



OPEN

DATA DESCRIPTOR

Biochemical data documenting variations in mucilage polysaccharides in a range of glycosyltransferase mutants

Yuki Aoi^{1,2}, Abdelilah Benamar¹, Luc Saulnier², Marie-Christine Ralet²✉ & Helen M. North¹✉

During *Arabidopsis* seed coat development, copious amounts of mucilage polysaccharides are produced in the epidermal cells. When hydrated on imbibition, these polysaccharides expand and are released to encapsulate the seed as a two-layered hydrogel. Polysaccharides are synthesized from UDP-sugars by glycosyltransferases (GTs) and several GTs, with differing activities, have been identified that contribute to mucilage polysaccharide synthesis. How these GTs orchestrate production of the complex polysaccharides found in mucilage remains to be determined. In this study, we generated a range of multiple GT mutants using either CRISPR/Cas9 targeted mutation or genetic crosses of existing T-DNA insertion mutants. Four traits for mucilage amounts or macromolecular properties were examined for four replicate seed lots from 31 different GT mutant combinations. This data provides a valuable resource for future genetic, biochemical, structural, and functional studies of the roles and properties of polysaccharides present in *Arabidopsis* mucilage and the relative contributions of different GTs to mucilage production.

Background & Summary

Mucilage is formed of polysaccharides which are produced in copious amounts in seed coat epidermal cells during seed development. The polysaccharides are hydrated and expand upon seed imbibition, leading to their release from the epidermal cells and the encapsulation of the seeds as a hydrogel^{1,2}. *Arabidopsis* mucilage is structured in two layers: an outer diffuse layer which can be removed easily by shaking with water and an inner adherent layer which remains attached to the seed coat³. Both layers are primarily composed of rhamnogalacturonan-I (RGI), a pectic domain. The inner layer also contains small amounts of another pectic domain, homogalacturonan (HG), together with cellulose and some hemicelluloses^{1,3,4}. As *Arabidopsis* mucilage shares common polysaccharides with cell walls in a more accessible format, it can be used as a simple proxy to investigate polysaccharide synthesis and physico-chemical properties.

Polysaccharides are synthesized from UDP- and GDP-sugars by a range of glycosyltransferases (GTs) having different activities, some of which have been implicated in the production of mucilage pectin based on mutant phenotypes. RGI: RHAMNOSYLTRANSFERASE1 (*RRT1*) was the first rhamnosyltransferase demonstrated to transfer Rha from UDP-Rha onto RGI primers⁵. While *RRT1* is highly expressed in the seed coat at the developmental stage when mucilage polysaccharides are being produced, *rrt1* mutant seeds have only slightly reduced mucilage RGI amounts. Expression analysis showed that two other GTs, *MUCILAGE-RELATED70* (*MUCI70*) and *GALACTURONOSYLTRANSFERASE-LIKE5* (*GATL5*), are also highly expressed in the seed coat^{6,7}. In addition to dramatically reduced mucilage polysaccharide amounts, mucilage from *muci70* mutant seeds is composed of shorter RGI polymer chains⁸. A recent study demonstrated the *in vitro* galacturonosyltransferase activity of *MUCI70* towards RGI, which was subsequently named RHAMNOGALACTURONAN GALACTURONOSYLTRANSFERASE1 (*RGGAT1*)⁹. In contrast, seed mucilage RGI polymers in *gat15* are larger than those of wild type (WT), even though the amount of mucilage polysaccharides is still reduced⁷. While

¹Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), 78000, Versailles, France.

²INRAE, UR1268 BIA, 3 impasse Yvette Cauchois, CS71627, 44316 Cedex3, Nantes, France. ✉e-mail: marie.ralet@inrae.fr; helen.north@inrae.fr

Genotype	genotype_code
WT	4000
<i>rrt1-1</i>	4001
<i>rrt1-2</i>	4002
<i>rrt1-3</i>	4003
<i>gaut11-4</i>	4004
<i>gat5-1</i>	4005
<i>gat5-3</i>	4006
<i>muci70-1</i>	4007
<i>muci70-3</i>	4008
<i>irx14-2</i>	4009
<i>irx14-4</i>	4010
<i>mum5-3</i>	4011
<i>rrt1-1 gaut11-4</i>	4012
<i>rrt1-3 gaut11-5</i>	4013
<i>rrt1-1 gat5-1</i>	4014
<i>rrt1-2 gat5-3</i>	4015
<i>rrt1-2 irx14-4</i>	4016
<i>rrt1-3 mum5-6</i>	4017
<i>gaut11-4 gat5-1</i>	4018
<i>gaut11-4 muci70-1</i>	4019
<i>gaut11-4 irx14-2</i>	4020
<i>gaut11-4 mum5-3</i>	4021
<i>gat5-1 muci70-1</i>	4022
<i>gat5-1 irx14-2</i>	4023
<i>gat5-3 irx14-4</i>	4024
<i>gat5-1 mum5-3</i>	4025
<i>gat5-3 mum5-4</i>	4026
<i>irx14-2 mum5-3</i>	4027
<i>irx14-6 mum5-4</i>	4028
<i>rrt1-2 gat5-3 irx14-4</i>	4029
<i>gat5-3 irx14-5 mum5-4</i>	4030
<i>rrt1-3 gaut11-5 mum5-6</i>	4031

Table 1. List of glycosyltransferase mutant genotypes used in this study and their corresponding genotype code in the dataset.

GALACTURONOSYLTRANSFERASE11 (GAUT11) was shown to catalyse the addition of UDP-galacturonic acid (GalA) to HG primers, *gaut11* mutants surprisingly exhibited reduced amounts of RGI in mucilage.

Other GT mutants show an altered distribution of mucilage polysaccharides between the outer and the inner layers of mucilage, with no effect on global amounts. This is the case for mutants of *MUCILAGE MODIFIED 5 (MUM5)/MUCI21* and *IRREGULAR XYLEM 14 (IRX14)*, which have more outer mucilage layer polysaccharides and a corresponding reduction in the inner mucilage layer^{10,11}. MUM5 and IRX14 have been proposed to be xylosyltransferases that synthesize xylan, a hemicellulose, which plays a pivotal role in the formation of the inner layer by anchoring RGI to cellulose attached to the seed surface¹¹. The six GT mutants cited above (*rrt1*, *muci70*, *gat5*, *gaut11*, *irx14*, *mum5*) have each been shown individually to impact mucilage polysaccharide characteristics. Nevertheless, it is currently difficult to appraise how these GTs collectively contribute to RGI synthesis and the nature of the RGI-xylan relationship, due to the lack of integrated studies.

