale et Fonctionnelle (UGSF), UMR 8576, CNRS, Université de Lille, Lille, France

*Author for communication: wout.boerjan@psb.vib-ugent.be

[†]Senior authors

M.W., R.V. and W.B. designed the study and analyzed the data; G.N. made the *ugt72e1,-e2,-e3* mutant. L.d.V. performed the *ccr1-6 x ugt72e1,-e2,-e3* cross and isolated the quadruple mutant; M.W. and H.B. extracted the genomes, M.W. and S.R. performed the ONT sequencing of the *ccr1-6* and *ugt72e1,-e2,-e3 epiccr1-6* mutants. T.D.P. and M.S. performed the ONT sequencing of the *ugt72e1,-e2,-e3* mutant. S.R. analyzed the ONT sequencing data; H.B. analyzed the DNA methylation and made the respective figures; M.W. isolated *epiccr1-6* and performed all other experiments; M.W., R.V. and W.B. wrote the manuscript.

ORCID IDs: 0000-0003-2254-0051 (M.W.), 0000-0001-7977-5420 (H.B.), 0000-0002-3985-4981 (S.R.), 0000-0002-1097-5464 (T.D.P), 0000-0001-8046-3408 (M.S.), 0000-0002-4017-6639 (G.N.),0000-0002-3245-9081 (L.d.V.), 0000-0001-5848-3138 (R.V.), 0000-0003-1495-510X (W.B.)

One sentence summary: The dwarfed phenotype of the Arabidopsis *ccr1-6* intronic T-DNA lignin mutant is restored by the introduction of additional T-DNAs sharing sequence homology.

Abstract

Arabidopsis thaliana T-DNA insertion collections are popular resources for fundamental plant research. Cinnamoyl-CoA reductase 1 (CCR1) catalyzes an essential step in the biosynthesis of the cell wall polymer lignin. Accordingly, the intronic transfer (T)-DNA insertion mutant *ccr1-6* has reduced lignin levels and shows a stunted growth phenotype. Here, we report restoration of the *ccr1-6* mutant phenotype and *CCR1* expression levels after a genetic cross with a *UDP-glucosyltransferase 72e1* (*ugt72e1*),*-e2*,*-e3* T-DNA mutant. We discovered that the phenotypic recovery was not dependent on the UGT72E family loss of function but due to an epigenetic phenomenon referred to as *trans* T-DNA suppression. Via *trans* T-DNA suppression, the gene function of an intronic T-DNA mutant is restored after introduction of an additional T-DNA sharing sequence homology, leading to heterochromatinization and splicing out of the T-DNA-containing intron. Consequently, the suppressed *ccr1-6* allele was named *epiccr1-6*. Long-read sequencing revealed that *epiccr1-6*, and not *ccr1-6*, carries dense cytosine methylation over the full-length of the T-DNA. We show that the SAIL T-DNA in the *UGT72E3* locus could trigger the *trans* T-DNA suppression of the GABI-Kat T-DNA in the *CCR1* locus. Furthermore, we scanned the literature for other potential cases of *trans* T-DNA suppression in Arabidopsis. These combined observations indicate that intronic T-DNA mutants need to be used with caution, since methylation of intronic T-DNA might derepress gene expression and can thereby confound results.

Introduction

Mutants are extensively used in fundamental plant research. In the model plant *Arabidopsis thaliana* (Arabidopsis), more than 260,000 transfer (T)-DNA insertion lines are available in the Columbia-0 accession (O'Malley et al., 2015). These lines were generated by random Agrobacterium-mediated T-DNA integration in the genome: T-DNA insertion in an exon causes interruption of the coding sequence, whereas T-DNA insertion in an intron can interfere with proper mRNA splicing. GABI-Kat (Rosso et al., 2003; Kleinboelting et al.,

2012), SALK (Alonso et al., 2003) and SAIL (Sessions et al., 2002) are the T-DNA consortia that make up the bulk of the mutant lines and differ each in the elements encoded on the T-DNA. A different source of mutants (48,000 lines) is provided by the John Innes Centre (JIC) Suppressor-mutator (SM) transposant library (Tissier et al., 1999). Initially, this library was built by T-DNA integration, after which a defective Suppressor-mutator (dSpm) element jumped out of the T-DNA sequence by means of a transposase also encoded by the T-DNA. The final JIC SM transposon lines are free of T-DNA sequences.

As for many other biological processes, unraveling the biosynthesis of lignin involved ample use of T-DNA insertion mutants (Sibout et al., 2003; Vanholme et al., 2012b; Van Acker et al., 2013; Wang et al., 2022). Lignin is a complex phenolic polymer in the plant secondary cell wall that confers both fiber strength, to allow plants to grow upright, and vessel hydrophobicity, to enable water transport. However, lignin is also the major factor determining recalcitrance to biomass processing into pulp and fermentable sugars in the biorefinery and, thus, plants are engineered to produce less lignin (Chen and Dixon, 2007; Ralph et al., 2019; Vanholme et al., 2019). Unfortunately, low-lignin plants often suffer from a yield penalty, making them less useful for industrial applications (Bonawitz et al., 2014; De Meester et al., 2018; Muro-Villanueva et al., 2019; Gallego-Giraldo et al., 2020; Panda et al., 2020). Therefore, engineering low-lignin biomass crops without yield penalty is an active field of research (Chanoca et al., 2019; De Meester et al., 2020; 2021; 2022).

Lignin is synthesized from phenylalanine through the general phenylpropanoid and monolignol-specific pathways (**Figure 1**), and is mainly made from the *p*-coumaryl, coniferyl and sinapyl alcohols (Vanholme et al., 2012a; Ralph et al., 2019; Vanholme et al., 2019). Oxidative coupling of these monolignols in the cell wall gives rise to the *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) units of the lignin polymer, respectively. In lignifying cells, also a small fraction of ferulic acid (FA) is translocated to the cell wall where it is incorporated into the lignin polymer as a FA unit (Ralph et al., 2008; Leplé et al., 2007; Van Acker et al., 2013). Nonetheless, the bulk of FA is metabolized into FA derivatives such as FA 4-*O*-glucoside and feruloyl glucose and stored in the vacuole (Bowles et al., 2006; Vanholme et al., 2012b; Dima et al., 2015; Le Roy et al., 2016).

Cinnamoyl-CoA reductase 1 (CCR1) catalyzes the first step of the monolignol-specific pathway (**Figure 1**). The Arabidopsis GABI-Kat intronic T-DNA insertion mutant *ccr1-6* (GK_622C01) shows a 52% reduction in lignin content and a plant biomass yield penalty of 34% (Van Acker et al., 2013). The stunted growth of *ccr1-6* mutants is most likely due to the lack of lignin in the vessels, because vessel-specific reintroduction of *CCR1* expression in *ccr1-6* mutants restores the plant biomass yield without reinstating the overall inflorescence stem lignin content to wild-type levels (De Meester et al., 2018). *ccr1-6* mutants also produce large amounts of FA (the hydrolysis product of CCR1's substrate feruloyl-CoA; **Figure 1**), resulting

in a relatively modest increase in FA units in lignin and a severe increase in FA sugar-derivatives in the vacuole (Vanholme et al., 2012a; De Meester et al., 2018).

The uridine diphosphate (UDP)-glucosyltransferase 72E (UGT72E) gene family, consisting of three members in Arabidopsis, is involved in the glucosylation of phenylpropanoids (Lim et al., 2001; Lim et al., 2005; Lanot et al., 2006; Speeckaert et al., 2020). More in particular, UGT72E2 has been shown to catalyze the glucosylation of FA to its respective 4-*O*-glucoside *in vitro* (Lim et al., 2001; Lim et al., 2005) and three FA-derived glucosides were identified as products of UGT72Es *in vivo* (Wu et al., 2022) (**Figure 1**). In the present study, we attempted to overcome the biomass yield penalty of *ccr1-6* by generating a quadruple mutant from a cross between *ccr1-6* and the *ugt72e1* (GK_340H02)*,-e2* (SM_3_20654)*,-e3* (SAIL_1279_D02) T-DNA mutant. According to our working hypothesis, blocking 4-*O*-glucosylation of FA by knocking out the *UGT72E* gene family would prevent the storage of those FA derivatives in the vacuole and, in turn, increase the amount of FA that can be incorporated as FA units in the lignin polymer. As such, compensating for the deficit in traditional lignin units could potentially restore the growth of the *ccr1-6* mutants. In addition to strengthening the cell wall, the increased amount of FA units in the lignin polymer would also make lignin easier to degrade, because its incorporation introduces acetal functionalities in the polymer that are readily cleavable in mild acid conditions (Ralph et al., 2008; Van Acker et al., 2013).

Here, we report the complete restoration of the *ccr1-6* mutant phenotype after a genetic cross with the *ugt72e1,-e2,-e3* T-DNA mutant. However, the suppression was not caused by knocking out the *UGT72E* genes, but rather the result of *trans* T-DNA suppression.

