

**Supplemental Figure 1. Structural elements and sequence similarities between SEX4 and LSF2.**

**A.** Sequence alignment of *Arabidopsis* LSF2 and SEX4. Secondary structures of SEX4, some of which are also predicted for LSF2, are displayed as ovals ( $\alpha$ -helices) and arrows ( $\beta$ -sheets) using the following colors: 1) tan, SEX4-specific elements, including the CBM; 2) black, LSF2-specific elements; 3) gray, common  $\alpha$ -helix in the cTP Green box, the predicted cTP cleavage site ; 4) pink,  $\alpha$ -helices and  $\beta$ -sheets in the DSP domain common to both proteins; 5) blue,  $\alpha$ -helices in the C-

terminal domain common to both proteins. The sub-domains of the DSP domain are labeled per standard nomenclature (Alonso et al., 2003) and colored in the following manner: green, substrate-recognition domain; black, delta-like; purple, variable loop; beige, D-loop; red, active site; black, AYLM motif; and salmon, R-motif. The Asp and Cys that participate in the nucleophilic attack of the phospho-substrate are boxed in red.

**B.** Percent similarity (in black) and identity (in red) between *Arabidopsis* LSF2 and SEX4 for the full-length proteins (left), for the dual specificity phosphatase domain (DSP), and carboxy-terminal domain (CT), as indicated.

**C.** The C-terminal domain is essential for soluble expression of LSF2. Coomassie-stained SDS-PAGE showing the purification of LSF2 (32 kDa) and LSF2 $\Delta$ CT (28 kDa), which lacks the C-terminal 35 residues. UI, uninduced cells; I, cells induced with IPTG; P, pellet of insoluble protein; S, soluble protein; E, eluted fraction.



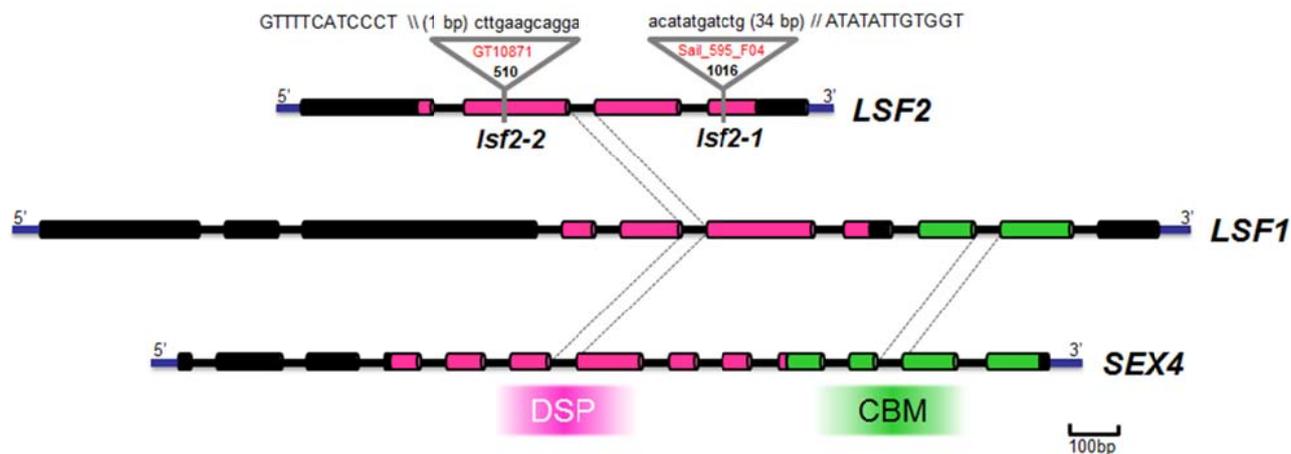
**Supplemental Figure 2. Temporal and spatial expression pattern of *LSF2*.**

**A-G.** GUS reporter gene expression in transgenic *Arabidopsis* plants carrying the  $\beta$ -glucuronidase gene fused downstream of the *LSF2* promoter. (A) Seven-day-old seedlings. After 6 h, staining was strongest in cotyledons, the vasculature, the lower part of the hypocotyl and the root-shoot junction. (B) Seven-day-old etiolated seedlings. Staining was observed only in the vasculature. (C) Floral organs. Strong staining was observed in the sepal vasculature, the stamen and the distal part of the style. (D and E) Roots of light-grown seven-day-old seedlings (as in (A)). Staining was detected in the central cylinder and the root tip and the lateral root primordia. (F) Developing siliques. (G) A decrease in staining was observed in cotyledons of light-grown seedlings after 72 h of dark treatment (samples were stained for 24 h).

**H.** Expression levels of *LSF2* in different organs at different developmental stages. Data were retrieved from the public eFP browser microarray dataset ‘Developmental map’ (<http://www.bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>). Each value is the mean  $\pm$  SD of three replicate samples.