Here, we have generated double and triple mutant combinations of the six GTs indicated above by a CRISPR/Cas9 targeted gene-editing approach to create novel, knock-out mutant alleles and by genetic crosses between existing T-DNA insertion mutant alleles. Mucilage from seeds of these mutants was then analysed to create a novel data set cataloguing the amounts and macromolecular properties of the different genotypes. These data will be a rich resource for determining how different GT activities impact RGI production and modulate polymer macromolecular properties, as well as for the choice of genotypes in studies examining the role of mucilage using seeds with different amounts and properties.

In total, data for 31 different GT mutants and their corresponding WT (Col-0) were obtained; 11 single, 17 double and 3 triple mutants (Table 1). These were used to generate data records which correspond to mucilage amounts quantified for the outer and inner layers using a medium-throughput method, and values for macromolecular characteristics of polymers present in the outer mucilage layer. The experimental procedures and different steps in data production are summarised in Fig. 1.

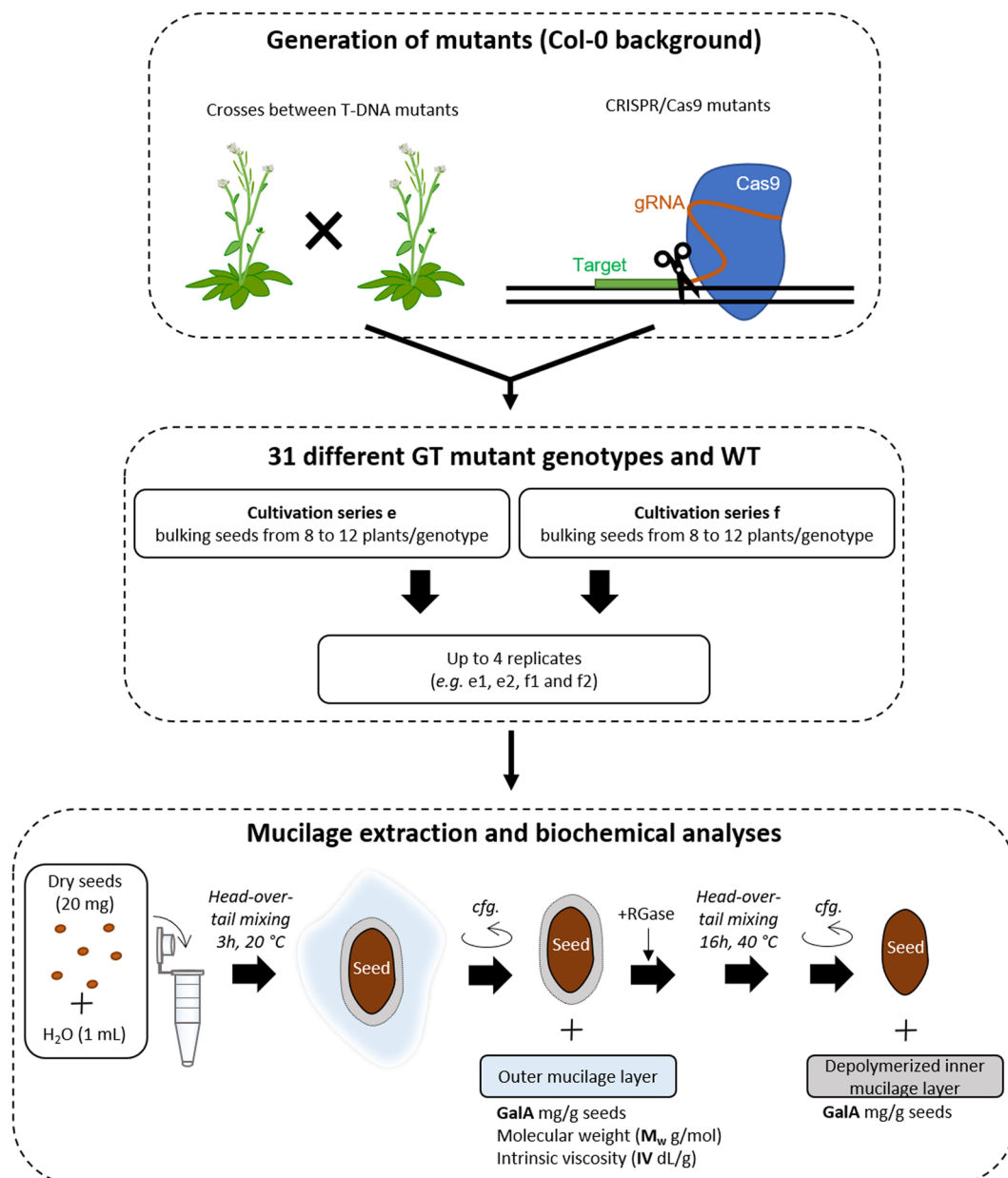


Fig. 1 Schematic representation of the experimental workflow used to generate data records for glycosyltransferase mutant mucilage traits. Mutants were generated by either a CRISPR/cas9 based-approach or crosses between previously characterised T-DNA mutants. Seeds had been produced at two different times corresponding to cultivation series e or f, and up to four replicates were used for biochemical analyses. Analyses of mucilage GalA contents and macromolecular characteristics for the 32 genotypes are available in two data records. cfg: centrifugation; RGase: rhamnogalacturonan hydrolase.

Methods

Plant materials and growth conditions. The *Arabidopsis thaliana* T-DNA insertion mutants *rrt1-1*, *gaut11-4*, *muci70-1*, *gatl5-1*, *irx14-2* and *mum5-3* were previously described^{5-7,11} (Table 2); all mutants are in the Col-0 accession background. The genotypes of homozygous lines, and homozygous double mutants derived from crosses between two T-DNA lines, were confirmed by PCR using the primers listed in Table 2 and Supplemental Table 1. Other mutant alleles and mutant combinations were generated by CRISPR/Cas9 targeted gene editing as described below. Surface-sterilized seeds were sown onto *Arabidopsis* Gamborg B5 growth medium (Duchefa; <https://www.duchefa-biochemie.com/>) pH 6 supplemented with 30mM Sucrose, 0.08% (w/v) bromocresol purple and 0.7% (w/v) agar. Following stratification for three days (4 °C in the dark), seeds were germinated in a growth chamber for 14 days (16h photoperiod, 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ light intensity, 60% relative humidity, 18 °C). Seedlings