Results

The *ccr1-6* stunted growth is completely restored after a genetic cross with the *ugt72e1,-e2,-e3* mutant

Homozygous *ugt72e1,-e2,-e3* and homozygous *ccr1-6* mutants were crossed to study whether blocking 4-*O*-glucosylation of FA by knocking out the *UGT72E* gene family could result in the restoration of the growth phenotype of *ccr1-6* mutants. As envisioned, the *ccr1-6* phenotype was fully recovered in the homozygous *ugt72e1,-e2,-e3 ccr1-6* mutant (**Figure 2A, Supplemental Table S1**). However, we noted a deviation from the Mendelian segregation in the F2 progeny of the original cross, as not a single plant out of the 480 F2 plants had the *ccr1-6* phenotype. Nevertheless, given that *ugt7e1,-e2* and *-e3* segregate independently, 27/256 of the F2 plants were expected to be homozygous for the *ccr1-6* mutation and to have at least one wild-type allele of each *UGT72E1, -E2* and *-E3*. To further

investigate this observation, homozygous *ccr1-6* mutants lacking the *ugt72e1,-e2* and *-e3* mutations were reisolated from the F3 progeny. Interestingly, the dwarfed phenotype of the parental *ccr1-6* was absent in the reisolated *ccr1-6* mutant homozygous for the *CCR1* T-DNA insertion (**Figure 2A, Supplemental Figure S1**). This observation showed that the growth restoration of *ccr1-6* is not due to UGT72E family loss of function, but suggested that the *ccr1-6* locus was epigenetically modified after crossing the *ugt72e1,-e2,-e3* and *ccr1-6* mutants. Therefore, we renamed the growth-restored *ugt72e1,-e2,-e3 ccr1-6* and *ccr1-6* mutants *'ugt72e1,-e2,-e3 epiccr1-6*' and '*epiccr1-6*', respectively, from here on. Subsequently, we measured the biomass yield. As compared with the wild type, the length and mass of the primary inflorescence stem, seed mass and total aerial plant biomass of *ccr1-6* were reduced by 47%, 57%, 74%, and 55%, respectively (**Supplemental Table S1**). In contrast, the biomass yield of *ugt72e1,-e2,-e3 epiccr1-6* mutants and *epiccr1-6* mutants were similar to those of wild-type plants (**Figure 2A**, **Supplemental Table S1**).

The *CCR1* expression and lignin content are partially restored in *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants

To examine whether the observed biomass yield restoration in *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants is due to relignification, we determined their lignin content. For lignin content analysis, inflorescence stem cross sections were stained with the Wiesner or Mäule reagent. Wiesner staining is used as a method to visualize coniferaldehyde residues in lignin. Consistent with previous analyses (De Meester et al., 2018), the Wiesner staining of *ccr1-6* cross sections was less intense as compared with those of the wild type (**Figure 2B**), implying that *ccr1-6* mutants have dramatically reduced lignin levels. The staining intensity of xylem of *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* sections resembled that of the wild type, whereas that of the interfascicular fibers was lighter than that of the wild type (**Figure 2B**). Cross sections of *ccr1-6* also revealed the typical irregularly shaped and collapsed vessels (Jones et al., 2001; De Meester et al., 2018), whereas those of *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants were large and open, similar as the wild type. Mäule staining visualizes S-unit rich lignin. With Mäule, interfascicular fibers stained brighter red than xylem in wild-type cross sections, whereas no red staining was observed in *ccr1-6*. The staining intensity of xylem of *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* cross sections resembled that of the wild type, but the red coloration of the interfascicular fibers appeared to be less dark than that of the wild type (**Figure 2B**). Together, these complementary staining techniques suggested that xylem lignification in *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutant inflorescence stems is similar to that in the wild type, but their lignin content is lower or their lignin composition is changed in interfascicular fibers.

To quantify lignin, a crude cell wall residue (CWR) was first prepared by removing solubles from the senesced inflorescence stem biomass (**Table 1, Supplemental Figure S2**). The wild-type plants had 79.0% CWR, while c*cr1-6* mutants had only 71.3% CWR (i.e., a relative reduction of 10%), indicating that *ccr1-6* biomass has more solubles. In contrast, the CWR of *epiccr1-6* (on average 77.0%) was higher than that of *ccr1-6* (i.e., a relative increase of 7%) and not significantly different from that of the wild type. Next, the fraction of lignin in these prepared CWRs was measured via the acetyl bromide method (**Table 1, Supplemental Figure S2**). Wild-type CWR had 14.3% lignin, whereas *ccr1-6* mutants had 9% lignin (a relative reduction of 37%). The lignin amounts of *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* (on average 10.6%) were recovered to levels higher than those of *ccr1-6* (i.e., a relative increase of 18%), but were still lower than those of the wild type (i.e., a relative decrease of 26%). Thus, although phenotypically indistinguishable, the lignin content of *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* was not completely restored to wild-type levels.

Finally, we measured *CCR1* expression levels. Correctly spliced *CCR1* transcript levels in *ccr1-6* inflorescence stems were 4% of those in the wild type (**Figure 2C**). On the contrary, correctly spliced *CCR1* transcript levels in *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants were on average 34% of the wild-type levels. Note that the *ccr1-6* mutants and the *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants were homozygous for the T-DNA insertion at the *CCR1* locus (**Supplemental Figure S1B**). We concluded that the T-DNA-containing intron can be spliced out correctly, but that correct splicing occurs less efficiently for the *ccr1-6* allele than for the *epiccr1-6* allele. Increased correctly spliced *CCR1* transcript levels, and not UGT72E family loss of function, must be the cause of the elevated lignin levels and, hence, the growth restoration observed in *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants.

Knocking out *UGT72E1,-E2* and *-E3* in *ccr1*-CRISPR is insufficient to enhance the incorporation of FA in lignin and restore growth

CCR1 transcript levels were partially restored in *ugt72e1,-e2,-e3 epiccr1-6* mutants, making them unsuitable to investigate our initial hypothesis that blocking 4-*O*-glucosylation of FA could result in higher levels of FA in lignin and as such restore the *ccr1-6* growth phenotype. Thus, to re-evaluate our hypothesis, we generated a full knockout of *ccr1* in the wild-type and *ugt72e1,-e2,-e3* mutant backgrounds through CRISPR-Cas9 gene editing (**Supplemental Figure S3**). The two independent *ccr1*-CRISPR single mutant lines were severely dwarfed at mature stage (on average 25.5 cm versus 59.0 cm for the wild type), as well as the two independent *ugt72e1,-e2,-e3 ccr1*-CRISPR mutant lines (on average 31.3 cm), although they were significantly taller (on average 5.8 cm, i.e., an increase of 23%) as compared with the *ccr1*-CRISPR mutants (**Figure 3, Supplemental Table S2**). However, the observed growth

207 restoration of *ugt72e1,-e2,-e3 ccr1*-CRISPR mutants was very small; the length and the
208 biomass of the primary inflorescence stems were, respectively, on average 47% shorter and
209 52% lighter than those of the wild type (**Supplemental Table S2**).
210 Next, we measured the lignin amount and composition in CWRs of *ccr1*-CRISPR and
211 *ugt72e1,-e2,-e3 ccr1*-CRISPR mutants. The proportion of CWR of *ccr1*-CRISPR line 1 and
212 both *ugt72e1,-e2,-e3 ccr1*-CRISPR lines was similar (on average 74%) and significantly
213 lower than that of the wild type (79%, **Supplemental Table S3**). Also the lignin amount in the
214 primary inflorescence stem of both *ccr1*-CRISPR and both *ugt72e1,-e2,-e3 ccr1*-CRISPR
215 lines was similar (on average 8%) and significantly less than that of the wild type (13%,
216 **Supplemental Table S3**). The minor growth restoration of *ugt72e1,-e2,-e3 ccr1-CRISPR*
217 mutants compared with *ccr1*-CRISPR could thus not be attributed to an increase in total
218 lignin amount.
219 Finally, we measured whether FA incorporation in lignin was increased in the *ugt72e1,-*
220 *e2,-e3 ccr1*-CRISPR mutants as compared with the *ccr1*-CRISPR mutants. The incorporation
221 of traditional monolignols and FA in lignin was quantified via the thioacidolysis reaction,
222 which releases three FA-derived units, two of which are linked via conventional β-*O*-4-
223 linkages (β-*O*-4-FA-I and β-*O*-FA-II) and the third derived from the bis-β–*O*–4-coupling of FA
224 (bis-β-*O*-4-FA) in lignin (Ralph et al., 2008). In accordance with previously published results
225 of Arabidopsis *CCR1*-deficient plants (Mir Derikvand et al., 2008; Van Acker et al., 2013; De
226 Meester et al., 2018), the released amounts of β-*O*-4-FA-I, β-*O*-4-FA-II and bis-β-*O*-4-FA
227 units increased on average about 7, 6 and 8 times, respectively, in *ccr1*-CRISPR mutants
228 compared with the wild type (**Table 2, Supplemental Figure S4, Supplemental Table S3**).
229 However, the relative abundances of each of the three FA-derived thioacidolysis units did not
230 increase in the *ugt72e1,-e2,-e3 ccr1*-CRISPR mutants compared with the *ccr1*-CRISPR
231 mutants. Altogether, these results disprove our initial hypothesis that knocking out the
232 *UGT72E* gene family in *ccr1* mutants could result in enhanced FA incorporation in the lignin
233 polymer and growth restoration.