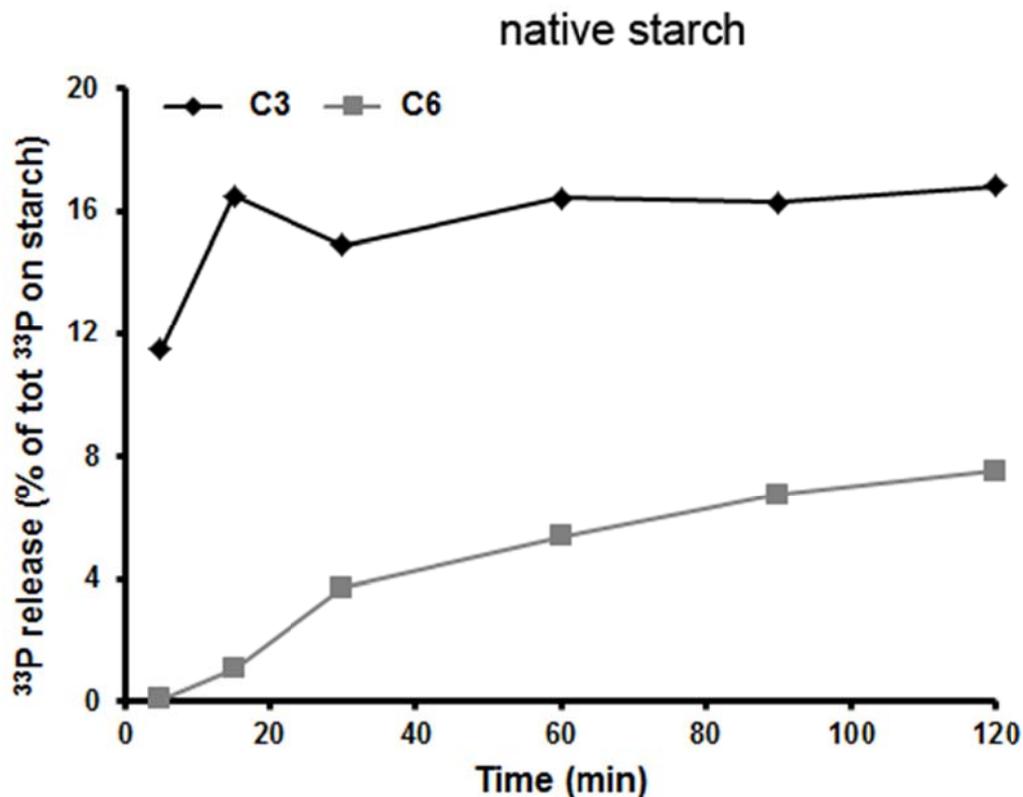
**I.** Comparison of expression profiles of *SEX4* and *LSF2* over a diurnal cycle (12 h dark/12 h light). Expression values were normalized to the median of all eleven time points for each gene. Data used in this analysis are from Smith *et al.* (2004) and were retrieved from the NASC website (<http://arabidopsis.info/>).

**J.** Numbers of peptides identified in different tissue proteomes of *Arabidopsis* (<http://fgcz-atproteome.unizh.ch/>). Overall, the number of identified peptides is representative of protein abundance.



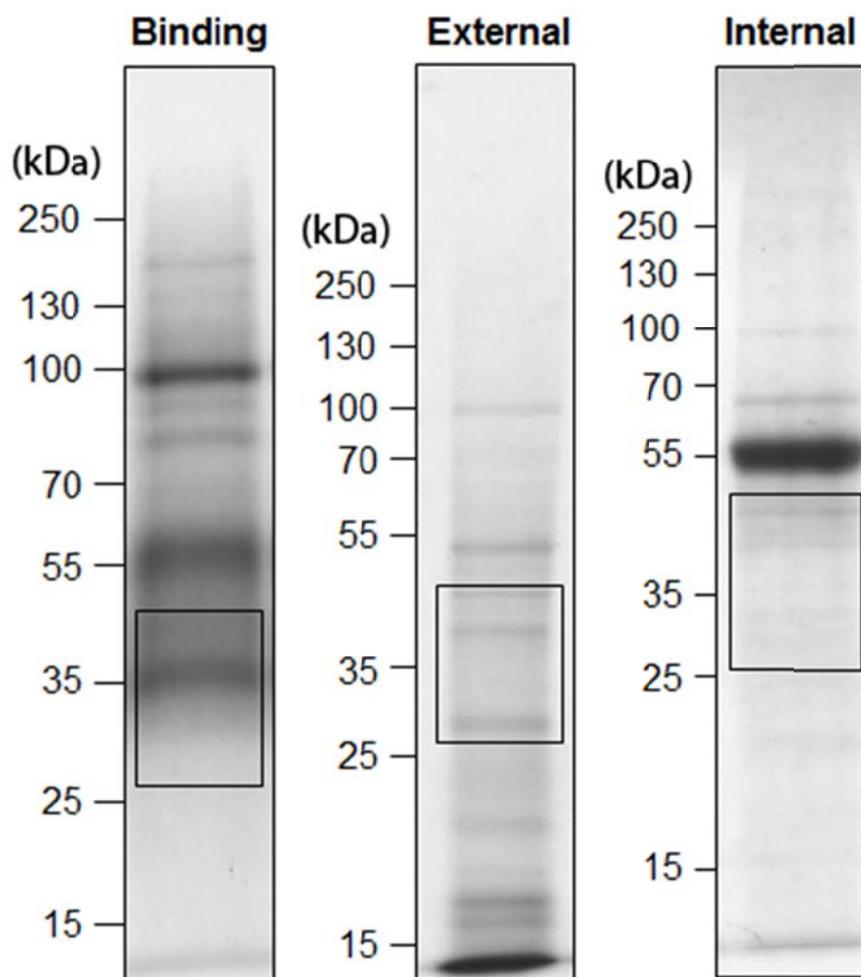
**Supplemental Figure 3. The intron-exon structure of the homologous genes *LSF2*, *LSF1* and *SEX4*.**

Exons (cylinders), introns (black lines, not to scale) and the 5' and 3' untranslated regions (blue lines, not to scale) are shown. Coloured exons encode the DSP domain and the CBM, as indicated. Dashed lines indicate conserved intron positions. The locations of the T-DNA and transposon insertions within *LSF2* are shown (510 and 1016 bp downstream of the ATG start codon for *lsf2-2* and *lsf2-1*, respectively). Line identifiers are given in red. The insertion site sequences are shown. The sequence is given above the insert, with the gene in lower case and the T-DNA or the Ds transposon in uppercase. The length of the intervening sequence not derived from either the T-DNA, Ds transposon or the gene is shown in parenthesis.