Mutant type	Genotype	References/ Mutation position	Primer 1	Primer 2	PCR product type
T-DNA insertion	<i>rrt1-1</i> (SALK_022924C)	Takenaka <i>et al. Nature Plants</i> (2018)	RRT1-TF	RRT1-TR	Gene specific
			RRT1-T	SALK-LBb1	T-DNA specific
	<i>gaut11-4</i> (SAIL_406_D01)	Voiniciuc <i>et al. Plant Physiol.</i> (2018)	GAUT11-TF	GAUT11-TR	Gene specific
			GAUT11-T	SAIL-IT-1	T-DNA specific
	<i>gatl5-1</i> (SALK_106615)	Kong <i>et al. Plant Physiol.</i> (2013)	GATL5-TF	GATL5-TR	Gene specific
			GATL5-T	SALK-LBb1	T-DNA specific
	<i>muci70-1</i> (SALK_129524)	Voiniciuc <i>et al. Plant Physiol.</i> (2018) Fabrissin <i>et al. Plant Physiol.</i> (2019)	MUCI70-TF	MUCI70-TR	Gene specific
			MUCI70-T	SALK-LBb1	T-DNA specific
	<i>irx14-2</i> (SALK_205399C)	Voiniciuc <i>et al. Plant Physiol.</i> (2015) Brown <i>et al.</i> (2007)	IRX14-TF	IRX14-TR	Gene specific
			IRX14-T	SALK-LBb1	T-DNA specific
	<i>mum5-3/muci21-1</i> (SALK_041744)	Ralet <i>et al. Plant Physiol.</i> (2016) Voiniciuc <i>et al. Plant Physiol.</i> (2015)	MUM5-TF	MUM5-TR	Gene specific
			MUM5-T	SALK-LBb1	T-DNA specific
CRISPR/Cas9 edited	<i>rrt1-2</i>	1/8 exons 37 bp deletion of G177-T213	RRT1-CF	RRT1-CR	Gene specific for sequencing
	<i>rrt1-3</i>	1/8 exons 26 bp deletion of T167-G192	RRT1-CF	RRT1-CR	Gene specific for sequencing
	<i>gaut11-5</i>	1/2 exons 1 bp insertion A between A52-G53	GAUT11-CF	GAUT11-CR	Gene specific for sequencing
	<i>gatl5-3</i>	1/1 exon 1 bp deletion of G312	GATL5-CF	GATL5-CR	Gene specific for sequencing
	<i>muci70-3</i>	1/9 exons 1 bp deletion of C79	MUCI70-CF	MUCI70-CR	Gene specific for sequencing
	<i>irx14-4</i>	1/3 exons 47 bp deletion of C385-G431	IRX14-CF	IRX14-CR	Gene specific for sequencing
	<i>irx14-5</i>	1/3 exons 1 bp insertion T between T402-G403	IRX14-CF	IRX14-CR	Gene specific for sequencing
	<i>irx14-6</i>	1/3 exons 1 bp deletion of G400	IRX14-CF	IRX14-CR	Gene specific for sequencing
	<i>mum5-4</i>	1/3 exons 1 bp insertion A between A599-T600	MUM5-CF	MUM5-CR	Gene specific for sequencing
<i>mum5-6</i>	1/3 exons 1 bp insertion A between A194-C195	MUM5-CF	MUM5-CR	Gene specific for sequencing	

Table 2. Summary of mutant alleles studied and the primers used for their genotyping. For full primer details see Supplemental Table 1.

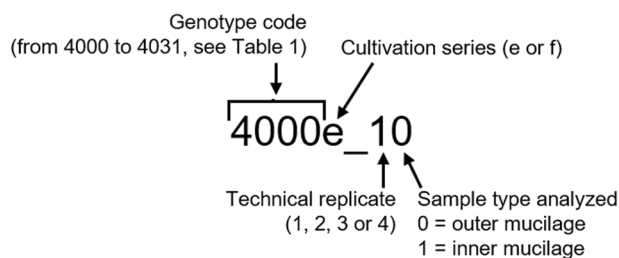


Fig. 2 Nomenclature for seed samples analyzed to generate dataset. This is based on, and compatible with, that used in Poulain *et al.*²⁰.

were transplanted to soil (Tref Substrates; <https://jiffygroup.com>) and transferred to the glasshouse (18°C–28°C, minimum photoperiod of 13 h provided by supplementary lighting) and watered for 3 weeks with tap water, then with Plan-Prod nutritive solution (Fertil; <https://www.fertil.fr/>). The plants were grown at two different times,

guideRNA	Target sequence with PAM site
GATL5_guideRNA1	ATGATCGGATAGCTGCCGC <u>CGG</u>
GATL5_guideRNA2	<u>CCT</u> CCAGCACTCAATGTGCCCT
GAUT11_guideRNA1	TAGAAGGAGATTGTCGAGT <u>TGG</u>
GAUT11_guideRNA2	<u>CCAT</u> CAACAAGATCCATCCCAG
IRX14_guideRNA1	<u>CCG</u> GAGTAATAGCTTCAGATCT
IRX14_guideRNA2	CCGCATCCGAATCCCGT <u>TGAGG</u>
MUC170_guideRNA1	<u>CCT</u> ATCCAGATCACTACCACGA
MUC170_guideRNA2	AAGATAGTCAGGAAGGTCA <u>AGG</u>
MUM5_guideRNA1	<u>CCCT</u> TACCCCTCTGCTCTGTTT
MUM5_guideRNA2	AATGATCTCACCACCGATCAGG
RRT1_guideRNA1	ATGGCTTTGGGAGAGATGT <u>GGG</u>
RRT1_guideRNA2	<u>CCT</u> ATTAGTTGGTCAAACATGT

Table 3. Glycosyltransferase gene guideRNA target sequences.

series 'e' was grown from February 2022 to April 2022 and series 'f' from March 2022 to May 2022. Seeds for each genotype were harvested as bulks from eight to twelve plants from a given series (e or f) so that each series represents a biological replicate. Subsequent analyses used up to four replicates (e.g. e_1, e_2, f_1 and f_2, where numbers indicate the technical replicates within a series) (Fig. 2).