234 **The T-DNA copies at the *CCR1* locus are hypomethylated in *ccr1-6* mutants and**
235 **hypermethylated in *ugt72e1,-e2,-e3 epiccr1-6* mutants**

236 To further investigate the causal mechanism of growth recovery in *epiccr1-6,* we examined
237 the genome and epigenome sequences of the T-DNA and dSpm element loci in the
238 *ugt72e1,-e2,-e3* and *ccr1-6* parental lines and in the resulting *ugt72e1,-e2,-e3 epiccr1-6*
239 mutant. For sequencing, we used the Oxford Nanopore Technologies (ONT) platform. First,
240 we found that all mutants carried additional unanticipated T-DNA insertions (**Supplemental**
241 **Figure S5**). Second, we learned that for SAIL and GABI-Kat, T-DNA insertions occurred not

as just one single copy, but as a concatemer of (stretches of) direct and inverse repeats of the T-DNA sequence that often also contained (part of) the vector backbone sequence (**Supplemental Figure S5**).

Because the growth restoration observed in *ugt72e1,-e2,-e3 epiccr1-6* mutants is not caused by UGT72E family loss of function (as proven by the *epiccr1-6* and *ugt72e1,-e2,-e3 ccr1*-CRISPR mutant phenotypes), we hypothesized that the epigenomic landscape of *ccr1-6* mutants was altered after the genetic cross of *ccr1-6* and *ugt72e1,-e2,-e3* mutant backgrounds. In plants, 5-methylcytosine in the CG dinucleotide context is a prominent epigenetic DNA modification. Because ONT technology directly sequences genomic DNA, i.e., without DNA amplification steps, DNA methylation is maintained and can be determined. We found that the T-DNAs integrated at the *CCR1* locus were differentially methylated in *ccr1-6* and *ugt72e1,-e2,-e3 epiccr1*-6 mutant backgrounds. Only a part near the right border of the T-DNA (a portion of the *pUC18* vector and the 35S promoter) of the *ccr1-6* allele was hypermethylated, whereas almost the entire T-DNA of the *epiccr1-6* allele was hypermethylated in the *ugt72e1,-e2,-e3 epiccr1*-6 mutant (**Figure 4, A and B**). Only the sulfonamide resistance gene on the T-DNAs, despite the presence of numerous CG dinucleotides, was less methylated in the *epiccr1-6* allele as compared with the other regions of the T-DNA (**Figure 4B**). In conclusion, the differential methylation status of the *ccr1-6* and *epiccr1-6* alleles is most likely the cause of the observed enhanced splicing efficiency of *CCR1* in the *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants, and thus also the cause of the observed growth restoration. Based on the restored growth phenotype of the F3, F4 and F5 generation of *epiccr1-6,* the methylation status of the *epiccr1-6* allele seems to be preserved over these subsequent generations.

In addition, we analyzed the methylation status of the other mutated loci (i.e., T-DNAs, T-DNA flanking sequences and vector backbone-derived fragments) in the different mutant backgrounds (**Supplemental Figure S6, 7 and 8**). The GABI-Kat T-DNAs in the *UGT72E1* and AT3G51430 loci were partially methylated in the *ugt72e1,-e2,-e3* mutant (**Supplemental Figure S6, C and D**). The mannopine synthase 1' 2' promoter, *pUC18* vector portion, ampicillin resistance gene and 35S promoter of the GABI-Kat T-DNA carried dense methylation, whereas the sulfonamide resistance gene was barely methylated. The methylation degree of these regions was increased in the *ugt72e1,-e2,-e3 epiccr1-6* mutant, especially in the left T-DNA region flanking the mannopine synthase 1' 2' promoter (**Figure 4C**, **Supplemental Figure S8D**). Notably, the methylation degree of these GABI-Kat T-DNAs was higher than that of the *ccr1-6* allele, but lower than that of the *epiccr1-6* allele. In *ccr1-6* mutants, the GABI-Kat T-DNAs in AT1G77990 showed a methylation pattern similar to that of the *ccr1-6* allele (**Figure 4A**, **Supplemental Figure S7B**). The SM dSpm element in the *UGT72E2* locus was not methylated in the *ugt72e1,-e2,-e3* background

(**Supplemental Figure S46H**) and remained as such in the *ugt72e1,-e2,-e3 epiccr1-6* background (**Figure 4D**). Copies of the SAIL T-DNA in the *UGT72E3* and AT5G35840 loci in the *ugt72e1,-e2,-e3* mutant were hypermethylated, with the exception of the BASTA resistance gene (**Supplemental Figure S6, E and F**). The methylation status of the SAIL T-DNAs in *UGT72E3* in the *ugt72e1,-e2,-e3 epiccr1-6* mutant was similar as that in the *ugt72e1,-e2*,*-e3* mutant, i.e., hypermethylated (**Figure 4E**). In short, we found that the T-DNAs integrated in *CCR1, UGT72E1* and *UGT72E3,* and not the dSpm element in *UGT72E2,* were already methylated before crossing and that the methylation was preserved, or increased after crossing. We also found that methylation did not spread into the T-DNA flanking sequences. A notable exception could be the GABI-Kat T-DNA inserted in the AT1G77990 locus in the *ccr1-6* mutant background, where methylation was observed in one of the flanking sequences (**Supplemental Figure S7B**). However, because this intronic region was also partially methylated in the *ugt72e1,-e2,-e3* (**Supplemental Figure S6B**) and *ugt72e1,-e2,-e3 epiccr1-6* (**Supplemental Figure S8B**) mutants, it is doubtful that the T-DNA insertion triggered the methylation of this particular T-DNA flanking sequence. Finally, the GABI-Kat (**Figure 4, A and B**) but not the SAIL (**Figure 4E, Supplemental Figure S6E**) vector backbone-derived sequences were more heavily methylated in the *ugt72e1,-e2,-e3 epiccr1-6* background (after *trans* T-DNA suppression) compared with the *ugt72e1,-e2,-e3* and *ccr1-6* backgrounds (before *trans* T-DNA suppression).

The *ugt72e3* allele triggers *trans* T-DNA suppression of the *ccr1-6* allele

To find out which of the *ugt72e* T-DNA mutant alleles is the trigger of *trans* T-DNA suppression of the *ccr1-6* allele, we crossed each of the individual *ugt72e1*, *ugt72e2* and *ugt72e3* T-DNA insertion mutants with the *ccr1-6* mutant. All plants homozygous for the *ccr1-6* allele in the F2 progeny of the reciprocal crosses of *ugt72e1* x *ccr1-6* (7 out of 47 plants; **Supplemental Figure S9, A and B**) and *ugt72e2* x *ccr1-6* (12 out of 72 plants; **Supplemental Figure S9, C and D**) were dwarfed. In contrast, all F2 progeny plants of the reciprocal crosses of *ugt72e3* x *ccr1-6* that were homozygous for the *ccr1-6* allele had a restored growth phenotype (7 out of 47 plants; **Supplemental Figure S9, E and F**). We concluded that the *ugt72e3* allele suppressed the *ccr1-6* mutant dwarfed phenotype in both the *ccr1-6* x *ugt72e3* and *ccr1-6* x *ugt72e1,-e2,-e3* cross.

To date, only five cases have been reported in which the gene function of a T-DNA insertion mutant was restored after a genetic cross with another T-DNA mutant (**Supplemental Table S4**). The phenomenon is termed *trans* T-DNA suppression and the minimal requirements for it to occur have been described as: (1) crossing two mutants sharing homologous T-DNAs (T-DNA lines from the same mutant collection) of which (2) the

suppressed T-DNA is inserted into an intronic region (Osabe et al., 2017). The suppressed T-DNA insertion in the *epiccr1-6* allele fulfills the second requirement: *ccr1-6* has an intronic T-DNA insertion. However, the trigger T-DNA in *ugt72e3* (SAIL) and the suppressed T-DNA in *ccr1-6* (GABI-Kat) originate from a different T-DNA collection and thus conflict with the first requirement. Nonetheless, we found that among all pairwise sequence alignments of T-DNA (and dSpm element) sequences of the most commonly used T-DNA consortia (and transponant) collections, the T-DNA sequences of the SAIL and GABI-Kat collections share the most sequence homology (**Figure 5**). The T-DNA sequences of *pDAP101* (SAIL) and *pAC161* (GABI-Kat) share, in sense orientation, the mannopine synthase 1' 2' promoter (484 bp; Velten et al., 1984; Ülker et al., 2008) and 35S terminator (233 bp) sequences. In antisense orientation, they share part of the *pUC18* vector sequence encoding among others ampicillin resistance (2,161 bp), and an additional vector sequence (158 bp). It thus seems that also one or more of these shorter stretches of identical sequences can trigger *trans* T-DNA suppression.