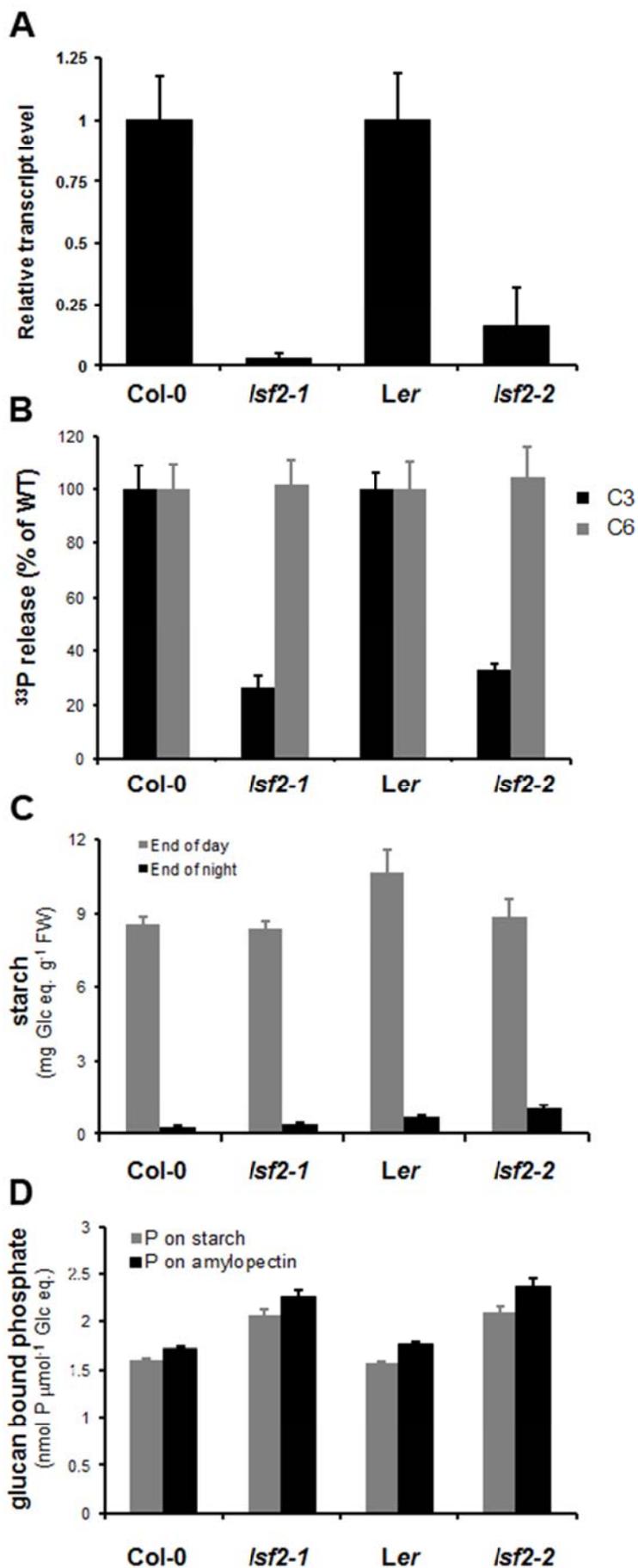
**Supplemental Figure 4. LSF2-mediated hydrolysis of C6- and C3-phosphate esters at native starch granules.**

Purified phosphate-free starch granules from the GWD-deficient *Arabidopsis* mutant *sex1-3* were pre-labeled with <sup>33</sup>P at either the C6- or C3-positions and incubated with 5 μg of LSF2 recombinant protein for 2 h. At intervals during the 2-h time course, the released <sup>33</sup>P was determined. After 15 min, LSF2 dephosphorylated exclusively C3-phospho esters, as expected. However, after 2 h, LSF2 also released small amounts of phosphate from the C6-position.



**Supplemental Figure 5. SDS-PAGE of proteins binding to starch granules.**

*Arabidopsis* proteins were incubated with amylose free potato starch and bound proteins were eluted with SDS (Binding). Proteins binding to isolated *Arabidopsis* starch were extracted with SDS (External), and granules were subsequently boiled to release proteins inside starch (Internal). The boxes indicate the regions of the gels that were subjected to in-gel tryptic digestion and analyzed by LC-MS/MS (see Table 1).