Generation of GT mutants using a CRISPR/Cas9 targeted gene editing strategy. Two different guide RNAs were designed for each of the six target GTs (twelve guide RNAs in total), using the CRISPOR Web site (<http://crispor.tefor.net>), and selecting from proposed guide sequences based on their predicted efficiency and off-target probability. The specificity of the guides for their corresponding genes was checked by a BLAST search (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the Arabidopsis genome. Guide sequences were also selected for their position close to the 5' end of each gene targeted, in order to increase the likelihood that induced mutations would produce truncated, inactive proteins. The target sequences of the twelve guide RNAs are listed in Table 3. Forward and reverse primers for each guide RNA were synthesized with additional sequences (Eurofins Genomics, <https://eurofinsgenomics.eu/>), that are complementary to the cohesive termini generated by *Bsa*I digestion of either the pUPD:*pU6-26* vector containing the Arabidopsis *U6* promoter or the pUPD:*psgRNA* vector containing a scaffold guideRNA. Both plasmids enabling the guide to be cloned directionally into the domestication vector pDGB3_α1 or pDGB3_α2, to generate a transcription unit (TU) through a GoldenBraid (GB) cyclic digestion/ligation level 0 reaction as detailed by Vazquez-Vilar *et al.*¹². These 12 guide RNA TUs were then stacked into a single vector through a 5 level pairwise cycle that alternated omega and alpha pDGB3 vectors between levels. The transcription units pro*RPS5A:hCas9:ter*, composed of the Ribosomal Protein 5A promoter (*pRPS5A*) with the hCAS9 CDS (GB0575) and the *RbcsE9* terminator, and pro*CMV:DsRED:terNos* (gifts of Lionel Gissot), with the final 12 guide RNA TUs-pro*RPS5A:hCas9:ter*-pro*CMV:DsRED:terNos* in pDGB3_α2. The correct assembly was confirmed by PCR and sequencing at each step in *E. coli* (DH10B) or *A. tumefaciens* (C58C1) using the primers listed in Supplemental Table S1. Arabidopsis (Col-0 accession) was transformed with the final binary vector by the agrobacterium-mediated floral dip method¹³. Transgenic seeds were identified by DsRED fluorescence and DNA extracts from resulting plants subsequently genotyped for edited genes by sequencing PCR products amplified with the GT-CF and GT-CR primers (e.g. RRT1-CF and RRT1-CR) listed in Table 2 and Supplemental Table S1. To fix the mutation, Cas9-containing constructs were removed from segregating lines by selecting for seeds that did not exhibit DsRED fluorescence in the progeny of selected edited lines¹⁴. A further round of genome sequence analysis was carried out to confirm the edited mutations indicated in Fig. 3 and Table 2.

Mucilage Extraction and biochemical analyses. A sequential extraction of the outer and the inner mucilage layers was performed essentially as previously described^{8,15}, but adapted to a much smaller quantity of seeds. Twenty mg of intact seeds were mixed head-over-tail in 1 mL of distilled water for 3 h at room temperature. The suspension was centrifuged (17,500 g, 10 min) and exactly 760 µl of supernatant was recovered and filtered through a disposable glass microfiber filter (13 mm diameter, 0.45 µm pore size; A.I.T. France FS13PVDF45) for analysis as an outer mucilage extract. Seeds were resuspended by adding exactly 800 µl of 50 mM sodium acetate buffer pH 4.5. Then, 10.5 µL of rhamnogalacturonan hydrolase (EC 3.2.1.171, glycoside hydrolase family 28) provided by Novozymes, at 0.1 mg/mL in 50 mM sodium acetate buffer pH 4.5, were added. The seed suspension was then incubated for 16 h at 40°C. Samples were centrifuged (17,500 g, 10 min), and the supernatants were collected for analysis as inner mucilage extracts.

The GalA content of the outer and inner mucilage extracts was determined colorimetrically by the automated *m*-hydroxybiphenyl method¹⁶. This quantification method is based on the ability of sugars to be converted into furfural derivatives in the presence of hot sulfuric acid. Furfural derivatives can then condense with various phenolic compounds to produce a coloured complex that can be quantified using a spectrometer. Acidic sugars can be quantified specifically using *meta*-hydroxy biphenyl (*m*phenyl-phenol or 3 phenyl-phenol; 530 nm)¹⁷. GalA solutions at 20, 40, 60, 80, and 100 µg/mL were used to generate a standard curve. To take residual outer