Prevalence of *trans* T-DNA suppression

To estimate the prevalence of *trans* T-DNA suppression in Arabidopsis research, we performed a literature search for the query 'T-DNA double mutant' (**Supplemental Table S5**). We focused on articles published between 2015 and August 2022, because, previously, a similar search was done for articles published in the period from 2002 till 2015 (Jia et al., 2015). From the 111 publications that matched our query, 24 (22%) reported double or higher-order T-DNA mutants that met the requirements for *trans* T-DNA suppression. We concluded that, since the first report of suppression of an intronic T-DNA mutant (Xue et al., 2012), many researchers still made double or higher-order mutants without considering the possibility of *trans* T-DNA suppression of intronic T-DNA mutants. Only two publications reported the suppression of the mutant phenotype in the double or higher-order mutant. The first paper describes the instability of the Arabidopsis *csn5a-2* mutant and proved this was indeed a case of *trans* T-DNA suppression (Jia et al., 2015). The second paper showed that the strong reduction of shoot growth associated with the severe phosphate deficiency of the *pho1-7* T-DNA mutant was alleviated in the T-DNA *tex1 pho1-7* double mutant as a consequence of the generation of a near full-length *PHO1* mRNA by read-through transcription and splicing out of the T-DNA-containing intron (Khan et al., 2020). Based on these observations, a role for the THO/TREX complex, of which TEX1 is a component, in transcription termination was proposed. However, one can argue that the growth restoration of *pho1-7* is not caused by the THO/TREX transcription termination complex, but merely resulted from *trans* T-DNA suppression of *pho1-7* intronic T-DNA.

Notably, the relatively low prevalence of reports on *trans* T-DNA suppression might be an underestimation, because the observations that follow from this suppression phenomenon might have been considered as 'aberrant' or 'out of scope' for research teams, and might therefore not have been reported.

Furthermore, we also screened all publications reporting a (partial) restoration of mutant phenotypes in our research field, i.e., in cell-wall engineering (**Supplemental Table S6**). A previous publication from our group reported the complete growth restoration of *ccr1-6* via reintroduction of the *CCR1*-encoding sequence under control of the artificial *SECONDARY NAC BINDING ELEMENT* (*SNBE*) promoter (*proSNBE*) that solely confers expression in the vessels (De Meester et al., 2018). Both the GABI-Kat T-DNA and the *proSNBE:CCR1* T-DNA contain the CaMV 35S minimal promoter sequence (46 bp). To investigate whether the growth restoration observed after the transformation of the *ccr1-6* mutant with the *proSNBE:CCR1* construct is caused by the generation of an *epiccr1-6*, we transformed *ccr1*-CRISPR line 2 with the *proSNBE:CCR1* construct. Because the *ccr1*-CRISPR line does not contain a T-DNA, the growth restoration observed in *ccr1*-CRISPR *proSNBE*:*CCR1* cannot be caused by *trans* T-DNA suppression and must thus be solely caused by the *proSNBE*-driven *CCR1* expression. The growth phenotype and length of the primary inflorescence stem of the T3 single-locus homozygous *ccr1*-CRISPR *proSNBE:CCR1* plants were similar to those of *ccr1-6 ProSNBE:CCR1* and wild-type plants **(Supplemental Figure S10, Supplemental Table S7)**, thus, also the growth restoration of *ccr1-6 ProSNBE:CCR1* mutants is most likely not a case of *trans* T-DNA suppression. Its mass was not completely restored to wild-type levels, although this was the case for *ccr1-6 proSNBE:CCR1* mutants **(Supplemental Table S7)**. A position effect of the *proSNBE:CCR1* construct might explain this minor discrepancy. Alternatively, the residual *CCR1* expression measured in *ccr1-6* mutants might account for this difference. In the meantime, vessel-specific relignification has independently been confirmed to be a valid strategy to fully restore growth of Arabidopsis *ccr1*-CRISPR mutants (Yu et al., 2021). For none of the other 14 reports of (partial) growth restoration, we did find evidence that *trans* T-DNA suppression would have caused the observed phenotypes.

Discussion

Multiple T-DNA insertion loci, each consisting of several T-DNA copies, and a deletion as a consequence of T-DNA insertion mutagenesis

Initially, the objective of this research was to overcome lignin-modification-induced dwarfism of Arabidopsis *ccr1* mutants by reducing the glucosylation of FA in order to enhance its

incorporation in lignin. We found that knocking out *UGT72E1,-E2* and *-E3* is insufficient to complement the *ccr1* dwarfed phenotype and, instead, investigated the cause of the complete growth restoration observed in *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* T-DNA mutants.

From the whole-genome ONT sequencing performed to study the methylation pattern of the T-DNA loci, we also obtained a detailed picture of the number of T-DNA inserts and their organization. In addition to the anticipated insertions, we found additional T-DNA insertions in the *ugt72e1,-e2,-e3*, *ccr1-6* and *ugt72e1,-e2,-e3 epiccr1-6* mutants. The integration of T-DNAs at multiple sites in the genome is a well-known phenomenon. The number of loci with T-DNA insertions per line is estimated to be on average 1.3 – 1.9 for GABI-Kat (Rosso et al., 2003; Pucker et al., 2021), and 1.5 – 2.0 for SAIL (McElver et al., 2001). In line with the expectations, the dSpm transposon element appeared as a single insert in the *ugt72e2* locus only (Tissier et al., 1999). In *ugt72e1,-e2,-e3*, additional SAIL T-DNAs were inserted in AT5G35840, encoding the apoprotein of phytochrome C (phyC). PhyC plays a role in the perception of daylight; *phyC* mutants flower early when grown in short-day photoperiods (Monte et al., 2003). Early flowering was also observed for the *ugt72e3* SAIL T-DNA mutant (Baldacci-Cresp et al., 2020) and the subsequent proposed relationship between *UGT72E3* and flowering is now questionable. We also found that three protein-encoding genes neighboring the *UGT72E3* gene were deleted. Recently, a role for UGT72E3 in lignification of secondary cell walls was proposed based on effects observed in the cell wall and transcriptome of *ugt72e3* (Baldacci-Cresp et al., 2020). Knowledge about the three deleted genes neighboring *ugt72e3* is incomplete, but so far these could not be linked with lignin biosynthesis nor growth, but we cannot exclude they do so in a *ccr1* mutant background.

Moreover, we found that several T-DNA insertion sites consist of concatemers of direct and inverse repeats of the T-DNA and even contain vector backbone sequence. These data are in line with recent studies of other Arabidopsis T-DNA mutants of the SALK, SAIL, and WISC collection (Jupe et al., 2019) and the GABI-Kat collection (Pucker et al., 2021). In addition, these recent studies also report large-scale genomic rearrangements. We did not observe the latter in the *ugt72e1,-e2,-e3*, *ccr1-6* and *ugt72e1,-e2,-e3 epiccr1-6* mutants. Nonetheless, we showed that the genomic backgrounds of the T-DNA mutants used for this analysis were more complex than initially anticipated.

Differential methylation of T-DNA copies before and after crossing

Nanopore sequencing-based methylation calling is the most recently developed technique to study the methylome (Gigante et al., 2019; Gouil and Keniry, 2019). In the *ccr1-6* mutant

418 background, only the right part of every GABI-Kat T-DNA copy of the *ccr1-6* allele was
419 methylated, whereas after crossing, in the *ugt72e1,-e2,-e3 epiccr1-6* mutant background, we
420 observed dense methylation throughout the entire GABI-Kat T-DNA copies of the
421 suppressed *epiccr1-6* allele. Methylation of T-DNA insertions has been previously described.
422 One study determined the methylation pattern of the T-DNA insertions of a SALK and a SAIL
423 line via bisulfite sequencing (Jupe et al., 2019). They found an enrichment of methylation of
424 the entire T-DNA region in the SALK line and only the reporter gene – and not the resistance
425 gene – in the SAIL line. The fact that the methylation pattern of a T-DNA can change is in
426 line with the findings of another study in which the DNA methylation status of two SALK T-
427 DNA insertion loci were determined via bisulfite sequencing before and after crossing their
428 respective genotypes (Osabe et al., 2017). Large portions of the T-DNA sequences were
429 found to be already heavily methylated before crossing, and from the F1 generation of the
430 reciprocal cross onward, methylation was induced further and covered the entire T-DNA
431 sequence after crossing. Similarly to the phenotype of *ugt72e1,-e2,-e3 epiccr1-6* mutants,
432 the mutant phenotype of the hypermethylated intronic T-DNA insertion mutation was
433 alleviated after the cross (Osabe et al., 2017).