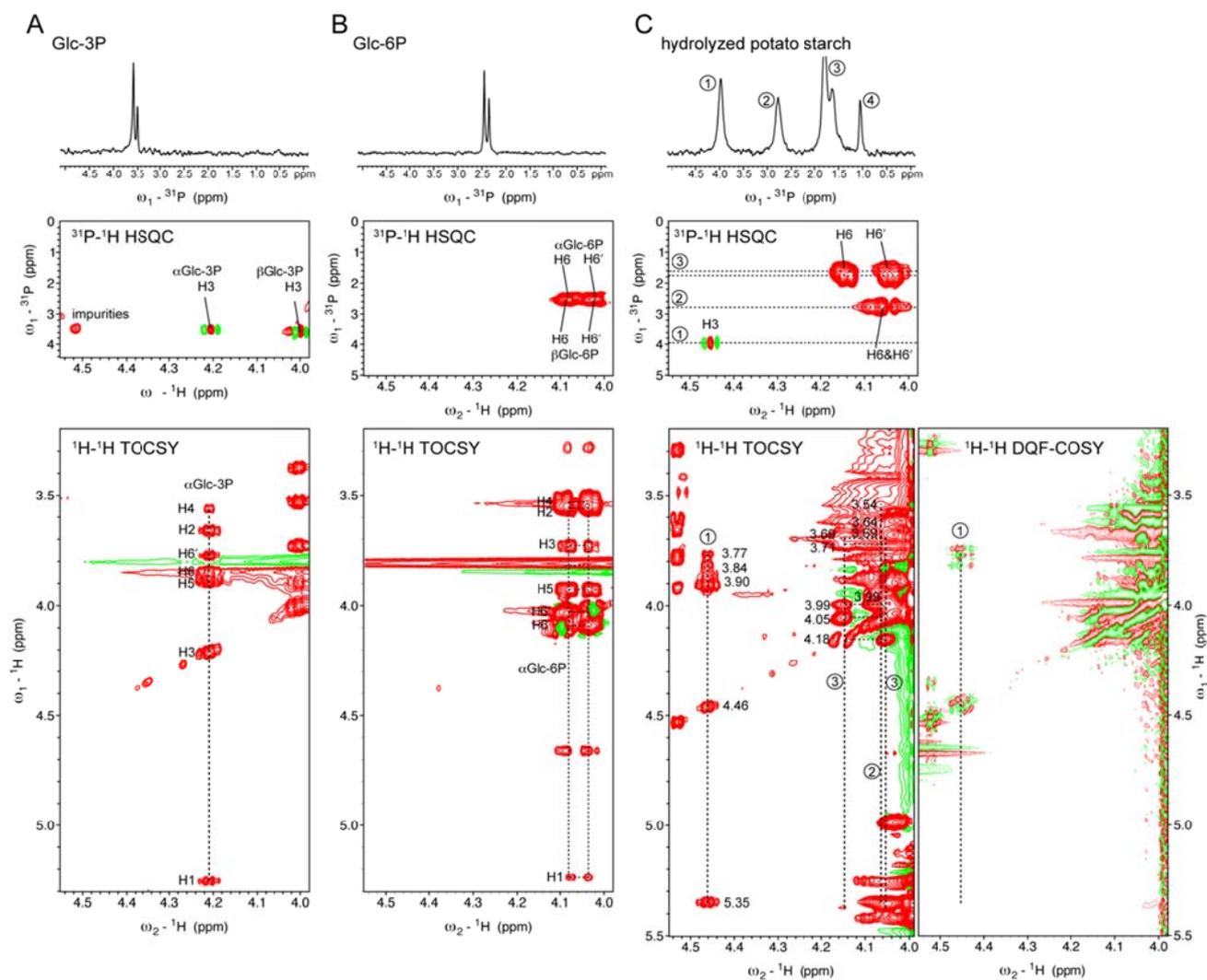
**Supplemental Figure 6. Phenotypic characterization of *lsf2* mutant alleles.**

**A.** Quantitative RT (Reverse Transcriptase)-PCR analysis of *LSF2* expression in leaves of 4-week-old plants. Transcript level for each line was normalized to the expression of the *PP2A* housekeeping gene. Transcript levels in *lsf2* plants are given relative to the respective wild-type plants. Reactions were run in triplicate with three different cDNA preparations and each value is the mean  $\pm$  SE.

**B.** Release of  $^{33}\text{P}$  from isolated starch granules by crude extracts of wild-type and *lsf2* leaves. Purified phosphate-free starch granules from the GWD-deficient *Arabidopsis* *sex1-3* mutant were pre-labeled with  $^{33}\text{P}$  at either the C6- or C3-positions and were then incubated with desalted leaf extracts. Phosphate release over time was linear under these conditions and is expressed relative to the phosphate released by the corresponding wild-type extracts. Each value is the mean  $\pm$  SE of four replicate samples.

**C.** Leaf starch content at the end of the day and the end of the night (as indicated) in the wild types Col-0 and *Ler-0* and in the *lsf2-1* and *lsf2-2* mutants. Each value is the mean  $\pm$  SE of eight replicate samples. FW, fresh weight.

**D.** Starch-bound phosphate content in *lsf2-1* and *lsf2-2* mutant alleles and their respective wild types. The phosphate content of starch purified from leaves of 4-week-old plants harvested at the end of the light period is shown as grey bars. The amylopectin content for the same starch preparations was determined to be  $92.6\% \pm 0.2\%$  for Col-0,  $91.4\% \pm 0.1\%$  for *lsf2-1*,  $88.7\% \pm 0.4\%$  for *Ler*, and  $88.4\% \pm 0.5\%$  for *lsf2-2*. Amylopectin-bound phosphate (black bars) was calculated assuming all the phosphate is bound to amylopectin. Each value is the mean  $\pm$  SE of four replicate samples.