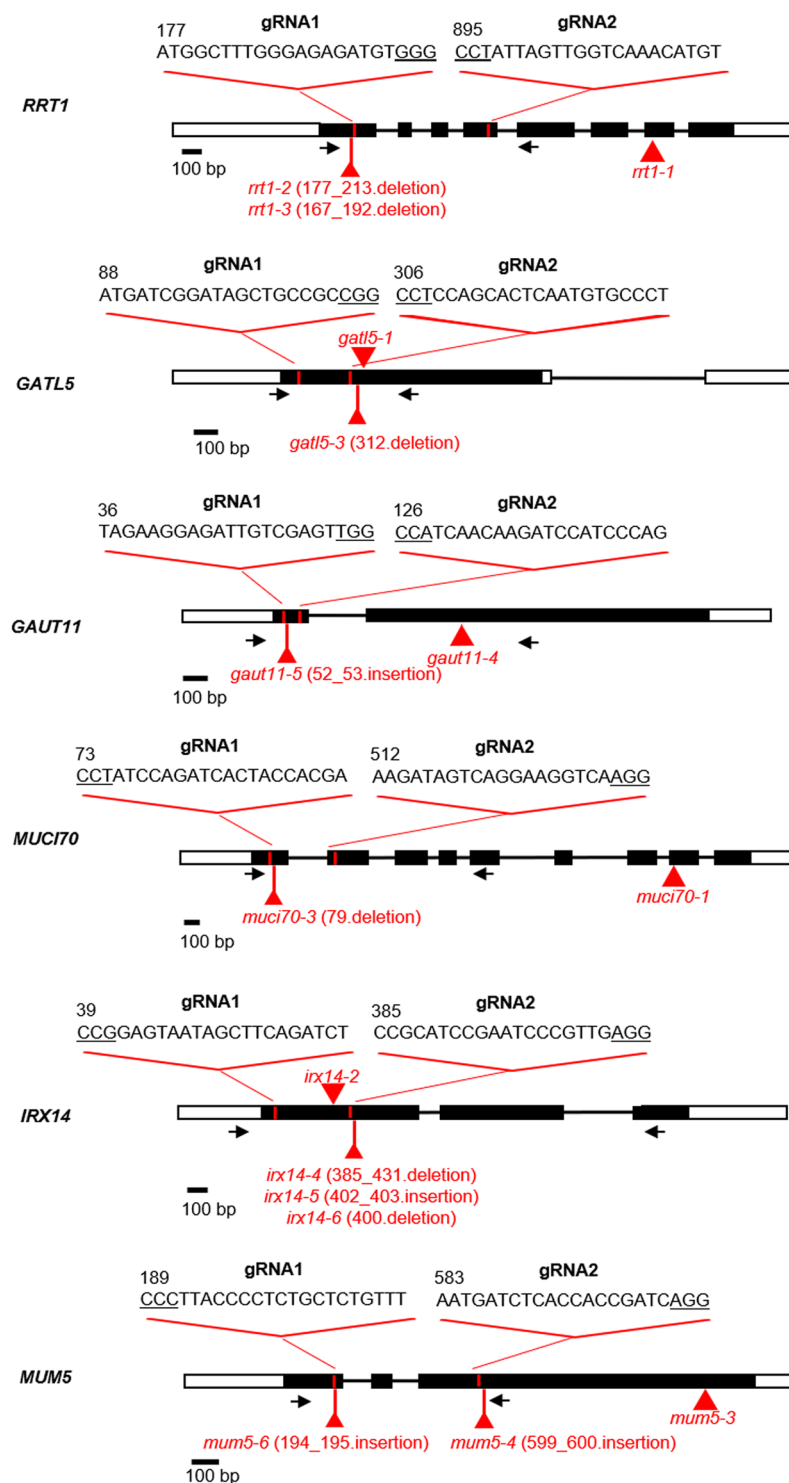


Fig. 3 Position of mutations in the glycosyltransferase mutants used in this study and the sequence of guide RNAs used for gene-editing. Graphic representation of six glycosyltransferase genes (*RRT1*, *GATL5*, *GAUT11*, *MUCI70*, *IRX14* and *MUM5*) with black boxes representing exons and red symbols and text the site and type of mutation in the corresponding mutants. The positions of the gene-specific primers used for genotyping gene-edited mutants are indicated with black arrows. For the precise position of mutations see Table 2. Guide RNAs, gRNA; bp, base pair.

mucilage extract present in inner mucilage extract into account, GalA amounts for the inner mucilage layer were calculated by subtraction of the GalA contents corresponding to this residual outer mucilage, as detailed in the following equation:

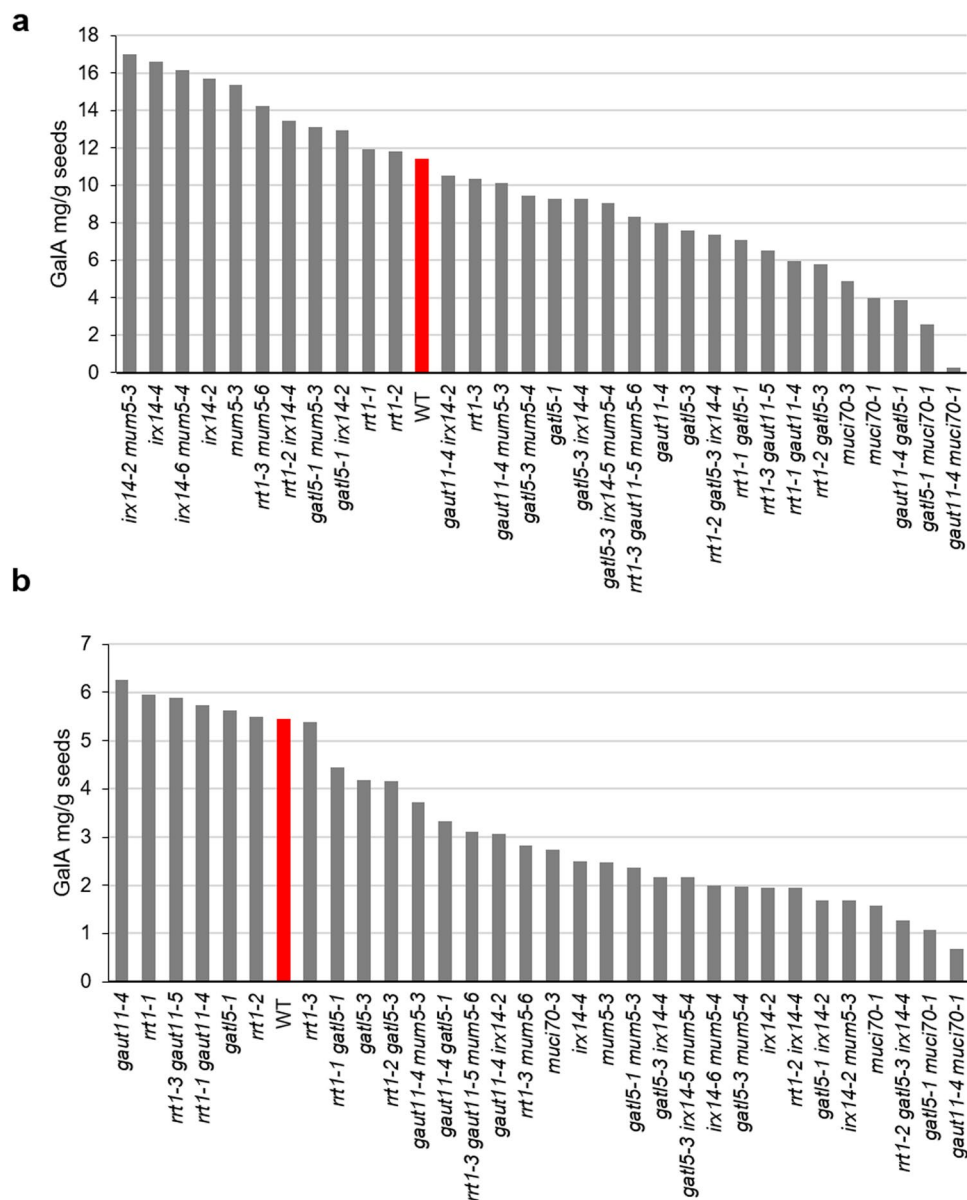


Fig. 4 Glycosyltransferase mutants exhibit a range of GalA amounts in (a) outer and (b) inner mucilage layers. Values represent one replicate from each of the 32 different genotypes. Red bar indicates the value for wild type (WT).

$$GR(\mu\text{g}) = [GO] (\mu\text{g}/\mu\text{L}) * 240(\mu\text{L})$$

$$GI(\mu\text{g}) = [GIE] (\mu\text{g}/\mu\text{L}) * 1050.5(\mu\text{L}) - GR(\mu\text{g})$$

where GR is the GalA contents remaining in the inner mucilage extraction, GO presents the GalA contents in outer mucilage extracts, GI presents the actual inner mucilage contents in the inner mucilage extraction, and GIE presents the GalA contents in inner mucilage extracts. The 240 μL corresponds to the volume of water added to the seed (1000 μL) after deduction of the volume recovered as outer mucilage extract (760 μL). The 1050.5 μL corresponds to 240 μL plus the volumes of buffer (800 μL) and enzyme (10.5 μL) added for adherent mucilage hydrolysis. The GalA contents for outer and inner mucilage layers were then expressed with regard to seed mass.