434 **Refining the existing and adding new requirements for *trans* T-DNA suppression**

435 The complete growth restoration of *ccr1* dwarfism observed in *ugt72e1,-e2,-e3 epiccr1-6* and
436 *epiccr1-6* mutants is caused by *trans* T-DNA suppression, an epigenetic phenomenon in
437 which gene function is restored after a genetic cross of T-DNA mutants (Xue et al., 2012;
438 Gao and Zhao, 2013; Sandhu et al., 2013; Jia et al., 2015; Osabe et al., 2017). The
439 molecular mechanism of *trans* T-DNA suppression is not yet fully understood, but a model for
440 the initiation and maintenance was proposed (Osabe et al., 2017). The initiation is similar to
441 the suppression of transposable elements in transcriptionally active regions, a natural plant
442 genome surveillance hack (Fultz et al., 2015; Saze, 2018). Our case of *trans* T-DNA
443 suppression of the *epiccr1-6* allele after a genetic cross of *ugt72e3* and *ccr1-6* T-DNA
444 mutants supports the proposed model for the initiation (**Figure 6**). Maintenance of *trans* T-
445 DNA suppression requires a heterochromatic state, maintained by methyltransferase 1
446 (MET1), a CG dinucleotide-specific DNA methyltransferase, and decreased DNA methylation
447 1 (DDM1), a chromatin remodeller (Osabe et al., 2017). As such, at the RNA level, both
448 intron splice sites are brought in physical proximity and splicing of the T-DNA-containing
449 intron is promoted by increased in bonsai methylation 2 (IBM2) and enhanced downy mildew
450 2 (EDM2) that recognize heterochromatin and enable splicing (Osabe et al., 2017).
451 To date, two requirements for *trans* T-DNA suppression to occur are defined (Osabe et
452 al., 2017; see above). However, our observation that the combination of T-DNAs of different

T-DNA collections also enables *trans* T-DNA suppression, instigates us to rephrase the first requirement as follows: the presence of two T-DNAs sharing stretches of identical sequence. To help researchers circumvent *trans* T-DNA suppression when creating a double or higher-order T-DNA mutant, we made the pairwise sequence alignment of all T-DNA sequences (and dSpm elements) of the most commonly used T-DNA mutant (and transposant) collections (**Figure 5**). All combinations showed some degree of sequence homology, requiring caution for *trans* T-DNA suppression any time an intronic T-DNA mutant is involved. A notable exception is the combination of GABI-Kat T-DNA with a JIC SM transposon line, as the respective sequences show basically no homology. Finally, we like to stress that not only the act of crossing T-DNA mutants but also adding multiple T-DNA sequences to the plant genome via transformation might result in *trans* T-DNA suppression of intronic T-DNA mutants. The second requirement: the suppressed T-DNA is inserted into an intronic region, remains intact.

The low prevalence of cases of *trans* T-DNA suppression in literature in combination with the absence of *trans* T-DNA suppression in the *ccr1-6 ugt72e1* mutant (where both *ccr1-6* and *ugt72e1* are GABI-Kat lines) shows that, even if the minimum requirements for *trans* T-DNA suppression are met, suppression is not necessarily triggered. Thus, in addition to the first two prerequisites, we propose two extra requirements for *trans* T-DNA suppression. Both additional requirements would be needed to trigger the initiation of T-DNA methylation, which is a key factor in T-DNA suppression. The third requirement is the presence of direct T-DNA-derived and reversed complement T-DNA-derived transcripts to enable the formation of dsRNA, implying that the T-DNA-containing genes should be simultaneously expressed in space and time. Both the T-DNA integration loci of *ugt72e3* and *ccr1-6* consist of direct and inverse repeats of their T-DNA sequences. After the genetic cross of *ugt72e3* with *ccr1-6* and *ugt72e1,-e2,-e3* with *ccr1-6*, primary transcripts (i.e., pre-mRNA) of both genes still contain the T-DNA sequences and can form dsRNA (**Figure 6**). This condition was also met in the study of Osabe et al. (2017). Unfortunately, the other reported cases of *trans* T-DNA suppression can currently not be inspected for this prerequisite because of the lack of genome sequence data. Notably, *ccr1-6* and *ccr1-6 ugt72e1* mutant backgrounds also had the ability to form T-DNA-derived dsRNA, but *trans* T-DNA suppression was not triggered. This notable exception lead us to speculate on the fourth requirement: the amount of T-DNA-derived dsRNA needs to exceed a threshold value to trigger the initiation of T-DNA methylation. Only after introducing copies of the T-DNA sequence that are also sufficiently transcribed (here: T-DNA copies in the *ugt72e3* allele) into the *ccr1-6* mutant background, the amount of T-DNA-derived dsRNA surpassed the threshold value to trigger methylation of the T-DNA in the *ccr1-6* locus.

In conclusion, we found that *trans* T-DNA suppression was at the origin of the growth recovery of the *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants. We reach out to the Arabidopsis research community to consider the possibility of *trans* T-DNA suppression when interpreting newly generated or reported data of intronic T-DNA mutants, especially when two (or more) T-DNA mutants are combined.

Materials and methods

For a detailed description, see **Supplemental Materials and Methods**. In brief, the Arabidopsis *ccr1-6* T-DNA mutant (GK_622C01) (Mir Derikvand et al., 2008) was crossed with the *ugt72e1* (GK_340H02),-*e2* (SM_3_20654),-*e3* (SAIL_1279_D02) T-DNA mutant, of which each of the individual *ugt72e* mutants was described previously (Baldacci-Cresp et al., 2020). Arabidopsis *ccr1*-CRISPR knockout mutants in wild-type and *ugt72e1,-e2,-e3* mutant background were generated with CRISPR-Cas9 gene editing technology. For cloning, we adopted the protocol of Decaestecker et al. (2019) that makes use of the GreenGate cloning toolbox (Lampropoulos et al., 2013). In the complementation study, the *ccr1*-CRISPR line 2 mutant was transformed with the *ProSNBE:CCR1* construct, in which *CCR1* is expressed under a vessel-specific promoter, previously described by De Meester et al. (2018).

For the stem phenotype, biomass measurements and analysis of the lignin content and composition, seeds were sown in soil and stratified at 4°C for 4 days prior to placing them under long-day conditions (16-h light/8-h dark photoperiods). For microscopy, seeds were sown in soil and stratified at 4°C for 4 days prior to placing them under short-day conditions (8-h light/16-h dark photoperiods). After 5 weeks in short-day conditions, plants were transferred to long-day conditions for 4 weeks and main stems were harvested.

Analysis of the lignin content was perfomed according to the acetyl bromide method as essentially described by Dence (1992). The general lignin composition and the quantification of FA-derived lignin units were determined using thioacidolysis as essentially described by Robinson and Mansfield (2009) and modified by De Meester et al. (2018). Response factors for H, G, S and FA-derived units were calculated according to Yue et al. (2012). For lignin microscopy, stem sections were stained with Wiesner and Mäule reagents (Pradhan Mitra and Loqué, 2014) and imaged with an Olympus BX51 microscope coupled to a Zeiss ICc1 camera with a Plan 10X 0.25 dry objective.

For transcript analysis, 1 rosette leaf of 3 individual 25-day-old plants were pooled. Total RNA was isolated using the ReliaPrep™ RNA Tissue Miniprep System (Promega). Total cDNA was synthesized using qScript® cDNA SuperMix (Quantabio). The relative expression level was determined with the Roche LightCycler 480 combined with the SYBR Green I Master Kit (Roche Diagnostics) in three technical repeats.

High-molecular weight genomic DNA was extracted from rosette leaves for whole-genome sequencing. From the obtained gDNA, a library was constructed using the 1D Sequencing Kit following the manufacturer's protocol (SQK-LSK109, ONT). Afterwards, the library was loaded onto a MinION Flow cell (R9.4.1, ONT) and sequenced using a MinION sequencing device (ONT). Base calling of the raw reads was done using guppy (version 5.0.7, http://www.nanoporetech.com). For each mutant, the ONT reads were assembled

using the Flye software (v2.8) (Kolmogorov et al., 2019) and each assembly was further polished using Nanopolish (Hu et al., 2021). The obtained polished genomes were then purged for haplotigs remaining in the assemblies (Roach et al., 2018). Contigs of the individual *de novo* genome assemblies were ordered and oriented with the 'Ragoo' tool (Alonge et al., 2019) using the reference Arabidopsis TAIR10 genome as a guide (Lamesch et al., 2012). Genes were predicted on individual polished genome assemblies using protein-coding genes from Araport11 (Cheng et al., 2017) and the Augustus gene prediction software (Stanke et al., 2008). Raw ONT sequencing data (FAST5 format) was deposited in the Sequence Read Archive (SRA) on NCBI under project PRJNA767312.

The methylome was obtained with Megalodon (v2.3.3, with guppy 5.0.7, https://github.com/nanoporetech/megalodon) and visualization of the methylation patterns of loci of interest was executed with the Methylartist tool (Cheetham et al., 2022). Complex T-DNA insertion loci, including gene prediction models, as well as ideograms of the individual Arabidopsis T-DNA mutant lines were plotted within RStudio using a custom script (R version 4.0.4). Some of those custom R plots were manually added to the upper track of the Methylartist plot.

For the pairwise sequence alignments of the complete T-DNA (dSpm element) sequences (**Supplemental Sequences**), the dotplot feature in Geneious Prime® 2022.0.1 (https://www.geneious.com) was used.

Statistical analyses were performed with RStudio, version 3.6.3.

Supplemental data

The following materials are available in the online version of this article.

Supplemental Materials and methods

Supplemental Figure S1: The *ccr1-6* mutants and the *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants are homozygous for the T-DNA insertion at the *CCR1* locus.

Supplemental Figure S2: Dot and box plot respresentation of Table 1.

Supplemental Figure S3: CRISPR-induced *ccr1* mutations in wild-type and *ugt72e1,-e2,-e3* mutant backgrounds.

Supplemental Figure S4: Dot and box plot representation of Table 2.

Supplemental Figure S5: Ideogram of the five chromosomes and zoom-in of the different T-DNA and vector backbone arrays integrated in different Arabidopsis mutants.