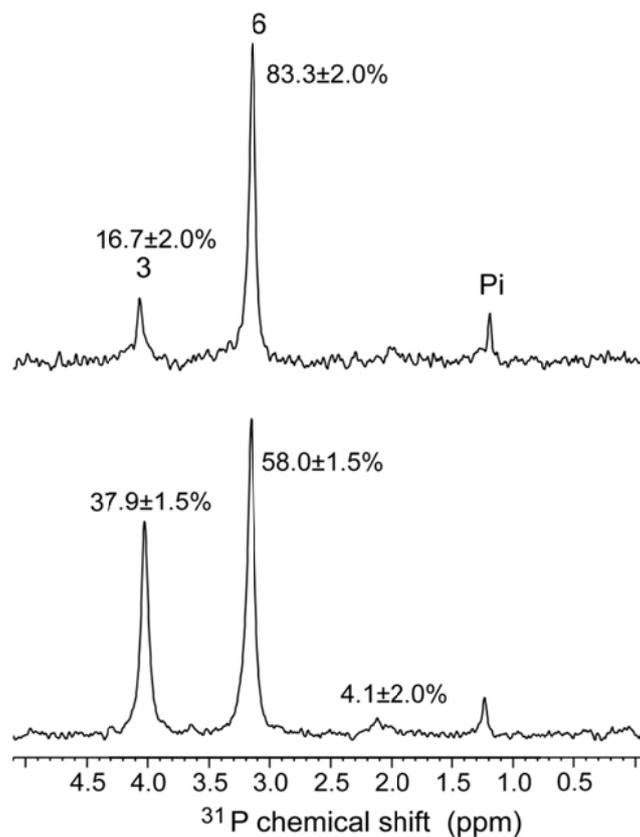
### Supplemental Figure 7. Assignment of $^{31}\text{P}$ signals using 2D NMR spectroscopy.

2D  $^{31}\text{P}$ - $^1\text{H}$  HSQC (Heteronuclear Single-Quantum Correlation) and 2D TOCSY (Total Correlation Spectroscopy) spectra of glucose-3-phosphate (Glc-3P, **A**), glucose-6-phosphate (Glc-6-P, **B**) and hydrolyzed potato starch (**C**).

The reference substances Glc-3P and Glc-6P at a concentration of 3.4 mM in 10 mM citrate/NaOH pH 6.0 (re-adjusted after dissolving) were lyophilized and dissolved in the same volume of  $\text{D}_2\text{O}$ . A solution of hydrolyzed potato starch (initially 25 mg) was adjusted to pH 6, lyophilized and dissolved in 500  $\mu\text{l}$   $\text{D}_2\text{O}$ . All 2D spectra were acquired at 303 K on a 600 MHz instrument (Avance III) equipped with a  $^1\text{H}/^{13}\text{C}/^{15}\text{N}/^{31}\text{P}$  quadruple-resonance CryoProbe (Bruker AG, Switzerland). Standard 2D  $^1\text{H}$ - $^1\text{H}$  TOCSY (Bax and Davis, 1985) and 2D DQF-COSY (Double Quantum Filtered - Correlation Spectroscopy) (Rance et al., 1983) spectra were recorded. A 2D  $^{31}\text{P}$ - $^1\text{H}$  HSQC (Bodenhausen et al., 1980) was applied with a transfer delay for evolution of the scalar coupling  $^3J_{\text{HP}}$  of 40 ms. During acquisition,  $^{31}\text{P}$  decoupling was applied using GARP (Globally optimized Alternating phase Rectangular Pulses; Shaka et al., 1985) at a field strength of 1.5 kHz. More