The outer mucilage extracts were also analyzed for their weight-average molar mass (M_w) and intrinsic viscosity (IV) on a high-performance size exclusion chromatography (HPSEC) system (OMNISEC RESOLVE-REVEAL - Malvern Panalytical). The filtered (0.45 μm PVDF filter) outer mucilage extracts were

Variable	Unit	Variable code	Sample type analyzed
Galacturonic acid content	mg/g seeds	GalA	0 and 1
Weight average molecular weight	g/mol	M_w	0
Intrinsic viscosity	dL/g	IV	0

Table 4. Nomenclature used for variables in the dataset and indication of sample type analysed for each trait variable. Each variable measured is listed with its corresponding code, unit of measure and the sample type analyzed (for sample code see Fig. 3).

sample_code	genotype_code	cultivation_series	technical_replicate	sample_type	variable	value
4000e_10	4000	e	1	0	GalA	11.4
4000e_20	4000	e	2	0	GalA	11.6
4000f_10	4000	f	1	0	GalA	13.4
4000f_20	4000	f	2	0	GalA	13.9
4000e_11	4000	e	1	1	GalA	5.4
4000e_21	4000	e	2	1	GalA	5.4
4000f_11	4000	f	1	1	GalA	7.1
4000f_21	4000	f	2	1	GalA	7.4
4000e_10	4000	e	1	0	M_w	478149
4000e_20	4000	e	2	0	M_w	445390
4000f_10	4000	f	1	0	M_w	524358
4000f_20	4000	f	2	0	M_w	546609
4000e_10	4000	e	1	0	IV	6.33
4000e_20	4000	e	2	0	IV	6.29
4000f_10	4000	f	1	0	IV	6.69
4000f_20	4000	f	2	0	IV	6.41

Table 5. Overview of the dataset comprising values for 4 seed mucilage traits from 32 genotypes.

boiled for 5 min, then samples were injected (50 μ L) onto the HPSEC system maintained at 30 °C comprising a Shodex OH SB-G precolumn and a Shodex OH-Pack SB-805 HQ column eluted with 50 mM sodium nitrate at a flow rate of 0.7 mL min⁻¹. Measurements were performed using a differential refractometer (OMNISEC REVEAL), a multi angle laser light scattering detector ($\lambda = 660$ nm, 44°, 60°, 76°, 90°, 108°, 124°, 140°, VISCOTEK SEC-MALS 9) and a differential pressure viscometer (OMNISEC REVEAL). Detectors were calibrated with a pullulan standard having narrow molecular mass distribution (weight-average molar mass = 40.611 Da, number-average molar mass = 38.931 Da, IV = 23.6 mL/g at 30 °C in 0.1 M sodium nitrate, refractive index increment [dn/dc] = 0.147 mL/g). Data analyses were carried out using OmniSec version 11.32 software (Malvern Panalytical) and a dn/dc value of 0.147 mL/g was used for mucilage extracts.

Data Records

The data record contains values for GalA content (mg/g seeds) in both the outer and inner layers of mucilage, as well as the weight-average molar mass (M_w) (g/mol) and intrinsic viscosity (IV) (dL/g) of the outer mucilage layer. The data were obtained from four replicates as described above, excepting three genotypes for M_w and IV and one genotype for GalA (genotypes coded 4004, 4012, 4030 and 4005, respectively, see Table 1), which had three replicates each. In addition, the four replicates analysed for the genotype coded 4010 were all from culture series f (f_1, f_2, f_3, and f_4), and for the genotype coded 4028 exclusively from culture series e (e_1, e_2, e_3, and e_4). All other genotypes were evaluated using two replicates from each culture series (e_1, e_2, f_1, and f_2) (Fig. 2). Note that certain values for M_w and IV were below the detection limit (B.D.L.) (genotypes coded 4007, 4008, 4018, 4019, and 4022).

The extent of variability in inner and outer mucilage GalA amounts observed between different glycosyltransferase mutants, compared to wild-type, is shown in Fig. 4 for one replicate from each of the 31 different mutant genotypes. This large diversity demonstrates the potential of these genotypes and the dataset for exploitation in future studies. The dataset is available at Data INRAE¹⁸. The sample nomenclature for input is outlined in Fig. 2, genotype codes in Table 1 and the description of variables in Table 4. An overview of the dataset is shown in Table 5 with the following seven columns:

Genotype code	Outer mucilage			Inner mucilage
	GalA	M _w	IV	GalA
4000	5.0	4.6	1.4	8.2
4001	2.3	2.1	1.1	3.2
4002	3.7	1.9	0.8	7.3
4003	4.0	3.0	0.8	3.9
4004	3.8	9.1	2.1	5.4
4005	1.4	2.7	3.1	3.4
4006	2.3	6.4	3.1	2.0
4007	1.9	—	—	4.5
4008	4.6	—	—	3.0
4009	8.0	1.3	1.2	1.2
4010	2.3	0.5	1.5	4.8
4011	5.1	1.5	1.5	4.8
4012	3.9	5.9	1.5	6.5
4013	1.5	4.8	1.6	4.2
4014	2.8	2.6	2.2	2.0
4015	2.1	5.0	3.2	2.4
4016	3.1	1.2	1.1	3.5
4017	4.0	1.3	0.3	2.4
4018	2.7	—	—	5.8
4019	0.0	—	—	4.3
4020	6.1	2.7	1.9	6.0
4021	6.8	4.9	3.3	3.9
4022	1.3	—	—	2.3
4023	3.5	2.0	1.1	4.9
4024	3.6	5.2	5.8	1.8
4025	3.4	3.2	3.1	3.1
4026	2.1	2.6	0.6	2.8
4027	7.1	1.7	1.4	2.7
4028	2.0	1.9	2.6	1.7
4029	2.7	2.0	2.4	2.3
4030	2.7	3.2	5.0	2.0
4031	3.6	1.7	1.6	4.5

Table 6. Variation between values from four replicates for the four traits measured in the 32 genotypes studied. Values are standard errors expressed as a % of the average value of the four replicates for wild type. GalA, galacturonic acid contents; M_w, weight average molecular weight; IV, intrinsic viscosity. —; not calculated as below the detection limit.