Supplemental Figure S6: Methylation profile plot of different loci in the *ugt72e1,-e2,-e3* mutant background.

Supplemental Figure S7: Methylation profile plot of different loci in the *ccr1-6* mutant background.

Supplemental Figure S8 Methylation profile plot of different loci in the *ugt72e1,-e2,-e3 epiccr1-6* mutant background.

Supplemental Figure S9: Phenotype of F2 progeny plants of the reciprocal crosses of *ccr1-6* with either of the *ugt72e1/ugt72e2/ugt72e3* mutants.

Supplemental Figure S10: Growth restoration of the *ccr1*-CRISPR mutant phenotype by stacking the *proSNBE:CCR1* construct and the *ccr1-6* mutation.

Supplemental Table S1: Biomass measurements of *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants.

Supplemental Table S2: Biomass measurements of the *ccr1*-CRISPR and *ugt72e1,-e2,-e3 ccr1*-CRISPR mutants.

Supplemental Table S3: Crude cell wall residue (CWR), lignin content and lignin composition of *ccr1*-CRISPR and *ugt72e1,-e2,-e3 ccr1*-CRISPR mutants.

Supplemental Table S4: Confirmed cases of *trans* T-DNA suppression.

Supplemental Table S5: Prevalence of *trans* T-DNA suppression in literature.

Supplemental Table S6: Prevalance of *trans* T-DNA suppression in the cell-wall engineering research field.

Supplemental Table S7: Biomass measurements of different *ccr1* mutant lines.

Supplemental Table S8: Primer sequences and primer pair-optimized annealing temperature for PCR-based genotyping of different T-DNA insertion loci.

Supplemental Table S9: gRNA target sequences.

Supplemental Table S10: Oligonucleotide sequences used for cloning, vector validation and genotyping of *ccr1*-CRISPR mutants.

Supplemental Table S11: Vector overview for Golden Gate cloning.

Supplemental Table S12: Summary of specific quantifier, qualifiers, retention time (RT) and response factor (RF) used for GC-MS-based analysis of lignin monomers released upon thioacidolysis.

Supplemental Table S13: Primer sequences for RT-qPCR.

Supplemental Sequences

Acknowledgments

The authors would like to thank the VIB Bioimaging core for assistance in microscopy, Sandrien Desmet of the VIB Metabolomics Core Ghent for the analysis of the thioacidolysis GC-MS data, and Véronique Storme and Annick Bleys (VIB-UGent, Belgium) for statistical support and critically reading the manuscript, respectively.

Funding

M.W. is supported by the Research Foundation Flanders under project number 3S019619. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-Adv-Grant agreement No. 834923).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC (2019) RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol **20:** 224

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301:** 653-657

Baldacci-Cresp F, Le Roy J, Huss B, Lion C, Créach A, Spriet C, Duponchel L, Biot C, Baucher M, Hawkins S, et al (2020) *UDP-GLYCOSYLTRANSFERASE 72E3* plays a role in lignification of secondary cell walls in Arabidopsis. Int J Mol Sci **21:** 6094

Bonawitz ND, Kim JI, Tobimatsu Y, Ciesielski PN, Anderson NA, Ximenes E, Maeda J, Ralph J, Donohoe BS, Ladisch M, et al (2014) Disruption of Mediator rescues the stunted growth of a lignin-deficient *Arabidopsis* mutant. Nature **509:** 376–380

Böhmdorfer G, Sethuraman S, Jordan Rowley M, Krzyszton M, Hafiz Rothi M, Bouzit L, Wierzbicki AT (2016) Long non-coding RNA produced by RNA polymerase V determines bounderies of heterochromatin. eLife **5**: e19092.

Bowles D, Lim E-K, Poppenberger B, Vaistij FE (2006) Glycosyltransferases of lipophilic small molecules. Annu Rev Plant Biol **57:** 567-597

Chanoca A, de Vries L, Boerjan W (2019) Lignin engineering in forest trees. Front Plant Sci **10:** 912

Cheetham SW, Kindlova M, Ewing AD (2022) Methylartist: tools for visualising modified bases from nanopore sequence data. Bioinformatics **38:** 3109-3112

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotechnol **25:** 759-761

Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. Plant J **89:** 789-804

De Meester B, de Vries L, Özparpucu M, Gierlinger N, Corneillie S, Pallidis A, Goeminne G, Morreel K, De Bruyne M, De Rycke R, et al (2018) Vessel-specific reintroduction of CINNAMOYL-COA REDUCTASE1 (CCR1) in dwarfed *ccr1* mutants restores vessel and xylary fiber integrity and increases biomass. Plant Physiol **176:** 611-633

De Meester B, Madariaga Calderón B, de Vries L, Pollier J, Goeminne G, Van Doorsselaere J, Chen M, Ralph J, Vanholme R, Boerjan W (2020) Tailoring poplar lignin without yield penalty by combining a null and haploinsufficient *CINNAMOYL-CoA REDUCTASE2* allele. Nat Commun **11:** 5020

De Meester B, Vanholme R, de Vries L, Wouters M, Van Doorsselaere J, Boerjan W (2021) Vessel-and ray-specific monolignol biosynthesis as an approach to engineer fiber-hypolignification and enhanced saccharification in poplar. Plant J **108:** 752-765

De Meester, B., Vanholme, R., Mota, T., and Boerjan, W. (2022). Lignin engineering in forest trees: from gene discovery to field trials. Plant Commun **3**:100465

Decaestecker W, Andrade Buono R, Pfeiffer ML, Vangheluwe N, Jourquin J, Karimi M, Van Isterdael G, Beeckman T, Nowack MK, Jacobs TB (2019) CRISPR-TSKO: a technique for efficient mutagenesis in specific cell types, tissues, or organs in Arabidopsis. Plant Cell **31:** 2868-2887

Dence CW (1992) The determination of lignin. *In* SY Lin, CW Dence, eds, Methods in Lignin Chemistry. Springer-Verlag, Berlin, Germany, pp 33-61

Dima O, Morreel K, Vanholme B, Kim H, Ralph J, Boerjan W (2015) Small glycosylated lignin oligomers are stored in Arabidopsis leaf vacuoles. Plant Cell **27:** 695-710

Fultz D, Choudury SG, Slotkin RK (2015) Silencing of active transposable elements in plants. Curr Opin Plant Biol **27:** 67-76

Gallego-Giraldo L, Liu C, Pose-Albacete S, Pattathil S, Peralta AG, Young J, Westpheling J, Hahn MG, Rao X, Knox JP, et al (2020) ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE 1 (ADPG1) releases latent defense signals in stems with reduced lignin content. Proc Natl Acad Sci USA **117:** 3281-3290

Gao Y, Zhao Y (2013) Epigenetic suppression of T-DNA insertion mutants in *Arabidopsis*. Mol Plant **6:** 539-545

Gigante S, Gouil Q, Lucattini A, Keniry A, Beck T, Tinning M, Gordon L, Woodruff C, Speed TP, Blewitt ME, et al (2019) Using long-read sequencing to detect imprinted DNA methylation. Nucleic Acids Res **47:** e46

Gouil Q, Keniry A (2019) Latest techniques to study DNA methylation. Essays Biochem **63:** 639-648

Hu K, Huang N, Zou Y, Liao X, Wang J (2021) MultiNanopolish: refined grouping method for reducing redundant calculations in Nanopolish. Bioinformatics **37:** 2757-2760

Jia X, Chanda B, Zhao M, Brunner AM, Beers EP (2015) Instability of the Arabidopsis mutant *csn5a-2* caused by epigenetic modification of intronic T-DNA. Plant Sci **238:** 53-63

Jones L, Ennos AR, Turner SR (2001) Cloning and characterization of *irregular xylem4* (*irx4*): a severely lignin-deficient mutant of *Arabidopsis*. Plant J **26:** 205-216

Jupe F, Rivkin AC, Michael TP, Zander M, Motley ST, Sandoval JP, Slotkin RK, Chen H, Castanon R, Nery JR, et al (2019) The complex architecture and epigenomic impact of plant T-DNA insertions. PLoS Genet **15:** e1007819

Khan GA, Deforges J, Reis RS, Hsieh Y-F, Montpetit J, Antosz W, Santuari L, Hardtke CS, Grasser KD, Poirier Y (2020) The transcription and export complex THO/TREX contributes to transcription termination in plants. PLoS Genet **16:** e1008732

Kleinboelting N, Huep G, Kloetgen A, Viehoever P, Weisshaar B (2012) GABI-Kat SimpleSearch: new features of the *Arabidopsis thaliana* T-DNA mutant database. Nucleic Acids Res **40:** D1211-D1215

Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol **37:** 540-546

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res **40:** D1202-D1210

Lampropoulos A, Sutikovic Z, Wenzl C, Maegele I, Lohmann JU, Forner J (2013) GreenGate - A novel, versatile, and efficient cloning system for plant transgenesis. PLoS ONE **8:** e83043

Lanot A, Hodge D, Jackson RG, George GL, Elias L, Lim E-K, Vaistij FE, Bowles DJ (2006) The glucosyltransferase UGT72E2 is responsible for monolignol 4-*O*-glucoside production in *Arabidopsis thaliana*. Plant J **48:** 286-295