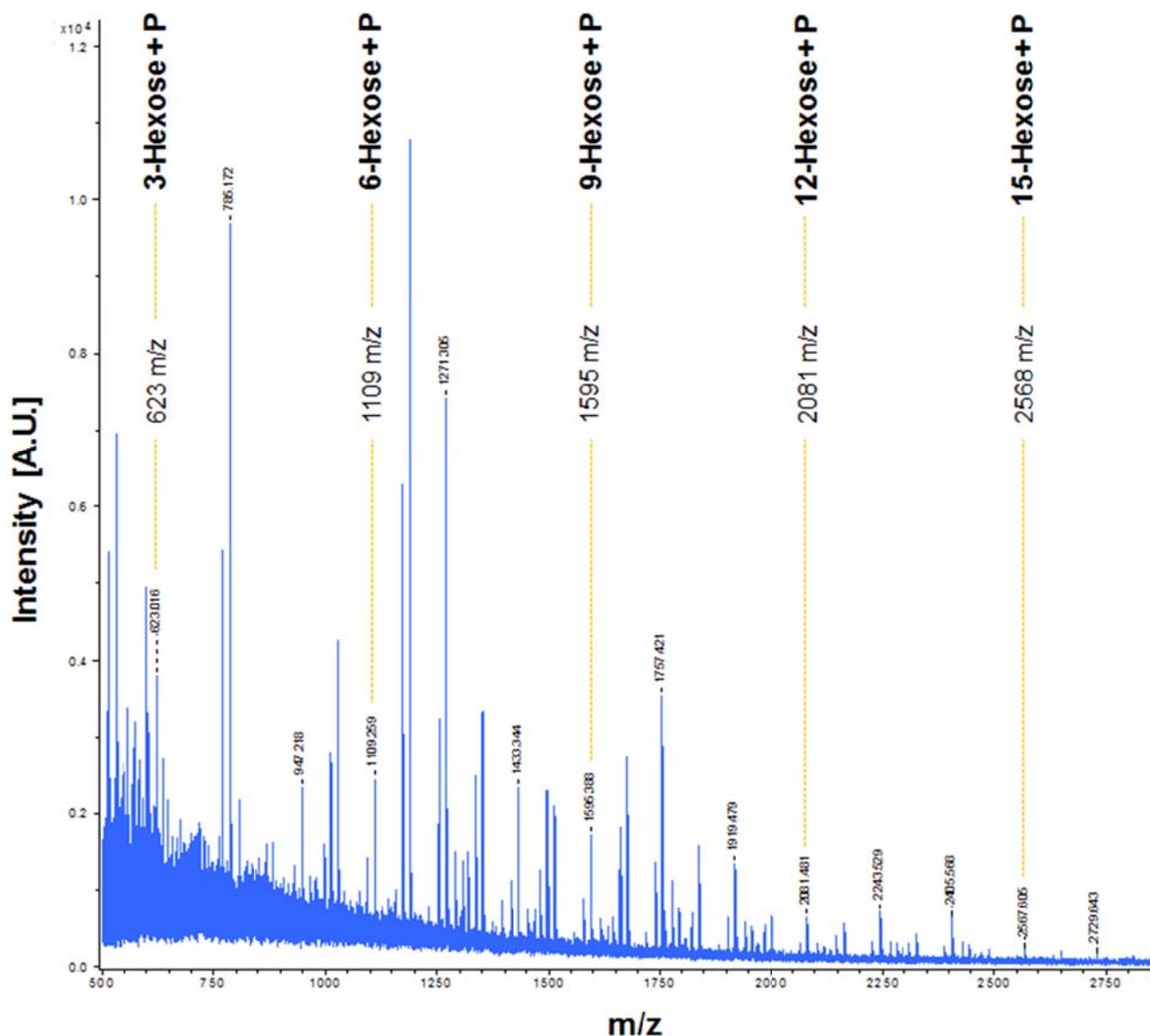
detailed acquisition parameters are given in Supplemental Table 1. Spectra were processed with Topspin 2.1 (Bruker AG) and analyzed with Sparky (Goddard and Kneller). The  $^1\text{H}$  axis was calibrated to DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) according to Markley et al. (1998) and the  $^{31}\text{P}$  axis was indirectly referenced to  $\text{H}_3\text{PO}_4$  (85% wt solution in  $\text{H}_2\text{O}$ ) using a  $\Xi$  value of 0.404807356 (Maurer and Kalbitzer, 1996).

The  $^{31}\text{P}$  chemical shifts of the hydrolyzed starch are numbered 1 to 4 and the chemical shifts of correlated protons within the phosphorylated glucose are indicated. Signal 1 is assigned to C3-bound phosphate because the  $^{31}\text{P}$ - $^1\text{H}$  HSQC shows a single correlation with a similar coupling pattern as in the reference substance Glc-3P and all the  $^1\text{H}$  chemical shifts of 2D TOCSY correlations within the same sugar ring correspond to calculated chemical shifts of a Glc-3P flanked by  $\alpha$ 1,4-linked Glc units on either side using the software CASPER (Lundborg and Widmalm, 2011). The lack of a correlation to H1 (5.35 ppm) in the 2D DQF-COSY excludes a C2-bound phosphate. Signals 2 and 3 show correlations to two protons, indicative that the phosphate is bound to a  $\text{CH}_2$  group and thus to C6. Signal 4 did not show any  $^{31}\text{P}$ - $^1\text{H}$  correlation and was assigned to inorganic phosphate. A TOCSY mixing time of 120 ms was used.



**Supplemental Figure 8.  $^{31}\text{P}$  NMR analysis of *Arabidopsis* starch-bound phosphate.**

1D  $^{31}\text{P}$  NMR spectra of *lsf2-2* (bottom) in comparison with wild type *Ler* (top). Spectra were recorded at 303 K, pH 6.0, with between 9216 and 16384 transients. Peak areas were used to calculate the relative amounts of phosphate at each position, which are given as a percentage on top of each peak.