1. sample_code: the sample code (see Fig. 2)
2. genotype_code: the genotype code (see Table 1)
3. cultivation_series: e or f
4. technical_replicate: the technical replicate (1, 2, 3, 4)
5. sample_type: the sample type analyzed (0 or 1) (see Fig. 2)
6. variable: the code of the variable (see Table 4 for the description)
7. value: the measured value

Technical Validation

The technical quality of the dataset was validated through the use of four replicates from two different culture series, series e grown from February 2022 to April 2022 and series f grown from March 2022 to May 2022, except for four genotype/variable combinations, as indicated above. The reproducibility of results was examined for biochemical analyses based on the variation between four replicates. The variation is presented as standard errors expressed as a % of the average value of the four replicates of WT, and the highest variation observed was less than 10% (Table 6). Furthermore, values obtained for WT mucilage were similar to those previously published from independent studies, validating the reproducibility of measurements (Fig. 5)^{3,11,15,19–23}. This technical validation also confirms that the calculation used to compensate for the absence of a wash step in the medium-throughput method used here produced robust inner mucilage GalA values.

For quantification of GalA concentrations, a standard curve was established both before and after a series of samples, using a serial dilution of GalA at 20, 40, 60, 80, and 100 µg/mL, to confirm technical rigour. HP-SEC

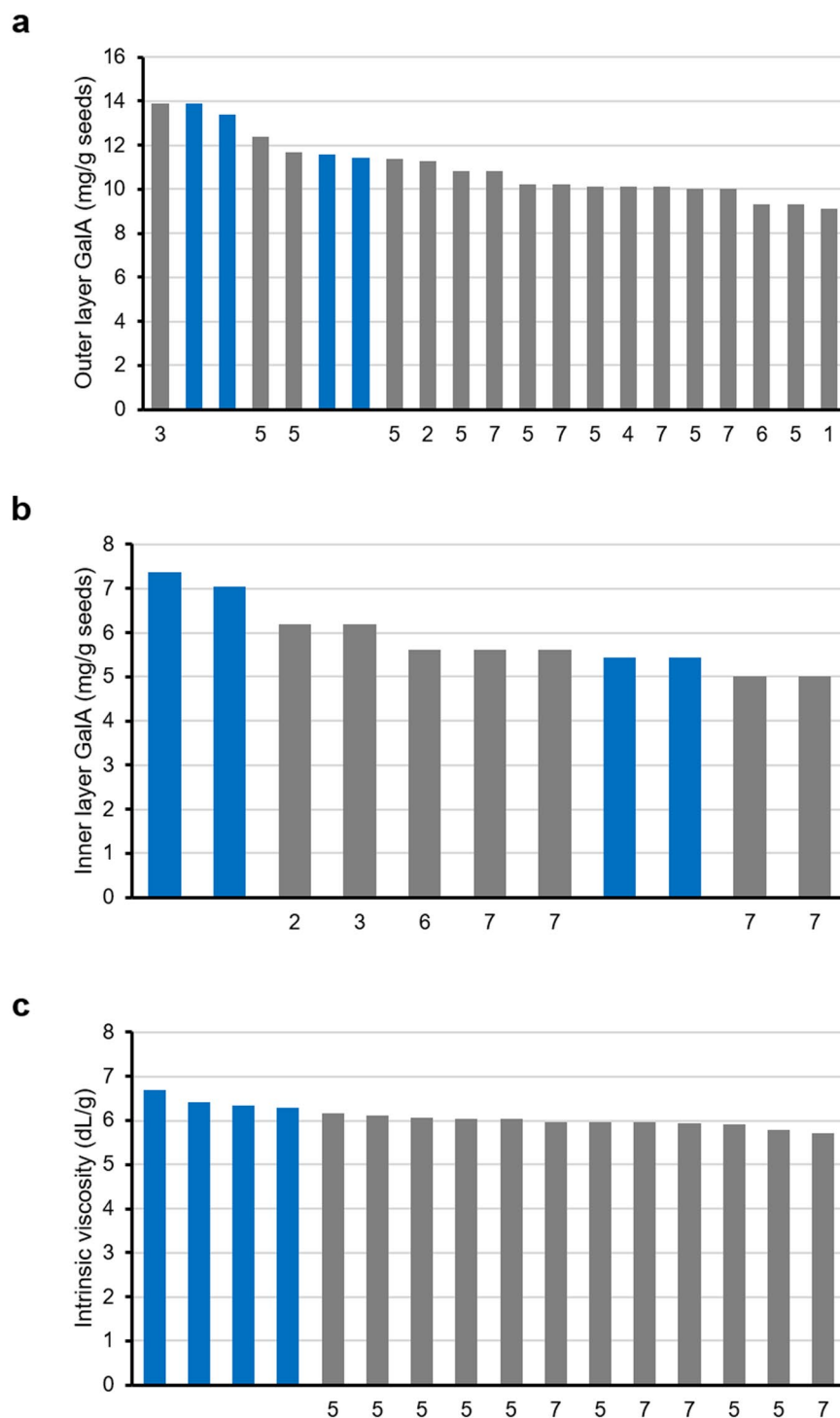


Fig. 5 Technical validation of values obtained by comparison of values for wild-type extracts in this study to previously obtained values. GalA amounts (mg/g seed) for **(a)** outer mucilage and **(b)** inner mucilage extracts, or **(c)** intrinsic viscosity (mL/g) of outer mucilage extracts. Blue bars indicate the values for two technical replicates obtained for samples analyzed in this study (culture e or f, respectively). Grey bars indicate published values corresponding to the following references: 1, Macquet *et al.*³; 2, Saez *et al.*¹⁹; 3, Ralet *et al.*¹¹; 4, Griffith *et al.*²²; 5, Poulain *et al.*²⁰; 6, Saez *et al.*²³; 7, Cambert *et al.*²¹.

columns were calibrated for IV using both a calibrant and a standard sample passed at the beginning, middle and end of a series of samples to check that no drift occurred over time.

