

**Supplemental figure 1: Workflow for protein identification in the** *Lotus* **seed globulin fraction.** The globulin fraction was solubilized in either 2 or 7 M urea, in the next step samples were split and *N*-glycan released using PNGase A, in half of the samples, subsequently half of these samples were desalted. The eight different samples for LC-MS are marked 1-8 which corresponds to the protein identification list given in Supplemental table 2. For a further description of the procedure, see materials and methods.

PNG A; PNGase A, DS; desalting.

## Supplemental figure 2A



Supplemental figure 2: Informative MS spectra of glycosylated and deglycosylated LCP2 and predicted lectin. A and B) The three ions at m/z 1646.8, 2304.1, and 2755.2 in the deglycosylated samples corresponding to peptides with an *N*-glycosylation motif with deaminated Asn (peptide sequences verified by MS/MS), whereas for the three glycosylated samples only noise is found in that m/z range. LCP2 was in-gel digested with Asp-N and the predicted lectin was *in-gel* digested with trypsin.





## **Supplemental figure 3**

**Supplemental figure 3: Structure and relative abundance of** *N***-glycans released from** *Lotus* **seed globulins using PNGase F.** The relative abundance (in molar %) of the 11 different *N*-glycans identified are indicated for each structure. CFG nomenclature has been used: Blue square; *N*-acetylglucosamine, red triangle; fucose, white star; xylose, green circle; mannose, yellow circle; galactose, blue circle; glucose.



Supplemental figure 4: MS spectra of HILIC enriched trypsin or Asp-N derived glycopeptides from LCP2 and the predicted lectin. A, B, and C) *m/z* values underlined corresponding to glycopeptide signals. The *N*-glycan structure corresponding to each signal is shown and the peak intensity is used for relative *N*-glycan occupancy for each glycosylation site shown in figure 5. Blue square; *N*-acetylglucosamine, red triangle; fucose, white star; xylose, green circle; mannose.

## Supplemental figure 4B

tryptic digestion of lectin







Supplemental figure 5: Fragmentation of the lectin *N*-glycopeptides carrying Man<sub>3</sub>Xyl<sub>1</sub>GlcNAc<sub>2</sub>. Fragment spectra of m/z 3326.9 and 1866.1 (precursors), which correspond to the predicted lectin *N*-glycopeptides with Man<sub>3</sub>Xyl<sub>1</sub>GlcNAc<sub>2</sub> attached to the NFT motif (insert). Peptide b- and y-ions together with glycopeptide fragments are indicated. CFG nomenclature: Blue square; *N*-acetylglucosamine, blue square with black line; 0,2X-ring cleavage of the innermost *N*-acetylglucosamine, green circle; mannose, star; xylose.

Intensity



## Supplemental table1

Accession	<i>N</i> -glycan motifs	Transmembrane region	Signal peptide	Localized
chr5.CM0200.930.r2.m (LCP2)	1	no	yes	secreted
chr5.LjT43D06.200.r2.d (predicted lectin)	2	no	yes	secreted
chr5.CM0200.910.r2.m (LCP1)	0	no	yes	secreted
chr5.LjT15N12.20.r2.m (LLP3)	0	no	yes	secreted
chr5.LjT15N12.10.r2.m (LLP2)	0	no	yes	secreted
chr1.CM0295.280.r2.d	0	no	yes	secreted
chr1.CM0295.230.r2.m	0	no	yes	secreted
LjSGA_063286.1 (glucosidase)	1	no	no	other
chr1.CM0544.570.r2.d	1	no	yes	secreted
chr1.CM0295.240.r2.d (LLP5)	0	no	yes	secreted
chr3.CM0241.200.r2.d	9	no	no	other
chr5.LjT43D06.110.r2.d (psedu lectin)	3	no	no	other

**Supplemental table 1: Identified proteins from the globulin fraction.** In total, 12 proteins were identified from eight different globulin fraction preparations (see Supplemental table 2 for detailed information of the protein identifications from the eight samples). *N*-glycan motifs are Asn-X-Ser/Thr where  $X \neq Pro$ . For transmembrane region prediction, the freely available TMHMM software (<u>http://www.cbs.dtu.dk/services/TMHMM/</u>) was used. For signal peptide prediction and localization of the protein, the freely available TargetP 1.1 software (<u>http://www.cbs.dtu.dk/services/TargetP/</u>) was used where the plant option was selected. The *Lotus* accessions correspond to the 2.5 version of the genome downloaded from http://www.kazusa.or.jp/lotus/. The accession numbers in bold correspond to the suggested proteins carrying complex *N*-glycans. LLP; *Lotus* legume storage protein, LCP; *Lotus* convicilin storage protein.