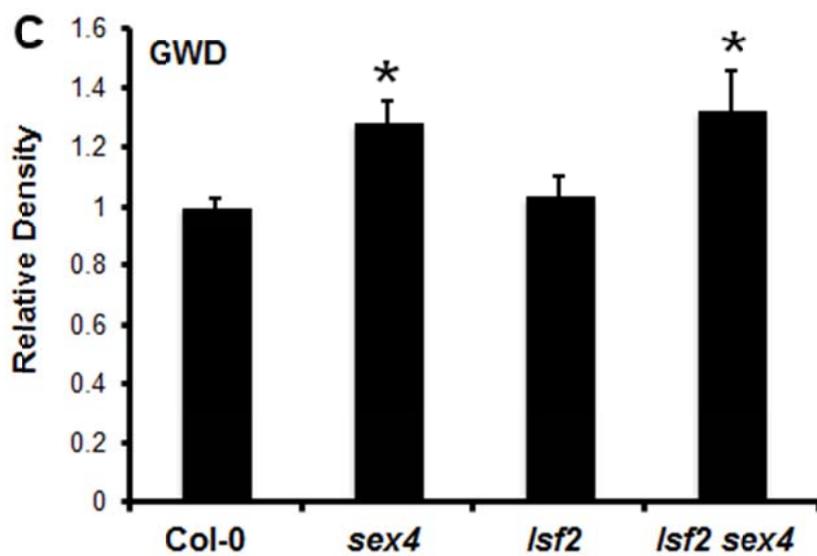
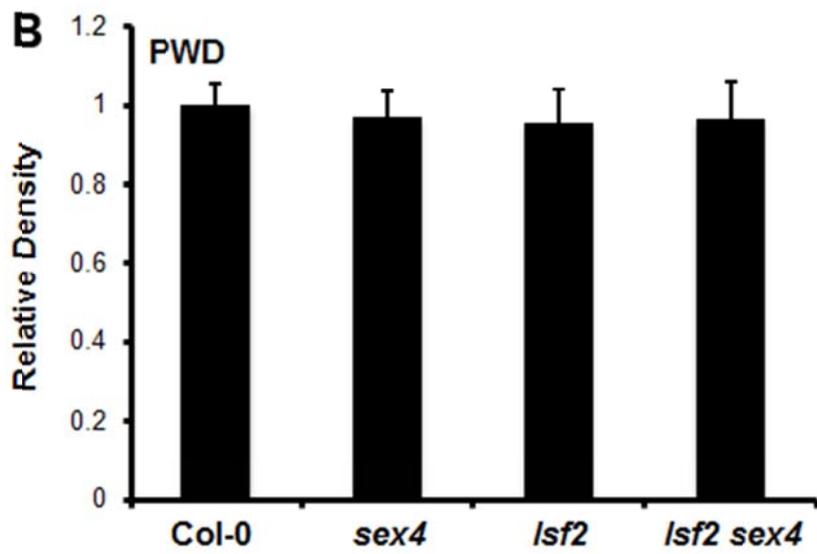
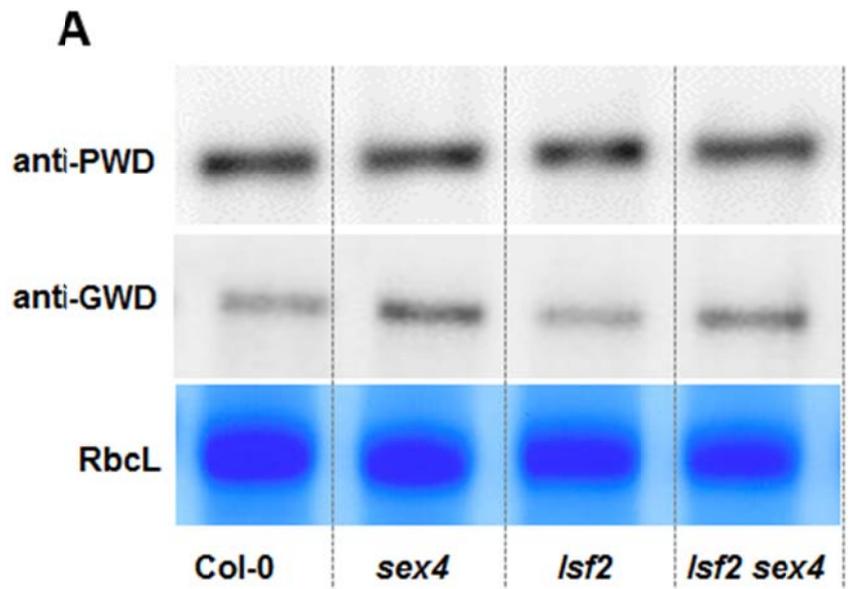
**Supplemental Figure 9. MALDI/MS/MS analysis of phosphorylated oligosaccharides.**

MALDI-TOF spectrum of phosphorylated oligosaccharides derived from digested Col-0 wild-type starch (see Methods). The spectrum displays a series of species differing by 162.1 m/z, indicative of a mixture of phosphorylated oligosaccharides ranging from 3 hexoses + 1 phosphate to 16 hexoses + 1 phosphate. Doubly phosphorylated glucans were also detected, but the signal intensities were very low. Mass calculations are as follows:

$$\text{Mass (m/z)} = 180 + 39 (\text{K}) + \text{N} * 162 + 80$$

N, polymerization degree - 1

\*, multiplied by



**Supplemental Figure 10. GWD and PWD protein levels in leaves of wild type Col-0, *lsf2*, *sex4*, and *lsf2 sex4* plants.**

Total protein was extracted from 4-week-old plants harvested at the end of the light period and equal amounts of protein were separated by SDS-PAGE. GWD and PWD were then detected by immunoblotting using an antibody against potato GWD and *Arabidopsis* PWD, respectively. The Rubisco large subunit (RbcL), which appeared as a dominant band when the extracts were visualized using Coomassie-stained SDS-PAGE, was used as internal loading control. Densitometry analysis of three replicate blots was used to quantify the band intensities. Values are expressed relative to the mean band intensity of the wild type. Both *sex4* and *lsf2 sex4* had small, but significant increases in GWD, whereas no differences were detected for PWD. Each value is the mean  $\pm$  SE of three biological samples. Each sample was blotted in three technical replicates. Asterisks indicate significant differences ( $p$  value  $< 0.05$ ).