	Quantification of GalA amounts	Macromolecular characteristics
Signal acquisition	San Plus Analyzer from Skalar analytical	VISCOTEK SEC-MALS 9
		OMNISEC REVEAL - Malvern Panalytical
Signal processing	Flowaccess v3	Omnisec v11.32

Table 7. Software versions used to acquire and process data.

Code availability

The different available software and the versions used to acquire and process data presented in the dataset are summarized in Table 7.

Received: 4 May 2023; Accepted: 26 September 2023;

Published online: 14 October 2023

References

- North, H. M., Berger, A., Saez-Aguayo, S. & Ralet, M.-C. Understanding polysaccharide production and properties using seed coat mutants: future perspectives for the exploitation of natural variants. *Ann. Bot.* **114**, 1251–1263 (2014).
- Francoz, E., Ranocha, P., Burlat, V. & Dunand, C. Arabidopsis seed mucilage secretory cells: regulation and dynamics. *Trends Plant Sci.* **20**, 515–524 (2015).
- Macquet, A., Ralet, M.-C., Kronenberger, J., Marion-Poll, A. & North, H. M. *In situ*, chemical and macromolecular study of the composition of *Arabidopsis thaliana* seed coat mucilage. *Plant Cell Physiol.* **48**, 984–999 (2007).
- Yu, L. *et al.* CELLULOSE SYNTHASE-LIKE A2, a glucomannan synthase, is involved in maintaining adherent mucilage structure in Arabidopsis seed. *Plant Physiol.* **164**, 1842–1856 (2014).
- Takenaka, Y. *et al.* Pectin RG-I rhamnosyltransferases represent a novel plant-specific glycosyltransferase family. *Nature Plants* **4**, 669–676 (2018).
- Voiniciuc, C. *et al.* Identification of Key Enzymes for Pectin Synthesis in Seed Mucilage. *Plant Physiol.* **178**, 1045–1064 (2018).
- Kong, Y. *et al.* GALACTURONOSYLTRANSFERASE-LIKE5 Is Involved in the Production of Arabidopsis Seed Coat Mucilage. *Plant Physiol.* **163**, 1203–1217 (2013).
- Fabrissin, I. *et al.* Natural Variation Reveals a Key Role for Rhamnogalacturonan I in Seed Outer Mucilage and Underlying Genes. *Plant Physiol.* **181**, 1498–1518 (2019).
- Amos, R. A. *et al.* Polymerization of the backbone of the pectic polysaccharide rhamnogalacturonan I. *Nature Plants* **8**, 1289–1303 (2022).
- Voiniciuc, C., Günl, M., Schmidt, M. H.-W. & Usadel, B. Highly Branched Xylan Made by IRREGULAR XYLEM14 and MUCILAGE-RELATED21 Links Mucilage to Arabidopsis Seeds. *Plant Physiol.* **169**, 2481–2495 (2015).
- Ralet, M.-C. *et al.* Xylans Provide the Structural Driving Force for Mucilage Adhesion to the Arabidopsis Seed Coat. *Plant Physiol.* **171**, 165–178 (2016).
- Vazquez-Vilar, M. *et al.* A modular toolbox for gRNA-Cas9 genome engineering in plants based on the GoldenBraid standard. *Plant Methods* **12**, 10 (2016).
- Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
- Aliaga-Franco, N. *et al.* Identification of Transgene-Free CRISPR-Edited Plants of Rice, Tomato, and Arabidopsis by Monitoring DsRED Fluorescence in Dry Seeds. *Front. Plant Sci.* **10**, 1150 (2019).
- Sullivan, S. *et al.* CESA5 is required for the synthesis of cellulose with a role in structuring the adherent mucilage of Arabidopsis seeds. *Plant Physiol.* **156**, 1725–1739 (2011).
- Thibault, J. F. Automatisation du dosage des substances pectiques par la méthode au méthahydroxydiphényle. *Lebensm Wiss Technol* **12**, 247–251 (1979).
- Blumenkrantz, N. & Asboe-Hansen, G. New method for quantitative determination of uronic acids. *Anal. Biochem.* **54**, 484–489 (1973).
- Aoi *et al.* Biochemical data documenting variations in mucilage polysaccharide amount and macromolecular properties in a range of glycosyltransferase mutants. *Portail Data INRAE* <https://doi.org/10.57745/DBSZ5X> (2023).
- Saez-Aguayo, S. *et al.* PECTIN METHYLESTERASE INHIBITOR6 promotes Arabidopsis mucilage release by limiting methylesterification of homogalacturonan in seed coat epidermal cells. *Plant Cell* **25**, 308–323 (2013).
- Poulain, D., Botran, L., North, H. M. & Ralet, M.-C. Composition and physicochemical properties of outer mucilage from seeds of Arabidopsis natural accessions. *AoB Plants* **11**, lz031 (2019).
- Cambert, M. *et al.* Datasets of seed mucilage traits for *Arabidopsis thaliana* natural accessions with atypical outer mucilage. *Scientific Data* **8**, 1–9 (2021).
- Griffiths, J. S., Crepeau, M.-J., Ralet, M.-C., Seifert, G. J. & North, H. M. Dissecting Seed Mucilage Adherence Mediated by FEI2 and SOS5. *Front. Plant Sci.* **7**, 1073 (2016).
- Saez-Aguayo, S. *et al.* Transport of UDP-rhamnose by URG2, URG4, and URG6 modulates rhamnogalacturonan-I length. *Plant Physiol.* **185**, 914–933 (2020).

Acknowledgements

We thank Fabien Nogue, Zsolt Kelemen and Anne Frey for technical advice on the CRISPR/Cas9 gene-editing technique used to generate mutants. We also thank Lionel Gissot for providing the pGBD3 TUs containing proRPS5A:hCas9:ter and proCMV:DsRED:terNos used in GoldenBraid cloning. Y.A. was supported by PhD fellowship funding from INRAE (BAP and TRANSFORM divisions) and the Pays de Loire region. The IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007, EUR SPS-GSR).

Author contributions

Y.A., M.-C.R. and H.M.N. conceived and designed the research. Y.A. and A.B. genotyped and produced seeds. Y.A. and L.S. performed biochemical analyses. Y.A., L.S., M.-C.R. and H.M.N. analyzed the data. Y.A., M.-C.R. and H.M.N. wrote the manuscript. All authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02604-2>.

Correspondence and requests for materials should be addressed to M.-C.R. or H.M.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023