Le Roy J, Huss B, Creach A, Hawkins S, Neutelings G (2016) Glycosylation is a major regulator of phenylpropanoid availability and biological activity in plants. Front Plant Sci **7:** 735

Leplé J-C, Dauwe R, Morreel K, Storme V, Lapierre C, Pollet B, Naumann A, Kang K-Y, Kim H, Ruel K, et al (2007) Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. Plant Cell **19:** 3669-3691

Lim E-K, Jackson RG, Bowles DJ (2005) Identification and characterisation of *Arabidopsis* glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde. FEBS Lett **579:** 2802-2806

Lim E-K, Li Y, Parr A, Jackson R, Ashford DA, Bowles DJ (2001) Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in *Arabidopsis*. J Biol Chem **276:** 4344-4349

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet **15:** 394-408

McElver J, Tzafrir I, Aux G, Rogers R, Ashby C, Smith K, Thomas C, Schetter A, Zhou Q, Cushman MA, et al (2001) Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. Genetics **159:** 1751-1763

Mir Derikvand M, Berrio Sierra J, Ruel K, Pollet B, Do C-T, Thévenin J, Buffard D, Jouanin L, Lapierre C (2008) Redirection of the phenylpropanoid pathway to feruloyl malate in *Arabidopsis* mutants deficient for cinnamoyl-CoA reductase 1. Planta **227:** 943-956

Monte E, Alonso JM, Ecker JR, Zhang Y, Li X, Young J, Austin-Phillips S, Quail PH (2003) Isolation and characterization of *phyC* mutants in Arabidopsis reveals complex crosstalk between phytochrome signaling pathways. The Plant Cell **15:** 1962-1980

Muro-Villanueva F, Mao X, Chapple C (2019) Linking phenylpropanoid metabolism, lignin deposition, and plant growth inhibition. Curr Opin Biotechnol **56:** 202-208

O'Malley RC, Barragan CC, Ecker JR (2015) A user's guide to the Arabidopsis T-DNA insertion mutant collections. Methods Mol Biol **1384:** 323-342

Osabe K, Harukawa Y, Miura S, Saze H (2017) Epigenetic regulation of intronic transgenes in *Arabidopsis*. Sci Rep **7:** 45166

Panda C, Li X, Wager A, Chen H-Y, Li X (2020) An importin-beta-like protein mediates lignin-modification-induced dwarfism in Arabidopsis. Plant J **102:** 1281-1293

Pradhan Mitra P, Loqué D (2014) Histochemical staining of *Arabidopsis thaliana* secondary cell wall elements. J Vis Exp **87:** e51381

Pucker B, Kleinbölting N, Weisshaar B (2021) Large scale genomic rearrangements in selected *Arabidopsis thaliana* T-DNA lines are caused by T-DNA insertion mutagenesis. BMC Genomics **22:** 599

Ralph J, Kim H, Lu F, Grabber JH, Leplé J-C, Berrio-Sierra J, Mir Derikvand M, Jouanin L, Boerjan W, Lapierre C (2008) Identification of the structure and origin of a thioacidolysis marker compound for ferulic acid incorporation into angiosperm lignins (and an indicator for cinnamoyl CoA reductase deficiency). Plant J **53:** 368-379

Ralph J, Lapierre C, Boerjan W (2019) Lignin structure and its engineering. Curr Opin Biotechnol **56:** 240-249

Roach MJ, Schmidt SA, Borneman AR (2018) Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics **19:** 460

Robinson AR, Mansfield SD (2009) Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling. Plant J **58:** 706-714

Rosso MG, Li Y, Strizhov N, Reiss B, Dekker K, Weisshaar B (2003) An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol **53:** 247-259

Sandhu KS, Koirala PS, Neff MM (2013) The *ben1-1* brassinosteroid-catabolism mutation is unstable due to epigenetic modifications of the intronic T-DNA insertion. G3: Genes, Genomes, Genet **3:** 1587-1595

Saze H (2018) Epigenetic regulation of intragenic transposable elements: a two-edged sword. J Biochem **164:** 323-328

Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C, et al (2002) A high-throughput Arabidopsis reverse genetics system. Plant Cell **14:** 2985-2994

Sibout R, Eudes A, Pollet B, Goujon T, Mila I, Granier F, Séguin A, Lapierre C, Jouanin L (2003) Expression pattern of two paralogs encoding cinnamyl alcohol dehydrogenases in Arabidopsis. Isolation and characterization of the corresponding mutants. Plant Physiol **132:** 848-860

Speeckaert N, Adamou NM, Hassane HA, Baldacci-Cresp F, Mol A, Goeminne G, Boerjan W, Duez P, Hawkins S, Neutelings G, et al (2020) Characterization of the UDP-glycosyltransferase UGT72 family in poplar and identification of genes involved in the glycosylation of monolignols. Int J Mol Sci **21:** 5018

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics **24:** 637-644

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JDG (1999) Multiple independent defective *Suppressor-mutator* transposon insertions in Arabidopsis: a tool for functional genomics. Plant Cell **11:** 1841-1852

Ülker B, Peiter E, Dixon DP, Moffat C, Capper R, Bouché N, Edwards R, Sanders D, Knight H, Knight MR (2008) Getting the most out of publicly available T-DNA insertion lines. Plant J **56:** 665-677

Van Acker R, Vanholme R, Storme V, Mortimer JC, Dupree P, Boerjan W (2013) Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in *Arabidopsis thaliana*. Biotechnol Biofuels **6:** 46

Vanholme R, De Meester B, Ralph J, Boerjan W (2019) Lignin biosynthesis and its integration into metabolism. Curr Opin Biotechnol **56:** 230-239

Vanholme R, Morreel K, Darrah C, Oyarce P, Grabber JH, Ralph J, Boerjan W (2012a) Metabolic engineering of novel lignin in biomass crops. New Phytol **196:** 978-1000

Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W (2012b) A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. Plant Cell **24:** 3506-3529

Velten J, Velten L, Hain R, Schell J (1984) Isolation of a dual plant promoter fragment from the Ti plasmid of *Agrobacterium tumefaciens*. EMBO J **3:** 2723-2730

Vermaas JV, Dixon RA, Chen F, Mansfield SD, Boerjan W, Ralph J, Crowley MF, Beckham GT (2019) Passive membrane transport of lignin-related compounds. Proc Natl Acad Sci USA **116:** 23117-23123

Wambui Mbichi R, Wang Q-F, Wan T (2020) RNA directed DNA methylation and seed plant genome evolution. Plant Cell Rep **39:** 983-996

Wang P, Guo L, Morgan J, Dudareva N, Chapple C (2022) Transcript and metabolite network perturbations in lignin biosynthetic mutants of Arabidopsis. Plant Physiol, in press (10.1093/plphys/kiac344).

Wu J, Zhu W, Shan X, Liu J, Zhao L, Zhao Q (2022) Glycoside-specific metabolomics combined with precursor isotopic labeling for characterizing plant glycosyltransferases. Mol Plant, in press (10.1016/j.molp.2022.08.003).

Xue W, Ruprecht C, Street N, Hematy K, Chang C, Frommer WB, Persson S, Niittyla T (2012) Paramutation-like interaction of T-DNA loci in Arabidopsis. PLoS ONE **7:** e51651

Yu H, Liu C, Dixon RA (2021) A gene-editing/complementation strategy for tissue-specific lignin reduction while preserving biomass yield. Biotechnol Biofuels **14:** 175

Yue F, Lu F, Sun R-C, Ralph J (2012) Syntheses of lignin-derived thioacidolysis monomers and their uses as quantitation standards. J Agric Food Chem **60:** 922-928

Zhai J, Bischof S, Wang H, Feng S, Lee T, Teng C, Chen X, Park SY, Liu L, Gallego-Bartolome J, et al (2015) One precursor One siRNA model for Pol IV-dependent siRNAs Biogenesis. Cell **163**: 445-455

Figure legends and tables

Figure 1. The general phenylpropanoid and monolignol-specific pathways in Arabidopsis wild type. Major fluxes are indicated by solid arrows, while minor fluxes are indicated by dashed arrows. General phenylpropanoid pathway: PAL: phenylalanine ammonia-lyase; C4H: cinnamate 4-hydroxylase; C3H: *p*-coumarate 3-hydroxylase; 4CL: 4-coumarate:CoA ligase; HCT: hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase; C3'H: *p*-coumaroyl shikimate 3'-hydroxylase; CSE: caffeoyl shikimate esterase; CCoAOMT: caffeoyl-CoA *O*-methyltransferase. Monolignol-specific pathway: CCR: cinnamoyl-CoA reductase (in blue); F5H: ferulate 5-hydroxylase; COMT: caffeic acid *O*-methyltransferase;

CAD: cinnamyl alcohol dehydrogenase. UGT72E: UDP-glycosyltransferase 72E (in red). UGT84A: UDP-glycosyltransferase 84A. HCALDH: hydroxycinnamaldehyde dehydrogenase. How feruloyl-CoA is converted into ferulic acid is currently unknown, it might be either spontaneous or catalyzed via a thioesterase. H: *p*-hydroxyphenyl, G: guaiacyl, S: syringyl, FA: ferulic acid.