**Supplemental Table 1.** Acquisition parameters for 2D NMR spectroscopy.

experiment	sample	sw	swl	ns	ni	at	mix	np	d1
1D <sup>31</sup> P	Glc-3P and Glc-6P	50.7		128		1328		32768	3
1D <sup>31</sup> P	hydrolyzed potato starch	50.7		14336		1328		32768	3.5
2D <sup>31</sup> P- <sup>1</sup> H HSQC	Glc-3P and Glc-6P	13.9	8.0	2	256	123	40	2048	1.5
2D <sup>31</sup> P- <sup>1</sup> H HSQC	hydrolyzed potato starch	13.9	20.0	128	128	123	40	2048	1.5
2D TOCSY	Glc-3P and Glc-6P	10.0	10.0	32	512	170	120	2048	1.0
2D TOCSY	hydrolyzed potato starch	10.0	10.0	32	512	170	120	2048	1.3
2D DQF-COSY	potato starch hydrolysate	10.0	10.0	56	670	340		4096	1.0

Abbreviations used:

sw: spectral width in the directly detected dimension in ppm

swl: spectral width in the indirectly detected dimension in ppm

ns: number of scans

ni: number of increments

at: acquisition time in ms

mix: mixing time for TOCSY or length of INEPT block in ms

np: number of points in the directly detected dimension

d1: interscan delay in s

**Supplemental Table 2.** Sequence of primers used in this work.

<b>Primer</b>	<b>Sequence</b>
<b>Primers used for genotyping homozygous mutant plants</b>	
<u>T-DNA left border primers</u>	
LBb1 Sail	GCCTTTTCAGAAATGGATAAAATAGCCTTGCTTCC
LBb1 Salk	GCGTGGACCGCTTGCTGCAACT
DS3-2 (for GT10871)	CCGGTATATCCCGTTTTTCG
<b><u>lsf2-1 (Sail 595 F04)</u></b>	
SAIL_595 F04 LP	ATATTGCGGTGCAACTTTACG
SAIL_595 F04 RP	CTGAGCATTTATCAGTTGGGG
<b><u>lsf2-2 (GT10871)</u></b>	
At3g10940_fw1	TGTGATTGGAAGCAAGAGCT
At3g10940_re1	CCGAACACGTTCTTGAATCAAC
<b><u>sex4-3 (Salk 102567)</u></b>	
Salk_102567 LP	AAGCTGATGCGTAATGAATCG
Salk_102567 RP	GAAATCCCCAACATCCTCAC
<b>Primer used for qRT-PCR</b>	
<u>PP2A houskeeping gene (At1g13320)</u>	
PP2A_F01	CTCTTACCTGCGGTAATAACTG
PP2A_R01	CATGGCCGTATCATGTTCTC
<u>LSF2 gene (At3g10940)</u>	
At3g10940_F01	GACATGAATCTTAACACGGCT
At3g10940_R01	ATCATATGTTGCACCACGG
<b>Primers used for recombinant cloning</b>	
<u>Subcellular localization (LSF2-GFP)</u>	
LSF2pGFP2 fw (KpnI site added)	<u>GGGGTACCATGAGTGTGATTGGAAGC</u>
LSF2pGFP2 rev (KpnI site added)	<u>GGGGTACCGGTTCCACGGAGGGCC</u>
<u>Construction of LSF2<sub>prom</sub>:GUS fusion gene</u>	
LSF2 <sub>prom</sub> fw	GATTGCATTATTGATTTGTTGCTCTTGTAG
LSF2 <sub>prom</sub> rev	CGTTCTCTATCTCTCGTTCTTCACCTG
<u>Cloning of recombinant proteins</u>	
<b>LSF2 wt</b>	
LSF2 full length cDNA fw	ATGAGTGTGATTGGAAGCAAGAGC
LSF2 full length cDNA rev	TCAGGTTCCACGGAGGGC
<b>Δ65-LSF2</b>	
Δ65-LSF2_fw	TTTCATATGAACAAAATGGAGGATTACAATACAGC
Δ65-LSF2_rev	AAACTCGAGTCATCAGGTTCCACGGAGGGCC
<b>Δ78-LSF2</b>	
Δ78-LSF2_fw	TTTCATATGATGAGAAGCCCTTATGAATATCATCATG
Δ78-LSF2_rev	AAACTCGAGTCATCAGGTTCCACGGAGGGCC
<b>LSF2ΔCT</b>	
LSF2-CT fw	GAATGATCCCTGAAAAGAGCCCTTTG
LSF2-CT rev	CAAAGGGCTCTTTTCAGGGATCATTC
<b>LSF2 C/S</b>	
LSF2 C193S fw	GGTAAAGGAAGAGTCTATGTGCATTCTTCAGCCGGATTGG
LSF2 C193S rev	CCAATCCGGCTGAAGAATGCACATAGACTCTTCCTTACC

## References

- Alonso, A., Rojas, A., Godzik, A., Mustelin, T.** (2003). The dual-specific protein tyrosine phosphatase family. *Top. Curr. Gen.* **5**, 333-358
- Bax, A., and Davis, D.G.** (1985) Mlev-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355-360.
- Bodenhausen, G., and Ruben, D.J.** (1980) Natural abundance N-15 Nmr by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185-189.
- Goddard, T.D., and Kneller, D.G.** SPARKY 3. University of California, San Francisco.
- Lundborg, M., and Widmalm, G.** (2011), Structural analysis of glycans by NMR chemical shift prediction. *Anal. Chem.* **83**, 1514-1517.
- Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., Wüthrich, K.** (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy. *J Biomol NMR* **12**, 1-23.
- Maurer, T., and Kalbitzer, H.R.** (1996). Indirect referencing of P-31 and F-19 NMR spectra. *J. Magn. Reson. B* **113**, 177-178.
- Rance, M., Sorensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., Wüthrich, K.** (1983). Improved spectral resolution in cosy <sup>1</sup>H NMR spectra of proteins via double quantum filtering. *Biochem. Bioph. Res. Co.* **117**, 479-485.
- Shaka, A.J., Barker, P.B., Freeman, R.** (1985). Computer-optimized decoupling scheme for wideband applications and low-level operation. *J. Magn. Reson.* **64**, 547-552.