Figure 2. The *ccr1-6* dwarfed phenotype is suppressed after a cross between *ccr1-6* and *ugt72e1,-e2,-e3* mutants. A, Phenotype of seven-week-old *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* Arabidopsis plants grown under long-day conditions. The picture is representative for multiple plants of each genotype/generation (n = 21, except for *epiccr1-6* F3: n = 3). Scale bar = 5 cm. B, Tissue anatomy and lignin deposition in *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants. Transverse stem sections of vascular bundles were stained with Wiesner or Mäule reagent. V: vessel, XF: xylary fiber, IF: interfascicular fiber, P: phloem. Scale bars = 50 µm. C, Relative *CCR1* expression levels in rosette leaves of 25-day-old *ugt72e1,-e2,-e3 epiccr1-6* and *epiccr1-6* mutants. The error bars designate the standard error of the mean (n = 5). Values on the Y-axis represent *CCR1* transcript levels compared with the wild type (set to 1). Different letters represent significant differences at the 0.05 significance level (ANOVA, Tukey *post-hoc*).

Figure 3. Minor growth restoration of *ugt72e1,-e2,-e3 ccr1*-CRISPR mutants. Phenotype of 7-week-old Arabidopsis plants grown under long-day conditions. The picture is representative for multiple plants of each genotype (n = 17). Scale bar = 5 cm.

Figure 4. Methylation profile plot of the T-DNA copies, dSpm element and vector backbone sequences at different loci in the *ccr1-6* and *ugt72e1,-e2,-e3 epiccr1-6* mutant backgrounds. A, *CCR1* locus in the *ccr1-6* mutant background. B, *CCR1* locus, C, *UGT72E1* locus, D, *UGT72E2* locus and E, *UGT72E3* locus in the *ugt72e1,-e2,-e3 epiccr1-6* mutant background. Windows of each plot were framed 1 kb upsteam and 1 kb downstream of the gene prediction of the *de novo* genome assembly. The top panel is informative of the position of the gene (5'UTR as white rectangle, exon as black rectangle, 3'UTR as white arrow), T-DNA and vector backbone (bb) (light and dark color arrows, respectively). Red arrows correspond to the GABI-KAT T-DNA collection, blue arrows to the SAIL T-DNA collection, of which the arrow points towards the right border, and green arrows to the JIC SM transposant library, of which the arrow points towards the 3' end. Gray elements annotated on the T-DNA copies and vector backbone sequences are I: mannopine synthase 1' 2' promoter, II: sulfonamide resistance gene, III: region of the pUC18 vector, IV: ampicillin resistance gene, V: 35S promoter, VI: BASTA resistance gene and VII: spectinomycin resistance gene, of which the arrow points towards the 3' end. The second panel represents the individual reads aligned onto the *de novo* genome assembly of the individual mutant lines, with methylated cytosines (in the CG dinucleotide context) as black dots and non-methylated cytosines as blue dots. The third panel shows the translation from genome coordinate space into a modified base space consisting only of instances of the methylated motif (methylated cytosines in the CG dinucleotide context). The fourth panel shows the raw log-likelihood ratio (LLR) plot of cytosine methylation. The bottom panel shows a smoothed sliding window plot showing methylation fraction across the region. Finally, shaded bands across the background of the plot highlight the position of the T-DNA copies, dSpm element or vector backbone sequences across the various panels.

Figure 5. Pairwise sequence alignments of the complete T-DNA (and dSpm element) sequences of the most commonly used T-DNA consortia (and transposant) collections. Sequence homology in the sense orientation is visible as a line from the upper left corner to the bottom right corner of the dotplot. Sequence homology in the antisense orientation

(reverse complement) is visible as a line from the bottom left corner to the upper right corner of the dotplot.

Figure 6. Proposed molecular mechanism for the initiation of *trans* T-DNA suppression. Example of *trans* T-DNA suppression of the *ccr1-6* allele after a genetic cross of the *ugt72e3* and *ccr1-6* T-DNA mutants sharing stretches of identical sequences. Initiation of *trans* T-DNA suppression is mediated by the canonical RNA-directed DNA methylation (RdDM) pathway (Matzke and Mosher, 2014; Wambui Mbichi et al., 2020). In short, double-stranded RNA (dsRNA) is processed by dicer-like 3 (DCL3) into 24-nucleotide (nt) small interfering RNAs (siRNAs) that are incorporated into argonaute 4 (AGO4) and base pair with the transcript produced by RNA polymerase V (Pol V). In turn, AGO4 recruits domains rearranged methyltransferase 2 (DRM2), which catalyzes the *de novo* CG methylation (Me) of complementary DNA. It is hypothesized that the RNA duplex originates either from (A) a Pol IV-produced transcript (e.g. *ugt72e3*) that is processed into dsRNA by RNA-dependent RNA polymerase 2 (RDR2), which is part of the canonical RdDM pathway or (B and C) from RNA-RNA annealing of transcript(s) made via Pol II. Primary transcripts generated by Pol II may still contain T-DNA sequence(s) if not prematurely terminated. A primary transcript (e.g. *ugt72e3*) containing a reverse complement T-DNA-derived sequence might interact (B) in *cis* with a direct T-DNA-derived sequence (e.g. *ugt72e3*) to form a hairpin structure (Osabe et al., 2017) or (C) in *trans* with a direct T-DNA-derived sequence from another locus (e.g. *ccr1-6*) to form dsRNA (Jia et al., 2015). Note that the *ugt72e3* locus has a big deletion, resulting in an intergenic region flanking the T-DNA insertion site. Also, the identical regions in the GABI-Kat and SAIL T-DNA are in reverse orientation, affecting how transcripts from the respective *ccr1-6* and *ugt72e3* loci can form dsRNA. In the gene models, black triangles, black boxes, white boxes, solid lines and dotted lines represent the T-DNAs, exons, UTRs, introns and intergenic regions respectively. Polymerases are in blue: Pol IV transcripts (30-40 nt long: Zhai et al., 2015) and Pol V transcripts (approximately 200 nt in length: Böhmdorfer et al., 2016), are much shorter than Pol II transcripts and are represented accordingly. Other proteins involved in the initiation of *trans* T-DNA suppression are in red. LB: T-DNA left border, RB: T-DNA right border and >/<: orientation from LB to RB.

Table 1. Crude cell wall residue (CWR) and lignin content of *ugt72e1,-e2,-e3 epiccr1-6* quadruple and *epiccr1-6* mutants. Measurements were performed on inflorescence stems of fully senesced plants. The CWR is expressed as a percentage of dry weight. Lignin content was determined with the acetyl bromide (AcBr) assay and expressed as a percentage of CWR. The data represent the mean values ± standard deviation, n = 9. Different letters represent significant differences at the 0.05 significance level (ANOVA, Tukey *post-hoc*).

Line	CWR (% dry weight)	AcBr lignin (% CWR)
wild type	79.0 ± 1.7 a	14.3 ± 0.1 a
ugt72e1,-e2,-e3	78.0 ± 2.4 a	13.8 ± 0.2 a
ccr1-6	71.3 ± 3.0 b	9.0 ± 0.1 b
ugt72e1,-e2,-e3 epiccr1-6	74.1 ± 4.6 b,c	10.7 ± 0.1 c
epiccr1-6 F4	76.9 ± 2.9 a,c	10.4 ± 0.1 c
epiccr1-6 F5	77.0 ± 2.0 a,c	10.8 ± 0.2 c

Table 2. FA-derived lignin units in *ccr1*-CRISPR and *ugt72e1,-e2,-e3 ccr1*-CRISPR mutant lines. Lignin composition was determined with thioacidolysis. The data represent the mean values ± standard deviation, n = 8. Different letters represent significant differences at the 0.05 significance level (ANOVA, Tukey *post-hoc*). β-*O*-4-FA-I: G-CH=CH-COOH, β-*O*-4-FA-II: G-CHR-CH2-COOH, bis-β-*O*-4-FA: G-CHR-CHR2, where 'R' is a thioethyl group.

Line	β-*O*-4-FA-I (nmol g^{-1} AcBr lignin)	β-*O*-4-FA-II (nmol g^{-1} AcBr lignin)	bis-β-*O*-4-FA (nmol g^{-1} AcBr lignin)
wild type	6.31 ± 0.75 a	3.47 ± 0.97 a	31.66 ± 13.12 a
ugt72e1,-e2,-e3	7.15 ± 1.14 a	3.55 ± 0.71 a	26.95 ± 14.37 a
ccr1-CRISPR line 1	43.25 ± 7.48 b	23.18 ± 3.61 b	244.04 ± 81.59 b
ccr1-CRISPR line 2	48.62 ± 9.24 b	22.61 ± 8.11 b	235.59 ± 121.13 b
ugt72e1,-e2,-e3 ccr1-CRISPR line 1	39.59 ± 10.73 b	12.64 ± 6.50 c	132.60 ± 76.66 b
ugt72e1,-e2,-e3 ccr1-CRISPR line 2	40.54 ± 8.76 b	13.53 ± 5.98 b,c	170.84 ± 84.25 b